In Silico Screening-Level Prioritization of 8468 Chemicals Produced in OECD Countries to Identify Potential Planetary Boundary Threats

Efstathios Reppas-Chrysovitsinos · Anna Sobek · Matthew MacLeod

Received: 23 October 2017 / Accepted: 11 December 2017 / Published online: 28 December 2017
© The Author(s) 2017. This article is an open access publication

Abstract
Legislation such as the Stockholm Convention and REACH aim to identify and regulate the production and use of chemicals that qualify as persistent organic pollutants (POPs) and very persistent and very bioaccumulative (vPvB) chemicals, respectively. Recently, a series of studies on planetary boundary threats proposed seven chemical hazard profiles that are distinct from the POP and vPvB profiles. We previously defined two exposure-based hazard profiles; airborne persistent contaminants (APCs) and waterborne persistent contaminants (WPCs) that correspond to two profiles of chemicals that are planetary boundary threats. Here, we extend our method to screen a database of chemicals consisting of 8648 substances produced within the OECD countries. We propose a new scoring scheme to disentangle the POP, vPvB, APC and WPC profiles by focusing on the spatial range of exposure potential, discuss the relationship between high exposure hazard and elemental composition of chemicals, and identify chemicals with high exposure hazard potential.

Keywords QSAR · Hazard screening · Planetary boundary · POPs · Persistence · Bioaccumulation · Long-range transport

The global production of synthetic chemicals is increasing exponentially by total volume (CEFIC 2014; ACC 2014) and in structural diversity (Binetti et al. 2008). The presence of synthetic chemicals in the environment is responsible for a plethora of effects that manifest themselves at scales from the molecular level up to the scale of earth system processes (Rockström et al. 2009; Steffen et al. 2015; Persson et al. 2013; Bernhardt et al. 2017; Diamond et al. 2015; MacLeod et al. 2014). To address the conceptual and data-related challenges associated with early identification of chemicals with unacceptable impacts, a number of databases (Boethling et al. 2004; Apodaca 2011), in silico physicochemical property estimation tools (Schüürmann et al. 2007; Sushko et al. 2011; Cao et al. 2008; Bennett et al. 2009) and models of chemical fate and transport (Arnot 2009) have been developed. Furthermore, a series of regulatory efforts that aim to identify chemicals of concern by screening against exposure and hazard criteria, and by defining profiles of contaminants of concern (European Parliament the Council of the European Union 2006; UNEP 2001) have been developed and deployed.

There is a growing number of computational studies in the scientific literature that aim to identify and prioritize chemicals that may have unacceptable environmental profiles (Walker and Carlsen 2002; Matthies et al. 2016; Strempel et al. 2012; Brown and Wania 2008; Muir and Howard 2006; Rorije et al. 2011; Scheringer et al. 2012). Among these studies, Screen-POP (Arnot et al. 2012; Breivik et al. 2012) focused on screening chemicals reportedly produced in the OECD countries to identify chemicals similar to persistent organic pollutants (POPs)—a chemical profile regulated under the Stockholm Convention—and persistent, bioaccumulative and toxic (PBT) substances, which are regulated under the European chemicals regulation, REACH. To this end, a database of 12,619 organic chemicals reportedly produced in the OECD countries was assembled from five national and international production volume lists, and a number of potential POP and PBT chemicals were identified (Arnot et al. 2012; Breivik et al. 2012). The Screen-POP project applied a “holistic” approach to quantitatively assess the exposure potential of the chemicals in the database,
taking into account their physicochemical properties, their overall environmental partitioning and behavior, and their emissions. Moreover, uncertainties associated with emission and property estimates were also considered. A few chemicals identified in Screen-POP that were previously unknown as contaminants in the environment were later detected in the atmosphere and sediments near Stockholm, Sweden (McLachlan et al. 2014).

Recently, MacLeod et al. (2014) defined a set of chemical profiles for planetary boundary threats. The potential for chemicals to be planetary boundary threats is not considered in current regulatory regimes for PBT chemicals at local and regional scales, or for POPs at the global scale (Persson et al. 2013), and should therefore be considered independently in screening and prioritization exercises. We subsequently proposed a screening-level prioritization method to identify chemicals that fit two novel exposure-based profiles we derived from the planetary boundary threat framework (Reppas-Chrysovitsinos et al. 2017). These novel profiles assign high priority to persistent and mobile chemicals, which are labelled as airborne persistent contaminants (APCs) and/or waterborne persistent contaminants (WPCs). We applied an exposure-based screening method using our two proposed novel profiles, in parallel to profiles for POPs and very persistent and very bioaccumulative chemicals (vPvB), and evaluated the method using a set of 464 chemicals identified in an Arctic Monitoring and Assessment Programme (AMAP) report (AMAP 2017) as emerging Arctic contaminants of concern.

In this study, we apply the same methodology to 8648 chemicals taken from the Screen-POP database of 12,619 substances produced within OECD countries. Our goal is to apply our hazard scoring method to identify potential POP, vPvB, APC and WPC candidates among these chemicals. The Screen-POP database is larger than the AMAP database, which is composed only of known environmental contaminants, and it is also more structurally diverse. Therefore, the application of our method to the Screen-POP set of chemicals provides an opportunity to further investigate its prospective assessment capabilities, and to identify and prioritize chemicals that are POPs, vPvB, and potential planetary boundary threats from among those in the database.

**Materials and Methods**

Our screening method is described in detail in our previous publication in which we screened known Arctic contaminants (Reppas-Chrysovitsinos et al. 2017). In brief, our method is based on deriving an exposure-based hazard scoring scheme for each of the POP, vPvB, APC and WPC profiles and using these scores to benchmark the case study chemicals as a percentile rank against a reference set of 148 chemicals with well-characterized environmental fate profiles. The four hazard scores are the product of selected combinations of hazard metrics for persistence (P), bioaccumulation (B) and long-range transport (LRT). The P and LRT metrics are calculated using the OECD Tool (Wegmann et al. 2009), a multimedia model that calculates overall persistence ($P_{op}$), and the three LRT metrics; transfer efficiency ($TE$), characteristic travel distance in air for emissions to air ($CTD_{air}$) and characteristic travel distance in water for emissions to water ($CTD_{water}$). Our B metric is calculated using EPI Suite’s (United States Environmental Protection Agency 2012) BCFAF QSAR to obtain $BAF$, and thus considers both the hydrophobicity of the chemicals and their estimated potential for biotransformation.

Our exposure-based hazard profiles are calculated as the product of intensive or quasi-intensive chemical properties (P, B and LRT) that allow a meaningful comparison of chemicals (Mackay et al. 2001) for their potential to cause environmental or ecosystem exposure (Scheringer 2002). The profiles translate legislative language that uses the word “and” (e.g. “persistent and bioaccumulative”) as a logical AND operator and therefore we ascribed this as a multiplicative combination of the metrics.

The vPvB profile assigns high priority to very persistent and very bioaccumulative compounds without considering potential for long-range transport and is therefore a hazard profile that does not prioritize exposure due to long-range transport to regions remote from sources. The POP profile also includes multiplicative terms for persistence and potential for bioaccumulation, but also a term for long-range transport potential. Our LRT metric of choice for the POP profile is $TE$ (%), which accounts for atmospheric transport and deposition in remote areas. Including deposition in the metric for LRT for the POP profile is appropriate since deposition is a prerequisite for exposure to aquatic biota, which is further enhanced by high potential for bioaccumulation. The choice of $TE$ as LRT metric renders the POP profile of regional-to-hemispheric coverage. The APC and WPC profiles prioritize chemicals with hemispheric-to-global scale spatial range, as they include multiplicative terms for persistence and for mobility in air ($CTD_{air}$ km) or water ($CTD_{water}$ km).

We benchmarked the Screen-POP chemicals against a reference set of 148 contaminants for each of the four hazard profiles and assigned them scores ($S_{PROFILE}$) as percentile ranks compared to the reference set. Most Screen-POP chemicals are thus assigned to a bin bounded by two reference set compounds that serve as the upper and lower limit of the bin. However, there are 149 bins per profile in total as the upper and lower bins are each bounded by one reference set chemical which is, respectively, the highest and the lowest scoring contaminant in this profile. Therefore, possible scores for the Screen-POP chemicals for each of the profiles
range from 0 (hazard score lower than all chemicals in the reference set) to 100 (hazard score greater than all chemicals in the reference set).

Details on the reference set and the benchmarking method can be found in our previous publication (Reppas-Chrysosvitosinos et al. 2017). In brief, the reference set consists of 148 well-characterized chemicals that are used to rank and contextualize the hazard profiles of chemicals with which we have less experience and information on a comparative scale.

The Screen-POP database was obtained by request from the authors of the original Screen-POP study (Arnot et al. 2012). Details on the initial compilation of each of the datasets can be found in these studies. The partition ratios and degradation half-lives for both the Screen-POP and reference databases were derived from EPI Suite, but in slightly different ways. The log $K_{AW}$ and log $K_{OW}$ values of the reference set were estimated by the HENRYWIN and KOWWIN modules of EPI Suite 4.11, respectively, and no further modifications were applied. EPI Suite 4.11 was also used to estimate degradation half-lives in air ($t_{1/2air}$), water ($t_{1/2water}$) and soil ($t_{1/2soil}$) that are required as input for the OECD Tool for both the Screen-POP chemicals and the reference set chemicals. The log $K_{AW}$ and log $K_{OW}$ values in the Screen-POP database were selected by the Screen-POP authors, and are a combination of experimental values (when available in the EPI Suite database), log $K_{AW}$ values calculated from estimated water solubility and vapor pressure (when HENRYWIN could not provide estimates), and HENRYWIN and KOWWIN estimates for any other case of log $K_{AW}$ and, respectively, log $K_{OW}$. Partition ratios with predicted values exceeding the extrema of the respective measured values that were used as EPI Suite’s training sets were considered to fall outside the model domain and were converted to the corresponding experimental extrema. Therefore, the log $K_{OW}$ range is (−4, 10) and the log $K_{AW}$ range is (−12, 3). The Tool outputs are known to become insensitive to changes of partition ratios outside the aforementioned range (Fenner et al. 2005) and therefore we incorporated this adjustment which also simplifies the comparison of the chemical spaces of the Screen-POP database and the reference set (−10 < log $K_{AW}$ < 2 and −2 < log $K_{OW}$ < 10).

Previous work with the Screen-POP database used an alternative method (Arnot et al. 2012) to obtain a more diverse set of degradation half-lives from the initial EPI Suite estimates. In this work, we used the default EPI Suite estimates for degradation half-lives of chemicals in both the reference and Screen-POP datasets. Although the half-life estimation method used in previous Screen-POP work has advantages, our goal in this study is to focus on the application of our method as presented earlier (Reppas-Chrysosvitosinos et al. 2017) and further explore its efficacy to screen for planetary boundary threats without introducing new corrections and assumptions- which is also in line with the screening-level character of our method.

We removed from the Screen-POP database any chemical compound that is likely to be ionized within an environmentally relevant pH range (i.e. $pK_a < 5$ or $pK_b > 8$) because of uncertainties associated with their partitioning and their estimated BAF. For the estimation of $pK_a$ and $pK_b$, we employed the ACD Labs Percepta software (Advanced Chemistry Development Inc 2015). After this data curation step, a total of 8648 compounds remained from the original set of 12,619 in the Screen-POP database. Therefore, in the following lines the term “Screen-POP chemicals” is used to refer to these 8648 chemicals and not to the original set.

**Results and Discussion**

A comparison of the structures contained in the Screen-POP and reference datasets is presented in Table 1. The molecular weight range of the Screen-POP chemicals is broader and more diverse than the reference set. The reference set has a higher percentage of halogenated compounds- which reflects its focus on known and well-characterized environmental contaminants- compared to the Screen-POP dataset which is representative of chemicals in commerce in OECD countries. Interestingly, differences in the number of halogen atoms between the two databases are not observed in a uniform manner; the occurrence of bromine and iodine atoms is similar in both sets while the percentage of chlorine is smaller and the percentage of compounds containing fluorine is larger in the Screen-POP database. It is also noteworthy that the reference set contains no organosilicons and organoborons since these compound groups were not represented among well-characterized environmental pollutants at the time the data for the reference set was collected.

A more detailed analysis of the structural composition of the Screen-POP chemicals is presented in Fig. 1, which illustrates the frequency distribution of the number of atoms included in the structures of the Screen-POP chemicals for each of the ten elements found in Table 1. Most frequently the fluorinated compounds in the Screen-POP database have three atoms of fluorine in their structure, which is presumably a terminal –CF$_3$ group. The second most common fluorination level is one fluorine atom and the third is two fluorine atoms. More than half of the 1132 chlorinated compounds contain two chlorine atoms with the second most frequent chlorination level being four chlorine atoms. The majority of the 253 brominated compounds, as well as the majority of the 27 iodinated chemicals include only one halogen atom in their structures. For fluorinated and iodinated chemicals an odd number of fluorine or iodine atoms is more frequent than an even number while for chlorinated and brominated chemicals an even number of halogens is more common.
The relationship between chemical partitioning properties (log $K_{OW}$) and persistence is associated with higher log $K_{OW}$ values. However, phosphorus (log $P_{ov}$, log $K_{OW}$, log $K_{AW}$, log $K_{TE}$, log $P_{vPvB}$, and log $P_{APC90}$) for the Screen-POP chemicals. In general, higher persistence is associated with higher $K_{OW}$ values. The most persistent Screen-POP chemicals are volatile and semi-volatile molecules (indicated with dark blue). Notably, these Screen-POP compounds score very high in every profile, as illustrated in Fig. 2. "BAF" is positively associated with log $K_{OW}$, as expected by the BCFBAF model in EPI Suite. Highly water soluble chemicals (log $K_{AW} < -9$ and log $K_{OW} < 1$) are not efficiently transported through the atmosphere in our model calculations and hydrophobic and volatile chemicals (log $K_{AW} > 1$ and log $K_{OW} > 6$) are not likely to be deposited and thus they have the lowest log $TE$ values, in agreement with the definition of this metric. Volatile chemicals have high log $CTD_{air}$ values and highly water soluble chemicals the lowest ones. In contrast to the log $TE$ case, hydrophobic and volatile chemicals have high log $CTD_{air}$ values for the same reason; they are likely to be transported without being deposited. Similarly, such chemicals—when emitted in water—favor partitioning to particles and are transported with them in the water. Apart from these chemicals, water-soluble molecules that favor the dissolved phase in water also have high log $CTD_{water}$ values.

Figure 4 illustrates the influence of the presence of different atoms in the structures of the Screen-POP molecules—represented by the average number of atoms of this element per structure—on their hazard scores. More specifically, we observe that the presence of halogens in the molecular structure is associated with higher scores in all profiles while the presence of nitrogen and sulfur in the molecular structure is associated with lower scores for all profiles except for WPC. The presence of nitrogen and sulfur in the molecular structure is associated with higher $S_{WPC}$ scores but lower $S_{POP}$, $S_{vPvB}$, and $S_{APC}$ scores. Phosphorus seems to be generally associated with lower scores in every profile, and the presence of silicon is associated with higher estimated $K_{OW}$ values and, therefore, higher $BAF$ values and, consequently, higher $S_{POP}$ and $S_{vPvB}$ scores. Interestingly, organoboron Screen-POP compounds appear within the $S_{APC90}$ and $S_{WPC90}$ chemicals but not among the higher scoring Screen-POP chemicals in the "traditional" profiles.

In general, the average score as well as the median of the scores for each hazard profile for the Screen-POP profile, the chemical space region occupied by highly water soluble chemicals (bottom left corner) has the lowest $S_{APC}$ values (light grey color) as highly water soluble chemicals are not likely to have high $CTD_{water}$ scores. Between partitioning and hazard and to identify general trends. For example, in the case of the APC profile, the horizontal $K_{OW}$ axis, which reflects the influence of $K_{OW}$ on $BAF$ that is built into the BCFBAF model in EPI Suite, and also illustrated in Fig. 3.

Figure 3 consists of five three dimensional (log $K_{AW}$, log $K_{OW}$, log $K_{TE}$, log $P_{vPvB}$) subplots, each illustrating the relationship between chemical partitioning properties (log $K_{AW}$, log $K_{OW}$) and each of the five individual hazard metrics ($log P_{ov}$, log $BAF$, log $TE$, log $CTD_{air}$ and log $CTD_{water}$) for the Screen-POP chemicals. In general, high persistence is associated with higher $K_{OW}$ values. The most persistent Screen-POP chemicals are volatile and semi-volatile molecules (indicated with dark blue). Notably, these Screen-POP compounds score very high in every profile, as illustrated in Fig. 2. "BAF" is positively associated with log $K_{OW}$, as expected by the BCFBAF model in EPI Suite. Highly water soluble chemicals (log $K_{AW} < -9$ and log $K_{OW} < 1$) are not efficiently transported through the atmosphere in our model calculations and hydrophobic and volatile chemicals (log $K_{AW} > 1$ and log $K_{OW} > 6$) are not likely to be deposited and thus they have the lowest log $TE$ values, in agreement with the definition of this metric. Volatile chemicals have high log $CTD_{air}$ values and highly water soluble chemicals the lowest ones. In contrast to the log $TE$ case, hydrophobic and volatile chemicals have high log $CTD_{air}$ values for the same reason; they are likely to be transported without being deposited. Similarly, such chemicals—when emitted in water—favor partitioning to particles and are transported with them in the water. Apart from these chemicals, water-soluble molecules that favor the dissolved phase in water also have high log $CTD_{water}$ values.

Figure 4 illustrates the influence of the presence of different atoms in the structures of the Screen-POP molecules—represented by the average number of atoms of this element per structure—on their hazard scores. More specifically, we observe that the presence of halogens in the molecular structure is associated with higher scores in all profiles while the presence of nitrogen and sulfur in the molecular structure is associated with lower scores for all profiles except for WPC. The presence of nitrogen and sulfur in the molecular structure is associated with higher $S_{WPC}$ scores but lower $S_{POP}$, $S_{vPvB}$, and $S_{APC}$ scores. Phosphorus seems to be generally associated with lower scores in every profile, and the presence of silicon is associated with higher estimated $K_{OW}$ values and, therefore, higher $BAF$ values and, consequently, higher $S_{POP}$ and $S_{vPvB}$ scores. Interestingly, organoboron Screen-POP compounds appear within the $S_{APC90}$ and $S_{WPC90}$ chemicals but not among the higher scoring Screen-POP chemicals in the "traditional" profiles.

In general, the average score as well as the median of the scores for each hazard profile for the Screen-POP
chemicals is below 50, as presented in Fig. 5. These scores do not have any regulatory or risk significance, but reference set chemicals that define the boundaries of the score can be used to provide comparative context. Chlordane, 1,2,3,4,7,8-hexachlorodibenzodioxin, chloropiricin and 2,3,7,8-tetrachlorodibenzodioxin are the reference set chemicals that mark the top 10% interval for $S_{\text{POP}}$, $S_{\text{vPvB}}$, $S_{\text{APC}}$ and $S_{\text{WPC}}$ respectively. We denote as $S_{\text{PROFILE90}}$ the Screen-POP chemicals that scored over 90, i.e. chemicals that scored over the aforementioned reference set chemicals in each of the respective profiles. Similarly, we denote as $S_{\text{PROFILE10}}$ the Screen-POP chemicals that scored below 10. In total, 577 Screen-POP compounds scored over 90 in at least one of the four profiles, 80 Screen-POP chemicals did not score over 10 in any of the profiles, and six chemicals (sorbitol, inositol, isomaltitol, lactitol, polyvinyl alcohol and acarbose) scored 0 in all four profiles.

Four fluorinated chemicals exhibit such high estimated degradation half-lives in air that they scored 100 in every profile (Table 2) and seven more fluorinated chemicals scored over 90 in all four profiles (see Supporting Information). These eleven chemicals- the high scores of which...
are mostly attributable to very high estimated $P_{ov}$ rather than their partitioning properties- indicate the potentially dominant role of extremely high persistence in profiling chemicals for high exposure hazard.

A further exploration of the results illustrated in Fig. 4 showed that from the 97 chemicals that scored within $S_{Pop90}$, 20 chemicals scored that high only in this profile. Similarly, 114 chemicals scored within $S_{vPvB90}$ and more than half of them (63) scored over 90 only in the $vPvB$
are expected to be mostly relevant for regulatory action at a national-to-supranational scale (far-field exposure) while chemicals that fit the vPvB profile are of particular interest predominantly on a national scale (near-field exposure). Table 4 presents examples of such Screen-POP chemicals. Disulfide, bis(3,3,4,4,5,5-hexafluorohexyl) is a commenced premanufacture notice substance under the US toxic substances control act (TSCA) along with the whole class of bis(gamma-omega-perfluoro-C6-20-alkyl) disulfides (National Center for Biotechnology Information 2017c), 2-phenyl-1,1-bis(1h-indol-1-yl)ethane is a fragrance (European Chemicals Agency 2017) and pentabromotoluene is mainly used as a flame retardant (National Center for Biotechnology Information 2017e). Tris((3-ethyl-3-oxetanyl) methyl) phosphate is an additive found in an inventory list of substances used for the manufacture of surface coatings intended to come into contact with foodstuffs as a substance not fully evaluated by an international body (Committee of Ministers-Council of Europe 1996). In total, 114 chemicals scored within $S_{vPvB90}, 46$ chemicals scored within $S_{POP90}$ and not $S_{APC90}, 118$ chemicals scored within $S_{APC90}$ and not $S_{POP90}, 363$ chemicals scored within $S_{WPC90}$ and not $S_{POP90}$. Lists of these Screen-POP chemical groups are available in Supporting Information.

The Screen-POP database consists of chemicals that were reportedly produced in the OECD countries as of the year 2012. The chemical production of the OECD countries is estimated to account for at least half of the global chemical production volume (Sigman et al. 2001). Given the size and diversity of the Screen-POP database, our goal was to screen this set of chemicals for potential planetary boundary threats, POPs and vPvBs. We successfully identified high scoring Screen-POP chemicals that have a high potential to be POP, vPvB, APC and/or WPC candidates and we further proposed a scoring scheme that can be used to disentangle the complementary nature of these four profiles by focusing on the spatial range of their exposure potential. However, this approach to identify chemicals of concern, and especially potential planetary boundary threats, does not regard the socioeconomic modes of transfer of pollutants as exposure pathways (MacLeod et al. 2014; Ng and Goetz 2016).

The majority of the Screen-POP chemicals examined in this study scored low relative to the reference set in all profiles as the Screen-POP database is not contaminant-focused. However, scores should not be interpreted as a direct proxy for “risk” because possible effects are not explicitly considered in this assessment. Moreover, the actual emissions of these chemicals are necessary to estimate their exposure and also, occasionally, there are significant differences between the actual and the estimated physicochemical properties which define these scores. More specifically, EPI Suite is known to underestimate degradation half-lives of persistent chemicals (Gouin et al. 2004) and the physicochemical
properties of some chemical classes, such as organosilicon and organoborons, which are underrepresented in EPI Suite’s training sets, are subject to extremely high uncertainty. Various Screen-POP chemicals are not expected to fall equally well within the applicability domain of all the different EPI Suite QSARs (HENRYWIN, KOWWIN, BCF-BAF, AOPWIN and BIOWIN) used in this assessment and, therefore, chemicals that scored high should be assigned high priority for experimental estimation of their physicochemical properties.

In addition to chlorofluorocarbon 123, our method assigned a high $S_{APC}$ score to decabromobiphenyl (National Center for Biotechnology Information 2017b) and dichloromethane (National Center for Biotechnology Information 2017d). A high $S_{APC}$ score for decabromobiphenyl is unexpected as this is a chemical with low vapor pressure and therefore its presence in air can only be attributed to its sorption to air particles. Although chemicals that are highly sorbed to atmospheric particles may be transported in air if they are emitted to air, it is questionable whether the atmospheric half-lives of aerosol particles are sufficient to render...
such low-volatility chemicals of concern as APCs. Therefore, we identify the need to further explore the relationship between partitioning properties and scores and specifically examine the appropriateness of the $S_{APC}$ score to capture potential APCs. On the other hand, dichloromethane, which is present both in the reference set and the Screen-POP database, was recently identified as an ozone depleting chemical (Hossaini et al. 2017). This finding illustrates the utility of our ranking method and highlights the relative advantages of using a reference set rather than adopting “bright line” criteria in prioritization assessments.

Screening-level studies are highly dependent on data availability and data quality. The in silico tools and methodologies developed thus far have enabled estimation of property data and provided a rigorous framework to identify and prioritize organic chemicals for unacceptable environmental behavior. Therefore, the main limitations of such studies are on the data availability side, both in terms of measurements

| Name/Structure | $\log K_{ow}$, $\log K_{aw}$, $t_{1/2air}$, $t_{1/2water}$, $t_{1/2soil}$ (all $t_{1/2}$ in hours), $\log k_M$ ($k_M$ in days$^{-1}$ for 10 g fish) |
|---------------|--------------------------------------------------------------------------------------------------|
| octadecafluorooctane | 6.17, 2.99, $10^5$, 4320, 8640, -2.19 |
| hentriacontafluorodecane-1-sulfonyl fluoride | 6.56, 2.99, $10^5$, 4320, 8640, -2.06 |
| Perfluorooctyl iodide | 6.99, 2.99, $10^5$, 4320, 8640, -1.78 |
| flutecpp11 | 7.45, 2.99, $10^5$, 4320, 8640, -2.4 |
of property data of diverse substances to train QSPR models, and representativeness of the databases of chemicals used in the screening exercise. The Screen-POP database is large, diverse, focused on chemicals that were actually produced at least up to 2011, and by design covers an important part of the global chemical production. However, this database is a snapshot. Since this dataset was compiled, several new individual chemicals have been introduced in the market while information on the chemical production of non-OECD countries is scarce. Therefore, field studies and analytical approaches are necessary to identify new contaminants that would otherwise remain undiscovered (Sobek et al. 2016; Blum et al. 2018).

### Table 3 Examples of chemicals that scored over 90 (i.e. \(S_{\text{PROFILE90}}\)) in only one of the profiles

| Name, Structure, Scoring scheme | \(\log K_{\text{OW}}, \log K_{\text{AW}}, t_{1/2}\text{air}, t_{1/2}\text{water}, t_{1/2}\text{soil}\) (all \(t_{1/2}\) in hours), \(\log k_M\) (\(k_M\) in days\(^{-1}\) for 10 g fish) |
|--------------------------------|---------------------------------------------------------------------------------------------------------------------------------|
| Guaiene \(S_{\text{POP90}}\) only | 6.56, -0.25, 0.21, 900, 1800, -2.32 |
| CAS RN: 88-84-6 | |
| Monocarbinol terminated \(S_{\text{POP90}}\) only | 6.54, -6.15, 8.1, 900, 1800, -1.81 |
| CAS RN: 104780-66-7 | |
| Chlorofluorocarbon 123 \(S_{\text{APC90}}\) only | 2.17, 0.02, 7050, 1440, 2880, 0.86 |
| CAS RN: 306-83-2 | |
| Cyhalothrin \(S_{\text{WPC90}}\) only | 7, -4.22, 7.8, 4320, 8640, -0.74 |
| CAS RN: 68085-85-8 | |
Future studies building on our approach to screen chemicals for high exposure hazard potential should aim to expand the reference set currently used to better account for more diverse chemical databases, expand the methodology to include ionic and ionizing chemicals, improve on the EPI Suite-estimated degradation half-lives, and further investigate the influence of the partitioning properties on the profile scores.

Acknowledgements Financial support for this study was provided by the European Union’s Seventh Framework Programme for research, technological development and demonstration under Grant Agreement No. 316665 (A-TEAM project). We thank Jon Arnot and the CEFIC-LRI ECO.13 research team for providing the ScreenPOP database.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

ACC Y-E 2014 Chemical Industry Situation and Outlook American Chemistry Builds Momentum, (2014). http://files.clickdimensions.com/americanchemistrycom-avo5d/files/year-end2014situationandoutlook6fc2.pdf

Advanced Chemistry Development Inc., ACD/AbsoLv (2015). http://www.acdlabs.com

AMAP (2017) Chemicals of emerging Arctic concern: summary for policy-makers. Arctic Monitoring and Assessment Programme (AMAP), Oslo. http://www.amap.no/documents/doc/chemicals-of-emerging-arctic-concern.-summary-for-policy-makers/1533
Apodaca R (2011) Sixty-four free chemistry databases. http://depth-first.com/articles/2011/10/12/sixty-four-free-chemistry-data-bases/. Accessed 7 Jul 2017

Arnot JA (2009) Mass balance models for chemical fate, bioaccumulation, exposure and risk assessment. In: Arnot JA (ed) Exposure and risk assessment of chemical pollution: contemporary methodology. Springer, Berlin, pp 69–91

Arnot JA, Brown TN, Wania F, Breivik K, McLachlan MS (2012) Prioritizing chemicals and data requirements for screening-level exposure and risk assessment. Environ Health Perspect 120:1565–1570. https://doi.org/10.1289/ehp.1205355

Bennett ER, Clausen J, Linkov E, Linkov I (2009) Predicting physical properties of emerging compounds with limited physical and chemical data: QSAR model uncertainty and applicability to military munitions. Chemosphere 77:1412–1418. https://doi.org/10.1016/j.chemosphere.2009.09.003

Bernhardt ES, Rosi EJ, Gessner MO (2017) Synthetic chemicals as agents of global change. Front Ecol Environ. https://doi.org/10.1002/fee.1450

Binetti R, Costamagna FM, Marcello I (2008) Exponential growth of persistent organic pollutants: a case study of Arctic contaminants. Environ Sci Technol. 42 (2008) 5202–5209. http://www.ncbi.nlm.nih.gov/pubmed/18754370

Binetti R, Costamagna FM, Marcello I (2008) Exponential growth of persistent organic pollutants: a case study of Arctic contaminants. Environ Sci Technol. 42 (2008) 5202–5209. http://www.ncbi.nlm.nih.gov/pubmed/18754370

Boethling RS, Howard PH, Meylan WM (2004) Finding and estimating chemical property data for environmental assessment. Environ Toxicol Chem 23:2290–2308. https://doi.org/10.1897/03-532

Breivik K, Arnot J, Brown TN, McLachlan MS, Wania F (2012) Screening organic chemicals in commerce for emissions in the context of environmental and human exposure. J Environ Monit 14:2028–2037. https://doi.org/10.1039/c2em30259d

Brown TN, Wania F, Screening chemicals for the potential to be persistent organic pollutants: a case study of Arctic contaminants. Environ Sci Technol. 42 (2008) 5202–5209. http://www.ncbi.nlm.nih.gov/pubmed/18754370

Cao Y, Charisi A, Cheng LC, Jiang T, Girke T (2008) ChemmineR: a compound mining framework for R. Bioinformatics 24:1733–1734. https://doi.org/10.1093/bioinformatics/btn307

CEFC (2014) European chemical industry facts and figures. http://fr.zone-secure.net/13451/106811/publication/contents/templates/Cefic_F_and_F_2014-Full_report_Blanc_150dpi.pdf

Committee of Ministers-Council of Europe (1999) Resolution AP (99) 5 On surface coatings intended to come into contact with food-stuffs. http://www.contactalimentaire.com/fileadmin/ImageFichier_Collection/Fichiers_Documents/Resolution_96-5_ang.pdf

Committee of Ministers-Council of Europe (1999) Resolution AP (99) 3 On silicones used for food contact applications. http://www.contactalimentaire.com/fileadmin/ImageFichier_Collection/contact_alimentaire/Fichiers_Documents/Resolution_anglais/resolution_99-3_ang.pdf

Committee of Ministers-Council of Europe (1999) Resolution AP (99) 3 On silicones used for food contact applications. http://www.contactalimentaire.com/fileadmin/ImageFichier_Collection/contact_alimentaire/Fichiers_Documents/Resolution_anglais/resolution_99-3_ang.pdf

Diamond ML, de Wit CA, Molander S, Scheringer M, Backhaus T, Kohlmum R, Arvidsson J, Bergman Å, Hauschild M, Holoubek I, Persson L, Suzuki N, Vighi M, Zetzsch C (2015) Exploring the planetary boundary for chemical pollution. Environ Int 78:8–15. https://doi.org/10.1016/j.envint.2015.02.001

European Chemicals Agency (2017) 1,1′-(2-phenylethylidene)bis(1H-indole). https://echa.europa.eu/substance-information/-/substan ceinfo/100061.034. Accessed 6 Dec 2017

European Parliament the Council of the European Union, Regulation (EC) No. 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/./, Off J Eur Union. 396 (2006)

Fenner K, Scheringer M, MacLeod M, Matthies M, McKone T, Stroeb M, Beyer A, Bonnell M, Le Gall AC, Klasmeier J, Mackay D, van de Meent D, Pennington D, Scharenberg B, Suzuki N, Wania F (2005) Comparing estimates of persistence and long-range transport potential among multimedia models. Environ Sci Technol 39:1932–1942. https://doi.org/10.1021/es048917b

Gouin T, Cousins I, Mackay D (2004) Comparison of two methods for obtaining degradation half-lives. Chemosphere 56:531–535. https://doi.org/10.1016/j.chemosphere.2004.04.018

Hossaini R, Chipperfield MP, Montzka SA, Leeson AA, Dhomse SS, Pyle JA (2017) The increasing threat to stratospheric ozone from dichloromethane. Nat Commun 8:15962. https://doi.org/10.1038/ncomms15962

Mackay D, McCarty LS, MacLeod M (2001) On the validity of classifying chemicals for persistence, bioaccumulation, toxicity, and potential for long-range transport., Environ Toxicol Chem 20:1491–1498. http://www.ncbi.nlm.nih.gov/pubmed/11343289

MacLeod M, Breitholtz M, Cousins IT, de Wit CA, Persson LM, Rudén C, McLachlan MS (2014) Identifying chemicals that are planetary boundary threats. Environ Sci Technol 48:11057–11063. https://doi.org/10.1021/es501893m

Matthies M, Solomon K, Vighi M, Gilman A, Tarazona JV (2016) The origin and evolution of assessment criteria for persistent, bioaccumulative and toxic (PBT) chemicals and persistent organic pollutants (POPs). Environ Sci Process Impacts 18:1114–1128. https://doi.org/10.1039/C6EM00311G

McLachlan MS, Kierkegaard A, Radke M, Sobek A, Malmvärn A, Alsberg T, Arnot JA, Brown TN, Wania F, Breivik K, Xu S (2014) Using model-based screening to help discover unknown environmental contaminants. Environ Sci Technol 48:7264–7271. https://doi.org/10.1021/es5010544

Muir DCG, Howard PH (2006) Are there other persistent organic pollutants? A challenge for environmental chemists, Environ Sci Technol. 40:7157–7166. http://www.ncbi.nlm.nih.gov/pubmed/17180962

National Center for Biotechnology Information (2017a) PubChem compound database; CID = 5281873. https://pubchem.ncbi.nlm.nih.gov/cmpd/5281873

National Center for Biotechnology Information (2017b) PubChem compound database; CID = 26164. https://pubchem.ncbi.nlm.nih.gov/cmpd/26164. Accessed 6 Dec 2017

National Center for Biotechnology Information (2017c) PubChem compound database; CID = 3086586. https://pubchem.ncbi.nlm.nih.gov/cmpd/3086586. Accessed 6 Dec 2017

National Center for Biotechnology Information (2017d) PubChem compound database; CID = 6344. https://pubchem.ncbi.nlm.nih.gov/cmpd/6344. Accessed 6 Dec 2017

National Center for Biotechnology Information (2017e) PubChem compound database; CID = 6906. https://pubchem.ncbi.nlm.nih.gov/cmpd/6906. Accessed 6 Dec 2017

National Center for Biotechnology Information (2017f) PubChem compound database; CID = 3086586. https://pubchem.ncbi.nlm.nih.gov/cmpd/3086586. Accessed 6 Dec 2017

Ng CA, von Goetz N (2016) The global food system as a transport pathway for hazardous chemicals: the missing link between emissions and exposure. Environ. Health Perspect. https://doi.org/10.1289/EHP168

Persson LM, Breitholtz M, Cousins IT, de Wit C, MacLeod M, McLachlan MS (2013) Confronting unknown planetary boundary threats. Environ Sci Technol 48:11057–11063. https://doi.org/10.1021/es501893m

Pyle JA (2017) The increasing threat to stratospheric ozone from dichloromethane. Nat Commun 8:15962. https://doi.org/10.1038/ncomms15962

Reppas-Chrysostovitis E, Sobek A, MacLeod M (2017) Screening-level exposure-based prioritization to identify potential POPs,
vPvBs and planetary boundary threats among Arctic contaminants. Emerg Contam 3:85–94. https://doi.org/10.1016/j.emcon.2017.06.001
Rockström J, Steffen W, Noone K, Persson A, Chapin FS, Lambin EF, Lenton TM, Scheffer M, Folke C, Schellnhuber HJ, Nykvist B, de Wit CA, Hughes T, van der Leeuw S, Rodhe H, Sörlin S, Snyder PK, Costanza R, Svedin U, Falkenmark M, Karlberg L, Corell RW, Fabry VJ, Hansen J, Walker B, Liverman D, Richardson K, Crutzen P, Foley JA (2009) A safe operating space for humanity. Nature 461:472–475
Rorije JMPM, Verbruggen E, Hollander EMJ, Traas ATP (2011) Identifying potential POP and PBT substances, Development of a New Persistence/Bioaccumulation-Score. RIVM Report 601356001/2011. http://www.rivm.nl/bibliotheek/rapporten/601356001.pdf. Accessed 22 Dec 2017
Scheringer M (2002) Persistence and spatial range of environmental chemicals, 1st edn. Wiley, Weinheim
Scheringer M, Strempel S, Hukari S, Ng Ca, Blepp M, Hungerbühler K (2012) How many persistent organic pollutants should we expect? Atmos Pollut Res 3:383–391. https://doi.org/10.5094/APR.2012.044
Schüürmann G, Ebert R-U, Nendza M, Dearden JC, Paschke A, Kühne R (2007) Predicting fate-related physicochemical properties, In: van Leeuwen CJ, Vermeire TG (eds) Risk assessment of chemicals, Springer, Berlin, pp 375–426
Sigman R, Ladeuille B, Grandy N (2001) OECD environmental outlook for the chemicals industry Siltech, Silmer® OH Di-10, (2016). http://polymer-additives.special-chem.com/product/m-siltech-silmer-oh-di-10. Accessed 22 Aug 2017
Sobek A, Beigarn S, Rudén C, Breitholtz M (2016) The dilemma in prioritizing chemicals for environmental analysis: known versus unknown hazards. Environ Sci Process Impacts 18:1042–1049. https://doi.org/10.1039/C6EM00163G
Steffen W, Richardson K, Rockström J, Cornell S, Fetzer I, Bennett E, Biggs R, Carpenter S (2015) Planetary boundaries: guiding human development on a changing planet. Science 348:1217. https://doi.org/10.1126/science.aaa9629
Strempel S, Scheringer M, Ng CA, Hungerbühler K (2012) Screening for PBT chemicals among the “existing” and “new” chemicals of the EU. Environ Sci Technol 46:5680–5687. https://doi.org/10.1021/es3002713
Sushko I, Novotarskyi S, Körner R, Pandey AK, Rupp M, Teetz W, Brandmaier S, Abdelaziz A, Prokopenko VV, Tanchuk VY, Todeschini R, Varnek A, Marcou G, Ertl P, Potemkin V, Grishina M, Gasteiger J, Schwab C, Baskin H, Palyulin V, Radchenko EV, Welsh WJ, Khodolovych V, Chekmarev D, Cherkasov A, Aires-de-Sousa J, Zhang Q-Y, Bender A, Nigsch F, Patiny L, Williams A, Tkachenko Y, Teko IV (2011) Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. J Comput Aided Mol Des 25:533–554. https://doi.org/10.1007/s10822-011-9440-2
The Good Scents Company, Guaiene (2017) http://www.thegoodscentscCompany.com/data/rw1012371.html. Accessed 6 Dec, 2017
UNEP (2001) Stockholm Convention on Persistent Organic Pollutants, United Nations Environment Program. https://chm.pops.int/. Accessed 27 Jan 2017
United States Environmental Protection Agency, Estimation Programs Interface Suite™ for Microsoft® Windows (2012). http://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface
Walker JD, Carlens L (2002) QSARs for identifying and prioritizing substances with persistence and bioconcentration potential. SAR QSAR Environ Res 13:713–725. https://doi.org/10.1080/1062936021000043454
Wegmann F, Cavin L, MacLeod M, Scheringer M, Hungerbuhler K (2009) The OECD software tool for screening chemicals for persistence and long-range transport potential. Environ Model Softw 24:228–237. https://doi.org/10.1016/j.envsoft.2008.06.014