

## Article

# Software-Assisted Pattern Recognition of Persistent Organic Pollutants in Contaminated Human and Animal Food

Wenjing Guo <sup>1</sup>, Jeffrey Archer <sup>2</sup>, Morgan Moore <sup>2</sup>, Sina Shojaee <sup>2</sup>, Wen Zou <sup>1</sup>, Weigong Ge <sup>1</sup>, Linda Benjamin <sup>3</sup>, Anthony Adeuya <sup>4</sup>, Russell Fairchild <sup>2</sup> and Huixiao Hong <sup>1,\*</sup> 

<sup>1</sup> National Center for Toxicological Research, U.S. Food & Drug Administration, 3900 NCTR Road, Jefferson, AR 72079, USA; Wenjing.guo@fda.hhs.gov (W.G.); Wen.Zou@fda.hhs.gov (W.Z.); Weigong.Ge@fda.hhs.gov (W.G.)

<sup>2</sup> Office of Regulatory Affairs, Office of Regulatory Science, Arkansas Laboratory, U.S. Food & Drug Administration, 3900 NCTR Road, Jefferson, AR 72079, USA; Jeffrey.Archer@fda.hhs.gov (J.A.); Morgan.Moore@fda.hhs.gov (M.M.); Sina.Shojaee@fda.hhs.gov (S.S.); Russell.Fairchild@fda.hhs.gov (R.F.)

<sup>3</sup> Center for Veterinary Medicine, U.S. Food & Drug Administration, 7500 Standish Place, Rockville, MD 20855, USA; Linda.Benjamin@fda.hhs.gov

<sup>4</sup> Center for Food Safety and Applied Nutrition, U.S. Food & Drug Administration, 5001 Campus Dr, College Park, MD 20740, USA; Anthony.Adeuya@fda.hhs.gov

\* Correspondence: huixiao.hong@fda.hhs.gov; Tel.: +1-(870)-543-7296

**Abstract:** Persistent Organic Pollutants (POPs) are a serious food safety concern due to their persistence and toxic effects. To promote food safety and protect human health, it is important to understand the sources of POPs and how to minimize human exposure to these contaminants. The POPs Program within the U.S. Food and Drug Administration (FDA), manually evaluates congener patterns of POPs-contaminated samples and sometimes compares the finding to other previously analyzed samples with similar patterns. This manual comparison is time consuming and solely depends on human expertise. To improve the efficiency of this evaluation, we developed software to assist in identifying potential sources of POPs contamination by detecting similarities between the congener patterns of a contaminated sample and potential environmental source samples. Similarity scores were computed and used to rank potential source samples. The software has been tested on a diverse set of incurred samples by comparing results from the software with those from human experts. We demonstrated that the software provides results consistent with human expert observation. This software also provided the advantage of reliably evaluating an increased sample lot which increased overall efficiency.

**Keywords:** persistent organic pollutant; software; similarity; congener pattern; contamination



**Citation:** Guo, W.; Archer, J.; Moore, M.; Shojaee, S.; Zou, W.; Ge, W.; Benjamin, L.; Adeuya, A.; Fairchild, R.; Hong, H. Software-Assisted Pattern Recognition of Persistent Organic Pollutants in Contaminated Human and Animal Food. *Molecules* **2021**, *26*, 685. <https://doi.org/10.3390/molecules26030685>

Academic Editors: Emanuela Zanardi and Lenka Husáková

Received: 28 December 2020

Accepted: 16 January 2021

Published: 28 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Environment pollution has become a significant food safety concern due to pollutants detected in the environment being transferred to the food supply. Exposure to environmental pollutants such as persistent organic pollutants (POPs) [1–6] may cause various serious health effects including mutagenic, carcinogenic, and cardiovascular disease as well as endocrine disruption [7–10]. POPs are a class of carbon-based chemicals that may contain chlorine/bromine/fluorine in their structure. This stable chemical structure makes POPs either resistant to environmental degradation or slow to degrade. Thus, humans and animals can be exposed to POPs over long periods of time. To protect public health, many international and national regulations have been issued to reduce the release of POPs into the environment such as the Stockholm Convention [11–14], the United Nations Economic Commission for Europe (UNECE) POPs Protocol [15], the Regulation for Registration, Evaluation, Authorization and Restriction of Chemicals (REACH, EC no. 1907/2006 as amended) [16], the EU regulation for the Placing of Plant Protection Products on the Market (EC 1107/2009) [17], and the U.S. Toxic Substances Control Act [18]. As a result of these

regulations and legal frameworks, levels of POPs in the environment have been decreasing. However, POPs continue to be of great environmental concern due to their persistence and significant negative impact on human and environmental health.

Human exposure to POPs is mostly through contaminated food, especially fatty foods due to the lipophilic nature of the contaminants. There are many pathways by which POPs can enter the food chain. For example, previously released POPs in the environment can easily be transferred to crops, livestock, and seafood which can eventually pose a significant threat to public health. Aside from the environment, POPs can be introduced to the food supply by human activity. There are many accidents reported such as Yusho disease in 1968 [19] and Yucheng poisoning in 1979 [20] where the public was accidentally exposed to high levels of POPs. Contaminated animal feed supplies can also introduce POPs into food as occurred in the Belgian dioxin crisis in 1999 [21] and the Ireland dioxin accident in 2008 [22]. To ensure food safety, many international and national bodies have set food and animal feed limits for commonly found POPs, and related surveillance programs have been developed to ensure that POPs contamination is below harmful level [23–25].

If a food sample is found to contain high concentrations of POPs, it is important to trace sources of contamination so that measures can be taken to remove them and prevent further contamination. Since POPs-contaminated samples exhibit specific congener profiles, the congener-specific concentrations can be used as a fingerprint to trace sources of contamination. For example, the source of dioxin-contaminated milk in the Netherlands was identified to be kaolinic clay used in the bovine feed supply because its congener pattern was similar to the known congener patterns of clays [26]. Recent studies have shown that the congener patterns method is a useful tool to identify sources of contamination [27–30]. Hoogenboom et al. reviewed some known incidents and showed how visual comparison of congener patterns was used to determine sources of contamination [27]. Besides visual comparison, statistical methods have been used to compare POPs congener patterns including permutational multivariate analysis of variance [31], canonical correlation [32], principal component analysis [28,33–36], linear discriminant analysis [34,35] and multidimensional scaling [31]. These statistical methods are general tools that have been used to assess the relationship between two datasets. Therefore, some characteristics of POPs are not considered. For example, these methods cannot be used to compare congener patterns from both the concentration and the toxicity level. In addition to this, low level POPs would be ignored in the samples. Therefore, to improve the similarity calculation, we developed a software with algorithms designed specifically for POPs.

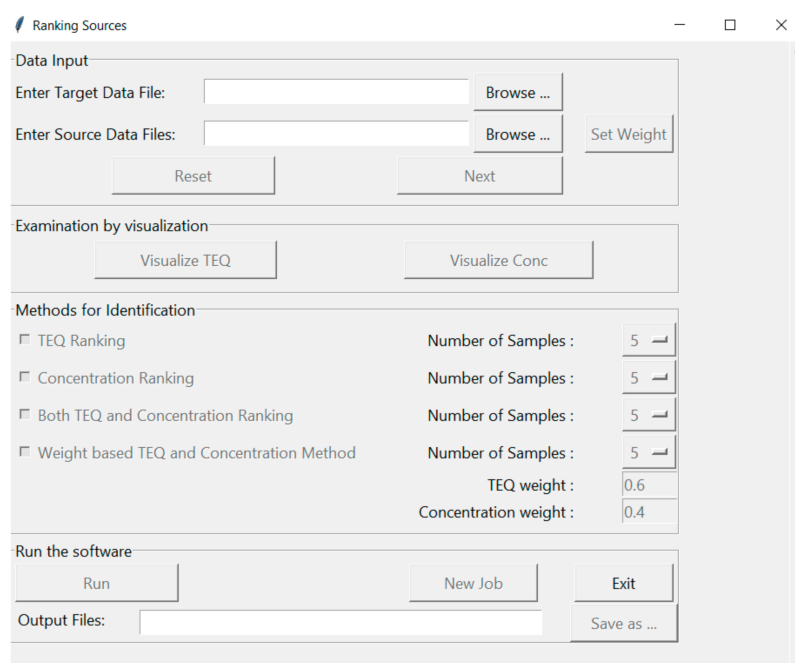
In the FDA laboratory, when a food sample is found to contain high levels of POPs, analysts examine the congener pattern of the food sample to infer the source of contamination. The concentration and congener pattern of POPs used in the examination is the excel and python processed targeted mass spectrometry (MS) data [37]. In many cases analysts need to compare the pattern of a contaminated food sample with samples collected by FDA investigators from the same farm or producer. If the pattern of the contaminated sample matches the pattern of a collected sample, the collected sample can be identified as a potential source of contamination. The contaminated sample is a sample that contains elevated naturally incurred levels of some POPs that we would like to find the source. The collected samples are historical samples or newly collected samples by the FDA that can be compared to “Contaminated Sample” in order to determine source of incurred levels in the “Contaminated Sample”. However, this visual comparison is dependent on human expertise. As the size and complexity of collected samples increase, it becomes increasingly difficult to conduct the comparison manually. To overcome this difficulty, we developed software to facilitate the identification of POPs-contaminated sources by computing similarities between a contaminated sample and potential source samples. Different similarity scores were computed and used to rank potential source samples. A similarity rank of potential source samples was provided to analysts for further investigation. Our results demonstrate that this software increased efficiency for the identification of potential sources of POPs contamination.

## 2. Results

### 2.1. Software Details and Interface

The software was developed in Python 3.6 (Python Software Foundation, Beaverton, OR, USA, <http://python.org>), and the graphical user interface was built using Tkinter module. The source code and Windows executable files are available upon request.

An illustration of the software interface, shown in Figure 1, contains four sections; (1) Data Input, (2) Examination by Visualization, (3) Method for Identification and (4) Run the Software. The first section, “Data Input”, prompts the user to select a contaminated sample of interest, “Enter Target Data File”, while the second prompt “Enter Source Data Files” contains the historical contaminated samples that are used for comparison (referred to as potential source samples hereafter). The sample files are csv files containing three columns: congener names, their concentrations and TEQs. The user can also manually weight results by selecting the “Set Weight” button and choosing congeners of interest. The default weight for each congener is 1, meaning that all congeners are considered equally. If there is a need or desire to focus on specific congeners, this can be accomplished by setting the weights of those congeners to 1 and the other values to zero. After selecting “Next” the user can examine data graphically in the second section, “Examination by Visualization”, by either concentration or by Toxic Equivalent (TEQ). All data are plotted for each congener for all samples provided in the first section. The third section, “Methods for Identification”, allows users to choose ranking methods. Based on the similarity scores computed, there are four ranking methods: (1) TEQ Ranking is based on the TEQ similarity score, (2) Concentration Ranking is based on the concentration similarity score, (3) Both TEQ and Concentration Ranking rank potential source samples equally by both TEQ and concentration similarity scores, (4) Weight-based TEQ and Concentration Ranking apply weights to both TEQ and concentration similarity scores to generate a combined similarity score for ranking samples. The “Number of Samples” ranked can be chosen by the user; however, the default value is set to 5 ranked samples. In the fourth section, “Run” can be selected to view similarity rankings and then saved “Save as . . . ” to a user specified location, or the program can be closed by selecting the “Exit” button.



**Figure 1.** Software interface. Functions are grouped into five sections. Section titles are shown in the top left corners of the section rectangles. Buttons with black text are functions that can be executed. Buttons with grey text are functions that need to be activated after users input data or parameters, or execute functions.

## 2.2. Performance Test

The software performance was evaluated by comparing results derived from human analysts to those obtained from the software. Rankings of the potential source samples from the software (some shown below) were compared with rankings made by human experts. Use of the software allows many samples to be quickly evaluated for similarities, thus significantly reducing the evaluation time.

### 2.2.1. Potential Source Sample Test Case

The software was tested on ten contaminated food samples. Here, animal tissue was used as an example to show software performance. Elevated levels of dioxins/furans were found in an animal tissue [38] and sent to FDA for follow-up studies. To trace the potential sources of the elevated dioxins/furans, FDA investigators collected 11 potential source samples such as wood slivers from a utility pole and soil from the same farm. The lab focused on seven dioxins, ten furans, and three dioxin-like PCBs while evaluating the patterns. Similarity scores comparing the contaminated animal tissue TEQ pattern to patterns of matrices collected at the farm are shown in Table 1. Wood slivers from a utility pole showed the highest similarity score at 0.63, indicating that this sample was a potential source of contamination.

**Table 1.** Ranking and TEQ similarity scores with the animal tissue sample of the 11 potential source samples.

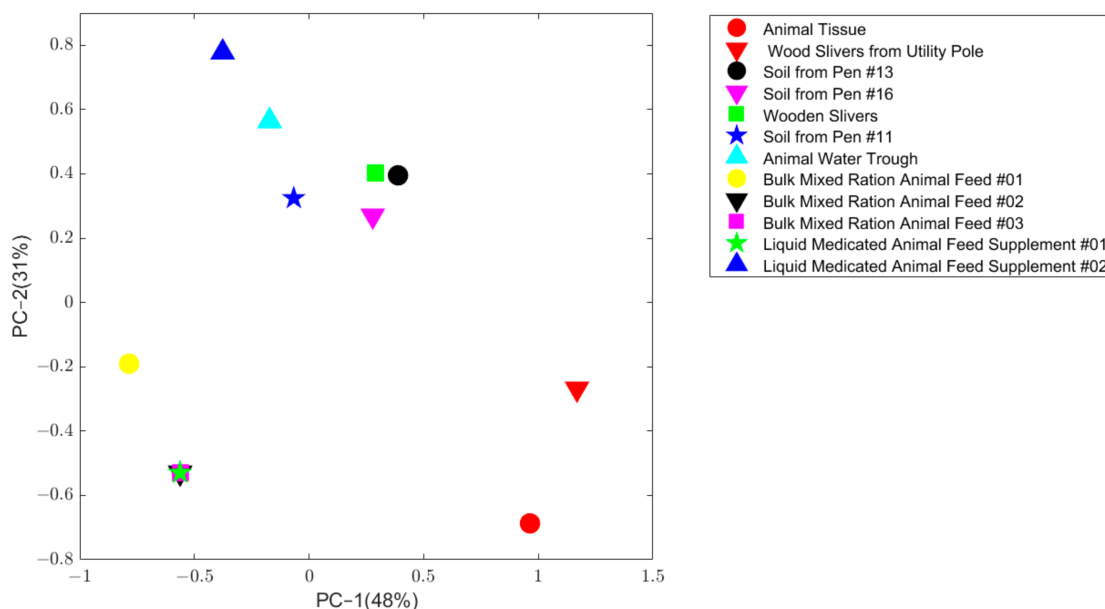
Ranking	Potential Source Sample	Similarity Score
1	Wood Slivers from Utility Pole	0.63
2	Soil from Pen #13	0.58
3	Soil from Pen #16	0.56
4	Wooden Slivers	0.54
5	Animal Water Trough	0.50
6	Bulk Mixed Ration Animal Feed #02	0.50
7	Bulk Mixed Ration Animal Feed #03	0.50
8	Liquid Medicated Animal Feed Supplement #01	0.50
9	Liquid Medicated Animal Feed Supplement #02	0.50
10	Soil from Pen #11	0.49
11	Bulk Mixed Ration Animal Feed #01	0.43

The TEQ pattern for the contaminated animal tissue sample was compared to the 11 FDA collected potential source samples using principal component analysis (PCA). As shown in Figure 2, wood sliver from a utility pole sample (solid red downward triangle) is the closest point to the contaminated animal tissue sample (solid red circle), indicating this sample has the most similar pattern as the contaminated sample, indicating the similarity score from the software could be used to represent the similarity in congeners pattern.

### 2.2.2. Potential Source Sample Test Results

For quality control purposes, duplicate samples are tested monthly to evaluate analytical precision in FDA Dioxin laboratory [37]. If the precision is acceptable, the original sample data should be statistically consistent with duplicate sample data and their similarity score should be relatively high. In this study, data from 19 original and duplicate sample pairs were used to test the performance of the software. The sample details for these 19 pairs are shown in Table 2. Some sample pairs with incomplete analytical results were not included in this study. These samples do not necessarily contain high levels of POPs. The software was not developed to show the similarity of duplicates, but historical duplicates were used here to show the reliability of the software in calculating similarity scores. The 38 potential source samples were ranked using weight-based similarity scores that were calculated with the “Concentration weight” of 0.4 and the “TEQ weight” of 0.6. The value of 0.4 and 0.6 are recommended by human expert based on their experience that TEQ played a more important role when they make the decision. Users can choose

any weight they like when using the software. The weight-based ranking methods was chosen so that both concentration and TEQ patterns were considered in the similarity calculation. The matrix of similarity scores between the 38 potential source samples are shown in Figure 3.



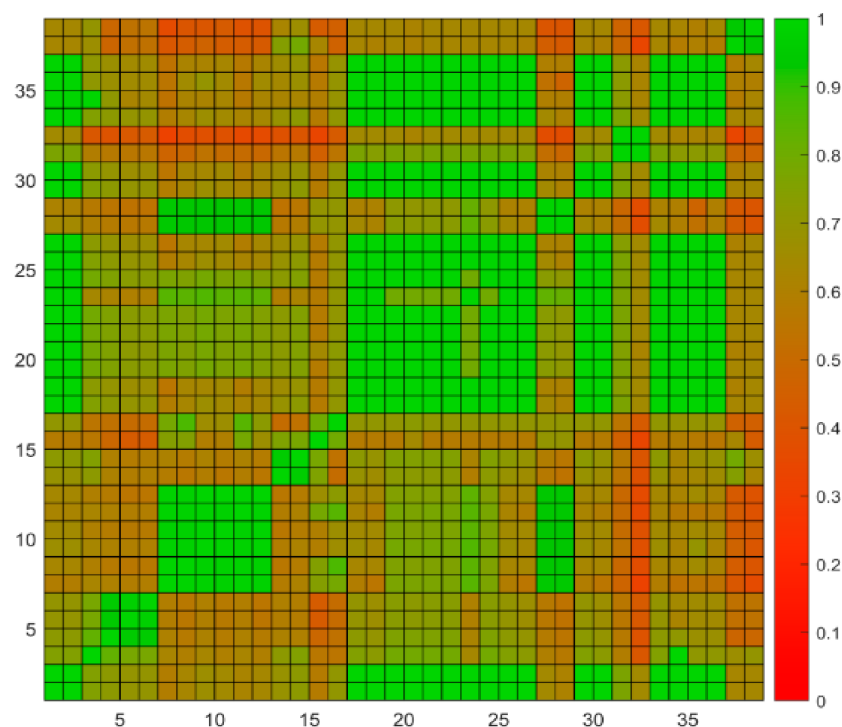
**Figure 2.** Principle component analysis on the normalized concentration of the animal tissue sample and 11 potential source samples. The x axis is the first principal component, and the y axis is the second principal component. Samples were plotted in different colors and shapes which are given in the legend located at top right.

**Table 2.** Samples details for duplicate and original samples.

Sample Number	Sample Matrix
Sample 1	Animal Feed ferrous sulfate duplicate
Sample 2	Animal Feed ferrous sulfate original
Sample 3	Animal feed finisher ration duplicate
Sample 4	Animal feed finisher ration original
Sample 5	Animal feed starter ration duplicate
Sample 6	Animal feed starter ration original
Sample 7	Bluefish 1 duplicate
Sample 8	Bluefish 1 original
Sample 9	Bluefish 2 duplicate
Sample 10	Bluefish 2 original
Sample 11	Bluefish 3 duplicate
Sample 12	Bluefish 3 original
Sample 13	Complete turkey ration duplicate
Sample 14	Complete turkey ration original
Sample 15	Crab 1 duplicate
Sample 16	Crab 1 original
Sample 17	Crab 2 duplicate
Sample 18	Crab 2 original
Sample 19	Cream Substitute duplicate
Sample 20	Cream Substitute original
Sample 21	Milk duplicate
Sample 22	Milk original
Sample 23	Poultry Mixed feed ration duplicate
Sample 24	Poultry Mixed feed ration original
Sample 25	Sweet rolls duplicate

Table 2. Cont.

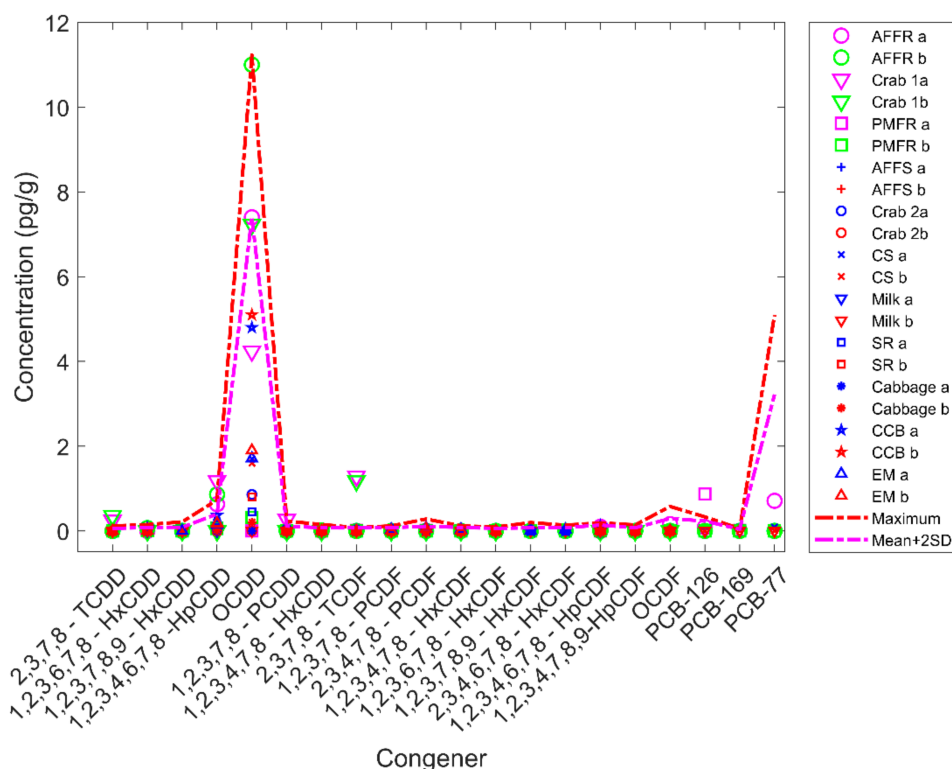
Sample Number	Sample Matrix
Sample 26	Sweet rolls original
Sample 27	Trout duplicate
Sample 28	Trout original
Sample 29	Cabbage duplicate
Sample 30	Cabbage original
Sample 31	Chocolate candy bar 1 duplicate
Sample 32	Chocolate candy bar 1 original
Sample 33	Chocolate candy bar 2 duplicate
Sample 34	Chocolate candy bar 2 original
Sample 35	English muffins duplicate
Sample 36	English muffins original
Sample 37	Frankfurters duplicate
Sample 38	Frankfurters original



**Figure 3.** Heatmap of similarity scores. The 19 original potential source samples are shown in odd numbers and their duplicates are shown in even numbers. Similarity scores are coded in colors defined in the right legend.

The similarity scores between the 19 potential source samples and their corresponding duplicates were high (green cells in Figure 3), indicating that the software reported the potential source samples at values very similar to their duplicates as expected. There are 52% of the duplicate (10 pairs) have their similarity scores equal to 1 while 78% (14 pairs) have scores greater than 0.98. We noticed that some similarity scores from non-paired samples were also high. For example, sample 1 is the duplicate of sample 2, which is animal feed ferrous sulfate. The similarity score between them is equal to 1 indicates that the experiments are of high quality. Interestingly, both samples 1 and 2 showed very high similarity scores at 1 with samples 17 and 18 which are duplicates of a crab sample, 19 and 20 which are duplicates of a cream substitute sample, 21 and 22 which are duplicates of a milk sample, 25 and 26 which are duplicates of a sweet rolls sample, 29 and 30 which are duplicates of a cabbage sample, 31 and 32 which are duplicates of a chocolate candy bar sample, 33 and 34 which are duplicates of another chocolate candy bar sample, 35

and 36 which are duplicates of a English muffin sample (Figure 3). A closer look at the POPs detected in these samples revealed that all of the samples contained very low POPs concentrations. Comparing these samples with randomly selected 100 blank samples tested in the FDA Dioxin laboratory found that the POPs detected were below the mean plus two times standard deviation of POPs levels of the blank samples (Figure 4), indicating these samples likely contain POPs near or below detection limit values. Human experts concluded that these samples were of minimal concern regarding levels of POPs detected. Again, the manual process and software reported consistently similar results.



**Figure 4.** Concentration (pg/g) of congeners acquired from mass spectrometry data. The congeners are labeled on the x-axis. The y-axis shows determined concentrations. Samples are plotted in different shapes with different colors for original and duplicate samples as shown in the right legend panel. The red line shows the maximum concentrations of congeners determined in the 100 randomly selected blank samples. The magenta line depicts mean concentrations plus two times the standard deviation for each of the 100 blank samples. Abbreviations used for sample names: AFFR-animal feed finisher ration; PMFR-poultry mixed feed ration; AFFS-animal feed ferrous sulfate; CS-cream substitute; SR-sweet rolls; CCB-chocolate candy bar; EM-English muffins.

On the other hand, the software reported similarity scores 0.8252, 0.8007, and 0.7951 for pairs 3 and 4 which are duplicates of an animal feed finisher ration sample, 15 and 16 which are duplicates of a crab sample, and 23 and 24 which are duplicates of a poultry mixed feed ration sample, respectively (Figure 3), indicating the concentrations of congeners for these three sample pairs were not perfectly consistent. Inspection of the data showed that a few congeners showed different concentrations in these pairs of samples (Figure 4). For example, 0.87 pg/g PCB-126 was determined in one duplicate of poultry mixed feed ration sample 23; but no PCB-126 was detected in the other duplicate sample 24. Although sample 23 showed a slightly higher PCB-126 concentration than the maximum PCB-126 concentration (0.3317 pg/g) among the 100 selected blank samples, levels of the other 19 congeners were below detection limits (compared to blank samples), thus the human experts concluded that the concentration pattern of samples 23 and 24 are similar. The software reported their similarity score at 0.7951, indicating the two samples exhibit a similar POPs pattern. In summary, the comparison between conclusions drawn from

the software reports and human expert evaluation showed no discrepancy for any of the 38 potential source samples tested.

### 3. Discussion

Comparison of congener patterns, acquired from MS data, is used to help identify sources of POPs contamination in food. Manual comparison by subject matter experts is time consuming and dependent on the experience of the analysts. The software developed here has increased sample capability, therefore, efficiency while identifying POPs contamination sources in regulatory science. To improve the efficiency of contamination source identification, we developed a software for automatic comparison of patterns of congener concentration determined using excel and python processed MS data [37]. The enhanced software greatly streamlines the process by reducing the data evaluation time. Our tests confirmed that the software can report results consistent with manual human expert evaluation.

The software was tested for comparing patterns of concentrations of dioxin/furan and dioxin-like PCBs in an FDA Dioxin laboratory. However, the software has no restrictions on the types of POPs or the number of potential source samples that can be processed. Once concentration/TEQ data is provided, the software can compute similarity scores based on common congeners for samples in one single run without user intervention. The different types of similarity scores increase the capability of this software well beyond the manual ability of human analysts. Multiple versions of the same data can be easily viewed which increases the opportunity to identify otherwise missed similarities.

There are some limitations of the software. The software ranks potential samples based on their similarity scores in congener concentrations and TEQs compared to a contaminated food sample. It is possible that the top ranked potential source samples are not always the sole source of contamination; this can be due to changes in the congener patterns as the congeners move through the food chain due to their different absorption, distribution, metabolism, and excretion rates [27]. A possible way to overcome this is to introduce animal bioconcentration factors. The software allows users to manually enter transfer rates as weights for congeners. However, due to the difficulty in obtaining the transfer rate for each food sample and each congener, we did not use this as its default setting in our software.

## 4. Materials and Methods

### 4.1. Study Design

This study aims to automate the process of identifying sources of POPs contamination by calculating the similarities between a contaminated sample and potential source samples. Our software is expected to calculate the similarity score between contaminated samples and potential source samples and suggest potential sources ranked by their similarity scores.

The concentration and calculated TEQ obtained from MS data are used to compute the similarity score between a contaminated sample and each potential source sample. TEQ is the product of concentration and toxic equivalency factor (TEF) [39] and is often used to measure the toxicity of a congener. The concentration/TEQ data provided to the software are compared with the concentration/TEQ data from 100 randomly selected blank samples. When a concentration/TEQ is smaller than the expected detection error determined using the 100 randomly selected blank samples (mean + 2 times the standard deviation), the congener should be considered absent from the potential source samples, thus the concentration/TEQ is set to zero. When a concentration/TEQ is larger than the expected detection error, the expected detection error is subtracted from the concentration/TEQ.

After preprocessing the input data, three steps are used to calculate similarity scores in the software. First, concentrations/toxic equivalency (TEQ) are normalized by their maximum value to show the relative contribution of each congener. Second, the normalized concentrations/TEQ are used to calculate the similarity between a contaminated sample



and each potential source sample. Then the computed similarity scores are used to rank the potential source samples.

#### 4.2. Similarity Calculation

##### 4.2.1. Normalized Concentration/TEQ

In this step, the concentration/TEQ for each congener is normalized by the maximum value in a sample. The software calculates both concentration and TEQ values to provide the user with three options to evaluate the similarity: concentration-based pattern, TEQ-based pattern, and the combined pattern. This normalization step converts the congener concentration/TEQ to a scale between 0 and 1. The normalized concentration/TEQ is calculated by Equation (1):

$$c_i^n = \frac{c_i^o}{c_i^{max}} \quad (1)$$

where  $c_i^n$  is the normalized concentration/TEQ for congener  $i$ ;  $c_i^o$  is the original concentration/TEQ for congener  $i$ ;  $c_i^{max}$  is the maximum concentration/TEQ of congeners in the sample.

##### 4.2.2. Calculate Similarity

The normalized concentration/TEQ is used to calculate similarity score. This similarity score calculation is based on three hypotheses. First, the overall similarity of two samples is based on the similarities of individual congeners. Second, congeners with higher concentration weights more than the low-level congeners since the source of high-level congeners are more important than that of low-level congeners. Third, the low-level congeners still contribute to the overall calculation. This would enable the software compare samples containing very low POPs concentrations. The similarity  $S_p^c$  between a potential source sample to a contaminated sample is calculated by Equation (2):

$$S_p^c = \sum_{i=1}^n \frac{w_i(c_i^c + c_i^p + 0.2566)}{\sum(c_i^c + c_i^p + 0.2566)} \left( \frac{2}{1 + |c_i^c - c_i^p|} - 1 \right) \quad (2)$$

where  $n$  is the number of congeners in the sample;  $w_i$  is the weight of congener  $i$ . The value of  $w_i$  is manually entered by users. Its default value is 1 meaning that all the congeners would be considered equally.  $c_i^c$  and  $c_i^p$  represent the normalized concentration/TEQ for congener  $i$  in a contaminated sample and a potential source sample, respectively. An adjustment factor of 0.2566 is the mean concentration of congeners from 100 randomly selected blank samples. The similarity score is between 0 and 1. High similarity values indicate a greater match is between a contaminated sample and a potential source sample. Here, we calculated two types of similarity score: the TEQ similarity score based on TEQ of each congener, and the concentration similarity score based on the concentration of each congener.

##### 4.2.3. Ranking Potential Source Samples

The software provides four methods for ranking potential source samples. In the first two methods, the potential source samples are ranked using TEQ similarity scores or concentration similarity scores, respectively. The third method ranks the potential source samples using consensus TEQ and concentration similarity scores. In this method, the potential source samples are first ranked separately using their TEQ and concentration similarities. Then samples common to both lists are selected and their average TEQ and concentration similarity scores are used to rank the potential source samples. For example, to find the top 5 potential source samples using the consensus ranking, the software first finds the two top 5 listings using TEQ and concentration ranking. Then potential source samples common to these two lists are ranked using the average value of TEQ and concentration similarity scores. The fourth method ranks potential source samples using

weight-based similarity scores that combine concentration and TEQ similarity scores. The weight-based similarity score is calculated using Equation (3):

$$S = (TEQ\ weight) \times S_{TEQ} + (Concentration\ weight) \times S_{conc} \quad (3)$$

where  $S_{TEQ}$  is the TEQ similarity score and  $S_{conc}$  is the concentration similarity score. If the weight-based method is chosen, the software then asks users to enter TEQ and *concentration weights*. If users want to focus on toxicity level, a larger value should be given to *TEQ weight*. After obtaining user input values, the software computes the weight-based similarity scores and ranks the potential source samples.

## 5. Conclusions

This software has increased the sample capability to compare large numbers of potential source samples using their similarity to a contaminated food sample. Similarity scores are calculated between a contaminated sample and potential source samples. Different types of similarity scores can be computed and used to rank potential source samples based on user needs. We tested the software on a diverse set of samples and compared the results from the software with results from human experts. The tests confirmed that the software yielded results consistent with human expert results. Our study results demonstrated that the software is reliable in ranking samples based on their POPs concentration patterns.

**Author Contributions:** Conceptualization, W.G. (Wenjing Guo), J.A. and H.H.; Data curation, J.A. and M.M.; Software, W.G. (Wenjing Guo), W.Z., W.G. (Weigong Ge) and H.H.; Validation, W.G. (Wenjing Guo), J.A., M.M., S.S., L.B., A.A. and R.F.; Writing—original draft, W.G. (Wenjing Guo) and H.H.; Writing—review & editing, J.A., L.B., A.A. and R.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The data are not publicly available due to data confidentiality requirements.

**Acknowledgments:** This research was supported in part by an appointment to the Research Participation Program at the National Center for Toxicological Research (Wenjing Guo), administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Food and Drug Administration. The authors thank Jeffrey Bruce and Michelle McLain in the Arkansas Laboratory, Office of Regulatory Affairs, FDA for technical assistance with data, and Gwenn Merry in the Office of Research at the National Center for Toxicological Research for her review and outstanding editing on this manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Disclaimer:** The views presented in this article do not necessarily reflect those of the U.S. Food and Drug Administration References.

**Sample Availability:** Samples of the compounds are not available from the authors.

## References

1. Ashraf, M.A. Persistent organic pollutants (POPs): A global issue, a global challenge. *Environ. Sci. Pollut. Res.* **2017**, *24*, 4223–4227. [[CrossRef](#)]
2. Gaur, N.; Narasimhulu, K.; Y, P. Recent advances in the bio-remediation of persistent organic pollutants and its effect on environment. *J. Clean. Prod.* **2018**, *198*, 1602–1631. [[CrossRef](#)]
3. El-Shahawi, M.; Hamza, A.; Bashammakh, A.; Al-Saggaf, W. An overview on the accumulation, distribution, transformations, toxicity and analytical methods for the monitoring of persistent organic pollutants. *Talanta* **2010**, *80*, 1587–1597. [[CrossRef](#)]
4. Jones, K.C.; De Voogt, P. Persistent organic pollutants (POPs): State of the science. *Environ. Pollut.* **1999**, *100*, 209–221. [[CrossRef](#)]
5. Beyer, A.; Mackay, D.; Matthies, M.; Wania, F.; Webster, E. Assessing Long-Range Transport Potential of Persistent Organic Pollutants. *Environ. Sci. Technol.* **2000**, *34*, 699–703. [[CrossRef](#)]
6. Kelly, B.C.; Ikononou, M.G.; Blair, J.D.; Morin, A.E.; Gobas, F.A.P.C. Food Web-Specific Biomagnification of Persistent Organic Pollutants. *Science* **2007**, *317*, 236–239. [[CrossRef](#)]

7. Guo, W.; Pan, B.; Sakkiah, S.; Yavas, G.; Ge, W.; Zou, W.; Tong, W.; Hong, H. Persistent Organic Pollutants in Food: Contamination Sources, Health Effects and Detection Methods. *Int. J. Environ. Res. Public Health* **2019**, *16*, 4361. [CrossRef]
8. Alharbi, O.M.; Basheer, A.A.; Khattab, R.A.; Ali, I. Health and environmental effects of persistent organic pollutants. *J. Mol. Liq.* **2018**, *263*, 442–453. [CrossRef]
9. Vorkamp, K.; Rigét, F.F. A review of new and current-use contaminants in the Arctic environment: Evidence of long-range transport and indications of bioaccumulation. *Chemosphere* **2014**, *111*, 379–395. [CrossRef]
10. Li, Q.Q.; Loganath, A.; Chong, Y.S.; Tan, J.; Obbard, J.P. Persistent Organic Pollutants and Adverse Health Effects in Humans. *J. Toxicol. Environ. Health Part A* **2006**, *69*, 1987–2005. [CrossRef]
11. Tang, H.P.-O. Recent development in analysis of persistent organic pollutants under the Stockholm Convention. *TrAC Trends Anal. Chem.* **2013**, *45*, 48–66. [CrossRef]
12. Xu, W.; Wang, X.; Cai, Z. Analytical chemistry of the persistent organic pollutants identified in the Stockholm Convention: A review. *Anal. Chim. Acta* **2013**, *790*, 1–13. [CrossRef] [PubMed]
13. Vallack, H.W.; Bakker, D.J.; Brandt, I.; Brostrom-Lunden, E.; Brouwer, A.; Bull, K.R.; Gough, C.; Guardans, R.; Holoubek, I.; Jansson, B.; et al. Controlling Persistent Organic Pollutants-What Next? *Environ. Toxicol. Pharmacol.* **1998**, *6*, 143–175. [CrossRef]
14. UNEP. *Stockholm Convention on Persistent Organic Pollutants*; Secretariat of the Stockholm Convention Report: Geneva, Switzerland, 2001; p. 43.
15. UNECE. Protocol to the 1979 Convention on Long-Range Transboundary Air Pollution on Persistent Organic Pollutants. *UNECE: Aarhus, Denmark*. 1998. Available online: [http://www.unece.org/env/lrtap/pops\\_h1.html](http://www.unece.org/env/lrtap/pops_h1.html) (accessed on 18 December 2020).
16. EC. Commission Regulation (EU) No 253/2011 of 15 March 2011 Amending Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) as Regards Annex XIII. *Off. J. Eur. Union.* **2011**, *69*, 7–12.
17. EP&C. Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 Concerning the Placing of Plant Protection Products on the Market and Repealing Council Directives 79/117/EEC and 91/414/EEC (91/414/EEC). *Off. J. Eur. Union.* **2009**, *309*, 1–50.
18. U.S. EPA. Category for Persistent, Bioaccumulative, and Toxic New Chemical Substances. *Fed. Regist.* **1999**, *64*, 20194–60204.
19. Yoshimura, T. Yusho in Japan. *Ind. Health* **2003**, *41*, 139–148. [CrossRef]
20. Hsu, S.-T.; Ma, C.-I.; Hsu, S.K.-H.; Wu, S.-S.; Hsu, N.H.-M.; Yeh, C.-C. Discovery and epidemiology of PCB poisoning in Taiwan. *Am. J. Ind. Med.* **1984**, *5*, 71–79. [CrossRef]
21. Bernard, A.; Broeckaert, F.; De Poorter, G.; De Cock, A.; Hermans, C.; Saegerman, C.; Houins, G. The Belgian PCB/Dioxin Incident: Analysis of the Food Chain Contamination and Health Risk Evaluation. *Environ. Res.* **2002**, *88*, 1–18. [CrossRef]
22. Pratt, I.; Anderson, W.A.; Crowley, D.; Daly, S.F.; Evans, R.I.; Fernandes, A.R.; Fitzgerald, M.; Geary, M.; Keane, D.P.; Malisch, R.; et al. Polychlorinated dibenzo-p-dioxins (PCDDs), polychlorinated dibenzofurans (PCDFs) and polychlorinated biphenyls (PCBs) in breast milk of first-time Irish mothers: Impact of the 2008 dioxin incident in Ireland. *Chemosphere* **2012**, *88*, 865–872. [CrossRef]
23. Van Leeuwen, F.; Feeley, M.; Schrenk, D.; Larsen, J.C.; Farland, W.; Younes, M. Dioxins: WHO's tolerable daily intake (TDI) revisited. *Chemosphere* **2000**, *40*, 1095–1101. [CrossRef]
24. European Commission. Commission Regulation (EC) No 1881/2006 of 19 December 2006 setting maximum levels for certain contaminants in foodstuffs. *Off. J. Eur. Union* **2006**, *364*, 324–365.
25. EC. Commission Regulation (EU) 2020/685 of 20 May 2020 Amending Regulation (EC) No 1881/2006 as Regards Maximum Levels of Perchlorate in Certain Foods. *Off. J. Eur. Union.* **2020**, *160*, 3–5.
26. Hoogenboom, R.L.; Zeilmaker, M.; Van Eijkeren, J.; Kan, K.; Mengelers, M.; Luyckx, D.; Traag, W. Kaolinic clay derived PCDD/Fs in the feed chain from a sorting process for potatoes. *Chemosphere* **2010**, *78*, 99–105. [CrossRef]
27. Hoogenboom, R.; Malisch, R.; van Leeuwen, S.P.J.; Vanderperren, H.; Hove, H.; Fernandes, A.; Schächtele, A.; Rose, M. Congener Patterns of Polychlorinated Dibenzo-P-Dioxins, Dibenzofurans and Biphenyls as a Useful Aid to Source Identification During a Con-tamination Incident in the Food Chain. *Sci. Total Environ.* **2020**, *746*, 141098. [CrossRef]
28. Johansson, A.-K.; Sellström, U.; Lindberg, P.; Bignert, A.; de Wit, C.A. Polybrominated Diphenyl Ether Congener Patterns, Hexabromocyclododecane, and Brominated Biphenyl 153 in Eggs of Peregrine Falcons (*Falco peregrinus*) Breeding in Sweden. *Environ. Toxicol. Chem.* **2009**, *28*, 9–17. [CrossRef]
29. Kim, K.S.; Hirai, Y.; Kato, M.; Urano, K.; Masunaga, S. Detailed PCB congener patterns in incinerator flue gas and commercial PCB formulations (Kanechlor). *Chemosphere* **2004**, *55*, 539–553. [CrossRef]
30. Santillo, D.; Fernandes, A.; Stringer, R.; Alcock, R.; Rose, M.; White, S.; Jones, K.C.; Johnston, P. Butter as an indicator of regional persistent organic pollutant contamination: Further development of the approach using polychlorinated dioxins and furans (PCDD/Fs), and dioxin-like polychlorinated biphenyls (PCBs). *Food Addit. Contam.* **2003**, *20*, 281–290. [CrossRef]
31. Gerig, B.S.; Chaloner, D.T.; Janetski, D.J.; Rediske, R.R.; O'Keefe, J.H.; Moerke, A.H.; Lamberti, G.A. Congener Patterns of Persistent Organic Pollutants Establish the Extent of Contaminant Biotransport by Pacific Salmon in the Great Lakes. *Environ. Sci. Technol.* **2015**, *50*, 554–563. [CrossRef]
32. Wu, M.-H.; Pei, J.-C.; Zheng, M.; Tang, L.; Bao, Y.-Y.; Xu, B.-T.; Sun, R.; Sun, Y.-F.; Xu, G.; Lei, J. Polybrominated diphenyl ethers (PBDEs) in soil and outdoor dust from a multi-functional area of Shanghai: Levels, compositional profiles and interrelationships. *Chemosphere* **2015**, *118*, 87–95. [CrossRef]

33. Pozo, K.; Harner, T.; Rudolph, A.; Oyola, G.; Estellano, V.H.; Ahumada-Rudolph, R.; Garrido, M.; Pozo, K.; Mabilia, R.; Focardi, S. Survey of Persistent Organic Pollutants (POPs) and Polycyclic Aromatic Hydrocarbons (PAHs) in the Atmosphere of Rural, Urban and Industrial Areas of Concepción, Chile, Using Passive Air Samplers. *Atmos. Pollut. Res.* **2012**, *3*, 426–434. [[CrossRef](#)]
34. Jin, R.; Fu, J.; Zheng, M.; Yang, L.; Habib, A.; Li, C.; Liu, G. Polychlorinated Naphthalene Congener Profiles in Common Vegetation on the Tibetan Plateau as Biomonitors of Their Sources and Transportation. *Environ. Sci. Technol.* **2020**, *54*, 2314–2322. [[CrossRef](#)]
35. Antignac, J.P.; Marchand, P.; Gade, C.; Matayron, G.; Qannari el, M.; Le Bizec, B.; André, F. Studying Variations in the PCDD/PCDF Profile across Various Food Products Using Multivariate Statistical Analysis. *Anal Bioanal Chem.* **2006**, *384*, 271–279. [[CrossRef](#)] [[PubMed](#)]
36. Alava, J.J.; Keller, J.M.; Wyneken, J.; Crowder, L.; Scott, G.; Kucklick, J.R. Geographical variation of persistent organic pollutants in eggs of threatened loggerhead sea turtles (*Caretta caretta*) from southeastern United States. *Environ. Toxicol. Chem.* **2011**, *30*, 1677–1688. [[CrossRef](#)] [[PubMed](#)]
37. Guo, W.; Archer, J.C.; Moore, M.; Bruce, J.; McLain, M.; Shojaee, S.; Zou, W.; Benjamin, L.A.; Adeuya, A.; Fairchild, R.; et al. QUICK: Quality and Usability Investigation and Control Kit for Mass Spectrometric Data from Detection of Persistent Organic Pollutants. *Int. J. Environ. Res. Public Health* **2019**, *16*, 4203. [[CrossRef](#)]
38. Dioxin FY2013 Survey: Dioxins and Dioxin-Like Compounds in the U.S. Domestic Meat and Poultry Supply. Available online: <https://www.fsis.usda.gov/wps/portal/fsis/topics/data-collection-and-reports/chemistry/dioxin-related-activities> (accessed on 18 December 2020).
39. Berg, M.V.D.; Birnbaum, L.S.; Denison, M.; De Vito, M.; Farland, W.; Feeley, M.; Fiedler, H.; Hakansson, H.; Hanberg, A.; Haws, L.; et al. The 2005 World Health Organization Reevaluation of Human and Mammalian Toxic Equivalency Factors for Dioxins and Dioxin-Like Compounds. *Toxicol. Sci.* **2006**, *93*, 223–241. [[CrossRef](#)]