FINAL REPORT

Machine Learning Pattern Recognition for Forensic Analysis of Detected Per- and Polyfluoroalkyl Substances in Environmental Samples

SERDP Project ER20-1205

OCTOBER 2020

Tohren Kibbey
University of Oklahoma

Rafal Jabrzemski
University of Oklahoma

Denis O'Carroll
University of New South Wales

Distribution Statement A
This document has been cleared for public release

SERDP
DOD ▪ EPA ▪ DOE

This report was prepared under contract to the Department of Defense Strategic Environmental Research and Development Program (SERDP). The publication of this report does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official policy or position of the Department of Defense. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the Department of Defense.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
10/14/2020	SERDP Final Report	11/5/2019 – 11/5/2020

4. TITLE AND SUBTITLE

Machine Learning Pattern Recognition for Forensic Analysis of Detected Per- and Polyfluoroalkyl Substances in Environmental Samples

5a. CONTRACT NUMBER
W912HQ20P0005

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Tohren Kibbey
Rafal Jabrzemski
Denis O'Carroll

5d. PROJECT NUMBER
ER20-1205

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Oklahoma
Norman, OK 73019
USA

University of New South Wales
Sydney, NSW 2052
Australia

8. PERFORMING ORGANIZATION REPORT NUMBER

ER20-1205

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Strategic Environmental Research and Development Program
4800 Mark Center Drive, Suite 17D03
Alexandria, VA 22350-3605

10. SPONSOR/MONITOR'S ACRONYM(S)
SERDP

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
ER20-1205

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution A: Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT This proof of concept project explored the use of modern machine learning algorithms for forensic source allocation of detected per- and polyfluoroalkyl substances (PFAS) in environmental samples. A challenge with allocating PFAS source by composition is that PFAS composition varies substantially even from a single source, as a result of differing compound mobility, as well as transformation of precursor compounds. Work involved compiling an extensive dataset from a range of worldwide PFAS concentration data sources, and then using a portion of the dataset to train supervised machine learning classifiers to recognize patterns that can distinguish between PFAS from AFFF and non-AFFF sources. Results show that a number of very different supervised learning classifiers work extremely well for this purpose, and are able to identify AFFF in samples even in difficult-to-recognize categories of samples. The results of this project show that supervised machine learning has significant promise as a means of distinguishing between PFAS from AFFF and non-AFFF sources.

15. SUBJECT TERMS
PFAS, Supervised Machine Learning, Pattern Recognition, Forensics, Source Allocation

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Tohren Kibbey
UNCLASS	UNCLASS	UNCLASS	UNCLASS	212	**19b. TELEPHONE NUMBER** *(include area code)* 405-325-0580

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

TABLE OF CONTENTS

List of Tables iii
List of Figures v
Acronyms and Abbreviations vii
Keywords viii
Acknowledgments ix
Abstract 1
Introduction and Objectives. 1
Technical Approach. 1
Results. 1
Benefits. 1
Executive Summary 3
Introduction 3
Objectives 3
Technical Approach 4
Results and Discussion 5
Data Collection and Preprocessing 5
Supervised Machine Learning for PFAS Classification 6
Clustering 10
Preliminary Data Mining 10
Implications for Future Research and Benefits 12
1. Objectives 13
1.A. Criteria for Success, and Risk Reduction for Future Work. 13
1.B. Relationship to SERDP SON. 14
2. Background 15
2.A. AFFF Formulations in the Environment 15
2.B. Machine Learning 17
2.C. Training and Assessing Machine Learning Algorithms 18
3. Materials and Methods 19
3.A. Approach 19
3.B. Methods 21
4. Results and Discussion 24
4.A. Data Collection and Preprocessing (Task 1). 24

4.A.1. Dataset Content. 24
4.A.2. Dataset Structure. 38
4.B. Supervised Machine Learning for Source Allocation (Subtasks 2a, 2b and 3b) 42
4.B.1. Dataset, Features and Training. 42
4.B.2. Source Allocation Results. 43
4.B.3. Difficult Subsets. 52
4.B.4. Identification of Sample Types Not in the Training Set 56
4.C. Feature Selection (Subtask 2b) 59
4.D. Clustering (Subtask 2c) 62
4.E. Preliminary Data Mining (Subtask 3a) 69
4.E.1. Prediction of PFAS Surfactant Parameters Driven by Hydrophobicity 69
4.E.2. The Effect of Distribution On PFAS Component Fingerprint 71
4.E.3. Machine Learning for non-Water Phases 78
5. Conclusions and Implications for Future Research 82
5.A. Conclusions 82
5.A.1. Supervised Learning 82
5.A.2. Feature Selection 82
5.A.3. Clustering 83
5.A.4. Preliminary Data Mining 83
5.B. Potential Next Steps 83
1. Creation of an Expanded Dataset 84
2. Selection of Methods for Rejection/Open Set Recognition 84
3. Exploration of Advanced Training Methods 84
4. Multiclass Classification 84
5. Simulations to Better Understand Classification Performance 84
6. Creation of the Final Tool 85
6. Literature Cited 86
Appendices 89
Appendix A. Data Source Information 89
Appendix B. Concentration-Averaged Component Distributions 120

LIST OF TABLES

Table ES-1. Overview of sample classification in the dataset. 5

Table ES-2. Contents of the dataset by country and classification. 6

Table ES-3. Possible outcomes in the binary classification of AFFF/non-AFFF samples. 7

Table ES-4. Logarithmic, Distribution: Performance metrics for classification of the 4053-sample test set, shown as percent for 4 classifiers. 8

Table ES-5. Classification accuracy for three subsets of the dataset. 9

Table 1. Overview of sample classification in the dataset. 24

Table 2. Contents of the dataset by country and classification. 25

Table 3. Detailed listing of data sources in the dataset, including sample classifications and concentration ranges. 27

Table 4. Components included in the dataset, and abbreviations selected for machine learning. 31

Table 5. Components analyzed at each site for groundwater and surface water. Blue indicates at least one sample analysis conducted for this compound. Only components through EtFOSAA in Table 4 are shown. 37

Table 6. Possible outcomes of binary classification. 43

Table 7. Possible outcomes in the binary classification of AFFF/non-AFFF samples. 43

Table 8. Performance measures for machine learning classifiers. 43

Table 9. Unscaled, Concentration: Performance metrics for classification of the 4053-sample test set, shown as percent for the 12 classifiers. 45

Table 10. Unscaled, Distribution: Performance metrics for classification of the 4053-sample test set, shown as percent for the 12 classifiers. 46

Table 11. Logarithmic, Concentration: Performance metrics for classification of the 4053-sample test set, shown as percent for the 12 classifiers. 47

Table 12. Logarithmic, Distribution: Performance metrics for classification of the 4053-sample test set, shown as percent for the 12 classifiers. 48

Table 13. Classification accuracy for three subsets of the dataset. 52

Table 14. Component combinations corresponding to the ***highest*** **classification accuracies** for each total number of components. Selected components are indicated by blue shading. 60

Table 15. Component combinations corresponding to the ***lowest*** **classification accuracies** for each total number of components. Selected components are indicated by blue shading. 60

Table 16. The accuracy (%) of clustering at distinguishing between AFFF and non-AFFF in the 10-component groundwater/surface water dataset. 62

Table 17. log K_{iw-0} coefficients determined from fits to surface tension data. Coefficients correspond K_{iw-0} in units of cm. 69

Table 18. Classification of soil samples using the transformed model, compared with the original (water) model. 79

Table 19. Classification of sediment samples using the transformed model, compared with the original (water) model..... 80

Table 20. Classification of biota samples using the transformed model, compared with the original (water) model..... 81

Table 21. Classification of concrete samples using the transformed model, compared with the original (water) model..... 81

LIST OF FIGURES

Figure ES-1. Fraction perfluorinated sulfonates vs. total PFAS concentration in 759 groundwater and surface water samples from a site where contamination is known to be primarily from PFOS-based AFFF. 4

Figure ES-2. Graphical representation of the confusion matrix for predictions conducted using the **Random Forest classifier** for a 4053 sample test data set. 9

Figure ES-3. Comparison between transformed groundwater composition data and data for other phases for RAAF Base Pearce (Australia, AFFF). 11

Figure 1. Conceptual illustration of the problem addressed by this work. 14

Figure 2. Fraction perfluorinated sulfonates vs. total PFAS concentration in 759 groundwater and surface water samples from a site where contamination is known to be primarily from PFOS-based AFFF. 16

Figure 3. Numbers of samples in the complete dataset for which individual compound analyses are available. 34

Figure 4. Numbers of AFFF samples in the dataset for which individual compound analyses are available. 35

Figure 5. Numbers of non-AFFF samples in the dataset for which individual compound analyses are available. 36

Figure 6. Graphical representation of the confusion matrix for predictions conducted using the **Random Forest classifier** for a 4053 sample test data set. 50

Figure 7. Graphical representation of the confusion matrix for predictions conducted using the **Deep Neural Network Ensemble classifier** for a 4053 sample test data set. 51

Figure 8. Classification of a subset of the test set corresponding to four sites where **non-PFOS-based AFFFs** are thought to be the major sources of PFAS present. 54

Figure 9. Classification of a subset of the test set corresponding to two sites where **perfluorinated-sulfonate-based non-AFFF formulations** are thought to be the major sources of PFAS present. 55

Figure 10. Effect of removing individual site data from training set on classification for AFFF data from military bases. Classification based on the Random Forest classifier. 56

Figure 11. Partial component distributions for 9 individual military bases, showing the similarity of composition between sites. 57

Figure 12. Partial component distributions for two sites that are unlike others in the dataset, so pose challenges for classification if samples for these sites are not included in the training set.. 58

Figure 13. Classification accuracy as a function of number of components selected for the 1023 individual combinations of components in the 10 component set used for analyses in this section. 59

Figure 14. Number of detects for each of the components in the training dataset. 61

Figure 15. Clustering of the groundwater samples at RAAF Base Williamtown using GMM. Colors indicated the two clusters. Total PFAS concentration is indicated by symbol size. 63

Figure 16. Clustering of the groundwater samples at RAAF Base Williamtown using K-Means. Colors indicated the two clusters. Total PFAS concentration is indicated by symbol size. 63

Figure 17. Two component clustering of a 10 component PFAS mixture transported from a single source. Clustering based on GMM. Symbol colors indicate identified clusters. 64

Figure 18. GMM clustering of three different sets of samples from the same simulation shown in Fig. 17, indicating that the compositions of the included points influence clustering, producing completely different spatial patterns. 65

Figure 19. GMM Clustering of randomly selected points in the simulation of transport from two distinct PFAS sources, one AFFF (lower) and one non-AFFF. 66

Figure 20. Clustering of randomly selected points in the simulation of transport from two distinct PFAS sources, one AFFF (lower) and one non-AFFF. A. Mean-shift clustering, and B. Spectral clustering. 67

Figure 21. Illustration of PFOS-normalized relationships. 70

Figure 22. Comparison between transformed groundwater composition data and data for other phases for RAAF Base Pearce (Australia, AFFF). 73

Figure 23. Comparison between transformed groundwater composition data and data for other phases for DND Site G (Canada, AFFF). 74

Figure 24. Comparison between transformed groundwater composition data and data for other phases for Brisbane Airport (Australia, AFFF). 75

Figure 25. Comparison between transformed groundwater composition data and data for other phases for Bennington (USA, non-AFFF). 76

Figure 26. Comparison between transformed groundwater composition data and data for other phases for Wolverine Tannery Site (USA, non-AFFF). 77

ACRONYMS AND ABBREVIATIONS

PFAS	Per- and polyfluoroalkyl substances
AFFF	Aqueous film forming foam
DoD	Department of Defense
P	Positive (In this work, AFFF)
N	Negative (In this work, non-AFFF)
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
ACC	Accuracy, (TP+TN)/(P+N), Fraction of all samples correctly identified
PPV	Positive Predictive Value (Precision), TP/(TP+FP), Fraction of all *identified* as P that are *actually* P
TPR	True Positive Rate (Recall), TP/(TP+FN), Fraction of all P that are *identified* as P
NPV	Negative Predictive Value, TN/(TN+FN), Fraction of all *identified* as N that are *actually* N
TNR	True Negative Rate, TN/(TN+FP), Fraction of all N that are *identified* as N
BA	Balanced Accuracy, (TPR+TNR)/2, Average of TPR, TNR
ANN	Artificial neural network
DNN	Deep neural network
PFPrA	Perfluoropropanoic acid
PFBA	Perfluorobutanoic acid
PFPeA	Perfluoropentanoic acid
PFHxA	Perfluorohexanoic acid
PFHpA	Perfluoroheptanoic acid
PFOA	Perfluorooctanoic acid
PFNA	Perfluorononanoic acid
PFDA	Perfluorodecanoic acid
PFUnDA	Perfluoroundecanoic acid
PFDoDA	Perfluorododecanoic acid
PFTrDA	Perfluorotridecanoic acid
PFTeDA	Perfluorotetradecanoic acid
PFHxDA	Perfluorohexadecanoic acid
PFOcDA	Perfluorooctadecanoic acid
PFBS	perfluorobutane sulfonate
PFPeS	perfluoropentane sulfonate
PFHxS	perfluorohexane sulfonate
PFHpS	perfluoroheptane sulfonate
PFOS	perfluorooctane sulfonate
PFNS	perfluorononane sulfonate
PFDS	perfluorodecane sulfonate
4:2FTS	4:2 Fluorotelomer sulfonate
6:2FTS	6:2 Fluorotelomer sulfonate

8:2FTS	8:2 Fluorotelomer sulfonate
10:2FTS	10:2 Fluorotelomer sulfonate
FOSA	Perfluorooctane sulfonamide
MeFOSA	n-Methyl perfluorooctane sulfonamide
EtFOSA	n-Ethyl perfluorooctane sulfonamide
MeFOSE	n-Methyl perfluorooctane sulfonamide ethanol
EtFOSE	n-Ethyl perfluorooctane sulfonamide ethanol
FOSAA	perfluorooctane sulfonamide acetic acid
MeFOSAA	n-Methyl perfluorooctane sulfonamide acetic acid
EtFOSAA	n-Ethyl perfluorooctane sulfonamide acetic acid
6:2FTCA	6:2 fluorotelomer carboxylic acid
8:2FTCA	8:2 fluorotelomer carboxylic acid
10:2FTCA	10:2 fluorotelomer carboxylic acid
3:3FTCA	3:3 fluorotelomer carboxylic acid
5:3FTCA	5:3 fluorotelomer carboxylic acid
7:3FTCA	7:3 fluorotelomer carboxylic acid
6:2FTUCA	6:2 fluorotelomer unsaturated carboxylic acid
8:2FTUCA	8:2 fluorotelomer unsaturated carboxylic acid
10:2FTUCA	10:2 fluorotelomer unsaturated carboxylic acid
6:2FTOH	6:2 Fluorotelomer alcohol
8:2FTOH	8:2 Fluorotelomer alcohol
10:2FTOH	10:2 Fluorotelomer alcohol
GenX	PFPrOPrA; HFPO-DA; 2,3,3,3-Tetrafluoro-2-(1,1,2,2,3,3,3-heptafluoropropoxy)-propanoic acid
PFMOAA	Perfluoro-2-methoxyacetic acid, C3HF5O3
PFO2HxA	Perfluoro-(3,5-dioxahexanoic) acid, C4HF7O4
PFMOPrA	Perfluoro-3-methoxy-propanoic acid
PFO3OA	Perfluoro-(3,5,7-trioxaoctanoic) acid
PFO4DA	Perfluoro-(3,5,7,9-tetraoxadecanoic) acid
PFO5DoDA	Perfluoro-3,5,7,9,11-pentaoxadodecanoic acid
PFMOBA	Perfluoro-4-methoxy-butanic acid
NfBP1	Nafion Byproduct 1, C7HF13SO5
NfBP2	Nafion Byproduct 2, C7H2F14SO5
ADONA	3H-Perfluoro-3-[(3-methoxy-propoxy)propanoic acid], ammonium salt
F53B	Potassium 9-chlorohexadecafluoro-3-oxanonane-1-sulfonate
F53Bmin	Potassium 11-Chloroeicosafluoro-3-oxaundecane-1-sulfonate (minor component of F53B)

KEYWORDS

PFAS; Supervised Machine Learning; Pattern Recognition; Forensics; Source Allocation

ACKNOWLEDGMENTS

This research was supported wholly by the U.S. Department of Defense, through the Strategic Environmental Research and Development Program (SERDP), project ER20-1205. This material is based upon work supported by the Humphrey's Engineer Centre Support Activity under Contract No. W912HQ-20-P-0005. Views, opinions, and/or findings contained in this report are those of the authors and should not be construed as an official Department of Defense position or decision unless so designated by other official documentation.

The investigators thank the following people for their efforts converting data for this project: Diana Ackerman Grunfeld, Jacob Shaughnessy, Yi Gao, Nicholas Wijaya, Su Zhou. The investigators also thank Aaron Button, Kenan Cetin, Pamela Cushing, Michael Ginsbach, Kasey Kathan, Jacki McManus, John Schmeltzer, Erin Schneider, and Ray Spaulding for providing data used in this work.

ABSTRACT

Introduction and Objectives. Aqueous film forming foam (AFFF) formulations based on per- and polyfluoroalkyl substances (PFAS) were used extensively at Department of Defense (DoD) sites, and contamination is a concern at hundreds of locations. Because PFAS have also been widely used in non-AFFF applications, it is important that the DoD be able to distinguish between detected PFAS contamination that originates from offsite non-AFFF sources, and contamination that originates from onsite DoD AFFF sources. This proof-of-concept project explored the use of modern machine learning algorithms to search for recognizable patterns in PFAS-containing samples to identify whether detected PFAS in environmental samples originates from AFFF sources.

Technical Approach. The project was focused around three main tasks: ***Task 1. Data collection and preprocessing.*** An extensive, machine-readable dataset was created from PFAS concentration data collected from around the world, specifically formatted to be used as an input to machine learning applications. ***Task 2. Evaluation of machine learning algorithms for source identification.*** A range of supervised and unsupervised learning activities were conducted to assess which methods exhibit the greatest performance for source identification. Work was also conducted to determine how PFAS component selection for machine learning impacts source identification. ***Task 3. Preliminary data mining and exploration of deep learning algorithms.*** Hybrid approaches integrating machine learning with an understanding of PFAS physicochemical behaviors were used to explore the relationship between compositions in different phases. Preliminary tests of deep neural networks were also conducted for source identification.

Results. An initial PFAS dataset containing more than 22,000 samples was compiled and used to explore machine learning classification. A total of 12 supervised machine learning classifiers were tested, and all were found to provide excellent classification performance. Classification is generally improved with greater numbers of components, but high performance could be achieved with as few as four components, provided they were selected such that they were present above detection limits in both the training dataset and the samples being classified. In contrast with supervised learning results, tests with unsupervised learning (clustering) found that the clustering methods tested were not able to distinguish between PFAS from different sources, but were rather more sensitive to the effects of transport on composition. Finally, work partially supported by this project was conducted to develop a group contribution model for estimation of physicochemical parameters describing interfacial and partitioning phenomena driven by hydrophobicity for individual PFAS components. The model was used with scaling relationships for this work to predict PFAS compositions in phases in equilibrium with water by transforming PFAS groundwater data. Overall, the transformed water data showed strong agreement with the distributions in non-water phases, and supervised learning classifiers trained on the transformed data exhibited improved ability to classify samples in non-water phases.

Benefits. The results of this project show that supervised machine learning has significant promise as a means of distinguishing between PFAS from AFFF and non-AFFF sources. A tool created using the methods from this proof-of-concept project could be of significant value, providing an immediate, quantitative means of assessing the origin of PFAS constituents in a

sample. However, additional work is needed to enhance the predictive capabilities and utility of the tool for use at DoD sites. Areas where additional work is needed include: Creation of an expanded dataset containing additional examples of specific source types; Selection of methods for rejection or open set recognition to handle cases of PFAS compositions not represented in the training set; Exploration of advanced training methods; Making use of the expanded dataset to allow multiclass classification; and Conducting simulations to better understand classification in complex environmental settings. Follow-up work would be designed to culminate in an user-level tool for PFAS source allocation that would provide immediate, low cost information that could guide allocation of resources by highlighting samples or areas for more detailed study, potentially providing substantial utility and cost savings to DoD.

EXECUTIVE SUMMARY

Introduction

Per- and polyfluoroalkyl substances (PFAS) have seen extensive use for decades over a large number of applications. Because of their persistence in the environment, environmental contamination by PFAS is a widespread concern. Uncovering the source of contamination can be difficult because of both their widespread use and persistence, because in many cases detected contamination could originate from multiple possible historical sources. The work of this project focused on forensic source allocation, using information in the analyzed samples to determine the origin of detected PFAS in environmental samples.

This 1-year proof-of-concept project centered around the use of modern machine learning algorithms to evaluate the likelihood that detected per- and polyfluoroalkyl substances (PFAS) in environmental samples come from aqueous film forming foam (AFFF) sources. The approach used machine learning to search for recognizable patterns in PFAS-containing samples, with the goal of determining whether contamination originates from specific sources.

The use of machine learning has exploded in recent years, and machine learning algorithms are ubiquitous in everyday applications. While machine learning algorithms can be applied to solve a range of different types of problems, the problem of identifying the origin of PFAS contamination based on the pattern of individual component concentrations in a set of samples is a ***classification*** problem. Machine learning algorithms excel at classification, and are capable of detecting complex patterns that can be virtually impossible to uncover through other means.

Objectives

The objective of this 1-year proof of concept project was to explore a range of machine learning techniques as a means of searching for patterns in detected PFAS components in sets of environmental samples, and using those patterns to assign probabilities to the likely origins of the detected PFAS.

The concept of the work is as follows: although different PFAS formulations had significantly different compositions at their time of manufacture (both across a range of different AFFF formulations, as well as across vastly different non-AFFF applications), information about original composition is not sufficient to identify the origin of PFAS contamination, because environmental mobility and transformation differs for the various PFAS components, so composition will differ with both age and distance (**Figure ES-1**). **The hypothesis driving this work** was that the environmental behavior of different formulations may create recognizable patterns in component distributions that can be identified at sites around the world. That is, the transport and transformation that occurs in environmental media for specific AFFF and non-AFFF PFAS formulations over time will show identifiable patterns that can be distinguished despite different site characteristics.

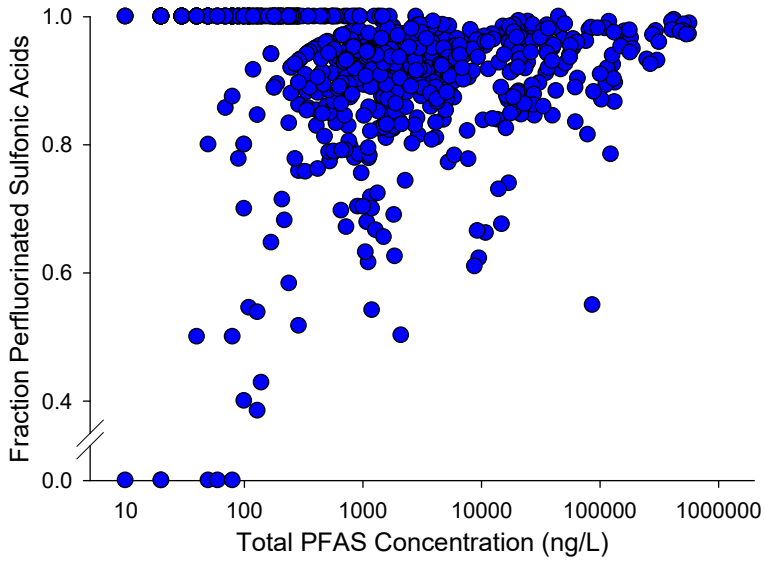


Figure ES-1. Fraction perfluorinated sulfonates vs. total PFAS concentration in 759 groundwater and surface water samples from a site where contamination is known to be primarily from PFOS-based AFFF.This figure illustrates how widely PFAS composition can vary at a site, even when a relatively narrow range of source formulations were used, a major challenge for source allocation.

Technical Approach

The project centered around the use of existing data from PFAS sites around the world as inputs to train machine learning algorithms to distinguish PFAS contaminants from different origins. Open source machine learning libraries were used to ensure that the results are accessible to practitioners and could be readily applied at SERDP sites beyond those considered in this work. The work was designed around three main tasks:

Task 1. Data collection and preprocessing. Machine learning algorithms require significant quantities of data for training; the greater the quantity of data and the more varied the data are, the higher the likelihood the algorithms will accurately identify underlying patterns and make accurate predictions. This task centered around collection of data from diverse PFAS sources worldwide, and then placing them into a custom format specifically designed to be used as input to machine learning analyses.

Task 2. Evaluation of machine learning algorithms for source identification. This task focused on the use of conventional machine learning classifiers as a means of source identification for PFAS sites. The task focused primarily on supervised learning (classification) techniques (**Subtask 2.a**), although selected unsupervised learning (clustering) analyses were also explored (**Subtask 2.c**). A part of the work also involved an examination of the roles of individual PFAS components in the classification accuracy, for purposes of feature selection (**Subtask 2.b**).

Task 3. Preliminary data mining and exploration of deep learning algorithms. This task centered on two areas to guide future research. **Subtask 3.a** focused on preliminary data mining aimed at uncovering information about physicochemical phenomena (transformations, adsorption) impacting PFAS compounds, while **Subtask 3.b** conducted a preliminary exploration of the use of deep learning for classification.

Results and Discussion

Data Collection and Preprocessing

In order to test the ability of supervised machine learning classifiers to distinguish between PFAS from AFFF and non-AFFF sources, it was first necessary to assemble a dataset. The target for Task 1 was to assemble a dataset containing more than ten thousand individual samples. Ultimately, we assembled a dataset containing data from **28,142 samples total, 22,068 of them with at least one detected PFAS compound**. (Note that although samples with no detected compounds contain no information about PFAS source origin, some were retained, particularly in cases where spatial information were available, for possible future site-specific spatial analyses such as plume delineation.) This section describes the contents and structure of the dataset.

Table ES-1 shows the sample breakdown of the dataset by phase (water, other) and classification (AFFF, non-AFFF, Unknown):

Table ES-1. Overview of sample classification in the dataset. Numbers shown correspond to numbers of individual samples collected.

Classification	Water	Other	Total
AFFF	7452	2762	10,214
non-AFFF	9288	678	9966
Unknown	1666	222	1888
Total	18,406	3662	**22,068**

The original objective was to collect at least 10,000 water samples for the dataset, but it became apparent early in the project that many of the data sources used also contained non-water sample data that could be useful for data mining and other exploratory project calculations, and there were substantial efficiencies to converting those data at the same time as the water data, so non-water sample data were compiled for selected sites as well. Non-water data include data for **soils**, **sediment**, **biota**, **sludge**, as well as other phases. **AFFF vs. non-AFFF** samples were classified based on available site information, and were scored on a certainty scale indicating the level of confidence in the classification. **Unknown** samples are those for which the source has not yet been identified. In some cases, the sources are likely unknowable, while in others classification is likely possible after future detailed study of available site and geographical information. Note that all duplicate samples (same location, same day) have been removed from the dataset, although samples from the same location from different sampling events have been retained. Overall, known **water samples** in the dataset are **44.5% from AFFF sources**, while known **samples in the entire dataset** are **50.6% from AFFF sources**. The reason for the difference is a significantly greater availability of data from non-water samples in available data for AFFF data sources.

As noted previously, data in the dataset were collected from sites around the world. **Table ES-2** shows the breakdown of samples by country and classification.

Table ES-2. Contents of the dataset by country and classification.

Country	AFFF	non-AFFF	Unknown	Total
Australia	6333	110	0	6443
Canada	2533	11	0	2544
China	0	163	263	426
Germany	0	20	63	83
Spain	0	6	0	6
Ireland	0	48	0	48
Malta	0	0	71	71
Sweden	208	119	124	451
United States	1140	9489	1367	11,996
Total	10,214	9966	1888	**22,068**

Note that the country-to country differences in part reflect the different data transparency and availability in the various countries, as well as ease with which data could be collected and converted rapidly for this 1-year proof of concept project. Substantial quantities of data across a wide range of sources remain available internationally for future collection.

Data collected from this work come largely from public sources, although some data have been provided through personal communication with specific agencies or investigators. With some exceptions, most of the data collected come from extracting PDF files or other publication data. A file-based data structure is used for this work. The format was developed specifically for this work, with the objective of putting the data into a format that can easily be read for machine learning applications. Each sub-site consists of two files, one containing header information, and a second containing data; the dataset is currently contained in 324 total files. Full details of the contents and structure of the dataset are provided in the main report.

Supervised Machine Learning for PFAS Classification

This section discusses work conducted for Subtask 2a. (Supervised learning – conventional classifiers), Subtask 3b. (Preliminary deep learning), and Subtask 2b. (Feature selection). Note that the deep learning work described in Subtask 3b was conducted in concert with the conventional classification work in Subtask 2a, so is described in this section.

Early in the project, a small 1197-sample subset of the dataset was used to conduct preliminary supervised machine learning. The results of that work are described in detail in the paper:

> **T. Kibbey, R. Jabrzemski and D. O'Carroll. Supervised Machine Learning for Source Allocation of Per- and Polyfluoroalkyl Substances (PFAS) in Environmental Samples, 2020, *Chemosphere*, 252, 126593**. [1]

The results of that early work supported the hypothesis of the work, suggesting that supervised machine learning is capable of distinguishing between AFFF and non-AFFF samples based on composition with high accuracy. This section provides an updated analysis of supervised

machine learning based on the final data set described in the preceding section; full details of the work are provided in the main report.

Dataset, Features and Training. Machine learning classifiers require that all samples used for training and testing have the same features (in this case, PFAS components). For this work, a **subset of 10 PFAS components was selected**: the perfluorinated carboxylic acids **PFHxA, PFHpA, PFOA, PFNA, PFDA, PFUnDA and PFDoDA**, and the perfluorinated sulfonates **PFBS, PFHxS and PFOS**. Selection of these samples allowed a dataset containing **8105 groundwater (GW) and surface water (SW) samples** to be extracted from the full dataset. The water sample data were divided evenly into separate test and training sets. **Classifiers were trained on the 4052 sample training set, and then evaluated on the *completely separate* 4053 sample test set.**

Source Allocation Results. Before discussing the source allocation results, it is necessary to provide definitions that are needed to understand the performance measures used to evaluate results. In the context of this work, which involves binary classification (i.e., classification into two classes), it is necessary to define a POSITIVE class and a NEGATIVE class. Here we define **POSITIVE** as **AFFF**, and **NEGATIVE** as **non-AFFF**. (The selection of this convention has no impact on the outcome of classification, but is necessary for interpreting binary classification.) To describe the outcomes of machine learning classification, it is typical to consider the four possible conditions for a classification result: **True Positive (TP)**, **True Negative (TN)**, **False Positive (FP)** and **False Negative (FN)**. These terms and their meanings are summarized in **Table ES-3**:

Table ES-3. Possible outcomes in the binary classification of AFFF/non-AFFF samples.

Condition	Description	Classified as:	Actually is:
TP	**AFFF** identified correctly as **AFFF**	AFFF	AFFF
TN	**non-AFFF** identified correctly as **non-AFFF**	non-AFFF	non-AFFF
FP	**non-AFFF** identified INCORRECTLY as **AFFF**	AFFF	non-AFFF
FN	**AFFF** identified INCORRECTLY as **non-AFFF**	non-AFFF	AFFF

Classification for the work examined a total of **12 different supervised learning classifiers**. **Table ES-4** shows the results for a **subset of 4 classifiers**, based on the common logs of component distributions, **the transformation that was found to produce the best classification performance**.

Also shown in the tables is the result of **naïve classification by ratio of perfluorinated sulfonates to total PFAS**. Because legacy PFOS-based AFFF formulations were dominated by perfluorinated sulfonates, a naïve first-cut identification is sometimes made by looking at this ratio. I.e., samples for which perfluorinated sulfinates are present in greater quantities than perfluorinated carboxylic acids are naïvely classified as being AFFF. However, this approach is unsatisfactory for the many different cases, including non-PFOS-based AFFF formulations (where perfluorinated sulfonates are absent, or are present at small concentrations), and for PFOS-based formulations far from their source (where more mobile non-sulfonated components

can dominate distributions). Furthermore, many examples of non-AFFF formulations based on perfluorinated sulfinates exist.

The performance metrics in **Table ES-4** include: Accuracy (ACC, the fraction of all samples correctly identified), Positive Predictive Value (PPV or Precision, the fraction of all samples identified as AFFF that are actually AFFF), True Positive Rate (TPR or Recall, the fraction of all AFFF that are identified as AFFF), the Negative Predictive Value (NPV, the fraction of all samples identified as non-AFFF that are actually non-AFFF), the True Negative Rate (TNR or Selectivity, the fraction of all non-AFFF that are identified as non-AFFF), and Balanced Accuracy (BA, the average of TPR and TNR). Multiple metrics are necessary because different metrics highlight different strengths and weaknesses in classification, particularly when the test set contains different numbers of AFFF and non-AFFF samples.

Table ES-4. Logarithmic, Distribution: Performance metrics for classification of the 4053-sample test set, shown as percent for 4 classifiers. Shading and font color indicates performance. Classification by ratio is shown for comparison. Results for all 12 tested classifiers are provided in the full report.

Classifier	Accuracy ACC	Pos. Pred. Value PPV (Precision)	True Pos. Rate TPR (Recall)	Neg. Pred. Value NPV	True Neg. Rate TNR (Selectivity)	Balanced Accuracy BA
Random Forest	**96.1**	**96.7**	**98.1**	94.6	90.8	94.5
K-neighbors	94.8	**96.0**	**97.0**	91.5	88.9	92.9
Support Vector (RBF)	94.1	**96.2**	**95.7**	88.4	89.7	92.7
DNN Ensemble	**95.5**	**95.8**	**98.2**	94.7	88.3	93.3
Classification by Ratio	79.8	87.5	84.5	61.4	67.1	75.8

In general, the work found that **all 12 tested classifiers performed better than naïve classification by ratio**, with the **highest performance observed by variants of the Random Forest method, followed closely by a deep neural network ensemble**. Classification of distributions provided higher performance than classification of concentrations on both a linear and logarithmic scale, and classification on a logarithmic scale was slightly better than classification on a linear scale.

Figure ES-2 shows a **graphical representations of the confusion matrix** for the best performing classifier for this dataset, the **Random Forest** classifier. Note that the analysis shown is based on a logarithmic distribution of contaminant concentrations, as that was found to provide the highest classification performance of the preprocessing methods tested. It is apparent from the figure that the method is able to correctly identify the vast majority of samples in the test set, despite the fact that AFFF and non-AFFF samples have a wide range of compositions in the environment.

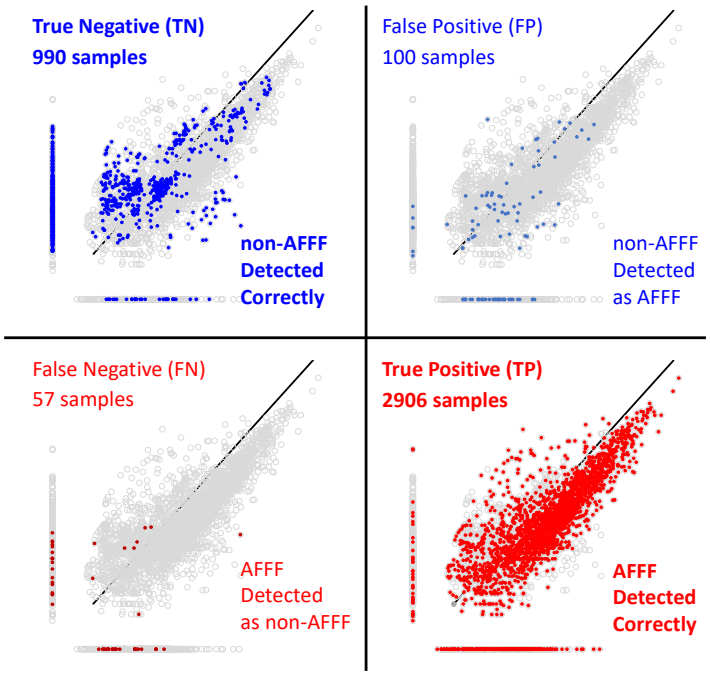


Figure ES-2. Graphical representation of the confusion matrix for predictions conducted using the **Random Forest classifier** for a 4053 sample test data set. The classifier was trained on a *completely separate* 4052 sample training set. Plots show total log concentrations of **perfluorinated sulfonates on the horizontal axis**, and **perfluorinated carboxylic acids on the vertical axis**; zero values cannot be displayed on a log scale, so are shown at the bottom and left of each plot. The line shown in each subfigure is a 1:1 line.

Difficult Subsets. Some categories of PFAS samples are easy to classify, while others can be particularly challenging because they mimic the other class. **Table ES-5** illustrates classification performance for three subsets of the dataset, samples from military bases, samples associated with sites where non-PFOS AFFF is known to be present, and samples associated with two sites where PFOS-based non-AFFF formulations are present. Note that the table shows accuracy. Because each subset is only AFFF or non-AFFF, accuracy provides a complete description of the performance of the classifiers.

Table ES-5. Classification accuracy for three subsets of the dataset.

	Accuracy, % (number correct/total in subset)		
	ALL Military Bases (AFFF)	**Portland, Clarendon, Marinette, Brisbane (likely non-PFOS AFFF)**	**Wolverine, Du-Wel (PFOS-based non-AFFF)**
Random Forest	**98.4** (2247/2284)	**96.2** (409/425)	**70.1** (157/224)
Deep Neural Network	**99.0** (2261/2284)	**92.9** (395/425)	**70.5** (158/224)
Classification by Ratio	**93.7** (2139/2284)	**26.1** (111/425)	**8.0** (18/224)

From **Table ES-5**, it is apparent that both machine learning methods do very well at classification of PFAS on military bases. In part, this is likely due to the large numbers of samples from military bases in the training set. However, the larger reason is that the patterns in that subset of samples are largely distinct from other subsets in the dataset. Note that classification by ratio also works well for this subset, primarily because despite variability in

composition, the majority of samples in this subset contain higher total concentrations of perfluorinated sulfonates than perfluorinated carboxylic acids. Nevertheless, both machine learning classifiers correctly identify many of the samples for which this is not true, providing accuracy that exceeds classification by ratio in both cases.

While classification performance on military sites is especially high, note that **source allocation and plume delineation depend on the ability to *distinguish* sources**; as such, it is **essential that more-difficult subsets of PFAS can be identified** so that they may be distinguished from AFFF.

Although contamination at many military sites is largely dominated by PFOS-based legacy AFFF, there are other **sites where there is a known history of non-PFOS-based AFFF contamination, or where low concentration off-site AFFF samples exhibit patterns that mimic non-AFFF sources**. (Portland, Clarendon and Brisbane Airports, and contamination associated with non-PFOS-based AFFF manufacturing and testing at Marinette, WI). From **Table ES-5**, it is apparent that **both classifiers identify data in this subset with high accuracy**, with the Random Forest method performing slightly better between the two. In contrast to the military bases, naïve classification by ratio performs poorly for this subset, entirely because nearly 74% of the samples have higher concentrations of perfluorinated carboxylic acids than perfluorinated sulfonates; as such, classification by ratio gives only a **26.1%** accuracy.

One of the most difficult subsets of the dataset for the classifiers to identify is the subset of non-AFFF samples that are dominated by PFOS and other perfluorinated sulfonates (e.g., Wolverine, Du-Wel). The performance of the classifiers on this subset is lower than other subsets (both are just over 70% accurate). However, in comparison to naïve classification by ratio (***8%* accuracy**), the performance of the classifiers is nevertheless a huge improvement.

Clustering

Clustering algorithms, also known as unsupervised learning algorithms, look for patterns in unlabeled data, and as such are often proposed for source delineation. The objective of this subtask was to conduct a preliminary exploration of the use of clustering as an alternate means of distinguishing between AFFF and non-AFFF samples. In general, **the clustering methods tested were not able to distinguish between PFAS from different sources, but were rather more sensitive to the effects of transport on composition**. Although it is likely there are some applications of clustering for source delineation, particularly for cases where very different source compositions are present, extreme care must be exercised interpreting the results. Based on the preliminary work conducted here, it is unlikely that clustering (unsupervised learning) would be satisfactory for more general PFAS source allocation.

Preliminary Data Mining

The work in this subtask focused on exploring the extent to which information about physicochemical phenomena (transformation, adsorption) could be uncovered from the dataset and machine learning results. The main emphasis of the work conducted was to use a hybrid method, combining machine learning with transformations based on a quantitative understanding of the behaviors of surfactants such as PFAS, to relate PFAS compositions in different phases.

Prediction of PFAS Surfactant Parameters Driven by Hydrophobicity. Work partially supported by this project was conducted to develop a group contribution model for estimation of individual PFAS physicochemical parameters describing interfacial and partitioning phenomena driven by hydrophobicity. The work was recently accepted for publication in the paper:

> **S.T. Le, T. C. G. Kibbey, K. P. Weber, W. C. Glamore, and D. M. O'Carroll. A Group-Contribution Model for Predicting the Physicochemical Behavior of PFAS Components for Understanding Environmental Fate, *Science of the Total Environment*, 142882. [2]**

Key components of that work as they relate to this project are described in the full report. In short, the group contribution model described in the paper uses linear combinations of constants corresponding to different functional groups in individual PFAS molecules to fit measured interfacial constants for air-water interfacial adsorption, and then normalized scaling relationships are used to predict other hydrophobicity-driven physicochemical parameters. The calculated constants provide **a tool for transforming water PFAS data to predict PFAS component compositions that would be expected for other phases if in equilibrium with water** (e.g., soil, sediment, biota). **The transformed data are then used to train a supervised learning machine classifier that can be used to classify samples from other phases.**

Figure ES-3 illustrates the transformation for one site, showing how the composition of the transformed groundwater looks very similar to compositions in both soil and biota from the same site. As described in the full report, **classifications based on the transformed data show improvements for soil, sediment and biota** compared with classification using a classifier trained on untransformed water data.

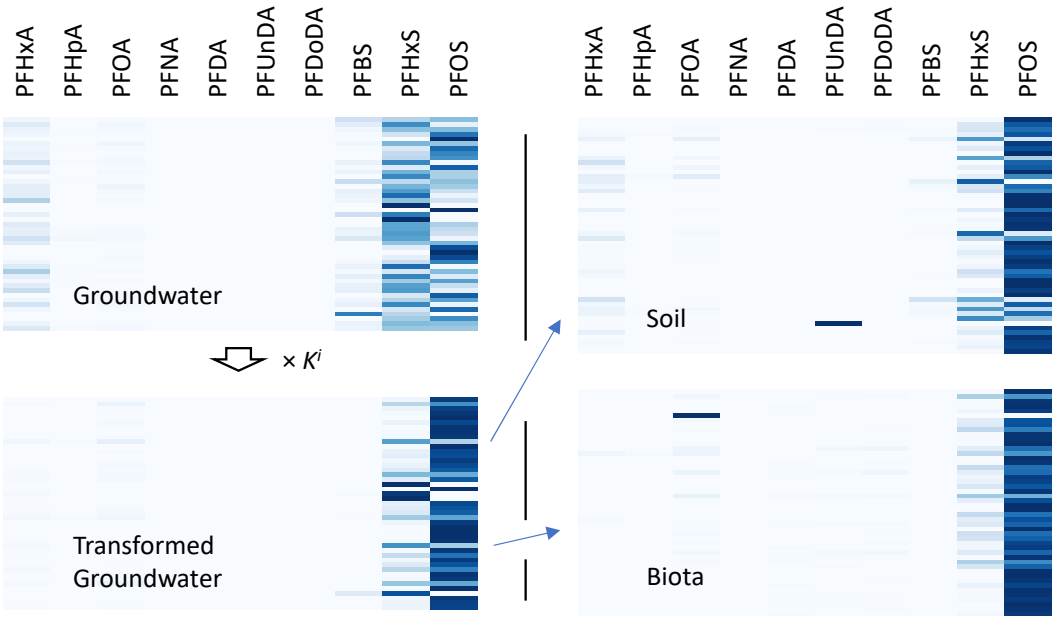


Figure ES-3. Comparison between transformed groundwater composition data and data for other phases for RAAF Base Pearce (Australia, AFFF). Darker bars correspond to a higher fraction present in each sample. Each row corresponds to an individual sample.

Implications for Future Research and Benefits

A tool created directly using the methods from this proof-of-concept project could be of significant value, providing an **immediate, quantitative means of assessing the origin of PFAS constituents in a sample**. However, additional work in the following areas is needed to enhance the forensic capabilities and utility of the tool for use at DoD sites.

1. Creation of an Expanded Dataset. The dataset created for this proof-of-concept project was a preliminary dataset, designed primarily to test the concept of supervised machine learning for classification of PFAS in environmental samples. Specific needs in an expanded dataset include the addition of more industry-specific non-AFFF PFAS examples, as well as more AFFF examples from the United States. These additions are important for increased classification certainty, particularly with difficult sample types.

2. Selection of Methods for Rejection/Open Set Recognition. A final tool needs to be able to identify cases where patterns in a new sample are unlike those in the training set. This could either take the form of rejection if the pattern is not a sufficient match, or open set recognition where new unknown samples are placed into their own class. Work in this area is essential to prevent non-AFFF samples with novel compositions from being mistaken for AFFF samples.

3. Exploration of Advanced Training Methods. The ability of a classifier to identify patterns in data depend in part on the use of a balanced training set. This extends to not just AFFF vs non-AFFF samples, but actual expected subsets within both categories and their environmental patterns. Strategies to do this might include the use of synthetic training data, or ensemble predictions trained on subsets of the dataset.

4. Multiclass Classification. With an expanded dataset containing additional examples of industry-specific PFAS formulations, multiclass classification (i.e., identification of specific subsets, rather than simply AFFF vs non-AFFF) should start to become practical. Only minimal changes to existing coding would be needed to conduct multiclass classification, but the ability to identify specific industrial sources for non-AFFF samples, as well as categories of AFFF samples, could further enhance the value of a tool for source allocation.

5. Simulations to Better Understand Classification Performance. Extensive transport simulation realizations coupled with supervised learning are needed to better understand how mixing of PFAS from diverse sources impacts classification, as well as to quantify the extent to which very similar sources can be distinguished in complex environmental settings.

6. Creation of the Final Tool. The final tool to be developed for the work would be created in consultation with DoD personnel and practitioners, and would be designed to provide simultaneous output from multiple classification approaches to give a full picture of the likelihood that a tested sample is of AFFF origin. A major part of the work would be designing methods of data entry that are well-suited to site applications. It is likely the tool would be web-based, to allow machine learning computations to be conducted in the cloud, and to provide a mechanism for the tool and underlying datasets to be updated as new information becomes available.

1. OBJECTIVES

This 1-year proof-of-concept project was designed to explore the use of modern **machine learning algorithms** to evaluate the probability that detected per- and polyfluoroalkyl substances (PFAS) in environmental samples come from aqueous film forming foam (AFFF) sources. The approach used machine learning to **search for recognizable patterns in PFAS-containing samples**, with the objective of **predicting whether contamination originates from specific sources**.

The use of machine learning has exploded in recent years, and machine learning algorithms are ubiquitous in everyday applications. While machine learning algorithms can be applied to solve a range of different types of problems, the problem of identifying the origin of PFAS contamination based on the pattern of individual component concentrations in a set of samples is a ***classification*** problem. Machine learning algorithms excel at classification, and are capable of detecting complex patterns that can be virtually impossible to uncover through other means [3-5].

One of the most powerful subsets of machine learning is known as deep learning [6]. Deep learning is the name for machine learning based on deep neural networks, a subset of artificial neural networks. Artificial neural networks function in a way that is patterned after networks of biological neurons in the brain. Deep neural networks are artificial neural networks with two or more hidden layers (layers of artificial neurons between the input and output layers). Although deep neural networks have been understood for some time, robust approaches to training them have only really been available for the past two to three decades [7]; computational power to apply them routinely in day-to-day applications has come more recently than that. Most of the greatest advances in machine learning in the past decade have been in the area of deep learning.

The objective of this 1-year proof of concept project was to explore a range of machine learning techniques as a means of searching for patterns in detected PFAS components in sets of environmental samples, and using those patterns to assign probabilities to the likely origins of the detected PFAS. The concept of the work is as follows: although different PFAS formulations had significantly different compositions at their time of manufacture (both across a range of different AFFF formulations, as well as across vastly different non-AFFF applications), information about original composition is not sufficient to identify the origin of PFAS contamination, because environmental mobility and transformation differs for the various PFAS components, so composition will differ with both age and distance. **The hypothesis driving this work** is that the environmental behavior of different formulations may create recognizable patterns in component distributions that can be identified at sites around the world. That is, the transport and transformation that occurs in environmental media for specific AFFF and non-AFFF PFAS formulations over time will show identifiable patterns that can be distinguished despite different site characteristics.

1.A. Criteria for Success, and Risk Reduction for Future Work.

Successful completion of this proof-of-concept project would involve **identification of one or more machine learning algorithms capable of distinguishing between AFFF and**

non-AFFF sources with high accuracy. Of particular interest are algorithms capable of distinguishing between AFFF and non-AFFF PFAS from sources that have similar compositions.

The outcomes of the work should provide **substantial risk reduction for subsequent work** by providing information about **the most successful machine learning techniques for source allocation**, and providing **quantitative insights into the level of performance that could be expected for source allocation** for different machine learning approaches. The results of the work highlight areas of needed future research and provide a roadmap for development of a tool for machine learning source allocation.

1.B. Relationship to SERDP SON.

This project was funded under the SERDP Statement of Need **ERSON-20-C5:** ***Forensic Methods for Source Tracking and Allocation of Per- and Polyfluoroalkyl Substances*****.** The proposed work was designed to primarily address the area of need:

- **Evaluation of conventional or novel analytical techniques or methodologies to differentiate PFAS from aqueous film forming foam (AFFF) versus non-AFFF sources.**

  The **primary focus of the work** was centered around addressing this area of need through exploration of machine learning classification methods for source allocation. An example of the type of scenario that could be addressed by this work is shown in **Figure 1**. In the figure, an offsite source releases PFAS compounds that are detected in water samples both on site, as well as at a down-gradient receptor. By providing an **immediate,**

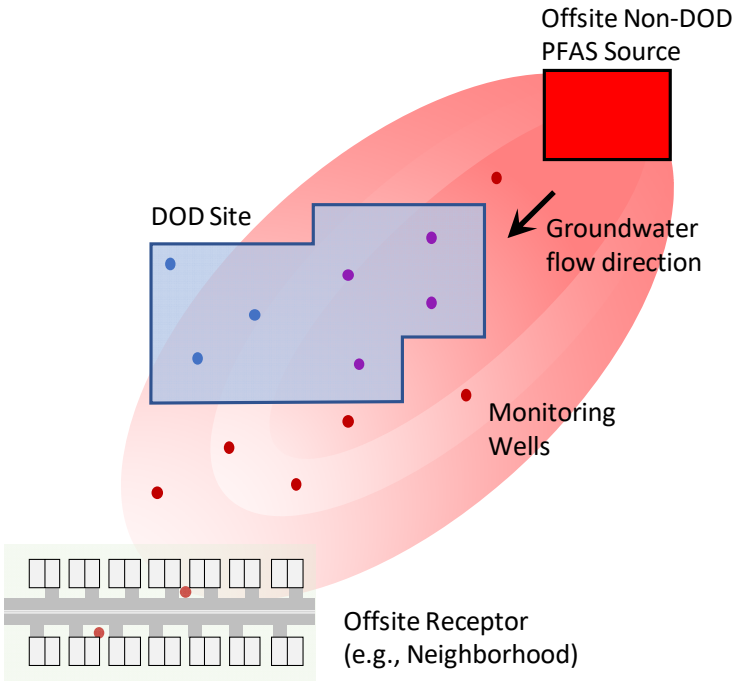


Figure 1. Conceptual illustration of the problem addressed by this work. In this illustration, an uncontained upgradient non-DOD PFAS source zone contaminates groundwater that ultimately reaches a downgradient receptor. This work is designed to use **machine learning** to identify the origin of PFAS components in individual PFAS samples.

quantitative means of assessing the likelihood that the PFAS constituents in a sample come from the offsite source, the results could be complementary to other more costly and labor-intensive techniques involving analytical methods. **A tool resulting from this work could guide allocation of resources by highlighting samples or areas for more detailed study**, providing substantial potential utility and cost savings to DoD.

A secondary objective of the work was to conduct preliminary exploratory work related to the area of need:

- **Improved analytical methods and/or validated models to predict changes to AFFF mixtures over time, including chemical pathways to the most toxic compounds.**

  Preliminary data mining activities were proposed to explore the use of data collected for classification as an input to better understand how physicochemical phenomena impact PFAS behaviors in the environment and their classification.

2. BACKGROUND

2.A. AFFF Formulations in the Environment

AFFF formulations have contained dozens, and potentially hundreds, of individual PFAS constituents with formulations varying with manufacturer as well as production year [8, 9]. For example, one study identified at least 103 individual PFAS with chain lengths ranging between C_3 and C_{15} in 10 different AFFF formulations with several unidentified PFAS constituents [10]. PFAS for AFFF applications has been mainly synthesized by two distinct synthesis processes: electrochemical fluorination and telomerization [11]. These synthesis processes result in a wide range of products including poorly characterized PFAS constituents (precursors) that can be transformed in the environment into chemicals of regulatory concern (e.g., PFOA and PFOS). For example Munoz, Labadie [12] and Anderson, Long [13] suggested the elevated concentrations of PFHxA, PFOS and PFHxS in groundwater was due to precursor transformation. Weber, Barber [14] indicated that PFBA, PFPeA and PFHxA were predominant products of the precursors 6:2 FTS and 8:2 FTS upon chemical oxidation using the total oxidative precursor oxidation assay.

While the fact that a large number of different PFAS formulations that have been used over time poses a significant challenge to source allocation based on composition, perhaps an even greater challenge comes from the fact that compositions in the environment often vary widely, even when they originate from the same formulation. Reasons for this include both differences in environmental mobility between individual PFAS compounds, as well as, in some cases, transformations of precursor compounds present in the original formulation to PFAS of regulatory concern (e.g., [9, 15]).

Figure 2 illustrates the magnitude of this problem, showing the fraction of perfluorinated sulfonic acids detected as a function of total concentration in 759 water samples at a site known to have been primarily contaminated by legacy PFOS-based AFFF. The site is the Williamtown RAAF Base in NSW Australia (one of more than 65 individual data sources included in the

dataset compiled for this work). Historical records at the site show more than 2 million gallons of legacy PFOS-based 3M Lightwater AFFF released at the site, compared with only 1800 gallons of a more modern non-PFOS-based AFFF. As such, the contamination at the site is more than 99.9% from legacy PFOS-based AFFF. (Note that this is a common pattern at many of the military sites studied, as the switch to non-PFOS-based AFFF also corresponded to a reduction in use and increased controls.) **Figure 2** clearly illustrates why knowing an initial source formulation of PFAS contamination is unlikely to be sufficient for forensic allocation, because a single source can give rise to a range of different component distributions as PFAS components are transported at different rates through the subsurface. This can lead to a situation where samples farthest from the source, which tend to have the lowest total concentrations, are also enriched in the most mobile components. In general, the most commonly-encountered perfluorinated carboxylic acids (e.g., PFOA) are more mobile than the most commonly encountered perfluorinated sulfonates (e.g., PFOS), so are often present in higher fractions in low concentration samples. This leads to a situation where the samples taken farthest from a source look the least like those close to the source (e.g., [15]), a serious problem if trying to use composition to identify a contamination source. (As noted previously, other mechanisms can also contribute to the variability of composition, including transformation of some precursor compounds to PFAS compounds of regulatory concern (e.g., [9, 15]).

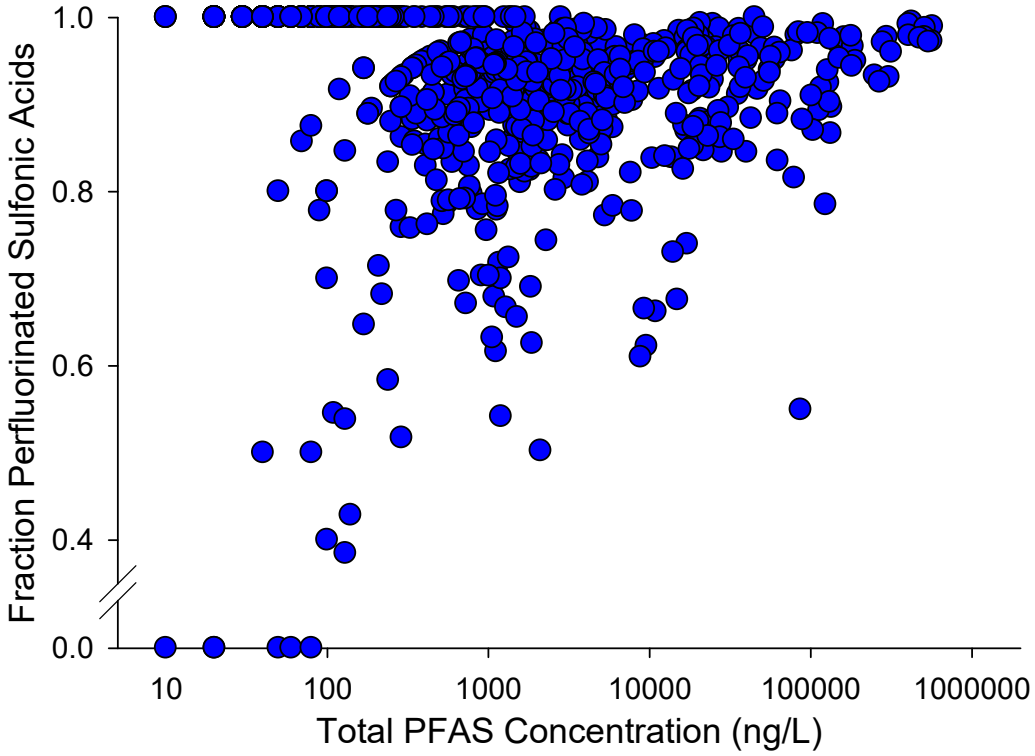


Figure 2. Fraction perfluorinated sulfonates vs. total PFAS concentration in 759 groundwater and surface water samples from a site where contamination is known to be primarily from PFOS-based AFFF.This figure illustrates how widely PFAS composition can vary at a site, even when a relatively narrow range of source formulations were used, a major challenge for source allocation.

2.B. Machine Learning

The term **machine learning** is used to describe a class of algorithms that are designed to allow computers to learn from data, and then use what they have learned to solve practical problems [3, 5]. A number of characterizations can be used to describe different machine learning applications. One major distinction is between **regression** and **classification** [5]. When machine learning is used for **regression**, it is used to predict a value based on a set of inputs. Although conceptually similar to regression using preselected equations (e.g., linear regression), many different machine learning approaches exist for making predictions based on training data, including a large number that are not tied to any specific underlying model. When machine learning is used for **classification**, it is used to place unknown data into categories based on features of the data. **Pattern recognition**, such as that to be used for this work, is a form of classification [3]. Machine learning classification algorithms can be described as **supervised** or **unsupervised** [3, 5]. In **unsupervised** machine learning applications, data used to train the model is unlabeled, and the algorithm searches for distinctions in the data to allow it to create unlabeled categories to put individual data points in; this approach is known as **clustering**. In **supervised** learning, **labels** are applied to all training data, and then test data are evaluated using the trained model to determine whether they match categories described by the labels. **This work primarily focuses on supervised learning for classification, but also conducts selected clustering analyses.**

Within the area of classification, algorithms can be designed to identify whether or not data fall into one of two classes (**binary classification**), or to determine which of several classes data fall into (**multiclass classification**) [3]. In the context of identifying AFFF sources in environmental samples, either approach could be valuable. Binary classification could be used to distinguish between AFFF and non-AFFF sources, while multiclass classification could be used to identify samples originating from a range of different specific sources (e.g., different AFFF formulations, specific non-AFFF formulations.)

Ensemble methods are often used to improve the performance of machine learning applications [5, 16]. A wide array of different individual machine learning algorithms exist, based on different underlying architectures. Some may be better for a specific application than others, and predictions made with one algorithm may differ from those made with another. One method of improving performance of methods, as well as increasing robustness and resistance to overfitting data is to use **ensemble methods**. In ensemble methods, multiple analyses are conducted, either with completely different algorithms, or with the same algorithm with different subsets of data used for training. The consensus of the methods is used to determine the classification based on either hard voting (the most common classification among the methods is chosen) or soft voting (the highest net probability is chosen). One commonly-used and extremely powerful ensemble method is the **random forest** method, which creates multiple decision trees using bootstrap aggregating (bagging) of subsets of the training data.

An important part of any machine learning application is determining which features (dimensions) to include and which to omit. If too many features are included in a model it can lead to misclassification, particularly if some of them are irrelevant to the application. **Feature selection** and **dimensionality reduction** are related concepts useful to determining which features are most critical to accurate classification.

Data mining is the process of using trained machine learning algorithms to uncover underlying information about the systems being studied [5]. Machine learning algorithms are able to classify data based on patterns they recognize through training. For many machine learning algorithms, it is then possible to go back after they are trained and determine what it is about the original data set that causes the algorithm to make its decisions. This powerful technique can allow users to discover fundamental physical phenomena that wouldn't be recognizable without machine learning. Among other uses, data mining can provide new hypotheses about a wide array of physical behavior that can subsequently be tested through experiments.

Artificial neural networks (**ANNs**) are a subset of machine learning algorithms based on networks of nodes whose mathematical behavior is designed to mimic biological neurons [6]. The most powerful ANNs are **deep neural networks (DNNs)**, a category of neural networks with two or more hidden layers. Deep neural networks form the basis of many of the most sophisticated commercial machine learning applications, and are often observed to exhibit greater performance than other methods. However, a complication with DNNs is that they can be difficult, if not impossible to reverse engineer, so are poorly suited to data mining [17]. This problem has hindered their use in applications such as for medical diagnoses, because the algorithm can make a prediction, but can't explain what caused the prediction to be made. In the context of this work, preliminary exploration of DNNs will be conducted to assess their performance relative to that of conventional machine learning algorithms. Although their opacity makes them less suitable the purpose of identifying AFFF in samples, DNNs may still have value as a second opinion, particularly if it can be shown that their predictive performance is substantially greater than that of conventional methods.

2.C. Training and Assessing Machine Learning Algorithms

Training a machine learning method involves providing a training data set which is used to automatically adjust internal method parameters. A typical approach is to take the entire available data set, and then split it into a **training data set** and a **test data set** [3, 5, 18]. Often something on the order of 20% - 50% of the data set will be set aside at random as a test data set. Methods such as **stratified sampling** can be used to ensure that the statistical properties of the two sets are equivalent. Once the test data set is created, it is not used in calibration of the model, but is only used at the end to test the model performance. Models can be tuned by using **cross validation** with multiple subsets of the training set to evaluate performance against remaining training set values and to adjust hyperparameters as needed [5].

A number of measures of performance can be used to assess how well a classification model is performing, both during cross validation tuning, and then in predictions with a test set. For a binary classifier, two common measures are the **precision** and the **recall** [5, 16]. **Precision** is defined as the ratio of the number of true positives (correct positive identification) to the sum of the number of true positives and true negatives (correct negative identification). **Recall** is defined as the ratio of the number of true positives (correct positive identification) to the sum of the number of true positives and false negatives (incorrect negative identification). Note that precision and recall cannot both be independently improved for a given model; tuning a model involves optimization of the balance between the two. Some applications may require higher

precision while others may require higher recall, depending on the significance of false positives vs. false negatives in the application. A more detailed way of looking at the performance of a model is to look at the **confusion matrix**, which is a tabular representation of the number of detections in each of the four states for a binary classifier (true positive (TP), true negative (TN), false positive (FP), false negative (FN)) [5]:

Confusion Matrix for a Binary Classifier:

TN	FP
FN	**TP**

The values on the diagonal – true negatives (TN) and true positives (TP) – are the desirable outcomes, in that they indicate the model has correctly predicted the states of the items being classified. Items off the diagonal are undesirable, and indicate misclassification by the model. (Note that confusion matrices can also be generated for multiclass classification problems.)

3. MATERIALS AND METHODS

3.A. Approach

The project centered around the use of existing data from PFAS sites around the world as inputs to train machine learning algorithms to distinguish PFAS contaminants from different origins. Open source machine learning libraries were used to ensure that the results are accessible to practitioners and could be readily applied at SERDP sites beyond those considered in this work. The work was designed around three main tasks:

Task 1. Data collection and preprocessing. Machine learning algorithms require significant quantities of data for training; the greater the quantity of data and the more varied the data are, the higher the likelihood the algorithms will accurately identify underlying patterns and make accurate predictions [3]. This task centered around collection of data from diverse PFAS sources worldwide. Note that the fact that the same sets of PFAS formulations have been used internationally means that PFAS data can provide a useful input to the algorithms, regardless of the country of origin, a fact that increases the available training data for this work. **The original objective of the task** was to assemble **more than ten thousand individual sample data points** over the duration of this limited-scope one-year project.

As an input to **Tasks 2** and **3**, all collected data from this task are placed into a consistent, machine readable format developed specifically for this work, converted to consistent units between sites, and tabulated using consistent naming conventions for all components and parameters. Because different data sets have significant differences in the number of components analyzed, all available components at each site are included in data sets, and then the machine learning component of the work (**Tasks 2 and 3**) will be used to decide which components to include in analyses.

Data are saved in formats designed to be read using the open source pandas library, ensuring they remain accessible into the future. Data from each site are stored in text-based files designed to be read by python, with the major parameter data files specifically designed to be read as pandas dataframes. Rather than create a single large database file to hold all collected

data, individual file sets will be created for each specific data source. This approach provides the greatest flexibility for the machine learning component of the work, and also yields a more compact dataset, because it won't be necessary to include blank spaces for parameters or components not available from a specific data source. Data formats developed as a part of this task are discussed in **Section *4A***. *Data Collection and Preprocessing*.

Task 2. Evaluation of machine learning algorithms for source identification. This task focused on the use of conventional machine learning classifiers as a means of source identification for PFAS sites. The task was focused primarily on supervised learning (classification) techniques, although selected unsupervised learning (clustering) analyses were also explored, as an alternate approach for distinguishing AFFF and non-AFFF sources at specific sites [3]. The work in this task was designed to be conducted using the open source scikit-learn toolbox [19], allowing a range of approaches to be explored rapidly, and making the results of the work highly extensible.

Subtask 2.a. **Supervised learning.** Machine learning algorithms for classification are trained using data collected in **Task 1**. For this work, a predetermined fraction of all samples at each site are set aside as a test set using stratified sampling, and then machine learning classifiers are be trained on the remaining data. For this subtask, work was focused primarily on classification based on composition, although strategies for tiered analysis based on additional information (water quality parameters, spatial information) were also planned if selectivity was insufficiently accurate using composition only.

Classification for this work focused on **binary classification** (AFFF vs. non-AFFF). Ensemble methods were used where appropriate, with emphasis on training using an ensemble of overlapping data subsets (bootstrap aggregating, aka bagging [20]) using the same underlying algorithm, to prevent overfitting and improve performance. Assessment and evaluation of analyses conducted for this task were based on conventional measures of performance, including evaluation of confusion matrices, as well as calculation of a range of specific metrics.

Subtask 2.b. **Feature selection.** An important component of the proposed work is to determine *which* specific component concentrations are necessary for accurate source identification. Many older data sets are based on analyses that looked for far fewer PFAS components, likely making them unsuitable for detection at some sites. However, discarding older data sets could eliminate thousands of potentially useful data points. While many machine learning classifiers can provide information on feature importance in trained models, the interpretation of this information can be complicated when multiple features are correlated (as is the case for many component concentrations in PFAS samples). For this preliminary work, the effect of component selection was explored with trained models to determine which components are most critical to prediction at specific types of sites, and their effect on prediction performance.

Subtask 2.c. **Clustering.** Clustering algorithms, also known as unsupervised learning algorithms, look for patterns in unlabeled data, and as such are often proposed for source delineation. This subtask centered on conducting a preliminary exploration of the use of clustering as an alternate means of distinguishing between AFFF and non-AFFF samples.

Task 3. Preliminary data mining and exploration of deep learning algorithms. The two emphases of this task were designed to provide insights that could guide future work.

Subtask 3.a. **Preliminary data mining.** This subtask focuses on preliminary data mining computations aimed at exploring the extent to which data mining can be used to uncover information about physicochemical phenomena (transformations, adsorption) impacting PFAS compounds.

Subtask 3.b. **Preliminary deep learning.** Deep learning through artificial neural networks can be significantly more powerful for predictions in some types of problems, although it is not without challenges. For this work, preliminary tests of deep learning are conducted using Google's open source TensorFlow library [21], an extremely robust library that can be used for deep learning, and which is optimized for GPU computing. This subtask was designed to provide preliminary insight into how well-suited deep learning is to the problem of identifying PFAS sources, potentially guiding future work. For this work, deep neural networks were created for binary classification, and evaluated using the same measures employed for conventional supervised machine learning classifiers in **Task 2**.

3.B. Methods

Supervised Machine Learning Classifiers. A total of **12 machine learning classifiers** were tested in this work. Classifiers included including **11 conventional machine learning classifiers**, and **one multilayer perceptron feedforward deep neural network**.

Four variants of the Random Forest method were tested: Random Forest, Extremely-randomized trees (also known as Extra Trees), AdaBoost, and GradientBoosting. The Random Forest method and its variations are ensemble methods based on decision trees [22]. Decision tree methods classify data by creating a network of choices based on the magnitudes of features. For decision trees based on binary choices, trees are trained by dividing training data sets by identifying the magnitude of a single feature that most completely separates the data set into two parts with the highest purity (the most complete separation of the classes possible with a single split). The process is recursively repeated for each subset until the training set is characterized, or until some preset limits are reached. Decision trees are extremely simple and powerful, but have the potential to overfit data. The Random Forest method significantly reduces risks of overfitting by generating an ensemble of multiple decision trees, each typically trained to a randomly selected overlapping subset of the training data. Extra Trees is a variation of Random Forest that further randomizes the process of creation of decision trees by selecting random thresholds for features, rather than searching for optimal values, potentially further reducing the likelihood of overfitting. AdaBoost and Gradient Boosting methods are variants of Random Forest that make adjustments during training in order to compensate for classification errors. AdaBoost does this by increasing the weight of misclassified data points during training, while Gradient Boosting successively trains on the residual error from the previous training step.

Support-Vector Machines work by searching for hyperplanes (planes in the *n*-dimensional space defined by the *n* features) that separate data into classes [3, 23]. For this work, three kernels were explored, a linear kernel, a cubic polynomial kernel, and a Gaussian radial basis function (RBF) kernel [3]. The kernel describes the shape of the hyperplanes that are

used to divide data into classes. The linear and cubic polynomial kernels have linear and polynomial shapes, respectively, while the RBF kernel is an exponential function that maps training data to higher dimensional space that describes the similarity between individual samples. In the context of this work, the RBF effectively transforms the features from the scaled PFAS concentrations themselves to measures of the lengths of vectors between the scaled PFAS concentrations of all pairs of components in the training data. The mapped data are then separated using hyperplanes, an approach that ultimately classifies samples based on their similarity to samples in the training set.

The K-Neighbors method is one of the simplest machine learning classification methods. Like Support-Vector Machines with the RBF kernel, the K-Neighbors method classifies samples based on their similarity to labeled samples in the training set [3]. However, the K-Neighbors method works by finding the distance in the *n*-dimensional space defined by the *n* features (PFAS components, in this work) from the unknown data point to the *k* (often 3-5) nearest labeled data points, and then classifying the unknown from the consensus of the classification of those *k* labeled training points.

The Gaussian Naïve Bayes, Quadratic Discriminant Analysis and Gaussian Process classifiers are all based on Bayes' Theorem, but with somewhat different internal assumptions. All three methods were tested in early preliminary work ([1]) and found to work poorly, but are evaluated again here with the larger dataset and different scaling approaches.

Artificial neural network machine learning methods are completely different from conventional machine learning methods in that they function in a way that is patterned after networks of biological neurons in the brain. For this work, a dense, feedforward deep network was used, meaning that the outputs of each node in one layer form inputs to all nodes in the subsequent layer [3]. The nodes in the hidden and output layers were perceptrons, a form of artificial neurons that create step outputs in response to a weighted sum of inputs. Artificial neural networks are considered deep neural networks if they have two or more hidden layers.

Conventional classifier analyses were conducted using the open-source scikit-learn toolbox [19] (ver. 0.22), while artificial neural network analyses were conducted using Google's open source TensorFlow library [24] (ver. 2.0.0).

Training. Training each supervised machine learning classifier involved labeling each component in the training set as coming from either an AFFF or non-AFFF source based on the characterization in the dataset, and then training the classifier on the training set. Testing was conducted by using trained classifiers to predict data in the test set portion of the main dataset. Hyperparameters were selected by initial validation using a subset of the training data. Hyperparameters used for the analyses were as follows: For Extra Trees, analyses were conducted based on an ensemble of 1000 estimators (individual generated decision trees). The number of features randomly selected for consideration for each split was 2. Class weights were balanced by scaling to the number of samples in each class (i.e., AFFF vs non-AFFF). No limit was placed on either maximum tree depth of number of leaf nodes. For Support-Vector Machines, linear, cubic, and RBF kernels were used. As with Extra Trees, class weights were balanced by scaling to the number of samples in each class. The C parameter, a hyperparameter

of Support-Vector Machines, was set to 10^7. Higher C values cause the method to locate hyperplanes to correctly classify more of the training set. The gamma value was set to the reciprocal of the number of features times the variance of the training data. For the K-Neighbors method, 3 nearest neighbors were used for classification.

For the deep neural network, a five-layer model was used, consisting of an input layer, an output layer, and three hidden layers. The input layer contained eight neurons (the number of features used), the output layer contained one neuron (because binary classification was conducted), and the hidden layers each contained 64 neurons. The hidden layers used ReLU activation functions, while the output layer used a sigmoidal activation function. Training was conducted using Nadam optimization, with the weights initialized using He Normal initialization. For this work, an ensemble of 10 deep neural networks was trained and used for predictions. Each was trained on a different randomly-selected 80% of the training set, with the remining 20% of the training set used as a validation set. Performance on the validation set was monitored and used to stop the training process for each network in the ensemble.

4. RESULTS AND DISCUSSION

4.A. Data Collection and Preprocessing (Task 1).

In order to test the ability of supervised machine learning classifiers to distinguish between PFAS from AFFF and non-AFFF origin, it was first necessary to assemble a dataset. The target for Task 1 was to assemble a dataset containing more than ten thousand individual samples. Ultimately, we assembled a dataset containing data from **28,142 samples total, 22,068 of them with at least one detected PFAS compound**. (Note that although samples with no detected compounds contain no information about PFAS source origin, some were retained, particularly in cases where spatial information were available, for possible future site-specific spatial analyses such as plume delineation.) This section describes the contents and structure of the dataset.

4.A.1. Dataset Content.

Table 1 shows the sample breakdown of the dataset by phase (water, other) and classification (AFFF, non-AFFF, Unknown):

Table 1. Overview of sample classification in the dataset.Numbers shown correspond to numbers of individual samples collected.

Classification	Water	Other	Total
AFFF	7452	2762	10,214
non-AFFF	9288	678	9966
Unknown	1666	222	1888
Total	18,406	3662	**22,068**

The original objective was to collect at least 10,000 water samples for the dataset, but it became apparent early in the project that many of the data sources used also contained non-water sample data that could be useful for data mining and other exploratory project calculations, and there were substantial efficiencies to converting those data at the same time as the water data, so non-water sample data were compiled for selected sites as well. Non-water data include data for **soils**, **sediment**, **biota**, **sludge**, as well as other phases. **AFFF vs. non-AFFF** samples were classified based on available site information, and were scored on a certainty scale indicating the level of confidence in the classification. **Unknown** samples are those for which the source has not yet been identified. In some cases, the sources are likely unknowable, while in others classification is likely possible after future detailed study of available site and geographical information. Note that all duplicate samples (same location, same day) have been removed from the dataset, although samples from the same location from different sampling events have been retained. Overall, known **water samples** in the dataset are **44.5% from AFFF sources**, while known **samples in the entire dataset** are **50.6% from AFFF sources**. The reason for the difference is a significantly greater availability of data from non-water samples in available data for AFFF data sources. Additional details of categories included in the dataset and the classification scale are included in the *Dataset Structure* section.

As noted previously, data in the dataset were collected from sites around the world. **Table 2** shows the breakdown of samples by country and classification.

Table 2. Contents of the dataset by country and classification.

Country	AFFF	non-AFFF	Unknown	Total
Australia	6333	110	0	6443
Canada	2533	11	0	2544
China	0	163	263	426
Germany	0	20	63	83
Spain	0	6	0	6
Ireland	0	48	0	48
Malta	0	0	71	71
Sweden	208	119	124	451
United States	1140	9489	1367	11,996
Total	10,214	9966	1888	**22,068**

Note that the country-to country differences in part reflect the different data transparency and availability in the various countries, as well as ease with which data could be collected and converted rapidly for this 1-year proof of concept project. Substantial quantities of data across a wide range of sources remain available internationally for future collection.

Data collected from this work come largely from public sources, although some data have been provided through personal communication with specific agencies or investigators. With some exceptions, most of the data collected come from extracting PDF files or other publication data. The lack of a uniform digital reporting format for PFAS data is a significant impediment to compiling datasets of historical data like this one.

Some of the major types of data sources in the dataset include:

Military Bases: Site investigations at military bases from Australia, Canada and the United States have been extracted. Many of these investigations contain extensive water and non-water sampling, and often include sampling over many years. More data are available from sites in all three of these countries.

Commercial Airports: Commercial airport data have primarily been collected from online agency publications from Australia, Canada, the United States and Sweden.

Specific Industrial Sites: A wide range of sits corresponding to specific industries have been collected, including data from areas surrounding coatings plants, metal platers, PFAS manufacturing sites, among others. Data in these categories generally comes from direct contact with agencies or site managers.

Landfills or Wastewater Treatment Plants: Data in this category come largely from direct contact with agencies or from published reports and other publications.

Studies in the Published Literature: Data in a range of different categories have also been collected from the published literature. Data from the published literature allow inclusion of information from countries where public domain data are not available, but are also often of less utility for this application because specific location information or other critical site details are often omitted.

Table 3 provides a detailed list of all of the sites included in the dataset. The table is organized by country and state/province, and includes all water and non-water sample data. Site structure and likely PFAS origin are included if known, as are the breakdown of compound detections by classification (AFFF, non-AFFF, Unknown). Concentration ranges for samples with detections are given, as is the breakdown of percentages of samples from different sample type categories. Note that additional details of the dataset are included in the accompanying Appendices. **Appendix A** provides more detailed data source information, while **Appendix B** provides concentration-averaged compositions for all data subsets in **Table 3**.

Table 3. Detailed listing of data sources in the dataset, including sample classifications and concentration ranges.

						Number of Detected Compounds				Conc. Range (ng/L)		Sample Type Percentages								
	Country	State/ Province	Short_name	Site_structure	Likely_origin	Total	AFFF	non-AFFF	Unknown	Min.	Max.	%_GW	%_SW	%_WAT_OTHER	%_SOIL	%_SEDIMENT	%_BIOTA	%_CONCRETE	%_ASPHALT	%_SLUDGE	%_OTHER
1	AU	JBT	JERVISBAY_BIOTA	MILITARY_BASE	AFFF	339	339	0	0	1	2730	0	0	0	0	0	100	0	0	0	0
2	AU	JBT	JERVISBAY_CONCRETE	MILITARY_BASE	AFFF	14	14	0	0	7.1	5648.1	0	0	0	0	0	0	100	0	0	0
3	AU	JBT	JERVISBAY_CONCRETE_LEACHATE	MILITARY_BASE	AFFF	15	15	0	0	66	200919	0	0	100	0	0	0	0	0	0	0
4	AU	JBT	JERVISBAY_GW	MILITARY_BASE	AFFF	125	125	0	0	7.1	433186	100	0	0	0	0	0	0	0	0	0
5	AU	JBT	JERVISBAY_SEDIMENT	MILITARY_BASE	AFFF	76	76	0	0	0.2	238.7	0	0	0	0	100	0	0	0	0	0
6	AU	JBT	JERVISBAY_SEDIMENT_LEACHATE	MILITARY_BASE	AFFF	28	28	0	0	11	8185.9	0	0	100	0	0	0	0	0	0	0
7	AU	JBT	JERVISBAY_SOIL	MILITARY_BASE	AFFF	245	245	0	0	0.2	17837.5	0	0	0	100	0	0	0	0	0	0
8	AU	JBT	JERVISBAY_SW	MILITARY_BASE	AFFF	114	114	0	0	0.3	14000.7	0	100	0	0	0	0	0	0	0	0
9	AU	JBT	JERVISBAY_TANK_SW	MILITARY_BASE	AFFF	7	7	0	0	630	40950	0	100	0	0	0	0	0	0	0	0
10	AU	NSW	ALBATROSS_GW	MILITARY_BASE	AFFF	79	79	0	0	10	697530	100	0	0	0	0	0	0	0	0	0
11	AU	NSW	ALBATROSS_LEACHATE	MILITARY_BASE	AFFF	235	235	0	0	10	563630	0	0	100	0	0	0	0	0	0	0
12	AU	NSW	ALBATROSS_SEDIMENT	MILITARY_BASE	AFFF	23	23	0	0	1.3	3059	0	0	0	0	100	0	0	0	0	0
13	AU	NSW	ALBATROSS_SOIL	MILITARY_BASE	AFFF	146	146	0	0	5	36096	0	0	0	100	0	0	0	0	0	0
14	AU	NSW	BLAMEY_GW	MILITARY_BASE	AFFF	8	8	0	0	20	580	100	0	0	0	0	0	0	0	0	0
15	AU	NSW	HOLSWORTHY_OFF_GW	MILITARY_BASE	AFFF	10	10	0	0	10	663100	100	0	0	0	0	0	0	0	0	0
16	AU	NSW	HOLSWORTHY_ON_GW	MILITARY_BASE	AFFF	32	32	0	0	10	663100	100	0	0	0	0	0	0	0	0	0
17	AU	NSW	RICHMOND_GW	MILITARY_BASE	AFFF	69	69	0	0	20	1043820	100	0	0	0	0	0	0	0	0	0
18	AU	NSW	SINGLETON_OFF_GW	MILITARY_BASE	AFFF	14	14	0	0	10	1060	100	0	0	0	0	0	0	0	0	0
19	AU	NSW	SINGLETON_ON_GW	MILITARY_BASE	AFFF	41	41	0	0	20	167530	100	0	0	0	0	0	0	0	0	0
20	AU	NSW	WAGGA_GW	MILITARY_BASE	AFFF	40	40	0	0	10	16201250	100	0	0	0	0	0	0	0	0	0
21	AU	NSW	WILLIAMTOWN_GW	MILITARY_BASE	AFFF	473	473	0	0	10	576460	100	0	0	0	0	0	0	0	0	0
22	AU	NSW	WILLIAMTOWN_SEDIMENT	MILITARY_BASE	AFFF	112	112	0	0	0.2	2921	0	0	0	0	100	0	0	0	0	0
23	AU	NSW	WILLIAMTOWN_SW	MILITARY_BASE	AFFF	369	369	0	0	7	97630	0	100	0	0	0	0	0	0	0	0
24	AU	NT	DARWIN_GW	MILITARY_BASE	AFFF	259	259	0	0	10	929300	100	0	0	0	0	0	0	0	0	0
25	AU	NT	ROBERTSON_DRY	MILITARY_BASE	AFFF	7	7	0	0	50	850	100	0	0	0	0	0	0	0	0	0
26	AU	NT	ROBERTSON_WET	MILITARY_BASE	AFFF	11	11	0	0	20	2160	100	0	0	0	0	0	0	0	0	0
27	AU	QLD	AMBERLEY_OFF_GWSW	MILITARY_BASE	AFFF	88	88	0	0	3	39630	8	92	0	0	0	0	0	0	0	0
28	AU	QLD	AMBERLEY_ON_GWSW	MILITARY_BASE	AFFF	126	126	0	0	10	455910	49	51	0	0	0	0	0	0	0	0
29	AU	QLD	CAIRNS_GW	MILITARY_BASE	AFFF	110	110	0	0	0.3	127240	100	0	0	0	0	0	0	0	0	0
30	AU	QLD	CAIRNS_SW	MILITARY_BASE	AFFF	21	21	0	0	0.1	1769	0	100	0	0	0	0	0	0	0	0
31	AU	QLD	LAVARACK_OFF_GW	MILITARY_BASE	AFFF	28	28	0	0	14	1146	100	0	0	0	0	0	0	0	0	0
32	AU	QLD	LAVARACK_OFF_SW	MILITARY_BASE	AFFF	61	61	0	0	2.1	1823	0	100	0	0	0	0	0	0	0	0
33	AU	QLD	LAVARACK_ON_GW	MILITARY_BASE	AFFF	58	58	0	0	34	804507	100	0	0	0	0	0	0	0	0	0
34	AU	QLD	LAVARACK_ON_SW	MILITARY_BASE	AFFF	38	38	0	0	11	39907	0	100	0	0	0	0	0	0	0	0
35	AU	QLD	OAKEY_OFF_GW	MILITARY_BASE	AFFF	57	57	0	0	10	59490	100	0	0	0	0	0	0	0	0	0
36	AU	QLD	OAKEY_ON_GW	MILITARY_BASE	AFFF	75	75	0	0	10	2158090	100	0	0	0	0	0	0	0	0	0
37	AU	QLD	OAKEY_ON_SOIL	MILITARY_BASE	AFFF	63	63	0	0	0.3	76197.7	0	0	0	100	0	0	0	0	0	0
38	AU	QLD	OAKEY_SW	MILITARY_BASE	AFFF	17	17	0	0	1.4	4713.4	0	100	0	0	0	0	0	0	0	0
39	AU	QLD	QH3_BIOTA	AIRPORT	AFFF	117	117	0	0	0.9	868.7	0	0	0	0	0	100	0	0	0	0
40	AU	QLD	QH3_CONCENTRATE	AIRPORT	AFFF	28	28	0	0	809810	1.343E+09	0	0	0	0	0	0	0	0	0	100
41	AU	QLD	QH3_CONCRETE	AIRPORT	AFFF	5	5	0	0	608.7	94856.8	0	0	0	0	0	0	100	0	0	0
42	AU	QLD	QH3_GW	AIRPORT	AFFF	33	33	0	0	19	858540	100	0	0	0	0	0	0	0	0	0
43	AU	QLD	QH3_PW	AIRPORT	AFFF	5	5	0	0	374	478610	100	0	0	0	0	0	0	0	0	0
44	AU	QLD	QH3_SEDIMENT	AIRPORT	AFFF	67	67	0	0	0.3	3106.5	0	0	0	0	100	0	0	0	0	0
45	AU	QLD	QH3_SEWER	AIRPORT	AFFF	168	168	0	0	1	98780	0	0	100	0	0	0	0	0	0	0
46	AU	QLD	QH3_SOIL	AIRPORT	AFFF	76	76	0	0	0.2	2039.7	0	0	0	100	0	0	0	0	0	0
47	AU	QLD	QH3_SW	AIRPORT	AFFF	179	179	0	0	0.4	600914	0	100	0	0	0	0	0	0	0	0
48	AU	QLD	QH3_WASTE_SOIL	AIRPORT	AFFF	2	2	0	0	87	125.8	0	0	0	100	0	0	0	0	0	0
49	AU	QLD	QH3_WWTP	AIRPORT	AFFF	348	348	0	0	20	6395870	0	0	100	0	0	0	0	0	0	0

Table 3, contd.

						Number of Detected Compounds				Conc. Range (ng/L)		Sample Type Percentages								
	Country	State/ Province	Short_name	Site_structure	Likely_origin	Total	AFFF	non-AFFF	Unknown	Min.	Max.	%_GW	%_SW	%_WAT_OTHER	%_SOIL	%_SEDIMENT	%_BIOTA	%_CONCRETE	%_ASPHALT	%_SLUDGE	%_OTHER
50	AU	QLD	TOWNSVILLE_OFF1_GW	MILITARY_BASE	AFFF	141	141	0	0	2.1	26880	100	0	0	0	0	0	0	0	0	0
51	AU	QLD	TOWNSVILLE_OFF2_GW	MILITARY_BASE	AFFF	27	27	0	0	2.5	3529.4	100	0	0	0	0	0	0	0	0	0
52	AU	QLD	TOWNSVILLE_ON_GW	MILITARY_BASE	AFFF	190	190	0	0	60	19739800	100	0	0	0	0	0	0	0	0	0
53	AU	Various	GALLEN_2017_LF_GW	LANDFILL	nonAFFF	97	0	97	0	211.8	45600	100	0	0	0	0	0	0	0	0	0
54	AU	VIC	BANDIANA_OFF_GW	MILITARY_BASE	AFFF	11	11	0	0	1.9	385.5	100	0	0	0	0	0	0	0	0	0
55	AU	VIC	BANDIANA_ON_GW	MILITARY_BASE	AFFF	41	41	0	0	10	171700	100	0	0	0	0	0	0	0	0	0
56	AU	VIC	BANDIANA_POU_SW	MILITARY_BASE	AFFF	4	4	0	0	0.7	109.4	100	0	0	0	0	0	0	0	0	0
57	AU	VIC	EASTSALE_OFF_GW	MILITARY_BASE	AFFF	5	5	0	0	40	770	100	0	0	0	0	0	0	0	0	0
58	AU	VIC	EASTSALE_ON_GW	MILITARY_BASE	AFFF	75	75	0	0	30	14243150	100	0	0	0	0	0	0	0	0	0
59	AU	VIC	WILLIAMS_GW	MILITARY_BASE	AFFF	10	10	0	0	70	210850	100	0	0	0	0	0	0	0	0	0
60	AU	VIC	HEPBURN_2019_LF_GW	LANDFILL	nonAFFF	13	0	13	0	25.57	5182.03	100	0	0	0	0	0	0	0	0	0
61	AU	WA	PEARCE_BIOTA	MILITARY_BASE	AFFF	48	48	0	0	0.2	2493.4	0	0	0	0	0	100	0	0	0	0
62	AU	WA	PEARCE_GW	MILITARY_BASE	AFFF	50	50	0	0	30	2973590	100	0	0	0	0	0	0	0	0	0
63	AU	WA	PEARCE_SOIL	MILITARY_BASE	AFFF	404	404	0	0	0.2	319755.5	0	0	0	100	0	0	0	0	0	0
64	AU	WA	PEARCE_LEACHABILITY	MILITARY_BASE	AFFF	10	10	0	0	6160	1486600	0	0	100	0	0	0	0	0	0	0
65	AU	WA	STIRLING_GW	NAVAL_BASE	AFFF	471	471	0	0	1	236410	100	0	0	0	0	0	0	0	0	0
66	AU	WA	STIRLING_SEDIMENT	NAVAL_BASE	AFFF	15	15	0	0	0.1	280.7	0	0	0	100	0	0	0	0	0	0
67	AU	WA	STIRLING_SOIL	NAVAL_BASE	AFFF	142	142	0	0	0.1	26860.6	0	0	0	100	0	0	0	0	0	0
68	AU	WA	STIRLING_SW	NAVAL_BASE	AFFF	28	28	0	0	1	512	0	100	0	0	0	0	0	0	0	0
69	CA	DND	DND_Site-B_GW	MILITARY_BASE	AFFF	101	101	0	0	18	3516580	100	0	0	0	0	0	0	0	0	0
70	CA	DND	DND_Site-C_GW	MILITARY_BASE	AFFF	57	57	0	0	5.8	358500	100	0	0	0	0	0	0	0	0	0
71	CA	DND	DND_Site-C_SW	MILITARY_BASE	AFFF	16	16	0	0	76	11018.1	0	100	0	0	0	0	0	0	0	0
72	CA	DND	DND_Site-E_GW	MILITARY_BASE	AFFF	161	161	0	0	1.5	373106	100	0	0	0	0	0	0	0	0	0
73	CA	DND	DND_Site-E_MH	MILITARY_BASE	AFFF	3	3	0	0	549.1	12077.1	0	0	100	0	0	0	0	0	0	0
74	CA	DND	DND_Site-E_SEDIMENT	MILITARY_BASE	AFFF	4	4	0	0	18	192.7	0	0	0	0	100	0	0	0	0	0
75	CA	DND	DND_Site-E_SOIL	MILITARY_BASE	AFFF	52	52	0	0	0.55	3925	0	0	0	100	0	0	0	0	0	0
76	CA	DND	DND_Site-E_SW	MILITARY_BASE	AFFF	8	8	0	0	43	10792	0	100	0	0	0	0	0	0	0	0
77	CA	DND	DND_Site-G_GW	MILITARY_BASE	AFFF	319	319	0	0	0.581	1404200	100	0	0	0	0	0	0	0	0	0
78	CA	DND	DND_Site-G_SOIL_SEDIMENT	MILITARY_BASE	AFFF	453	453	0	0	0.1	10560.38	0	0	0	83	17	0	0	0	0	0
79	CA	DND	DND_Site-G_GW	MILITARY_BASE	AFFF	374	374	0	0	2.3	240910	0	100	0	0	0	0	0	0	0	0
80	CA	DND	DND_Site-H_GW	MILITARY_BASE	AFFF	205	205	0	0	3.2	64050	100	0	0	0	0	0	0	0	0	0
81	CA	DND	DND_Site-H_STORMWATER	MILITARY_BASE	AFFF	45	45	0	0	21	1925	0	0	100	0	0	0	0	0	0	0
82	CA	DND	DND_Site-H_SW	MILITARY_BASE	AFFF	408	408	0	0	2.1	5365	0	100	0	0	0	0	0	0	0	0
83	CA	DND	DND_Site-I_GW	MILITARY_BASE	AFFF	112	112	0	0	147	2130000	100	0	0	0	0	0	0	0	0	0
84	CA	DND	DND_Site-I_SEDIMENT	MILITARY_BASE	AFFF	17	17	0	0	0.18	59.7	0	0	0	0	100	0	0	0	0	0
85	CA	DND	DND_Site-I_SOIL	MILITARY_BASE	AFFF	172	172	0	0	1	44730	0	0	0	100	0	0	0	0	0	0
86	CA	DND	DND_Site-I_SW	MILITARY_BASE	AFFF	17	17	0	0	20	10390	0	100	0	0	0	0	0	0	0	0
87	CA	ON	HAMILTON_AIRPORT_GWSW	AIRPORT	AFFF	9	9	0	0	50	206830	67	33	0	0	0	0	0	0	0	0
88	CA	Various	BENSKIN_2012_LF_GW	LANDFILL	nonAFFF	11	0	11	0	2546.4	36207	100	0	0	0	0	0	0	0	0	0
89	CN	HE	BOHAISEA_SEDIMENT	SEA	ATM	30	0	0	30	0.703	4.128	0	0	0	0	100	0	0	0	0	0
90	CN	HE	BOHAISEA_SW	SEA	ATM	86	0	0	86	10.46	13528.65	0	100	0	0	0	0	0	0	0	0
91	CN	JS	WEI_2018_IND_GW	INDUSTRY	INDUSTRY	10	0	10	0	607.42	1224.191	100	0	0	0	0	0	0	0	0	0
92	CN	LN	BAO_2019_GW	RESIDENCES	PFAS_MFG	5	0	5	0	217	26733	100	0	0	0	0	0	0	0	0	0
93	CN	LN	GAO_2020_IND_SEDIMENT	RIVER	INDUSTRY	18	0	18	0	1.9389	40.1267	0	0	0	0	100	0	0	0	0	0
94	CN	LN	DALING_RIVER_BASIN_SW	RIVER	INDUSTRY	43	0	43	0	47.498	9952.061	0	100	0	0	0	0	0	0	0	0
95	CN	SC	FANG_2020_SOIL	RIVER	INDUSTRY	3	0	3	0	321.81	475.44	0	0	0	100	0	0	0	0	0	0
96	CN	SC	FANG_2020_SW	RIVER	INDUSTRY	54	0	54	0	16.66	3816.29	0	100	0	0	0	0	0	0	0	0
97	CN	Various	WANG_2020_WWTP	WWTP	UNK	33	0	0	33	145.53345	5084.239	0	0	100	0	0	0	0	0	0	0
98	CN	JS	WEI_2018_nonIND_GW	NON_INDUSTRY	ATM + LF	102	0	0	102	2.698	555.94	100	0	0	0	0	0	0	0	0	0

Table 3, contd.

						Number of Detected Compounds				Conc. Range (ng/L)		Sample Type Percentages								
	Country	State/ Province	Short_name	Site_structure	Likely_origin	Total	AFFF	non-AFFF	Unknown	Min.	Max.	%_GW	%_SW	%_WAT_OTHER	%_SOIL	%_SEDIMENT	%_BIOTA	%_CONCRETE	%_ASPHALT	%_SLUDGE	%_OTHER
99	CN	JS	WEI_2018_nonIND_SW	NON_INDUSTRY	ATM + LF	12	0	0	12	9.912	588.59	0	100	0	0	0	0	0	0	0	0
100	CN	Various	YAN_2015_LF_GW	LANDFILL	nonAFFF	5	0	5	0	7271.4	292149.51	100	0	0	0	0	0	0	0	0	0
101	CN	Various	YAN_2015_LF_SLUDGE	LANDFILL	nonAFFF	5	0	5	0	74.77	1689.8	0	0	0	0	0	0	0	0	100	0
102	CN	Various	YAN_2015_LF_WWTP	LANDFILL	nonAFFF	20	0	20	0	98.4	282746	0	0	100	0	0	0	0	0	0	0
103	DE	BY	BAVARIA_WTP	WTP	UNK	63	0	0	63	4	410	0	0	100	0	0	0	0	0	0	0
104	DE	Various	BUSCH_2010_LF_WWTP	LANDFILL	nonAFFF	20	0	20	0	4.72	8048.74	0	0	100	0	0	0	0	0	0	0
105	ES	Various	FUERTES_2017_LF_GW	LANDFILL	nonAFFF	6	0	6	0	639.5	3163.3	100	0	0	0	0	0	0	0	0	0
106	IE	Various	HARRAD_2019_LF_GW	LANDFILL	nonAFFF	48	0	48	0	13.736	17524.2	100	0	0	0	0	0	0	0	0	0
107	MT	Various	UNK-SAMMUT_2017_SW	ENV	ATM	41	0	0	41	0.89	35.02	0	100	0	0	0	0	0	0	0	0
108	MT	Various	UNK-SAMMUT_2019_GW	ENV	ATM	6	0	0	6	1.2	11.2	100	0	0	0	0	0	0	0	0	0
109	MT	Various	UNK-SAMMUT_2019_SOIL	ENV	ATM	24	0	0	24	0.2	8.5	0	0	0	100	0	0	0	0	0	0
110	SE	SIG	STOCKHOLM-ARLANDA_GW	AIRPORT	AFFF	26	26	0	0	182.02	797515	100	0	0	0	0	0	0	0	0	0
111	SE	Various	GOBELIUS_2018_FIRE_GWSW	FIRE_TRAINING	AFFF	182	182	0	0	0.07	12904.4	20	80	0	0	0	0	0	0	0	0
112	SE	Various	GOBELIUS_IND_GWSW	INDUSTRY	nonAFFF	88	0	88	0	0.1	214.8	48	52	0	0	0	0	0	0	0	0
113	SE	Various	GOBELIUS_LF_GWSW	LANDFILL	UNK	23	0	23	0	0.3	758.1	13	87	0	0	0	0	0	0	0	0
114	SE	Various	GOBELIUS_SKIING	SKI_AREA	UNK	8	0	8	0	0.09	56.1	25	75	0	0	0	0	0	0	0	0
115	SE	Various	GOBELIUS_OTHER_GWSW	UNK	UNK	89	0	0	89	0.05	345.3	31	69	0	0	0	0	0	0	0	0
116	SE	Various	GOBELIUS_URBAN	URBAN	UNK	20	0	0	20	0.08	22	60	40	0	0	0	0	0	0	0	0
117	SE	Various	GOBELIUS_WWTP	WWTP	UNK	15	0	0	15	0.2	35	0	0	100	0	0	0	0	0	0	0
118	US	AL	CENTRE_SW	WTP	COATINGS	97	0	97	0	130	870	0	100	0	0	0	0	0	0	0	0
119	US	AL	GADSDEN_SW	WTP	COATINGS	175	0	175	0	115.8	3024	0	100	0	0	0	0	0	0	0	0
120	US	AZ	DAVIS-MONTHAN_GW	MILITARY_BASE	AFFF	6	6	0	0	83.8	27039.4	100	0	0	0	0	0	0	0	0	0
121	US	AZ	DAVIS-MONTHAN_SEDIMENT	MILITARY_BASE	AFFF	3	3	0	0	1.07	1131.454	0	0	0	0	100	0	0	0	0	0
122	US	AZ	DAVIS-MONTHAN_SOIL	MILITARY_BASE	AFFF	10	10	0	0	0.7	27120.3	0	0	0	100	0	0	0	0	0	0
123	US	CA	CALIFGAMA_DOMESTIC_GW	DOMESTIC_WELLS	UNK	119	0	0	119	0.65	122.6	100	0	0	0	0	0	0	0	0	0
124	US	CA	CALIFGAMA_MON_GW	MONITORING_WELLS	UNK	251	0	0	251	0.29	278676.93	100	0	0	0	0	0	0	0	0	0
125	US	MI	DU-WEL_DBS_VAS_GW	METAL_PLATING	METAL_PLATING	14	0	14	0	3.91	14171.59	100	0	0	0	0	0	0	0	0	0
126	US	MI	DU-WEL_MW_GW	METAL_PLATING	METAL_PLATING	18	0	18	0	2.22	8808.3	100	0	0	0	0	0	0	0	0	0
127	US	MI	DU-WEL_RES_OFF_GW	METAL_PLATING	METAL_PLATING	53	0	53	0	2	900	100	0	0	0	0	0	0	0	0	0
128	US	MI	DU-WEL_VAS_OFF_GW	METAL_PLATING	METAL_PLATING	40	0	40	0	2.28	5247.46	100	0	0	0	0	0	0	0	0	0
129	US	MI	DU-WEL_VAS_ON_GW	METAL_PLATING	METAL_PLATING	102	0	102	0	2.59	264803.71	100	0	0	0	0	0	0	0	0	0
130	US	MI	WOLVERINE_HS_GW	TANNERY	SCOTCHGARD	99	0	99	0	1.8	598670	100	0	0	0	0	0	0	0	0	0
131	US	MI	WOLVERINE_HS_SOIL	TANNERY	SCOTCHGARD	280	0	280	0	4.9	84500	0	0	0	100	0	0	0	0	0	0
132	US	MI	WOLVERINE_TA_GW	TANNERY	SCOTCHGARD	108	0	108	0	5.6	451320	100	0	0	0	0	0	0	0	0	0
133	US	MI	WOLVERINE_TA_SOIL	TANNERY	SCOTCHGARD	182	0	182	0	3.7	4540	0	0	0	100	0	0	0	0	0	0
134	US	MI	WOLVERINE_TA_SW	TANNERY	SCOTCHGARD	14	0	14	0	8.1	2020	0	100	0	0	0	0	0	0	0	0
135	US	MN	EASTMETRO_A_GW	WELLS	PFAS_MFG	6518	0	6518	0	3	108011810	100	0	0	0	0	0	0	0	0	0
136	US	NC	CAPEFEAR_SW	RIVER	CHEMOURS	456	0	456	0	5.79	2971	0	100	0	0	0	0	0	0	0	0
137	US	OK	ALTUS_GW	MILITARY_BASE	AFFF	36	36	0	0	384.9	2481790	100	0	0	0	0	0	0	0	0	0
138	US	OK	ALTUS_SEDIMENT	MILITARY_BASE	AFFF	3	3	0	0	7.56	121.15	0	0	0	0	100	0	0	0	0	0
139	US	OK	ALTUS_SOIL	MILITARY_BASE	AFFF	94	94	0	0	0.24	8538.95	0	0	0	100	0	0	0	0	0	0
140	US	OK	ALTUS_SW	MILITARY_BASE	AFFF	3	3	0	0	245.9	50516.8	0	100	0	0	0	0	0	0	0	0
141	US	OR	PDX_GW	AIRPORT	AFFF	118	118	0	0	1.83	347870	100	0	0	0	0	0	0	0	0	0
142	US	OR	PDX_SOIL	AIRPORT	AFFF	4	4	0	0	54.29	121.94	0	0	0	100	0	0	0	0	0	0
143	US	OR	PDX_SW	AIRPORT	AFFF	24	24	0	0	4.62	19194.3	0	100	0	0	0	0	0	0	0	0
144	US	TX	ALY_2020_SW	SW	AFFF	52	52	0	0	41.539908	2044.0766	0	100	0	0	0	0	0	0	0	0
145	US	Various	HUSET_2011_LF_GW	LANDFILL	nonAFFF	6	0	6	0	2608.2	6971.3	100	0	0	0	0	0	0	0	0	0
146	US	Various	LANG_2017_LF_GW	LANDFILL	nonAFFF	87	0	87	0	224.3872	33500.48	100	0	0	0	0	0	0	0	0	0
147	US	Various	EPA_UCMR3_GWSW	WTP	UNK	598	0	0	598	10	9450	72	28	0	0	0	0	0	0	0	0

Table 3, contd.

	Country	State/ Province	Short_name	Site_structure	Likely_origin	Number of Detected Compounds				Conc. Range (ng/L)		Sample Type Percentages								
						Total	AFFF	non-AFFF	Unknown	Min.	Max.	%_GW	%_SW	%_WAT_OTHER	%_SOIL	%_SEDIMENT	%_BIOTA	%_CONCRETE	%_ASPHALT	%_SLUDGE	%_OTHER
148	US	VT	CLARENDON_GW_OFF	AIRPORT	AFFF	39	39	0	0	1.8	344.1	100	0	0	0	0	0	0	0	0
149	US	VT	CLARENDON_GW_ON	AIRPORT	AFFF	6	6	0	0	108.2	10926.2	100	0	0	0	0	0	0	0	0
150	US	VT	CLARENDON_SOIL_ON	AIRPORT	AFFF	10	10	0	0	0.7	98.51	0	0	0	100	0	0	0	0	0
151	US	VT	COVENTRY_LF_SLUDGE	LANDFILL	WWTP	31	0	0	31	0.28	273.3	0	0	0	0	0	0	0	100	0
152	US	VT	COVENTRY_LF_SOIL	LANDFILL	MIXED	10	0	0	10	0.043	26.07	0	0	0	100	0	0	0	0	0
153	US	VT	BENNINGTON_GW	RESIDENTIAL_WELLS	COATINGS	1042	0	1042	0	0.78	4653	100	0	0	0	0	0	0	0	0
154	US	VT	BENNINGTON_SOIL	SOIL	COATINGS	135	0	135	0	0.38	86.86	0	0	0	100	0	0	0	0	0
155	US	VT	COVENTRY_LF_SOURCE	LANDFILL	COATINGS	54	0	54	0	0.071	6750	0	0	0	0	0	0	0	0	100
156	US	VT	VTWWTF_EFF_WWTP	WWTP_EFFLUENT	UNK	127	4	4	119	3.83	1408.25	0	0	100	0	0	0	0	0	0
157	US	VT	VTWWTF_INF_WWTP	WWTP_INFLUENT	UNK	120	4	4	112	2	1902.06	0	0	100	0	0	0	0	0	0
158	US	VT	VTWWTF_SLUDGE	WWTP_SLUDGE	UNK	62	0	1	61	0.901	361.9	0	0	0	0	0	0	0	100	0
159	US	VT	VT_SHALLOW_SOIL	ENV	ATM	66	0	0	66	0.615	35.048	0	0	0	100	0	0	0	0	0
160	US	WI	MARINETTE_OFF_GW	RESIDENTIAL_WELLS	AFFF	634	634	0	0	0.33	2203	100	0	0	0	0	0	0	0	0
161	US	WI	MARINETTE_ON_GWSW	PFAS_MFG	AFFF	72	72	0	0	2.06	60736	44	56	0	0	0	0	0	0	0
162	US	WI	MARINETTE_ON_SOIL	PFAS_MFG	AFFF	18	18	0	0	3.285	1346	0	0	0	100	0	0	0	0	0

Overall, the dataset includes data for a total of **58 different PFAS components**, although the included components vary by site, depending on the analyses conducted at each site. **Table 4** shows the specific components included in the dataset. Note that individual data sources often use slightly different abbreviations for specific compounds, so it was necessary to convert these to a consistent naming convention for this work. **Table 4** shows the selected abbreviations used internally for machine learning.

Table 4. Components included in the dataset, and abbreviations selected for machine learning.

Abbreviation	Name	CAS number
PFPrA	Perfluoropropanoic acid	422-64-0
PFBA	Perfluorobutanoic acid	375-22-4
PFPeA	Perfluoropentanoic acid	2706-90-3
PFHxA	Perfluorohexanoic acid	307-24-4
PFHpA	Perfluoroheptanoic acid	375-85-9
PFOA	Perfluorooctanoic acid	335-67-1
PFNA	Perfluorononanoic acid	375-95-1
PFDA	Perfluorodecanoic acid	335-76-2
PFUnDA	Perfluoroundecanoic acid	2058-94-8
PFDoDA	Perfluorododecanoic acid	307-55-1
PFTrDA	Perfluorotridecanoic acid	72629-94-8
PFTeDA	Perfluorotetradecanoic acid	376-06-7
PFHxDA	Perfluorohexadecanoic acid	67905-19-5
PFOcDA	Perfluorooctadecanoic acid	16517-11-6
PFBS	perfluorobutane sulfonate	375-73-5
PFPeS	perfluoropentane sulfonate	2706-91-4
PFHxS	perfluorohexane sulfonate	355-46-4
PFHpS	perfluoroheptane sulfonate	375-92-8
PFOS	perfluorooctane sulfonate	1763-23-1
PFNS	perfluorononane sulfonate	68259-12-1
PFDS	perfluorodecane sulfonate	335-77-3
4:2FTS	4:2 Fluorotelomer sulfonate	757124-72-4
6:2FTS	6:2 Fluorotelomer sulfonate	27619-97-2
8:2FTS	8:2 Fluorotelomer sulfonate	39108-34-4
10:2FTS	10:2 Fluorotelomer sulfonate	120226-60-0
FOSA	Perfluorooctane sulfonamide	754-91-6
MeFOSA	n-Methyl perfluorooctane sulfonamide	31506-32-8
EtFOSA	n-Ethyl perfluorooctane sulfonamide	4151-50-2
MeFOSE	n-Methyl perfluorooctane sulfonamide ethanol	24448-09-7
EtFOSE	n-Ethyl perfluorooctane sulfonamide ethanol	1691-99-2

Table 4, contd.

Abbreviation	Name	CAS number
FOSAA	perfluorooctane sulfonamide acetic acid	2806-24-8
MeFOSAA	n-Methyl perfluorooctane sulfonamide acetic acid	2355-31-9
EtFOSAA	n-Ethyl perfluorooctane sulfonamide acetic acid	2991-50-6
6:2FTCA	6:2 fluorotelomer carboxylic acid	53826-12-3
8:2FTCA	8:2 fluorotelomer carboxylic acid	27854-31-5
10:2FTCA	10:2 fluorotelomer carboxylic acid	53826-13-4
3:3FTCA	3:3 fluorotelomer carboxylic acid	1169706-83-5
5:3FTCA	5:3 fluorotelomer carboxylic acid	914637-49-3
7:3FTCA	7:3 fluorotelomer carboxylic acid	812-70-4
6:2FTUCA	6:2 fluorotelomer unsaturated carboxylic acid	70887-88-6
8:2FTUCA	8:2 fluorotelomer unsaturated carboxylic acid	70887-84-2
10:2FTUCA	10:2 fluorotelomer unsaturated carboxylic acid	70887-94-4
6:2FTOH	6:2 Fluorotelomer alcohol	647-42-7
8:2FTOH	8:2 Fluorotelomer alcohol	678-39-7
10:2FTOH	10:2 Fluorotelomer alcohol	865-86-1
GenX	PFPrOPrA; HFPO-DA; 2,3,3,3-Tetrafluoro-2-(1,1,2,2,3,3,3-heptafluoropropoxy)-propanoic acid	13252-13-6
PFMOAA	Perfluoro-2-methoxyacetic acid, C3HF5O3	674-13-5
PFO2HxA	Perfluoro-(3,5-dioxahexanoic) acid, C4HF7O4	39492-88-1
PFMOPrA	Perfluoro-3-methoxy-propanoic acid	377-73-1
PFO3OA	Perfluoro-(3,5,7-trioxaoctanoic) acid	39492-89-2
PFO4DA	Perfluoro-(3,5,7,9-tetraoxadecanoic) acid	39492-90-5
PFO5DoDA	Perfluoro-3,5,7,9,11-pentaoxadodecanoic acid	39492-91-6
PFMOBA	Perfluoro-4-methoxy-butanic acid	863090-89-5
NfBP1	Nafion Byproduct 1, C7HF13SO5	29311-67-9
NfBP2	Nafion Byproduct 2, C7H2F14SO5	749836-20-2
ADONA	3H-Perfluoro-3-[(3-methoxy-propoxy)propanoic acid], ammonium salt	958445-44-8
F53B	Potassium 9-chlorohexadecafluoro-3-oxanonane-1-sulfonate	73606-19-6
F53Bmin	Potassium 11-Chloroeicosafluoro-3-oxaundecane-1-sulfonate (minor component of F53B)	83329-89-9

Figures 3-5 show the number of samples in the dataset for which individual compounds have been analyzed. This information is provided for the overall dataset (**Fig. 3**), the AFFF samples (**Fig. 4**) and the non-AFFF samples (**Fig. 5**). From **Figures 3-5**, it is apparent that there are significant differences in the compounds analyzed in samples both overall, and between site classes (i.e., AFFF, nonAFFF). In general, newer data sources contain data based on analyses of more compounds, while older sources contain fewer. Components that for which very few analyses are available include PFPrA, PFHxDA, PFOcDA, PFNS, FOSAA, as well as all of the compounds beyond EtFOSAA in **Table 4**. In many cases these compounds have not been analyzed because they are generally present in very low concentrations or not expected to be present. For example, most analyses of PFHxDA are found to be below detection limits. Similarly, although GenX is present in high concentrations at sites related to its manufacture, it might not be expected at many other categories of sites, so would not have been included in analyses conducted at those sites.

Table 5 shows the components for which analyses are available at each of the groundwater and surface water sites. (Only compounds through EtFOSAA are shown in **Table 5** to allow the information to fit on one page.) **Table 5** illustrates some of the same information as **Figures 3-5**, but on a site-by-site basis: some compound analyses are common to all data sources, while others are only analyzed infrequently.

The fact that analyses differ by site is important for the reason that **machine learning classifiers require the same features (in this case, PFAS components) to be present for all samples**. As such, a common subset of components must be selected for machine learning. That means that classification based on a larger number of components will likely mean that fewer data can be used to train models. Although there is likely a benefit to inclusion of more components for classifier performance, all other things being equal inclusion of low frequency components in machine learning analyses will likely remove representative site data from the training set, a significant problem for classification. While there are other important reasons to analyze for larger numbers of compounds in future sampling activities, from the standpoint of machine learning from historical data, it would likely take decades of new sampling to provide sufficient training data to allow the use of additional compounds. Furthermore, with remediation activities progressing at many known PFAS sites, the opportunity to collect representative samples for machine learning at some types of sites, at least in the US, has likely passed.

Quantitative information related to component selection for machine learning is provided in *Section 4.C Feature Selection.*

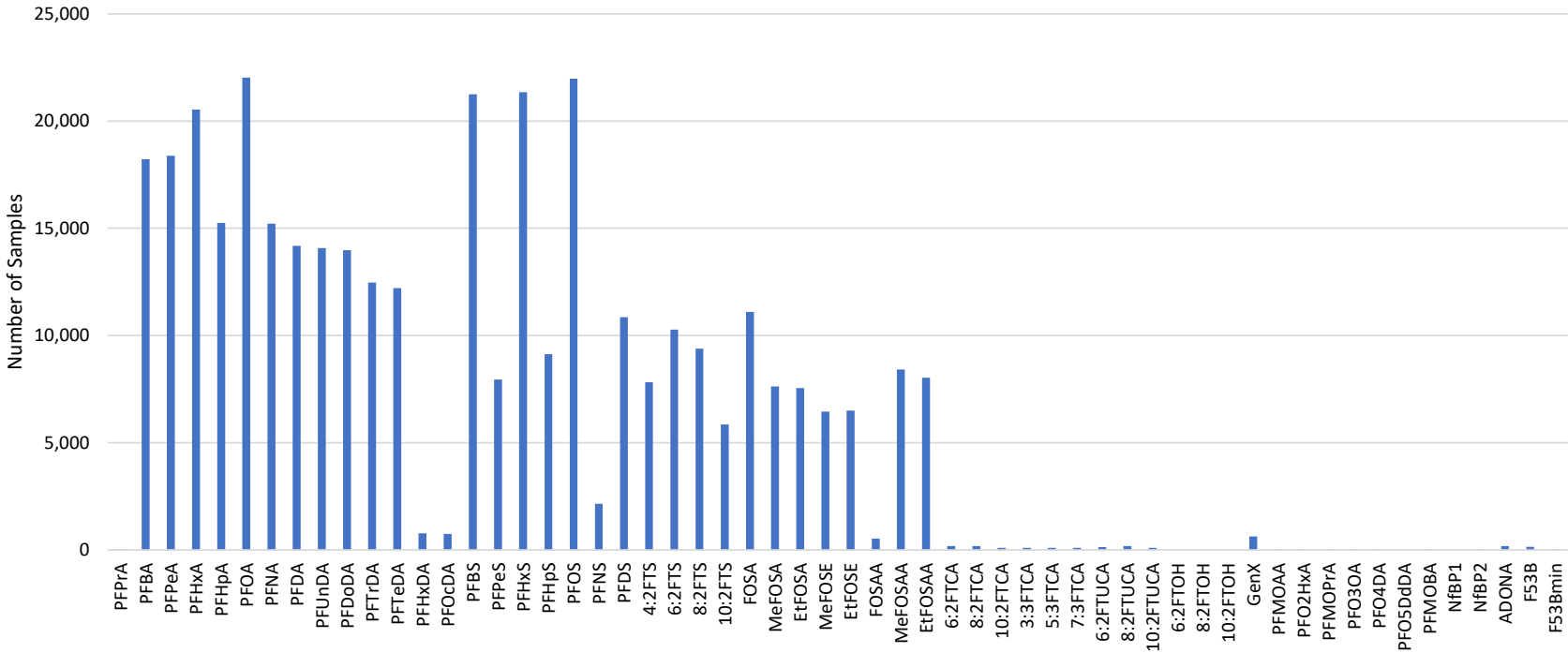


Figure 3. Numbers of samples in the complete dataset for which individual compound analyses are available.

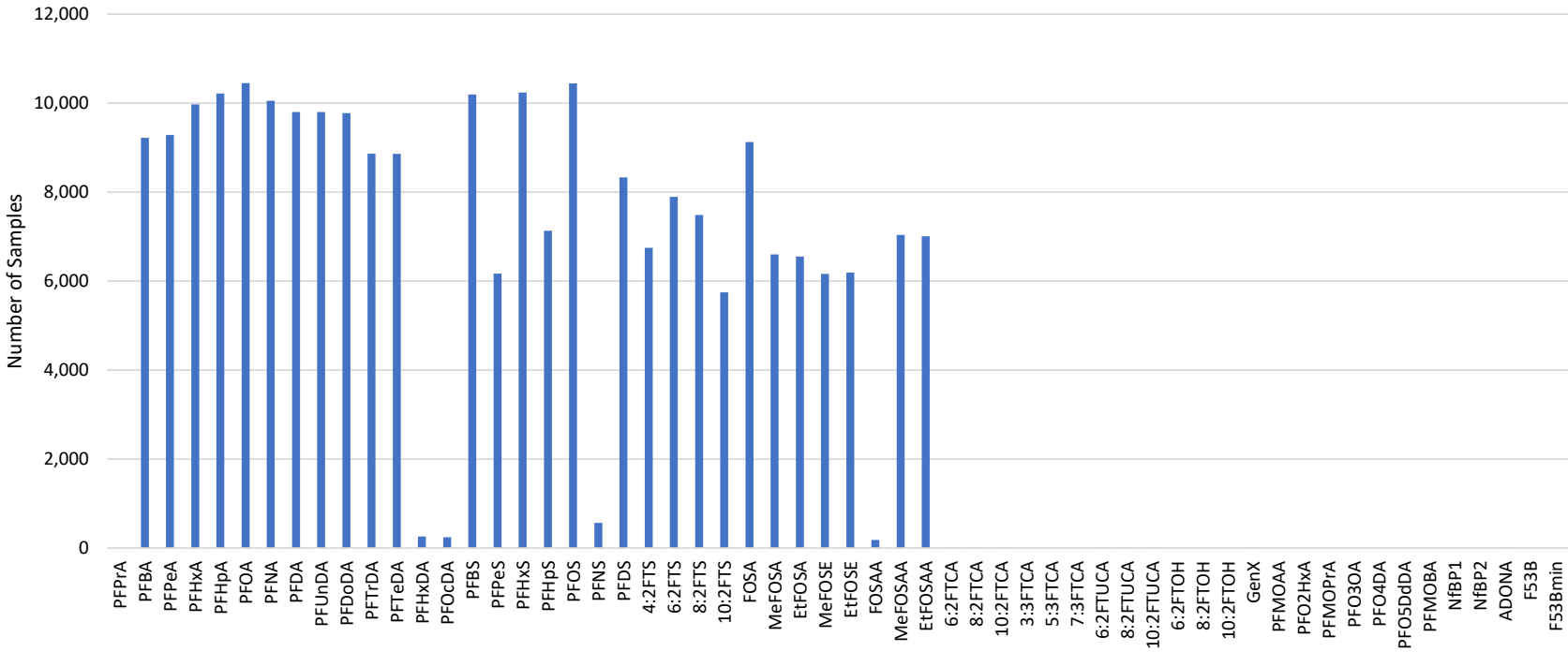


Figure 4. Numbers of AFFF samples in the dataset for which individual compound analyses are available.

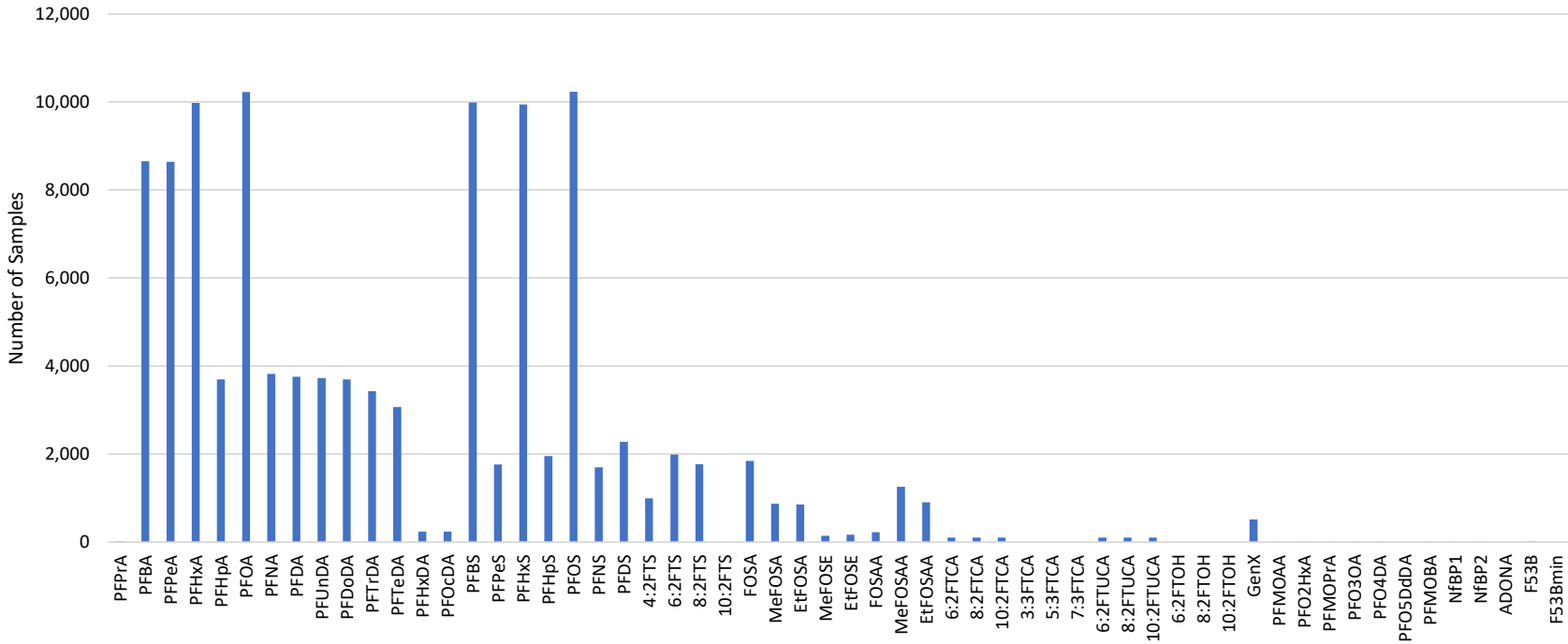


Figure 5. Numbers of non-AFFF samples in the dataset for which individual compound analyses are available.

Table 5. Components analyzed at each site for groundwater and surface water. Blue indicates at least one sample analysis conducted for this compound. Only components through EtFOSAA in Table 4 are shown.

	Country	State/ Province	Short_name	AFFF
4	AU	JBT	JERVISBAY_GW	X
8	AU	JBT	JERVISBAY_SW	X
9	AU	JBT	JERVISBAY_TANK_SW	X
10	AU	NSW	ALBATROSS_GW	X
14	AU	NSW	BLAMEY_GW	X
15	AU	NSW	HOLSWORTHY_OFF_GW	X
16	AU	NSW	HOLSWORTHY_ON_GW	X
17	AU	NSW	RICHMOND_GW	X
18	AU	NSW	SINGLETON_OFF_GW	X
19	AU	NSW	SINGLETON_ON_GW	X
20	AU	NSW	WAGGA_GW	X
21	AU	NSW	WILLIAMTOWN_GW	X
23	AU	NSW	WILLIAMTOWN_SW	X
24	AU	NT	DARWIN_GW	X
25	AU	NT	ROBERTSON_DRY	X
26	AU	NT	ROBERTSON_WET	X
27	AU	QLD	AMBERLEY_OFF_GWSW	X
28	AU	QLD	AMBERLEY_ON_GWSW	X
29	AU	QLD	CAIRNS_GW	X
30	AU	QLD	CAIRNS_SW	X
31	AU	QLD	LAVARACK_OFF_GW	X
32	AU	QLD	LAVARACK_OFF_SW	X
33	AU	QLD	LAVARACK_ON_GW	X
34	AU	QLD	LAVARACK_ON_SW	X
35	AU	QLD	OAKEY_OFF_GW	X
36	AU	QLD	OAKEY_ON_GW	X
38	AU	QLD	OAKEY_SW	X
42	AU	QLD	QH3_GW	X
43	AU	QLD	QH3_PW	X
47	AU	QLD	QH3_SW	X
50	AU	QLD	TOWNSVILLE_OFF1_GW	X
51	AU	QLD	TOWNSVILLE_OFF2_GW	X
52	AU	QLD	TOWNSVILLE_ON_GW	X
53	AU	Various	GALLEN_2017_LF_GW	
54	AU	VIC	BANDIANA_OFF_GW	X
55	AU	VIC	BANDIANA_ON_GW	X
56	AU	VIC	BANDIANA_POU_SW	X
57	AU	VIC	EASTSALE_OFF_GW	X
58	AU	VIC	EASTSALE_ON_GW	X
59	AU	VIC	WILLIAMS_GW	X
60	AU	VIC	HEPBURN_2019_LF_GW	
62	AU	WA	PEARCE_GW	X
65	AU	WA	STIRLING_GW	X
68	AU	WA	STIRLING_SW	X
69	CA	DND	DND_Site-B_GW	X
70	CA	DND	DND_Site-C_GW	X
71	CA	DND	DND_Site-C_SW	X
72	CA	DND	DND_Site-E_GW	X
76	CA	DND	DND_Site-E_SW	X
77	CA	DND	DND_Site-G_GW	X
79	CA	DND	DND_Site-G_GW	X
80	CA	DND	DND_Site-H_GW	X
82	CA	DND	DND_Site-H_SW	X
83	CA	DND	DND_Site-I_GW	X
86	CA	DND	DND_Site-I_SW	X
87	CA	ON	HAMILTON_AIRPORT_GWSW	X
88	CA	Various	BENSKIN_2012_LF_GW	
90	CN	HE	BOHAISEA_SW	
91	CN	JS	WEI_2018_IND_GW	
92	CN	LN	BAO_2019_GW	
94	CN	LN	DALING_RIVER_BASIN_SW	
96	CN	SC	FANG_2020_SW	
98	CN	JS	WEI_2018_nonIND_GW	
99	CN	JS	WEI_2018_nonIND_SW	
100	CN	Various	YAN_2015_LF_GW	
105	ES	Various	FUERTES_2017_LF_GW	
106	IE	Various	HARRAD_2019_LF_GW	
107	MT	Various	UNK-SAMMUT_2017_SW	
108	MT	Various	UNK-SAMMUT_2019_GW	
110	SE	SIG	STOCKHOLM-ARLANDA_GW	X
111	SE	Various	GOBELIUS_2018_FIRE_GWSW	X
112	SE	Various	GOBELIUS_IND_GWSW	
118	US	AL	CENTRE_SW	
119	US	AL	GADSDEN_SW	
120	US	AZ	DAVIS-MONTHAN_GW	X
125	US	MI	DU-WEL_DBS_VAS_GW	
126	US	MI	DU-WEL_MW_GW	
127	US	MI	DU-WEL_RES_OFF_GW	
128	US	MI	DU-WEL_VAS_OFF_GW	
129	US	MI	DU-WEL_VAS_ON_GW	
130	US	MI	WOLVERINE_HS_GW	
132	US	MI	WOLVERINE_TA_GW	
134	US	MI	WOLVERINE_TA_SW	
135	US	MN	EASTMETRO_A_GW	
136	US	NC	CAPEFEAR_SW	
137	US	OK	ALTUS_GW	X
140	US	OK	ALTUS_SW	X
141	US	OR	PDX_GW	X
143	US	OR	PDX_SW	X
144	US	TX	ALY_2020_SW	X
145	US	Various	HUSET_2011_LF_GW	
146	US	Various	LANG_2017_LF_GW	
148	US	VT	CLARENDON_GW_OFF	X
149	US	VT	CLARENDON_GW_ON	X
153	US	VT	BENNINGTON_GW	
160	US	WI	MARINETTE_OFF_GW	X
161	US	WI	MARINETTE_ON_GWSW	X

4.A.2. Dataset Structure.

A file-based data structure is used for this work. The format was developed specifically for this work, with the objective of putting the data into a format that can easily be read for machine learning applications. Each sub-site consists of two files, one containing header information, and a second containing data. (Each line in **Table 3** corresponds to a pair of files, so the dataset currently consists of **324 total files**.) This section describes the structure of those files.

Header Files

Header files consist of multiple lines with one or more entries on each line, separated by commas. The first entry on each line is a code indicating the contents of the line. Codes in header files include:

Site:	A description of the site
Document:	The specific document from which data were extracted
URL:	If available, the URL of the data source
Short_name:	A short name for the site, all capital letters. The naming convention includes an abbreviated site name, whether the data come from on or off site (where appropriate), and an indication of the phase present. In most cases, on and off-site data are not mixed in files, and files contain only one phase, although in some cases groundwater and surface water have been included in the same file.) Example: OAKEY_ON_GW
Country:	The country where the site is located.
State/Province:	The state or province where the site is located, two letter abbreviation. If location is unknown, this is reported as UNK. If data source includes multiple sites, this is reported as Various.
File_version:	The input file version. Over the course of this project, the input file format was revised, so this line was added to track format updates over the course of the project. Version 6 is the final version.
Site_location:	Three entries for latitude, longitude, and UTM Zone of site as a whole. Reported as UNK if unknown, or a representative country location if various sites present.
Site_structure:	A description of the type of site.
Likely_origin:	PFAS origin if known (e.g., type of use).

ON/OFF: Data for ON or OFF site, or BOTH. Corresponds to location where PFAS entered environment, so WWTP and WTP sites would be OFF, for example.

Prepared_by: Name of researcher who did the file conversion.

Notes: Any important notes about the data or conversion.

History: Details of conversion and revisions.

Data Files

The structure of the data files is as follows:

- The **top row** contains column headings, which are described below.
- One (or more) rows below the top row contain detection limits, with values listed in the PFAS component columns. Note that in some cases, as indicated in the header notes for specific sites, detection limits in the dataset are approximate. This is true for data sources where detection limits are not given (so are taken as the lowest detected value), or in cases where detection limits vary slightly by sample event due to different dilutions or other factors. Note that detection limits are included to allow reconstruction of original data sources, but are generally not useful for the machine learning analyses conducted for this project.
- All subsequent rows contain data corresponding to individual environmental samples

Column headers include the following. Headers marked with an asterisk (*) below are optional, and are only included if data are available to populate them:

Code An identifier for the sample location

Alt_Code* An additional identifier, if available (could be a lab analysis code, for example.)

Date The sampling date

Easting* Location of sample if known. na if not. (Line is omitted if location information is available for the entire site)

Northing* Location of sample if known. na if not. (Line is omitted if location information is available for the entire site)

Depth_Min* For samples collected over a depth range, minimum depth in meters

Depth_Max* For samples collected over a depth range, maximum depth in meters

Site_type AFFF, nonAFFF, UNK

Certainty A code from 1 to 5 indicating the certainty of the Site_type characterization. On-site at a remote military base would be 5, off site near the same base would be 3, for example. Cases where a classification has been suggested by authors or investigators but can't be verified, or for which questions remain, would be coded as 1.

Sub-type Additional information about the PFAS formulation present if known from historical information.

Sample A code describing the type of sample collected. Codes include:

Water codes:

- GW – Groundwater
- SW – Surface water
- WTP – Water sample taken in a water treatment plant, either influent or effluent. Also used for treated tap water if not from a residential well.
- WWTP – Influent or effluent from a wastewater treatment plant
- SEWER – Water samples from a storm sewer, manhole or other storm water
- LEACHATE_SYN – Leachate from a synthetic leaching procedure. (Note: Landfill leachate is coded as GW)
- CONCENTRATE – Concentrate from an original formulation

Non-water codes:

- SOIL – Soil
- SEDIMENT – Sediment from a stream or river
- BIOTA – Any kind of biota. Type specified in header file, Code or Alt_Code column
- CONCRETE – A concrete sample
- ASPHALT – An asphalt sample
- SLUDGE – Sludge from WWTP
- SOURCE – A solid sample that serves as a PFAS source, such as samples taken from objects in a landfill

DL_code If multiple detection limits are included, this is an integer numeric code (1, 2, …) that indicates which detection limit row corresponds to a specific sample.

PFAS component acronyms Acronyms from **Table 4**, in the order shown in **Table 4**. Only PFAS components analyzed are included. If a specific component is below detection limits, it is reported as **bdl**. If analyses were not conducted for a specific sample, it is reported as **na**. (This latter situation occurs in sites with a long history of analyses, for example, where earlier data analyzed fewer compounds.)

As mentioned previously, all sample duplicates (same day, same location) have been removed from the dataset.

Units in the dataset are **ng/L for water samples**, and **µg/kg for non-water samples**.

Filename Convention

File names have the following components, separated by dashes: Two letter country code, two or three letter state/province code, indication of whether the samples in the set are AFFF, nonAFFF or unknown, the short site name, and then 0_header.csv for the header files, or 1_data.csv for data files. (Note that the 0 and 1 are included to control the sort order in file lists.)

Data Extraction and Summary

Two tools have been developed to allow summarization of the contents of the data file, as well as extraction of subsets based on specified criteria for machine learning purposes. The data summary includes the information shown in **Tables 1, 2, 3** and **5** of this report. The extraction tool allows specification of desired phases (Sample codes), Site_types (AFFF, nonAFFF, UNK), certainty ranges, and specific PFAS components to include, and outputs files specifically for use as input to machine learning analyses.

4.B. Supervised Machine Learning for Source Allocation (Subtasks 2a, 2b and 3b)

This section discusses work conducted for Subtask 2a. (Supervised learning – conventional classifiers), Subtask 2b. (Feature selection), and Subtask 3b. (Preliminary deep learning). Note that the deep learning work described in Subtask 3b was conducted in concert with the conventional classification work in Subtask 2a, so is described in this section.

Early in the project, a small 1197-sample subset of the dataset was used to conduct preliminary supervised machine learning. The results of that work are described in detail in the paper:

> **T. Kibbey, R. Jabrzemski and D. O'Carroll. Supervised Machine Learning for Source Allocation of Per- and Polyfluoroalkyl Substances (PFAS) in Environmental Samples, 2020, *Chemosphere*, 252, 126593.** [1]

The results of that early work supported the hypothesis of the work, suggesting that supervised machine learning is capable of distinguishing between AFFF and non-AFFF samples based on composition with high accuracy. This section provides an updated analysis of supervised machine learning based on the final data set described in the preceding section.

4.B.1. Dataset, Features and Training.

Machine learning classifiers require that all samples used for training and testing have the same features (in this case, PFAS components). For this work, a **subset of 10 PFAS components was selected**: the perfluorinated carboxylic acids **PFHxA, PFHpA, PFOA, PFNA, PFDA, PFUnDA and PFDoDA**, and the perfluorinated sulfonates **PFBS, PFHxS and PFOS**. Selection of these samples allowed a dataset containing **8105 groundwater (GW) and surface water (SW) samples** to be extracted from the full dataset. Note that **only AFFF and non-AFFF samples with a certainty of 3 or better** were extracted.

Note that because of the different components analyzed in the different subsets of the dataset, the selected components provided the largest 10 component subset that could be extracted (**Table 5**). Including additional components or a different set of components would have led to a smaller extracted dataset size. This selection of components produced a set that contained **5929 AFFF samples**, and **2176 non-AFFF samples**, i.e., **73.1% AFFF**. The water sample data were divided evenly into separate test and training sets. **Classifiers were trained on the 4052 sample training set, and then evaluated on the *completely separate* 4053 sample test set.**

Note that in previous work, the features for machine learning were taken as the raw PFAS component concentrations. In this work, four transformations are explored: component concentrations, component distributions (i.e., fractions of each detected compound in each sample), common logs of component concentrations, and common logs of component distributions. As before, components found to be below detection limits were set to zero for this work. For analyses requiring a log transformation, zero values of concentration were set to 10^{-4} ng/L, a value well below the lowest value in the dataset in the dataset; testing found that the exact value selected did not impact classification, as all non-detects are essentially binned with this approach.

4.B.2. Source Allocation Results.

Before discussing the source allocation results, it is necessary to provide definitions that are needed to understand the performance measures used to evaluate results. In the context of this work, which involves binary classification (i.e., classification into two classes), it is necessary to define a POSITIVE class and a NEGATIVE class. Here we define **POSITIVE** as **AFFF**, and **NEGATIVE** as **non-AFFF**. (The selection of this convention has no impact on the outcome of classification, but is necessary for interpreting binary classification.) To describe the outcomes of machine learning classification, it is typical to consider the four possible conditions for a classification result: **True Positive (TP)**, **True Negative (TN)**, **False Positive (FP)** and **False Negative (FN)**. These terms and their meanings are summarized in **Table 6**:

Table 6. Possible outcomes of binary classification.

Condition	Description	Classified as:	Actually is:
TP	Positive identified correctly as positive	P	P
TN	Negative identified correctly as negative	N	N
FP	Negative identified INCORRECTLY as positive	P	N
FN	Positive identified INCORRECTLY as negative	N	P

Table 7 shows the information from **Table 6**, translated to AFFF and non-AFFF.

Table 7. Possible outcomes in the binary classification of AFFF/non-AFFF samples.

Condition	Description	Classified as:	Actually is:
TP	**AFFF** identified correctly as **AFFF**	AFFF	AFFF
TN	**non-AFFF** identified correctly as **non-AFFF**	non-AFFF	non-AFFF
FP	**non-AFFF** identified INCORRECTLY as **AFFF**	AFFF	non-AFFF
FN	**AFFF** identified INCORRECTLY as **non-AFFF**	non-AFFF	AFFF

Table 8 shows the performance measured used to evaluate the supervised machine learning classifiers tested for this work.

Table 8. Performance measures for machine learning classifiers.

Performance Measure	Acronym	Alt Name	Formula	Description
Accuracy	ACC	--	(TP+TN)/(P+N)	Fraction of all samples correctly identified
Positive Predictive Value	PPV	Precision	TP/(TP+FP)	Fraction of all *identified* as AFFF that are *actually* AFFF
True Positive Rate	TPR	Recall	TP/(TP+FN)	Fraction of all AFFF that are *identified* as AFFF
Negative Predictive Value	NPV	--	TN/(TN+FN)	Fraction of all *identified* as non-AFFF that are *actually* non-AFFF
True Negative Rate	TNR	Selectivity	TN/(TN+FP)	Fraction of all non-AFFF that are *identified* as non-AFFF
Balanced Accuracy	BA	--	(TPR+TNR)/2	Average of TPR, TNR

While at first glance it might seem that accuracy would be sufficient to describe the performance of a classifier. However, in a test dataset where the two classes are not perfectly balanced in numbers, accuracy can give a skewed picture of performance. For example, in this case, where the samples in the test set are 73% AFFF, a classifier that erroneously classifying the samples as *all* AFFF would indicate a 73% accuracy, a situation that clearly overstates the performance of the classifier. In that case, the TNR (selectivity) would actually be 0%, indicating that there is a problem with the classification. The balanced accuracy, an average of TPR and TNR, and a useful metric for evaluating classification of unbalanced data sets, would be 50% in that case.

Tables 9, 10, 11 and **12** show these performance measures for the **12 tested classifiers** on the **4053 sample test set**. Calculations are based on the four transformation used for the component concentrations: **Table 9.** concentration, unscaled, **Table 10.** distribution, unscaled, **Table 11.** concentration, logarithmic scale, **Table 12.** distribution, logarithmic scale. Also shown in the tables is the result of **naïve classification by ratio of perfluorinated sulfonates to total PFAS**. Because legacy PFOS-based AFFF formulations were dominated by perfluorinated sulfonates, a naïve first-cut identification is sometimes made by looking at this ratio. I.e., samples for which perfluorinated sulfinates are present in greater quantities than perfluorinated carboxylic acids are naïvely classified as being AFFF. However, this approach is unsatisfactory for the many different cases, including non-PFOS-based AFFF formulations (where perfluorinated sulfonates are absent, or are present at small concentrations), and for PFOS-based formulations far from their source (where more mobile non-sulfonated components can dominate distributions). Furthermore, many examples of non-AFFF formulations based on perfluorinated sulfinates exist.

The data in **Tables 9-12** have been shaded and highlighted as follows to make it easier to discern performance differences between methods: Values of **95% or above** are shown in a **bold blue font, with a light blue shaded background**, values **90% to 94.9%** are shown in a medium blue font, values **80% to 89%** are shown in black, values between **40% and 79.9%** are shown in red, and **values less than 40%** are shown in bold dark red with a gray shaded background.

Table 9. Unscaled, Concentration: Performance metrics for classification of the 4053-sample test set, shown as percent for the 12 classifiers. Classifiers are grouped by category, and shading and font color indicates performance. Classification by ratio is shown for comparison.

Classifier	Accuracy ACC	Pos. Pred. Value PPV (Precision)	True Pos. Rate TPR (Recall)	Neg. Pred. Value NPV	True Neg. Rate TNR (Selectivity)	Balanced Accuracy BA
Extra Trees	**95.7**	**96.2**	**97.9**	94.0	89.5	93.7
Random Forest	**95.5**	**96.1**	**97.8**	93.6	89.3	93.5
AdaBoost	92.0	93.5	**95.6**	87.4	81.9	88.8
Gradient Boosting	**95.7**	**96.4**	**97.8**	93.8	90.0	93.9
K-neighbors	92.9	94.8	**95.6**	87.7	85.8	90.7
Support Vector (Linear)	86.8	**95.3**	86.2	70.2	88.3	87.3
Support Vector (Cubic)	**28.6**	**100.0**	**2.3**	**27.4**	**100.0**	51.1
Support Vector (RBF)	63.3	**97.1**	51.3	42.0	**95.9**	73.6
Gaussian Naïve Bayes	**33.9**	90.1	**10.7**	**28.5**	**96.8**	53.7
Quadratic Discriminant	**39.3**	93.6	**18.2**	**30.3**	**96.6**	57.4
Gaussian Process Classifier	59.5	93.8	47.8	**39.2**	91.4	69.6
DNN Ensemble	94.7	**95.4**	**97.5**	92.9	87.2	92.3
Classification by Ratio	79.8	87.5	84.5	61.4	67.1	75.8

Table 10. Unscaled, Distribution: Performance metrics for classification of the 4053-sample test set, shown as percent for the 12 classifiers. Classifiers are grouped by category, and shading and font color indicates performance. Classification by ratio is shown for comparison.

Classifier	Accuracy ACC	Pos. Pred. Value PPV (Precision)	True Pos. Rate TPR (Recall)	Neg. Pred. Value NPV	True Neg. Rate TNR (Selectivity)	Balanced Accuracy BA
Extra Trees	**95.1**	**95.6**	**97.9**	93.9	87.6	92.8
Random Forest	**95.1**	**95.7**	**97.7**	93.3	88.1	92.9
AdaBoost	92.1	93.1	**96.3**	89.0	80.6	88.5
Gradient Boosting	94.6	**95.3**	**97.5**	92.7	86.8	92.1
K-neighbors	93.7	94.5	**97.0**	91.3	84.6	90.8
Support Vector (Linear)	82.8	**95.9**	79.8	62.3	90.8	85.3
Support Vector (Cubic)	86.8	**97.9**	83.7	68.2	**95.2**	89.4
Support Vector (RBF)	92.5	94.6	**95.2**	86.6	85.1	90.2
Gaussian Naïve Bayes	79.1	**97.9**	73.0	56.6	**95.7**	84.3
Quadratic Discriminant	72.2	**97.5**	63.6	49.2	**95.6**	79.6
Gaussian Process Classifier	87.8	88.9	**95.2**	83.8	67.7	81.4
DNN Ensemble	94.4	94.7	**97.9**	93.6	85.0	91.5
Classification by Ratio	79.8	87.5	84.5	61.4	67.1	75.8

Table 11. Logarithmic, Concentration: Performance metrics for classification of the 4053-sample test set, shown as percent for the 12 classifiers. Classifiers are grouped by category, and shading and font color indicates performance. Classification by ratio is shown for comparison.

Classifier	Accuracy ACC	Pos. Pred. Value PPV (Precision)	True Pos. Rate TPR (Recall)	Neg. Pred. Value NPV	True Neg. Rate TNR (Selectivity)	Balanced Accuracy BA
Extra Trees	**95.5**	**96.3**	**97.6**	93.2	89.9	93.7
Random Forest	**95.4**	**96.1**	**97.8**	93.6	89.1	93.4
AdaBoost	92.0	93.5	**95.6**	87.4	81.9	88.8
Gradient Boosting	**95.7**	**96.4**	**97.8**	93.8	90.0	93.9
K-neighbors	**95.1**	**96.5**	**96.8**	91.3	90.4	93.6
Support Vector (Linear)	84.5	90.0	88.6	70.3	73.2	80.9
Support Vector (Cubic)	91.8	94.7	94.1	84.3	85.6	89.9
Support Vector (RBF)	93.7	**96.2**	**95.2**	87.3	89.7	92.5
Gaussian Naïve Bayes	79.9	89.3	82.4	60.5	73.2	77.8
Quadratic Discriminant	84.7	90.2	88.8	70.7	73.8	81.3
Gaussian Process Classifier	94.4	95.0	**97.5**	92.6	86.0	91.7
DNN Ensemble	**96.2**	**96.8**	**98.0**	94.4	91.3	94.6
Classification by Ratio	79.8	87.5	84.5	61.4	67.1	75.8

Table 12. Logarithmic, Distribution: Performance metrics for classification of the 4053-sample test set, shown as percent for the 12 classifiers. Classifiers are grouped by category, and shading and font color indicates performance. Classification by ratio is shown for comparison.

Classifier	Accuracy ACC	Pos. Pred. Value PPV (Precision)	True Pos. Rate TPR (Recall)	Neg. Pred. Value NPV	True Neg. Rate TNR (Selectivity)	Balanced Accuracy BA
Extra Trees	**95.8**	**96.6**	**97.7**	93.5	90.6	94.2
Random Forest	**96.1**	**96.7**	**98.1**	94.6	90.8	94.5
AdaBoost	93.5	**95.0**	**96.2**	89.4	86.2	91.2
Gradient Boosting	**95.5**	**96.6**	**97.3**	92.6	90.6	94.0
K-neighbors	94.8	**96.0**	**97.0**	91.5	88.9	92.9
Support Vector (Linear)	85.0	90.0	89.5	71.8	72.9	81.2
Support Vector (Cubic)	91.2	**95.3**	92.6	81.3	87.5	90.1
Support Vector (RBF)	94.1	**96.2**	**95.7**	88.4	89.7	92.7
Gaussian Naïve Bayes	82.0	89.5	85.4	64.7	72.8	79.1
Quadratic Discriminant	86.0	90.1	90.9	74.7	72.8	81.8
Gaussian Process Classifier	94.7	**95.3**	**97.6**	92.9	86.8	92.2
DNN Ensemble	**95.5**	**95.8**	**98.2**	94.7	88.3	93.3
Classification by Ratio	79.8	87.5	84.5	61.4	67.1	75.8

From the results in **Tables 9-12**, a number of things are apparent:

1. The performance of the methods varies significantly by scaling method, with the **worst performance observed for the unscaled concentrations (Table 9)**, and the **best performance observed for the logarithmic distributions (Table 12)**. In general, for both concentrations and distributions, moving from unscaled to logarithmic improved performance, possibly due to the fact that concentrations in the data set range over many orders of magnitude. In addition, for both unscaled and logarithmic, moving from concentration to distributions improved performance. This is an interesting result given that conversion from concentration to distribution removes information. The reason for the improved performance could be due to the fact that it also reduces the range of the data.

2. Looking at the **unscaled concentration results** (**Table 9**), it is apparent that six of the methods perform worse than naïve classification by ratio with this scaling, some by a large margin: Support Vector (Linear, Cubic, RBF), Gaussian Naïve Bayes, Quadratic Discriminant, Gaussian Process Classifier. All of these except Support Vector (RBF) were found in our previous preliminary work [1] to perform poorly, and were not considered further in that work. Interestingly, **looking at the logarithmic distribution results (Table 12), all 12 methods outperform classification by ratio, although some only by a small margin.**

3. Overall, the **best performance is observed from the variants of the Random Forest method**, particularly **Extra Trees**, **Random Forest** and **Gradient Boosting**. The **Deep Neural Network Ensemble** is close behind, followed by **K-Neighbors**, **Support Vector (RBF)** and then **Gaussian Process**. All of these methods provide substantial improvements over naïve classification by ratio when a logarithmic scaling of distributions is used as preprocessing for the machine learning features.

Figures 6 and 7 show **graphical representations of the confusion matrix** for two of the best-performing classifiers: **Random Forest (Fig. 6)** and the **Deep Neural Network Ensemble (Fig. 7)**. Note that these and all subsequent analyses are based on a logarithmic distribution of contaminant concentrations, as that was found to provide the highest classification performance of the preprocessing methods tested (**Tables 9-12**). It is apparent from both figures that both methods are able to correctly identify the vast majority of samples, despite the fact that AFFF and non-AFFF samples have a wide range of compositions in the environment. Furthermore, it is interesting to note that the predictions made by both classifiers are quite similar, a fact that supports the assertion that patterns exist in the data that allow classification; both classifiers are able to identify those patterns.

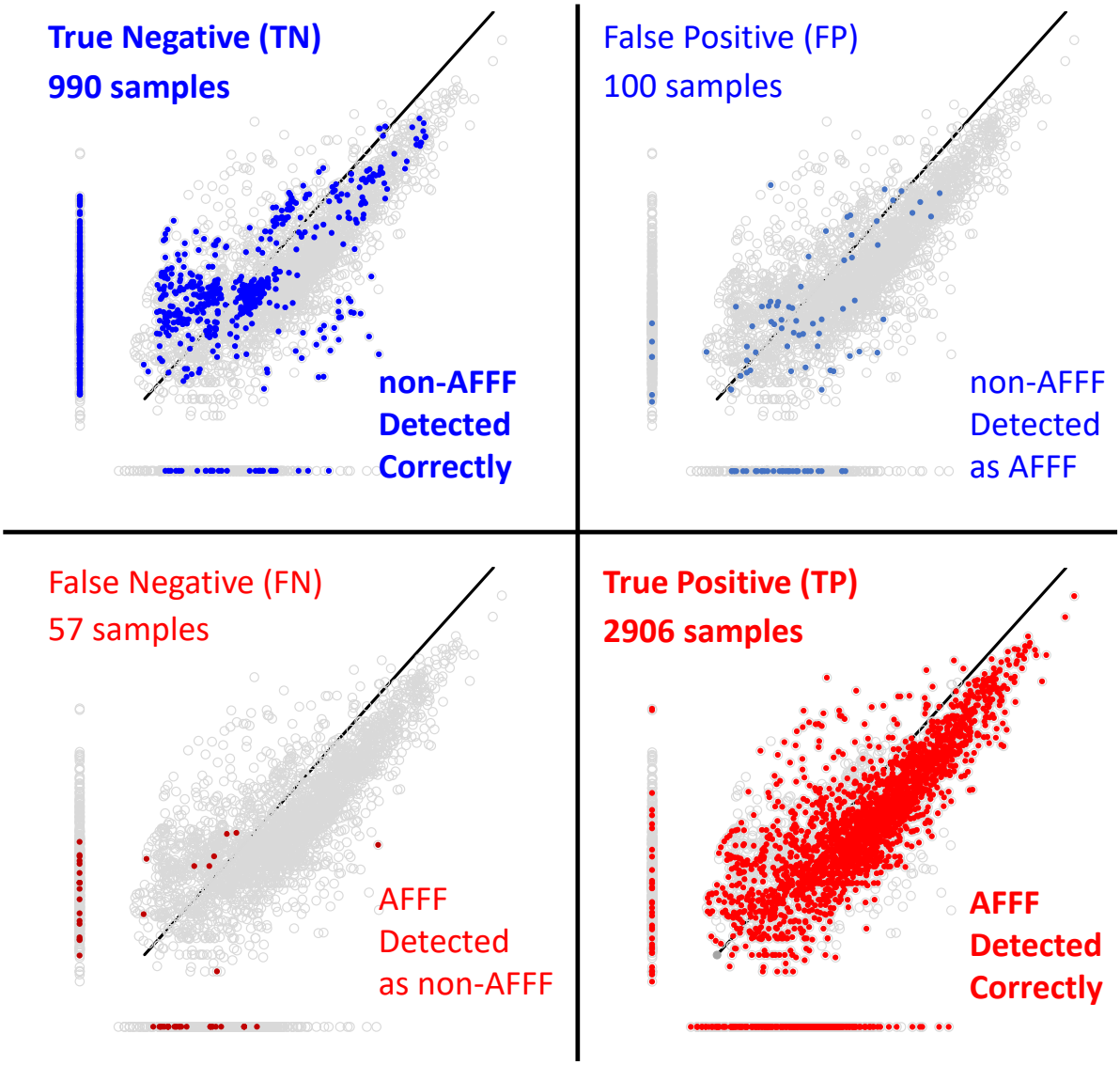


Figure 6. Graphical representation of the confusion matrix for predictions conducted using the **Random Forest classifier** for a 4053 sample test data set. The classifier was trained on a *completely separate* 4052 sample training set. Plots show total log concentrations of **perfluorinated sulfonates on the horizontal axis**, and **perfluorinated carboxylic acids on the vertical axis**; zero values cannot be displayed on a log scale, so are shown at the bottom and left of each plot. The line shown in each subfigure is a 1:1 line.

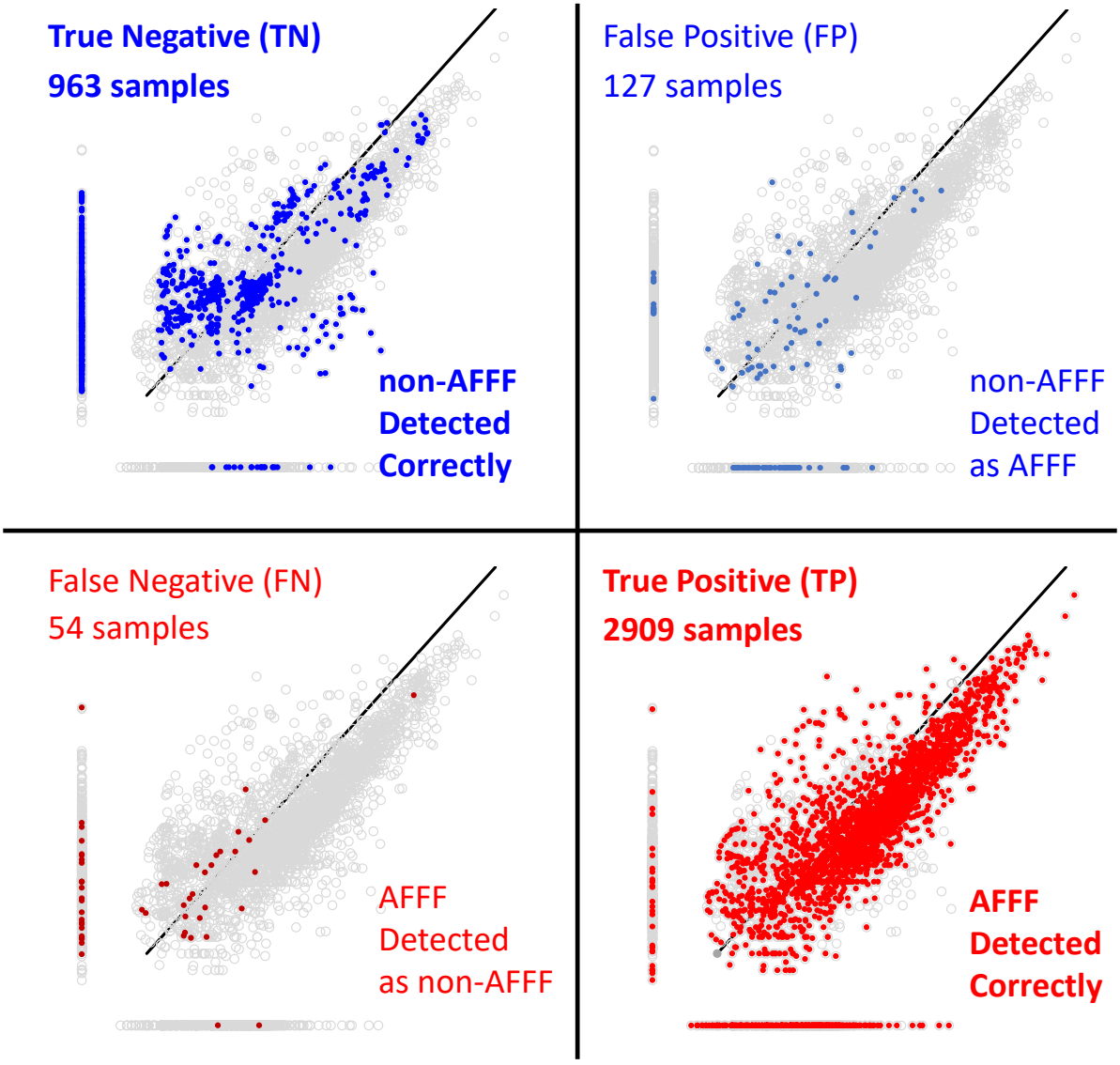


Figure 7. Graphical representation of the confusion matrix for predictions conducted using the **Deep Neural Network Ensemble classifier** for a 4053 sample test data set. The classifier was trained on a *completely separate* 4052 sample training set. Plots show total log concentrations of **perfluorinated sulfonates on the horizontal axis**, and **perfluorinated carboxylic acids on the vertical axis**; zero values cannot be displayed on a log scale, so are shown at the bottom and left of each plot. The line shown in each subfigure is a 1:1 line.

4.B.3. Difficult Subsets.

It is important to note that the absolute values in **Tables 9-12** and the results in **Figures 6 and 7** can not necessarily be directly compared with values for any other test set, such as the much smaller set examined in our earlier work [1], for the simple reason that the overall performance of the classifiers depends on what kinds of samples are present in the dataset. Some categories of samples are very easy to classify, while others can be particularly challenging because they mimic the other class. **Table 13** illustrates classification performance for three subsets of the dataset, samples from military bases, samples associated with sites where non-PFOS AFFF is known to be present, and samples associated with two sites where PFOS-based non-AFFF formulations are present. Note that the table shows accuracy. Because each subset is only AFFF or non-AFFF, accuracy provides a complete description of the performance of the classifiers.

Table 13. Classification accuracy for three subsets of the dataset.

	Accuracy, % (number correct/total in subset)		
	ALL Military Bases (AFFF)	**Portland, Clarendon, Marinette, Brisbane (likely non-PFOS AFFF)**	**Wolverine, Du-Wel (PFOS-based non-AFFF)**
Random Forest	**98.4** (2247/2284)	**96.2** (409/425)	**70.1** (157/224)
Deep Neural Network	**99.0** (2261/2284)	**92.9** (395/425)	**70.5** (158/224)
Classification by Ratio	**93.7** (2139/2284)	**26.1** (111/425)	**8.0** (18/224)

From **Table 13**, it is apparent that both methods do very well at classification of PFAS on military bases. In part, this is likely due to the large numbers of samples from military bases in the training set. However, the larger reason is that the patterns in that subset of samples are largely distinct from other subsets in the dataset. Note that classification by ratio also works well for this subset, primarily because despite variability in composition, the vast majority of samples in this subset contain higher total concentrations of perfluorinated sulfonates than perfluorinated carboxylic acids. Nevertheless, both classifiers correctly identify many of the samples for which this is not true, providing accuracy that exceeds classification by ratio in both cases.

While classification performance on military sites is especially high, note that **source allocation and plume delineation depend on the ability to *distinguish* sources**; as such, it is **essential that more difficult subsets of PFAS can be identified** so that they may be distinguished from AFFF.

Although contamination at many military sites is largely dominated by PFOS-based legacy AFFF, there **sites where there is a known history of non-PFOS-based AFFF contamination, or where low concentration off-site AFFF samples exhibit patterns that mimic non-AFFF sources**. Fire training activities associated with the Portland International Airport (Oregon) made use of Ansulite AFFF formulations for decades, and in at least one location there is evidence that land application was used for disposal. Contamination at the Rutland airport in Clarendon, Vermont likely includes non-PFOS-based AFFF contamination from a plane crash, and also includes low concentration data for off-site downgradient residential wells, where compositions are largely dominated by perfluorinated carboxylic acids, possibly due to transport. Contamination in Marinette, Wisconsin comes from manufacturing and testing

of Ansulite AFFFs, and include both on- and offsite data. Finally, contamination from a hanger fire at the Brisbane Airport is documented to have been due to non-PFOS-based legacy AFFF. From **Table 13**, it is apparent that **both machine learning classifiers identify data in this subset with high accuracy**, with the Random Forest method performing slightly better between the two. In contrast to the military bases, naïve classification by ratio performs poorly for this subset, entirely because nearly 74% of the samples have higher concentrations of perfluorinated carboxylic acids than perfluorinated sulfonates; as such, classification by ratio gives only a **26.1%** accuracy.

One of the most difficult subsets of the dataset for the classifiers to identify is the subset of non-AFFF samples that are dominated by PFOS and other perfluorinated sulfonates. Two sites for which this is true are the Wolverine Tannery site and the Du-Wel metal plating site, both in Michigan. The Wolverine Tannery site used early formulations of Scotchgard, a PFOS-based formulation. The formulation used at Du-Wel is unknown, but the environmental samples are largely dominated by PFOS. The performance of the classifiers on this subset is lower than other subsets (both are just over 70% accurate). However, in comparison to naïve classification by ratio (***8%* accuracy**), the performance of the classifiers is nevertheless a huge improvement.

Figures 8 and 9 show plots of correctly- and incorrectly- identified samples from the two difficult subsets in **Table 13**. Each figure shows the total perfluorinated sulfonates on the horizontal axis, and total perfluorinated carboxylic acids on the vertical axis. As before, those for which a total (or perfluorinated sulfonates or perfluorinated carboxylic acids) is below detection limits are shown to the left and below the figures because they can't be shown on a log scale. Classifications shown correspond to the **Random Forest** method in both cases. From the figures, it is apparent that data from the two subsets do not follow the pattern of AFFF having a greater fraction of perfluorinated sulfonates. In **Figure 8**, which corresponds to sites likely to have substantial contamination from non-PFOS-based AFFF the a large fraction of the samples are *above* the 1:1 line, indicating that they are **dominated by perfluorinated carboxylic acids**, and yet **more than 95% are correctly identified**. In **Figure 9**, which corresponds to sites likely to have substantial contamination from PFOS-based non-AFFF formulations, samples are almost entirely *below* the 1:1 line, indicating that they are **dominated by perfluorinated sulfonates**, and yet **more than 70% are correctly identified**.

It should be noted that many of the misclassified samples in **Figure 9** are for cases where only one or two components are above detection limits. This produces an especially difficult classification situation, as the distribution of a sample with only PFOS detected, for example, will necessarily look very much like many AFFF samples in the training set from legacy-PFOS-based AFFF, particularly those near a source, where perfluorinated sulfonates dominate. Future work is needed to consider the number of detected compounds to provide a weighting for the probability of classification. However, it is important to emphasize that samples with only a few component detections must necessarily be included in training data for the machine learning classifiers, because omitting them would skew the model classifications so that only higher concentration samples from near the source would be correctly classified, since some of the effects of different mobilities on composition would be lost.

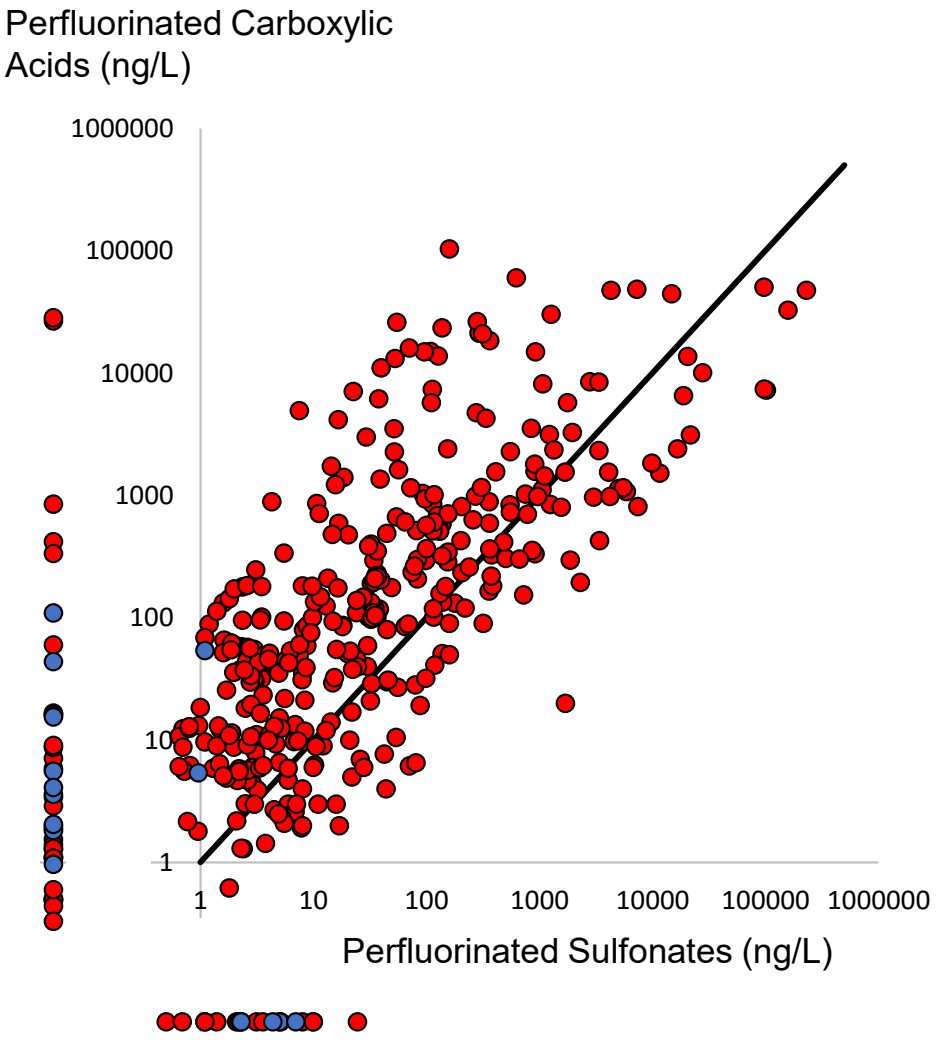


Figure 8. Classification of a subset of the test set corresponding to four sites where **non-PFOS-based AFFFs** are thought to be the major sources of PFAS present. **Red** = samples correctly classified, **Blue** = samples incorrectly classified. Plots show total log concentrations of perfluorinated sulfonates on the horizontal axis, and perfluorinated carboxylic acids on the vertical axis; zero values cannot be displayed on a log scale, so are shown at the bottom and left of each plot. The line shown is a 1:1 line. Classification based on the Random Forest method.

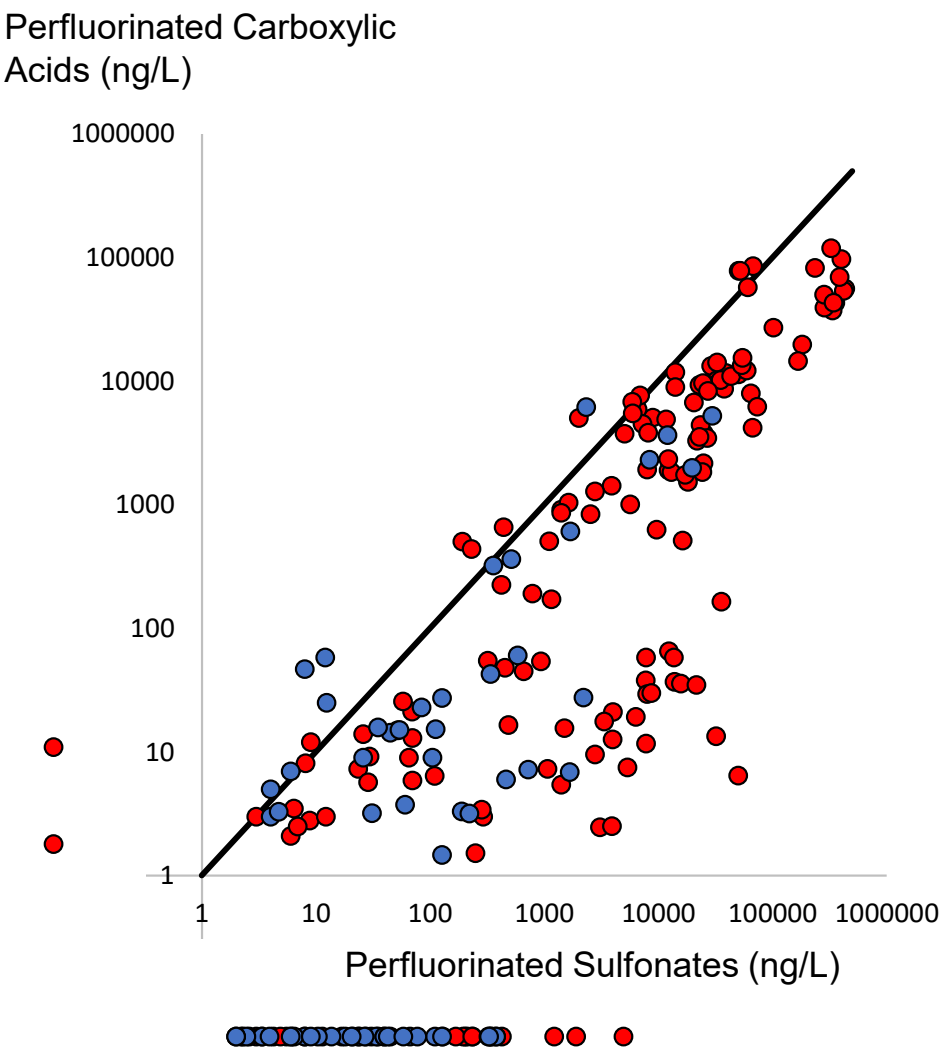


Figure 9. Classification of a subset of the test set corresponding to two sites where **perfluorinated-sulfonate-based non-AFFF formulations** are thought to be the major sources of PFAS present. **Red** = samples correctly classified, **Blue** = samples incorrectly classified. Plots show total log concentrations of perfluorinated sulfonates on the horizontal axis, and perfluorinated carboxylic acids on the vertical axis; zero values cannot be displayed on a log scale, so are shown at the bottom and left of each plot. The line shown is a 1:1 line. Classification based on the Random Forest method.

4.B.4. Identification of Sample Types Not in the Training Set

An inherent challenge with most supervised machine learning classifiers, including those tested for this work, is that they conduct *closed set recognition*, which means that they are only able to recognize patterns contained within training set data. The implication is that if a new pattern is encountered, the resulting classification will not necessarily be predictable. As such, it is necessary to have a training dataset that incorporates all anticipated PFAS compositions and their environmental patterns. Although the preliminary dataset compiled for this work contains data from 22,068 samples with detections, it is likely that substantially more data are needed, especially for specific types of sites likely to be encountered in classification, to ensure that new unknown non-AFFF samples can be distinguished from AFFF samples.

One way to explore the extent to which the patterns observed at a particular site are represented in the dataset is to systematically remove individual sites from the training set, and classify test set data corresponding to the removed site based on a training set based on the remaining sites. When this is done for AFFF samples from military bases around the world, there is virtually no reduction in classification performance (**Figure 10**). The reason for this result is quite simply that the distribution of compositions at the military bases studied is remarkably similar – a fact that further supports the hypothesis of this work that recognizable patterns exist that can be distinguished despite differences in site characteristics. **Figure 11** shows partial component distributions (up to 50 random samples) for a selection of the military bases examined, illustrating the **remarkable similarity across sites**. It is important to emphasize, however, that **this observation does not address the potential problem of false positives, i.e., non-AFFF offsite samples erroneously detected as being of AFFF origin.**

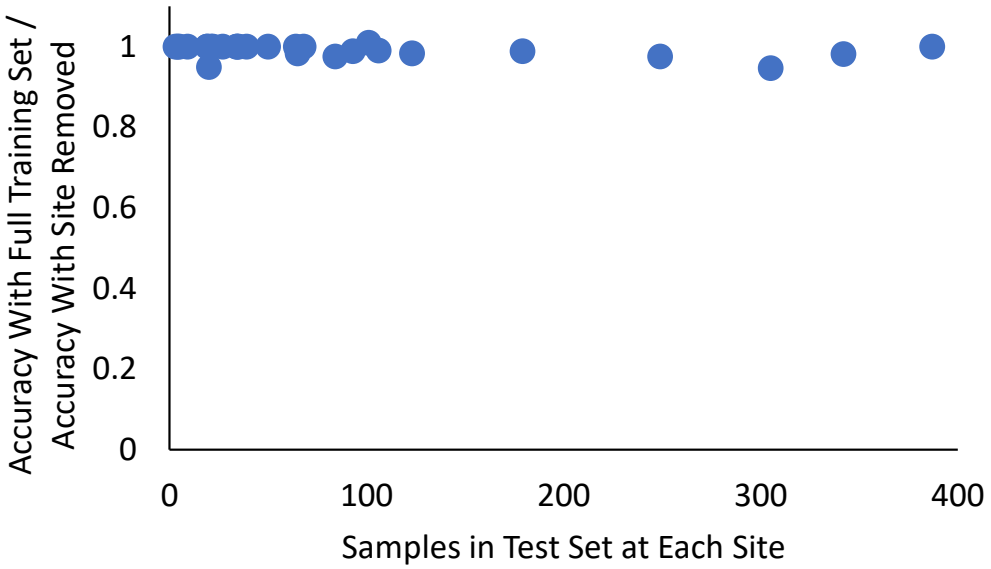


Figure 10. Effect of removing individual site data from training set on classification for AFFF data from military bases. Classification based on the Random Forest classifier.

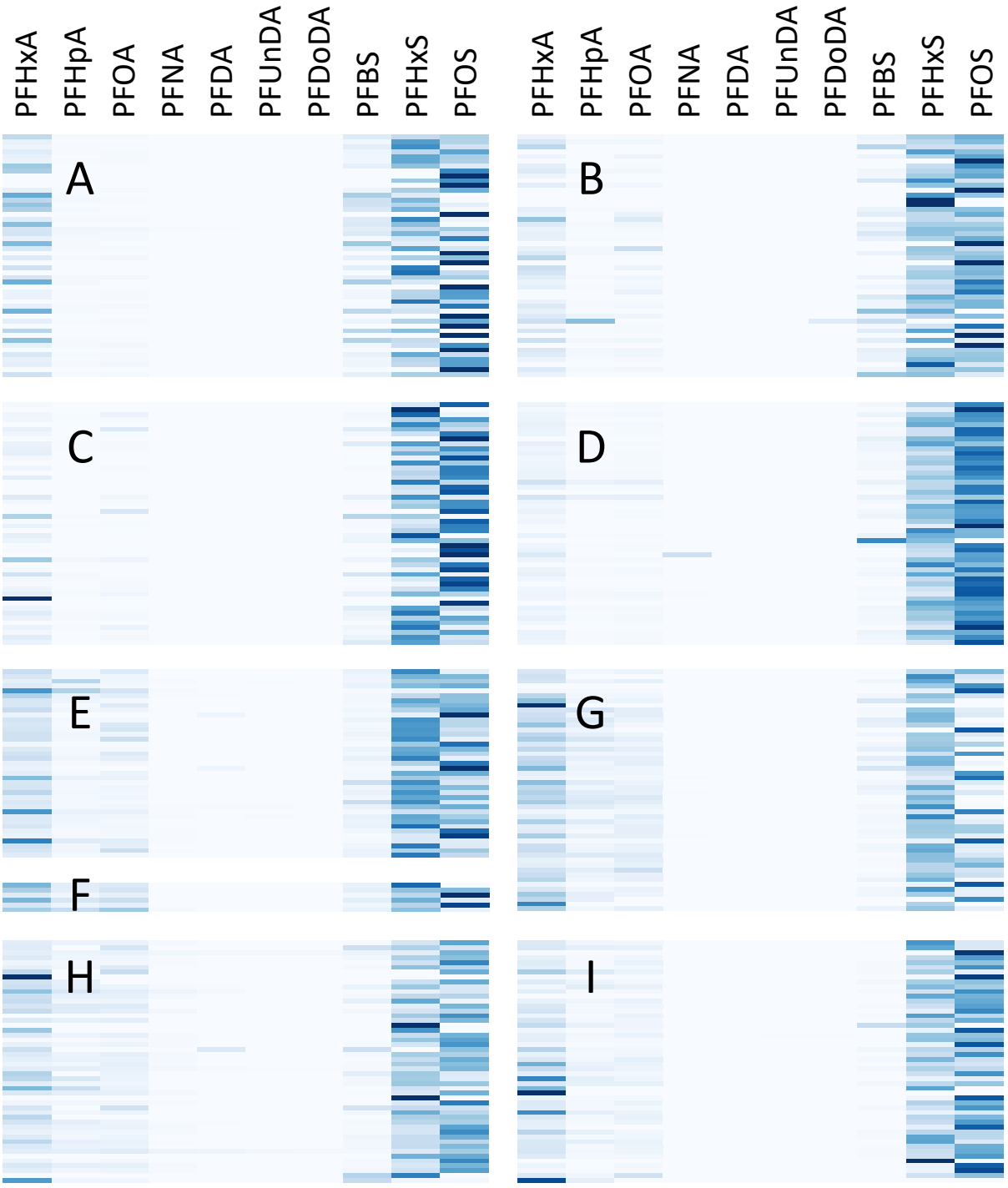


Figure 11. Partial component distributions for 9 individual military bases, showing the similarity of composition between sites. Up to 50 randomly selected samples are shown for each site. Darker bars correspond to higher fractions in each distribution. Sites shown are: A. Albatross (AU), B. Amberly (AU), C. Williamtown (AU), D. Darwin (AU), E. Altus (US), F. Davis-Monthan (US), G. DND Site B (CA), H. DND Site G (CA), I. DND Site I (CA).

Two examples of sites in the dataset for which recognition is not good if all corresponding site data are removed from the training set are Bennington (non-AFFF, coatings), and Du-Wel (non-AFFF metal plating). The distributions at these sites simply look entirely different from all other sites in the dataset, so removing samples from these sites from the training set means the classification algorithms have nothing similar to learn the patterns from. **Figure 12** shows partial distributions of samples at the two sites.

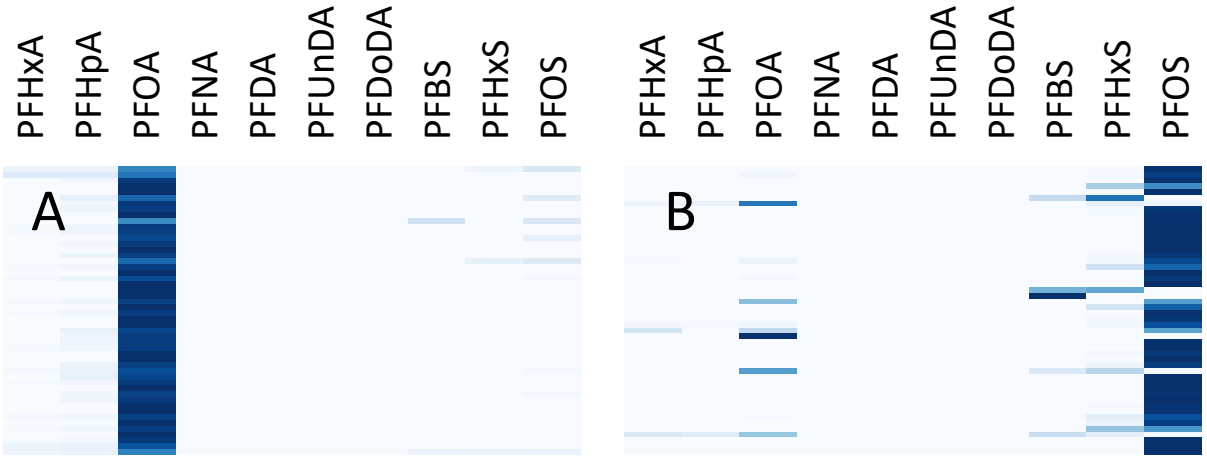


Figure 12. Partial component distributions for two sites that are unlike others in the dataset, so pose challenges for classification if samples for these sites are not included in the training set. Up to 50 randomly selected samples are shown for each site. Darker bars correspond to higher fractions in each distribution. Sites shown are: A. Bennington (US), B. Du-Wel (US).

The Bennington data are dominated by PFOA, while the Du-Wel data are dominated by PFOS. While both sites correspond to non-AFFF PFAS, **if samples from these sites are omitted from the training data, both sites are erroneously identified almost entirely as being of AFFF origin**. In the case of the Bennington and Du-Wel sites, accuracy drops from 97% and 77%, respectively, based on the full training set, to 2.5% and 4.3%, respectively, if samples from these sites are omitted from the training set. This result clearly **highlights the need for a comprehensive training set**, as well as **other strategies to ensure either accurate classification of novel compositions, or rejection of samples that do not look like known samples**. (In this vein, it should be noted that preliminary explorations with synthetic training data were able to increase the accuracy of Du-Wel samples to 82.5% without including any data from the site in the training data, an improvement over even the 77% classification based on the full training set.)

Looking at the entire dataset, results suggest that **sites based on specific industrial non-AFFF sources are likely currently underrepresented**. It is important to emphasize that the presence of sites such as Bennington and Du-Wel in the training set *does* mean that new sites with these same composition distributions will likely be correctly classified based on a model trained on the current dataset. But **the fact that many industrial compositions are represented by only a single site also means that there is risk that some patterns for a specific composition will not be detected**. Furthermore, it is likely that sites representing other industrial PFAS uses are still not represented in the dataset.

4.C. Feature Selection (Subtask 2b)

A critical part of machine learning classification is identifying how many and which features are needed for accurate classification. For this work, this question was explored using a brute force approach, **training a total of 1023 different machine learning classifiers**, one on each of the **1023 possible combinations of components** ($\sum_{k=1}^{10} C_k^{10} = 1023$), and then exploring how the selected components impact classification accuracy.

Figure 13 shows the resulting accuracies for Random Forest classifiers trained on **all 1023 different combinations of the 10 PFAS components** used here. Note that classifiers were trained using the component concentration on a logarithmic scale.

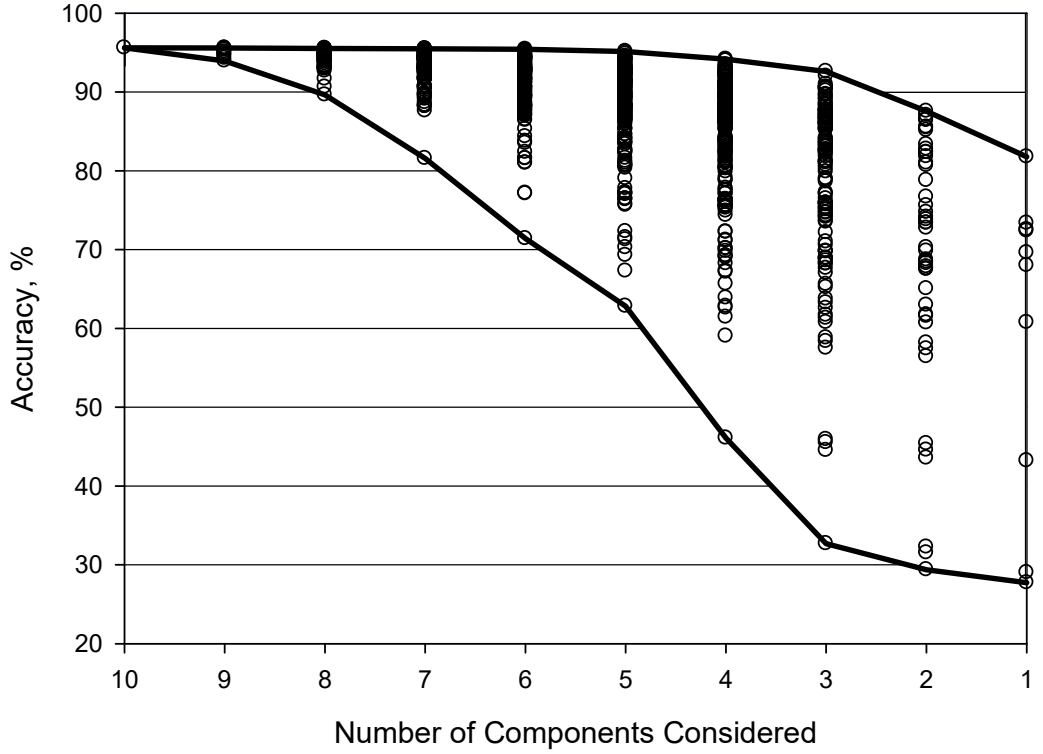


Figure 13. Classification accuracy as a function of number of components selected for the 1023 individual combinations of components in the 10 component set used for analyses in this section. Values shown correspond to Random Forest classifiers. Symbols correspond to the 1023 individual classification results, while lines show the envelope of possible outcomes.

From **Figure 13**, it is apparent that for each number of selected components, there is a range of accuracies. In general, difference between the best and worst accuracies increases as the number of components decreases, suggesting that **even if the best subset of components is not known, having more components is likely to lead to more accurate classification**. (Note that other performance metrics parallel accuracy here, so all trends that apply to **Figure 13** also apply to other performance metrics.) It is also apparent from **Figure 13** that the worst accuracy for a number of components decreases steadily with decreasing numbers of components, while the **best accuracy remains quite high down to 4 components**, and then begins to drop below that.

Tables 14 and **15** show the combinations of components that provide the best and worst accuracies. In each case, shaded blue cells indicate the included components in the trained model that gives the indicated accuracy.

Table 14. Component combinations corresponding to the ***highest*** **classification accuracies** for each total number of components. Selected components are indicated by blue shading.

# Comp.	PFHxA	PFHpA	PFOA	PFNA	PFDA	PFUnDA	PFDoDA	PFBS	PFHxS	PFOS	ACC (%)
10	■	■	■	■	■	■	■	■	■	■	95.6
9	■	■	■	■	■		■	■	■	■	95.6
8	■	■	■	■			■	■	■	■	95.5
7	■	■	■	■				■	■	■	95.5
6	■	■	■					■	■	■	95.4
5	■		■					■	■	■	95.1
4	■		■						■	■	94.2
3		■	■						■		92.6
2									■	■	87.6
1									■		81.8

Table 15. Component combinations corresponding to the ***lowest*** **classification accuracies** for each total number of components. Selected components are indicated by blue shading.

# Comp.	PFHxA	PFHpA	PFOA	PFNA	PFDA	PFUnDA	PFDoDA	PFBS	PFHxS	PFOS	ACC (%)
10	■	■	■	■	■	■	■	■	■	■	95.6
9	■	■		■	■	■	■	■	■	■	93.9
8	■	■		■	■	■	■	■	■		89.6
7	■	■		■	■	■	■	■			81.6
6		■		■	■	■	■	■			71.4
5		■		■	■	■	■				62.8
4				■	■	■	■				46.1
3					■	■	■				32.7
2						■	■				29.4
1							■				27.7

Overall from **Table 14**, it can be seen that the most important component for classification is **PFHxS**, as it is a feature in all of the highest accuracy trained models. Note that even considering the concentration of *only* PFHxS (i.e., a single feature) produces accuracies of 81.5% - slightly better than naïve classification by ratio (79.8%). Furthermore, classification based on only the concentrations of PFHxS and PFOS gives an accuracy of 87.6%, significantly better than classification by ratio.

Figure 14 provides some insight into why the different features are important for classification, showing the number of samples with detected concentrations of each component in the training dataset. Note that the four components with the fewest detects (PFNA, PFDA, PFUnDA, PFDoDA) feature prominently in combinations with the worst accuracies (**Table 15**). This is for the simple reason that it is not possible to distinguish the classes of two samples based on a component that hasn't been detected. I.e., if one AFFF sample and one non-AFFF sample each show no detection of PFDoDA, for example, then looking at PFDoDA concentration is not a useful way to distinguish between the two samples. Similarly, the remaining six most-detected compounds figure prominently in the combinations that provide the best accuracy (**Table 14**). This result highlights the importance of including components that are present above detection limits both in the samples being classified, as well as in other examples of environmental PFAS, to provide the machine learning classifiers sufficient information to allow classes to be distinguished.

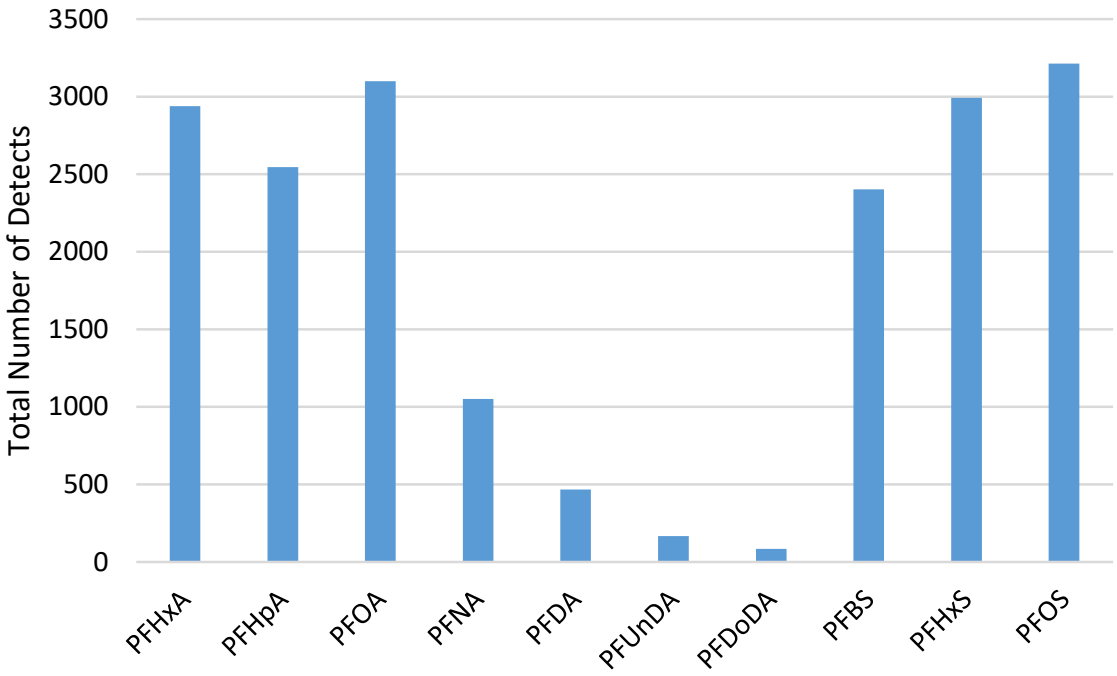


Figure 14. Number of detects for each of the components in the training dataset.

4.D. Clustering (Subtask 2c)

Clustering algorithms, also known as unsupervised learning algorithms, look for patterns in unlabeled data, and as such are often proposed for source delineation. The objective of this subtask was to conduct a preliminary exploration of the use of clustering as an alternate means of distinguishing between AFFF and non-AFFF samples. It is important to note that many clustering algorithms do not allow direct specification of the number of clusters to look for, but rather are intended for use in cases when the number of clusters is unknown. **Table 16** shows accuracy for two commonly used clustering algorithms for which the number of clusters can be specified, the K-Means algorithm, and Gaussian Mixture Models (GMM).

Table 16. The accuracy (%) of clustering at distinguishing between AFFF and non-AFFF in the 10-component groundwater/surface water dataset.

Pre-processing				
Logarithmic			**X**	**X**
Distribution		**X**		**X**
Clustering Method				
GMM	**59.6**	**69.4**	**54.8**	**68.3**
Kmeans	**73.1**	**82.0**	**65.0**	**67.5**

It is important to note that clustering cannot identify which cluster is AFFF and which is non-AFFF, but rather simply separates the data into two clusters, so for purposes of **Table 16**, the accuracy in each case is reported based on the class assignment that gives the best match with the two classes in the data. Overall, it is apparent that the two methods do worse at separating the samples in the dataset into AFFF and non-AFFF categories than all of the supervised learning methods tested, and worse, even, than naïve classification by ratio of perfluorinated sulfonates.

To understand why this is, it is worth exploring spatial behavior of clustering. **Figures 15 and 16** illustrate the two methods from **Table 16** applied to groundwater data from a single site, RAAF Base Williamtown in Australia, a site that is almost certainly dominated by legacy PFOS-based AFFF. Symbol size in the figures corresponds to total PFAS concentration. Note that while the different preprocessing methods and different clustering methods yield different results, it is apparent that clusters are generally separated into groups that follow total concentration, with the higher concentration samples being clustered differently from the lower concentration samples. This suggests that the effects of transport on the individual PFAS components are creating differences in composition that cause transported samples to be identified by the methods into separate clusters. (It should be noted that while slightly different formulations were likely used over the years at the site, the differences are subtle compared with the differences caused by transport.)

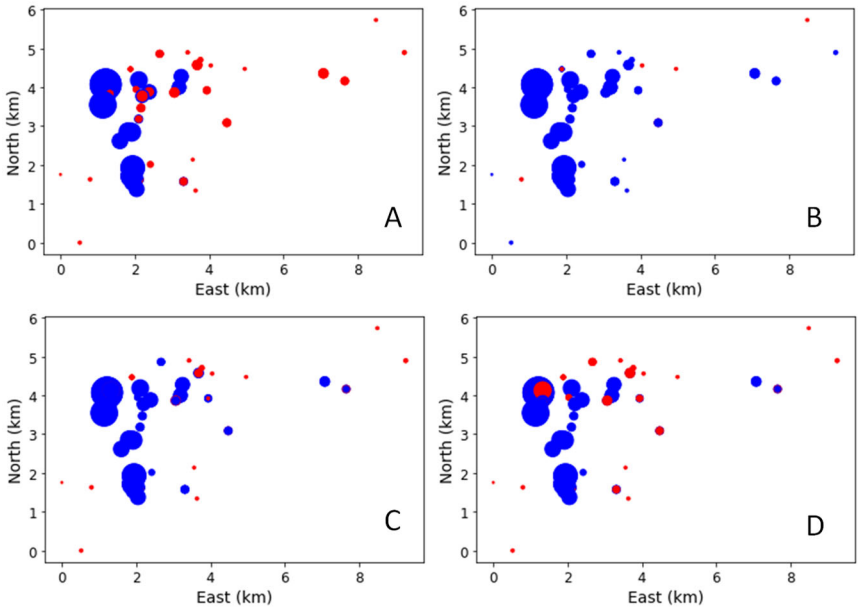


Figure 15. Clustering of the groundwater samples at RAAF Base Williamtown using GMM. Colors indicated the two clusters. Total PFAS concentration is indicated by symbol size.

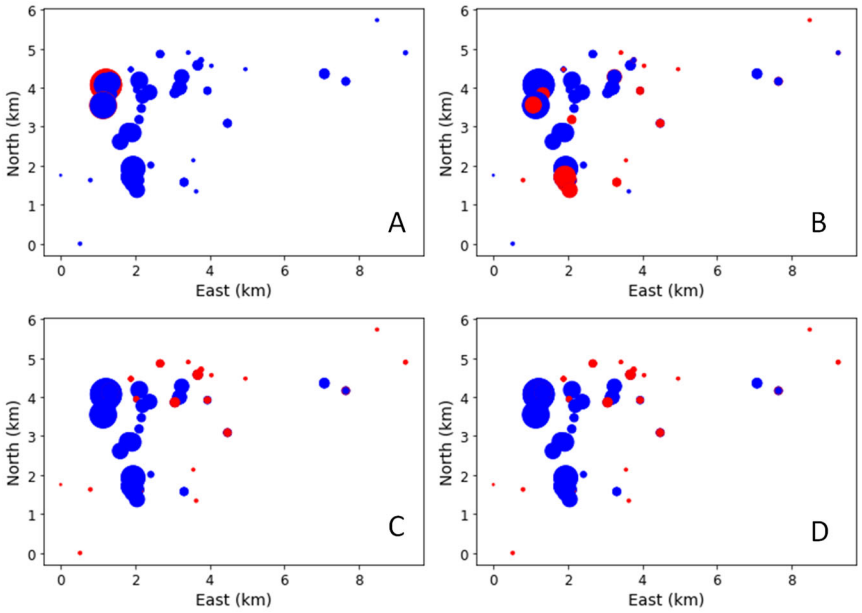


Figure 16. Clustering of the groundwater samples at RAAF Base Williamtown using K-Means. Colors indicated the two clusters. Total PFAS concentration is indicated by symbol size.

Note that the same behavior is observed in simulated transport. **Figures 17 and 18** show simulated transport from a constant source component concentration distribution taken from a military base. The simulation is a based on a finite difference code written for this work based on implicit solution of dispersion coupled with upstream weighted advection. The simulation and others in the remainder of this section tracks transport of the 10 components used in supervised learning (i.e., PFHxA, PFHpA, PFOA, PFNA, PFDA, PFUnDA, PFDoDA, PFBS, PFHxS, PFOS). Linear adsorption is assumed, with constants determined using the method developed for this work, and described in section ***4.E. Preliminary Data Mining***. All simulations in this section are based on parameters similar to those at RAAF Base Williamtown, and simulate ten years of transport from a constant concentration source.

Figure 17 shows the results of GMM clustering of 186 random points taken from the simulated plume, with different colors indicating different assigned clusters. Note that samples close to the source are identified as being in a different cluster from those far from the source. (Note that the shaded contour shown behind the symbols is that of PFBS).

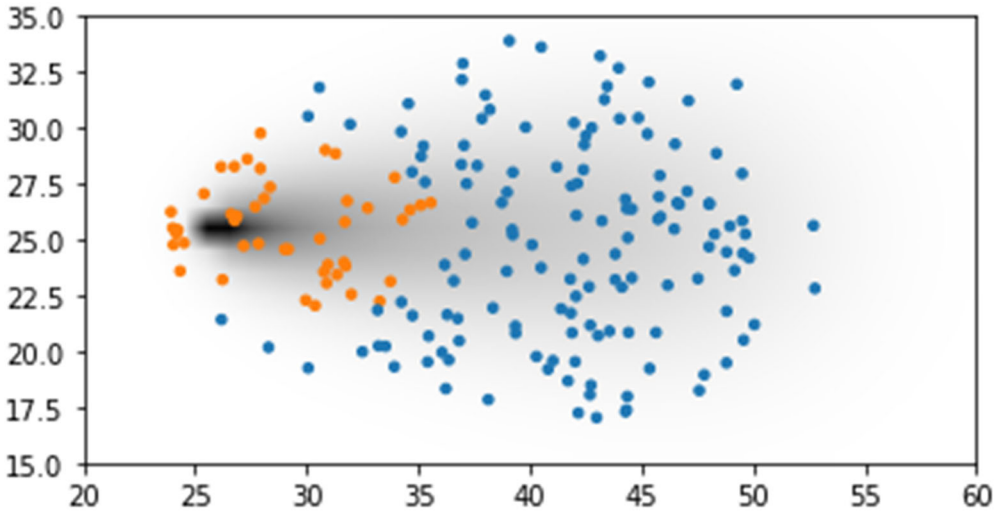


Figure 17. Two component clustering of a 10 component PFAS mixture transported from a single source. Clustering based on GMM. Symbol colors indicate identified clusters.

Figure 18 shows clustering of three different selections of points for the same simulation, highlighting a challenge with clustering, the fact that the **identified clusters depend on which points are selected to cluster**. Note the very different spatial patterns of clustering in the three subfigures, despite the same PFAS composition in all three subfigures. Clustering is comparing samples against others that have been collected, and subtle differences can cause completely different spatial patterns to be observed. The implication for source delineation is that the **sources indicated by clustering could in fact depend strongly on where samples were collected and how many samples were collected**.

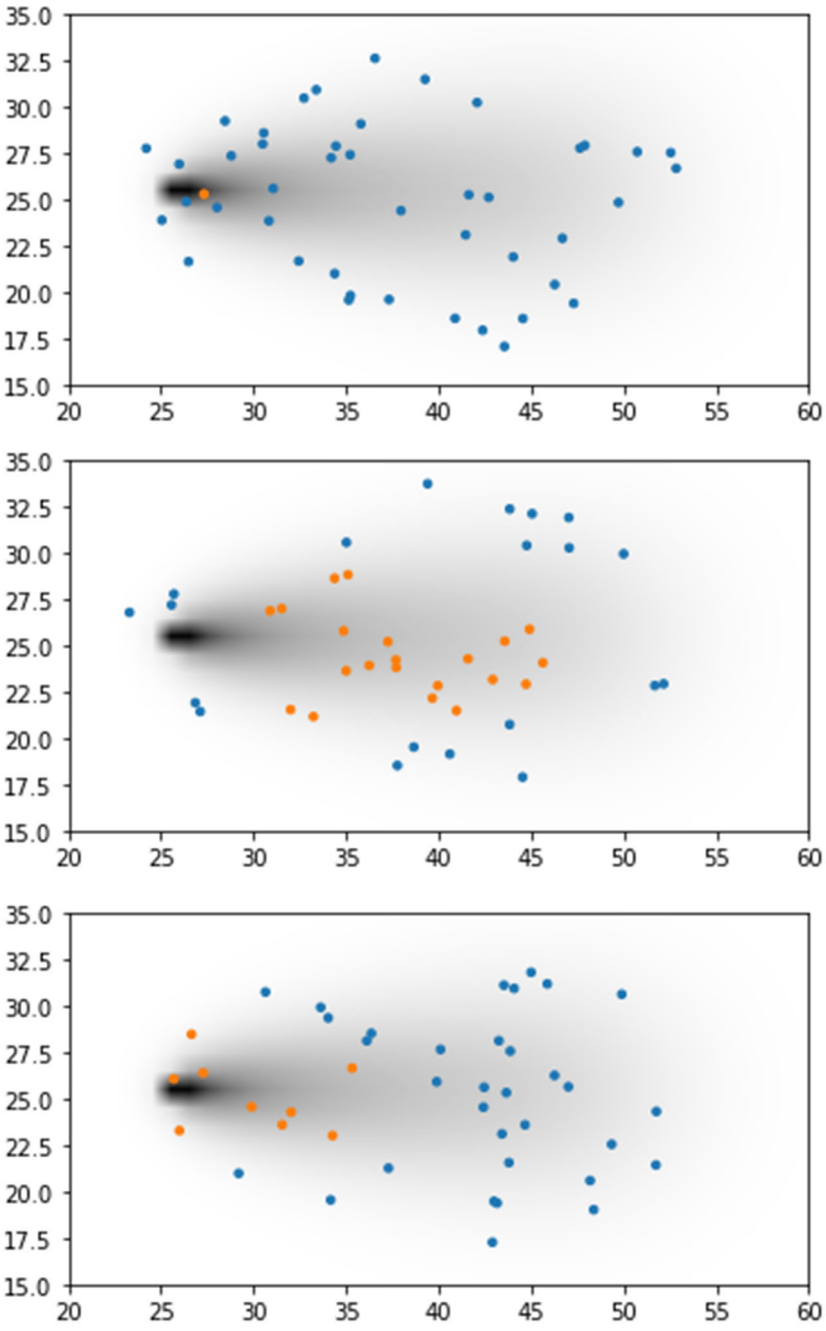


Figure 18. GMM clustering of three different sets of samples from the same simulation shown in **Fig. 17**, indicating that the compositions of the included points influence clustering, producing completely different spatial patterns.

This result is further apparent in **Figure 19**, which shows **simulated transport from two distinct sources**. The lower source in the figure is the same AFFF composition from **Figures 17 and 18**, while the upper source is taken from landfill leachate data, and is completely different in composition from the AFFF. As before, transport is simulated over a 10-year period. Note again in **Figure 19** how the clustering assignments are completely different for the different selections of points. **Fig. 19A** does appear to be identifying the non-AFFF source as being separate from the AFFF source, but large regions at low concentration around both sources are identified as being consistent with the AFFF source. In **Fig. 19B**, a large selection of points out from the two sources are identified as belonging to one cluster, while the lower concentrations farther from the sources are identified as belonging to another. The different results are entirely due to the compositions of the different randomly-selected points.

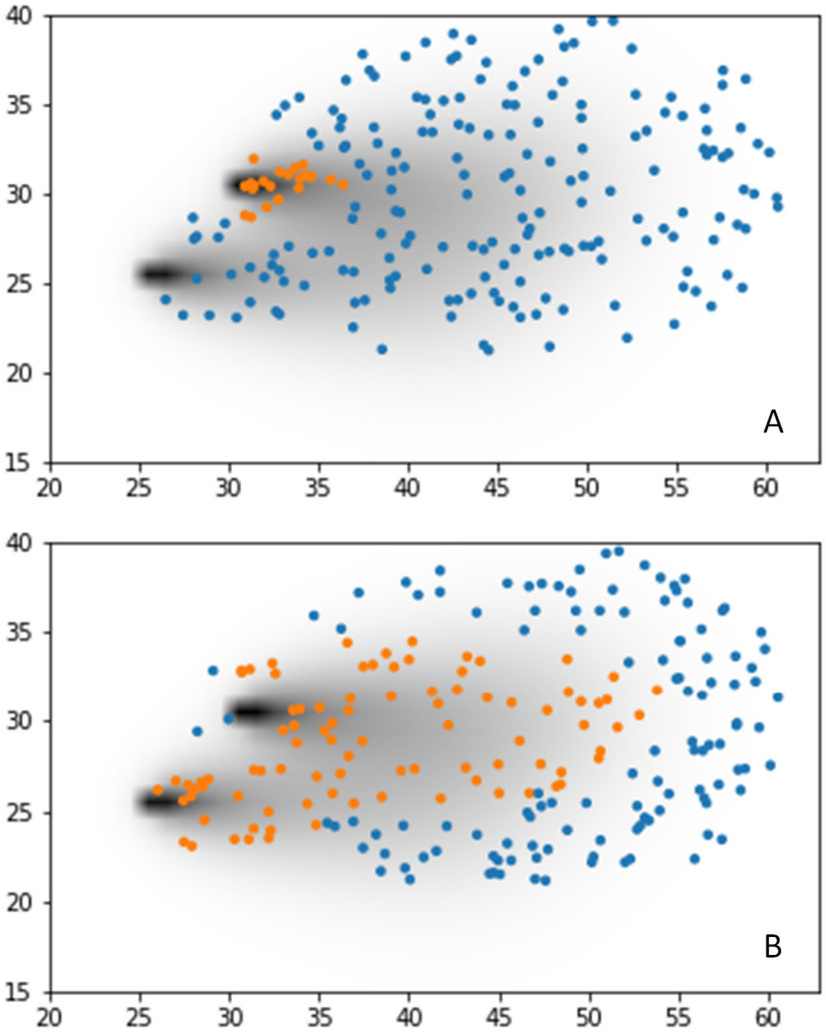


Figure 19. GMM Clustering of randomly selected points in the simulation of transport from two distinct PFAS sources, one AFFF (lower) and one non-AFFF. Differences in clustering are entirely due to the subset of points selected for clustering.

Finally, **Figure 20** illustrates clustering of the same two-source simulation using two different algorithms that identify the number of clusters present, Mean-shift clustering, and Spectral clustering. In this case, Mean-shift clustering identifies three clusters, while Spectral clustering identifies ten, but **in both cases, the identified clusters are not consistent with the two distinct known sources**. (It should be noted that the number of clusters can be influenced in both of these algorithms by adjusting algorithm parameters.)

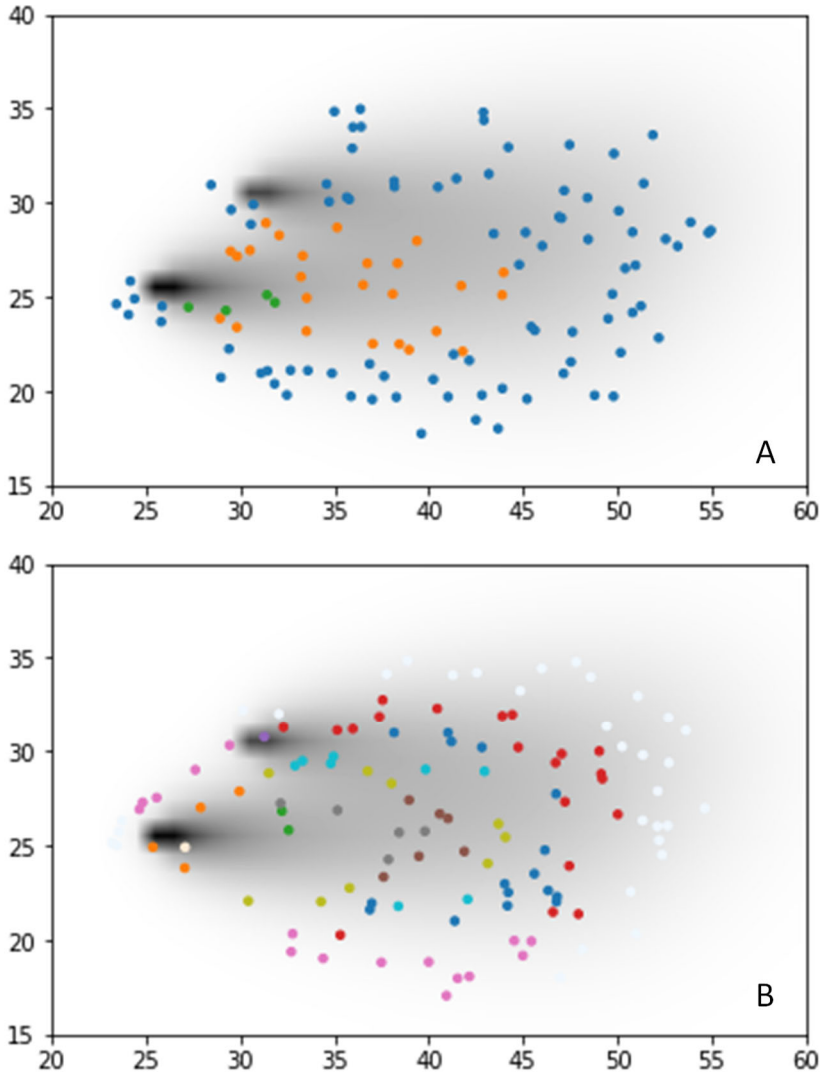


Figure 20. Clustering of randomly selected points in the simulation of transport from two distinct PFAS sources, one AFFF (lower) and one non-AFFF. A. Mean-shift clustering, and B. Spectral clustering.

Overall, it is important to emphasize that the fact that clustering does not distinguish between AFFF and non-AFFF sources in the context of this work is not because of a problem with the algorithms, but rather simply because of the fact that the differences created by transport of PFAS components create patterns that can differ more than the differences between the sources themselves. Although it is likely there are some applications of clustering for source delineation, particularly for cases where very different source compositions are present, extreme care must be exercised interpreting the results. Based on the preliminary work conducted here, it is unlikely that clustering (unsupervised learning) would be satisfactory for more general PFAS source allocation.

4.E. Preliminary Data Mining (Subtask 3a)

The work in this subtask focused on exploring the extent to which information about physicochemical phenomena could be uncovered from the dataset and machine learning results. The main emphasis of the work conducted was to use a hybrid method, combining machine learning with mathematical transformations based on a quantitative understanding of the behaviors of surfactants such as PFAS, to relate PFAS compositions in different phases.

4.E.1. Prediction of PFAS Surfactant Parameters Driven by Hydrophobicity

Work partially supported by this project was conducted to develop a group contribution model for estimation of individual PFAS physicochemical parameters describing interfacial and partitioning phenomena driven by hydrophobicity. The work was recently accepted for publication in the paper:

> **S.T. Le, T. C. G. Kibbey, K. P. Weber, W. C. Glamore, and D. M. O'Carroll. A Group-Contribution Model for Predicting the Physicochemical Behavior of PFAS Components for Understanding Environmental Fate, *Science of the Total Environment*, 142882. [2]**.

This section describes some of the key components of that work as they relate to this project. In short, the group contribution model described in the paper uses linear combinations of constants corresponding to different functional groups in individual PFAS molecules to fit measured interfacial constants for air-water interfacial adsorption, and then normalized scaling relationships are used to predict other hydrophobicity-driven physicochemical parameters.

Table 17 shows constants for determining the asymptotic low concentration Szyszkowski interfacial partition coefficient for air-water adsorption. (The paper also includes constants for calculating standard free energy of adsorption.) The constants were determined from experimental data from 46 PFAS and 43 hydrocarbon surfactants.

Table 17. log K_{iw-0} coefficients determined from fits to surface tension data. Coefficients correspond K_{iw-0} in units of cm. Constants are from ref.[2].

Functional Group	*log K_{iw-0} Coefficients*
CFx	0.60
CHx	0.36
COO^-	-2.42
COOH	-0.47
$SO3^-$	-2.35
$R4N^+$	-4.30
OH	-0.79
$OSO3^-$	-2.39
-O-	-0.41
-S-	-0.21
N(CH3)2–CH2–COO	-1.07
Intercept	-5.19

Calculation of log $K_{iw\text{-}0}$ from the coefficients in **Table 17** is done with **Eq. 1**, where log $K_{iw\text{-}0}(X)$ is the coefficient for component X in **Table 17**, and n_x is the number of group X in the PFAS component of interest.

$$\mathbf{\log K_{iw-0} = \log K_{iw-0}(A).\, n_A + \log K_{iw-0}(B).\, n_B + \cdots + \log K_{iw-0}(N).\, n_N + Constant} \quad (1)$$

Strictly speaking, $K_{iw\text{-}0}$ calculated with **Eq. 1** is the theoretical low concentration asymptote for air water adsorption ($K_{iw-0} = \lim_{C_w \to 0} \frac{\Gamma}{C_w}$, where Γ is surface excess, and C_w is water concentration) for a no-salt added condition, and assuming a Langmuir adsorption isotherm. Note that issues exist with both conditions, specifically that ionic surfactants (like many of the most widely-detected PFAS) are strongly influenced by ionic strength, and that the Langmuir equation is likely not accurate for the no-salt-added condition, as a result of the required ion pairing. However, **the purpose of the model is not necessarily to predict air-water interfacial adsorption, but rather to use the calculated $K_{iw\text{-}0}$ values as indicators of PFAS hydrophobicity for prediction of other behaviors driven by hydrophobicity**. This approach is based on the recognition that anything that influences hydrophobicity in surfactants, such as PFAS, tends to produce a parallel influence on all hydrophobicity-driven phenomena.

This is illustrated in **Figure 21**, which shows PFOS-normalized properties (log CMC, log solubility, log K_D, log K_{oc}, log K_{ow}) plotted against PFOS-normalized calculated log $K_{iw\text{-}0}$ values.

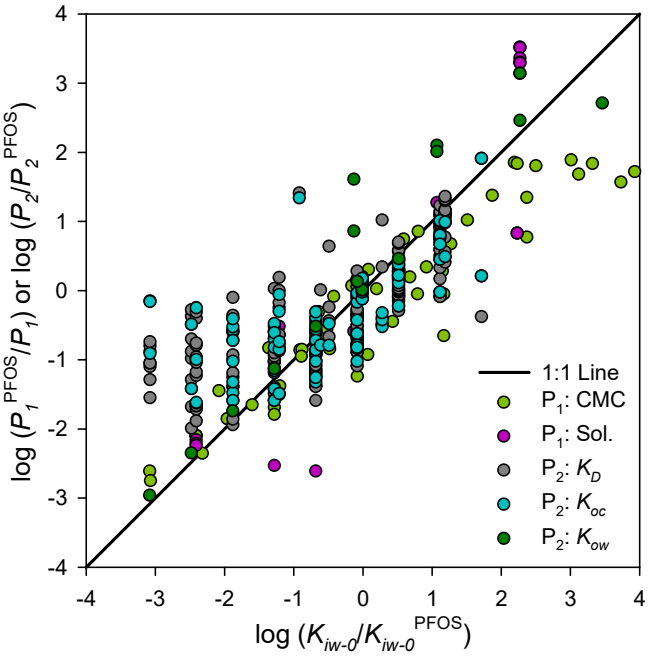


Figure 21. Illustration of PFOS-normalized relationships. Note that data are plotted differently depending on whether the property decreases with hydrophobicity (Property Group 1: *CMC*, Solubility) or increases with hydrophobicity (Property Group 2: K_D, K_{OC}, K_{OW}). Figure is reproduced from ref. [2].

It is important to emphasize that the relationship in **Figure 21** is not a fit, but rather an illustration of how surfactant phenomena that are driven by hydrophobicity tend to move in parallel for surfactants, so things that influence one (either structural factors, or environmental factors) tend to influence them all in a predictable way. Note that the line in **Fig. 21** is **a 1:1 line**.

To use the information in **Fig. 21** to predict distribution between phases, **Eq. 2** can be used:

$$\boldsymbol{K_X^i = K_X^{PFOS} \frac{K_{iw-0}^i}{K_{iw-0}^{PFOS}}} \quad (2)$$

where K_X^i refers to any distribution constant for equilibrium with water, and *i* corresponds to an individual PFAS component. Note that the selection of PFOS for normalization is arbitrary (any component could be used), but convenient for calculations involving PFAS, where historical constants at a site may have been determined for only a few specific components, such as PFOS. Note that **Equation 2** was used to determine K_D values for all of the compounds in the 10-component formulation for the simulations in Sec. *4.D. Clustering* based on a K_D^{PFOS} of 1.0 L/kg, a value consistent with reported measurements at RAAF Base Williamtown.

It should be emphasized that **Eq. 2** is only appropriate in cases where hydrophobicity is the primary driver of phase distribution; as such, it is generally not appropriate for determining soil adsorption coefficients for cationic or zwitterionic PFAS components, where charge-charge interactions with negatively-charged soil are expected to be important, but should provide a source of realistic predicted constants for anionic and nonionic PFAS components.

4.E.2. The Effect of Distribution On PFAS Component Fingerprint

In the context of this work, **Eq. 2** and the constants in the preceding section are of interest for what they can tell us about how distribution between phases impacts the fractions of PFAS in different phases in equilibrium. Of particular interest is how these relationships can be used to explore whether partitioning is a dominant mechanism relating the PFAS compositional signatures in different phases, and the extent to which that information can be used for machine learning classification.

Equations 3-6 illustrate the calculation of the mass fraction of PFAS component *i* in a sample in non-water phase *X* ($\boldsymbol{\alpha_X^i}$). Note that *X* could be any phase in equilibrium with water where the equilibrium can be described by a distribution coefficient whose magnitude is dictated by hydrophobicity. The mass fraction $\boldsymbol{\alpha_X^i}$ is given by Eq. 3:

$$\boldsymbol{\alpha_X^i = \frac{m_X^i}{\sum_j m_X^j}} \quad (3)$$

where m_X^i is the mass of component *i* in phase X. If the concentration in phase X is given by C_X^i and the volume of phase X is given by V_X, then:

$$\boldsymbol{\alpha_X^i = \frac{C_X^i \times V_X}{\sum_j (C_X^j \times V_X)}} \quad (4)$$

Note that this notation is appropriate for a partitioning relationship, but for sorption C_X^i would need to be replaced by surface excess or a measure of sorption density, and V_X by area or

phase mass, depending on the units of sorption. Dividing through by V_X and substituting the relationship between C_X^i and C_w^i ($C_X^i = K_X^i \times C_w^i$) gives **Eq. 5**:

$$\boldsymbol{\alpha_X^i = \frac{K_X^i \times C_w^i}{\sum_j (K_X^j \times C_w^j)}} \quad (5)$$

Finally, substituting **Eq. 2** into **Eq. 5** for K_X^i gives **Eq. 6**, which can be used with the constants in **Table 17** to calculate the mass fractions of components in any phase in equilibrium with water:

$$\boldsymbol{\alpha_X^i = \frac{K_{iw-0}^i \times C_w^i}{\sum_j (K_{iw-0}^j \times C_w^j)}} \quad (6)$$

Of particular interest from **Eq. 6** is the fact that ***the mass fraction in any given phase is independent of the actual affinity for that phase***, assuming the observation in **Figure 21** remains valid. The implication of that is that for a given PFAS composition in water, the **composition in all phases in equilibrium with water should be the same**. This suggests a way of exploring whether a phase is, in fact, in equilibrium with an aqueous PFAS sample, or whether some other mechanisms are responsible for the observed composition, such as transformations. Furthermore, **Eq. 6** suggests an approach that can be used for machine learning to make classifications of samples in different phases by using **Eq. 6** to mathematically transform training data to train a classifier for non-water phases based on water data.

Figures 22-26 show transformed groundwater data for several sites for which both water and non-water data are available. Note that the distributions of 50 randomly-selected samples are shown in plots, except in cases where fewer than 50 samples are available. **Figures 22** and **23** correspond to legacy PFOS-based AFFF at military bases, while **Figure 24** corresponds to an AFFF fire extinguishing event at a commercial airport where non-legacy AFFF was used. **Figures 25** and **26** correspond to non-AFFF sources, both where a single component dominates the formulation.

From the figures, it is apparent that the **transformed water data look very much like soil, sediment and biota data at most of the sites**, supporting the approach described here. One notable exception is the **concrete data** at the Brisbane Airport (**Figure 24**), which appears to be dominated by PFHxA, unlike most of the other phases at that site. A possible explanation for this is the PFAS associated with the concrete samples were likely never in equilibrium with water, but rather resulted from direct contact with AFFF material (e.g., a spill onto the concrete). Another notable exception is the Bennington site, where the soil data appears to have observable quantities of other compounds not detected in the water. It should be noted that the Bennington soil data are shallow residential garden soils, so were likely contaminated by atmospheric deposition. While their contamination does appear to be dominated by PFOA, consistent with water samples from the local source, other contributions are likely present as well.

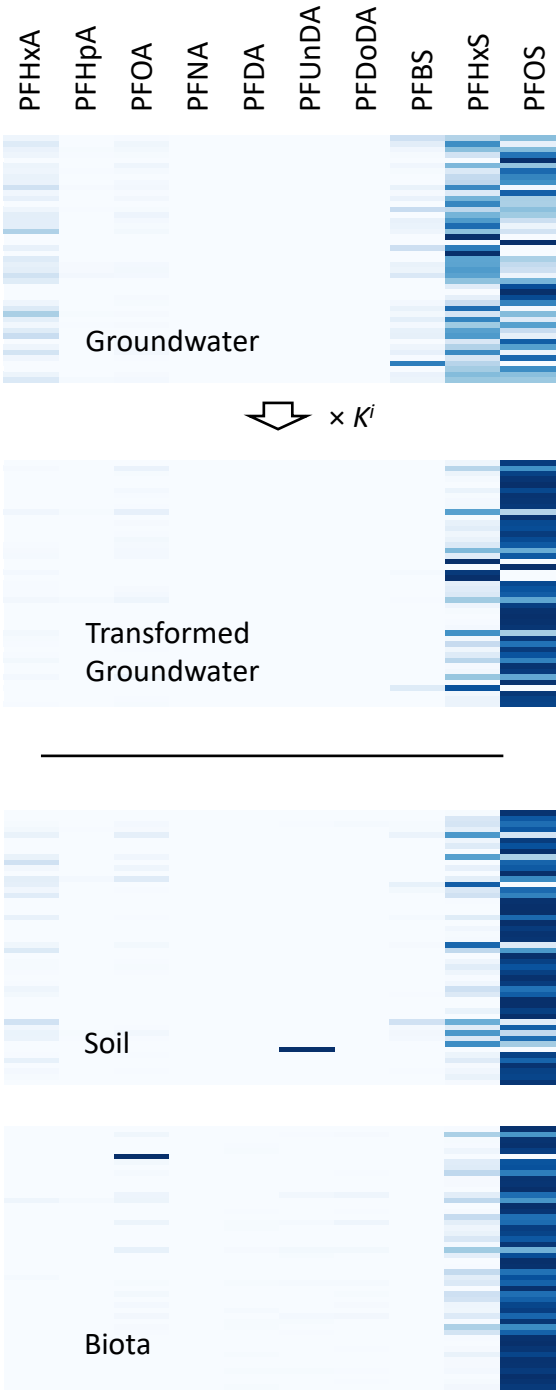


Figure 22. Comparison between transformed groundwater composition data and data for other phases for RAAF Base Pearce (Australia, AFFF). Darker bars correspond to a higher fraction present in each sample.

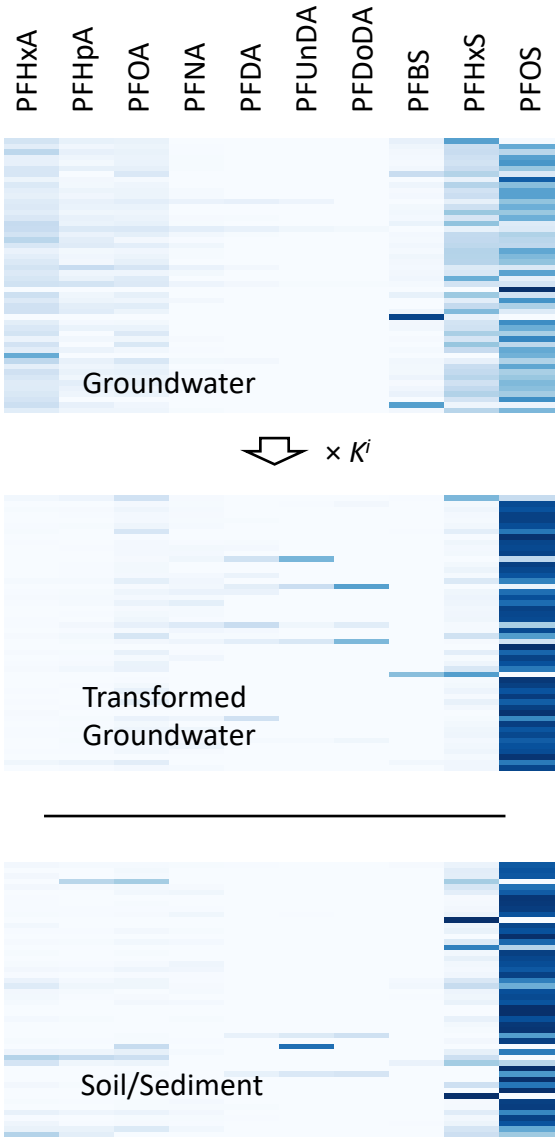


Figure 23. Comparison between transformed groundwater composition data and data for other phases for DND Site G (Canada, AFFF). Darker bars correspond to a higher fraction present in each sample.

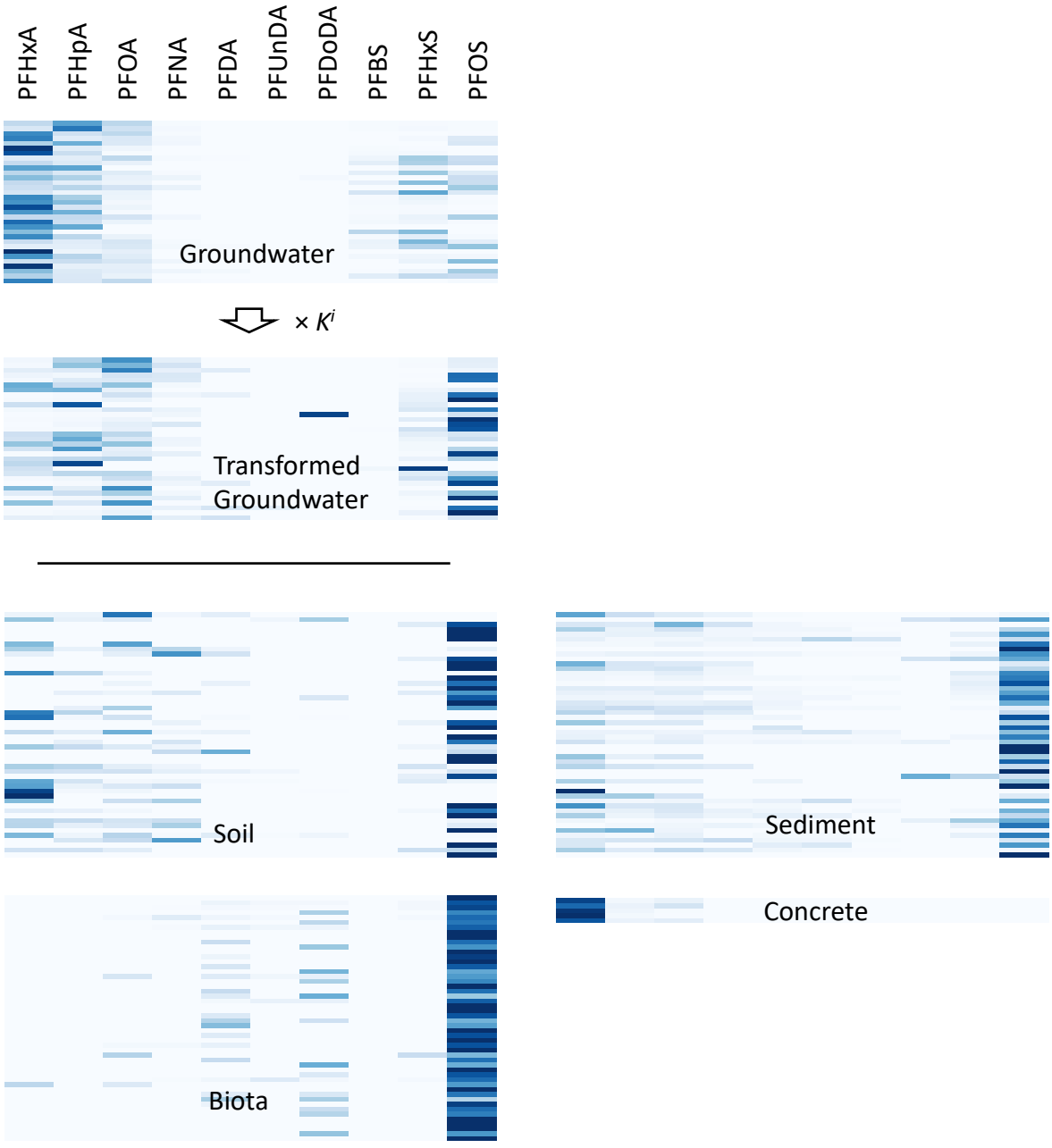


Figure 24. Comparison between transformed groundwater composition data and data for other phases for Brisbane Airport (Australia, AFFF). Darker bars correspond to a higher fraction present in each sample.

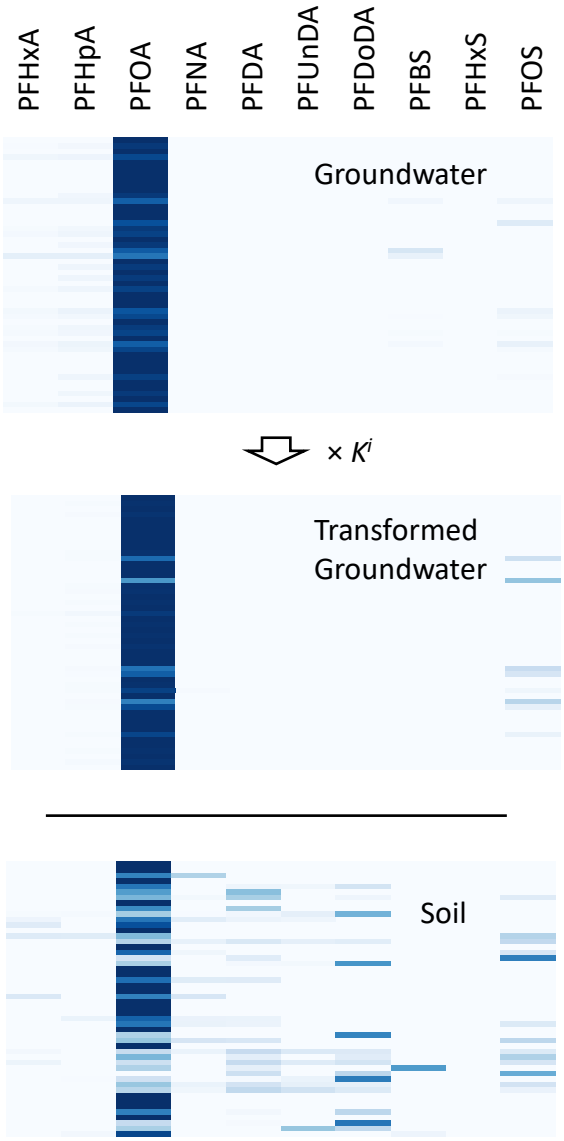


Figure 25. Comparison between transformed groundwater composition data and data for other phases for Bennington (USA, non-AFFF). Darker bars correspond to a higher fraction present in each sample.

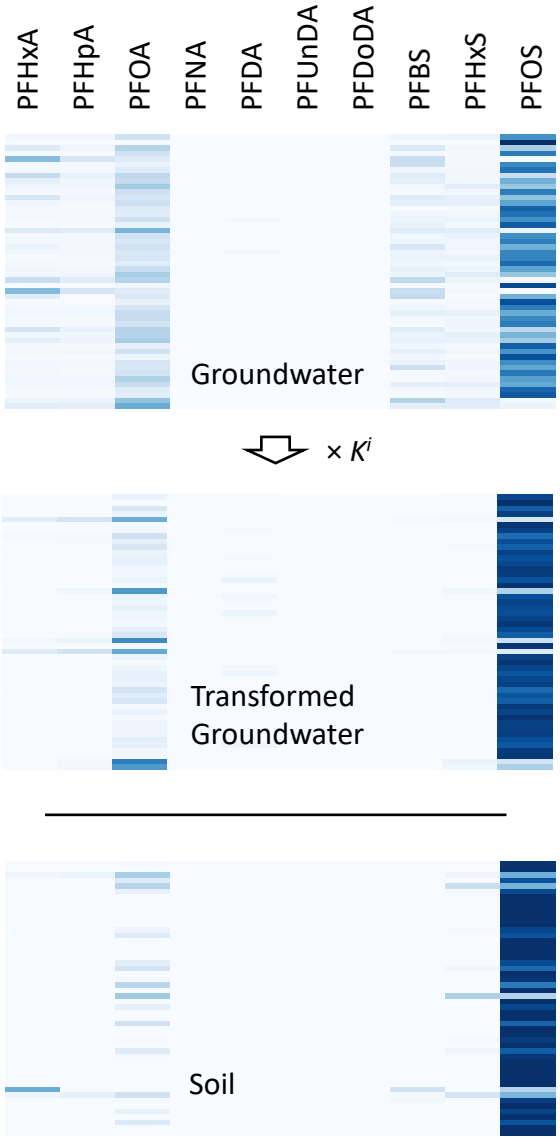


Figure 26. Comparison between transformed groundwater composition data and data for other phases for Wolverine Tannery Site (USA, non-AFFF). Darker bars correspond to a higher fraction present in each sample.

4.E.3. Machine Learning for non-Water Phases

Equation 6 was used to transform the water data in the 10-component dataset used for supervised machine learning (***Sec. 4.B*** of this report), and then a new Random Forest classifier was trained on the transformed water data. Both the original and transformed Random Forest classifiers were used to classify all available data from other phases at sites corresponding water data in the training set. Results are shown in **Tables 18-21**.

For **soils data** (**Table 18**), it is apparent that the original, water phase, Random Forest model does a reasonably good job of classifying soil samples, particularly for the AFFF data associated with military bases. However, the model based on the transformed water data improves classification in soils in all but one case in **Table 18**. Overall, classification of the AFFF samples is improved to the greatest extent. The two non-AFFF data sources – the Bennington Coatings data, and the Wolverine Tannery data – provide especially difficult classification problems for the samples analyzed, in part because both contain very few detected compounds for classification. In the case of Wolverine, the average number of compounds detected in the soil samples is 1.38, meaning that most of the samples are dominated by a single compound. This is in part simply because of the nature of the contamination present and the fact that PFAS is detected in the soil samples at low concentration, so few compounds are above detection limits, but it is also partly the result of the amplification of hydrophobic compounds by adsorption (**Eq. 6**). In the case of the Wolverine data, this causes PFOS to be even more dominant, resulting in poor classification of the soil samples. In the case of the Bennington data, classification is better (74.8 % in the transformed model), but not as good as classification of water data from the same source (97.1% from the analyses of water samples in **Sec. 4.B**). The reason for this could be in part that the samples come from shallow soils so are likely not in equilibrium with groundwater, but likely include contributions from atmospheric deposition from other sources, consistent with the other compounds detected in the soil samples compared with the water samples from the same data source (**Figure 25**).

For the **sediment samples** (**Table 19**), classification of all samples is very good overall with both models, so there is little room for improvement with the transformed model, although all sediment classification are either the same or improved by the transformation.

The greatest impact of the transformation is apparent in the **biota data** (**Table 20**).The three sites for which biota data are available have more than 500 separate samples. Samples include a wide range of plant and animal data, including fish, worms, insects, and a range of vertebrate land animal samples. Overall the transformation yields the greatest classification improvement for the military base data, with smaller improvement for AFFF data from a commercial airport.

Finally, **Table 21** shows concrete data for two sites. In both cases both the original and transformed models classify the concrete data with 100% accuracy. This is likely due to the fact that the concrete contamination likely resulted from spills onto the concrete rather than equilibrium with water, so distributions look more like the original formulation than the transformed water data.

Table 18. Classification of soil samples using the transformed model, compared with the original (water) model.

Site	Country	Site type	Source	Sample Type	Tot. Samples	Detected as AFFF		% Accuracy		% Improved
						Orig. Model	Trans. Model	Orig. Model	Trans. Model	
Albatross	AU	Military Base	AFFF	Soil	144	127	142	88.2	98.6	10.4
Altus	US	Military Base	AFFF	Soil	94	91	92	96.8	97.9	1.1
Clarendon	US	Airport	AFFF	Soil	10	7	8	70.0	80.0	10.0
Davis-Monthan	US	Military Base	AFFF	Soil	10	9	10	90.0	100.0	10.0
DND E	CA	Military Base	AFFF	Soil	52	46	52	88.5	100.0	11.5
DND G	CA	Military Base	AFFF	Soil	302	281	289	93.0	95.7	2.6
DND I	CA	Military Base	AFFF	Soil	146	122	144	83.6	98.6	15.1
Jervis Bay	AU	Military Base	AFFF	Soil	245	233	245	95.1	100.0	4.9
Marinette	US	AFFF Mfg.	AFFF	Soil	18	18	18	100.0	100.0	0.0
Oakey	AU	Military Base	AFFF	Soil	63	62	63	98.4	100.0	1.6
PDX	US	Airport	AFFF	Soil	4	2	4	50.0	100.0	50.0
Pearce	AU	Military Base	AFFF	Soil	267	218	265	81.6	99.3	17.6
Brisbane Airport	AU	Airport	AFFF	Soil	74	59	67	79.7	90.5	10.8
Stirling	AU	Military Base	AFFF	Soil	141	114	133	80.9	94.3	13.5
Bennington	US	Coatings	non-AFFF	Soil	135	38	34	71.9	74.8	3.0
Wolverine	US	Tannery	non-AFFF	Soil	459	413	432	10.0	5.9	-4.1
									Avg. Improvement (%):	9.9

Table 19. Classification of sediment samples using the transformed model, compared with the original (water) model.

Site	Country	Site type	Source	Sample Type	Tot. Samples	Detected as AFFF		% Accuracy		% Improved
						Orig. Model	Trans. Model	Orig. Model	Trans. Model	
Albatross	AU	Military Base	AFFF	Sediment	23	20	23	87.0	100.0	13.0
Altus	US	Military Base	AFFF	Sediment	3	3	3	100.0	100.0	0.0
Davis-Monthan	US	Military Base	AFFF	Sediment	3	3	3	100.0	100.0	0.0
DND E	CA	Military Base	AFFF	Sediment	4	4	4	100.0	100.0	0.0
DND G	CA	Military Base	AFFF	Sediment	57	57	57	100.0	100.0	0.0
DND I	CA	Military Base	AFFF	Sediment	13	12	12	92.3	92.3	0.0
Jervis Bay	AU	Military Base	AFFF	Sediment	71	67	71	94.4	100.0	5.6
Brisbane Airport	AU	Airport	AFFF	Sediment	67	65	65	97.0	97.0	0.0
Stirling	AU	Military Base	AFFF	Sediment	15	15	15	100.0	100.0	0.0
Williamtown	AU	Military Base	AFFF	Sediment	99	88	99	88.9	100.0	11.1
									Avg. Improvement (%):	3.0

Table 20. Classification of biota samples using the transformed model, compared with the original (water) model.

Site	Country	Site type	Source	Sample Type	Tot. Samples	Detected as AFFF		% Accuracy		% Improved
						Orig. Model	Trans. Model	Orig. Model	Trans. Model	
Jervis Bay	AU	Military Base	AFFF	Biota	358	304	358	84.9	100.0	15.1
Pearce	AU	Military Base	AFFF	Biota	48	29	47	60.4	97.9	37.5
Brisbane Airport	AU	Airport	AFFF	Biota	116	108	109	93.1	94.0	0.9
									Avg. Improvement (%):	17.8

Table 21. Classification of concrete samples using the transformed model, compared with the original (water) model.

Site	Country	Site type	Source	Sample Type	Tot. Samples	Detected as AFFF		% Accuracy		% Improved
						Orig. Model	Trans. Model	Orig. Model	Trans. Model	
Jervis Bay	AU	Military Base	AFFF	Concrete	14	14	14	100.0	100.0	0.0
Brisbane Airport	AU	Airport	AFFF	Concrete	5	5	5	100.0	100.0	0.0
									Avg. Improvement (%):	0.0

5. CONCLUSIONS AND IMPLICATIONS FOR FUTURE RESEARCH

5.A. Conclusions

Results of this one-year proof-of-concept project strongly support the idea that supervised machine learning based on PFAS composition can form the basis of a tool be used for source allocation. The approach is likely to prove complementary to approaches based on advanced analytical techniques or other methods, and could be extremely valuable for initial forensic assessments and allocation of site investigation resources. Results of the work come from the following four major areas:

5.A.1. Supervised Learning

The primary objective of the work was to conduct a preliminary study of supervised learning for source allocation. A total of **12 different supervised learning classifiers were tested** for their ability to distinguish between AFFF and non-AFFF PFAS in environmental samples on the basis of composition. Of the methods, variations of the Random Forest algorithm exhibited the highest performance, followed closely by Deep Neural Network ensembles. Other promising methods include the K-neighbors classifier (one of the simplest classifiers tested), and Support Vector Machines with the RBF kernel. **All tested classifiers performed significantly better than naïve classification by ratio of perfluorinated sulfonates.** Different pre-processing methods were found to yield different classification performance, with classification of data on a log scale providing greater performance than on a linear scale, and classification of distributions providing greater performance than raw concentrations.

Classification performance was explored on difficult subsets, specifically AFFF sources that have compositions that are reminiscent of non-AFFF, and non-AFFF sources that have compositions that are reminiscent of AFFF. Overall, even these difficult subsets are accurately identified, provided representative training data are used to train the classifiers. One challenge with the supervised learning methods used here, however, is that they can only accurately classify samples containing patterns they have seen before in training data. **As such, a comprehensive training dataset is essential, as are strategies for rejecting poor matches, both areas where future work is needed.**

5.A.2. Feature Selection

Supervised machine learning requires that all samples used for training and testing have the same number of features (in this case, PFAS components); however, the fact that historical data from different sources have analyzed for different subsets of components mean that increasing the number of components selected as features necessarily decreases the availability of training data. As such, strategic selection of features is necessary to maximize classification performance. For this work, all 1023 possible component combinations of the 10-component dataset used for supervised learning were extracted and used to train separate individual machine learning classifiers, which were then used to evaluate classification of a separate test set. For each number of components included there was a range of classification accuracies, and for larger numbers of components, the range was smaller. The best classification performance dropped gradually down to 4 components, with a 3 component classification still capable of exceeding 90% accuracy on the test set. The worst classification performance dropped more steeply with decreasing numbers of components, however, with the worst-case accuracies in the 30% range for 3 components. This result highlights the facts that the number of components

selected is important to classification performance, with more components having a greater chance of yielding better performance, but that a small subset of carefully-selected components can still produce high classification performance, in some cases higher performance than a poorly-selected larger subset of components. In general, a major factor in determining the classification performance was the number of samples for which components were above detection limits. Components which were present above detection limits in a relatively small number of samples (e.g., PFUnDA, PFDoDA) also contributed the least to classification performance. For this dataset, PFHxS was the single most important component to include for classification.

5.A.3. Clustering

Clustering algorithms, also known as unsupervised learning algorithms, look for patterns in unlabeled data, and as such are often proposed for source delineation. The objective of this subtask was to conduct a preliminary exploration of the use of clustering as an alternate means of distinguishing between AFFF and non-AFFF samples. In contrast with supervised methods, the clustering methods tested were not able to distinguish between PFAS from different sources, but were rather more sensitive to the effects of transport on composition. Although it is likely there are some applications of clustering for source delineation, particularly for cases where very different source compositions are present, extreme care must be exercised interpreting the results. Based on the preliminary work conducted here, it is unlikely that clustering (unsupervised learning) would be satisfactory for more general PFAS source allocation.

5.A.4. Preliminary Data Mining

For this work, hybrid methods were considered where knowledge of PFAS surfactant physicochemical behavior was combined with supervised learning as a means of classifying PFAS data from non-water sources (soil, sediment, biota, concrete). A group contribution model was developed for calculating the affinities of individual PFAS components for different phases, and then used to mathematically transform water data to provide a predicted signature for phases in equilibrium with water. Comparison between the transformed water data and non-water data from the same sites shows strong agreement for phases likely to have been contaminated through partitioning from water. Transformed water data were used to train a new supervised machine learning classifier which was used to classify data from other phases. **The resulting classifications showed great promise both as a way of classifying PFAS in other phases, and identifying situations where other mechanisms may be responsible for the observed PFAS composition.**

5.B. Potential Next Steps

The focus of future work would be on **creation of an end-user tool** that could be used to **conduct screening classifications of PFAS in environmental samples**. By providing an **immediate, quantitative means of assessing the origin of PFAS constituents in a sample,** the results could be complementary to other more costly and labor-intensive techniques involving analytical methods. A tool resulting from this work could **guide allocation of resources** by highlighting samples or areas for more detailed study, potentially providing **substantial utility and cost savings to DoD**.

While it is clear that a tool created directly using the methods from this proof-of-concept project could be of significant value, additional work in the following areas would further

enhance the forensic capabilities and utility of the tool for use at DoD sites. Future work is needed in the following areas:

1. Creation of an Expanded Dataset

The dataset created for this proof-of-concept project was a preliminary dataset, designed primarily to test the concept of supervised machine learning for classification of PFAS in environmental samples. As such, the dataset described in this report only begins to scratch the surface of available PFAS data. Specific needs in an expanded dataset include the **addition of more industry-specific non-AFFF PFAS examples**, as well as **more AFFF examples from the United States**. Overall, hundreds of sites from around the US and elsewhere likely have accessible data that could be added to the dataset. Furthermore, some data in the preliminary dataset await classification based on available site and location information. These additions are important for increased classification certainty, particularly with difficult sample types.

2. Selection of Methods for Rejection/Open Set Recognition

A final tool needs to be able to identify cases where patterns in a new sample are unlike those in the training set. This could either take the form of rejection if the pattern is not a sufficient match, or open set recognition where unknown samples are placed into their own class. In the area of rejection, one-class recognition methods and outlier detection methods could take advantage of the fact that AFFF samples are well-represented in the dataset. Work in this area is essential to prevent non-AFFF samples with novel compositions from being mistaken for AFFF samples.

3. Exploration of Advanced Training Methods

The ability of a classifier to identify patterns in data depend in part on the use of a balanced training set. This means that representative samples from the various expected patterns must be present in sufficient quantities to influence the classifications of a model trained on the dataset. This extends to not just AFFF vs non-AFFF samples, but actual expected subsets within both categories and their environmental patterns. Balancing strategies have yet to be explored for this application, but it is likely that performance can be improved with balancing of training subsets. Strategies to do this might include the use of synthetic training data, or ensemble predictions trained on subsets of the dataset.

4. Multiclass Classification

With an expanded dataset containing additional examples of industry-specific PFAS formulations, multiclass classification (i.e., identification of specific subsets, rather than simply AFFF vs non-AFFF) should become practical. Only minimal changes to existing coding would be needed to conduct multiclass classification, but the ability to identify specific industrial sources for non-AFFF samples, as well as categories of AFFF samples, could further enhance the value of a tool for source allocation.

5. Simulations to Better Understand Classification Performance

It is likely that the most effective machine learning-based source allocation strategies can benefit from some combination of machine learning with a physicochemical understanding of PFAS behavior. One example of this is the transformations explored in Section 4.E of this report for classification of PFAS in non-water phases. In the area of classification, transport

simulations are needed to better understand how mixing of PFAS from diverse sources impacts classification, as well as to quantify the extent to which very similar sources can be distinguished in complex environmental settings. Simulations could be useful both for method validation and for increasing understanding of method behavior, and may provide additional insights that could ultimately further improve classifications.

6. Creation of the Final Tool

The final tool to be developed for the work would be created in consultation with DoD personnel and practitioners, and would be designed to provide simultaneous output from multiple classification approaches to give a full picture of the likelihood that a tested sample is of AFFF origin. A major part of the work would be designing methods of data entry that are well-suited to site applications, and building in internal quality control checks to screen for data entry errors in the input data. It is likely the tool would be web-based, to allow machine learning computations to be conducted in the cloud, and to provide a mechanism for the tool and underlying datasets to be updated as new information becomes available.

6. LITERATURE CITED

1. Kibbey, T.C.G., R. Jabrzemski, and D.M. O'Carroll, Supervised machine learning for source allocation of per- and polyfluoroalkyl substances (PFAS) in environmental samples. *Chemosphere*, 2020. **252**: 126593, DOI: 10.1016/j.chemosphere.2020.126593.

2. Le, S.T., T.C.G. Kibbey, K.P. Weber, W.C. Glamore, and D.M. O'Carroll, A group-contribution model for predicting the physicochemical behavior of PFAS components for understanding environmental fate. *Science of the Total Environment*: 142882, DOI: 10.1016/j.scitotenv.2020.142882.

3. Bishop, C.M., *Pattern recognition and machine learning*. Information science and statistics. 2006, New York: Springer. 738.

4. Abu-Mostafa, Y.S., *Learning from data : a short course*. 2012, New York: AMLBooks. 213.

5. Marsland, S., *Machine Learning : an algorithmic perspective*. Second edition. ed. Chapman et Hall/CRC machine learning et pattern recognition series. 2015, Boca Raton, FL: CRC Press. 437.

6. LeCun, Y., Y. Bengio, and G. Hinton, Deep learning. *Nature*, 2015. **521**(7553): 436-444, DOI: 10.1038/nature14539.

7. Rumelhart, D.E., G.E. Hinton, and R.J. Williams, Learning Representations by Back-Propagating Errors. *Nature*, 1986. **323**(6088): 533-536, DOI: 10.1038/323533a0.

8. Baduel, C., J.F. Mueller, A. Rotander, J. Corfield, and M.J. Gomez-Ramos, Discovery of novel per- and polyfluoroalkyl substances (PFASs) at a fire fighting training ground and preliminary investigation of their fate and mobility. *Chemosphere*, 2017. **185**: 1030-1038, DOI: 10.1016/j.chemosphere.2017.06.096.

9. Houtz, E.F., C.P. Higgins, J.A. Field, and D.L. Sedlak, Persistence of perfluoroalkyl acid precursors in AFFF-impacted groundwater and soil. *Environmental Science and Technology*, 2013. **47**(15): 8187-8195, DOI: 10.1021/es4018877.

10. Krafft, M.P. and J.G. Riess, Per- and polyfluorinated substances (PFASs): Environmental challenges. *Current Opinion in Colloid & Interface Science*, 2015. **20**(3): 192-212, DOI: 10.1016/j.cocis.2015.07.004.

11. Buck, R.C., J. Franklin, U. Berger, J.M. Conder, I.T. Cousins, P. de Voogt, A.A. Jensen, K. Kannan, S.A. Mabury, and S.P.J. van Leeuwen, Perfluoroalkyl and Polyfluoroalkyl Substances in the Environment: Terminology, Classification, and Origins. *Integrated Environmental Assessment and Management*, 2011. **7**(4): 513-541, DOI: 10.1002/ieam.258.

12. Munoz, G., P. Labadie, F. Botta, F. Lestremau, B. Lopez, E. Geneste, P. Pardon, M.H. Devier, and H. Budzinski, Occurrence survey and spatial distribution of perfluoroalkyl

and polyfluoroalkyl surfactants in groundwater, surface water, and sediments from tropical environments. *Science of the Total Environment*, 2017. **607**: 243-252, DOI: 10.1016/j.scitotenv.2017.06.146.

13. Anderson, R.H., G.C. Long, R.C. Porter, and J.K. Anderson, Occurrence of select perfluoroalkyl substances at U.S. Air Force aqueous film-forming foam release sites other than fire-training areas: Field-validation of critical fate and transport properties. *Chemosphere*, 2016. **150**: 678-685, DOI: 10.1016/j.chemosphere.2016.01.014.

14. Weber, A.K., L.B. Barber, D.R. LeBlanc, E.M. Sunderland, and C.D. Vecitis, Geochemical and Hydrologic Factors Controlling Subsurface Transport of Poly- and Perfluoroalkyl Substances, Cape Cod, Massachusetts. *Environmental Science & Technology*, 2017. **51**(8): 4269-4279, DOI: 10.1021/acs.est.6b05573.

15. O'Carroll, D.M., T.C. Jeffries, M.J. Lee, S.T. Le, A. Yeung, S. Wallace, N. Battye, D.J. Patch, M.J. Manefield, and K.P. Weber, Developing a Roadmap to determine Per- and Polyfluoroalkyl Substances-Microbial Population Interactions. *Science of the Total Environment*, 2020. **12**: 135994, DOI: 10.1016/j.scitotenv.2019.135994.

16. Géron, A., *Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems*. 2017, Sebastopol (CA): O'Reilly Media. 543.

17. Strogatz, S., *One Giant Step for a Chess-Playing Machine*, in *The New York Times*. Dec. 26, 2018.

18. Müller, A.C. and S. Guido, *Introduction to machine learning with Python a guide for data scientists*. 2017, Sebastopol (CA): O'Reilly.

19. Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 2011. **12**: 2825-2830.

20. Breiman, L., Bagging predictors. *Machine Learning*, 1996. **24**(2): 123-140, DOI: 10.1007/Bf00058655.

21. Martín Abadi, A.A., Paul Barham, Eugene Brevdo,, C.C. Zhifeng Chen, Greg S. Corrado, Andy Davis,, M.D. Jeffrey Dean, Sanjay Ghemawat, Ian Goodfellow,, G.I. Andrew Harp, Michael Isard, Rafal Jozefowicz, Yangqing Jia,, M.K. Lukasz Kaiser, Josh Levenberg, Dan Mané, Mike Schuster,, S.M. Rajat Monga, Derek Murray, Chris Olah, Jonathon Shlens,, I.S. Benoit Steiner, Kunal Talwar, Paul Tucker,, V.V. Vincent Vanhoucke, Fernanda Viégas,, P.W. Oriol Vinyals, Martin Wattenberg, Martin Wicke,, and a.X.Z. Yuan Yu, TensorFlow: Large-scale machine learning on heterogeneous systems. *Software available from tensorflow.org.*, 2015.

22. Geurts, P., D. Ernst, and L. Wehenkel, Extremely randomized trees. *Machine Learning*, 2006. **63**(1): 3-42, DOI: 10.1007/s10994-006-6226-1.

23. Chang, C.C. and C.J. Lin, LIBSVM: A Library for Support Vector Machines. *Acm Transactions on Intelligent Systems and Technology*, 2011. **2**(3): Artn 27, DOI: 10.1145/1961189.1961199.

24. Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, R. Jozefowicz, Y. Jia, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, M. Schuster, R. Monga, S. Moore, D. Murray, C. Olah, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems. *Software available from tensorflow.org.*, 2015.

APPENDICES

Appendix A. Data Source Information

0. JERVISBAY_BIOTA
Jervis Bay Range Facility (JBRF)
Jervis Bay Range Facility & HMAS Creswell Detailed Site Investigation REPORT July 2019
https://www.defence.gov.au/Environment/PFAS/docs/JervisBay/Reports/201907.JBRF.DSI.pdf
AU, JBT
MILITARY_BASE
AFFF

1. JERVISBAY_CONCRETE
Jervis Bay Range Facility (JBRF)
Jervis Bay Range Facility & HMAS Creswell Detailed Site Investigation REPORT July 2019
https://www.defence.gov.au/Environment/PFAS/docs/JervisBay/Reports/201907.JBRF.DSI.pdf
AU, JBT
MILITARY_BASE
AFFF

2. JERVISBAY_CONCRETE_LEACHATE
Jervis Bay Range Facility (JBRF)
Jervis Bay Range Facility & HMAS Creswell Detailed Site Investigation REPORT July 2019
https://www.defence.gov.au/Environment/PFAS/docs/JervisBay/Reports/201907.JBRF.DSI.pdf
AU, JBT
MILITARY_BASE
AFFF

3. JERVISBAY_GW
Jervis Bay Range Facility (JBRF)
Jervis Bay Range Facility & HMAS Creswell Detailed Site Investigation REPORT July 2019
https://www.defence.gov.au/Environment/PFAS/docs/JervisBay/Reports/201907.JBRF.DSI.pdf
AU, JBT
MILITARY_BASE
AFFF

4. JERVISBAY_SEDIMENT
Jervis Bay Range Facility (JBRF)
Jervis Bay Range Facility & HMAS Creswell Detailed Site Investigation REPORT July 2019
https://www.defence.gov.au/Environment/PFAS/docs/JervisBay/Reports/201907.JBRF.DSI.pdf
AU, JBT
MILITARY_BASE
AFFF

5. JERVISBAY_SEDIMENT_LEACHATE
Jervis Bay Range Facility (JBRF)

Jervis Bay Range Facility & HMAS Creswell Detailed Site Investigation REPORT July 2019
https://www.defence.gov.au/Environment/PFAS/docs/JervisBay/Reports/201907.JBRF.DSI.pdf
AU, JBT
MILITARY_BASE
AFFF

6. JERVISBAY_SOIL
Jervis Bay Range Facility (JBRF)
Jervis Bay Range Facility & HMAS Creswell Detailed Site Investigation REPORT July 2019
https://www.defence.gov.au/Environment/PFAS/docs/JervisBay/Reports/201907.JBRF.DSI.pdf
AU, JBT
MILITARY_BASE
AFFF

7. JERVISBAY_SW
Jervis Bay Range Facility (JBRF)
Jervis Bay Range Facility & HMAS Creswell Detailed Site Investigation REPORT July 2019
https://www.defence.gov.au/Environment/PFAS/docs/JervisBay/Reports/201907.JBRF.DSI.pdf
AU, JBT
MILITARY_BASE
AFFF

8. JERVISBAY_TANK_SW
Jervis Bay Range Facility (JBRF)
Jervis Bay Range Facility & HMAS Creswell Detailed Site Investigation REPORT July 2019
https://www.defence.gov.au/Environment/PFAS/docs/JervisBay/Reports/201907.JBRF.DSI.pdf
AU, JBT
MILITARY_BASE
AFFF

9. ALBATROSS_GW
Royal Australian Navy HMAS Albatross
Detailed Site Investigation – Groundwater Monitoring: January 2017 to November 2017
https://www.defence.gov.au/Environment/PFAS/docs/Albatross/Reports/20171116HMASAlbatrossDSIFullReport.pdf
AU, NSW
MILITARY_BASE
AFFF

10. ALBATROSS_LEACHATE
Royal Australian Navy HMAS Albatross; Nowra Hill NSW; Australian Department of Defense
Detailed Site Investigation Groundwater Monitoring: January 2017 to November 2017
https://www.defence.gov.au/Environment/PFAS/docs/Albatross/Reports/20171116HMASAlbatrossDSIFullReport.pdf
AU, NSW
MILITARY_BASE

AFFF

11. ALBATROSS_SEDIMENT
Royal Australian Navy HMAS Albatross
Detailed Site Investigation â€“ Groundwater Monitoring: January 2017 to November 2017
https://www.defence.gov.au/Environment/PFAS/docs/Albatross/Reports/20171116HMASAlbatrossDSIFullReport.pdf
AU, NSW
MILITARY_BASE
AFFF

12. ALBATROSS_SOIL
Royal Australian Navy HMAS Albatross; Nowra Hill NSW; Australian Department of Defense
Detailed Site Investigation â€“ Groundwater Monitoring: January 2017 to November 2017
https://www.defence.gov.au/Environment/PFAS/docs/Albatross/Reports/20171116HMASAlbatrossDSIFullReport.pdf
AU, NSW
MILITARY_BASE
AFFF

13. BLAMEY_GW
Blamey Barracks
Detailed Site Investigation - Groundwater Monitoring: December 2018 - September 2019
https://www.defence.gov.au/Environment/PFAS/docs/Kapooka/Reports/201909BlameyBarracksKapookaDetailedSiteInvestigationReport.pdf
AU, NSW
MILITARY_BASE
AFFF

14. HOLSWORTHY_OFF_GW
Holsworthy Barracks
Detailed Site Investigation - Groundwater Monitoring: August - November 2018
https://www.defence.gov.au/Environment/PFAS/docs/Holsworthy/Reports/201811HolsworthyDSIFullReport.pdf
AU, NSW
MILITARY_BASE
AFFF

15. HOLSWORTHY_ON_GW
Holsworthy Barracks
Detailed Site Investigation - Groundwater Monitoring: August - November 2018
https://www.defence.gov.au/Environment/PFAS/docs/Holsworthy/Reports/201811HolsworthyDSIFullReport.pdf
AU, NSW
MILITARY_BASE
AFFF

16. RICHMOND_GW
RAAF Base Richmond
Detailed Site Investigation - Groundwater Monitoring: February - June 2018
https://www.defence.gov.au/Environment/PFAS/docs/Richmond/Reports/201806DetailedSiteInvestigationReportBody.pdf
AU, NSW
MILITARY_BASE
AFFF

17. SINGLETON_OFF_GW
Singleton Military Area
Detailed Site Investigation - Groundwater Monitoring: June 2018 - November 2019
https://www.defence.gov.au/Environment/PFAS/docs/Singleton/Reports/201912DetailedSiteInvestigationReport.pdf
AU, NSW
MILITARY_BASE
AFFF

18. SINGLETON_ON_GW
Singleton Military Area
Detailed Site Investigation - Groundwater Monitoring: June 2018 - November 2019
https://www.defence.gov.au/Environment/PFAS/docs/Singleton/Reports/201912DetailedSiteInvestigationReport.pdf
AU, NSW
MILITARY_BASE
AFFF

19. WAGGA_GW
RAAF Base Wagga
Detailed Site Investigation - Groundwater Monitoring: March - June 2018
https://www.defence.gov.au/Environment/PFAS/docs/wagga/Reports/DetailedSiteInvestigationReport.pdf
AU, NSW
MILITARY_BASE
AFFF

20. WILLIAMTOWN_GW
RAAF Base Williamtown
Interim Monitoring Event Report - June 2019
https://www.defence.gov.au/Environment/PFAS/docs/Williamtown/Reports/201909WilliamtownInterimMonitoringEventReport.pdf
AU, NSW
MILITARY_BASE
AFFF

21. WILLIAMTOWN_SEDIMENT
RAAF Base Williamtown
Interim Monitoring Event Report - June 2019
https://www.defence.gov.au/Environment/PFAS/docs/Williamtown/Reports/201909WilliamtownInterimMonitoringEventReport.pdf
AU, NSW
MILITARY_BASE
AFFF

22. WILLIAMTOWN_SW
RAAF Base Williamtown
Interim Monitoring Event Report - June 2019
https://www.defence.gov.au/Environment/PFAS/docs/Williamtown/Reports/201909WilliamtownInterimMonitoringEventReport.pdf
AU, NSW
MILITARY_BASE
AFFF

23. DARWIN_GW
RAAF Base Darwin
Detailed Site Investigation - February 2018
https://www.defence.gov.au/environment/pfas/Darwin/publications.asp
AU, NT
MILITARY_BASE
AFFF

24. ROBERTSON_DRY
Robertson Barracks (RB)
Detailed Site Investigation - 31 May 2018
https://www.defence.gov.au/Environment/PFAS/docs/Robertson/Reports/201806DetailedSiteInvestigationReport.pdf
AU, NT
MILITARY_BASE
AFFF

25. ROBERTSON_WET
Robertson Barracks (RB)
Detailed Site Investigation - 31 May 2018
https://www.defence.gov.au/Environment/PFAS/docs/Robertson/Reports/201806DetailedSiteInvestigationReport.pdf
AU, NT
MILITARY_BASE
AFFF

26. AMBERLEY_OFF_GWSW
RAAF Base Amberley

Detailed Site Investigation - Groundwater and Surface water Monitoring: November 2018
https://www.defence.gov.au/Environment/PFAS/docs/Amberley/Reports/201811DSIReportBody.pdf
AU, QLD
MILITARY_BASE
AFFF

27. AMBERLEY_ON_GWSW
RAAF Base Amberley
Detailed Site Investigation - Groundwater and Surface water Monitoring: November 2018
https://www.defence.gov.au/Environment/PFAS/docs/Amberley/Reports/201811DSIReportBody.pdf
AU, QLD
MILITARY_BASE
AFFF

28. CAIRNS_GW
HMAS Cairns
Detailed Site Investigation - Groundwater Monitoring: December 2019
https://www.defence.gov.au/Environment/PFAS/docs/Cairns/Reports/201912CairnsDetailedSiteInvestigationReportBody.pdf
AU, QLD
MILITARY_BASE
AFFF

29. CAIRNS_SW
HMAS Cairns
Detailed Site Investigation - Groundwater Monitoring: December 2019
https://www.defence.gov.au/Environment/PFAS/docs/Cairns/Reports/201912CairnsDetailedSiteInvestigationReportBody.pdf
AU, QLD
MILITARY_BASE
AFFF

30. LAVARACK_OFF_GW
Lavarack Barracks
Detailed Site Investigation - Groundwater Monitoring: February - December 2019
https://www.defence.gov.au/Environment/PFAS/docs/Lavarack/Reports/20191203LavarackBarracksDetailedSiteInvestigationReport.pdf
AU, QLD
MILITARY_BASE
AFFF

31. LAVARACK_OFF_SW
Lavarack Barracks
Detailed Site Investigation - Groundwater Monitoring: February - December 2019

https://www.defence.gov.au/Environment/PFAS/docs/Lavarack/Reports/20191203LavarackBarracksDetailedSiteInvestigationReport.pdf
AU, QLD
MILITARY_BASE
AFFF

32. LAVARACK_ON_GW
Lavarack Barracks
Detailed Site Investigation - Groundwater Monitoring: February - December 2019
https://www.defence.gov.au/Environment/PFAS/docs/Lavarack/Reports/20191203LavarackBarracksDetailedSiteInvestigationReport.pdf
AU, QLD
MILITARY_BASE
AFFF

33. LAVARACK_ON_SW
Lavarack Barracks
Detailed Site Investigation - Groundwater Monitoring: February - December 2019
https://www.defence.gov.au/Environment/PFAS/docs/Lavarack/Reports/20191203LavarackBarracksDetailedSiteInvestigationReport.pdf
AU, QLD
MILITARY_BASE
AFFF

34. OAKEY_OFF_GW
Army Aviation Centre Oakey (AACO); Oakey QLD; Australian Department of Defence
Stage 2C Environmental Investigation. Groundwater and Surface Water Monitoring: July 2017 to May 2018
https://www.defence.gov.au/Environment/PFAS/Docs/Oakey/Reports/201810OakeyGME.pdf
AU, QLD
MILITARY_BASE
AFFF

35. OAKEY_ON_GW
Army Aviation Centre Oakey (AACO); Oakey QLD; Australian Department of Defence
Stage 2C Environmental Investigation. Groundwater and Surface Water Monitoring: July 2017 to May 2018
https://www.defence.gov.au/Environment/PFAS/Docs/Oakey/Reports/201810OakeyGME.pdf
AU, QLD
MILITARY_BASE
AFFF

36. OAKEY_ON_SOIL
Army Aviation Centre Oakey (AACO); Oakey QLD; Australian Department of Defence
Environmental Site Assessment Report (vol 2). 2018

https://www.defence.gov.au/Environment/PFAS/Docs/Oakey/Reports/20171206AACOESAVolume2.pdf
AU, QLD
MILITARY_BASE
AFFF

37. OAKEY_SW
Army Aviation Centre Oakey (AACO); Oakey QLD; Australian Department of Defence
Stage 2C Environmental Investigation. Groundwater and Surface Water Monitoring: July 2017 to May 2018; Also converted the 2019 sampling report. Both included
https://www.defence.gov.au/Environment/PFAS/Docs/Oakey/Reports/201810OakeyGME.pdf
AU, QLD
MILITARY_BASE
AFFF

38. QH3_BIOTA
Qantas Hangar 3 (QH3); Brisbane Airport East Precinct; Brisbane QLD; AU: May - June 2017
Qantas Hangar 3; AFFF LOC Event; Brisbane Airport; Environmental Evaluation Notice (EEN)
AU, QLD
AIRPORT
AFFF

39. QH3_CONCENTRATE
Qantas Hangar 3 (QH3); Brisbane Airport East Precinct; Brisbane QLD; AU: April - May 2017
Qantas Hangar 3; AFFF LOC Event; Brisbane Airport; Environmental Evaluation Notice (EEN)
AU, QLD
AIRPORT
AFFF

40. QH3_CONCRETE
Qantas Hangar 3 (QH3); Brisbane Airport East Precinct; Brisbane QLD; AU: April 2017
Qantas Hangar 3; AFFF LOC Event; Brisbane Airport; Environmental Evaluation Notice (EEN)
AU, QLD
AIRPORT
AFFF

41. QH3_GW
Qantas Hangar 3 (QH3); Brisbane Airport East Precinct; Brisbane QLD; AU: May - June 2017
Qantas Hangar 3; AFFF LOC Event; Brisbane Airport; Environmental Evaluation Notice (EEN)
AU, QLD
AIRPORT
AFFF

42. QH3_PW
Qantas Hangar 3 (QH3); Brisbane Airport East Precinct; Brisbane QLD; AU: May 2017
Qantas Hangar 3; AFFF LOC Event; Brisbane Airport; Environmental Evaluation Notice (EEN)

AU, QLD
AIRPORT
AFFF

43. QH3_SEDIMENT
Qantas Hangar 3 (QH3); Brisbane Airport East Precinct; Brisbane QLD; AU: April - June 2017
Qantas Hangar 3; AFFF LOC Event; Brisbane Airport; Environmental Evaluation Notice (EEN)
AU, QLD
AIRPORT
AFFF

44. QH3_SEWER
Qantas Hangar 3 (QH3); Brisbane Airport East Precinct; Brisbane QLD; AU: April - July 2017
Qantas Hangar 3; AFFF LOC Event; Brisbane Airport; Environmental Evaluation Notice (EEN)
AU, QLD
AIRPORT
AFFF

45. QH3_SOIL
Qantas Hangar 3 (QH3); Brisbane Airport East Precinct; Brisbane QLD; AU: April - June 2017
Qantas Hangar 3; AFFF LOC Event; Brisbane Airport; Environmental Evaluation Notice (EEN)
AU, QLD
AIRPORT
AFFF

46. QH3_SW
Qantas Hangar 3 (QH3); Brisbane Airport East Precinct; Brisbane QLD; AU: April - July 2017
Qantas Hangar 3; AFFF LOC Event; Brisbane Airport; Environmental Evaluation Notice (EEN)
AU, QLD
AIRPORT
AFFF

47. QH3_WASTE_SOIL
Qantas Hangar 3 (QH3); Brisbane Airport East Precinct; Brisbane QLD; AU: June 2017
Qantas Hangar 3; AFFF LOC Event; Brisbane Airport; Environmental Evaluation Notice (EEN)
AU, QLD
AIRPORT
AFFF

48. QH3_WWTP
Qantas Hangar 3 (QH3); Brisbane Airport East Precinct; Brisbane QLD; AU: April - June 2017
Qantas Hangar 3; AFFF LOC Event; Brisbane Airport; Environmental Evaluation Notice (EEN)
AU, QLD
AIRPORT
AFFF

49. TOWNSVILLE_OFF1_GW
RAAF Base Townsville
RAAF BASE TOWNSVILLE - ECOLOGICAL RISK ASSESSMENT (ERA): 2019
https://www.defence.gov.au/Environment/PFAS/docs/Townsville/Reports/201912EcologicalRiskAssessmentReportAppendices.pdf
AU, QLD
MILITARY_BASE
AFFF

50. TOWNSVILLE_OFF2_GW
RAAF Base Townsville
RAAF BASE TOWNSVILLE - ECOLOGICAL RISK ASSESSMENT (ERA): 2019
https://www.defence.gov.au/Environment/PFAS/docs/Townsville/Reports/201912EcologicalRiskAssessmentReportAppendices.pdf
AU, QLD
MILITARY_BASE
AFFF

51. TOWNSVILLE_ON_GW
RAAF Base Townsville
RAAF BASE TOWNSVILLE - ECOLOGICAL RISK ASSESSMENT (ERA): 2019
https://www.defence.gov.au/Environment/PFAS/docs/Townsville/Reports/201912EcologicalRiskAssessmentReportAppendices.pdf
AU, QLD
MILITARY_BASE
AFFF

52. GALLEN_2017_LF_GW
Gallen et al.; Australia-wide assessment of perfluoroalkyl substances (PFASs) in landfill leachates; Journal of Hazardous Materials; Volume 331, 5 June 2017, Pages 132-141
https://www.sciencedirect.com/science/article/pii/S0304389417300870
AU, Various
LANDFILL
nonAFFF

53. BANDIANA_OFF_GW
Bandiana Military Area Vic
PFAS Detailed Site Investigation Bandiana Military Area
https://www.defence.gov.au/Environment/PFAS/docs/Cerberus/Reports/201810CerberusDetailedSiteInvestigationReportAppendices.pdf
AU, VIC
MILITARY_BASE
AFFF

54. BANDIANA_ON_GW
Bandiana Military Area Vic

PFAS Detailed Site Investigation Bandiana Military Area
https://www.defence.gov.au/Environment/PFAS/docs/Cerberus/Reports/201810CerberusDetailedSiteInvestigationReportAppendices.pdf
AU, VIC
MILITARY_BASE
AFFF

55. BANDIANA_POU_SW
Bandiana Military Area Vic
PFAS Detailed Site Investigation Bandiana Military Area
https://www.defence.gov.au/Environment/PFAS/docs/Cerberus/Reports/201810CerberusDetailedSiteInvestigationReportAppendices.pdf
AU, VIC
MILITARY_BASE
AFFF

56. EASTSALE_OFF_GW
RAAF Base East Sale VIC Australian Department of Defence
RAAF Base East Sale – Per- and Poly-fluoroalkyl Substances (PFAS) Investigations
https://www.defence.gov.au/Environment/PFAS/docs/EastSale/Reports/RAAFEastSalePFASInvestigationDSIReportFull.pdf
AU, VIC
MILITARY_BASE
AFFF

57. EASTSALE_ON_GW
RAAF Base East Sale VIC Australian Department of Defence
RAAF Base East Sale – Per- and Poly-fluoroalkyl Substances (PFAS) Investigations
https://www.defence.gov.au/Environment/PFAS/docs/EastSale/Reports/RAAFEastSalePFASInvestigationDSIReportFull.pdf
AU, VIC
MILITARY_BASE
AFFF

58. WILLIAMS_GW
RAAF Williams (Laverton) VIC Australian Department of Defence
DEPARTMENT OF DEFENCE PFAS BRANCH Preliminary Site Investigation for PFAS
RAAF Base Williams Laverton VIC (0927)
https://www.defence.gov.au/Environment/PFAS/docs/Williams/Reports/RAAFBaseWilliamsPSIReport.pdf
AU, VIC
MILITARY_BASE
AFFF

59. HEPBURN_2019_LF_GW
Fisherman Bend

Hepburn et al. Contamination of groundwater with per- and polyfluoroalkylsubstances (PFAS) from legacy landfills in an urban re-development precinct. Environ. Pollution, 2019, 248, 101-113.
https://www.sciencedirect.com/science/article/pii/S0269749118327994
AU, VIC
LANDFILL
nonAFFF

60. PEARCE_BIOTA
RAAF Base Pearce
Detailed Site Investigation - Appendix D to F
https://www.defence.gov.au/Environment/PFAS/docs/Pearce/Reports/RAAFBasePearceDetailedSiteInvestigationAppendixDtoF.pdf
AU, WA
MILITARY_BASE
AFFF

61. PEARCE_GW
RAAF Base Pearce
Detailed Site Investigation - Appendix D to F
https://www.defence.gov.au/Environment/PFAS/docs/Pearce/Reports/RAAFBasePearceDetailedSiteInvestigationAppendixDtoF.pdf
AU, WA
MILITARY_BASE
AFFF

62. PEARCE_SOIL
RAAF Base Pearce
Detailed Site Investigation - Appendix D to F
https://www.defence.gov.au/Environment/PFAS/docs/Pearce/Reports/RAAFBasePearceDetailedSiteInvestigationAppendixDtoF.pdf
AU, WA
MILITARY_BASE
AFFF

63. PEARCE_LEACHABILITY
RAAF Base Pearce
Detailed Site Investigation - Appendix D to F
https://www.defence.gov.au/Environment/PFAS/docs/Pearce/Reports/RAAFBasePearceDetailedSiteInvestigationAppendixDtoF.pdf
AU, WA
MILITARY_BASE
AFFF

64. STIRLING_GW
HMAS Stirling Fleet Base West

Detailed Site Investigation
https://defence.gov.au/Environment/PFAS/docs/Stirling/Reports/1805StirlingDetailedSiteInvestigationReport.pdf
AU, WA
NAVAL_BASE
AFFF

65. STIRLING_SEDIMENT
HMAS Stirling Fleet Base West
Detailed Site Investigation
https://defence.gov.au/Environment/PFAS/docs/Stirling/Reports/1805StirlingDetailedSiteInvestigationReport.pdf
AU, WA
NAVAL_BASE
AFFF

66. STIRLING_SOIL
HMAS Stirling Fleet Base West
Detailed Site Investigation
https://defence.gov.au/Environment/PFAS/docs/Stirling/Reports/1805StirlingDetailedSiteInvestigationReport.pdf
AU, WA
NAVAL_BASE
AFFF

67. STIRLING_SW
HMAS Stirling Fleet Base West
Detailed Site Investigation
https://defence.gov.au/Environment/PFAS/docs/Stirling/Reports/1805StirlingDetailedSiteInvestigationReport.pdf
AU, WA
NAVAL_BASE
AFFF

68. DND_Site-B_GW
CFB in Western Canada; DND Site B
O'Carroll et al.; Developing a roadmap to determine per- and polyfluoroalkyl substances-microbial population interactions; STOTEN, 2020, 712, 135994.
https://www.sciencedirect.com/science/article/pii/S004896971935990X
CA, DND
MILITARY_BASE
AFFF

69. DND_Site-C_GW
DND Site C; Canada: February 2009 - December 2017
DND Site C

Personal communication DND personnel.
CA, DND
MILITARY_BASE
AFFF

70. DND_Site-C_SW
DND Site C; Canada: July 2009 - December 2016
DND Site C
Personal communication DND personnel.
CA, DND
MILITARY_BASE
AFFF

71. DND_Site-E_GW
DND Site G; Canada: Oct 2016 - Jan 2020
DND Site E
Personal communication DND personnel.
CA, DND
MILITARY_BASE
AFFF

72. DND_Site-E_MH
DND Site G; Canada: Dec 2017 - Nov 2018
DND Site E
Personal communication DND personnel.
CA, DND
MILITARY_BASE
AFFF

73. DND_Site-E_SEDIMENT
DND Site G; Canada: Feb 2016 - Nov 2018
DND Site E
Personal communication DND personnel.
CA, DND
MILITARY_BASE
AFFF

74. DND_Site-E_SOIL
DND Site G; Canada: Nov 2016 - Nov 2018
DND Site E
Personal communication DND personnel.
CA, DND
MILITARY_BASE
AFFF

75. DND_Site-E_SW

DND Site G; Canada: Nov 2016 - Nov 2018
DND Site E
Personal communication DND personnel.
CA, DND
MILITARY_BASE
AFFF

76. DND_Site-G_GW
DND Site G; Canada: January 2015 - Feb 2020
DND Site G
Personal communication DND personnel.
CA, DND
MILITARY_BASE
AFFF

77. DND_Site-G_SOIL_SEDIMENT
DND Site G; Canada: January 2015 - January 2020
DND Site G
Personal communication DND personnel.
CA, DND
MILITARY_BASE
AFFF

78. DND_Site-G_GW
DND Site G; Canada: January 2015 - Feb 2020
DND Site G
Personal communication DND personnel.
CA, DND
MILITARY_BASE
AFFF

79. DND_Site-H_GW
DND Site G; Canada: January 2015 - Feb 2020
DND Site H
Personal communication DND personnel.
CA, DND
MILITARY_BASE
AFFF

80. DND_Site-H_STORMWATER
DND Site G; Canada: Nov 2012 - Jan 2020
DND Site H
Personal communication DND personnel.
CA, DND
MILITARY_BASE
AFFF

81. DND_Site-H_SW
DND Site G; Canada: Jul 2012 - Jan 2020
DND Site H
Personal communication DND personnel.
CA, DND
MILITARY_BASE
AFFF

82. DND_Site-I_GW
DND Site G; Canada: Feb 2014 - Nov 2019
DND Site I
Personal communication DND personnel.
CA, DND
MILITARY_BASE
AFFF

83. DND_Site-I_SEDIMENT
DND Site G; Canada: Dec 2013 - Nov 2014
DND Site I
Personal communication DND personnel.
CA, DND
MILITARY_BASE
AFFF

84. DND_Site-I_SOIL
DND Site G; Canada: Nov 2013 - Nov 2019
DND Site I
Personal communication DND personnel.
CA, DND
MILITARY_BASE
AFFF

85. DND_Site-I_SW
DND Site G; Canada: Dec 2013 - Oct 2015
DND Site I
Personal communication DND personnel.
CA, DND
MILITARY_BASE
AFFF

86. HAMILTON_AIRPORT_GWSW
Former Fire Training Facility
Initial Subsurface investigation - Perfluorooctane Sulfonate (PFOS) and Perfluorooctanoate (PFOA): June - December 2011
https://cdn.flyhamilton.ca/wp-content/uploads/2019/05/Initial-Subsurface-Investigation.pdf

CA, ON
AIRPORT
AFFF

87. BENSKIN_2012_LF_GW
Municipal solid waste landfill treated and raw leachate in 2 locations. Canada: August 2009 - June 2010
Benskin et al., Per- and Polyfluoroalkyl Substances in Landfill Leachate: Patterns…, ES&T, 2012, 46, 21, 11532-11540
https://pubs.acs.org/doi/pdf/10.1021/es302471n
CA, Various
LANDFILL
nonAFFF

88. BOHAISEA_SEDIMENT
Sediment samples in the Bohai Sea and surrounding rivers; China: August - September 2017
Zhao et al., Emerging and legacy per- and polyfluoroalkyl substances in water, 2020, EP, 263, 114391.
https://www.sciencedirect.com/science/article/pii/S0269749119374573
CN, HE
SEA
ATM

89. BOHAISEA_SW
Surface water samples in the Bohai Sea and surrounding rivers; China: August - September 2017
Zhao et al., Emerging and legacy per- and polyfluoroalkyl substances in water, 2020, EP, 263, 114391.
https://www.sciencedirect.com/science/article/pii/S0269749119374573
CN, HE
SEA
ATM

90. WEI_2018_IND_GW
Non-industrial areas in Jiangsu province China: July - February 2017
Wei et al. Distribution source identification and health risk assessment of PFASs and two PFOS alternatives in groundwater from non-industrial areas, 2018, 152, 141-150.
https://doi.org/10.1016/j.ecoenv.2018.01.039
CN, JS
INDUSTRY
INDUSTRY

91. BAO_2019_GW
Fluorochemical industrial park
Bao et al. Perfluoroalkyl substances in groundwater and home-produced vegetables and eggs around a fluorochemical industrial park in China. Ecotoxicology and Environ. Safety. 2019, 171, 199-205.

https://www.sciencedirect.com/science/article/abs/pii/S0147651318313824
CN, LN
RESIDENCES
PFAS_MFG

92. GAO_2020_IND_SEDIMENT
Sediment from Xi and Daling rivers adjacent to FIP; Liaoning; China: September 2018 - August 2020
Gao et al. Multicompartment occurrence and partitioning of alternative and legacy per- and polyfluoroalkyl substances in an impacted river in China, STOTEN, 2020, 138753.
https://www.sciencedirect.com/science/article/pii/S0048969720322701
CN, LN
RIVER
INDUSTRY

93. DALING_RIVER_BASIN_SW
Overlying and pore water from Xi and Daling rivers adjacent to FIP; Liaoning; China: September 2018 - August 2020
Gao et al. Multicompartment occurrence and partitioning of alternative and legacy per- and polyfluoroalkyl substances in an impacted river in China, STOTEN, 2020, 138753.
https://www.sciencedirect.com/science/article/pii/S0048969720322701
CN, LN
RIVER
INDUSTRY

94. FANG_2020_SOIL
Soil samples from Fushun County; Zigong City; Sichuan: January - April 2019
Fang et al. Environment occurrence of perfluoroalkyl acids and associated human health risks near a major fluorochemical manufacturing park in southwest of China, JHM, 2020, 122617.
https://www.sciencedirect.com/science/article/pii/S0304389420306063
CN, SC
RIVER
INDUSTRY

95. FANG_2020_SW
River and tap water samples from Fushun County; Zigong City; Sichuan: January - April 2019
Fang et al. Environment occurrence of perfluoroalkyl acids and associated human health risks near a major fluorochemical manufacturing park in southwest of China, JHM, 2020, 122617.
https://www.sciencedirect.com/science/article/pii/S0304389420306063
CN, SC
RIVER
INDUSTRY

96. WANG_2020_WWTP
Influent and effluent from municipal WWTP in 17 cities January 2019 - September 2020

Wang et al. Non-target and suspect screening of per- and polyfluoroalkyl substances in Chinese municipal wastewater treatment plants. Water Research, 2020, 183, 115989
https://www.sciencedirect.com/science/article/pii/S0043135420305261
CN, Various
WWTP
UNK

97. WEI_2018_nonIND_GW
Non-industrial areas in Jiangsu province China: July - February 2017
Wei et al. Distribution source identification and health risk assessment of PFASs and two PFOS alternatives in groundwater from non-industrial areas. Ecotoxicology and Environmental Safety, 2018, 152, 141-150.
https://doi.org/10.1016/j.ecoenv.2018.01.039
CN, JS
NON_INDUSTRY
ATM + LF

98. WEI_2018_nonIND_SW
Non-industrial areas in Jiangsu province China: July - October 2016
Wei et al. Distribution source identification and health risk assessment of PFASs and two PFOS alternatives in groundwater from non-industrial areas. Ecotoxicology and Environmental Safety, 2018, 152, 141-150.
https://doi.org/10.1016/j.ecoenv.2018.01.039
CN, JS
NON_INDUSTRY
ATM + LF

99. YAN_2015_LF_GW
Raw and treated landfill leachates in 5 locations
Yan et al., Perfluoroalkyl acids in municipal landfill leachates from China: Occurrence…, STOTEN, 2015, 524-525, 23-31.
https://www.sciencedirect.com/science/article/pii/S0048969715003952
CN, Various
LANDFILL
nonAFFF

100. YAN_2015_LF_SLUDGE
Raw and treated landfill leachates in 5 locations
Yan et al., Perfluoroalkyl acids in municipal landfill leachates from China: Occurrence…, STOTEN, 2015, 524-525, 23-31.
https://www.sciencedirect.com/science/article/pii/S0048969715003952
CN, Various
LANDFILL
nonAFFF

101. YAN_2015_LF_WWTP

Raw and treated landfill leachates in 5 locations
Yan et al., Perfluoroalkyl acids in municipal landfill leachates from China: Occurrence…, STOTEN, 2015, 524-525, 23-31.
https://www.sciencedirect.com/science/article/pii/S0048969715003952
CN, Various
LANDFILL
nonAFFF

102. BAVARIA_WTP
Altötting district
Measured values of the LGL on perfluorinated substances in drinking water
https://www.lgl.bayern.de/lebensmittel/warengruppen/wc_59_trinkwasser/ue_2016_trinkwasser_pft_messwerte.htm
DE, BY
WTP
UNK

103. BUSCH_2010_LF_WWTP
Municipal solid waste of 22 treated and untreated landfill leachates. GERMANY: August - December 2009
Busch et al., Polyfluoroalkyl compounds in landfill leachates, EP, 2010, 158, 1467-1471.
https://www.sciencedirect.com/science/article/pii/S0269749109006393
DE, Various
LANDFILL
nonAFFF

104. FUERTES_2017_LF_GW
Municipal solid waste landfill treated and raw leachate in 4 locations. Northern Spain: March 2015 - February 2017
Fuertes et al., Perfluorinated alkyl substances (PFASs) in northern Spain municipal solid waste landfill leachates, Chemosphere, 2017, 399-407.
https://www.sciencedirect.com/science/article/pii/S0045653516314564
ES, Various
LANDFILL
nonAFFF

105. HARRAD_2019_LF_GW
Municipal solid waste of 40 landfill sites. Republic of Ireland: June - November 2017
Harrad et al., Brominated flame retardants and perfluoroalkyl substances in landfill leachate from Ireland, STOTEN, 2019, 695, 133810.
https://www.sciencedirect.com/science/article/pii/S0048969719337519
IE, Various
LANDFILL
nonAFFF

106. UNK-SAMMUT_2017_SW

Various locations on the Maltese Islands: February - March 2015
Sammut et al., Perfluoroalkyl substances in the Maltese environment - (I) surface water and rain water, STOTEN, 2017, 589, 182-190.
https://reader.elsevier.com/reader/sd/pii/S0048969717303765
MT, Various
ENV
ATM

107. UNK-SAMMUT_2019_GW
Various locations on the Maltese Islands: December 2015 - January 2016
Sammut et al., Perfluoroalkyl substances in the Maltese environment - (II) sediments, STOTEN, 2019, 682, 180-189.
https://www.sciencedirect.com/science/article/pii/S0048969719319540
MT, Various
ENV
ATM

108. UNK-SAMMUT_2019_SOIL
Various locations on the Maltese Islands: May - August 2015
Sammut et al., Perfluoroalkyl substances in the Maltese environment - (II) sediments, STOTEN, 2019, 682, 180-189.
https://www.sciencedirect.com/science/article/pii/S0048969719319540
MT, Various
ENV
ATM

109. STOCKHOLM-ARLANDA_GW
Stockholm-Arlanda Airport
Spridning av högfluorerade ämnen i mark från Stockholm Arlanda Airport
https://www.ivl.se/download/18.1369484715f59ce4babf2/1509550871689/B2289.pdf
SE, SIG
AIRPORT
AFFF

110. GOBELIUS_2018_FIRE_GWSW
Various
Gobelius et al., Per- and Polyfluoroalkyl Substances in Swedish Groundwater and Surface Water: Implications for Environmental Quality Standards and Drinking Water Guidelines, ES&T, 2018, 52, 4340-4349.
https://pubs.acs.org/doi/pdf/10.1021/acs.est.7b05718
SE, Various
FIRE_TRAINING
AFFF

111. GOBELIUS_IND_GWSW
Various

Gobelius et al., Per- and Polyfluoroalkyl Substances in Swedish Groundwater and Surface Water: Implications for Environmental Quality Standards and Drinking Water Guidelines, ES&T, 2018, 52, 4340-4349.
https://pubs.acs.org/doi/pdf/10.1021/acs.est.7b05718
SE, Various
INDUSTRY
nonAFFF

112. GOBELIUS_LF_GWSW
Various
Gobelius et al., Per- and Polyfluoroalkyl Substances in Swedish Groundwater and Surface Water: Implications for Environmental Quality Standards and Drinking Water Guidelines, ES&T, 2018, 52, 4340-4349.
https://pubs.acs.org/doi/pdf/10.1021/acs.est.7b05718
SE, Various
LANDFILL
UNK

113. GOBELIUS_SKIING
Various
Gobelius et al., Per- and Polyfluoroalkyl Substances in Swedish Groundwater and Surface Water: Implications for Environmental Quality Standards and Drinking Water Guidelines, ES&T, 2018, 52, 4340-4349.
https://pubs.acs.org/doi/pdf/10.1021/acs.est.7b05718
SE, Various
SKI_AREA
UNK

114. GOBELIUS_OTHER_GWSW
Various
Gobelius et al., Per- and Polyfluoroalkyl Substances in Swedish Groundwater and Surface Water: Implications for Environmental Quality Standards and Drinking Water Guidelines, ES&T, 2018, 52, 4340-4349.
https://pubs.acs.org/doi/pdf/10.1021/acs.est.7b05718
SE, Various
UNK
UNK

115. GOBELIUS_URBAN
Various
Gobelius et al., Per- and Polyfluoroalkyl Substances in Swedish Groundwater and Surface Water: Implications for Environmental Quality Standards and Drinking Water Guidelines, ES&T, 2018, 52, 4340-4349.
https://pubs.acs.org/doi/pdf/10.1021/acs.est.7b05718
SE, Various
URBAN

UNK

116. GOBELIUS_WWTP
Various
Gobelius et al., Per- and Polyfluoroalkyl Substances in Swedish Groundwater and Surface Water: Implications for Environmental Quality Standards and Drinking Water Guidelines, ES&T, 2018, 52, 4340-4349.
https://pubs.acs.org/doi/pdf/10.1021/acs.est.7b05718
SE, Various
WWTP
UNK

117. CENTRE_SW
Centre Alabama
Analytical Report Job Number: 280-123857-1 Job Description: ADEM PFC Sampling. Centre
http://www.adem.state.al.us/default.cnt
US, AL
WTP
COATINGS

118. GADSDEN_SW
Gadsden Alabama
Analytical Report Job Number:280-131137-1 Job Description: ADEM PFC Sampling. Gadsden
http://lf.adem.alabama.gov/WebLink/DocView.aspx?id=103684837&dbid=0&cr=1
US, AL
WTP
COATINGS

119. DAVIS-MONTHAN_GW
Davis Monthan AFB
Final Site Inspection Report Davis-Monthan Air Force Base
https://ar.afcec-cloud.af.mil/
US, AZ
MILITARY_BASE
AFFF

120. DAVIS-MONTHAN_SEDIMENT
Davis Monthan AFB
Final Site Inspection Report Davis-Monthan Air Force Base
https://ar.afcec-cloud.af.mil/
US, AZ
MILITARY_BASE
AFFF

121. DAVIS-MONTHAN_SOIL
Davis Monthan AFB

Final Site Inspection Report Davis-Monthan Air Force Base
https://ar.afcec-cloud.af.mil/
US, AZ
MILITARY_BASE
AFFF

122. CALIFGAMA_DOMESTIC_GW
California GAMA system
Q4 2019 GW Monitoring Report.pdf
https://gamagroundwater.waterboards.ca.gov/gama/gamamap/public/Default.asp#
US, CA
DOMESTIC_WELLS
UNK

123. CALIFGAMA_MON_GW
California GAMA system
Q4 2019 GW Monitoring Report.pdf
https://gamagroundwater.waterboards.ca.gov/gama/gamamap/public/Default.asp#
US, CA
MONITORING_WELLS
UNK

124. DU-WEL_DBS_VAS_GW
Du-Wel Metals Hartford Michigan
Du-wel Metals 2019 Laboratory Analytical (PFAS).xlsx
Personal communication State of Michigan
US, MI
METAL_PLATING
METAL_PLATING

125. DU-WEL_MW_GW
Du-Wel Metals Hartford Michigan
Du-wel Metals 2019 Laboratory Analytical (PFAS).xlsx
Personal communication State of Michigan
US, MI
METAL_PLATING
METAL_PLATING

126. DU-WEL_RES_OFF_GW
Du-Wel Metals Hartford Michigan
Du-Wel Metals - REsidential PFAS Results Summary Tables.xlsx
Personal communication State of Michigan
US, MI
METAL_PLATING
METAL_PLATING

127. DU-WEL_VAS_OFF_GW
Du-Wel Metals Hartford Michigan
Du-wel Metals 2019 Laboratory Analytical (PFAS).xlsx
Personal communication State of Michigan
US, MI
METAL_PLATING
METAL_PLATING

128. DU-WEL_VAS_ON_GW
Du-Wel Metals Hartford Michigan
Du-wel Metals 2019 Laboratory Analytical (PFAS).xlsx
Personal communication State of Michigan
US, MI
METAL_PLATING
METAL_PLATING

129. WOLVERINE_HS_GW
Wolverine Tannery
IMPLEMENTATION OF 2018 WORK PLAN SUMMARY REPORT: February 2018 - May 2019; House Street Site
https://www.epa.gov/sites/production/files/2019-06/documents/wolverine-implementation-2018-workplan.pdf
US, MI
TANNERY
SCOTCHGARD

130. WOLVERINE_HS_SOIL
Wolverine Tannery
IMPLEMENTATION OF 2018 WORK PLAN SUMMARY REPORT: February 2018 - May 2019; House Street Site
https://www.epa.gov/sites/production/files/2019-06/documents/wolverine-implementation-2018-workplan.pdf
US, MI
TANNERY
SCOTCHGARD

131. WOLVERINE_TA_GW
Wolverine Tannery
IMPLEMENTATION OF 2018 WORK PLAN SUMMARY REPORT: January 2018; Tannery 2018 Work
https://www.epa.gov/sites/production/files/2019-06/documents/wolverine-implementation-2018-workplan.pdf
US, MI
TANNERY
SCOTCHGARD

132. WOLVERINE_TA_SOIL
Wolverine Tannery
IMPLEMENTATION OF 2018 WORK PLAN SUMMARY REPORT: January 2018; Tannery 2018 Work
https://www.epa.gov/sites/production/files/2019-06/documents/wolverine-implementation-2018-workplan.pdf
US, MI
TANNERY
SCOTCHGARD

133. WOLVERINE_TA_SW
Wolverine Tannery
IMPLEMENTATION OF 2018 WORK PLAN SUMMARY REPORT: January 2018; Tannery 2018 Work
https://www.epa.gov/mi/wolverine-world-wide-tannery
US, MI
TANNERY
SCOTCHGARD

134. EASTMETRO_A_GW
Minnesota East Metro PFAS Superfund Site
https://webapp.pca.state.mn.us/cleanup/search
US, MN
WELLS
PFAS_MFG

135. CAPEFEAR_SW
Cape Fear River
GenX_Results-CapeFear.csv
https://deq.nc.gov/
US, NC
RIVER
CHEMOURS

136. ALTUS_GW
Altus AFB
Altus.pdf
https://ar.afcec-cloud.af.mil/
US, OK
MILITARY_BASE
AFFF

137. ALTUS_SEDIMENT
Altus AFB
Altus.pdf
https://ar.afcec-cloud.af.mil/

US, OK
MILITARY_BASE
AFFF

138. ALTUS_SOIL
Altus AFB
Altus.pdf
https://ar.afcec-cloud.af.mil/
US, OK
MILITARY_BASE
AFFF

139. ALTUS_SW
Altus AFB
Altus.pdf
https://ar.afcec-cloud.af.mil/
US, OK
MILITARY_BASE
AFFF

140. PDX_GW
Portland International Airport
Q4 2019 GW Monitoring Report.pdf
https://www.deq.state.or.us/Webdocs/Forms/Output/FPController.ashx?SourceId=3324&SourceIdType=11
US, OR
AIRPORT
AFFF

141. PDX_SOIL
Portland International Airport
Q4 2019 GW Monitoring Report.pdf
https://www.deq.state.or.us/Webdocs/Forms/Output/FPController.ashx?SourceId=3324&SourceIdType=11
US, OR
AIRPORT
AFFF

142. PDX_SW
Portland International Airport
Q4 2019 GW Monitoring Report.pdf
https://www.deq.state.or.us/Webdocs/Forms/Output/FPController.ashx?SourceId=3324&SourceIdType=11
US, OR
AIRPORT
AFFF

143. ALY_2020_SW
Surface water samples from ITC fire in Deer Park Houston; Texas; USA: March - August 2019
Aly et al., Temporal and spatial analysis of per and polyfluoroalkyl substances in surface waters of Houston ship channel following a large-scale industrial fire incident. EP, 2020, 265, 115009.
https://www.sciencedirect.com/science/article/pii/S0269749120321254
US, TX
SW
AFFF

144. HUSET_2011_LF_GW
Municipal solid waste landfill leachate in 4 locations. USA: 2006 - March 2011
Huset et al., Quantitative determination of fluorochemicals in municipal landfill leachates. Chemosphere, 2011, 82, 1380-1386.
https://www.sciencedirect.com/science/article/pii/S0045653510013809
US, Various
LANDFILL
nonAFFF

145. LANG_2017_LF_GW
Lang et al., National Estimate of Per- and Polyfluoroalkyl Substance (PFAS) Release to U.S. Municipal Landfill Leachate. ES&T, 51, 2197-2205.
https://pubs.acs.org/doi/10.1021/acs.est.6b05005
US, Various
LANDFILL
nonAFFF

146. EPA_UCMR3_GWSW
EPA UCMR3 Sampling of Drinking Water
ucmr-3-occurrence-data.zip
https://www.epa.gov/dwucmr/occurrence-data-unregulated-contaminant-monitoring-rule
US, Various
WTP
UNK

147. CLARENDON_GW_OFF
Rutland Rermont Airport
20184763.SVRA.2018.DW.SamplingSummary.rpt.pdf (+ more detailed summary from web with all analytes)
https://anrweb.vt.gov/PubDocs/DEC/Hazsites/20184763.SVRA.2018.DW.SamplingSummary.rpt.pdf
US, VT
AIRPORT
AFFF

148. CLARENDON_GW_ON

Rutland Rermont Airport
20184763.RutlandSVRA.SI.2018.rpt.pdf
https://anrweb.vt.gov/PubDocs/DEC/Hazsites/20184763.RutlandSVRA.SI.2018.rpt.pdf
US, VT
AIRPORT
AFFF

149. CLARENDON_SOIL_ON
Rutland Rermont Airport
20184763.RutlandSVRA.SI.2018.rpt.pdf
https://anrweb.vt.gov/PubDocs/DEC/Hazsites/20184763.RutlandSVRA.SI.2018.rpt.pdf
US, VT
AIRPORT
AFFF

150. COVENTRY_LF_SLUDGE
New England Waste Services (NEWSVT) Coventry
PFAS WASTE SOURCE TESTING REPORT NEW ENGLAND WASTE SERVICES OF VERMONT
https://legislature.vermont.gov/Documents/2020/WorkGroups/House Natural/Contaminants/W~Matt Chapman~PFAS Waste Source Testing Report~1-16-2020.pdf
US, VT
LANDFILL
WWTP

151. COVENTRY_LF_SOIL
New England Waste Services (NEWSVT) Coventry
PFAS WASTE SOURCE TESTING REPORT NEW ENGLAND WASTE SERVICES OF VERMONT
https://legislature.vermont.gov/Documents/2020/WorkGroups/House Natural/Contaminants/W~Matt Chapman~PFAS Waste Source Testing Report~1-16-2020.pdf
US, VT
LANDFILL
MIXED

152. BENNINGTON_GW
Bennington Vermont
2019.0510.Data.bennington.water.supplies - JS.xlsx
https://dec.vermont.gov/pfas/pfoa/communities
Detailed data provided by personal communication.
US, VT
RESIDENTIAL_WELLS
COATINGS

153. BENNINGTON_SOIL
Bennington Vermont

07-2016-Final-Draft-Shallow-Soil-Sampling.pdf
https://dec.vermont.gov/pfas/pfoa/communities
US, VT
SOIL
COATINGS

154. COVENTRY_LF_SOURCE
New England Waste Services (NEWSVT) Coventry
PFAS WASTE SOURCE TESTING REPORT NEW ENGLAND WASTE SERVICES OF VERMONT
https://legislature.vermont.gov/Documents/2020/WorkGroups/House Natural/Contaminants/W~Matt Chapman~PFAS Waste Source Testing Report~1-16-2020.pdf
US, VT
LANDFILL
COATINGS

155. VTWWTF_EFF_WWTP
Vermont Statewide
Poly- and Perfluoroalkyl Substances at Wastewater Treatment Facilities and Landfill Leachate 2019 Summary Report
https://dec.vermont.gov/sites/dec/files/wmp/SolidWaste/Documents/02.03.20_PFAS in LF and WWTF Final Report.pdf
US, VT
WWTP_EFFLUENT
UNK

156. VTWWTF_INF_WWTP
Vermont Statewide
Poly- and Perfluoroalkyl Substances at Wastewater Treatment Facilities and Landfill Leachate 2019 Summary Report
https://dec.vermont.gov/sites/dec/files/wmp/SolidWaste/Documents/02.03.20_PFAS in LF and WWTF Final Report.pdf
US, VT
WWTP_INFLUENT
UNK

157. VTWWTF_SLUDGE
Vermont Statewide
Poly- and Perfluoroalkyl Substances at Wastewater Treatment Facilities and Landfill Leachate 2019 Summary Report
https://dec.vermont.gov/sites/dec/files/wmp/SolidWaste/Documents/02.03.20_PFAS in LF and WWTF Final Report.pdf
US, VT
WWTP_SLUDGE
UNK

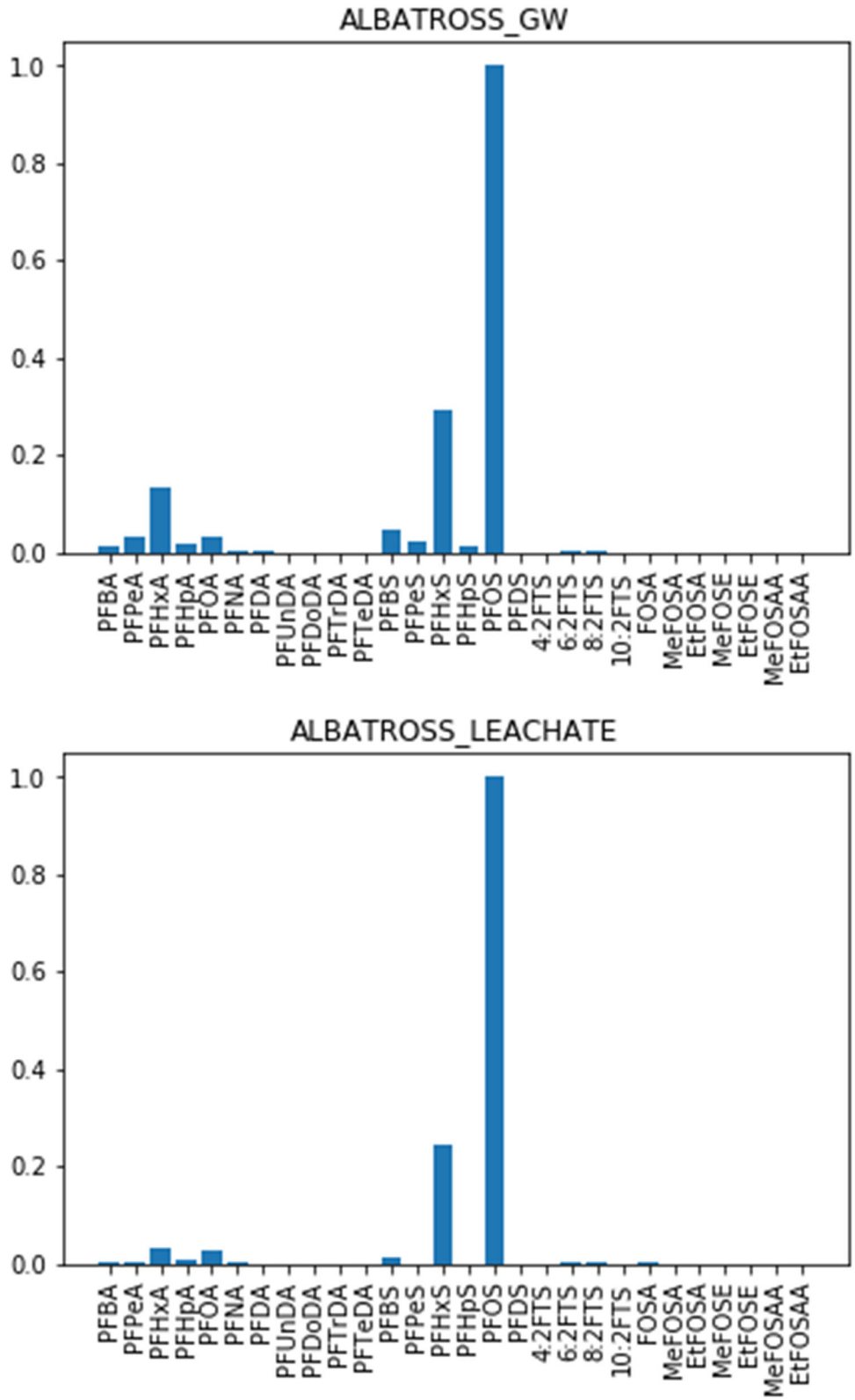
158. VT_SHALLOW_SOIL
Various locations in Vermont
PFAS BACKGROUND IN VERMONT SHALLOW SOILS
https://anrweb.vt.gov/PubDocs/DEC/PFOA/Soil-Background/PFAS-Background-Vermont-Shallow-Soils-03-24-19.pdf
US, VT
ENV
ATM

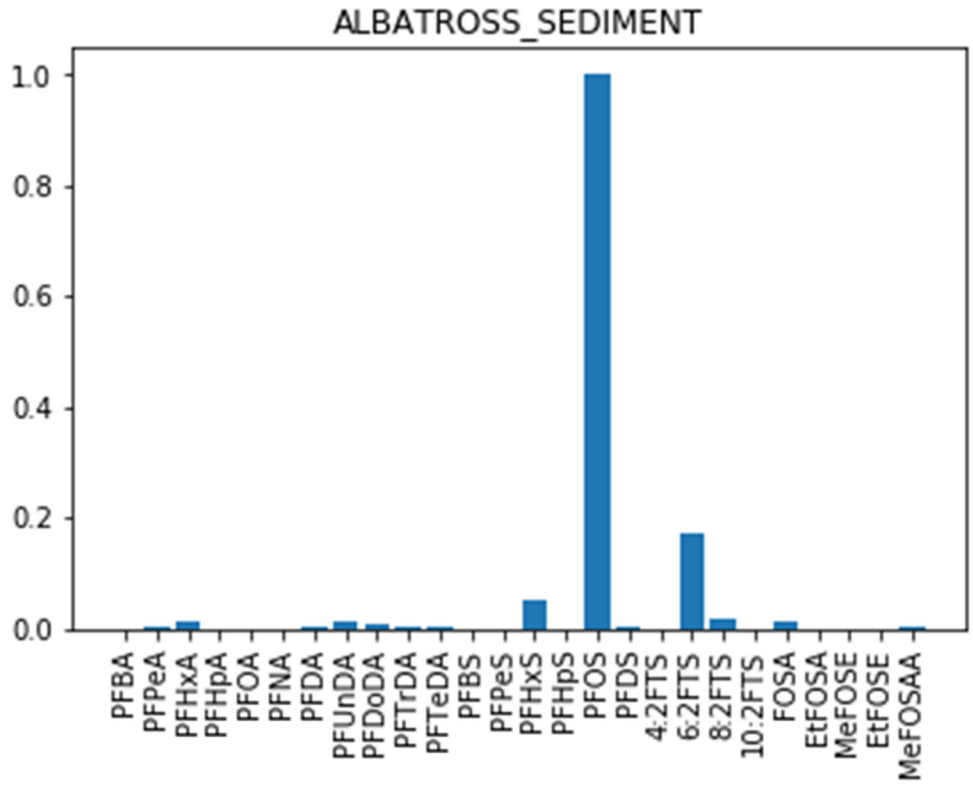
159. MARINETTE_OFF_GW
Residential wells surrounding the Tyco Fire Products
TOWN OF PESHTIGO FINAL WATER SAMPLE RESULTS: DECEMBER 2017 - SEPTEMBER 2020
https://dnr.wi.gov/botw/GetActivityDetail.do;jsessionid=4JHfhvuzoRA2P36uArdpzcwujX6s-T_vIQPHFGU_xfZ9y-_CtUUX!1987276291?detailSeqNo=581955&siteId=643000&crumb=1&search=b
US, WI
RESIDENTIAL_WELLS
AFFF

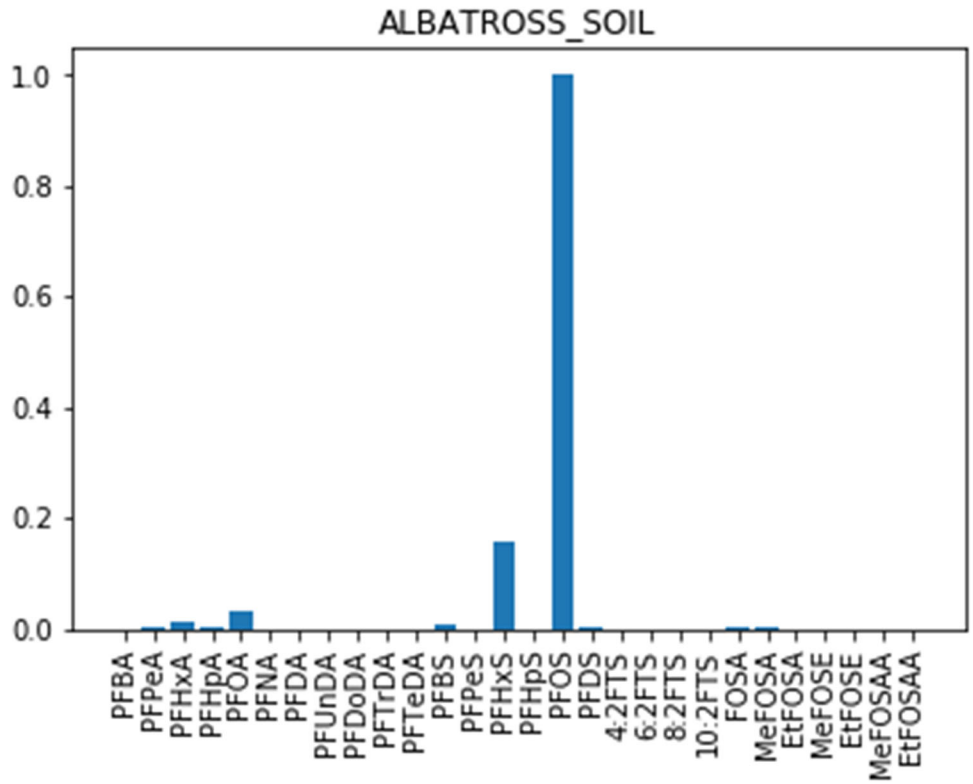
160. MARINETTE_ON_GWSW
GW and SW samples from the area surrounding the Tyco Fire Products
20191231_43_Data_Summary_Rpt.pdf and 20200204_43_Stanton_Status_Update.pdf
https://dnr.wi.gov/botw/GetActivityDetail.do;jsessionid=4JHfhvuzoRA2P36uArdpzcwujX6s-T_vIQPHFGU_xfZ9y-_CtUUX!1987276291?detailSeqNo=581955&siteId=643000&crumb=1&search=b
US, WI
PFAS_MFG
AFFF

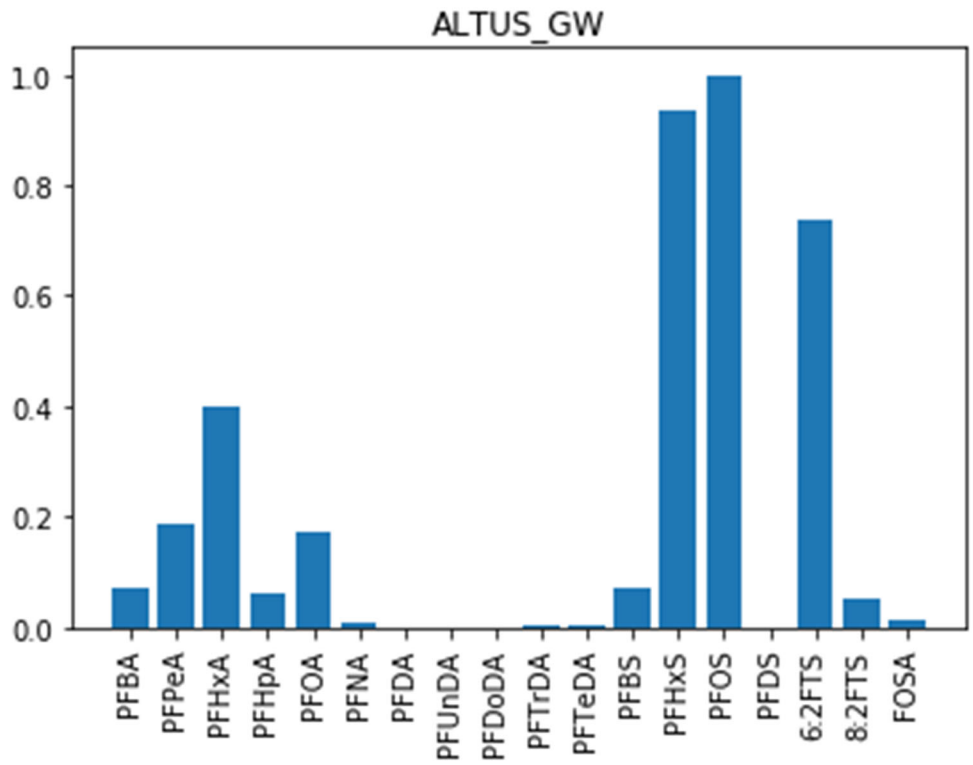
161. MARINETTE_ON_SOIL
SOIL samples from the area surrounding the Tyco Fire Products
20191231_43_Data_Summary_Rpt.pdf and 20200204_43_Stanton_Status_Update.pdf
https://dnr.wi.gov/botw/GetActivityDetail.do;jsessionid=4JHfhvuzoRA2P36uArdpzcwujX6s-T_vIQPHFGU_xfZ9y-_CtUUX!1987276291?detailSeqNo=581955&siteId=643000&crumb=1&search=b
US, WI
PFAS_MFG
AFFF

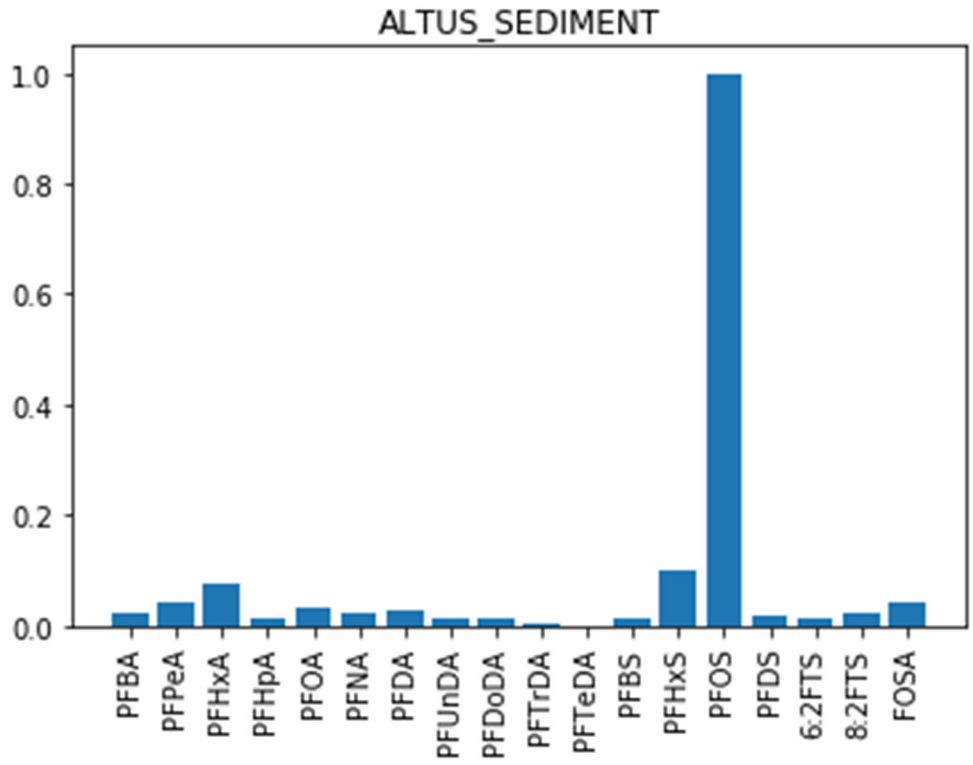
Appendix B. Concentration-Averaged Component Distributions

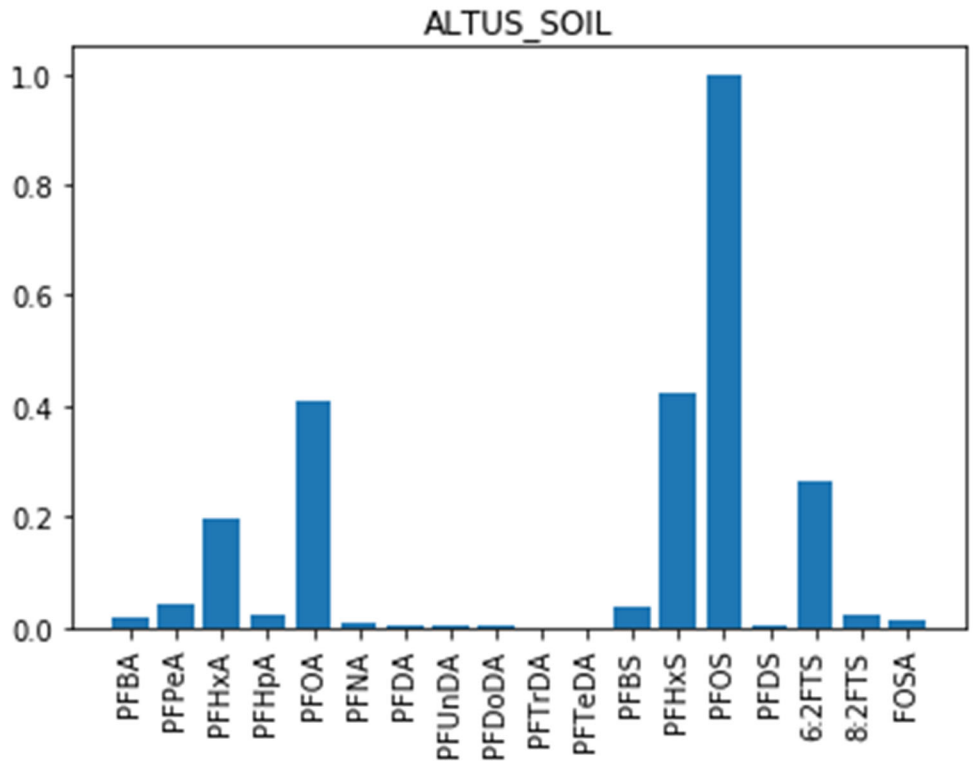


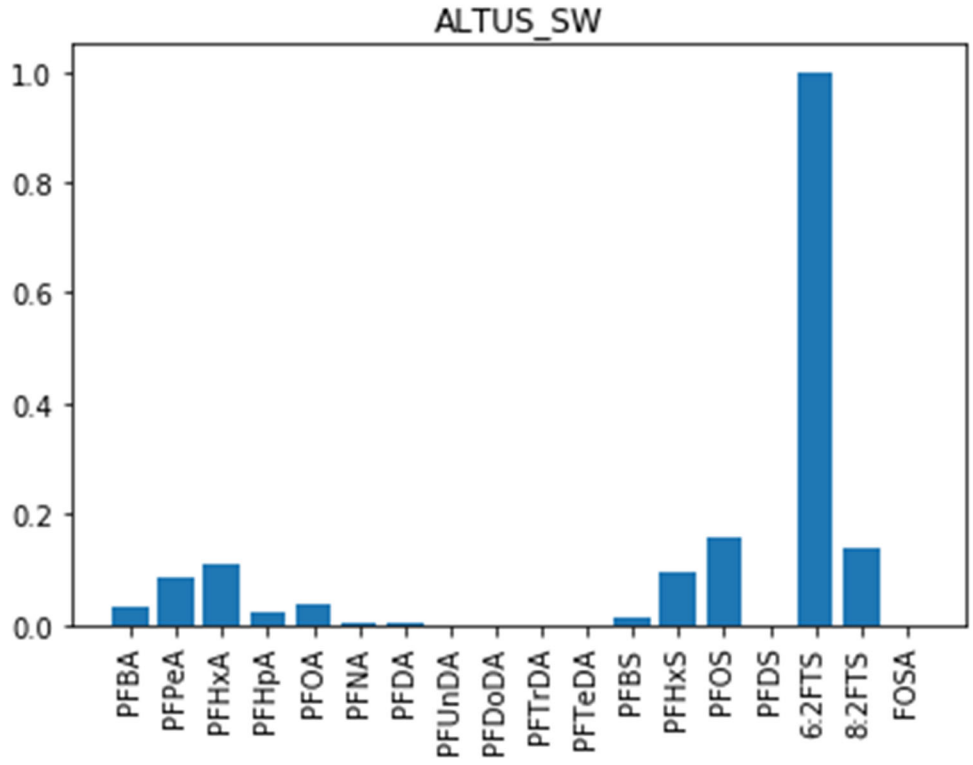

ALBATROSS_SEDIMENT
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA

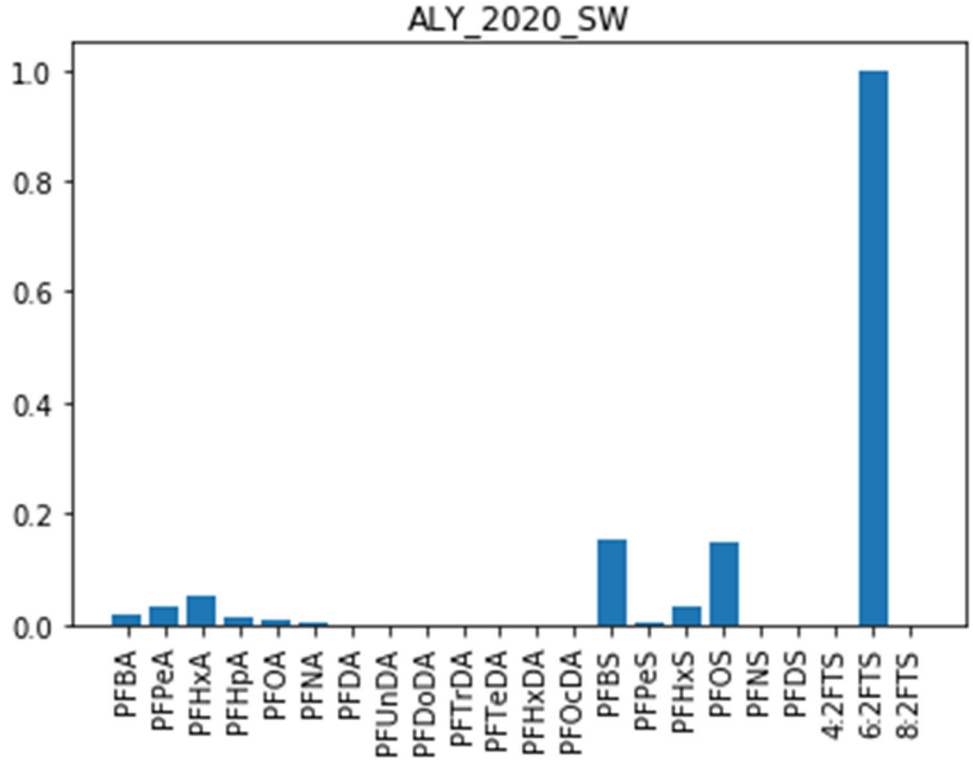

ALBATROSS_SOIL
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

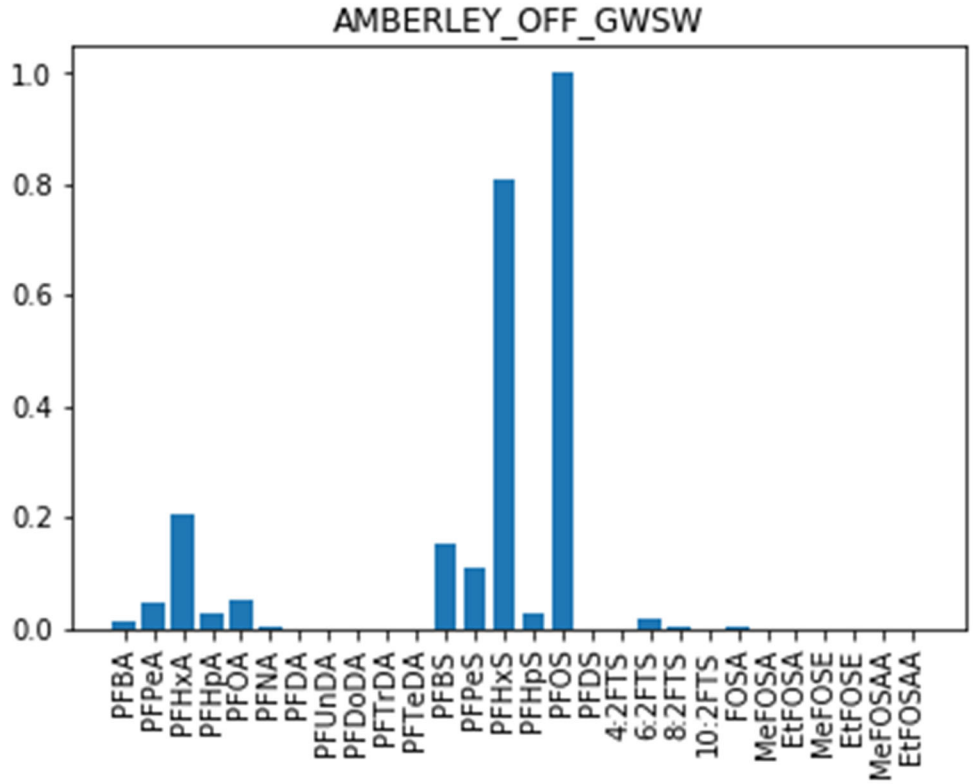

ALTUS_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFOS
PFDS
6:2FTS
8:2FTS
FOSA

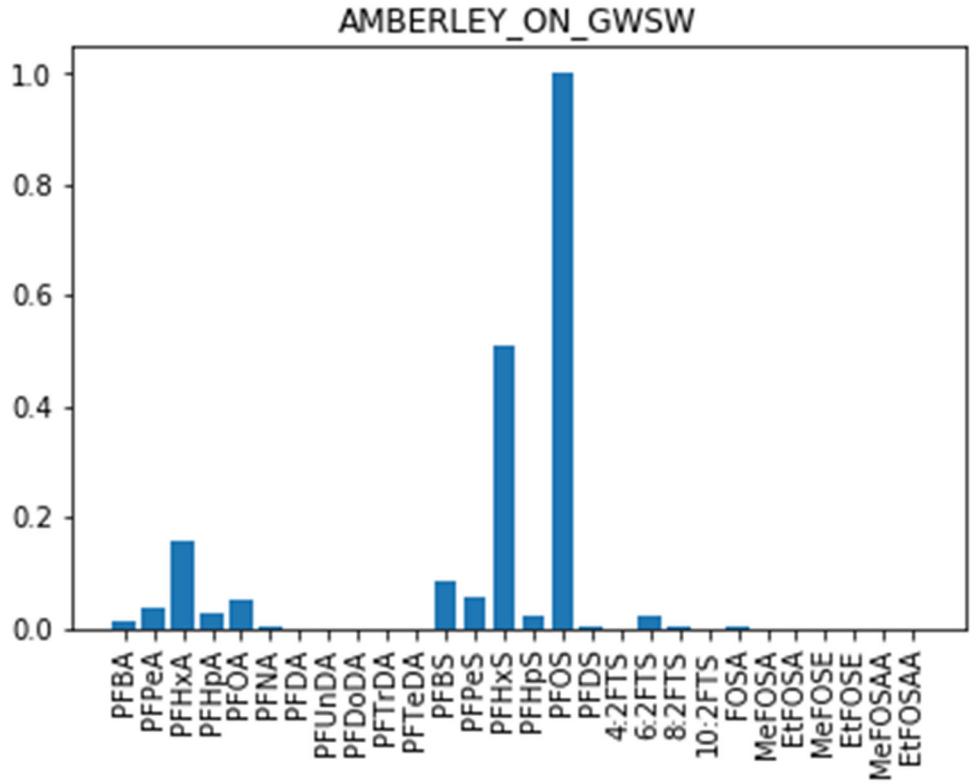

ALTUS_SEDIMENT
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFOS
PFDS
6:2FTS
8:2FTS
FOSA

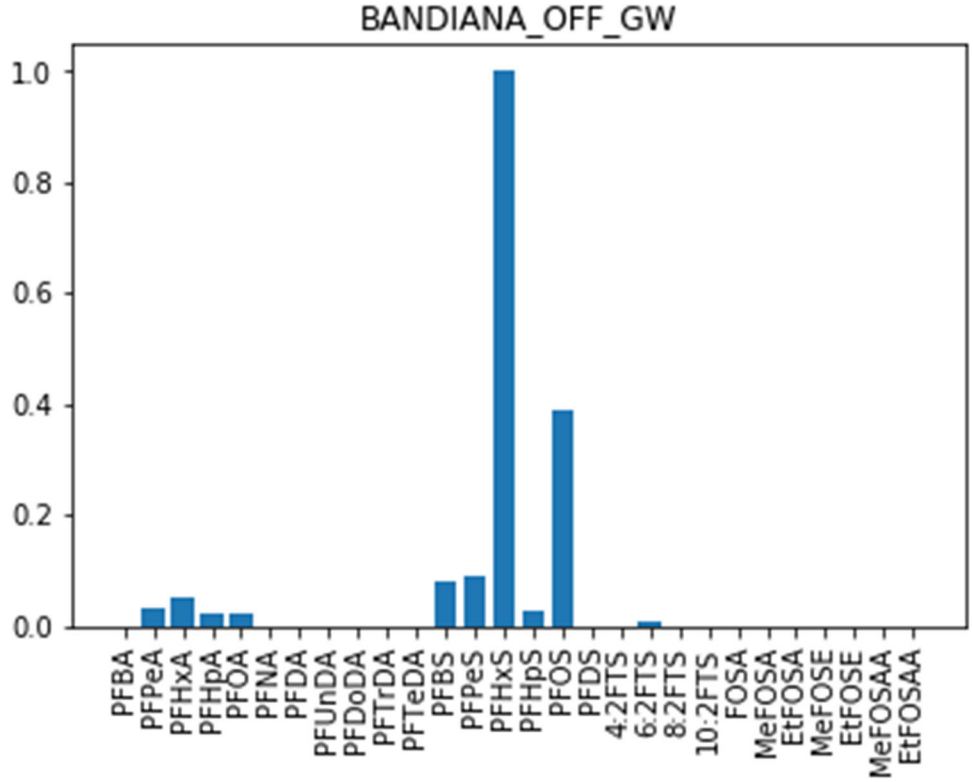

ALTUS_SOIL
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFOS
PFDS
6:2FTS
8:2FTS
FOSA

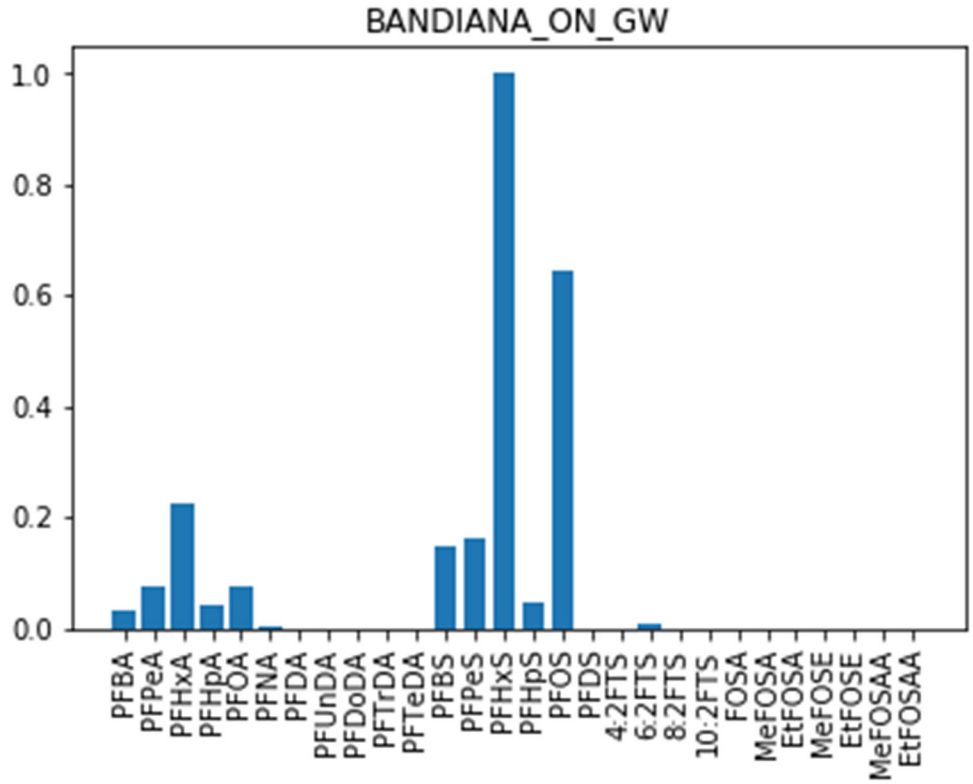

ALTUS_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFOS
PFDS
6:2FTS
8:2FTS
FOSA

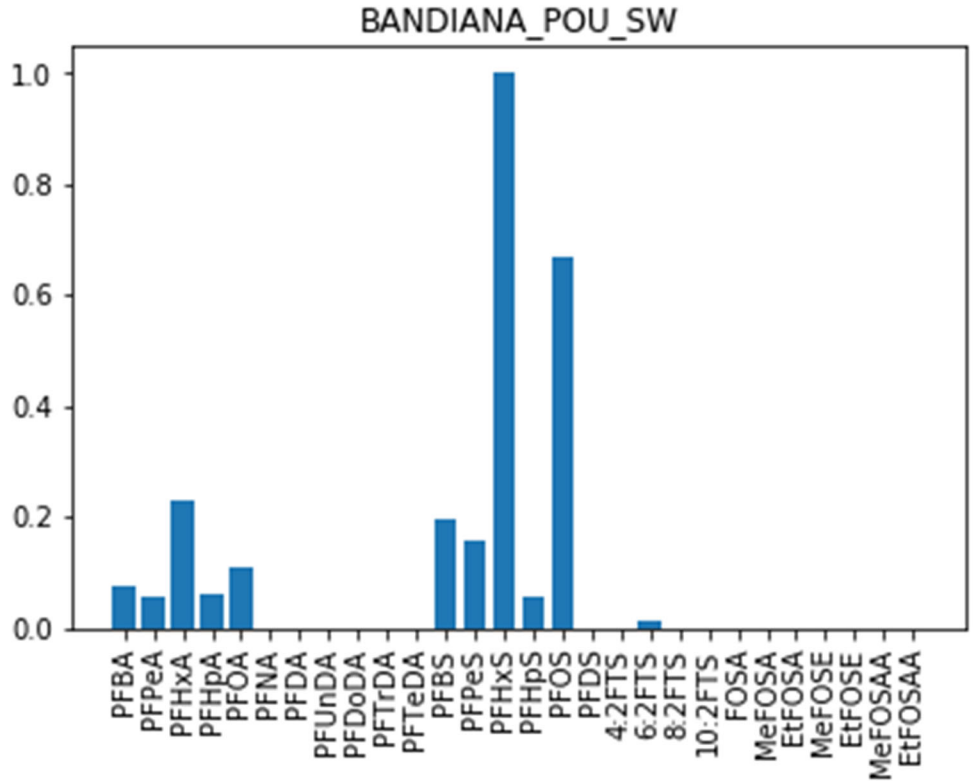

ALY_2020_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFHxDA
PFOcDA
PFBS
PFPeS
PFHxS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS

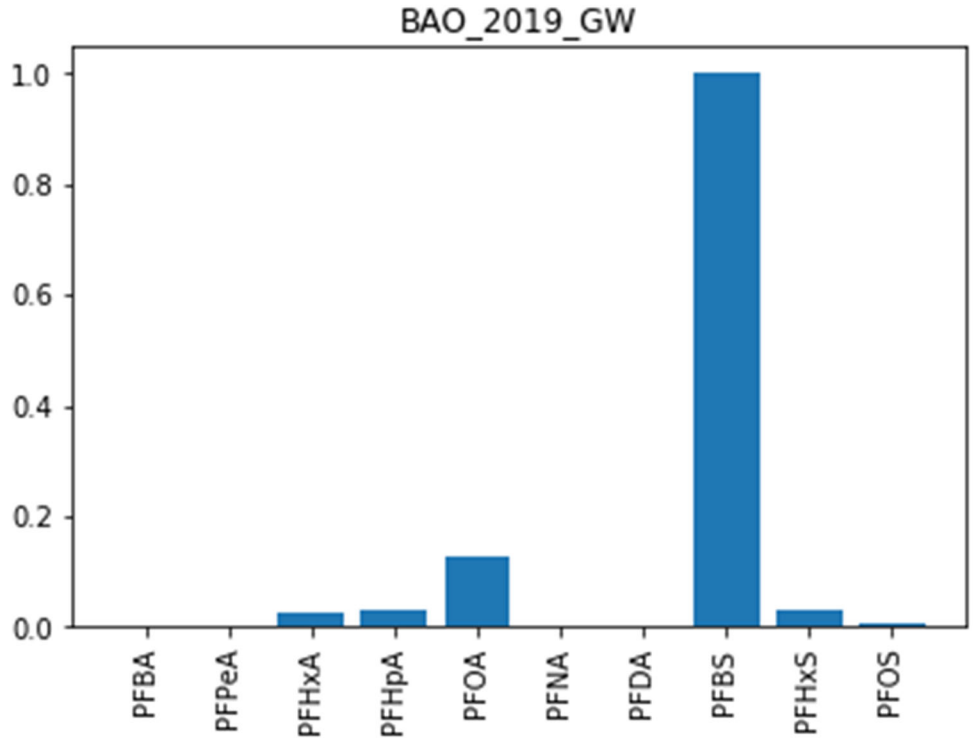

AMBERLEY_OFF_GWSW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

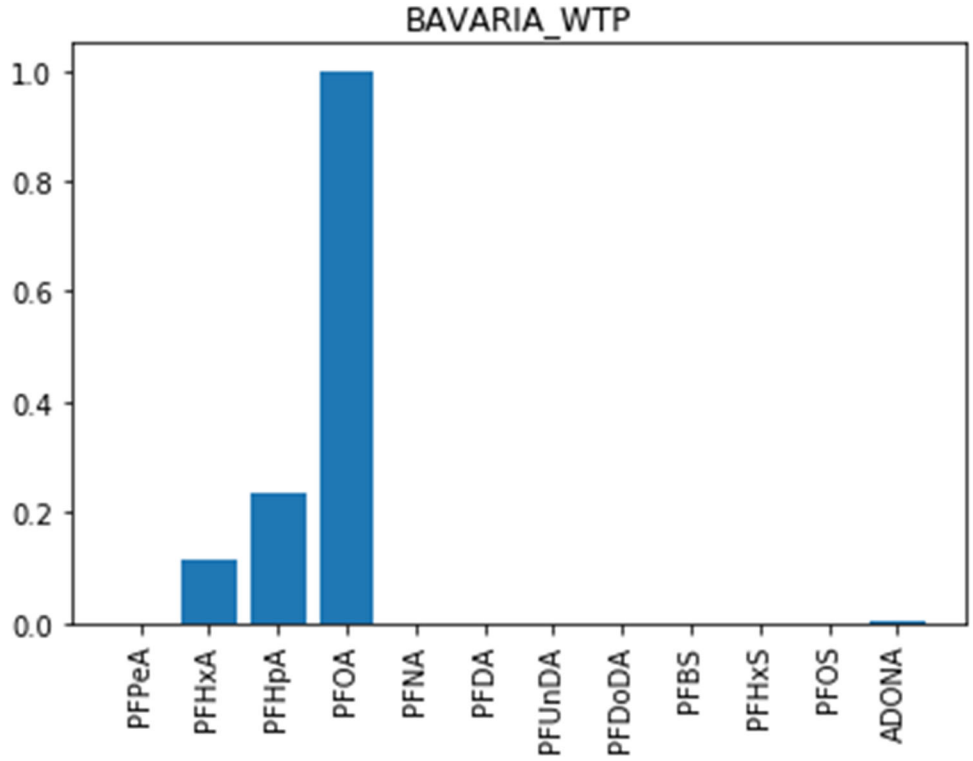

AMBERLEY_ON_GWSW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

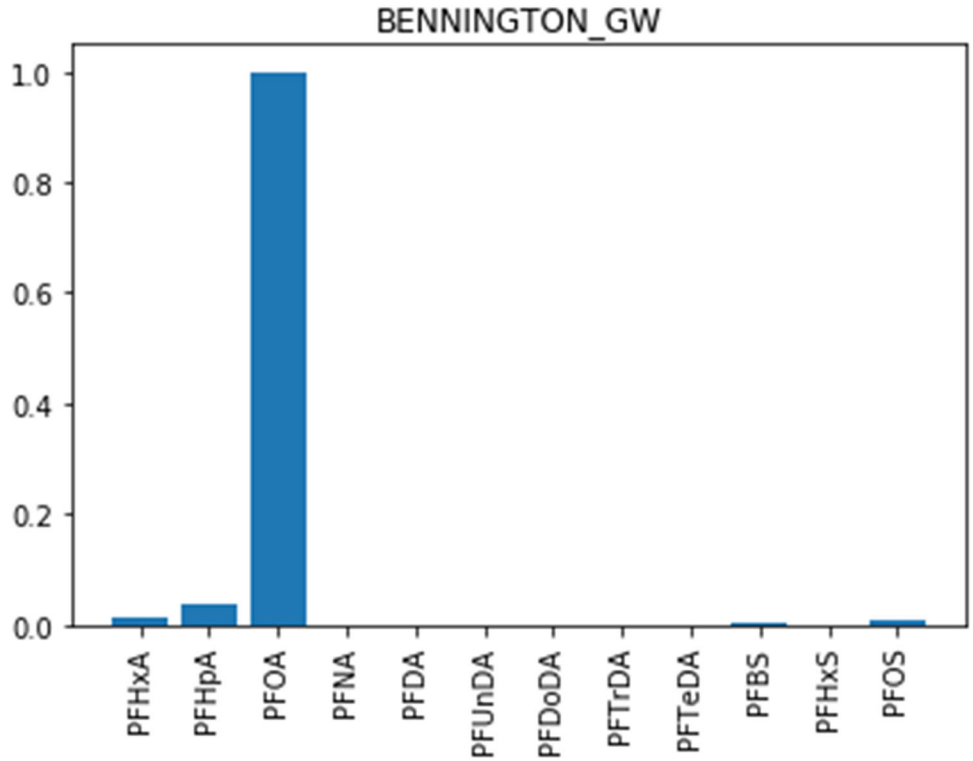

BANDIANA_OFF_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

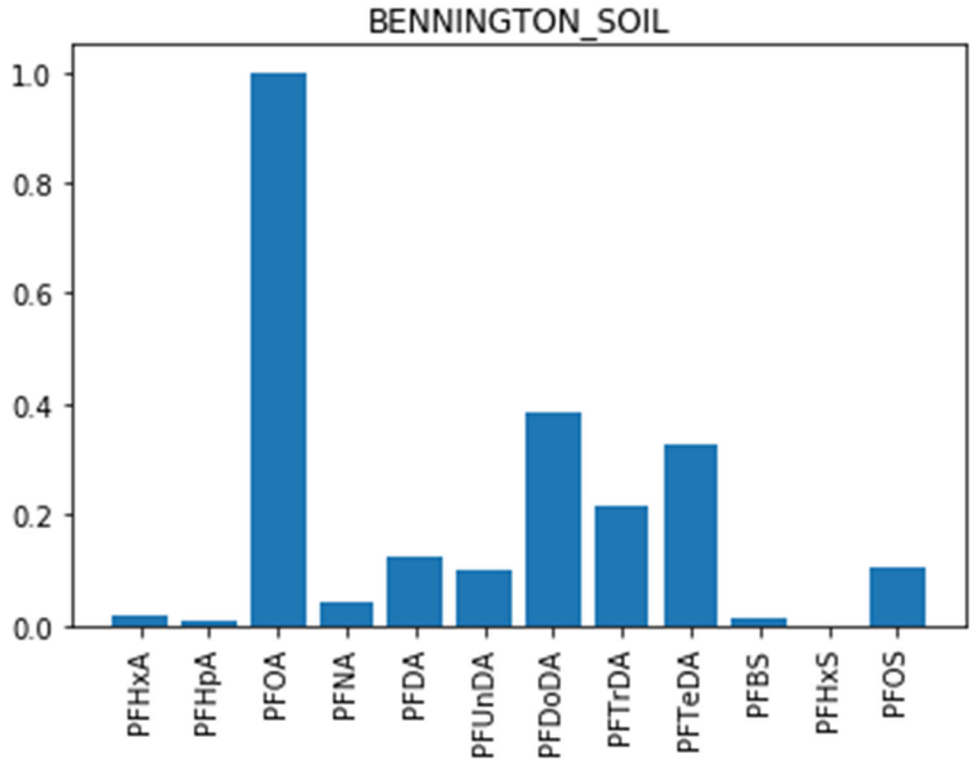

BANDIANA_ON_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

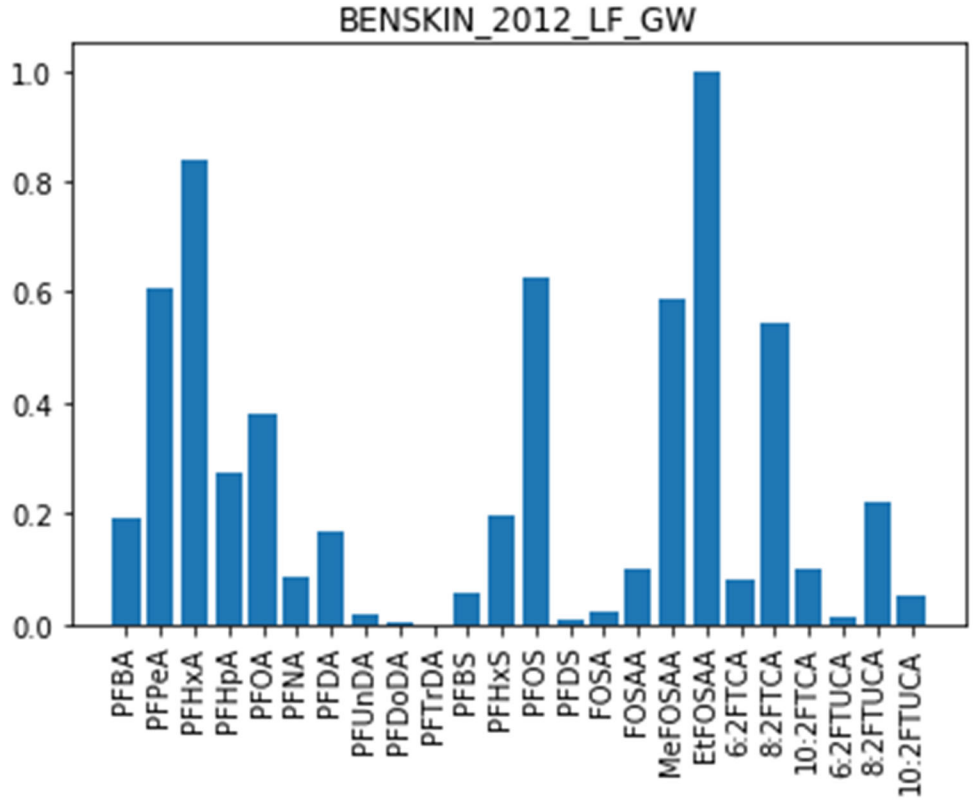

BANDIANA_POU_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

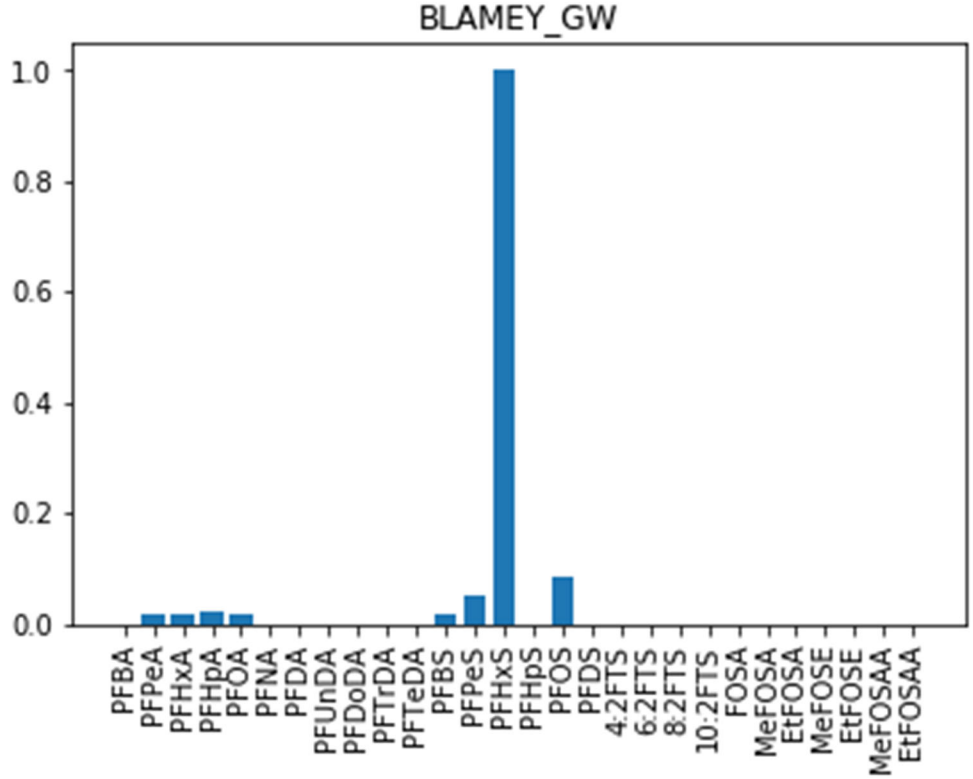

BAO_2019_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFBS
PFHxS
PFOS

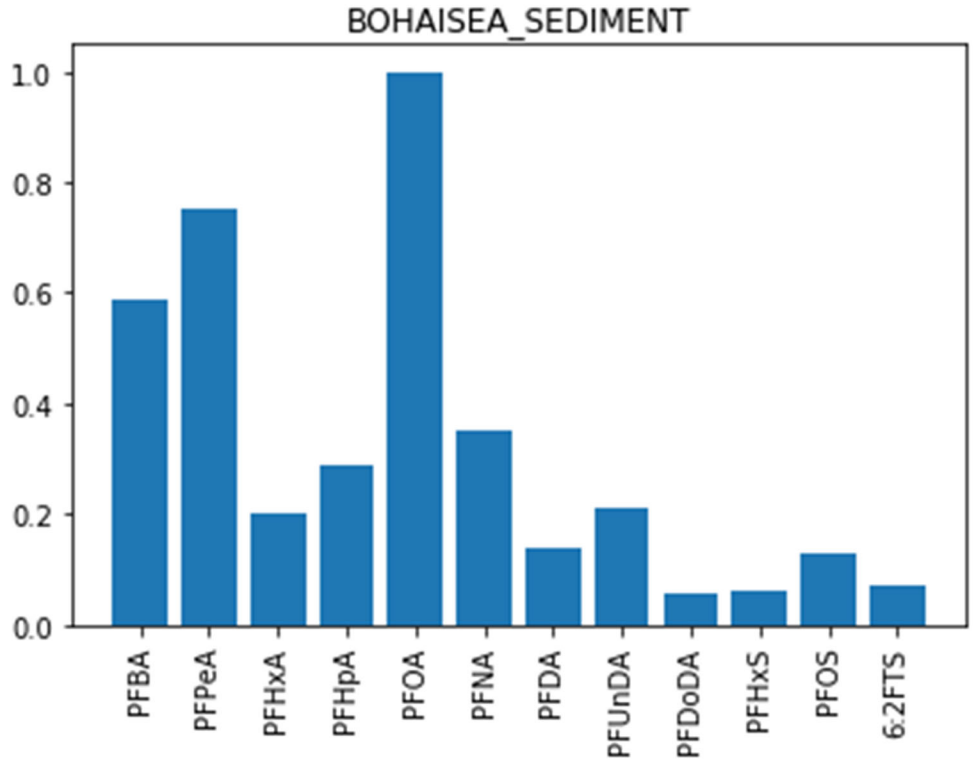

BAVARIA_WTP
1.0
0.8
0.6
0.4
0.2
0.0
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFBS
PFHxS
PFOS
ADONA

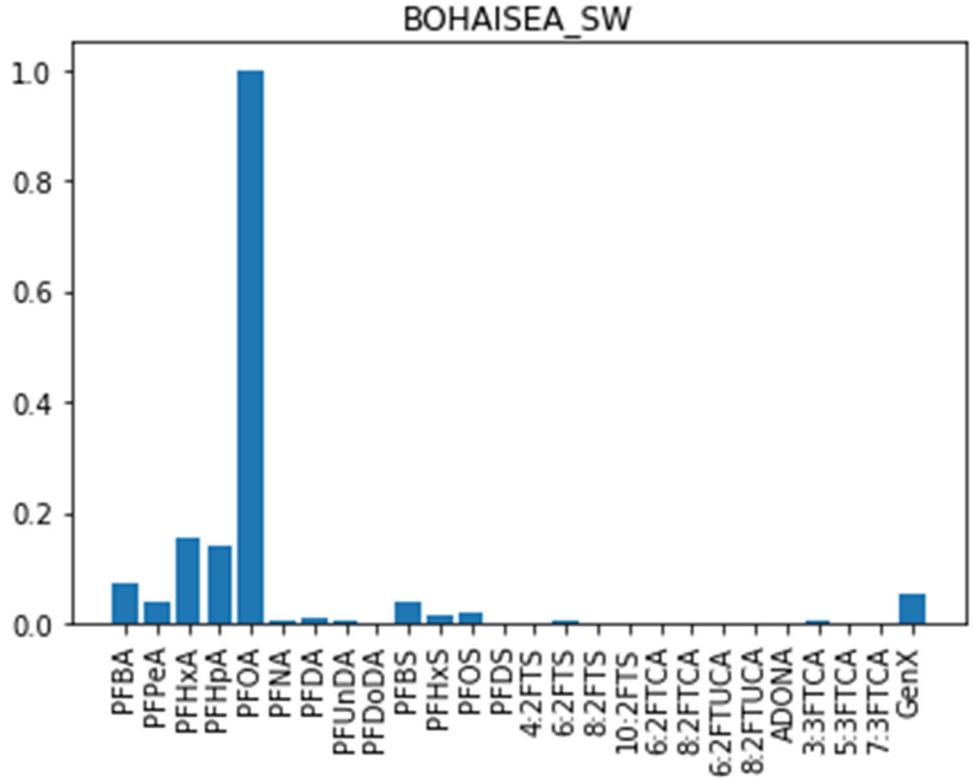

BENNINGTON_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFOS

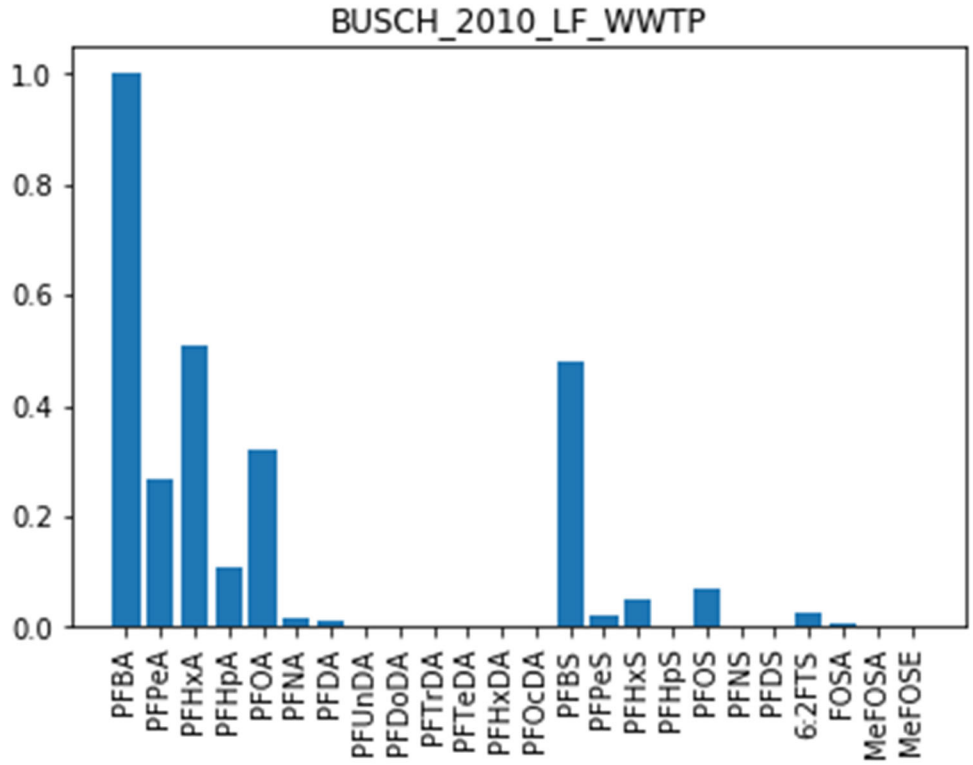

BENNINGTON_SOIL
1.0
0.8
0.6
0.4
0.2
0.0
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFOS

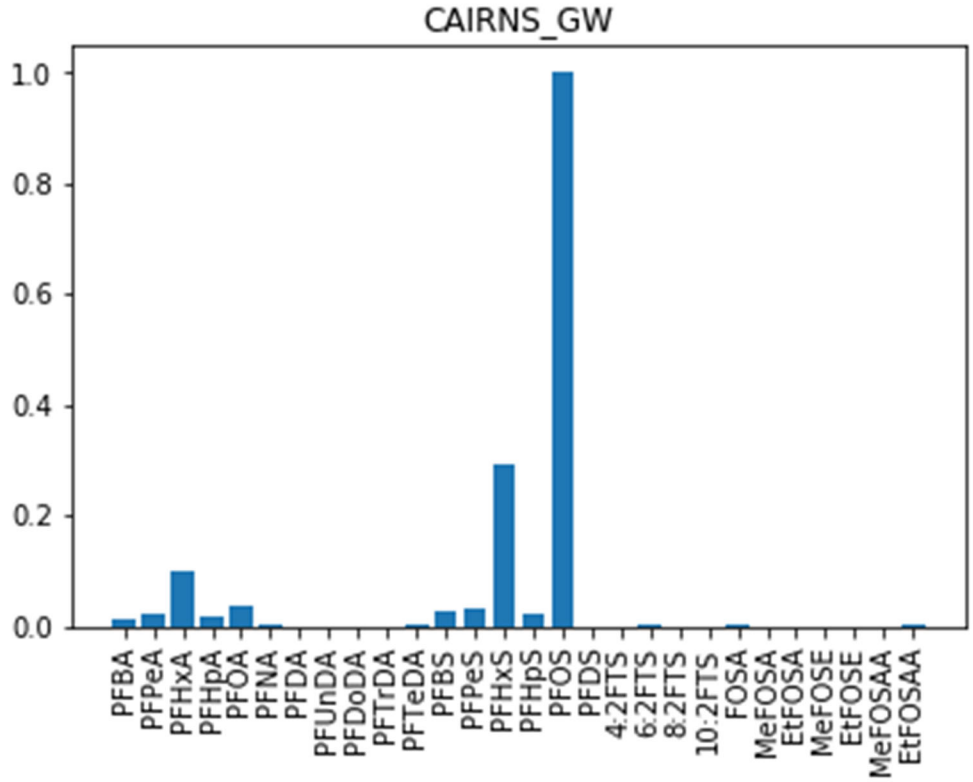

BENSKIN_2012_LF_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFBS
PFHxS
PFOS
PFDS
FOSA
FOSAA
MeFOSAA
EtFOSAA
6:2FTCA
8:2FTCA
10:2FTCA
6:2FTUCA
8:2FTUCA
10:2FTUCA

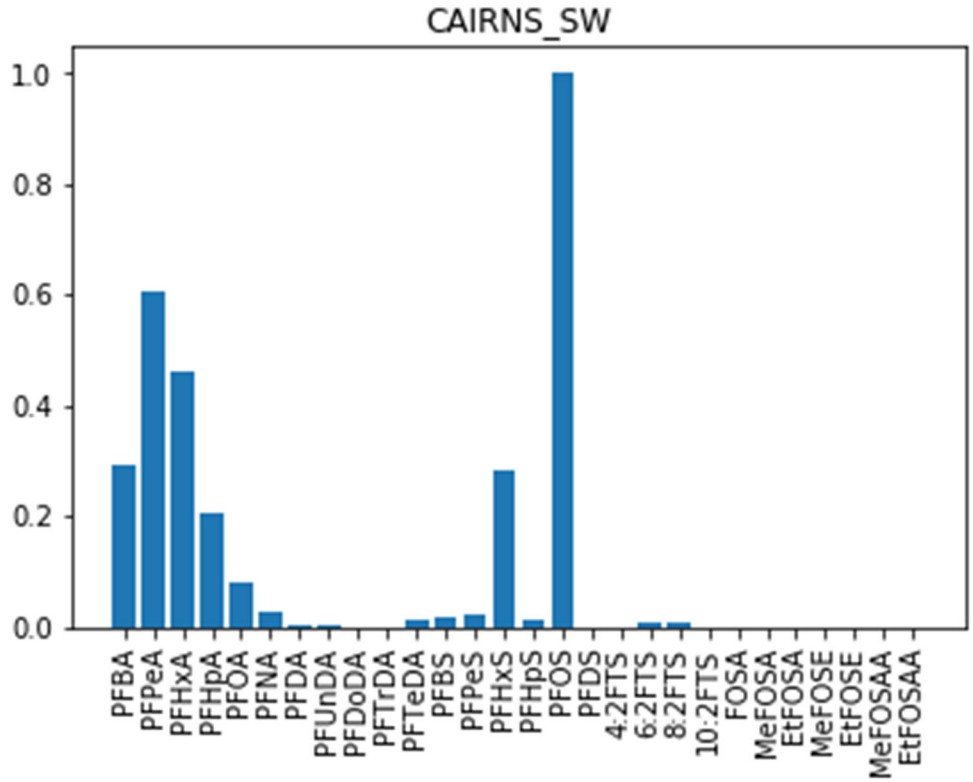

BLAMEY_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

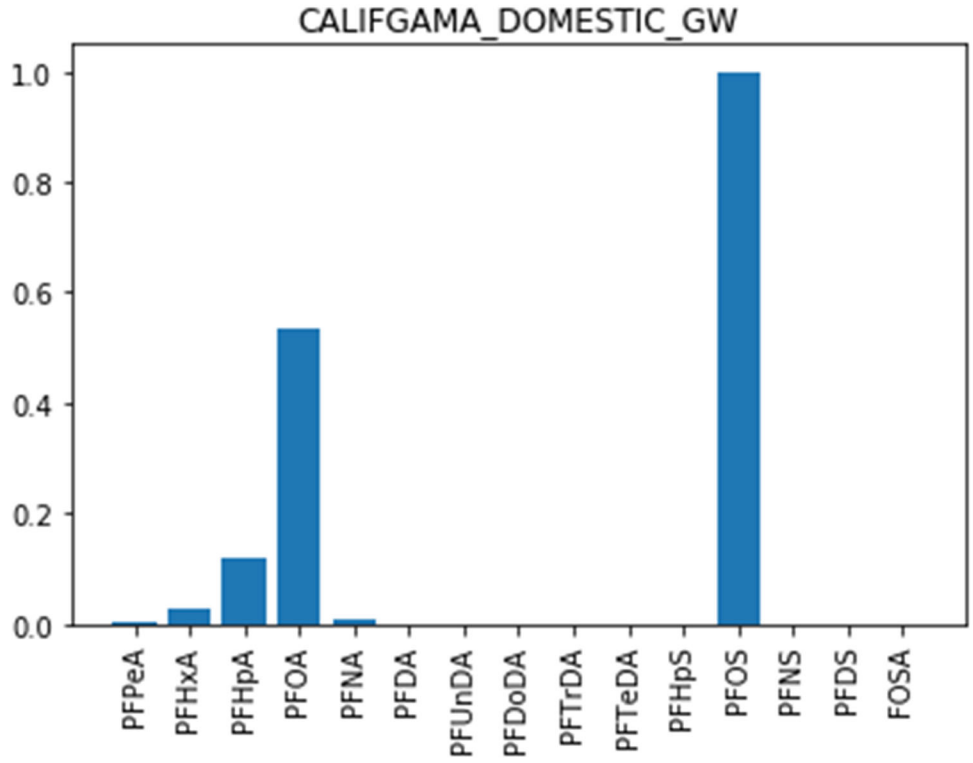

BOHAISEA_SEDIMENT
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFHxS
PFOS
6:2FTS

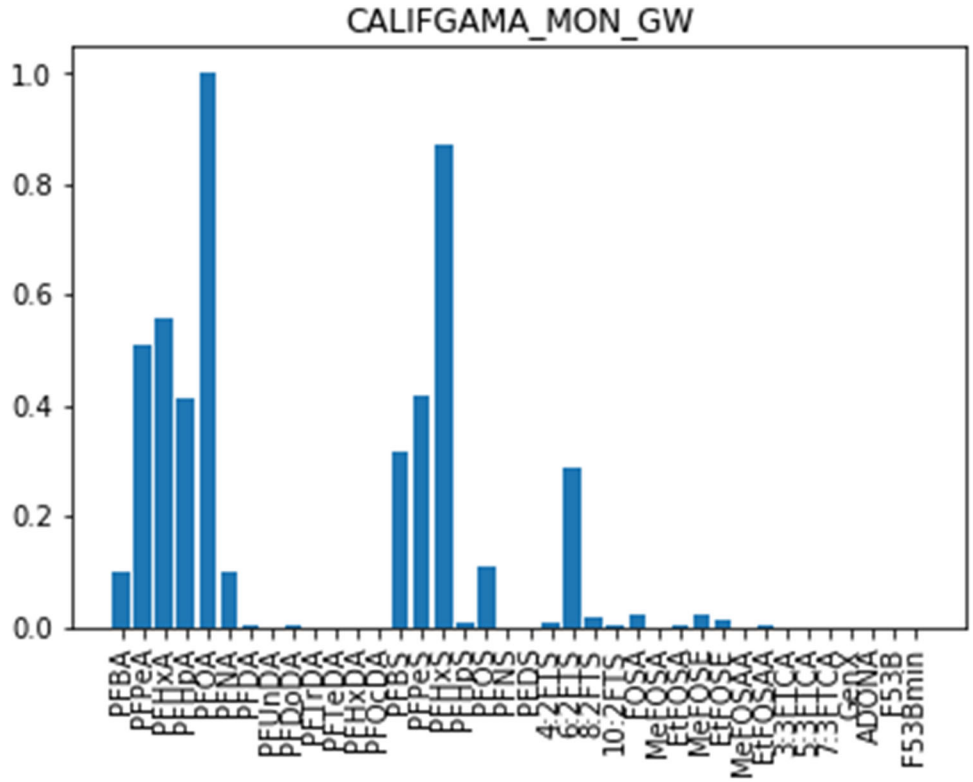

BOHAISEA_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFBS
PFHxS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
6:2FTCA
8:2FTCA
6:2FTUCA
8:2FTUCA
ADONA
3:3FTCA
5:3FTCA
7:3FTCA
GenX

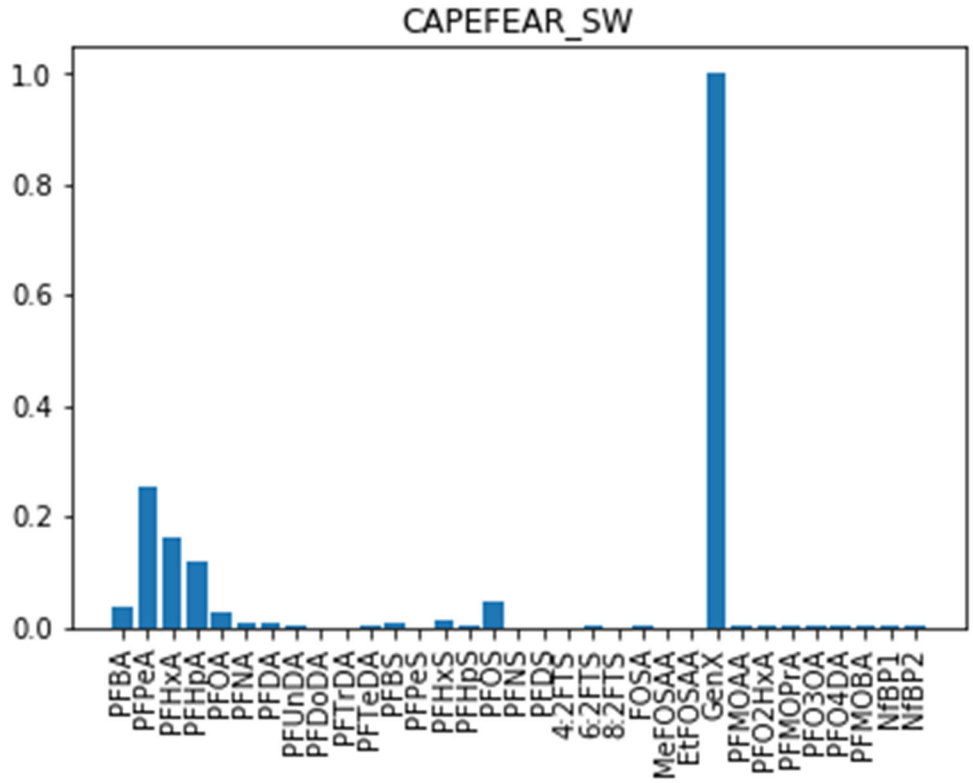

BUSCH_2010_LF_WWTP
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFHxDA
PFOcDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
6:2FTS
FOSA
MeFOSA
MeFOSE

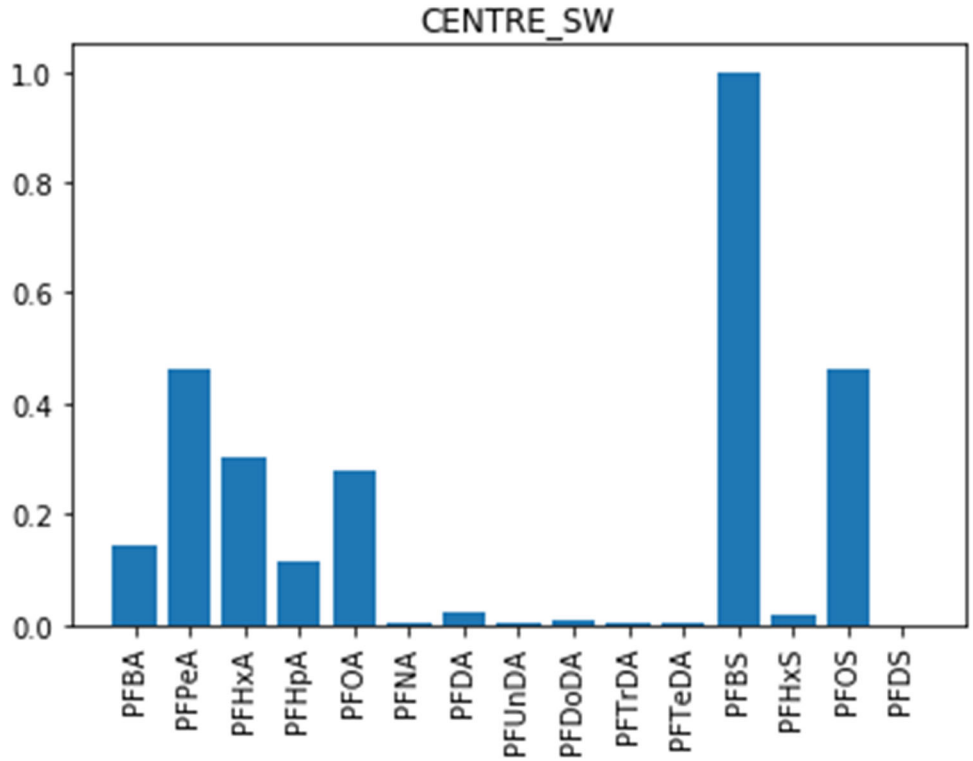

CAIRNS_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

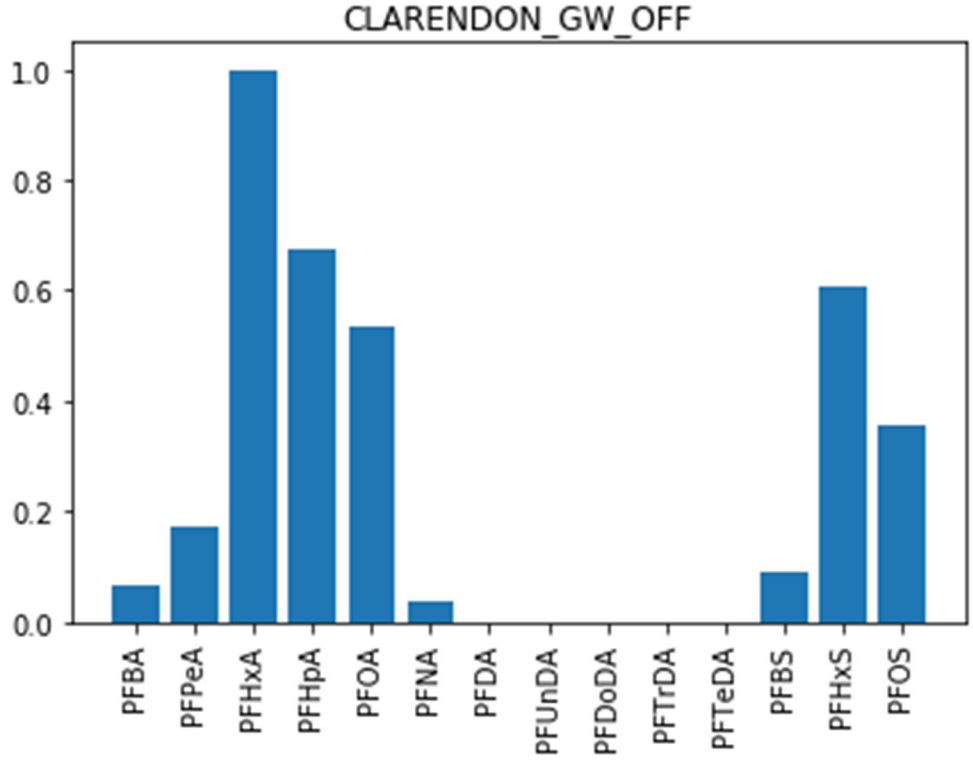

CAIRNS_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

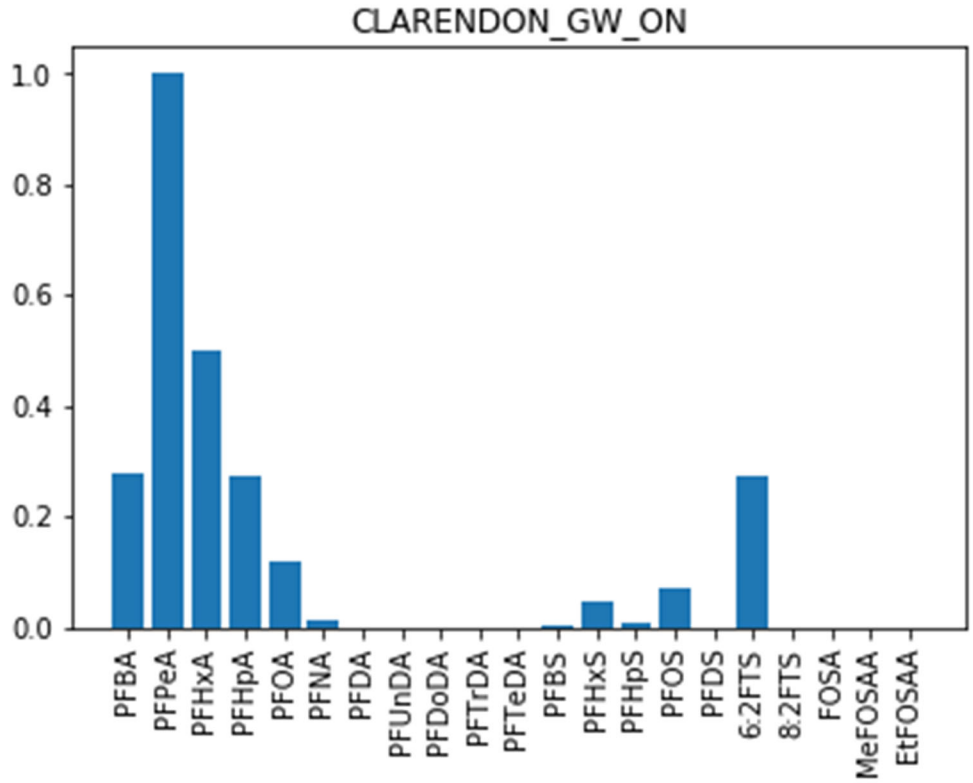

CALIFGAMA_DOMESTIC_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFHpS
PFOS
PFNS
PFDS
FOSA

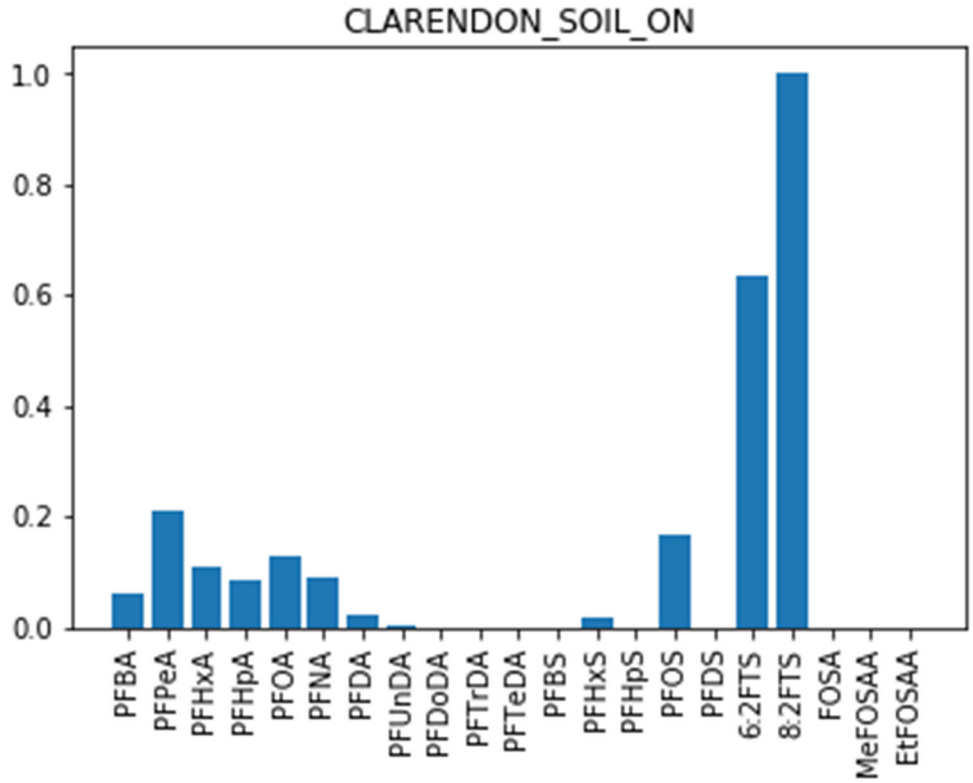

CALIFGAMA_MON_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFHxDA
PFOcDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSAA
EtFOSAA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA
3:3FTCA
5:3FTCA
7:3FTCA
GenX
ADONA
F53B
F53Bmin

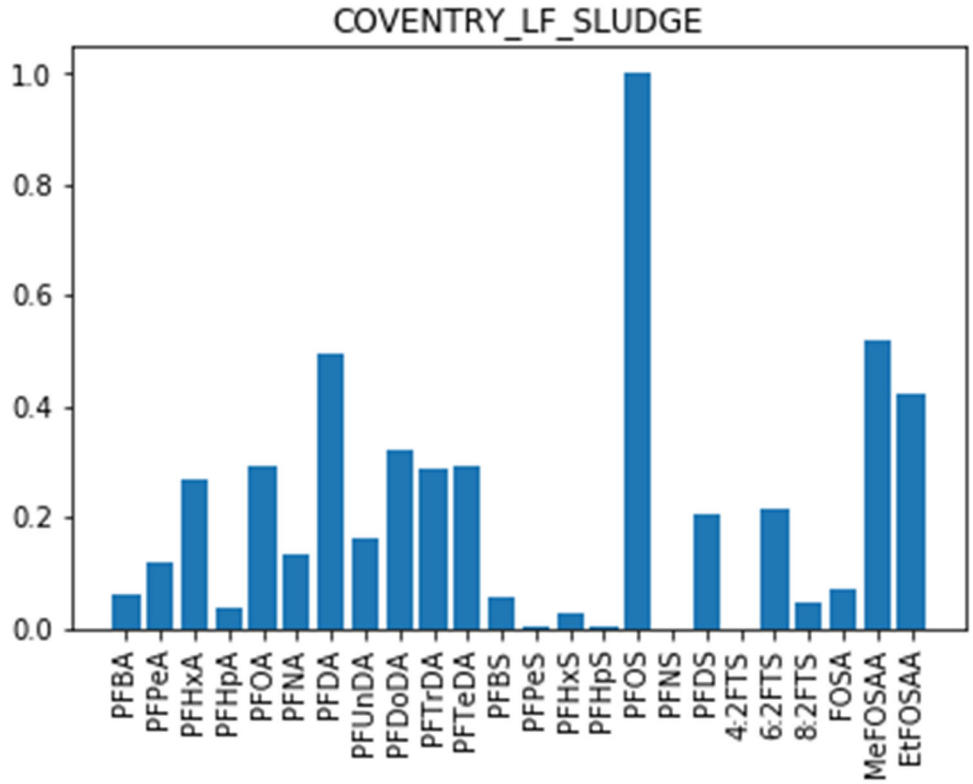

CAPEFEAR_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
FOSA
MeFOSAA
EtFOSAA
GenX
PFMOAA
PFO2HxA
PFMOPrA
PFO3OA
PFO4DA
PFMOBA
NfBP1
NfBP2

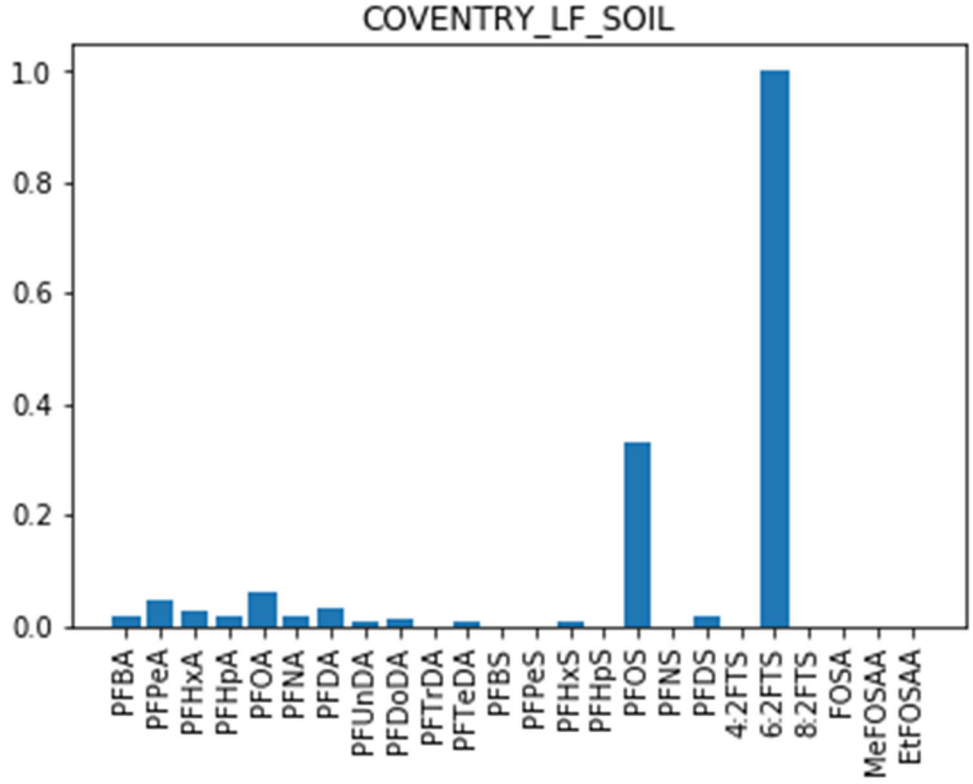

CENTRE_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFOS
PFDS

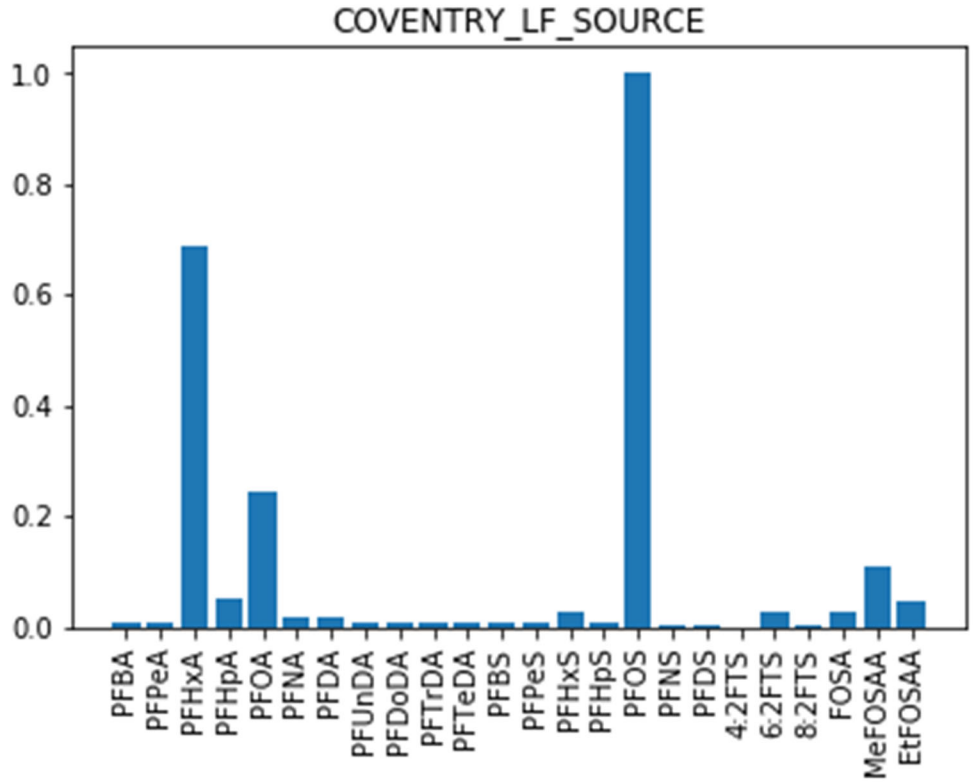

CLARENDON_GW_OFF
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFOS

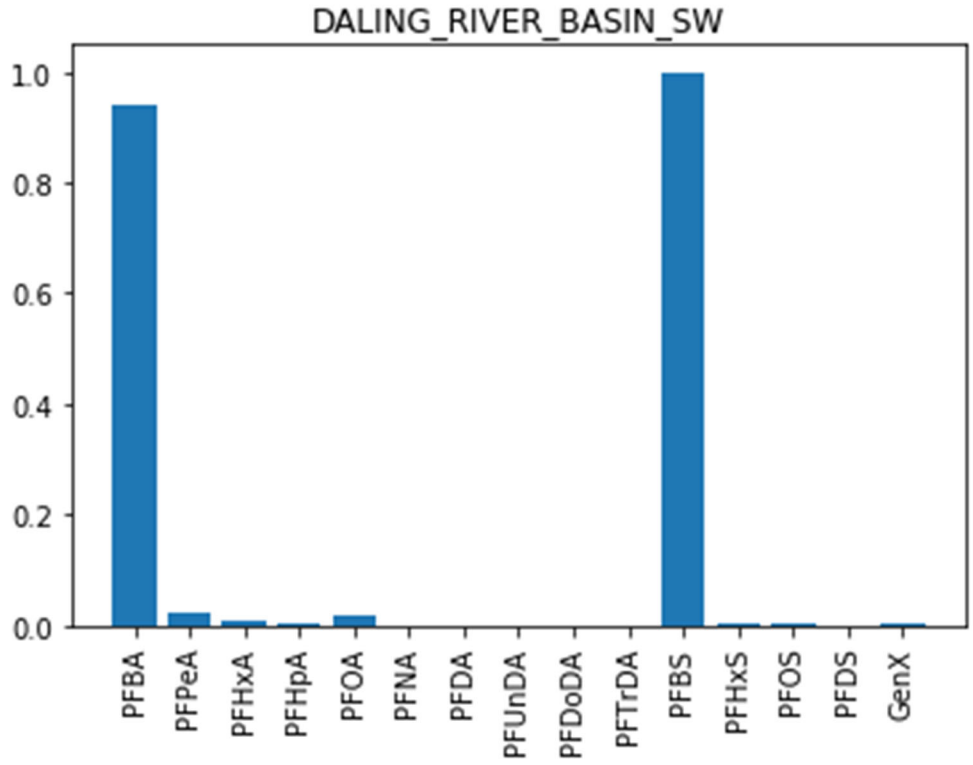

CLARENDON_GW_ON
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFHpS
PFOS
PFDS
6:2FTS
8:2FTS
FOSA
MeFOSAA
EtFOSAA

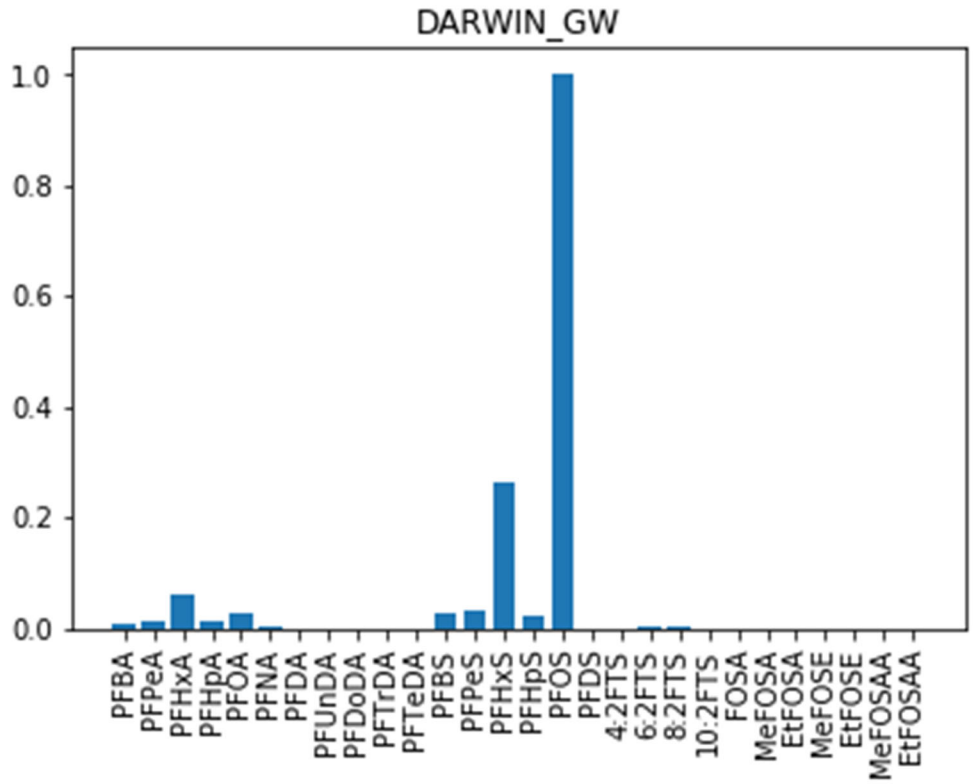

CLARENDON_SOIL_ON
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFHpS
PFOS
PFDS
6:2FTS
8:2FTS
FOSA
MeFOSAA
EtFOSAA

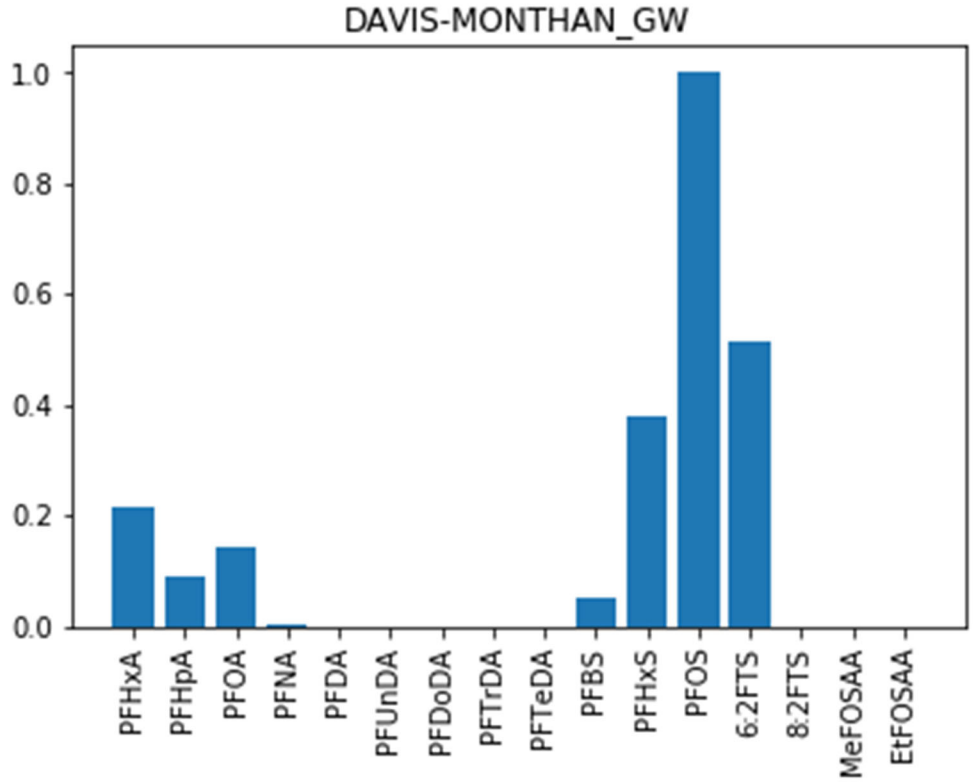

COVENTRY_LF_SLUDGE
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
FOSA
MeFOSAA
EtFOSAA

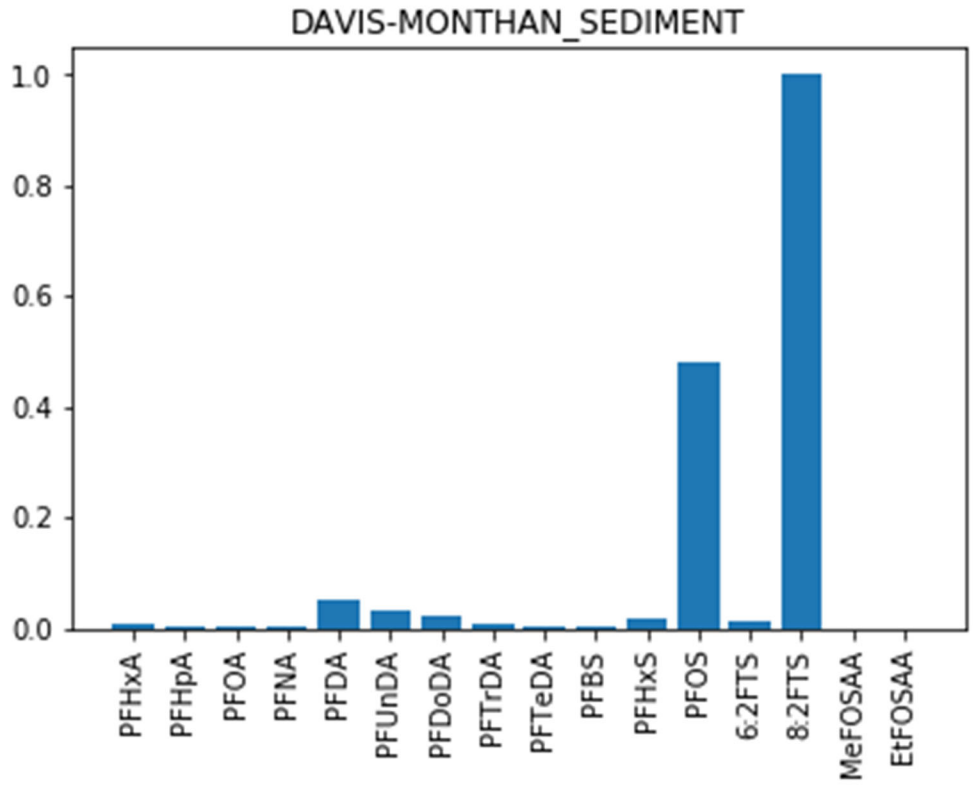

COVENTRY_LF_SOIL
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
FOSA
MeFOSAA
EtFOSAA

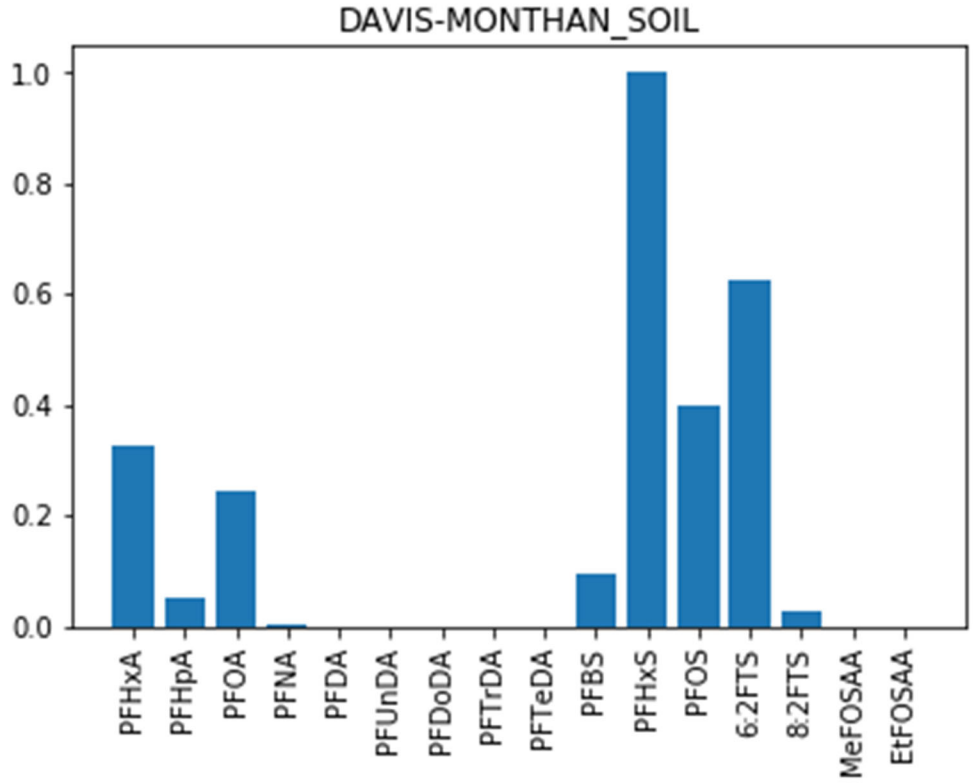

COVENTRY_LF_SOURCE
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
FOSA
MeFOSAA
EtFOSAA

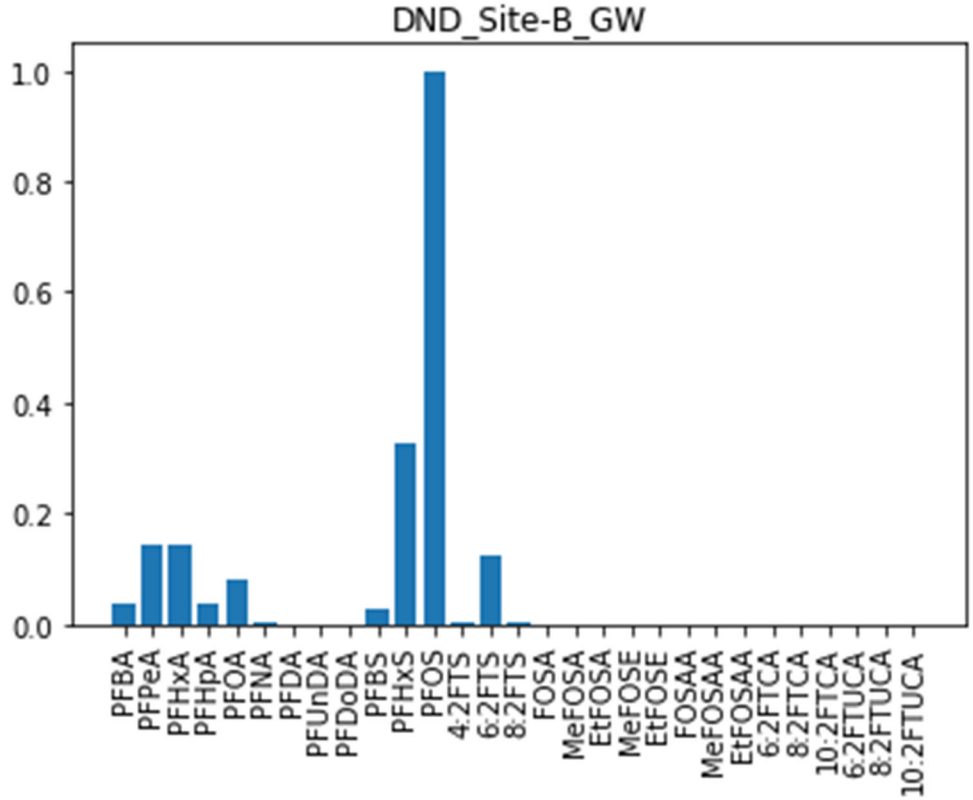

DALING_RIVER_BASIN_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFBS
PFHxS
PFOS
PFDS
GenX

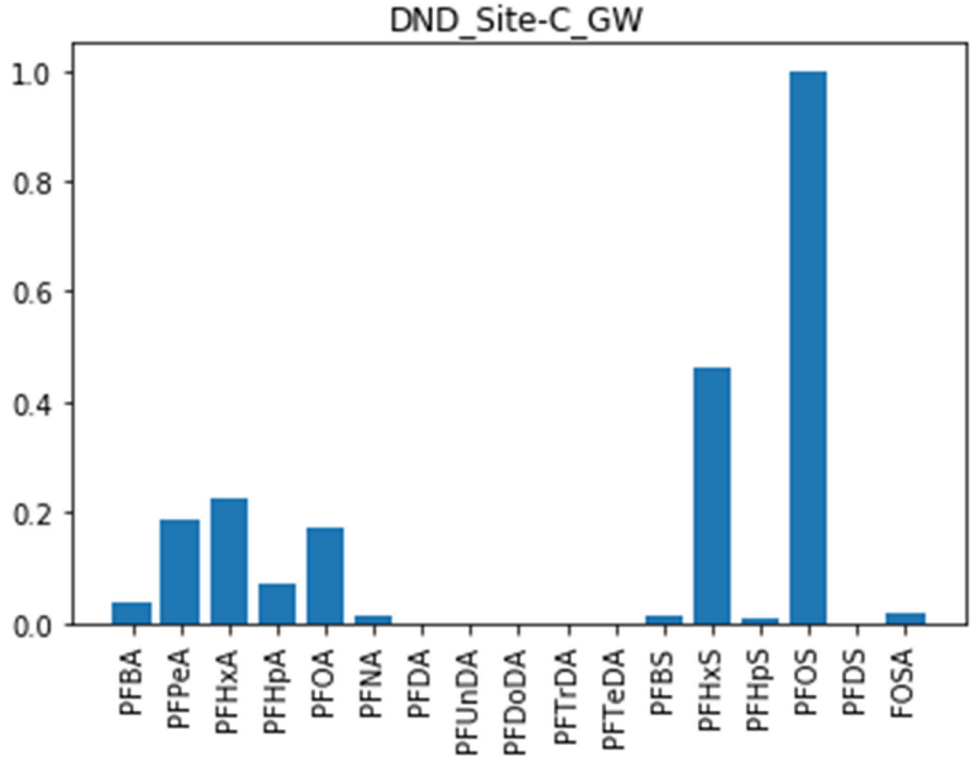

DARWIN_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

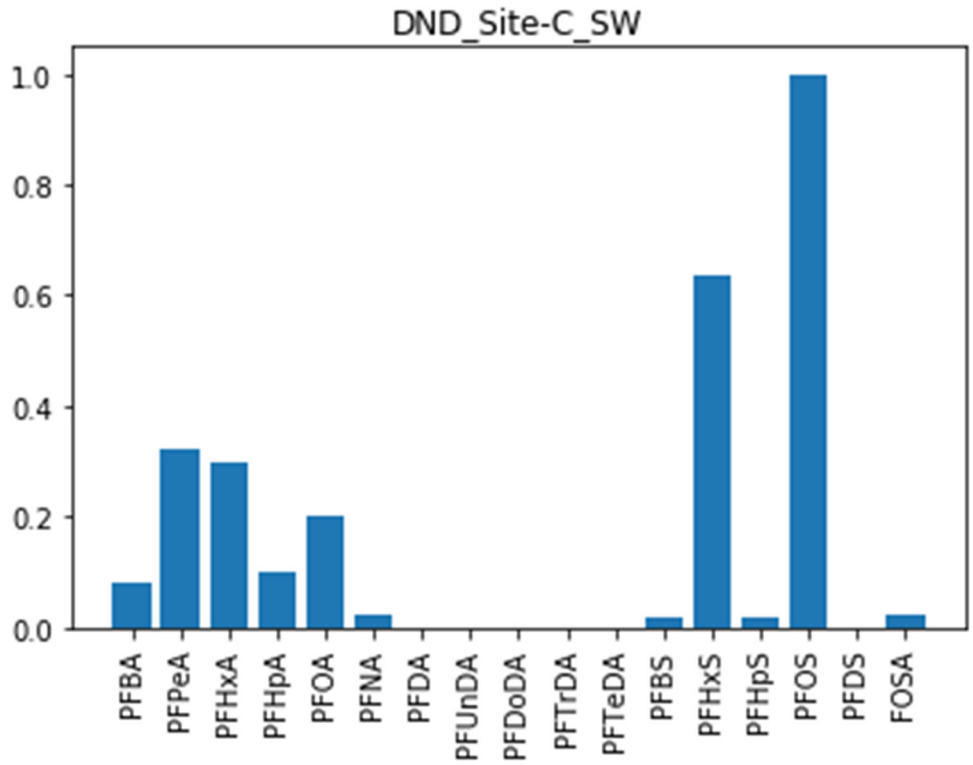

DAVIS-MONTHAN_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFOS
6:2FTS
8:2FTS
MeFOSAA
EtFOSAA

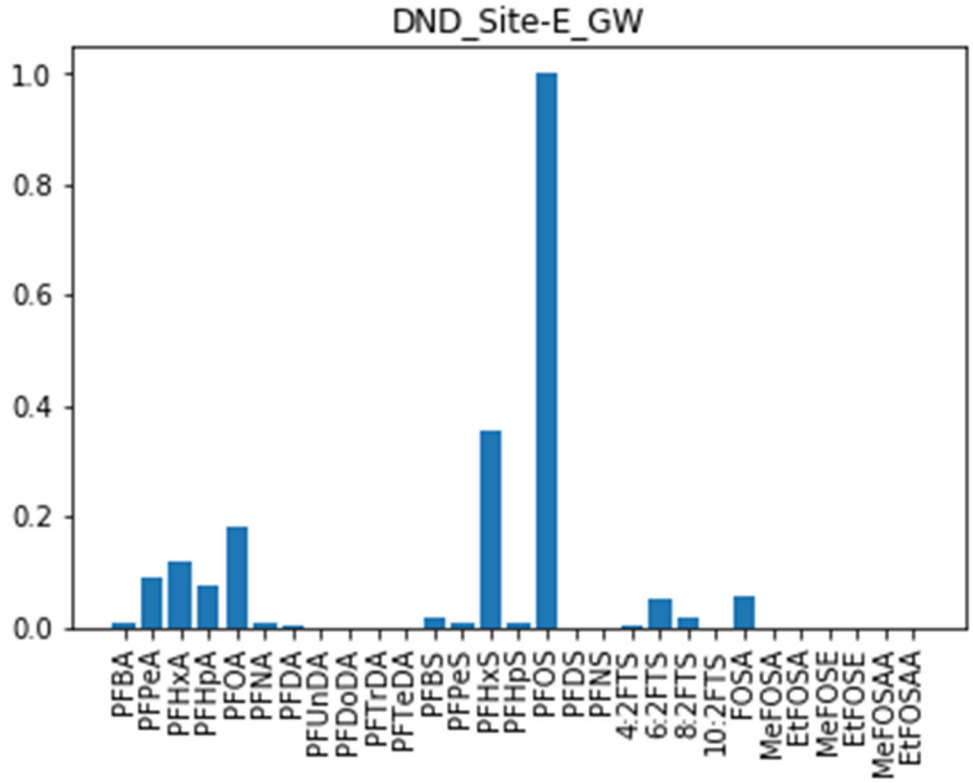

DAVIS-MONTHAN_SEDIMENT
1.0
0.8
0.6
0.4
0.2
0.0
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFOS
6:2FTS
8:2FTS
MeFOSAA
EtFOSAA

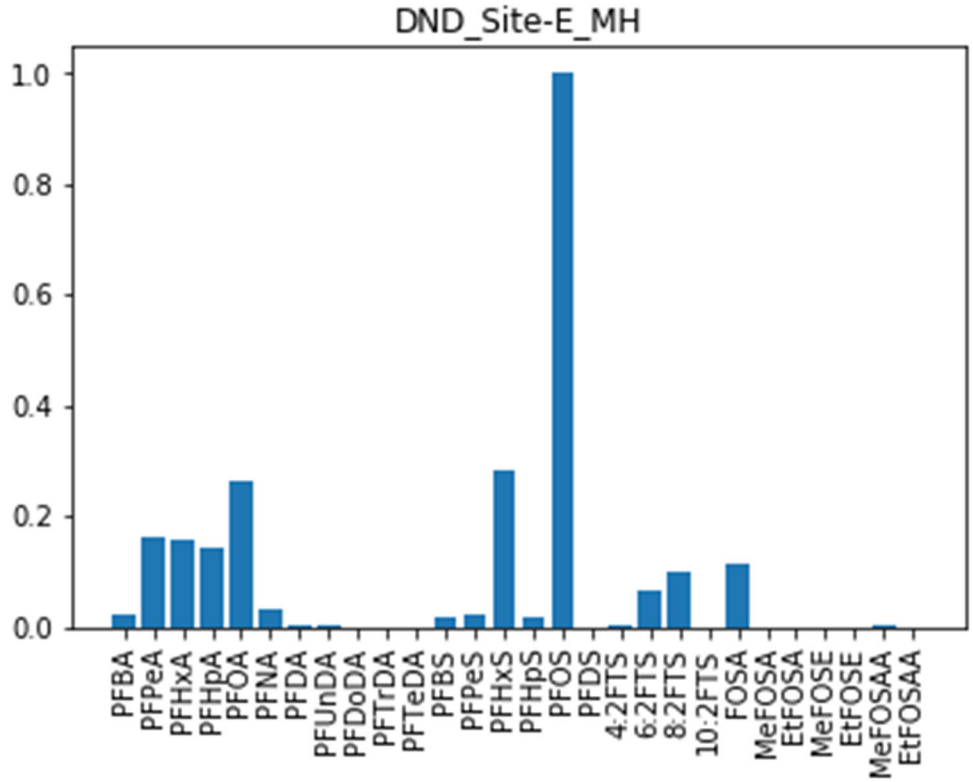

DAVIS-MONTHAN_SOIL
1.0
0.8
0.6
0.4
0.2
0.0
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFOS
6:2FTS
8:2FTS
MeFOSAA
EtFOSAA

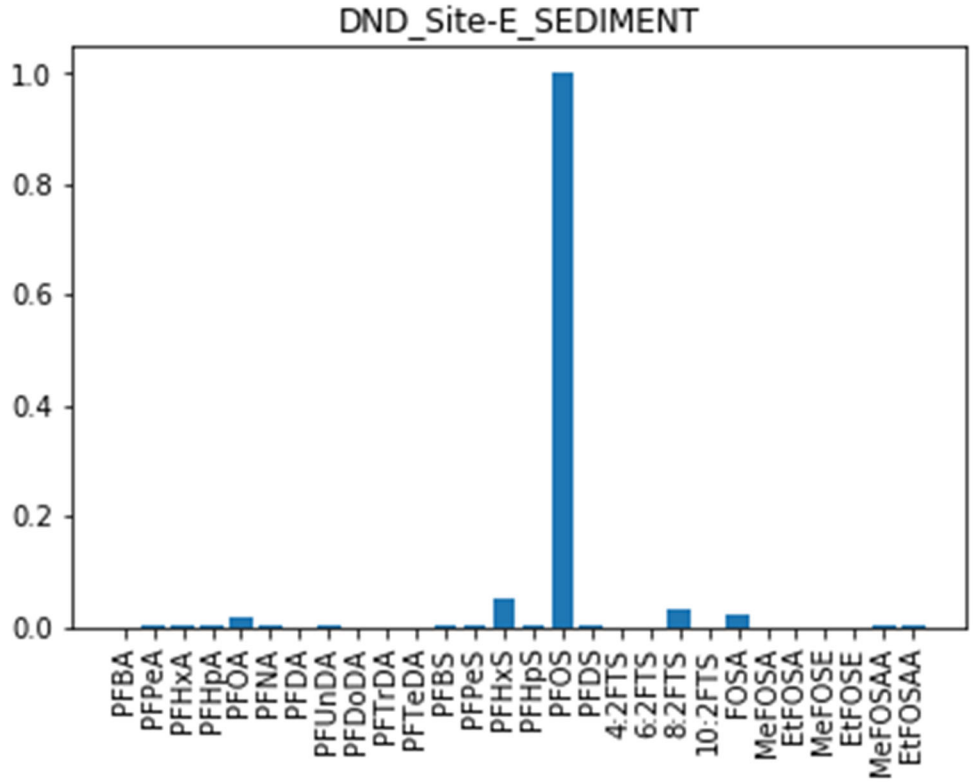

DND_Site-B_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFBS
PFHxS
PFOS
4:2FTS
6:2FTS
8:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
FOSAA
MeFOSAA
EtFOSAA
6:2FTCA
8:2FTCA
10:2FTCA
6:2FTUCA
8:2FTUCA
10:2FTUCA

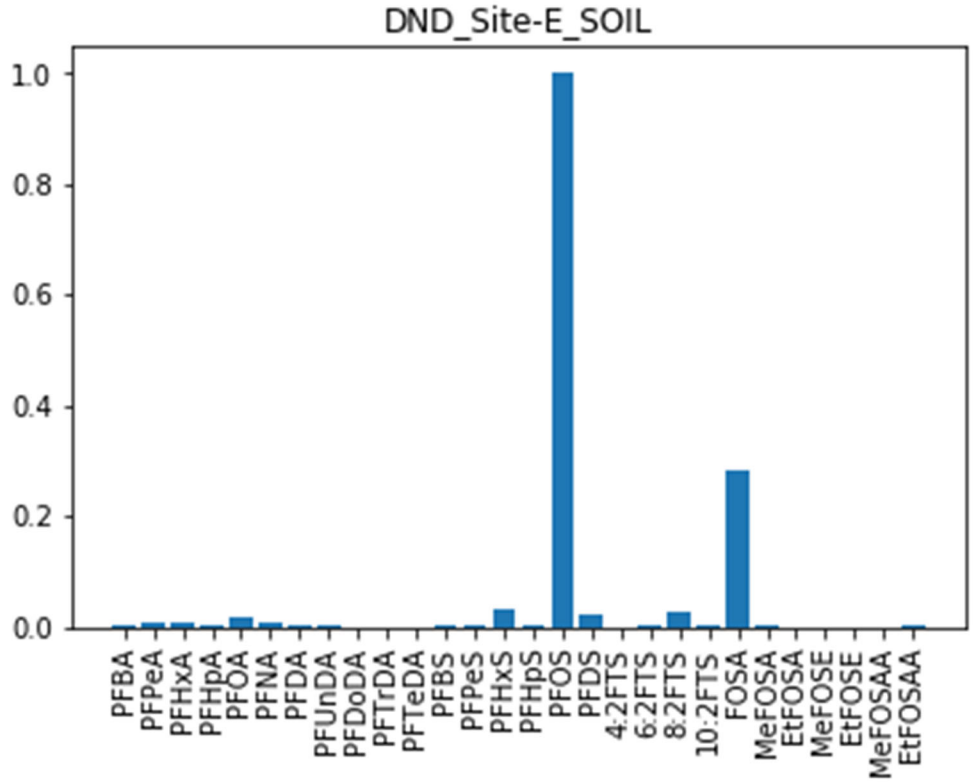

DND_Site-C_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFHpS
PFOS
PFDS
FOSA

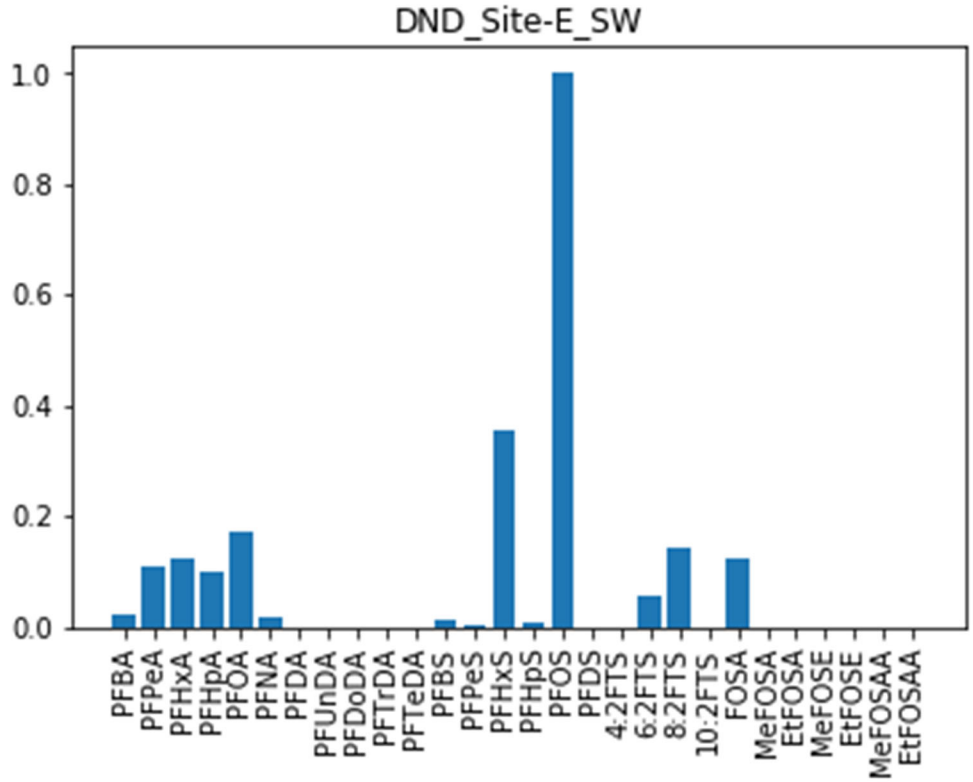

DND_Site-C_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFHpS
PFOS
PFDS
FOSA

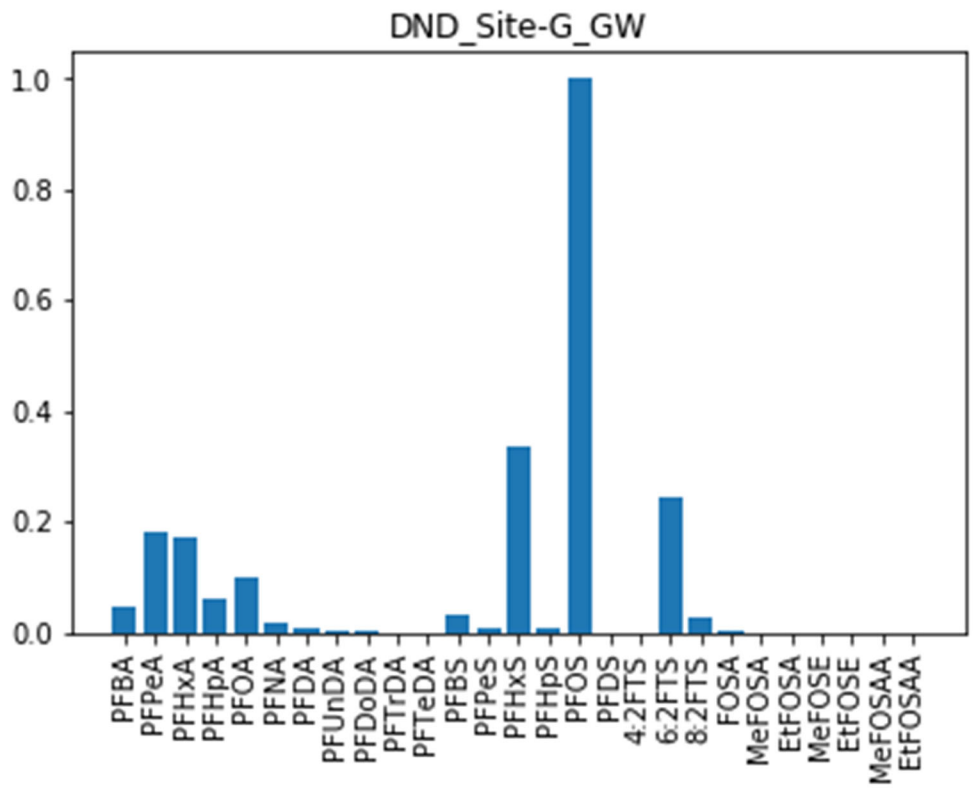

DND_Site-E_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
PFNS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

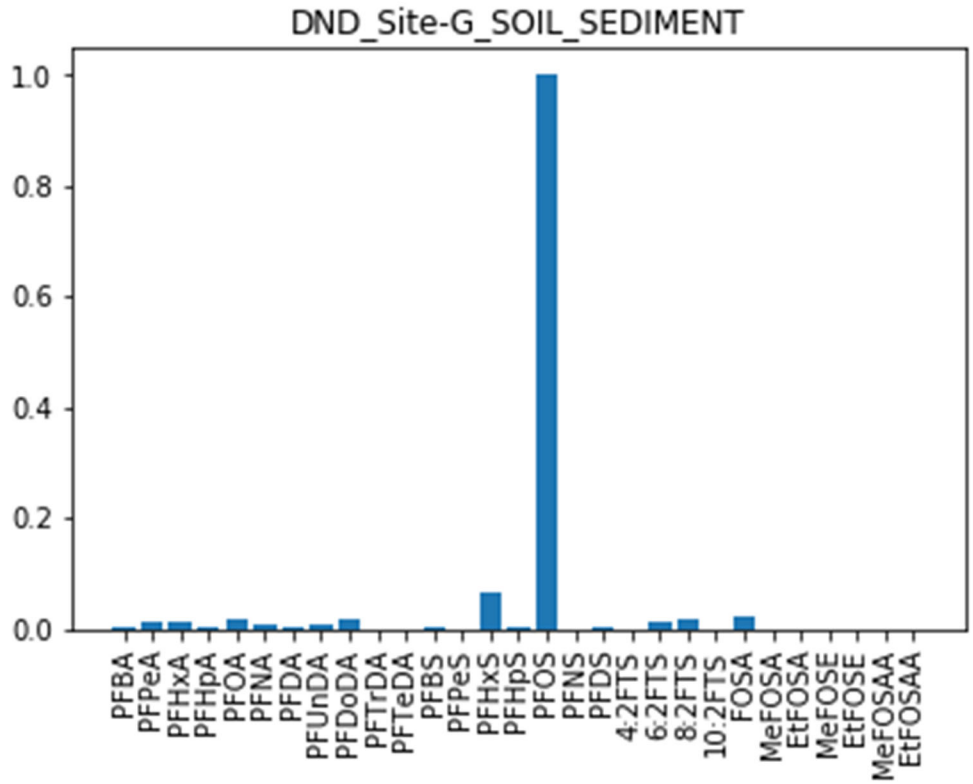

DND_Site-E_MH
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

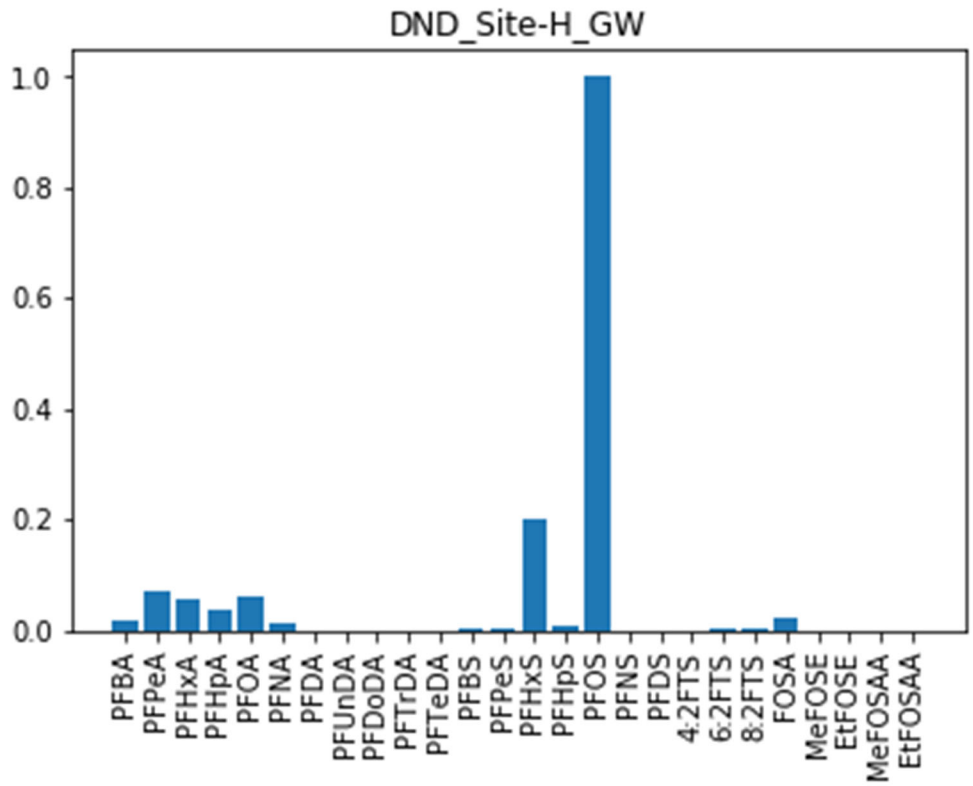

DND_Site-E_SEDIMENT
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

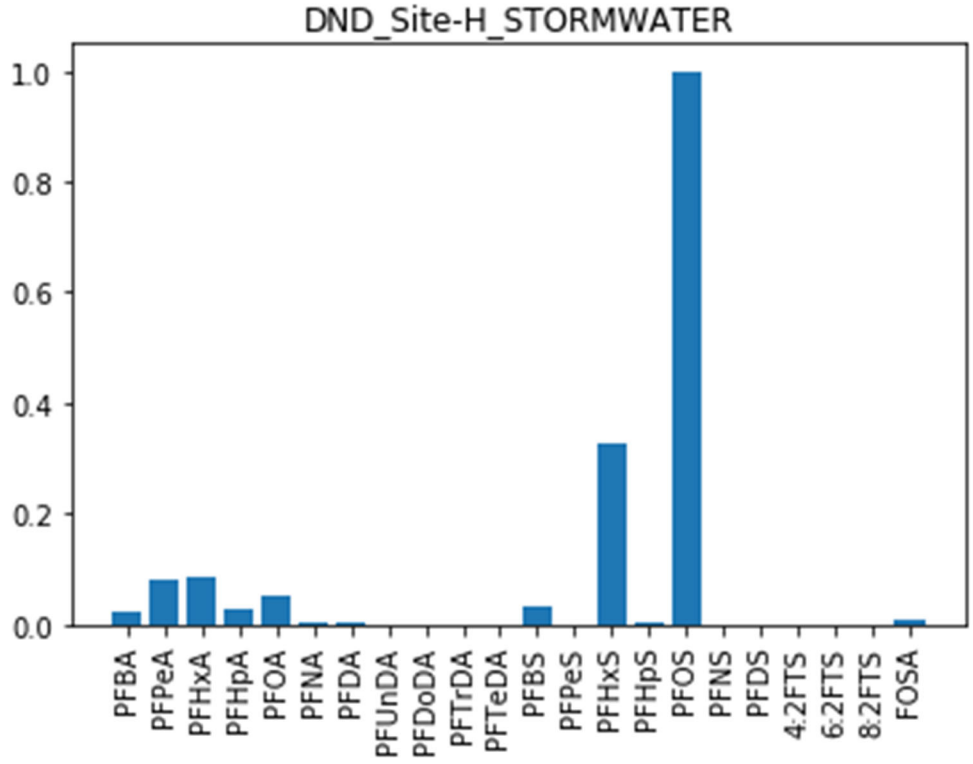

DND_Site-E_SOIL
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

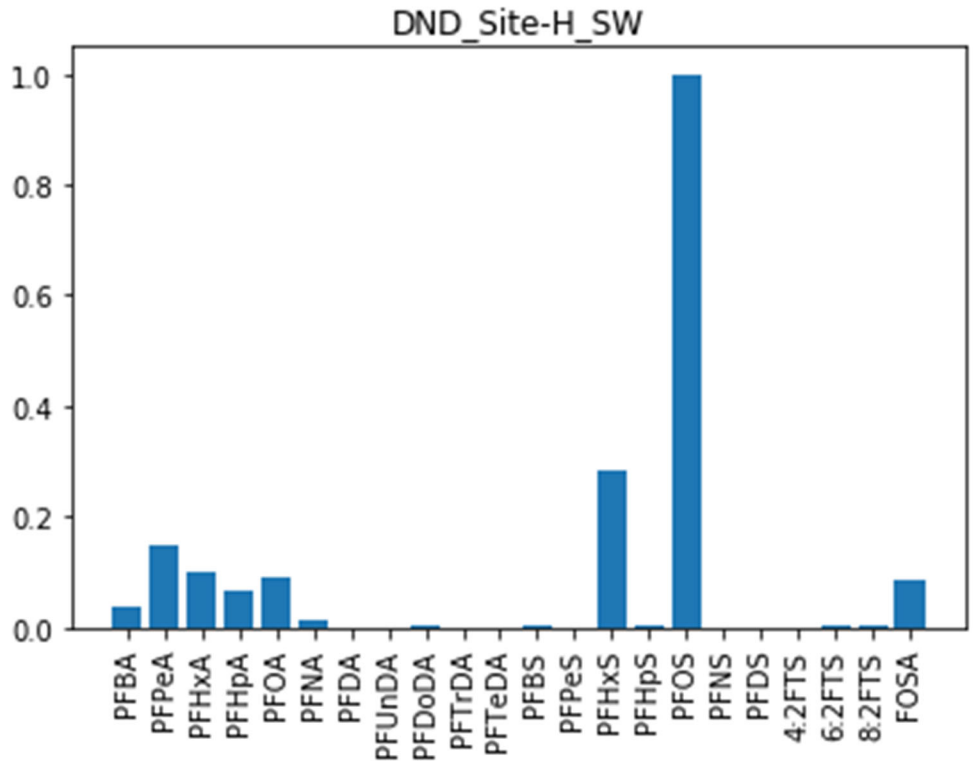

DND_Site-E_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

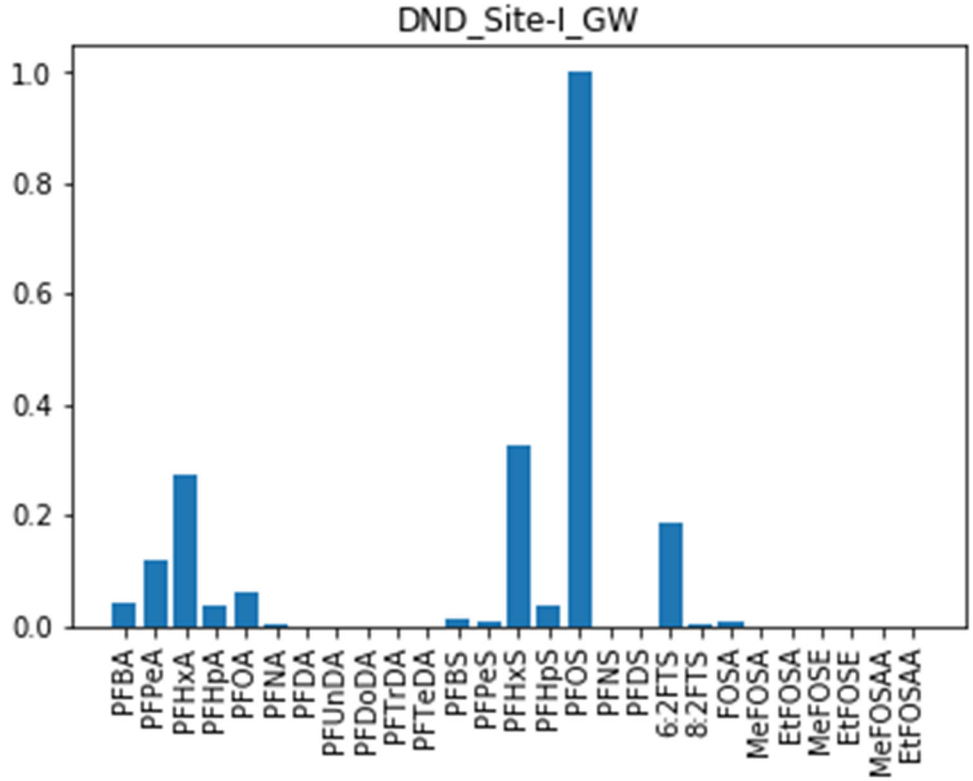

DND_Site-G_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

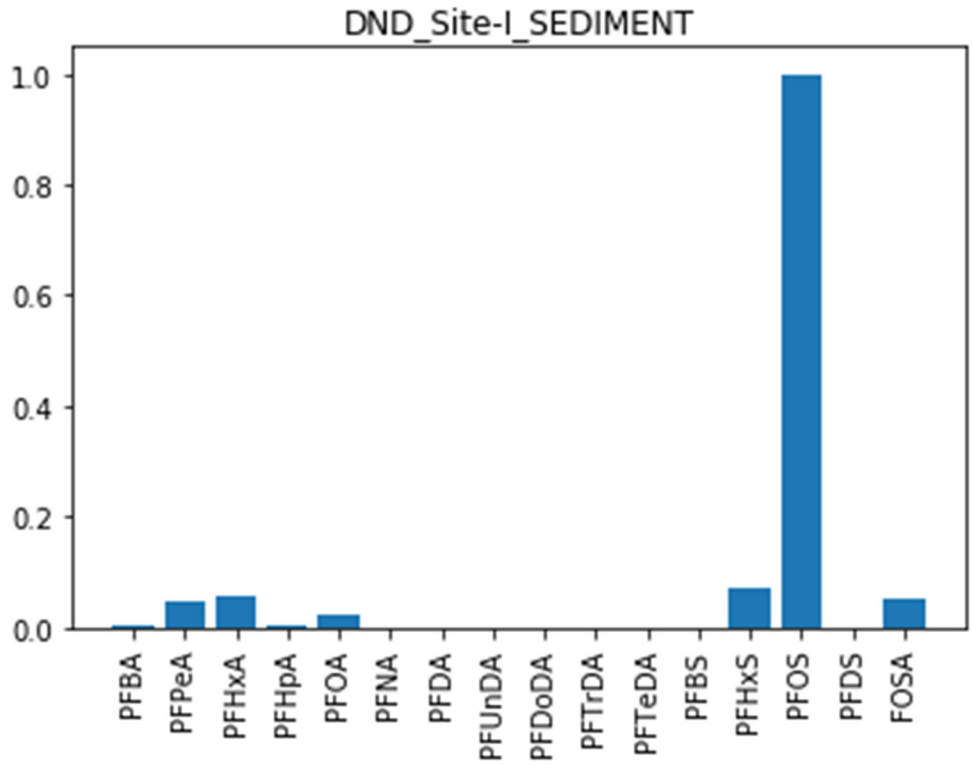

DND_Site-G_SOIL_SEDIMENT
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

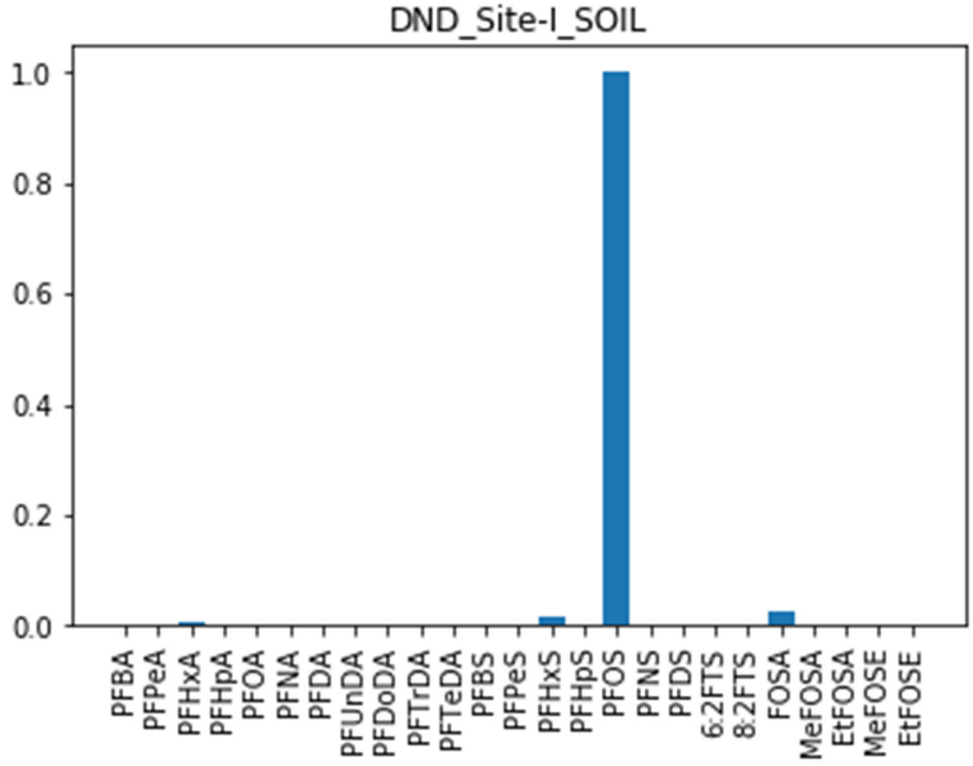

DND_Site-H_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
FOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

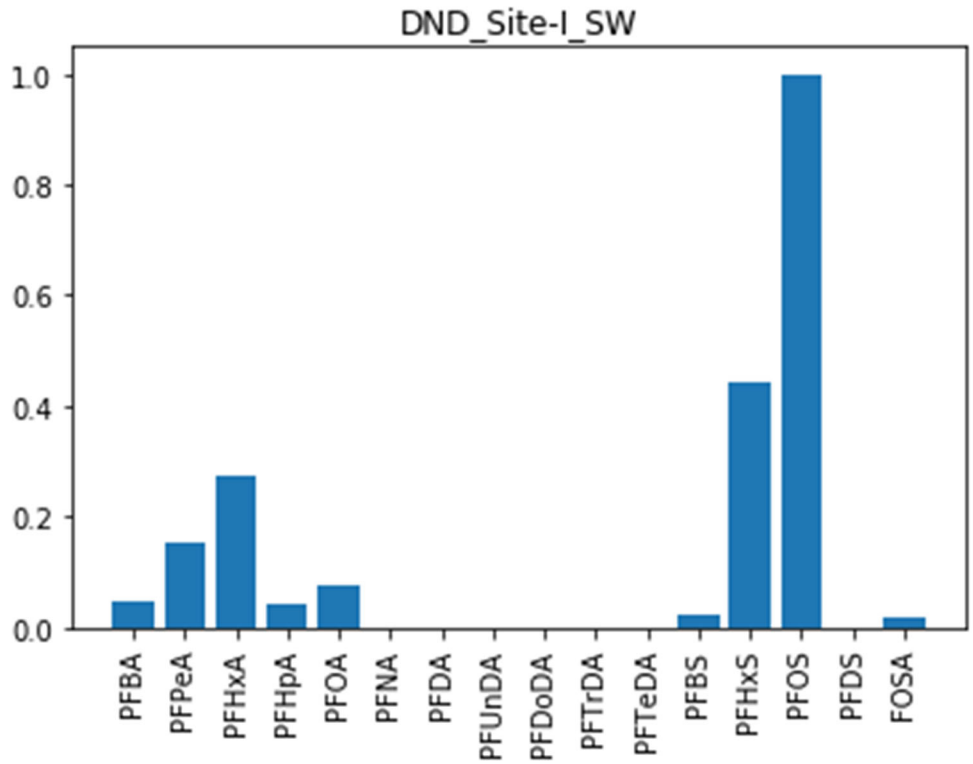

DND_Site-H_STORMWATER
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
FOSA

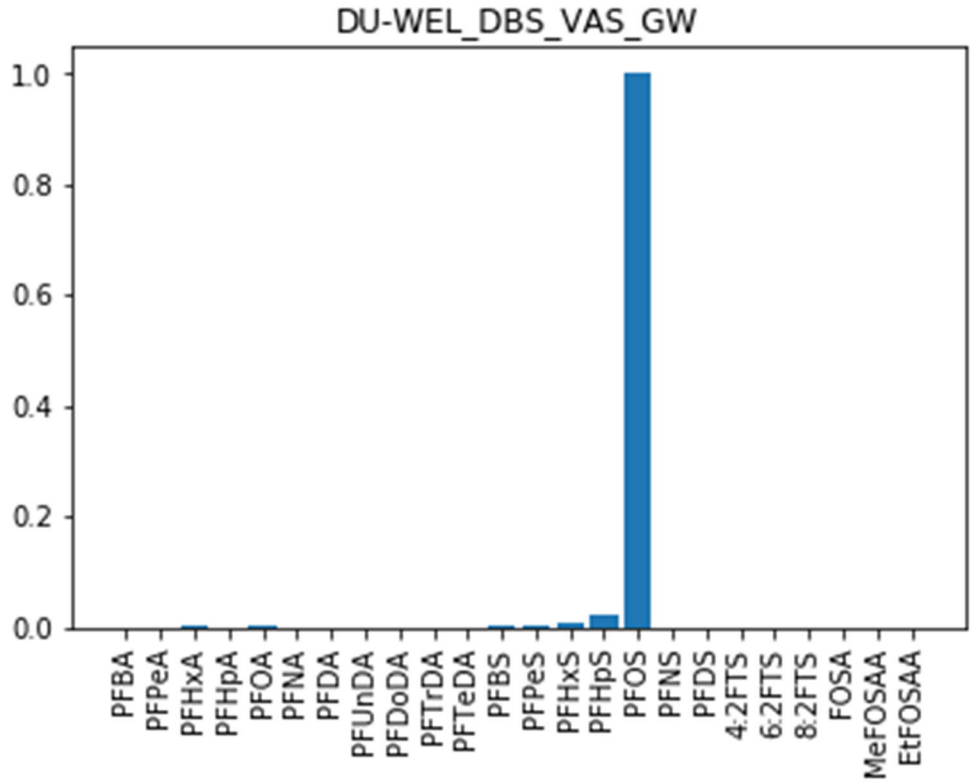

DND_Site-H_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
FOSA

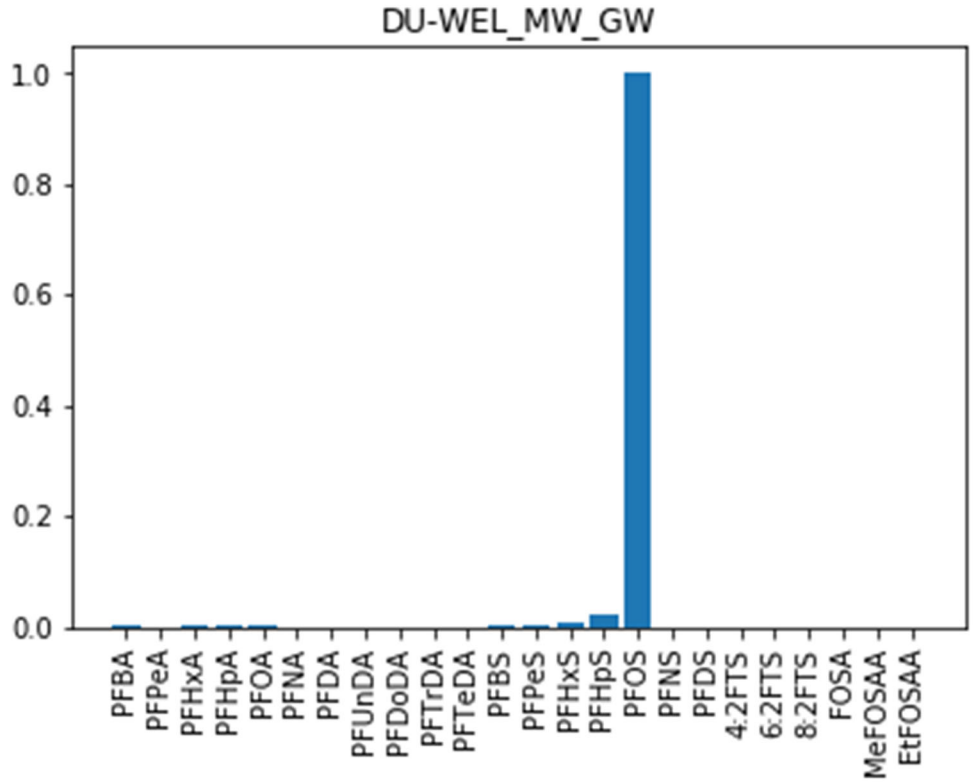

DND_Site-I_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
6:2FTS
8:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

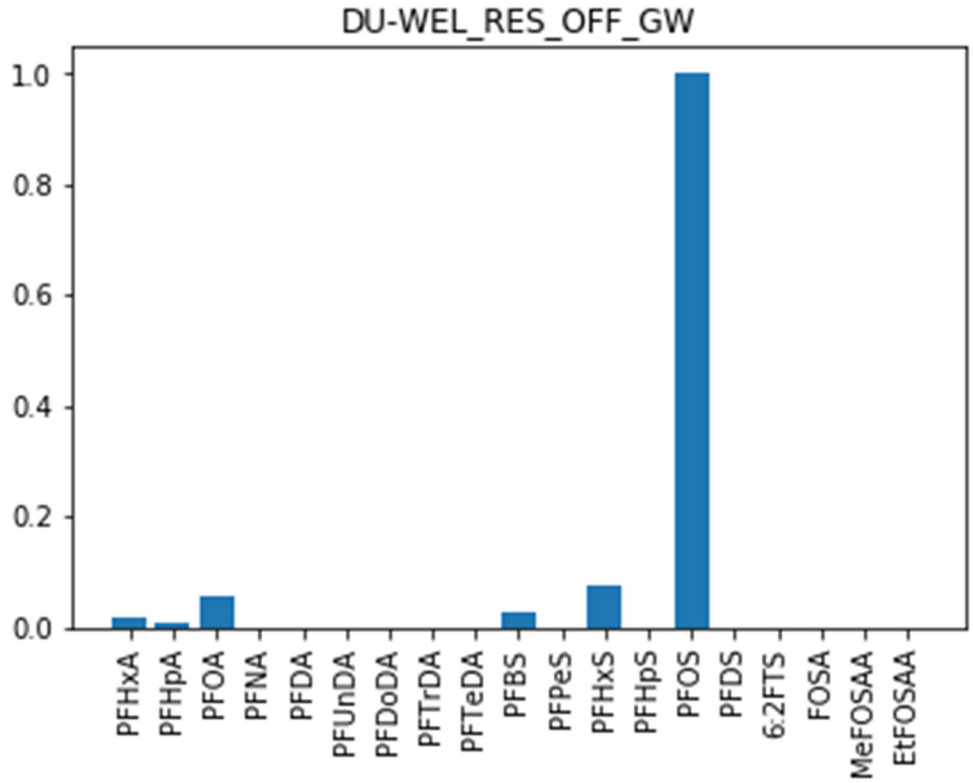

DND_Site-I_SEDIMENT
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFOS
PFDS
FOSA

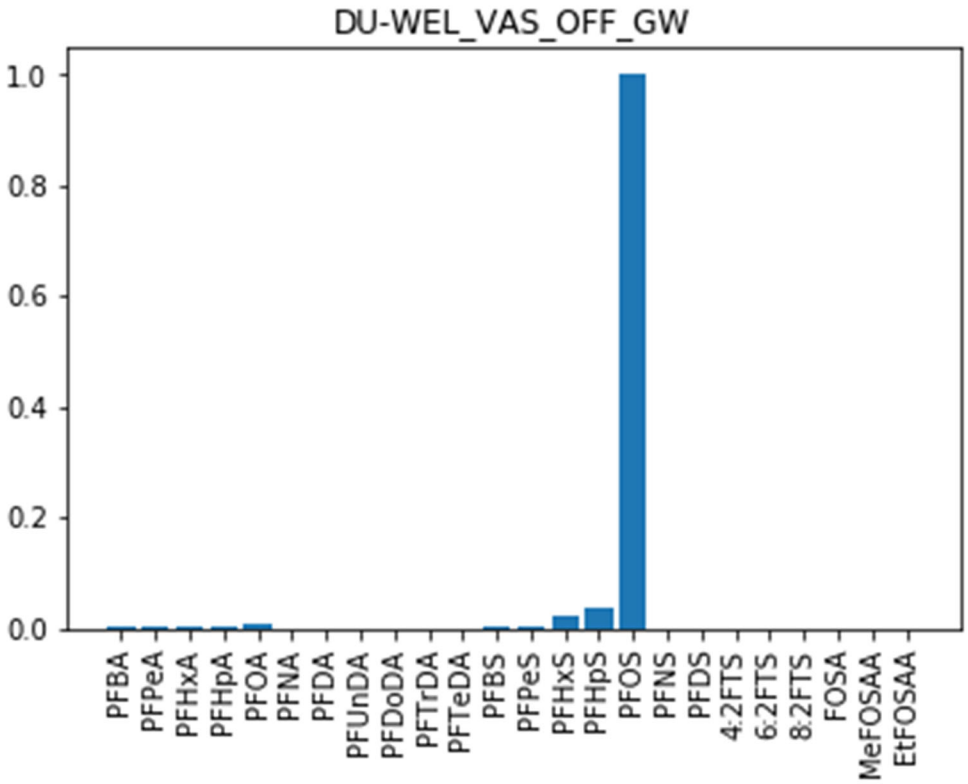

DND_Site-I_SOIL
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
6:2FTS
8:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE

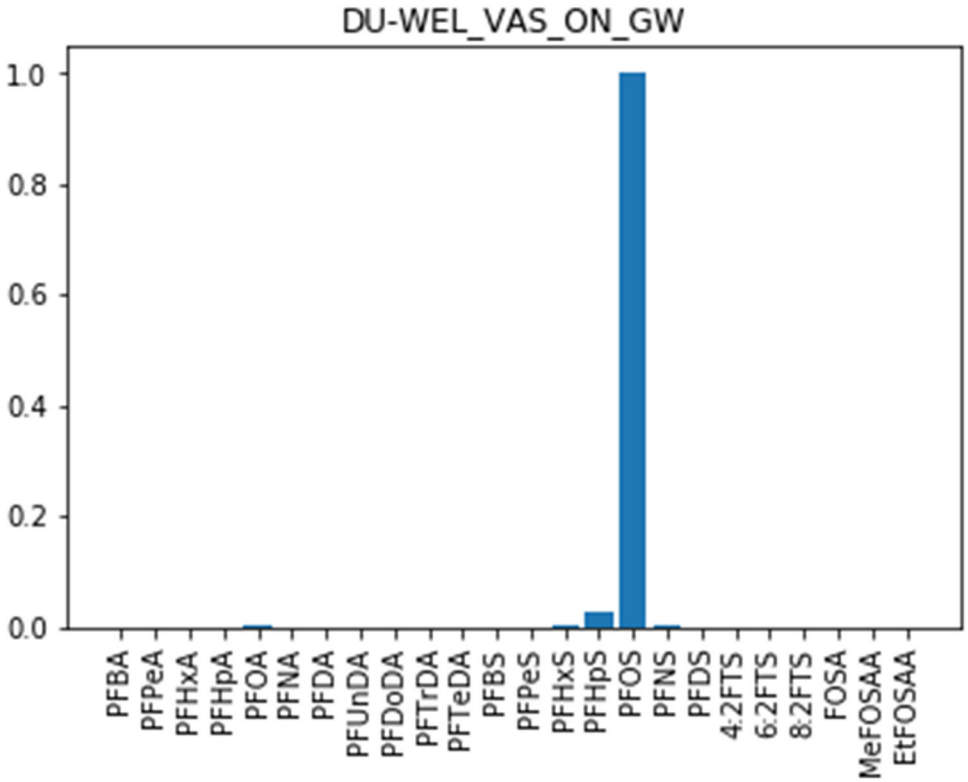

DND_Site-I_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFOS
PFDS
FOSA

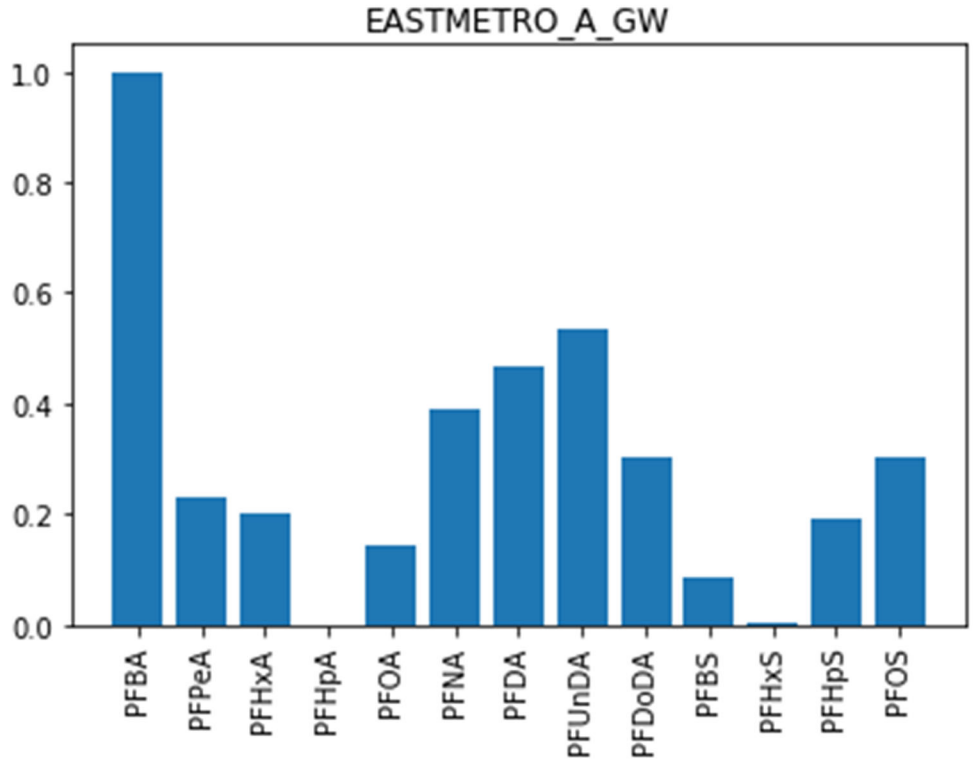

DU-WEL_DBS_VAS_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
FOSA
MeFOSAA
EtFOSAA

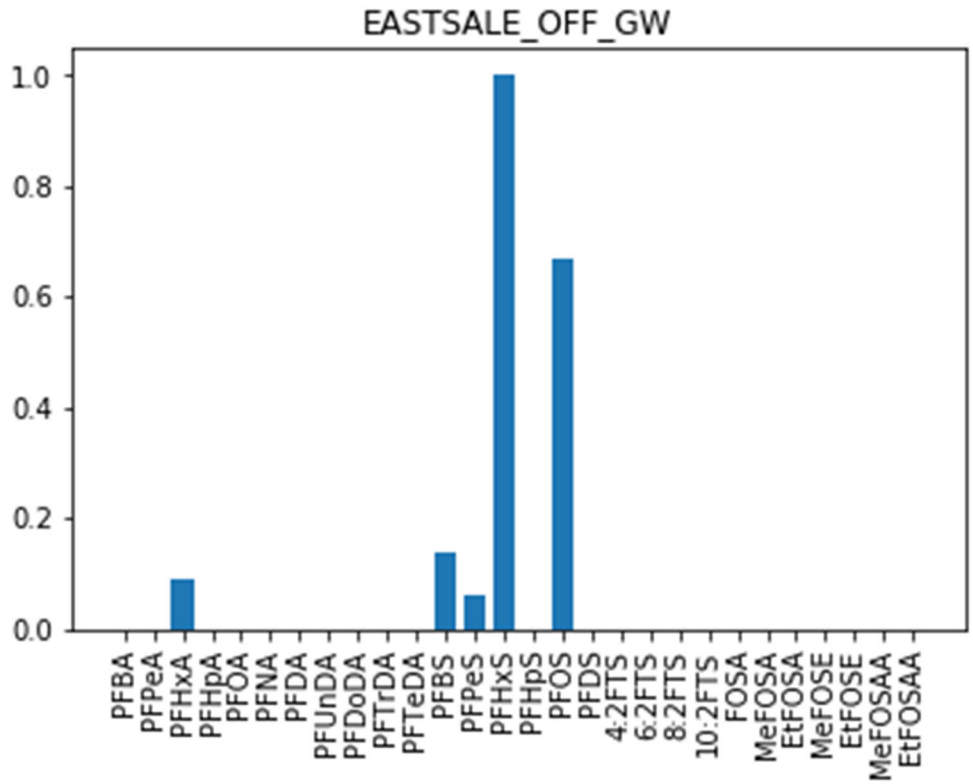

DU-WEL_MW_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
FOSA
MeFOSAA
EtFOSAA

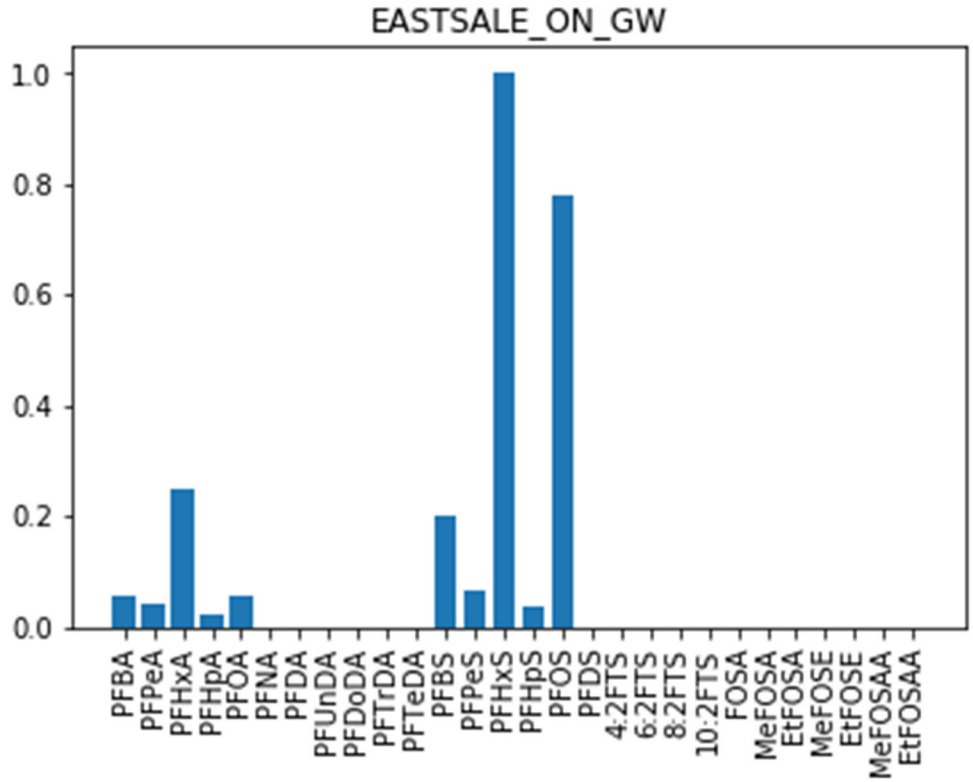

DU-WEL_RES_OFF_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
6:2FTS
FOSA
MeFOSAA
EtFOSAA

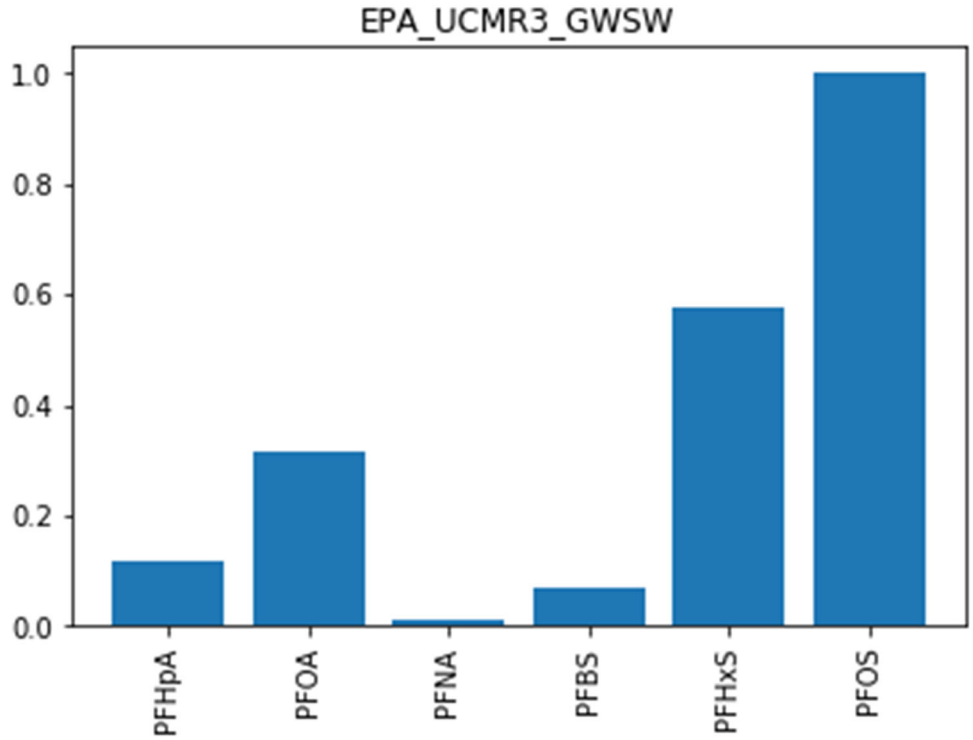

DU-WEL_VAS_OFF_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
FOSA
MeFOSAA
EtFOSAA

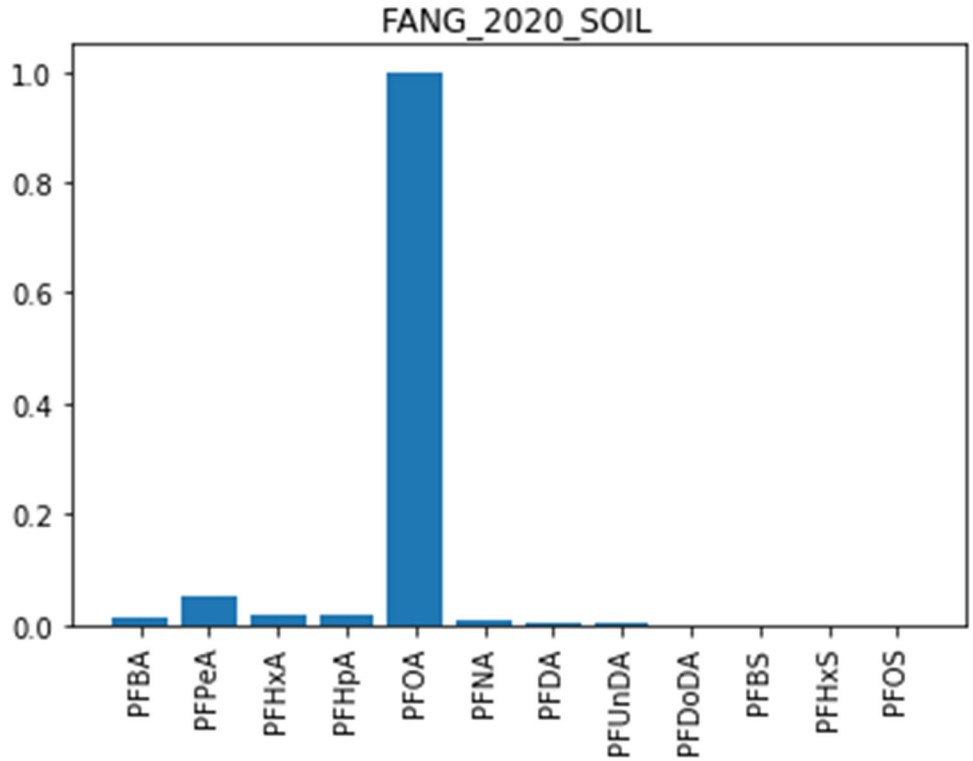

DU-WEL_VAS_ON_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
FOSA
MeFOSAA
EtFOSAA

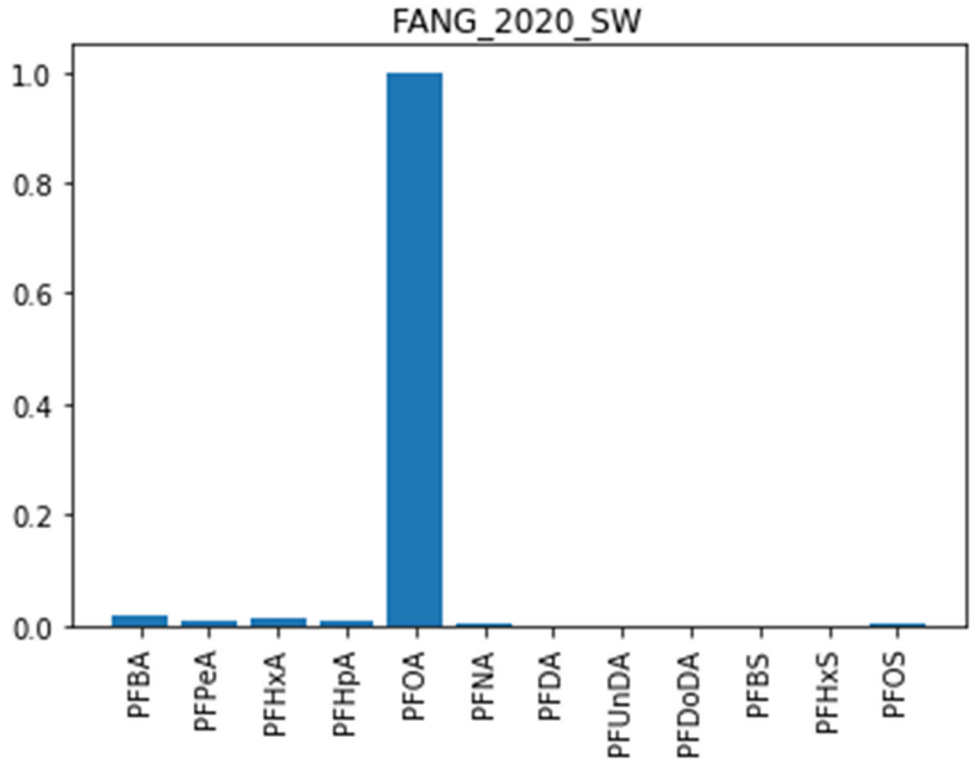

EASTMETRO_A_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFBS
PFHxS
PFHpS
PFOS

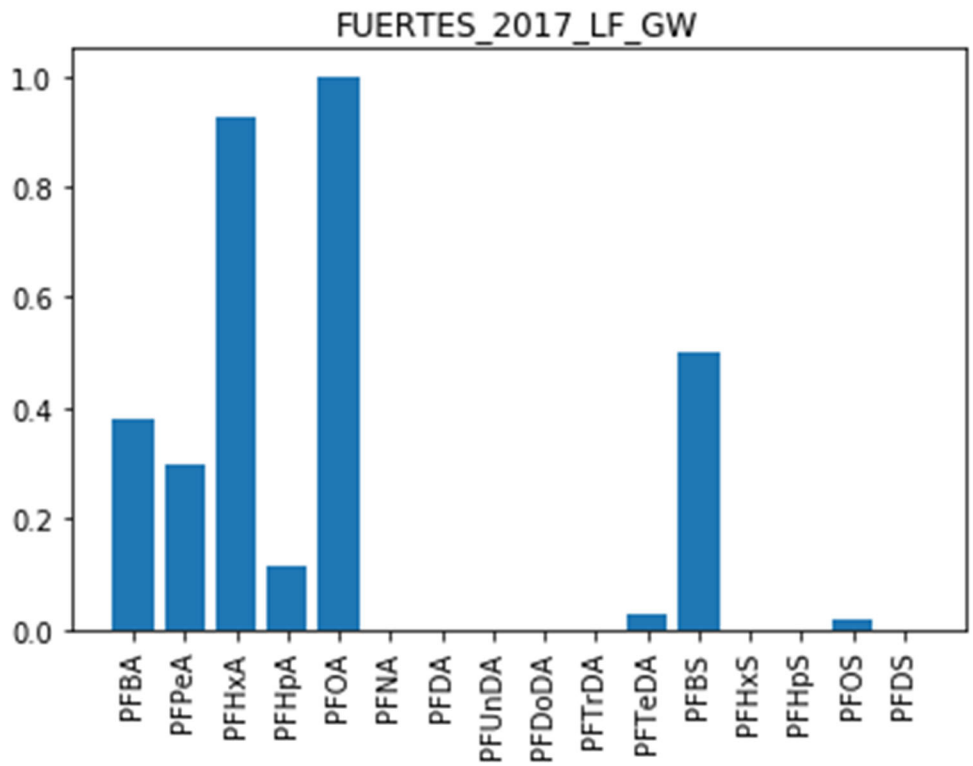

EASTSALE_OFF_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

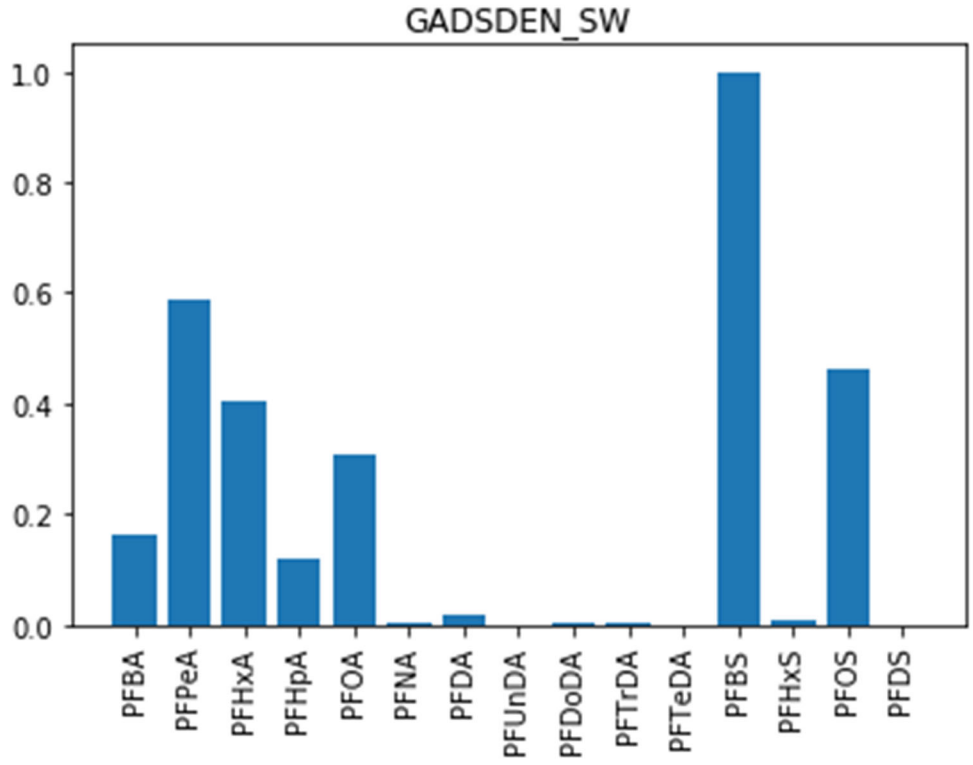

EASTSALE_ON_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

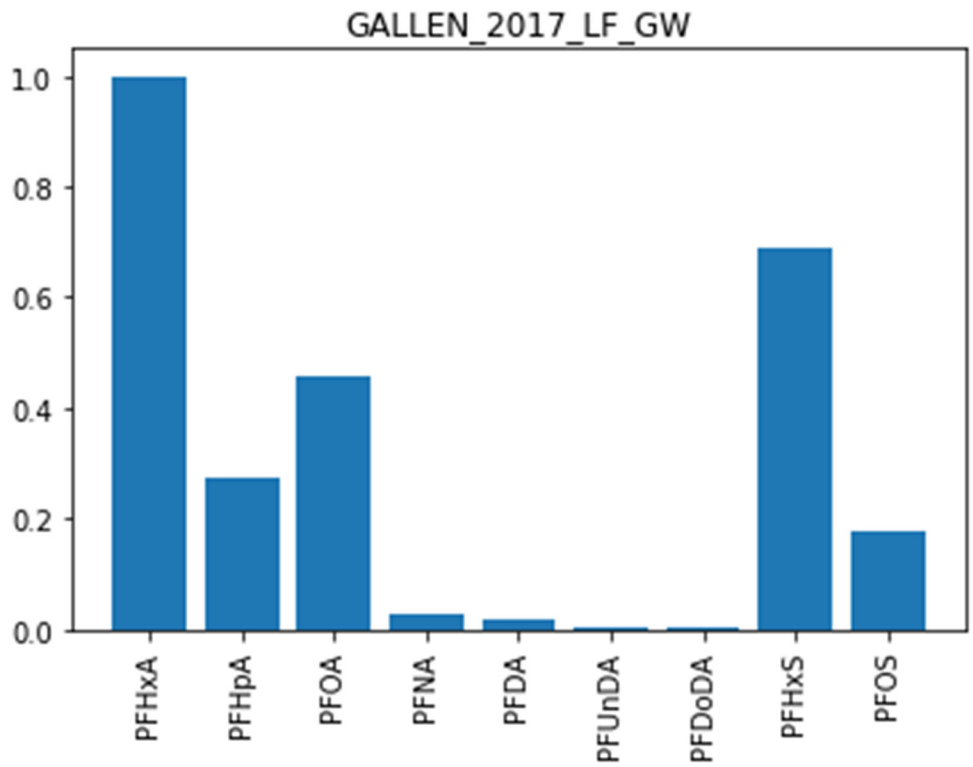

EPA_UCMR3_GWSW
1.0
0.8
0.6
0.4
0.2
0.0
PFHpA
PFOA
PFNA
PFBS
PFHxS
PFOS

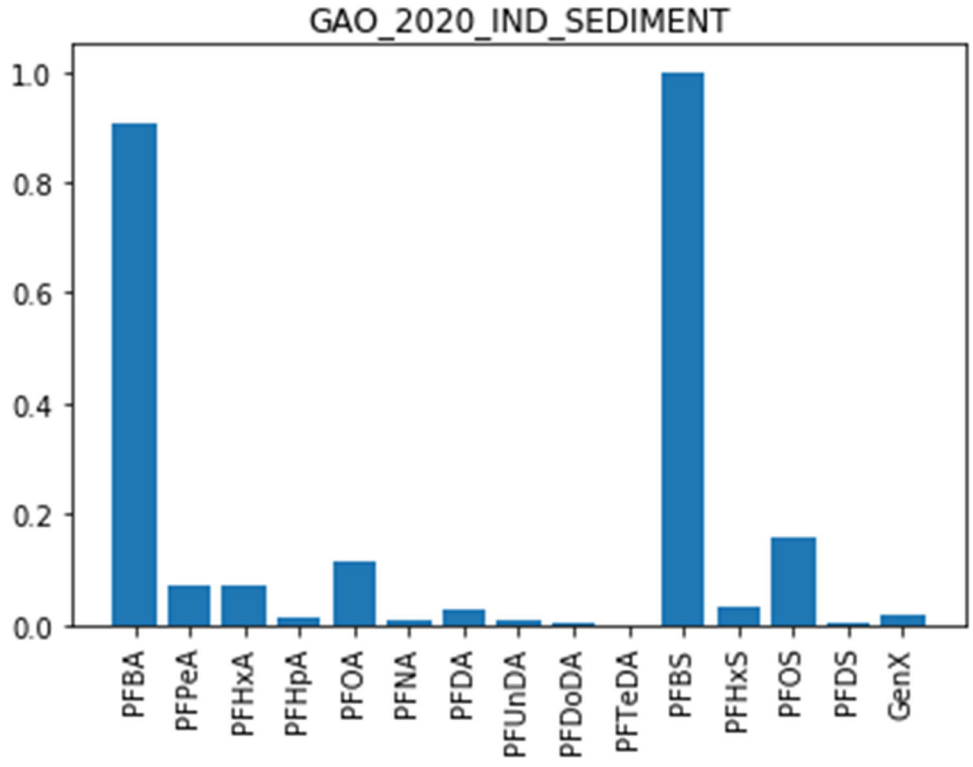

FANG_2020_SOIL
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFBS
PFHxS
PFOS

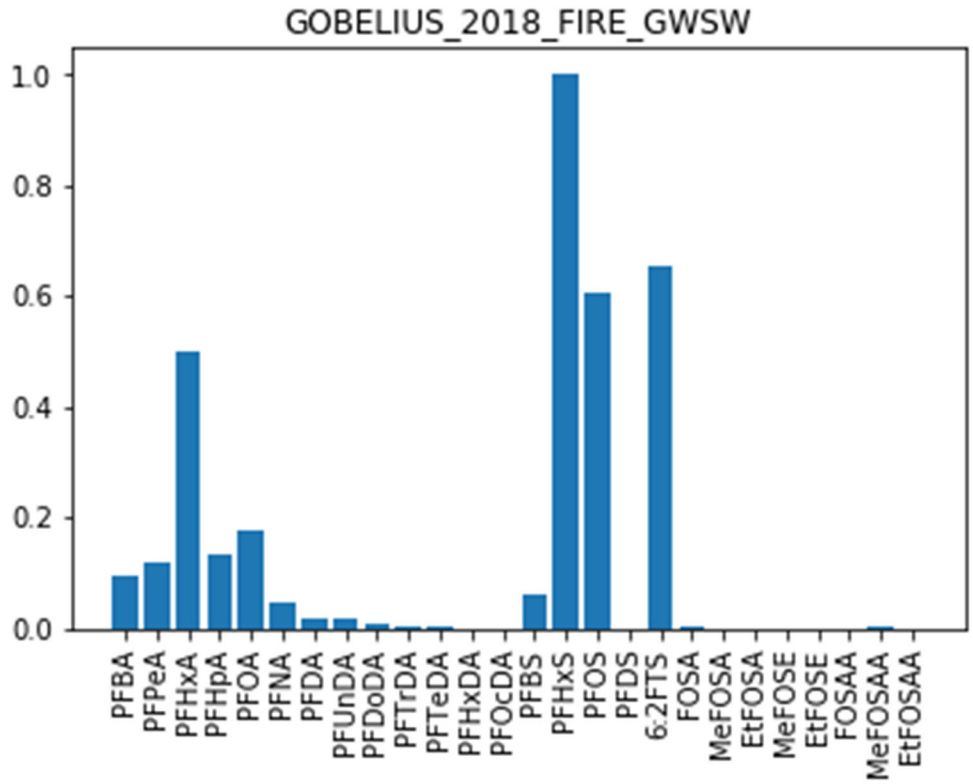

FANG_2020_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFBS
PFHxS
PFOS

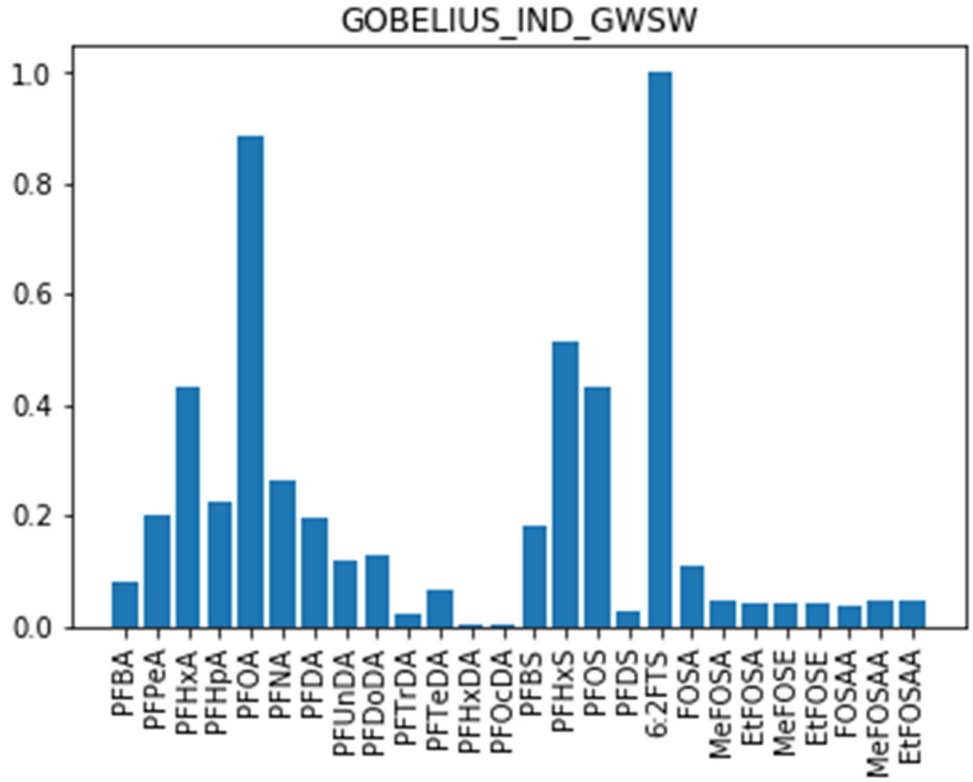

FUERTES_2017_LF_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFHpS
PFOS
PFDS

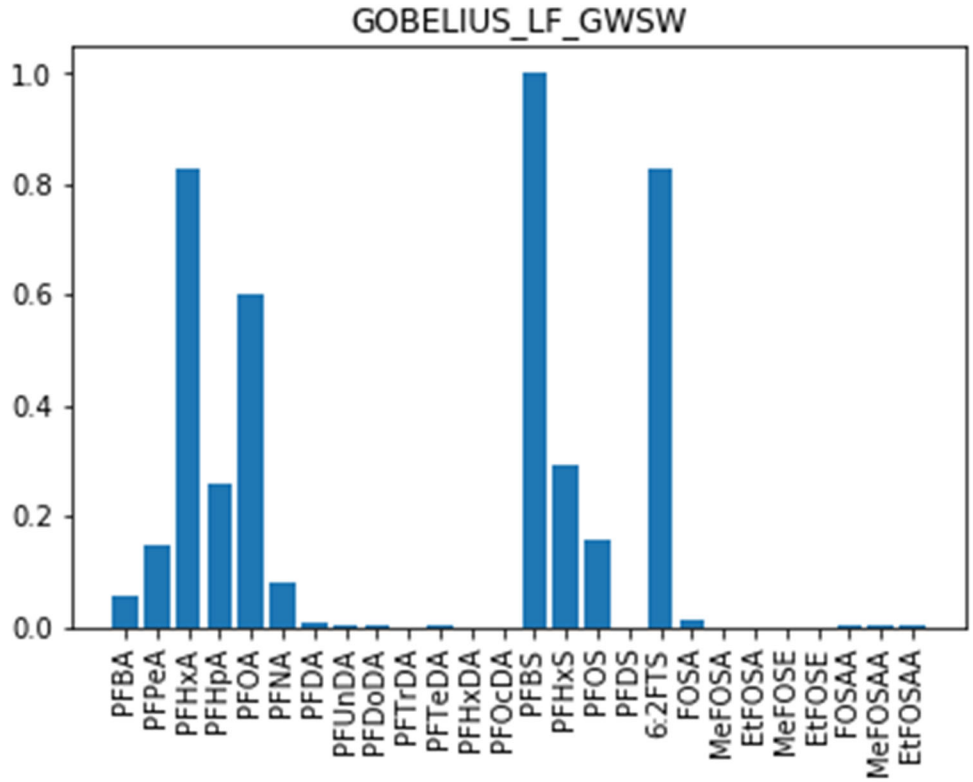

GADSDEN_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFOS
PFDS

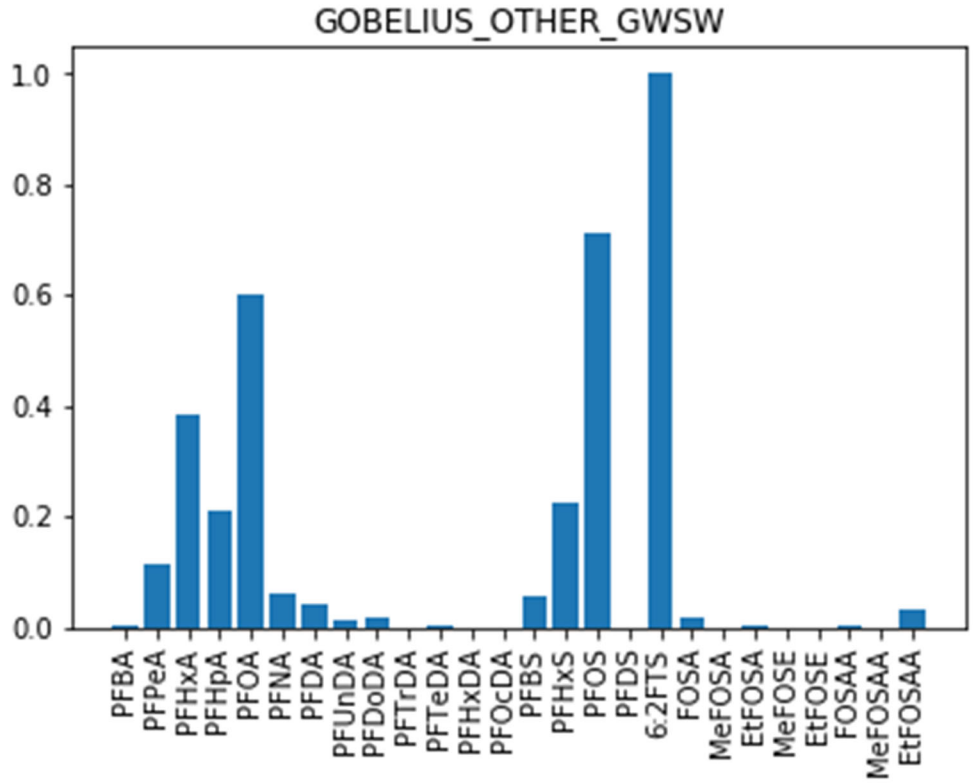

GALLEN_2017_LF_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFHxS
PFOS

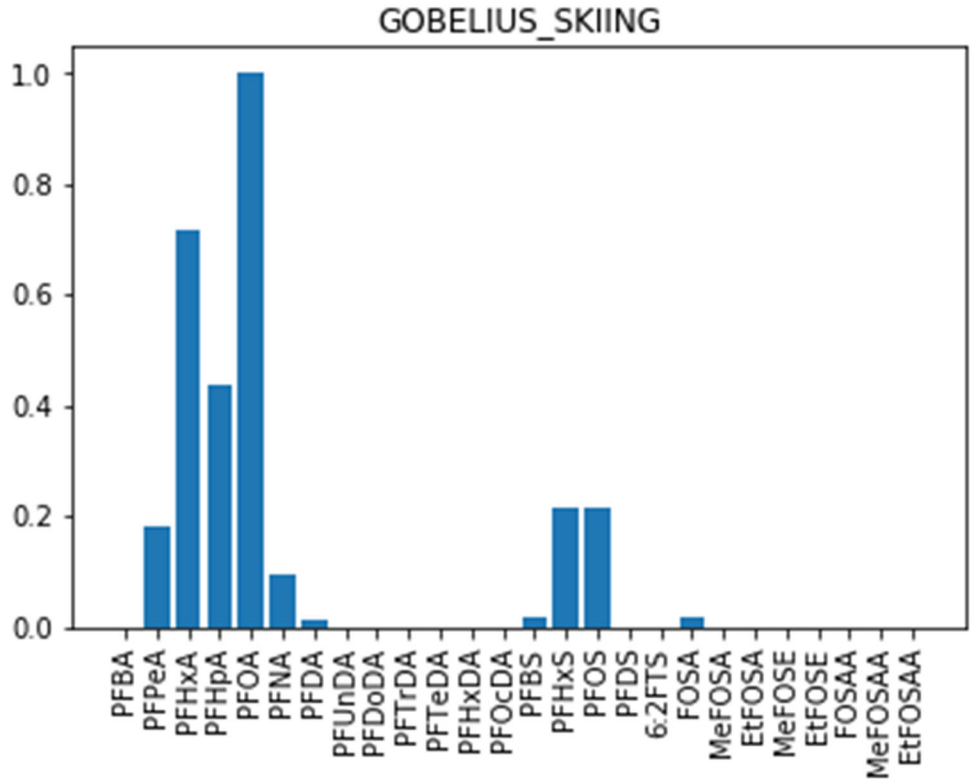

GAO_2020_IND_SEDIMENT
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTeDA
PFBS
PFHxS
PFOS
PFDS
GenX

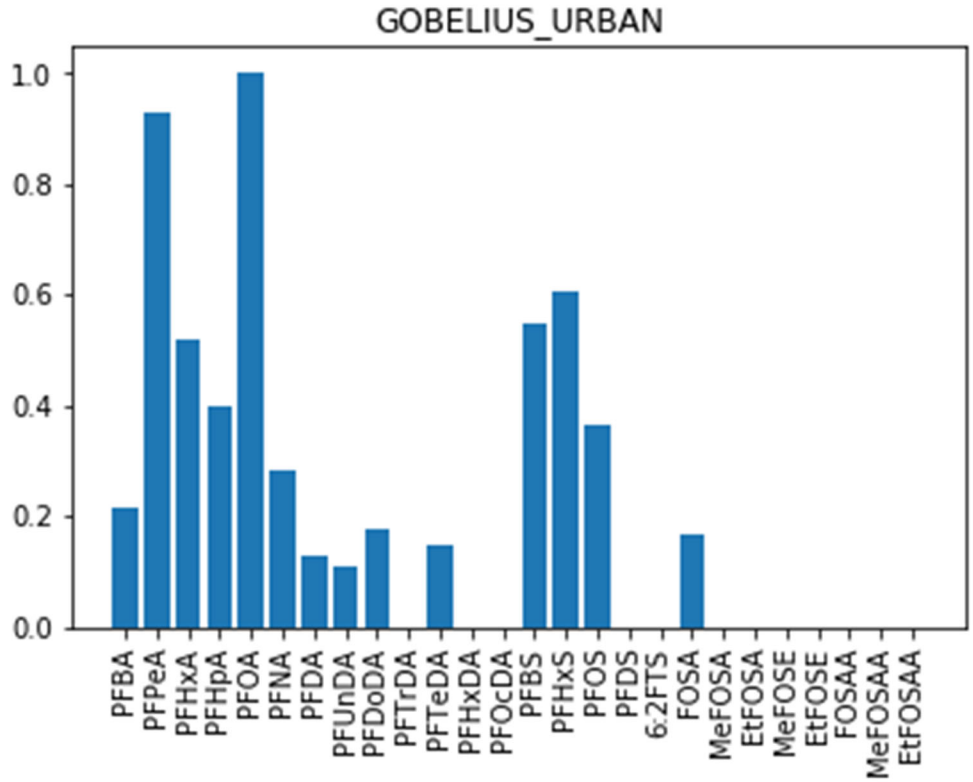

GOBELIUS_2018_FIRE_GWSW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFHxDA
PFOcDA
PFBS
PFHxS
PFOS
PFDS
6:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
FOSAA
MeFOSAA
EtFOSAA

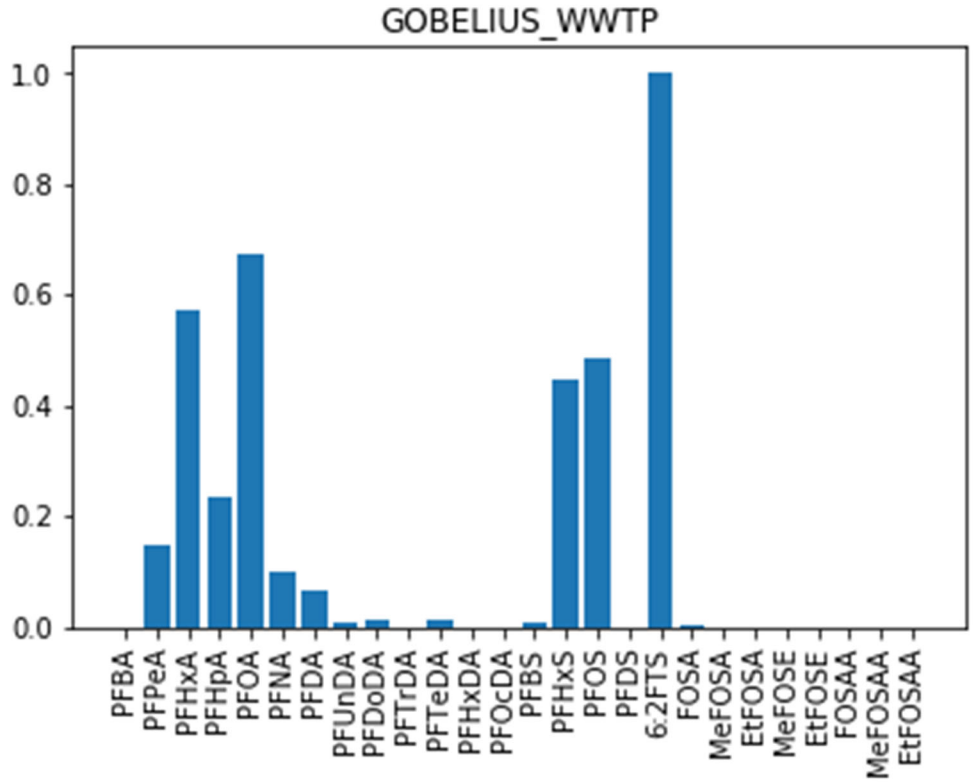

GOBELIUS_IND_GWSW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFHxDA
PFOcDA
PFBS
PFHxS
PFOS
PFDS
6:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
FOSAA
MeFOSAA
EtFOSAA

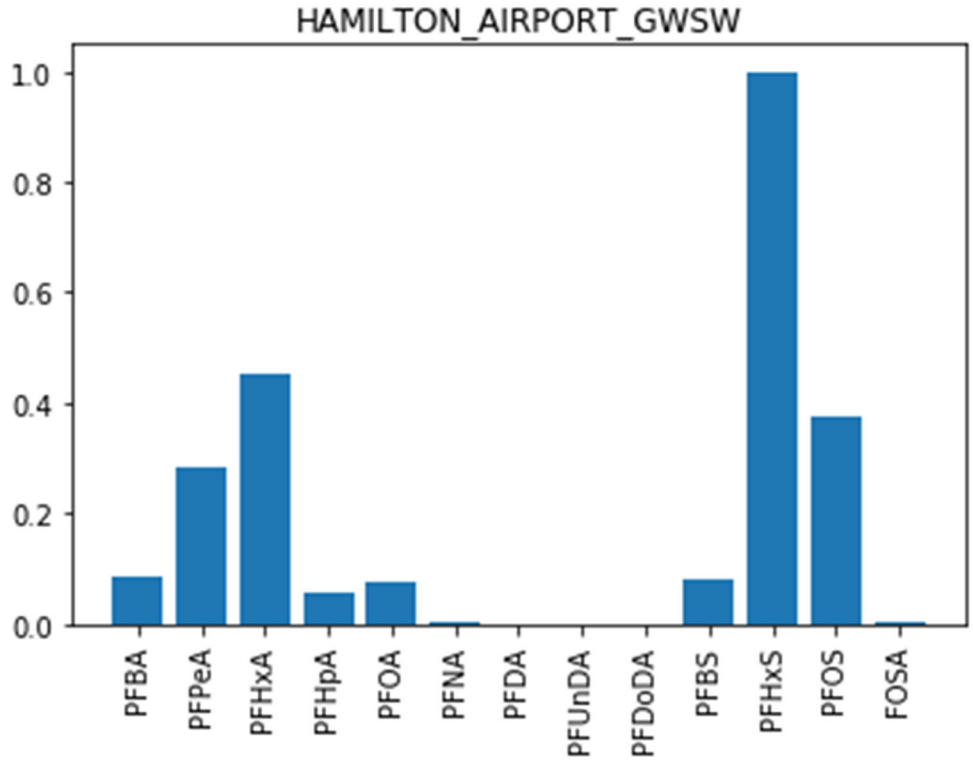

GOBELIUS_LF_GWSW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFHxDA
PFOcDA
PFBS
PFHxS
PFOS
PFDS
6:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
FOSAA
MeFOSAA
EtFOSAA

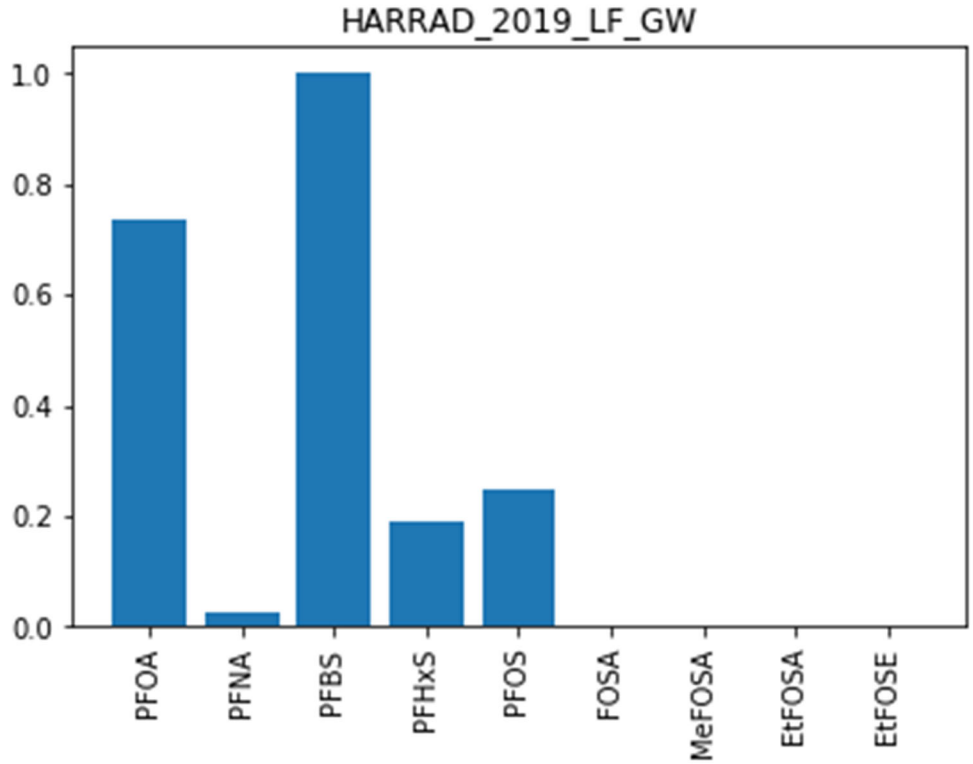

GOBELIUS_OTHER_GWSW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFHxDA
PFOcDA
PFBS
PFHxS
PFOS
PFDS
6:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
FOSAA
MeFOSAA
EtFOSAA

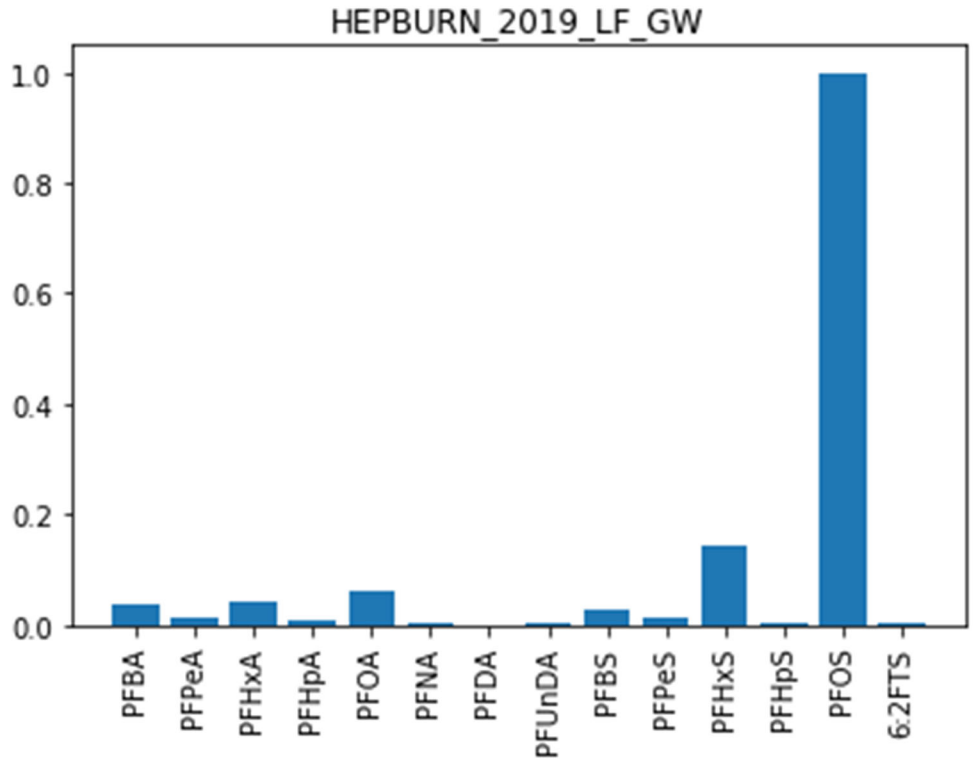

GOBELIUS_SKIING
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFHxDA
PFOcDA
PFBS
PFHxS
PFOS
PFDS
6:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
FOSAA
MeFOSAA
EtFOSAA

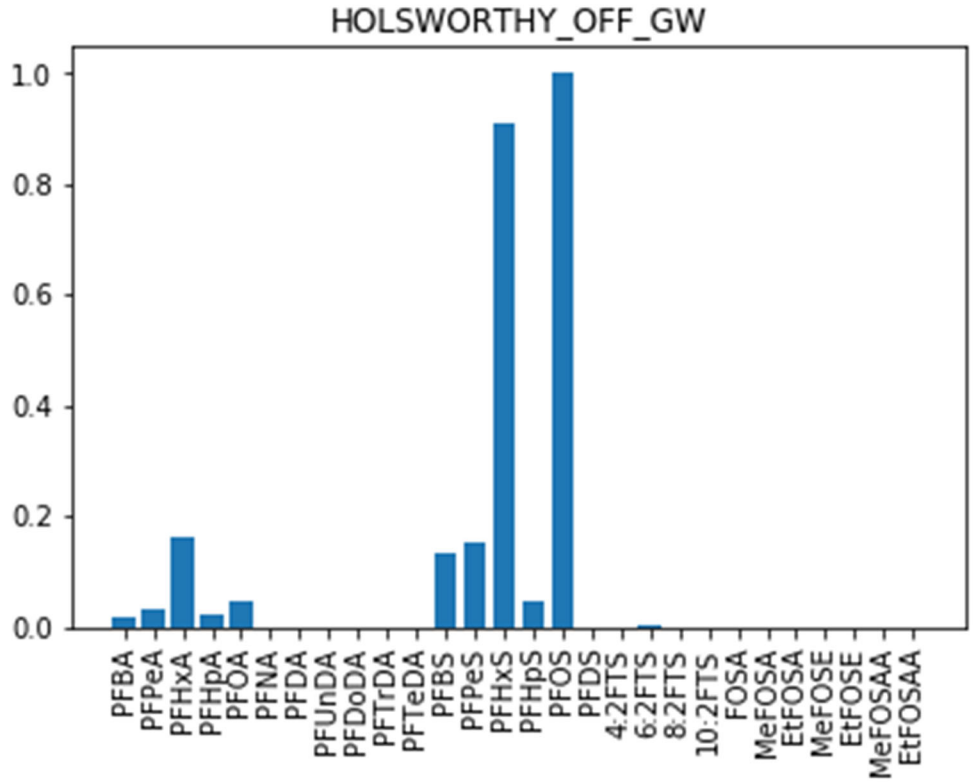

GOBELIUS_URBAN
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFHxDA
PFOcDA
PFBS
PFHxS
PFOS
PFDS
6:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
FOSAA
MeFOSAA
EtFOSAA

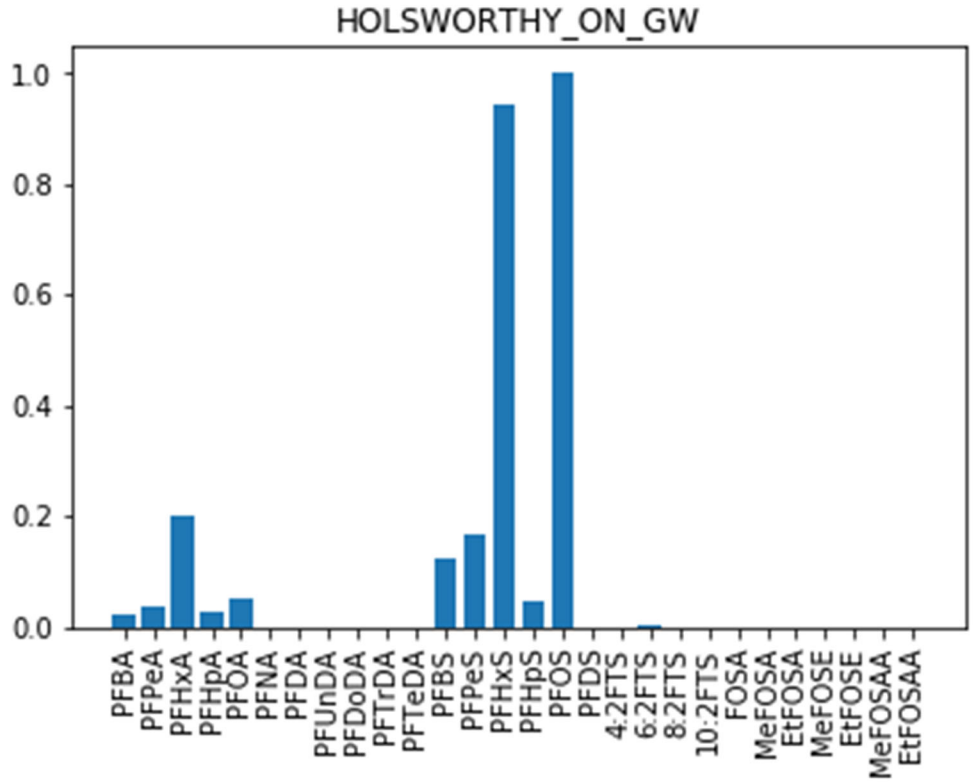

GOBELIUS_WWTP
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFHxDA
PFOcDA
PFBS
PFHxS
PFOS
PFDS
6:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
FOSAA
MeFOSAA
EtFOSAA

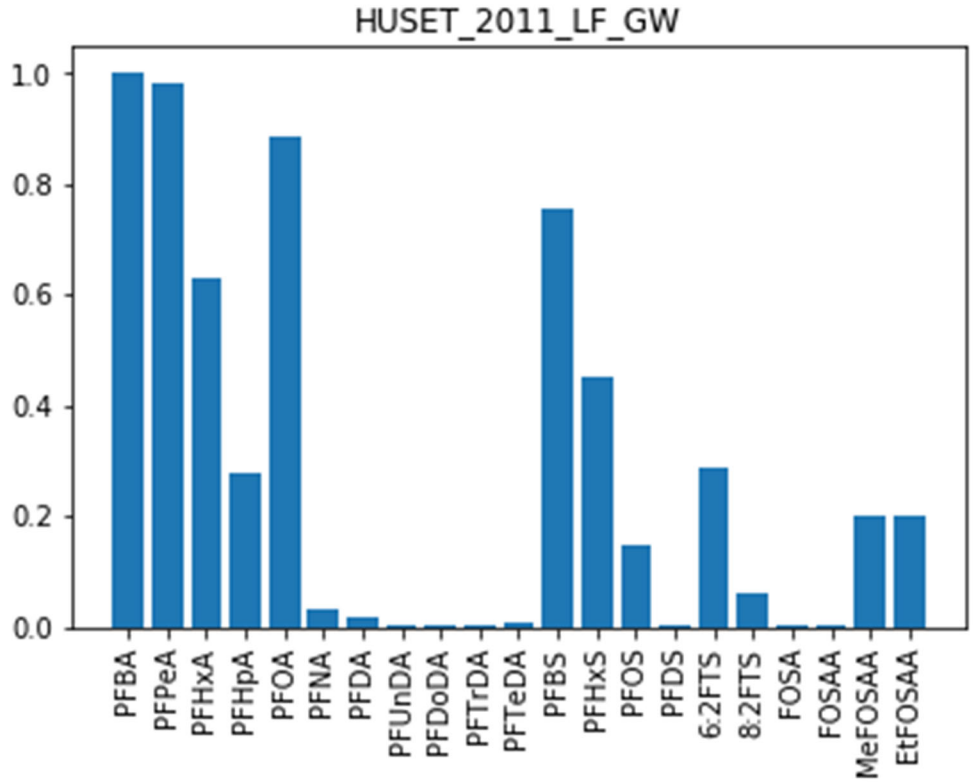

HAMILTON_AIRPORT_GWSW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFBS
PFHxS
PFOS
FOSA

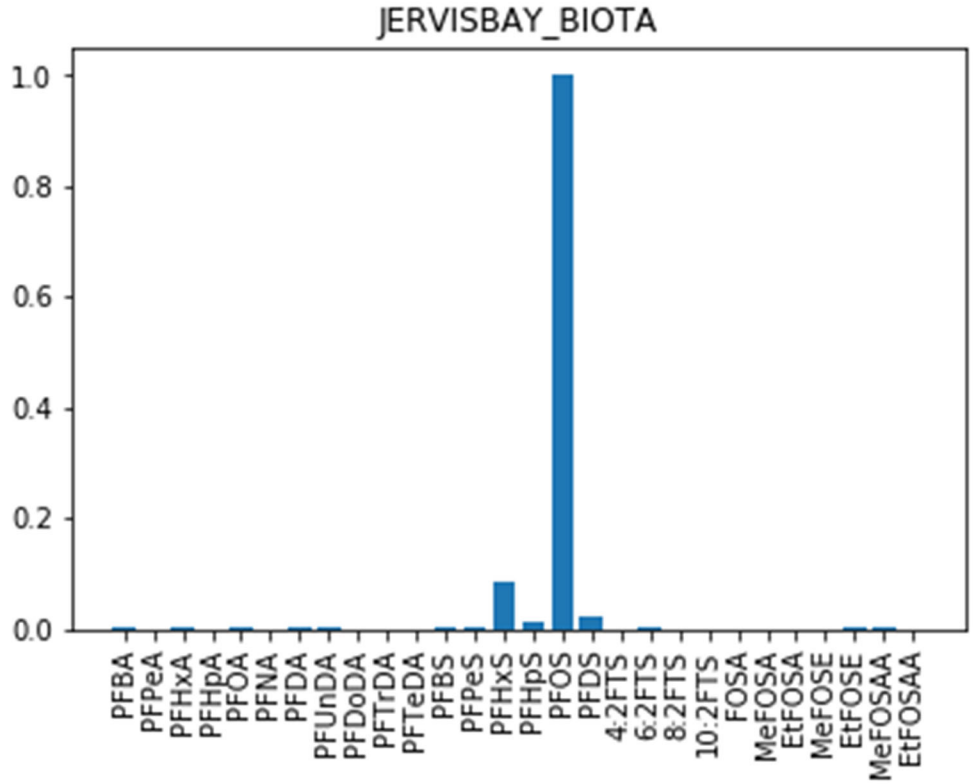

HARRAD_2019_LF_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFOA
PFNA
PFBS
PFHxS
PFOS
FOSA
MeFOSA
EtFOSA
EtFOSE

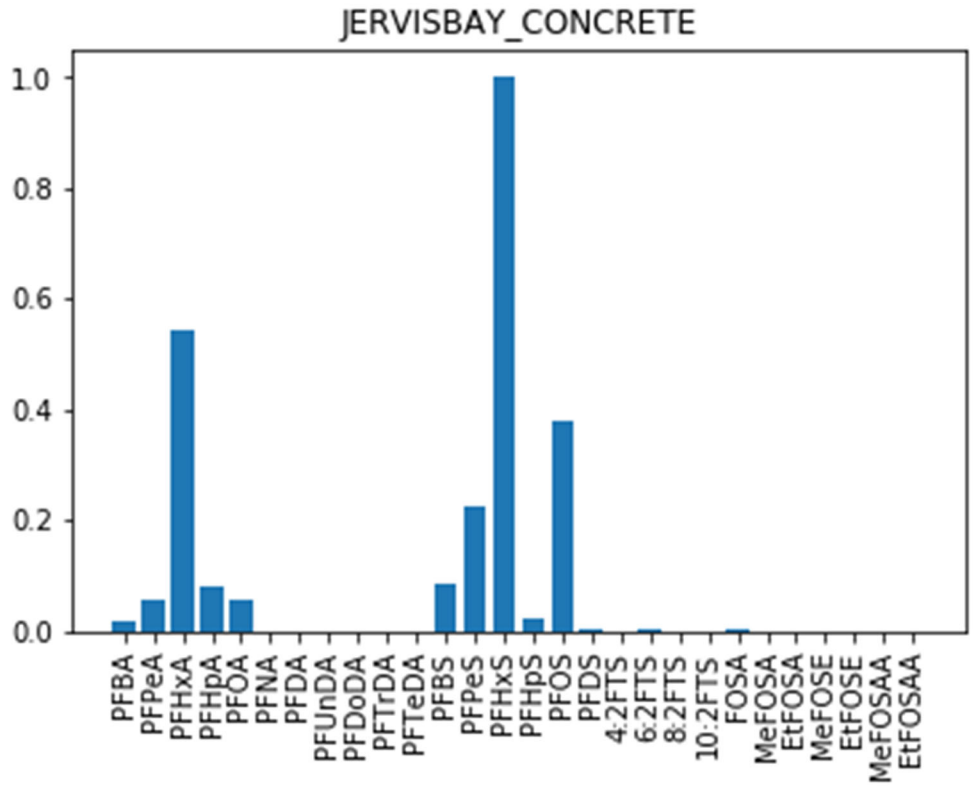

HEPBURN_2019_LF_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
6:2FTS

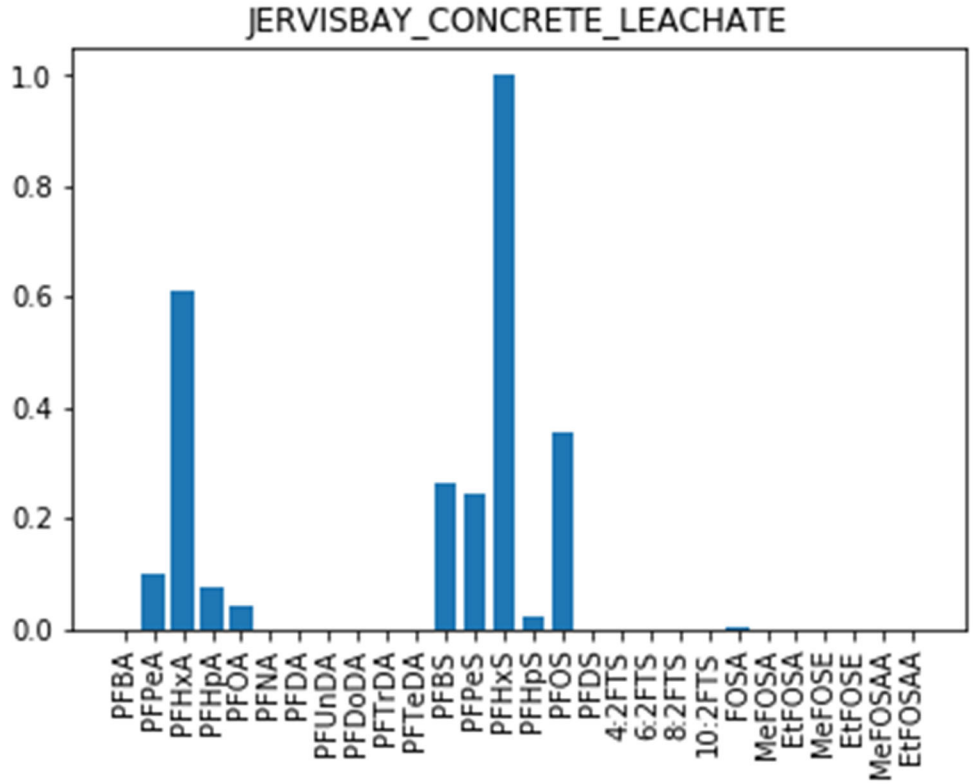

HOLSWORTHY_OFF_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

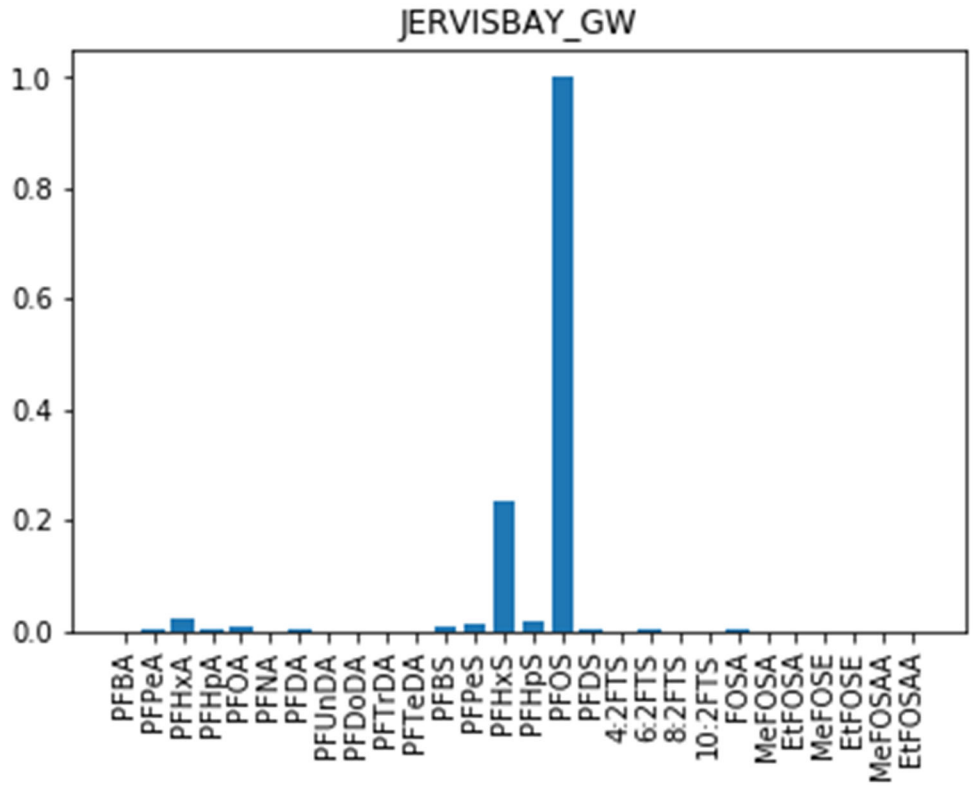

HOLSWORTHY_ON_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

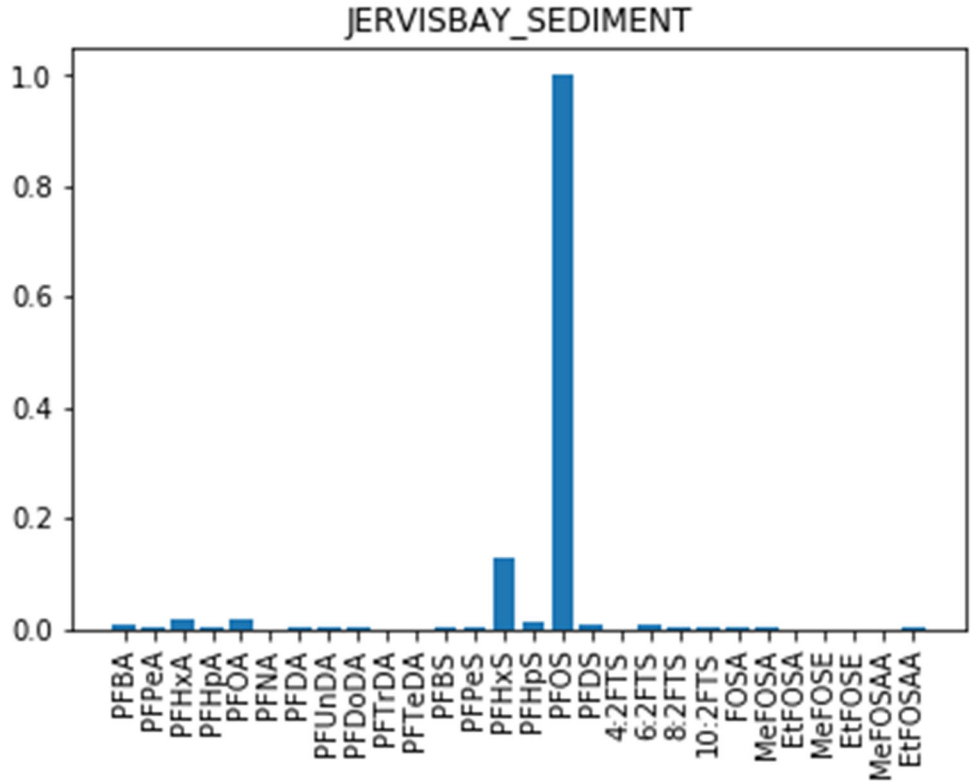

HUSET_2011_LF_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFOS
PFDS
6:2FTS
8:2FTS
FOSA
FOSAA
MeFOSAA
EtFOSAA

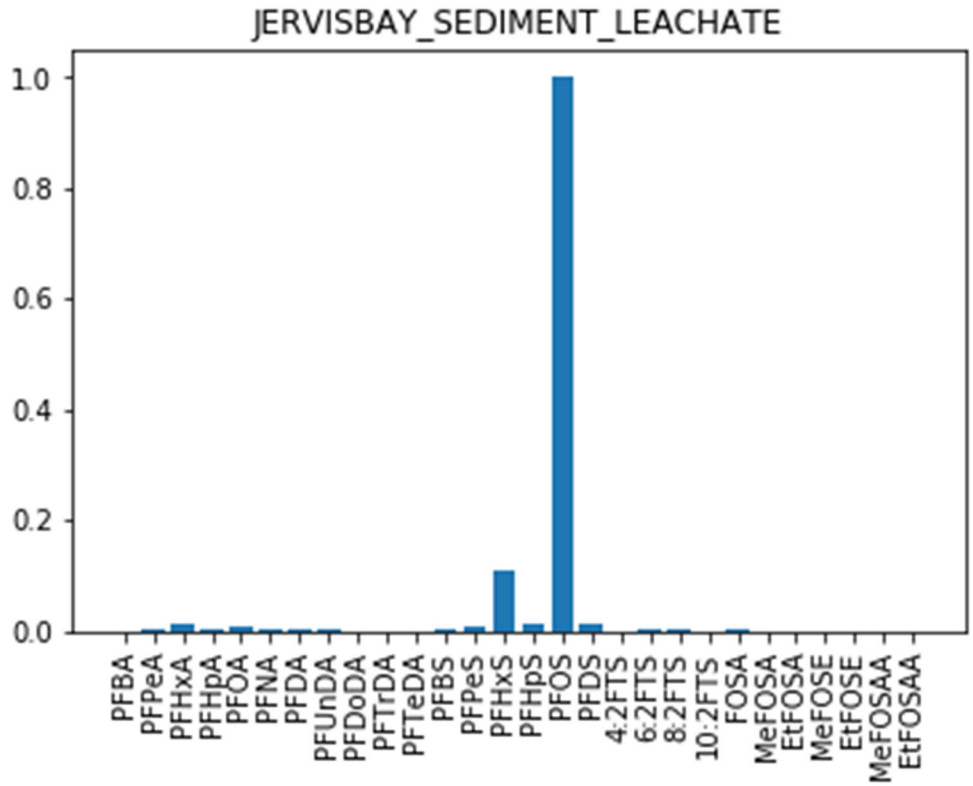

JERVISBAY_BIOTA
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

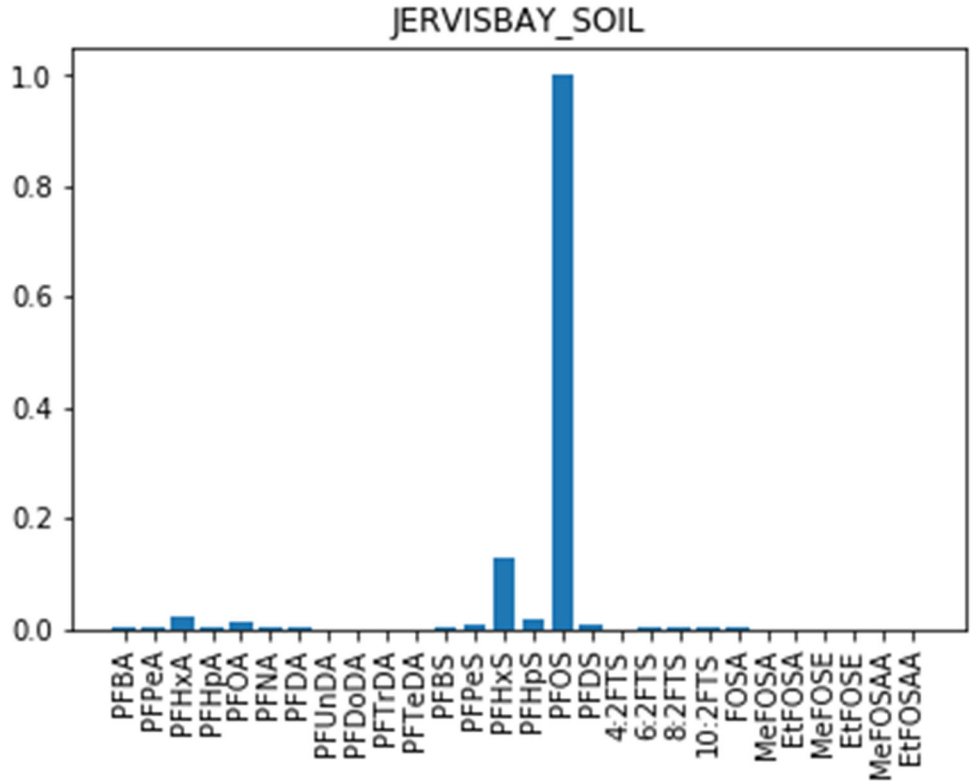

JERVISBAY_CONCRETE
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

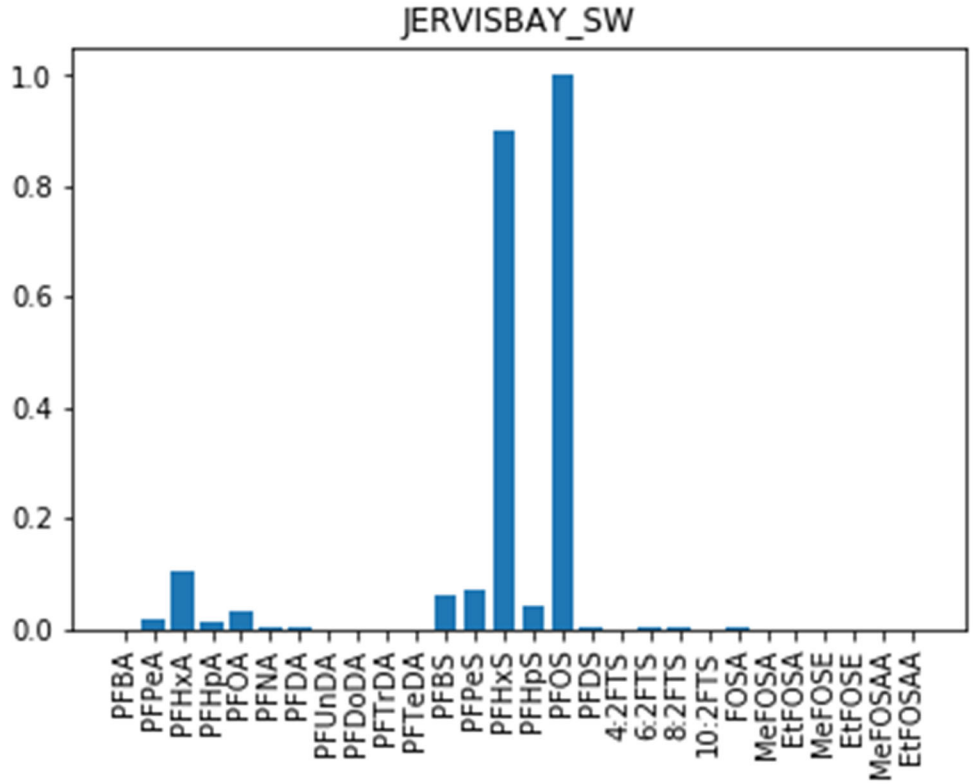

JERVISBAY_CONCRETE_LEACHATE
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

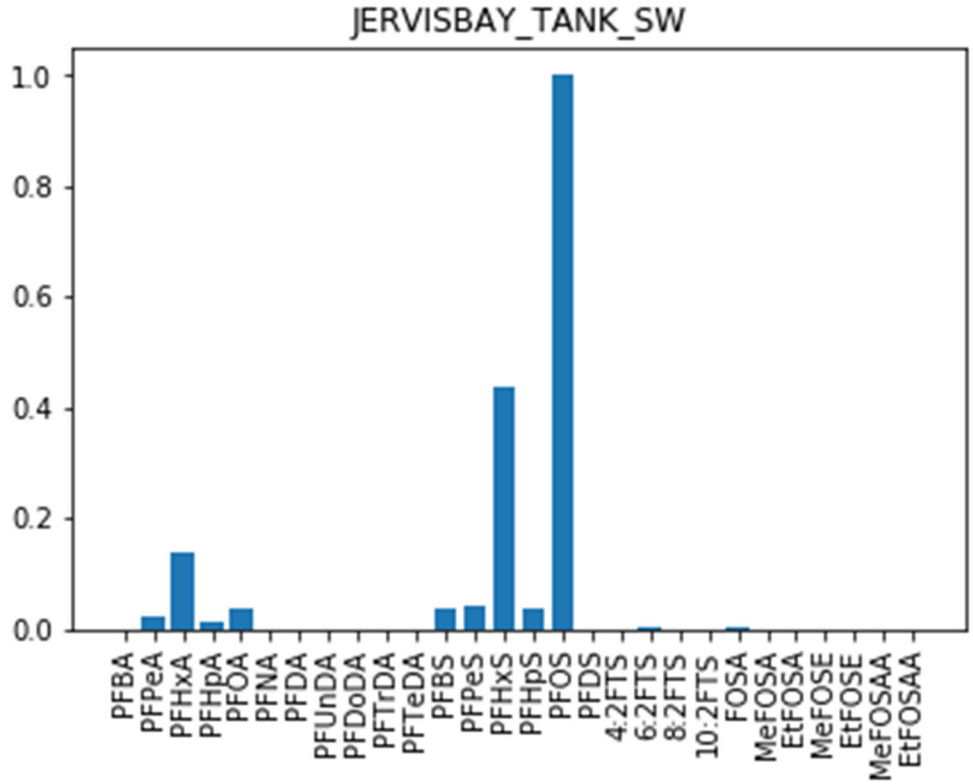

JERVISBAY_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

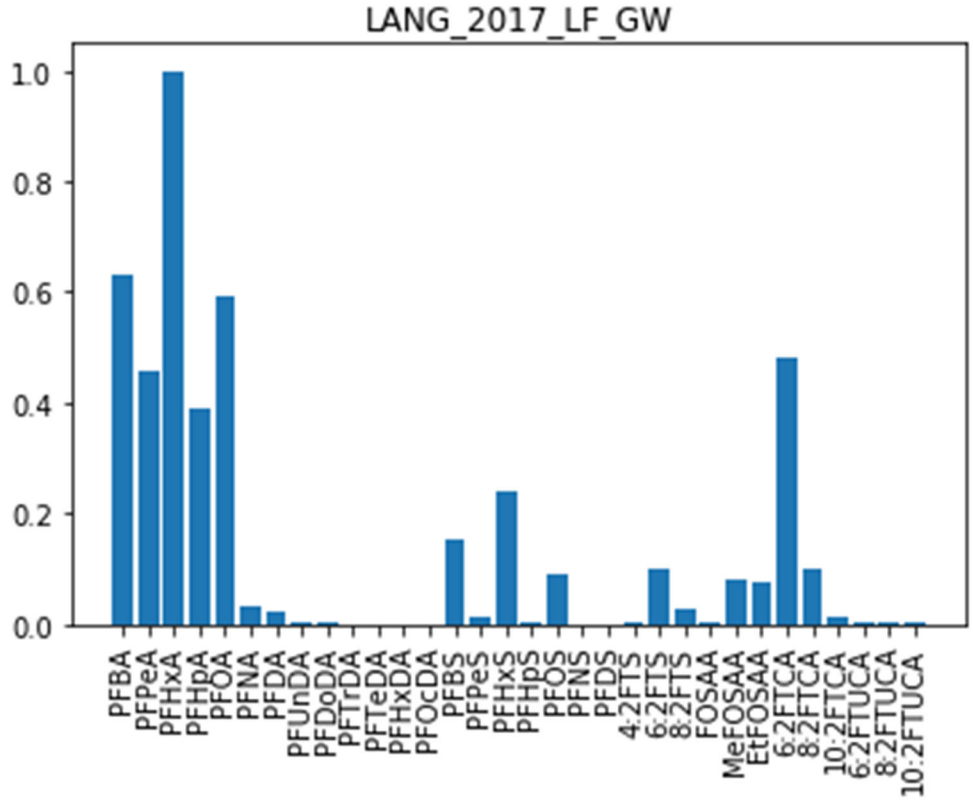

JERVISBAY_SEDIMENT
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

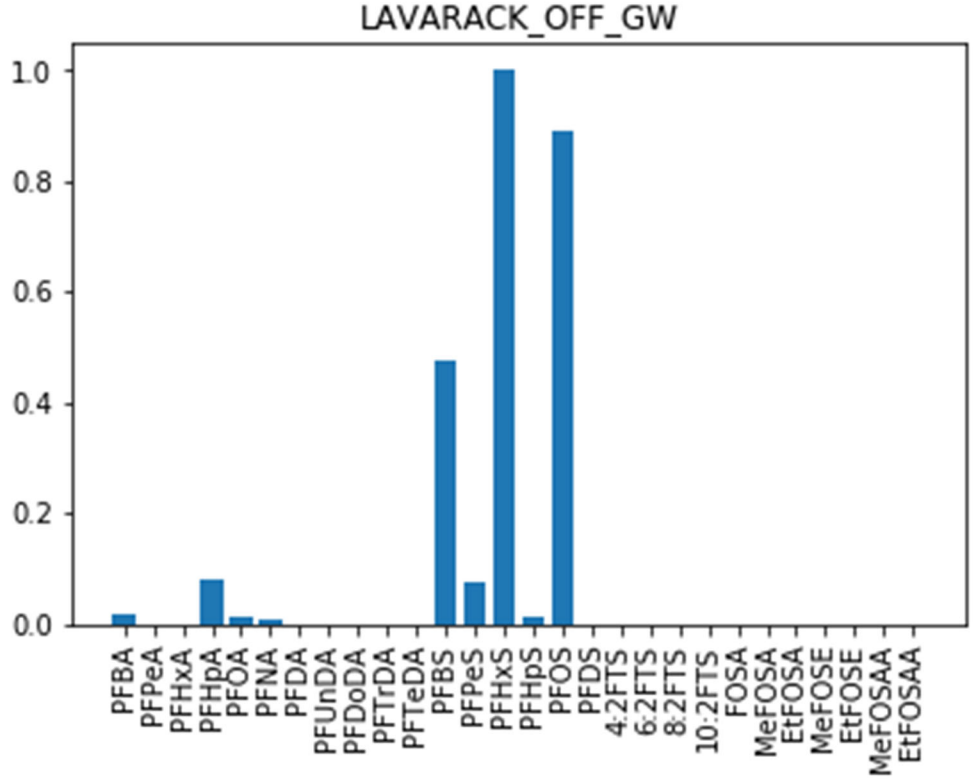

JERVISBAY_SEDIMENT_LEACHATE
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

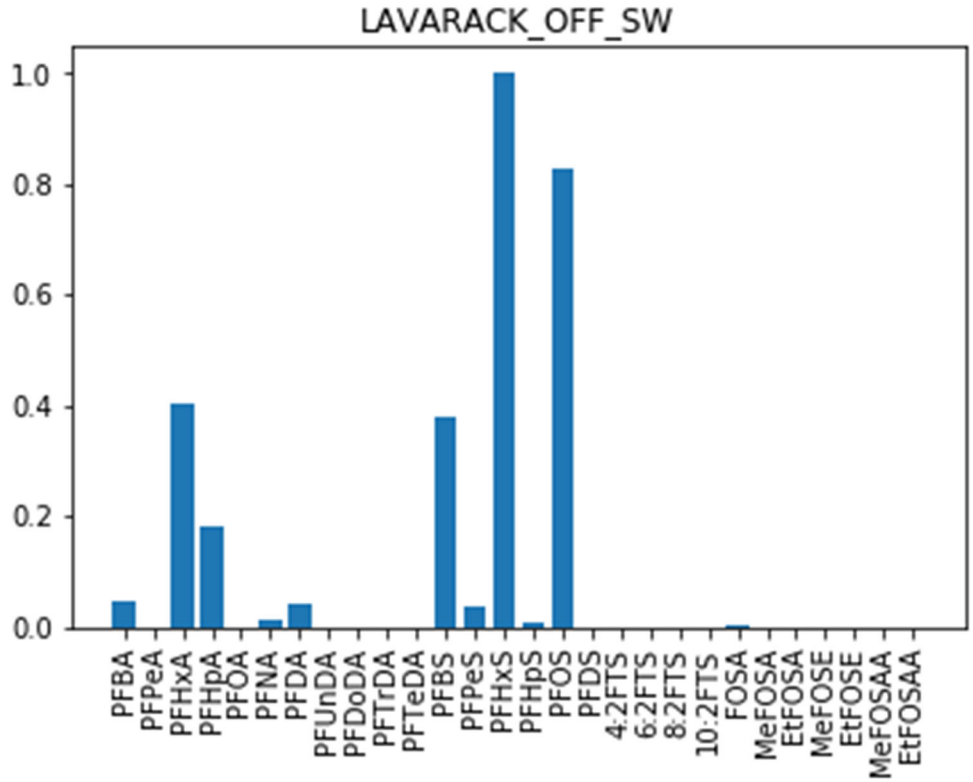

JERVISBAY_SOIL
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

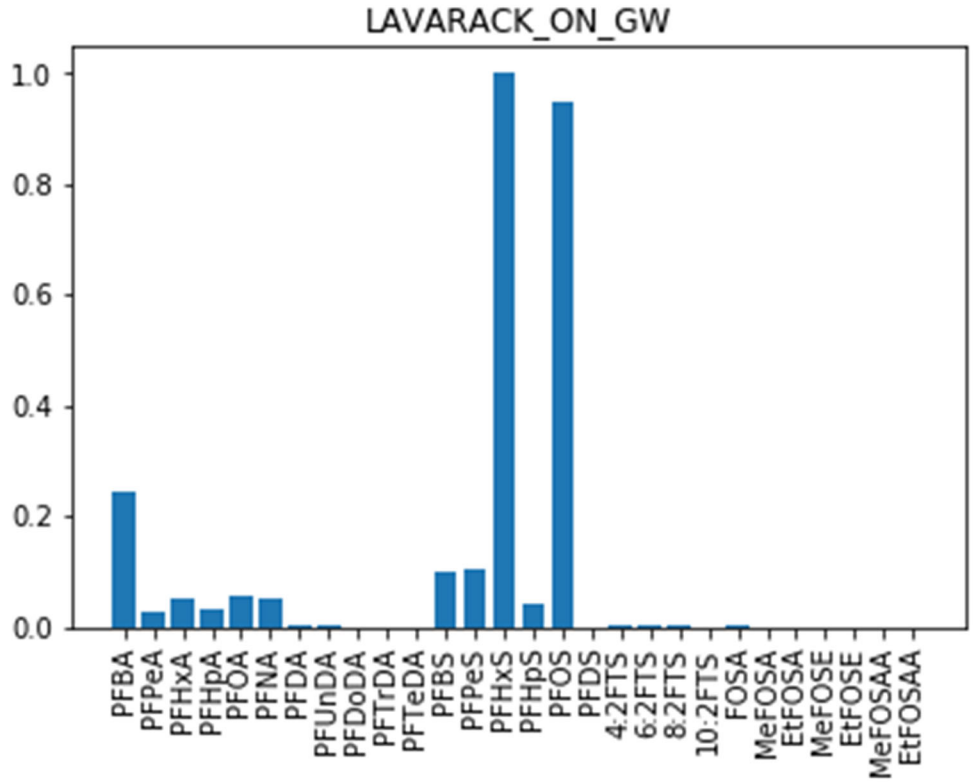

JERVISBAY_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

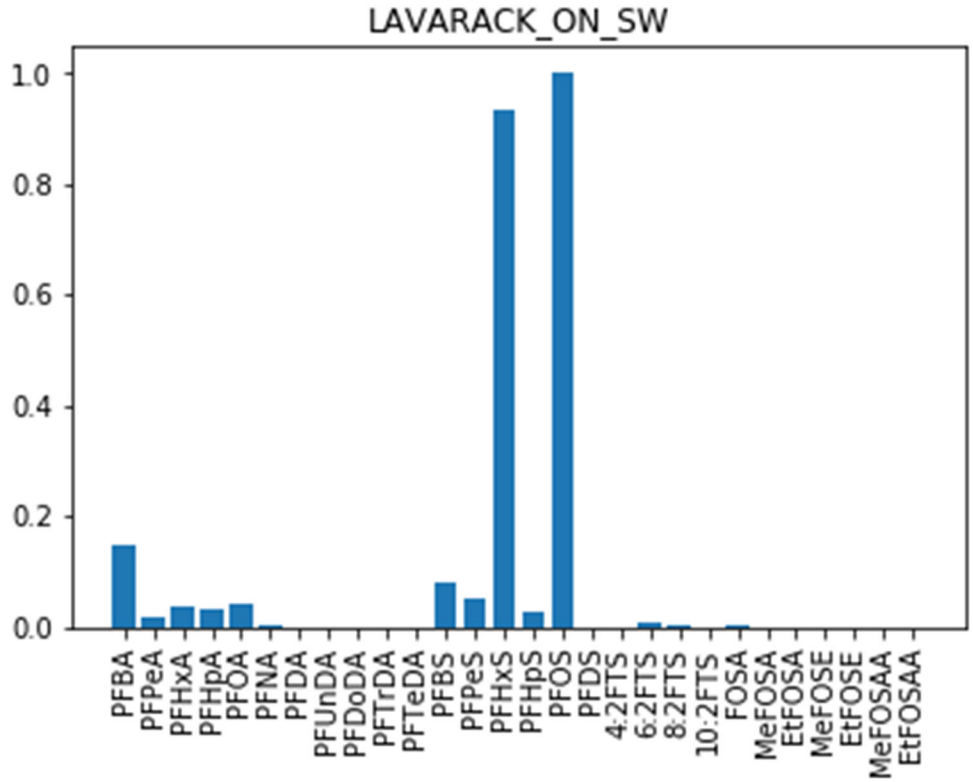

JERVISBAY_TANK_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

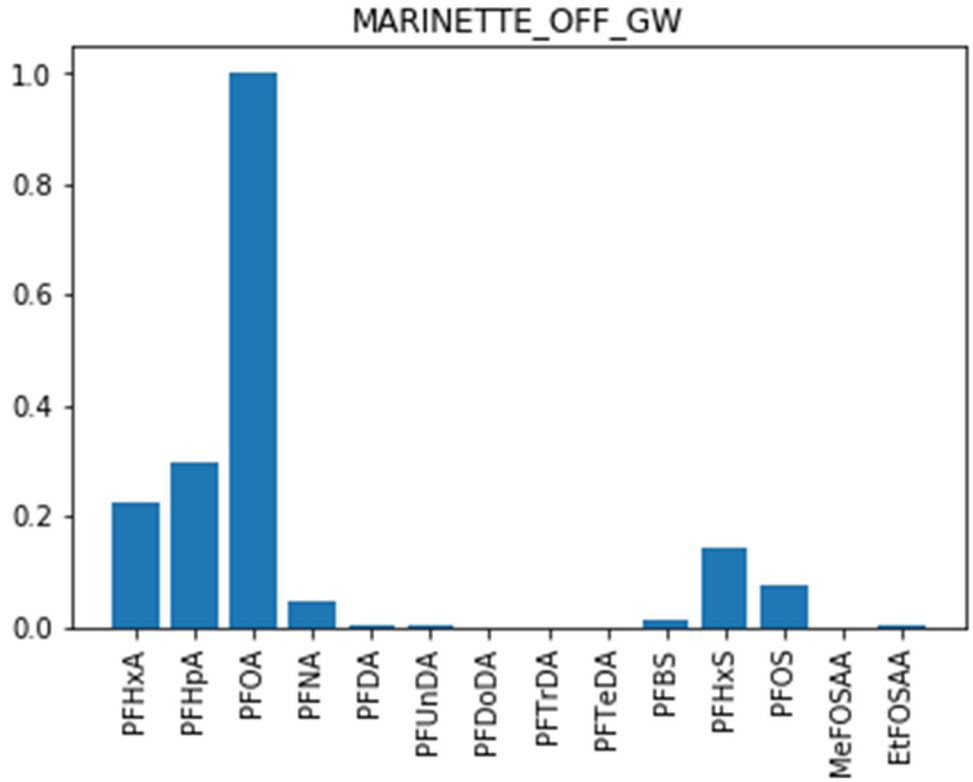

LANG_2017_LF_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFHxDA
PFOcDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
FOSAA
MeFOSAA
EtFOSAA
6:2FTCA
8:2FTCA
10:2FTCA
6:2FTUCA
8:2FTUCA
10:2FTUCA

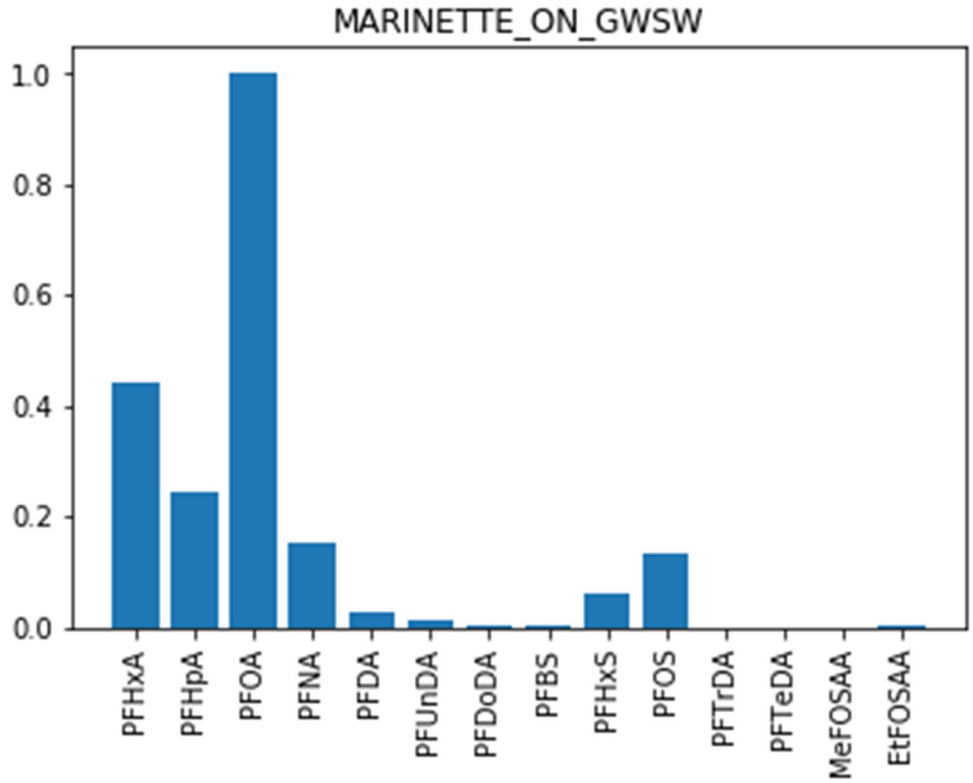

LAVARACK_OFF_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

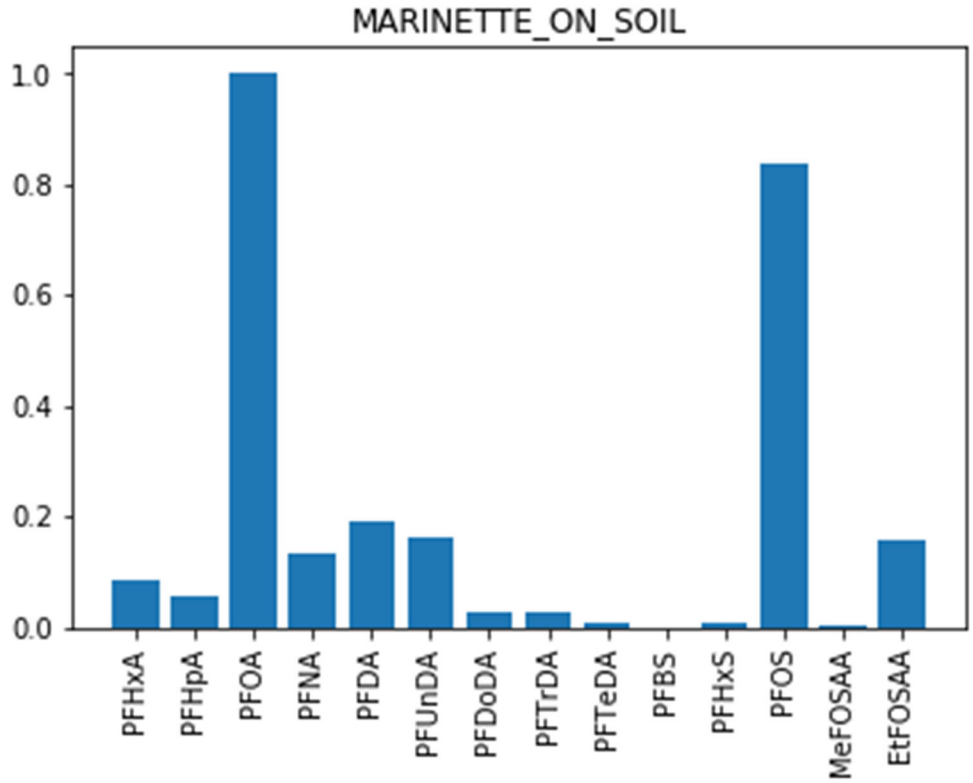

LAVARACK_OFF_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

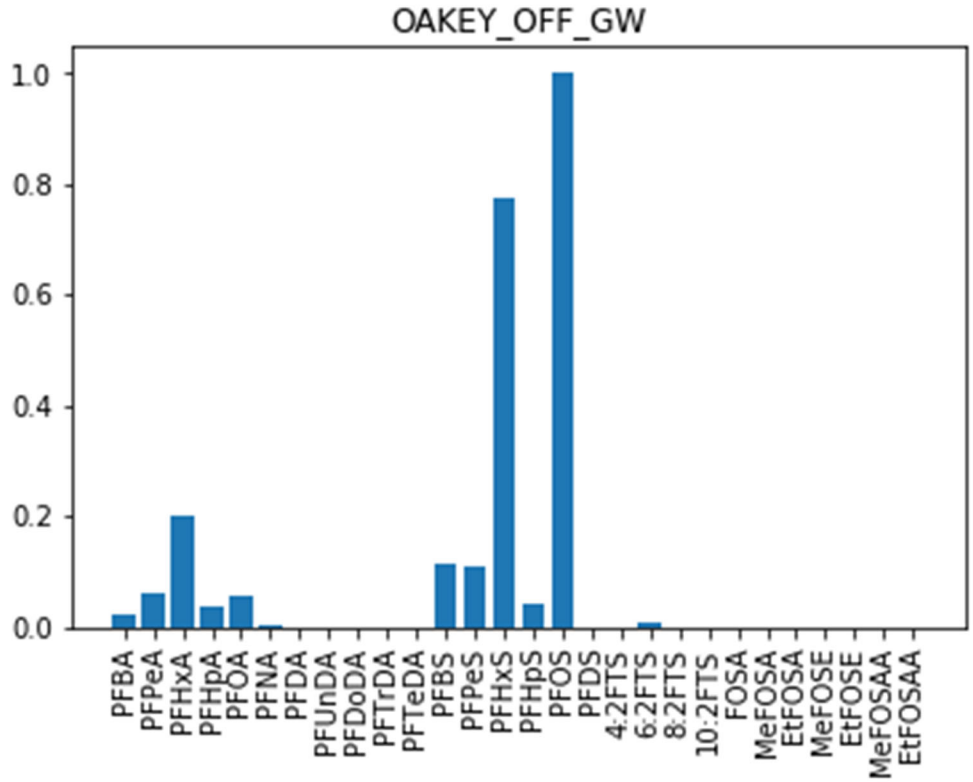

LAVARACK_ON_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

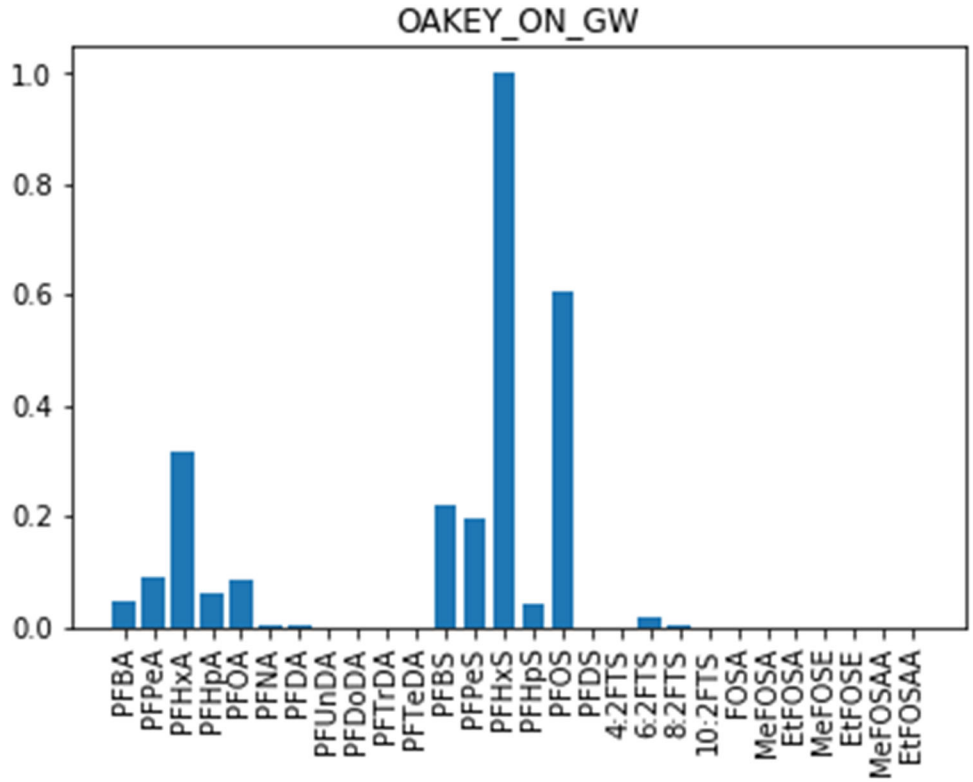

LAVARACK_ON_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

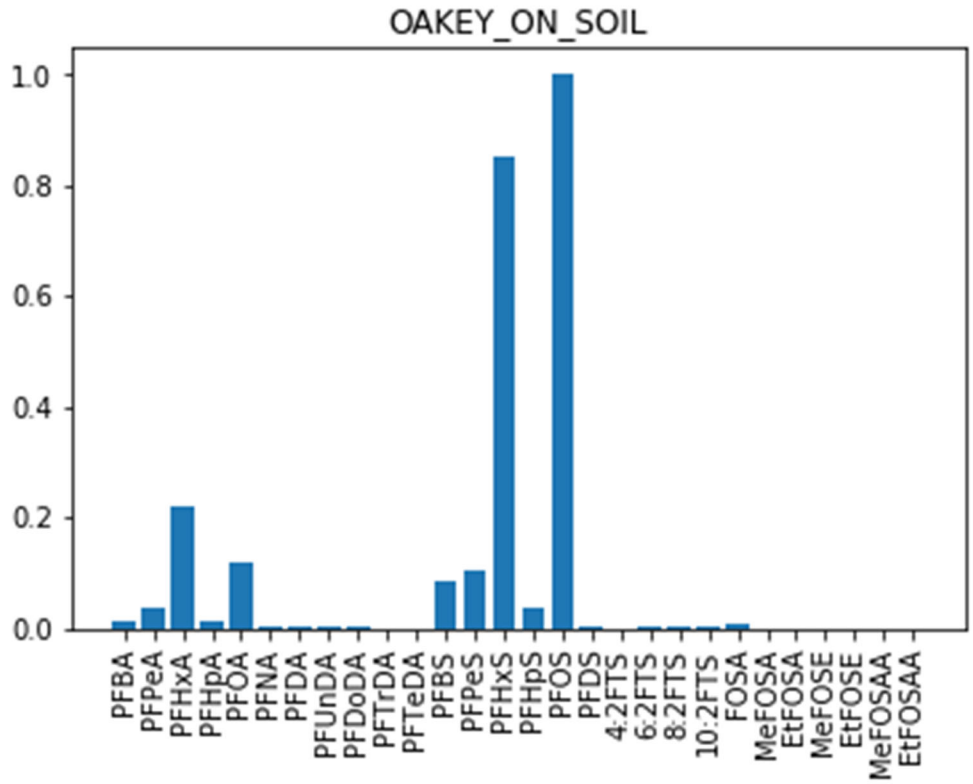

MARINETTE_OFF_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFOS
MeFOSAA
EtFOSAA

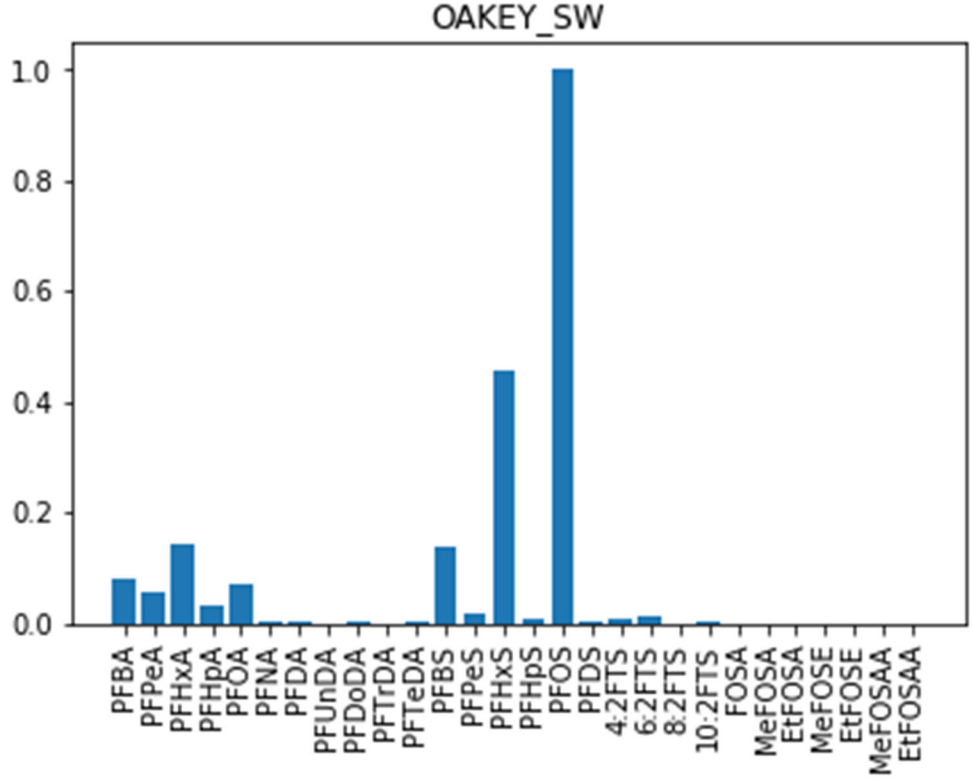

MARINETTE_ON_GWSW
1.0
0.8
0.6
0.4
0.2
0.0
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFBS
PFHxS
PFOS
PFTrDA
PFTeDA
MeFOSAA
EtFOSAA

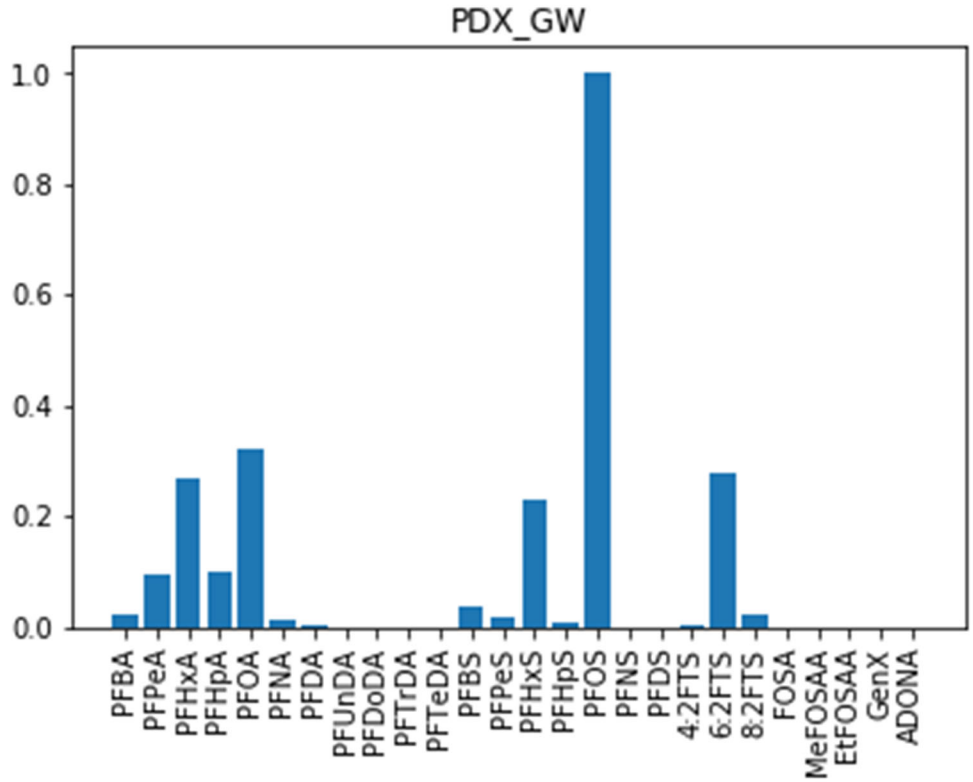

MARINETTE_ON_SOIL
1.0
0.8
0.6
0.4
0.2
0.0
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFOS
MeFOSAA
EtFOSAA

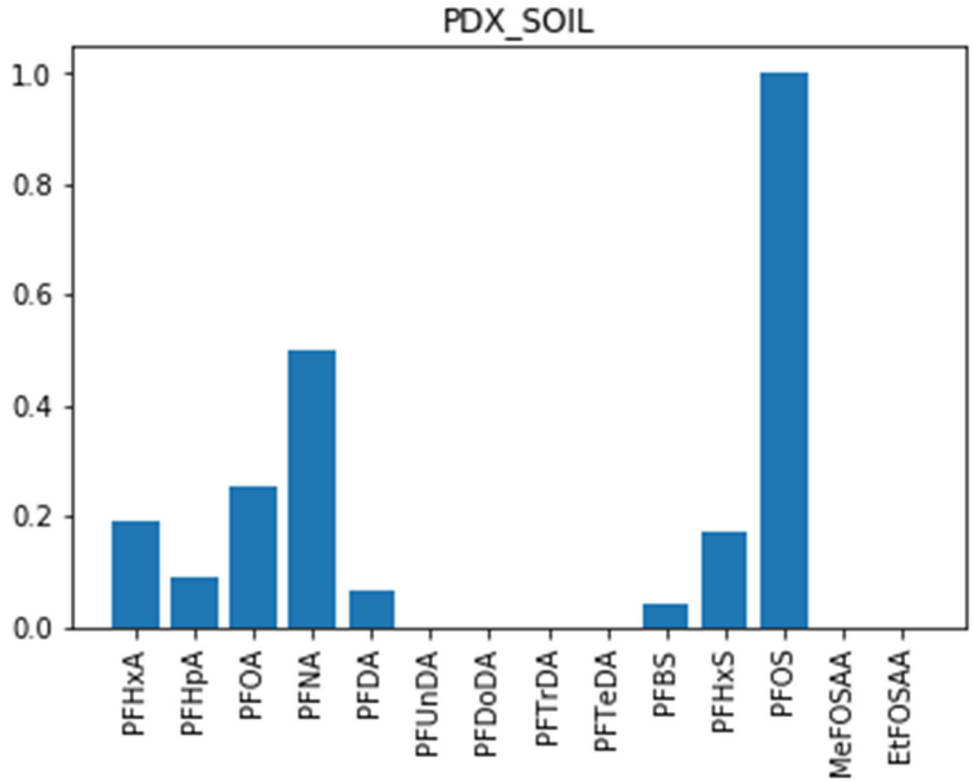

OAKEY_OFF_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

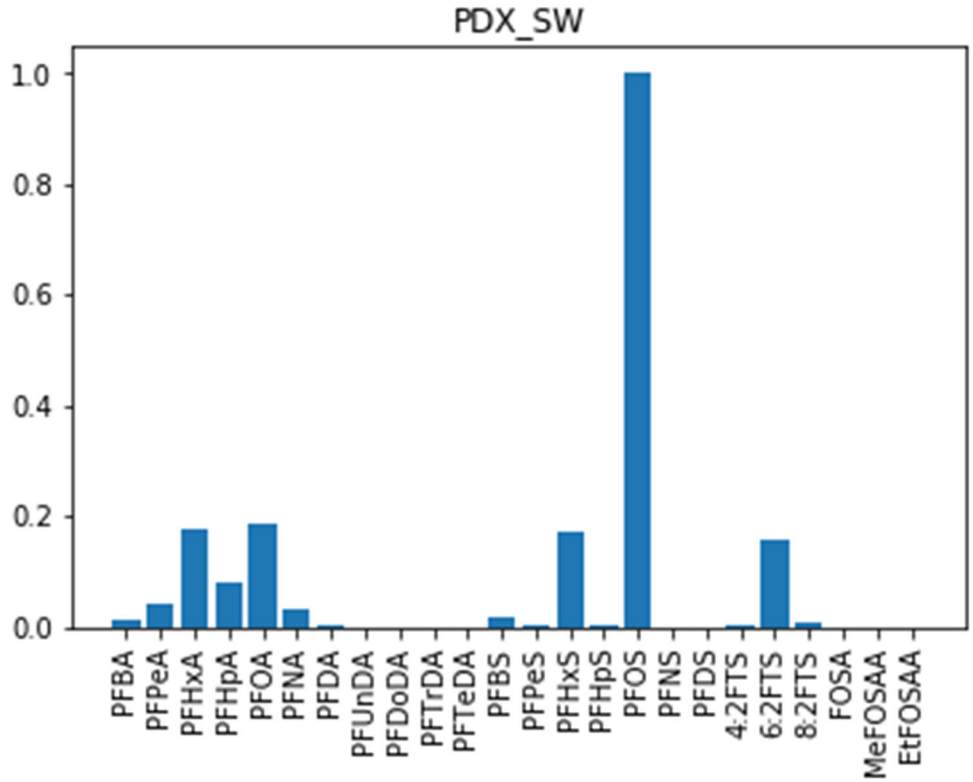

OAKEY_ON_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

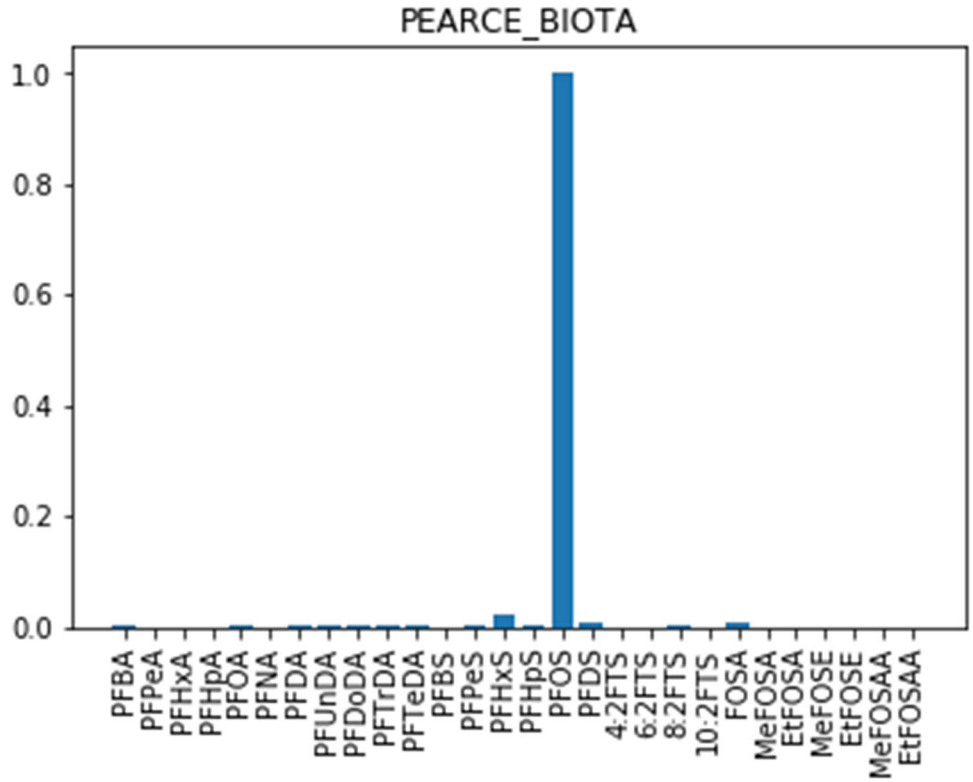

OAKEY_ON_SOIL
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

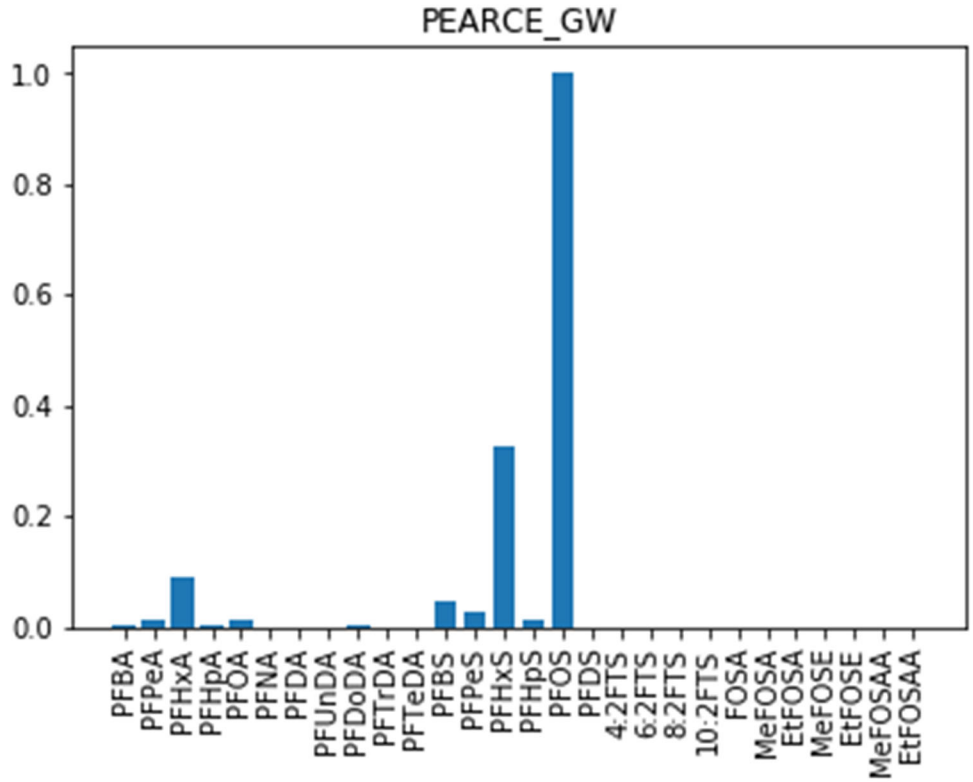

OAKEY_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

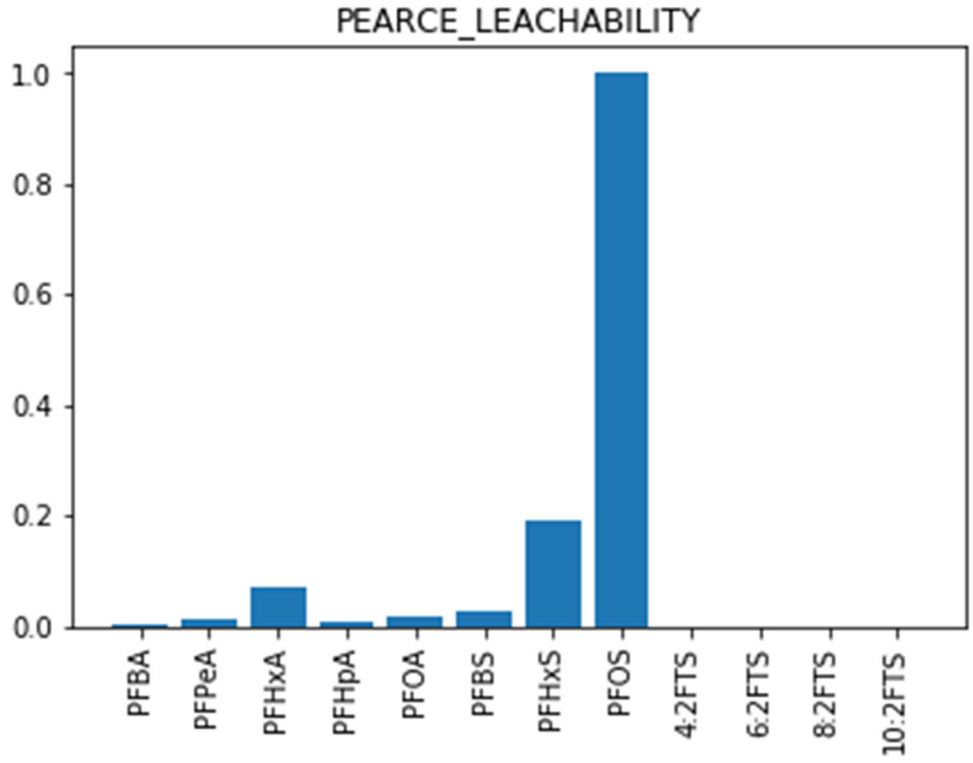

PDX_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
FOSA
MeFOSAA
EtFOSAA
GenX
ADONA

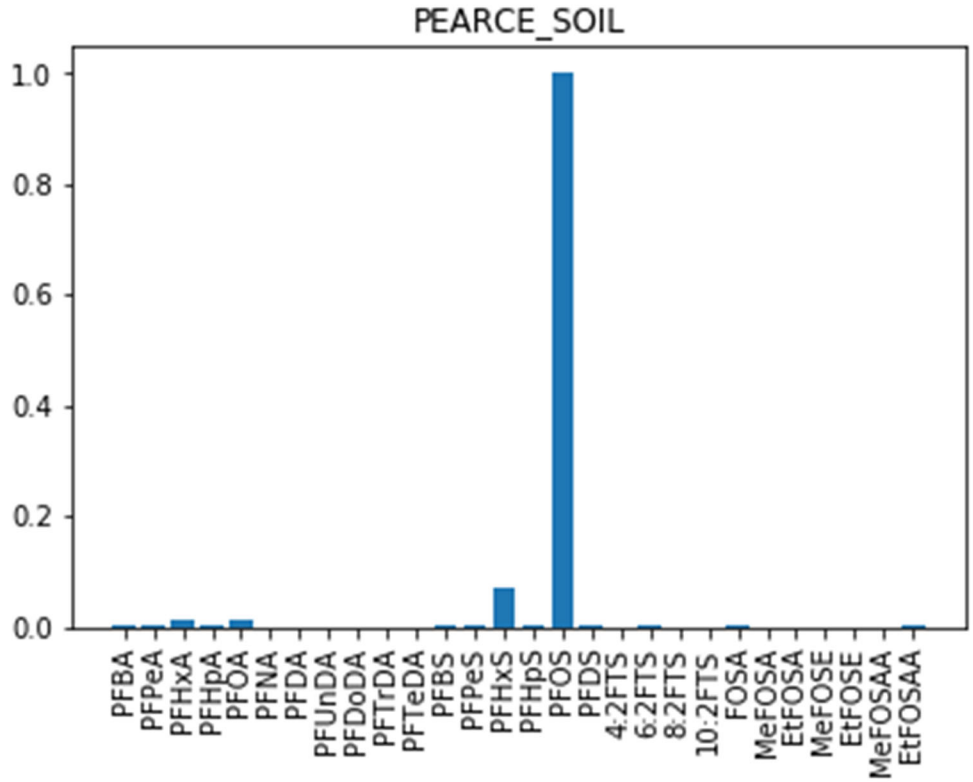

PDX_SOIL
1.0
0.8
0.6
0.4
0.2
0.0
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFHxS
PFOS
MeFOSAA
EtFOSAA

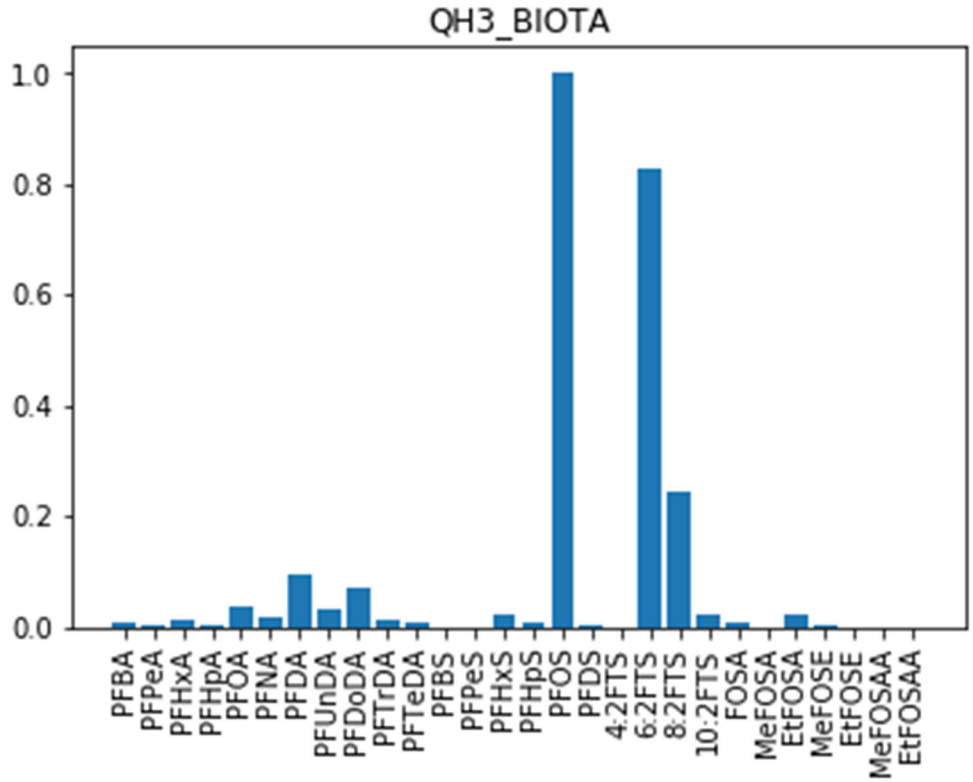

PDX_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
FOSA
MeFOSAA
EtFOSAA

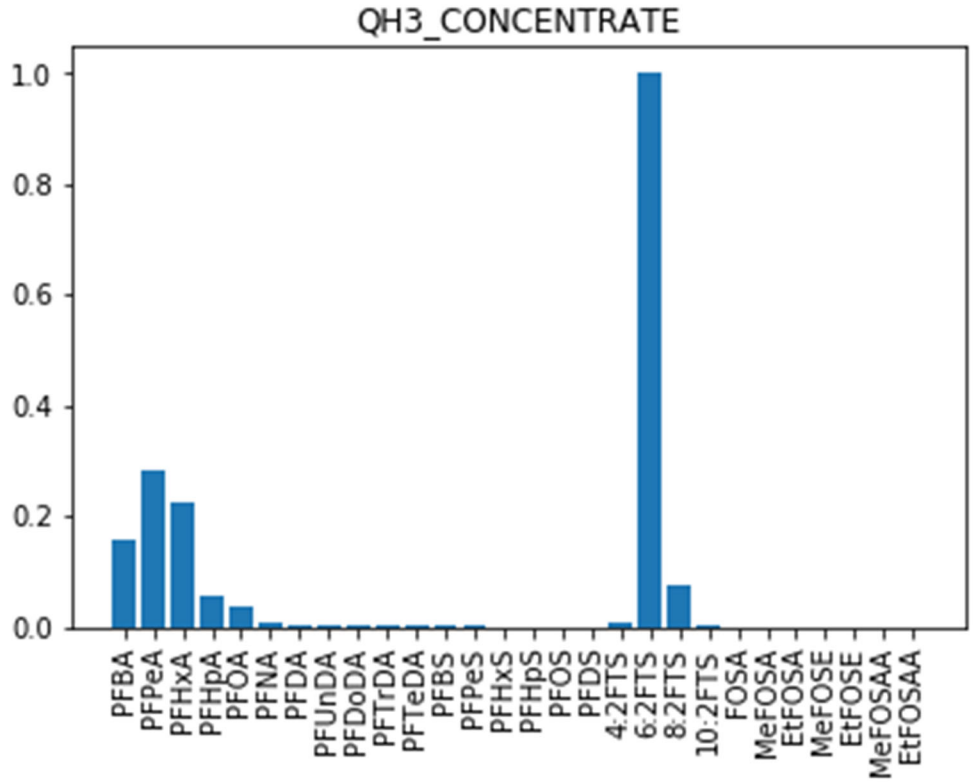

PEARCE_BIOTA
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

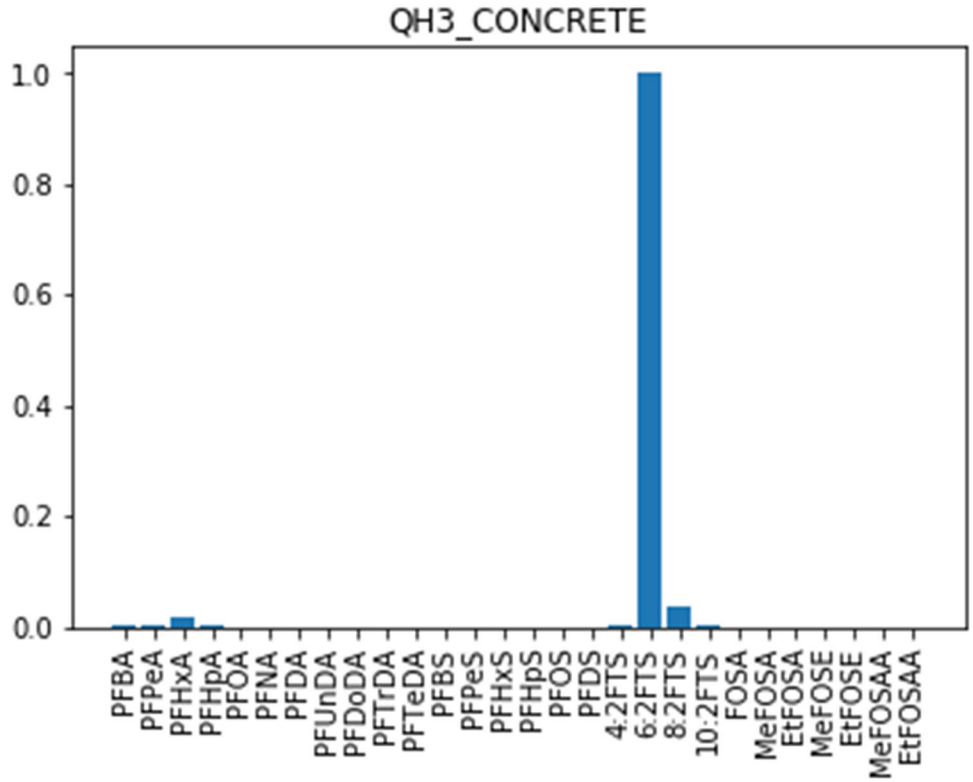

PEARCE_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

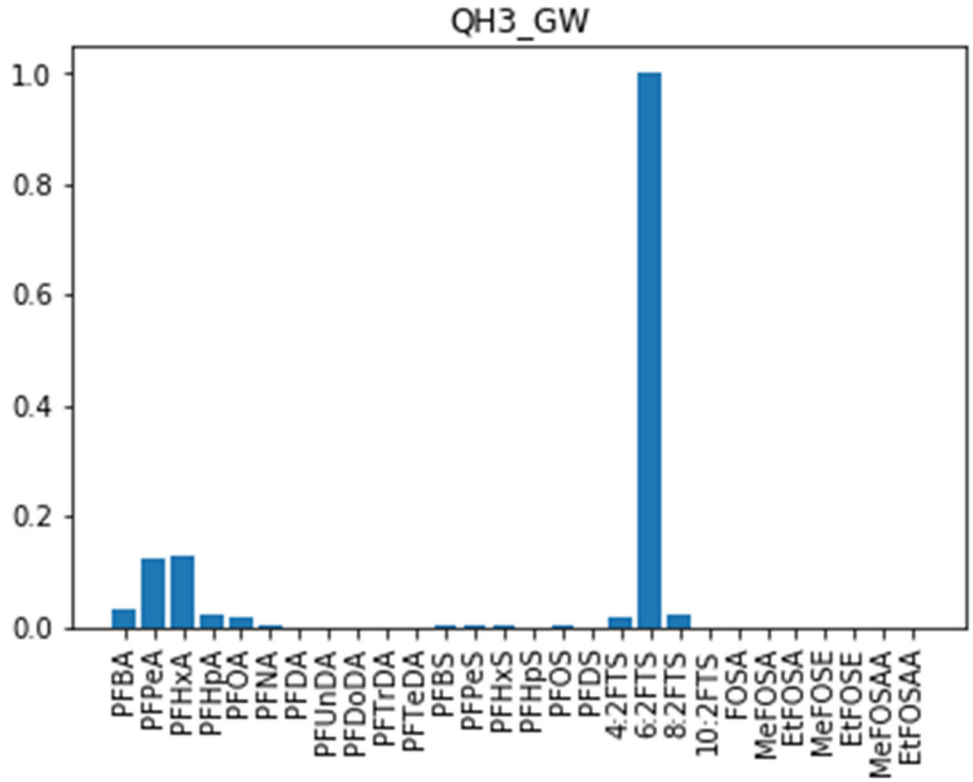

PEARCE_LEACHABILITY
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFBS
PFHxS
PFOS
4:2FTS
6:2FTS
8:2FTS
10:2FTS

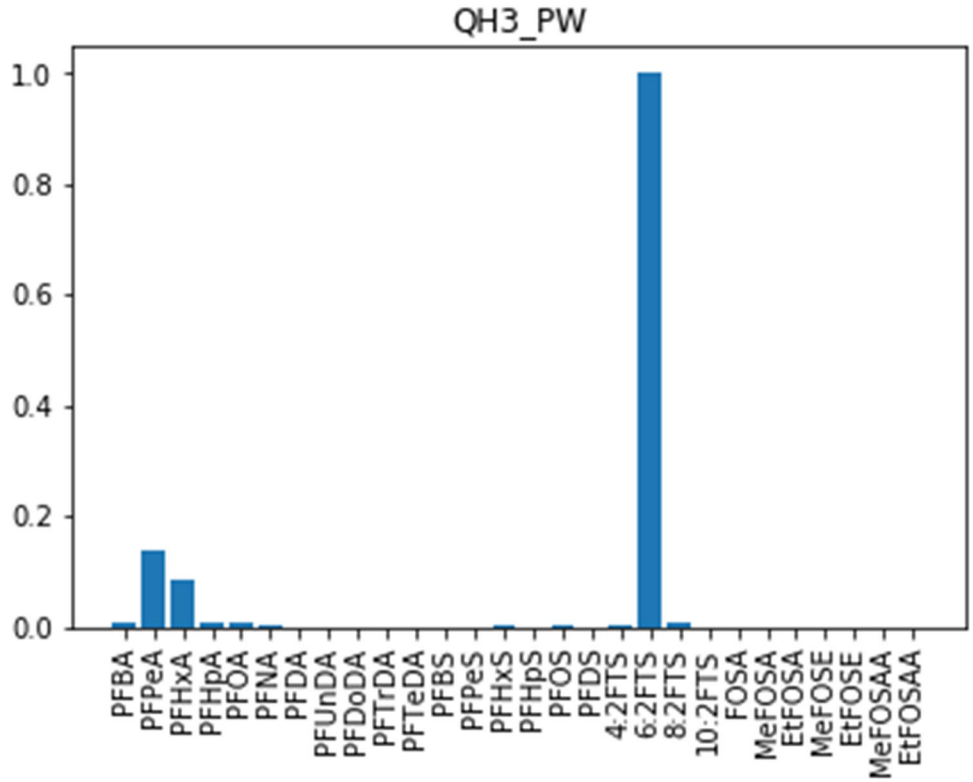

PEARCE_SOIL
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

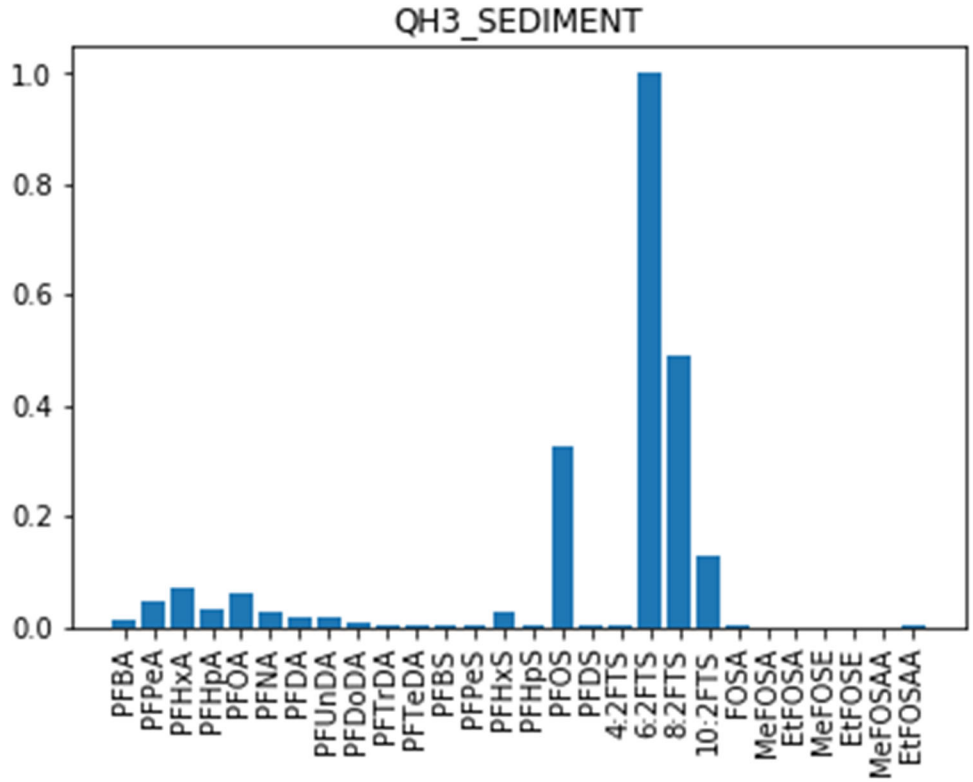

QH3_BIOTA
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

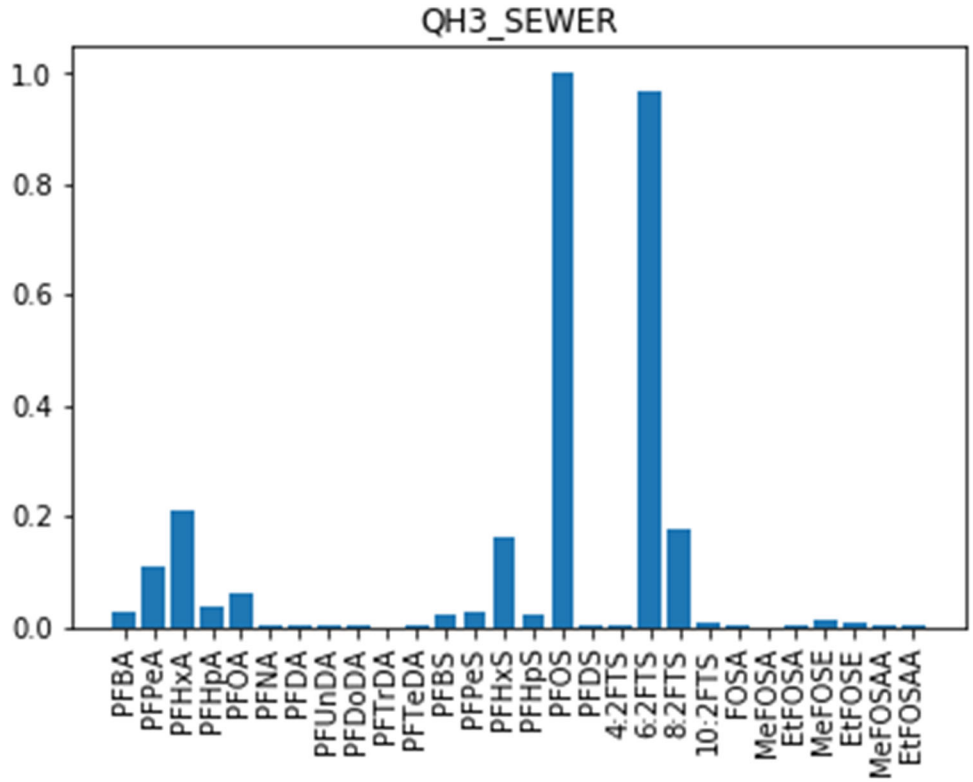

QH3_CONCENTRATE
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

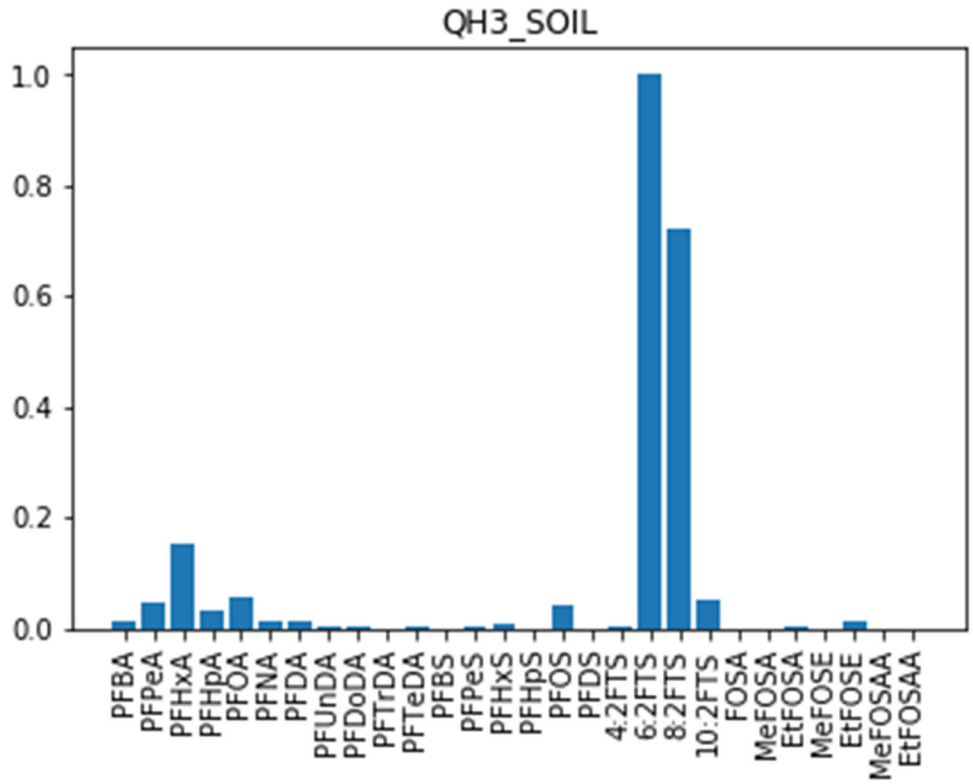

QH3_CONCRETE
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

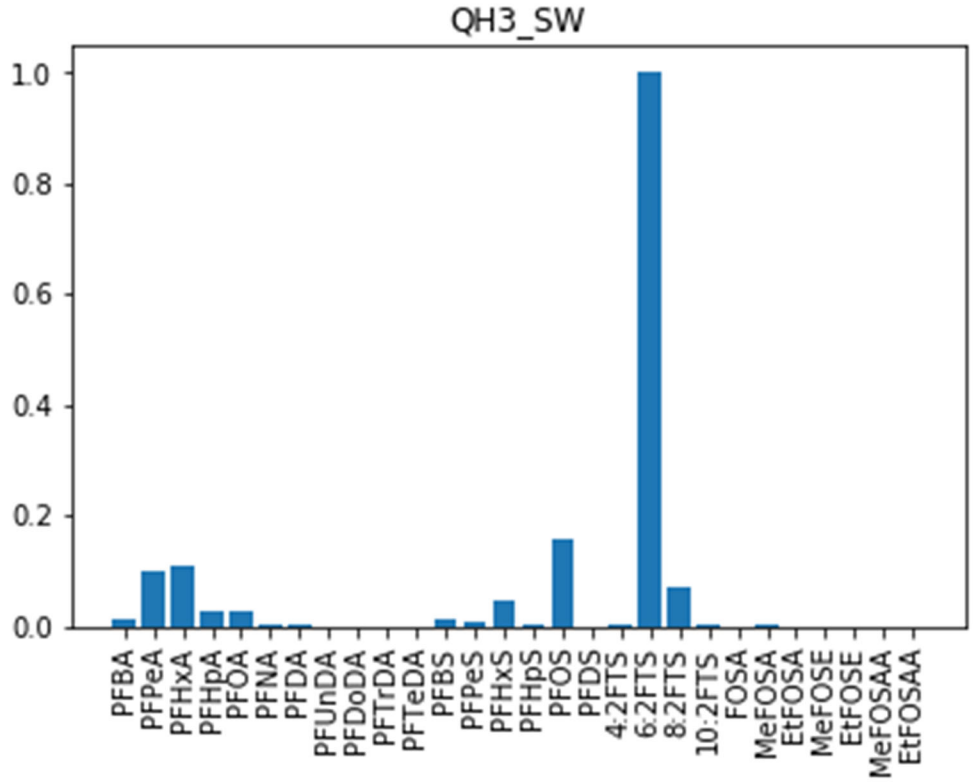

QH3_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

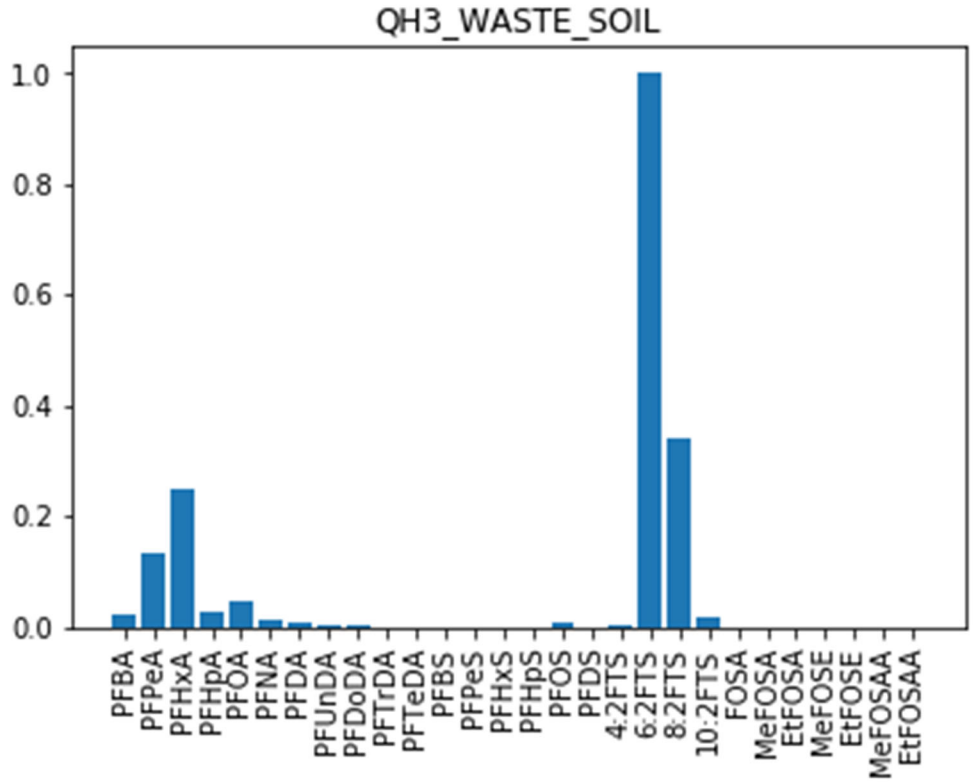

QH3_PW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

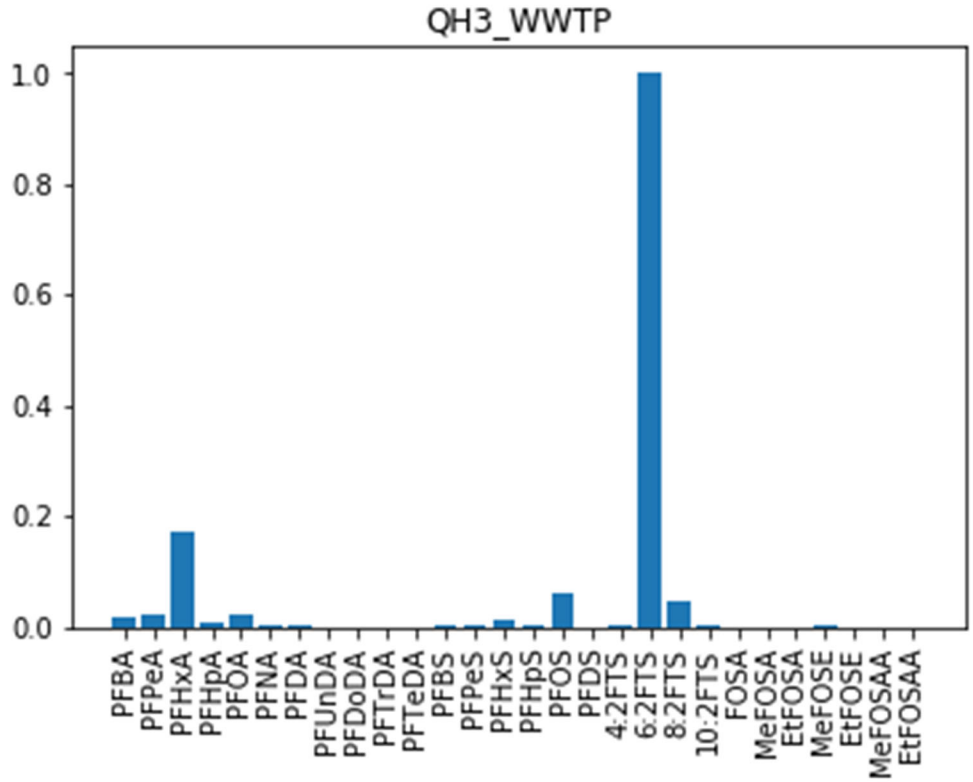

QH3_SEDIMENT
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

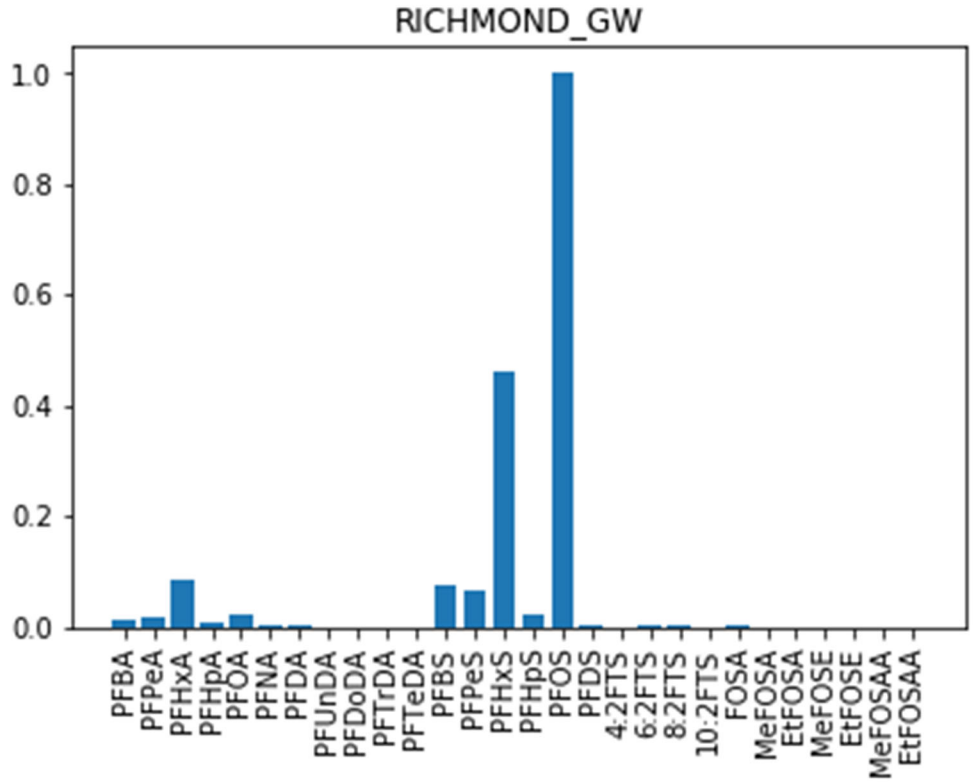

QH3_SEWER
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

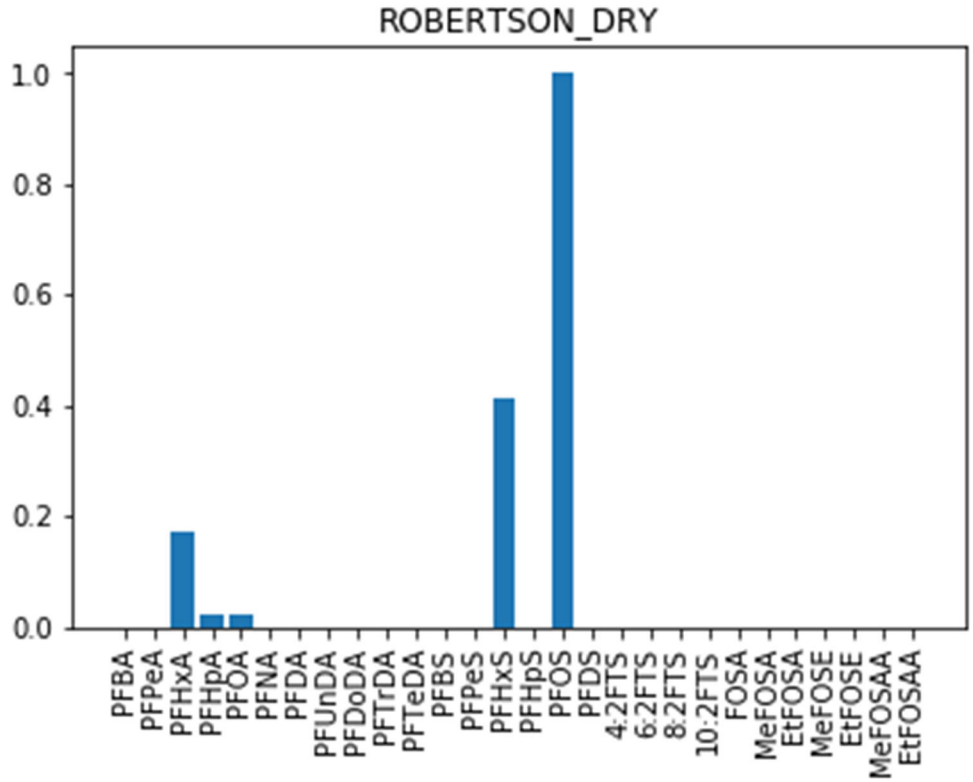

QH3_SOIL
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

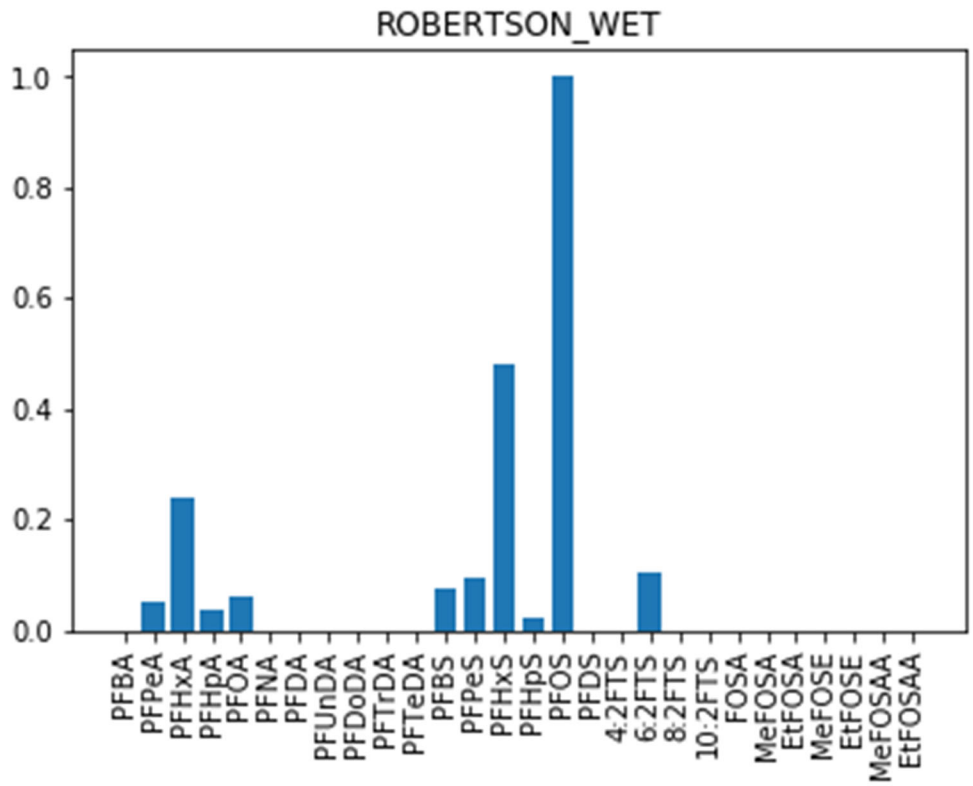

QH3_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

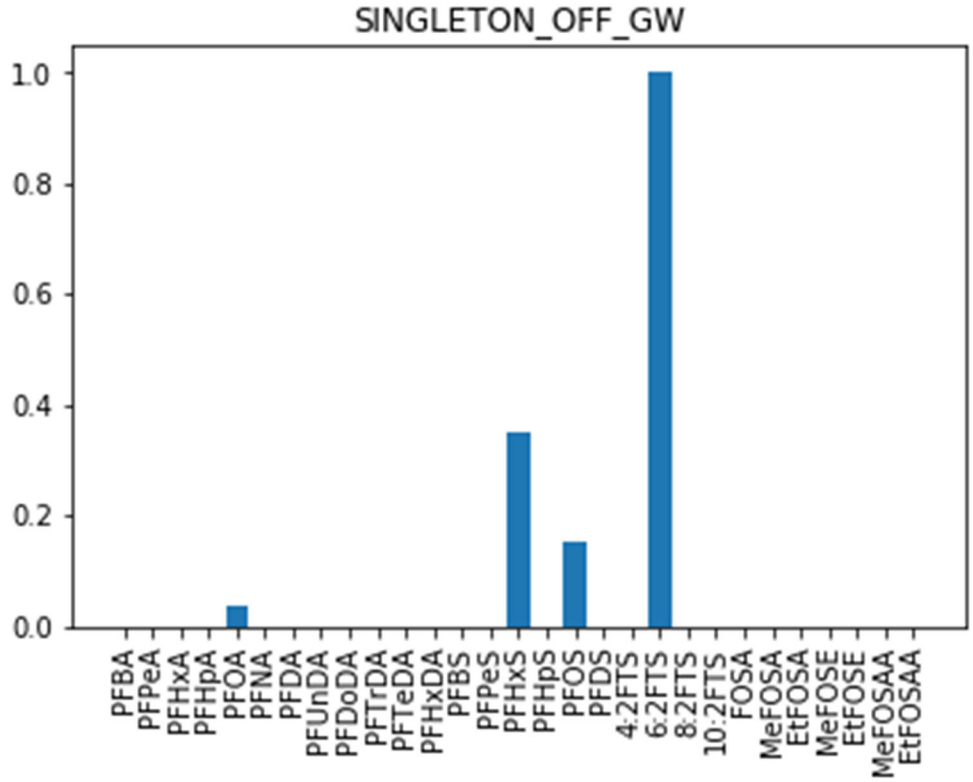

QH3_WASTE_SOIL
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

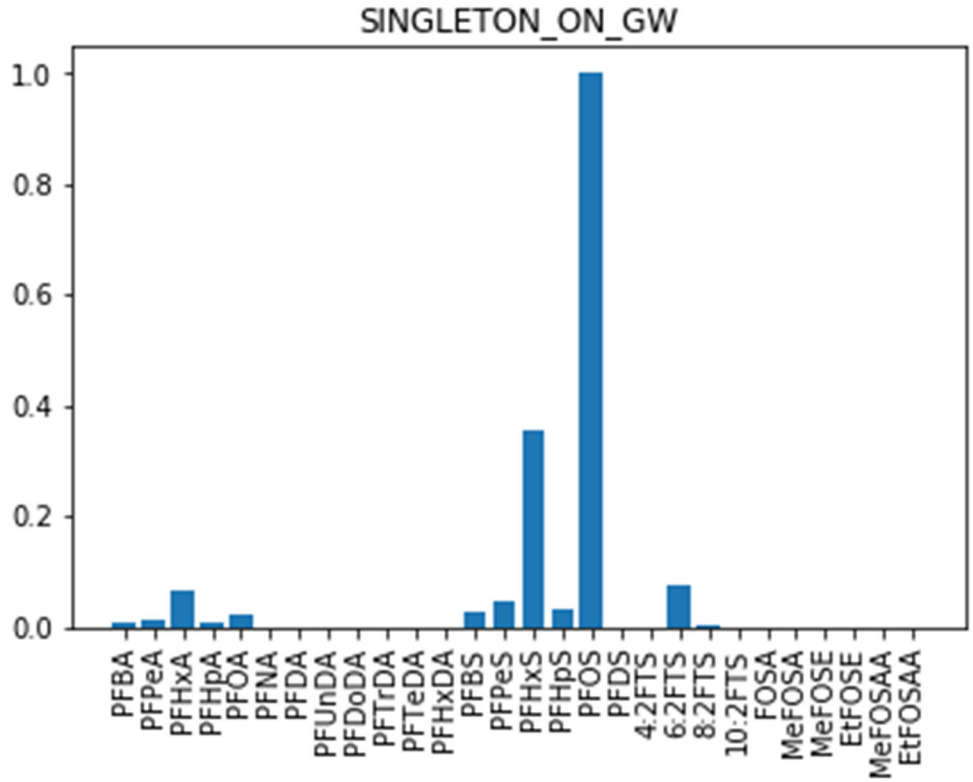

QH3_WWTP
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

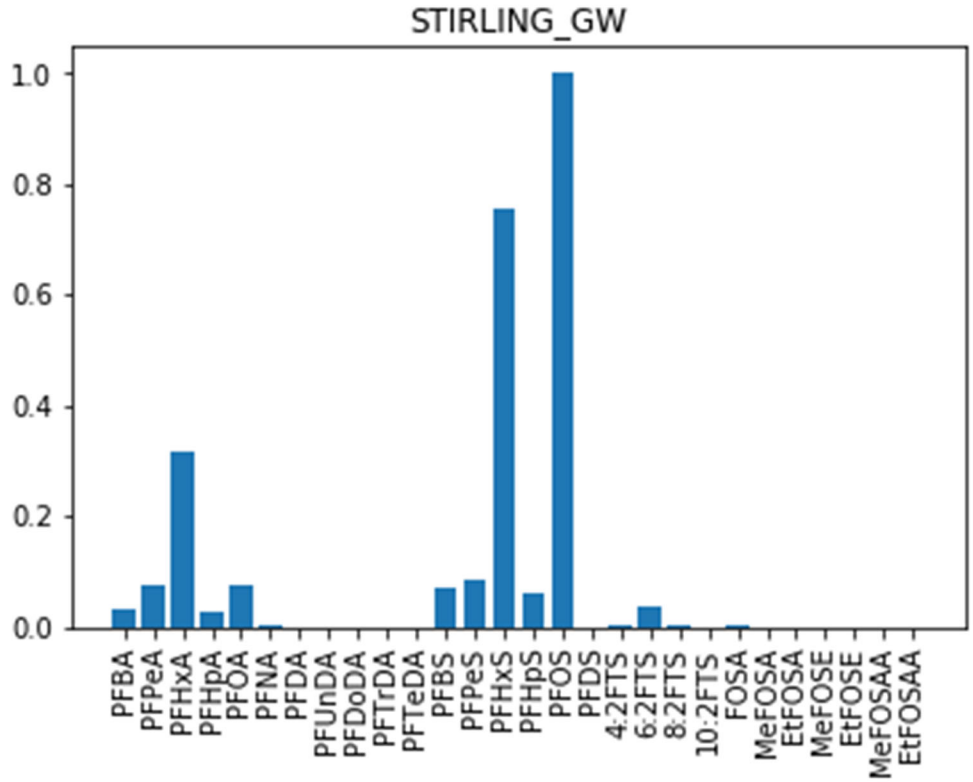

RICHMOND_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

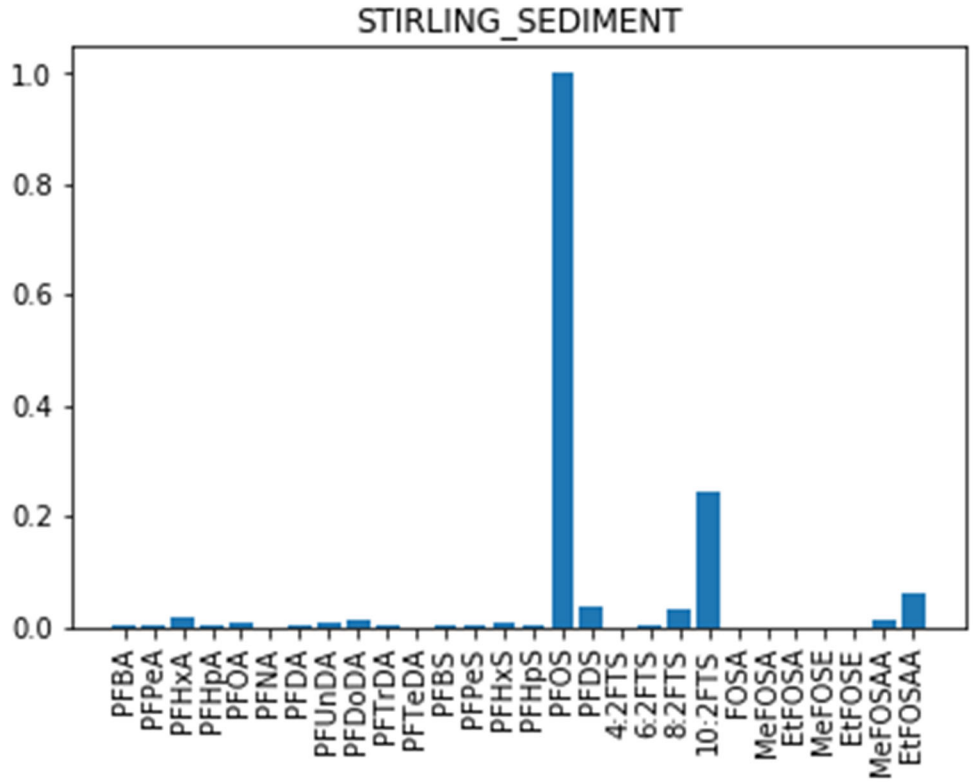

ROBERTSON_DRY
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

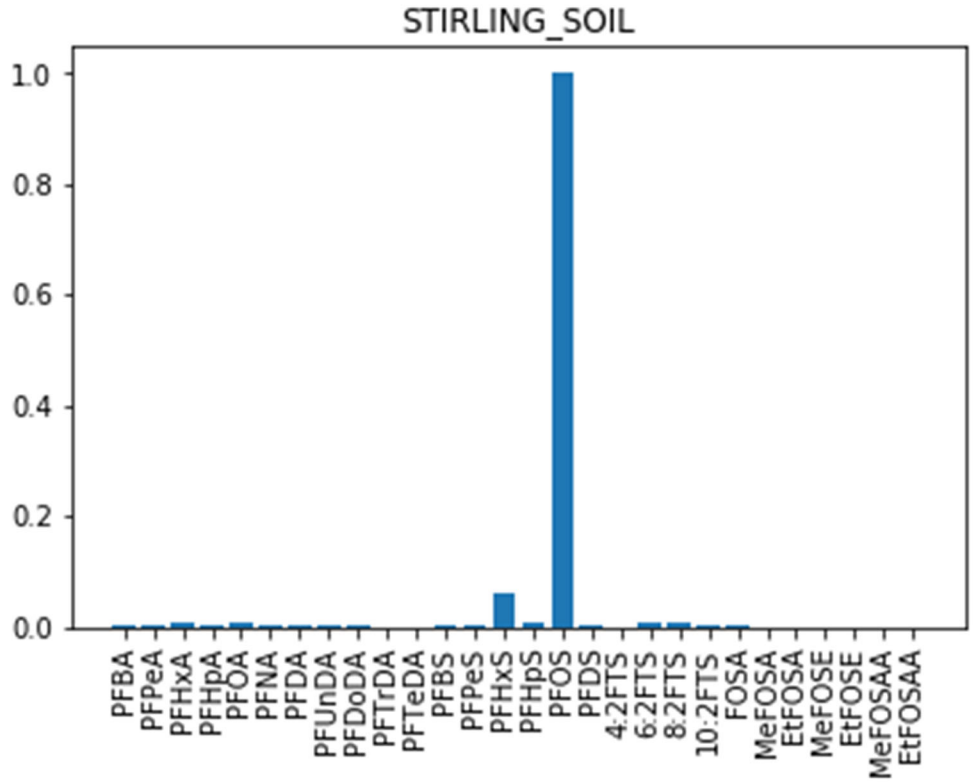

ROBERTSON_WET
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

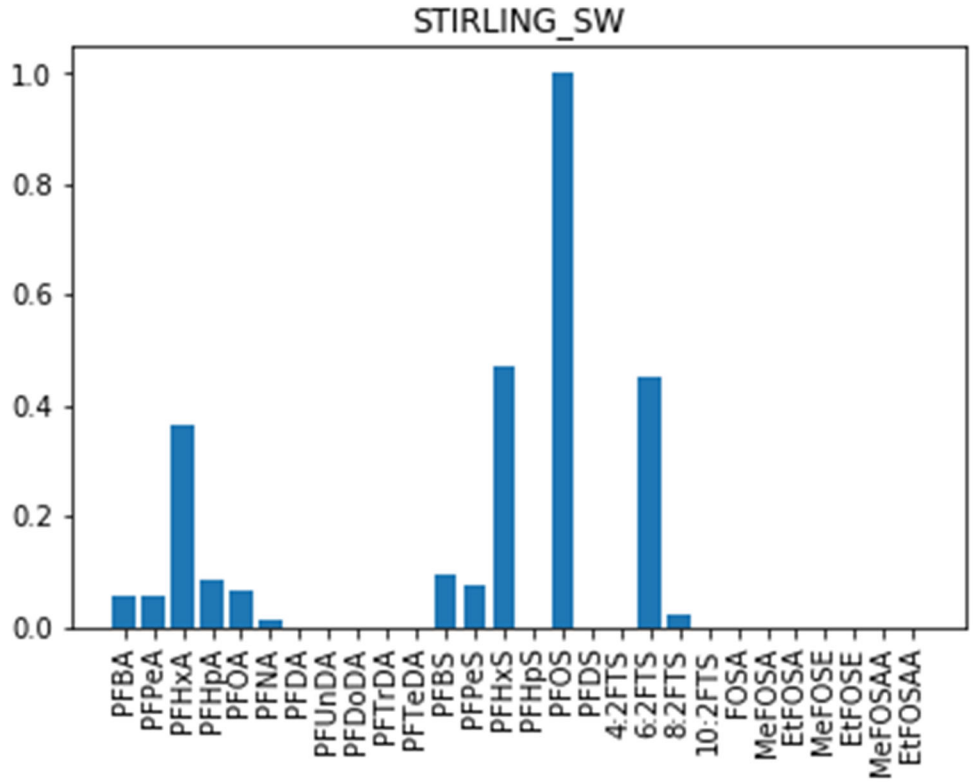

SINGLETON_OFF_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFHxDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

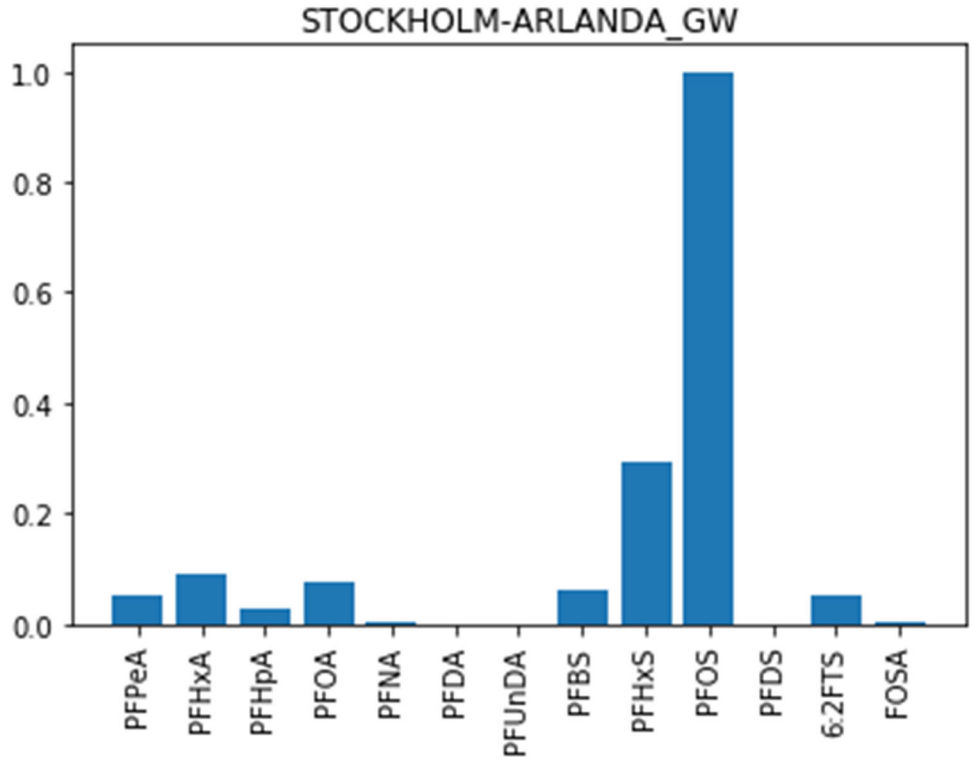

SINGLETON_ON_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFHxDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

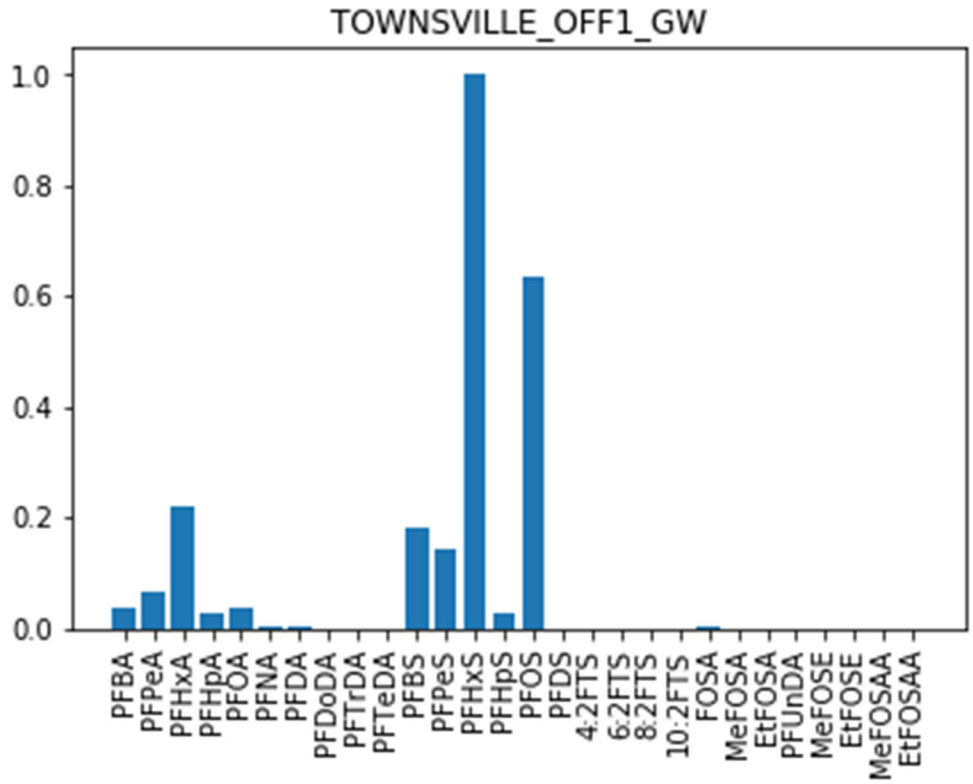

STIRLING_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

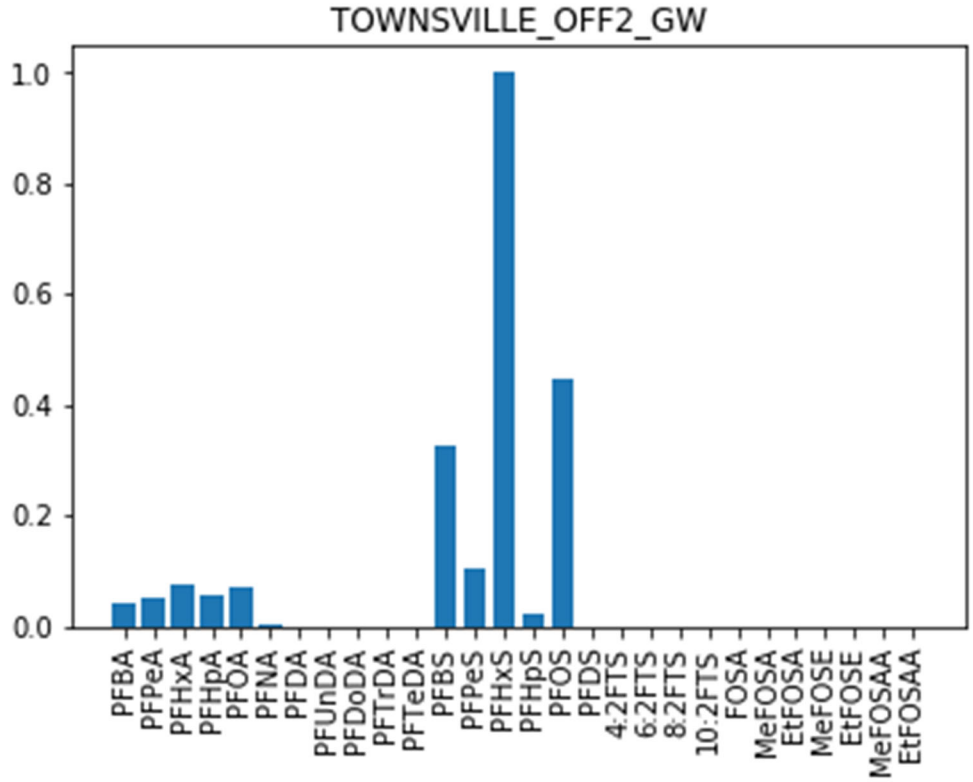

STIRLING_SEDIMENT
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

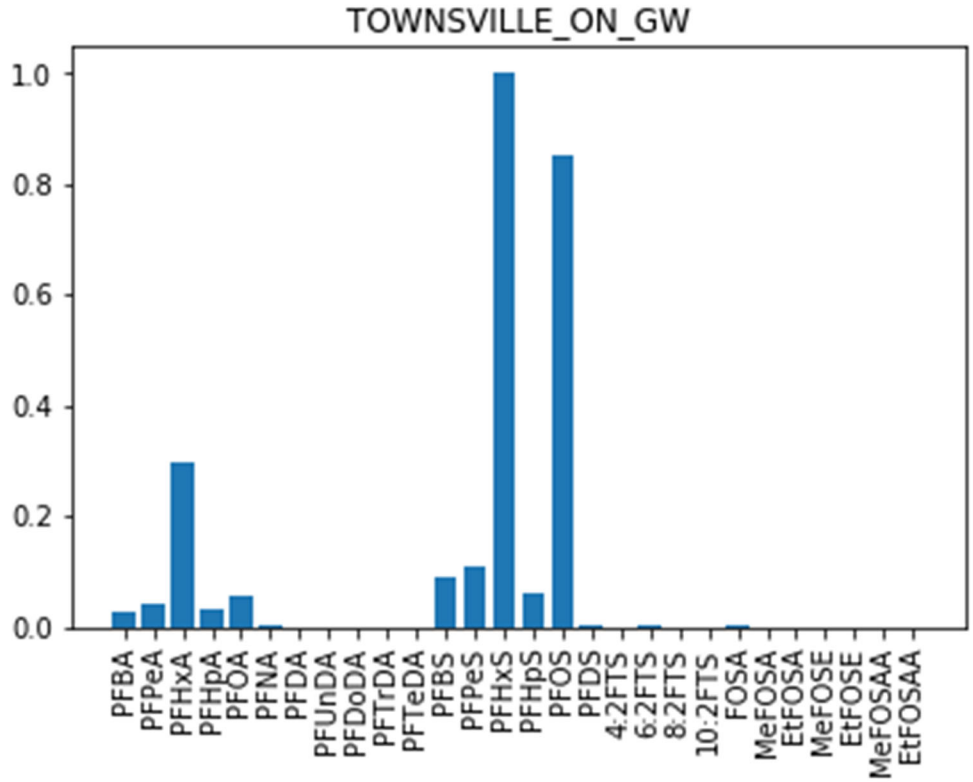

STIRLING_SOIL
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

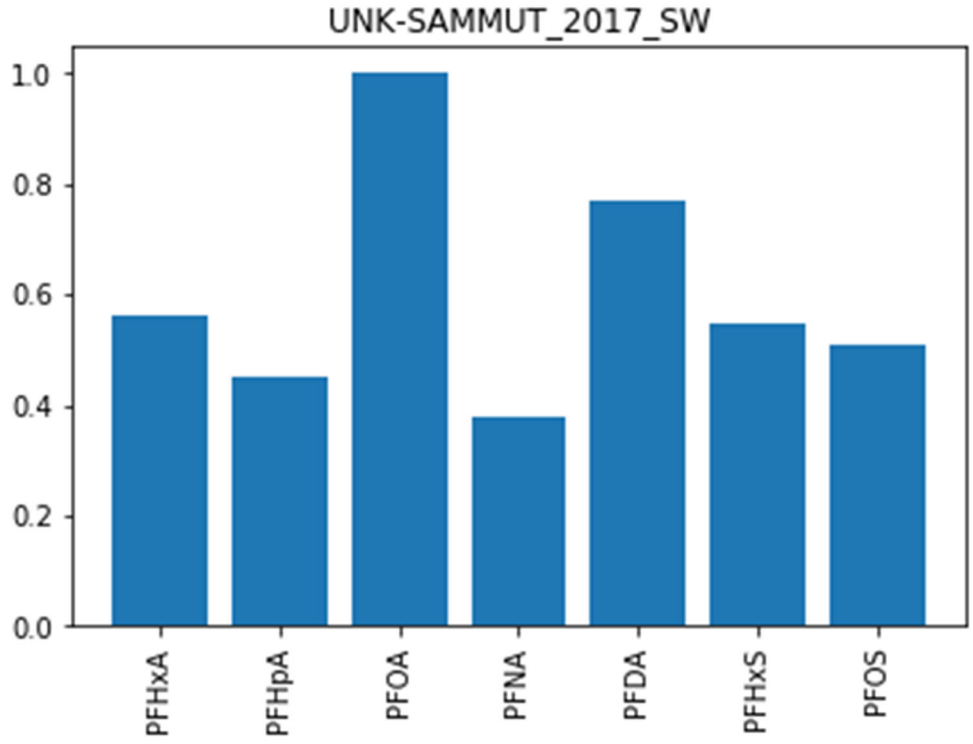

STIRLING_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

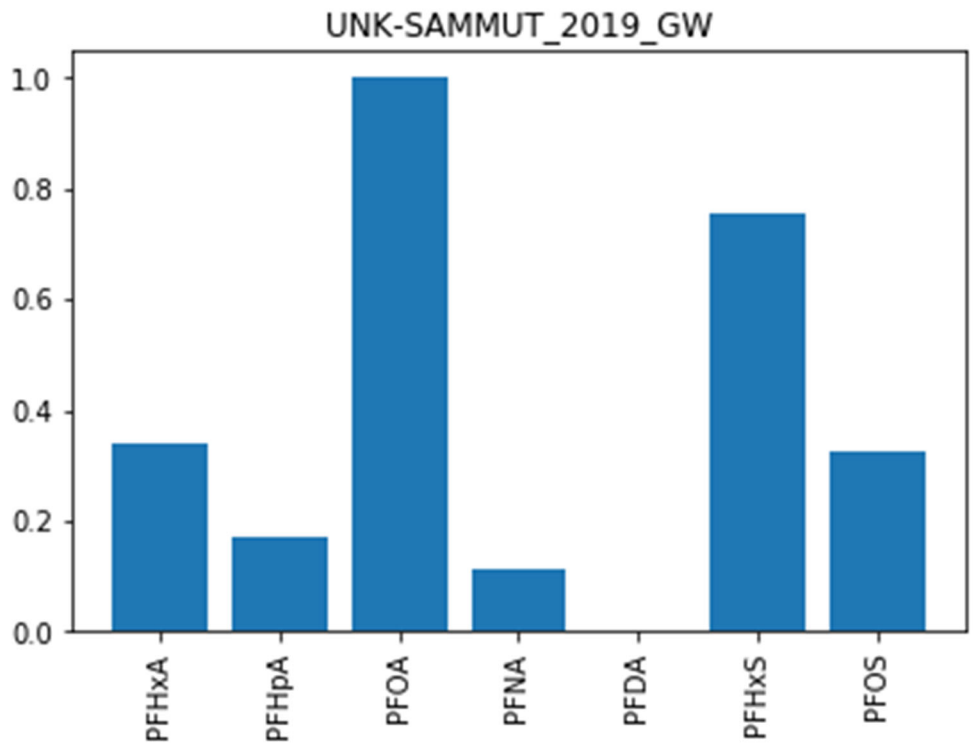

STOCKHOLM-ARLANDA_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFBS
PFHxS
PFOS
PFDS
6:2FTS
FOSA

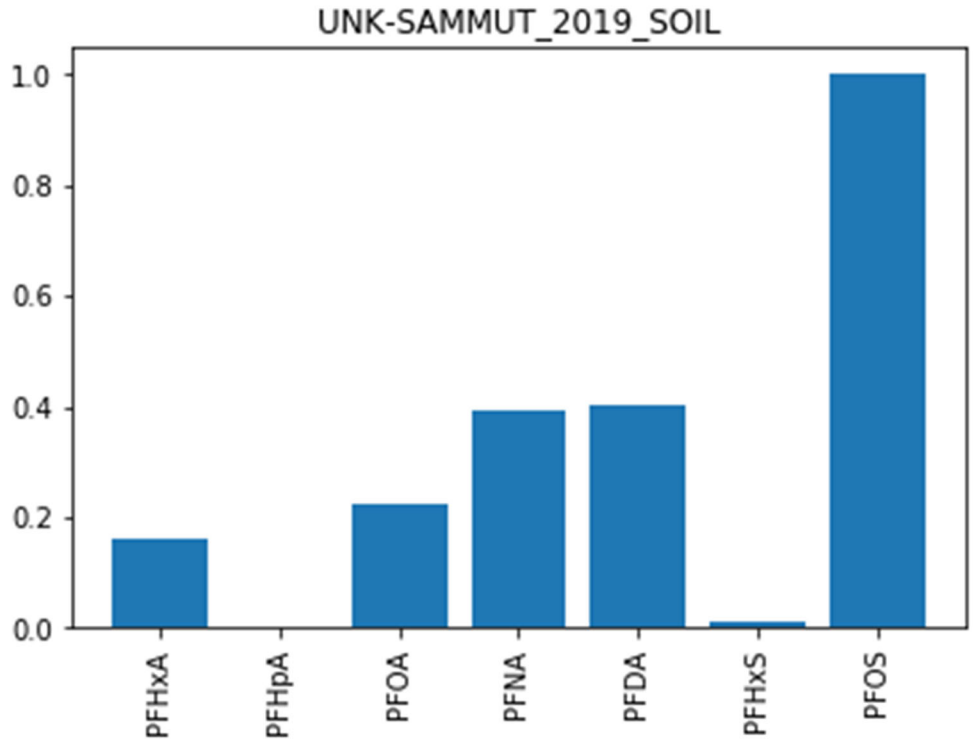

TOWNSVILLE_OFF1_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
PFUnDA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

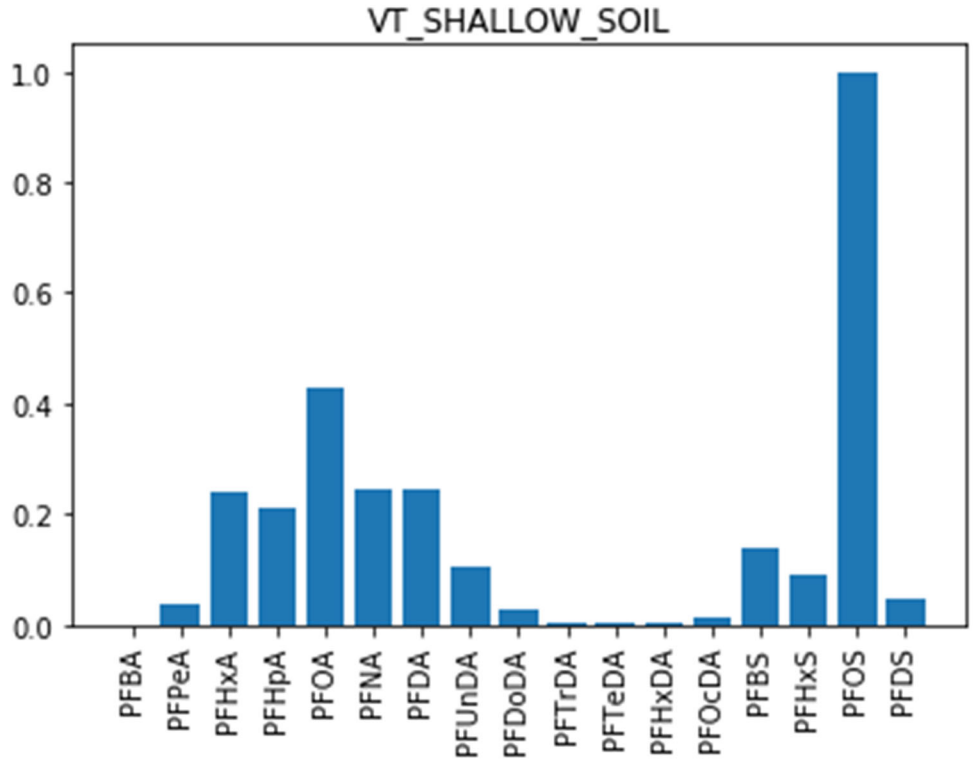

TOWNSVILLE_OFF2_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

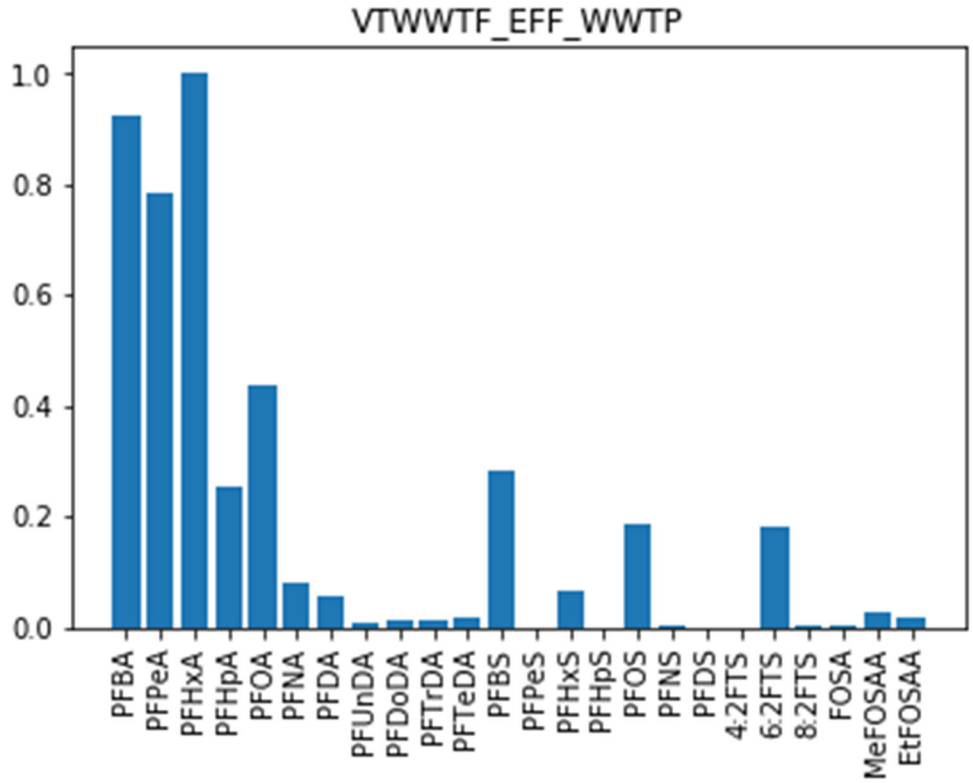

TOWNSVILLE_ON_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

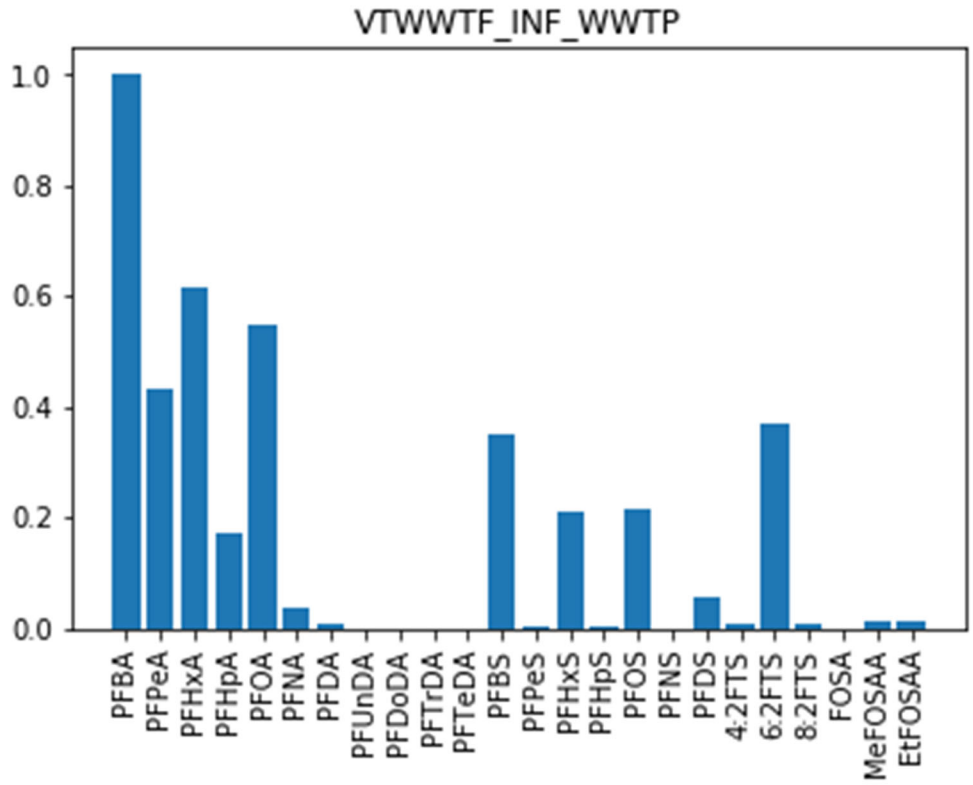

UNK-SAMMUT_2017_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFHxS
PFOS

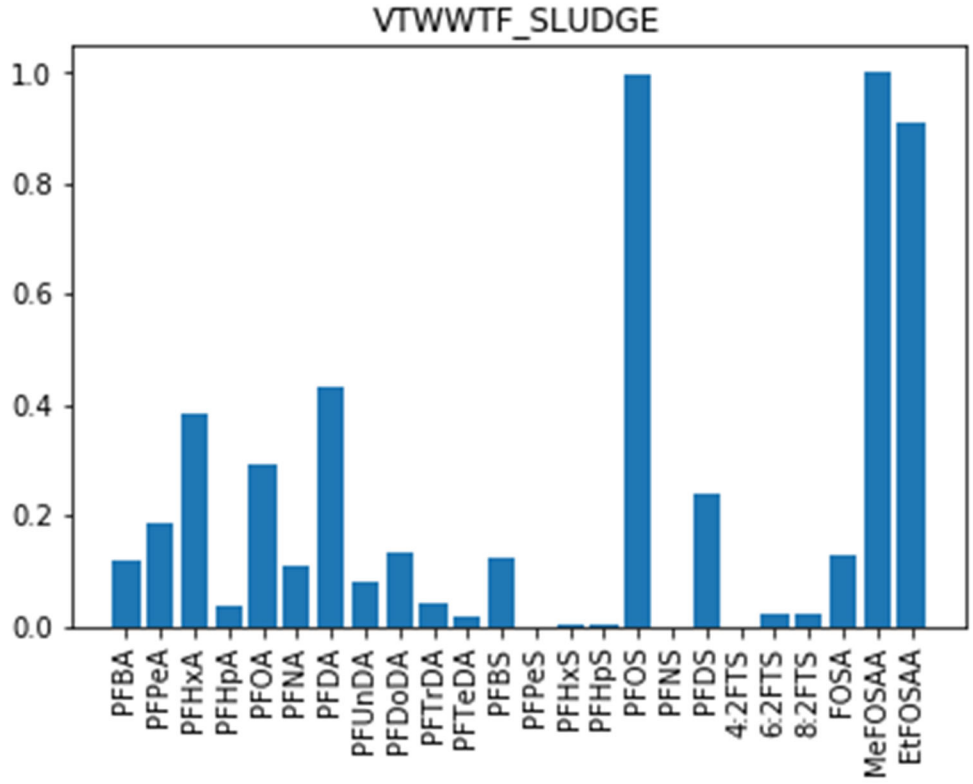

UNK-SAMMUT_2019_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFHxS
PFOS

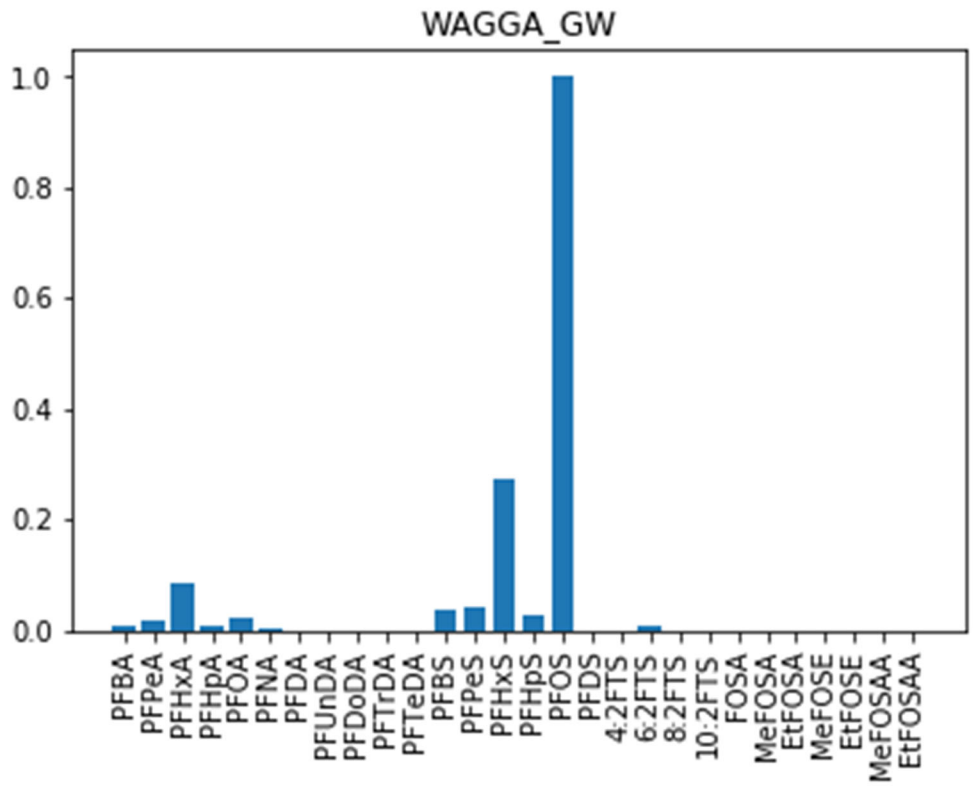

UNK-SAMMUT_2019_SOIL
1.0
0.8
0.6
0.4
0.2
0.0
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFHxS
PFOS

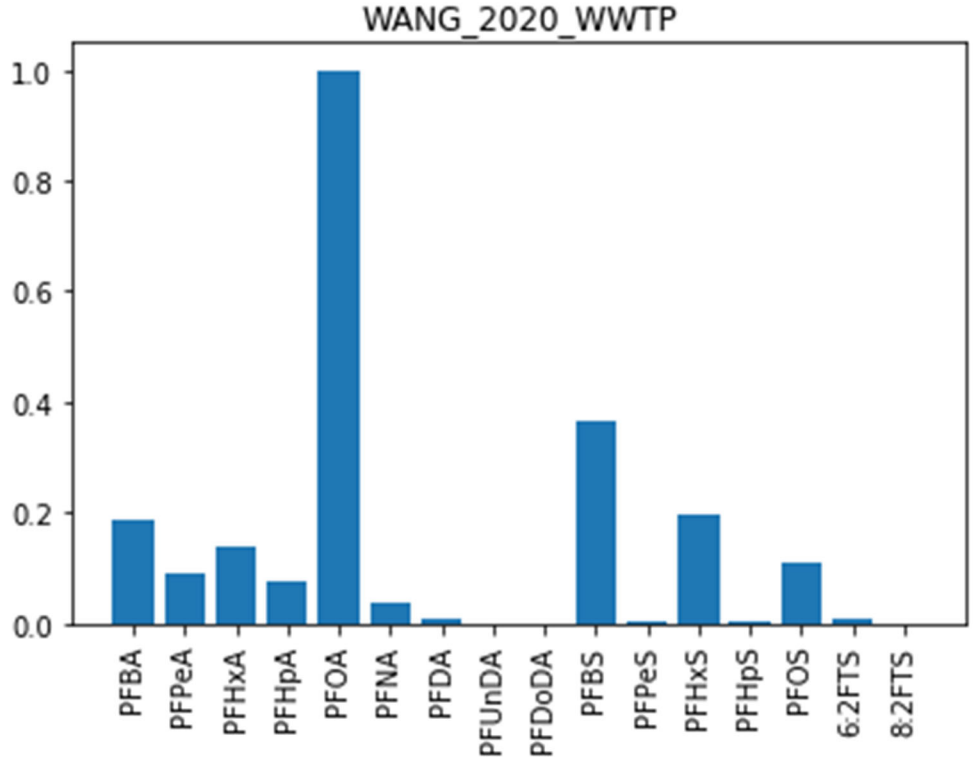

VT_SHALLOW_SOIL
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFHxDA
PFOcDA
PFBS
PFHxS
PFOS
PFDS

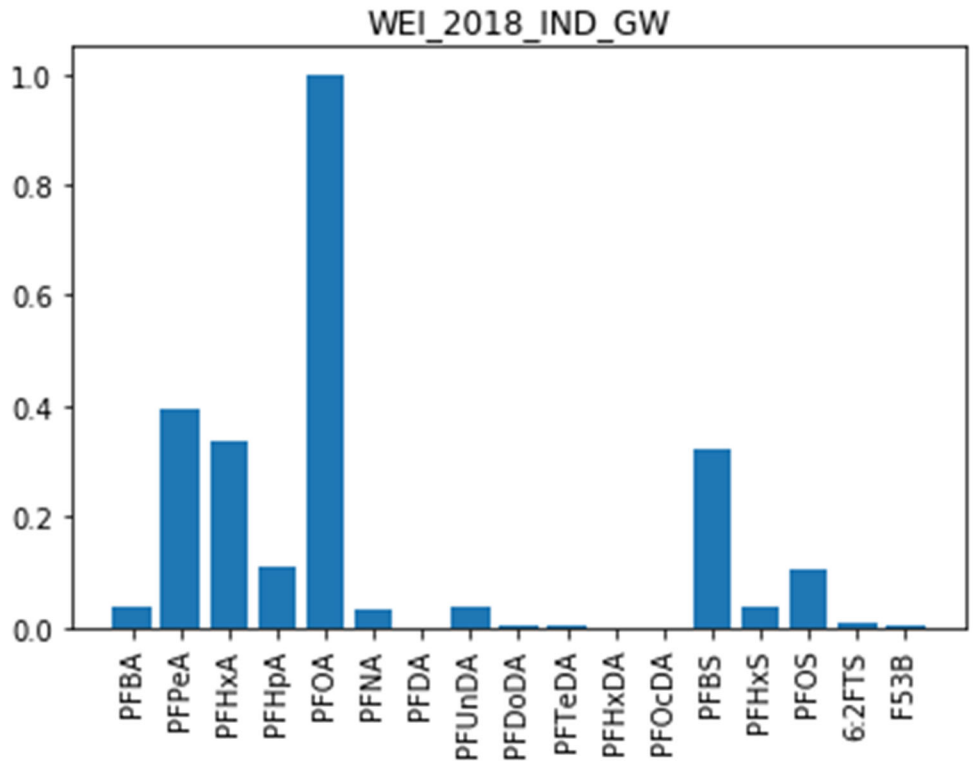

VTWWTF_EFF_WWTP
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
FOSA
MeFOSAA
EtFOSAA

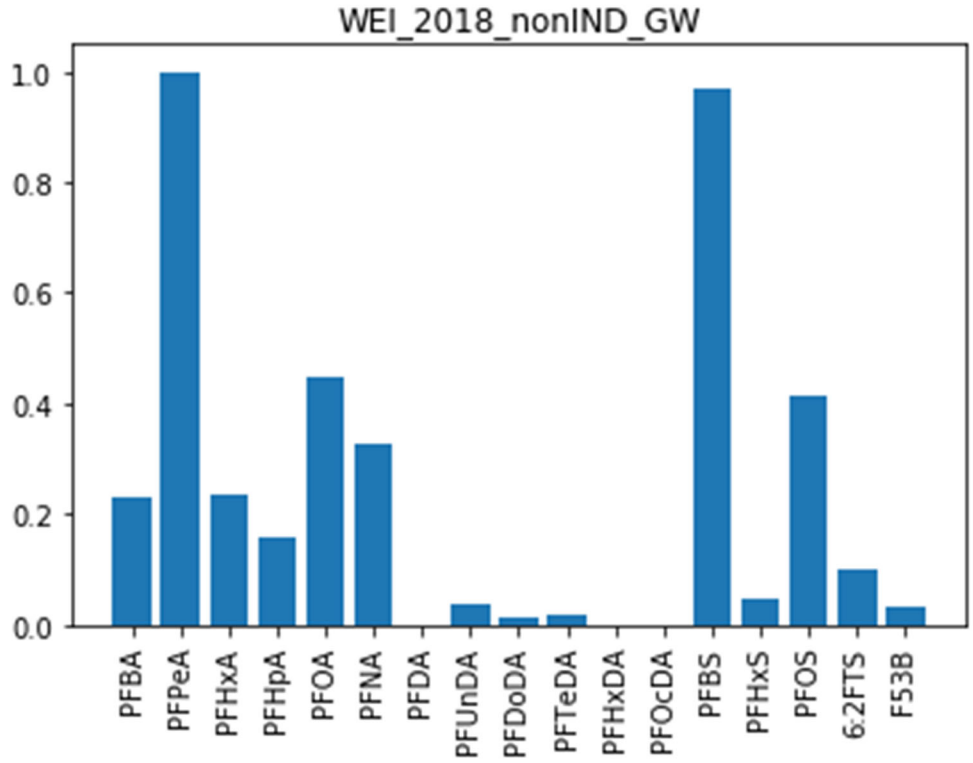

VTWWTF_INF_WWTP
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
FOSA
MeFOSAA
EtFOSAA

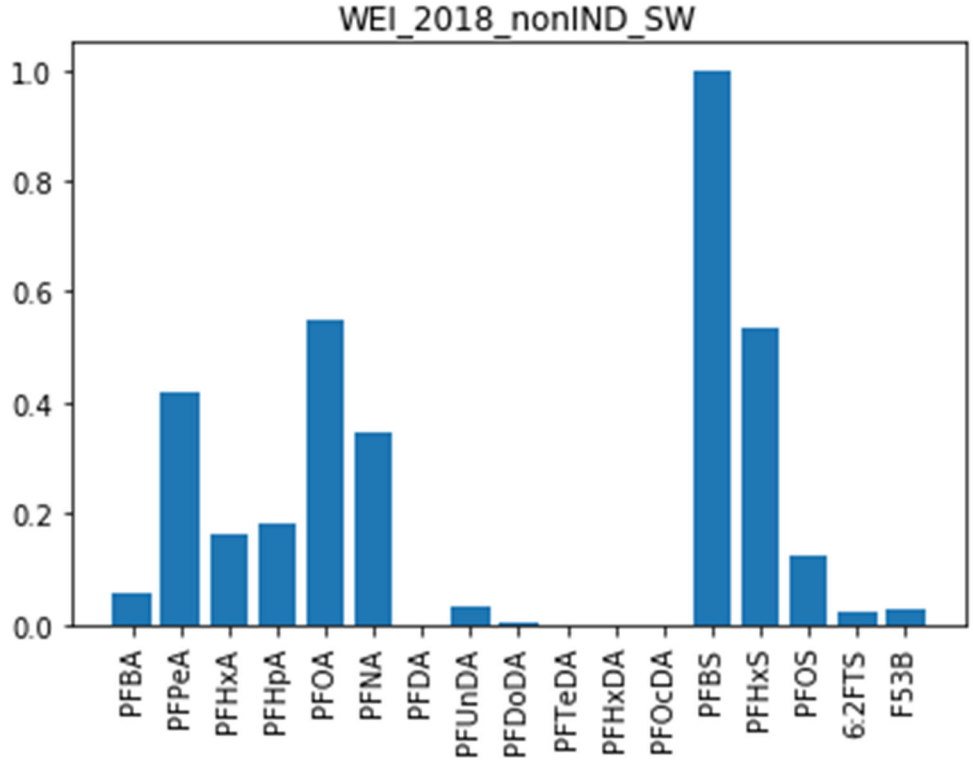

VTWWTF_SLUDGE
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
FOSA
MeFOSAA
EtFOSAA

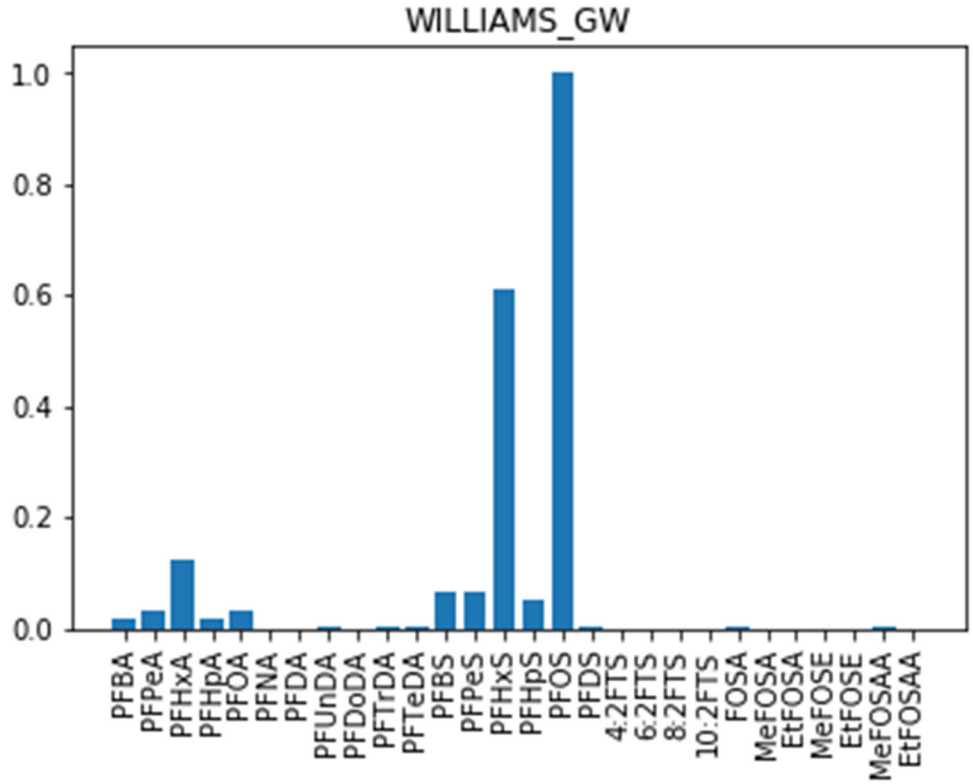

WAGGA_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

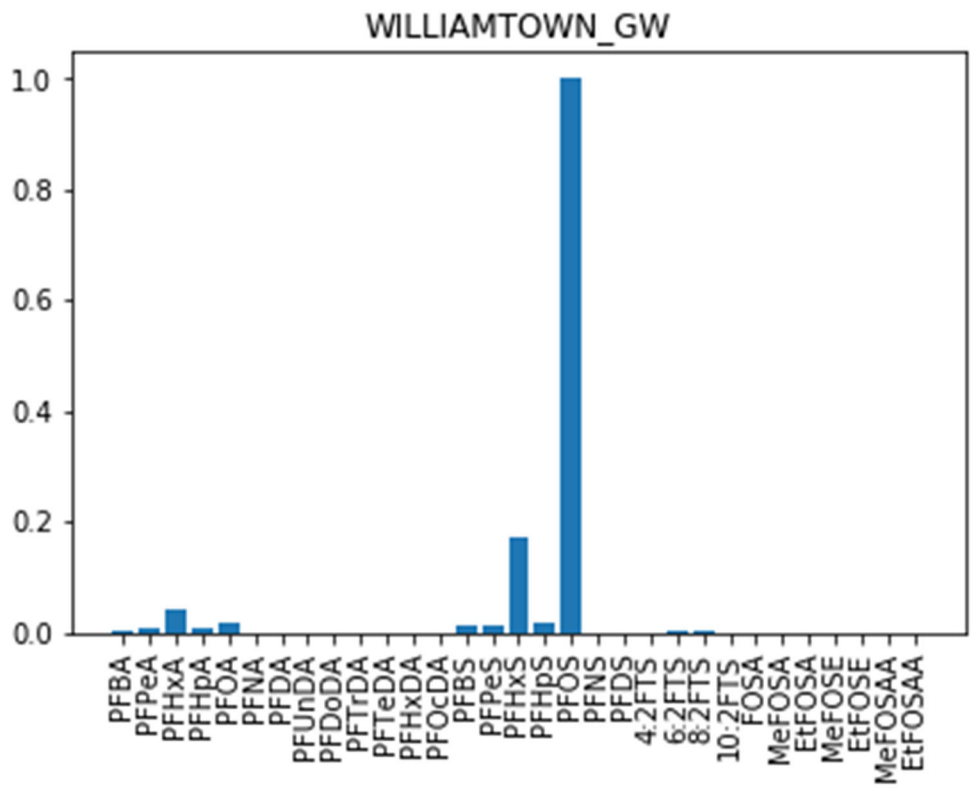

WANG_2020_WWTP
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
6:2FTS
8:2FTS

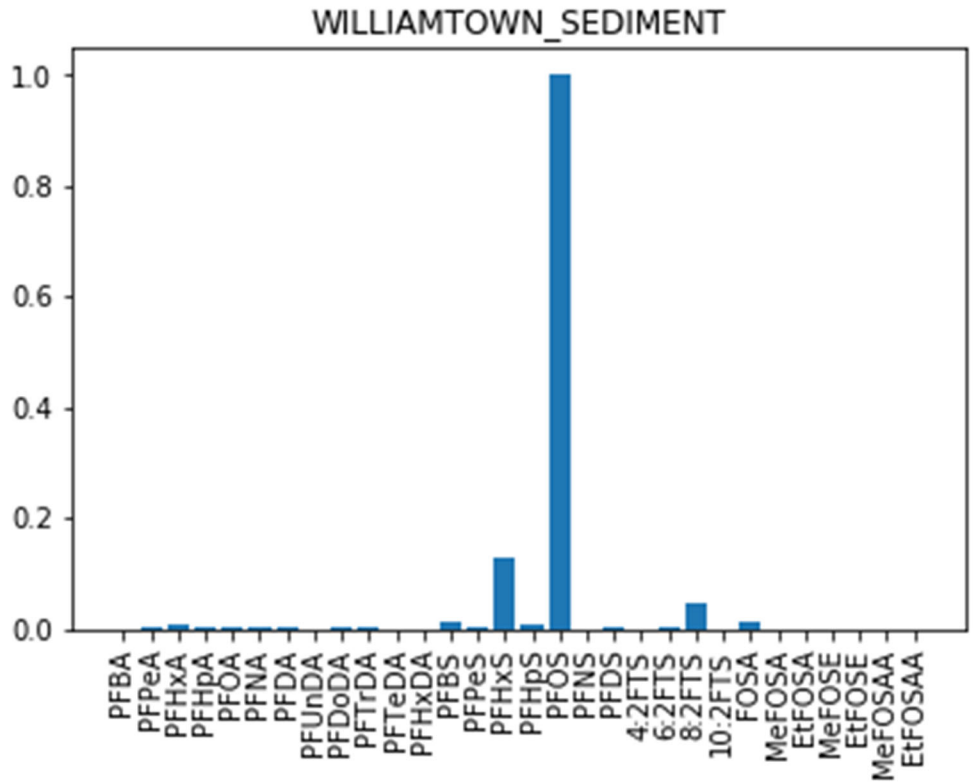

WEI_2018_IND_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTeDA
PFHxDA
PFOcDA
PFBS
PFHxS
PFOS
6:2FTS
F53B

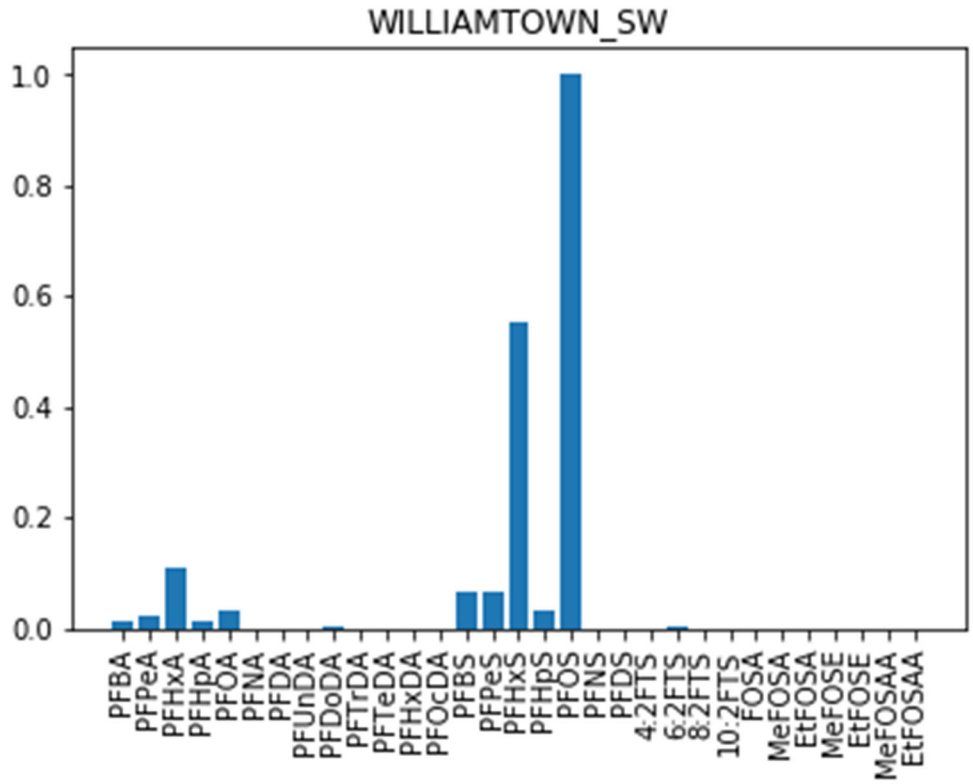

WEI_2018_nonIND_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTeDA
PFHxDA
PFOcDA
PFBS
PFHxS
PFOS
6:2FTS
F53B

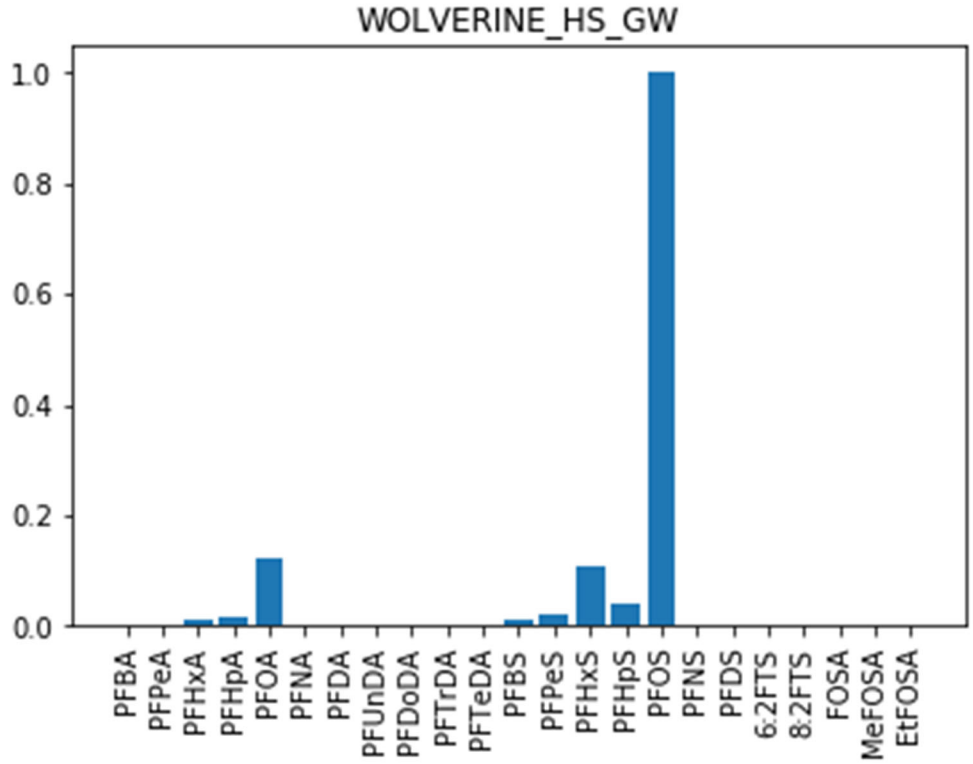

WEI_2018_nonIND_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTeDA
PFHxDA
PFOcDA
PFBS
PFHxS
PFOS
6:2FTS
F53B

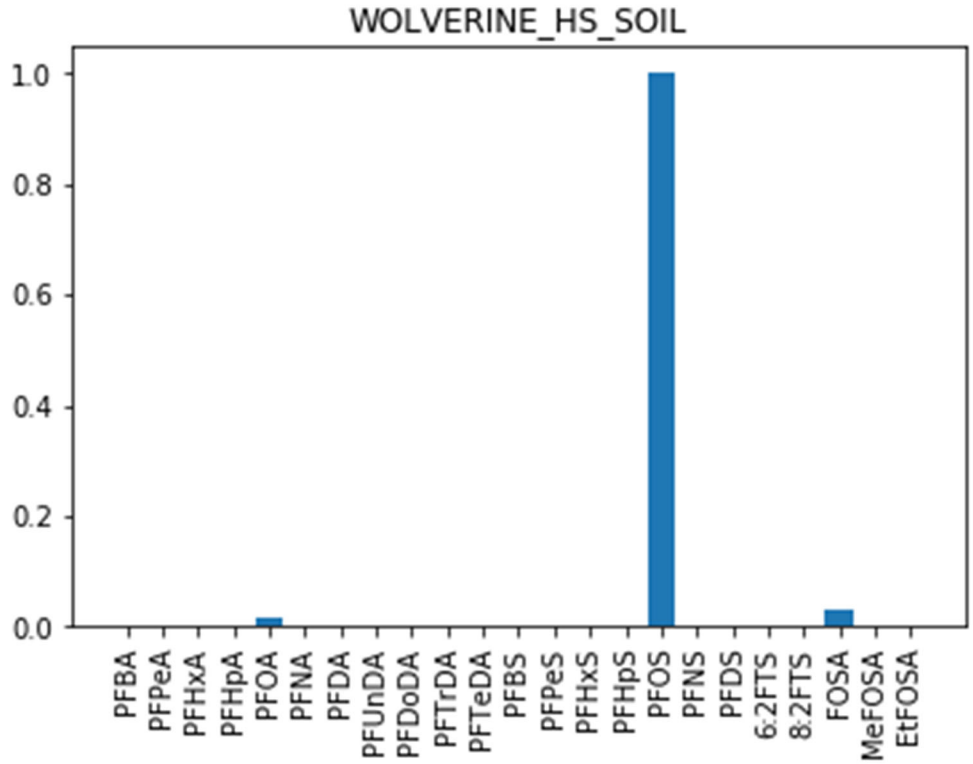

WILLIAMS_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

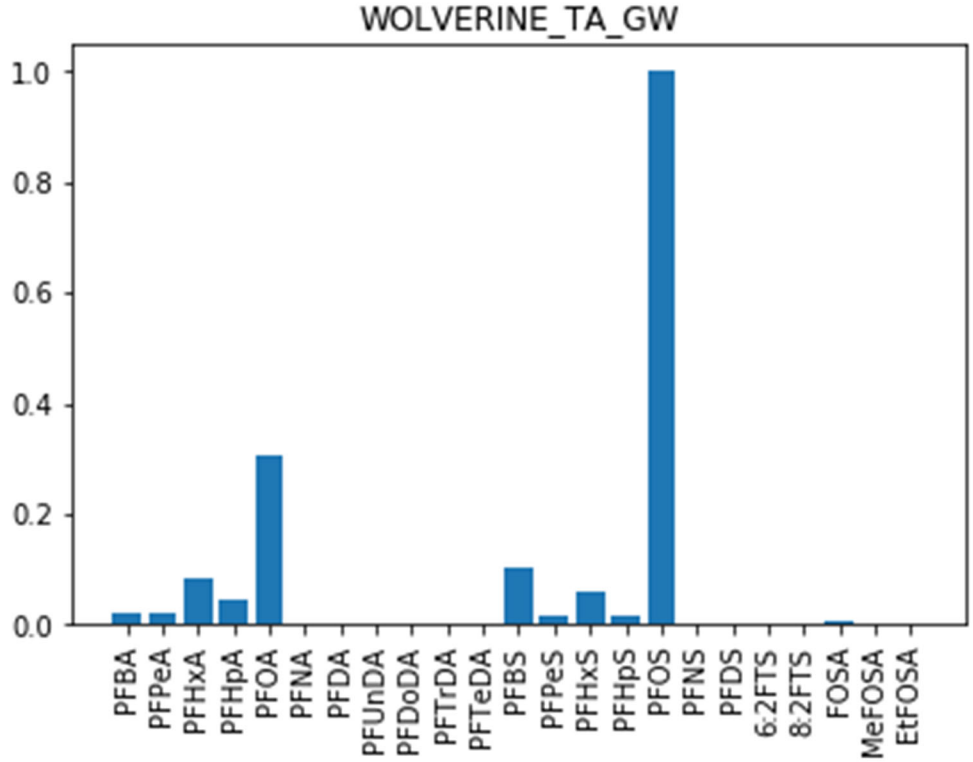

WILLIAMTOWN_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFHxDA
PFOcDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

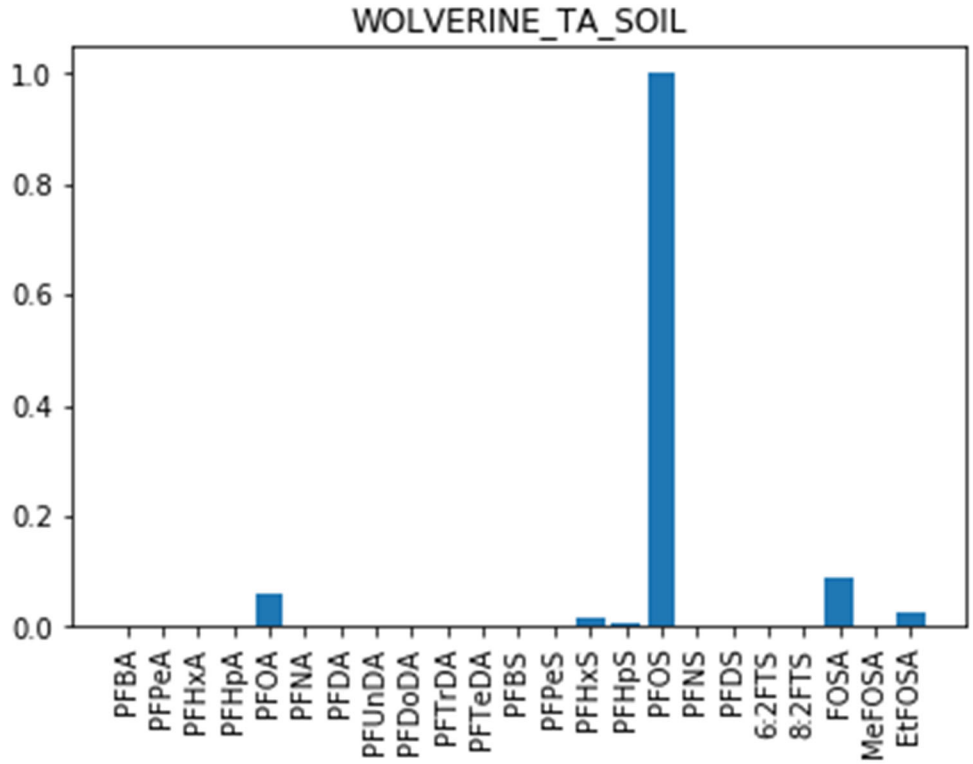

WILLIAMTOWN_SEDIMENT
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFHxDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

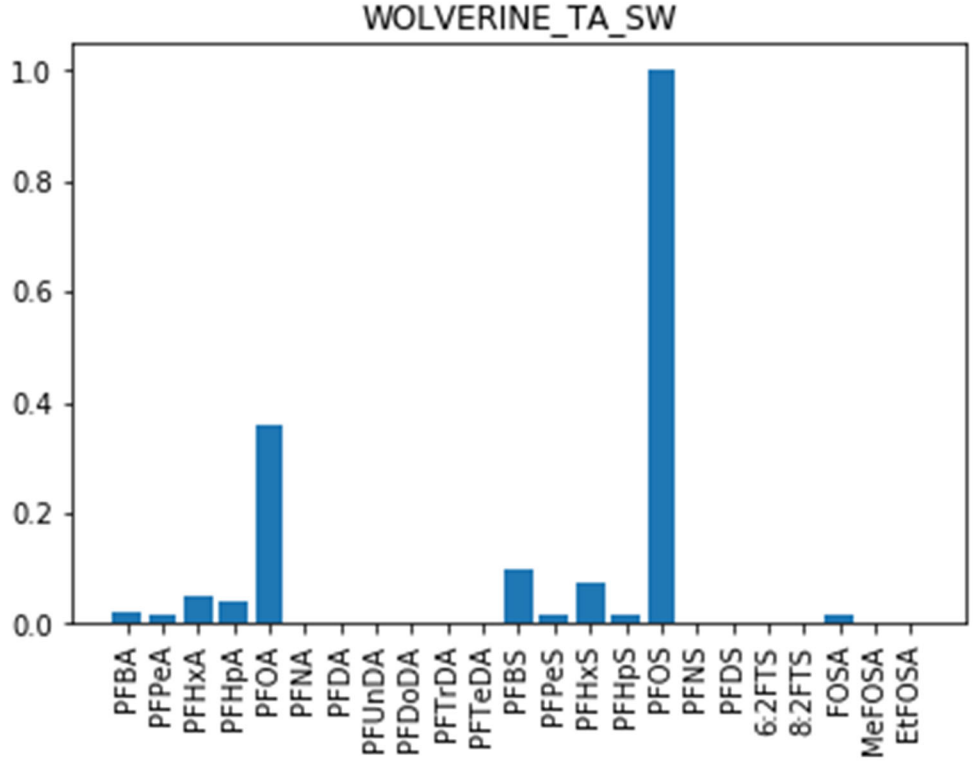

WILLIAMTOWN_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFHxDA
PFOcDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
4:2FTS
6:2FTS
8:2FTS
10:2FTS
FOSA
MeFOSA
EtFOSA
MeFOSE
EtFOSE
MeFOSAA
EtFOSAA

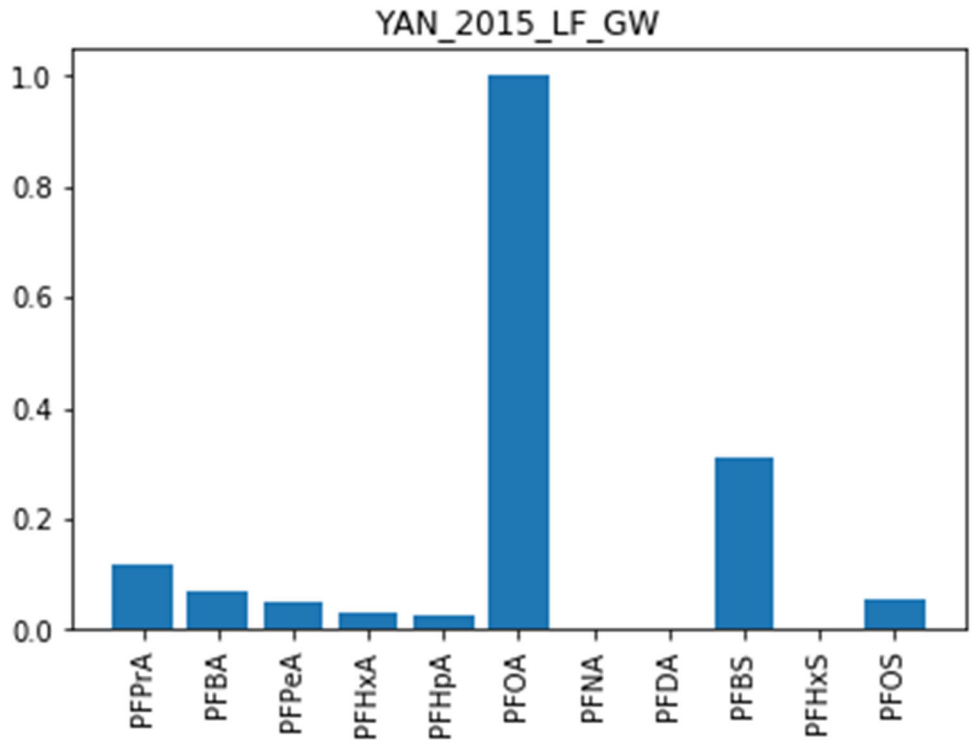

WOLVERINE_HS_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
6:2FTS
8:2FTS
FOSA
MeFOSA
EtFOSA

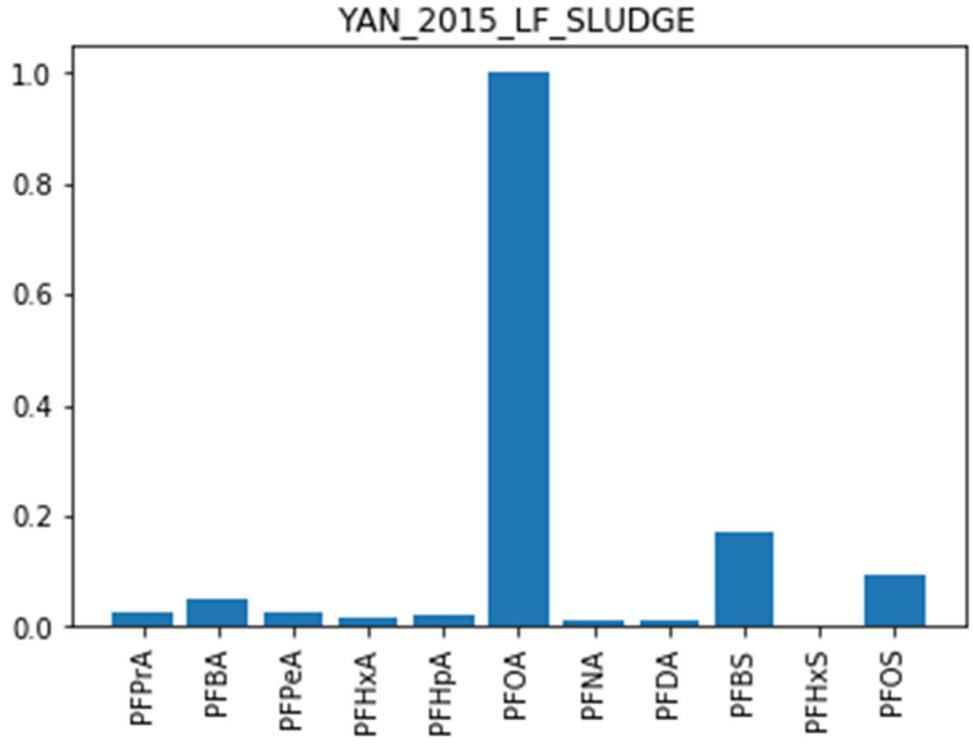

WOLVERINE_HS_SOIL
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
6:2FTS
8:2FTS
FOSA
MeFOSA
EtFOSA

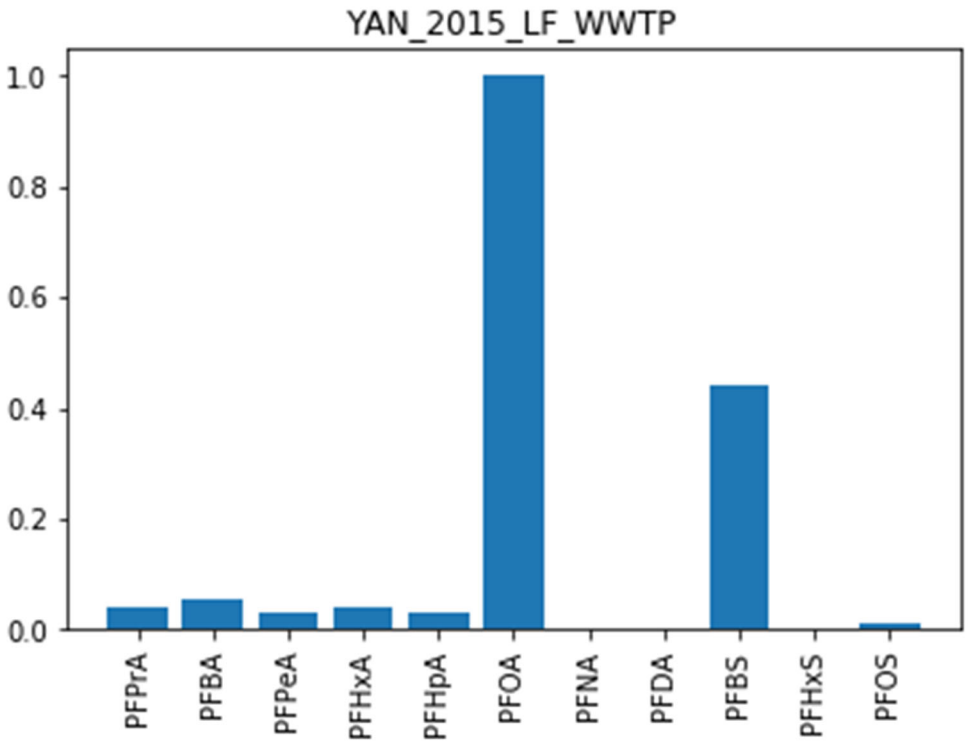

WOLVERINE_TA_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
6:2FTS
8:2FTS
FOSA
MeFOSA
EtFOSA


WOLVERINE_TA_SOIL
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
6:2FTS
8:2FTS
FOSA
MeFOSA
EtFOSA


WOLVERINE_TA_SW
1.0
0.8
0.6
0.4
0.2
0.0
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFUnDA
PFDoDA
PFTrDA
PFTeDA
PFBS
PFPeS
PFHxS
PFHpS
PFOS
PFNS
PFDS
6:2FTS
8:2FTS
FOSA
MeFOSA
EtFOSA


YAN_2015_LF_GW
1.0
0.8
0.6
0.4
0.2
0.0
PFPrA
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFBS
PFHxS
PFOS


YAN_2015_LF_SLUDGE
1.0
0.8
0.6
0.4
0.2
0.0
PFPrA
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFBS
PFHxS
PFOS


YAN_2015_LF_WWTP
1.0
0.8
0.6
0.4
0.2
0.0
PFPrA
PFBA
PFPeA
PFHxA
PFHpA
PFOA
PFNA
PFDA
PFBS
PFHxS
PFOS