OPAL (Observational PFAS Access paneL)

Initial Data File Processing Documentation

OPAL (Observational PFAS Access paneL)

Initial Data File Processing Documentation

Initial Data File Processing, aka Data Munging

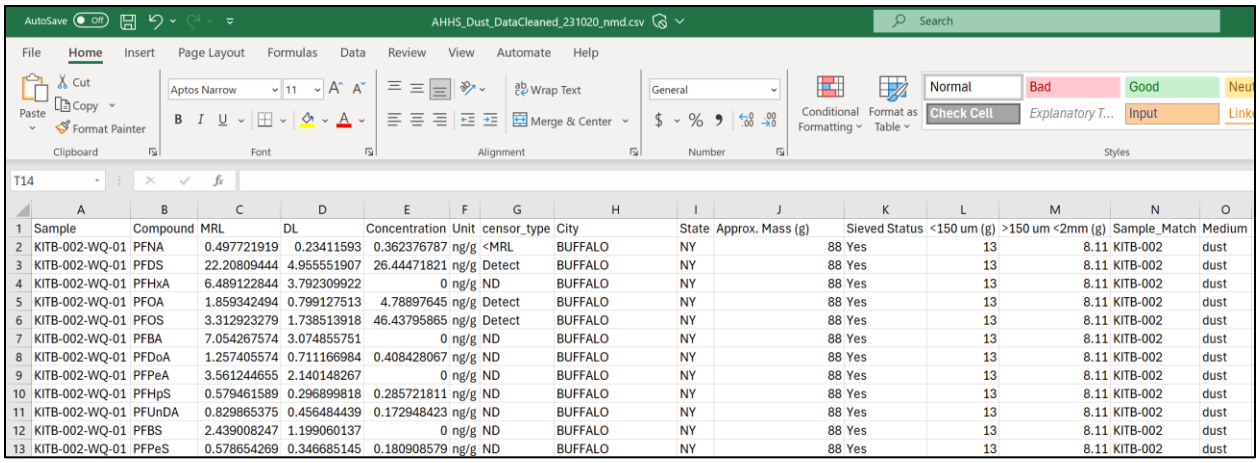
The data processing detailed in this document describes the steps taken to process the original input data files, the initial data files received from EPA, to prepare them for ingest into the PFAS Observations Data Model.

The basic set of steps includes:

1. Transposing the data, i.e. columns to rows
2. Update the attribute names, i.e. column headers, to pre-pend the relevant compound abbreviation. Ex. Update from "Flag" to "PFNA_Flags".

AHHS Dust

Orig file from EPA: AHHS_Dust_DataCleaned_231020.nmd.csv

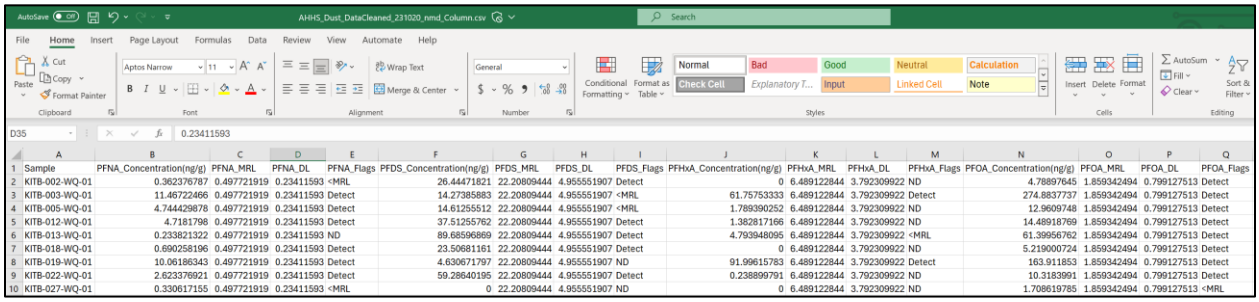


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Sample	Compound	MRL	DL	Concentration	Unit	censor_type	City	State	Approx. Mass (g)	Sieved Status	<150 um (g)	>150 um <2mm (g)	Sample_Match	Medium
2	KITB-002-WQ-01	PFNA	0.497721919	0.23411593	0.362376787	ng/g	<MRL	BUFFALO	NY	88	Yes	13	8.11	KITB-002	dust
3	KITB-002-WQ-01	PFDS	22.20809444	4.955551907	26.44471821	ng/g	Detect	BUFFALO	NY	88	Yes	13	8.11	KITB-002	dust
4	KITB-002-WQ-01	PFHxA	6.489122844	3.792309922	0	ng/g	ND	BUFFALO	NY	88	Yes	13	8.11	KITB-002	dust
5	KITB-002-WQ-01	PFOA	1.859342494	0.799127513	4.78897645	ng/g	Detect	BUFFALO	NY	88	Yes	13	8.11	KITB-002	dust
6	KITB-002-WQ-01	PFOS	3.312923279	1.738513918	46.43795865	ng/g	Detect	BUFFALO	NY	88	Yes	13	8.11	KITB-002	dust
7	KITB-002-WQ-01	PFBA	7.054267574	3.074855751	0	ng/g	ND	BUFFALO	NY	88	Yes	13	8.11	KITB-002	dust
8	KITB-002-WQ-01	PFDoA	1.257405574	0.711166984	0.408428067	ng/g	ND	BUFFALO	NY	88	Yes	13	8.11	KITB-002	dust
9	KITB-002-WQ-01	PFPeA	3.561244655	2.140148267	0	ng/g	ND	BUFFALO	NY	88	Yes	13	8.11	KITB-002	dust
10	KITB-002-WQ-01	PFHpS	0.579461589	0.296899818	0.285721811	ng/g	ND	BUFFALO	NY	88	Yes	13	8.11	KITB-002	dust
11	KITB-002-WQ-01	PFUnDA	0.829865375	0.456484439	0.172948423	ng/g	ND	BUFFALO	NY	88	Yes	13	8.11	KITB-002	dust
12	KITB-002-WQ-01	PFBS	2.439008247	1.199060137	0	ng/g	ND	BUFFALO	NY	88	Yes	13	8.11	KITB-002	dust
13	KITB-002-WQ-01	PFPeS	0.578654269	0.346685145	0.180908579	ng/g	ND	BUFFALO	NY	88	Yes	13	8.11	KITB-002	dust

R code (PFAS_Data_Processing.R) to process the original file: 1) Transposed data 2) For each compound abbreviation, add concentration, MRL, DL, and Flag as suffix

```
###AHHS_Dust_Cleaned_Data###
data <- read.csv("C:/Publications/PFAS database/Data Files Working Copies/AHHS_Dust_DataCleaned_231020_nmd.csv", header = TRUE) #head(data)
length(unique(data$Sample))
length(unique(data$Compound))

Tdata <- matrix(data = NA, nrow = length(unique(data$Sample)), ncol = 1+4*length(unique(data$Compound)), byrow =FALSE, dimnames =NULL) #head(Tdata)
for (i in 1:length(unique(data$Sample))) { # i=2
  datasub <- data[data$Sample == unique(data$Sample)[i],]
  for (j in 1:length(unique(data$Compound))) { #j=2
    Tdata[i,1] <- unique(data$Sample)[i]
    Tdata[i,4*j-2] <- datasub[datasub$Compound == unique(data$Compound)[j],5]
    Tdata[i,4*j-1] <- datasub[datasub$Compound == unique(data$Compound)[j],3]
    Tdata[i,4*j] <- datasub[datasub$Compound == unique(data$Compound)[j],4]
    Tdata[i,4*j+1] <- datasub[datasub$Compound == unique(data$Compound)[j],7]
  }
}

colnames(Tdata) <- c("Sample","PFNA_Concentration(ng/g)","PFNA_MRL","PFNA_DL","PFNA_Flags","PFDS_Concentration(ng/g)","PFDS_MRL","PFDS_DL","PFDS_Flags","
                     ,"PFOS_Concentration(ng/g)","PFOS_MRL","PFOS_DL","PFOS_Flags","PFBA_Concentration(ng/g)","PFBA_MRL","PFBA_DL","PFBA_Flags","PFDoA_Co
                     ,"PFHpS_Concentration(ng/g)","PFHpS_MRL","PFHpS_DL","PFHpS_Flags","PFUnDA_Concentration(ng/g)","PFUnDA_MRL","PFUnDA_DL","PFUnDA_Flags
                     ,"PFNS_Concentration(ng/g)","PFNS_MRL","PFNS_DL","PFNS_Flags","PFHpA_Concentration(ng/g)","PFHpA_MRL","PFHpA_DL","PFHpA_Flags","PFHx
Tdata <- as.data.frame(Tdata)
write.csv(Tdata,"C:/Publications/PFAS database/Data Files Working Copies/Processed_data/AHHS_Dust_DataCleaned_231020_nmd_Column.csv", row.names=FALSE)
```

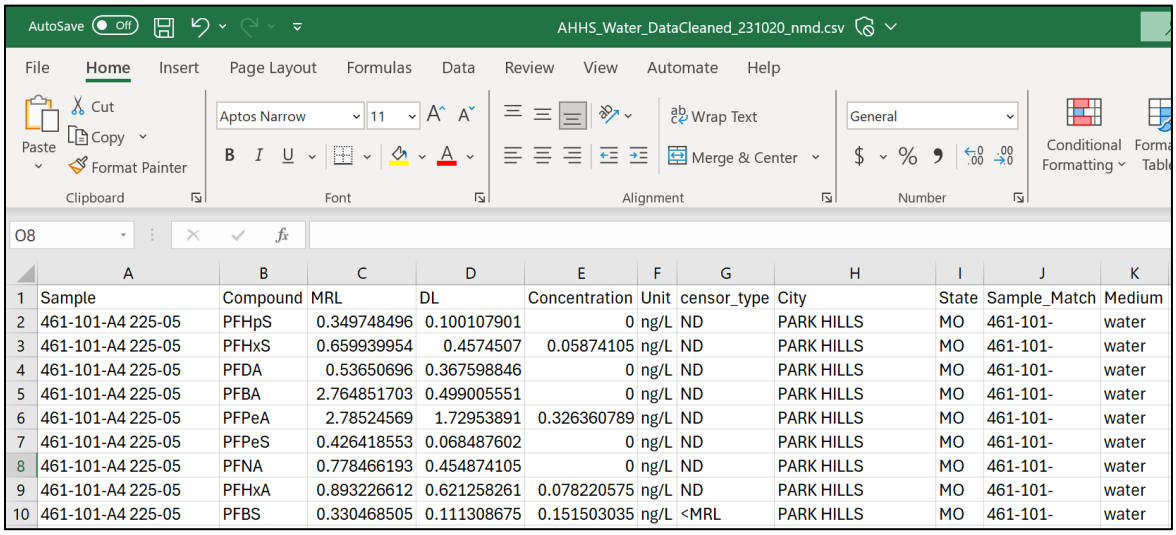
Bin processed file (16 PFAS compounds):
AHHS_Dust_DataCleaned_231020_nmd_Column.csv



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Sample	PFNA_Concentration(ng/g)	PFNA_MRL	PFNA_DL	PFNA_Flags	PFDS_Concentration(ng/g)	PFDS_MRL	PFDS_DL	PFDS_Flags	PFHxA_Concentration(ng/g)	PFHxA_MRL	PFHxA_DL	PFHxA_Flags	PFOA_Concentration(ng/g)	PFOA_MRL	PFOA_DL	PFOA_Flags
2	KITB-002-WQ-01	0.362376787	0.497721919	0.23411593	<MRL	26.44471821	22.20809444	4.955551907	Detect	0	6.489122844	3.792309922	ND	4.78897645	1.859342494	0.799127513	Detect
3	KITB-003-WQ-01	11.46722466	0.497721919	0.23411593	Detect	14.27385883	22.20809444	4.955551907	<MRL	61.75753333	6.489122844	3.792309922	Detect	274.8837737	1.859342494	0.799127513	Detect
4	KITB-005-WQ-01	4.744429878	0.497721919	0.23411593	Detect	14.61255512	22.20809444	4.955551907	<MRL	1.789390252	6.489122844	3.792309922	ND	12.9609748	1.859342494	0.799127513	Detect
5	KITB-012-WQ-01	4.7181798	0.497721919	0.23411593	Detect	37.51255762	22.20809444	4.955551907	Detect	1.382817166	6.489122844	3.792309922	ND	14.48918769	1.859342494	0.799127513	Detect
6	KITB-013-WQ-01	0.233821322	0.497721919	0.23411593	ND	89.68596869	22.20809444	4.955551907	Detect	4.793948095	6.489122844	3.792309922	<MRL	61.39956762	1.859342494	0.799127513	Detect
7	KITB-018-WQ-01	0.690258196	0.497721919	0.23411593	Detect	23.50681161	22.20809444	4.955551907	Detect	0	6.489122844	3.792309922	ND	5.219000724	1.859342494	0.799127513	Detect
8	KITB-019-WQ-01	10.06186343	0.497721919	0.23411593	Detect	4.630671797	22.20809444	4.955551907	ND	91.99615783	6.489122844	3.792309922	Detect	163.911853	1.859342494	0.799127513	Detect
9	KITB-022-WQ-01	2.623376921	0.497721919	0.23411593	Detect	59.28640195	22.20809444	4.955551907	Detect	0.238899791	6.489122844	3.792309922	ND	10.3183991	1.859342494	0.799127513	Detect
10	KITB-027-WQ-01	0.330617155	0.497721919	0.23411593	<MRL	0	22.20809444	4.955551907	ND	0	6.489122844	3.792309922	ND	1.708619785	1.859342494	0.799127513	<MRL

AHHS Water

Orig file from EPA: AHHS_Water_DataCleaned_231020_nmd.csv

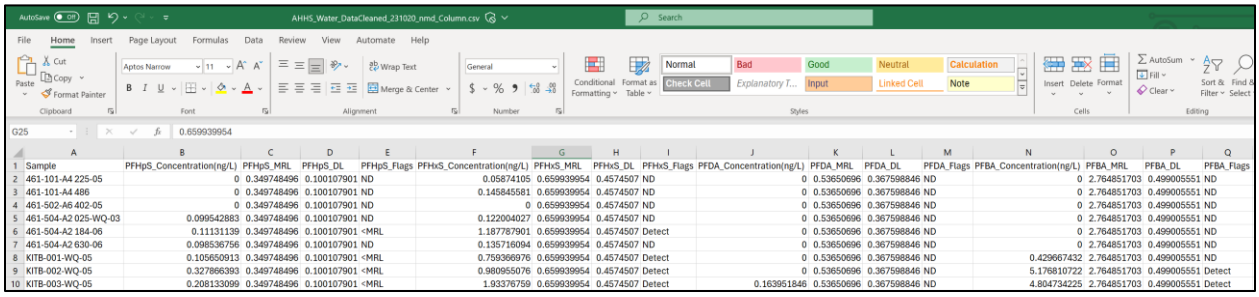


	A	B	C	D	E	F	G	H	I	J	K
1	Sample	Compound	MRL	DL	Concentration	Unit	censor_type	City	State	Sample_Match	Medium
2	461-101-A4 225-05	PFHpS	0.349748496	0.100107901	0	ng/L	ND	PARK HILLS	MO	461-101-	water
3	461-101-A4 225-05	PFHxS	0.659939954	0.4574507	0.05874105	ng/L	ND	PARK HILLS	MO	461-101-	water
4	461-101-A4 225-05	PFDA	0.53650696	0.367598846	0	ng/L	ND	PARK HILLS	MO	461-101-	water
5	461-101-A4 225-05	PFBA	2.764851703	0.499005551	0	ng/L	ND	PARK HILLS	MO	461-101-	water
6	461-101-A4 225-05	PFPeA	2.78524569	1.72953891	0.326360789	ng/L	ND	PARK HILLS	MO	461-101-	water
7	461-101-A4 225-05	PFPeS	0.426418553	0.068487602	0	ng/L	ND	PARK HILLS	MO	461-101-	water
8	461-101-A4 225-05	PFNA	0.778466193	0.454874105	0	ng/L	ND	PARK HILLS	MO	461-101-	water
9	461-101-A4 225-05	PFHxA	0.893226612	0.621258261	0.078220575	ng/L	ND	PARK HILLS	MO	461-101-	water
10	461-101-A4 225-05	PFBS	0.330468505	0.111308675	0.151503035	ng/L	<MRL	PARK HILLS	MO	461-101-	water

R code (PFAS_Data_Processing.R) to process the original file: 1) Removed all the NA rows 2) Transposed data 3) For each compound abbreviation, add concentration, MRL, DL, and Flag as suffix

```
###AHHS_Water_Cleaned_Data###
data <- read.csv("C:/Publications/PFAS database/Data Files Working Copies/AHHS_Water_DataCleaned_231020_nmd.csv", header = TRUE) #head(data)
data <- na.omit(data)
length(unique(data$Sample))
length(unique(data$Compound))
#data <- as.data.frame(data)
#write.csv(data,"C:/Publications/PFAS database/Data Files Working Copies/AHHS_Water_DataCleaned_231020_nmd_NO_NA.csv", row.names=FALSE)


Tdata <- matrix(data = NA, nrow = length(unique(data$Sample)), ncol = 1+4*length(unique(data$Compound)), byrow =FALSE, dimnames =NULL) #head(Tdata)
for (i in 1:length(unique(data$Sample))) { # i=2
  datasub <- data[data$Sample == unique(data$Sample)[i],]
  for (j in 1:length(unique(data$Compound))) { #j=2
    Tdata[i,1] <- unique(data$Sample)[i]
    Tdata[i,4*j-2] <- datasub[datasub$Compound == unique(data$Compound)[j],5]
    Tdata[i,4*j-1] <- datasub[datasub$Compound == unique(data$Compound)[j],3]
    Tdata[i,4*j] <- datasub[datasub$Compound == unique(data$Compound)[j],4]
    Tdata[i,4*j+1] <- datasub[datasub$Compound == unique(data$Compound)[j],7]
  }
}

colnames(Tdata) <- c("Sample","PFHpS_Concentration(ng/L)","PFHpS_MRL","PFHpS_DL","PFHpS_Flags","PFHxS_Concentration(ng/L)","PFHxS_MRL","PFHxS_DL","PFHxS_F
                     ,"PFPeA_Concentration(ng/L)","PFPeA_MRL","PFPeA_DL","PFPeA_Flags","PFPeS_Concentration(ng/L)","PFPeS_MRL","PFPeS_DL","PFPeS_Flags","P
                     ,"PFBS_Concentration(ng/L)","PFBS_MRL","PFBS_DL","PFBS_Flags","PFOS_Concentration(ng/L)","PFOS_MRL","PFOS_DL","PFOS_Flags","PFHpA_Cor
                     ,"PFNS_Concentration(ng/L)","PFNS_MRL","PFNS_DL","PFNS_Flags")
Tdata <- as.data.frame(Tdata)
write.csv(Tdata,"C:/Publications/PFAS database/Data Files Working Copies/Processed_data/AHHS_Water_DataCleaned_231020_nmd_Column.csv", row.names=FALSE)
```

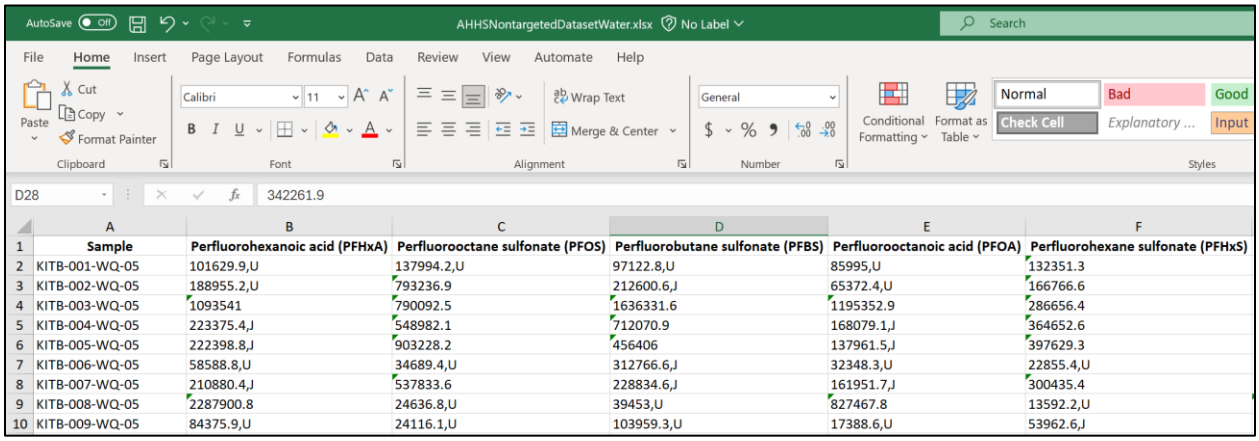
Bin processed file (13 PFAS compounds):
AHHS_Water_DataCleaned_231020_nmd_Column.csv



Sample	PFHpS_Concentration(ng/L)	PFHpS_MRL	PFHpS_DL	PFHpS_Flags	PFHxS_Concentration(ng/L)	PFHxS_MRL	PFHxS_DL	PFHxS_Flags	PFDA_Concentration(ng/L)	PFDA_MRL	PFDA_DL	PFDA_Flags	PFBA_Concentration(ng/L)	PFBA_MRL	PFBA_DL	PFBA_Flags
461-101-A4 225-05	0	0.349748496	0.100107901	ND	0.05874105	0.659939954	0.4574507	ND	0	0.53650696	0.367598846	ND	0	2.764851703	0.499005551	ND
461-101-A4 486	0	0.349748496	0.100107901	ND	0.145845581	0.659939954	0.4574507	ND	0	0.53650696	0.367598846	ND	0	2.764851703	0.499005551	ND
461-502-A6 402-05	0	0.349748496	0.100107901	ND	0	0.659939954	0.4574507	ND	0	0.53650696	0.367598846	ND	0	2.764851703	0.499005551	ND
461-504-A2 025-WQ-03	0.099542883	0.349748496	0.100107901	ND	0.122004027	0.659939954	0.4574507	ND	0	0.53650696	0.367598846	ND	0	2.764851703	0.499005551	ND
461-504-A2 184-06	0.11131139	0.349748496	0.100107901	<MRL	1.187787901	0.659939954	0.4574507	Detect	0	0.53650696	0.367598846	ND	0	2.764851703	0.499005551	ND
461-504-A2 630-06	0.098536756	0.349748496	0.100107901	ND	0.135716094	0.659939954	0.4574507	ND	0	0.53650696	0.367598846	ND	0	2.764851703	0.499005551	ND
KITB-001-WQ-05	0.105650913	0.349748496	0.100107901	<MRL	0.759366976	0.659939954	0.4574507	Detect	0	0.53650696	0.367598846	ND	0.429667432	2.764851703	0.499005551	ND
KITB-002-WQ-05	0.327866393	0.349748496	0.100107901	<MRL	0.980955076	0.659939954	0.4574507	Detect	0	0.53650696	0.367598846	ND	5.176810722	2.764851703	0.499005551	Detect
KITB-003-WQ-05	0.208133099	0.349748496	0.100107901	<MRL	1.93376759	0.659939954	0.4574507	Detect	0.163951846	0.53650696	0.367598846	ND	4.804734225	2.764851703	0.499005551	Detect

AHHS Nontargeted Water

Orig file from EPA: AHHSNontargetedDatasetWater.xlsx

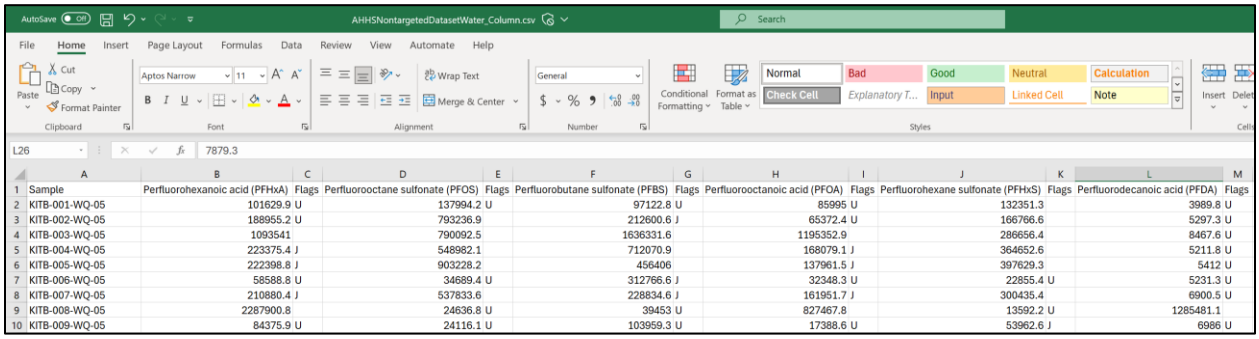


Sample	Perfluorohexanoic acid (PFHxA)	Perfluorooctane sulfonate (PFOS)	Perfluorobutane sulfonate (PFBS)	Perfluorooctanoic acid (PFOA)	Perfluorohexane sulfonate (PFHxS)
KITB-001-WQ-05	101629.9,U	137994.2,U	97122.8,U	85995,U	132351.3
KITB-002-WQ-05	188955.2,U	793236.9	212600.6,J	65372.4,U	166766.6
KITB-003-WQ-05	1093541	790092.5	1636331.6	1195352.9	286656.4
KITB-004-WQ-05	223375.4,J	548982.1	712070.9	168079.1,J	364652.6
KITB-005-WQ-05	222398.8,J	903228.2	456406	137961.5,J	397629.3
KITB-006-WQ-05	58588.8,U	34689.4,U	312766.6,J	32348.3,U	22855.4,U
KITB-007-WQ-05	210880.4,J	537833.6	228834.6,J	161951.7,J	300435.4
KITB-008-WQ-05	2287900.8	24636.8,U	39453,U	827467.8	13592.2,U
KITB-009-WQ-05	84375.9,U	24116.1,U	103959.3,U	17388.6,U	53962.6,J

R code (PFAS_Data_Processing.R) to process the original file: 1) For each compound, separated measurements values with flag and make measurements values and flag as two separate columns

```r
###AHHSNontargetedDatasetWater###
data <- read.csv("C:/Publications/PFAS database/Data Files Working Copies/AHHSNontargetedDatasetWater.csv", header = TRUE) #head(data)

Tdata <- matrix(data = NA, nrow = length(unique(data$Sample)), ncol = 2*dim(data)[2]-1, byrow =FALSE, dimnames =NULL) #head(Tdata) tail(Tdata)
for (i in 1:length(unique(data$Sample))) {
  datasub <- data[data$Sample == unique(data$Sample)[i],]
  for (j in 1:(dim(data)[2]-1)) { #i=3 j=1
    Tdata[i,1] <- unique(data$Sample)[i]
    a <- length(unlist(strsplit(as.character(datasub[j+1]), ",")))
    if (a  > 1) {
      Tdata[i,2*j] <- unlist(strsplit(as.character(datasub[j+1]), ","))[1]
      Tdata[i,2*j+1] <- unlist(strsplit(as.character(datasub[j+1]), ","))[2]
    }
    else {
      Tdata[i,2*j] <- unlist(strsplit(as.character(datasub[j+1]), ","))
      Tdata[i,2*j+1] <- " "
    }
  }
}

header <- matrix(data = NA, nrow = 150, ncol = 1, byrow =FALSE, dimnames =NULL)
for (i in 1:75) {
  header[2*i-1] <- colnames(data)[i+1]
  header[2*i] <- "Flags"
}

colnames(Tdata) <- rbind("Sample",header)


Tdata <- as.data.frame(Tdata)
write.csv(Tdata,"C:/Publications/PFAS database/Data Files Working Copies/Processed_data/AHHSNontargetedDatasetWater_Column.csv", row.names=FALSE)
```

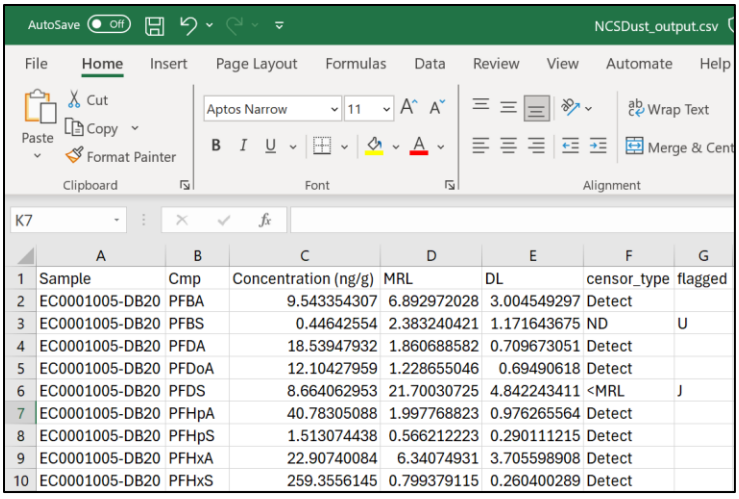
Bin processed file (75 PFAS compounds): AHHSNontargetedDatasetWater_Column.csv



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Sample	Perfluorohexanoic acid (PFHxA)	Flags	Perfluorooctane sulfonate (PFOS)	Flags	Perfluorobutane sulfonate (PFBS)	Flags	Perfluorooctanoic acid (PFOA)	Flags	Perfluorohexane sulfonate (PFHxS)	Flags	Perfluorodecanoic acid (PFDA)	Flags
2	KITB-001-WQ-05	101629.9	U	137994.2	U	97122.8	U	85995	U	132351.3		3989.8	U
3	KITB-002-WQ-05	188955.2	U	793236.9		212600.6	J	65372.4	U	166766.6		5297.3	U
4	KITB-003-WQ-05	1093541		790092.5		1636331.6		1195352.9		286656.4		8467.6	U
5	KITB-004-WQ-05	223375.4	J	548982.1		712070.9		168079.1	J	364652.6		5211.8	U
6	KITB-005-WQ-05	222398.8	J	903228.2		456406		137961.5	J	397629.3		5412	U
7	KITB-006-WQ-05	58588.8	U	34689.4	U	312766.6	J	32348.3	U	22855.4	U	5231.3	U
8	KITB-007-WQ-05	210880.4	J	537833.6		228834.6	J	161951.7	J	300435.4		6900.5	U
9	KITB-008-WQ-05	2287900.8		24636.8	U	39453	U	827467.8		13592.2	U	1285481.1	
10	KITB-009-WQ-05	84375.9	U	24116.1	U	103959.3	U	17388.6	U	53962.6	J	6986	U

NCS Dust

Orig file from EPA: NCSDust_output.csv

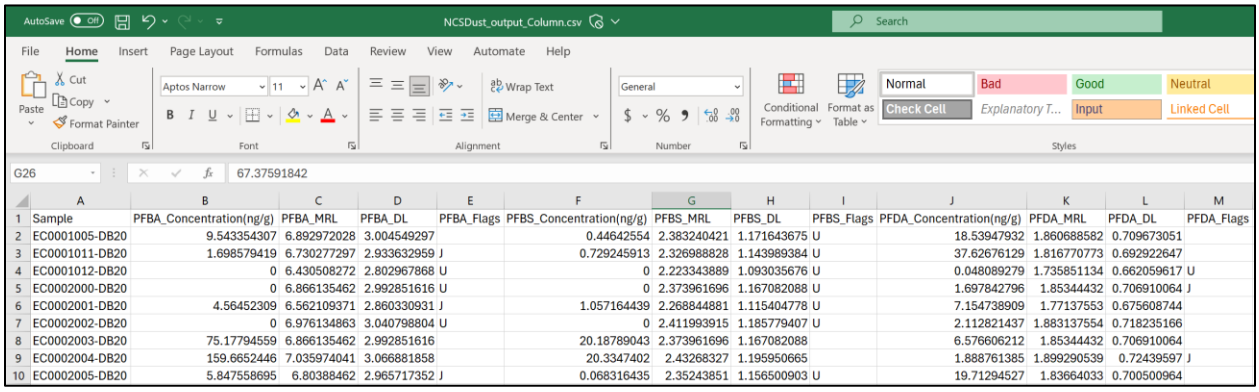


	A	B	C	D	E	F	G
1	Sample	Cmp	Concentration (ng/g)	MRL	DL	censor_type	flagged
2	EC0001005-DB20	PFBA	9.543354307	6.892972028	3.004549297	Detect	
3	EC0001005-DB20	PFBS	0.44642554	2.383240421	1.171643675	ND	U
4	EC0001005-DB20	PFDA	18.53947932	1.860688582	0.709673051	Detect	
5	EC0001005-DB20	PFDoA	12.10427959	1.228655046	0.69490618	Detect	
6	EC0001005-DB20	PFDS	8.664062953	21.70030725	4.842243411	<MRL	J
7	EC0001005-DB20	PFHpA	40.78305088	1.997768823	0.976265564	Detect	
8	EC0001005-DB20	PFHpS	1.513074438	0.566212223	0.290111215	Detect	
9	EC0001005-DB20	PFHxA	22.90740084	6.34074931	3.705598908	Detect	
10	EC0001005-DB20	PFHxS	259.3556145	0.799379115	0.260400289	Detect	

R code (PFAS_Data_Processing.R) to process the original file: 1) Transposed data 2) For each compound abbreviation, add concentration, MRL, DL, and Flag as suffix

```
###NCSDust###
data <- read.csv("C:/Publications/PFAS database/Data Files Working Copies/NCSDust_output.csv", header = TRUE) #head(data)
length(unique(data$Sample))
length(unique(data$Cmp))

Tdata <- matrix(data = NA, nrow = length(unique(data$Sample)), ncol = 1+4*length(unique(data$Cmp)), byrow =FALSE, dimnames =NULL) #head(Tdata)
for (i in 1:length(unique(data$Sample))) { # i=46
  datasub <- data[data$Sample == unique(data$Sample)[i],]
  for (j in 1:length(unique(data$Cmp))) { #j=1
    Tdata[i,1] <- unique(data$Sample)[i]
    Tdata[i,4*j-2] <- mean(datasub[datasub$Cmp == unique(data$Cmp)[j],3])
    Tdata[i,4*j-1] <- mean(datasub[datasub$Cmp == unique(data$Cmp)[j],4])
    Tdata[i,4*j] <- mean(datasub[datasub$Cmp == unique(data$Cmp)[j],5])
    Tdata[i,4*j+1] <- datasub[datasub$Cmp == unique(data$Cmp)[j],7][1]
  }
}

colnames(Tdata) <- c("Sample","PFBA_Concentration(ng/g)","PFBA_MRL","PFBA_DL","PFBA_Flags","PFBS_Concentration(ng/g)","PFBS_MRL","PFBS_DL","PFBS_Flags","P
                     ,"PFDS_Concentration(ng/g)","PFDS_MRL","PFDS_DL","PFDS_Flags","PFHpA_Concentration(ng/g)","PFHpA_MRL","PFHpA_DL","PFHpA_Flags","PFHpS
                     ,"PFHxS_Concentration(ng/g)","PFHxS_MRL","PFHxS_DL","PFHxS_Flags","PFNA_Concentration(ng/g)","PFNA_MRL","PFNA_DL","PFNA_Flags","PFNS_
                     ,"PFOS_Concentration(ng/g)","PFOS_MRL","PFOS_DL","PFOS_Flags","PFPEA_Concentration(ng/g)","PFPEA_MRL","PFPEA_DL","PFPEA_Flags","PFPeS
Tdata <- as.data.frame(Tdata)
write.csv(Tdata,"C:/Publications/PFAS database/Data Files Working Copies/Processed_data/NCSDust_output_Column.csv", row.names=FALSE)
```

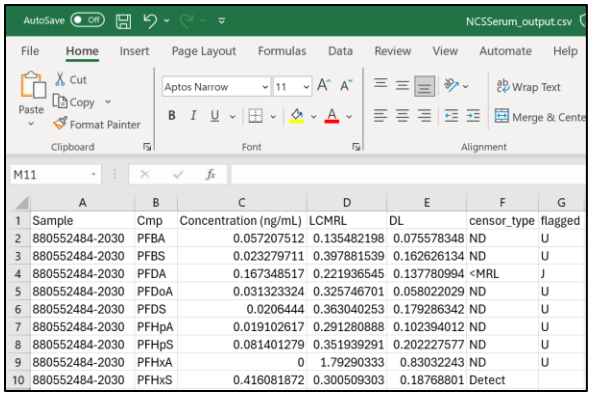
Bin processed file (16 PFAS compounds): NCSDust_output_Column.csv



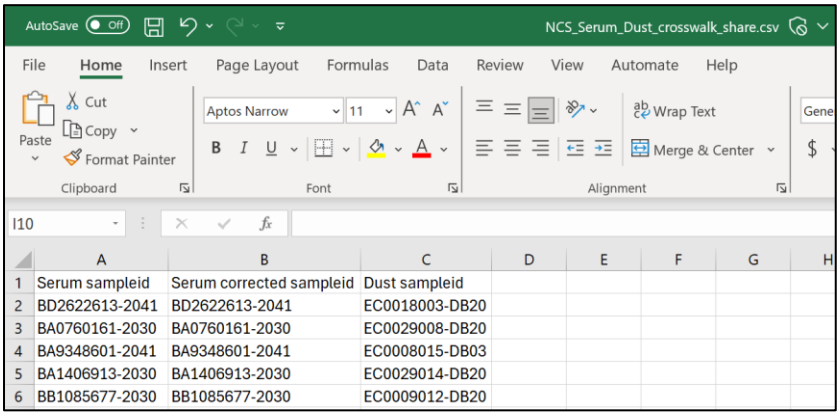
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Sample	PFBA_Concentration(ng/g)	PFBA_MRL	PFBA_DL	PFBA_Flags	PFBS_Concentration(ng/g)	PFBS_MRL	PFBS_DL	PFBS_Flags	PFDA_Concentration(ng/g)	PFDA_MRL	PFDA_DL	PFDA_Flags
2	EC0001005-DB20	9.543354307	6.892972028	3.004549297		0.44642554	2.383240421	1.171643675	U	18.53947932	1.860688582	0.709673051	
3	EC0001011-DB20	1.698579419	6.730277297	2.933632959	J	0.729245913	2.326988828	1.143989384	U	37.62676129	1.816770773	0.692922647	
4	EC0001012-DB20	0	6.430508272	2.802967868	U	0	2.223343889	1.093035676	U	0.048089279	1.735851134	0.662059617	U
5	EC0002000-DB20	0	6.866135462	2.992851616	U	0	2.373961696	1.167082088	U	1.697842796	1.85344432	0.706910064	J
6	EC0002001-DB20	4.56452309	6.562109371	2.860330931	J	1.057164439	2.268844881	1.115404778	U	7.154738909	1.77137553	0.675608744	
7	EC0002002-DB20	0	6.976134863	3.040798804	U	0	2.411993915	1.185779407	U	2.112821437	1.883137554	0.718235166	
8	EC0002003-DB20	75.17794559	6.866135462	2.992851616		20.18789043	2.373961696	1.167082088		6.576606212	1.85344432	0.706910064	
9	EC0002004-DB20	159.6652446	7.035974041	3.066881858		20.3347402	2.43268327	1.195950665		1.888761385	1.899290539	0.72439597	J
10	EC0002005-DB20	5.847558695	6.80388462	2.965717352	J	0.068316435	2.35243851	1.156500903	U	19.71294527	1.83664033	0.700500964	

NCS Serum

Orig file from EPA: NCSSerum_output.csv + NCS_Serum_Dust_crosswalk_share.csv



	A	B	C	D	E	F	G
1	Sample	Cmp	Concentration (ng/mL)	LCMRL	DL	censor_type	flagged
2	880552484-2030	PFBA	0.057207512	0.135482198	0.075578348	ND	U
3	880552484-2030	PFBS	0.023279711	0.397881539	0.162626134	ND	U
4	880552484-2030	PFDA	0.167348517	0.221936545	0.137780994	<MRL	J
5	880552484-2030	PFDoA	0.031323324	0.325746701	0.058022029	ND	U
6	880552484-2030	PFDS	0.0206444	0.363040253	0.179286342	ND	U
7	880552484-2030	PFHpA	0.019102617	0.291280888	0.102394012	ND	U
8	880552484-2030	PFHpS	0.081401279	0.351939291	0.202227577	ND	U
9	880552484-2030	PFHxA	0	1.79290333	0.83032243	ND	U
10	880552484-2030	PFHxS	0.416081872	0.300509303	0.18768801	Detect	



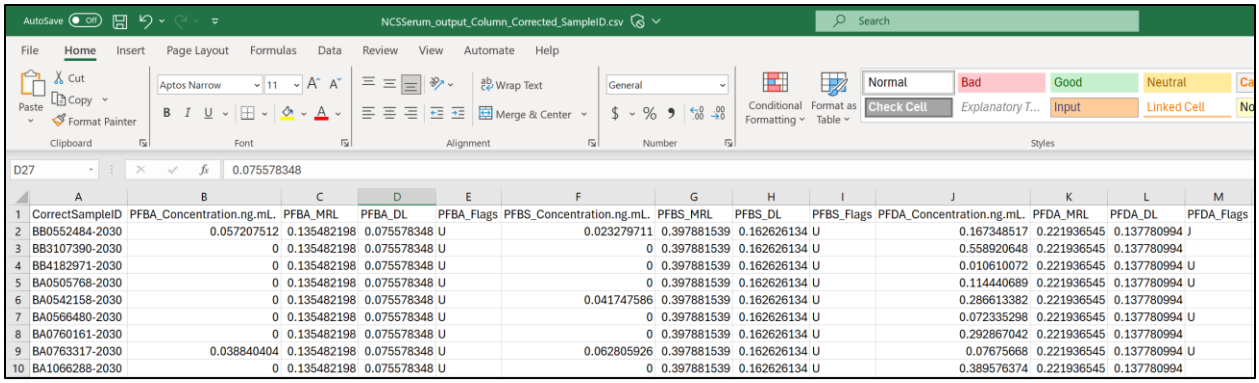
	A	B	C
1	Serum sampleid	Serum corrected sampleid	Dust sampleid
2	BD2622613-2041	BD2622613-2041	EC0018003-DB20
3	BA0760161-2030	BA0760161-2030	EC0029008-DB20
4	BA9348601-2041	BA9348601-2041	EC0008015-DB03
5	BA1406913-2030	BA1406913-2030	EC0029014-DB20
6	BB1085677-2030	BB1085677-2030	EC0009012-DB20

R code (PFAS_Data_Processing.R) to process the original file: 1) Transposed data 2) Corrected the sample ID 3) For each compound abbreviation, add concentration, MRL, DL, and Flag as suffix

```
###NCSSerum###
data <- read.csv("C:/Publications/PFAS database/Data Files Working Copies/NCSSerum_output.csv", header = TRUE) #head(data)
length(unique(data$Sample))
length(unique(data$Cmp))

Tdata <- matrix(data = NA, nrow = length(unique(data$Sample)), ncol = 1+4*length(unique(data$Cmp)), byrow =FALSE, dimnames =NULL) #head(Tdata)
for (i in 1:length(unique(data$Sample))) { # i=2
  datasub <- data[data$Sample == unique(data$Sample)[i],]
  for (j in 1:length(unique(data$Cmp))) { #j=1
    Tdata[i,1] <- unique(data$Sample)[i]
    Tdata[i,4*j-2] <- datasub[datasub$Cmp == unique(data$Cmp)[j],3]
    Tdata[i,4*j-1] <- datasub[datasub$Cmp == unique(data$Cmp)[j],4]
    Tdata[i,4*j] <- datasub[datasub$Cmp == unique(data$Cmp)[j],5]
    Tdata[i,4*j+1] <- datasub[datasub$Cmp == unique(data$Cmp)[j],7]
  }
}

colnames(Tdata) <- c("Sample","PFBA_Concentration(ng/mL)","PFBA_MRL","PFBA_DL","PFBA_Flags","PFBS_Concentration(ng/mL)","PFBS_MRL","PFBS_DL","PFBS_Flags","PFDA_Concentr
                     ,"PFDS_Concentration(ng/mL)","PFDS_MRL","PFDS_DL","PFDS_Flags","PFHpA_Concentration(ng/mL)","PFHpA_MRL","PFHpA_DL","PFHpA_Flags","PFHpS_Concentrat
                     ,"PFHxS_Concentration(ng/mL)","PFHxS_MRL","PFHxS_DL","PFHxS_Flags","PFNA_Concentration(ng/mL)","PFNA_MRL","PFNA_DL","PFNA_Flags","PFNS_Concentrati
                     ,"PFOS_Concentration(ng/mL)","PFOS_MRL","PFOS_DL","PFOS_Flags","PFPEA_Concentration(ng/mL)","PFPEA_MRL","PFPEA_DL","PFPEA_Flags","PFPeS_Concentrat
Tdata <- as.data.frame(Tdata)
write.csv(Tdata,"C:/Publications/PFAS database/Data Files Working Copies/Processed_data/NCSSerum_output_Column.csv", row.names=FALSE)
```

```
###Correct NCS transposed Serum data Sample ID###
NCSSerum <- read.csv("C:/Publications/PFAS database/Data Files Working Copies/Processed_data/NCSSerum_output_Column.csv", header = TRUE) #head(NCSSerum)
NCSDust <- read.csv("C:/Publications/PFAS database/Data Files Working Copies/Processed_data/NCSDust_output_Column.csv", header = TRUE) #head(NCSDust)
Cross <- read.csv("C:/Publications/PFAS database/Data Files Working Copies/Replies from Jason Boettger/NCS_Serum_Dust_crosswalk_share.csv", header = TRUE) #head(Cross)
NCSSerum$CorrectSampleID <- NA
for (i in 1:105) { #i =1
  NCSSerum$CorrectSampleID[i] <- Cross[Cross$Serum.sampleid == NCSSerum$Sample[i],2]
}
write.csv(NCSSerum,"C:/Publications/PFAS database/Data Files Working Copies/Processed_data/NCSSerum_output_Column_Corrected_SampleID.csv", row.names=FALSE)
```

Bin processed file (16 PFAS compounds): NCSSerum_output_Column_Corrected_SampleID.csv



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	CorrectSampleID	PFBA_Concentration.ng.mL.	PFBA_MRL	PFBA_DL	PFBA_Flags	PFBS_Concentration.ng.mL.	PFBS_MRL	PFBS_DL	PFBS_Flags	PFDA_Concentration.ng.mL.	PFDA_MRL	PFDA_DL	PFDA_Flags
2	BB0552484-2030	0.057207512	0.135482198	0.075578348	U	0.023279711	0.397881539	0.162626134	U	0.167348517	0.221936545	0.137780994	J
3	BB3107390-2030	0	0.135482198	0.075578348	U	0	0.397881539	0.162626134	U	0.558920648	0.221936545	0.137780994	
4	BB4182971-2030	0	0.135482198	0.075578348	U	0	0.397881539	0.162626134	U	0.010610072	0.221936545	0.137780994	U
5	BA0505768-2030	0	0.135482198	0.075578348	U	0	0.397881539	0.162626134	U	0.114440689	0.221936545	0.137780994	U
6	BA0542158-2030	0	0.135482198	0.075578348	U	0.041747586	0.397881539	0.162626134	U	0.286613382	0.221936545	0.137780994	
7	BA0566480-2030	0	0.135482198	0.075578348	U	0	0.397881539	0.162626134	U	0.072335298	0.221936545	0.137780994	U
8	BA0760161-2030	0	0.135482198	0.075578348	U	0	0.397881539	0.162626134	U	0.292867042	0.221936545	0.137780994	
9	BA0763317-2030	0.038840404	0.135482198	0.075578348	U	0.062805926	0.397881539	0.162626134	U	0.07675668	0.221936545	0.137780994	U
10	BA1066288-2030	0	0.135482198	0.075578348	U	0	0.397881539	0.162626134	U	0.389576374	0.221936545	0.137780994	