

BILAG 1 DATADOKUMENTATION

1.1 ORGANISERING AF DATA

De 5 danske regioner har undersøgt for PFAS i forbindelse med forureningsundersøgelser siden 2014, hvor arbejdet har været systematisk siden 2015. For de fleste undersøgelser har kortlægning af PFAS forurening ikke været det primære formål med undersøgelsen, så prøverne er taget i områder hvor der var mistanke om andre forureningsstoffer.

Der har før 2022 været få undersøgelser hvor PFAS har været det primære fokuskomponent for undersøgelsen.

Datasættet, der er anvendt i dette projekt er data, der er indrapporteret af regionerne i GEOGIS.

Der er der anvendt data leveret af Region Hovedstaden, Region Midt, Region Syd, Region Nord og Region Sjælland.

Datasættet indeholder data fra perioden 2013 til januar 2023. For 10 datapunkter er dog angivet en dato, der ligger før 2013.

Data er leveret fra de enkelte regioner som csv filer.

Der er foretaget en opdeling af brancher i forhold til potentiel PFAS-brancher jf. liste fra regionerne fra april 2022.

Data er samlet og organiseret i en Access-database, så hver lokalitet kun forekommer én gang med samme branche, uafhængigt af evt. flere driftsperioder. I denne organisering er alle branchekoder relateret til lossepladser og fyldpladser lagt sammen til én branchegruppe:

BRANCHEGRUPPER UNDER LOSSEPLADS

Branche 90.02 Indsamling og behandling af andet affald	Losseplads
Branche 90.02.00 Indsamling og behandling af andet affald	Losseplads
Branche 90.02.10 Indsamling af affald	Losseplads
Branche 90.02.20 Drift af affaldsbehandlingsanlæg	Losseplads
Branche 90.03.10 Renovation, snerydning mv.	Losseplads

Brancherne er yderligere grupperet ved indlæsning i Power BI. Som tommelfingerregel efter branchehovedgruppe, men nogle brancher er dog ikke trukket ind under hovedgruppen, da indsigt i forholdene omkr. brancheundergruppen har været ønsket, fx for 28.51 Overfladebehandling af metal, 17.51 Tæppefabrikker, mm. Grupperingen i branchegrupper kan ses i Bilag 2 ("Brancheliste_grupperede.xlsx").

1.2 DATAFILER

Fra Access-databasen er der udtrukket en række datasæt, der er anvendt til visualisering af data i Power BI.

Ved visualiseringer og analyser, der referer til brancher er følgende dataudtræk anvendt:

DATAFILER MED BRANCHEINFORMATION

Filnavn	Beskrivelse
Udtræk_Branche_alle_grundvand_medLossepladser.xlsx	Udtræk med brancheinformation med lokaliteter uanset antal PFAS-brancher i forhold til analyser i grundvand. Lokaliteter er opført med branchekoder, 1 gang pr. tilknyttet branche.
Udtræk_Branche_alle_jord_medLossepladser.xlsx	Udtræk med brancheinformation med lokaliteter uanset antal PFAS-brancher i forhold til analyser i grundvand. Lokaliteter er opført med branchekoder, 1 gang pr. tilknyttet branche.

Ved visualiseringer og analyser på tværs af alle lokaliteter, uagtet branche, er anvendt følgende dataudtræk:

DATAFILER UDEN BRANCHEINFORMATION

Filnavn	Beskrivelse
Udtræk_Alle lokaliteter_grundvand_max dybde konc.xlsx	Udtræk med den boring med det dybeste filter pr. lokalitet med tilhørende koncentration for prøver i grundvand.
Udtræk_Alle lokaliteter_grundvand_max konc_dybde.xlsx	Udtræk med højeste koncentration for prøver i grundvand, der er målt på lokaliteten samt tilhørende dybde af filteret.
Udtræk_Alle lokaliteter_grundvand.xlsx	Alle lokaliteter uagtet brancher, dvs. lokaliteter optræder med tilhørende analyseresultater for grundvand én gang i datasættet.

Dataudtrækket med prøver i grundvand for alle lokaliteter indeholder data fra 1501 lokaliteter, bestående af 11632 prøver og 223.474 analyseresultater for 38 PFAS forbindelser.

1.3 LOKALITETER MED FLERE BRANCHER

Der kan være flere brancher knyttet til én lokalitet. Det betyder at én prøve med tilhørende analyseresultater kan optræde fx tre gange i datasættet, under tre forskellige brancher. Dette kan ses for fx lokalitet 101-00087, prøvepunkt B104.

Denne datastruktur giver den udfordring, at lokaliteter som har mange prøver og/eller meget høje koncentrationer, vil skævvride billedet for tilknyttede brancher, der ikke nødvendigvis er drivende for PFAS-forureningen på lokaliteten. Dette er især problematisk hvis lokaliteten bidrager med mange analysedata i det samlede datasæt.

Et eksempel er Københavns Lufthavn (185-00011). Lokaliteten er registreret med brancher: 63.23 Lufthavne mv. samt 93.01.30 Renserier. De meget høje PFAS-koncentration forårsaget af aktiviteter knyttet til branche 63.23 forefindes også for branche 93.01.30, selvom renserivirkomheden på lokaliteten ikke vurderes som den drivende for PFAS forureningen.

For at undgå fejlslutninger har vi gjort følgende:

- For alle brancher tjekkes manuelt de 5 lokaliteter med højeste max koncentrationer (enkeltstof) eller dem med max koncentrationer >100 ng/L, hvis der er få lokaliteter i branchen.
 - Derefter er der foretaget en vurdering af, hvilken af de registrerede brancher, der er drivende for PFAS-forureningen på lokaliteten. Det har kun været muligt at foretage denne vurdering for én lokalitet, Københavns Lufthavn (185-00011). Her er lokaliteten fjernet fra branchen "Renserier" for alle branchespecifikke analyser i Power BI.

- For de øvrige lokaliteter er der beregnet et datasæt, hvor lokaliteter hvor antallet af prøver udgør mere end 10% af det samlede antal prøver for branchen er udeladt.
Dette datasæt er eksporteret fra Power BI og anvendt til at beregne histogrammer og sumkurver for de 10 brancher med flest lokaliteter (antal lokaliteter >50) i Excel. Kun for to af de brancher der er lavet histogrammer og sumkurver for, er der frasortet lokaliteter. Én lokalitet under branchen 28.51 Overfladebehandling af metal (657-00019) og to lokaliteter tilhørende branchen 24 Kemisk industri (169-0001 og 169-05131).

1.4 DATA-RENSNING

Alle skridt i den oprensning af data, der er beskrevet nedenfor, kan ses i Power Query i "Avanceret Editor" for de enkelte datafiler.

ENSRETNING AF ATTRIBUTTER:

Vi er som udgangspunkt interesseret i at analysere data, hvor der er påvist PFAS koncentrationer.

Værdier angivet som tomme celler eller NULL er ændret til "=" for at ensrette angivelsen af påviste værdier.

Værdier angivet som x, A, D eller < er alle ændret til "<" for at ensrette angivelsen af ikke-påviste værdier.

STOFKODE OG STOFTRIVIALNAVN

Stofkoder 20315 og 20314 er rettet til hhv. 20115 og 20114 på baggrund af det angivne StofTrivialnavn (20115: Sum 4 PFAS, 20114: Sum 12 PFAS).

Skrivemåden for StofTrivialnavn for Sum 4, 12, 22 i det indlæste data fra regionerne er ensrettet med udgangspunkt i den angivne stofkode, så der kun findes ét tilsvarende StofTrivialnavn til hver stofkode.

PFAS-navne der ikke var forkortet i kolonnen StofTrivialnavn, fx Perfluordodecansulfonsyre er erstattet med deres respektive forkortelser, her PFDoDS, så et stof ikke optræder under to (eller flere) StofTrivialnavne.

Stoftrivialnavne der var angivet "incl. precursor" er reduceret til navnet på stofnavnet, ex "PFHpA incl precursor" ændret til "PFHpA", for at simplificere kategoriseringen. Den tilhørende Stofkode er ikke ændret.

ENHEDER

Alle koncentrationer er omregnet til ng/L.

Datapunkter hvor enheden ikke er oplyst er manuelt tjekket igennem. Kun ét datapunkt har attribut "=" og værdi forskellig fra 0.

Dette datapunkt er en værdi for (Lab) Sum af 12 PFAS. Denne parameter bruges dog ikke i dataarbejdet og den manglende enhed har ikke betydning for det videre dataarbejde.