Data Analyse Landelijke Achtergrondwaarden bodem voor PFOA en PFOS

RIVM

Contents


```
## Skipping install of 'datafile' from a github remote, the SHA1 (eb19a689) has not changed since last i
##   Use `force = TRUE` to force installation
```

```
## Skipping install of 'achtergrondwaarden' from a github remote, the SHA1 (2d159208) has not changed si
##   Use `force = TRUE` to force installation
```

1 Inleiding

Om de analyse van de PFAS data reproduceerbaar te maken is gebruik gemaakt van een methode die bekend staat als 'literate programming'. De programma code voor de analyse is gecombineerd met documentatie en documentatie, code en output is gecombineerd in dit document.

Dit document geeft de gevolgde procedure en resultaten weer. Interpretatie van de resultaten wordt zoveel mogelijk vermeden. Deze interpretatie vind plaats in de afzonderlijke notitie *Tijdelijke landelijke achtergrondwaarden bodem voor PFOS en PFOA* (RIVM rapport nr 245 2019). Dit document maakt gebruik van de data die openbaar zijn gemaakt. Deze data verschilt met de data die gebruikt zijn voor de notitie. Op verzoek van de bronhouders van de data zijn de exacte coordinaten van een aantal locaties afgeschermd.

Voor die analyses waar de locatie van belang is, kunnen de resultaten in dit document dus afwijken van die uit bovengenoemde notitie.

Voor de analyse is gebruik gemaakt van de software R. Naast algemeen beschikbare packages, is voor dit project een specifiek package ontworpen voor het afleiden van achtergrondwaarden. Dit document maakt ook gebruik van het datafile package.

Dit document is niet geschreven voor het algemene publiek maar bedoeld voor experts op het gebied van statistiek en bodem chemie. Om dit document te kunnen volgen is het noodzakelijk dat men de programeerdcode kan lezen. In verband met de omvang is niet alle uitvoer volledig gegeven in het document.

2 Data import, tidy en transform

Alle data die naar het RIVM is opgestuurd is verwerkt en gecombineerd in een enkele dataset.

```
# pad en bestandsnaam van import bestand
fname.import <- datafile("bronbestand.csv")
x <- fname.import %>% read_delim(delim=";")
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   boorpunt = col_double(),
##   data_nr = col_double(),
##   kwadrant = col_double(),
##   chemours = col_double(),
##   dist_chemours = col_double(),
##   angle_chemours = col_double(),
##   galvano = col_double(),
##   dist_galvano = col_double(),
##   angle_galvano = col_double(),
##   textiel = col_double(),
##   dist_textiel = col_double(),
##   angle_textiel = col_double(),
##   brand = col_double(),
##   dist_brand = col_double(),
##   angle_brand = col_double(),
##   vlieg = col_double(),
##   dist_vlieg = col_double(),
##   angle_vlieg = col_double(),
##   stort = col_double(),
##   dist_stort = col_double()
##   # ... with 12 more columns
## )
```

```
## See spec(...) for full column specifications.
```

```
## Warning: 4 parsing failures.
##  row col               expected actual                                    file
## 1032   y no trailing characters    .68 '~/rivm/work/pfas/data/bronbestand.csv'
## 1033   y no trailing characters    .69 '~/rivm/work/pfas/data/bronbestand.csv'
## 1034   y no trailing characters    .68 '~/rivm/work/pfas/data/bronbestand.csv'
## 1035   y no trailing characters    .69 '~/rivm/work/pfas/data/bronbestand.csv'
```

```
saveRDS(x,datafile("brondata.rds"))
```

De verschillende veldnamen uit het bronbestand worden gegroepeerd en opgeslagen in een list. Met deze list kunnen selecties gemaakt worden van kolommen uit de brondata

```
dvar <- list(all=names(x))
dvar[["meta"]] <- c("monster","boorpunt","dataset","meng","methode",
                "dh","dl","diepte","grondlaag","landgebruik","lab","orig_set",
                "x","y","data_nr")
dvar[["metaselect"]] <- c("monster","boorpunt","dataset","methode","dh","dl","diepte","x","y","data_nr")
dvar[["cats"]] <- c("chemours","galvano","textiel","brand","vlieg","stort","avi", "kwadrant",
                    "use", "use_cat","type")
dvar[["catsnum"]] <- c("dist_chemours","angle_chemours","angle_galvano","dist_galvano",
                    "angle_textiel","dist_textiel","angle_brand","dist_brand",
                    "angle_vlieg","dist_vlieg","angle_stort","dist_stort","angle_avi",
                    "dist_avi")
dvar[["other"]] <- c("temp_ph","caco3","totc","gloei","ph","ds", "lutum","om")
dvar[["params"]] <- setdiff(dvar$all,c(dvar$meta,dvar$cats,
                                       dvar$catsnum))
dvar[["pfparams"]] <- setdiff(dvar$params,dvar$other)
dvar[["numfields"]] <- c(dvar$params,dvar$catsnum,"x","y","dh","dl")
dvar
```

```
## $all
##  [1] "monster"        "boorpunt"       "dataset"        "orig_set"
##  [5] "data_nr"        "kwadrant"       "use"            "use_cat"
##  [9] "type"           "chemours"       "dist_chemours"  "angle_chemours"
## [13] "galvano"        "dist_galvano"   "angle_galvano"  "textiel"
## [17] "dist_textiel"   "angle_textiel"  "brand"          "dist_brand"
## [21] "angle_brand"    "vlieg"          "dist_vlieg"     "angle_vlieg"
## [25] "stort"          "dist_stort"     "angle_stort"    "avi"
## [29] "dist_avi"       "angle_avi"      "meng"           "methode"
## [33] "dh"             "dl"             "diepte"         "lab"
## [37] "temp_ph"        "caco3"          "pfbs"           "pfds"
## [41] "pfhps"          "pfhxs"          "pfba"           "pfdea"
## [45] "pfdoda"         "pfhpa"          "pfhxa"          "pfna"
## [49] "pfoa_lin"       "pfos_lin"       "pfosa"          "pfpea"
## [53] "pftrda"         "pfteda"         "pfunda"         "pfes"
## [57] "pfos_tak"       "pfoa_tak"       "pfhxda"         "pfoda"
## [61] "mrfbsa"         "pfbsa"          "genx"           "mefbsaa"
## [65] "etfosaa"        "hhpfds8"        "hhpfds10"       "pfpes"
## [69] "pfosaa"         "chdfda"         "f53b"           "hhpfua"
## [73] "hhpfhsa"        "hpfhpa"         "dipap8"         "adona"
## [77] "etfosa"         "mefosa"         "pfdmoa"         "fts62"
## [81] "totc"           "gloei"          "ph"             "ds"
## [85] "lutum"          "om"             "x"              "y"
## [89] "som_pfos"       "som_pfoa"       "Atlas"
##
## $meta
##  [1] "monster"     "boorpunt"    "dataset"     "meng"        "methode"
##  [6] "dh"          "dl"          "diepte"      "grondlaag"   "landgebruik"
## [11] "lab"         "orig_set"    "x"           "y"           "data_nr"
##
## $metaselect
##  [1] "monster"  "boorpunt" "dataset"  "methode"  "dh"       "dl"
```

```
##  [7] "diepte"   "x"        "y"        "data_nr"
##
## $cats
##  [1] "chemours" "galvano"  "textiel"  "brand"    "vlieg"    "stort"
##  [7] "avi"      "kwadrant" "use"      "use_cat"  "type"
##
## $catsnum
##  [1] "dist_chemours"  "angle_chemours" "angle_galvano"  "dist_galvano"
##  [5] "angle_textiel"  "dist_textiel"   "angle_brand"    "dist_brand"
##  [9] "angle_vlieg"    "dist_vlieg"     "angle_stort"    "dist_stort"
## [13] "angle_avi"      "dist_avi"
##
## $other
## [1] "temp_ph" "caco3"   "totc"    "gloei"   "ph"      "ds"      "lutum"
## [8] "om"
##
## $params
##  [1] "temp_ph"  "caco3"    "pfbs"     "pfds"     "pfhps"    "pfhxs"
##  [7] "pfba"     "pfdea"    "pfdoda"   "pfhpa"    "pfhxa"    "pfna"
## [13] "pfoa_lin" "pfos_lin" "pfosa"    "pfpea"    "pftrda"   "pfteda"
## [19] "pfunda"   "pfes"     "pfos_tak" "pfoa_tak" "pfhxda"   "pfoda"
## [25] "mrfbsa"   "pfbsa"    "genx"     "mefbsaa"  "etfosaa"  "hhpfds8"
## [31] "hhpfds10" "pfpes"    "pfosaa"   "chdfda"   "f53b"     "hhpfua"
## [37] "hhpfhsa"  "hpfhpa"   "dipap8"   "adona"    "etfosa"   "mefosa"
## [43] "pfdmoa"   "fts62"    "totc"     "gloei"    "ph"       "ds"
## [49] "lutum"    "om"       "som_pfos" "som_pfoa" "Atlas"
##
## $pfparams
##  [1] "pfbs"     "pfds"     "pfhps"    "pfhxs"    "pfba"     "pfdea"
##  [7] "pfdoda"   "pfhpa"    "pfhxa"    "pfna"     "pfoa_lin" "pfos_lin"
## [13] "pfosa"    "pfpea"    "pftrda"   "pfteda"   "pfunda"   "pfes"
## [19] "pfos_tak" "pfoa_tak" "pfhxda"   "pfoda"    "mrfbsa"   "pfbsa"
## [25] "genx"     "mefbsaa"  "etfosaa"  "hhpfds8"  "hhpfds10" "pfpes"
## [31] "pfosaa"   "chdfda"   "f53b"     "hhpfua"   "hhpfhsa"  "hpfhpa"
## [37] "dipap8"   "adona"    "etfosa"   "mefosa"   "pfdmoa"   "fts62"
## [43] "som_pfos" "som_pfoa" "Atlas"
##
## $numfields
##  [1] "temp_ph"        "caco3"          "pfbs"           "pfds"
##  [5] "pfhps"          "pfhxs"          "pfba"           "pfdea"
##  [9] "pfdoda"         "pfhpa"          "pfhxa"          "pfna"
## [13] "pfoa_lin"       "pfos_lin"       "pfosa"          "pfpea"
## [17] "pftrda"         "pfteda"         "pfunda"         "pfes"
## [21] "pfos_tak"       "pfoa_tak"       "pfhxda"         "pfoda"
## [25] "mrfbsa"         "pfbsa"          "genx"           "mefbsaa"
## [29] "etfosaa"        "hhpfds8"        "hhpfds10"       "pfpes"
## [33] "pfosaa"         "chdfda"         "f53b"           "hhpfua"
## [37] "hhpfhsa"        "hpfhpa"         "dipap8"         "adona"
## [41] "etfosa"         "mefosa"         "pfdmoa"         "fts62"
## [45] "totc"           "gloei"          "ph"             "ds"
## [49] "lutum"          "om"             "som_pfos"       "som_pfoa"
## [53] "Atlas"          "dist_chemours"  "angle_chemours" "angle_galvano"
## [57] "dist_galvano"   "angle_textiel"  "dist_textiel"   "angle_brand"
## [61] "dist_brand"     "angle_vlieg"    "dist_vlieg"     "angle_stort"
```

```
## [65] "dist_stort"     "angle_avi"      "dist_avi"       "x"
## [69] "y"              "dh"             "dl"
```

2.1 Tidy data

Alle niet numerieke velden worden omgezet in numerieke velden. Als er geen numerieke waarde aanwezig is wordt het veld op NA gezet. Waarden beneden de detectielimiet worden in de dataset weergegeven met een 'kleiner dan' teken . Deze detectielimieten worden gecodeerd als negatieve waarden.

```
x <- datafile("brondata.rds") %>%
    readRDS()

x[dvar$params] <- sapply(x[dvar$params],replacelt)

x[dvar$numfields] <- sapply(x[dvar$numfields],as.numeric)
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
x$dataset <- sub("Zl1","ZL1",x$dataset) # correctie naamgeving
x <- bind_cols(id=1:nrow(x),x)
dvar$metaselect <- append(dvar$metaselect,"id")

saveRDS(x,datafile("pfas_tidy.rds"))
write_csv(x,datafile("pfas_tidy.csv"))
```

2.2 Transform data

2.2.1 Diepte profielen

We definieren de 3 diepte profielen:

1) *top* de laag 0-0.1 m-mv;
2) *sub* de laag 0-0.5 m-mv;
3) *diep* de laag dieper dan 0.5m-mv;

Monsters die vallen in de toplaag (0-0.1 m-mv) worden niet toegekend aan de sub laag (0-0.5 m-mv)

Monsters waarvan de diepte (dl) onbekend is, wordt het profiel *onbekend* aan toegekend.

```
xd <- readRDS(datafile("pfas_tidy.rds")) %>%
    select(id,dh,dl) %>%
    mutate(top=ifelse(dl<=0.1,"top","")) %>%
    mutate(sub=ifelse(top==""&dl<=.5,"sub","")) %>%
    mutate(diep=ifelse(dl>.5,"diep","")) %>%
    mutate(topsubdiep=paste0(top,sub,diep)) %>%
    mutate(profiel=str_replace(topsubdiep,"NANANA","onbekend"))

x  <- xd %>% select(id,profiel) %>%
    right_join(x,by="id")
x
```

```
## # A tibble: 1,461 x 93
##       id profiel monster boorpunt dataset orig_set data_nr kwadrant use
##    <int> <chr>   <chr>      <dbl> <chr>   <chr>      <dbl>    <dbl> <chr>
##  1     1 top     3-MM01         3 NH1     <NA>           1        1 Over~
##  2     2 sub     3-MM02         3 NH1     <NA>           2        1 Over~
```

```
##  3     3 diep    3-MM03         3 NH1     <NA>           3        1 Over~
##  4     4 top     5-MM01         5 NH1     <NA>           4        1 Over~
##  5     5 sub     5-MM02         5 NH1     <NA>           5        1 Over~
##  6     6 diep    5-MM03         5 NH1     <NA>           6        1 Over~
##  7     7 top     07-MM01        7 NH1     <NA>           7        1 Open~
##  8     8 sub     07-MM02        7 NH1     <NA>           8        1 Open~
##  9     9 diep    07-MM03        7 NH1     <NA>           9        1 Open~
## 10    10 top     8-MM01         8 NH1     <NA>          10        1 Over~
## # ... with 1,451 more rows, and 84 more variables: use_cat <chr>, type <chr>,
## #   chemours <dbl>, dist_chemours <dbl>, angle_chemours <dbl>, galvano <dbl>,
## #   dist_galvano <dbl>, angle_galvano <dbl>, textiel <dbl>, dist_textiel <dbl>,
## #   angle_textiel <dbl>, brand <dbl>, dist_brand <dbl>, angle_brand <dbl>,
## #   vlieg <dbl>, dist_vlieg <dbl>, angle_vlieg <dbl>, stort <dbl>,
## #   dist_stort <dbl>, angle_stort <dbl>, avi <dbl>, dist_avi <dbl>,
## #   angle_avi <dbl>, meng <chr>, methode <chr>, dh <dbl>, dl <dbl>,
## #   diepte <chr>, lab <chr>, temp_ph <dbl>, caco3 <dbl>, pfbs <dbl>,
## #   pfds <dbl>, pfhps <dbl>, pfhxs <dbl>, pfba <dbl>, pfdea <dbl>,
## #   pfdoda <dbl>, pfhpa <dbl>, pfhxa <dbl>, pfna <dbl>, pfoa_lin <dbl>,
## #   pfos_lin <dbl>, pfosa <dbl>, pfpea <dbl>, pftrda <dbl>, pfteda <dbl>,
## #   pfunda <dbl>, pfes <dbl>, pfos_tak <dbl>, pfoa_tak <dbl>, pfhxda <dbl>,
## #   pfoda <dbl>, mrfbsa <dbl>, pfbsa <dbl>, genx <dbl>, mefbsaa <dbl>,
## #   etfosaa <dbl>, hhpfds8 <dbl>, hhpfds10 <dbl>, pfpes <dbl>, pfosaa <dbl>,
## #   chdfda <dbl>, f53b <dbl>, hhpfua <dbl>, hhpfhsa <dbl>, hpfhpa <dbl>,
## #   dipap8 <dbl>, adona <dbl>, etfosa <dbl>, mefosa <dbl>, pfdmoa <dbl>,
## #   fts62 <dbl>, totc <dbl>, gloei <dbl>, ph <dbl>, ds <dbl>, lutum <dbl>,
## #   om <dbl>, x <dbl>, y <dbl>, som_pfos <dbl>, som_pfoa <dbl>, Atlas <dbl>
```

```
dvar$metaselect <- append(dvar$metaselect,"profiel")
```

2.2.2 Berekenen van som parameters

Concentratie linaire en vertakte PFOS en concentraties van lineaire en vertakte PFOA worden bij elkaar opgeteld. Er worden twee soorten somwaarden afgeleid:

1. somwaarden waarbij de detectielimieten (LOQ) worden vervangen door 0.7 * LOQ en vervolgens worden opgeteld;
2. somwaarden worden bij elkaar opgeteld en gecodeerd als negatieve waarden indien 1 van 2 concentraties onder de detectielimiet valt.

Als er geen somwaarde berekend kan worden dan wordt de somwaarde gebruikt uit de dataset, indien deze aanwezig is.

```
c_som <- x %>% select(pfos_tak,pfos_lin,pfoa_tak,pfoa_lin, som_pfos,som_pfoa) %>%
    as_tibble() %>%
    mutate(s_som_pfos=ifelse(pfos_tak<0|pfos_lin<0,-1,1)) %>%
    mutate(s_som_pfoa=ifelse(pfoa_tak<0|pfoa_lin<0,-1,1)) %>%
    mutate(i_pfos_tak=replaceNegative(pfos_tak,0.7)) %>%
    mutate(i_pfos_lin=replaceNegative(pfos_lin,0.7)) %>%
    mutate(i_pfoa_tak=replaceNegative(pfoa_tak,0.7)) %>%
    mutate(i_pfoa_lin=replaceNegative(pfoa_lin,0.7)) %>%
    mutate(i_som_pfoa=replaceNegative(som_pfoa,0.7)) %>%
    mutate(i_som_pfos=replaceNegative(som_pfos,0.7)) %>%
    mutate(t_som_pfos=i_pfos_tak+i_pfos_lin,t_som_pfoa=i_pfoa_tak+i_pfoa_lin) %>%
    mutate(l_som_pfos=s_som_pfos*(abs(pfos_tak)+abs(pfos_lin)),
           l_som_pfoa=s_som_pfoa*(abs(pfoa_tak)+abs(pfoa_lin))) %>%
```

```
    mutate(c_som_pfos=ifelse(is.na(t_som_pfos),i_som_pfos,t_som_pfos),
           c_som_pfoa=ifelse(is.na(t_som_pfoa),i_som_pfoa,t_som_pfoa)) %>%
    mutate(lc_som_pfos=ifelse(is.na(l_som_pfos),som_pfos,l_som_pfos),
           lc_som_pfoa=ifelse(is.na(l_som_pfoa),som_pfoa,l_som_pfoa))

dvar$somparams <- c("c_som_pfoa","c_som_pfos","lc_som_pfoa","lc_som_pfos")
dvar$pfparams <- append(dvar$pfparams,dvar$somparams)
xd2 <- c_som %>% select(c_som_pfoa,c_som_pfos,lc_som_pfoa,lc_som_pfos) %>% bind_cols(x) %>%
    select(dvar$metaselect,dvar$other,dvar$pfparams)

xd2 %>% saveRDS(datafile("pfas_transform.rds"))
xd2 %>% write_csv(datafile("pfas_getransformeerdeWaarden.csv"))
dvar %>% saveRDS(datafile("pfas_params.rds"))
c_som %>% write_csv(datafile("c_som.csv"))
```

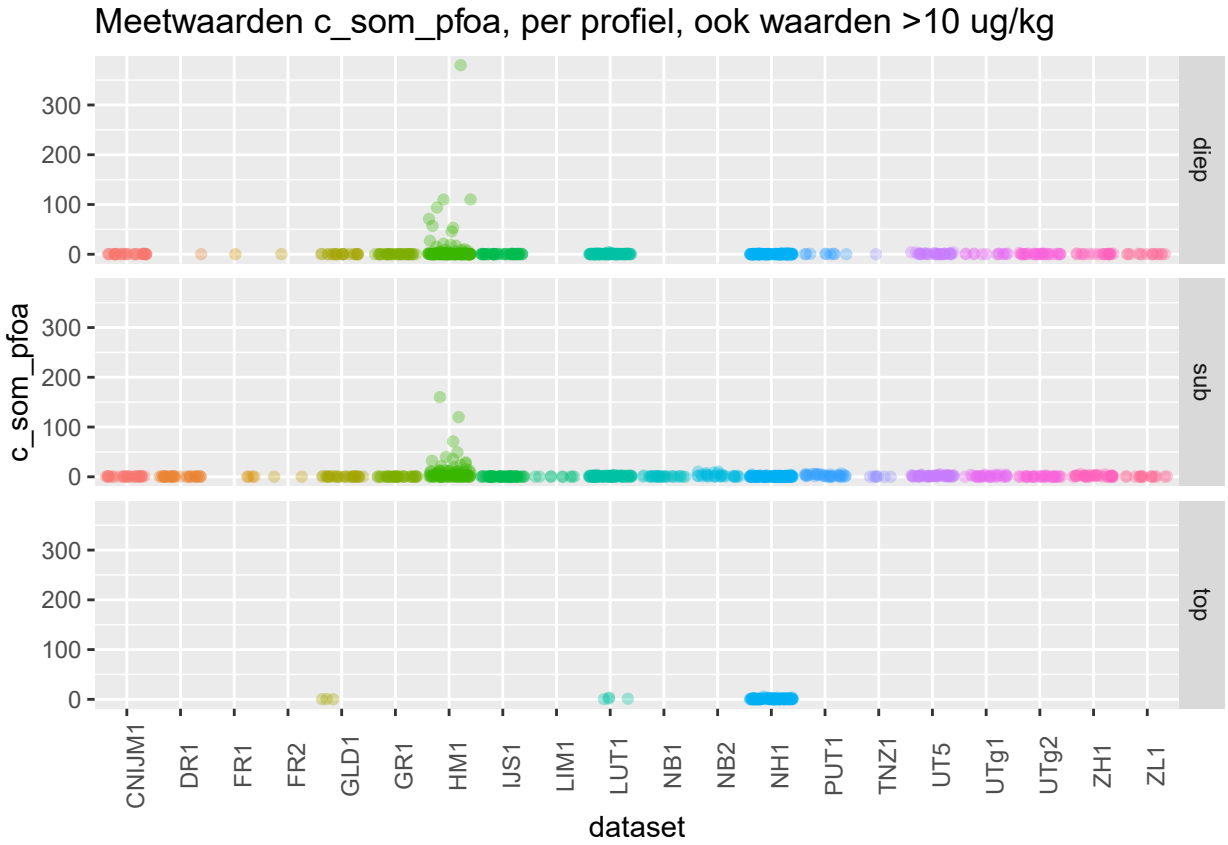
3 Overzicht data

Om een indruk te krijgen van de data maken we een paar visuele overzichten met jitter plots. Alleen de data waarvan de diepte van de meting bekend was, wordt getoond.

Voor de afzonderlijke notitie worden figuren gegenereerd waarbij waarden boven de 10 ug/kg niet worden weergegeven in verband met de visualisatie. In dit document worden zowel figuren gegeven met alle waarden, en alleen de waarden beneden de 10 ug/kg.

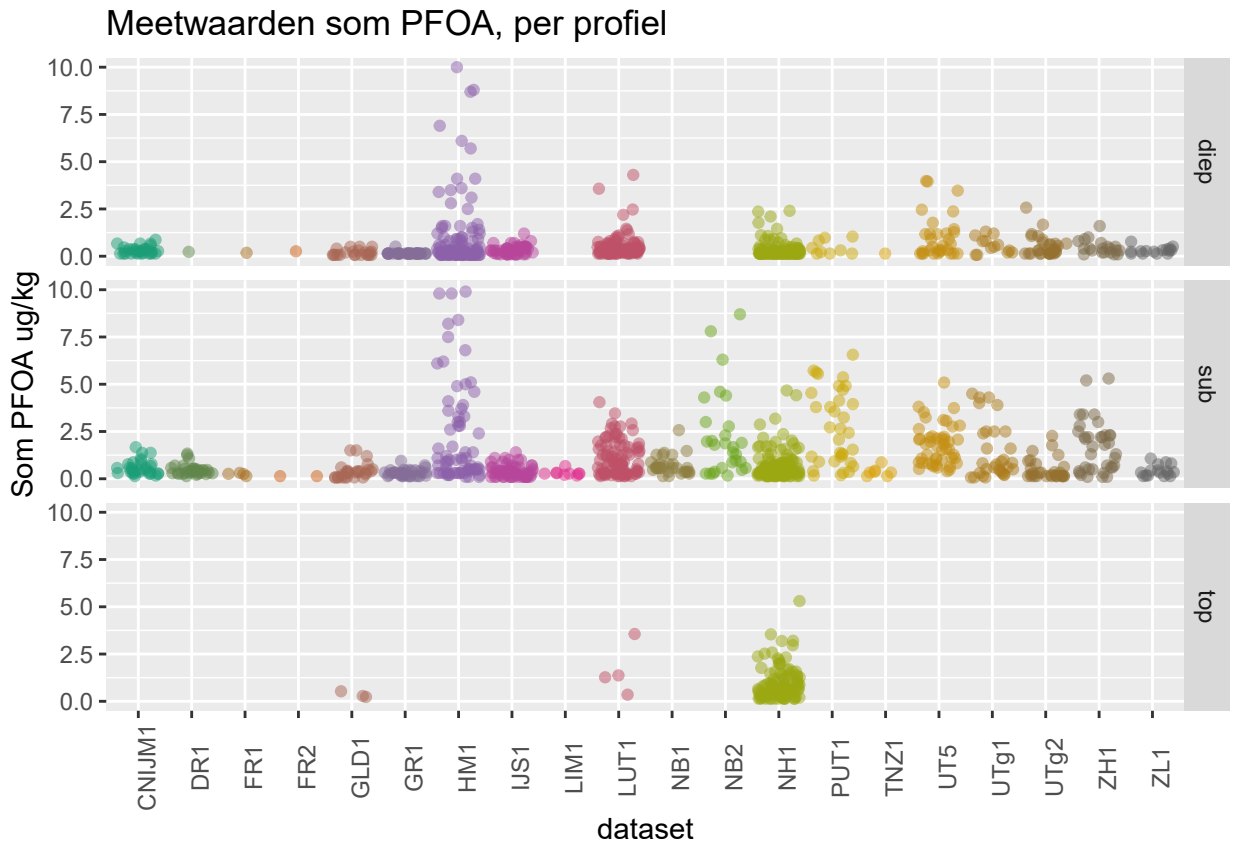
```
somparams <- dvar$somparams[1:2]
d <- readRDS(datafile("pfas_transform.rds")) %>%
    select("dataset","profiel",somparams) %>%
    filter(profiel!="onbekend")

d[somparams] <- sapply(d[somparams],replaceNegative,repl=.7)
colspal  <-  colorRampPalette(brewer.pal(8, "Dark2"))
ncols <- length(unique(d$dataset))

p1 <- qplot(dataset,c_som_pfoa,data=d,facets=profiel~.,geom="jitter",
      alpha=I(1/3),color=dataset) +
    theme(legend.position = "none") +
    theme(axis.text.x = element_text(angle=90))+
    ggtitle("Meetwaarden c_som_pfoa, per profiel, ook waarden >10 ug/kg")
p1
```

```
## Warning: Removed 17 rows containing missing values (geom_point).
```



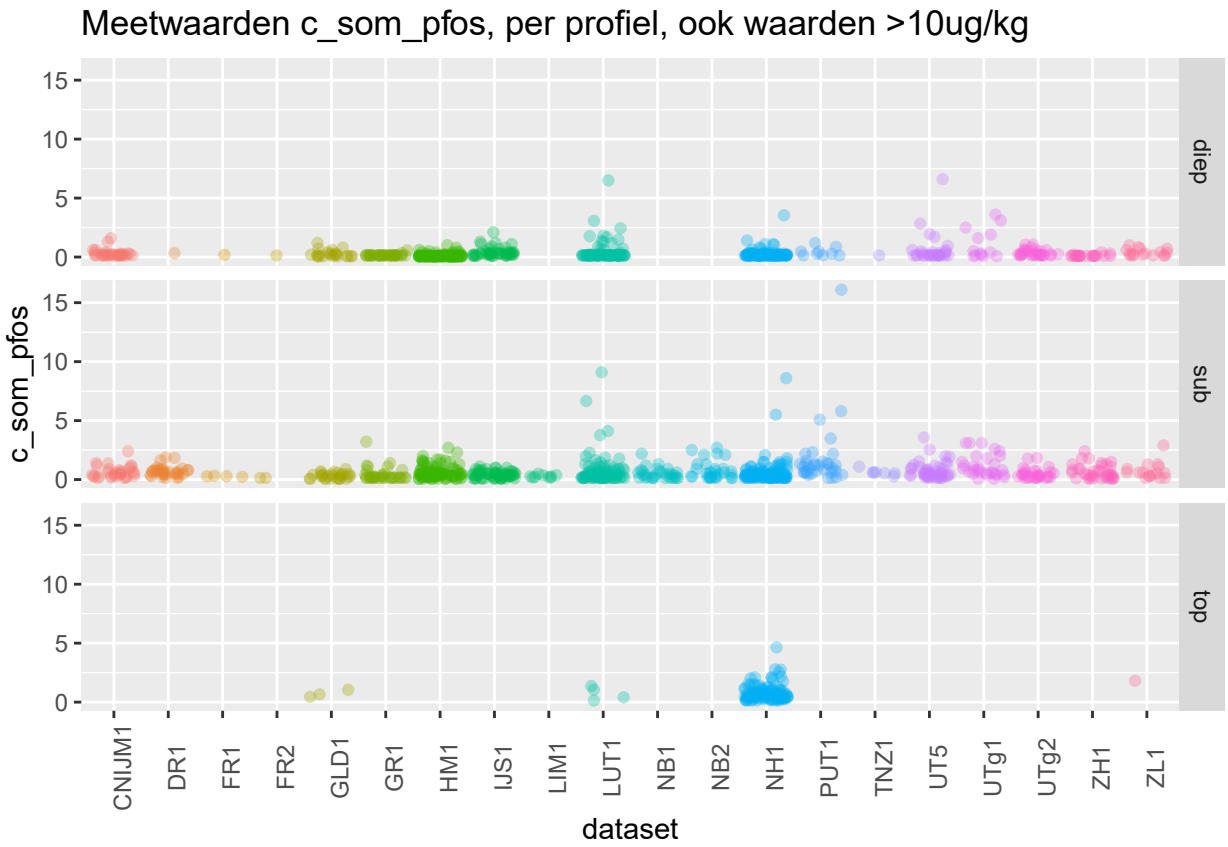

```
p2 <- qplot(dataset,c_som_pfoa,data=d,facets=profiel~.,geom="jitter",
      alpha=I(1/2),color=dataset) +
    theme(legend.position = "none") +
    theme(axis.text.x = element_text(angle=90))+
    scale_colour_manual(values = colspal(ncols)) +
    ylim(c(0,10)) +
    ylab("Som PFOA ug/kg") +
    ggtitle("Meetwaarden som PFOA, per profiel" )
p2
```

```
## Warning: Removed 56 rows containing missing values (geom_point).
```



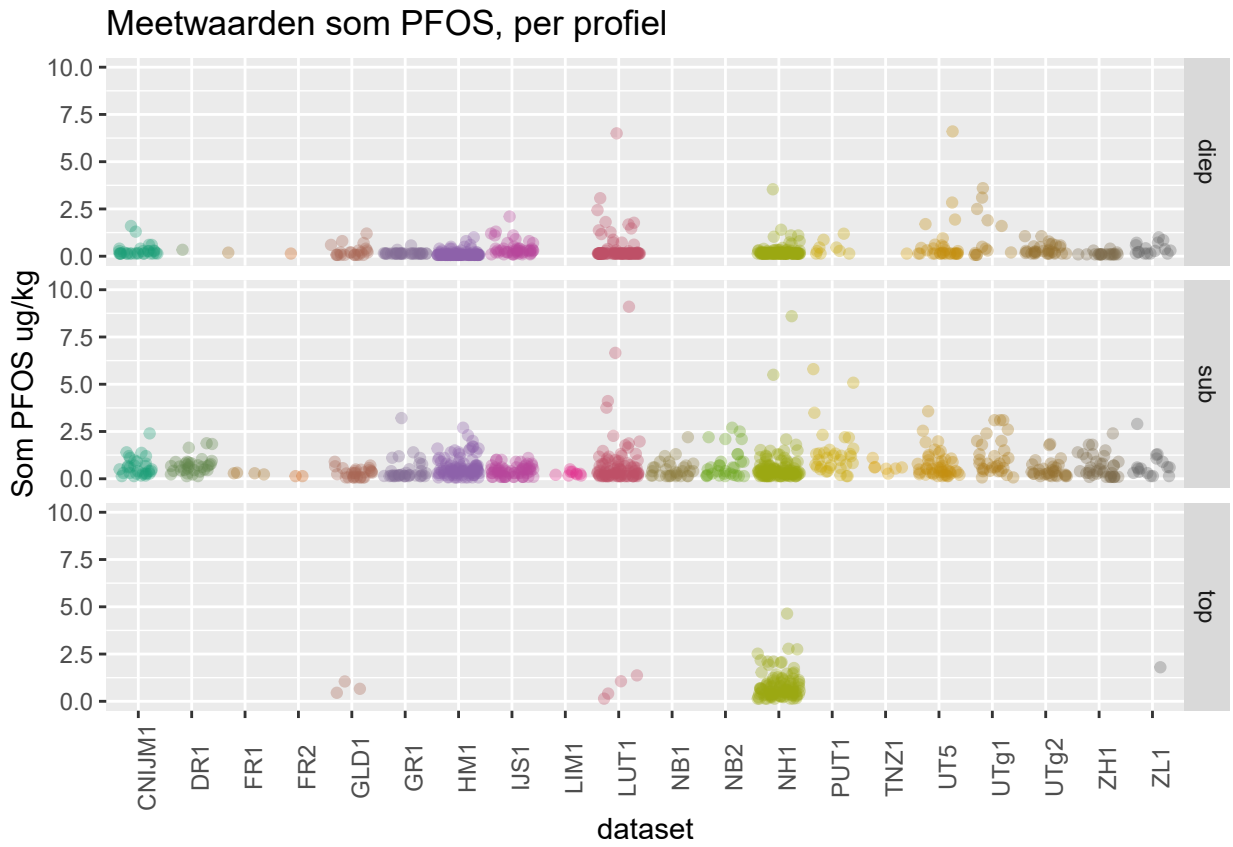

```
p3 <- qplot(dataset,c_som_pfos,data=d,facets=profiel~.,geom="jitter",
      alpha=I(1/3),color=dataset) +
    theme(legend.position = "none") +
    theme(axis.text.x = element_text(angle=90))+
    ggtitle("Meetwaarden c_som_pfos, per profiel, ook waarden >10ug/kg")
p3
```

```
## Warning: Removed 15 rows containing missing values (geom_point).
```




```
p4 <- qplot(dataset,c_som_pfos,data=d,facets=profiel~.,geom="jitter",
      alpha=I(1/3),color=dataset) +
    theme(legend.position = "none") +
    theme(axis.text.x = element_text(angle=90))+
    scale_colour_manual(values = colspal(ncols)) +
    ylim(c(0,10)) +
    ylab("Som PFOS ug/kg") +
    ggtitle("Meetwaarden som PFOS, per profiel")
p4
```

```
## Warning: Removed 16 rows containing missing values (geom_point).
```




```
# geselecteerde figuren opslaan voor rapportage
ggsave(datafile("figJitterPFOAperProfiel.png"),p2,scale=1,dpi="print")
```

```
## Saving 6.5 x 4.5 in image
```

```
## Warning: Removed 56 rows containing missing values (geom_point).
```

```
ggsave(datafile("figJitterPFOSperProfiel.png"),p4,scale=1,dpi="print")
```

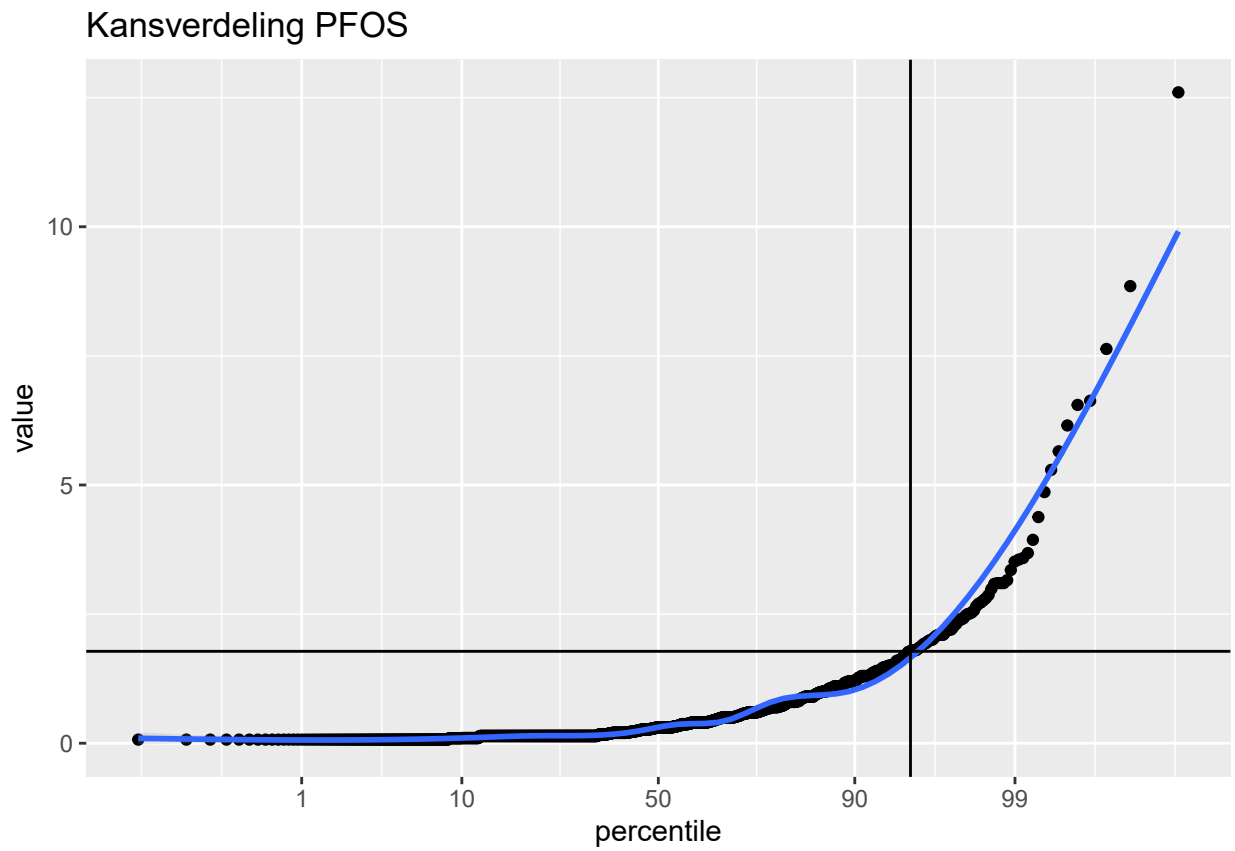
```
## Saving 6.5 x 4.5 in image
```

```
## Warning: Removed 16 rows containing missing values (geom_point).
```

De volgende figuren worden ook opgenomen in de rapportage

```
p1 <- cfp(na.omit(as.data.frame(select(d,"c_som_pfos")))[,1]) +
    ggtitle("Kansverdeling PFOS")
p1
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

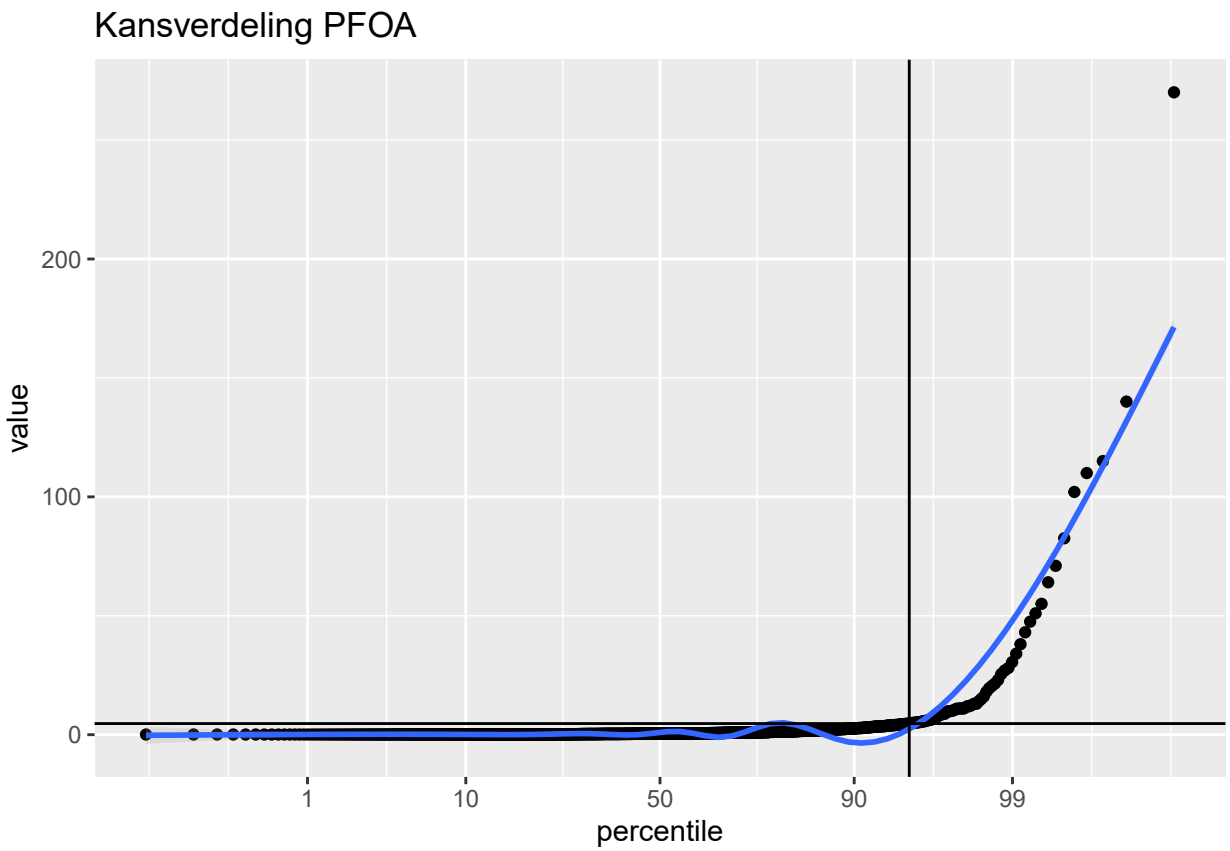



```
ggsave(datafile("figKansverdelingPFOS.png"),p1,scale=1,dpi="print")
```

```
## Saving 6.5 x 4.5 in image
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

```
p2 <- cfp(na.omit(as.data.frame(select(d,"c_som_pfoa")))[,1]) +
    ggtitle("Kansverdeling PFOA")
p2
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



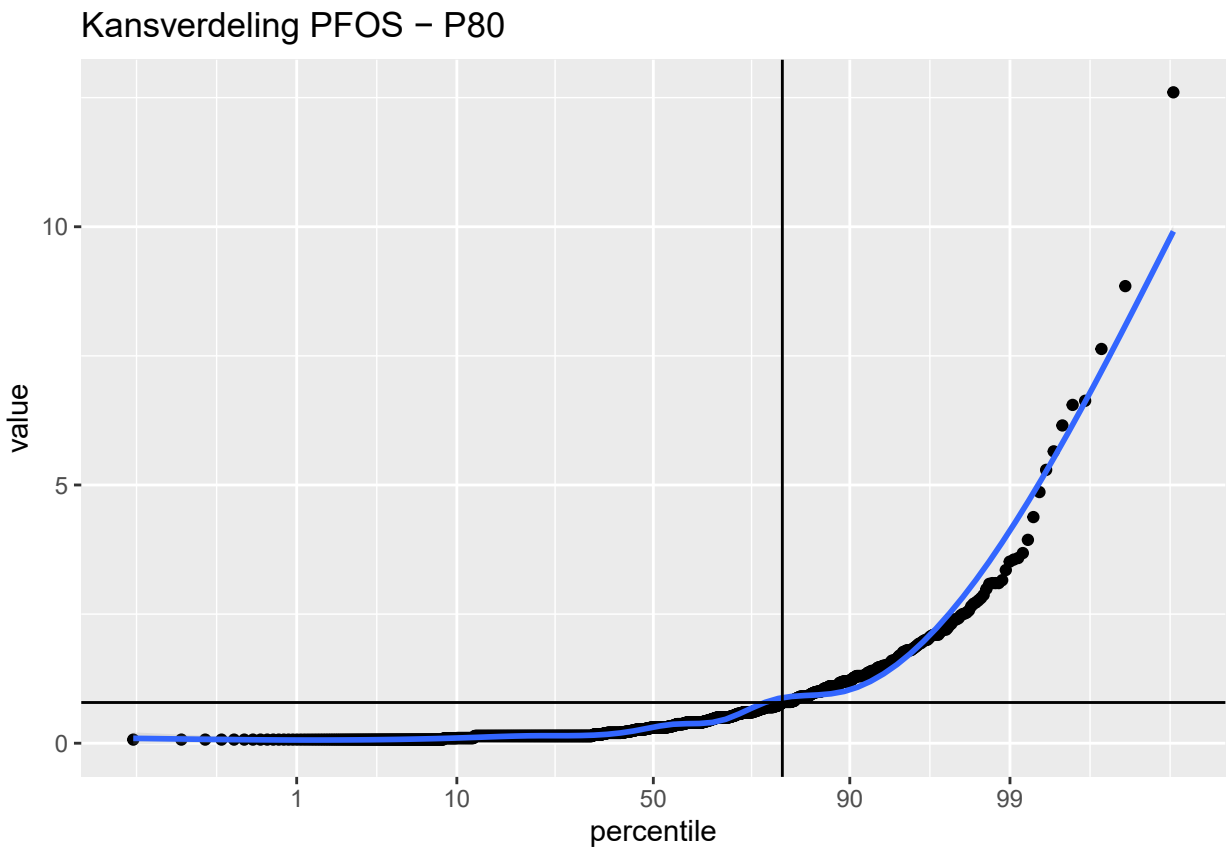

```
ggsave(datafile("figKansverdelingPFOA.png"),p2,scale=1,dpi="print")
```

```
## Saving 6.5 x 4.5 in image
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

Voor controle, dezelfde figuren als hierboven, nu met indicatie 80 percentiel.

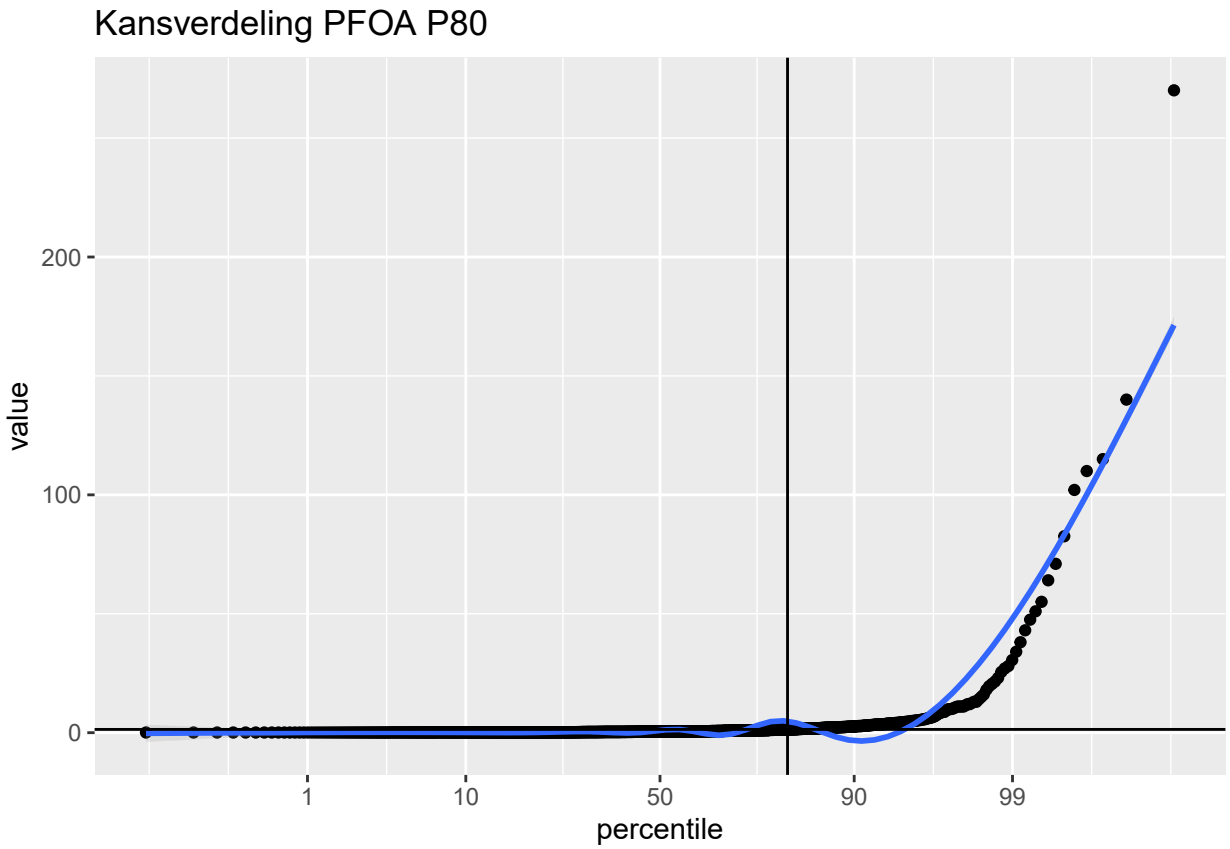
```
cfp(na.omit(as.data.frame(select(d,"c_som_pfos")))[,1],p=.8) +
    ggtitle("Kansverdeling PFOS - P80")
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```




```
cfp(na.omit(as.data.frame(select(d,"c_som_pfoa")))[,1],p=.8) +
    ggtitle("Kansverdeling PFOA P80")
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



Naast figuren maken we ook tabellen met statistische kengetallen van PFOS en PFOA

```
somparams <- dvar$somparams[1:2]
d <- readRDS(datafile("pfas_transform.rds")) %>%
    select("dataset","profiel",somparams)
d[somparams] <- sapply(d[somparams],replaceNegative,repl=.7)

summaryDataset.pfos <- d %>% select(dataset,profiel,c_som_pfos) %>%
    group_by(dataset,profiel) %>%
    summarise(n=n(),
              min=min(c_som_pfos,na.rm=TRUE),
              p25=quantile(c_som_pfos,p=.25,na.rm=TRUE),
              median=quantile(c_som_pfos,p=.5,na.rm=TRUE),
              p75=quantile(c_som_pfos,p=.75,na.rm=TRUE),
              p80=quantile(c_som_pfos,p=.8,na.rm=TRUE),
              max=max(c_som_pfos,na.rm=TRUE))
knitr::kable(summaryDataset.pfos)
```

dataset	profiel	n	min	p25	median	p75	p80	max
CNIJM1	diep	30	0.140	0.14000	0.155	0.2700	0.3200	1.600
CNIJM1	sub	30	0.140	0.27000	0.500	0.8750	0.9960	2.400
DR1	diep	1	0.338	0.33800	0.338	0.3380	0.3380	0.338
DR1	sub	32	0.140	0.45275	0.656	0.8285	0.8832	1.879
FR1	diep	1	0.190	0.19000	0.190	0.1900	0.1900	0.190
FR1	sub	4	0.230	0.27500	0.290	0.2975	0.3020	0.320
FR2	diep	1	0.140	0.14000	0.140	0.1400	0.1400	0.140

dataset	profiel	n	min	p25	median	p75	p80	max
FR2	sub	2	0.140	0.14000	0.140	0.1400	0.1400	0.140
GLD1	diep	22	0.070	0.07000	0.200	0.3000	0.3800	1.200
GLD1	onbekend	3	0.400	0.40000	0.400	0.4000	0.4000	0.400
GLD1	sub	28	0.070	0.07000	0.300	0.4650	0.4860	0.900
GLD1	top	3	0.450	0.55500	0.660	0.8550	0.8940	1.050
GR1	diep	35	0.140	0.14000	0.140	0.1400	0.1400	0.570
GR1	sub	35	0.140	0.14000	0.189	0.4035	0.5572	3.210
HM1	diep	115	0.070	0.07000	0.070	0.0700	0.1200	1.000
HM1	sub	92	0.070	0.30000	0.500	0.8000	0.9800	2.700
IJS1	diep	48	0.100	0.20000	0.300	0.4000	0.4800	2.100
IJS1	sub	59	0.100	0.20000	0.400	0.6000	0.6000	1.100
LIM1	sub	9	0.170	0.21000	0.260	0.3700	0.3700	0.500
LUT1	diep	91	0.140	0.14000	0.140	0.1400	0.2300	6.500
LUT1	sub	89	0.140	0.14000	0.370	0.8200	0.9700	9.100
LUT1	top	4	0.140	0.34250	0.735	1.1375	1.1840	1.370
NB1	sub	34	0.140	0.17000	0.400	0.6750	0.7200	2.200
NB2	sub	40	0.140	0.20000	0.535	0.9000	1.3000	2.700
NH1	diep	113	0.140	0.14000	0.140	0.2100	0.2420	3.540
NH1	sub	113	0.140	0.20000	0.360	0.6300	0.7260	8.600
NH1	top	106	0.140	0.37000	0.605	0.9975	1.1200	4.640
PUT1	diep	9	0.140	0.14000	0.280	0.4600	0.6240	1.190
PUT1	sub	34	0.140	0.64000	1.150	1.4925	1.8260	16.100
TNZ1	diep	1	0.140	0.14000	0.140	0.1400	0.1400	0.140
TNZ1	sub	7	0.270	0.53500	0.600	0.6100	0.6160	1.100
UT5	diep	31	0.140	0.14000	0.140	0.4650	0.5700	6.600
UT5	sub	47	0.140	0.30500	0.490	0.9500	1.0600	3.570
UTg1	diep	13	0.070	0.17000	0.400	1.9000	2.2600	3.600
UTg1	sub	30	0.070	0.51000	0.800	1.7250	2.0000	3.100
UTg2	diep	30	0.140	0.14750	0.245	0.5100	0.5740	1.060
UTg2	sub	30	0.140	0.19750	0.335	0.6250	0.7360	1.840
ZH1	diep	20	0.098	0.09800	0.098	0.1000	0.1000	0.400
ZH1	sub	36	0.098	0.18250	0.420	0.9500	1.2000	2.400
ZL1	diep	16	0.140	0.14000	0.290	0.6125	0.7100	1.000
ZL1	sub	16	0.140	0.30750	0.570	0.9425	1.0700	2.900
ZL1	top	1	1.800	1.80000	1.800	1.8000	1.8000	1.800

```
summaryDataset.pfos %>% write_csv(datafile("summaryDatasetPfos.csv"))

summaryProfiel.pfos <- d %>% select(profiel,c_som_pfos) %>%
    group_by(profiel) %>%
    summarise(n=n(),
              min=min(c_som_pfos,na.rm=TRUE),
              p25=quantile(c_som_pfos,p=.25,na.rm=TRUE),
              median=quantile(c_som_pfos,p=.5,na.rm=TRUE),
              p75=quantile(c_som_pfos,p=.75,na.rm=TRUE),
              p80=quantile(c_som_pfos,p=.8,na.rm=TRUE),
              max=max(c_som_pfos,na.rm=TRUE))
knitr::kable(summaryProfiel.pfos)
```

profiel	n	min	p25	median	p75	p80	max
diep	577	0.07	0.14	0.140	0.2700	0.300	6.60

profiel	n	min	p25	median	p75	p80	max
onbekend	3	0.40	0.40	0.400	0.4000	0.400	0.40
sub	767	0.07	0.20	0.410	0.8000	0.932	16.10
top	114	0.14	0.37	0.615	1.0425	1.148	4.64

```
summaryProfiel.pfos %>% write_csv(datafile("summaryProfielPfos.csv"))


summaryDataset.pfoa <- d %>% select(dataset,profiel,c_som_pfoa) %>%
    group_by(dataset,profiel) %>%
    summarise(n=n(),
              min=min(c_som_pfoa,na.rm=TRUE),
              p25=quantile(c_som_pfoa,p=.25,na.rm=TRUE),
              median=quantile(c_som_pfoa,p=.5,na.rm=TRUE),
              p75=quantile(c_som_pfoa,p=.75,na.rm=TRUE),
              p80=quantile(c_som_pfoa,p=.8,na.rm=TRUE),
              max=max(c_som_pfoa,na.rm=TRUE))
knitr::kable(summaryDataset.pfoa)
```

dataset	profiel	n	min	p25	median	p75	p80	max
CNIJM1	diep	30	0.140	0.14000	0.2700	0.37000	0.390	0.870
CNIJM1	sub	30	0.140	0.27750	0.4700	0.77000	0.790	1.670
DR1	diep	1	0.231	0.23100	0.2310	0.23100	0.231	0.231
DR1	sub	32	0.140	0.29575	0.4005	0.52175	0.537	1.330
FR1	diep	1	0.180	0.18000	0.1800	0.18000	0.180	0.180
FR1	sub	4	0.140	0.22250	0.2550	0.27250	0.280	0.310
FR2	diep	1	0.260	0.26000	0.2600	0.26000	0.260	0.260
FR2	sub	2	0.140	0.14000	0.1400	0.14000	0.140	0.140
GLD1	diep	22	0.070	0.07000	0.0700	0.24000	0.288	0.500
GLD1	onbekend	3	0.070	0.07000	0.0700	0.07000	0.070	0.070
GLD1	sub	28	0.070	0.07000	0.3350	0.43250	0.488	1.500
GLD1	top	3	0.230	0.25500	0.2800	0.40500	0.430	0.530
GR1	diep	35	0.140	0.14000	0.1400	0.14000	0.140	0.510
GR1	sub	35	0.140	0.15750	0.3000	0.40000	0.434	0.950
HM1	diep	115	0.070	0.10000	0.4000	1.60000	3.420	380.000
HM1	sub	92	0.100	0.50000	2.8000	11.00000	11.800	160.000
IJS1	diep	48	0.100	0.12000	0.3000	0.50000	0.500	1.200
IJS1	sub	59	0.100	0.10000	0.3000	0.60000	0.640	1.400
LIM1	sub	9	0.170	0.27000	0.2700	0.29000	0.302	0.670
LUT1	diep	91	0.140	0.19000	0.3300	0.54000	0.670	4.300
LUT1	sub	89	0.140	0.32000	0.8700	1.77000	1.868	4.050
LUT1	top	4	0.350	1.04000	1.3200	1.91750	2.246	3.560
NB1	sub	34	0.140	0.37000	0.5700	0.77000	0.950	2.570
NB2	sub	40	0.170	0.64500	1.8350	3.97500	4.440	10.600
NH1	diep	113	0.140	0.14000	0.1900	0.30000	0.380	2.400
NH1	sub	113	0.140	0.27000	0.4400	0.87000	1.070	4.670
NH1	top	106	0.140	0.36250	0.6700	1.27000	1.470	5.300
PUT1	diep	9	0.140	0.14000	0.3100	0.83000	0.890	1.040
PUT1	sub	34	0.170	0.98250	2.5400	4.43750	4.790	6.560
TNZ1	diep	1	0.140	0.14000	0.1400	0.14000	0.140	0.140
TNZ1	sub	7	0.140	0.23500	0.3300	0.40000	0.418	0.870
UT5	diep	31	0.140	0.18500	0.5600	1.23500	1.430	3.980

dataset	profiel	n	min	p25	median	p75	p80	max
UT5	sub	47	0.400	0.92500	1.7100	2.20000	2.554	5.090
UTg1	diep	13	0.070	0.20000	0.5000	0.80000	0.980	1.300
UTg1	sub	30	0.070	0.32500	0.7850	2.47500	2.500	4.500
UTg2	diep	30	0.140	0.14000	0.3350	0.66750	0.760	2.570
UTg2	sub	30	0.140	0.14000	0.2050	0.60750	0.820	2.270
ZH1	diep	20	0.098	0.20000	0.4000	0.55000	0.720	1.600
ZH1	sub	36	0.098	0.57500	1.7500	2.30000	2.500	5.300
ZL1	diep	16	0.140	0.19000	0.2700	0.34000	0.370	0.770
ZL1	sub	16	0.140	0.17000	0.3400	0.52000	0.760	1.070
ZL1	top	1	Inf	NA	NA	NA	NA	-Inf

```
summaryDataset.pfoa %>% write_csv(datafile("summaryDatasetPfoa.csv"))

summaryProfiel.pfoa <- d %>% select(profiel,c_som_pfoa) %>%
    group_by(profiel) %>%
    summarise(n=n(),
              min=min(c_som_pfoa,na.rm=TRUE),
              p25=quantile(c_som_pfoa,p=.25,na.rm=TRUE),
              median=quantile(c_som_pfoa,p=.5,na.rm=TRUE),
              p75=quantile(c_som_pfoa,p=.75,na.rm=TRUE),
              p80=quantile(c_som_pfoa,p=.8,na.rm=TRUE),
              max=max(c_som_pfoa,na.rm=TRUE))
knitr::kable(summaryProfiel.pfoa)
```

profiel	n	min	p25	median	p75	p80	max
diep	577	0.07	0.14	0.27	0.515	0.700	380.00
onbekend	3	0.07	0.07	0.07	0.070	0.070	0.07
sub	767	0.07	0.30	0.57	1.570	1.966	160.00
top	114	0.14	0.35	0.67	1.270	1.450	5.30

```
summaryProfiel.pfoa %>% write_csv(datafile("summaryProfielPfoa.csv"))
```

4 Berekening relatie met Organisch Stof

Eerst selecteren we de paramers die iets met organisch stof (OM) te maken hebben en alle PFAS stoffen, als groepen selecteren we de dataset en profiel.

```
d <- readRDS(datafile("pfas_transform.rds"))
somparams <- dvar$somparams[1:2]
omparams <- c("om","totc","gloei")
params <- c(somparams,omparams)

d[params] <- sapply(d[params],replaceNegative,repl=.7)
d[params] <- sapply(d[params],replaceNegative,repl=.7)
d.sel <- d %>% select(dataset,profiel,omparams,somparams) %>%
    filter(profiel!="onbekend")
d.sel
```

```
## # A tibble: 1,458 x 7
```

```
##    dataset profiel     om  totc gloei c_som_pfoa c_som_pfos
##    <chr>   <chr>    <dbl> <dbl> <dbl>      <dbl>      <dbl>
##  1 NH1     top     10.3    NA    NA        1.17       0.6
##  2 NH1     sub     12.4    NA    NA        1.72       0.55
##  3 NH1     diep    54.5    NA    NA        1.77       0.29
##  4 NH1     top      1.7    NA    NA        0.39       0.21
##  5 NH1     sub      1.9    NA    NA        0.44       0.220
##  6 NH1     diep     0.35   NA    NA        0.140      0.140
##  7 NH1     top      1.1    98.8   6.3      0.140      0.67
##  8 NH1     sub      0.490  99.5   7.7      0.37       0.7
##  9 NH1     diep     0.490  99.6   8.1      0.67       0.3
## 10 NH1     top     31.3    NA    NA        3.19       1.42
## # ... with 1,448 more rows
```

```
d.sel %>% select(omparams) %>% summary()
```

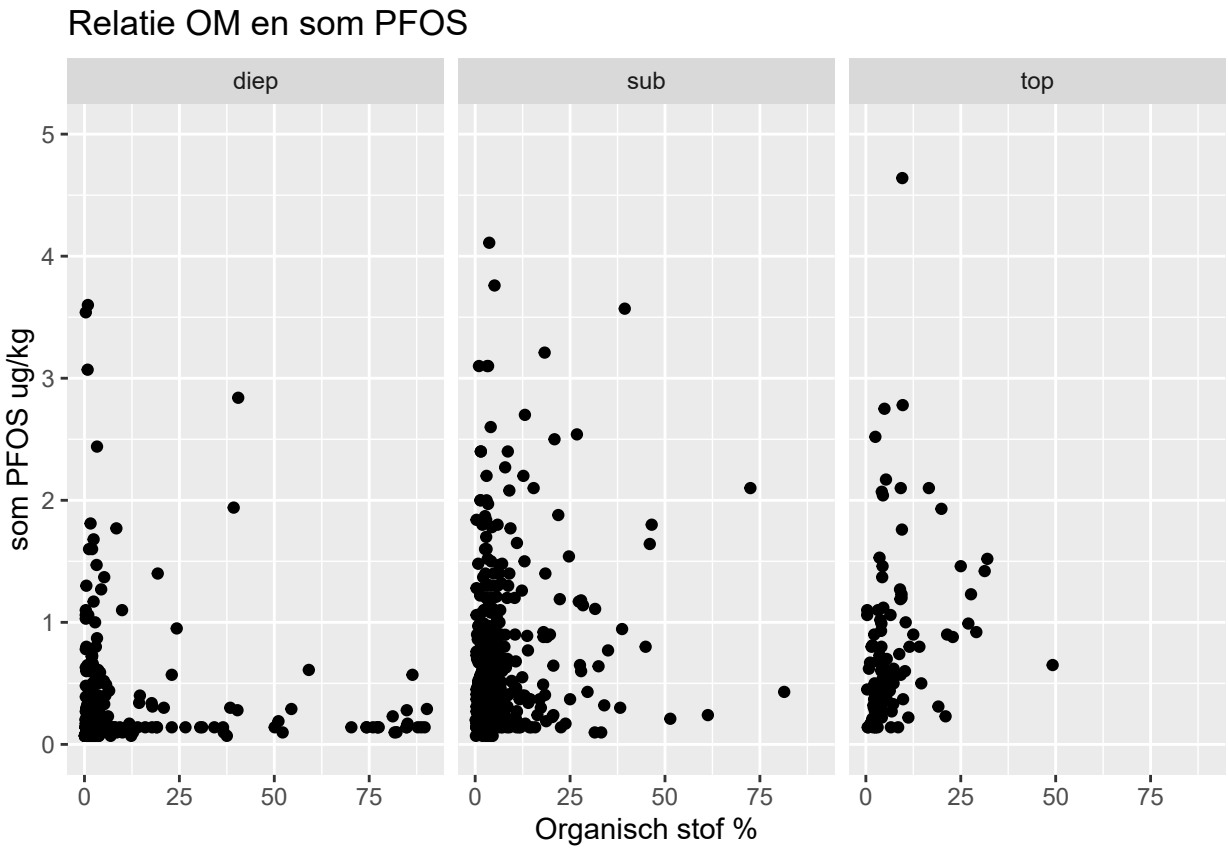
```
##        om              totc           gloei
##  Min.   : 0.140   Min.   : 3.50   Min.   : 4.60
##  1st Qu.: 1.500   1st Qu.:10.75   1st Qu.: 8.00
##  Median : 3.000   Median :21.00   Median :94.20
##  Mean   : 7.078   Mean   :44.84   Mean   :69.38
##  3rd Qu.: 5.600   3rd Qu.:91.10   3rd Qu.:97.45
##  Max.   :90.200   Max.   :99.70   Max.   :99.70
##  NA's   :395      NA's   :1418    NA's   :1403
```

De parameters 'totc' en 'gloei' hebben te veel missing values voor een analyse. We beperken ons tot 'om'.

Vervolgens maken we een eerste plot met overzicht van de relatie OM en c_som_pfos. We verkorten de x-as tot de range 0-5 in verband met aanwezigheid outliers.

```
p <-  qplot(om,c_som_pfos,data=d.sel,facets=.~profiel) +
    ylim(c(0,5)) +
    xlab("Organisch stof %") + ylab("som PFOS ug/kg")+
    ggtitle("Relatie OM en som PFOS")
p
```

```
## Warning: Removed 403 rows containing missing values (geom_point).
```



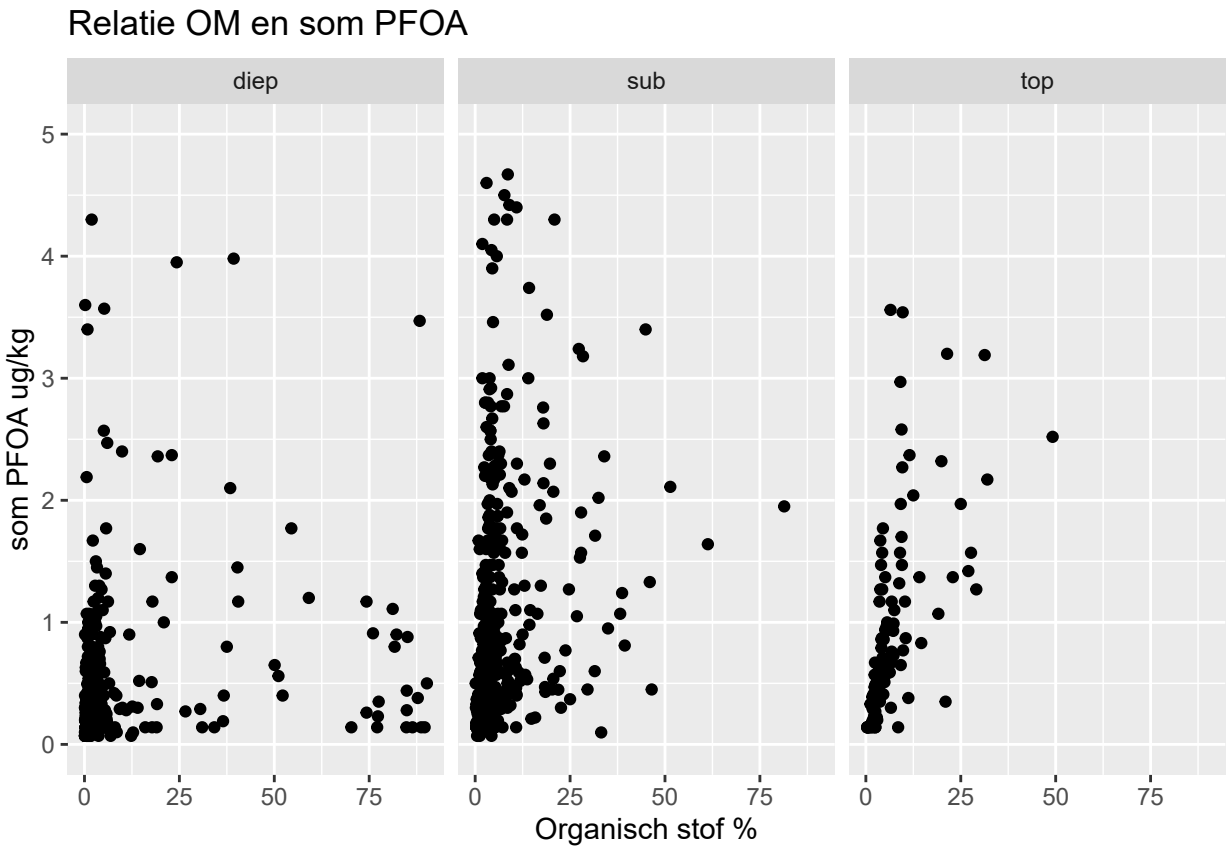

```
ggsave(datafile("figOMvsPFOS.png"),p,scale=1,dpi="print")
```

```
## Saving 6.5 x 4.5 in image
```

```
## Warning: Removed 403 rows containing missing values (geom_point).
```

```
p <-  qplot(om,c_som_pfoa,data=d.sel,facets=.~profiel) +
    ylim(c(0,5)) +
    xlab("Organisch stof %") + ylab("som PFOA ug/kg")+
    ggtitle("Relatie OM en som PFOA")
p
```

```
## Warning: Removed 418 rows containing missing values (geom_point).
```




```
ggsave(datafile("figOMvsPFOA.png"),p,scale=1,dpi="print")
```

```
## Saving 6.5 x 4.5 in image
```

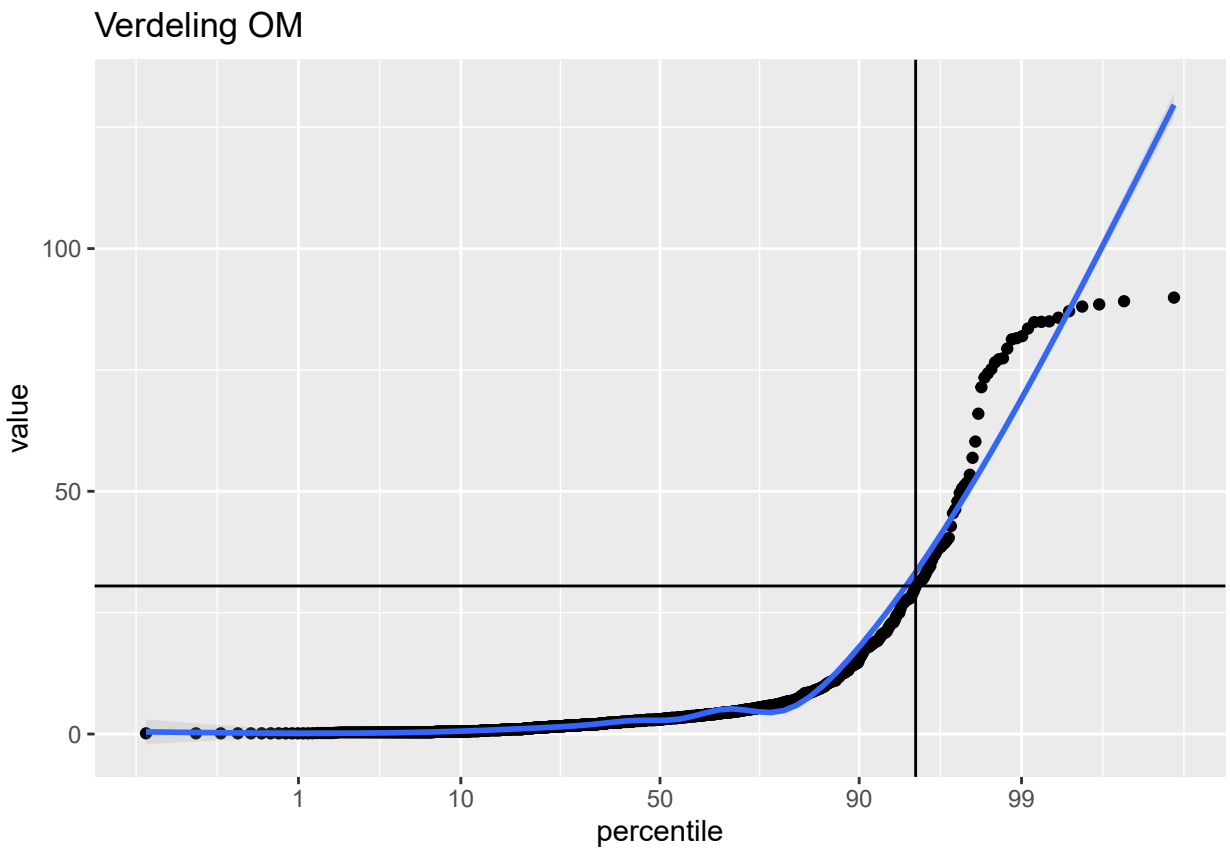
```
## Warning: Removed 418 rows containing missing values (geom_point).
```

Op basis van een visuele inspectie is er geen reden om aan te nemen dat er een relatie bestaat tussen om en som PFOS of som PFOA.

We proberen enkele regressie methoden uit om te kijken of we een model kunnen opstellen voor de relatie om en som PFOS. We maken een standaard OLS regressie model en een robuust LQS model.

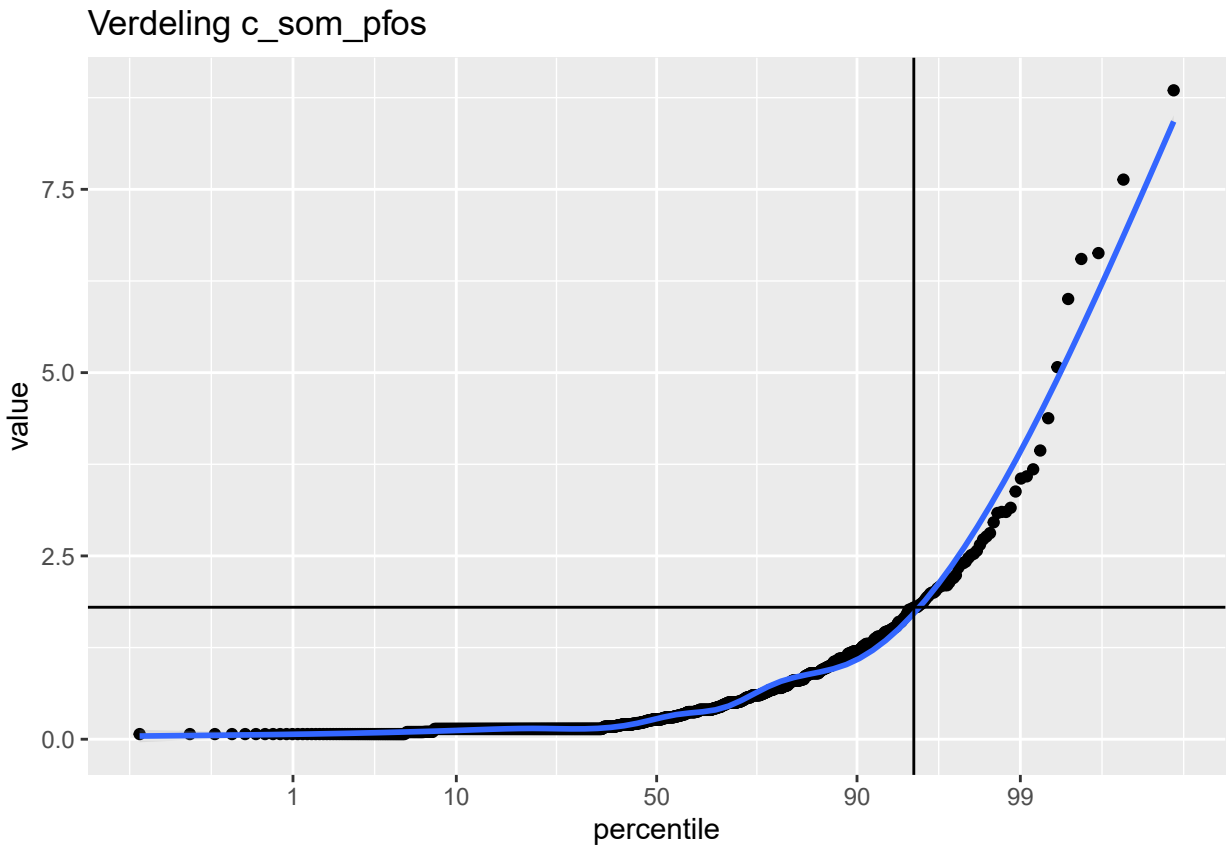
```
ompfos <- d.sel %>% select(profiel,om,c_som_pfos) %>%
    filter(c_som_pfos>0) %>%
    na.omit()
cfp(ompfos$om) + ggtitle("Verdeling OM")
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```




```
cfp(ompfos$c_som_pfos) + ggtitle("Verdeling c_som_pfos")
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```




```
m <- lm(c_som_pfos~profiel/om,data=ompfos)

summary(m)
```

```
##
## Call:
## lm(formula = c_som_pfos ~ profiel/om, data = ompfos)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.1731 -0.3258 -0.1442  0.0258  8.4830
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.273721   0.039294   6.966 5.73e-12 ***
## profielsub     0.306827   0.055664   5.512 4.46e-08 ***
## profieltop     0.396904   0.104348   3.804 0.000151 ***
## profieldiep:om 0.003612   0.001945   1.857 0.063544 .
## profielsub:om  0.012562   0.003631   3.459 0.000563 ***
## profieltop:om  0.018491   0.008818   2.097 0.036244 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7413 on 1055 degrees of freedom
## Multiple R-squared:  0.07948,    Adjusted R-squared:  0.07512
## F-statistic: 18.22 on 5 and 1055 DF,  p-value: < 2.2e-16
```

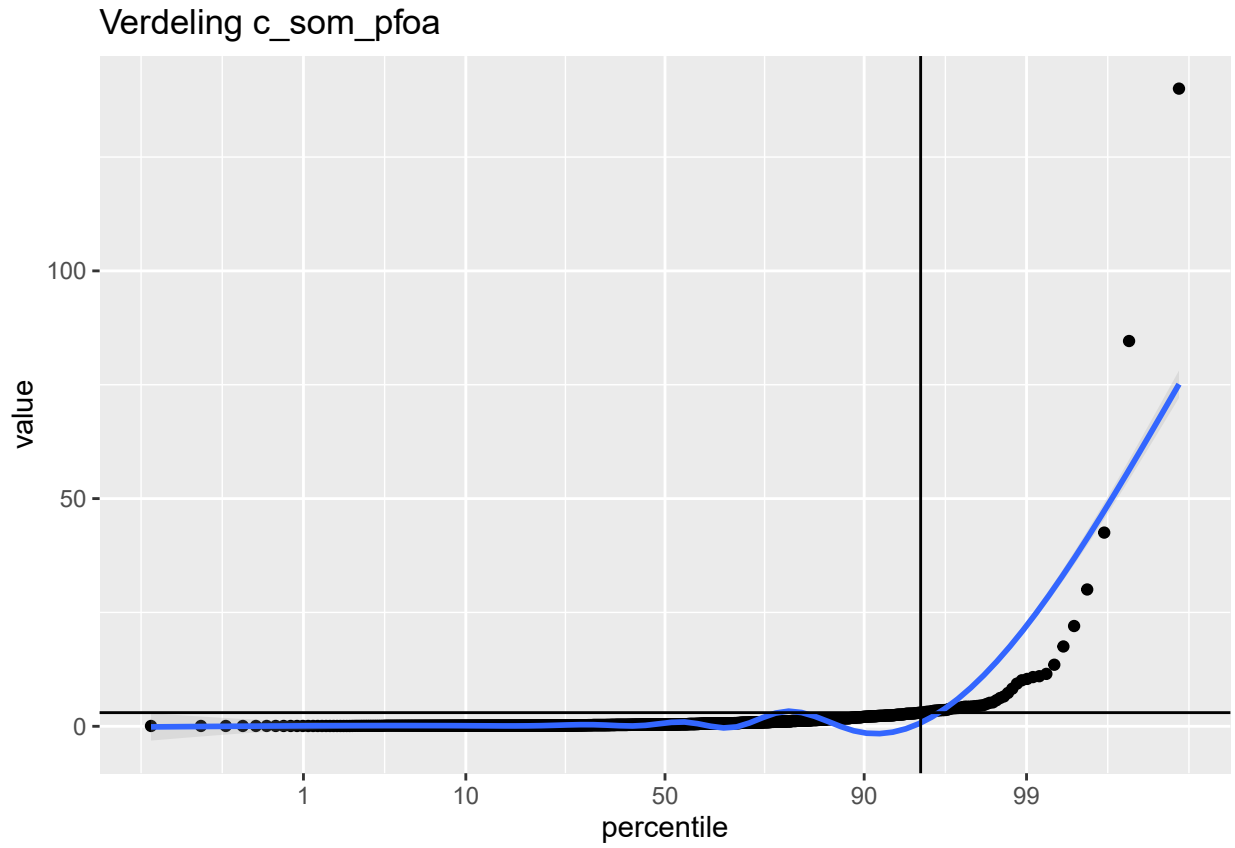
```
m<-lqs(c_som_pfos~profiel/om,data=ompfos)
m
```

```
## Call:
## lqs.formula(formula = c_som_pfos ~ profiel/om, data = ompfos)
##
## Coefficients:
##    (Intercept)      profielsub      profieltop  profieldiep:om   profielsub:om
##      1.462e-01      -3.333e-02       1.556e-01       4.786e-16       8.130e-03
##  profieltop:om
##      2.552e-02
##
## Scale estimates 0.1161 0.1087
```

De output van beide modellen laten zien dat het model weinig relevant is. Het OLS model heeft een verwaarloosbare verklaarde variantie. In het robuuste model is de richtingcoefficient van OM erg klein, met name voor de diepere lagen.

```
ompfoa <- d.sel %>% select(profiel,om,c_som_pfoa) %>%
    filter(c_som_pfoa>0) %>%
    na.omit()
ompfoa$c_som_pfoa <- replaceNegative(ompfoa$c_som_pfoa,0.7)
cfp(ompfoa$c_som_pfoa) + ggtitle("Verdeling c_som_pfoa")
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```




```
m <- lm(c_som_pfoa~profiel/om,data=ompfoa)
summary(m)
```

```
##
## Call:
## lm(formula = c_som_pfoa ~ profiel/om, data = ompfoa)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -1.930  -1.267  -0.315  -0.074 158.166
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.436368   0.346701   1.259  0.20844
## profielsub     1.379488   0.491139   2.809  0.00507 **
## profieltop     0.064490   0.920690   0.070  0.94417
## profieldiep:om 0.005555   0.017158   0.324  0.74618
## profielsub:om  0.006388   0.032039   0.199  0.84201
## profieltop:om  0.063498   0.077806   0.816  0.41463
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.54 on 1055 degrees of freedom
## Multiple R-squared:  0.01064,    Adjusted R-squared:  0.005956
## F-statistic:  2.27 on 5 and 1055 DF,  p-value: 0.04563
```

```
m<-lqs(c_som_pfoa~profiel/om,data=ompfoa)
m
```

```
## Call:
## lqs.formula(formula = c_som_pfoa ~ profiel/om, data = ompfoa)
##
## Coefficients:
##    (Intercept)      profielsub      profieltop  profieldiep:om   profielsub:om
##      1.856e-01      -1.021e-01       9.411e-01       1.913e-16       1.458e-01
##  profieltop:om
##     -2.222e-01
##
## Scale estimates 0.2904 0.2834
```

Voor PFOA lijken de modellen iets relevanter, de verklaarde variantie op basis van OLS ligt tussen de 10 en 20%. Het robuuste model laat echter nog steeds een minimale richtingscoefficient zien.

Nog een optie is om de modellen op basis van log-transformatie van de data op te stellen

```
m.pfoa <- lm(log(c_som_pfoa)~profiel/log(om),data=ompfoa)
m.pfoa
```

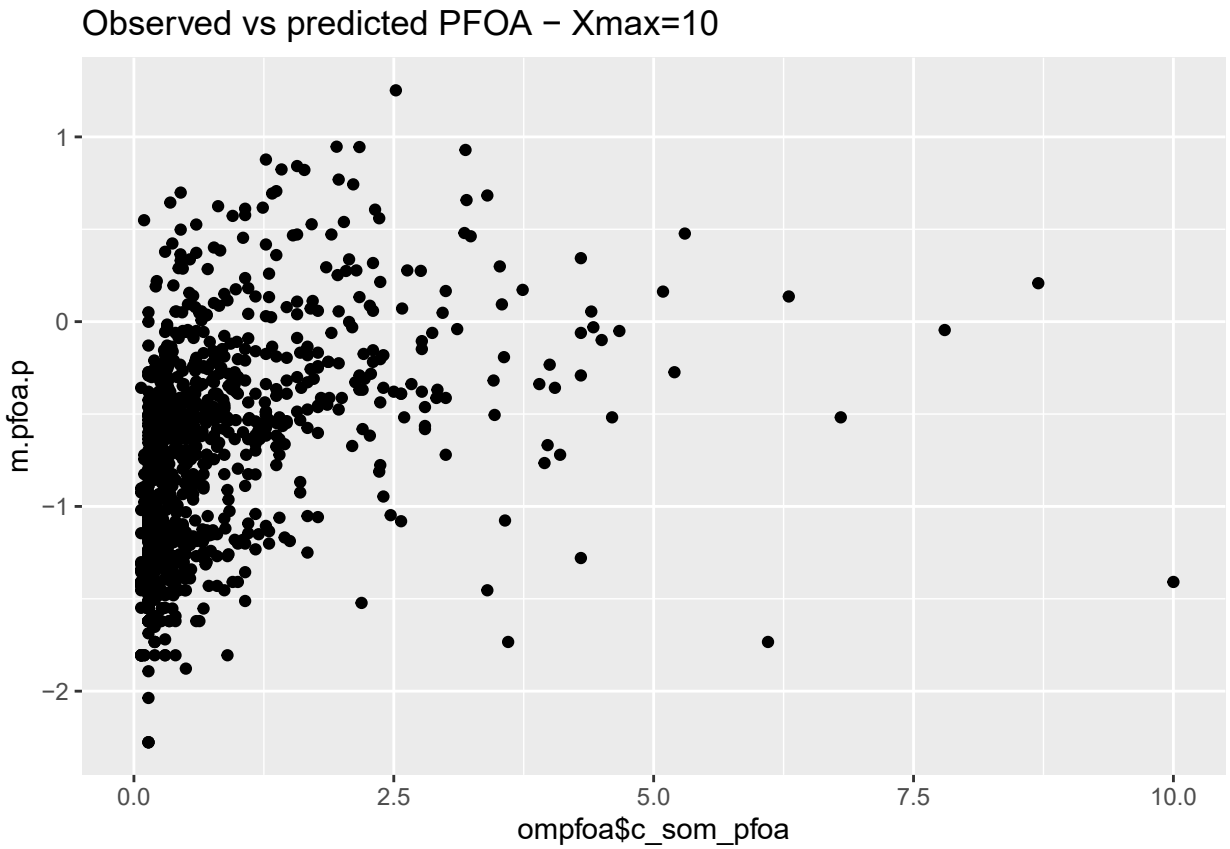
```
##
## Call:
## lm(formula = log(c_som_pfoa) ~ profiel/log(om), data = ompfoa)
##
## Coefficients:
##         (Intercept)           profielsub           profieltop
##             -1.4089               0.4035              -0.1188
## profieldiep:log(om)   profielsub:log(om)   profieltop:log(om)
##              0.2017               0.4438               0.7135
```

```r
summary(m.pfoa)
```

```
##
## Call:
## lm(formula = log(c_som_pfoa) ~ profiel/log(om), data = ompfoa)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8718 -0.6271 -0.1053  0.5051  5.6080
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -1.40889    0.05032 -28.000  < 2e-16 ***
## profielsub           0.40352    0.08214   4.913 1.04e-06 ***
## profieltop          -0.11877    0.17578  -0.676    0.499
## profieldiep:log(om)  0.20174    0.03028   6.663 4.30e-11 ***
## profielsub:log(om)   0.44378    0.03959  11.209  < 2e-16 ***
## profieltop:log(om)   0.71349    0.09132   7.813 1.34e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9223 on 1055 degrees of freedom
## Multiple R-squared:  0.2925, Adjusted R-squared:  0.2892
## F-statistic: 87.25 on 5 and 1055 DF,  p-value: < 2.2e-16
```

```r
m.pfoa.p <- predict(m.pfoa)

qplot(ompfoa$c_som_pfoa,m.pfoa.p) +
    xlim(0,10) +
    ggtitle("Observed vs predicted PFOA - Xmax=10")
```

```
## Warning: Removed 12 rows containing missing values (geom_point).
```




```
m.pfos <- lm(log(c_som_pfos)~profiel/(om),data=ompfos)
m.pfos
```

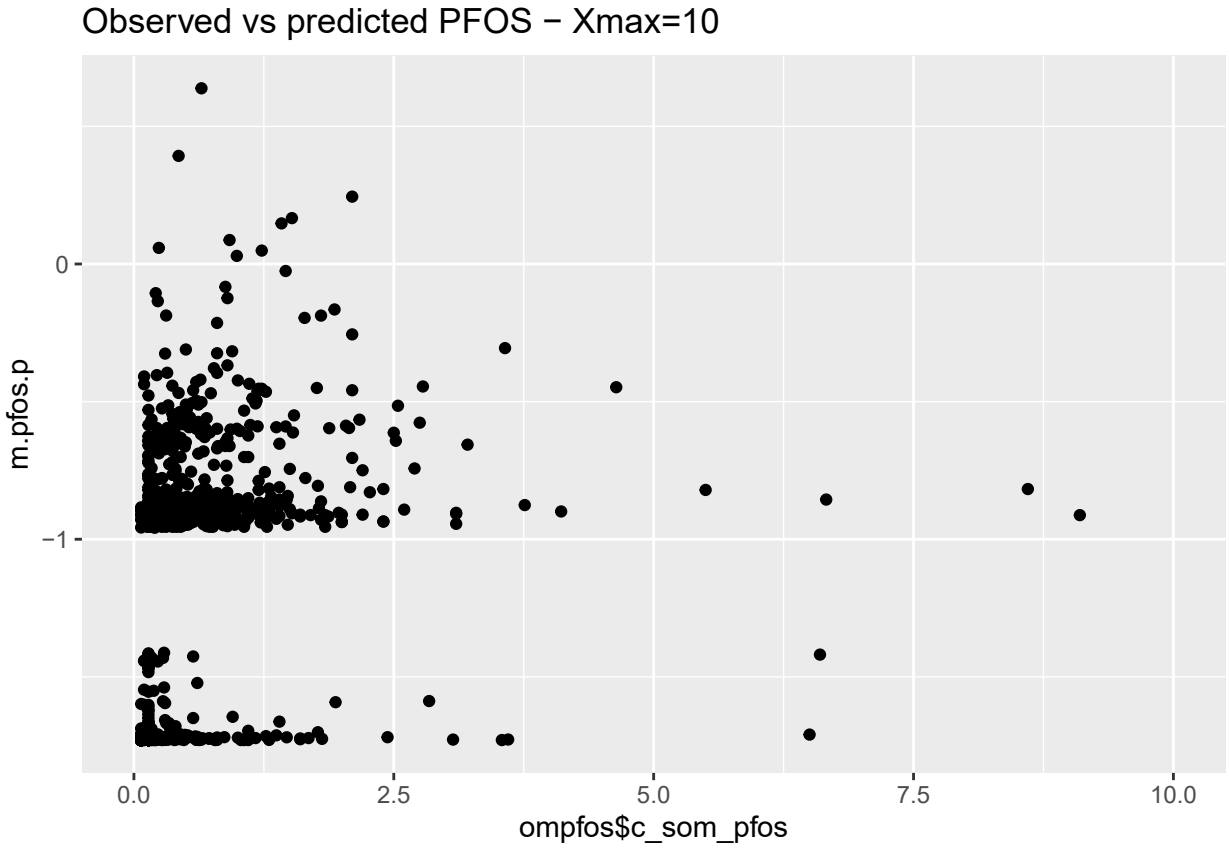
```
##
## Call:
## lm(formula = log(c_som_pfos) ~ profiel/(om), data = ompfos)
##
## Coefficients:
##    (Intercept)      profielsub      profieltop  profieldiep:om   profielsub:om
##      -1.730802        0.770206        1.019801        0.003529        0.016625
##  profieltop:om
##       0.027424
```

```
summary(m.pfos)
```

```
##
## Call:
## lm(formula = log(c_som_pfos) ~ profiel/(om), data = ompfos)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9141 -0.5338 -0.2365  0.4775  3.5814
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.730802   0.044052 -39.290  < 2e-16 ***
## profielsub      0.770206   0.062404  12.342  < 2e-16 ***
```

```
## profieltop      1.019801   0.116983   8.718  < 2e-16 ***
## profieldiep:om  0.003529   0.002180   1.619  0.10579
## profielsub:om   0.016625   0.004071   4.084 4.76e-05 ***
## profieltop:om   0.027424   0.009886   2.774  0.00563 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.831 on 1055 degrees of freedom
## Multiple R-squared:  0.2468, Adjusted R-squared:  0.2432
## F-statistic: 69.12 on 5 and 1055 DF,  p-value: < 2.2e-16
```

```
m.pfos.p <- predict(m.pfos)
qplot(ompfos$c_som_pfos,m.pfos.p) +
    xlim(0,10) +
    ggtitle("Observed vs predicted PFOS - Xmax=10")
```



De modellen lijken wel relevanter te worden in termen van verklaarde variantie. Echter inspectie van de richtingscoefficienten en observed vs fitted values geven een grote spreiding aan in de predictie.

5 Data in relatie tot landgebruik en bronnen

5.1 Natuur en landbouw

Data punten met een coordinaat zijn gebruikt om een overlay te maken met de LGN7 landgebruikskaart en de bodemkaart. Ook zijn de afstanden tot bekende bronnen (Chemours, brandweer oefenplaatsen etc.) te

berekenen.

Het landgebruik en afstand tot bronnen is opgenomen als aparte velden in het bronbestand. De Afleiding van deze velden heeft buiten deze analyse plaatsgevonden.

We visualiseren de data onderverdeeld in landgebruikscategorieen:

```
meta <- readRDS(datafile("pfas_tidy.rds")) %>%
    select("id",dvar$cats,dvar$catsnum) %>%
    na.omit

d <- readRDS(datafile("pfas_transform.rds")) %>%
    filter(id%in%meta$id) %>%
    right_join(meta)
```

```
## Joining, by = "id"
```

```
colspal  <-  colorRampPalette(brewer.pal(8, "Dark2"))
ncols <- length(unique(d$dataset))


p1 <- qplot(as.factor(use_cat),c_som_pfos,data=d,geom="jitter",
      color=dataset,alpha=I(1/3)) +
    xlab("Landgebruik") + ylab("Som PFOS ug/kg")+
    scale_colour_manual(values = colspal(ncols)) +
    ggtitle("Som PFOS per landgebruikscategorie")

p2 <- qplot(as.factor(use_cat),c_som_pfoa,data=d,geom="jitter",
      color=dataset,alpha=I(1/3)) +
    xlab("Landgebruik") + ylab("Som PFOA ug/kg")+
    scale_colour_manual(values = colspal(ncols)) +
    ggtitle("Som PFOA per landgebruikscategorie")

ggsave(datafile("fitJitterPFOSperLandgebruik.png"),p1,scale=1,dpi="print")
```

```
## Saving 6.5 x 4.5 in image
```

```
## Warning: Removed 15 rows containing missing values (geom_point).
```

```
ggsave(datafile("fitJitterPFOAperLandgebruik.png"),p2,scale=1,dpi="print")
```

```
## Saving 6.5 x 4.5 in image
```

```
## Warning: Removed 17 rows containing missing values (geom_point).
```

5.2 Bodemtype

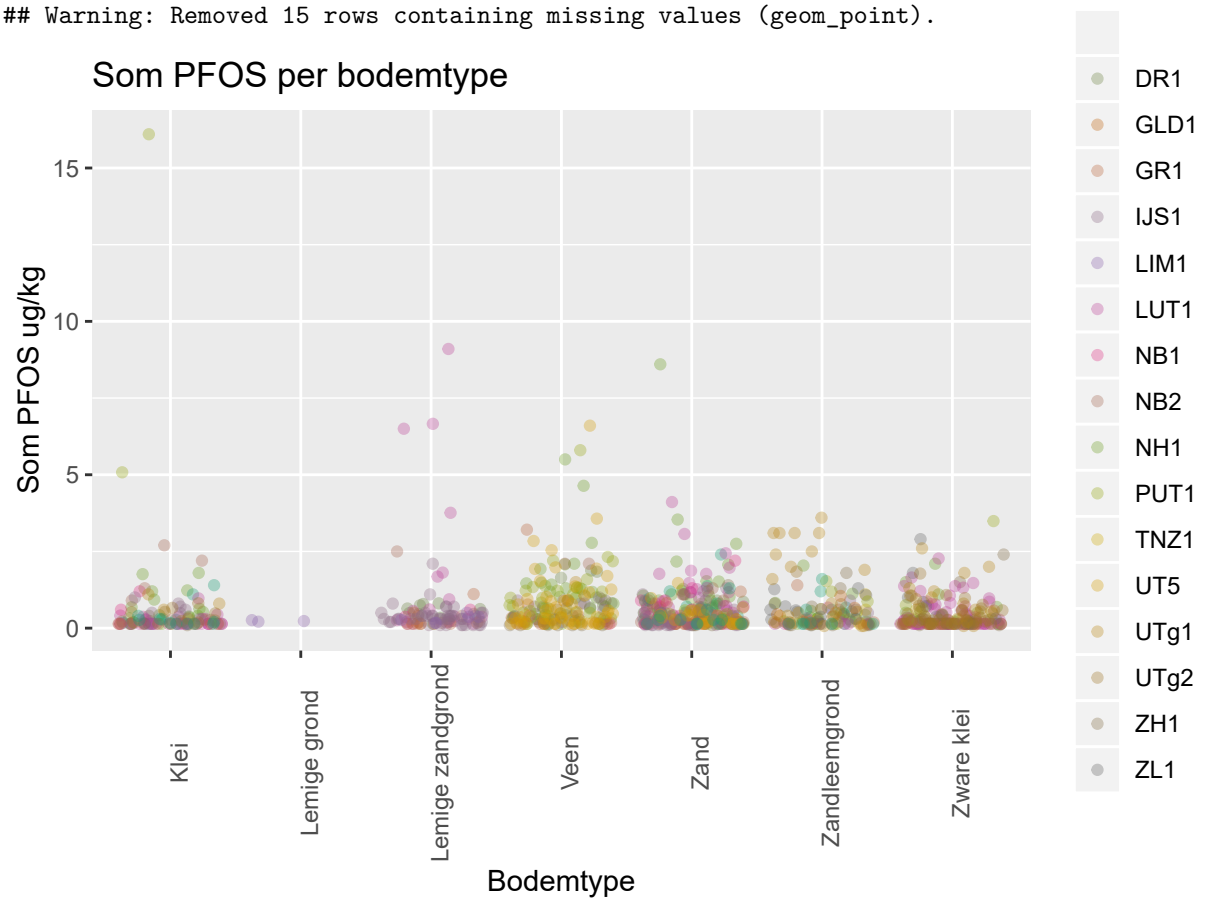
Visualisatie van data onderverdeeld naar bodemtype

```
meta <- readRDS(datafile("pfas_tidy.rds")) %>%
    select("id",dvar$cats,dvar$catsnum) %>%
    na.omit

d <- readRDS(datafile("pfas_transform.rds")) %>%
    filter(id%in%meta$id) %>%
    right_join(meta)
```

```
## Joining, by = "id"
```

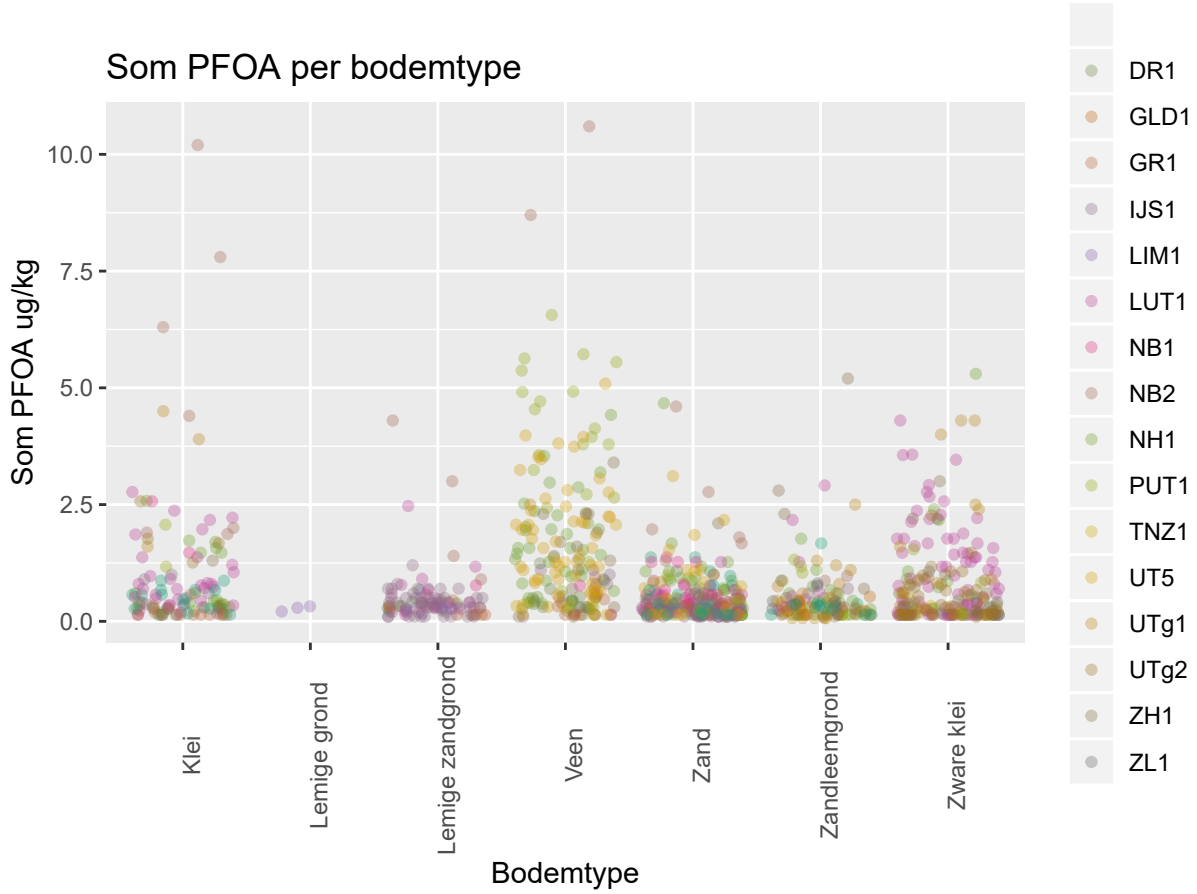
```
colspal  <-  colorRampPalette(brewer.pal(8, "Dark2"))
ncols <- length(unique(d$dataset))


p1 <- qplot(as.factor(type),c_som_pfos,data=d,geom="jitter",
      color=dataset,alpha=I(1/3)) +
    scale_colour_manual(values = colspal(ncols)) +
    theme(axis.text.x = element_text(angle=90))+
    xlab("Bodemtype") + ylab("Som PFOS ug/kg")+
    ggtitle("Som PFOS per bodemtype")

p1
```

```
## Warning: Removed 15 rows containing missing values (geom_point).
```




```
p2 <- qplot(as.factor(type),c_som_pfoa,data=d,geom="jitter",
      color=dataset,alpha=I(1/3)) +
    xlab("Bodemtype") + ylab("Som PFOA ug/kg")+
    theme(axis.text.x = element_text(angle=90))+
    scale_colour_manual(values = colspal(ncols)) +
    ggtitle("Som PFOA per bodemtype")
p2
```

```
## Warning: Removed 17 rows containing missing values (geom_point).
```




```
ggsave(datafile("fitJitterPFOSperBodemtype.png"),p1,scale=1,dpi="print")
```

```
## Saving 6.5 x 4.5 in image
```

```
## Warning: Removed 15 rows containing missing values (geom_point).
```

```
ggsave(datafile("fitJitterPFOAperBodemtype.png"),p2,scale=1,dpi="print")
```

```
## Saving 6.5 x 4.5 in image
```

```
## Warning: Removed 17 rows containing missing values (geom_point).
```

5.3 Invloed bronnen: Chemours

Visualisatie van de data met de afstand tot Chemours

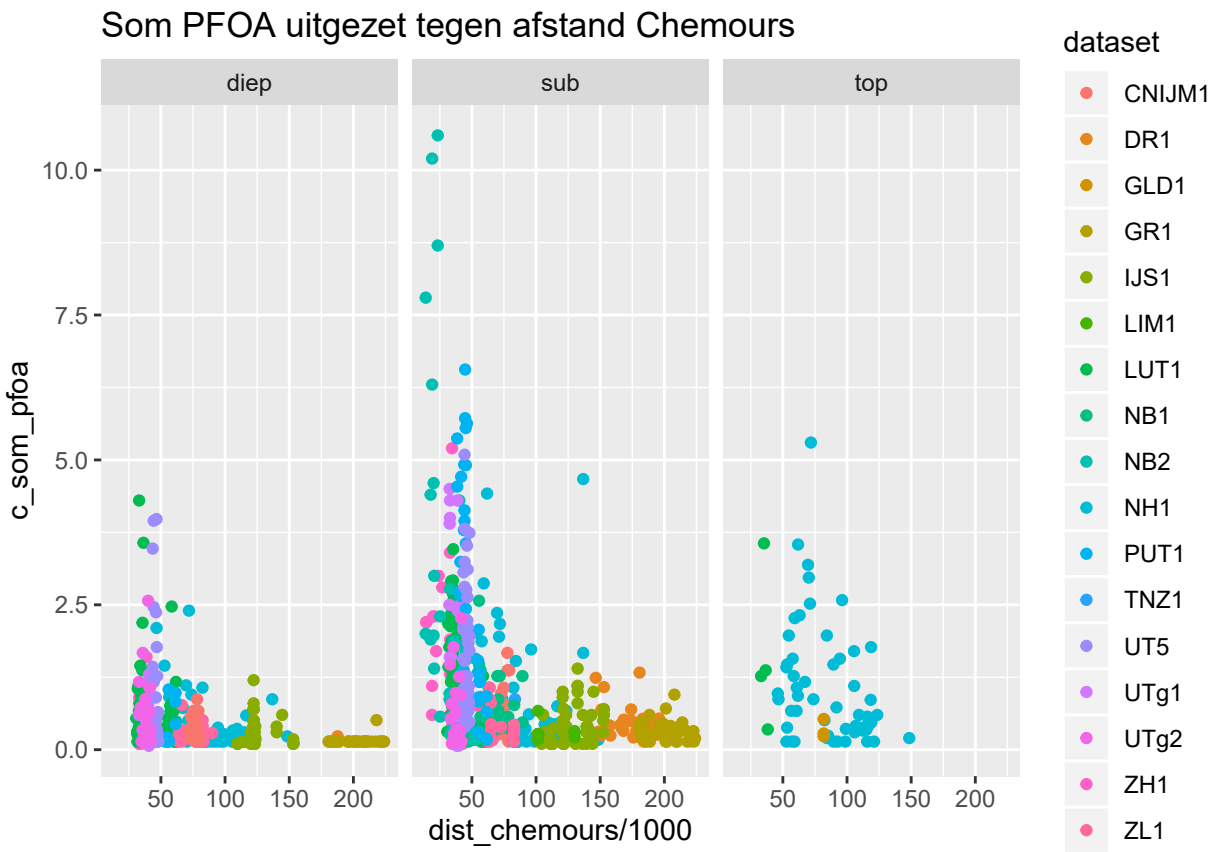
```
meta <- readRDS(datafile("pfas_tidy.rds")) %>%
    select("id",dvar$cats,dvar$catsnum) %>%
    na.omit

d <- readRDS(datafile("pfas_transform.rds")) %>%
    filter(id%in%meta$id) %>%
    right_join(meta)
```

```
## Joining, by = "id"
```

```
qplot(dist_chemours/1000,c_som_pfoa,data=d,color=dataset,
      facets=.~profiel) +
```

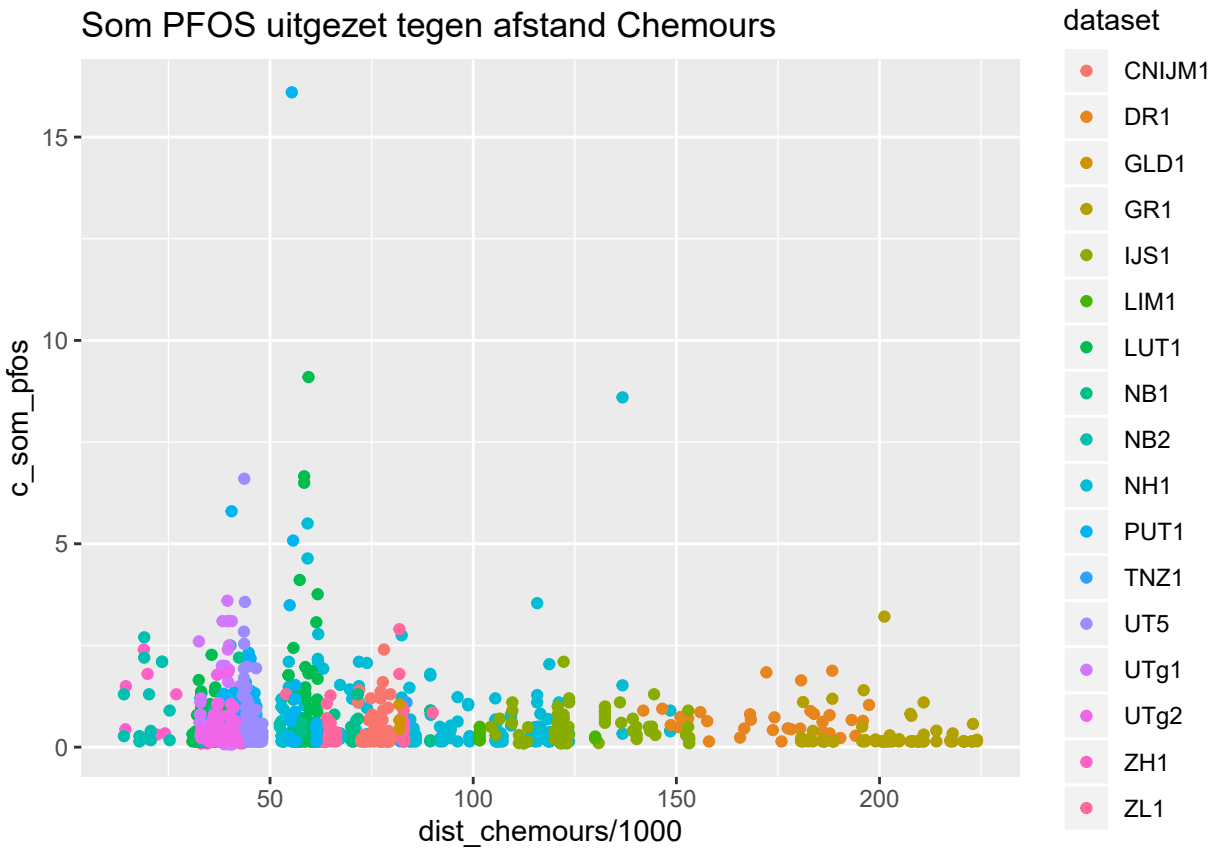
```
    ggtitle("Som PFOA uitgezet tegen afstand Chemours")
```

```
## Warning: Removed 17 rows containing missing values (geom_point).
```




```
qplot(dist_chemours/1000,c_som_pfos,data=d,color=dataset) +
    ggtitle("Som PFOS uitgezet tegen afstand Chemours")
```

```
## Warning: Removed 15 rows containing missing values (geom_point).
```



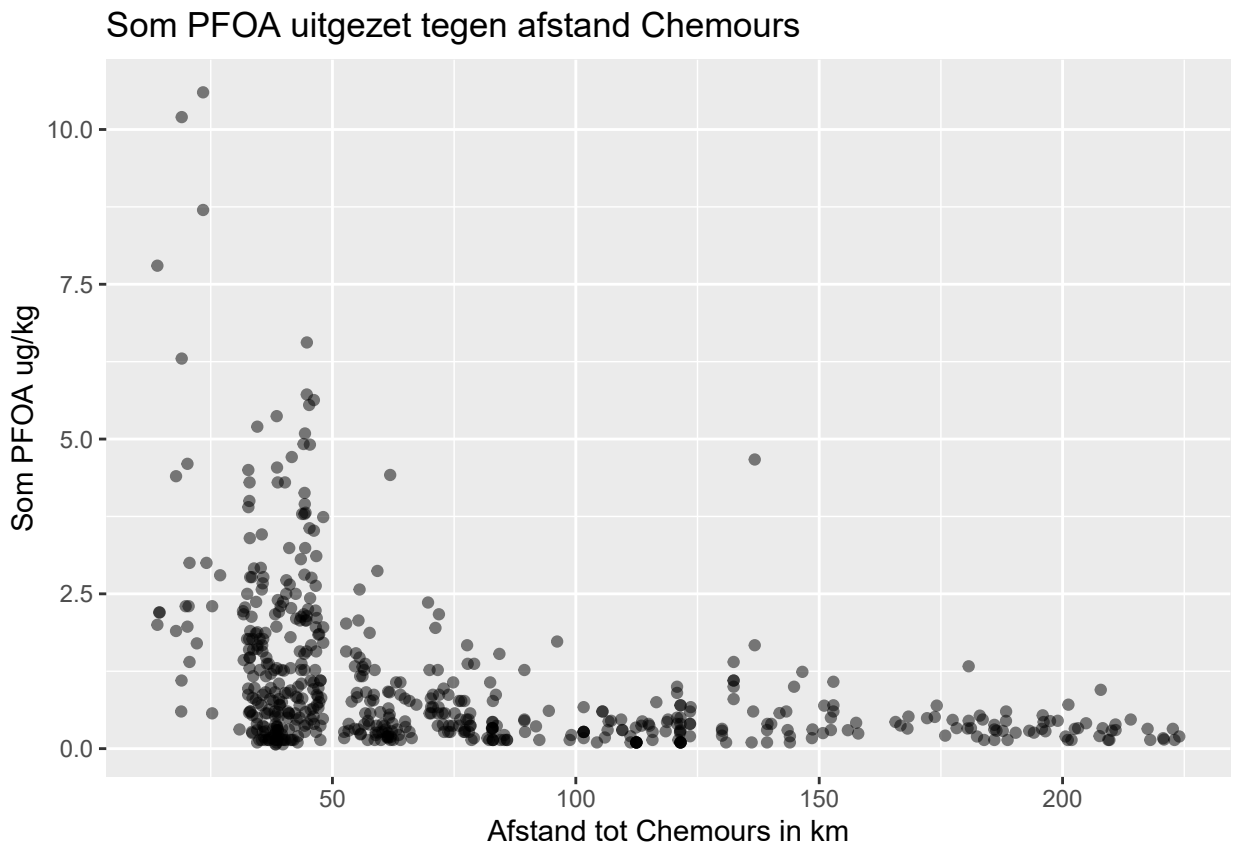
Maken figuur voor de rapportage

```
meta <- readRDS(datafile("pfas_tidy.rds")) %>%
    select("id",dvar$cats,dvar$catsnum) %>%
    na.omit

d <- readRDS(datafile("pfas_transform.rds")) %>%
    filter(id%in%meta$id) %>%
    right_join(meta) %>%
    filter(profiel=="sub")
```

```
## Joining, by = "id"
```

```
colspal  <-  colorRampPalette(brewer.pal(8, "Dark2"))
ncols <- length(unique(d$dataset))

p1 <- qplot(dist_chemours/1000,c_som_pfoa,data=d,alpha=I(1/2)) +
    xlab("Afstand tot Chemours in km") +
    ylab("Som PFOA ug/kg") +
    scale_colour_manual(values = colspal(ncols)) +
    ggtitle("Som PFOA uitgezet tegen afstand Chemours")
p1
```

```
## Warning: Removed 14 rows containing missing values (geom_point).
```



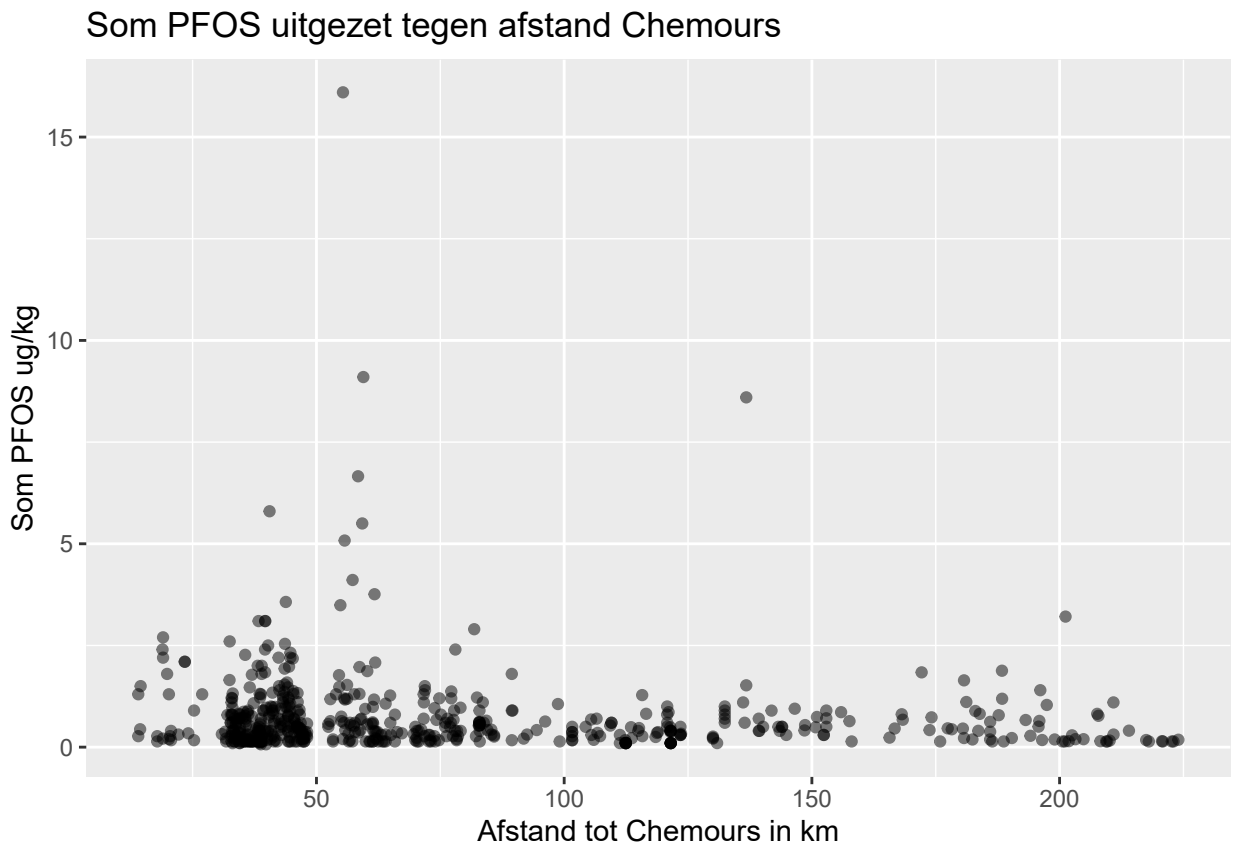

```
ggsave(datafile("figAfstandChemoursPFOA.png"),p1,scale=1,dpi="print")
```

```
## Saving 6.5 x 4.5 in image
```

```
## Warning: Removed 14 rows containing missing values (geom_point).
```

```
p2 <- qplot(dist_chemours/1000,c_som_pfos,data=d,alpha=I(1/2)) +
    xlab("Afstand tot Chemours in km") +
    ylab("Som PFOS ug/kg") +
    scale_colour_manual(values = colspal(ncols)) +
    ggtitle("Som PFOS uitgezet tegen afstand Chemours")
p2
```

```
## Warning: Removed 14 rows containing missing values (geom_point).
```




```
ggsave(datafile("figAfstandChemoursPFOS.png"),p2,scale=1,dpi="print")
```

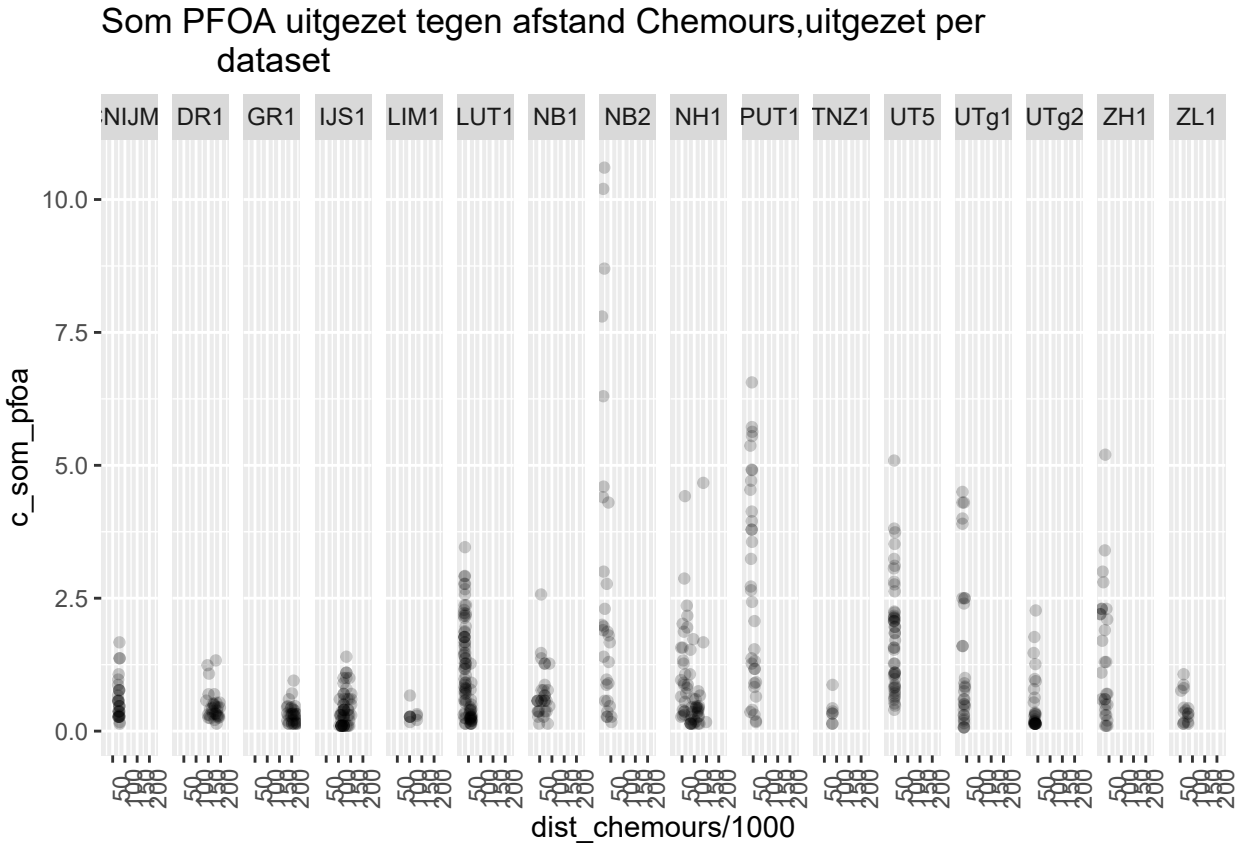
```
## Saving 6.5 x 4.5 in image
```

```
## Warning: Removed 14 rows containing missing values (geom_point).
```

Welke datasets zijn beinvloed:

```
qplot(dist_chemours/1000,c_som_pfoa,data=d,alpha=I(2/10),
      facets=.~dataset) +
    theme(axis.text.x = element_text(angle=90)) +
    ggtitle("Som PFOA uitgezet tegen afstand Chemours,uitgezet per
            dataset")
```

```
## Warning: Removed 14 rows containing missing values (geom_point).
```

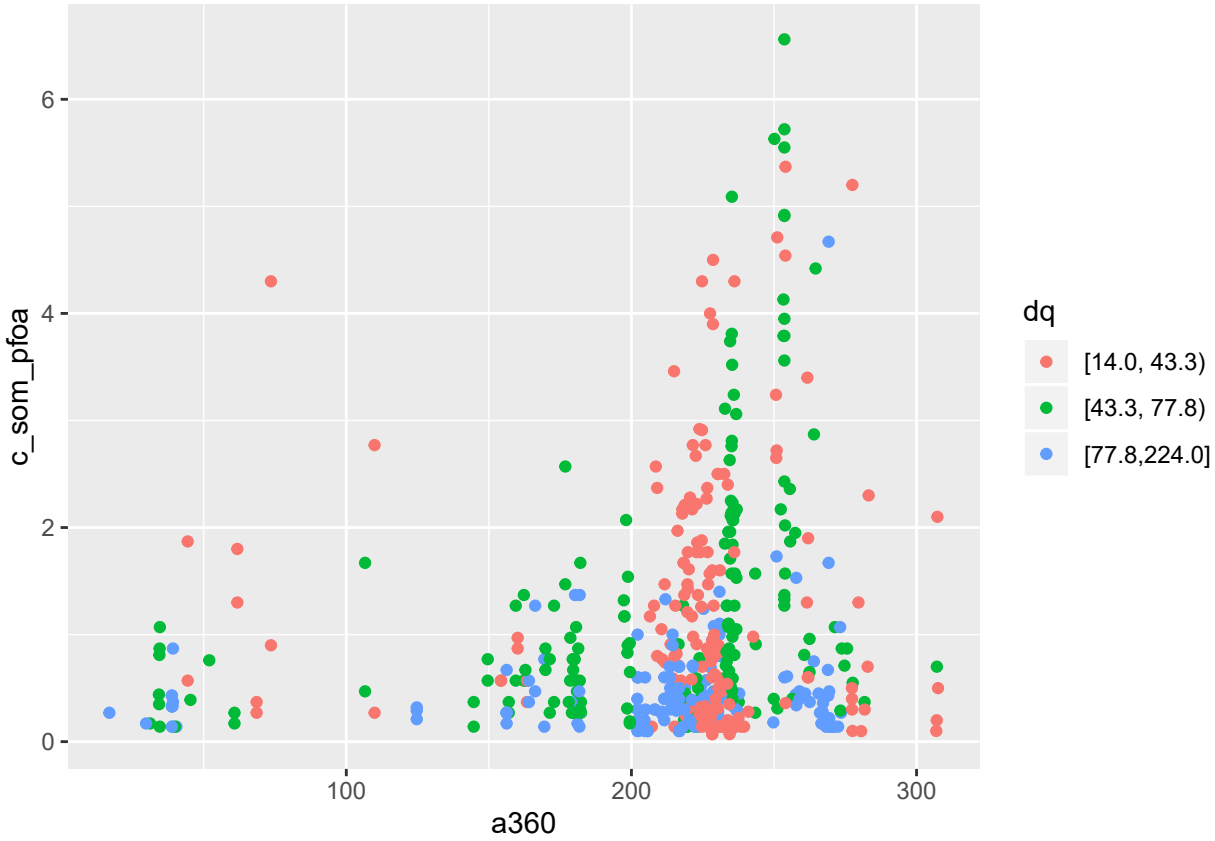


Om te onderzoeken of er een relatie met de hoek ten opzicht van chemours is, hebben we onderstaande plot gemaakt. Hierin zijn alle locatie op afstand tussen 0 en 30 km van Chemours verwijderd. De overgebleven afstanden zijn vervolgens in 5 groepen ingedeeld (bins). De concentratie is uitgezet tegen de hoek, de kleuren van de punten komen overeen met de afstand bins.

De hoek is de hoek van meetlocatie ten opzichte van de bron, het Oosten is 0 graden, Noord is 90 graden. Een hoek van 270 graden betekent dat de bron ten zuiden ligt van het meetpunt.

```
angles <- d %>% select(dataset,profiel,c_som_pfoa,
                       angle_chemours,dist_chemours) %>%
            mutate(a360=ifelse(angle_chemours<0,angle_chemours+360,angle_chemours)) %>%
            mutate(dq=Hmisc::cut2(dist_chemours/1000,bins=5)) %>%
            filter(dist_chemours>30000)

qplot(a360,c_som_pfoa,data=angles,colour=dq)
```

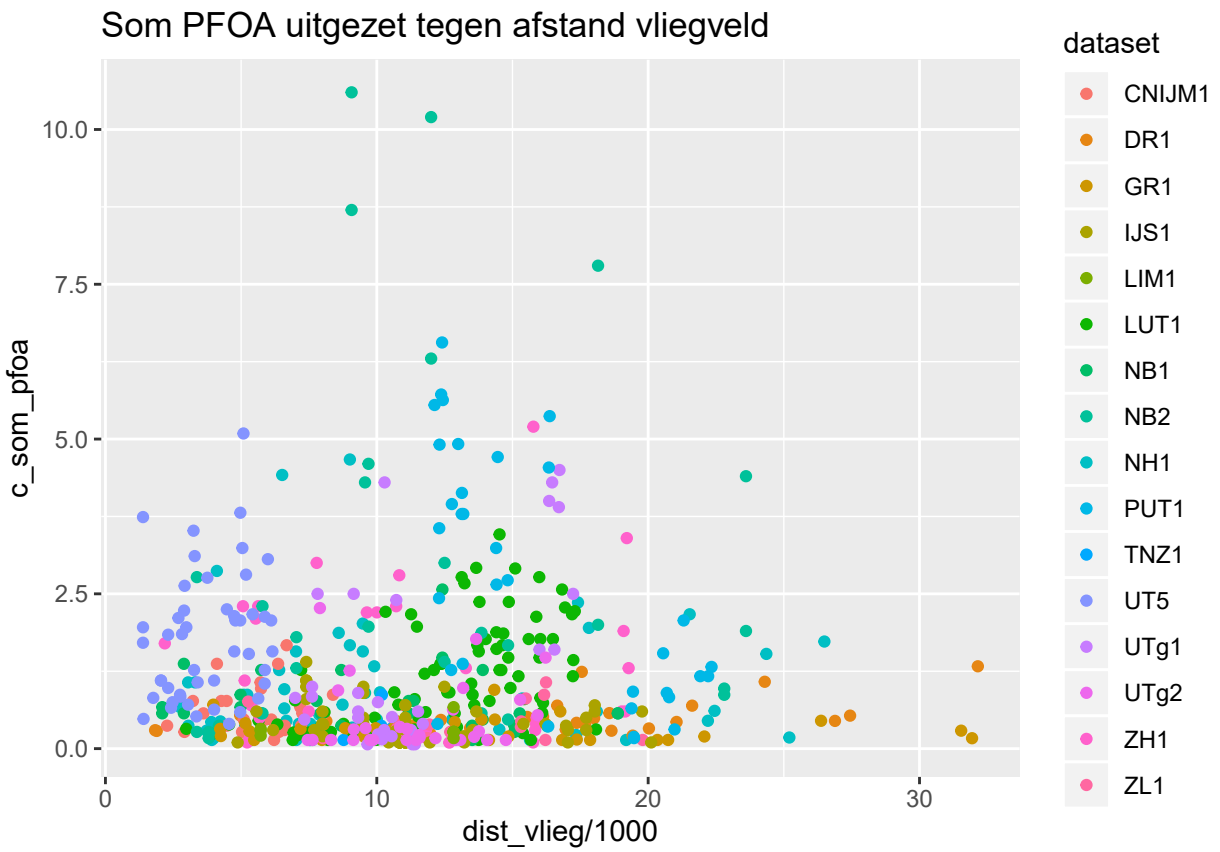
```
## Warning: Removed 14 rows containing missing values (geom_point).
```



5.4 Invloed bronnen: overige bronnen

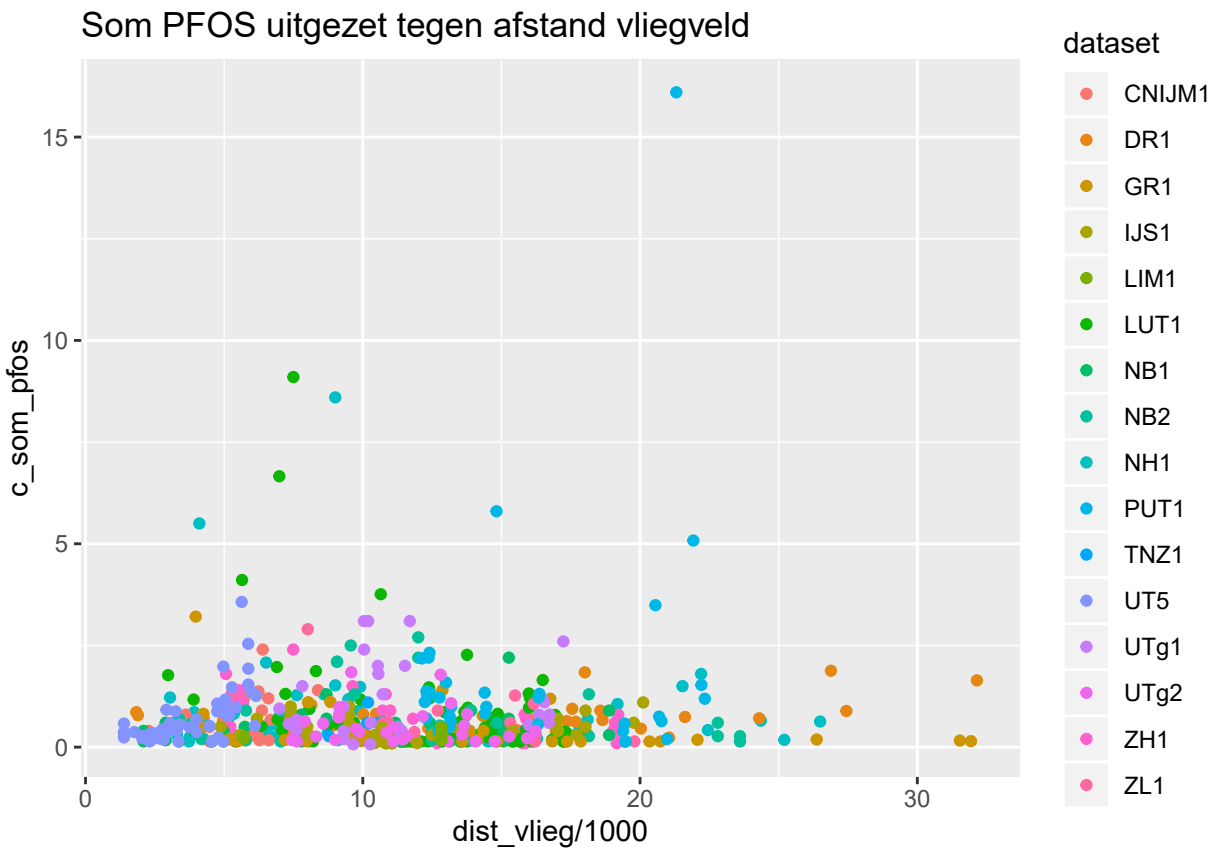
```
distances <- d %>% select(id,contains("dist_"))

qplot(dist_vlieg/1000,c_som_pfoa,data=d,color=dataset) +
    ggtitle("Som PFOA uitgezet tegen afstand vliegveld")
```

```
## Warning: Removed 14 rows containing missing values (geom_point).
```



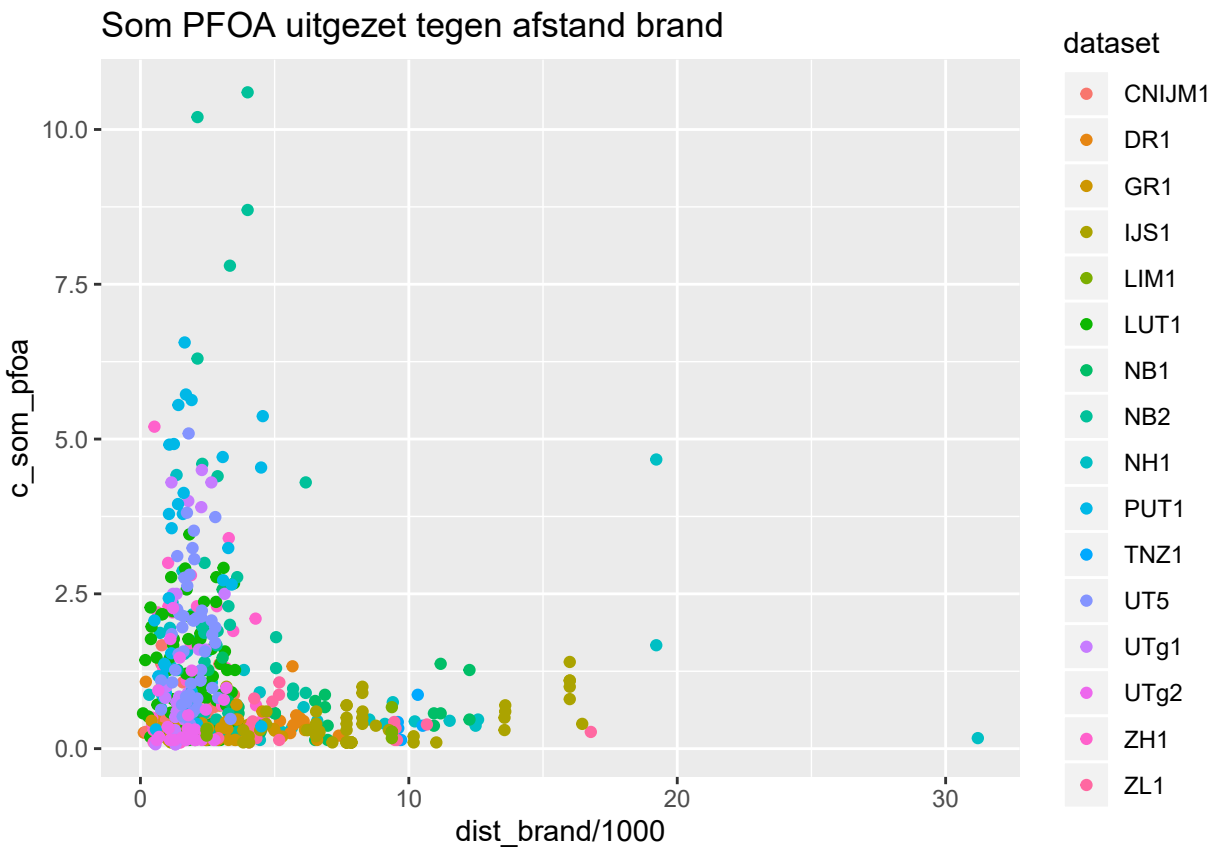

```
qplot(dist_vlieg/1000,c_som_pfos,data=d,color=dataset) +
    ggtitle("Som PFOS uitgezet tegen afstand vliegveld")
```

```
## Warning: Removed 14 rows containing missing values (geom_point).
```



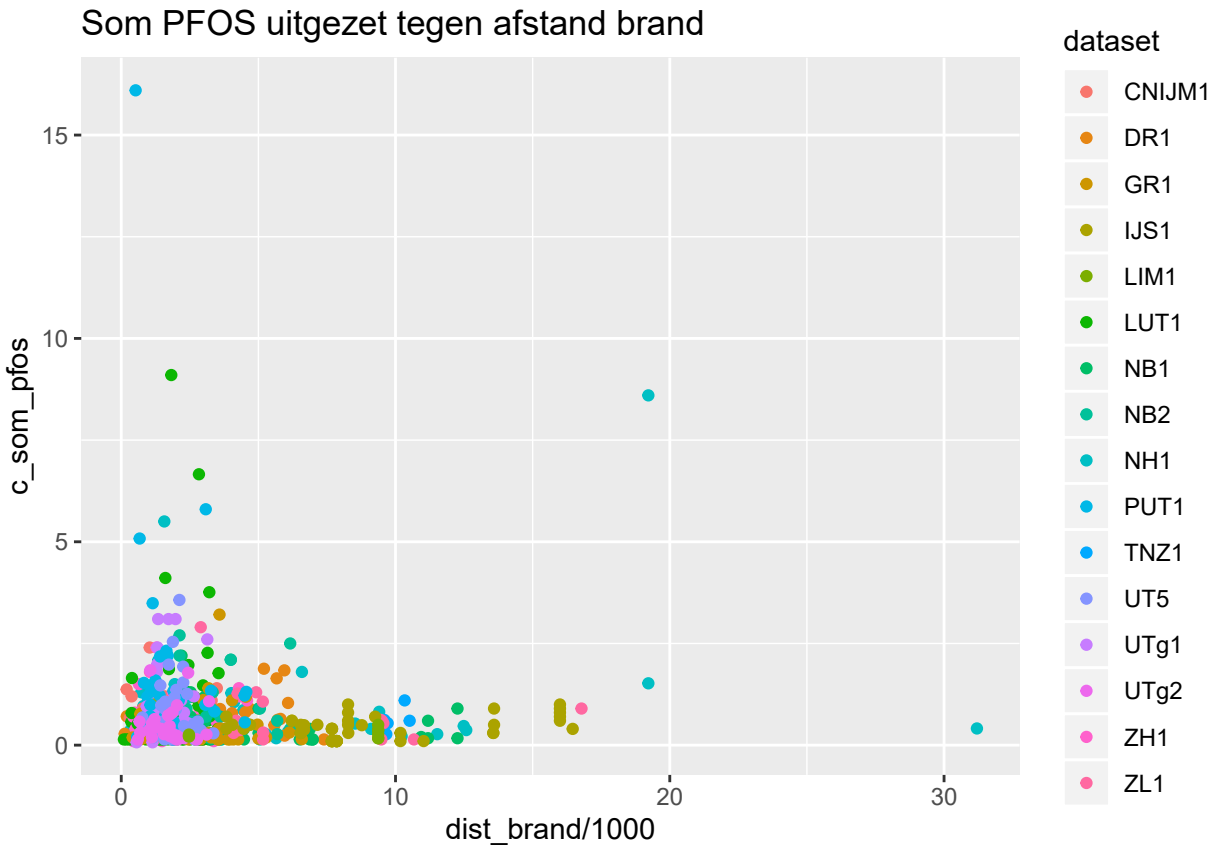

```
qplot(dist_brand/1000,c_som_pfoa,data=d,color=dataset) +
    ggtitle("Som PFOA uitgezet tegen afstand brand")
```

```
## Warning: Removed 14 rows containing missing values (geom_point).
```



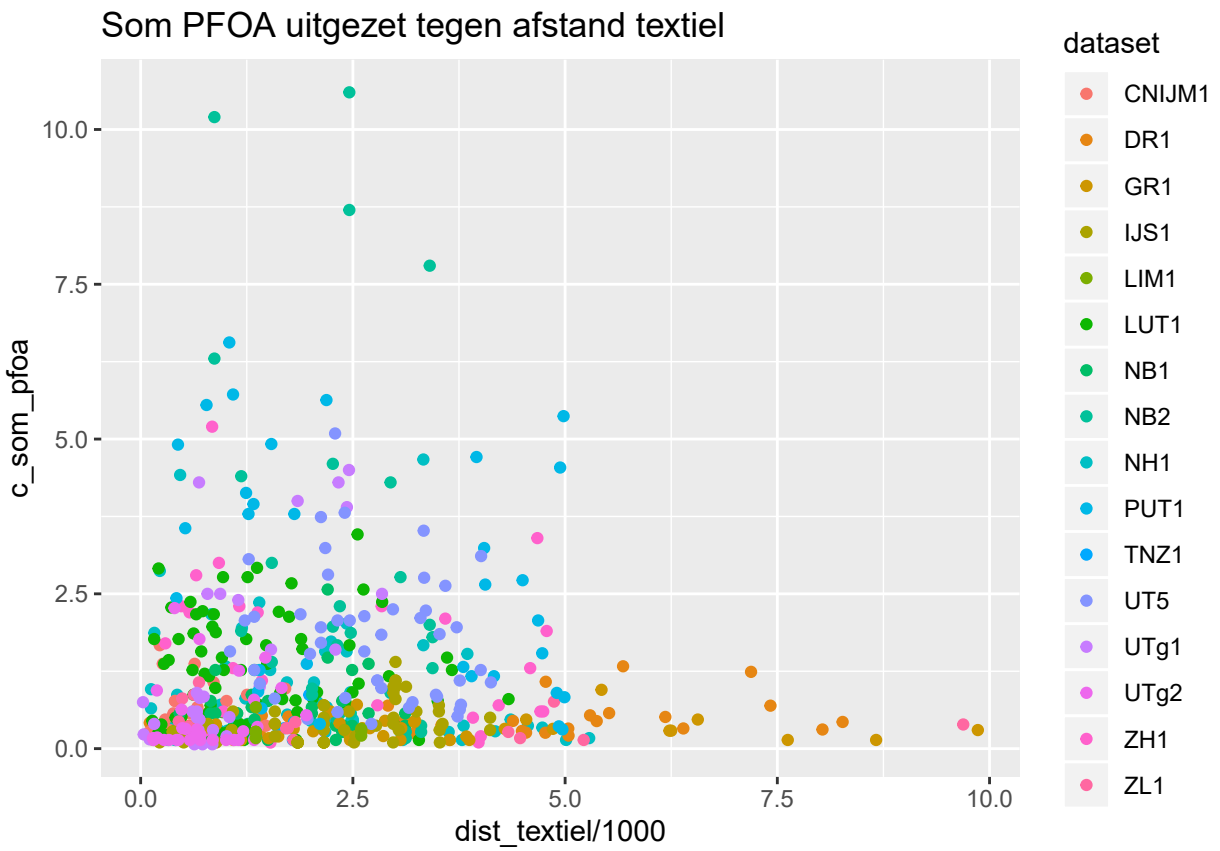

```
qplot(dist_brand/1000,c_som_pfos,data=d,color=dataset) +
    ggtitle("Som PFOS uitgezet tegen afstand brand")
```

```
## Warning: Removed 14 rows containing missing values (geom_point).
```



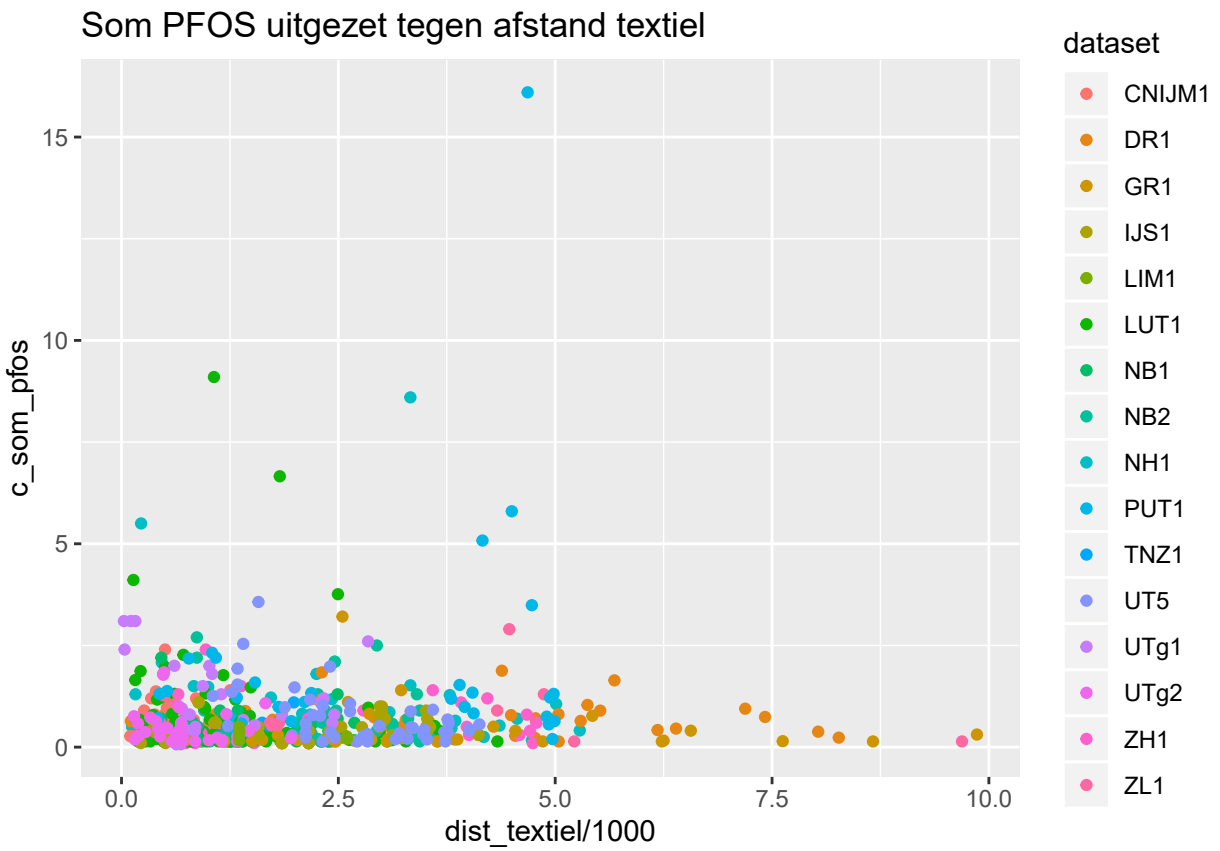

```
qplot(dist_textiel/1000,c_som_pfoa,data=d,color=dataset) +
    ggtitle("Som PFOA uitgezet tegen afstand textiel")
```

```
## Warning: Removed 14 rows containing missing values (geom_point).
```



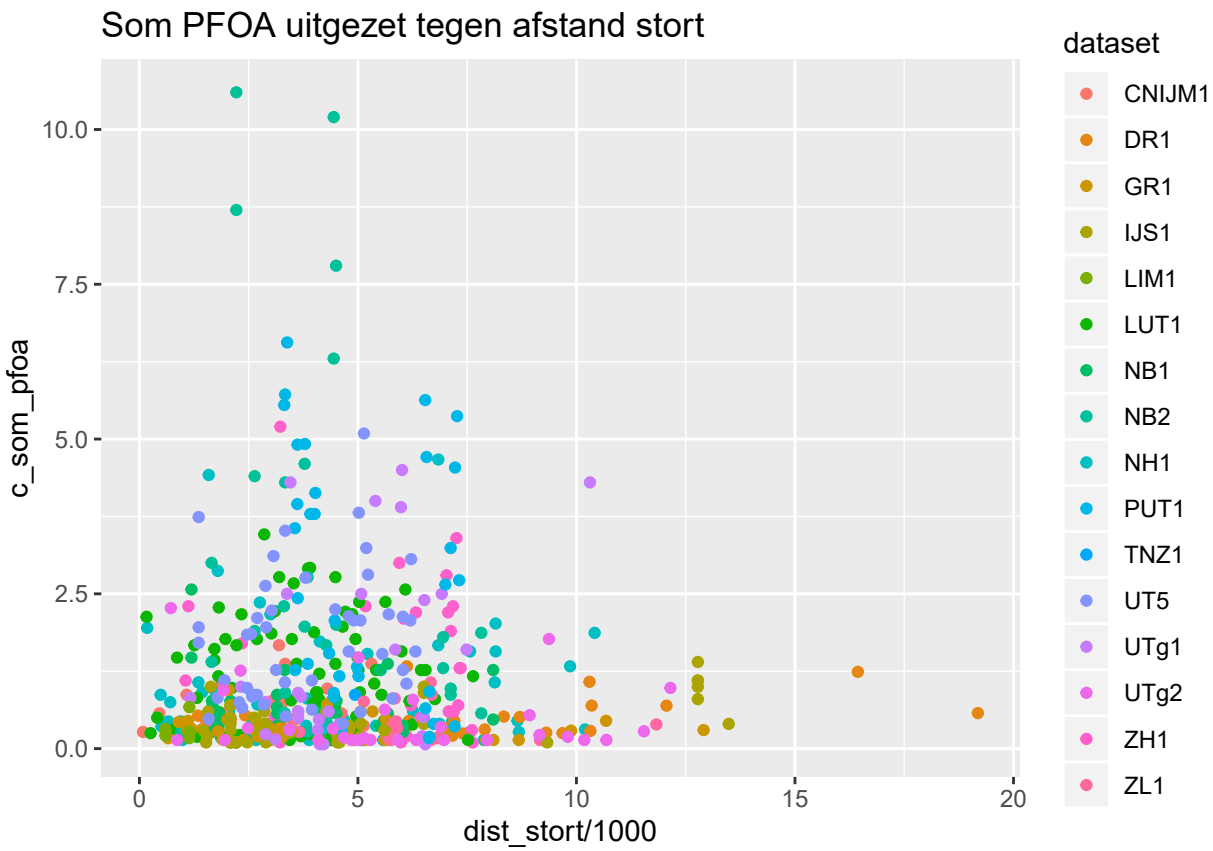

```
qplot(dist_textiel/1000,c_som_pfos,data=d,color=dataset) +
    ggtitle("Som PFOS uitgezet tegen afstand textiel")
```

```
## Warning: Removed 14 rows containing missing values (geom_point).
```



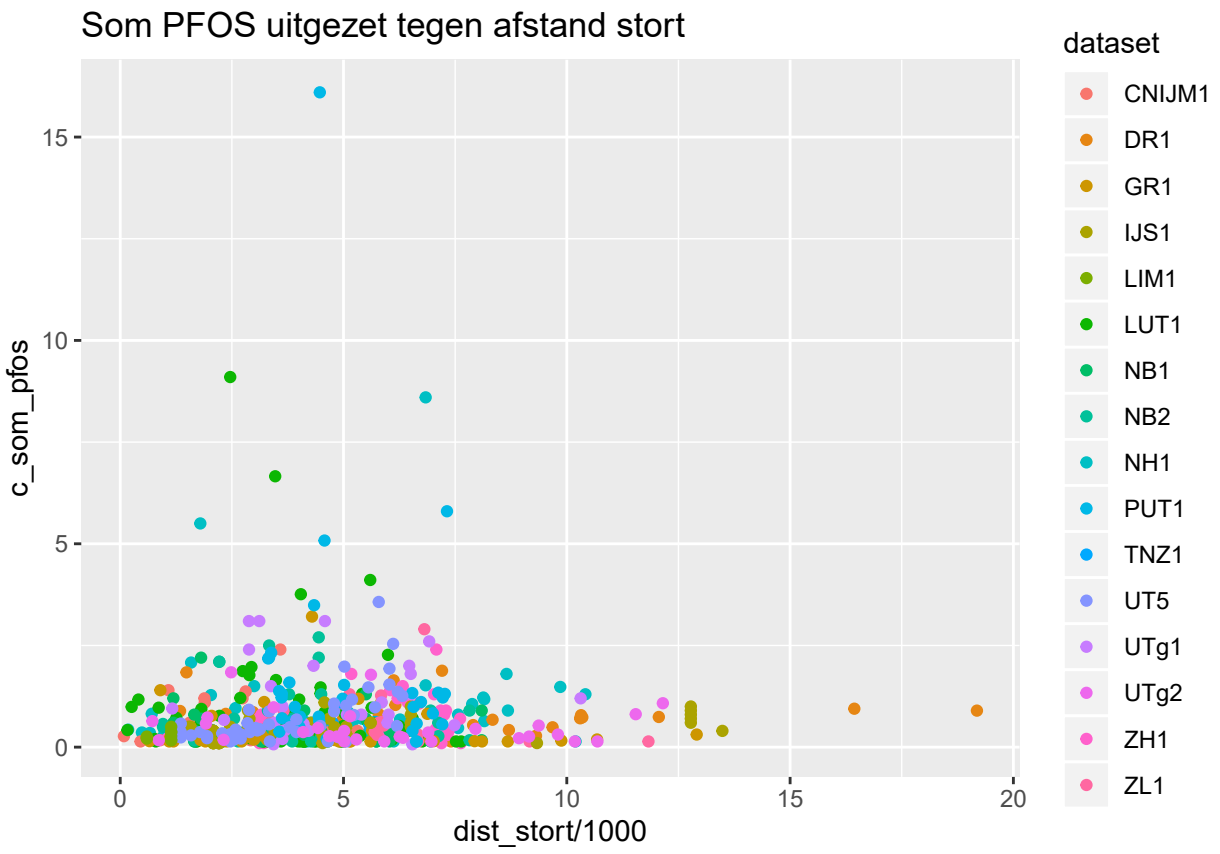

```
qplot(dist_stort/1000,c_som_pfoa,data=d,color=dataset) +
    ggtitle("Som PFOA uitgezet tegen afstand stort")
```

```
## Warning: Removed 14 rows containing missing values (geom_point).
```



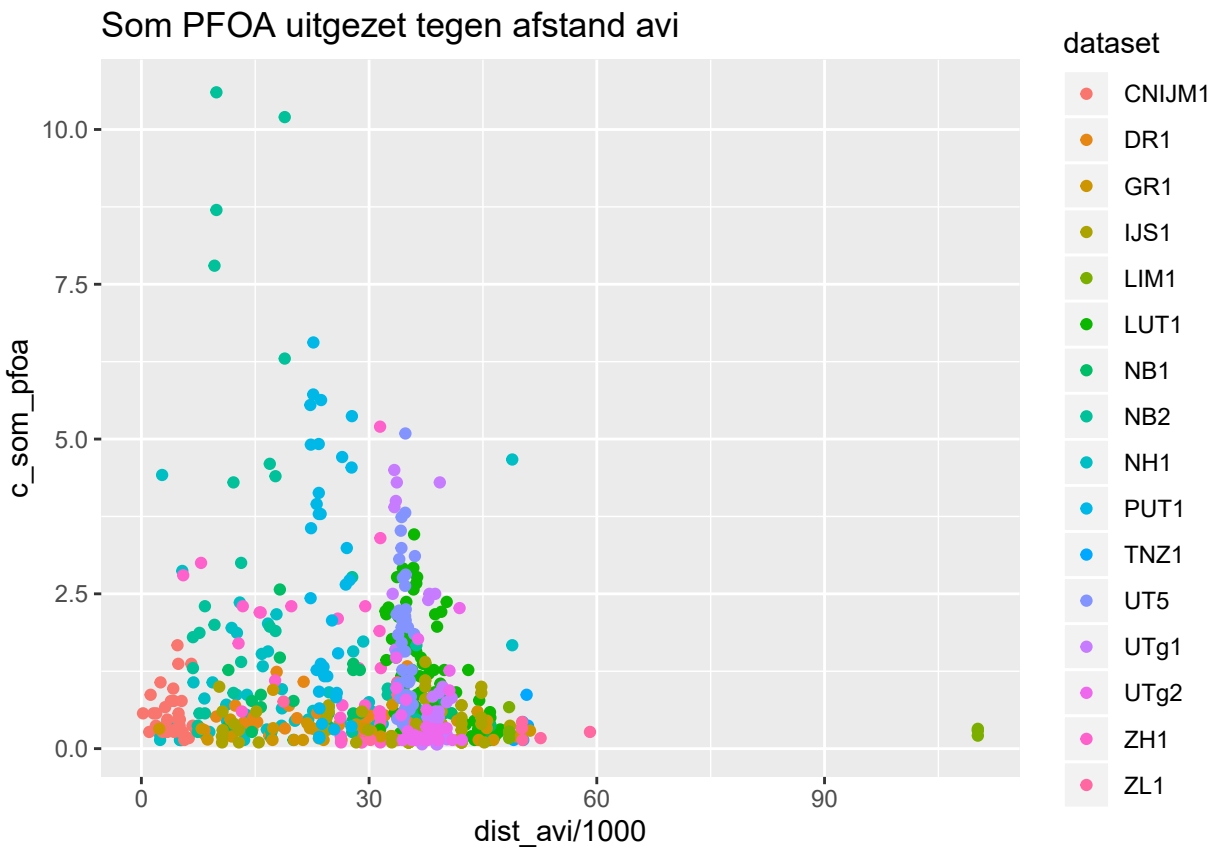

```
qplot(dist_stort/1000,c_som_pfos,data=d,color=dataset) +
    ggtitle("Som PFOS uitgezet tegen afstand stort")
```

```
## Warning: Removed 14 rows containing missing values (geom_point).
```



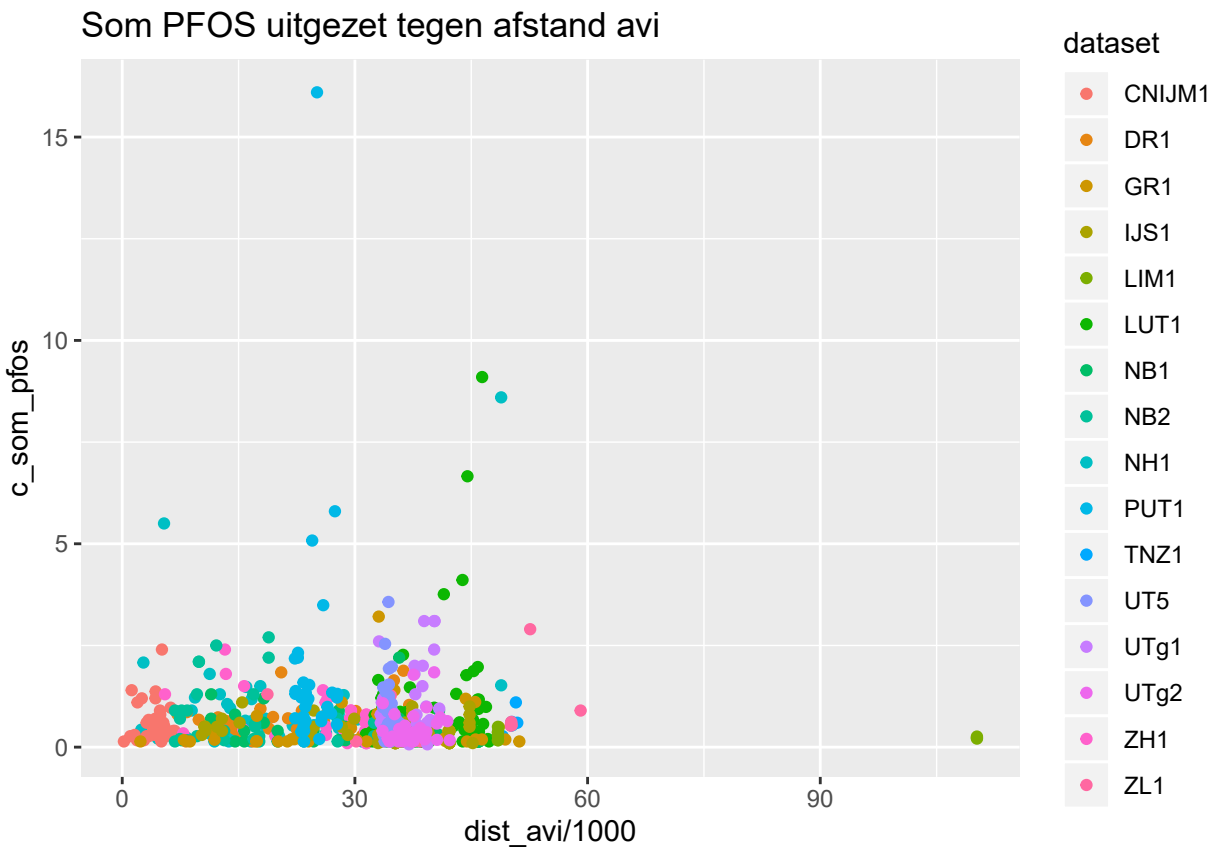

```
qplot(dist_avi/1000,c_som_pfoa,data=d,color=dataset) +
    ggtitle("Som PFOA uitgezet tegen afstand avi")
```

```
## Warning: Removed 14 rows containing missing values (geom_point).
```




```
qplot(dist_avi/1000,c_som_pfos,data=d,color=dataset) +
    ggtitle("Som PFOS uitgezet tegen afstand avi")
```

```
## Warning: Removed 14 rows containing missing values (geom_point).
```



6 Berekening percentielwaarden

6.1 Berekening 80 percentiel, gehele dataset

Berekening 80 percentiel volledige dataset

```
d <- readRDS(datafile("pfas_transform.rds")) %>%
    select("dataset","profiel",dvar$somparams) %>%
    filter(profiel=="sub")

if(!file.exists(datafile("resultaatTot3.rds"))) {
    resultaatTot3 <- data.frame()
    qobjs <- list()

    for(j in dvar$somparams) {
        cat("running",j,"\n")

        d.sel <- d %>% select(j) %>%
            na.omit()
        d.sel[[j]][d.sel[[j]]==0] <- NA
        d.sel <- na.omit(d.sel)

        r1 <- qestimates(d.sel[[j]],p=.80)
```

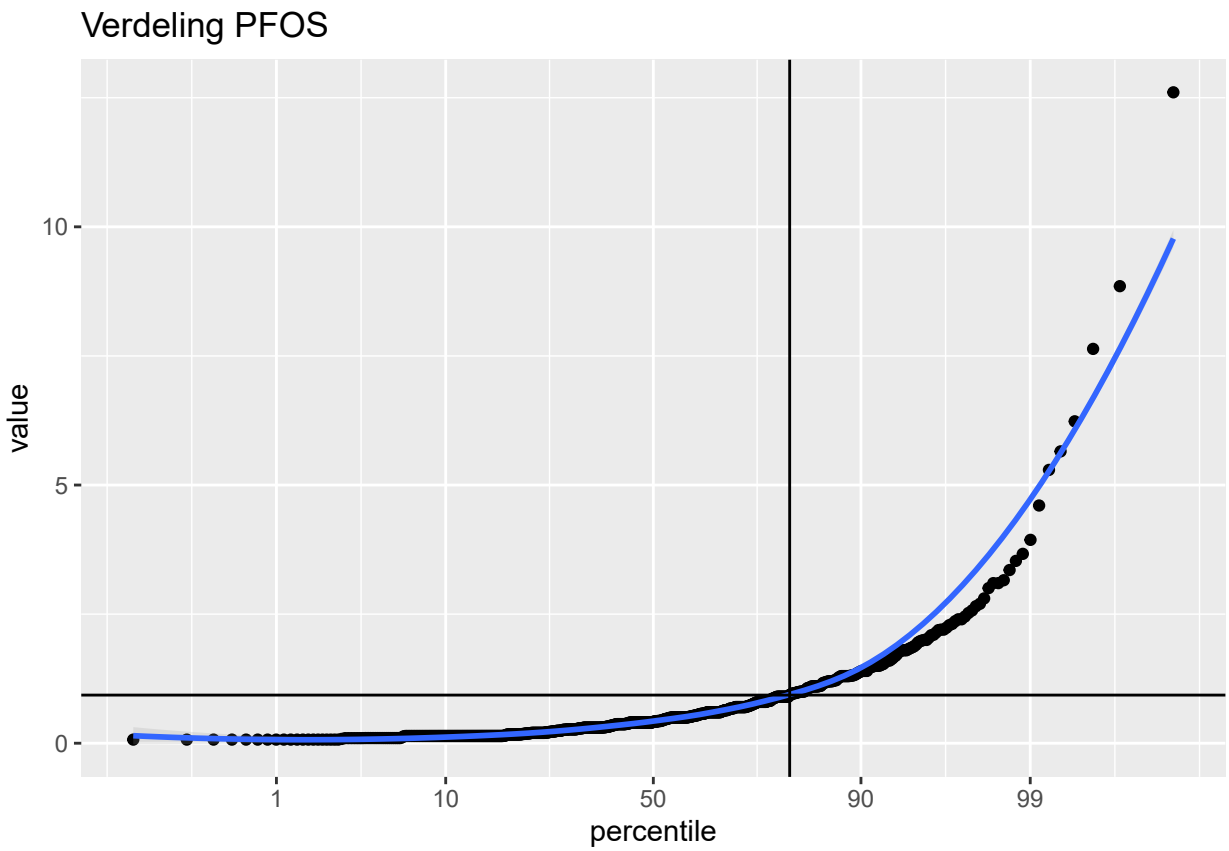
```
        res <- cbind(data.frame( param=j,
                                 n=nrow(d.sel)),
                     r1$result)

        resultaatTot3 <- rbind(resultaatTot3,res)

        obj <- list()
        obj[[paste0("C",j)]] <- r1
        qobjs <- append(qobjs,obj)
    }
    resultaatTot3 %>% saveRDS(datafile("resultaatTot3.rds"))
    write_csv(resultaatTot3,datafile("resultaatTot3.csv"))
    qobjs %>% saveRDS(datafile("resultaatTot3objs.rds"))
} else {
    resultaatTot3 <- readRDS(datafile("resultaatTot3.rds"))
}

resultaatTot3
```

```
##            param   n ro.n   ro.floq ro.quant   bs.mean    bs.bias   bs.error
## 80%   c_som_pfoa 753  753 0.0000000    1.966 1.9707000  0.0047000 0.14356432
## 80%1  c_som_pfos 753  753 0.0000000    0.932 0.9378108  0.0058108 0.04648096
## 80%2 lc_som_pfoa 753  753 0.5617530    1.984 1.9807720 -0.0032280 0.14566861
## 80%3 lc_som_pfos 753  753 0.3930943    0.938 0.9450746  0.0070746 0.05021353
##      bs.lci bs.uci fit.meanlog fit.sdlog fit.qnorm   fit.lci  fit.uci
## 80%    1.77  2.170  -0.3689609 1.2582683 1.9937641 1.8020896 2.206414
## 80%1   0.90  0.996  -0.8433740 0.9194345 0.9328057 0.8669743 1.006825
## 80%2   1.80  2.188  -0.3036581 1.1995865 2.0257488 1.8374406 2.226346
## 80%3   0.90  1.000  -0.7624001 0.8341894 0.9414551 0.8775262 1.008900
##      fcens.meanlog fcens.sdlog fcens.qnorm fcens.lci fcens.uci
## 80%     -0.3689609   1.2582683   1.9937641 1.7519499 2.2780123
## 80%1    -0.8433740   0.9194345   0.9328057 0.8637093 1.0160963
## 80%2    -1.3602210   2.0067193   1.3891285 1.1566083 1.6538028
## 80%3    -1.1197285   1.1621310   0.8679212 0.7933010 0.9461613
```

Data verdeling PFOS

```
cfp(na.omit(d$c_som_pfos),p=.8) + ggtitle("Verdeling PFOS")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



6.2 Berekening 80 percentiel, Chemours locaties uitgezonderd

Eerst kijken welke datasets er dan uit vallen.

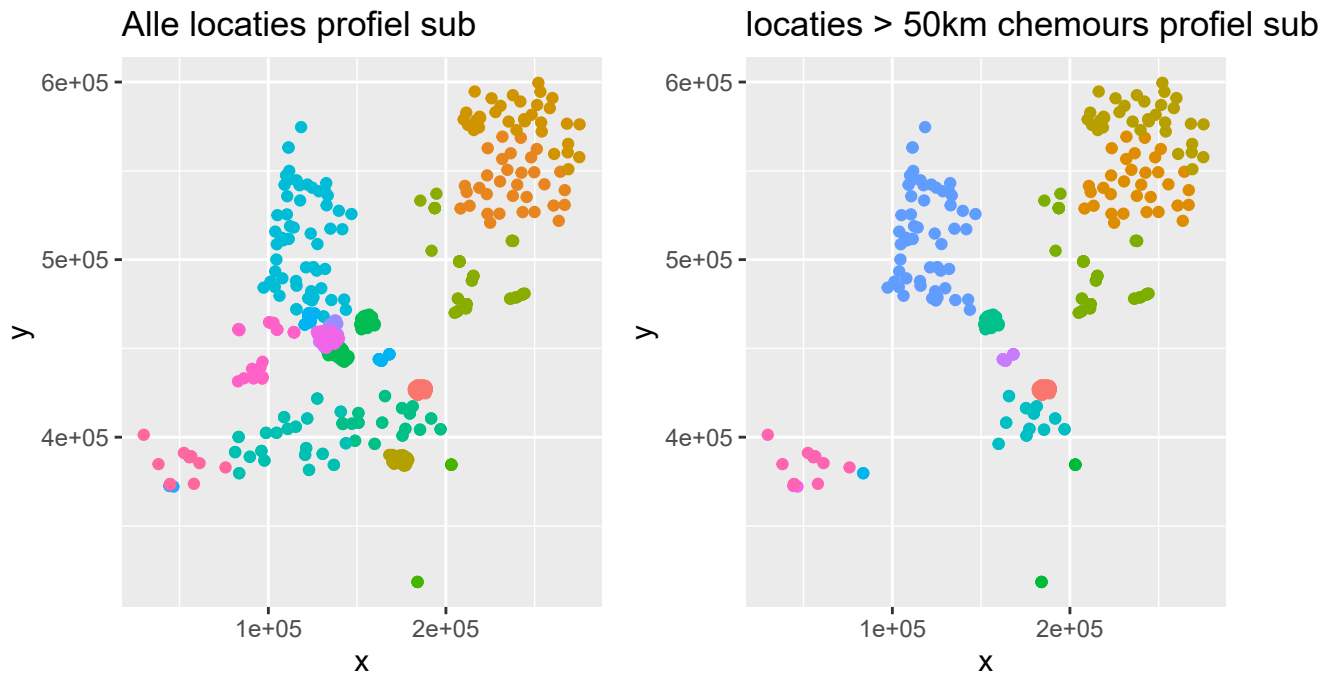
```
d1 <- readRDS(datafile("pfas_transform.rds")) %>%
    select("dataset","profiel","x","y",dvar$somparams) %>%
    filter(y<600000) %>%
    filter(profiel=="sub")
ds1<-unique(d1$dataset)

maxdist <- 50
meta <- readRDS(datafile("pfas_tidy.rds")) %>%
    select("id",dvar$cats,dvar$catsnum)
d2 <- readRDS(datafile("pfas_transform.rds")) %>%
    filter(y<600000) %>%
    filter(id%in%meta$id) %>%
    right_join(meta) %>%
    filter(dist_chemours/1000>maxdist|dataset%in%c("FR1","FR2","GLD1")) %>%
    filter(profiel=="sub")
```

```
## Joining, by = "id"
```

```
ds2<-unique(d2$dataset)

setdiff(ds1,ds2)
```

```
## [1] "HM1"  "ZH1"  "UTg1" "UT5"  "UTg2"
```

```
p1 <- qplot(x,y,data=d1,color=dataset) +
    coord_fixed() +
    ggtitle("Alle locaties profiel sub") +
    theme(legend.position = "none")


p2 <- qplot(x,y,data=d2,color=dataset) +
    coord_fixed() +
    ggtitle("locaties > 50km chemours profiel sub") +
    theme(legend.position = "none")

gridExtra::grid.arrange(p1,p2,ncol=2)
```



Dan berekenen we de 80 percentiel.

```
source("pfas_fun.R")

maxdist <- 50

meta <- readRDS(datafile("pfas_tidy.rds")) %>%
    select("id",dvar$cats,dvar$catsnum)
d <- readRDS(datafile("pfas_transform.rds")) %>%
    filter(id%in%meta$id) %>%
    right_join(meta) %>%
    filter(dist_chemours/1000>maxdist|dataset%in%c("FR1","FR2","GLD1")) %>%
```

```
    filter(profiel=="sub")
```

```
## Joining, by = "id"
```

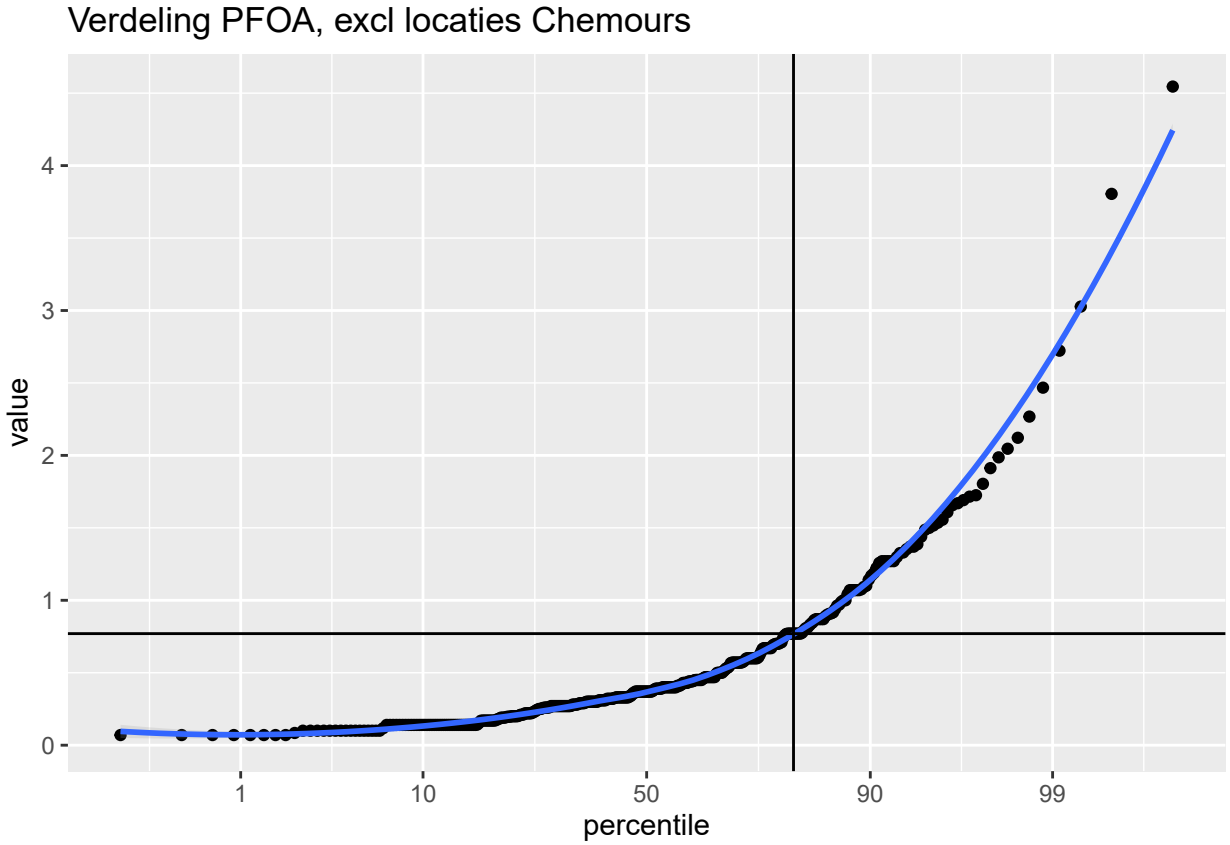
```
if(!file.exists(datafile("resultaatTotDist3.rds"))) {
    resultaatTotDist3 <- data.frame()
    qobjs <- list()

    for(j in dvar$somparams) {
        cat("running",j,"\n")

        d.sel <- d %>% select(j) %>%
            na.omit()
        d.sel[[j]][d.sel[[j]]==0] <- NA
        d.sel <- na.omit(d.sel)

        r1 <- qestimates(d.sel[[j]],p=.80)

        res <- cbind(data.frame( param=j,
                                 n=nrow(d.sel)),
                     r1$result)

        resultaatTotDist3 <- rbind(resultaatTotDist3,res)

        obj <- list()
        obj[[paste0("C",j)]] <- r1
        qobjs <- append(qobjs,obj)
    }
    resultaatTotDist3 %>% saveRDS(datafile("resultaatTotDist3.rds"))
    write_csv(resultaatTotDist3,datafile("resultaatTotDist3.csv"))
    qobjs %>% saveRDS(datafile("resultaatTotDist3objs.rds"))
} else {
    resultaatTotDist3 <- readRDS(datafile("resultaatTotDist3.rds"))
}

resultaatTotDist3
```

```
##            param   n ro.n   ro.floq ro.quant   bs.mean    bs.bias   bs.error
## 80%   c_som_pfoa 389  389 0.0000000     0.77 0.7657436 -0.0042564 0.05000336
## 80%1  c_som_pfos 389  389 0.0000000     0.80 0.7997708 -0.0002292 0.06320096
## 80%2 lc_som_pfoa 389  389 0.7557841     0.80 0.7916526 -0.0083474 0.04560903
## 80%3 lc_som_pfos 389  389 0.4524422     0.80 0.8068190  0.0068190 0.06444887
##        bs.lci bs.uci fit.meanlog fit.sdlog fit.qnorm   fit.lci   fit.uci
## 80%  0.700000 0.8300  -0.9610215 0.8101798 0.7564212 0.6886911 0.8303486
## 80%1 0.709301 0.8988  -0.9224403 0.9011644 0.8487410 0.7707568 0.9338369
## 80%2 0.725000 0.8400  -0.8652475 0.7428156 0.7865664 0.7219617 0.8589618
## 80%3 0.714000 0.8988  -0.8325654 0.8190629 0.8665593 0.7866680 0.9585763
##      fcens.meanlog fcens.sdlog fcens.qnorm fcens.lci fcens.uci
## 80%     -0.9610215   0.8101798   0.7564212 0.6819384 0.8315987
## 80%1    -0.9224403   0.9011644   0.8487410 0.7449417 0.9647382
## 80%2    -2.1033143   1.2646637   0.3538272 0.2989334 0.4150696
## 80%3    -1.2911158   1.1934992   0.7507808 0.6415574 0.8711136
```

Data verdeling PFOA

```
cfp(d$c_som_pfoa,p=.8) + ggtitle("Verdeling PFOA, excl locaties Chemours")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



6.3 Gevoeligheids analyse PFOS

PFOS bevat de grootste uitschieters tussen de verschillende datsets. Om te kijken wat de effecten zijn van de verschillende datasets voeren we een gevoeligheidsanalyse uit volgens het Leave-One-Out principe.

```
d <- readRDS(datafile("pfas_transform.rds")) %>%
    select("dataset","profiel","c_som_pfos") %>%
    filter(profiel=="sub")
looPFOS <- data.frame()
qobjs <- list()
ds <- unique(d$dataset)
for(i in ds) {
    d.sel <- subset(d,dataset!=i)
    x <- na.omit(d.sel$c_som_pfos)
    loo <- qestimates(x,p=.80)

    looPFOS <- rbind(looPFOS,data.frame(dataset=i,loo$result))
    obj <- list()
    obj[[paste0("C",i)]] <- loo
    qobjs <- append(qobjs,obj)
```

```
}
```

```
looPFOS %>% select(dataset,ro.n,ro.quant,bs.mean,bs.lci,bs.uci)
```

```
##       dataset ro.n ro.quant   bs.mean bs.lci bs.uci
## 80%       NH1  640    0.972 0.9667226 0.9000 1.0444
## 80%1      DR1  721    0.950 0.9444990 0.9000 1.0000
## 80%2      HM1  661    0.920 0.9366240 0.9000 1.0000
## 80%3   CNIJM1  723    0.912 0.9350092 0.9000 1.0000
## 80%4     LUT1  664    0.900 0.9320888 0.9000 1.0000
## 80%5      ZH1  717    0.900 0.9293802 0.8996 0.9900
## 80%6     TNZ1  746    0.940 0.9391520 0.9000 1.0000
## 80%7      NB1  723    0.956 0.9487818 0.9000 1.0000
## 80%8      NB2  723    0.912 0.9283708 0.8890 0.9900
## 80%9      GR1  718    0.948 0.9455038 0.9000 1.0000
## 80%10     ZL1  737    0.916 0.9324092 0.8996 0.9980
## 80%11    IJS1  694    0.974 0.9791666 0.9000 1.0700
## 80%12    PUT1  719    0.900 0.8909998 0.8200 0.9400
## 80%13    UTg1  723    0.900 0.9076362 0.8660 0.9700
## 80%14     UT5  706    0.900 0.9237150 0.8980 0.9900
## 80%15    UTg2  723    0.948 0.9455836 0.9000 1.0000
## 80%16    GLD1  725    0.970 0.9621286 0.9000 1.0200
## 80%17     FR1  749    0.942 0.9416216 0.9000 1.0000
## 80%18     FR2  751    0.940 0.9394490 0.9000 1.0000
## 80%19    LIM1  744    0.947 0.9451272 0.9000 1.0000
```