

# Verkenning svi, influent en effluent

Auke Beeksma

23 november 2018

## Synopsis

Dit is een eerste verkenning om antwoord te kunnen geven op de vraag hoe pieken in svi ontstaan. Vanzelfsprekend is hier eerst gekeken naar de waarden van svi. In combinatie met het verkennen van de zogenoemde influence(water voor zuivering) en effluence (water na zuivering) geeft dit waarschijnlijk eerste inzichten.

### Inladen data, bibliotheken etc.

```
setwd("C:/Users/auke.beeksma/Desktop/Kennisnetwerk/Waterlaboratorium-Noord-IoT")

library(openxlsx)
library(tibble)
library(janitor)
library(tidyr)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(ggplot2)
library(ggthemes)
library(data.table)

##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
##   between, first, last

influent <- read.xlsx("Kennisnetwerk_procesdata_kwaliteitsdata_WLN.xlsx",
  sheet = 1)
effluent <- read.xlsx("Kennisnetwerk_procesdata_kwaliteitsdata_WLN.xlsx",
  sheet = 2)
```

```
svi <- read.xlsx("Kennisnetwerk_procesdata_kwaliteitsdata_WLN.xlsx", sheet = 3)
```

### Eerste inspectie

```
str(influent)
```

```
## 'data.frame':    170 obs. of  17 variables:
## $ Datum          : num  41913 41918 41923 41928 41933 ...
## $ zwst.(mg/l)    : num  172 91 66 174 74 113 207 68 54 161 ...
## $ egv.(mS/m)     : num  716 876 3755 2495 6308 ...
## $ cl.(mg/l)      : num  1920 2480 12800 8310 25100 29100 17500 5250 1530
3060 ...
## $ so4.(ug/l)     : num  57 67 65 92 73 150 97 164 49 61 ...
## $ nh4.(mg.N/l)   : num  91 54.4 8 14.9 2.6 0.3 23.2 17.1 48.2 1.5 ...
## $ nkj.(mg.N/l)   : num  110 64 13 31 12 4.3 46 24 66 310 ...
## $ no2.(mg.N/l)   : chr  "0.58399999999999996" "0.24199999999999999" "6.37"
"6.44" ...
## $ no3.(mg.N/l)   : chr  "1.3" "1.7" "57.1" "27.5" ...
## $ n.(mg.N/l)     : num  112 66 76 65 130 156 130 43 66 311 ...
## $ po4o.(mg.P/l)  : num  0.31 1.3 1.6 2.2 2.5 1.4 2.4 1.1 0.9 0.2 ...
## $ po4.(mg.P/l)   : num  2 2.7 2.8 4.2 4 2.4 4.8 1.9 2 1.2 ...
## $ czv.(mg.O2/l)  : num  1580 1510 2080 1400 2460 1780 2210 1640 2460 3410
...
## $ bzv-5.(mg/l)   : num  680 NA 1020 NA 1230 NA 940 NA 1190 NA ...
## $ na.(mg/l)      : num  NA NA NA 5600 16300 NA 10400 3300 1200 1500 ...
## $ k.(mg/l)       : num  NA NA NA 17 17 NA 7.3 41 6.4 9 ...
## $ ph.(pH)        : num  7.5 7.5 7.4 7.3 7.4 7 7.5 7.5 7.6 8.4 ...
```

```
str(effluent)
```

```
## 'data.frame':    262 obs. of  17 variables:
## $ Datum          : num  41913 41918 41923 41928 41933 ...
## $ zwst.(mg/l)    : chr  "17" "7" "39" "15" ...
## $ egv.(mS/m)     : chr  "3650" "2105" "2623" "3210" ...
## $ cl.(mg/l)      : num  12400 6560 8760 11100 13500 15500 17700 19100 12900
17300 ...
## $ so4.(mg/l)     : num  666 687 698 471 750 726 451 292 165 119 ...
## $ nh4.(mg.N/l)   : chr  "14.4" "26.8" "10.8" "8.3000000000000007" ...
## $ nkj.(mg.N/l)   : chr  "21" "32" "13" "12" ...
## $ no2.(mg.N/l)   : chr  "0.65100000000000002" "2.4500000000000002"
"0.72399999999999998" "0.69" ...
## $ no3.(mg.N/l)   : chr  "8.1" "9.5" "7.3" "7.5" ...
## $ n.(mg.N/l)     : chr  "29" "44" "21" "20" ...
## $ po4o.(mg.P/l)  : chr  "0.25" "9.7000000000000003E-2"
"6.5000000000000002E-2" "4.4999999999999998E-2" ...
## $ po4.(mg.P/l)   : chr  "0.7" "0.42" "0.36" "0.18" ...
## $ czv.(mg.O2/l)  : chr  "128" "113" "89" "49" ...
## $ bzv-5.(mg/l)   : num  9 6 8 5 4 5 6 9 5 6 ...
## $ na.(mg/l)      : num  NA NA NA 7700 9600 NA 11700 11800 9000 10400 ...
## $ k.(mg/l)       : num  NA NA NA 7.4 7.6 NA 11 11 14 6.6 ...
## $ ph.(pH)        : num  8.2 8.2 8.1 8.1 8.2 8.2 8.2 8.2 8.2 8.2 ...
```

```
str(svi)
```

```
## 'data.frame': 351 obs. of 8 variables:
## $ datum : num 41908 41912 41915 41919 41922 ...
## $ D.STOF.AT: num 4400 4600 4600 4100 4200 4600 4800 5600 5600 5200 ...
## $ SVI.5 : num 91.9 54.8 85.7 112.9 114.5 ...
## $ SVI.10 : num 75.7 52.4 74.9 80.6 76.3 ...
## $ SVI.15 : num 64.9 50 64.2 69.9 66.2 ...
## $ SVI.20 : num 54.1 47.6 60 61.8 61.1 ...
## $ SVI.25 : num 48.6 45.2 55.7 53.8 56 ...
## $ SVI.30 : num 48.6 42.9 45 51.1 50.9 ...
```

### Datums goed

```
influent$Datum <- excel_numeric_to_date(influent$Datum)
effluent$Datum <- excel_numeric_to_date(effluent$Datum)
svi$datum <- excel_numeric_to_date(svi$datum)
```

### Betekenis termen geven

```
setnames(influent, c("datum", "zwevendestof", "geleidbaarheid", "chloride",
"sulfaat", "ammonium", "kjeldahl", "nitriet",
"nitraat", "kjeldahlnitrietnitraat", "opgelostfosfaat", "totaalfosfaat",
"chemischzuurstofverbruik", "biologischzuurstofverbruik", "natrium", "kalium",
"zuurgraad"))
```

### Na controle identiek voor data effluent, idem dito

```
setnames(effluent, c("datum", "zwevendestof", "geleidbaarheid", "chloride",
"sulfaat", "ammonium", "kjeldahl", "nitriet",
"nitraat", "kjeldahlnitrietnitraat", "opgelostfosfaat", "totaalfosfaat",
"chemischzuurstofverbruik", "biologischzuurstofverbruik", "natrium", "kalium",
"zuurgraad"))
```

Voor data svi staat het redelijk informatief. D.STOF.AT even aanpassen (naar dichtheid zwevende stof o.id.? Heb het niet exact meer voor ogen) en SVI.5 voor “intuïtieve volgordelijkheid”.

```
setnames(svi, c("datum", "hoeveelheidzwevendestof", "SVI.05", "SVI.10",
"SVI.15", "SVI.20", "SVI.25", "SVI.30"))
```

Eerste voorbereidingen zijn getroffen, maar even wat tekenen :)

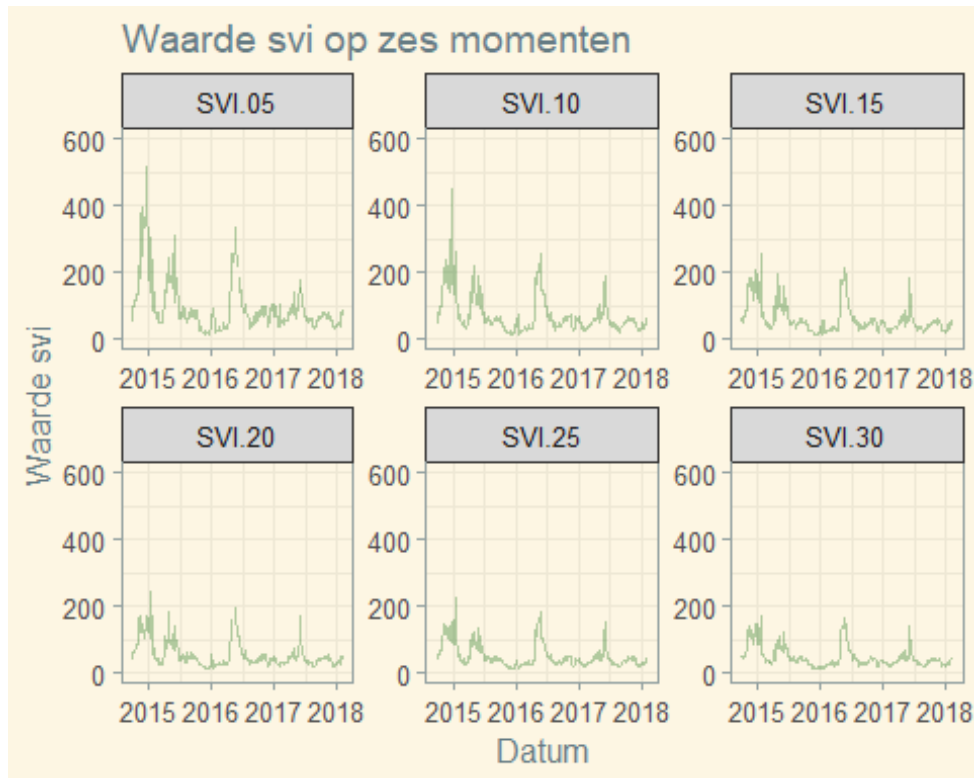
## Grafieken

### De trend in SVI inzichtelijk maken voor eind 2015 - begin 2018

```
svig <- svi %>%
  gather(-datum, - hoeveelheidzwevendestof, key="var", value = "value")

ggplot(svig, aes(x = datum, y = value)) +
  geom_line(alpha = 0.3, colour = "darkgreen") +
  facet_wrap(~var, scales = "free") +
  labs(y = "Waarde svi", x = "Datum", title = "Waarde svi op zes momenten") +
```

```
ylim(0, 600) +  
theme_solarized()
```



```
ggsave("svi.pdf")
```

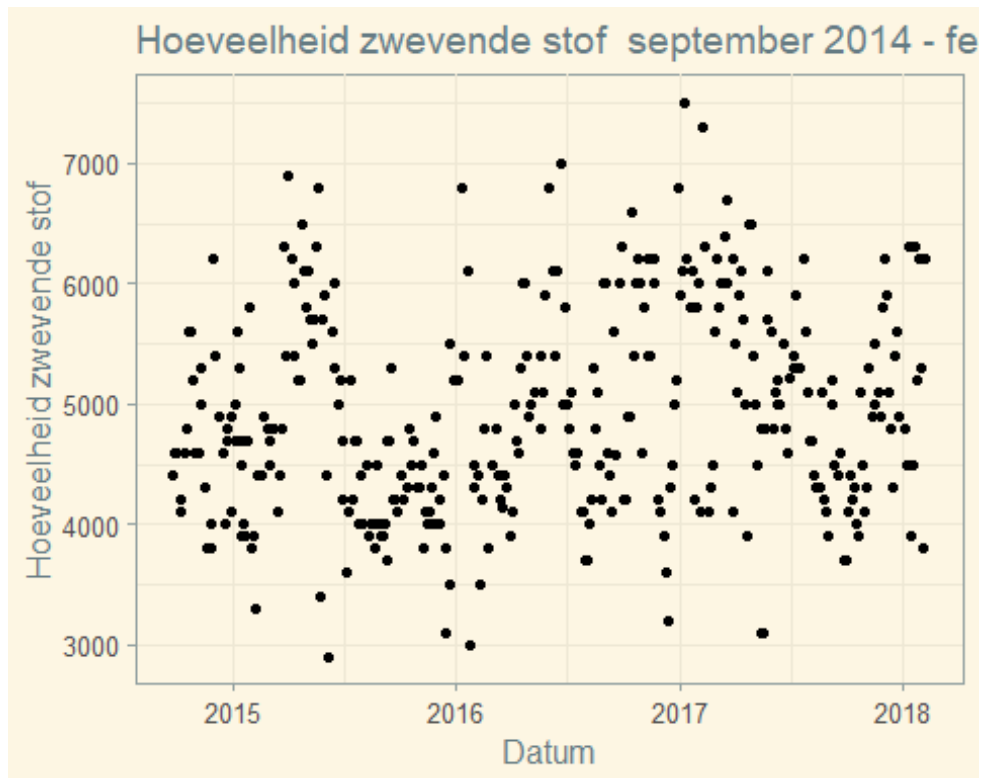
```
## Saving 5 x 4 in image
```

Duidelijk waarneembaar is naarmate de tijd vordert de waarde van SVI daalt. Op het oog lijkt dit sneller te gaan voor hogere waarden van SVI.

Om geen zaken te missen, voor verkenning nog hoeveelheid stof

```
ggplot(svi, aes(x = datum, y = hoeveelheidzwevendestof)) +  
  geom_point() +  
  labs(y = "Hoeveelheid zwevende stof", x = "Datum", title = "Hoeveelheid  
zwevende stof september 2014 - februari 2018") +  
  theme_solarized()
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```



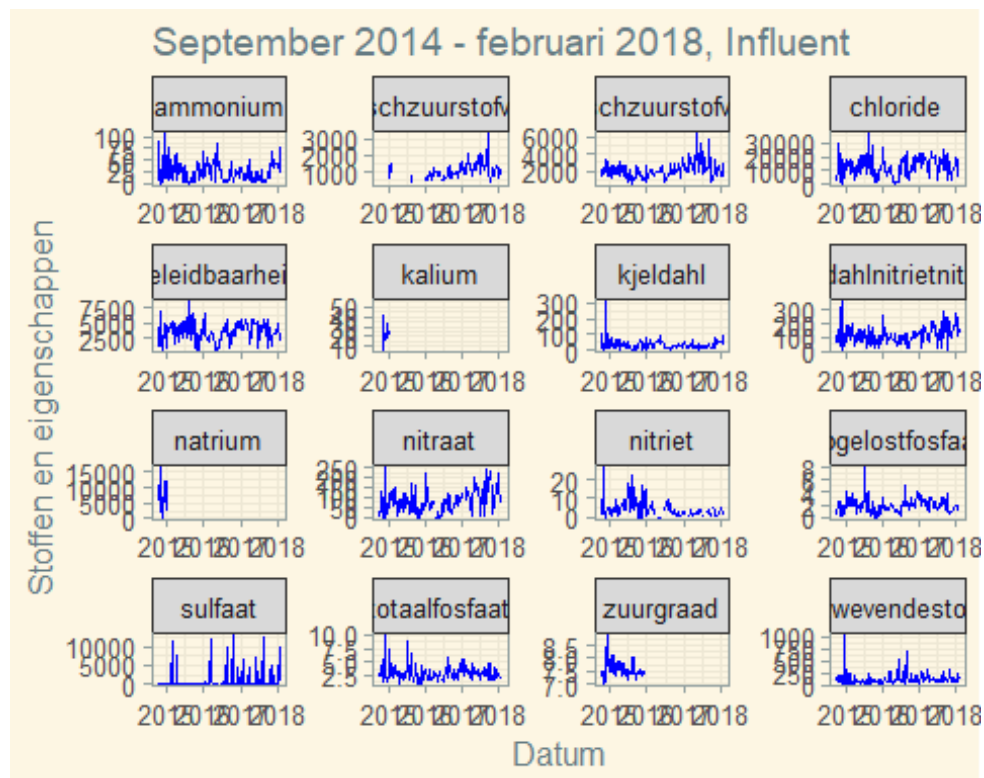
Op het oog is hier geen chocolade van te maken. Wellicht iets met clustering te doen.

## Influent

```
influentg <- influent %>%
  gather(-datum, key = "var", value = "value")

ggplot(influentg, aes(x = datum, y = as.numeric(value))) +
  geom_line(colour = "blue") +
  labs(y = "Stoffen en eigenschappen", x = "Datum", title = "September 2014 -
februari 2018, Influent") +
  facet_wrap(~var, scales = "free") +
  theme_solarized()

## Warning in FUN(X[[i]], ...): NAs introduced by coercion
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```



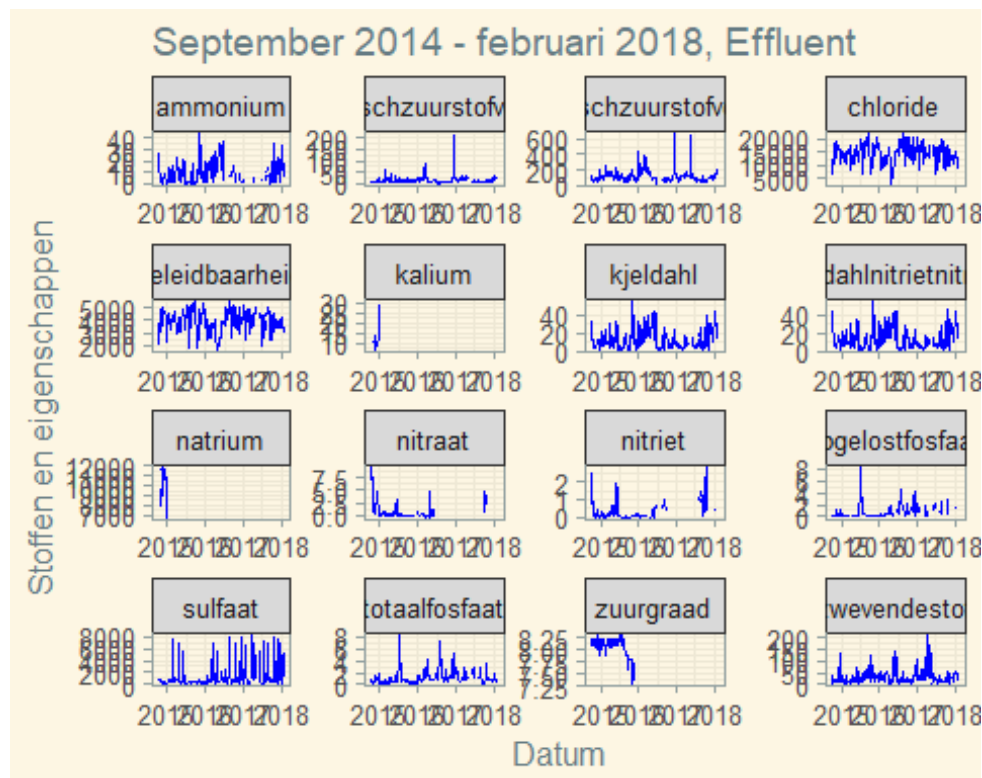
Opvallend is het grote aantal niet aanwezige metingen van kalium, natrium, zuurgraad. Deze lijken mij zodoende slechte voorspellers om mee te nemen in een eventueel model.

## Effluent

```
effluentg <- effluent %>%
  gather(-datum, key = "var", value = "value")

ggplot(effluentg, aes(x = datum, y = as.numeric(value))) +
  geom_line(colour = "blue") +
  labs(y = "Stoffen en eigenschappen", x = "Datum", title = "September 2014 -
februari 2018, Effluent") +
  facet_wrap(~var, scales = "free") +
  theme_solarized()

## Warning in FUN(X[[i]], ...): NAs introduced by coercion
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```



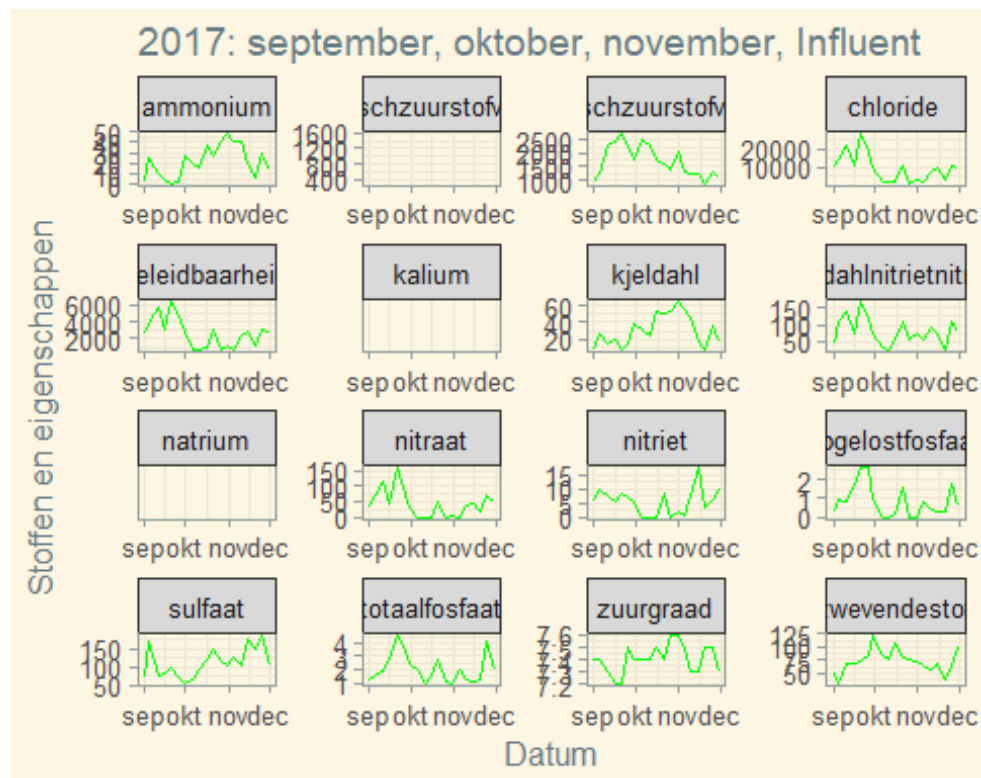
Niet geheel toevallig zijn dezelfde stoffen vaak afwezig. Het lijkt erop dat men vanaf bepaalde datums systematisch is gestopt met het meten van bepaalde stoffen.

Afgaande op de svi waarden zijn er drie pieken waarneembaar: in december 2015, mei 2016 en mei 2017. Ik stel voor om te kijken of er veranderingen in hoeveelheid van een bepaalde stof zijn in de 3 maanden voorafgaand aan dergelijke pieken in svi.

## september, oktober en november 2015

```
influentson2015 <- influent %>%
  filter(datum >= as.Date("2015-09-01") & datum < as.Date("2015-12-01")) %>%
  gather(-datum, key= "var", value = "value")

ggplot(influentson2015, aes(x = datum, y = as.numeric(value))) +
  geom_line(colour = "green") +
  facet_wrap(~var, scales = "free") +
  labs(y = "Stoffen en eigenschappen", x = "Datum", title = "2015: september,
oktober, november, Influent") +
  theme_solarized()
```



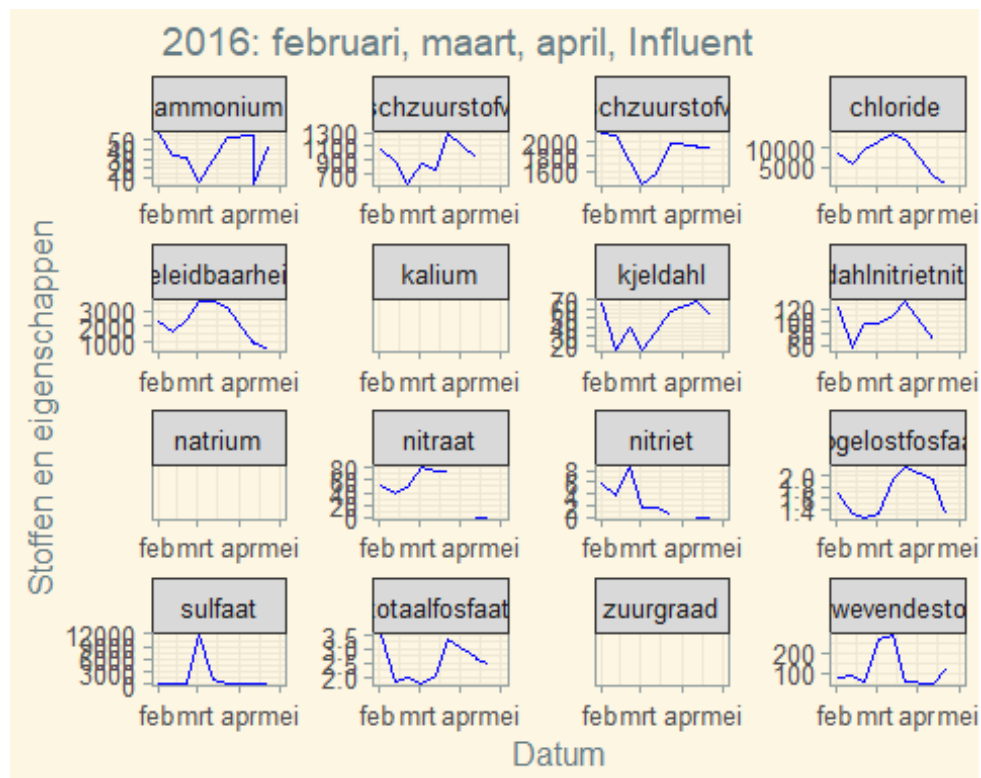
## februari, maart en april 2016

```
influentfma2016 <- influent %>%
  filter(datum >= as.Date("2016-02-01") & datum < as.Date("2016-05-01")) %>%
  gather(-datum, key= "var", value = "value")

ggplot(influentfma2016, aes(x = datum, y = as.numeric(value))) +
  geom_line(colour = "blue") +
  facet_wrap(~var, scales = "free") +
  labs(y = "Stoffen en eigenschappen", x = "Datum", title = " 2016: februari,
maart, april, Influent") +
  scale_x_date(breaks = "1 month", date_labels = "%b", limits =
as.Date(c("2016-02-01", "2016-05-01"))) +
  theme_solarized()

## Warning in FUN(X[[i]], ...): NAs introduced by coercion
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```





## februari, maart en april 2017

```
influentfma2017 <- influent %>%
  filter(datum >= as.Date("2017-02-01") & datum < as.Date("2017-05-01")) %>%
  gather(-datum, key= "var", value = "value")

ggplot(influentfma2017, aes(x = datum, y = as.numeric(value))) +
  geom_line(colour = "orange") +
  facet_wrap(~var, scales = "free") +
  labs(y = "Stoffen en eigenschappen", x = "Datum", title = " 2017: februari,
maart, april, Influent") +
  scale_x_date(breaks = "1 month", date_labels = "%b", limits =
as.Date(c("2017-02-01", "2017-05-01")))
```

## 2017: februari, maart, april, Influent

