

Luchtkwaliteitsdata als CSV bestanden

06 OKTOBER 2022 (VERSIE 5)

Inleiding

De website www.luchtmeetnet.nl is sinds 2014 beschikbaar en is een initiatief van het Rijksinstituut voor Volksgezondheid en Milieu (RIVM), het ministerie van Infrastructuur en Waterstaat (IenW), GGD Amsterdam, DCMR Milieudienst Rijnmond, Regionale Uitvoeringsdienst Zuid-Limburg (RUDZL), Omgevingsdienst Midden- en West-Brabant (OMWB) en Omgevingsdienst regio Arnhem (ODRA).

Via de [rapportagetool](#) op www.luchtmeetnet.nl kunnen jaarlijkse statistieken gedownload worden van de metingen van bovengenoemde meetnetten. Het downloaden van de ruwe data loopt via data.rivm.nl/data/luchtmeetnet/. Deze notitie geeft een toelichting op de gegevens die hier opgehaald kunnen worden.

Automatische metingen versus niet-automatische metingen

Grofweg kan er binnen de meetnetten een onderscheid worden gemaakt tussen twee verschillende soorten metingen, de automatische en de niet-automatische metingen. Voor de automatische metingen geldt dat de gemeten data van verschillende componenten uurlijks of dagelijks automatisch digitaal verstuurd worden naar de meetnetten en opgeslagen worden in een database. Hierdoor kunnen de concentraties van deze componenten getoond worden via verschillende kanalen, zoals bijvoorbeeld [Teletekst pagina 711](#) en www.luchtmeetnet.nl.

De niet-automatische metingen bevatten componenten die door een analyselaboratorium worden geanalyseerd en op een later tijdstip worden gerapporteerd, dit geldt onder andere voor de chemische samenstelling van fijnstof filters en de chemische samenstelling van regenwater. Ook zijn er automatische metingen die pas later in definitieve vorm beschikbaar zijn, zoals de concentratie van ozonprecursors en ammoniak. Door de langere doorlooptijd zijn de resultaten van de niet-automatische metingen van een kalenderjaar later beschikbaar, veelal zo rondom de zomerperiode.

Actueel jaar versus vastgestelde data

In het dataoverzicht wordt een onderscheid gemaakt tussen data van automatische metingen van het actuele jaar (maandbestanden) en vastgestelde data (jaarbestanden). Data van automatische metingen voor het actuele jaar zijn voorlopige gevalideerde cijfers. De uitgevoerde validatie voor de maandbestanden bestaat uit een automatische en handmatige validatie. De automatische validatie wordt uitgevoerd op basis van de statusvlaggen van het meetapparaat. Sommige statusvlaggen leiden tot een directe afkeuring, andere vlaggen tot een waarschuwing. Tijdens de handmatige validatie, die uitgevoerd wordt door een valideur, worden deze statusvlaggen en de gemeten waarden gecontroleerd.

In sommige gevallen kunnen meetwaarden afwijken vanwege (tijdelijke) veranderingen in de omgeving van een meetstation of door (tijdelijke) bijdragen van een of meerdere grootschalige bronnen. Een omleiding in het verkeer of een mesthoop in de nabijheid van een meetstation zijn voorbeelden van lokale veranderingen c.q. bronnen die de meetwaarden kunnen beïnvloeden. Vuurwerk, paasvuren en Saharazand zijn voorbeelden van grootschalige bronnen die invloed kunnen hebben op de resultaten van meerdere meetstations in de omgeving. Meer informatie over de validatie van luchtkwaliteitsdata en hoe we omgaan met (tijdelijke) afwijkingen van meetresultaten is te vinden op [deze pagina van luchtmeetnet.nl](#).

Jaarlijks wordt uiterlijk 15 maart data van automatische metingen van het voorgaande jaar vrijgegeven voor download en pas dan worden deze gegevens als definitief aangemerkt. Concreet betekent dit dat deze data voor het kalenderjaar 2020 pas op 15 maart 2021 definitief is. Voor 15 maart wordt de jaarlijkse data nogmaals uitvoerig bekeken. Een belangrijk onderdeel van deze controle is de vergelijking van de equivalente methode met de referentiemethode, bijvoorbeeld bij metingen van fijnstof. Door middel van deze vergelijking zorgen we ervoor dat na toepassing van kalibratiefactoren de uitgevoerde metingen vergelijkbaar zijn met de referentiemethode. In de voorlopige maandelijkse rapportage worden de kalibratiefactoren van het afgelopen jaar gebruikt voor het actueel jaar.

Aangezien de definitieve kalibratiefactoren pas worden vastgesteld aan het eind van het jaar is het mogelijk dat de definitieve concentraties naar boven of naar beneden worden bijgesteld ten opzichte van de eerdere rapportage in de maandbestanden. Nadat deze controles zijn afgerond zal de definitieve data als jaarbestanden in een eigen folder worden toegevoegd aan de vastgestelde data. Tot die tijd blijven de voorlopige cijfers van de automatische metingen over het recent afgeronde jaar beschikbaar.

In de map "[Actueel-jaar](#)" worden voor de automatische metingen de CSV bestanden en ZIP bestanden per maand geplaatst. Het formaat van de bestandsnaam voor de CSV bestanden is "*yyyy_MM_component.csv*" met als *yyyy* het actuele jaar en *MM* als maand. Om het downloaden makkelijker te maken, is er ook een maandelijkse zip met daarin alle componenten. Deze heeft als naamgeving "*yyyy_MM.zip*".

Vastgestelde data voor automatische metingen wordt opgeslagen in de map "[Vastgesteld-jaar](#)" waarbij elk jaar zijn eigen *yyyy*-map heeft. In deze map staan zowel de CSV bestanden ("*yyyy_component.csv*") als een ZIP bestand met alle componenten voor dat jaar ("*yyyy.zip*").

Op 1 januari vallen de metingen van het afgelopen jaar niet meer onder de map "[Actueel-jaar](#)". Deze metingen worden tijdelijk verplaatst naar de map "[Voorlopig-jaar](#)". Wanneer alle metingen van het afgelopen jaar definitief zijn, wordt de data beschikbaar gesteld via "[Vastgesteld-jaar](#)" en zal de map "[Voorlopig-jaar](#)" verdwijnen.

Voor het beschikbaar stellen van de resultaten voor de niet-automatische metingen wordt een andere route gevolgd aangezien deze resultaten pas later beschikbaar zijn. Net als bij automatische metingen worden hier ook jaarlijkse controles uitgevoerd, maar worden voorlopige maandcijfers niet gerapporteerd in de map "[Actueel-jaar](#)". Na het afronden van de controles, veelal in de zomerperiode van het actueel jaar, worden de resultaten van de niet-automatische metingen geplaatst in de desbetreffende jaarmap in de hoofdmap "[Vastgesteld-jaar](#)".

Formaat van de geleverde CSV data

Het gebruikte scheidingsteken voor de CSV bestanden is een *puntkomma* (“;”), het gebruikte decimaal teken is een *punt* (“.”). Het inlezen van de CSV bestanden in Microsoft Excel met andere instellingen voor het scheidings- en decimaalteken kan leiden tot onverwachte resultaten.

In de eerste twee kolommen is additionele metadata over het bestand opgeslagen. De eerste vier rijen van deze kolommen bevatten de volgende gegevens:

- **Datum export:** datum van de export van de gegevens in het CSV bestand.
- **Periode:** voor welke periode data is opgeslagen in het CSV bestand. De datumtijd notatie bestaat uit *yyyyMMdd hh:mm - yyyyMMdd hh:mm*, die moet worden gelezen als “vanaf – tot”.
- **Bron:** url naar de website waar dit bestand is gedownload
- **Beschrijving data:** Link naar dit bestand.

Vanaf kolom 5 wordt aanvullende meta-informatie weergegeven over het meetstation en de betreffende meetopstellingen (regels 1 t/m 7). De volgende meta-informatie is beschikbaar:

- **StationsCode:** unieke identificatiecode voor elk station
- **StationsNaam:** naam van het station
- **Latitude, Longitude:** coördinaten van het station
- **StationsGebied:** locatie van een station (regionaal, voorstad, stad, onbekend)
- **StationsType:** type station (achtergrond, verkeer, industrie)
- **Meetprincipe:** met welk principe een component bepaald wordt
- **Meetopstelling:** welke apparatuur gebruikt wordt voor het meten.

Regel 8 bevat de headers van de data, waarbij er voor gekozen is om voor de meetdata het unieke identificatiecode van elk station als header te gebruiken. Vanaf regel 9 is de data beschikbaar, beginnend met een blok van 5 kolommen met daarin de componentnaam, beschouwingsperiode, eenheid, begin- en einddatum in het *yyyyMMdd hh:mm* formaat. De weergegeven begin- en einddatums zijn altijd volgens wintertijd (ETC/GMT+1) en moeten gelezen worden als “van-tot”.

Voor de rapportage van de data onderscheiden we twee verschillende datastromen, de zogenaamde automatische en niet-automatische metingen. Automatische metingen zijn metingen waarbij de gemeten concentratieniveaus via een acquisitieproces “automatisch” in de database geladen worden en vervolgens direct openbaar gemaakt worden via diverse kanalen zoals bijvoorbeeld www.luchtmeetnet.nl. Niet-automatische metingen zijn indirecte metingen in een matrix als fijnstof of regenwater. Deze matrices worden na een periodieke bemonstering geanalyseerd in een analyselaboratorium, zoals de chemische samenstelling van fijnstof filters of regenwater. Voor de automatische datastroom bestaan er ook metingen die achteraf nog uitgebreid gecontroleerd moeten worden, zoals bijvoorbeeld ozonprecursors in lucht. Daarnaast zijn er componenten waarbij nog een omrekening moet plaatsvinden, zoals ammoniakmetingen. Voor ammoniakmetingen geldt dat deze onder omgevingscondities gemeten worden maar onder standaardcondities gerapporteerd dienen te worden.

Beide datastromen hebben een vergelijkbaar, maar toch een iets ander CSV formaat. Bij de automatische metingen worden na het informatieblok de gemeten waarden voor de betreffende component op elke individuele station weergegeven. Dit is het zogenaamde “wide” formaat. De meeste automatische metingen bevatten alleen kolommen waarin de metingen per station worden weergegeven. Uitzondering daarop is de component NH₃. Het bestand met NH₃ meetwaarden bevat

sinds 2020 naast een stationskolom ook een kolom met aanvullende opmerkingen over bron specifieke bijdragen tijdens een of meerdere meetperioden. Hiermee kunnen, afhankelijk van het gebruik, bepaalde perioden uitgesloten worden tijdens de verwerking/analyse van de data.

Bij de niet-automatische metingen worden de gegevens van elk station (voor zover van toepassing) onder elkaar geplaatst, ook wel bekend als het “long” formaat. Het inlezen van deze twee datastromen vereist dan ook een iets andere aanpak om te gebruiken in het R-package “[openair](#)”.

Voor beide datastromen geldt dat wanneer er geen meetwaarde zichtbaar is, deze waarde niet beschikbaar is. Dit kan veroorzaakt zijn door technische problemen (bijv. storing apparatuur) of door de uitgevoerde validatie. In het laatste geval voldoet een meetwaarde niet aan de kwaliteitseisen van het betreffende meetnet en is de meetwaarde om die reden afgekeurd.

In de bijlage wordt een voorbeeld script gegeven voor het inlezen van de data in R ([The R Project for Statistical Computing](#)).

Daarnaast is in de bijlage van deze notitie ook een verkorte Engelstalige versie opgenomen van de beschrijving van de data.

Negatieve concentraties

In sommige gevallen bevatten de CSV bestanden negatieve concentraties. Dit komt onder andere door de meetonnauwkeurigheid. Voor gedetailleerde informatie hierover verwijzen we u graag naar [de uitleg over negatieve waarden op www.luchtmeetnet.nl](#). Hierin wordt onder andere uitgelegd waarom wij ook negatieve meetwaarden publiceren en hoe het validatieproces van de meetwaarden in zijn werk gaat.

Overzicht veranderingen

Een compleet overzicht van alle veranderingen aan de basisdata van de CSV bestanden is te vinden op [deze pagina op www.luchtmeetnet.nl](#). Op deze pagina staat een link naar een document waarin alle wijzigingen aan vastgestelde data worden weergegeven, inclusief de datum waarop deze wijzigingen zijn doorgevoerd.

Jaarlijkse statistieken

Op de site [www.luchtmeetnet.nl](#) worden de officiële en vrijgegeven jaarlijkse statistieken, zoals bijvoorbeeld jaargemiddelden, gepubliceerd. Via de [rapportagetool](#) kunnen de jaarlijkse statistieken opgevraagd worden. Deze jaarlijkse statistieken zijn berekend op technisch en inhoudelijk gevalideerde meetdata en kunnen mogelijk bron gerelateerde verhogingen bevatten, bijvoorbeeld van vuurwerk, paasvuren, Saharazand, tijdelijke werkzaamheden nabij een meetpunt, en composthopen. Deze bron gerelateerde bijdragen worden meegenomen in de berekening van de jaarlijkse statistieken. Meer informatie hierover is te vinden op [deze pagina op www.luchtmeetnet.nl](#).

Contact

Voor vragen over de downloadbare CSV bestanden kunt u contact opnemen met het RIVM via luchtmeetnet@rivm.nl. Voor vragen over de informatie op [www.luchtmeetnet.nl](#) of de jaarlijkse statistieken kunt u contact opnemen met info@luchtmeetnet.nl.

Disclaimer

De op deze site gepresenteerde resultaten zijn ten behoeve van onderzoeksdoeleinden niet afgerond. Hierbij volgen we het [Europese uitvoeringsbesluit 2011/850/EG](#)¹ waarin staat beschreven dat de afronding de allerlaatste stap van elke berekening dient te zijn. Daarnaast dient de publicatie van deze data niet gezien te worden als een eindrapportage zoals beschreven in [NEN-EN-ISO/IEC 17025 \(2018\)](#) en bevat daarom niet alle daarvoor vereiste kenmerken van de geaccrediteerde verrichtingen van de verschillende deelnemende meetnetten.

Voor meer informatie over het toepassen van afronden in het kader van vergelijkingen met milieudoelstellingen verwijzen we u naar het [Europese uitvoeringsbesluit 2011/850/EG](#), Bijlage I, A.2.

¹ 2011/850/EU: Uitvoeringsbesluit van de Commissie van 12 december 2011 houdende uitvoeringsbepalingen van Richtlijnen 2004/107/EG en 2008/50/EG van het Europees Parlement en de Raad met betrekking tot de onderlinge uitwisseling van informatie en de verslaglegging over de luchtkwaliteit (Kennisgeving geschied onder nummer C(2011) 9068)

Appendix A: Overzicht van beschikbare componenten (Overview of available pollutants)

Bestandsnaam (file name)	Component (pollutant)	Matrix	Eenheid (units)	Tijdseenheid (time unit)
yyyy_BC	BC (black carbon)	Buitenlucht (ambient air)	µg/m ³	Uur (hour)
yyyy_CO	CO (koolmonoxide / carbon oxide)	Buitenlucht (ambient air)	µg/m ³	Uur (hour)
yyyy_H2S	H ₂ S (waterstofsulfide / hydrogen sulfide)	Buitenlucht (ambient air)	µg/m ³	Uur (hour)
yyyy_NH3	NH ₃ (ammoniak / ammonia)	Buitenlucht (ambient air)	µg/m ³	Uur (hour)
yyyy_NO	NO (stikstofoxide / nitrogen oxide)	Buitenlucht (ambient air)	µg/m ³	Uur (hour)
yyyy_NO2	NO ₂ (stikstofdioxide / nitrogen dioxide)	Buitenlucht (ambient air)	µg/m ³	Uur (hour)
yyyy_NOx	NO _x (stikstofoxiden / nitrogen oxides)	Buitenlucht (ambient air)	µg/m ³	Uur (hour)
yyyy_O3	O ₃ (ozon / ozone)	Buitenlucht (ambient air)	µg/m ³	Uur (hour)
yyyy_PM10	PM ₁₀ (fijnstof / particulate matter < 10 µm)	Buitenlucht (ambient air)	µg/m ³	Uur (hour)
yyyy_PM25	PM _{2.5} (fijnstof / particulate matter < 2.5 µm)	Buitenlucht (ambient air)	µg/m ³	Uur (hour)
yyyy_SO2	SO ₂ (zwaveldioxide / sulphur dioxide)	Buitenlucht (ambient air)	µg/m ³	Uur (hour)
yyyy_groep_BTX_lucht	benzeen (benzene) meta-paraxyleen (meta-paraxylene) tolueen (toluene)	Buitenlucht (ambient air)	µg/m ³	Uur (hour)
yyyy_groep_HC_aerosol	Cl NH ₄ NO ₃ SO ₄	Aerosol (PM ₁₀ filters)	µg/m ³	Dag (day)
yyyy_groep_HC_regenwater	Ca Cl F K25 (geleidbaarheid/conductivity) K Mg Na NH ₄ NO ₃ PO ₄ SO ₄	Regenwater (rainwater)	µmol/l	variabel (variable)
yyyy_groep_Koolstof_aerosol	EC OC	Aerosol (PM _{2.5} filters)	µg/m ³	Dag (day)

Bestandsnaam (file name)	Component (pollutant)	Matrix	Eenheid (units)	Tijdseenheid (time unit)
yyyy_groep_Lindaan_regenwater	lindaan (lindane)	Regenwater (rainwater)	µmol/l	variabel (variable)
yyyy_groep_PAK_aerosol	benz[a]-antraceen (benz[a]-anthracene) benz[a]pyreen (benz[a]pyrene) benz[b,j,k]fluorantheen (benz[b,j,k]fluoranthene) benz[g,h,i]peryleen (benz[g,h,i]perylene) chrysene (chrysene) dibenz[a,h]antraceen (dibenz[a,h]anthracene) indeno[1,2,3-cd]pyreen (indeno[1,2,3-cd]pyrene)	Aerosol (PM ₁₀ filters)	ng/m ³	Dag (day)
yyyy_groep_ZM_aerosol	As Ca Cd K Mg Na Ni Pb Zn	Aerosol (PM ₁₀ filters)	ng/m ³	Dag (day)
yyyy_groep_ZM_regenwater	As Ca Cd Co Cr Cu Fe K Mg Na Ni Pb V Zn	Regenwater (rainwater)	µmol/l	variabel (variable)
yyyy_groep_ZVOC_lucht	1,2,3-trimethylbenzeen (1,2,3-trimethylbenzene) 1,2,4-trimethylbenzeen (1,2,4-trimethylbenzene) 1,3,5-trimethylbenzeen (1,3,5-trimethylbenzene)	Buitenlucht (ambient air)	ng/m ³	Uur (hour)

Bestandsnaam (file name)	Component (pollutant)	Matrix	Eenheid (units)	Tijdseenheid (time unit)
yyyy_groep_ZVOC_lucht	1,3-butadieen (1,3-butadiene) 1-buteen (1-butene) 1-penteen (1-pentene) 2-penteen (2-pentene) acetyleen (acetylene) benzeen (benzene) cis-2-buteen (cis-2-butene) ethaan (ethane) ethylbenzeen (ethylbenzene) ethyleen (ethylene) i-butaan (i-butane) i-hexaan (i-hexane) i-octaan (i-octane) i-pentaan (i-pentane) isoprene (isoprene) meta-paraxyleen (meta-paraxylene) n-butaan (n-butane) n-heptaan (n-heptane) n-hexaan (n-hexane) n-octaan (n-octane) n-pentaan (n-pentane) orthoxyleen (orthoxylyene) propaan (propane) propeen (propene) tolueen (toluene) trans-2-buteen (trans-2-butene)	Buitenlucht (ambient air)	ng/m ³	Uur (hour)

Appendix B: R script voor het inlezen van de CSV bestanden (R script for reading the CSV files)

In this section we will provide some example R code to read the provided CSV files for the automatic measurements. In this example we will use the packages `tidyverse`, `lubridate` and `openair` to read the data, clean things up and produce a graph.

First step is to load the packages:

```
library(tidyverse)
library(lubridate)
library(openair)
```

If these packages are not available on your system, you will need to install them. Installing these packages is outside the scope of this example.

The first step we will do is load the metadata, which is located in the first 7 rows of the CSV file and consists of two separate blocks of data.

```
# store the filename in a variable
filename <- "https://data.rivm.nl/data/luchtmeetnet/Vastgesteld-jaar/2019/2019_N02.csv"

# first read the block with metadata, consisting of the first 7 rows
csv.metadata <- read.table(file = filename,
                           sep = ";",
                           header = FALSE,
                           nrows = 7)

# show contents of the variable
head(csv.metadata)

##           V1                               V2 V3 V4
## 1 Datum export                20201202 16:34 NA NA
## 2 Periode                20190101 00:00 - 20200101 00:00 NA NA
## 3 Bron                https://www.rivm.data.nl/luchtmeetnet NA NA
## 4 Beschrijving data https://www.rivm.data.nl/luchtmeetnet/info NA NA
## 5                               NA NA
## 6                               NA NA
# Note: we removed some columns for the sake of brevity.
```

We now have the metadata in R, but we need to clean things up first in order to make the data usable. Let us grab the first section, the metadata associated with the file:

```
# we only grab the data in column 2, rows 1-4.
# the next step is to transpose the data and turn it in a tibble
file.metadata <- as_tibble(t(csv.metadata[1:4,2]))

# to name the columns we grab the descriptive names from the
# metadata and use make.names to create valid names
names(file.metadata) <- make.names(tolower(csv.metadata[1:4,1]))
# this gives us a nice tibble with all the information
file.metadata

## # A tibble: 1 x 4
##   datum.export periode bron beschrijving.data
```

```
## <chr> <chr> <chr> <chr>
## 1 20201202 16:~ 20190101 00:00 ~ https://www.rivm.d~ https://www.rivm.data.nl/
~
```

Please note that we left processing the dates as an exercise to the reader.

The next step is to transform all the metadata associated with the stations. We use a similar approach to creating the variable `file.metadata`.

```
# we only grab the data from column 6 to the last column for all 7 rows.
# the next step is to transpose the data and turn it in a tibble
station.metadata <- as_tibble(t(csv.metadata[ ,6:ncol(csv.metadata)]))

# to name the columns we grab the descriptive names from the
metadata and use make.names to create valid names
names(station.metadata) <- make.names(tolower(csv.metadata[ ,5]))

# this gives us a nice tibble with all the information
head(station.metadata)

## # A tibble: 6 x 7
##   stationscode stationsnaam latitude.longit~ stationsgebied stationstype
##   <chr> <chr> <chr> <chr> <chr>
## 1 NL01485 Hoogvliet-L~ (51.867411,4.35~ stad achtergrond
## 2 NL01487 Rotterdam Z~ (51.891147,4.48~ regionaal verkeer
## 3 NL01488 Rotterdam Z~ (51.893617,4.48~ stad achtergrond
## 4 NL01489 Ridderkerk~ (51.869431,4.58~ stad verkeer
## 5 NL01491 Rotterdam-O~ (51.938472,4.43~ stad verkeer
## 6 NL01493 Rotterdam N~ (51.927111,4.46~ stad verkeer
## # ... with 2 more variables: meetprincipe <chr>, meetopstelling <chr>
```

With the metadata in place we can now read the actual data and transform it into the openair format. Please note that transforming data is not needed for the non-automatic pollutants.

```
# first we read the data, skipping the first 7 rows.
# we defined row 8 as our header row, making sure our columns
have a name.
csv.data <- read.table(file = filename,
                      sep = ";",
                      header = TRUE,
                      skip = 7)

# to name the values to the pollutant, we grab the pollutant name
pollutant <- tolower(unique(csv.data$Component))

# now we can cleanup things and turn the data in something suitable
# for use in openair
mydata <- csv.data %>%
  dplyr::mutate(
    date = lubridate::ymd_hm(Begindatumtijd, tz = "Etc/GMT-1"),
    date_end = lubridate::ymd_hm(Einddatumtijd, tz = "Etc/GMT-1")) %>%
  dplyr::select(-Component,
               -Bep.periode,
               -Eenheid,
               -Begindatumtijd,
               -Einddatumtijd) %>%
  tidyr::pivot_longer(!c("date", "date_end"),
                     names_to = "site",
                     values_to = pollutant)
```

Let us explain the things we do in the last statement. The dates are read from the CSV file as a character string. Using `lubridate` we mutate these dates towards an actual date, renaming the variables to `date` and `date_end` in the process as well. As can be seen we also apply the correct time zone (Note: Etc/GMT-1 is actually Etc/GMT+1 in R).

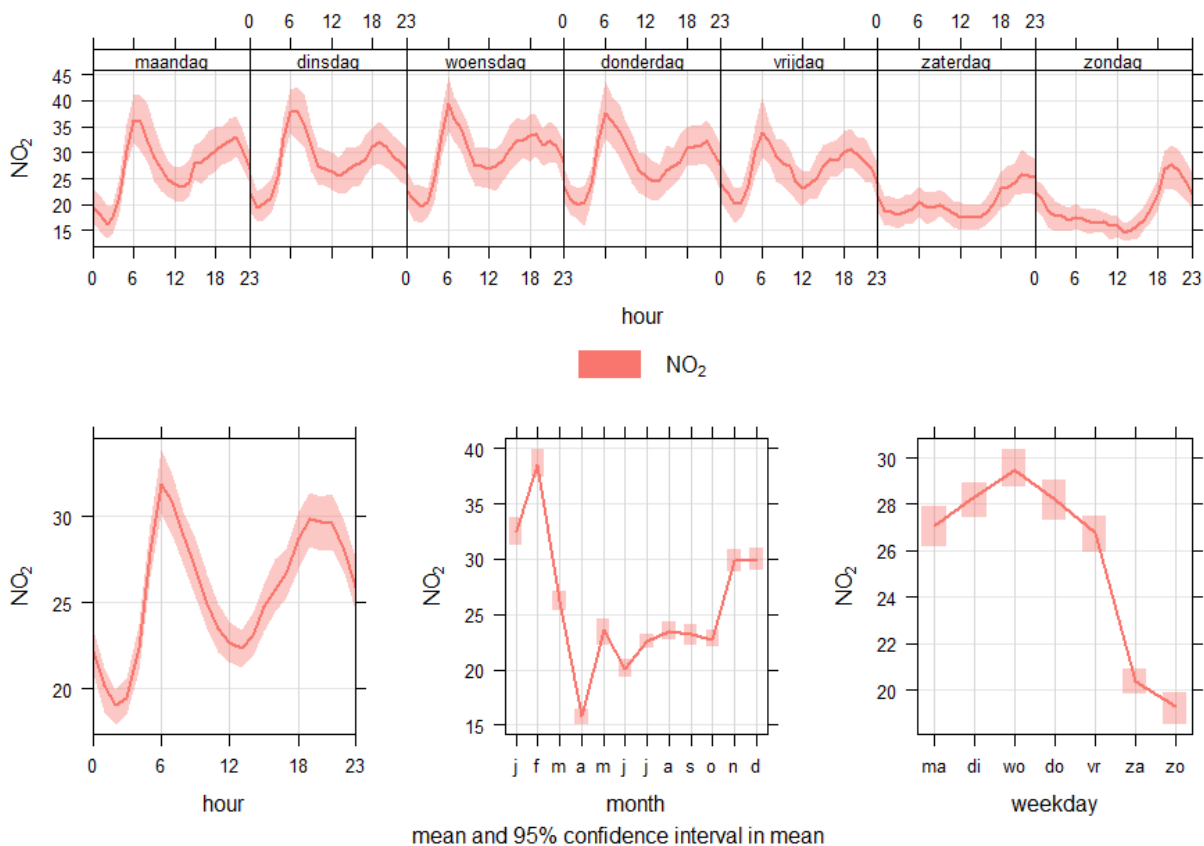
The next step, using a `select`, is used to drop some unwanted columns. After this we transform the data from wide to long, putting the column names from the wide data set in the `site` column and the values in a column with the pollutant name, in this case `no2`.

This leaves us with the following data structure:

```
head(mydata)
## # A tibble: 6 x 4
##   date           date_end           site      no2
##   <dtm>          <dtm>          <chr>    <dbl>
## 1 2019-01-01 00:00:00 2019-01-01 01:00:00 NL01485  17.2
## 2 2019-01-01 00:00:00 2019-01-01 01:00:00 NL01487  48.3
## 3 2019-01-01 00:00:00 2019-01-01 01:00:00 NL01488  54.4
## 4 2019-01-01 00:00:00 2019-01-01 01:00:00 NL01489  30.8
## 5 2019-01-01 00:00:00 2019-01-01 01:00:00 NL01491  27.9
## 6 2019-01-01 00:00:00 2019-01-01 01:00:00 NL01493  32.9
```

The data is now suitable to use in several functions of the `openair` package. As an example we will show a `timeVariation` plot for the NO₂ concentrations at station NL10641:

```
openair::timeVariation(mydata = mydata %>% filter(site == "NL10641"),
  pollutant = "no2")
```



Appendix C: Description of the data set

Automatic measurements versus non-automatic measurements

A distinction within the monitoring networks can be made between two different types of measurements, the automatic and the non-automatic measurements. For the automatic measurements, the measured data of various components are sent automatically on an hourly or daily basis to the monitoring networks and stored in a database. This allows the concentrations of these components to be shown through various channels, such as [Teletext page 711](#) and www.luchtmeetnet.nl.

The non-automatic measurements include components that are reported through an analytical laboratory, such as the chemical composition of particulate filters or rainwater, or automatic measurements that are not available in final form until later, such as the concentration of ozone precursors. Because of the longer lead time, the results of the non-automatic measurements from a calendar year are available later, usually around the summer period.

Current data set versus established data set

In the data overview of the website, a distinction is made between data from automatic measurements of the current year (monthly files) and established data sets (annual files). Data for the current year are provided as provisional data. Each year, no later than March 15, data from the previous year is released for download and only then is this data considered final. In other words, this means that the data for the year 2020 will not be final until March 15, 2021. This data will then be added to the established data sets folder as annual files in its own yearly folder. Until then, the provisional data for the recently completed year will remain available.

In the folder "[Actueel-jaar](#)" the CSV files and ZIP files related to the current year are placed per month. The filename format of the CSV files is constructed as "*yyyy_MM_component.csv*" with the current year as *yyyy* and *MM* as month. To make bulk downloading easier, there is also a monthly ZIP file containing all components. The name of this ZIP file is constructed as "*yyyy_MM.zip*".

Established data is stored in the folder "[Vastgesteld-jaar](#)" where each year has its own *yyyy* folder. This year folder contains both the CSV files ("*yyyy_component.csv*") and a ZIP file with all components for that year ("*yyyy.zip*").

A different route for making results available for non-automatic measurements is followed since these results are not available until later in the year. As with automatic measurements, annual checks are also performed, but preliminary monthly data is not reported in the "[Actueel-jaar](#)" folder. After completing the checks, usually in the summer period of the current year, the results of the non-automatic measurements are placed in the annual folder in the main "[Vastgesteld-jaar](#)" folder.

Format of the supplied CSV data

The separator used for the CSV files is a semicolon (“;”), the decimal character used is a period (“.”). Importing the CSV files into Microsoft Excel with different separator and decimal separator settings can lead to unexpected results.

Additional metadata about the file is stored in the first two columns. The first four rows in these columns contain the following information:

- **Datum export:** date of the export of the data in the CSV file.
- **Periode:** for which period data is stored in the CSV file. The date time format consists of *yyyyMMdd hh:mm - yyyyMMdd hh:mm*, which should be read as “from - to”.
- **Bron:** url to the website from which this file has been downloaded
- **Beschrijving data:** Link to this file.

From column 5, additional meta-information about the measuring station and the relevant measuring setups is displayed (lines 1 to 7). The following meta-information is available:

- **StationsCode:** unique identification code for each station
- **StationsNaam:** name of the station
- **Latitude, Longitude:** coordinates of the station
- **StationsGebied:** surrounding area of a station (regional, suburb, city, unknown)
- **StationsType:** station type (background, traffic, industry)
- **Meetprincipe:** principle used for measuring a component
- **Meetopstelling:** which equipment is used for measurement.

Line 8 contains the headers of the data. We have chosen to use the unique identification number of each station as the header for the measurement data. From line 9 the data is available, starting with a block of 5 columns containing the component name, period, measurement units, start and end date in the *yyyyMMdd hh:mm* format. The start and end dates shown are always winter time (ETC / GMT + 1) and should be read as “from-to”.

For the reporting of the data, we distinguish two different dataflows, containing the so-called automatic and non-automatic measurements. Automatic measurements are measurements with a monitor giving concentration levels instantaneous. Non-automatic measurements are measurements that take place afterwards in a laboratory, such as the composition of filters of particulate matter, or measurements requiring additional checks (measurements of ozoneprecursors in air).

The CSV format for both dataflows are similar, but also slightly different. The CSV files for the automatic measurements contain different columns, representing the air quality stations, after the information block. This is the so-called “wide” format. Most CSV files containing the results of automatic measurements only contain columns that display the measurements per air quality station. An exception to this format is the format of the NH₃ measurements. In addition to having station columns, the file with NH₃ measurement values also contains columns with additional comments about source-specific contributions during one or more measurement periods. Depending on the use of the data, this allows certain periods to be excluded during the processing/analysis.

With the format of the non-automatic measurements, the data from each station (if applicable) is stacked, also known as the “long” format. Slightly different approaches are thus needed to read the data for usage in the R-package “[openair](#)”.

If no measured value is visible, this value is not available. This can be caused by technical problems (e.g., equipment failure) or as a product of the validation process. In the latter case, a measured value does not meet the quality requirements of the relevant measuring network and the measured value is rejected for that reason.

An example script is given in the appendix for reading the data into R ([The R Project for Statistical Computing](#)).

Negative Concentrations

In some cases the CSV files contain negative concentrations. This is partly due to the measurement inaccuracy. For detailed information about this, please refer to [the explanation of negative values on www.luchtmeetnet.nl \(in Dutch\)](#). This page explains why we publish negative measured values and how the validation process of the measured values works.

Annual aggregated data

Official and publicly available annual aggregated data, such as annual averages, are published on the site [www.luchtmeetnet.nl](#). The aggregated data can be requested using the [reporting tool](#). The official and publicly available yearly aggregated data are calculated on air quality measurements which have been validated technically and based on the measured concentrations with possible source-related contributions such as fireworks, bonfires, Saharan dust, temporary work near an air quality station and compost piles. At present, these source-related contributions are included in the calculation of the annual aggregated data. More information about this can be found on [this page at www.luchtmeetnet.nl \(in Dutch\)](#).

Contact

For questions about the downloadable CSV files, please contact RIVM via luchtmeetnet@rivm.nl. For questions related to the webcontent of [www.luchtmeetnet.nl](#) or the annual aggregated data, please contact Luchtmeetnet.nl via info@luchtmeetnet.nl.

Disclaimer

For research purposes the results presented on this site have not been rounded. We follow the [European Commission Implementing Decision 2011/850/EC](#)², which states that rounding should be the very last step of every calculation. In addition, the publication of this data should not be seen as a final report as described in [NEN-EN-ISO/IEC 17025 \(2018\)](#) and therefore does not contain all the required features of the accredited operations of the various participating monitoring networks.

For more information about the application of rounding in the context of comparisons with environmental objectives, we refer you to the [European Implementing Decree 2011/850/EC](#), Annex I, A.2.

² 2011/850/EU: Commission Implementing Decision of 12 December 2011 laying down rules for Directives 2004/107/EC and 2008/50/EC of the European Parliament and of the Council as regards the reciprocal exchange of information and reporting on ambient air quality (notified under document C(2011) 9068)