

Documentatie

Programma

ALVA versie 2.5

Algemene Lucht

Vervuiling Analyse

LV2 (Lucht voor Leidschendam-
Voorburg)

Auteur: Frans Kets

De functie van ALVA is:

- Het plotten van tijdreeksen van luchtvervuiling tezamen met relevante andere parameters
- Het vergelijken van tijdreeksen, en van statistisch relevante grootheden
- Berekeningen uitvoeren op tijdreeksen, en met combinaties van tijdreeksen
- Plotten van de gemiddelden van de waarnemingen als functie van één of twee relevante parameters

Het doel is:

- Analyseren of meetapparaten en/of metingen betrouwbaar zijn
- Analyseren van door RIVM aangebrachte kalibratie procedures
- Vergelijken van vervuiling op verschillende locaties
- Het detecteren van vervuilingspieken
- Het afleiden van wetmatigheden en relaties tussen vervuiling en meteorologische parameters en/of tijd
- Controleren of de vervuiling binnen de wettelijke grenzen valt en of de vervuiling kleiner is dan de door de WHO gestelde aanbevolen bovengrenzen

Inhoud

Inleiding	5	Verkennen van de dataset; kwaliteitscontrole	45
Nieuw in ALVA 2.2	7	Tijdsafhankelijkheden: Plotten van de tijdreeksen	54
Software configuratie	9	Gedetailleerde vergelijkingen tussen twee stations	57
Runnen van het programma	10	Analyse van relaties tussen metingen .	62
De cardfile.	10	Functieafhankelijkheden	64
De interne logica van het programma	15	Definitie van de 1-D en 2-D functies.	64
Inputfiles met meetgegevens.....	17	Definitie van het toepassingsbereik: de afhankelijke parameters	74
De attributen, wat voor een soort gegevens worden gebruikt?	Fout!	Dankbetuigingen.....	77
Bladwijzer niet gedefinieerd.		Opmerkingen en Commentaar	77
Bewerking van de meetgegevens	24	Appendix A. Opzetten Python omgeving voor ALVA	78
Verwijderen van (onbetrouwbare of foutieve) meetgegevens	24	Appendix B. Voorbeelden van carddecks	80
Specifieke selecties maken van gegevens	28	Appendix C. Inlezen meetgegevens.....	85
Overbrengen van gegevens van één station naar andere stations	29	Conventies csv files.....	85
Filteren in tijd –gemiddeldes e.d.	Fout!	Programma's om gegevens op te halen.	87
Bladwijzer niet gedefinieerd.		Appendix D. Stationsgegevens in StationDirectory22.py	90
Combineren van tijdreeksen.....	32	Appendix E. Attriboot gegevens in StationDirectory22.py	93
Bijhouden van wettelijke grenzen	34	Appendix F. Belangrijkste rijksomgevingswaarden luchtkwaliteit ...	94
Berekenen van pseudostations	35		
Analyse en Plotten.....	39		
Instellingen	39		
Output van de analyses.....	44		

Kaarten gebruikt in ALVA carddecks

Algemeen:

9	DEBUG	Specificeert genereren van diagnostische output
	STOP	Beëindigt het inlezen van een kaartdeck
11	DATADIRECTORY	Specificeert de directory waar de input files staan
12	KADER	Specificeert de stations met gegevens van sensormetingen
	MEETNET	Specificeert gecertificeerde metingen van het Landelijke Lucht Meet Netwerk.
	CLUSTER	Specificeert verzamelingen van stations die onder één naam bekend zijn.
13	NIETGEBRUIKEN	Specificeert uitzonderingen op de lijst van de stations die op de KADER kaart werden vermeld.
13	KNMI_GEGEVENS	Specificeert de file in de inputdirectory met KNMI meetgegevens
	KNMI_EXTRA	Specificeert de file met meteorologische gegevens die niet in de KNMI files staan.
14	PERIODE	Specificeert het tijdsbereik waarover berekeningen worden gedaan.
14	ANALYSEPERIODE	Specificeert de tijdsperiode voor de analyses en de plots.
14	PERIODE4STAT	Specificeert de begin- en eindtijd van de meetreeksen voor één station
	SUBSET	Specificeert een selectie van stations gebaseerd op een aantal eigenschappen

Bewerken van tijdreeksen (meetreeksen/attributen) van een station en combineren van deze tijdreeksen.

17	ELIMINEER	Wissen van onbetrouwbare gedeeltes van een tijdserie
17	GRENSWAARDEN	Specificeert het bereik waarbinnen de waarden van de ingelezen meetreeksen (attributen) zich moeten bevinden.
18	UITSCHIETER	Specificeert het bereik waarbinnen de waarden van de ingelezen meetreeksen (attributen) zich moeten bevinden in vergelijking met een glijdende mediaan.
18	DAGUITSCHIETER	specificeert het bereik waarbinnen de waarden van de ingelezen meetreeksen (attributen) zich moeten bevinden in vergelijking met een glijdende mediaan, genomen over een periode van drie uur per dag over een aantal dagen.
18	VERVANG_WAARDE	Vervangt één specifieke waarde van een tijdreeks door een andere.
	OPSCHONEN	
19	SCENARIO	Maakt een nieuwe meetgegevensreeks aan met een selectie gebaseerd op specifieke tijden
	MOMENTOPNAME	Diagnostische tekst, met waarden van meetreeksen op een bepaalde tijd, of waarde van gemiddelde en standaarddeviatie van de tijdreeksen in een bepaalde periode.
19	KOPIER_ALS	Maakt een nieuwe meetgegevensreeks aan met een selectie van gegevens gebaseerd op een criterium
20	OVERBRENGEN	Kopieert een meetgegevensreeks van één enkel station naar een meetgegevensreeks op een ander station, of naar meerdere andere stations
21	DAG	Berekenen van een gemiddelde of een andere statistiek over de gemeten waarden gedurende één dag
21	VLOEIEND	Berekend een tijdreeks waarbij de variaties binnen een bepaald tijdvenster zijn geëlimineerd of deze juist worden berekend.
22	DAGRITME	Berekent een meetwaarde waarbij gegevens van verschillende dagen worden gecombineerd om een beste schatting te maken.
22	BEREKEN_ATTR	Berekent een meetreeks/attribuut uit een combinatie van andere meetreeksen/attributen.

23	OVERSCHRIJDING	Maakt een tijdreeks aan die oploopt van 0 naar het aantal geconstateerde overschrijdingen van een grenswaarde.
----	----------------	--

Combineren van stations; aanmaken van pseudostations:

25	SAMENVOEGEN	Specificeert de berekening van een pseudostations met de gemiddelden, standaard deviaties e.d. van meetreeksen voor een stel stations.
26	T_SHIFT	Specificeert een verschuiving in tijd die op de meetreeksen van een station wordt toegepast.
27	BEREKEN_STAT	Berekent een pseudostation uit een combinatie van andere stations.

Analyseren en plotten:

30	PLOTPARAM	Specificeert of op het scherm wordt geplot of dat de plots in een directory worden opgeslagen
30	BEMONSTERING	Specificeert de bemonsteringsfrequentie (sampling rate) van de te plotten tijdreeksen.
31	ANALYSEPERIODE	Specificeert de tijdsperiode voor de analyses en de plots.
	AANPASSING	Specificeert veranderingen in formaat, kleuren e.d. van plots
32	OVERZICHT	Maakt een heatmap die aangeeft welke gemeten grootheden (de attributen) verkregen zijn per meetstation.
35	DOOSDIAGRAM	Specificeert doosdiagrammen (boxdiagrams).
	VIOOLPLOT	Specificeert vioolplots (violinplots).
36	KRUISCORRELATIE	Specificeert berekening van een kruiscorrelatie matrix
37	PLOT	Specificeert tijdreeksplots voor een station
40	PLOT_SOORT	Specificeert tijdreeksplots voor een type meetreeks
40	VERGELIJK_PAREN	Specificeert een plot waarbij een aantal attributen wordt geplot voor twee verschillende stations bij elkaar
44	VERGELIJK_ATTR2STAT	Specificeert een plot waarbij één attribuut wordt geplot voor twee verschillende stations in een kruisdiagram.
44	PUNTGRAFIEK	Specificeert een scatterplot waarbij een tweetal attributen tegen elkaar worden uitgezet
46	1DFUNCTIE_VAN	Specificeert de onafhankelijke parameter voor 1D functies.
49	2DFUNCTIE_VAN	Specificeert de 2 onafhankelijke parameters voor 2D functies.
55	KNMI_FUNCTIE_ITEMS	Specificeert de KNMI meetgegevens als de afhankelijke parameters voor functie plots.
56	FUNCTIE_ITEMS	Specificeert afhankelijke parameters voor functie plots (per station).
57	SOORT_FUNCTIE_ITEMS	Specificeert afhankelijke parameters voor functie plots (per meetreeks).

Inleiding

ALVA - de acroniem staat voor “Algemene Lucht Vervuiling Analyse” - is een programma waarmee reeksen meetgegevens afkomstig van Citizen Science en RIVM luchtkwaliteit-metstations in samenhang met meetreeksen van een nabij KNMI weerstation kunnen worden bewerkt en geanalyseerd.

De functie van ALVA is:

- Het plotten van tijdreeksen van luchtvervuiling tezamen met relevante andere parameters
- Het vergelijken van tijdreeksen, en van statistisch relevante grootheden
- Berekeningen uitvoeren op tijdreeksen, en met combinaties van tijdreeksen
- Plotten van de gemiddelden van de waarnemingen als functie van één of twee relevante parameters

Het doel is:

- Analyseren of meetapparaten en/of metingen betrouwbaar zijn
- Analyseren van door RIVM aangebrachte kalibratie procedures
- Vergelijken van vervuiling op verschillende locaties
- Het detecteren van vervuilingsspieken
- Het afleiden van wetmatigheden en relaties tussen vervuiling en meteorologische parameters en/of tijd
- Controleren of de vervuiling binnen de wettelijke grenzen valt en of de vervuiling kleiner is dan de door de WHO gestelde aanbevolen bovengrenzen

Al deze bewerkingen worden aangestuurd door een reeks instructies in een tekstfile – hier aangeduid als carddeck of cardfile. Elke lijn, ofwel elke kaart in deze cardfile is één instructie voor één soort bewerking op tijdreeksen (attributen) van één of meerdere stations. Zo heb je instructies om de stations te selecteren die je in de analyse wil meenemen, instructies om duidelijk foutieve gegevens in meetreeksen ofwel te elimineren ofwel te corrigeren, instructies om een datareeks met een andere datareeks te combineren, instructies om gegevens te plotten, en instructies om de relatie tussen een aantal meetreeksen duidelijk te maken.

Als eenvoudig voorbeeld het volgende carddeck, waarmee we kijken naar de meetgegevens van een aantal RIVM en CSLV stations (CSLV is de afkorting voor de Paddenstoelen in gebruik in Leidschendam-Voorburg):

#	← # op eerste positie betekent – geen instructie.
Datadirectory 2020CumDatafile	← Dit geeft de folder aan met de meetgegevens
Periode 1/12/2018 31/1/2020 Jaaroverzicht 2019	← We verzamelen gegevens van deze periode
Analyseperiode 1/1/2019 31/12/2019	← Periode beperking voor de uiteindelijke plots
KNMI_gegevens uurgeg_215_2011-2020.txt	← De file met gegevens van KNMI station Voorschoten
Kader CSLV RIVM	← Alle CSLV stations (paddenstoelen) en RIVM stations
#	
# data processing - zet vooralsnog ruime grenzen	← # betekent – geen instructie; wel commentaar
#	
Grenswaarden 0 200 NO2	← Alleen NO2 meetwaarden tussen 0 en 200 toegestaan
Grenswaarden 0 200 P10	← Alleen PM10 meetwaarden tussen 0 en 200 toegestaan
#	
# geef bemonstering aan voor de plots	
#	
Plotparam SAVEONLY	← Plots als plaatsjes in files, niet op het scherm
Bemonstering GEEN1HPLOT	← Geen bemonstering (sampling) met 1 uur interval
Bemonstering D	← Eén meetpunt per dag op de plots
#	
Overzicht NO2 P10 ALL	← Maak een overzichtsplot voor zowel NO2 als PM10
Plot NO2 P10 RIVM 1	← Maak een plot voor RIVM station 1 met NO2 als PM10

Het zal duidelijk zijn dat de gegevens achter het pijltje geen deel uitmaken van de kaarten. Ook is het zo dat de KNMI gegevens in het carddeck hierboven eigenlijk niet nodig zijn.

Door het carddeck te veranderen naar gelang van wat men ziet in de resulterende plots, kan men focussen op de interessante gegevens in de dataset. Zo kan men bijvoorbeeld controleren of een daggemiddelde een wettelijke grenswaarde heeft overschreden. Voor PM10 geldt dat een 24 uur gemiddelde van $50 \mu\text{g}/\text{m}^3$ maximaal 35 keer per jaar mag worden overschreden.

Toevoeging van de “OVERSCHRIJDING” kaart leidt ertoe dat een nieuwe tijdreeks wordt berekend, die het aantal overschrijdingen bijhoudt (met een naam die door de gebruiker gekozen wordt – hieronder P10X). Ook die tijdreeks kunnen wij in een OVERZICHT plot laten zien, en ook in een plot, waarbij we één en dezelfde attribuut (meetreeks) voor alle stations tegelijk laten zien (PLOT_SOORT kaart).

```
.....
Grenswaarden          P10 0 200
Overschrijding  DAG 50 P10X P10 ALL
#      ...
#
Plotparam ...
#
Overzicht P10X ALL
Plot_soort P10X ALL
```

← Overschrijdingen berekend voor alle stations

← Maak een overzichtspot voor de overschrijdingen

Het aantal overschrijdingen van de $50 \mu\text{g}/\text{m}^3$ grens blijkt heel gevoelig te zijn voor fouten en afwijkingen in de tijdreeksen. De bron van fouten: de levensduur van de apparatuur is relatief kort, en het is lang niet altijd a priori duidelijk dat de apparatuur nog een juist signaal afgeeft. Analyses en correcties zijn daarom nodig. RIVM doet dat voor een deel: er is een kalibratieprocedure voor NO₂ metingen met de zogenoemde Paddenstoelen, en voor de P10 en P25 metingen voor de meest recente meetstations. Voor het merendeel van de resterende afwijkingen zit er niets anders op om een gedeelte van de tijdreeks ongeldig te verklaren en te verwijderen uit de berekeningen en de plots. Na dit werk kan de verdere analyse beginnen: het herkennen van afhankelijkheden van weersomstandigheden, het ontdekken van patronen, het detecteren van abnormaal gedrag, de hoge pieken of de periodes dat het vuil uit de lucht is verdwenen, het controleren of er wettelijke grenzen zijn overschreden, het correleren met weer en wind.

Nieuw in ALVA 2.2

- Carddecks worden gelezen vanuit een subdirectory van de directory waarin de ALVA programma modules aanwezig zijn. De debugfiles worden ook in een subdirectory geplaatst.
- Files met meetgegevens, waarbij de tijd van meten is aangeduid in Universal Time (UTC) eenheid zoals gedownload van luchtmeetnet.nl en gedownload van de meest recente versie van de RIVM website kunnen worden ingelezen.
- Identificatie van gekalibreerde en ruwe metingen bij het inlezen afhankelijk van type csv file. Ruwe metingen van PM₁₀, PM₂₅, NO₂ worden voorzien van het suffix “_m”: P10_m, P25_m, NO2_m.
- De software is voorbereid op het inlezen van KNMI gegevens voor de periode 2021-2030.
- Met de bewerking ELIMINEER verwijdert men een gedeelte van een tijdreeks op basis van een begin- en een eindtijd.
- Met de bewerking OVERBRENGEN wordt een tijdreeks van één station naar een ander station overgebracht.
- De toegestane afwijking op de UITSCHIETER en DAGUIJSCHIETER kaart is nu een absolute grootte (was relatief).
- Een functie van windrichting (WINDRICHTING of DWIND) wordt nu geplot in een polaire plot, met aanduiding van de windstreken.
- Aangeven van onzekerheidsmarges in FUNCTIE plots is als 95% betrouwbaarheidsinterval (was standaard deviatie).
- Met de kaart “FUNCTIE_ITEMS” krijgt men een functieafhankelijkheid van verschillende attributen van een station op één plot getekend.
- Met de windroosplots wordt drie extra plots geleverd: een polaire plot van de optelsom van de gemiddelde vervuiling per windrichting en windsnelheid als functie van de windsnelheid, alsook de cumulatieve plot van de frequentie van voorkomen en de cumulatieve plot van de som van de vervuiling. De eerste plot geeft de richting aan waar de vervuiling vandaan zou komen als elke windrichting even vaak zou voorkomen, de tweede de heersende windrichting, de derde de vervuiling over de meetperiode.
- De stijl van de plot kan worden veranderd (plots met witte achtergrond in plaats van plots met donkere achtergrond). Standaard zijn plots met lichte achtergrond.
- Ook de soort plotfile kan worden veranderd.

Nieuw in ALVA 2.3

- Plotfiles worden niet langer geplaatst in een subdirectory van een datadirectory, maar in een subdirectory van een directory waarvan de naam refereert naar de naam van de datadirectory.
- Gegevens met metingen van Palmesbuisjes (één meting per tijdperiode) worden gelezen uit een directory “Extra” (PALMES kaart). Ook gegevens van meteorologische tijdreeksen die niet van KNMI afkomstig zijn kunnen uit deze directory worden ingelezen (KNMI_EXTRA kaart).
- Bij het inlezen kunnen clusters van stations worden gespecificeerd. Daardoor kan men bij de kaart invoer één enkele naam specificeren waardoor een aantal stations worden geselecteerd (CLUSTER kaart).
- Het inlezen van een carddeck kan worden onderbroken door het invoeren van een STOP kaart. Dit kan nuttig zijn als men stap voor stap debugfiles wil genereren en bekijken.
- Om te zorgen dat een serie tijdreeksen van een bepaald meettype maar van verschillende stations beter vergeleken kunnen worden, kan men met de OPSCHONEN kaart ervoor zorgen dat voor alle tijden waarop één van die meetreeksen een NaN (Not a Number) waarde heeft, de andere meetreeksen in die verzameling van stations ook op NaN worden gezet.

- Met de VIOOLPLOT kaart kan men violinplots oftewel Vioolplots maken. Deze geven de mediane waarde aan van een tijdreeks en de geschatte distributie in amplitude.
- Met de VERGELIJK_AMPL kaart krijgt men statistische gegevens voor de vergelijking van tijdreeksen van hetzelfde type maar van verschillende stations.

Nieuw in ALVA 2.5

- Kaarten waarin plots worden gespecificeerd kunnen nu gegeven worden tussen andere kaarten – er hoeft niet meer gewacht te worden tot alle processing stappen zijn afgerond.
 - Layout van plots en kleurkeuze kan worden gewijzigd door middel van specificaties op een AANPASSING kaart. Wat precies kan worden gewijzigd hangt van de soort plot af.
 - Diagnostische operatie MOMENTOPNAME – geeft waarde van tijdreeks op een bepaalde tijd (of gemiddelde over een tijdvak)
 - Tijdsperiodes kunnen nu begrensd worden met een precisie van uren (PERIODE, ANALYSEPERIODE, PERIODE4STAT kaarten, op eenzelfde manier als ELIMINEER kaarten. Omdat men ook een tijdsperiode kan definiëren die buiten een tijdsinterval valt is de noodzaak van de PERIODE4STAT kaart vervallen.
 - De tijdsperiodes worden gedefinieerd in één bepaald formaat, dat voor alle kaarten geldt.
 - De aanduiding ALL voor de selectie van stations hoeft niet meer gegeven te worden op een groot aantal kaarten. Als deze aanduiding ontbreekt wordt de aanduiding ALL aangenomen. De aanduiding ALL+ is nog steeds noodzakelijk als alle stations plus de pseudostations moeten worden meegenomen.
 - De volgorde van de stations in de definitie van stations op de KADER kaart of in een selectie op een kaart waarmee plots worden gemaakt is nu bepalend voor de volgorde van stations op de plot.
 - Men kan op de kaarten variabelen definiëren die moeten worden vervangen door vooraf gedefinieerde woord reeksen (VERVANG kaart).
- *ALVA 2.5 is work in progress. De afleiding van functies is veranderd, met een nieuwe kaartlayout. In de code zijn een aantal nieuwe functies toegevoegd die nog niet zijn gedocumenteerd (bijv. Bayes). De functionaliteit voor de berekening van 2D functies en Windroses is nog niet aangepast aan de nieuwe versie en is dus afwezig.*

Software configuratie

ALVA is geschreven in Python 3, en kan worden opgestart via een commando in een console, waaruit Python 3 aanroepbaar is, of worden gedraaid vanuit de tekst editor Spyder. Het opzetten van een software omgeving waarbij dit mogelijk is wordt beschreven in Appendix A.

Er wordt uitgegaan van een specifieke directory structuur.

Het programma ALVA bestaat uit vier modules, StationDirectory25.py, ALVAUtilities.py, ALVA25INPUTMOD.py en ALVA25MAIN.py, beschikbaar in een hoofddirectory (Current Working Directory – CWD).

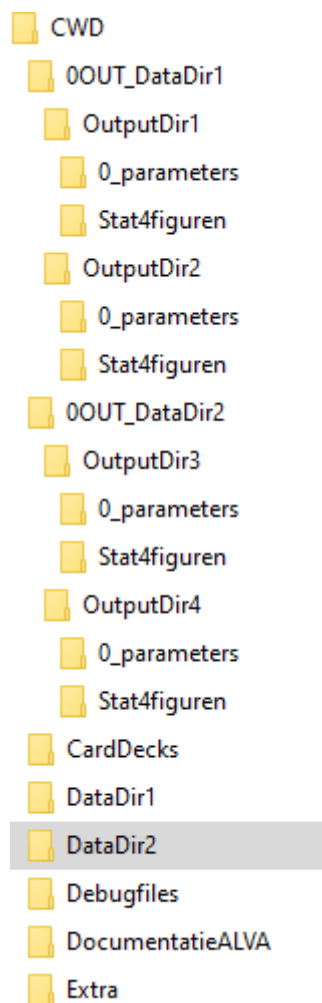


Fig. 1 – Directory Structuur

De gebruiker draait het programma door ALVA25MAIN.py aan te roepen. StationDirectory25.py bevat de namen en details van de workstations en de meetreeksen, waarmee ALVA werkt – in principe is dit de enige module die

door gebruikers van ALVA veranderd moet worden als stations toegevoegd moeten worden of als andere types metingen beschikbaar komen. Ook kan men in deze modules het type plotfile en de stijl van de plots aanpassen.

Bij het draaien van het programma wordt eerst een cardfile met opdrachten voor ALVA ingelezen vanuit een subdirectory van CWD /Carddecks. De naam van de cardfile wordt gespecificeerd in een tekstfile ALVAINPUTFILE.txt (ook in /Carddecks) of wordt gespecificeerd als een parameter in de aanroep in de command line. Daarna wordt gecontroleerd of inputfiles met meetgegevens van RIVM en van KNMI aanwezig zijn in een datadirectory (subdirectory van CWD – hier aangegeven met DataDir1 of DataDir2). De KNMI file wordt, indien nodig, omgezet naar een kleinere file waarbij alleen de meteogegevens van de voor de analyse relevante tijdsperiode zijn geselecteerd. Meetgegevens van Palmesbuisjes of enkele meetreeksen verkregen vanuit andere bronnen dan RIVM en KNMI moeten aanwezig zijn in een directory “Extra” en worden daaruit ingelezen. Bij het inlezen van de cardfile wordt gecontroleerd of de opdrachten in de cardfile fouten bevatten; pas daarna worden de opdrachten uitgevoerd.

De berekeningen in ALVA25MAIN kent in principe drie fases:

1. Het inlezen van de relevante RIVM en KNMI gegevens (plus eventueel gegevens uit directory Extra)
2. Het bewerken van de tijdreeksen en het berekenen van tijdreeksen uit bestaande tijdreeksen indien nodig.
3. Het op verschillende manieren plotten van de oorspronkelijke meetresultaten en van berekende tijdreeksen.

De instructies voor al deze bewerkingen worden in de cardfile gespecificeerd. De volgorde van de uitvoering van de bewerkingen is in principe gegeven door de volgorde van de instructies in het cardfile.

Met elke run van het programma kunnen plots worden geproduceerd in een directory (de outputdirectory) in een directory waarvan de naam refereert naar de datadirectory met inputfiles zowel als naar de naam van het carddeck. De outputdirectories gemaakt met gegevens van

DataDir1 komen terecht in subdirectories van directory OUT_DataDir1. Indien de outputdirectory (nog) niet bestaat, wordt deze door ALVA aangemaakt. Alle plots hebben begeleidende tekst met statistische informatie over de getoonde gegevens. Deze teksten worden apart verzameld

in txt files in een subdirectory van de output directory. Ook in een extra subdirectory: een kopie van het oorspronkelijke carddeck
Diagnostische gegevens (debugfiles) worden tijdens het draaien van het programma verzameld in een subdirectory van de CWD.

Runnen van het programma

Men kan ALVA22MAIN runnen via een command line (in de current working directory):

- 1 met het commando “python ALVA25MAIN.py”, waarbij de naam van de cardfile gespecificeerd moet zijn in een tekstfile ALVAINPUTFILE.txt (in subdirectory /Carddecks), of
- 2 met het commando “python ALVA25MAIN.py parameterfile_bla bla.txt”, met parameterfile_bla bla.txt als voorbeeld van de naam van een carddeck met de instructies. De outputdirectory krijgt dan de naam blabla.

Als alternatief:

- 3 men kan de modules laden in Spyder, en ALVA22MAIN runnen vanuit Spyder. In dat geval krijgt men ook de Python 3 source code van de modules van ALVA te zien. Ook dan kan de carddeck worden gespecificeerd in “Command line options” of in de tekstfile ALVAINPUTFILE.txt.

De naam van de cardfile is een txt file, met suffix “.txt”. Deze suffix kan in de commandline of in de specificatie in file ALVAINPUTFILE.txt worden weggelaten.

Terwijl het programma loopt, kan diagnostische output gegenereerd worden in een aantal tekst files, debug0.txt t/m debug9.txt. Deze worden geplaatst in een subdirectory van de current working directory (/Debugfiles), en worden bij elke run overschreven.

Relevante kaart in het carddeck:

DEBUG. Specificeert welke debugfiles worden gecreëerd. Aanbevolen is alle debugfiles te laten aanmaken, voor het geval er iets mis gaat: de gegevens in deze files zijn nuttig bij de foutopsporing. Debugfiles worden bij een volgende run van het programma overschreven.

DEBUG kaart	Geeft specificatie waar in het programma diagnostische output wordt gegenereerd
Velden	
Kop	DEBUG : definieert de kaart
Optioneel: Parameter 1 – 10	Getal (0,1,2,..9) aangevend welke debug files met diagnostische output worden gegeneerd
• Indien geen DEBUG kaart wordt gegeven, wordt er geen diagnostische output gegenereerd. • Indien er een DEBUG kaart wordt gegeven zonder paramaters, wordt volledige diagnostische output gegenereerd.	

De cardfile.

Indien de naam van de cardfile niet wordt gespecificeerd in de commandline (optie 2 hierboven) moet deze worden gespecificeerd in een tekstfile ALVAINPUTFILE.txt. Deze file moet zich bevinden in de subdirectory van de current working directory, net zoals de cardfile (subdirectory: /Carddecks).

De tekstfile hoeft maar één regel te bevatten, met de naam van de tekstfile, gevolgd door een mogelijk korte tekst ter verdere karakterisering.

Regels in de tekstfile ALVAINPUTFILE.txt beginnend met een blank of met een # worden overgeslagen.

Indien de cardfile gespecificeerd wordt als een parameter in de commandline hoeft er geen file ALVAINPUTFILE.txt aanwezig te zijn.

De cardfile bevat de instructies voor het programma (één instructie per kaart, of beter gezegd, één instructie per regel) op grond waarvan het programma een bewerking uitvoert – bijvoorbeeld een analyse of het maken van een plot. Een trefwoord als eerste woord op de regel (de kop) bepaalt de actie die ondernomen moet worden. Additionele parameters worden hierna gegeven, en zijn van elkaar gescheiden door spaties. Dit document beschrijft welke soort instructies bestaan, en welke parameters kunnen worden gebruikt. De parameters kunnen additionele informatie bevatten over de soort bewerking, zij kunnen aanduiden welke tijdreeksen moeten worden behandeld (de attributen), of zij kunnen aangeven voor welke stations de actie moet worden ondernomen.

De instructies worden in dit document beschreven: elke instructie wordt beschreven in een kader. In het carddeck staan alle parameters van zo'n instructie op één kaart. De kop van de instructie is in dit document aangegeven in hoofdletters. In het carddeck kan men ook kiezen voor kleine letters, of voor een hoofdletter gevolgd door kleine letters. Dat kan de leesbaarheid van het carddeck vergroten.

Het carddeck kan lege regels bevatten, of regels die beginnen met het karakter #. Deze regels worden bij het inlezen genegeerd.

Als een kaart van het carddeck begint met een onbekende kop, geeft dit een foutmelding.

Instructies die het inlezen beïnvloeden

De kaartfile wordt in zijn geheel ingelezen. Als er een probleem is waarbij men naar de debugfiles wil gaan kijken, is het nuttig het inlezen te onderbreken. Dit kan men forceren door een STOP kaart te geven.

Relevante kaart in het carddeck:

STOP. Beëindigt het inlezen van het kaartdeck.

STOP kaart	Beëindigt het inlezen van het kaartdeck.
Velden	
Kop	STOP : definieert de kaart

Velden in een kaartdeck kunnen vervangen worden door een naam, die eerder in het kaartdeck is gedefinieerd. Deze naam moet beginnen met een dollarteken. Om deze naam te definiëren wordt de VERVANG kaart gebruikt.

VERVANG kaart	Specificeert een parameter die, als hij gevonden wordt in een kaartdeck, vervangen wordt door een reeks woorden
Velden	
Kop	VERVANG : definieert de kaart
Eerste veld	Parametrisatie naam: het woord dat staat voor de woorden als gedefinieerd in de verdere velden. Het woord moet beginnen met een dollarteken : \$
Tweede veld en verder	Definitie van de parametrisatie: reeks woorden die worden gerepresenteerd door het woord in het eerste veld. Deze woorden mogen niet beginnen met een dollarteken.
- Als er maar één veld aanwezig is wordt de parametrisatie opgeheven. - Een tweede VERVANG kaart met dezelfde parametrisatie naam maar een andere definitie vervangt de eerste definitie door de tweede. - Voorbeeld: VERVANG \$AQ NO2 P10 P25 OVERZICHT \$AQ PLOT \$AQ	

Voorbeelden van cardfiles zijn te vinden in Appendix B.

Weergave van stations, meetreeksen en tijdgegevens op de kaarten van het carddeck

Kaarten in het carddeck bevatten de instructies voor een taak. Die taak is specifiek voor datareeksen van bepaalde stations in een tijdvak. De kaart is daarom ingedeeld in velden die specifiek zijn voor de taak, en velden die de tijdgegevens, meetreeksen en stations weergeven. Bij plottaken wordt ook informatie over het bemonsteringsinterval van de tijdreeks vermeld.

In de meeste gevallen is de indeling als volgt:

Kop	Taak info	Tijdinfo Blok	Meetreeks Blok	Station Blok
-----	-----------	---------------	----------------	--------------

Taakinfo:

De velden die de taak informatie beschrijven kunnen teksten en getallen bevatten.

Tijdinfo blok:

Het tijdinfo blok omvat informatie over gewenste bemonstering (bij een aantal taken, in het bijzonder de plottaken) en over het tijdbereik. Het maakt niet uit of eerst de bemonstering wordt opgegeven en daarna het tijdbereik of omgekeerd.

Bemonstering:

De taken worden in het algemeen uitgevoerd met een bemonsteringsfrequentie van 1 uur. Men kan (bij sommige taken) overgaan naar een andere bemonstering door een andere bemonsteringsfrequentie (sampling interval) op te geven. Deze nieuwe bemonsteringsfrequenties worden als volgt aangegeven:

- nS: samoling interval wordt n seconden, met n (optioneel) een geheel getal (bijv. 20S – 20 seconden)
- nT: sampling interval wordt n minuten, met n (optioneel) een geheel getal
- nH: sampling interval wordt n uren, met n (optioneel) een geheel getal
- nD: sampling interval wordt n dagen, met n (optioneel) een geheel getal
- nW: sampling interval wordt n weken, met n (optioneel) een geheel getal
- nM: sampling interval wordt n maanden, met n (optioneel) een geheel getal

Tijdbereik:

Het tijdbereik gedeelte van het tijdinfo blok (**Tijdblok**) bevat of één of tijdgrenzen, mogelijk voorafgegaan met aanduidingen als TOT,VOOR,NA,VANAF, PUNT of twee tijdsgrenzen, mogelijk voorafgegaan met aanduidingen als BUITEN, TUSSEN en BINNEN.

Velden Tijdblok	
Eerste parameter [niet verplicht]	Optie: TOT,VOOR,NA,VANAF, PUNT als er één tijdgrens wordt gegeven; BUITEN, TUSSEN BINNEN met twee grenzen Indien niet gegeven, wordt BINNEN verondersteld of PUNT.
Tweede parameter	Datum voor de eerste tijdgrens (in formaat (d)d/m(m)/jjjj)
Derde parameter [niet verplicht]	Tijd van de dag van eerste tijdaanduiding (getal van 0 t/m 23) [indien niet gegeven wordt 0 uur Nederland lokale tijd aangenomen]
Vierde parameter	Datum voor de tweede tijdaanduiding (in formaat (d)d/m(m)/jjjj)

Vijfde parameter [niet verplicht]	Tijd van de dag voor tweede tijdaanduiding (getal van 0 t/m 23) [indien niet gegeven wordt 23 uur Nederland lokale tijd aangenomen]
Bij BINNEN en TUSSEN worden de grenzen meegenomen in de berekeningen van de taak. Bij BUITEN omvatten de berekeningen de tijdgrenzen niet. Bij VOOR, TOT en NA reikt het rekenbereik van de taak tot de tijdgrens, maar bevat de tijdgrens niet. Bij VANAF en PUNT omvat het rekenbereik de tijdgrens.	

Meetreeks Blok:

Het meetreeks blok bevat de namen van meetreeksen die bij het programma bekend zijn. Namen zijn bekend als de meetreeksen zijn ingelezen vanaf de datadirectory (files met station gegevens en de gegevens van een KNMI station) of als de tijdreeks is gegenereerd in een eerdere stap in ALVA (dus als output is gegenereerd als gevolg van een eerdere actie / kaart in het carddeck).

Sensor gegenereerde gegevens

De sensor gegevens die als csv files per station worden ingelezen als gespecificeerd op de KADER/NIETGEBRUIKEN kaarten, zijn door de sensoren gegenereerd, of door RIVM toegevoegd als vergelijkingsmateriaal of als additionele versies van dezelfde metingen.

De naamgeving van de metingen varieert. Op de kaarten wordt alleen gewerkt met de namen van de parameters die in de berekeningen worden gebruikt. NO2 staat dan voor de beste NO2 waarde die wordt aangeleverd. Deze kan op een csv file als tijdreeks NO2_kal bekend staan. Intern wordt deze tijdreeks in de tijdreeks NO2 omgezet, terwijl de tijdreeks NO2 in dezelfde csv file wordt omgezet in een tijdreeks NO2_m (NO2_gemeten). Dit hangt af van het type station (voor details; zie Appendix C).

In de plots wordt een uitgebreidere beschrijving van de grootte gehanteerd.

De namen van de metingen staan gecodeerd in StationDirectory25.py, subroutine attribdir() – zie Appendix E.

De volgende waarden worden herkend en kunnen worden gebruikt in de verschillende displays:

NO2, NO2_ref1, NO2_ref2, NO2_m	Stikstofoxide concentratie.
P10, P10_ref1, P10_ref2, P10_m	PM10 concentratie. Ref1 en ref2 zijn referentie stations, P10 is inclusief een door RIVM gegeven kalibratie.
P25, P25_ref1, P25_ref2, P25_m	PM2.5 concentratie. Ref1 en ref2 zijn referentie stations, P2.5 is inclusief een door RIVM gegeven kalibratie.
temp, rh, ld	Temperatuur (Celsius) als gemeten bij de sensor: temp Relatieve luchtvochtigheid gemeten bij de sensor in percent: rh Luchtdruk (millibar) gemeten bij de sensor: ld
FN, O3, NO en SO2	Metingen verkregen bij sommige RIVM stations. FN: fumée noire - zwarte rook - staat voor Roet O3: ozon NO: stikstofmonoxide SO2: zwaveldioxide

Gegevens verkregen van een KNMI station

De volgende waarden worden herkend en kunnen worden gebruikt in de verschillende displays:

- Dwind: Windrichting (in graden) gemiddeld over de laatste 10 minuten van het afgelopen uur (Conventie in ALVA is: 0 = Oost, 90 = Noord, 180 = West, 270 = Zuid, 360 = Oost, NaN = veranderlijk. Deze verschilt van de KNMI conventie: 360=noord, 90=oost, 180=zuid, 270=west, 0=windstil 990=veranderlijk).

- Dwind10 : Dwind/10
- Dwind9: Dwind/9
- Vwind: Uurgemiddelde windsnelheid (in 0.1 m/s).
- Windstoot: Hoogste windstoot (in 0.1 m/s) over het afgelopen uurvak.
- TempKNMI: Temperatuur (in 0.1 graden Celsius) op 1.50 m hoogte tijdens de waarneming.
- DauwTemp: Dauwpuntstemperatuur (in 0.1 graden Celsius) op 1.50 m hoogte tijdens de waarneming.
- LVocht: Relatieve vochtigheid (in procenten) op 1.50 m hoogte tijdens de waarneming.
- RegenDuur: Duur van de neerslag (in 0.1 uur) per uurvak.
- Neerslag: Uursom van de neerslag (in 0.1 mm) (-1 voor <0.05 mm)
- Ldruk: Luchtdruk (in 0.1 hPa) herleid naar zeeniveau, tijdens de waarneming.
- Ldruk960: Ldruk – 960 mbar
- Zicht: Horizontaal zicht tijdens de waarneming (0=minder dan 100m, 1=100-200m, 2=200-300m,..., 49=4900-5000m, 50=5-6km, 56=6-7km, 57=7-8km, ..., 79=29-30km, 80=30-35km, 81=35-40km,..., 89=meer dan 70km). *Deze waarde kan worden gebruikt om mist te herkennen.*
- Instraling: Globale straling (in J/cm2) per uurvak.
- T_gemidd: Temperatuur gemiddeld over een tijdsbereik van 1 dag. Dit daggemiddelde wordt berekend over de periode van één dag in UTC tijd, (zeg Greenwich tijd zonder zomertijd). Het gemiddelde wordt dus genomen van ongeveer 2 uur 's nachts tot 2 uur 's nachts in Nederlandse tijd (zomer).
Opmerking: *Met de instructie op kaart DAG kan de gemiddelde temperatuur over een dag ook worden berekend. Het gemiddelde wordt dan genomen over de metingen op één dag gerekend in lokale Nederlandse tijd.*
- T_stdev: Standaard deviatie van ware temperatuur in het tijdsbereik van 1 dag. Hiervoor geldt hetzelfde tijdsbereik als T_gemidd.
Opmerking: *Met de instructie op kaart DAG kan de standaard deviatie van de temperatuur over een dag ook worden berekend. De standaarddeviatie wordt dan genomen van de metingen op één dag gerekend in lokale Nederlandse tijd.*

Datum en Tijd

In ALVA wordt de dag van de week bijgehouden (getal 0 t/m 6: 0 = maandag – “WeekDayNr”); het uur van de dag (“Tijd”), het uur in de week (“Weektijd”), het nummer van de week in het jaar (“WeekNr”) en het nummer van de maand in het jaar (“Maand”), en het nummer van de dag in het jaar (getal van 1 t/m 365): “Jaardag”.

Deze waarden kunnen worden gebruikt in de analyses en in berekeningen.

Door ALVA gegenereerde (tussen)resultaten

ALVA genereert altijd een tijdreeks met de constante waarde 1. Deze heeft de naam “_1_”.

De meeste acties die op grond van een kaart uit het carddeck worden uitgevoerd geven een tijdreeks als resultaat. Deze naam krijgt een door de gebruiker gekozen naam. De naam van dit resultaat (de uitvoer van de taak) wordt altijd gegeven voorafgaand aan het Meetreeks blok.

De naam van dit resultaat kan in volgende berekeningen als invoer worden gegeven.

Station Blok:

Stations worden gedefinieerd door de naam van het type station en door een nummer. De naam van het typenummer wordt voorafgegaan door een tilde (~), als de meetgegevens van dit station zijn ingelezen met behulp van de RIVM API (waarbij een csv file wordt gebruikt waarbij de tijdinformatie in UTC wordt gegeven). Als er geen tilde wordt gegeven zijn de meetgegevens ingelezen met oudere software (getrivm).

De volgende types meetstations worden herkend:

Op een kaart wordt één meetstation opgegeven als type meetstation gevolgd door een nummer. Als men alle meetstations van een bepaald type moet hebben, volstaat de opgave van het type. Als men alle stations moet hebben kan men in plaats van types op te geven volstaan met de aanduiding ALL.

De verzameling meetstations wordt altijd aangegeven als de laatste set parameters op een kaart. In principe geeft men op de KADER kaart de verzameling aan. Als de verzameling meetstations eenmaal is gedefinieerd door de KADER kaart, kan men bij volgende kaarten het veld voor de meetstations leeg laten.

Op de kaartbeschrijvingen worden de velden waar de station definities staan aangegeven als “Station Blok”. Hierbij wordt dan bedoeld:

Velden Station blok	
Eerste Parameter	Meetstation type. Toegestaan zijn: • CSLV - paddenstoel • ~CSLV - paddenstoel, ingelezen met RIVM API, csv files met tijdaanduiding in UTC • XCSLV - paddenstoel – ongekalibreerd 2018 • GEM - station van gemeente Leidschendam-Voorburg (2018) • RIVM - officieel RIVM station • ~RIVM – officieel RIVM station, csv met tijdaanduiding in UTC • LTD - Lufdaten station • ~LTD - Lufdaten station, csv met tijdaanduiding in UTC • SB - SPPS fijnstof meter • ~SB - SPPS fijnstofmeter, csv met tijdaanduiding in UTC • PM - Palmesbuisjes Eventueel de aanduiding van een pseudostation’ • _PS
Tweede parameter en verder	Stationnummer: Getal dat meetstation identificeert voor vorige meetstation type (bij pseudostations een naam i.p.v. een stationnummer). Óf Type Meetstation, gevolgd door stationnummer of een Type meetstation
Opmerkingen	• ALL of ALL+ kan de bovenstaande parameters vervangen. De combinatie Type Meetstation en ALL of ALL+ is niet toegestaan. ALL+ omvat pseudostations, ALL omvat deze niet. • Als er na een Type Meetstation geen stationnummer wordt gegeven, worden alle stations van dat type meegenomen. • Het hele blok kan worden weggelaten mits de stations eerder zijn gedefinieerd op de KADER kaart. Dit komt overeen met de aanduiding ALL. Uitdrukkelijk: voor pseudostations moet of de pseudostation naam worden gebruikt of de aanduiding ALL+.

Uitzonderingen

Men kan meetreeksen overbrengen van één station naar een ander station, of men kan gegevens van verschillende stations combineren. Het resultaat van bewerkingen op verschillende stations kan dan een “pseudostation” opleveren. De layout van de kaarten is dan anders dan hier aangegeven.

Men kan ook analyseren hoe een bepaalde parameter afhangt van een andere parameter (bijvoorbeeld – is er een relatie tussen luchtvervuiling en windsnelheid?). Ook dan is de layout wat anders.

De interne logica van het programma

De berekeningen verlopen in verschillende stappen:

1. Inlezen van de cardfile.
 - a. Het controleren van de gegevens voor zover mogelijk. Als er problemen worden gevonden kan dit aanleiding geven tot een foutmelding. In sommige gevallen zal de fout zo belangrijk worden gevonden, dat het programma wordt afgebroken.
2. Inlezen van meetresultaten van meetstations. De in te lezen waarden worden gespecificeerd in KADER, NIETGEBRUIKEN, PERIODE, PERIODE4STAT en KNMI_GEGEVENS kaarten.
 - a. Inlezen van KNMI data, en omzetten naar een tijdelijke file.
 - b. Inlezen van meetgegevens van meetstations (per maand).
 - c. Inlezen van de gegevens van het KNMI en koppelen aan de gegevens van de meetstations.
3. Bewerken en combineren van de meetgegevens zoals die in stap 2 zijn aangegeven, of voor een subset van deze groep gegevens. Deze bewerkingen worden uitgevoerd in de volgorde van de opdrachten in het kaartdeck.
 - a. Bewerking van de meetgegevens per station en het berekenen van door de gebruiker gedefinieerde attributen/meetreeksen (bijvoorbeeld een attribuut dat verkregen wordt door een gewogen optelling van twee attributen van een station, of een vloeiende versie van een tijdreeks).
 - b. Creëren van pseudo stations met meetreeksen die berekend zijn door het combineren van meetreeksen van verschillende stations.
 - c. Bewerking van meetgegevens van pseudostations (alleen voor enkele instructies toegestaan).
4. Resampling, of in het Nederlands: herbemonstering naar andere tijdsintervallen.
5. Het maken van diagnostische plots van de meetreeksen, van berekende meetreeksen/attributen, het vergelijken van meetreeksen. Deze stap wordt uitgevoerd voor de tijdreeks zoals gespecificeerd op de PERIODE kaart, maar kan verder worden beperkt door een specificatie op een ANALYSEPERIODE kaart. Dit gebeurt voor alle gespecificeerde sampling intervallen.

Deze bewerkingen worden uitgevoerd in de volgorde van de opdrachten in het kaartdeck.
6. Het berekenen van gemiddeld gedrag van de metingen als functie van tijd van de dag o.i.d. en als functie van meteorologische omstandigheden (metingen van het KNMI). Dit gebeurt voor de tijdreeks als gespecificeerd op de PERIODE kaart, eventueel beperkt door de ANALYSEPERIODE kaart.

De logica is veranderd in versie 2.5 – stappen als resampling, gebruik van gemiddelden en het maken van plots worden nu gezien als normale bewerkingen en kunnen worden gespecificeerd tussen alle andere bewerkingen.

Modificeren van het programma

ALVA is modulaair opgezet, en, omdat het python programma altijd is meegeleverd, kan worden uitgebreid door de gebruiker. Elke module in het programma wordt in principe bestuurd met één kaart van het kaartdeck. Een aantal aanwijzingen hoe het programma is gestructureerd is gegeven in Appendix G.

Inlezen van de files met meetgegevens

Type files met meetgegevens

De meetgegevens van de meetstations moeten beschikbaar zijn in een subdirectory van de current working directory (de directory waar de modules ALVA25MAIN.py, ALVA25INPUTMOD.py en StationDirectory25.py aanwezig zijn). De naam van de subdirectory wordt gespecificeerd op één van de kaarten van de cardfile – de DATADIRECTORY kaart. De datadirectory kan via een aantal subdirectories gekoppeld zijn aan de hoofddirectory.

De gegevens van de meetstations in een bepaalde maand moeten beschikbaar worden gemaakt in csv (Comma separated values) files. Deze csv files hebben een naam, afgeleid van de naam van het station, met een aanduiding van de maand. De gegevens voor verschillende maanden worden door ALVA aan elkaar gekoppeld. Voor een jaaroverzicht van één station moeten dus zo'n twaalf files worden ingelezen.

De header van de csv file geeft de namen van de metingen. Csv files geven in feite een tabel, met kolommen waarvan de naam gegeven wordt in de bovenste rij van de tabel. Als voorbeeld: de figuur hieronder geeft het begin van de file file _NL10446 - Den Haag-Bleriotlaan_201912.csv met de metingen van RIVM station NL10446. In de kolom met de naam kit staat de naam van het station (dezelfde waarde voor de hele kolom); de naam periode identificeert de kolom met het tijdstip van meting (er is één meting op elk uur); de namen NO2, P10, P25, temp, rh en ld geven de type metingen aan.

kit,periode,NO2,P10,P25,temp,rh,ld NL10446,2019-12-01 00:00:00,54.25,33.87,,,,,

Deze csv files worden gemaakt door hulpprogramma's die de gegevens van de website van RIVM lezen. Bovenstaand voorbeeld is geproduceerd door een programma getRIVM.exe (versie Diederik van Hemert) dat de gegevens van samenmeten.rivm.nl uitleest (zie voor beschrijving van formaten en programma's: Appendix C).

Het scheidingsteken in deze csv files volgt de internationale (Amerikaanse) conventie. Het scheidingsteken is een komma (",").

De selectie van de files gebeurt door de specificatie van de datadirectory, de types stations en voor elk type de specifieke stations. De groep stations die ingelezen moeten worden is gespecificeerd in een KADER kaart. Het kan nuttig zijn om op zo'n KADER kaart een groep stations van een type aan te geven ook indien die groep meer stations bevat dan moeten worden gebruikt in de analyses; bijvoorbeeld door aan te geven dat men de stations van het type "RIVM" wil inlezen. Welke stations van deze groep niet moeten worden gebruikt wordt vervolgens gespecificeerd op een NIETGEBRUIKEN kaart. Alleen die files worden ingelezen waarin zinnige gegevens staan voor de tijdsperiode van de analyses. De tijdsperiode wordt aangegeven op de PERIODE kaart. Sommige stations hebben alleen zinnige gegevens in specifieke periodes. Om niet zinnige data weg te gooien kan men de PERIODE4STAT of de ELIMINEREN kaart gebruiken.

Het aantal meetstations is groot, en daarom zijn opties ingebouwd om de specificatie van meetstations te vergemakkelijken. Men kan verzamelingen van stations met één naam aanduiden (CLUSTER kaart), of men een keuze maken van de stations op basis van een eigenschap (SUBSET kaart).

Meetgegevens verkregen met metingen met palmesbuisjes zijn niet georganiseerd per maand per meetstation. De files met deze meetgegevens zijn georganiseerd per meetperiode, en één file bevat de gegevens van meerdere meetstations.

Metingen met Palmesbuisjes leveren één meetpunt van gemiddelde stikstofdioxide concentratie gedurende een bepaalde meetperiode. De meetmethode is gebaseerd op diffusie van de stikstofdioxide in het buisje.

Een gespecialiseerd laboratorium berekent uit de verdeling van de stikstofdioxide in het buisje de gemiddelde concentratie van stikstofdioxide gedurende de meetmethode. Normaliter wordt per meetstation twee buisjes gebruikt. Het laboratorium rapporteert metingen van een project met een aantal meetstations in één enkele Excel file voor elke meetperiode.

LV2 Lucht voor Leidschendam Voorburg				project: 10295		
Resultaten NO2 analyse set 518						
Blauw code	Omschrijving	Startdatum	Starttijd	Einddatum	Eindtijd	Concentratie NO2 [µg.m-3]
518_1	LVW_1	11-2-2021	11:40	10-3-2021	11:02	21
518_2	LVW_1	11-2-2021	11:40	10-3-2021	11:02	21
518_3	LVW_2	11-2-2021	12:08	10-3-2021	11:11	23
518_4	LVW_2	11-2-2021	12:08	10-3-2021	11:11	21
518_5	LVW_3	11-2-2021	12:15	10-3-2021	11:15	24
518_6	LVW_3	11-2-2021	12:15	10-3-2021	11:15	23
518_7	LVW_4	11-2-2021	12:23	10-3-2021	11:17	25
518_8	LVW_4	11-2-2021	12:23	10-3-2021	11:17	25
518_9	LVW_5	11-2-2021	12:36	10-3-2021	11:23	26
518_10	LVW_5	11-2-2021	12:36	10-3-2021	11:23	26
518_11	LVW_6	11-2-2021	12:43	10-3-2021	11:28	26
518_12	LVW_6	11-2-2021	12:43	10-3-2021	11:28	28
518_13	LVW_7	11-2-2021	12:50	10-3-2021	11:34	25
518_14	LVW_7	11-2-2021	12:50	10-3-2021	11:34	26
518_15	LVW_8	11-2-2021	13:03	10-3-2021	11:40	27
518_16	LVW_8	11-2-2021	13:03	10-3-2021	11:40	27

Een voorbeeld van zo'n Excel file wordt hierboven gegeven. De meetstations zijn LVW-1, LVW-2, enz.

Een ander project kan in plaats van de hierboven gegeven omschrijving een omschrijving hebben van de vorm van een getal gevolgd door een spatie en een adres. De stationnaam krijgt dan de vorm van de projectnaam gevolgd door een underscore en het getal dat aan het adres voorafgaat.

De Excel file moet als csv file worden opgeslagen, en kan dan worden ingelezen in ALVA. Deze csv file moet aanwezig zijn in de directory "Extra", een subdirectory van de Current Working Directory (CWD) waar het ALVA programma is opgeslagen.

De stations (onder de naam "omschrijving") staan beschreven in de StationDirectorySpec, en hebben als karakteristiek dat de stationsnaam altijd begint met drie letters die het project aangeven (hierboven "LVW") gevolgd door een underscore en een getal. De Excel files (in feite csv files) gaan per meetperiode, en de naamgeving moet ook beginnen met de naam van het project (dus ook "LVW").

De stations met Palmesbuisjes worden met PM aangeduid. Indien PM is aangegeven op de KADER kaart, worden de csv files door ALVA opgehaald uit de "Extra" directory.

Verder geldt:

- Het programma leest csv files met het scheidingsteken volgens de Europese conventie (scheidingsteken is een puntkomma (";")) maar kan ook files volgens de internationale (Amerikaanse) conventie aan (scheidingsteken is een komma (",").
- De conventie voor het aangeven van een reëel getal op de csv file: het decimale scheidingsteken is een punt (Amerikaanse conventie).

De KNMI gegevens worden verkregen door downloads van het adres <https://www.knmi.nl/nederland-nu/klimatologie/uurgegevens> . Deze downloads zijn van een lange periode (2011-2020 of 2021-2030), en de gegevens van elk van deze periodes wordt als één geheel ingelezen.

De file(s) met KNMI gegevens wordt intern omgezet in een andere file, waarbij alleen de relevante gegevens voor de gevraagde tijdsperiode wordt bewaard. Het is mogelijk om aan te geven bij volgende runs van het programma dat de omzetting al is gebeurd, en dat een tijdelijke file aanwezig is (niet aanbevolen).

Meteorologische gegevens vanuit een andere bron (www.eumetrain.org) worden op een andere manier ingelezen. Het inlezen van deze gegevens wordt beschreven bij de beschrijving van de KNMI_EXTRA kaart.

Relevante kaarten in het carddeck:

DATADIRECTORY

Deze kaart specificeert de subdirectory van de cwd waarin de meetgegevens en de KNMI gegevens te vinden zijn.

DATADIRECTORY kaart	Specificeert de directory waar de input files staan (files genoemd op KAART, NIETGEBRUIKEN en KNMI_GEGEVENS kaarten)
Velden	
Kop	DATADIRECTORY : definieert de kaart
Parameter 1	Folder in Current Working Directory
Optioneel: Parameter 2 en verder	Folder in folder gegeven door vorige parameter
- De inputdirectory wordt gegeven door: Cwd \\ parameter 1 \\ parameter 2 \\ parameter 3 - Indien er geen DATADIRECTORY kaart wordt gegeven, of een DATADIRECTORY kaart zonder parameters, is de inputdirectory gelijk aan de Current Working Directory	

KADER

De KADER kaart specificeert de luchtmeetstations met relevante meetgegevens. De namen waaronder de stations bekend staan zijn gecodeerd in StationDirectory25.py, en bestaan uit een mnemonic voor het type station en een nummer. Er worden op het ogenblik tien types meetstation onderscheiden (definities zijn gecodeerd in file StationDirectory25.py en kunnen naar behoefte van de gebruiker worden aangepast – zie Appendix D).

De volgorde van de parameters op de kaart is: type meetstation, gevolgd door nul, óf één of meer nummers die het station van dit type aangeven. Indien er geen nummer wordt gegeven worden alle stations van dit type voor zover ze beschikbaar zijn in de datadirectory meegenomen. Na zo'n specificatie kan een tweede type station worden aangegeven, etc. Indien men alle stations van welk type dan ook wil specificeren, kan men als eerste parameter ALL invullen.

KADER kaart	Specificeert de stations met gegevens van sensormetingen
Velden	
Kop	KADER : definieert de kaart
Volgende veld tot eind van de kaart	Station Blok. <i>ALL of ALL+ of een blanco veld i.p.v. ALL is niet toegestaan.</i>
- KADER kaart is verplicht. Er is maar één KADER kaart toegelaten. - DATADIRECTORY kaart, indien aanwezig, moet voorafgaan aan KADER kaart. - KADER kaart moet voorafgaan aan display en analyse kaarten - Men kan ook de naam van één of meer clusters gebruiken in de definitie van de stations, mits de clusters eerder in het carddeck zijn gedefinieerd. Zie de CLUSTER kaart hieronder. - Stationtype en nummer moet bekend zijn in StationDirectory25.py (subroutine stationdir()). Zie Appendix D.	

INLEZEN VAN GEGEVENS VAN PALMESBUIJSJES

Metingen met Palmesbuisjes leveren één meetpunt van gemiddelde stikstofdioxide concentratie gedurende een bepaalde meetperiode. De meetmethode is gebaseerd op diffusie van de stikstofdioxide in het buisje.

Een gespecialiseerd laboratorium berekent uit de verdeling van de stikstofdioxide in het buisje de gemiddelde concentratie van stikstofdioxide gedurende de meetmethode. Normaliter wordt per meetstation twee buisjes gebruikt. Het laboratorium rapporteert metingen van een project met een aantal meetstations in één enkele Excel file voor elke meetperiode.

LV2 Lucht voor Leidschendam Voorburg				project: 10295		
Resultaten NO2 analyse set 518						
Blauw code	Omschrijving	Startdatum	Starttijd	Einddatum	Eindtijd	Concentratie NO2 [µg.m-3]
518_1	LVW_1	11-2-2021	11:40	10-3-2021	11:02	21
518_2	LVW_1	11-2-2021	11:40	10-3-2021	11:02	21
518_3	LVW_2	11-2-2021	12:08	10-3-2021	11:11	23
518_4	LVW_2	11-2-2021	12:08	10-3-2021	11:11	21
518_5	LVW_3	11-2-2021	12:15	10-3-2021	11:15	24
518_6	LVW_3	11-2-2021	12:15	10-3-2021	11:15	23
518_7	LVW_4	11-2-2021	12:23	10-3-2021	11:17	25
518_8	LVW_4	11-2-2021	12:23	10-3-2021	11:17	25
518_9	LVW_5	11-2-2021	12:36	10-3-2021	11:23	26
518_10	LVW_5	11-2-2021	12:36	10-3-2021	11:23	26
518_11	LVW_6	11-2-2021	12:43	10-3-2021	11:28	26
518_12	LVW_6	11-2-2021	12:43	10-3-2021	11:28	28
518_13	LVW_7	11-2-2021	12:50	10-3-2021	11:34	25
518_14	LVW_7	11-2-2021	12:50	10-3-2021	11:34	26
518_15	LVW_8	11-2-2021	13:03	10-3-2021	11:40	27
518_16	LVW_8	11-2-2021	13:03	10-3-2021	11:40	27

Een voorbeeld van zo'n Excel file wordt hierboven gegeven. De meetstations zijn LVW-1, LVW-2, enz.

Een ander project kan in plaats van de hierboven gegeven omschrijving een omschrijving hebben van de vorm van een getal gevolgd door een spatie en een adres. De stationnaam krijgt dan de vorm van de projectnaam gevolgd door een underscore en het getal dat aan het adres voorafgaat.

De Excel file moet als csv file worden opgeslagen, en kan dan worden ingelezen in ALVA. Deze csv file moet aanwezig zijn in de directory "Extra", een subdirectory van de Current Working Directory (CWD) waar het ALVA programma is opgeslagen.

De stations (onder de naam "omschrijving") staan beschreven in de StationDirectorySpec, en hebben als karakteristiek dat de stationsnaam altijd begint met drie letters die het project aangeven (hierboven "LVW") gevolgd door een underscore en een getal. De Excel files (in feite csv files) gaan per meetperiode, en de naamgeving moet ook beginnen met de naam van het project (dus ook "LVW").

De stations met Palmesbuisjes worden met PM aangeduid. Indien PM is aangegeven op de KADER kaart, worden de csv files door ALVA opgehaald uit de "Extra" directory.

Verder geldt:

- Het programma leest csv files met het scheidingsteken volgens de Europese conventie (scheidingsteken is een puntkomma (";")) maar kan ook files volgens de internationale (Amerikaanse) conventie aan (scheidingsteken is een komma (",")).
- De conventie voor het aangeven van een reëel getal op de csv file: het decimale scheidingsteken is een punt (Amerikaanse conventie).

Men selecteert de gewenste Palmesbuisjes stations met PM gevolgd door het nummer van het station. Men kan met de CLUSTER kaart ook zo'n selectie maken: bijvoorbeeld door te definiëren CLUSTER PMS PM 1 5 7 12 19, en dan overall waarbij men deze specifieke stations wil specificeren de afkorting PMS te gebruiken.

CLUSTER

Bij het invullen van kaarten in het carddeck moet men vaak een lijst stations opgeven. Men kan ervoor kiezen een veel voorkomende lijst stations van te voren te definiëren als een cluster. Dan kan men in plaats van de lijst stations één enkele naam aangeven, de clusternaam. Deze clusternaam moet dan echter wel uniek zijn. De clusternaam en de aanduidingen van de stations in een cluster worden gegeven op een CLUSTER kaart. Er kunnen meerdere clusterkaarten worden opgegeven. Een clusterkaart kan ook namen van eerder opgegeven clusters bevatten.

CLUSTER kaart	Specificeert een cluster van stations met gegevens van sensormetingen
Velden	
Kop	CLUSTER : definieert de kaart
Eerste veld	Naam van de cluster. Moet uniek zijn.
Verdere velden	Station Blok + eerder gedefinieerde clusternamen
• De aanduiding ALL is niet toegestaan. • DATADIRECTORY kaart moet voorafgaan aan CLUSTER kaart. • Een CLUSTER kaart kan ook verwijzen naar eerder gedefinieerde clusters: Voorbeeld: CLUSTER Voorburg LTD 8 10 CLUSTER Leidschendam LTD 12 18 CLUSTER LV Voorburg Leidschendam • De CLUSTER kaart met een bepaalde clusternaam moet voorafgaan aan alle kaarten waarin de clusternaam wordt gebruikt om een verzameling meetstations aan te geven. • Een clusternaam kan maar één keer worden gedefinieerd. • Stationtype en nummer van genoemde stations moet bekend zijn in StationDirectory25.py (subroutine stationdir()). Zie Appendix D.	

NIETGEBRUIKEN

De KADER kaart geeft een verzameling stations aan. Als er één of meer stations niet meegenomen hoeven te worden, kan men dit aangeven op één of meer NIETGEBRUIKEN kaarten.

NIETGEBRUIKEN kaart	Specificeert uitzonderingen op de lijst van de stations die op de KADER kaart werden vermeld.
Velden	
Kop	NIETGEBRUIKEN : definieert de kaart
Volgende veld tot eind van de kaart	Station Blok. <i>ALL of ALL+ of een blanco veld i.p.v. ALL is niet toegestaan.</i>
• KADER kaart moet voorafgaan aan NIETGEBRUIKEN kaart(en). • NIETGEBRUIKEN kaarten moet voorafgaan aan display en analyse kaarten • Stationtype en nummer moet bekend zijn in StationDirectory25.py (subroutine stationdir()). Indien een station niet bekend is wordt een waarschuwing gegeven. • In tegenstelling tot de conventie bij de KADER kaart mag men niet een hele groep van één type meetstation uitsluiten met deze kaart. Er moet altijd een stationnummer worden gespecificeerd. • De aanduiding ALL is niet toegestaan	

SUBSET

De KADER kaart geeft een verzameling stations aan. Men kan uit de verzameling van deze stations een keuze maken op basis van bepaalde eigenschappen die in de database van ALVA bekend zijn. Men kan zowel data uitsluiten of data selecteren.

SUBSET kaart	Specificeert een conditie op basis waarvan een selectie kan worden gemaakt de lijst van de stations die op de KADER kaart werden vermeld.
Velden	
Kop	SUBSET : definieert de kaart
Parameter 1	òf "Niet" of "Geen" : data uitsluiten op basis van eigenschap òf data selecteren op basis van eigenschap. Toegestaan: 'Organisatie' 'PM' 'Cluster'

Parameter 2	Als Parameter 1 gelijk is aan 'Niet' of 'Geen' dan moet hier de soort eigenschap staan waarop de selectie wordt gebaseerd. Anders: Als Parameter 1 gelijk is aan PSensor (type fijnstofsensor): 'SPS30', 'SDS011', 'PPD42', 'Onbekend' Als Parameter 1 gelijk is aan Organisatie: 'LV2', 'DCMR', 'RIVM', 'CLAIRE', 'Lansingerland', 'Gemeente', 'Onbekend', 'Test' Als Parameter 1 gelijk is aan Cluster: 'Voorburg-West', 'Voorburg station', 'Sluis', 'The Mall', 'Het Lien', 'Rijswijk', 'Leidschenveen', 'Benoordenhout', 'Nieuwerkerk aan den IJssel', 'Rotterdam', 'Ypenburg', 'Den Haag', 'Nootdorp', 'Onbekend'
Parameter 3	Als Parameter 1 gelijk is aan 'Geen' dan komt afhankelijk van de soort eigenschap in parameter 2 de specificaties zoals vermeld in parameter 2
- KADER kaart moet voorafgaan aan SUBSET kaart(en). - SUBSET kaarten moet voorafgaan aan display en analyse kaarten - Voorbeeld: KADER SUBSET Niet PSensor PPD42	

KNMI_GEGEVENS

Met deze kaart geeft men aan wat de naam is van file met KNMI gegevens. De KNMI gegevens worden verkregen door een download van het adres <https://www.knmi.nl/nederland-nu/klimatologie/uurgegevens>. Er zijn meerdere stations in Nederland, en gegevens van elk station kunnen worden gevonden op bovenstaand adres.

KNMI_GEGEVENS kaart	Specificeert de file in de inputdirectory met KNMI meetgegevens
Velden	
Kop	KNM_GEGEVENS : definieert de kaart
Optioneel: Parameter 1	Een file met de download van KNMI-gegevens. Of de tekst: BESTAAND Een eerder gegenereerde tussenfile wordt gebruikt – zie Waarschuwing
- Indien er geen KNMI kaart wordt gegeven worden geen KNMI gegevens meegenomen in de analyse. - De naam van de KNMI file is "uurgeg_xxx_2011-2020.txt" waarbij xxx het KNMI weerstation aangeeft. Voor parameter 1 kan worden ingevuld "uurgeg_xxx_2011-2020.txt" of "uurgeg_xxx_2011-2020". De suffix wordt dan door ALVA ingevuld. - Indien een file "uurgeg_xxx_2021-2030.txt" naast "uurgeg_xxx_2011-2020.txt" aanwezig is, wordt deze ook ingelezen. Er hoeft geen additionele parameter op de KNMI_GEGEVENS kaart te worden gespecificeerd. - Als Parameter 1 niet is gespecificeerd, wordt gelezen van uurgeg_215_2011-2020.txt (station Voorschoten), en eventueel van uurgeg_215_2021-2030.txt.	
Waarschuwing: Indien optie BESTAAND is gespecificeerd wordt een file gebruikt die alleen gegevens heeft voor het tijdsbereik van de run van ALVA waarin deze bestaande file is gegenereerd. Het tijdsbereik van deze gegevens kan potentieel niet kloppen met het tijdsbereik waarin men geïnteresseerd is.	

KNMI_EXTRA

De meetgegevens bij de KNMI stations vormt de basis voor (berekende) weermodellen. Uitkomsten van de weermodellen zijn bijvoorbeeld de windsnelheid op een bepaalde hoogte. Sommige van deze parameters zijn mogelijk van belang in relatie tot luchtvervuiling.

Kaarten van dit soort gegevens kunnen worden bekeken op weerkundige websites, bijv. http://eumetrain.org/ePort_MapViewer/index.html. Door op een locatie op een kaart te klikken kan je de getalwaarden vinden. De metingen worden per drie uur gegeven.

Door handmatig deze gegevens over te brengen kan men een getallenreeks construeren die het verloop van één grootte op een bepaalde locatie weergeeft. Deze getallenreeks kan worden opgeslagen in een csv of txt file, die met de KNMI_EXTRA kaart door ALVA kan worden ingelezen.

KNMI_EXTRA kaart	Specificeert de file in de EXTRA directory met meteorologische gegevens
Velden	
Kop	KNMI_EXTRA : definieert de kaart
Optioneel: Parameter 1	naam van de te maken attribuut (kan worden overgeslagen - in dat geval wordt de naam uit de datafile gehaald)
Parameter 2	Een file in de directory Extra met de download van meteorologische gegevens-gegevens.
De file in Extra moet beginnen met de aanduiding van de inhoud, daarna een lijn met uur, naamcurve, datum/tijd eerste sample {in notatie 2020-11-01T00:00Z } en een multiplier daarna: datum/tijdaanduiding in UTC, gemeten waarde, 2 lege kolommen	
Voorbeeld: Boundary Layer uur,BL,2018-09-01T06:00Z,0.1 2018-09-01T06:00Z,72.4,, 2018-09-01T09:00Z,669.0,, 2018-09-01T12:00Z,1265.6,, 2018-09-01T15:00Z,,, 2018-09-01T18:00Z,217.3,, 2018-09-01T21:00Z,160.4,,	

PERIODE

Deze kaart specificeert de tijdsperiode waarvoor de berekeningen worden gedaan.

Pas op: de PERIODE kaart geeft een tijdsbereik inclusief de tijdspanne die nodig is om de gegevens te filteren, een tijdsverschuiving toe te passen, enz. Deze periode is typisch groter dan de periode waarover de plots en de analyses moeten worden uitgevoerd. Zie de ANALYSEPERIODE kaart.

PERIODE kaart	Specificeert het tijdsbereik waarover berekeningen worden gedaan.
Velden	
Kop	PERIODE : definieert de kaart
	Tijdblok Geen beschrijvingen zoals "BINNEN" of "TUSSEN" toegestaan; alleen datum en tijdaanduiding. Men kan hier geen resampling aangeven
Optioneel: Verdere velden	Tekst die als verhelderend commentaar bij plots wordt gevoegd
Het tijdsbereik op de PERIODE kaart bepaalt welke files worden ingelezen.	

De ANALYSEPERIODE kaart definieert de tijdsperiode voor plots en de analyses van afhankelijkheden van onafhankelijke parameters. Deze kan korter zijn dan de periode vermeld op de PERIODE kaart.

Relevante kaart:

ANALYSEPERIODE

ANALYSEPERIODE kaart	Specificeert de tijdsperiode voor de analyses en de plots.
Velden	
Kop	ANALYSEPERIODE : definieert de kaart
Velden	Tijdblok Toegestane beschrijving "TUSSEN", "BINNEN" Men kan hier geen resampling aangeven
Als er geen ANALYSEPERIODE kaart wordt gegeven wordt de periode genomen als gegeven door de PERIODE kaart.	

PERIODE4STAT

Als een station slechts in een beperkte tijdsperiode relevante data heeft, kan men met de PERIODE4STAT de begin- en eindtijd van de meetreeksen voor dat station specificeren.

PERIODE4STAT kaart	Specificeert de begin- en eindtijd van de meetreeksen voor één station
Velden	
Kop	PERIODE44STAT : definieert de kaart
Eerste velden	Tijdblok Toegestane beschrijving "TUSSEN", "BINNEN" Men kan hier geen resampling aangeven
	Station blok Alleen één station mag worden aangegeven
Slechts één kaart per station is toegestaan.	

Bewerking van de meetgegevens

Standaard kaartlayout

De kaarten definiëren een actie voor meetreeksen voor meetstations. De layout van een kaart hangt af van de verlangde actie, maar er is wel een redelijk vaste volgorde.

Eén input meetreeks – één output meetreeks

Eerst is er een kop – die definieert de kaart (en de actie). Vervolgens volgen een aantal parameters die bij de actie horen. Deze parameters bevatten als laatste item ook de naam van de meetreeks/attribuut die als gevolg van de actie wordt geproduceerd (de output). Dan volgt de naam van de meetreeks waarop de actie wordt toegepast (de input). Als de output meetreeks de input meetreeks moet overschrijven hoeft maar één meetreeks te worden aangegeven. Daarna volgen de namen van de stations. De namen van stations kunnen worden weggelaten als de gehele verzameling stations die met CLUSTER/KADER / NIETGEBRUIKEN /SUBSET kaarten is gedefinieerd is bedoeld.

Een voorbeeld van zo'n layout is de layout van de INTERPOLEER kaart hieronder.

Verwijderen van (onbetrouwbare of foutieve) meetgegevens

Met de PERIODE4STAT kaart (hierboven beschreven) kan men de bewerkingen op de meetreeksen van een station beperken. Dit is nuttig als er aanloop periodes te zien zijn, of als het meetstation geen valide resultaten meer geeft aan het eind van zijn levensduur.

Als er bij bepaalde metingen specifieke tijden zijn waarin de gegevens niet betrouwbaar lijken kan men per meetreeks en per station aangeven dat deze metingen verwijderd moeten worden. Dat gebeurt met de ELIMINEER kaart. De selectie van de te verwijderen gegevens is gebaseerd op een begin- en een eindtijd.

Het gebruik van deze kaart kan ook nuttig zijn om een gedeelte van de metingen weg te snijden waar geen belangstelling voor is. Een voorbeeld is het wegsnijden van 31 december en 1 januari wanneer men geïnteresseerd is in de gemiddelde fijnstof vervuiling in wintermaanden (en dat te vergelijken met de vervuiling inclusief het vuurwerk bij de jaarwisseling).

Relevante kaart:

ELIMINEER

ELIMINEER kaart	Geeft de zones in de meetreeksen aan die zo onbetrouwbaar zijn dat ze gewist worden (vervangen door “Geen Data” of in computertaal: NaN – Not a Number).
Velden	
Kop	ELIMINEER : definieert de kaart
Eerste velden	Tijdblok Toegestane beschrijving: "TOT", "VOOR", "NA", "VANAF", "BUITEN", "TUSSEN", "BINNEN", "PUNT" Met tijdstip van begintijd en tijdstip van eindtijd
Verder	Meetreeks Blok
Verder	Station Blok
Als geen station wordt opgegeven worden voor alle stations de aangegeven meetreeksen verwijderd	

Eventuele uitschieters (outliers) kunnen worden afgesneden door grenzen aan te geven waarbinnen het signaal moet vallen: dan worden deze uitschieters vervangen door “Geen Data” of in computer taal; “NaN” – Not a Number of door de minimale of maximale waarde die is aangegeven (in computer taal: clippen). Er zijn verschillende methodes om te bepalen wat een uitschieter zou zijn. De meest simpele methode is grenzen te stellen aan de grootte van het signaal. Meer complexe methodes bepalen eerst hoe het signaal er normaliter uit zou moeten zien, en zetten daarna pas de grenzen waarbinnen het signaal zou moeten vallen.

De eenvoudigste manier gebruikt grenswaarden als grenzen via de GRENSWAARDEN kaart.

Relevante kaart:

GRENSWAARDEN

GRENSWAARDEN kaart	Specificeert het bereik waarbinnen de waarden van de ingelezen meetreeksen (attributen) zich moeten bevinden.
Velden	
Kop	GRENSWAARDEN : definieert de kaart
Eerste één of twee velden (Optioneel: indien geen tekst is ingevuld wordt “GEENDATA” verondersteld)	Aanduiding: “GEEN” “DATA” of “GEENDATA” of “NaN” of “CLIP”
Eerstvolgende veld	Minimum toegestane waarde

Volgende veld	Maximum toegestane waarde
Veld (optioneel – indien afwezig wordt de input tijdreeks overschreven)	Output tijdreeks - naam tijdreeks na toepassing van taak
Veld	Input tijdreeks - Naam tijdreeks waarop taak wordt toegepast
Verdere velden	Station Blok
Meerdere GRENSWAARDEN kaarten zijn toegestaan.	

De meer geavanceerde methodes voor het elimineren van uitschieters zijn gebaseerd op het trekken van een vloeiende lijn door de tijdreeksen en het vergelijken van de uitschieters met deze vloeiende lijn. Met de optie gegeven in de UITSCIETER kaart worden de pieken vergeleken met een glijdende mediaan van het signaal. De UITSCIETER kaart is geschikt wanneer er weinig dag/nacht variatie is in het signaal. Indien er wel veel variatie is, kan beter de DAGUITSCIETER kaart worden gebruikt.

Relevante kaart:

UITSCIETER

UITSCIETER kaart	Specificeert het bereik waarbinnen de waarden van de ingelezen meetreeksen (attributen) zich moeten bevinden.
Velden	
Kop	UITSCIETER : definieert de kaart
Eerste één of twee velden (Optioneel: indien geen tekst is ingevuld wordt "GEENDATA" verondersteld)	Aanduiding: "GEEN" "DATA" of "GEENDATA" of "NaN" of "CLIP"
Eerstvolgende veld	Maximaal toegestane afwijking
Volgende twee velden (optioneel)	In één veld het woord "@Filterlengte". Het volgende veld de filterlengte in uren. De mediaan van de waarden binnen de filterlengte wordt gebruikt als referentie. Indien niet aanwezig wordt een waarde van 7 uur gebruikt.
Veld (optioneel – indien afwezig wordt de input tijdreeks overschreven)	Output tijdreeks - naam tijdreeks na toepassing van taak
Veld	Input tijdreeks - Naam tijdreeks waarop taak wordt toegepast
Verdere velden	Station Blok
Meerdere UITSCIETER kaarten zijn toegestaan.	

Relevante kaart:

DAGUITSCIETER

UITSCIETER kaart	specificeert het bereik waarbinnen de waarden van de ingelezen meetreeksen (attributen) zich moeten bevinden in vergelijking met een glijdende mediaan, genomen over een periode van drie uur per dag over een aantal dagen. De glijdende mediaan kan eventueel afzonderlijk worden uitgerekend met de DAGRITME kaart.
Velden	
Kop	DAGUITSCIETER : definieert de kaart
Eerste één of twee velden (Optioneel: indien geen tekst is	Aanduiding: "GEEN" "DATA" of "GEENDATA" of "NaN" of "CLIP"

ingevuld wordt "GEENDATA" verondersteld)	
Eerstvolgende veld	Maximaal toegestane afwijking
Volgende twee velden (optioneel)	In één veld het woord "@Dagen". In het volgende veld: aantal dagen vóór en na het tijdstip van vergelijking meegenomen in de berekening. Voorbeeld: de waarde 3 geeft aan dat de mediaan wordt berekend met gegevens van totaal 7 dagen. Indien niet aanwezig wordt een waarde van 3 dagen gebruikt.
Veld (optioneel – indien afwezig wordt de input tijdreeks overschreven	Output tijdreeks - naam tijdreeks na toepassing van taak
Veld	Input tijdreeks - Naam tijdreeks waarop taak wordt toegepast
Verdere velden	Station Blok
Meerdere UITSCHIETER kaarten zijn toegestaan.	

Sommige meetreeksen kunnen specifieke numerieke waarden bevatten die een niet valide meting aanduiden. Deze waarden moeten worden omgezet naar NaN (Not a Number).

Relevante kaart:

VERVANG_WAARDE

VERVANG_WAARDE kaart	Vervangt één specifieke waarde van een tijdreeks door een andere.
Velden	
Kop	VERVANG_WAARDE : definieert de kaart
Parameter 1	Waarde die vervangen moet worden (een getal)
Parameter 2	Vervangingswaarde (getal of NaN)
Parameter 3 - n	Namen van de attributen/meetreeksen (bijv. NO2 of P10)
Parameter n+1 - eind	Stations (zelfde conventie als KADER kaart)
- RIVM heeft bij metingen met Paddenstoelen de waarde 2.5 gebruikt als aanduiding van een niet valide meting. Met deze kaart kan deze waarde worden omgezet in NaN (Not a Number). - De tijdreeks wordt vervangen: er wordt geen nieuwe tijdreeks gecreëerd.	

Gaten - "gevuld met " onbekende waarden (NaNs) - in tijdreeksen kunnen worden opgevuld met waarden die afgeleid worden uit valide waarden vóór en na het gat. In dat geval is het verstandig om niet de tijdreeks te vervangen, maar om een nieuwe tijdreeks te definiëren.

Relevante kaart:

INTERPOLEER

INTERPOLEER kaart	Vult gaten op in tijdreeksen
Velden	
Kop	INTERPOLEER : definieert de kaart
Veld (optioneel) Als niet aanwezig wordt Lineaire interpolatie verondersteld	Interpolatie methode: ORDE1 of LINEAIR – het gat wordt opgevuld met een lineaire functie (rechte lijn) tussen de waarde vóór en na het gat ORDE2 of KWADRATISCH – – het gat wordt opgevuld met een tweede orde functie (kwadratische functie) tussen de waarde vóór en na het gat ORDE3 of KUBISCH – – het gat wordt opgevuld met een derde orde functie (kubische functie) tussen de waarde vóór en na het gat
Veld (optioneel) indien niet aanwezig wordt 6 uur verondersteld	Maximale overbrugging met interpolatie

Meetreks veld (indien niet aanwezig wordt de input meetreeks overschreven)	Output tijdreeks - naam tijdreeks na toepassing van taak
Meetreks veld	Input tijdreeks - Naam tijdreeks waarop taak wordt toegepast
Verdere velden	Station Blok
- Voor elk van de genoemde stations wordt interpolatie toegepast - De bemonstering van de output curve is dezelfde als die van de inputcurve. - Gegevens worden niet geëxtrapoleerd.	

Specifieke selecties maken van gegevens

Om te zorgen dat een serie tijdreeksen van een bepaald meettype maar van verschillende stations beter vergeleken kunnen worden, kan men met de OPSCHONEN kaart ervoor zorgen dat voor alle tijden waarop één van die meetreeksen een NaN (Not a Number) waarde heeft, de andere meetreeksen in die verzameling van stations ook op NaN worden gezet. De te vergelijken meetreeksen hebben dan allemaal hetzelfde aantal punten.

OPSCHONEN

OPSCHONEN kaart	Elimineert meetwaarden/attributen van een aantal meetstations op de tijdwaarden waarop de meetwaarde/attribuut van één of meer van de meetstations niet gegeven is (de waarde NaN – Not a Number heeft).
Velden	
Kop	OPSCHONEN : definieert de kaart
Eerste veld (optioneel, indien niet aanwezig wordt de input meetreeks overschreven)	Output tijdreeks - naam tijdreeks na toepassing van taak
Volgende veld	Input tijdreeks - Naam tijdreeks waarop taak wordt toegepast
Verdere velden	Station Blok
Nieuwe tijdreeksen met een nieuwe naam worden gemaakt voor alle aangegeven stations.	

Indien men alleen geïnteresseerd is in een bepaalde periode van een dag of gedurende een week kan men deze data uitfilteren. Deze selectie gaat per meetgegevensreeks, en er wordt een nieuwe gegevensreeks (attribuut) aangemaakt met de geselecteerde gegevens. Gegevens in deze nieuwe meetgegevensreeks voor andere tijden staan dan op NaN (Not a Number).

Relevante kaart:

SCENARIO

SCENARIO kaart	Maakt een nieuwe meetgegevensreeks aan met een selectie gebaseerd op specifieke tijden (gegeven door een scenario-naam).
Velden	
Kop	SCENARIO : definieert de kaart
Eerste veld	Naam van het scenario.
Tweede veld	Output tijdreeks - naam tijdreeks na toepassing van taak
Derde veld	Input tijdreeks - Naam tijdreeks waarop taak wordt toegepast
Verder	Station Blok
- De namen van de scenario's zijn de volgende, en beslaan de volgende tijden: - ochtendspits : maandag – vrijdag 7,8,9h - avondspits : maandag-vrijdag 17, 18, 19h	

- werkdag : maandag-vrijdag 10, 11 ... 16h
- zaterdagochtend : zaterdag 9,10,11,12,13h
- vroege ochtend : 0,1,2,...6h
- avond : 20, 21,22,23 h
- weekend : zaterdag, zondag, hele dag van 0 h tot en met 23h
- geheel : alles.

- Indien men scenario's wil aanpassen en andere scenario's wil inbouwen, moet men veranderingen aanbrengen in de subroutine READSCENARIO in ALVA22INPUTMOD.py.
- Met de optie (scenario) "geheel" maakt men een duplicaat van een gegevensreeks.

Een andere selectieprocedure is op basis van een voorwaarde die voortkomt uit een vergelijking van waarden van specifieke, door de gebruiker gekozen, attributen. Men kan dan de metingen selecteren die gedaan zijn bij bepaalde weersomstandigheden (bijvoorbeeld alleen die gegevens die waargenomen zijn bij hoge windsnelheden, of bij lage luchtvochtigheid, of in een bepaalde windrichting). De selectieperiode is gebaseerd op een vergelijking van een attribuut met een getal, of met een tweede attribuut. Ook in deze berekening wordt een nieuw attribuut aangemaakt. Door de selectieprocedure twee keer toe te passen kan men selecteren op basis van een bereik van windsnelheden of windrichtingen (bijv: selecteren van een windrichting waarvoor geldt $45 < D_{wind} < 135$ vereist twee stappen waarbij de tweede stap wordt toegepast op het resultaat van de eerste stap.

Relevante kaart:

KOPIEER_ALS

KOPIEER_ALS kaart	Maakt een nieuwe meetgegevensreeks aan met een selectie van gegevens gebaseerd op een criterium
Velden	
Kop	KOPIEER_ALS : definieert de kaart
Parameter 1	Voorwaarde: een uitdrukking als "(x operator y)" waarbij: x en y attributen(waarden van meetreeksen) zijn of getallen, en operator één van de volgende vergelijkingsoperatoren zijn: "<" "<=" ">" ">=" of "==" (kleiner dan, kleiner of gelijk dan, groter dan, groter of gelijk dan, gelijk aan).
Parameter 2	Naam van de te berekenen meetgegevensreeks
Parameter 3	Input tijdreeks - Naam tijdreeks waarop taak wordt toegepast
Parameter 4 - eind	Station Blok
• De voorwaarde moet de haakjes bevatten, en de operator. Spaties binnen de uitdrukking die de voorwaarde aangeven zijn niet verplicht. • Als de te berekenen gegevensreeks al bestaat, wordt de reeks (gedeeltelijk) overschreven door de te kopiëren attribuut indien aan de voorwaarde wordt voldaan; de andere getallen worden niet veranderd. • Als de te berekenen gegevensreeks niet bestaat, wordt NaN (Not a Number) ingevuld als niet aan de voorwaarde wordt voldaan. • De selectie gaat op basis van een meetreeks / een aantal meetreeksen van hetzelfde station. Ook de te berekenen reeks hoort bij datzelfde station. Indien gewenst kan men meetreeksen van of naar een ander meetstation overbrengen met behulp van de OVERBRENGEN kaart.	

Overbrengen van gegevens van station naar stations

De meeste kaarten geven een bewerking aan op meetreeksen met als eindresultaat een andere meetreeks. Al deze meetreeksen behoren tot hetzelfde station. Om gegevens van twee stations goed te kunnen vergelijken of berekeningen te kunnen doen op basis van gegevens van twee stations, kan het nuttig zijn een meetreeks van één station naar een ander station over te brengen.

Een andere toepassing is dat het gemiddelde dat wordt berekend met de SAMENVOEGEN kaart en is opgeslagen in een pseudostation wordt overgebracht naar de oorspronkelijke stations.

Relevante kaart:

OVERBRENGEN

OVERBRENGEN kaart	Kopieert een meetgegevensreeks van één enkel station naar een meetgegevensreeks op een ander station, of naar meerdere andere stations.
Velden	
Kop	OVERBRENGEN : definieert de kaart
Parameter 1	Naam van de meetgegevensreeks op andere stations na het kopiëren (bijv NO2_kopie).
Parameter 2	Naam van de te kopiëren meetgegevensreeks (bijv: NO2).
Parameter 3 - 4	Naam van het station waaruit gekopieerd moet worden (format als op KADER kaart voor individuele stations: klasse station plus het getal dat het station aangeeft). <i>Het station kan een pseudostation zijn (aangeduid met _PS en een naam).</i>
Parameter 5 - eind	Namen van de stations waarnaar gekopieerd moet worden (format als op KADER kaart). ALL is niet toegestaan. <i>De stations kunnen een pseudostation zijn (aangeduid met _PS en een naam)</i>
- Als de gegevensreeks waarnaar gekopieerd wordt al bestaat, wordt de reeks (gedeeltelijk) overschreven door de meetwaarden van de te kopiëren meetreeks. Indien de te kopiëren meetreeks gaten vertoont (waardes NaN: Not a Number), worden deze gaten genegeerd, en de oude waarden van de gegevensreeks waarnaar gekopieerd wordt blijven bestaan. - Als het de bedoeling is om de tijdreeksen van twee stations in één plot te vergelijken, kan dat eenvoudig met de VERGELIJK_PAREN plot optie.	

Er zijn op dit ogenblik twee bronnen van gegevens voor een station: het inlezen via samenmeten.rivm.nl (de oude getrivm) en het inlezen via de RIVM API (vereist een tilde voor de specificatie). De verzamelingen gegevens die zo kunnen worden opgehaald verschillen van elkaar (zie Appendix).

Men kan tijdreeksen van de éne verzameling naar de andere verzameling overbrengen door de bewerking TILDETRANSFER.

TILDETRANSFER kaart	Kopieert een meetgegevensreeks van één verzameling stations naar een meetgegevensreeks naar de andere verzameling.
Velden	
Kop	TILDETRANSFER : definieert de kaart
Eerste veld	Naam van de meetgegevensreeks op andere stations na het kopiëren (bijv NO2_kopie).
Tweede veld (niet noodzakelijk)	Naam van de te kopiëren meetgegevensreeks (bijv: NO2). Indien niet gegeven is de naam van de meetgegevensreeks na kopiëren gelijk aan die van de oorspronkelijke tijdreeks (gelijk aan de naam ingevuld in het eerste veld.
Volgende veld	Of: "NAARTILDE" of "VANTILDE" <i>Bij NAARTILDE gaan gegevens van de oude getrivm naar stations gelezen met de RIVM API. Bij VANTILDE gebeurt het omgekeerde.</i>
Volgende velden	Stationblok (moet stations bevatten van zowel input als output)
Als de gegevensreeks waarnaar gekopieerd wordt al bestaat, wordt de reeks (gedeeltelijk) overschreven door de meetwaarden van de te kopiëren meetreeks. Indien de te kopiëren meetreeks gaten vertoont (waardes NaN: Not a Number), worden deze gaten genegeerd, en de oude waarden van de gegevensreeks waarnaar gekopieerd wordt blijven bestaan.	

Berekeningen met een enkele tijdreeks

Men kan een datareeks een aantal uur naar voren of naar achteren verschuiven. Tijd-verschoven data reeksen kunnen dan gebruikt worden in combinaties met andere data reeksen (zie hoofdstukje “Combineren van tijdreeksen”). Deze verschuiving gebeurt met de T_SHIFT_ATTR kaart.

Relevante kaart:

T_SHIFT_ATTR

T_SHIFT_ATTR kaart	Verschuift een enkele datareeks in tijd
Velden	
Kop	T_SHIFT_ATTR : definieert de kaart
Parameter 1	Tijdverschuiving in uren. Als positief – de nieuwe tijdreeks is naar achter verschoven – gebeurtenissen komen later. Als negatief – de nieuwe tijdreeks is naar voren verschoven – gebeurtenissen komen eerder
Parameter 2	Naam van de te berekenen meetgegevensreeks
Parameter 3	Input tijdreeks - Naam tijdreeks waarop taak wordt toegepast
Parameter 4 - eind	Station Blok
	- Voor elk van de genoemde stations in het Station Blok wordt de tijdverschuiving toegepast - De bemonstering van de output curve is dezelfde als die van de inputcurve. - Afhankelijk van de richting van tijdverschuiving worden de begin- of de eindwaarden van de reeksen op NaN (=onbekend) gezet. Met een langere periode gespecificeerd op de PERIODE kaart kan men een voldoende tijdspanne overhouden voor de analyse waarbij de periode is bepaald met de ANALYSEPERIODE kaart.

Men kan daggemiddelden, weekgemiddelden en andere statistieken voor een dag berekenen met de BLOK kaart. Dit kan nuttig zijn om een snel overzicht te krijgen van verloop van de metingen over een langere tijdsperiode.

Relevante kaart:

BLOK

BLOK kaart	Berekenen van een gemiddelde of een andere statistiek over de gemeten waarden gedurende één periode (lokale tijd Nederland)
Velden	
Kop	BLOK : definieert de kaart
Eerste veld	Aanduiding periode: “Dag”, “Week”, “Maand”, “Jaar”
Volgende veld	Statistiek die moet worden berekend: “Gemiddelde” - gemiddelde van de waarden “Mediaan” - mediaan van de waarden “StDev” - standaard deviatie van de waarden “Minimum” - minimum van de waarden “Maximum” - maximum van de waarden “Variantie” - variantie (σ^2) – het kwadraat van de standaard deviatie “Aantal” - aantal metingen in de periode
Volgende veld	Naam van de te berekenen meetgegevensreeks
Volgende veld	Input tijdreeks - Naam tijdreeks waarop taak wordt toegepast
Volgende velden - eind	Station Blok
	De bemonstering van de output curve is dezelfde als die van de inputcurve.

Een vloeiende lijn door de tijdreeks kan worden berekend als glijdend gemiddelde of als glijdende mediaan, en ook de standaard deviatie.

Hiervoor gebruikt men de kaart VLOEIEND, met de onder-aanduidingen GEMIDDELDE, MEDIAAN en STDEV. Voor het specifieke geval dat er een sterk dag/nacht verschil bestaat gebruikt men de kaart DAGRITME.

Relevante kaart:

VLOEIEND

VLOEIEND kaart	Berekend een tijdreeks waarbij de variaties binnen een bepaald tijdvenster zijn geëlimineerd of deze juist worden berekend. Deze tijdreeks wordt als nieuwe attribuut opgeslagen.
Velden	
Kop	VLOEIEND: definieert de kaart
Eerste veld	Aanduiding van proces in het vloeiende tijdvenster: "Gemiddelde" "Mediaan" "Stdev"
Optioneel – volgende twee velden	Aanduiding "@Filterlengte" (Optioneel) Filterlengte in uren (Filterlengte moet een oneven getal zijn). Indien niet gegeven is de filterlengte 25 uur.
Volgende veld	Naam van de te berekenen meetgegevensreeks
Volgende veld	Input tijdreeks - Naam tijdreeks waarop taak wordt toegepast
Rest	Station Blok
	- Indien geen naam wordt opgegeven voor de te berekenen tijdreeks, wordt de input tijdreeks overschreven - Indien de input reeks NaN waarden bevat (NaN – not a number / dus onbepaald), dan bevat de output reeks op diezelfde tijden ook NaN waarden.

Als er een sterk dag/nacht verschil bestaat is een gemiddelde over een aaneengesloten tijdvenster niet de goede maat om een gemiddelde waarde te berekenen. In plaats daarvan kan men beter een gemiddelde nemen over overeenkomende gedeeltes van een dag over een aantal dagen. Deze benadering wordt gevolgd met de berekening van een dagritme.

Relevante kaart:

DAGRITME

DAGRITME kaart	Berekent een tijdreeks waarbij gegevens van verschillende dagen worden gecombineerd om een beste schatting te maken van een gemiddelde (mediane) waarde op een tijdstip . Deze tijdreeks wordt als een nieuwe tijdreeks/attribuut opgeslagen.
Velden	
Kop	DAGRITME: definieert de kaart
Parameter 1	Aantal dagen vóór en na het tijdstip van de te berekenen waarde meegenomen in de berekening. Voorbeeld: de waarde 3 geeft aan dat de mediaan wordt berekend met gegevens van totaal 7 dagen. Van elke dag worden drie metingen gebruikt, de meting op het uur waar men interesse voor heeft, en de metingen van het uur ervoor en erna.
Parameter 2	Naam van de te berekenen meetgegevensreeks
Parameter 3	Input tijdreeks - Naam tijdreeks waarop taak wordt toegepast
Parameter 4 - eind	Station Blok
	- Indien parameter 2 en 3 hetzelfde zijn, wordt de tijdreeks vervangen: er wordt dan geen nieuwe tijdreeks gecreëerd. - Indien de input reeks NaN waarden bevat (NaN – not a number / dus onbepaald), dan bevat de output reeks op diezelfde tijden ook NaN waarden.

Combineren van tijdreeksen

Men kan ook verschillende attributen (meetreeksen) combineren tot een nieuw attribuut (meetreeks). Met de BEREKEN_ATTR kaart kan men formules opgeven in Reverse Polish Notation¹. Het voordeel van Reverse Polish Notation is dat de prioriteit van de berekeningen wordt gegeven door de volgorde, en niet door haakjes. Een operator werkt altijd op de twee daarvoor gegeven grootheden (waarvan een grootheden mogelijkerwijs nog moet worden afgeleid uit een combinatie van grootheden en een operator). Een combinatie van twee grootheden en een operator vormt altijd een nieuwe grootheid.

Relevante kaart:

BEREKEN_ATTR

BEREKEN_ATTR kaart	Berekent een meetreeks/attribuut uit een combinatie van andere meetreeksen/attributen.
Velden	
Kop	BEREKEN_ATTR: definieert de kaart
Parameter 1	Naam van de te maken meetreeks/attribuut.
Parameter 2	Een input meetreeks/attribuut of een getal
Parameter 3	Een tweede input meetreeks/attribuut of een getal
Parameter 4	De operator of een derde getal (zie voorbeeld in voetnoot)
Parameter 5 – parameter x	Operators en getallen
Parameter x - eind	Station Blok <i>De stations kunnen pseudostations zijn (_PS gevolgd door een naam).</i>
• Er wordt altijd een nieuwe tijdreeks gedefinieerd. • Het aantal input attributen / input getallen is altijd 1 groter dan het aantal operators. Het laatste item in de opdrachtenreeks is altijd een operator, behalve in het geval dat er alleen één input attribuut/getal is gegeven zonder een operator. In dat geval wordt het input attribuut / getal gekopieerd naar het te maken attribuut. Dit geeft de mogelijkheid om attributen van een andere naam te voorzien, of een constante grens te definiëren. • Toegestane operators: + - * / min max ^ log Bij de operator ^ is de eerste grootheid de basis, en de tweede de exponent Bij de operator log is de eerste grootheid het getal waarvan de log moet worden genomen, het tweede getal geeft de basis aan. • De operators min en max zijn beperkt tot het minimum of maximum van twee getallen • Voorbeelden: 1. BEREKEN_ATTR NO2corr NO2 0.4 Vwind * + T_gemidd 0.2 * - ALL Dit is equivalent met $NO2 + 0,4 * Vwind - 0,2 * T_{gemidd}$ voor alle stations 2. BEREKEN_ATTR NO2norm 40 ALL Dit definieert een attribuut NO2norm met de waarde van de EU norm voor NO2 luchtvervuiling. • De meetreeksen waarmee gerekend wordt behoren allemaal bij hetzelfde station. Ook de te berekenen reeks hoort bij datzelfde station. Indien gewenst kan men meetreeksen van of naar een ander meetstation overbrengen met behulp van de OVERBRENGEN kaart.	

¹ Overgenomen uit Wikipedia: In reverse Polish notation, the [operators](#) follow their [operands](#); for instance, to add 3 and 4, one would write 3 4 + rather than 3 + 4. If there are multiple operations, operators are given immediately after their second operands; so the expression written 3 – 4 + 5 in conventional notation would be written 3 4 – 5 + in reverse Polish notation: 4 is first subtracted from 3, then 5 is added to it. An advantage of reverse Polish notation is that it removes the need for parentheses that are required by [infix notation](#). While 3 – 4 × 5 can also be written 3 – (4 × 5), that means something quite different from (3 – 4) × 5. In reverse Polish notation, the former could be written 3 4 5 × –, which unambiguously means 3 (4 5 ×) – which reduces to 3 20 – (which can further be reduced to -17); the latter could be written 3 4 – 5 × (or 5 3 4 – ×, if keeping similar formatting), which unambiguously means (3 4 –) 5 ×.

Bijhouden van wettelijke grenzen

Grenswaarden voor luchtvervuiling zijn meestal gedefinieerd in termen van jaaroverschrijdingen (zie Appendix F, waarin de tabellen gekopieerd zijn uit de website van de Omgevingswet). Overschrijding van deze normen is meestal gemakkelijk te controleren met het definiëren van een constante waarde en deze te plotten tezamen met de meetwaardes (voor definiëren van een constante waarde zie bijvoorbeeld het voorbeeld bij de BEREKEN_ATTR kaart: "BEREKEN_ATTR NO2norm 40 ALL "). Ook kan men bij de onderschriften van de plots de gemiddelden over een tijdsperiode aflezen.

Minder eenvoudig is het bijhouden van het aantal overschrijdingen van een norm binnen een tijdsperiode. In de wet kent men twee van deze regels – een aantal overschrijdingen van uurgemiddeldes per jaar en een aantal overschrijdingen van daggemiddelden.

Met de OVERSCHRIJDING kaart berekent men een cumulatie curve die aan het begin van de tijdsperiode met 0 begint, en die telkens als er een overschrijding plaats vindt met 1 wordt opgehoogd. De maximum van deze curve is dan het aantal overschrijdingen over de tijdsperiode waarover is gemeten.

Waarschuwing: zeker met het tellen van uuroverschrijdingen, maar ook met het tellen van dagoverschrijdingen kan men de mist in gaan. Er kunnen toevallige uitschieters plaatsvinden door communicatiestoringen, apparatuur problemen e.d. Als men de gegevens filtert met de GLIJDEND kaart of de DAGRITME kaart, kan men uitschieters elimineren. In elk geval moet men nauwkeurig nagaan of de overschrijdingen reëel zijn, en geen artefact.

Relevante kaart:

OVERSCHRIJDING

OVERSCHRIJDING kaart	Maakt een nieuwe gegevensreeks aan die oploopt van 0 naar het aantal geconstateerde overschrijdingen van een grenswaarde.
Velden	
Kop	OVERSCHRIJDING : definieert de kaart
Parameter 1	"DAG" of "UUR". Bij "DAG" wordt het gemiddelde over de dag bepaald, en die wordt vergeleken met een gestelde norm.
Parameter 2	De te stellen norm (bijvoorbeeld 200 voor de grens van het uurgemiddelde). Zie appendix F.
Parameter 3	Naam van de te berekenen cumulatie curve
Parameter 4	Naam van de attribuut (bijv. NO2 of P10) die wordt getoetst
Parameter 5 - eind	Stations (zelfde conventie als KADER kaart)
- Voor elk van de genoemde stations wordt een cumulatie curve berekend - De cumulatieve curve begint op 0 op het tijdstip genoemd op de ANALYSEPERIODE kaart.	

Berekenen van pseudostations

Meetreeksen van verschillende stations kunnen worden samengevoegd en worden dan verzameld in pseudostations (zie voorbeeld in figuur hieronder). Men kan verschillende soorten van middeling gebruiken bij dit samenvoegen – de standaard gemiddelde, en de mediaan van elke soort meting van de tijdreeksen van de samen te voegen stations kunnen worden berekend en die vormen dan de tijdreeksen van het pseudostation. Ook kan men maximum , minimum en standaard deviatie berekenen en overbrengen naar een pseudostation.



Voorbeeld: Meetreeksen van vijf stations worden gecombineerd tot meetreeksen in één pseudostation.

Men kan ook pseudostations definiëren die een versie van de gegevens van een station bevatten na een verschuiving in tijd. Ook is het mogelijk stations te combineren door middel van een formule. Deze optie kan interessant zijn als men het verschil tussen twee stations wil duidelijk maken.

De attributen/meetreeksen van de pseudostations kunnen geplot worden. Wil men deze plotten moet als type van het pseudostation worden ingevuld `_PS` gevolgd door de naam van het pseudostation.

Relevante kaart:

SAMENVOEGEN

SAMENVOEGEN kaart	Specificeert de gegevens voor het berekenen van de gemiddelden, standaard deviaties of andere berekeningen op de meetreeksen van dezelfde soort voor een stel stations. Het resultaat wordt als een pseudostation opgeslagen.
Velden	
Kop	SAMENVOEGEN : definieert de kaart
Veld 1 (optioneel: als geen van de genoemde opties wordt ingevuld, wordt "GEMIDDELDE" verondersteld.	Methode waarbij de gegevens worden gecombineerd. Mogelijkheden zijn: GEMIDDELDE MEDIAAN MAXIMUM MINIMUM STDEV QUANTIEL
Volgende veld (alleen als "QUANTIEL" ingevuld is). Indien	Percentage (getal tussen 0 en 100) die het quantiel aangeeft. 25 geeft het eerste kwartiel, 50 de mediaan, 75 het derde kwartiel.

niet ingevuld wordt 75 (derde quantiel) verondersteld	
Volgende veld	Naam van het te creëren pseudo station. Moet uniek zijn.
Volgende velden - eind	De stations waarvoor dit proces wordt toegepast. Dezelfde conventies worden gebruikt als op de KADER kaart.
• Alle numerieke attributen/meetreeksen van de stations worden gecombineerd / gemiddeld (ook al zou het niet zinnig zijn). • Men krijgt bijvoorbeeld het gemiddelde van alle NO2 reeksen, P10 reeksen, en van eventueel berekende attributen. • Op verdere kaarten wordt naar het pseudostation XXX gerefereerd met : _PS XXX indien men dit pseudostation mee wilt nemen in de berekeningen.	

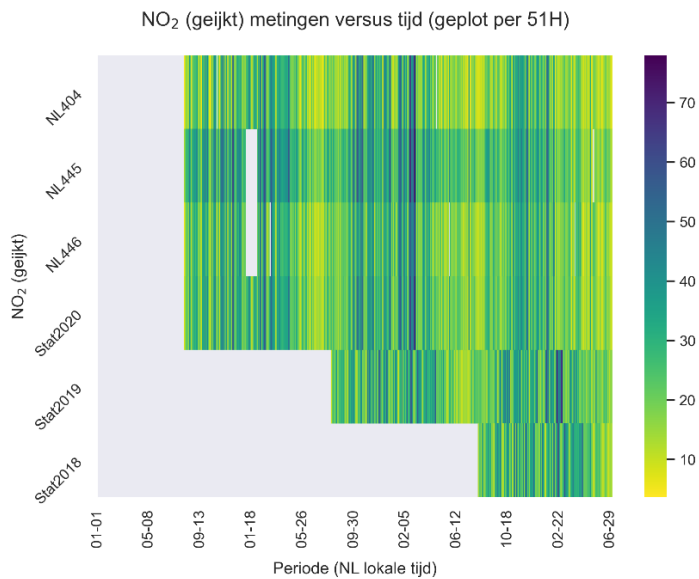
Relevante kaart:

T_SHIFT

T_SHIFT kaart	Specificeert een verschuiving in tijd die op de meetreeksen van een station wordt toegepast. Een nieuw station wordt gecreëerd, waarin de tijdverschoven meetreeksen worden opgeslagen.
Velden	
Kop	T_SHIFT : definieert de kaart
Parameter 1	Naam van het te creëren pseudo station. Moet uniek zijn.
Parameter 2	De gewenste tijdverschuiving in dagen. Een positieve tijdverschuiving is een verschuiving naar een later tijdstip.
Parameter 3 + (indien nodig voor specificatie parameter 4)	Het station waarvoor dit proces wordt toegepast. De conventies om het station aan te geven is identiek aan de conventie als gehanteerd op de KADER kaart. ALL is niet toegestaan. Een naam van een station als gecreëerd met de SAMENVOEGEN kaart is toegestaan. In dat geval is er geen parameter 4.
• De tijdverschuiving wordt toegepast op alle numerieke attributen van het station (ook al zou het niet zinnig zijn).	

Het tweede voorbeeld in Appendix B is een toepassing waarbij gebruik gemaakt wordt van SAMENVOEGEN zowel als van tijdverschuiving. Het resultaat wordt getoond in een plot gemaakt met de optie OVERZICHT (zie het volgende hoofdstuk).

Doel was het vergelijken van de hoeveelheid vervuiling in het voorjaar van de jaren 2018, 2019 en 2020. In dit voorbeeld wordt SAMENVOEGEN gebruikt om het gemiddelde te berekenen van de meetreeksen van de Haagse RIVM meetstations over de periode 2017 t/m midden 2018. Het resultaat wordt opgeslagen in een pseudostation Stat2020. De tijdreeksen van dit station wordt tweemaal een jaar naar achteren verschoven. Daarmee worden de pseudostations Stat2019 en Stat2018 gegenereerd. De tijdreeksen van Stat2018 in de periode van 1 januari 2020 t/m eind Juni zijn dus in feite de meetreeksen van twee jaar eerder.



De eerste drie rijen geven in kleur de NO₂ vervuiling aan in de periode 1 januari 2017 t/m 30 juni 2020. Er blijken geen metingen op de RIVM website te staan van de stations NL404, NL445 en NL446 in de periode voor augustus 2017. Het gemiddelde van de metingen in de rij Stat2020 heeft dus ook geen gegevens voor die datum.

De pseudostations Stat2018 en Stat2019 zijn de verschoven versies. Als men nu de NO₂ gegevens over het eerste half jaar van 2020 plot voor de drie stations Stat2018, Stat2019 en 2020, heeft men de gezochte vergelijking.

Heatmap: NO₂ metingen van drie RIVM stations en drie pseudostations.

Zoals men tijdreeksen kan combineren met de BEREKEN_ATTR kaart, kan men ook stations en pseudostations combineren. Dit gebeurt met behulp van de BEREKEN_STAT kaart.

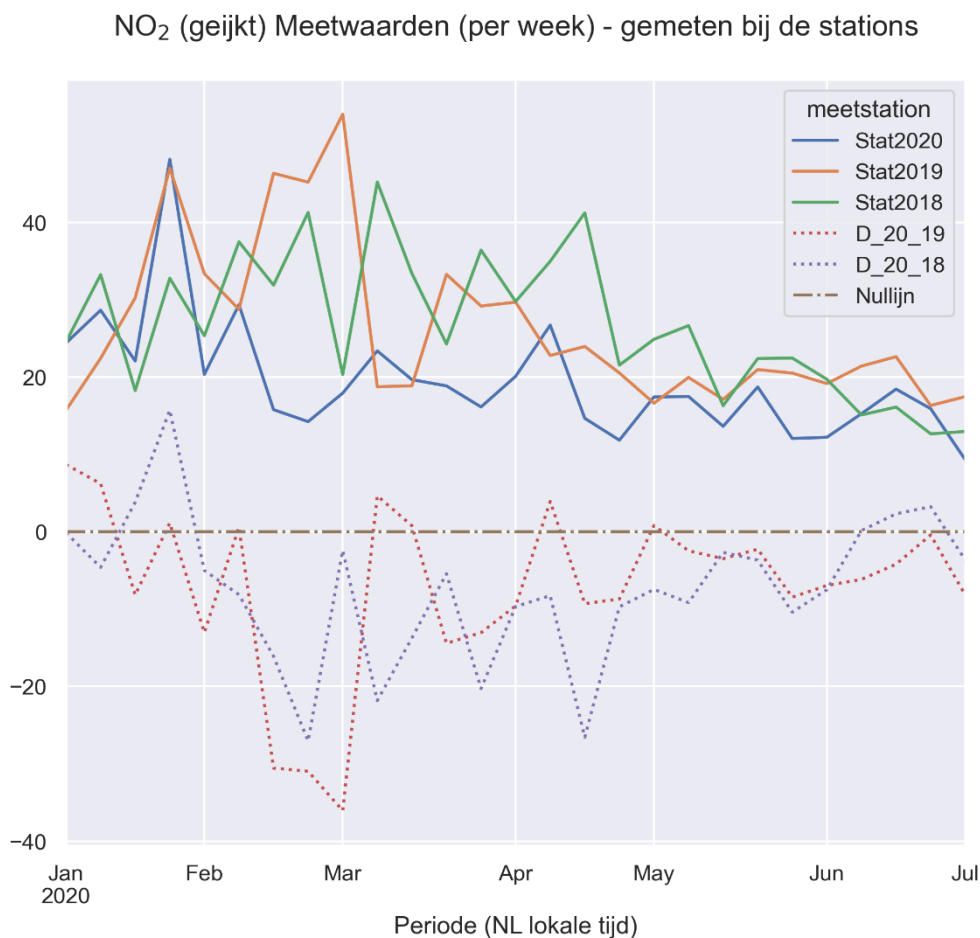
Relevante kaart:

BEREKEN_STAT

BEREKEN_STAT kaart	Berekent een pseudostation uit een combinatie van andere stations.
Velden	
Kop	BEREKEN_STAT: definieert de kaart
Parameter 1	Naam van de te maken pseudostation.
Parameter 2	De naam van een station of een getal - zie opmerkingen hieronder.
Parameter 3	Een tweede naam van een station of een getal
Parameter 4	De operator of een derde getal (zie voorbeeld in voetnoot)
Parameter 5 – eind	Operators en getallen / stations
- Een station moet worden gegeven door het soort station en de bijbehorende identifier (bijvoorbeeld CSLV 7). Om de kaart leesbaar te houden kan deze specificatie met haakjes worden omringd. Ook eerder berekende pseudostations kunnen worden gespecificeerd. In dat geval is de eerder opgegeven naam voldoende. Dit wijkt af van de conventie voor plotten en functieberekeningen: in die gevallen worden pseudostations weergegeven met _PS gevolgd door de naam van het station. Voorbeeld: BEREKEN_STAT Verschil (CSLV 7) Avg_CSLV – 10 x 20 + - De berekening wordt toegepast op alle numerieke attributen van de stations (ook al zou het niet zinnig zijn). - De formule wordt gegeven in Reversed Polish Notation (RPN). Zie de uitleg bij de BEREKEN_ATTR kaart. - Het aantal stationsnamen / input getallen is altijd 1 groter dan het aantal operators. Het laatste item in de opdrachtenreeks is altijd een operator, behalve in het geval dat er alleen één input attribuut/getal is gegeven zonder een operator. In dat geval wordt het station / getal gekopieerd naar het te maken pseudostation. Dit geeft de mogelijkheid om stations van een andere naam te voorzien, of een constante lijn te definiëren. - Toegestane operators: + - * / min max ^ - De operators min en max zijn beperkt tot het minimum of maximum van attributen van twee stations	

Met BEREKENSTAT kunnen we het voorbeeld hierboven, waarbij we een vergelijking maakten tussen de vervuiling in 2018, 2019 en 2020 verder uitwerken. We kunnen het verschil berekenen tussen de meetreeksen van 2020 en die van 2018 en 2019 respectievelijk (we berekenen pseudostations voor het verschil 2020 – 2018 en voor het verschil 2020 - 2019). Als we deze verschillen plotten samen met de tijdreeksen, zien we dat de vervuiling tussen midden februari en eind mei 2020 lager lijkt dan de vervuiling in de overeenkomstige periodes in 2018 en 2019. Het effect van weer is echter heel groot – duidelijk een onderwerp voor verder onderzoek. Ook de Nullijn is berekend met BEREKENSTAT.

De onderstaande plot is gemaakt met PLOT_SOORT.



PLOT_SOORT Plot van de gemiddelden van de NO₂ metingen als gemeten bij de RIVM stations in 2018, 2019 en 2020 (pseudostations Stat2018, Stat2019 en Stat2020), het verschil tussen de metingen van 2018 en 2020, en die van 2019 en 2020. Een nullijn is toegevoegd om aan te geven dat de vervuiling in 2020 over het algemeen laag was in vergelijking tot 2018 en 2019 als gevolg van weersomstandigheden en de Covid-19 lockdown.

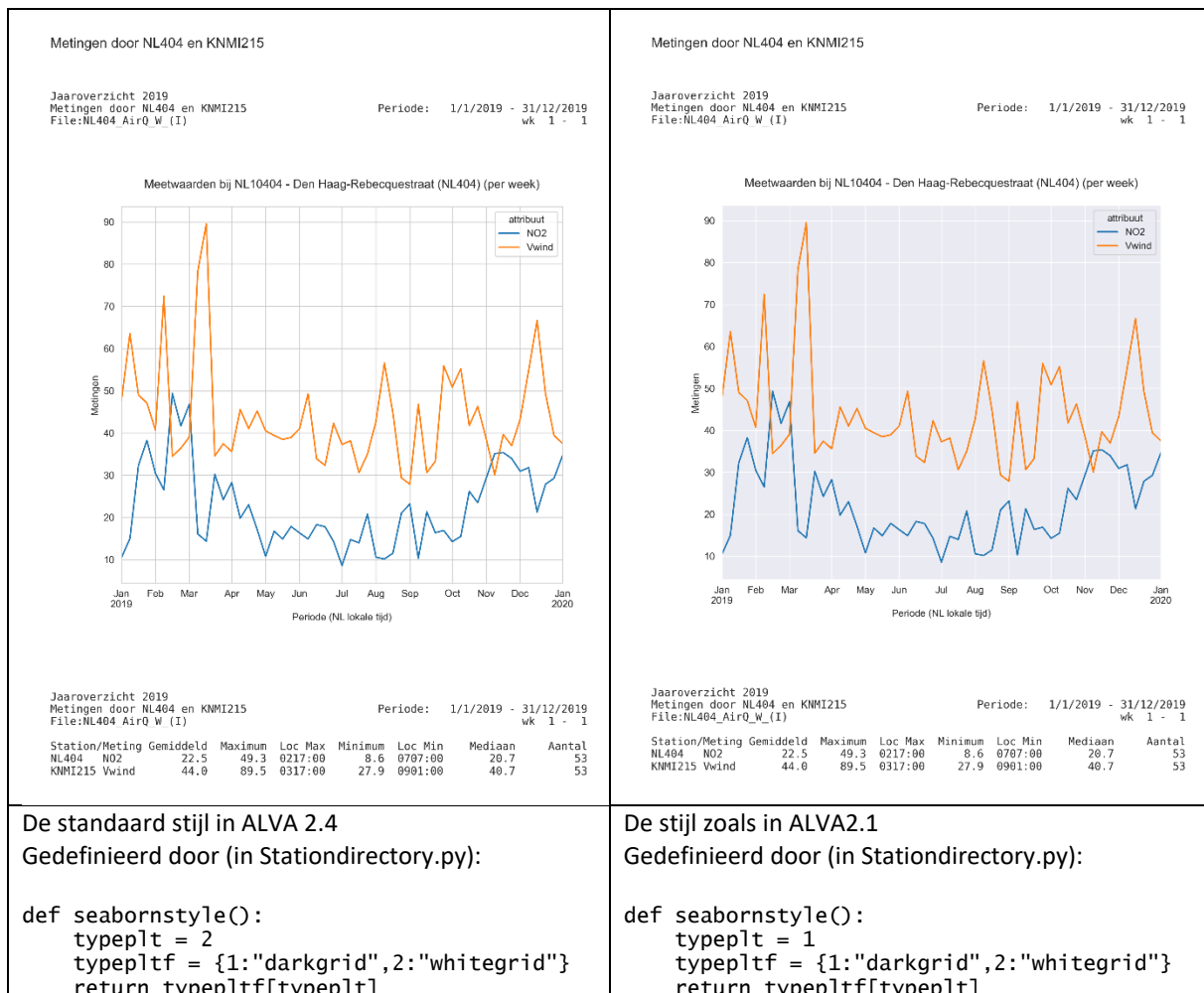
Analyse en Plotten

Instellingen

ALVA genereert veel plots, die als jpg, png of andere soort graphic files kunnen worden opgeslagen, of op het scherm kunnen worden vertoond. Omdat plots op het scherm veel geheugen vergen, is het aanbevolen geen plots op het scherm te tonen.

Men kan het type plots beïnvloeden door een wijziging aan te brengen in een tweetal subroutines in StationDirectory22.py. In subroutine typeplotfile() kan men het type plotfile bepalen (png, jpg, pdf) en in seabornstyle() kan men kiezen voor een standaard stijl uit een tweetal stijlen van display. Er is geen kaart gedefinieerd om deze standaarden te definiëren. Als men wil afwijken van de standaard plotstijl voor een enkele plot dan kan dit door middel van een AANPASSING kaart te specificeren voor de desbetreffende plot kaart.

De stijl is een kwestie van smaak. Ter vergelijking hierbij dezelfde plot met twee verschillende stijlen.



Alle plots in deze handleiding zijn gemaakt met de oude stijl (zoals in ALVA2.1). De standaard stijl is echter de stijl zoals getoond in de linkerkolom.

Afhankelijk van de soort plot kan verder breedte en hoogte van de plot worden aangepast, de soort lijnen (vast, stippeltjes,...), de kleur van de lijnen en het bereik van de plot in de verticale richting. De statistische informatie in de tabellen onder de plot zijn per (verkorte) stationnaam of per (verkort) adres. Het verkorte adres is soms

niet uniek – in dat geval wordt verdere informatie verschaft over station, verkorte stationnaam en het adres in een aparte tabel.

Er worden plots gemaakt van geselecteerde meetgegevens en KNMI gegevens met verschillende bemonsteringsfrequenties voor de tijdseries van elk station.

De attributen van de pseudostations kunnen geplotted worden. Wil men deze plotten moet als type van het pseudostation worden ingevuld `_PS` gevolgd door de naam van het pseudostation.

De plots en de analyses worden gedaan over een tijdsperiode als gegeven op de ANALYSEPERIODE kaart. Als er geen ANALYSEPERIODE kaart wordt gegeven wordt de periode genomen als gegeven door de PERIODE kaart.

Relevante kaarten:

PLOTPARAM (niet verplicht!)

PLOTPARAM kaart	Specificeert of op het scherm wordt geplotted of dat de plots in een directory worden opgeslagen
Velden	
Kop	PLOTPARAM : definieert de kaart
Parameter 1	Optie voor plotten - ALL : zowel op het scherm als in een directory - PLOTONLY : alleen op het scherm - SAVEONLY : alleen opslaan in een directory - NONE : geen plotoutput
Optioneel: Parameter 2 - eind	Subfolder in de default outputdirectory waarin de gegevens worden opgeslagen. De subdirectory kan zijn: <code>..outputdir/param2/param3</code> enz.
• De standaard outputdirectory is een subdirectory van de inputdirectory. De naam van deze directory wordt afgeleid van de naam van de inputcarddeck. Als de parameterfile een naam heeft beginnend met "parameterfile_", dan wordt dit begin afgekapt. • Indien er geen PLOTPARAM kaart wordt gegeven, worden de plots in de standaard directory zoals hierboven gedefinieerd geplaatst. • In de standaard output directory worden twee additionele directories gegenereerd. De eerste heeft de naam <code>0__parameters</code> en bevat een kopie van de input card file. De tweede is een directory <code>Stats4figuren</code> en bevat de tekst files met de tekst die aan een plot wordt toegevoegd.	

De meetgegevens worden gegeven per uur. Indien gewenst kan men andere bemonsteringsintervallen kiezen. Het programma genereert dan voor plotting en analyse meerdere datasets met andere sampling.

Relevante kaart:

BEMONSTERING

BEMONSTERING kaart	Specificeert de bemonsteringsfrequentie (sampling rate) van de te plotten tijdreeksen. Alle berekeningen gebeuren met tijdreeksen met een bemonsteringsfrequentie van één uur. De csv files die worden ingelezen als gespecificeerd op de KADER/NIETGEBRUIKEN kaarten hebben deze bemonsteringsfrequentie.
Velden	
Kop	BEMONSTERING : definieert de kaart
Parameter 1	Drie opties: • 1HPLOT (Tijdplots met een sampling rate van één uur worden gemaakt) • GEEN1HPLOT (Plots met een sampling rate van één uur worden overgeslagen en worden niet gemaakt) • Een indicatie van de gewenste bemonsteringen voor plots. Dit kan zijn: 1. nH: sampling interval wordt n uren, met n (optioneel) een geheel getal 2. nD: sampling interval wordt n dagen, met n (optioneel) een geheel getal

	3. nW: sampling interval wordt n weken, met n (optioneel) een geheel getal 4. nM: sampling interval wordt n maanden, met n (optioneel) een geheel getal
• Meerdere BEMONSTERING kaarten zijn toegestaan. • De BEMONSTERING kaart heeft maar één parameter. Om met meerdere sampling rates te plotten, moeten meerdere BEMONSTERING kaarten worden gebruikt. • Indien geen BEMONSTERING kaart wordt gegeven, worden alleen plots gemaakt met een bemonstering van één uur. • De BEMONSTERING kaart beïnvloedt het plotten van de gegevens, maar niet de berekeningen in het bewerken van de meetgegevens en de berekeningen die gemaakt worden om pseudo stations te maken. Al deze berekeningen gebeuren met een één uur bemonsteringsfrequentie. • De BEMONSTERING kaart heeft geen invloed op de 1DFUNCTIE en 2DFUNCTIE berekeningen. Al deze berekeningen worden gedaan met een bemonsteringsfrequentie van één uur. • De optie GEEN1HPLOT zorgt dat er geen plots worden gemaakt op basis van de kaarten: PLOT, PLOT_SOORT,VERGELIJK_PAREN, VERGELIJK_ATTR2STAT • De bemonsteringsfrequenties op de bemonsteringskaart worden gebruikt in de plots gemaakt op basis van de volgende kaarten: OVERZICHT, DOOSDIAGRAM, KRUISCORRELATIE, PLOT, PLOT_SOORT,VERGELIJK_PAREN, VERGELIJK_ATTR2STAT,PUNTGRAFIEK	

De analyses worden beperkt tot de reeks stations zoals die zijn afgebakend in KADER, NIETGEBRUIKEN EN PERIODE kaarten, en beslaan de attributen zoals die aanwezig zijn in elk station, inclusief de attributen die door berekening of dergelijke stappen ter beschikking zijn gekomen.

De ANALYSEPERIODE kaart definieert de tijdspanne voor plots en de analyses van afhankelijkheden van onafhankelijke parameters.

Wat betreft de positie van de kaart in het carddeck: de ANALYSEPERIODE kan ook vlak achter de PERIODE kaart geplaatst worden.

Relevante kaart:

ANALYSEPERIODE

ANALYSEPERIODE kaart	Specificeert de tijdsperiode voor de analyses en de plots.
Velden	
Kop	ANALYSEPERIODE : definieert de kaart
Parameter 1	Begindatum van het bereik voor de analyses en de plots. Formaat: d(d)/m(m)/jjjj
Parameter 2	Twee mogelijkheden: Uur van de dag op de begindatum (getal van 0 t/m 23) of Einddatum van het bereik voor de analyses en de plots. Formaat: d(d)/m(m)/jjjj
Parameter 3	Als parameter 2 een uur aangaf: Eind datum van het bereik voor de analyses en de plots. Formaat: d(d)/m(m)/jjjj Of Als parameter 2 de einddatum aangaf: Uur van de dag voor de einddatum (niet verplicht)
Parameter 4	Niet verplichten alleen als parameter 3 een einddatum aangaf: uur van de dag voor de einddatum
• De ANALYSEPERIODE kaart is niet noodzakelijk, maar indien hij wordt gebruikt, moet deze na de PERIODE kaart in het carddeck voorkomen. • Als er geen ANALYSEPERIODE kaart wordt gegeven wordt de periode genomen als gegeven door de PERIODE kaart. • Voorbeelden: ANALYSEPERIODE 3/11/2020 6 7/11/2020 23 ANALYSEPERIODE 2/10/2020 25/11/2020	

Het aanzien van de plot kan worden gewijzigd met één of meer AANPASSING kaarten, die gespecificeerd moeten worden voor een kaart waarop de plot zal worden gemaakt.

Relevante kaart:

AANPASSING

AANPASSING kaart	Specificeert aanpassingen voor plotparameters voor een volgende plot kaart of volgende plotkaarten. Afhankelijk van de soort plot kan een specifieke aanpassing wel of niet worden uitgevoerd.
Velden	
Kop	AANPASSING : definieert de kaart
Eerste veld: Soort Aanpassing	Soort aanpassing. Kan zijn: ACHTERGROND FORMAAT STIJL KLEUR YLIMIET LIGGING KADER
Op "Soort Aanpassing" volgende specificatie. Als soort aanpassing = ACHTERGROND	Kiezen uit DONKER, LICHT, WIT en GRIJS Standaard is LICHT: witte achtergrond met gridlijnen. DONKER staat voor donkere achtergrond met gridlijnen. WIT en GRIJS hebben geen gridlijnen.
Op "Soort Aanpassing" volgende specificatie. Als soort aanpassing = FORMAAT	Twee getallen: breedte (een getal, vrij gewoon is: 8) hoogte (een getal, vrij gewoon is: 7)
Op "Soort Aanpassing" volgende specificatie. Als soort aanpassing = STIJL	Reeks aanduidingen van soort lijnen die geplotted moeten worden. De eerste aanduiding geeft het formaat van de eerste te plotten lijn, de volgende van de tweede, etc. Aanduidingen zijn (hier aangegeven tussen aanhalingstekens): "-" of 'solid' --- continue lijn "." of 'dotted' ----stippeltjes "--" of 'dashed' ---- streep streep "-." Of 'dashdot' ----- Streep punt Ook mogelijk: 'None' De standaard waarden worden door het plotting pakket bepaald.
Op "Soort Aanpassing" volgende specificatie. Als soort aanpassing = KLEUR	Aanduidingen van de kleur lijnen die geplotted moeten worden. Parameter 2 geeft de kleur van de eerste te plotten lijn, parameter 3 van de tweede, etc. Aanduidingen worden hieronder in tabellen getoond. Een speciale aanduiding is "Palmes". Daarmee wordt een palet gecreëerd waarbij Palmesbuisjes onderscheiden worden van RIVM stations en van citizen science stations.
Op "Soort Aanpassing" volgende specificatie. Als soort aanpassing = YLIMIET Parameter 2 en 3	laag – laagste waarde van de verticale as van de plot (bij verticale LIGGING) hoog – hoogste waarde van de verticale as
Op "Soort Aanpassing" volgende specificatie.	Keuze uit ADRES en STATIONNAAM. Als ADRES wordt gespecificeerd, worden (verkorte) adressen getoond in de statistische gegevens onder de plot (en soms ook in de plot zelf).

Als soort aanpassing = LABEL	Omdat verkorte adressen niet uniek hoeven te zijn, wordt de relatie nog in een aparte tabel onder de plot uitgelegd. Standaard waarde is STATIONNAAM.
Op "Soort Aanpassing" volgende specificatie. Als soort aanpassing = LIGGING	LIGGING geeft de richting aan van de data-as. Keuze uit "liggend", "staand", "horizontaal", "verticaal", "h", "v". Standaard waarde is verticaal.
Op "Soort Aanpassing" volgende specificatie. Als soort aanpassing KADER	Keuze uit VOL of BEPERKT
Op "Soort Aanpassing" volgende "Soort Aanpassing" of einde van de kaart	Terugbrengen van Soort Aanpassing naar standaard waarden
Na specificatie van Soort Aanpassing: andere Soort Aanpassing	Soort aanpassing. Kan zijn: ACHTERGROND, FORMAAT etc.
Vervolgens	Aanpassingen afhankelijk van Soort Aanpassing
• Men kan op één kaart meerdere soortaanspassingen specificeren. • Volgende AANPASSING kaarten nemen de aanpassingen van vorige kaarten over, tenzij de aanpassingen ongedaan worden gemaakt. Men zet de aanpassing terug naar standaard waarden door: 1. Als na een aanduiding van Soort Aanpassing op het volgende veld geen nadere specificatie volgt maar een andere Soort Aanpassing, dan worden de plotparameters teruggebracht naar standaard waarden. 2. Indien de kaart alleen bestaat uit de kop "AANPASSING", dan worden alle aanpassingen ongedaan gemaakt en teruggezet naar standaard waarden. • Het aantal specificaties van STIJL en KLEUR hoeft niet gelijk te zijn aan het aantal te plotten lijnen. Bij een kleiner aantal worden de overige lijnen geplot met de standaard lijn-stijl of kleur. • Niet elke plot accepteert alle aanpassingen	

Kleurtabellen

De volgende namen van kleuren worden herkend:

Base Colors



en:

CSS Colors

black	bisque	forestgreen	slategrey
dimgray	darkorange	limegreen	lightsteelblue
dimgrey	burlywood	darkgreen	cornflowerblue
gray	antiquewhite	green	royalblue
grey	tan	lime	ghostwhite
darkgray	navajowhite	seagreen	lavender
darkgrey	blanchedalmond	mediumseagreen	midnightblue
silver	papayawhip	springgreen	navy
lightgray	moccasin	mintcream	darkblue
lightgrey	orange	mediumspringgreen	mediumblue
gainsboro	wheat	mediumaquamarine	blue
whitesmoke	oldlace	aquamarine	slateblue
white	floralwhite	turquoise	darkslateblue
snow	darkgoldenrod	lightseagreen	mediumslateblue
rosybrown	goldenrod	mediumturquoise	mediumpurple
lightcoral	cornsilk	azure	rebeccapurple
indianred	gold	lightcyan	blueviolet
brown	lemonchiffon	paleturquoise	indigo
firebrick	khaki	darkslategray	darkorchid
maroon	palegoldenrod	darkslategrey	darkviolet
darkred	darkkhaki	teal	mediumorchid
red	ivory	darkcyan	thistle
mistyrose	beige	aqua	plum
salmon	lightyellow	cyan	violet
tomato	lightgoldenrodyellow	darkturquoise	purple
darksalmon	olive	cadetblue	darkmagenta
coral	yellow	powderblue	fuchsia
orangered	olivedrab	lightblue	magenta
lightsalmon	yellowgreen	deepskyblue	orchid
sienna	darkolivegreen	skyblue	mediumvioletred
seashell	greenyellow	lightskyblue	deeppink
chocolate	chartreuse	steelblue	hotpink
saddlebrown	lawngreen	aliceblue	lavenderblush
sandybrown	honeydew	dodgerblue	palevioletred
peachpuff	darkseagreen	lightslategray	crimson
peru	palegreen	lightslategrey	pink
linen	lightgreen	slategray	lightpink

Output van de analyses

De output van de analyses bestaat uit plots van verschillende soort. Aan elke plot wordt tekst toegevoegd met wat statistische gegevens: de plot en de bijbehorende statistische gegevens worden in één plaatje gecombineerd. Bovendien wordt een aparte tekstfile met die statistische gegevens gegenereerd en apart opgeslagen in de subdirectory Stats4figures. De plotfiles, en de aparte tekstfiles, krijgen een unieke naam waaruit het soort gegevens die zijn geploteerd kan worden afgeleid. Als voorbeeld, bij het plotten van attributen van een station tegen tijd bestaat de naam van de plotfile uit de korte alias van het station gevolgd door “_AirQ_” gevolgd door de mnemonic van het resampling interval. Voorbeeld: NL404_AirQ_W.png geeft een plot voor metingen gedaan met RIVM station NL0404 met één sample per week. Indien deze laatste toevoeging er niet is, is de oorspronkelijke sampling aangehouden.

Indien een kaart meerdere malen wordt gegeven, worden als gevolg daarvan evenveel meer plots geproduceerd. De plotfiles kunnen dezelfde naam hebben: daarom wordt de naam van de plotfile verlengd met een teller. De teller wordt gegeven als een romeins cijfer (I, II, III, IV enz.)

Verkennen van de dataset; kwaliteitscontrole

Om een indruk te krijgen van de eigenschappen van de dataset, kan men plots maken van statistische eigenschappen van de tijdreeksen.

Het is te verwachten dat de metingen voor de verschillende stations ongeveer hetzelfde gedrag vertonen. De invloed van weer is groot, en zal voor alle stations ongeveer hetzelfde zijn. Ook komt het voor dat vervuiling optreedt in de vorm van vervuilingswolken – die kunnen door alle verschillende stations gedetecteerd worden als een anomalie. Als één of meer stations zich anders gedragen dan andere stations, moet daar een reden voor zijn: de detector kan – bijvoorbeeld – niet goed meer functioneren.

Voor elk verschillende type vervuiling kunnen de corresponderende tijdreeksen vergeleken worden.

1. Men kan een overzicht krijgen van de hele dataset met een *heatmap* of *warmtekaart*. Een heatmap is een grafische weergave van gegevens waarbij de individuele waarden als kleuren in een matrix worden weergegeven. Om de tijdreeksen van verschillende stations te vergelijken wordt een heatmap gegenereerd bestaande uit verschillende rijen, één rij per station. In de horizontale richting worden de meetwaarden aangegeven als functie van tijd. Met deze techniek kunnen afwijkingen snel worden gedetecteerd.

Relevante kaart:

OVERZICHT

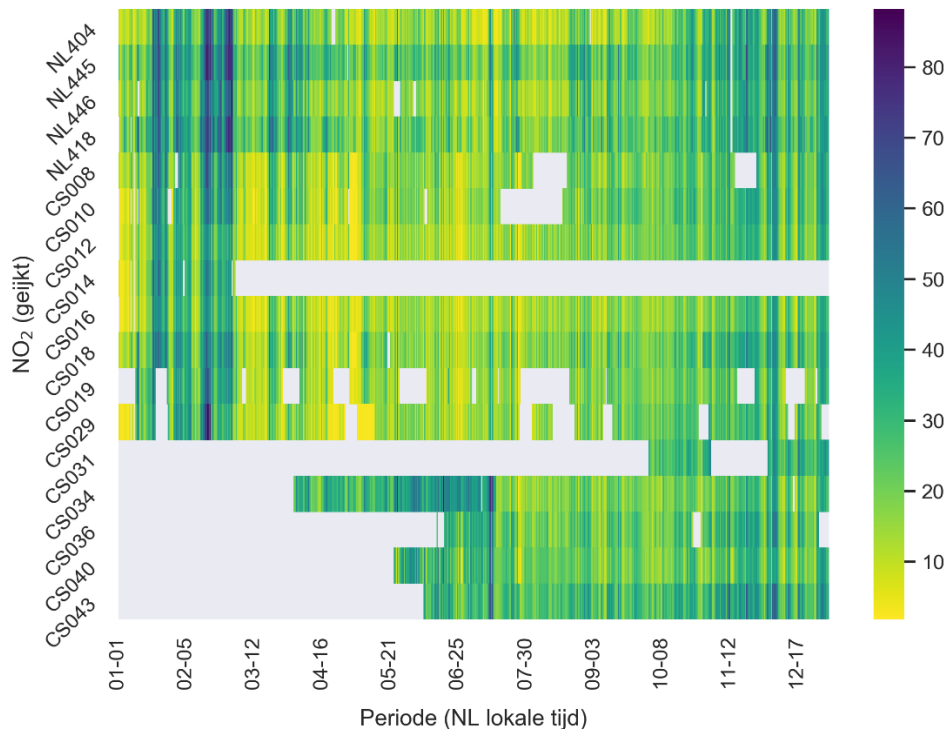
OVERZICHT kaart	Maakt een heatmap die aangeeft welke gemeten grootheden (de attributen) verkregen zijn per meetstation. Elke plot geeft de vergelijking voor een enkele attribuut voor alle opgegeven stations.
Velden	
Kop	OVERZICHT: definieert de kaart
Eerste velden (facultatief)	Tijdinfoblok (tijdbereik en sampling)
Volgende velden	Meetreeks blok - per type attribuut wordt één plot gemaakt
Volgende velden	Stationblok
• Alle KADER/NIETGEBRUIKEN/BEREIK kaarten met stations en de KNMI_GEGEVENS kaart moeten al eerder zijn ingelezen. • Indien er geen OVERZICHT kaart wordt gegeven, worden geen plots gegenereerd. • Toegestane AANPASSING opties: FORMAAT/ADRES/YLIMIET. Standaard formaat = 8.5, 6. Ylimiet bepaalt de waarden aan de uiterste zijden van de kleurenbalk.	
Naam plotfile: "attribuut" _tijd_overz_ "teller" of "attribuut" _tijd_overz_sampling_ "teller"	

NO2 Overzicht Metingen

Jaaroverzicht 2019
N02 Overzicht Metingen
File:N02_tijd_overz_(I)

Periode: 1/1/2019 - 31/12/2019
wk 1 - 1

NO₂ (geijkt) metingen versus tijd (geplot per 14H)



Jaaroverzicht 2019
N02 Overzicht Metingen
File:N02_tijd_overz_(I)

Periode: 1/1/2019 - 31/12/2019
wk 1 - 1

Station/Meting	Gemiddeld	Maximum	Loc Max	Minimum	Loc Min	Mediaan	Aantal
NL404 N02	22.3	138.6	0121:08	1.7	0901:16	16.9	7985
NL445 N02	30.0	137.3	0121:07	3.0	0102:03	25.9	8477
NL446 N02	24.2	143.6	0121:10	2.2	0630:16	19.5	8159
NL418 N02	28.6	113.6	0121:10	3.0	0707:11	24.7	8417
CS008 N02	22.3	116.9	0121:09	0.90	0416:19	18.8	7178
CS012 N02	23.7	99.8	0118:10	0.84	0427:17	21.0	6791
CS014 N02	19.4	89.7	0118:11	0.84	0828:13	17.5	7728
CS016 N02	28.8	123.6	0121:09	0.88	0101:13	26.0	1218
CS018 N02	18.4	86.8	0118:11	0.53	0412:16	15.7	7687
CS019 N02	23.9	98.8	0118:09	0.88	0522:18	20.9	7691
CS029 N02	21.7	106.1	0118:09	0.90	0411:17	18.8	5734
CS031 N02	20.9	115.5	0215:20	0.86	0426:11	18.3	6664
CS034 N02	30.9	72.0	1216:19	8.0	1027:05	29.5	1417
CS036 N02	27.5	115.2	0524:00	2.1	0724:16	25.1	6119
CS040 N02	26.6	76.4	0828:07	2.1	0724:20	24.3	4186
CS043 N02	25.5	78.9	0602:06	1.2	0626:19	23.6	4842
CS052 N02	31.2	89.6	1203:08	2.5	0623:11	28.0	4703

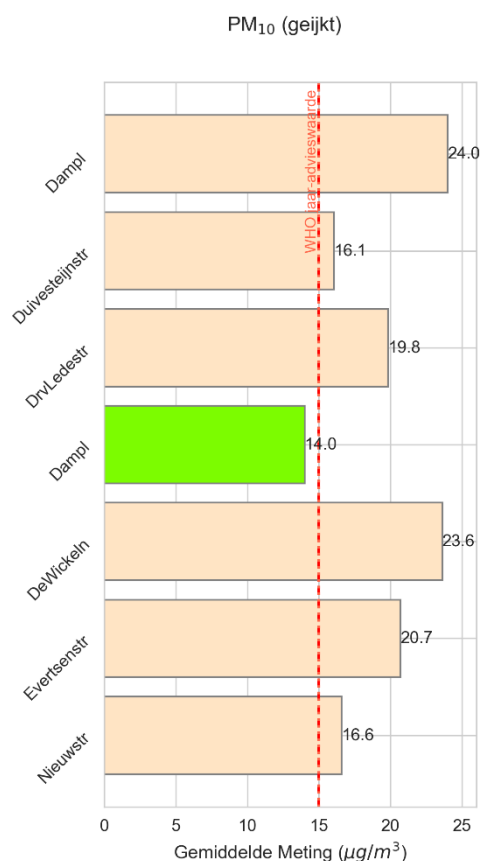
Output van een OVERZICHT opdracht voor stikstofmetingen. De bemonstering (oorspronkelijk per uur) is aangepast aan het oplossend vermogen van de figuur. Statistische gegevens (bemonstering 1H) zijn onderin aangegeven.

2. Om het niveau van de metingen van vervuiling tussen stations te vergelijken kunnen staafdiagrammen,

boxplots of doosdiagrammen worden gegenereerd. De staafdiagram is het meest eenvoudig. Men krijgt alleen de waarde van de gemiddelden van de tijdreeksen te zien.

In de standaard uitvoering ziet een plot gemaakt met CLSTAAFDIAGRAM er als volgt uit:

De kleurkeuze is eenvoudig: als de gemiddelde waarde onder de jaarnorm van de vervuiling is, is de staaf groen. Boven de dagnorm wordt de staaf rood. Daartussen is hij geel. In de plot hieronder is SB936 een gunstige uitzondering. Dit is geen toeval – het meetstation staat binnen, en er is een luchtreiniger.



Men kan de verschijningsvorm van de plot wijzigen met AANPASSING kaarten, die aan de STAAFDIAGRAM kaart moeten voorafgaan.

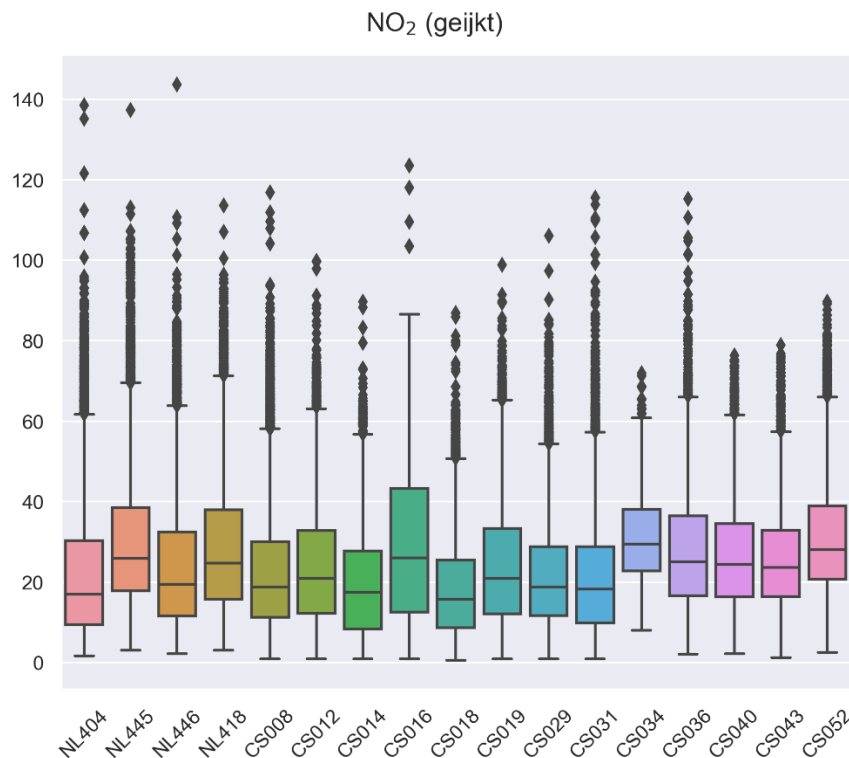
Relevante kaart:

CLSTAAFDIAGRAM

CLSTAAFDIAGRAM kaart	Geeft aan welke gemeten grootheden (de attributen) en welke gegevens die van het KNMI meetstation zijn verkregen worden vergeleken in doosdiagrammen. Elke plot geeft de vergelijking voor een enkele attribuut voor alle opgegeven stations.
Velden	
Kop	CLSTAAFDIAGRAM: definieert de kaart
Parameters 1-n	Lijst van grootheden. Toegestane namen: alle mogelijke sensormetingen: NO2, NO2_ref1, NO2_ref2, NO2_gemeten P10, P10_ref1, P10_ref2, P10_gemeten P25, P25_ref1, P25_ref2, P25_gemeten temp, rh, ld
Parameters n+1 tot eind	De stations waarvoor plots worden gemaakt. Dit kunnen ook de pseudo stations zijn die eerder zijn gegenereerd. Dezelfde conventies worden gebruikt als op de KADER kaart. Bovendien kan ALL of

	ALL+ worden gespecificeerd. ALL staat voor alle stations waarmee gemeten is; ALL+ omvat ook de stations die zijn gegenereerd in dit programma.
• Alle KADER/NIETGEBRUIKEN/BEREIK kaarten met stations en de KNMI_GEGEVENS kaart moeten al eerder zijn ingelezen. • Indien er geen STAAFDIAGRAM kaart wordt gegeven, worden geen plots gegenereerd. • Als er een STAAFDIAGRAM kaart wordt gegeven zonder dat er stations worden gespecificeerd, wordt de kaart genegeerd. • Als er een STAAFDIAGRAM kaart wordt gegeven zonder dat attributen worden opgegeven, worden plots gemaakt voor NO₂, P10 en P25. • Toegestane AANPASSING kaarten voor de STAAFDIAGRAM kaart. Opties: ACHTERGROND, FORMAAT, KLEUR, ADRES	
Naam plotfile: "attribuut" _AirQ_BarP_"teller" of "attribuut" _AirQ_BoxP_samplng_"teller"	

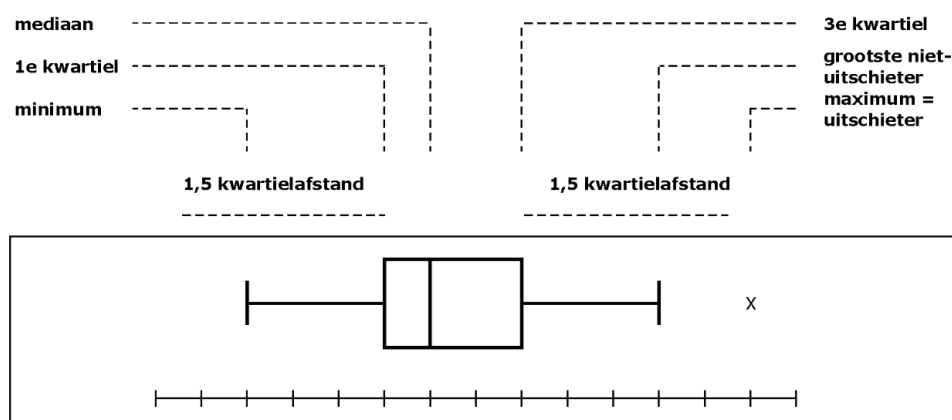
Een boxplot geeft een beeld van de statistische verdeling van de waarden in een tijdreeks. In het voorbeeld hieronder kan men per station de mediaan van de NO₂ metingen zien: het horizontale streepje binnen de gekleurde doos. 50% van de metingen in de tijdsperiode ligt boven de mediaan, en 50 % daaronder. De grenzen van de doos wordt gegeven door de waarden van vervuiling waar 75% van de metingen erboven ligt, respectievelijk 25%. De balken boven en onder de doos geven de rest van de metingen aan, het bereik van de bovenste 25% en het bereik van de onderste 25%. Ook worden outliers of uitschieters aangegeven.



Voorbeeld: De "dozen" in de boxplot geven het bereik aan van NO₂ metingen in de tijdsperiode zoals gedefinieerd op de ANALYSEPERIODE kaart. CS016 lijkt meer variatie te tonen dan de andere stations.

Exacte definities worden gegeven in deze figuur van Wikipedia (Door Madyno - Eigen werk, CC BY-SA 3.0,

<https://commons.wikimedia.org/w/index.php?curid=17013933>).



In ALVA worden doosdiagrammen van verschillende tijdreeksen naast elkaar geplot, en statistische informatie wordt standaard toegevoegd. Deze informatie is identiek aan de informatie gegenereerd op basis van de OVERZICHT kaart: gemiddelde, minimum en maximum van de metingen over de meetperiode.

Relevante kaart:

DOOSDIAGRAM

DOOSDIAGRAM kaart	Geeft aan welke gemeten grootheden (de attributen) en welke gegevens die van het KNMI meetstation zijn verkregen worden vergeleken in doosdiagrammen. Elke plot geeft de vergelijking voor een enkele attribuut voor alle opgegeven stations.
Velden	
Kop	DOOSDIAGRAM: definieert de kaart
Eerste velden (facultatief)	Tijdinfo blok sampling en tijdbereik
Volgende velden	Meetreeksblok Per type attribuut wordt één plot gegenereerd
Volgende velden	Stationblok
	- Alle KADER/NIETGEBRUIKEN/BEREIK kaarten met stations en de KNMI_GEGEVENS kaart moeten al eerder zijn ingelezen. - Indien er geen DOOSDIAGRAM kaart wordt gegeven, worden geen plots gegenereerd. - Toegestane AANPASSING kaarten voor de DOOSDIAGRAM kaart. Opties: ACHTERGROND, FORMAAT, KLEUR, YLIMIET, ADRES
	Naam plotfile: "attribuut" _AirQ_BoxP_ "teller" of "attribuut" _AirQ_BoxP_ sampling_ "teller"

3. Een Vioolplot geeft, net als een doosdiagram, een idee van de mediane waarde en de verdeling van de waardes van een meetreeks. De plot ziet er anders uit.

Relevante kaart:

VIOOLPLOT

VIOOLPLOT kaart	Geeft aan welke gemeten grootheden (de attributen) en welke gegevens die van het KNMI meetstation zijn verkregen worden vergeleken in doosdiagrammen. Elke plot geeft de vergelijking voor een enkele attribuut voor alle opgegeven stations.
Velden	
Kop	VIOOLPLOT: definieert de kaart
Eerste velden (facultatief)	Tijdinfoblok Sampling en tijdbereik
Volgende velden	Meetreeksblok Per type attribuut wordt één plot gegenereerd

Volgende velden	Stationblok
- Alle KADER/NIETGEBRUIKEN/BEREIK kaarten met stations en de KNMI_GEGEVENS kaart moeten al eerder zijn ingelezen. - Toegestane AANPASSING kaarten voor de DOOSDIAGRAM kaart. Opties: ACHTERGROND, FORMAAT, KLEUR, YLIMIT,ADRES Standaard formaat = 8,6	
Naam plotfile: "attribuut " _AirQ_ViolP_ "teller" of "attribuut" _AirQ_BoxP_ sampling_ "teller"	

Het kan nuttig zijn om één run een aantal vrijwil identieke doosdiagrammen te produceren, waarbij het enige verschil ligt in de kleurenkeuze. Met behulp van de CLDOOSDIAGRAM kaart wordt een serie doosdiagrammen geproduceerd, waarbij alle doosdiagram figuren dezelfde kleur hebben, behalve één die gehighlight wordt met een andere kleur. Met behulp van de AANPASSING kaart kan men het plaatje naar keuze veranderen.

Het plaatje hieronder is geproduceerd met de reeks kaarten :

AANPASSING FORMAAT 8 3

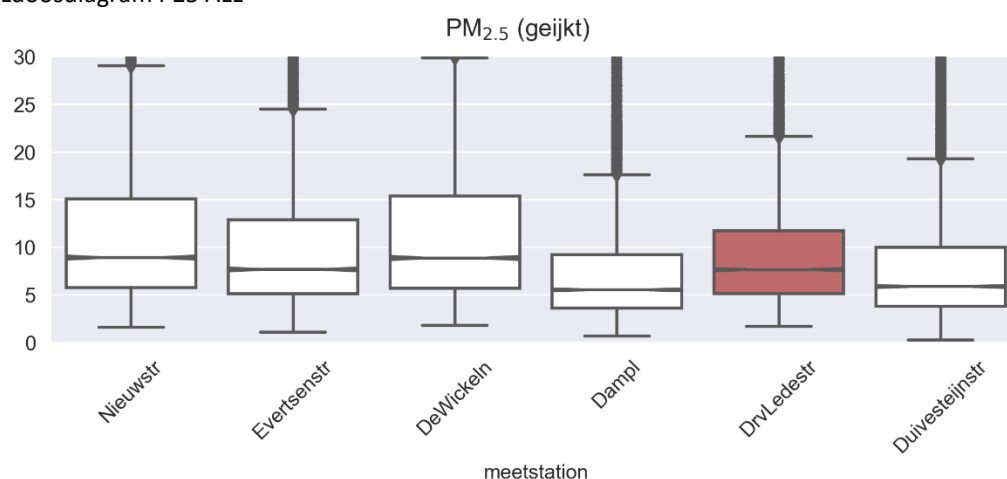
AANPASSING YLIMIT 0 30

AANPASSING ADRES

AANPASSING KLEUR white indianred green

AANPASSING ACHTERGROND DONKER

CLdoosdiagram P25 ALL



Relevante kaart:

CLDOOSDIAGRAM

CLDOOSDIAGRAM kaart	Geeft aan welke gemeten grootheden (de attributen) en welke gegevens die van het KNMI meetstation zijn verkregen worden vergeleken in doosdiagrammen. Elke plot geeft de vergelijking voor een enkele attribuut voor alle opgegeven stations.
Velden	
Kop	CLDOOSDIAGRAM: definieert de kaart
Parameters 1-n	Lijst van grootheden. Toegestane namen: alle mogelijke sensormetingen: NO2, NO2_ref1, NO2_ref2, NO2_gemeten P10, P10_ref1, P10_ref2, P10_gemeten P25, P25_ref1, P25_ref2, P25_gemeten temp, rh, ld
Parameters n+1 tot eind	De stations waarvoor plots worden gemaakt. Dit kunnen ook de pseudo stations zijn die eerder zijn gegenereerd. Dezelfde conventies worden gebruikt als op de KADER kaart. Bovendien kan ALL of ALL+ worden gespecificeerd. ALL staat voor alle stations waarmee gemeten is; ALL+ omvat ook de stations die zijn gegenereerd in dit programma.

• Alle KADER/NIETGEBRUIKEN/BEREIK kaarten met stations en de KNMI_GEGEVENS kaart moeten al eerder zijn ingelezen. • Indien er geen CLDOOSDIAGRAM kaart wordt gegeven, worden geen plots gegenereerd. • Als er een CLDOOSDIAGRAM kaart wordt gegeven zonder dat er stations worden gespecificeerd, wordt de kaart genegeerd. • Als er een CLDOOSDIAGRAM kaart wordt gegeven zonder dat attributen worden opgegeven, worden plots gemaakt voor NO2, P10 en P25. • Toegestane AANPASSING kaarten voor de DOOSDIAGRAM kaart. Opties: ACHTERGROND, FORMAAT, KLEUR, YLIMIET, ADRES Standaard formaat = 8,6 De normale kleur van het doosdiagram voor een meting van een station is bepaald door de eerste opgegeven kleur op de AANPASSING KLEUR kaart. Het door een kleur gehighlighte station heeft de kleur zoals aangegeven als de tweede kleur op de AANPASSING KLEUR kaart.
Naam plotfile: "attribuut" _AirQ_BoxP_ "teller" of "attribuut" _AirQ_BoxP_ sampling_ "teller"

4. Vergelijking van amplitude waarden kan ook anders. De vioolplot en het doosdiagram geven een vergelijking van de amplitudes en de amplitudeverdeling en tonen de resultaten naast elkaar. Ook van belang is in hoeverre tijdreeksen op elkaar lijken en of de waarden van de ene tijdreeks een relatie heeft met de andere tijdreeks. De kruiscorrelatie geeft een goede maat of er een verband is tussen de tijdreeksen (zie hieronder); zij geeft aan hoe twee signalen zich tot elkaar verhouden. De kruiscorrelatie waarden worden in kleur aangegeven in een matrix.

Men kan ook een vergelijking van amplitude waarden in een matrix tonen. Dit gebeurt als men de kaart VERGELIJK_AMPL invult. Dan worden de volgende waarden tussen een aantal stations twee aan twee vergeleken :

- De meetwaarden worden twee aan twee vergeleken, en er wordt een lijn getrokken die het best de relatie van de set meetpunten en de andere set meetpunten weergeeft. Dit zal in het algemeen een lineaire relatie zijn in de vorm $y = ax + b$, waarbij b niet 0 is. De term a is de richtingscoëfficiënt, en geeft de relatieve stijging aan van y als x toeneemt. De waarde van a wordt voor elke combinatie van a in kleur weergegeven als de optie "relatief" wordt gebruikt.
- Eenzelfde lineaire relatie kan worden afgeleid, waarbij men de waarde b op nul stelt. Dit gebeurt onder de optie "relatief0".
- Bij de optie "absoluut" worden de gemiddelde waarden vergeleken.

Relevante kaart:

VERGELIJK_AMPL

VERGELIJK_AMPL kaart	Vergelijkt gemeten grootheden (de attributen) in matrixvorm. Elke plot geeft de vergelijkingen voor een enkele attribuut voor alle opgegeven stations.
Velden	
Kop	VERGELIJK_AMPL: definieert de kaart
Parameter 1	Type vergelijking Kan zijn: relatief relatief0 absoluut
Parameters 2-n	Lijst van namen van meetreeksen/attributen. Toegestane namen: alle mogelijke sensormetingen: NO2, NO2_ref1, NO2_ref2, NO2_gemeten P10, P10_ref1, P10_ref2, P10_gemeten P25, P25_ref1, P25_ref2, P25_gemeten temp, rh, ld
Parameters n+1 tot eind	De stations waarvoor plots worden gemaakt. Dit kunnen ook de pseudo stations zijn die eerder zijn gegenereerd. Dezelfde conventies worden gebruikt als op de KADER kaart. Bovendien kan ALL of

	ALL+ worden gespecificeerd. ALL staat voor alle stations waarmee gemeten is; ALL+ omvat ook de stations die zijn gegenereerd in dit programma.
• Alle KADER/NIETGEBRUIKEN/BEREIK kaarten met stations en de KNMI_GEGEVENS kaart moeten al eerder zijn ingelezen. • Indien er geen KRUISCORRELATIE kaart wordt gegeven, worden geen plots gegenereerd. • Als er een KRUISCORRELATIE kaart wordt gegeven zonder dat er stations worden gespecificeerd, wordt de kaart genegeerd. • Toegestane AANPASSING kaarten voor de VERGELIJK_AMPL kaart. Opties: FORMAAT, ADRES Standaard formaat = 8.5,6.5 De ACHTERGROND kan niet worden veranderd: deze is altijd grijs.	
Naam plotfile: "attribuut" _airQ_ Vgl"type (R,R0 of Q)" _"teller"	

5. Ook een maat voor de gelijkvormigheid van tijdreeksen kan worden aangegeven. Dit gebeurt door kruiscorrelatie. Een kruiscorrelatie tussen twee tijdreeksen is gelijk aan 1 als twee tijdreeksen identiek zijn (op een constante factor na), en 0 als er geen gelijkheid is te ontdekken. Bij een kruiscorrelatie van -1 zijn de twee tijdreeksen gelijkvormig maar verschilt het teken. In ALVA worden drie verschillende maten gebruikt, de rekenkundige kruiscorrelatie (Pearson) en een kruiscorrelatie die gelijkvormigheid ook erkent als toppen en dalen gelijk lopen, maar als de verhouding tussen toppen en dalen verschilt (Spearman rank correlatie en Kendalls Tau correlatie). In alle gevallen is het slecht nieuws als de correlatie tussen twee stations beduidend lager is dan 1.

Een kruiscorrelatie geeft aan hoe twee signalen zich tot elkaar verhouden. De kruiscorrelatie waarden worden in kleur aangegeven in een matrix. De rijen geven de kruiscorrelatie van één station met de andere stations aan, en er zijn evenveel rijen en kolommen als er stations zijn. De matrix is symmetrisch, vanzelfsprekend is de kruiscorrelatie van een tijdreeks van een station met zichzelf gelijk aan 1. Omdat de matrix symmetrisch is, hoeft alleen maar de onderste helft te worden aangegeven.

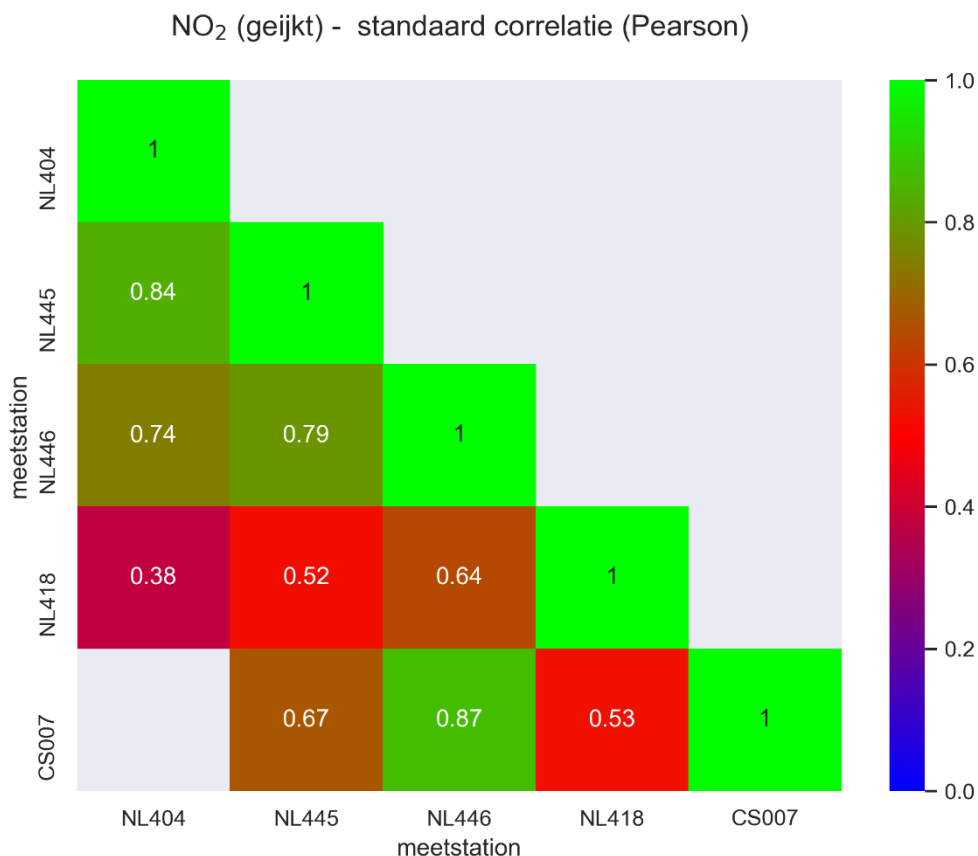
De kleurenschaal in onderstaand voorbeeld loopt van blauw tot groen via rood en komt overeen met het bereik van 0 tot 1. Het tijdsbereik in dit voorbeeld is erg kort, 2 dagen, en er missen data punten (dit laatste is bewust gedaan om opties in het programma uit te testen).

Als er veel stations met elkaar worden vergeleken, worden de correlatie waarden in de vierkantjes niet aangegeven.

NO2 standaard correlatie (Pearson)

Nieuwjaarsnacht 2020
 NO2 standaard correlatie (Pearson)
 File:NO2_AirQ_CorP_(IV)

Periode: 31/12/2019 - 1/1/2020
 wk 1 - 1



Nieuwjaarsnacht 2020
 NO2 standaard correlatie (Pearson)
 File:NO2_AirQ_CorP_(IV)

Periode: 31/12/2019 - 1/1/2020
 wk 1 - 1

Station/Meting	Gemiddeld	Maximum	Loc Max	Minimum	Loc Min	Mediaan	Aantal
NL404 NO2	29.6	65.1	1231:19	9.8	1231:07	26.3	24
NL445 NO2	36.2	72.2	1231:19	14.4	1231:06	34.1	48
NL446 NO2	33.8	71.1	1231:18	15.7	1231:14	29.9	48
NL418 NO2	33.0	64.4	1231:18	15.8	0101:13	31.1	48
CS007 NO2	39.4	45.7	0101:18	33.2	0101:21	38.3	22

Relevante kaart:

KRUISCORRELATIE

KRUISCORRELATIE kaart	Geeft aan welke gemeten grootheden (de attributen) en welke gegevens die van het KNMI meetstation zijn verkregen worden vergeleken in een kruiscorrelatie matrix. Elke plot geeft de vergelijkingen voor een enkele attribuut zoals gemeten op alle opgegeven stations of de vergelijking tussen attributen voor de stations.
-----------------------	---

Velden	
Kop	KRUISCORRELATIE: definieert de kaart
Eerste veld (facultatief – als niet opgegeven worden kruiscorrelaties tussen stations voor een enkele meting gegeven)	0 of “Meting” : kruiscorrelaties tussen stations voor een enkele attribuut 1 of “Station” : kruiscorrelaties tussen verschillende metingen van een station
Volgende veld (facultatief; indien niet gegeven wordt PEARSON gebruikt)	Type kruiscorrelatie. Kan zijn: PEARSON (- standaard lineaire correlatie) SPEARMAN KENDALL
Volgende velden	Meetreeksblok
Volgende velden	Stationblok
- Alle KADER/NIETGEBRUIKEN/BEREIK kaarten met stations en de KNMI_GEGEVENS kaart moeten al eerder zijn ingelezen. - Indien er geen KRUISCORRELATIE kaart wordt gegeven, worden geen plots gegenereerd. - Toegestane AANPASSING kaarten voor de KRUISCORRELATIE kaart. Opties: FORMAAT, YLIMIET, ADRES Standaard formaat = 8.5, 6.5 De YLIMIET geeft het minimum en het maximum van de te plotten correlatie aan. De ACHTERGROND kan niet worden veranderd: deze is altijd grijs.	
Naam plotfile: “attribuut”_airQ_Cor “type (P,S of K)”_”teller” of “station”_AirQ_Cor etc.	

Tijdsafhankelijkheden: Plotten van de tijdreeksen.

Per station

De meetgegevens komen als functie van tijd binnen. De tijdreeksen van metingen van verschillende aard voor één station kunnen op één plot worden vertoond.

Relevante kaart:

PLOT

PLOT kaart	Geeft aan welke gemeten grootheden (de attributen) en welke gegevens die van het KNMI meetstation zijn verkregen worden geplotted op tijdreeksplots. Elke plot geeft attributen voor een enkel station.
Velden	
Kop	PLOT: definieert de kaart
Parameters 1-n	Lijst van grootheden. Toegestane namen: alle mogelijke sensormetingen: NO2, NO2_ref1, NO2_ref2, NO2_kal P10, P10_ref1, P10_ref2, P10_kal P25, P25_ref1, P25_ref2, P25_kal temp, rh, ld Alle berekende attributen voor dit station Alle KNMI gegevens: Dwind, Vwind, Windstoot, TempKNMI, DauwTemp, LVocht, RegenDuur, Neerslag, Ldruk, Zicht, Instraling Twee afgeleide gegevens: Dwind10, Dwind9, waarbij Dwind10 de windrichting in graden is gedeeld door 10, en Dwind9 dezelfde windrichting gedeeld door 9.

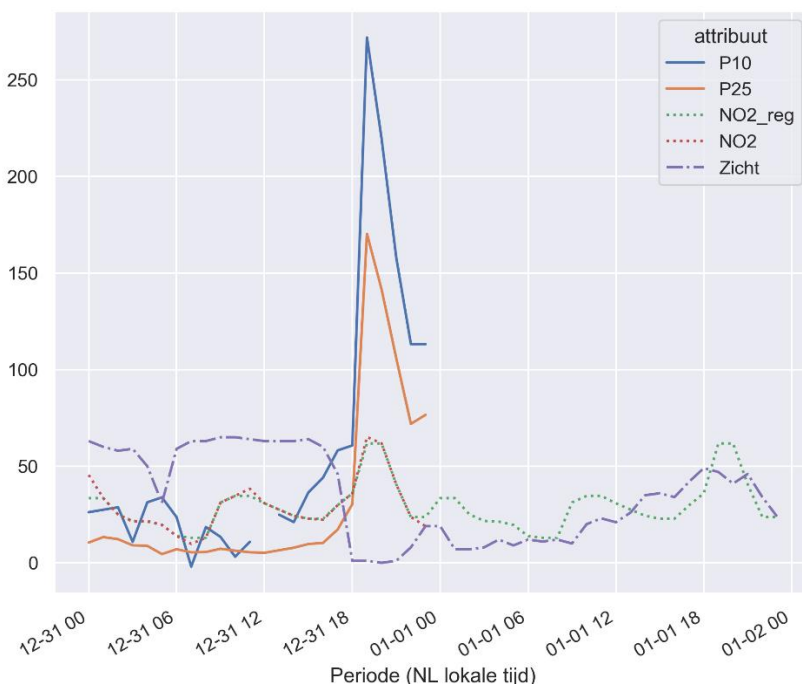
Parameters n+1 tot eind	De stations waarvoor plots worden gemaakt. Dit kunnen ook de pseudo stations zijn die eerder zijn gegenereerd – deze kaart kan dus ook gebruikt worden om gemiddelde waardes van meetreeksen te laten plotten. Dezelfde conventies worden gebruikt als op de KADER kaart. Bovendien kan ALL of ALL+ worden gespecificeerd. ALL staat voor alle stations waarmee gemeten is; ALL+ omvat ook de stations die zijn gegenereerd in dit programma.
• Alle KADER/NIETGEBRUIKEN/BEREIK kaarten met stations en de KNMI_GEGEVENS kaart moeten al eerder zijn ingelezen. • Indien er geen PLOT kaart wordt gegeven, worden geen plots gegenereerd. • Als er een PLOT kaart wordt gegeven zonder dat er stations worden gespecificeerd, wordt de kaart genegeerd. • Als er een PLOT kaart wordt gegeven zonder dat attributen worden opgegeven, worden NO2, P10, P25, temp en rh als mee te nemen parameters genomen. • Toegestane AANPASSING kaarten voor de PLOT kaart: alle opties- ACHTERGROND, KLEUR, FORMAAT, YLIMIT, STIJL, ADRES Standaard formaat = 8,7	
Naam plotfile: "station_AirQ_"teller" of "station_AirQ_sampling_"teller"	

Metingen door NL404 en KNMI215

Nieuwjaarsnacht 2020
Metingen door NL404 en KNMI215
File:NL404_AirQ_(V)

Periode: 31/12/2019 - 1/1/2020
wk 1 - 1

Meetwaarden bij NL10404 - Den Haag-Rebecquestraat (NL404)



Nieuwjaarsnacht 2020
Metingen door NL404 en KNMI215
File:NL404_AirQ_(V)

Periode: 31/12/2019 - 1/1/2020
wk 1 - 1

Station/Meting	Gemiddeld	Maximum	Loc Max	Minimum	Loc Min	Mediaan	Aantal
NL404 P10	58.6	271.9	1231:19	-2.0	1231:07	28.8	23
NL404 P25	31.2	170.3	1231:19	4.5	1231:05	9.4	24
NL404 NO2_reg	29.2	61.8	1231:19	13.0	1231:07	26.3	48
NL404 NO2	29.6	65.1	1231:19	9.8	1231:07	26.3	24
KNMI215 Zicht	34.9	65.0	1231:09	0.00	1231:20	34.5	48

Voorbeeld van PLOT: tijdreeksen van metingen op station NL10404 voor 31/12/19, en berekende attribuut P10_reg (berekend met DAGRITME), en de aanduiding van mist (KNMI: Horizontaal Zicht) voor die dag en voor 1/1/20.

Per attribuut

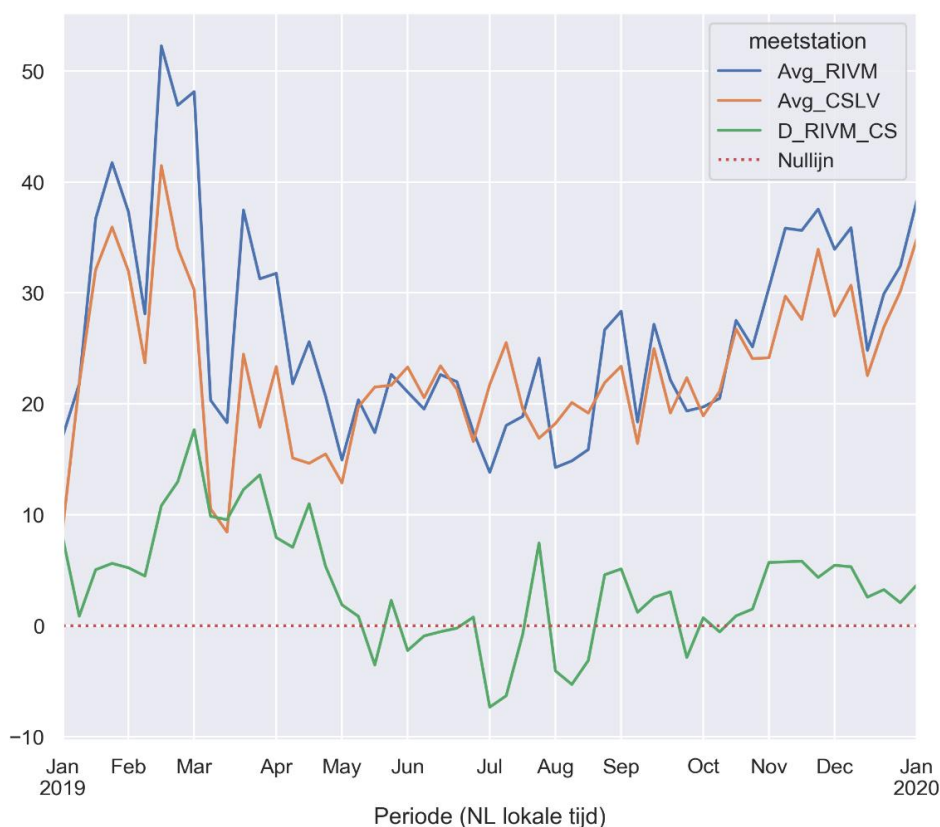
De meetgegevens voor attributen zoals NO2, PM10 en PM2,5 kunnen per vervuilingsoort tussen geselecteerde stations vergeleken worden. Dit kunnen ook pseudostations zijn.

NO2 gemeten bij de stations

Jaaroverzicht 2019
NO2 gemeten bij de stations
File:NO2_AirQ_W_(VI)

Periode: 1/1/2019 - 31/12/2019
wk 1 - 1

NO₂ (geijkt) Meetwaarden (per week) - gemeten bij de stations



Jaaroverzicht 2019
NO2 gemeten bij de stations
File:NO2_AirQ_W_(VI)

Periode: 1/1/2019 - 31/12/2019
wk 1 - 1

Station/Meting	Gemiddeld	Maximum	Loc Max	Minimum	Loc Min	Mediaan	Aantal
vg_RIVM NO2	26.5	52.3	0217:00	13.8	0707:00	24.1	53
vg_CSLV NO2	23.0	41.5	0217:00	8.4	0317:00	22.4	53
RIVM_CS NO2	3.5	17.7	0303:00	-7.3	0707:00	3.3	53
Nullijn NO2	0.00	0.00	0106:00	0.00	0106:00	0.00	53

Voorbeeld: het pseudostation Avg_RIVM bevat het gemiddelde van de NO2 metingen van een aantal RIVM stations, Avg_CSLV bevat die van een aantal Citizen Science stations. Met BEREKEN_STAT is het

verschil tussen deze twee berekend (Pseudostation D_RIVM_CS). Ook Nullijn is een pseudostation, waarbij de waarde van de tijdreeks NO2 op 0 is gezet.
Plotting is met PLOT_SOORT (na resampling naar wekelijkse intervallen).

Relevante kaart:

PLOT_SOORT

PLOT_SOORT kaart	Geeft aan welke op de stations gemeten grootheden (meetreeksen/attributen) en welke gegevens die van het KNMI meetstation zijn verkregen worden geplot op de tijdreeksplots. Elke plot geeft de waarden van een enkele attribuut/meetreeks voor alle gespecificeerde stations.
Velden	
Kop	PLOT_SOORT: definieert de kaart
Parameters 1-n	Lijst van grootheden (attributen). Toegestane namen: alle mogelijke sensormetingen: NO2, NO2_ref1, NO2_ref2, NO2_kal P10, P10_ref1, P10_ref2, P10_kal P25, P25_ref1, P25_ref2, P25_kal temp, rh, ld en de attributen die eerder zijn gegenereerd.
Parameters n+1 tot eind	De stations waarvoor plots worden gemaakt. Dit kunnen ook de pseudo stations zijn die eerder zijn gegenereerd. Dezelfde conventies worden gebruikt als op de KADER kaart. Bovendien kan ALL of ALL+ worden gespecificeerd. ALL staat voor alle stations waarmee gemeten is; ALL+ omvat ook de stations die zijn gegenereerd in dit programma.
• Alle KADER/NIETGEBRUIKEN/BEREIK kaarten met stations en de KNMI_GEGEVENS kaart moeten al eerder zijn ingelezen. • Indien er geen PLOT_SOORT kaart wordt gegeven, worden geen plots gegenereerd. • Als er een PLOT_SOORT kaart wordt gegeven zonder dat er stations worden gespecificeerd, wordt de kaart genegeerd. • Als er een PLOT_SOORT kaart wordt gegeven zonder dat attributen worden opgegeven, worden NO2, P10, P25 als mee te nemen parameters genomen. • Toegestane AANPASSING kaarten voor de PLOT_SOORT kaart: alle opties- ACHTERGROND, KLEUR, FORMAAT, YLIMIET, STIJL, ADRES Standaard formaat = 8,7	
Naam plotfile: "attribuut" _AirQ_ "teller" of "attribuut" _AirQ_ sampling_ "teller"	

Gedetailleerde vergelijkingen tussen twee stations

Individuele meetwaarden van paren van stations worden vergeleken op twee manieren. Ten eerste worden tijdreeksen vergeleken, ten tweede worden crossplots (kruisdiagrammen) en distributies gemaakt van de metingen over het volledige tijdsinterval.

De hier beschreven functies kunnen ook gebruikt worden om de meetreeksen van een individueel station te vergelijken met die van een pseudostations met gemiddelde waarden, of door twee verschillende genomen gemiddeldes te vergelijken.

Relevante kaart:

VERGELIJK_PAREN

VERGELIJK_PAREN kaart	Specificeert de gegevens voor plots waarbij een aantal attributen wordt geplot voor twee verschillende stations bij elkaar
Velden	
Kop	VERGELIJK_PAREN : definieert de kaart

Parameter 1-N	Lijst van grootheden waarvoor vergelijkingen worden gemaakt tussen paren van stations. Toegestane namen: alle mogelijke sensormetingen: NO2, NO2_ref1, NO2_ref2, NO2_gemeten P10, P10_ref1, P10_ref2, P10_gemeten P25, P25_ref1, P25_ref2, P25_gemeten temp, rh, ld en berekende attributen Alle KNMI gegevens: Dwind, Vwind, Windstoot, TempKNMI, DauwTemp, LVocht, RegenDuur, Neerslag, Ldruk, Zicht, Instraling Twee afgeleide gegevens: Dwind10, Dwind9, waarbij Dwind10 de windrichting in graden is gedeeld door 10, en Dwind9 dezelfde windrichting gedeeld door 9.
Parameter N+1 -eind	Lijst van paren van stations (Dezelfde conventies worden gebruikt als op de KADER kaart.). Het gebruik van ALL of ALL+ is niet toegestaan. Pseudostations worden aangegeven door _PS gevolgd door een naam.
• Het aantal stations dat wordt opgegeven moet even zijn. • Toegestane AANPASSING kaarten voor de VERGELIJK_PAREN kaart: alle opties- ACHTERGROND, KLEUR, FORMAAT, YLIMIT, STIJL, ADRES Standaard formaat = 8,7	
Naam plotfile: "station1"_"station2"_tijd_"teller"	

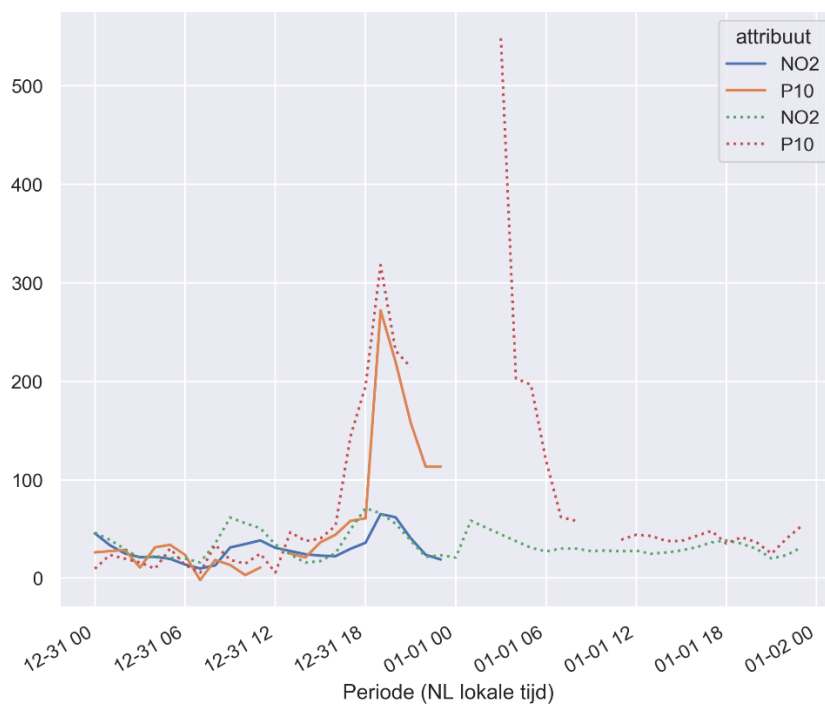
Voorbeeld:

Vergelijking metingen bij NL404 & NL446

Nieuwjaarsnacht 2020
Vergelijking metingen bij NL404 & NL446
File:NL404_NL446_tijd_(VI)

Periode: 31/12/2019 - 1/1/2020
wk 1 - 1

Metingen bij NL404 & NL446



Nieuwjaarsnacht 2020
Vergelijking metingen bij NL404 & NL446
File:NL404_NL446_tijd_(VI)

Periode: 31/12/2019 - 1/1/2020
wk 1 - 1

Station/Meting	Gemiddeld	Maximum	Loc Max	Minimum	Loc Min	Mediaan	Aantal
NL404 NO2	29.6	65.1	1231:19	9.8	1231:07	26.3	24
NL446 NO2	33.8	71.1	1231:18	15.7	1231:14	29.9	48
NL404 P10	58.6	271.9	1231:19	-2.0	1231:07	28.8	23
NL446 P10	81.9	548.4	0101:03	5.7	1231:07	40.3	42

Voorbeeld: vergelijkingen metingen van RIVM stations NL10404 en NL10446 bij de jaarwisseling.

Een crossplot tezamen met de distributies van de amplitudes van de metingen kan een indicatie geven van de kwaliteit van de metingen bij de stations. Paarsgewijs vergelijken van de meetwaarden van de NO2 meetreeksen van het vorige voorbeeld geeft het volgende plaatje (gegenereerd met de VERGELIJK_ATTR2STAT kaart):

NO2 distributie bij stations NL404,NL446

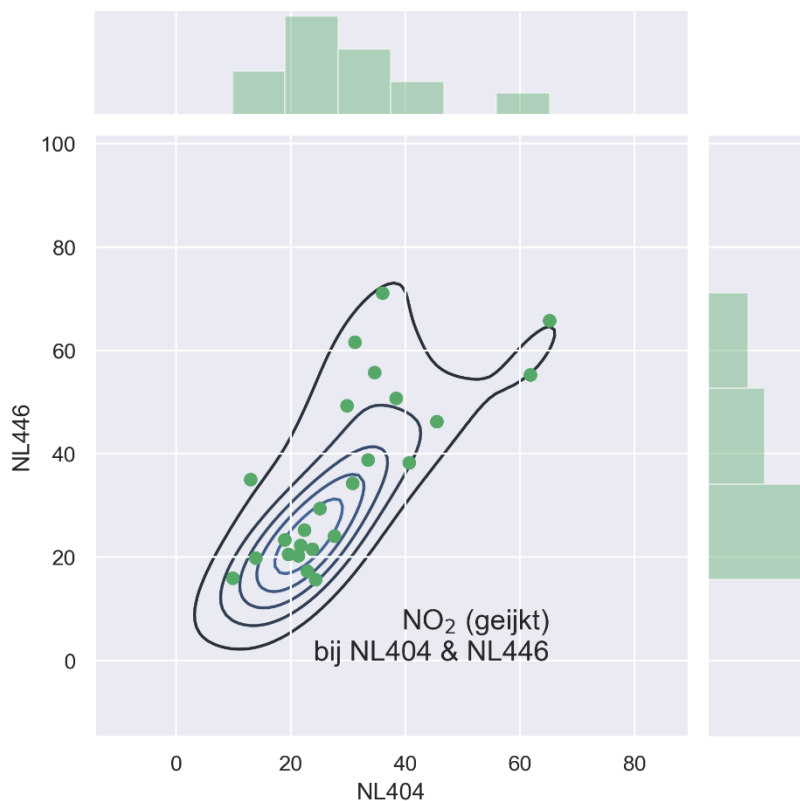
Nieuwjaarsnacht 2020

NO2 distributie bij stations NL404,NL446

File:NO2_NL404_NL446_(VII)

Periode: 31/12/2019 - 1/1/2020

wk 1 - 1



Nieuwjaarsnacht 2020

NO2 distributie bij stations NL404,NL446

File:NO2_NL404_NL446_(VII)

Periode: 31/12/2019 - 1/1/2020

wk 1 - 1

Station/Meting	Gemiddeld	Maximum	Loc Max	Minimum	Loc Min	Mediaan	Aantal
NL404 NO2	29.6	65.1	1231:19	9.8	1231:07	26.3	24
NL446 NO2	33.8	71.1	1231:18	15.7	1231:14	29.9	48

Informatie beste lineaire benadering (kleinste kwadraten)

Onafhank	Afhanklk	op X-as	op Y-as	Y-doorsn	helling	R ²
NL404	NL446	NL404	NL446	7.98	0.94	0.55
NL446	NL404	NL404	NL446	-14.57	1.70	0.55
	Rekenkundig gemiddelde:			-3.30	1.32	
	Harmonisch gemiddelde:			-2.15	1.21	

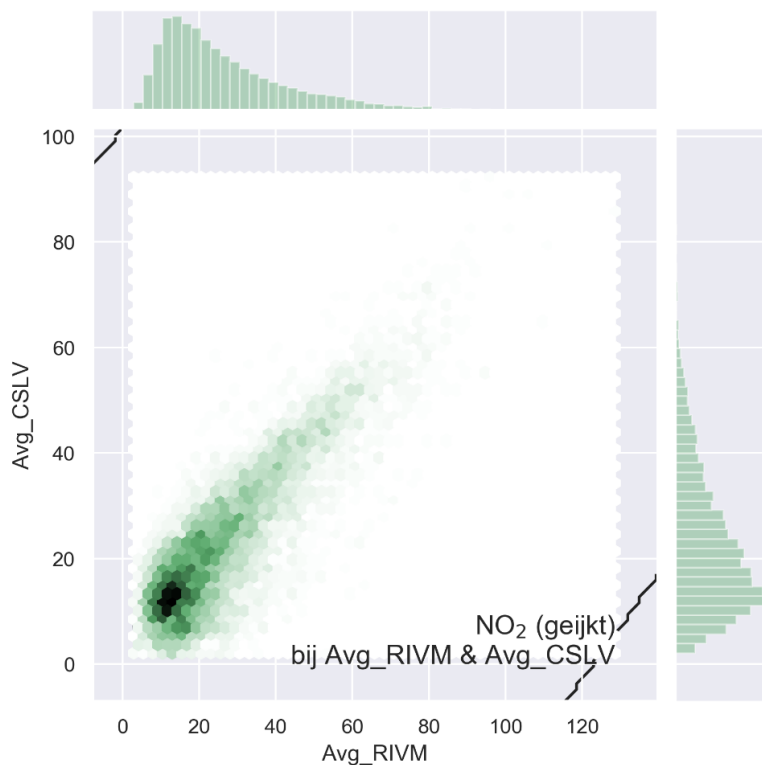
Voorbeeld: Kruisdiagram (crossplot) gegenereerd door VERGELIJK_ATTR2STAT instructie van de amplitudes van twee korte meetreeksen van NO2 (metingen op 31 december 2019) van station NL10404 en NL10446. Een schatting van de formule voor een lineaire relatie tussen de reeksen is onder de plot aangegeven.

In deze vergelijking worden niet alleen een kruisdiagram met distributies getoond, maar ook wordt de beste lijn door de distributie van punten berekend. Omdat beide tijdseries fouten bevatten, wordt het gemiddelde genomen van de twee kleinste kwadraten berekeningen – dat levert waarschijnlijk de “beste” lijn.

Als er veel punten in een vergelijking worden meegenomen wordt het kruisdiagram weergegeven als een dichtheidsplot, waarbij een kleurenschaal aangeeft hoeveel punten aanwezig zijn in een zeshoekje vlakje.

NO2 distributie bij stations Avg_RIVM,Avg_CSLV

Jaaroverzicht 2019
 NO2 distributie bij stations Avg_RIVM,Avg_CSLV Periode: 1/1/2019 - 31/12/2019
 File:NO2_Avg_RIVM_Avg_CSLV_(VII) wk 1 - 1



Jaaroverzicht 2019
 NO2 distributie bij stations Avg_RIVM,Avg_CSLV Periode: 1/1/2019 - 31/12/2019
 File:NO2_Avg_RIVM_Avg_CSLV_(VII) wk 1 - 1

Station/Meting	Gemiddeld	Maximum	Loc Max	Minimum	Loc Min	Mediaan	Aantal
vg_RIVM NO2	26.3	128.6	0121:10	2.9	0102:03	21.9	8669
vg_CSLV NO2	22.9	92.3	0118:09	2.1	0324:17	20.0	8397

Informatie beste lineaire benadering (kleinste kwadraten)

Onafhank	Afhanklk	op X-as	op Y-as	Y-doorsn	helling	R ²
Avg_RIVM	Avg_CSLV	Avg_RIVM	Avg_CSLV	3.88	0.72	0.73
Avg_CSLV	Avg_RIVM	Avg_RIVM	Avg_CSLV	-3.30	0.99	0.73
Rekenkundig gemiddelde:					0.29	0.86
Harmonisch gemiddelde:					0.59	0.84

Voorbeeld: Kruisdiagram (crossplot) gegenereerd door VERGELIJK_ATTR2STAT instructie van de amplitudes van twee lange meetreeksen van NO2 (metingen over geheel 2019). Het voorbeeld hierboven geeft de vergelijking tussen twee pseudostations. Het eerste pseudostation bevat het gemiddelde van de NO2 metingen van RIVM stations, de tweede die van Citizen Science stations.

Relevante kaart:

VERGELIJK_ATTR2STAT

VERGELIJK_ATTR2STAT kaart	Specificeert de gegevens voor plots waarbij één attribuut wordt geplot voor twee verschillende stations in een kruisdiagram.
---------------------------	--

Velden	
Kop	VERGELIJK_ATTR2STAT : definieert de kaart
Parameter 1-N	Lijst van grootheden waarvoor vergelijkingen worden gemaakt tussen paren van stations. Toegestane namen: alle mogelijke sensormetingen: NO2, NO2_ref1, NO2_ref2, NO2_gemeten P10, P10_ref1, P10_ref2, P10_gemeten P25, P25_ref1, P25_ref2, P25_gemeten temp, rh, ld en berekende attributen.
Parameter N+1 -eind	Lijst van paren van stations (Dezelfde conventies worden gebruikt als op de KADER kaart.). Het gebruik van ALL of ALL+ is niet toegestaan. Pseudostations zijn toegestaan (_PS gevolgd door naam).
• Het aantal stations dat wordt opgegeven moet even zijn. • Voor elk attribuut wordt één aparte plot gemaakt. • Toegestane AANPASSING kaarten voor de VERGELIJK_ATTR2STAT kaart: ACHTERGROND, KLEUR, FORMAAT, ADRES Alleen de eerste kleur op de AANPASSING KLEUR kaart wordt gebruikt. Standaard formaat: 8,6	
Naam plotfile: "station1"_"station2"_"attrib"_"teller" of "station1"_"station2"_"attrib"_"sampling_"teller"	

Analyse van relaties tussen metingen

Scatterplots of puntgrafieken

Gemeten grootheden kunnen gerelateerd zijn aan andere metingen. De invloed van weer op vervuiling is groot, en dat kan op verschillende manieren zichtbaar worden gemaakt. Zo zijn er een groot aantal relaties te bedenken.

Als een grootheid afhankelijk is van een andere grootheid, is dat vaak te herkennen op een scatterplot of puntgrafiek. Een puntgrafiek geeft één meetgegeven van één soort op een tijdstip uitgezet tegen een ander van een ander soort op datzelfde tijdstip. Elk meetgegeven kan worden opgegeven, en vergeleken met elk ander meetgegeven, zolang de meetgegevens via een tijdreeks met elkaar in verband kunnen worden gebracht.

Er kan een puntgrafiek gemaakt worden waarbij metingen van twee grootheden met elkaar worden vergeleken, voor een aantal stations (mogelijkerwijs alle stations op één plot). Dit gebeurt voor alle sampling intervallen die eerder zijn gespecificeerd. Ook hier wordt een "beste" lijn berekend zoals bij VERGELIJK_ATTR2STAT, maar dit is vaak niet zinnig.

Relevante kaart:

PUNTGRAFIEK

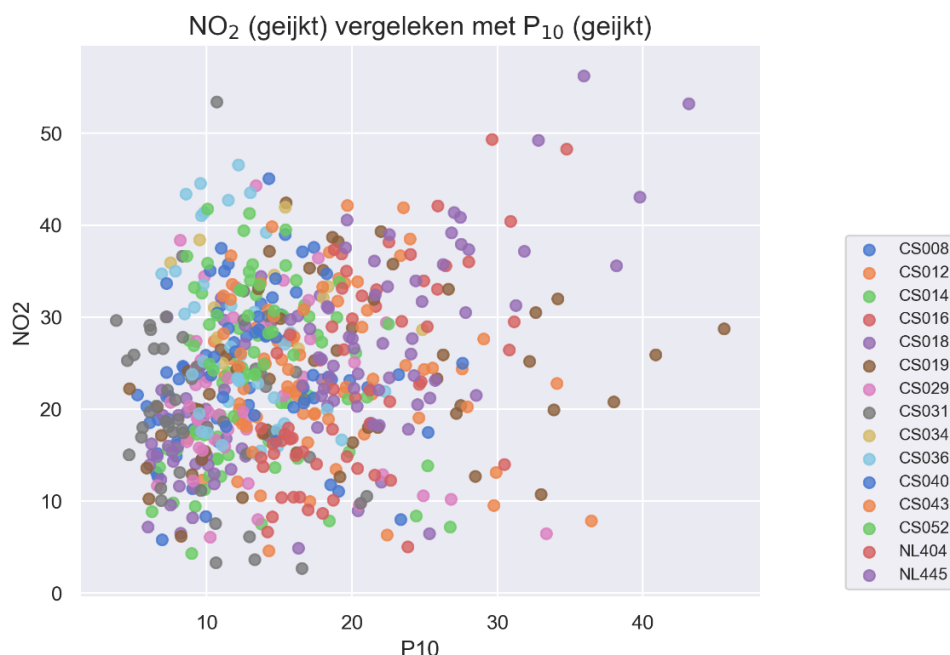
PUNTGRAFIEK kaart	Specificeert de gegevens voor plots waarbij een tweetal attributen tegen elkaar worden uitgezet in een scatterplot
Velden	
Kop	PUNTGRAFIEK : definieert de kaart
Parameter 1 & 2 – t/m 2N-1 & 2N	Lijst van paren van attributen die met elkaar moeten worden vergeleken Toegestane namen: alle mogelijke sensormetingen: NO2, NO2_ref1, NO2_ref2, NO2_kal P10, P10_ref1, P10_ref2, P10_kal P25, P25_ref1, P25_ref2, P25_kal

	temp, rh, ld Alle KNMI gegevens: Dwind, Vwind, Windstoot, TempKNMI, DauwTemp, LVocht, RegenDuur, Neerslag, Ldruk, Zicht, Instraling Twee afgeleide gegevens: Dwind10, Dwind9, waarbij Dwind10 de windrichting in graden is gedeeld door 10, en Dwind9 dezelfde windrichting gedeeld door 9. Alle gegevens die eerder berekend zijn
Parameter 2N+1 - eind	Lijst stations die in de scatterplot moeten worden meegenomen (gegeven door de station specificatie zoals op de KADER kaart wordt gegeven – bijv CSLV 9). ALL en ALL+ zijn toegestaan.
- Het aantal attributen dat wordt opgegeven moet even zijn. - Toegestane AANPASSING kaarten voor de VERGELIJK_ATTR2STAT kaart: ACHTERGROND, KLEUR, FORMAAT, ADRES, KLEUR, YLIMIET In tegenstelling tot normaal gebruik: de eerste waarde op de AANPASSING YLIMIET kaart geeft de maximum te plotten x waarde aan, de tweede is de maximum te plotten y waarde Standaard formaat: 8,7	
Naam plotfile: "attrib1"_"attr2"_"scatter_"teller"	

NO2 versus P10

Jaaroverzicht 2019
NO2 versus P10
File:NO2_P10_scatter_W_(IX)

Periode: 1/1/2019 - 31/12/2019
wk 1 - 1



Jaaroverzicht 2019
NO2 versus P10
File:NO2_P10_scatter_W_(IX)

Periode: 1/1/2019 - 31/12/2019
wk 1 - 1

Station/Meting	Gemiddeld	Maximum	Loc Max	Minimum	Loc Min	Mediaan	Aantal
CS008 P10	11.7	25.2	0113:00	5.2	0616:00	9.9	50
CS008 NO2	22.2	45.1	0217:00	5.8	0505:00	21.5	50
CS010 P10	18.6	24.2	0113:00	15.8	0106:00	17.2	4
CS012 P10	18.3	36.4	0317:00	8.7	0721:00	16.9	50
CS012 NO2	23.2	42.2	0217:00	4.6	0505:00	22.5	50
CS014 P10	13.5	26.7	0317:00	6.2	0421:00	12.2	53
CS014 NO2	19.3	34.0	0217:00	4.3	0505:00	17.7	53
CS016 P10	24.0	30.5	0113:00	19.3	0203:00	23.8	9
CS016 NO2	27.6	38.2	0217:00	5.0	0106:00	33.1	9
CS018 P10	11.4	25.3	0317:00	6.0	0421:00	10.1	53
CS018 NO2	18.2	34.4	0217:00	4.9	0106:00	16.3	53
CS019 P10	18.8	45.6	1222:00	4.7	0519:00	17.4	53
CS019 NO2	23.6	42.4	0217:00	6.2	0505:00	23.3	53

Voorbeeld: Scatterplot van NO₂ en PM₁₀ metingen in 2019 (weekgemiddelden). Kleuren geven de verschillende stations aan. Er is kennelijk weinig tot geen correlatie tussen de twee.

Functieafhankelijkheden

Definitie van de 1-D en 2-D functies

De gegevens die bij de verschillende stations worden gemeten zijn functies van tijd. De puntgrafieken geven aan dat er ook een relatie kan zijn tussen een bepaalde grootheid en een andere grootheid. Er kunnen ook relaties bestaan tussen meerdere grootheden.

Om de afhankelijkheid van een grootheid (bijvoorbeeld PM₁₀ metingen) ten opzichte van andere grootheden (bijvoorbeeld windsnelheid of tijd van de dag als “onafhankelijke” grootheden) duidelijk te maken gebruikt ALVA de kaarten “FUNCTIE1D” en de oude functiekaarten 1DFUNCTIE_VAN”,

“2DFUNCTIE_VAN”, “KNMI_FUNCTIE_ITEMS”, “FUNCTIE_ITEMS” en “SOORT_FUNCTIE_ITEMS”. De eerste twee van de vier kaarten van de oudere reeks definiëren de onafhankelijke parameters en daarmee het type analyse dat moet worden uitgevoerd; de laatste drie kaarten definiëren de parameters die afhangen van de onafhankelijke parameters en voor welke stations de analyses moeten worden uitgevoerd.

De functieafleiding veronderstelt dat de onafhankelijke parameter geen meetfout heeft, en de afhankelijke parameter wel. Als voorbeeld nemen we de uur van de dag van een meting als onafhankelijke parameter. Gedurende een maand wordt elke dag de fijnstof gemeten. We kunnen nu per uur van de dag alle fijnstofmetingen verzamelen en deze middelen. Als we dat voor elk uur van de dag doen, krijgen we een functie van de hoeveelheid fijnstof tegen het uur van de dag – een functie die we kunnen plotten. We verwachten een dag/nacht patroon, dat voor elk station iets anders kan zijn. Een weerkundig voorbeeld: we meten de luchtvochtigheid in de stations. De luchtvochtigheid vertoont een afhankelijkheid van de windsnelheid. We doen de metingen over een periode, verzamelen de luchtvochtigheidsmetingen die horen bij één bepaalde windsnelheid, middelen die, en plotten nu de gemiddelde luchtvochtigheidsmetingen tegen de windsnelheden. Wij middelen hier over de metingen van één station en kunnen dat plotten: luchtvochtigheid versus windsnelheid. Middeling over een ander station levert een ander resultaat op, dat ook op dezelfde plot kan worden getoond. De afwijkingen kunnen een indicatie zijn dat de luchtvochtigheidsmeting bij een station niet klopt (aangezien de luchtvochtigheidsmeters niet erg betrouwbaar zijn is dat zeer wel mogelijk).

De FUNCTIE1D kaart

Relevante kaart:

FUNCTIE1D

FUNCTIE1D kaart	Specificeert de gegevens voor plots waarbij gemiddelde attributen worden geplot als afhankelijk van een andere attribuut, waarvan verondersteld wordt dat hij geen meetfout heeft.
Velden	
Kop	FUNCTIE1D: definieert de kaart
Eerste velden (facultatief)	Tijdinfoblok (tijdbereik en sampling)
Één volgend veld	Naam van onafhankelijk attribuut (veronderstelling – geen meetfout)
Volgend veld	Scheidingsteken : “VOOR”
Volgende velden	Meetreeks blok - per type attribuut wordt één plot gemaakt
Volgend veld	Type plot: “SAMEN” - er wordt een plot gemaakt waarbij de gegevens van alle stations wordt gecombineerd – naast het gemiddelde worden de kwantilen 2,5%, 25%, 50%, 75% en 97,25% aangegeven. “APART” – er wordt een plot gemaakt waarbij het gemiddelde van elk station wordt gegeven voor de waarde van de onafhankelijke parameter. “CI” – niet operationeel. Functionaliteit is aanwezig in de oude versie
Volgende velden	Stationblok
	• Alle KADER/NIETGEBRUIKEN/BEREIK kaarten met stations en de KNMI_GEGEVENS kaart moeten al eerder zijn ingelezen. • Indien er geen OVERZICHT kaart wordt gegeven, worden geen plots gegenereerd. • Toegestane AANPASSING opties: FORMAAT/ADRES/YLIMIET. Standaard formaat = 8.5, 6. Ylimiet bepaalt de waarden aan de uiterste zijden van de kleurenbalk.
	Naam plotfile: “attribuut”_tijd_overz_”teller” of “attribuut”_tijd_overz_sampling_”teller”

De volgende plots zijn gegenereerd door de specificatie van de functie op de 1-DFUNCTIE_VAN kaart met de specificatie van het toepassingsbereik op de SOORT_FUNCTIE_ITEMS kaart.

Relevante kaart:

1DFUNCTIE_VAN

1DFUNCTIE_VAN kaart	Specificeert de gegevens voor plots waarbij gemiddelde attributen worden geplot als afhankelijk van een attribuut die vaak voorkomt. Op deze kaart wordt alleen deze laatste attribuut (de onafhankelijke parameter) gedefinieerd.
Velden	
Kop	1DFUNCTIE_VAN : definieert de kaart
Parameter 1-n	Definitie van de gevraagde plot(s): Ten eerste zijn alle attributen van tijd en weer toegestaan: Meer in het bijzonder kunnen de volgende acronyemen worden gebruikt: UpW, UpD, DpW, Windkracht, Windrichting, Zicht, LVocht, Instraling Deze acronyemen staan voor: UpW : De meetreeks wordt getoond als functie van het uur van de week. UpD : De meetreeks wordt getoond als functie van het uur van de dag. DpW : De meetreeks wordt getoond als functie van de Dag van de Week (0 = maandag, 6= zondag). Windkracht: De meetreeks wordt getoond als functie van windkracht. Windrichting: De meetreeks wordt getoond als functie van windrichting. Zicht: De meetreeks wordt getoond als functie van horizontaal zicht. LVocht: De meetreeks wordt getoond als functie van luchtvochtigheid, zoals gemeten door het KNMI. TempKNMI: de meetreeks wordt getoond als functie van de temperatuur zoals die in het KNMI station is gemeten.
• Meerdere 1DFUNCTIE_VAN kaarten zijn toegestaan. • De plots worden gemaakt voor de afhankelijke eigenschappen en de stations genoemd op de FUNCTIE_ITEMS, SOORT_FUNCTIE_ITEMS kaart en de KNMI_FUNCTIE_ITEMS kaart. • Als er geen FUNCTIE_ITEMS, SOORT_FUNCTIE_ITEMS kaart of de KNMI_FUNCTIE_ITEMS kaart aanwezig is worden geen plots geproduceerd. • Een functie van windrichting wordt in een polaire plot gepresenteerd.	

Luchtvochtigheid sensor gemeten bij meetstations
als functie van luchtvochtigheid

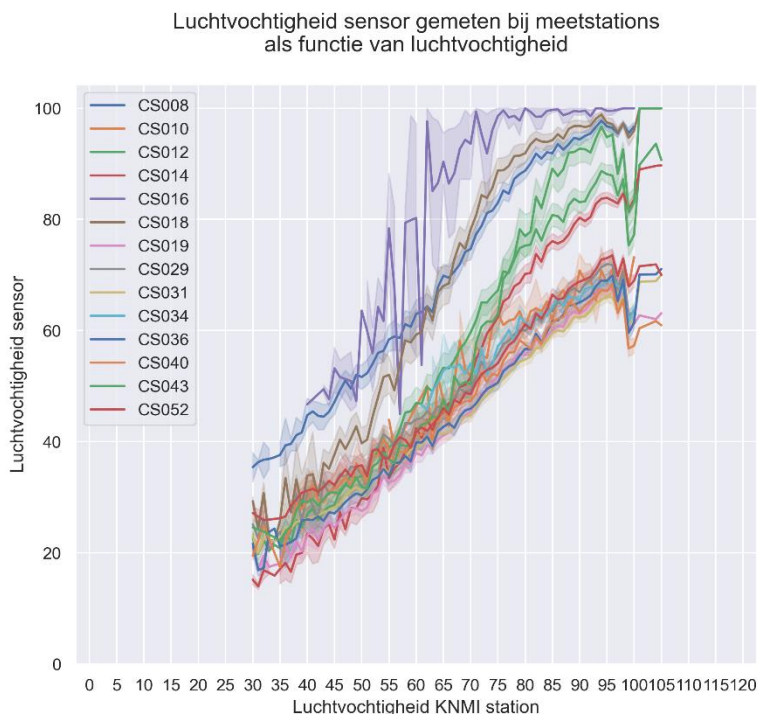


Station/Meting	Gemiddeld	Maximum	Loc Max	Minimum	Loc Min	Mediaan	Aantal
CS008 rh	86.6	100.0	0101:11	31.2	0825:16	92.2	7857
CS010 rh	65.0	81.7	0112:12	38.7	0120:16	66.5	453
CS012 rh	74.0	100.0	0316:09	17.0	0420:17	74.8	7705
CS014 rh	69.5	99.7	0113:06	12.6	0420:17	73.8	8623
CS016 rh	97.4	100.0	0101:00	43.5	0224:17	100.0	1370
CS018 rh	88.1	100.0	0101:00	19.9	0420:18	100.0	8623
CS019 rh	56.4	89.4	1209:15	15.0	0420:18	58.4	8553
CS029 rh	60.8	100.0	1226:14	17.9	0629:18	61.4	6374
CS031 rh	55.6	97.3	1112:08	11.9	1025:00	57.0	7335
CS034 rh	64.5	83.7	1208:10	33.5	1029:17	64.5	1448
CS036 rh	56.8	87.4	1206:16	15.7	0825:16	57.6	6410
CS040 rh	58.3	92.8	1206:12	17.4	0629:16	59.5	4402
CS043 rh	78.1	100.0	0526:22	22.2	0629:16	89.9	5264
CS052 rh	62.6	88.8	1127:07	24.3	0825:16	64.6	5244
KNMI215 LVocht	82.6	105.0	1124:06	30.0	0824:17	85.0	8760

Voorbeeld: metingen gedurende 2019. Luchtvochtigheidsmetingen bij de meetstations als functie van luchtvochtigheid bij het KNMI station. De trend van de metingen voor de verschillende stations is hetzelfde; het algehele niveau niet. Een betrouwbaarheidsprobleem.

Deze grafiek is het resultaat van een 1D_FUNCTIE kaart waarop LVocht is aangegeven als onafhankelijke parameter gecombineerd met een SOORT_FUNCTIE_ITEMS kaart met rh als afhankelijke parameter.

De functie leiden we af van een groot aantal metingen over een langere tijdperiode door te middelen. Met dit middelen verliezen we informatie over de variatie van de meetpunten. Daarom wordt ook een 95% betrouwbaarheidsinterval berekend tijdens het middelingsproces. Naast de plot van de afgeleide functies wordt ook een plot gemaakt waar het betrouwbaarheidsinterval als een band om de functie zichtbaar is. Bij deze plot wordt dezelfde statistische informatie in het onderschrift geleverd als bij de plot zonder standaard deviatie. Het onderschrift is hier niet getoond.



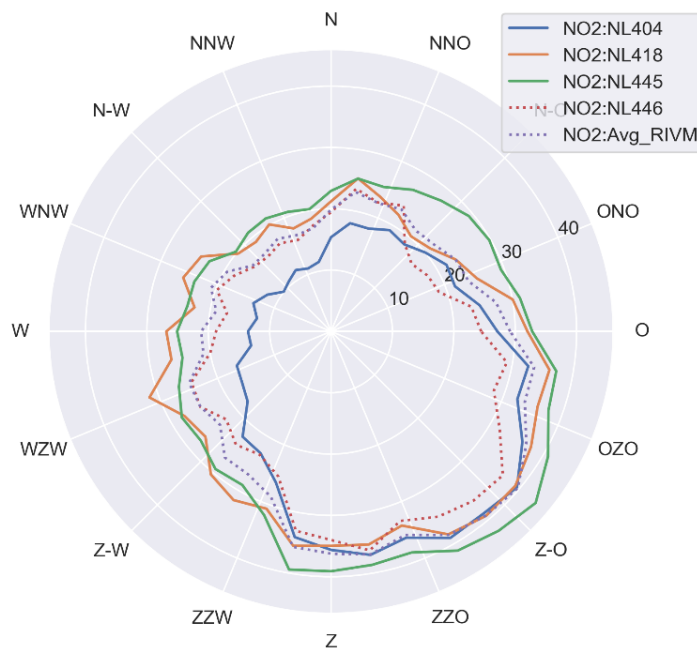
Voorbeeld: Zelfde grafiek als de vorige, maar nu met indicatie van de betrouwbaarheidsinterval. Luchtvochtigheidsmetingen bij de meetstations gedurende 2019 als functie van luchtvochtigheid gemeten bij het KNMI station. Met name bij station CS16 is de variatie in de metingen hoog, en is de band die het betrouwbaarheidsinterval aangeeft breed.

Sommige relaties zijn min of meer lineair: een toename van de onafhankelijke parameter leidt tot een toename van de afhankelijke parameter die alleen afhangt van de waarde van de toename van de onafhankelijke parameter : dat levert een rechte lijn op een plot op. De beste rechte lijn door de gemeten functie waarden kan worden afgeleid en deze lijn wordt afgebeeld als een aparte lijn, een trendlijn. De betrouwbaarheid van zo'n afleiding wordt gegeven door een determinatiecoëfficiënt, een waarde tussen 0 en 1. 1 betekent een betrouwbare relatie; 0 een onbetrouwbare. In ALVA wordt de rechte lijn afgeleid van de oorspronkelijke meetpunten (waar meetfouten een rol spelen), niet van de gemiddelde functiewaarden die op de plot zichtbaar zijn. De determinatiecoëfficiënt R^2 kan daarom heel wat lager zijn dan men op het eerste gezicht zou verwachten. Een relatief hoge waarde van R^2 kan ook het gevolg zijn van een te kleine hoeveelheid meegenomen meetpunten.

Omdat men niet altijd een lineaire relatie kan verwachten, worden intercept, richtingscoëfficiënt en determinatiecoëfficiënt alleen bij de statistische informatie van de plot toegevoegd als de gemiddelde determinatiecoëfficiënt van alle metingen groter is dan 0.1, of als er maar één relatie wordt geplot met een bijbehorende trendlijn. Indien dat niet het geval is, is de lijst van alle regressie waarden op dat ogenblik alleen beschikbaar op de aparte tekstfile die tegelijkertijd met de plot wordt gemaakt.

Plots gegenereerd op basis van de specificatie van de functie op de 1-DFUNCTIE_VAN kaart met "windrichting" als onafhankelijke parameter zien er anders uit. Het zijn polaire plots, met een annotatie in termen van Noord, Zuid, West en Oost. De waarde die in de radiale richting wordt afgelezen geeft de waarde voor de afhankelijke component.

NO₂ (geijkt) gemeten bij meetstations
als functie van windrichting 0=O;90=N;180=W;270=Z;360=O



Voorbeeld: Polaire plot van (gemiddelde) NO₂ vervuiling als functie van windrichting voor verschillende RIVM stations.

Het is aantrekkelijk windrichting te associëren met de locatie van de bron van de vervuiling ten opzichte van het meetstation. Windrichting en windsnelheid zijn echter vaak gekoppeld. Harde wind komt meestal uit het zuidwesten. De gemeten hoeveelheid vervuiling is ook afhankelijk van hoe de wind zich gedraagt, windsnelheid, turbulentie en andere meteorologische factoren. Het is daarom van belang om de vervuiling te bestuderen als functie van beide variabelen, windrichting en windsterkte.

Het is, zoals in het voorbeeld van windrichting en windsterkte, mogelijk dat een meetreeks afhangt van twee onafhankelijke parameters, die niet los van elkaar kunnen worden gezien. Om die relatie te onderzoeken, verzamelen we de waarden die bij één combinatie van de twee onafhankelijke parameters horen, en die middelen we. Om dat te doen tellen we alle metingen die bij zo'n combinatie horen bij elkaar op, en delen dat door het aantal metingen. Zowel de som van de metingen, als het aantal metingen en ook het gemiddelde zijn van belang. Het aantal metingen bij een combinatie van twee parameters geeft aan hoe vaak deze combinatie van de onafhankelijke parameters voorkomt, en het gemiddelde geeft de verwachtingswaarde aan voor de meting van de afhankelijke parameter als deze combinatie van onafhankelijke parameters voorkomt.

Relevante kaart:

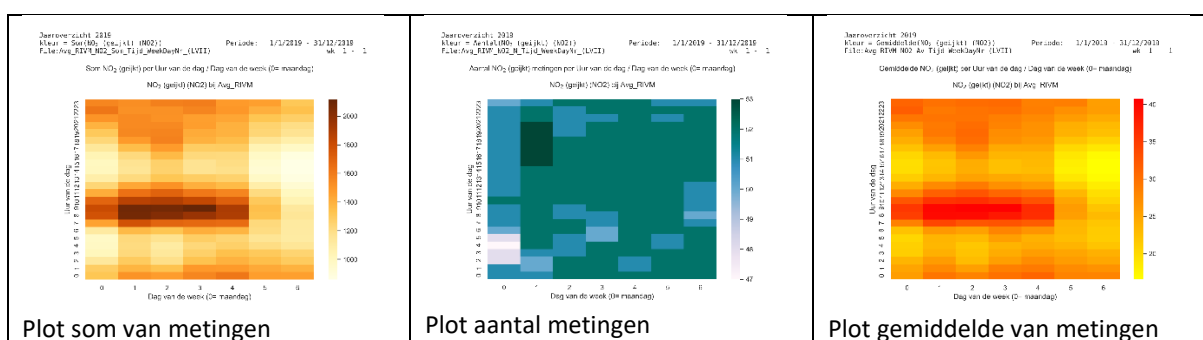
2DFUNCTIE_VAN

2DFUNCTIE_VAN kaart	Specificeert de gegevens van de onafhankelijke parameters voor plots waarbij gemiddelde attributen worden geplot als functie van twee attributen. De afhankelijke attribuut wordt getoond in kleur op een kader van de onafhankelijke attributen en wordt gedefinieerd op SOORT_FUNCTIE_ITEMS kaart of de KNMI_FUNCTIE_ITEMS kaarten.
Velden	
Kop	2DFUNCTIE_VAN : definieert de kaart
Parameter 1 Of Parameter 1&2	Definitie van de gevraagde plot: Of met de mnemonic: Windroos, Weekoverzicht, Periodeoverzicht - Windroos: de meetreeks wordt getoond als functie van windrichting en windkracht. - Weekoverzicht: de meetreeks wordt getoond als functie van dag van de week en uur van de dag.

	• Periodeoverzicht: de meetreeks wordt vertoond als functie van nummer van de week en uur van de dag Of door een combinatie van sensor metingen, KNMI metingen of tijdaanduidingen. Alle KNMI gegevens: Dwind, Vwind, Windstoot, TempKNMI, DauwTemp, LVocht, RegenDuur, Neerslag, Ldruk, Zicht, Instraling Twee afgeleide gegevens: Dwind10, Dwind9, waarbij Dwind10 de windrichting in graden is gedeeld door 10, en Dwind9 dezelfde windrichting gedeeld door 9. De tijdgegevens: • “WeekDayNr” (getal 0 t/m 6: 0 = maandag) • “Tijd” – het uur van de dag • “Weektijd”- het uur in de week, • “WeekNr” - het nummer van de week in het jaar • “Maand” - het nummer van de maand in het jaar, • “Jaardag” - het nummer van de dag van het jaar.
• Meerdere 2DFUNCTIE_VAN kaarten zijn toegestaan. • Indien een mnemonic wordt gebruikt, mag Parameter 2 niet zijn ingevuld. • De plots worden gemaakt voor de afhankelijke eigenschappen en de stations genoemd op de SOORT_FUNCTIE_ITEMS kaart en de KNMI_FUNCTIE_ITEMS kaart. • Als er geen SOORT_FUNCTIE_ITEMS kaart of de KNMI_FUNCTIE_ITEMS kaart aanwezig is worden geen plots geproduceerd.	

Als tweedimensionale plots worden gevraagd, worden altijd drie plots gemaakt. Ten eerste een plot van de som van alle meetpunten voor de combinaties van de twee onafhankelijke parameters, ten tweede een plot van het aantal meetpunten die bijdragen aan de som, en ten derde een plot van het gemiddelde van de meetpunten. Welk van de twee plots, som of gemiddelde, nuttig is, hangt af van de doel van het onderzoek. Het aantal meetpunten geeft een indicatie of de plots statistisch relevant zijn.

Om een functie van twee parameters te tonen wordt gebruik gemaakt van een heatmap, een kleurweergave op een rechthoekig bord, waarvan de assen gegeven worden door de onafhankelijke parameters. De gemiddelde, som of aantal metingen van de meetreeks is een derde dimensie, die in kleur wordt getoond op de achtergrond van de twee dimensies van de onafhankelijke parameters. Voor de drie soorten plots worden drie onderscheiden kleurschalen gebruikt.



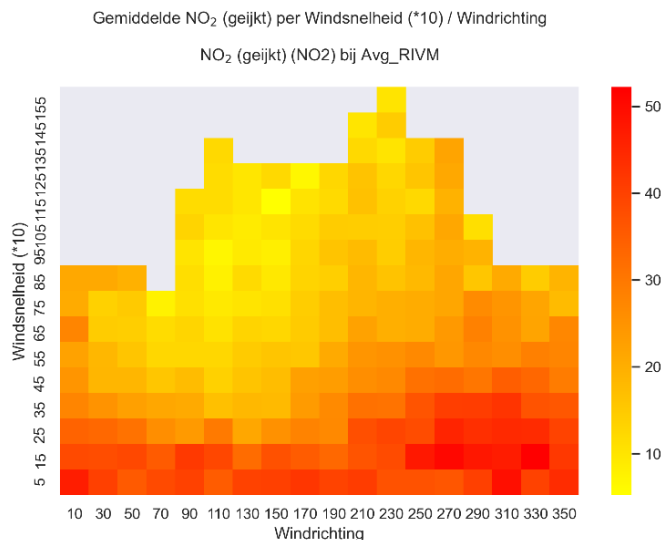
Voorbeeld: Overzicht vervuiling per dag van de week (horizontaal) en uur van de dag (verticaal). Links: som van de metingen; midden: het aantal metingen; rechts: het gemiddelde.

Zo kan men bijvoorbeeld een overzicht krijgen van de gemiddelde vervuiling per dag van de week en van de uur van de dag (voorbeeld hier boven).

Een ander voorbeeld - de vervuiling bij bepaalde windrichting en een bepaalde windsnelheid kan ernstig zijn, maar het kan zijn dat die windrichting niet vaak voorkomt. De som van de metingen voor die windrichting zal dan niet erg hoog zijn omdat het aantal metingen laag is. Omgekeerd: bij heersende windrichtingen zal de som

van de vervuiling hoog zijn, maar dat heeft niets te maken met de richting waar de belangrijkste vervuiler zich kan vinden. Voor dat laatste heeft men voor een redelijke schatting de gemiddelde van de metingen in een bepaalde windrichting (en per windsterkte) nodig.

Jaaroverzicht 2019
 kleur = Gemiddelde(NO₂ (geijkt) (NO₂)) Periode: 1/1/2019 - 31/12/2019
 File:Avg_RIVM_NO2_Av_Vwind_Dwind_(VIII) wk 1 - 1



Jaaroverzicht 2019
 kleur = Gemiddelde(NO₂ (geijkt) (NO₂)) Periode: 1/1/2019 - 31/12/2019
 File:Avg_RIVM_NO2_Av_Vwind_Dwind_(VIII) wk 1 - 1

Station/Meting	Gemiddeld	Maximum	Loc Max	Minimum	Loc Min	Mediaan	Aantal
vg RIVM NO ₂	26.2	107.0	0121:09	2.3	0930:03	21.8	8730
KNMI215 Vwind	44.1	150.0	0307:17	0.00	0120:22	40.0	8760
KNMI215 Dwind	186.6	350.0	0119:20	-10.0	0103:04	200.0	8760

Gemiddelde(NO ₂ (geijkt) (NO ₂))						
Maximum	52.2	bij	Dwind	330.0	Vwind	15.0
Zwaartepunt bij			Dwind	196.1	Vwind	46.4

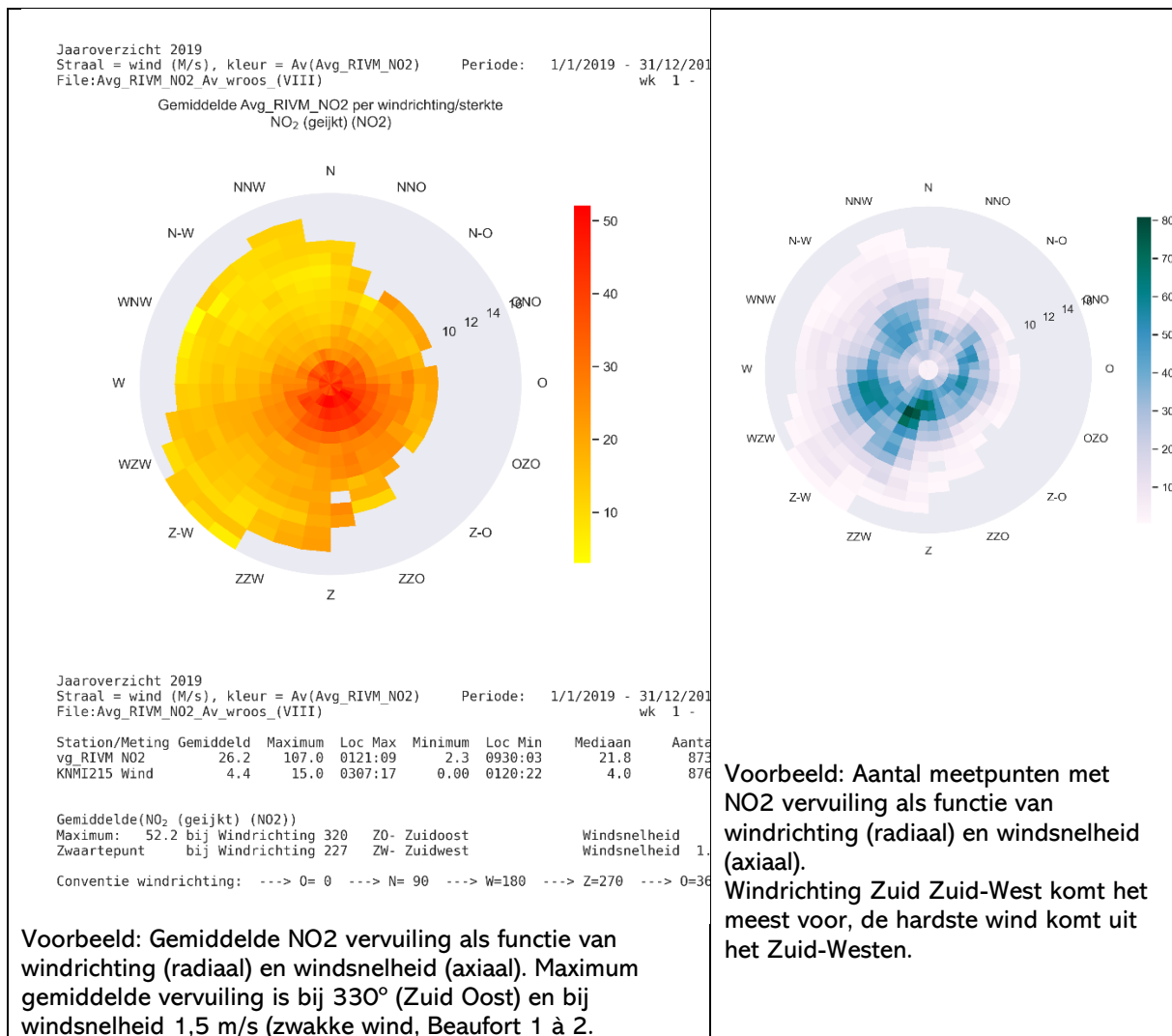
Parameters Kleinste kwadraten			
R ² :	0.32	Intercept:	31.53
		Helling:	Vwind -0.33
			Dwind 0.05

Voorbeeld: Gemiddelde NO₂ vervuiling als functie van windrichting (horizontaal) en windsnelheid (verticaal). Maximum gemiddelde vervuiling is bij 330° (Zuid Oost) en bij windsnelheid 1,5 m/s (zwakke wind, Beaufort 1 à 2).

Ook bij deze plots wordt een “beste” lineaire benadering berekend: een vlak dat het onregelmatige vlak van de meetpunten ongeveer beschrijft (Kleinste kwadraten oplossing).

Een speciale versie van een tweedimensionale plot is de windroos. In de plot hierboven waren de twee dimensies gespecificeerd als Vwind en Dwind op de 2DFUNCTIE_VAN kaart. Als in plaats van deze specificatie (Vwind en Dwind) de specificatie Windroos wordt gegeven, of de specificatie Wind en Dwind, zijn de onafhankelijke dimensies wederom de windrichting en de windsnelheid, maar de plot wordt anders gepresenteerd. Windrichting wordt gegeven door de richting vanuit het nulpunt in een cirkelvormig plot, de windsnelheid wordt gegeven door de radiale afstand ten opzichte van het nulpunt. De vervuiling wordt in een kleur als de derde dimensie getoond. De windroos hieronder bevat dezelfde informatie als de 2D functie plot hierboven.





De rechterplot geeft het aantal meetpunten per windrichting en windsnelheid (de frequentie van voorkomen van deze windcondities) aan: deze is verre van uniform. Ook is de verdeling over de windsnelheden gegeven een bepaalde windrichting niet uniform.

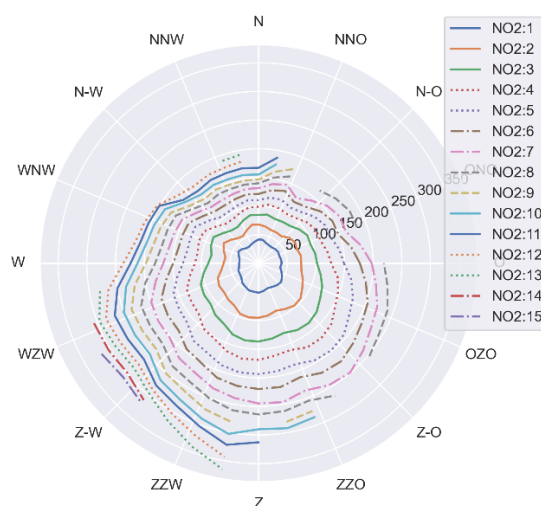
De drie windrozen geven andersoortige informatie:

- Het aantal meetpunten (Aantal, N) per windrichting en windsnelheid – de hoeveelheid keer dat deze specifieke combinatie van windrichting en windsnelheid voorkomt in de periode van meten.
- De som van de vervuiling (Som) per windrichting en windsnelheid – het totaal van alle metingen van de vervuiling bij deze combinatie van windrichting en windsnelheid. Als in een periode de wind alleen uit het westen komt, en de bron van vervuiling ligt in het oosten, zal de som in de windroosdiagram lage waarden hebben (en bij het meetstation hebben we dan weinig vervuiling gemeten). Als de bron van vervuiling in het westen ligt daarentegen, zal de som hoge waarden hebben.
- Het gemiddelde van de vervuiling (Gemiddelde, Av) per windrichting en windsnelheid. Als de hoeveelheid metingen groot genoeg is, geeft het gemiddelde de verwachtingswaarde aan van de vervuiling van die combinatie van windrichting en windsnelheid: gegeven die combinatie van parameters kun je verwachten dat de vervuiling die gemiddelde waarde heeft.

De vraag “waar komt de vervuiling vandaan?” is in feite alleen oplosbaar als men weet hoe de vervuiling van een puntbron wordt gedistribueerd. Als men aanneemt dat de vervuiling zich alleen voortplant in een bepaalde windrichting en dat de gemeten vervuiling dan alleen afhankelijk is van de afstand tussen puntbron en meetpunt (en dus op een zelfde manier afhankelijk is van windsnelheid), dan is de gemeten vervuiling bij een

meetpunt ook afhankelijk van de (omgekeerde) windrichting en van dezelfde windsnelheid. Aannemend dat de vervuilende puntbronnen bij dezelfde soort weercondities (bijv. stormachtig slecht weer) op dezelfde manier hun vervuiling verspreiden, de één in de éne richting, de andere in de andere richting, dan kan de emissie van elke vervuiler worden bepaald door de totale vervuiling die van die puntbron komt, gedeeld door het aantal keren dat de vervuiling bij het meetpunt komt – dus het gemiddelde van de vervuiling bij een gegeven windsnelheid en windrichting. Dit geldt voor alle vervuilers bij die weersomstandigheden. Voor vervuilers op een andere afstand en bij andere weerscondities zal de verdeling van de vervuiling als functie van de windsnelheid anders liggen, maar ook daar het relatieve belang van de emissie in de windrichting gedefinieerd door het gemiddelde bij een gegeven windsnelheid en richting.

Onder de aanname dat de manier van distribueren van vervuiling onafhankelijk is van windrichting, wordt de vergelijking van de gemiddelden per windrichting zinvol. Op basis van deze redenering is er een cumulatieve windroosplot van gemiddelden gedefinieerd, waarbij de som van de gemiddelde waarden van de vervuiling in een bepaalde windrichting worden geplot. In de plot hieronder is curve NO2:1 de curve die de gemiddelde waarde van de NO2 vervuiling aangeeft voor windsnelheden tussen 0 en 2 m/s (de twee bereiken 0-1 m/s en 1-2 m/s zijn samengevoegd), de curve NO2:2 geeft de som aan van de gemiddelde waarde voor windsnelheden tussen 2 en 3 m/s opgeteld bij de waarden van curve NO2:1, de curve NO2:3 geeft de som aan van de gemiddelde waarde voor windsnelheden tussen 3 en 4 m/s opgeteld bij de waarden van curve NO2:2, en zo verder. De windverdeling is zo dat bij een bepaalde windrichting en windsnelheid geen metingen beschikbaar zijn – dat geeft de gaten in de curves NO2:8 en hoger, en die curves kunnen dan, bij gebrek aan statistiek, worden genegeerd. De richting waaruit de meest forse vervuiling komt is bij zo'n 310 graden: Zuid Oosten wind.



Jaaroverzicht 2019
 Straal = NO2(Avg_RIVM_NO2)
 File:Avg_RIVM_NO2_wrcum0_Av_(II)

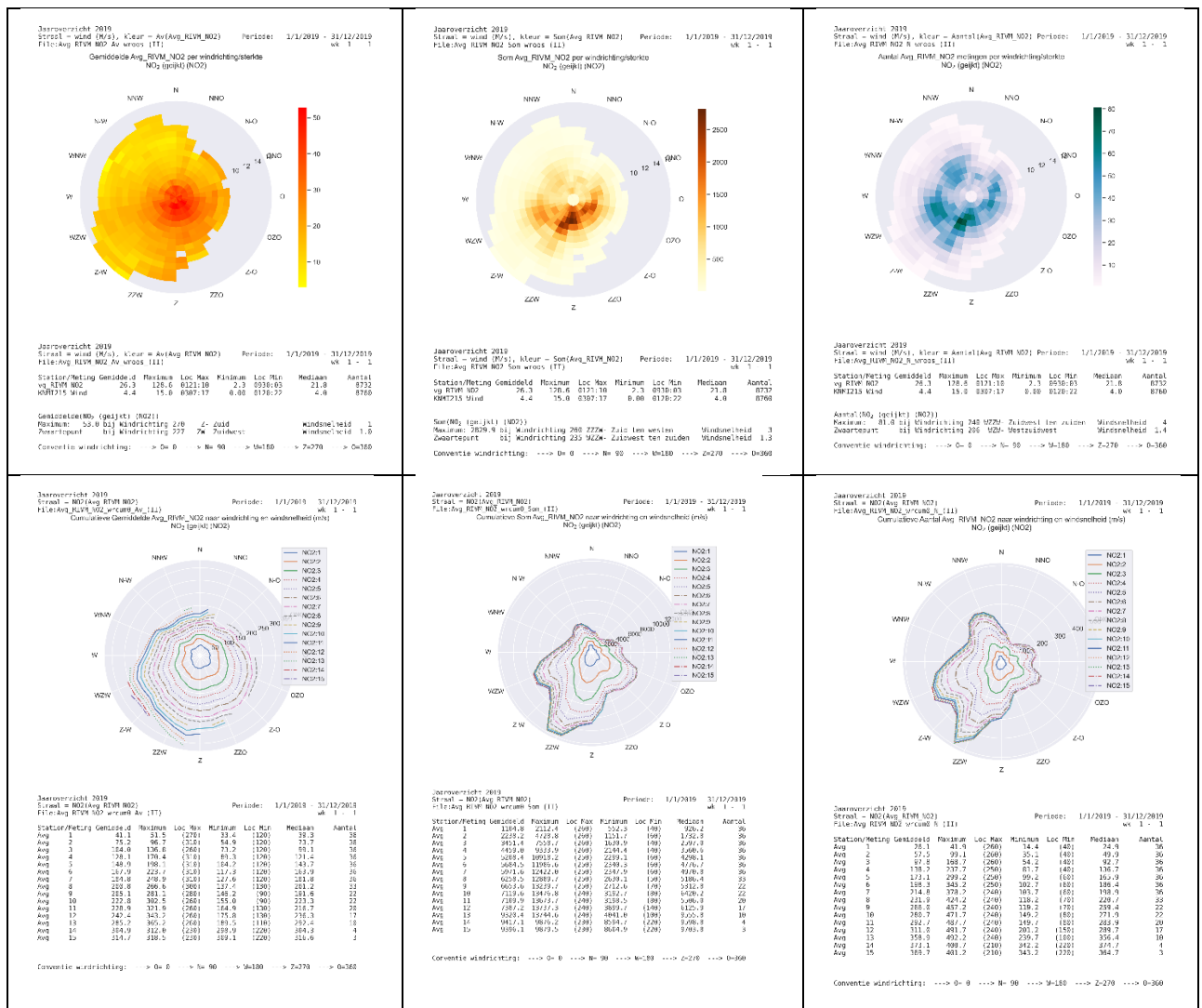
Periode: 1/1/2019 - 31/12/2019
 wk 1 - 1

Station/Meting	Gemiddeld	Maximum	Loc Max	Minimum	Loc Min	Mediaan	Aantal
Avg 1	41.1	51.5	(270)	33.4	(120)	39.3	36
Avg 2	75.2	96.7	(310)	54.9	(120)	73.7	36
Avg 3	104.0	136.8	(260)	73.2	(120)	99.1	36
Avg 4	128.1	170.4	(310)	89.3	(120)	121.4	36
Avg 5	148.9	198.1	(310)	104.2	(120)	143.7	36
Avg 6	167.9	223.7	(310)	117.3	(120)	163.9	36
Avg 7	184.8	248.9	(310)	127.6	(120)	181.8	36

Voorbeeld: Gemiddelde NO2 vervuiling per windrichting (radiaal) en intervallen van windsnelheid (axiaal), gesommeerd over de intervallen van windsnelheid (cumulatieve windroos).

De informatie van een cumulatieve windroos en de veelkleurige windrozen die eerder zijn beschreven is hetzelfde. De gemiddelde windrozen worden beide berekend door de totale vervuiling per windsnelheid/windrichting te delen door de frequentie van voorkomen.

Gemiddelde =	Som metingen /	Aantal metingen
--------------	----------------	-----------------



Men kan ook redeneren dat het gemiddelde van de vervuiling per windrichting een aanduiding geeft voor de richting waarin de vervuulende bron te vinden is (de plot op pagina 44). In die benadering weegt de bijdrage van de meest voorkomende windsnelheid in een bepaalde windrichting het zwaarst. In de benadering hierboven weegt elke windsnelheid in die windrichting even zwaar.

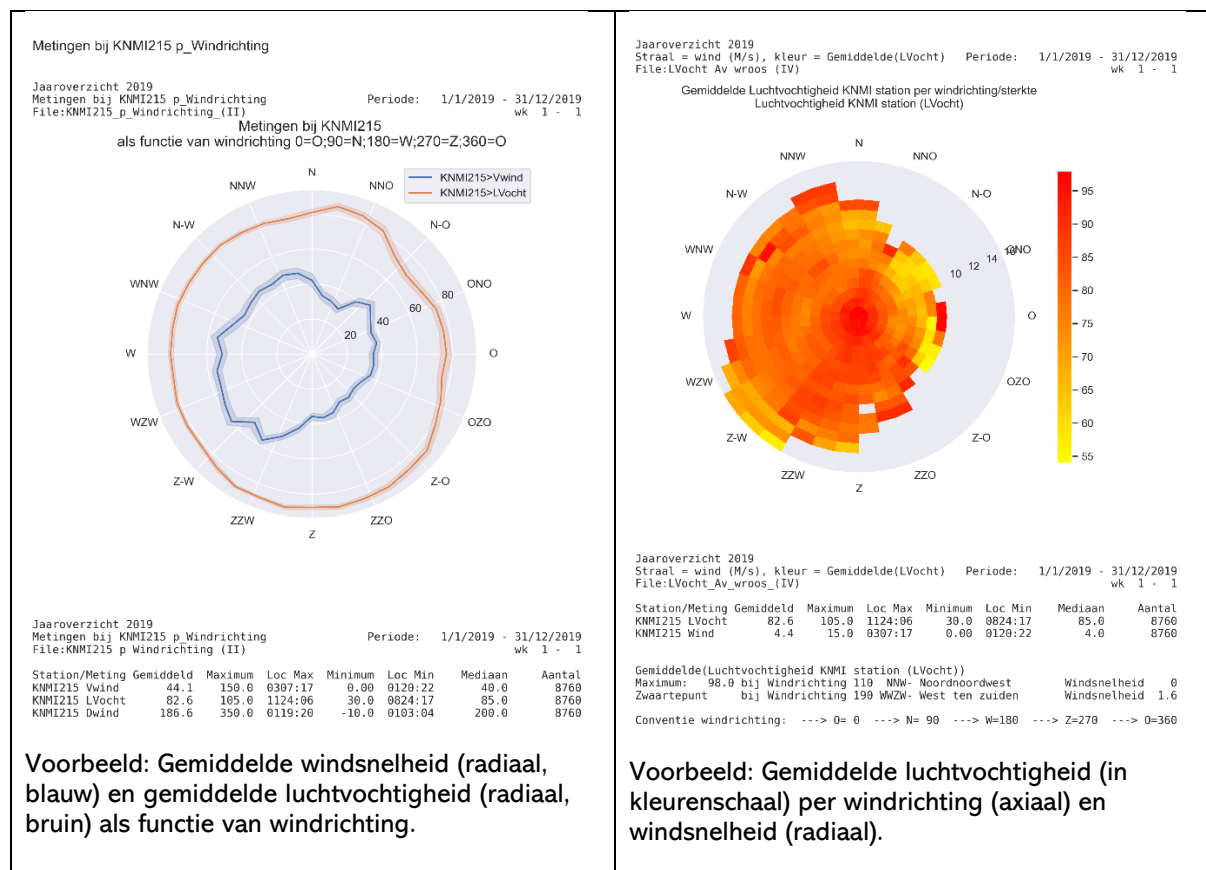
Bij beide methodes blijft het probleem spelen dat de vervuiling niet alleen afhankelijk is van windrichting en windsnelheid. De dikte van de menglaag (de atmosferische laag waarin de vervuiling zich verspreidt) is ook een belangrijke factor, en die hangt weer af van dag en nacht patronen, jaargetijden, enz....

Definitie van het toepassingsbereik: de afhankelijke parameters

Om de afhankelijkheid van een grootheid ten opzichte van andere grootheden duidelijk te maken gebruikt ALVA de kaarten “1DFUNCTIE_VAN”, “2DFUNCTIE_VAN”, “KNMI_FUNCTIE_ITEMS”, “FUNCTIE_ITEMS” en “SOORT_FUNCTIE_ITEMS”. De eerste twee van deze vijf kaarten definiëren de onafhankelijke parameters en daarmee het type analyse dat moet worden uitgevoerd en zijn hierboven beschreven; de laatste drie kaarten definiëren de parameters die afhangen van de onafhankelijke parameters en voor welke stations de analyses moeten worden uitgevoerd.

De afhankelijke parameters zullen vaak stations-afhankelijke parameters zoals NO₂, P₂₅ of P₁₀ zijn. Deze worden gedefinieerd op de FUNCTIE_ITEMS of de SOORT_FUNCTIE_ITEMS kaart. Ook KNMI gegevens kunnen als “afhankelijke” parameters worden gespecificeerd. Dit gebeurt op de KNM_FUNCTIE_ITEMS kaart.

Als voorbeeld hieronder: het zal niemand verbazen dat de luchtvochtigheid in Nederland afhangt van de windrichting. Dat kan worden getoond in een 1D functie plot en in een 2D-functie plot (windroos).



De KNMI_FUNCTIE_ITEMS kaart geeft aan welke gegevens die bij het KNMI station zijn gemeten moeten worden geplotted als functies van de onafhankelijke variabelen die in de 1DFUNCTIE_VAN en de 2DFUNCTIE_VAN kaarten zijn gespecificeerd.

Relevante kaart:

KNMI_FUNCTIE_ITEMS

KNMI_FUNCTIE_ITEMS kaart	Geeft aan welke gegevens die van het KNMI meetstation zijn verkregen worden geplotted als afhankelijke parameter in de FUNCTIE en 2DFUNCTIE plots
Velden	
Kop	KNMI_FUNCTIE_ITEMS : definieert de kaart
Parameters	Lijst van gemeten grootheden. Toegestaan: alle KNMI gegevens: Dwind, Vwind, Windstoot, TempKNMI, DauwTemp, LVocht, RegenDuur, Neerslag, Ldruk, Zicht, Instraling Twee afgeleide gegevens: Dwind10, Dwind9, waarbij Dwind10 de windrichting in graden is gedeeld door 10, en Dwind9 dezelfde windrichting gedeeld door 9.
- Als onafhankelijke parameters worden de gegevens gebruikt gedefinieerd op 1DFUNCTIE_VAN en 2DFUNCTIE_VAN kaarten. - Bij 2D functies: voor elk van de op deze kaart genoemde grootheden worden drie plots gemaakt. Indien de 2D functie een “windroos” is, komt hier nog een cumulatieve plot bij.	

- Bij 1D functies: er worden twee plots gemaakt waarbij de eerste de gemiddelde van de gemeten grootte als functie van de onafhankelijke parameter weergeeft, en de tweede naast het gemiddelde ook het betrouwbaarheidsinterval vertoont.
- Bij 1D functies wordt voor elke functie een rechte lijn berekend die de functie zou kunnen beschrijven (gesteld dat de afhankelijke parameter een lineaire relatie heeft met de onafhankelijke parameter). De intercept en richtingscoëfficiënt worden bij de statistische gegevens vermeld. De determinatiecoëfficiënt R^2 wordt ook vermeld (0 = slecht, 1 = goed).
Bij een axiale 1D functie plot wordt geen trendlijn berekend.
- Als er maar één functie wordt weergegeven in een 1D functie plot, wordt de berekende rechte lijn ook getoond (aangeduid als Trend_1).
- Bij 2D functies wordt voor elke plot een lineaire functie berekend die de afhankelijke parameter het "best" zou representeren. Ook hier worden de parameters vermeld onder de plot, inclusief een determinatiecoëfficiënt.

Relevante kaart:

FUNCTIE_ITEMS

FUNCTIE_ITEMS kaart	Specificeert de afhankelijke parameters en de stations waarvoor plots moeten worden gemaakt als functie van onafhankelijke parameters als gedefinieerd op de 1DFUNCTIE_VAN kaarten. Er wordt één plot gemaakt per station met meerdere afhankelijke parameters.
Velden	
Kop	FUNCTIE_ITEMS : definieert de kaart
Parameter 1 - n	De afhankelijke parameters (de attributen, bijv. NO2) waarvoor plots moeten worden gemaakt. Toegestaan: alle attributen zoals gemeten bij de stations of zoals berekend bij deze stations of bij de pseudostations. Ook toegestaan: alle KNMI parameters.
Parameter n+1 - eind	Lijst van stations (gegeven door de station specificatie zoals op de KADER kaart wordt gegeven – bijv CSLV 9) of de tekst "ALL" / "ALL+"
• Als onafhankelijke parameters worden de gegevens gebruikt gedefinieerd op de gebruikte 1DFUNCTIE_VAN kaarten. De combinatie van een 2DFUNCTIE_VAN kaart en deze FUNCTIE_ITEMS kaart levert geen 2D functieplot op. Om een 2D functie plot voor een attribuut als NO2 te plotten moet men de SOORT_FUNCTIE_ITEMS kaart gebruiken. • Voor elk gespecificeerd station wordt een plot gemaakt. Ook de naam van pseudostations kan worden aangegeven. • Bij 1D functies: voor elk van de op deze kaart genoemd station worden twee plots gemaakt. De plots bevatten de gegevens van alle op deze kaart genoemde attributen. • Bij 1D functies wordt voor elke functie per station een rechte lijn berekend die de functie zou kunnen beschrijven (gesteld dat de afhankelijke parameter een lineaire relatie heeft met de onafhankelijke parameter). De intercept en richtingscoëfficiënt worden bij de statistische gegevens vermeld. De determinatiecoëfficiënt R^2 wordt ook vermeld (0 = slecht, 1 = goed). • Als er maar één functie voor één station wordt weergegeven in een 1D functie plot, wordt de berekende rechte lijn ook getoond (aangeduid als Trend).	
1D plotfile: "station"_p_"onafh attrib"_"teller" ; en een tweede plot met toevoeging stdev	

De SOORT_FUNCTIE_ITEMS kaart definieert 1-D plots waarbij per plot één afhankelijke parameter / attribuut wordt geplot voor mogelijk meerdere stations. Deze kaart wordt ook gebruikt voor het definiëren van 2-D plots waarbij per plot één afhankelijke parameter / attribuut wordt geplot voor elke gedefinieerde functie. Voor deze 2D plots geldt: voor elk station dat op deze kaart wordt genoemd wordt een nieuwe set plots gemaakt (één set houdt in: één plot voor gemiddelden, één plot voor aantal metingen, en één plot voor de som van de metingen; bij windroosplots worden dan nog 3 cumulatieve plots toegevoegd). Het aantal plots kan dus erg groot worden indien men niet voorzichtig is.

Relevante kaart:

SOORT_FUNCTIE_ITEMS

SOORT_FUNCTIE_ITEMS kaart	Specificeert de afhankelijke parameters en de stations waarvoor plots moeten worden gemaakt als functie van onafhankelijke parameters als gedefinieerd op de 1DFUNCTIE_VAN en 2DFUNCTIE_VAN kaarten.
Velden	
Kop	SOORT_FUNCTIE_ITEMS : definieert de kaart
Parameter 1 - n	De attributen waarvoor plots moeten worden gemaakt
Parameter n+1 - eind	Lijst van stations (gegeven door de station specificatie zoals op de KADER kaart wordt gegeven – bijv CSLV 9) of de tekst “ALL” / “ALL+”
- De functies gedefinieerd op 1DFUNCTIE_VAN en 2DFUNCTIE_VAN kaarten worden toegepast op de afhankelijke attributen als gemeten (en berekend) voor de stations genoemd op deze kaart. Ook de naam van pseudostations kan worden aangegeven. - Bij 2D functies: voor elk van de op deze kaart genoemde attributen worden drie plots gemaakt. Voor elk station op deze kaart wordt per attribuut een onafhankelijk set van drie plots gemaakt. - Bij 1D functies: voor elk van de op deze kaart genoemde attributen worden twee plots gemaakt. De plots bevatten de gegevens van alle op deze kaart genoemde stations. - Bij 1D functies wordt voor elke functie per station een rechte lijn berekend die de functie zou kunnen beschrijven (gesteld dat de afhankelijke parameter een lineaire relatie heeft met de onafhankelijke parameter). De intercept en richtingscoëfficiënt worden bij de statistische gegevens vermeld. De determinatiecoëfficiënt R^2 wordt ook vermeld (0 = slecht, 1 = goed). - Als er maar één functie voor één station wordt weergegeven in een 1D functie plot, wordt de berekende rechte lijn ook getoond (aangeduid als Trend). - Bij 2D functies wordt voor elke plot een lineaire functie berekend die de afhankelijke parameter het “best” zou representeren. Ook hier worden de parameters vermeld onder de plot, inclusief de determinatiecoëfficiënt.	
1D plotfile:	“afh attrib”_p_“onafh attrib”_“teller” ; en een tweede plot met toevoeging stdev
2D	“station”_“afh_attrib”_“typeplot”_“onafh attrib1”_“onafh attrib2”_“teller”, waarbij typeplot gelijk is aan Av, Som of N (gemiddelde, som, aantal meegenomen punten).
2D	“station”_“afh_attrib”_“typeplot”_windroos_“teller”
1D plotfile:	“afh attrib”_p_“onafh attrib”_“teller” ; en een tweede plot met toevoeging stdev
2D	“afh_attrib”_“typeplot”_“onafh attrib1”_“onafh attrib2”_“teller”, waarbij typeplot gelijk is aan Av, Som of N (gemiddelde {average}, som, aantal meegenomen punten).
2D	“afh_attrib”_“typeplot”_windroos_“teller”

Dankbetuigingen

Opmerkingen van Diederik van Hemert, Karen de Mos en Henk Thomas waren zeer nuttig om de functionaliteit van dit programma vorm te geven. Diederik van Hemert heeft de eerste versie van getRIVM geschreven – een essentiële stap om de analyse mogelijk te maken.

Ook dank aan de Citizen Science groepen in Gouda, Lansingerland en Zuidplas. Omdat zij ALVA versie 1 - in feite een uitgebouwd prototype - gingen gebruiken, won het herschrijven van het programma aan prioriteit.

Opmerkingen en Commentaar

Commentaar op de software en op deze beschrijving graag naar:

frans.kets@kpnmail.nl (met kopie aan lv2@kpnmail.nl)

Appendix A. Opzetten Python omgeving voor ALVA

STAP 1: DOWNLOAD ANACONDA

Selecteer: <https://www.anaconda.com/distribution/>

En download de laatste Python 3.8 versie

Anaconda is een environment manager, waarin verschillende versies van python en bijbehorende pakketten naast elkaar kunnen draaien. Het basis pakket van anaconda levert zo'n 300 verschillende pakketten. Omdat alle pakketten moeten kunnen samenwerken, is het een wat minder recente versie van python (bijv: python 3.7.3 in plaats van 3.7.x of 3.8). ALVA gebruikt een beperkt aantal pakketten, en kan daarom ook in een andere environment met een hogere versie van python werken.

Anaconda bestaat uit conda, een command gedreven interface, en een interactief front-end, Anaconda Navigator. Het opzetten van environments is het gemakkelijkst (en betrouwbaarst) in conda.

Om verschillende versies efficiënt te kunnen benaderen, is het handig om verschillende command windows te definiëren. Dit kan het beste in Anaconda Navigator.

STAP 2: PYTHON ENVIRONMENT VOOR ALVA

Om ALVA te kunnen gebruiken zijn drie dingen nodig:

- Met behulp van conda: opzetten special environment met python 3.8
- Met behulp van conda: in dat environment laden: de verschillende modules die in ALVA worden gebruikt
- Creëer een directory om de python programma's te kunnen runnen

Stapsgewijs:

1. Open vanuit het Windows Start menu : Anaconda prompt.
2. Een window wordt vertoond. U bent in de base environment. Gebruik dit window voor de commando's:
 - a. Check dat conda beschikbaar is:
`conda --version`
 - b. Update conda naar de nieuwste versie:
`conda update conda`
 - c. als nieuwe versie beschikbaar, type y om te installeren bij prompt Proceed ([y]/n)?
 - d. maak een nieuwe environment aan (willekeurige naam: hier :ALVA)
`conda create --name ALVA python=3.8`
 - e. activeer deze environment:
`conda activate ALVA`
 - f. verifieer:
`conda info --envs`
`python --version`
 - g. kijk of package requests aanwezig is en installeer. Daarna hetzelfde voor pandas en andere modules.
`conda search requests`
`conda search pandas`
`conda install seaborn matplotlib pandas numpy scikit-learn requests`
 - h. verifieer:
`conda list`
 - i. om naar de base environment terug te gaan:
`conda deactivate`

- j. afsluiten van de Anaconda browser:
`exit`
- 3. Plaats de files van ALVA in een directory, bijv:
/Documenten/Calc/Analyse
- 4. De ALVA software is nu benaderbaar via de Anaconda prompt (in Start menu) met de volgende stappen:
 - a. Activeren van de environment:
`conda activate ALVA`
 - b. Navigeren naar de directory waar de software aanwezig is (Documents\Calc\Analyse):
`cd Documents`
`cd Calc`
`cd Analyse`
 - c. Met de input parameterfile bijvoorbeeld gegeven door parameterfile_blabla.txt kan men het programma runnen met
`python ALVA22MAIN.py` (een file ALVAINPUTFILE.txt is dan nodig) of met
`python ALVA22MAIN.py parameterfile_blabla.txt`
of, indien er meer aanroepen zijn, met commando's in een batfile.bat file
 - d. Aan het eind afsluiten van de environment: `conda activate`

3: ALVA SPECIFIEKE SETUP GEBRUIKEN

- 5. Om verschillende environments gemakkelijk te kunnen benaderen, kan met Anaconda Navigator speciale versies van applicaties worden gemaakt.
 - a. Start Anaconda Navigator van Start menu
 - b. Onder de linker kolom, selecteer Home. Je krijgt een overzicht welke standaard python programma's aanwezig zijn.
 - c. Selecteer bovenin in de box tussen "Applications in" "Channels" de environment ALVA. Installeer daarna: Cmd.exe prompt en Spyder
 - d. In de Windows Start menu zijn deze applicaties nu los van de base environment applicaties zichtbaar. Ze kunnen worden vastgemaakt aan het Start menu.
- 6. Als Spyder is geïnstalleerd kan ALVA worden gerund vanuit Spyder.
 - a. Laad ALVA22MAIN.py , ALVA22INPUTMOD.py, StationDirectory22.py
 - b. Run ALVA22MAIN
 - i. De inputfile wordt gespecificeerd in ALVAINPUTFILE.txt (zie beschrijving ALVA).
- 7. De base environment wordt in deze opzet niet gebruikt. De python versie van deze environment kan achter lopen.
- 8. Een nieuwe versie van de ALVA environment kan worden verkregen met
`conda update -n ALVA --all`

Appendix B. Voorbeelden van carddecks

VERKENNING VAN DE DATASET

CARDDECK

```
#
Datadirectory 2020CumDatafile
Periode 1/12/2018 31/1/2020 Jaaroverzicht 2019
KNMI_gegevens
Kader CSLV RIVM
Nietgebruiken RIVM 4 CSLV 9
Periode4stat CSLV 12 5/8/2018
Periode4stat CSLV 10 5/8/2018 24/01/2019
Periode4stat CSLV 7 1/1/2020
Periode4stat CSLV 34 20/7/2019
Periode4stat CSLV 36 1/4/2019
Periode4stat CSLV 40 1/4/2019
Periode4stat CSLV 43 1/4/2019
Periode4stat RIVM 0 1/1/2018 31/12/2019
#
# data processing - zet vooralsnog ruime grenzen
#
Grenswaarden          NO2 0 200
Grenswaarden          P10_m 0 200
#
# bereken gemiddelde waardes voor alle parameters
#
Samenvoegen Gemiddelde Avg_ALL ALL
Samenvoegen Gemiddelde Avg_RIVM RIVM
Samenvoegen Gemiddelde Avg_CSLV CSLV
#
# bereken het verschil tussen de RIVM en de CSLV stations
#
Bereken_stat D_RIVM_CS Avg_RIVM Avg_CSLV -
Bereken_stat Nullijn 0
#
# welke bemonstering willen wij zien?
#
Plotparam SAVEONLY
Bemonstering GEEN1HLOT
Bemonstering 3H
Bemonstering W
#
# data display - tijdsbereik kleiner dan die op PERIODE kaart
#
Analyseperiode 1/1/2019 31/12/2019
#
Overzicht NO2 P10_m ALL
Doosdiagram NO2 ALL
Doosdiagram P10_m ALL
Kruiscorrelatie Pearson NO2 ALL
Kruiscorrelatie Pearson P10_m ALL
Plot_soort NO2 P10_m _PS D_RIVM_CS Avg_RIVM Avg_CSLV _PS Nullijn
```

VERLOOP VAN HET PROGRAMMA

Twee en twintig plots worden gegenereerd. Het verloop van het programma (na het lezen van files) is als volgt:

```
*****
*
*
*
*
*
*****
Verwerken      : Periode4stat
Verwerken      : Periode4stat
```


81

VOORBEELD 2: VERSCHUIVINGEN IN TIJD

In dit voorbeeld vergelijken we resultaten van metingen gedaan in 2018, 2019 en 2020. De resultaten zijn te zien in het hoofdstuk over Combineren van Tijdreeksen.

```
Debug 0 1 2 3 4 5 6 7 8 9
#
#   bepaal de dataverzameling
#
Datadirectory 2020CumDatafile
#
#       We willen 2020 vergelijken met gegevens van 2017 t/m 2019
#       In deze voorbeeld run gebruiken we geen KNMI gegevens - wordt toch ingelezen
#
Periode 1/1/2017 10/7/2020 Januari - Juni 2020
KNMI_gegevens
Kader RIVM
Nietgebruiken RIVM 3 4
#
#   data processing
#
Grenswaarden      N02 0 150
Grenswaarden      P10 0 150
Uitschieter       9    P10 4
Daguitschieter    3    P10 3
#
#   bereken het gemiddelde van de RIVM stations
#
Samenvoegen Gemiddelde Stat2020   RIVM
#
#       Bereken de tijdverschuivingen
#
T_shift           Stat2019 365 Stat2020
T_shift           Stat2018 730 Stat2020
T_shift           Stat2017 1095 Stat2020
#
#       Bereken de verschillen en maak een Nullijn voor display
#
Bereken_stat D_20_19 Stat2020 Stat2019 -
Bereken_stat D_20_18 Stat2020 Stat2018 -
Bereken_stat D_20_17 Stat2020 Stat2017 -
Bereken_stat Nullijn 0
#
#       Bepaal welke plots we willen zien (tijdintervallen)
#
Bemonstering GEEN1HPLOT
Bemonstering D
Bemonstering W
#
#   data display - we zijn alleen geïnteresseerd in het voorjaar 2020
#
Analyseperiode 1/1/2020 30/6/2020
Plotparam SAVEONLY
#
#
#
Overzicht N02 P10 _PS Stat2017 Stat2018 Stat2019 Stat2020
#
#   Deze plot geeft het verschil goed aan
#
Plot_Soort N02 P10 _PS Stat2017 Stat2018 Stat2019 Stat2020 D_20_17 D_20_18 D_20_19 Nullijn
```

VOORBEELD 3: ONDERZOEK HOEKAFHANKELIJKHEID

Dit is een meer specialistische studie. De vraagstelling is of er een statistisch verband bestaat tussen windsnelheid en vervuiling. Een verband kan worden verwacht, omdat bij fikse wind de vervuiling door verkeer snel verwaait en naar hogere luchtlagen wordt gebracht.

CARDDECK

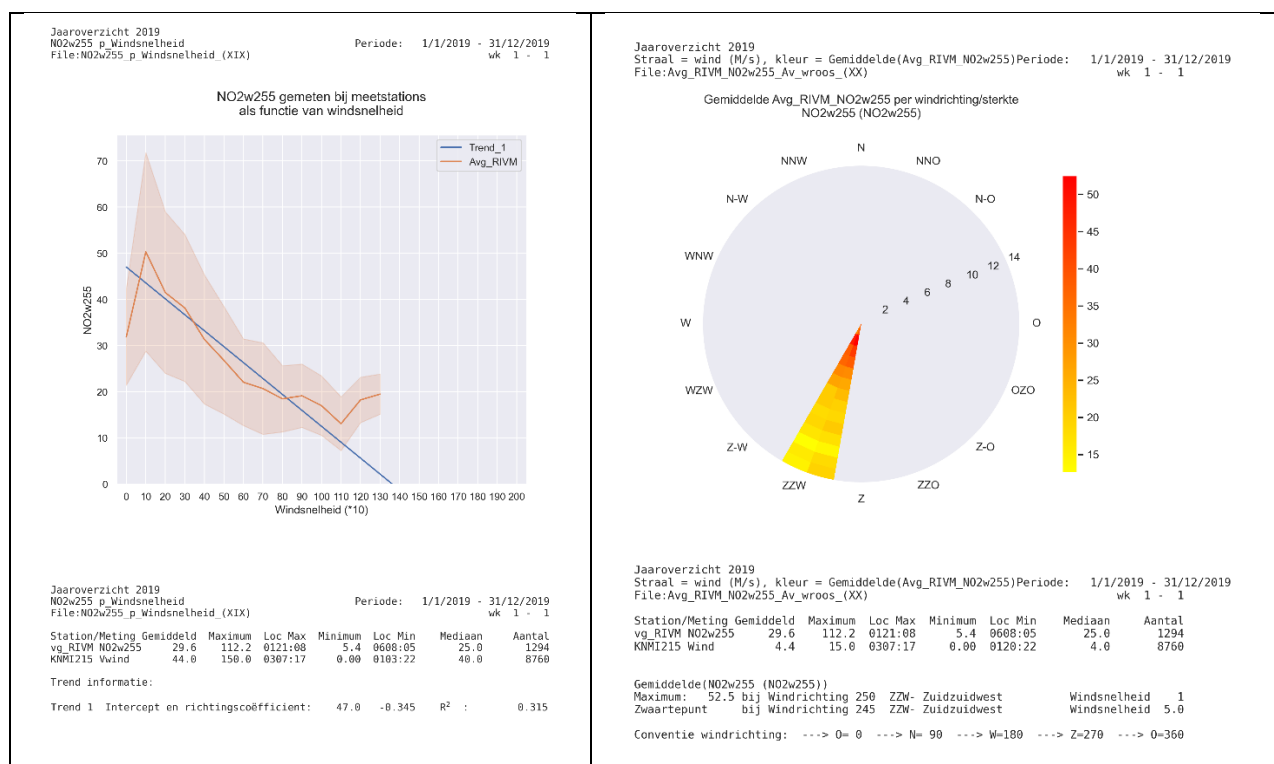
```
#
#   Doelstelling: onderzoek hoekafhankelijkheid van NO2 luchtvervuiling
#                   op RIVM stations
#
Debug 0 1 2 3 4 5 6 7 8 9
#
#   bepaal de dataverzameling - hele jaar 2019
#
Datadirectory 2020CumDatafile
#
Periode 1/12/2018 31/1/2020 Jaaroverzicht 2019
KNMI_gegevens
Kader RIVM
Nietgebruiken RIVM 4
#
#   Periode4stat kaarten kunnen blijven staan
#
Periode4stat CSLV 12 5/8/2018
Periode4stat CSLV 10 5/8/2018 24/01/2019
Periode4stat CSLV 7 1/1/2020
Periode4stat CSLV 34 20/7/2019
Periode4stat CSLV 36 1/4/2019
Periode4stat CSLV 40 1/4/2019
Periode4stat CSLV 43 1/4/2019
Periode4stat RIVM 0 1/1/2018 31/12/2019
#
#   data processing   - om daguitschieters te bepalen
#                       moeten meerdere dagen worden meegenomen
#
Grenswaarden          NO2 0 150
Daguitschieter        3    NO2 3
#
#   maak kopieën van de NO2 tijdreeks, afh van windrichting
#                       (12 taartpunten met een hoek van 30 graden)
#   doe dit voor alle stations
#
Kopieer_als (Dwind >= 0) xx NO2 RIVM
Kopieer_als (Dwind < 30) NO2w15 xx RIVM
Kopieer_als (Dwind >= 30) xx NO2 RIVM
Kopieer_als (Dwind < 60) NO2w45 xx RIVM
Kopieer_als (Dwind >= 60) xx NO2 RIVM
Kopieer_als (Dwind < 90) NO2w75 xx RIVM
Kopieer_als (Dwind >= 90) xx NO2 RIVM
Kopieer_als (Dwind < 120) NO2w105 xx RIVM
Kopieer_als (Dwind >= 120) xx NO2 RIVM
Kopieer_als (Dwind < 150) NO2w135 xx RIVM
Kopieer_als (Dwind >= 150) xx NO2 RIVM
Kopieer_als (Dwind < 180) NO2w165 xx RIVM
Kopieer_als (Dwind >= 180) xx NO2 RIVM
Kopieer_als (Dwind < 210) NO2w195 xx RIVM
Kopieer_als (Dwind >= 210) xx NO2 RIVM
Kopieer_als (Dwind < 240) NO2w225 xx RIVM
Kopieer_als (Dwind >= 240) xx NO2 RIVM
Kopieer_als (Dwind < 270) NO2w255 xx RIVM
Kopieer_als (Dwind >= 270) xx NO2 RIVM
Kopieer_als (Dwind < 300) NO2w285 xx RIVM
Kopieer_als (Dwind >= 300) xx NO2 RIVM
Kopieer_als (Dwind < 330) NO2w315 xx RIVM
Kopieer_als (Dwind >= 330) xx NO2 RIVM
```

```

Kopieer_als (Dwind < 360) NO2w345 xx RIVM
#
#      voeg de stations samen
#
Samenvoegen GEMIDDELDE Avg_RIVM RIVM
#
#      data display - eerst de plot van de windrichtingafhankelijke NO2 luchtvervuiling
#                      over tijd
#
Analyseperiode 1/1/2019 31/12/2019
Plotparam SAVEONLY
Bemonstering W
Plot NO2w15 NO2w45 NO2w75 NO2w105 NO2w135 NO2w165 NO2w195 NO2w225 NO2w255 NO2w285 NO2w315
NO2w345 _PS Avg_RIVM
#
#      Functiedefinitie - we willen weten wat het gedrag is van windrichtingafhankelijke
#                      NO2 vervuiling als functie van windkracht
#                      Als controle maken we ook een windroos, per hoek, en ook voor het totaal
#
1DFunctie_van Windkracht
2DFunctie_van Windroos
#
Soort_Functie_Items NO2 NO2w15 NO2w45 NO2w75 NO2w105 NO2w135 NO2w165 NO2w195 NO2w225 NO2w255
NO2w285 NO2w315 NO2w345 _PS Avg_RIVM

```

TWEE VAN DE GEGENEREERDE PLOTS



Appendix C. Inlezen meetgegevens

Conventies csv files

De meetgegevens worden betrokken van RIVM en worden ingelezen in ALVA in de vorm van csv files (comma separated volumes). ALVA veronderstelt de volgende conventies:

- **Ordering van de files:**
 - Er is één, en niet meer dan één, csv file per station per maand. Alle csv files staan in één directory. In ALVA wordt een analyse gemaakt van de inhoud van deze data directory zodat alléén de gegevens worden ingelezen waarnaar gevraagd wordt in het carddeck.
 - Uitzondering op deze regel is als er twee reeksen files bestaan, waarvan de éne reeks de tijdaanduiding in lokale Nederlandse tijd heeft, en de andere reeks de tijdaanduiding in UTC (Universele Tijd = Greenwich (winter) tijd zonder wisseling van winter- naar zomertijd). ALVA beschouwt een station waarvoor deze twee tijdaanduidingen bestaan als twee verschillende stations.
- **Naamgeving van de files:**
 - De naam van het station en de maand staan gecodeerd in de naam van de file. Bovendien kan een precursor zijn gespecificeerd. Deze precursor kan noodzakelijk zijn om aan te geven hoe de inhoud van de csv moet worden geïnterpreteerd, met name de aanduiding van de tijd, of dat de gegevensreeksen gekalibreerd zijn of niet (of dat zowel gekalibreerd als niet gekalibreerde data aanwezig zijn). Het eerste teken is altijd een underscore, daarna de eventuele precursor, dan de stationnaam, dan een underscore, dan de maand aanduiding (formaat yyyymm), dan .csv.
 - De aanwezigheid van precursor ~ geeft aan dat de tijdaanduiding niet in lokaal Nederlandse tijd wordt gegeven maar in UTC (Universele tijd = Greenwich (winter) tijd zonder wisseling van winter- naar zomertijd).
 - De precursor X is specifiek voor paddenstoelen, waarbij de stikstofdioxide meting niet is gekalibreerd (bijv de file _XSPPS014_201812.csv). Normaliter doet RIVM een zogenaamde nachtkalibratie.
 - De maand wordt gegeven in de notatie jjjjmm
 - Welke stations door ALVA worden herkend is gespecificeerd in StationDirectory22.py (subroutine stationdir()). Deze python subroutine kan naar de wens van de gebruiker worden aangepast. Daarvoor gelden de volgende regels:
 - Elk station is deel van een familie van stations. Het station wordt in het carddeck aangeduid met de naam "familie van station" en een getal.
 - Deze combinatie van familienaam en getal (bijvoorbeeld ~RIVM 0) wordt gebruikt in het bepalen van:
 - een stationsnaam, zoals aanwezig in de filenaam (in dit voorbeeld: NL10404)
 - een alias, die in de plots van ALVA en in diagnostische teksten van ALVA wordt gebruikt (in dit voorbeeld: NL404)
 - een adres (in dit voorbeeld: Den Haag-Rebecquestraat)
 - een eigenaar (niet relevant in dit geval)
 - een email adres (niet relevant in dit geval)
 - In het voorbeeld hierboven bevat de file met filenaam _~NL10404_202009.csv de gegevens van 1 september 2009 middernacht (op UTC tijd) tot het eind van de maand. In Nederlandse

lokale tijd begint deze tijdreeks op 30 augustus 10 uur 's avonds (verschil Nederlandse zomertijd met UTC).

- De kolommen van de csv files.
 - Scheidingsteken, en conventie voor decimaal teken:
Scheidingsteken, dat de scheiding tussen kolommen aangeeft, is een komma (",").
Het decimaalteken is een punt "." .
Hiermee volgt ALVA de Amerikaanse conventie.
 - Tijdaanduiding:
De naam van de tijd kolom is altijd "periode".
Metingen zijn altijd op hele uren.
Er wordt onderscheid gemaakt tussen de NL lokale tijd, die wordt ingelezen in een formaat zonder aanduiding van tijdzone (naïeve tijd), en de tijd met tijdzone aanduiding in UTC vorm (UTC = Universele tijd = Greenwich (winter) tijd zonder wisseling van winter- naar zomertijd).
De volgende formaten worden herkend:
Naïef: jjjj-mm-dd uu:mm:ss (voorbeeld: 2020-09-01 12:00:00)
UTC : jjjj-mm-ddTuu:mm:ss+00:00 (voorbeeld: 2017-01-01T01:00:00+00:00)
 jjjj-mm-dd uu:mm:ss+00:00 (voorbeeld: 2020-04-18 11:00:00+00:00)
Het naïeve formaat wordt geleverd door de getrivm.exe van Diederik van Hemert en de webinterface van Eddy Cornelisse. Het eerste UTC formaat komt van de interface naar luchtmeetnet.nl geschreven door Walter Takens (RIVMplain) , het laatste formaat komt van de getrivm.exe geschreven door Walter Takens. Deze laatste software benadert samenmeten.rivm.nl via de officiële RIVM API. Al deze programma's, behalve de webinterface van Eddy Cornelisse, worden hieronder beschreven.
 - De meetwaarden:
Er wordt een beperkt aantal types metingen herkend. De namen van deze meetwaardes zijn vermeld in het hoofdstuk "De attributen, wat voor een soort gegevens worden gebruikt?". De daar genoemde namen worden gebruikt in de carddecks. In de carddecks worden geen andere namen geaccepteerd (behalve de namen van grootheden die door één of andere bewerking in ALVA worden gegenereerd). Deze namen zijn over het algemeen kort, wat handig is wanneer lijstjes worden geprint. Als voorbeeld: in ALVA wordt met P10 het grove fijnstof PM₁₀ aangeduid. De completere aanduiding (bijvoorbeeld PM₁₀) of een wat explicietere beschrijving van de attribuut (bijvoorbeeld Roet in plaats van FN) wordt als bijschrift bij plots gegeven.
De namen van de kolommen in de csv files kunnen verschillen van de in het bovengenoemde hoofdstuk genoemde namen. De vertaling van de naam als gebruikt in de csv file naar de in ALVA geaccepteerde namen gebeurt bij het inlezen en is verschillend voor de verschillende families van meetstations en van hoe de file is geproduceerd.
Er is dus sprake van drie benamingen van de meetwaarden; deze worden vermeld in de file StationDirectory22.py in de subroutine attribdir().

Naam in csv file (afhankelijk van type file)	Naam in ALVA [te gebruiken in carddeck]	Betekenis / "mooiere" naam
NO2	NO2_m	NO ₂ (gemeten)
NO2_ref1	NO2_ref1	NO ₂ (ref 1)
NO2_ref2	NO2_ref2	NO ₂ (ref 2)
NO2_kal, NO2_KAL NO2	NO2	NO ₂ (geijkt)
P10, PM10	P10_m	PM ₁₀ (gemeten)
P10_ref1, PM10_ref1	P10_ref1	PM ₁₀ (ref 1)
P10_ref2, PM10_ref2	P10_ref2	PM ₁₀ (ref 2)
P10_kal, PM10_kal, P10_KAL, PM10_KAL, P10	P10	PM ₁₀ (geijkt)

P25, PM25	P25_m	PM ₂₅ (gemeten)
P25_ref1, PM25_ref1	P25_ref1	PM ₂₅ (ref 1)
P25_ref2, PM25_ref2	P25_ref2	PM ₂₅ (ref 2)
P25_kal, PM25_kal, P25_KAL, PM25_KAL, P25	P25	PM ₂₅ (geijkt)
temp, TEMP	temp	Temperatuur sensor
rh, RH	rh	Luchtvochtigheid sensor
ld, LD	ld	Luchtdruk sensor
FN	FN	Roet
O3	O3	Ozon
NO	NO	NO
SO2	SO2	SO ₂

- Andere in de csv files genoemde attributen:
Kolommen in de csv files met namen die niet in de linker kolom van bovengenoemde lijst zijn vermeld worden genegeerd.

Programma's om gegevens op te halen.

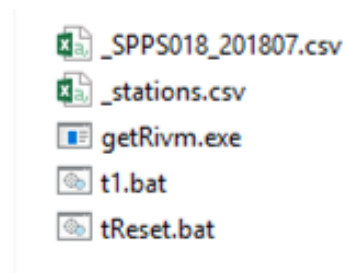
GetRIVM (oude versie)

Diederik van Hemert (LV2, Leidschendam-Voorburg) heeft de eerste versie van getRIVM gebouwd. Dit is een vb.net applicatie en draait alleen op Windows of een Windows emulator. De applicatie kan van elke Windows console worden gedraaid (te openen via Start – Systeem – Opdrachtprompt).

Deze versie leest de gegevens van de oude versie van `samenmeten.rivm.nl`

Versie 1 van GetRivm levert gegevens per maand op voor elk station. ALVA veronderstelt deze indeling in zijn inleesprocedure.

GEBRUIKSAANWIJZING GETRIVM VERSIE 1:



In de directory moet het programma `getRivm.exe` aanwezig zijn. Er wordt gebruik gemaakt van een commandline interface. Bij het downloaden van RIVM ga je door de volgende stappen

- Verzamel de gegevens van de stations op de RIVM website.
Met het commando **getRivm.exe reset** maak je het bestand `_stations.csv` aan. Deze bevat alle meetstations & meta gegevens die via de site beschikbaar zijn. De batfile `Reset.bat` bevat dit commando.
- Het bestand `_stations.csv` bevat per station de stationsnaam en de kitnaam. Deze zijn nodig bij de aanroep van `getRivm` om de gegevens van het station te downloaden.
- Omdat de file `_stations.csv` erg groot is, is het aan te raden een programmatje te schrijven om deze gegevens automatisch te selecteren.
- De aanroep **getRivm.exe station kit jaar maand** (jaar en maand als getallen) download gegevens van de website en slaat deze op in een csv file. De naamgeving van de csv file is:
underscore stationsnaam underscore jaar maand punt csv.
Voorbeeld: voor de gegevens van station SPPS018, met kitnaam RTD_15708092, voor jaar 2018

maand 7 gebruik je **getRivm.exe SPPS018 RTD_15708092 2018 7** . Het resultaat komt dan in file **_SPPS018_201807.csv** . Alle files worden geplaatst in een subdirectory **_SRC**.

- e. Aangezien men typisch veel files moet downloaden is het verstandig deze commando's in een programmatje te genereren en in een batfile op te slaan.

Deze getrivm versie (getrivm.exe) is meegeleverd met de ALVA 2.2 release in een subdirectory /getrivm_oud.

GetRIVM (nieuwe versie)

Walter Takens (Gouda) heeft een versie van getRIVM gebouwd op de door RIVM gebouwde API (Application Programmers Interface) van de RIVM website. Deze software heeft twee bronnen van informatie: de samenmeten.rivm API, en de API van luchtmeetnet.nl. De laatste website geeft de gegevens van de officiële RIVM stations.

Het programma is geschreven in python, en vereist daarom een environment waarin python bekend is. Net als in Appendix A kan men de software benaderen via de Anaconda Prompt (via Start menu). De opzet van de ALVA omgeving in Appendix A bevat alle python pakketten die voor getRIVM nodig zijn. getRIVM gebruikt een SQLite database als een tussentijdse opslag tussen de opslag op de RIVM website en de opslag als csv files. De SQLite database kan apart worden bestudeerd via een DB browser (te downloaden van het adres <https://sqlitebrowser.org/>).

De getRIVM software is nu benaderbaar via de Anaconda prompt (in Start menu) met de volgende stappen:

1. Activeren van de environment:
`conda activate ALVA`
2. Navigeren naar de directory waar de software aanwezig is (bijvoorbeeld :Documents\Calc\getRIVM):
`cd Documents`
`cd Calc`
`cd getrivm`
3. Een handleiding hoe getrivm te gebruiken is te verkrijgen door in te typen:
`python getrivm.py -man > handleiding.txt`

Deze versie van getRivm is erg flexibel. Men kan er voor kiezen om de Sqlite database geheel te vullen, of men kan elke keer selectief gegevens van de website halen en die dan vervolgens door te sluizen naar een csv file. Het is aan te bevelen scripts te ontwikkelen voor vaste taken, en deze als batfiles uit te voeren.

De tijden in de csv files zijn in UTC tijd.

De software is op een dusdanige manier geschreven dat het niet gemakkelijk aan te passen is door anderen dan de auteurs. Modules horend bij de software zijn: *create_db.py*, *getrivm.py*, *gerivmexport.py*, *man.py*, *nl.py*, *replicator.py* en de Unix shell *cleardb.sh*, en deze modules zijn bij de release van ALVA2.2 gevoegd. Twee directories worden door getrivm.py gebruikt: /db en /log. De SQLite database zit in /db. Indien deze niet aanwezig is moet hij gemaakt worden met het programma *create_db.py*. De log files opgeslagen in de /log directory moeten van tijd tot tijd verwijderd worden.

Scripts voor GetRIVM (nieuwe versie)

Bij de ALVA2.2 release is GetRIVM.py toegevoegd met de bijbehorende software.

Daarbovenop zijn de volgende procedures bijgevoegd:

Script:

rebuild_Db.bat

1. Bouwt de database (weer) op, en maakt een subdirectory aan met de naam yymmdd (jaar-maand-dag) waarin twee files komen te staan – één file met welke stations in de RIVM database zijn gevonden, en één file van stations die niet in de database aanwezig waren. In de eerste van deze files staat vermeld

wanneer de laatste meting met het station is gedaan. De namen van de te onderzoeken files worden geput uit *stationdirectory22.py*.

2. Dit script haalt alleen de Citizen Science stations op, niet de officiële RIVM stations.
3. De namen van de te onderzoeken files worden geput uit *stationdirectory22.py*. Verder gebruikt dit script de python files *tijdverschil.py* en *FileSelector.py*.
4. Er zijn geen command line parameters.

rivm2csv.bat

1. Haalt gegevens voor een specifieke maand uit de RIVM database en schrijft ze uit in csv files die door ALVA kunnen worden gelezen. De namen van de te onderzoeken files worden geput uit *stationdirectory22.py*.
2. Er is één command line parameter: jjjjmm (jaar, maand) . Voorbeeld:
`rivm2csv.bat 202011`
3. Dit script haalt alleen de Citizen Science stations op, niet de officiële RIVM stations. De officiële RIVM stations worden opgehaald door *RIVMplainkts.py*.

rivm2csv_year.bat

1. Haalt gegevens voor een specifieke jaar uit de RIVM database en schrijft ze uit in csv files die door ALVA kunnen worden gelezen. De namen van de te onderzoeken files worden geput uit *stationdirectory22.py*.
2. Er is één command line parameter: jjjj (jaar) . Voorbeeld:
`rivm2csv_year.bat 2020`
3. Dit script haalt alleen de Citizen Science stations op, niet de officiële RIVM stations. De officiële RIVM stations worden opgehaald door *RIVMplainkts.py*.

RIVMplain & RIVMplainkts

RIVMplain.py is python software ontwikkeld door Walter Takens (Gouda) waarbij gegevens van luchtmeetnet.nl worden gelezen. De software is aan te passen voor eigen taken en is bij de release van ALVA2.2 toegevoegd.

Output (csv files met meetgegevens) geschikt voor ALVA wordt gegenereerd in een versie ***RIVMplainkts.py***, met een functionaliteit die beperkt is ten opzichte van de oorspronkelijke *RIVMplain.py*.

RIVMplainkts.py kan worden aangeroepen in een commandline interface als volgt, met station (stationnaam), year (in formaat yyyy) en month (getal van 1 – 12) als variabelen.

```
python rivmplainkts.py --s station --y year
python rivmplainkts.py --s station --y year --m month
```

De gegenereerde outputfiles worden geplaatst in een subdirectory _SRC. De tijden in de csv files zijn in UTC tijd.

RIVMplainkts.py is toegevoegd bij de release van ALVA2.2. De software is benaderbaar via de Anaconda prompt (in Start menu) waarbij de python environment geactiveerd moet worden (soortgelijke procedure als beschreven voor getRIVM.py.).

Appendix D. Stationsgegevens in StationDirectory22.py

ALVA2.2 werkt alleen met gegevens van stations die in het programma zijn gespecificeerd. De stations worden opgeslagen in een Pandas dataframe (een soort Excelsheet met gegevens). De velden van dit dataframe worden ingevuld in een subroutine `stationdir()` in de module `StationDirectory22.py` (in de Current Working Directory). Het is aan de gebruiker om deze routine in te vullen voor zijn/haar verzameling stations die hij/zij wil bestuderen.

Introdactie van eventuele nieuwe filetypes kunnen veranderingen in de code van ALVA vereisen. In dit geval moet contact gelegd worden met lv2 (lv2@kpnmail.nl).

Alle stationsgegevens worden opgeslagen als functie van het stationtype en stationnummer, zoals die op de KADER en andere kaarten worden gebruikt:

- Stationtype ofwel familie van station. Naam van het stationtype is vrij, maar men moet wel rekening houden met een eventuele precursor (zie Appendix C)
- Stationnummer (keuze is vrij, maar moet voor de gebruiker logisch zijn).

De verdere gegevens per stationtype en stationnummer zijn

- Stationnaam - de naam die gespecificeerd staat op de csv files (inclusief precursor)
- Alias – een naam die intern in ALVA wordt gebruikt en die op de plots zichtbaar is
- Adres – wordt niet in ALVA gebruikt, kan wel in andere programma's worden gebruikt. Wel handig als opzoeklijstje voor als men wil weten welk station het is.
- Eigenaar – wordt niet in ALVA gebruikt, kan wel in andere programma's worden gebruikt. Wel handig als opzoeklijstje voor als men wil weten wie verantwoordelijk is. Indien onbekend wordt np.nan (Not a Number) of "Onbekend" genoteerd.
- Email– wordt niet in ALVA gebruikt, kan wel in andere programma's worden gebruikt. Wel handig als opzoeklijstje voor de communicatie. Indien onbekend wordt np.nan (Not a Number) of "Onbekend" genoteerd.

In de subroutine `stationdir()` wordt per stationtype de lijstjes aangemaakt als zogenaamde dictionaries: een lijstje met een key, en de bijbehorende grootheid, beide gespecificeerd tussen twee van de volgende soort haakjes `{}`. Voor de lijstjes met RIVM stations worden de dictionaries `RIVMst`, `aliasNL`, `adresNL`, `eigenaarNL` en `emailNL` aangemaakt. Met een aanroep van een subroutine `makeDFstat` worden deze vijf lijstjes samengesmeed tot een kleine dataframe `DFNL`. Deze kleine dataframe wordt gevoegd bij een lijstje van dataframes van andere stationtypes (met de functie `DFs.append(DFNL)`). Ook de naam van het stationtype moet nog aan een lijstje worden toegevoegd (met de statement `stationtype.append("RIVM")`).

```
# RIVMstations, stationtype RIVM
#
RIVMst = { 0 : "NL10404 - Den Haag-Rebecquestraat",
1 : "NL10445 - Den Haag-Veerakade",
2 : "NL10446 - Den Haag-Bleriotlaan",
3 : "NL10418 - Rotterdam-Schiedamsevest",
4 : "NL10448 - Rotterdam-Bentinckplein",
5 : "NL01493 - DCMR-Rotterdam Statenweg"}
aliasNL = \
{ 0 : "NL404",
1 : "NL445",
2 : "NL446",
3 : "NL418",
4 : "NL448",
5 : "NL1493"}
adresNL = \
{ 0 : "Den Haag-Rebecquestraat",
1 : "Den Haag-Veerakade",
2 : "Den Haag-Bleriotlaan",
3 : "Rotterdam-Schiedamsevest",
4 : "Rotterdam-Bentinckplein",
5 : "Rotterdam-Statenweg"}
eigenaarNL = { 0: np.nan, 1 : np.nan, 2 : np.nan, 3 : np.nan, 4: np.nan, 5: np.nan}
```

```

emailNL = { 0: np.nan, 1 : np.nan, 2 : np.nan, 3 : np.nan, 4: np.nan, 5: np.nan}
DFNL = makeDFstat(RIVMst, aliasNL, adresNL, eigenaarNL,emailNL)
DFs.append(DFNL)
stationtype.append("RIVM")

```

De naam van de lijstjes (RIVMst, aliasNL, adresNL, eigenaarNL,emailNL) is aan de gebruiker. Men kan op deze manier ook het stationtype ~RIVM definiëren, en elke ander stationtype. Aan het eind van alle lijstjes voor alle stationtypes wordt het geheel aan elkaar geknoopt met de statement `StationDF = pd.concat(DFs,keys= stationtype,axis=0).`

Omdat de lijstjes als dataframes worden verwerkt, kan men het lijstje ~RIVM ook afleiden van het lijstje RIVM, met wat minder statements. Dit vereist echter kennis van de programmeertaal. Dit ziet er dan als volgt uit:

```

# RIVM met UTC tijd      ~RIVM
DFNNL = DFNL.copy()
DFNNL["station"]="~" +DFNNL["station"].str[0:7]
DFNNL.loc[:,"alias"]="~" + DFNNL.loc[:,"alias"]
DFs.append(DFNNL)
stationtype.append("~RIVM")

```

De subroutine stationdir()

```

def stationdir():
    import numpy as np
    import pandas as pd

    #####
    # Gemeentestations
    #####
    DFs = []
    stationtype = []
    # stationtype GEM
    gemst = \
    { 1 : "NSCLL_1305167546887086",
      2 : "NSCLL_1305167546879497" ,
      3 : "NSCLL_1305167546892843",
      4 : "NSCLL_1305167546883668" ,
      5 : "NSCLL_1305167546888618",
      6 : "NSCLL_1305167546887753",
      7 : "NSCLL_1305167546907389",
      8 : "NSCLL_1305167546887076",
      9 : "NSCLL_1305167546887775",
      10 : "NSCLL_1305167546887914" }
    aliasGM = \
    { 1 : "GEM01", 2 : "GEM02" , 3 : "GEM03", 4 : "GEM04" , 5 : "GEM05",
      6 : "GEM06", 7 : "GEM07" , 8 : "GEM08", 9 : "GEM09" , 10 : "GEM10"}
    adresGM = \
    { 1 : "LV", # 20050001
      2 : "LV" , # 20050002
      3 : "LV", # 20050003
      4 : "LV" , # 20050004
      5 : "LV", # 20050005
      6 : "LV", # 20050006
      7 : "LV", # 20050007
      8 : "LV", # 20050008
      9 : "LV",
      10 : "LV" } # 20050010 -
    eigenaarGM = { 1 : np.nan, 2 : np.nan, 3 : np.nan, 4: np.nan, 5: np.nan,
                   6: np.nan, 7: np.nan, 8: np.nan, 9: np.nan, 10: np.nan}
    emailGM = { 1 : np.nan, 2 : np.nan, 3 : np.nan, 4: np.nan, 5: np.nan,
                6: np.nan, 7: np.nan, 8: np.nan, 9: np.nan, 10: np.nan}
    DFGEM = makeDFstat(gemst, aliasGM, adresGM, eigenaarGM,emailGM)
    DFs.append(DFGEM)
    stationtype.append("GEM")
    #####
    # RIVMstations, stationtype RIVM
    #####
    RIVMst = { 0 : "NL10404 - Den Haag-Rebecquestraat",
               1 : "NL10445 - Den Haag-Veerkade",
               2 : "NL10446 - Den Haag-Bleriotlaan",
               3 : "NL10418 - Rotterdam-Schiedamsevest",
               4 : "NL10448 - Rotterdam-Bentinkplein",
               5 : "NL01493 - DCMR-Rotterdam Statenweg"}
    aliasNL = \

```

```

{ 0 : "NL404",
1 : "NL445",
2 : "NL446",
3 : "NL418",
4 : "NL448",
5 : "NL1493"}
adresNL =\
{ 0 : "Den Haag-Rebecquestraat",
1 : "Den Haag-Veerkaade",
2 : "Den Haag-Bleriotlaan",
3 : "Rotterdam-Schiedamsevest",
4 : "Rotterdam-Bentinckplein",
5 : "Rotterdam-Statenvweg"}
eigenaarNL = { 0: np.nan, 1 : np.nan, 2 : np.nan, 3 : np.nan, 4: np.nan, 5: np.nan}
emailNL = { 0: np.nan, 1 : np.nan, 2 : np.nan, 3 : np.nan, 4: np.nan, 5: np.nan}
DFNL = makeDFstat(RIVMst, aliasNL, adresNL, eigenaarNL,emailNL)
DFs.append(DFNL)
stationtype.append("RIVM")
##### RIVM met UTC tijd ~RIVM
DFNNL = DFNL.copy()
#print(DFNNL)
DFNNL["station"]="~" +DFNNL["station"].str[0:7]
DFNNL.loc[:, "alias"]="~" + DFNNL.loc[:, "alias"]
#print (DFNNL)
DFs.append(DFNNL)
stationtype.append("~RIVM")
#####
# Paddenstoelen LV2 CSLV
#####
CSLVst = {
1 : "SPPS001", # testfile only
2 : "SPPS002", # testfile only
} #
aliasCS = {
1 : "CS001", 2 : "CS002"}
adresCS = {
1 : np.nan ,
2 : np.nan
}
eigenaarCS = {
1 : np.nan,
2 : np.nan,
}
emailCS = {
1 : np.nan,
2 : np.nan}
DFCS = makeDFstat(CSLVst, aliasCS, adresCS, eigenaarCS,emailCS)
DFs.append(DFCS)
stationtype.append("CSLV")
##### oude ongekalibreerde metingen XCSLV
DFXCS = DFCS.copy()
DFXCS.loc[:, "station"]="X" + DFXCS.loc[:, "station"]
DFXCS.loc[:, "alias"]="X" + DFXCS.loc[:, "alias"]
DFs.append(DFXCS)
stationtype.append("XCSLV")
##### nieuwe aanduiding, met UTC tijd ~CSLV
DFNCS = DFCS.copy()
DFNCS.loc[DFNCS.loc[:, "station"]=="SPPS001", "station"]= "NSPPS001"
DFNCS.loc[DFNCS.loc[:, "station"]=="SPPS002", "station"]= "NSPPS002"

DFNCS.loc[:, "alias"]="~" + DFNCS.loc[:, "alias"]
DFs.append(DFNCS)
stationtype.append("~CSLV")
#####
# Fijnstofstations LV2 type SB
#####
sbst = {
916: "NBI_SB916"
}
aliasSB = {
916: "SB916"
}
adresSB = {
916: "LV"
}
eigenaarSB = {

```

```

        916: "NN"
    }
    emailSB = {
        916: "NN"
    }
    DFsB = makeDFstat(sbst, aliasSB, adresSB, eigenaarSB,emailSB)
    DFs.append(DFsB)
    stationtype.append("SB")
    ##### SB stations met UTC aanduiding ~SB
    DFNSB = DFsB.copy()
    DFNSB["station"]="~" +DFNSB["station"]
    DFNSB.loc[:,"alias"]="~" + DFNSB.loc[:,"alias"]
    #print (DFNNL)
    DFs.append(DFNSB)
    stationtype.append("~SB")
    #####
    #   Luftdaten      LTD
    #####
    ltst = {
        18: "LTD_27329",
        19: "LTD_37452"
    }
    aliasLT = {
        18: "L27329",
        19: "L37452"
    }
    adresLT = {
        18: "Ijsselkade",
        19: "Ijsselkade"
    }
    eigenaarLT = {
        18: "Onbekend",
        19: "Onbekend"
    }
    emailLT = {
        18: "Onbekend",
        19: "Onbekend"
    }
    DFLT = makeDFstat(ltst, aliasLT, adresLT, eigenaarLT,emailLT)
    DFs.append(DFLT)
    stationtype.append("LTD")
    ##### Luftdaten met UTC ~LTD
    DFNLT = DFLT.copy()
    DFNLT["station"]="~" +DFNLT["station"]
    DFNLT.loc[:,"alias"]="~" + DFNLT.loc[:,"alias"]
    #print (DFNNL)
    DFs.append(DFNLT)
    stationtype.append("~LTD")
    StationDF = pd.concat(DFs,keys= stationtype,axis=0)
    return StationDF

```

Appendix E. Attriboot gegevens in StationDirectory22.py

De namen van de meetwaardes zijn vermeld in het hoofdstuk “De attributen, wat voor een soort gegevens worden gebruikt?”.

Introductie van eventuele nieuwe attributen kunnen veranderingen in de code van ALVA vereisen. In dit geval moet contact gelegd worden met lv2 (lv2@kpnmail.nl).

Appendix F. Belangrijkste rijksomgevingswaarden luchtkwaliteit

Overgenomen uit <https://aandeslagmetdeomgevingswet.nl/thema/lucht/belangrijkste-rijksomgevingswaarden-luchtkwaliteit/>

Voor de kwaliteit van de buitenlucht bestaan verschillende soorten normen. Deze pagina geeft informatie over luchtkwaliteitsnormen op leefniveau (de buitenlucht), die zijn gericht op de bescherming van mensen.

Luchtkwaliteitsnormen

De tabel geeft een overzicht van de belangrijkste luchtkwaliteitsnormen voor luchtkwaliteit. Waaronder de [rijksomgevingswaarden](#) voor luchtkwaliteit uit het Besluit kwaliteit leefomgeving. De rijksomgevingswaarden met een [resultaatsverplichting](#) zijn opgenomen in onderstaande tabel. De rijksomgevingswaarden met een [inspanningsverplichting](#), fijnstof (PM_{2,5}), ozon (O₃), benzo[a]pyreen, arseen, cadmium en nikkel staan niet in de tabel.

Ook bevat de tabel de [advieswaarden van de Wereldgezondheidsorganisatie \(WHO\)](#) voor fijnstof (PM₁₀ en PM_{2,5}). In de praktijk zijn er in Nederland nog [enkele overschrijdingen \(opent externe website\)](#) van de omgevingswaarden voor fijnstof (PM₁₀) en stikstofdioxide (NO₂).

OVERZICHT VAN DE BELANGRIJKSTE LUCHTKWALITEITSNORMEN

STOF	Soort norm	Concentratie	Status	Wanneer toetsen
NO ₂	Jaargemiddelde	40 µg/m ³	omgevingswaarde / WHO advieswaarde	omgevingsplan, projectbesluit en omgevingsvergunning
NO ₂	Uurgemiddelde (mag max. 18 keer per jaar worden overschreden)	200 µg/m ³	omgevingswaarde / WHO advieswaarde	omgevingsvergunning projectbesluit
NO _x	Jaargemiddelde	30 µg/m ³	omgevingswaarde	omgevingsvergunning
PM ₁₀	Jaargemiddelde	40 µg/m ³	omgevingswaarde	omgevingsplan en omgevingsvergunning
PM ₁₀	Jaargemiddelde	20 µg/m ³	WHO advieswaarde	optioneel: omgevingsverordening omgevingsplan
PM ₁₀	24-uurgemiddelde (mag max. 35 keer per jaar worden overschreden)	50 µg/m ³	omgevingswaarde	omgevingsplan, projectbesluit en omgevingsvergunning
PM _{2,5}	Jaargemiddelde	25 µg/m ³	omgevingswaarde	omgevingsvergunning
PM _{2,5}	Jaargemiddelde	10 µg/m ³	WHO advieswaarde	optioneel: omgevingsverordening en omgevingsplan
BEN-	Jaargemiddelde	5 µg/m ³	omgevingswaarde	omgevingsvergunning
ZEEN				
SO ₂	Jaargemiddelde	20 µg/m ³	omgevingswaarde	omgevingsvergunning
SO ₂	Uurgemiddelde (mag max. 24 keer per jaar worden overschreden)	350µg/m ³	omgevingswaarde	omgevingsvergunning

SO₂	24-uurgemiddelde (mag max. 3 keer per jaar worden overschreden)	125µg/m ³	omgevingswaarde	omgevingsvergunning
SO₂	Winterhalfjaargemiddelde (periode 1 oktober t/m 31 maart)	20 µg/m ³	omgevingswaarde	omgevingsvergunning
LOOD	Jaargemiddelde in PM ₁₀	0,5µg/m ³	omgevingswaarde	omgevingsvergunning

Toelichting normen voor fijnstof

De omgevingswaarden voor PM_{2,5} en PM₁₀ komen uit de [Europese richtlijn \(2008/50/EG\) \(opent externe website\)](#) voor luchtkwaliteit.

Bij een jaargemiddelde concentratie PM₁₀ van 32,5 µg/m³ en hoger vindt overschrijding van de dagnorm (24-uurgemiddelde) plaats.

Relatie PM₁₀ - PM_{2,5}

Een bijdrage aan de concentratie van PM_{2,5} hoeft niet apart te worden beoordeeld. De toetsing aan PM₁₀ maakt voldoende aannemelijk dat de omgevingswaarden voor PM_{2,5} in acht worden genomen. PM₁₀ en PM_{2,5} concentraties zijn namelijk [sterk aan elkaar gerelateerd \(opent externe website\)](#). Als aan de omgevingswaarden voor PM₁₀ wordt voldaan, wordt daarmee ook aan de omgevingswaarden voor PM_{2,5} voldaan.

Rijksverplichtingen voor ozon en PM_{2,5}

De omgevingswaarden die in de bovenstaande tabel staan gelden voor decentrale overheden. Daarnaast zijn er ook omgevingswaarden waar alleen het Rijk voor verantwoordelijk is.

Voor het terugdringen van ozonconcentraties is internationale samenwerking van belang, daarom heeft het Rijk een resultaatsverplichting om de omgevingswaarden te behalen.

Naast de genoemde waarden van PM_{2,5} is er ook een blootstellingsconcentratieverplichting. De blootstellingsconcentratie is de concentratie waaraan de stedelijke bevolking wordt blootgesteld. Deze blootstellingsconcentratieverplichting geldt voor de rijksoverheid. Lokale overheden hoeven hier niet aan te toetsen. Voor het Rijk gelden voor PM_{2,5} de volgende omgevingswaarden voor de 3-jaargemiddelde blootstellingsconcentratieverplichting:

- Resultaatsverplichting van 20 µg/m³
- Inspanningsverplichting van 14,4 µg/m³

Appendix G. Structuur ALVA23

ALVA23 bestaat uit drie modules:

1. ALVA23Main.py. Doet alle berekeningen en roept ALVA22INPUTMod aan.
2. ALVA22INPUTMOD regelt het inlezen van het carddeck en controleert of de input valide is. Roept daarbij StationDirectory22.py aan.
3. StationDirectory22.py bevat lijsten met toegestane types metingen en de lijsten van toegestane en gebruikte meetstations.

In de eerste regels van ALVA23Main wordt de parameterfile gevonden, of via parameters op de commandline, of via de file ALVAINPUTFILE.txt die in de carddecks directory staat.

Daarna worden een groot aantal subroutines gedefinieerd, die verderop in ALVA23Main worden aangeroepen en gebruikt worden voor de berekeningen en het plotten.

Het hoofdprogramma gaat pas na de definitie van deze subroutines door.

De tussenliggende code zijn subroutines. ALVA werkt met subroutines. De subroutines horen intern consistent te zijn – er worden geen globale variabelen gebruikt. Variabelen worden als parameters doorgegeven naar de subroutine in de call, en aan het eind van de subroutine worden de resultaten bij de return teruggegeven. Dat leidt ertoe dat sommige calling sequences vrij lang zijn. Vrijwel overal wordt in een call een debugparameter doorgegeven, met de naam van een debugfile. Diagnostische output wordt gestuurd naar de debugfile als de debugparameter de waarde True heeft.

Als één van de eerste calls in ALVA23Main wordt de subroutine readComm in de ALVA22INPUTMOD module aangeroepen. Deze subroutine readComm, met de andere subroutines in ALVA22INPUTMOD regelt het inlezen van alle parameterkaarten, controleert welke stations en welke tijdreeksen in de berekeningen moeten worden meegenomen (en welke additionele stations en tijdreeksen in de berekeningen worden geproduceerd), en levert aan ALVA23Main een aantal lijstjes af die nodig zijn om de datastructuur waarvan ALVA gebruik maakt op te zetten [met name om te bepalen welke datafiles met meetstation-gegevens moeten worden ingelezen].

De datastructuur in ALVA bestaat uit tijdreeksen, met een sampling van één uur. De reeksen zijn per station en per type meting (attribuut) geïndexeerd. Er wordt onderscheid tussen stationstypes en tussen verschillende types metingen (attributen). De definities worden gegeven in StationDirectory22.py, en zijn tijdens de berekeningen beschikbaar in de vorm van tabellen onder de naam StationDF en AttribDF. Tijdreeksen van een meetstation kunnen in plots en in berekeningen gecombineerd worden met tijd-gerelateerde tijdreeksen en metingen van een KNMI station. De KNMI metingen en de tijd-gerelateerde tijdreeksen kunnen niet worden veranderd – zij vormen als het ware de ruggengraat van de datastructuur.

Men kan nieuwe tijdreeksen (dus nieuwe attributen) definiëren voor de meetstations, en men kan nieuwe “meetstations” definiëren, bijvoorbeeld een meetstation die tijdreeksen bevat die de gemiddelde zijn van tijdreeksen van een aantal stations. Deze “meetstations” ofwel pseudostations worden als een apart type meetstation gecatalogiseerd (type_PS).

Zoals gezegd, readComm in ALVA22INPUTMOD en de daarin opgeroepen routines controleert of de invoer in het carddeck correct is en of er geen problemen bij de uitvoering te verwachten zijn. Tijdens deze fase worden fouten en waarschuwingen aangegeven als printoutput.

De parameterkaarten in het carddeck worden ingelezen in een loop (regel 163 in ALVA22INPUTMOD tot regel 763). In deze loop wordt gekeken of de soort parameterkaart bekend is, en dan wordt afhankelijk van de parameterkaart actie ondernomen. Deze actie heeft de vorm van een call naar een subroutine, eventueel gevolgd door een aantal stappen om gegevens toe te voegen aan één of meer van de lijstjes die na afloop van readComm aan het hoofdprogramma moet worden doorgegeven. De subroutine die de parameterkaart analyseert draagt – om de functie duidelijk te maken – de naam “readXXXXcard”, waarbij XXXX de naam van de parameterkaart is.

Als voorbeeld:

```
# begin van de loop – er wordt ingelezen van een file aangegeven met infile
for line in infile:
#..... de ingelezen regel wordt gesplitst in woorden (opgeslagen in een list “words”).....
#..... het eerste item van words (words[0]) geeft de soort parameterkaart
#..... controle is door een aantal if, elif statements
    elif words[0].capitalize() == “VLOEIEND”.capitalize():
        #.... kaart VLOEIEND is gevonden....
        rollist,pollval,AttribDF =
        readVLOEIENDcard(words,stationlist,StationDF,AttribDF,idebug0,debugfile)
        #.... de subroutine readVLOEIENDcard genereert een lijst instructies met naam rollist,
        #.... een lijst pollval met de attributen (soorten meetreeksen) die gebruikt worden
        #.... in de operatie VLOEIEND en een gewijzigde definitie van AttribDF, waarin
        #.... eventueel nieuwe attributen zijn opgenomen
        #.... de operatie VLOEIEND geeft geen wijzigingen in de StationDF, dus dat is geen
        #.... output parameter
        #.... Input parameters zijn: de instructies in words,
        #.... de lijst stations die beschikbaar zijn en zijn gespecificeerd door eerder aangeroepen
        #.... operaties, de definities in StationDF, AttribDF, de debugparameter idebug0 en de
        #.... naam van de debugfile.
        if len(rollist) > 0:
            #.... als de gegevens in words geleid hebben tot een zinnig instructie om later uit te voeren
            #.... worden wat lijstjes bijgewerkt. De instructies worden toegevoegd aan een
            #.... lijst processlist, en de namen van de attributen worden toegevoegd aan een lijst
            #.... pollution. De functie verzamel zorgt er voor dat er geen duplicaten in de lijst
            #.... pollution voorkomen
            processlist.append(rollist)
            pollution = verzamel(pollution,pollval)
        elif words[0].capitalize() == ....
```

De subroutine readVLOEIENDcard ziet er als volgt uit

```
def readVLOEIENDcard(words,stationlist, StationDF, AttribDF, idebug, debugfile) :
#
# controleren foutcondities
#
if len(words)> 5:
    #.... Eerste parameter op VLOEIEND kaart moet het soort gemiddelde aangeven
    if words[1].upper() not in ['GEMIDDELDE','MEDIAAN','STDEV']:
        # geef waarschuwing en breek het programma af (routine “waarschuwing”)
        waarschuwing("niet_herkend",words,words[1],True)
    #.... tweede parameter moet als integer kunnen worden gelezen
    try :
        filterlength = int(words[2])
    #.... als dat niet lukt – geef een foutmelding en breek het programma af
    except :
        #.... roep de standaard foutmelding routine aan
        foutmelding("positie",words,words[2],True)
    #.... Vind op de kaart waar de definities beginnen van de stations die meegenomen
    #.... moeten worden. Dit is standaard aan het eind van de kaart. Het begin van de
    #.... reeks in dit geval moet op positie 5 liggen (lenuarly)
    lenuarly,MaxStatlist = vindlijststationsALL(words,StationDF,1, idebug, debugfile)
```

```

#.... de gevonden lijst met stations kan groter zijn dan de lijst die eerder is geformuleerd
#.... en waaruit kan worden gekozen --- beperk de lijst
MaxStationList2 = inter_section(MaxStatlist,stationlist)
#.... het aantal stations moet groter zijn dan 0
if len(MaxStationList2) > 0:
    #.... begin van reeks stations moet op positie 5 liggen (lenearly)
    if lenearly != 5 :
        #.... fatale fout – stop het programma
        foutmelding("kaart_indeling",words,"attributen",True)
    pollution = words[4]
    #.... controleer of words[4] (pollution) al bekend is als attribuut
    #.... pollcontrol geeft een fatale fout als dit niet het geval is
    pollcontrol(words, pollution ,["kit","KNMI","calc"],AttribDF)
    #.... aan alle condities is voldaan
    rollst = [words[0],words[3],words[1],filterlength, pollution,MaxStationList2]
    #.... we weten nu dat er twee attributen (twee types meetreeksen) meespelen
    pollution=[words[3],words[4]]
    #.... en voegen de nieuwe attribuut toe aan de database AttribDF
    AttribDF = voegattribtoe(words[3],AttribDF,words)
else:
    #.... geen stations gevonden: fout, maar niet fataal
    #.... resultaat – geen actie (rollst is leeg)
    foutmelding("nietgevonden",words,"stations",False)
    rollst = []
else:
    #.... te weinig parameters: fout, programma stopt
    foutmelding("nr_parameters",words,words,True)
# geef standaard diagnostische output in de debugfile als daarom wordt gevraagd
kaartdiagnostiek(words,1,rollst,idebug,debugfile)
# ga terug naar readComm
return rollst,pollution, AttribDF

```

De gegevens in de list rollst worden in de bovenliggende subroutine toegevoegd aan een list processlist, die, nadat alle kaarten zijn ingelezen, wordt doorgegeven aan het hoofdprogramma in ALVA23MAIN. De list proceslist bepaalt de processingstappen in het hoofdprogramma (in een loop van regel 3730 tot en met regel 3768. Weer wordt gebruik gemaakt van een “Case” constructie, in python vorm gegeven door if en elif statements, waarbij gecontroleerd wordt op de naam van de parameterkaart. Weer is er een conventie voor de processing routines: procXXXX waarbij XXXX de naam van de parameterkaart is.

for polprocess in processlist:

```

# polprocessid is de naam van de parameterkaart
polprocessid = polprocess[0].capitalize():
if any(debugParam): print("Verwerken    :",polprocessid)
# hier begint de case constructie met if, elif statements
if polprocessid == "GRENSWAARDEN".capitalize():
    dfObj = procGRENSWAARDEN(polprocess,dfObj,idebug,debugfile)
....
# in deze loop wordt ook de processingstap “VLOEIEND” uitgevoerd:
elif polprocessid == "VLOEIEND".capitalize() :
    #
    # dfObj is de datastructuur waarin alle stations en alle attributen zijn opgeslagen -
    # aangezien dfObj door de processing routine wordt veranderd, is deze zowel input als output
    # polprocess bevat hier de instructies voor procVLOEIEND (en heette rollst in de inleesphase)

```

```
# in de list pollprocess worden alle stations genoemd. Omdat ook gegevens uit de ruggengraat
# worden gebruikt, moet het knmistation station bekend zijn.
dfObj = procVLOEIEND(polprocess,dfObj,knmistation,idebug,debugfile)
#
# er zijn ook processingstappen als gevolg van parameters ingevoerd bij andere kaarten
elif polprocessid == "OVERBRENGEN".capitalize():
....
```

In de routine readComm worden meer lijstjes bijgehouden. Naast processlist wordt ook een lijst stattoevoeglijst bijgehouden: deze geeft informatie door aan de hoofdroutine over processingstappen die resulteren in nieuwe stations – deze stappen worden in het hoofdprogramma uitgevoerd nadat de processingstappen in proclist zijn uitgevoerd. Daarnaast is er nog een resamplist (waarbij bijgehouden wat voor een soort bemonsteringen moeten worden gebruikt in de analyses en de displays), en een lijst actionlist (waarbij bijgehouden wordt wat voor een plots moeten worden gemaakt). Tenslotte zijn er nog lijstjes gerelateerd aan functieberekeningen.

De processinglijsten voor deze laatste opties zijn wat ingewikkelder, en om niet nodeloos code te dupliceren wordt soms een enkele routine gebruikt voor twee verschillende opties:

```
elif actionid == "VIOOLPLOT".capitalize():
    #.. bij deze routine wordt geen output naar de hoofdroutine teruggegeven. Er wordt
    # alleen een plot (of meerdere plots) gemaakt
    actDOOSDIAGRAM(action,dfObj,AttribDF,knmistation,"","",periodebeperking, \
        OutputDir, plotsave, plotyes, idebug, debugfile, suffixpl)
```

Tegelijkertijd worden in de inleesroutine readComm ook de lijstjes bijgehouden van toegestane stations en attributen (welke zijn bij het programma bekend), en lijstjes van gebruikte stations en attributen (welke worden gebruikt, en welke worden eventueel toegevoegd).

De reden van al deze lijstjes: er moet bepaald worden welke van de ingelezen attributen gebruikt zullen worden en welke niet. De laatste worden bij het inlezen meteen weggegooid om ruimte te besparen – dfObj, de datastructuur met alle ingelezen en geprocesste data moet zo klein mogelijk zijn.

Het programmeren van een nieuwe optie in ALVA houdt dus in principe in:

- Het maken van een routine waarbij de parameterkaart wordt ingelezen en geïnterpreteerd
- Het maken van een processing routine, waarbij dfObj wordt gemanipuleerd, of het maken van een routine waarbij de gegevens in dfObj worden gebruikt om plots te maken.