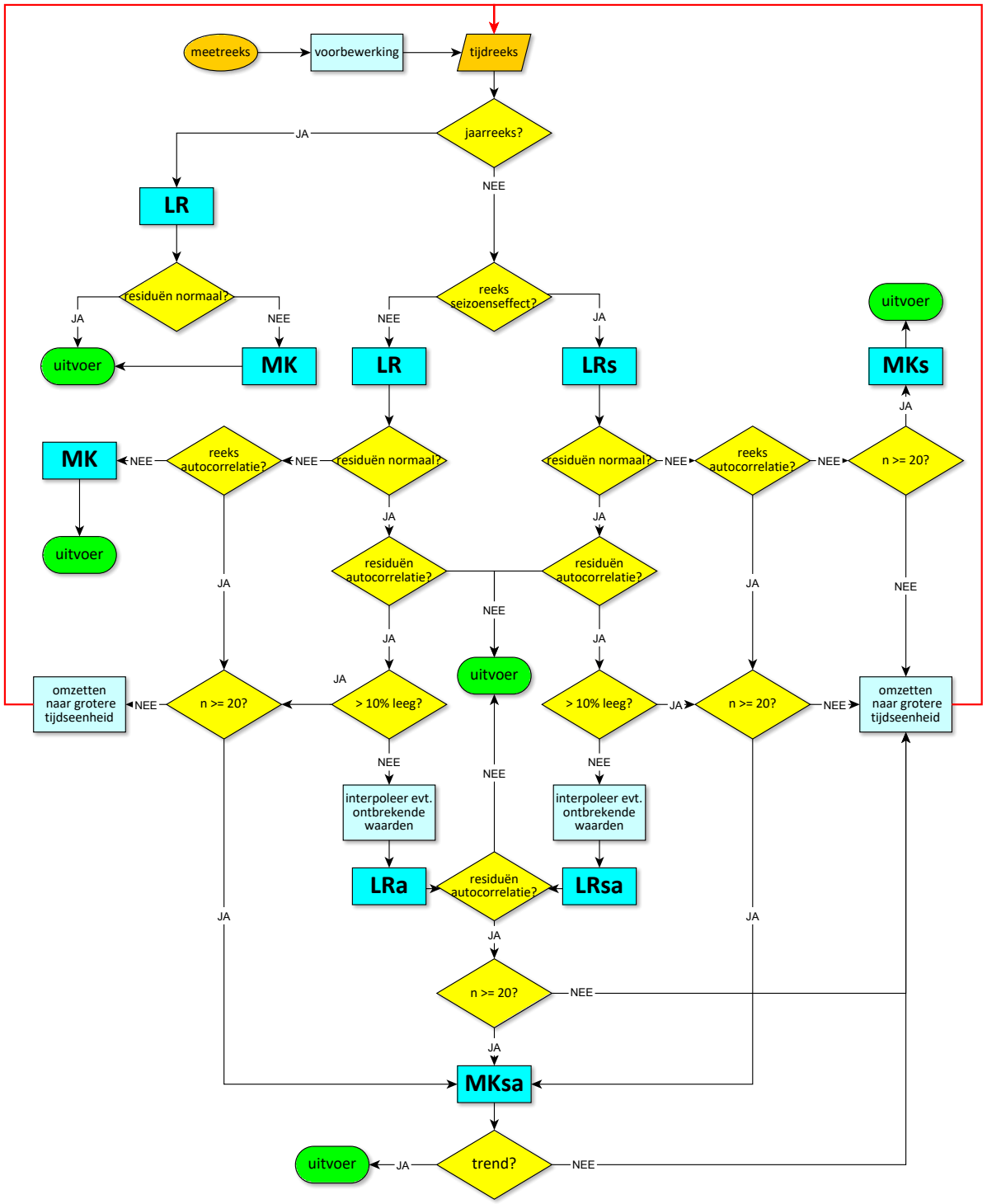
Trendanalist

Gebruikershandleiding



April 2026

Trendanalist

Gebruikershandleiding

Auteurs:
drs. Paul K. Baggelaar
ir. Eit C.J. van der Meulen (AMO)

April 2026

Deldenerstraat 81
7551 AC Hengelo
Tel.: +31624865321
trendanalist@amo-nl.com
www.amo-nl.com

INHOUD

1 INLEIDING 3

1.1 MOGELIJKHEDEN VAN *TRENDANALIST* 3
1.2 LEESWIJZER 3

2 DE BENODIGDE INVOER VAN *TRENDANALIST* 4

2.1 BEDOELING 4
2.2 VEREISTE INFORMATIEINHOUD VAN MEETREEKS 4
2.3 VEREISTE FORMAAT INVOERBESTAND 4
2.3.1 Vereiste formaat invoerbestand met extensie '.tai' 4
2.3.2 Vereiste formaat invoerbestand met extensie '.csv' 6

2.3.2.1 IBEVER CSV-DATAFORMAAT 6

2.3.2.2 AQUO-KIT CSV-DATAFORMAAT 7

2.3.3 Vereiste formaat invoerbestand met extensie '.dia' 9

3 HET REKENHART VAN *TRENDANALIST* 11

3.1 DE INSPIRATIEBRONNEN VAN *TRENDANALIST* 11
3.2 TRENDDETECTIE MET *TRENDANALIST* 11
3.3 LINEAIRE REGRESSIETOETS EN ZIJN UITBREIDINGEN 12
3.3.1 Lineaire regressietoets 12
3.3.2 Lineaire regressietoets met verdiscontering seizoenseffecten en/of autocorrelatie 13
3.4 MANN-KENDALL-TOETS EN ZIJN UITBREIDINGEN 14
3.4.1 Mann-Kendall-toets 14
3.4.2 Mann-Kendall-toets met verdiscontering van autocorrelatie 15
3.4.3 Seizoenale Mann-Kendall-toets 16
3.4.4 Seizoenale Mann-Kendall-toets met verdiscontering van autocorrelatie 17
3.5 AUTOMATISCHE SELECTIEPROCEDURE VAN *TRENDANALIST* 18
3.5.1 Stroomschema van de selectieprocedure 18
3.5.2 Voorbewerking meetreeks 20
3.5.3 Vertoont de tijdreeks seizoenseffecten? 24
3.5.4 Zijn de modelresiduën normaal verdeeld? 25
3.5.5 Vertonen de modelresiduën autocorrelatie? 26
3.5.6 Vertoont de tijdreeks autocorrelatie? 27
3.6 TRENDKWANTIFICERING MET *TRENDANALIST* 28
3.6.1 Theilhelling 28
3.6.2 Kendall-seizoenshelling 29
3.6.3 Trendlijn 29

4 DE INSTELOPTIES VAN *TRENDANALIST* 30

4.1 STANDAARDMAPPEN 30
4.2 MIDDELINGSMETHODE 30
4.3 ALTERNATIEVE HYPOTHESE 30
4.4 BETROUWBAARHEID TRENDTOETS 31
4.5 STARTPUNT TREND 31
4.6 GECENSUREERDE WAARDEN 31
4.7 GECENSUREERDE VRACHTEN 32
4.8 AUTOCORRELATIEPLOTS 32
4.9 SMOOTH 33
4.10 VOORBEWERKEN MEETREEKSEN 33
4.11 OVERZICHT OPTIES 33
4.12 MELDINGEN EN UITVOER 33
4.13 GROOTTE GRAFIEKEN 34
4.14 FONT GRAFIEKEN 34
4.15 TAAL 34

5 PRESENTATIEMOGELIJKHEDEN 35

5.1 MEETREEKSPLOTS 35
5.2 JAARBOXPLOTS 36
5.3 MEETDICHTHEIDSMATRIX 36
5.4 KENGETALLENMATRIX 37
5.5 MEETLOCATIEKAART 38
5.6 BEKIJK SESSIE 38
5.7 TRENDPLOTS 38
5.8 RESULTATEN 40
5.9 TRENDPALET 43
5.10 TRENDKAART 44
5.10.1.1 GERAADPLEEGDE LITERATUUR 46
5.10.1.2 BIJLAGE – VERKLARING VAN EEN AANTAL TERMEN 48

1 Inleiding

1.1 Mogelijkheden van *Trendanalist*

Trendanalist is een programma om trendanalyse uit te voeren. Trendanalyse is een statistische techniek om objectief vast te kunnen stellen of een bepaalde meetreeks[1] een trend vertoont, waarbij tevens een schatting voor de grootte van de trend wordt gegeven. Onder het begrip trend verstaan we hier een verandering van het centrum van de kansverdeling van de meetwaarden over tenminste enkele jaren. Seizoensmatige veranderingen en kortstondige calamiteiten vallen daar dus niet onder.

Trends kunnen door *Trendanalist* objectief worden vastgesteld als er meerjarige meetreeksen beschikbaar zijn. Daarbij houdt het programma rekening met de bijzondere karakteristieken die vaak kenmerkend zijn voor meetreeksen van milieuparameters, zoals niet-normale kansverdelingen, seizoenseffecten, autocorrelatie[2], ontbrekende meetwaarden en gecensureerde meetwaarden[3]. Het programma kan de analyse volledig automatisch uitvoeren, zodanig dat er voor de analyse van grote aantallen meetreeksen – zoals bijvoorbeeld verzameld met een meetnet – slechts één druk op de knop nodig is. Maar de gebruiker kan een meetreeks desgewenst ook handmatig analyseren, aangezien *Trendanalist* daarvoor een uitgebreide statistische toolbox bevat.

Trendanalist kan de basisgegevens en de resultaten presenteren in de vorm van tabellen, figuren en kaarten. Voor de kaartpresentatie wordt Google Earth door *Trendanalist* aangeroepen.

1.2 Leeswijzer

De eerste hoofdstukken van deze handleiding beschrijven de werking van de belangrijkste onderdelen van het programma *Trendanalist*, namelijk:

- hoofdstuk 2: de benodigde invoer;
- hoofdstuk 3: het rekenhart;
- hoofdstuk 4: de instelopties;
- hoofdstuk 5: de presentatiemogelijkheden.

Het hoofddeel van deze handleiding sluit af met de alfabetisch gerangschikte lijst van referenties.

De bijlage bevat een verklaring van enkele technische – veelal statistische - termen die in deze handleiding worden gebruikt.

[1] Een meetreeks is een chronologisch gerangschikte reeks meetwaarden van een bepaald verschijnsel, zoals de concentratie van een bepaalde stof, de grondwaterstandstand, of de afvoer.
[2] Het verschijnsel dat opeenvolgende meetwaarden doorgaans minder van elkaar verschillen dan meetwaarden die op grotere tijdsintervallen van elkaar liggen.
[3] Een gecensureerde meetwaarde is bijvoorbeeld '< 1'. De '1' staat hier voor de rapportagegrens.

2 De benodigde invoer van *Trendanalist*

2.1 Bedoeling

Als het programma is opgestart, verschijnt het hoofdscherm. Klik op 'Bestand' en dan op 'Inlezen invoerbestand' om een bestand met één of meer meetreeksen in te lezen. In § 2.2 is aangegeven waar een meetreeks minimaal aan moet voldoen om geanalyseerd te kunnen worden. En in § 2.3 is aangegeven hoe het invoerbestand moet zijn opgebouwd.

2.2 Vereiste informatieinhoud van meetreeks

Trendanalist zal eerst van elke ingelezen meetreeks nagaan of deze de minimale informatieinhoud heeft om trendanalyse te kunnen uitvoeren. Die vereisten zijn:

1. meetwaarden over minstens 4,5 jaar (periode tussen start en eind meetwaarden);
2. minstens 5 meetwaarden, of anders minstens 3 verschillende jaarwaarden;
3. minder dan 80% gecensureerde meetwaarden, uitgezonderd vrachten (product meetwaarden en debiet) dan minder dan 30% gecensureerde waarden.

Trendanalist interpreteert meetwaarden als vrachten als de meeteenheid eindigt op de tijdseenheid: '/s', '/sec', '/uur', '/dag', '/week', '/maand', '/kwartaal' of '/jaar'. Als een ingelezen meetreeks niet voldoet aan deze drie criteria, dan wordt de reeks door *Trendanalist* verder niet beschikbaar gesteld voor de trendanalyse.

2.3 Vereiste formaat invoerbestand

Trendanalist kan alleen meetreeksen inlezen die zijn opgeslagen als een tekstbestand (ASCII-formaat). De extensie van de bestandsnaam moet zijn: '.tai', '.csv', '.dia', '.riw' en '.hyc'. Het standaard invoerbestand voor *Trendanalist* is het tai-bestand (zie § 2.3.1). *Trendanalist* gaat ervan uit dat een bestand met de extensie '.csv' afkomstig is van de applicatie *iBever* of *Aquo-kit*. De applicatie programma *iBever* is ontwikkeld door Rijkswaterstaat om een gestandaardiseerde verwerking en uitwisseling van 'natte' meetgegevens mogelijk te maken (zie § 2.3.2). IBever wordt inmiddels niet meer ondersteund door Rijkswaterstaat, maar is opgevolgd door de applicatie Aquo-kit. Trendanalist kan beide csv-dataformaten inlezen. Verder kan Trendanalist hyca-bestanden (extensie '.hyc') inlezen van het programma HYCA, '.riw' uit de RIWA-Base en dia-bestanden (extensie '.dia', zie § 2.3.3) uit de database Donar van Rijkswaterstaat.

2.3.1 Vereiste formaat invoerbestand met extensie '.tai'

Het standaard-invoerbestand voor *Trendanalist* heeft de extensie '.tai', een afkorting van '**t**rend**a**nalist **i**nvoer'. Dit bestand moet op de volgende wijze zijn opgebouwd (zie ook de voorbeelden in de map '\Trendanalist\Data'):

meetlocatie *X-coordinaat Y-coordinaat* parameter datum *tijdstip* meetwaarde meeteenheid

Optioneel zijn daarin X-coordinaat, Y-coordinaat en tijdstip.

Voorbeeld van een standaard invoerbestand van Trendanalist, met als extensie '.tai'. De velden zijn hier gescheiden door een tab, maar mogen ook zijn gescheiden door een punt-komma (';').

```
Buitenpoort  O2-verzadiging  21/08/1980 09:45:00  46  %
Buitenpoort  O2-verzadiging  20/09/1980 09:40:00  32  %
Buitenpoort  O2-verzadiging  21/10/1980 09:49:00  41  %
Buitenpoort  O2-verzadiging  21/11/1980 09:41:00  52  %
Buitenpoort  O2-verzadiging  20/12/1980 09:46:00  26  %
```

Als ook ruimtelijke uitvoer via Google Earth is gewenst, dient de invoer te worden aangevuld met de X- en Y-Rijksdriehoekscoördinaten[4] (in meter): meetlocatie X-coörd Y-coörd parameter datum (en eventueel tijdstip) meetwaarde meeteenheid

Voorbeeld van een standaard invoerbestand van Trendanalist als ook uitvoer via Google Earth is gewenst. De X en Y-coördinaten zijn volgens het Rijksdriehoekscoördinatenstelsel (in meter).

```
Buitenpoort  157912 499867 O2-verzadiging  21/08/1980 09:45:00  46  %
Buitenpoort  157912 499867 O2-verzadiging  20/09/1980 09:40:00  32  %
Buitenpoort  157912 499867 O2-verzadiging  21/10/1980 09:49:00  41  %
Buitenpoort  157912 499867 O2-verzadiging  21/11/1980 09:41:00  52  %
Buitenpoort  157912 499867 O2-verzadiging  20/12/1980 09:46:00  26  %
```

Verdere opmaakspecificaties

1. Het bestand mag meerdere meetreeksen onder elkaar bevatten.
   De velden met achtereenvolgens meetlocatie (en eventueel X-coördinaat en Y-coördinaat), parameter, datum (en eventueel tijdstip), meetwaarde en meeteenheid moeten alle zijn gescheiden door een tab of door een punt-komma (';'). Het scheidingsteken (een tab of een ';') mag echter niet worden voorafgegaan door spaties, of worden gevolgd door spaties.
2. Als decimaal scheidingsteken van een meetwaarde worden zowel de punt ('.') als de komma (',') geaccepteerd. Maar scheidingstekens voor duizendtallen worden niet geaccepteerd.
3. Elke meetreeks moet zijn gesorteerd op datum (oplopend). De gebruiker kan er voor kiezen dat Trendanalist dit uitvoert via 'Voorbewerken meetreeksen' onder 'Opties' (zie § 4.9). De datum heeft als formaat dag/maand/jaar, waarbij alleen nummers mogen worden gebruikt (dus voor januari een '1' en geen 'jan'). Het scheidingsteken binnen de datum moet zijn '/' (voorbeeld: 26/1/2002), of '-' (voorbeeld: 26-1-2002). Het jaartal mag met 4 cijfers zijn (voorbeeld: 1999), of met 2 cijfers (voorbeeld: 99). Het tijdstip *hoeft niet* te zijn vermeld. Als dit wel is vermeld, moet dit met het formaat uur:minuut, waarbij elk uit 1 of 2 cijfers bestaat (voorbeelden: 13:21 en 7:5).
4. Als de meeteenheid eindigt met ('uur', 'dag', 'week', 'maand', 'kwartaal' of 'jaar') dan interpreteert Trendanalist de meetwaarden als zijnde vrachten, een product van meetwaarden en debieten.
5. Een meetwaarde kan ook worden voorafgegaan door een '<'-teken, maar dan zonder spatie. De meetwaarde wordt in dit geval door *Trendanalist* geïnterpreteerd als een gecensureerde waarde, die vervolgens voor verdere kwantitatieve verwerkingen standaard op de helft van de hoogste rapportagegrens wordt gezet (bij vrachten op nul), evenals eventuele niet-gecensureerde waarden die lager zijn dan de hoogste rapportagegrens. Zo wordt vermeden dat er

[4] De Rijksdriehoekscoördinaten geven de ligging ten opzichte van een punt op 120 km ten zuidoosten van Parijs.

trends worden gedetecteerd die louter het gevolg zijn een veranderende rapportagegrens. De gebruiker kan er overigens via 'Gecensureerde waarden' en 'Gecensureerde vrachten' onder 'Opties' (zie § 4.6) ook voor kiezen dat een gecensureerde waarde op de helft van zijn eigen rapportagegrens wordt gezet. Als u zelf al vóóraf gecensureerde waarden heeft vervangen door één of andere waarde, verstrek dan geen meetreeksen die oorspronkelijk 80% (bij vrachten 30%) of meer gecensureerde meetwaarden bevatten. Bij te veel gecensureerde meetwaarden zijn de resultaten van de trendanalyse namelijk zeer onbetrouwbaar en door de vooraf zelf uitgevoerde substitutie kan *Trendanalist* bij het inlezen niet meer checken op het percentage gecensureerde meetwaarden (de derde in § 2.2 vermelde eis aan de informatieinhoud van de meetreeks).

6. Het bestand mag bovenin ook commentaarregels bevatten. De eerste kolom van een commentaarregel moet echter het **%**-teken bevatten.
7. Er mogen meetwaarden ontbreken in het invoerbestand. Het betreffende veld moet dan leeg zijn.

Vervaardigen 'tai'-invoerbestand met Excel

Een tabblad van een Excel-bestand kan eenvoudig worden opgeslagen als een tekstbestand (via het menu 'Bestand' en dan 'Opslaan als'). Kies daarbij als standaarduitgang '.txt'. Er zijn drie mogelijkheden om dit te bewerkstelligen, namelijk 'Tekst (tab gescheiden)', 'Unicode tekst', of 'Tekst (MS-DOS)'. Gebruik hiervan alleen de eerste optie, dus 'Tekst (tab gescheiden)' en verander na het opslaan de uitgang '.txt' in 'tai'.

Let op dat er precisie verloren kan gaan bij het omzetten van een Excelbestand naar een tekstbestand. Een waarde in Excel kan immers meer decimalen bevatten dan weergegeven. Zorg er daarom voor dat er voldoende decimalen in Excel zijn weergegeven.

2.3.2 Vereiste formaat invoerbestand met extensie '.csv'

De applicatie iBever wordt niet meer door Rijkswaterstaat ondersteund. De applicatie iBever is opgevolgd door de applicatie Aquo-kit. *Trendanalist* kan zowel een csv-bestand uitgevoerd door *iBever* als uitgevoerd door Aquo-kit inlezen.

Een csv-bestand kan één of meer meetreeksen bevatten. Het heeft een kopregel met de kolomnamen en de daarop volgende regels bevatten dan de gegevens voor elke kolom. Zowel de kolomnamen als de gegevens zijn gescheiden door puntkomma's of door tabs. Trendanalist bepaalt op basis van de kolomnamen in de kopregel of het een iBever of een Aquokit csv-bestand betreft.

2.3.2.1 IBever csv-dataformaat

Trendanalist leest tien van de kolommen van het iBever csv-bestand in (zie onderstaande tabel), waarbij de volgorde van deze kolommen niet uitmaakt.

De kolommen van het iBever csv-bestand die Trendanalist inleest (hun volgorde is niet belangrijk).

Kolomnaam	Voorbeeld	Omschrijving
mpn_mpnident	BOVSS	Code meetlocatie
mps_domgwcod	P	Code parameter
mep_domgwcod	mg/l	Code meeteenheid
mwa_mwadtme, of anders *mwa_mwadtmb*	10-1-2003	Eind-, of anders begindatum

mrsinovs_domafkrt	<	Meetwaarde is gecensureerd
mwa_mwawrden	76	Meetwaarde
mpn_mrfxcoor	156912	X-coördinaat [m]
mpn_mrfycoor	499867	Y-coördinaat [m]
hoe_domgwcod	NVT	Hoedanigheid
mco_domgwcod	10	Compartiment

Opmerking aanpassing: Trendanalist doet niets met de indicatie kwaliteit meetwaarde code.

Trendanalist sorteert de ingelezen gegevens op meetlocatie, parameternaam, meeteenheid, hoedanigheid, compartiment en meetdatum voor het maken van tijdreeksen. Meetreeksen worden onderscheiden op basis van meetlocatie, parameternaam, meeteenheid, hoedanigheid en compartiment. Deze laatste twee kenmerken, hoedanigheid en compartiment, worden bij de uitvoer alleen vermeld als deze nodig zijn om de meetreeks te onderscheiden van een andere meetreeks van dezelfde combinatie van meetlocatie, parameternaam en meeteenheid.

Verdere opmaakspecificaties

1. Elke meetreeks moet oplopend zijn gesorteerd op datum. Als een gesorteerde meetreeks twee of meer gelijke waarden bevat met dezelfde datum (en indien vermeld ook hetzelfde tijdstip), dan wordt de eerste van deze waarden meegenomen. Trendanalist sorteert de meetreeks oplopend op meetlocatie, parameternaam, meeteenheid en datum.
2. Als het csv-bestand geen kolom met de einddatum bevat (*mwa_mwadtme*), dan zal *Trendanalist* voor de meetdatum uitgaan van de begindatum (*mwa_mwadtmb*).
3. Als de meeteenheid in de kolom *mep_domgwcod* eindigt met (‘uur’, ‘dag’, ‘week’, ‘maand’, ‘kwartaal’ of ‘jaar’) dan interpreteert Trendanalist de meetwaarden als zijnde vrachten, een product van een meetwaarden en debieten.
4. Als de kolom *mrsinovs_domafkrt* een ‘<’-teken bevat, dan wordt de meetwaarde door *Trendanalist* geïnterpreteerd als een gecensureerde waarde, die vervolgens voor verdere kwantitatieve verwerkingen standaard op de helft van de hoogste rapportagegrens wordt gezet (bij vrachten op nul), evenals eventuele niet-gecensureerde waarden die lager zijn dan de hoogste rapportagegrens. Zo wordt vermeden dat er trends worden gedetecteerd die louter het gevolg zijn van een veranderende rapportagegrens. De gebruiker kan er overigens via ‘Gecensureerde waarden’ en ‘Gecensureerde vrachten’ onder ‘Opties’ (zie § 4.6) ook voor kiezen dat een gecensureerde waarde op de helft van zijn eigen rapportagegrens wordt gezet.
5. Als de kolom *mrsinovs_domafkrt* ontbreekt in het csv-bestand, dan worden alle meetwaarden als niet gecensureerd beschouwd.

2.3.2.2 Aquo-kit csv-dataformaat

Trendanalist leest dertien van de kolommen van het aquo-kit csv-bestand in (zie onderstaande tabel), waarbij de volgorde van deze kolommen niet uitmaakt.

De kolommen van het Aquo-kit csv-bestand die Trendanalist inleest (hun volgorde is niet belangrijk).

Kolomnaam	Voorbeeld	Omschrijving
Meetpunt.identificatie	240049	Code meetlocatie
Typering.code		Code typering
Grootheid.code	CONCTTE	Code grootheid
Parameter.code	Pb	Code parameter
Eenheid.code	ug/l	Code meeteenheid
Begindatum	7-1-2020	Begindatum
Begintijd	12:42:00	Begintijd
Limietsymbool	<	Meetwaarde is gecensureerd
Numerieke waarde	1.5	Meetwaarde
GeometricPunt.X_RD	169231	X-coördinate [m]
GeometricPunt.Y_RD	377801	Y-coördinate [m]
Hoedanigheid.code	NVT	Hoedanigheid
Compartiment.code	OW	Compartiment

De 10 kolomnamen van iBever zijn compatibel met de kolomnamen van Aquo-kit. De drie nieuwe kolomnamen zijn de typeringscode, de grootheidscode en ook de begintijd. De typeringscode en grootheidscode kwamen niet voor in iBever en worden in dit concept toegevoegd om altijd een uniek onderscheid tussen waarnemingssoorten/meetreeksen te bewerkstelligen. De tijd is extra toegevoegd waarmee het mogelijk is om ook meetreeksen met een kortere tijdsduur dan een dag tussen metingen te verwerken.

Trendanalist sorteert de ingelezen gegevens op meetlocatie, parameternaam en meetdatum voor het maken van tijdreeksen. De parameternaam wordt bepaald door de kolommen: Typering.code, Grootheid.cod, Parameter.code, Hoedanigheid.code, de Eenheid.code en de Compartiment.code. De parameternaam is daarmee uniek voor elke waarnemingssoort/meetreeks van een monster. Een parameternaam in *Trendanalist* heeft de vorm parametercombinatiecode [eenheid][hoedanigheid][compartiment], waarbij de parametercombinatiecode wordt opgebouwd uit de kolommen Typering.code, Grootheid.code en Parameter.code. De parametercombinatiecode is Typering.code als die gegeven is anders de Parameter.code als die gegevens is en anders de Grootheid.code. In bovenstaande voorbeeld is de parameternaam Pb [ug/l][NVT][OW].

Verdere opmaakspecificaties

1) Elke meetreeks moet oplopend zijn gesorteerd op datum. Als een gesorteerde meetreeks twee of meer gelijke waarden bevat met dezelfde datum (en indien vermeld ook hetzelfde tijdstip), dan wordt de eerste van deze waarden meegenomen. Trendanalist sorteert de meetreeksen oplopend op meetlocatie, parameternaam, meeteenheid en datum.
2) Als de meeteenheid in de kolom *Eenheid.code* eindigt met ('uur', 'dag', 'week', 'maand', 'kwartaal' of 'jaar') dan interpreteert Trendanalist de meetwaarden als zijnde vrachten, een product van een meetwaarden en debieten.
3) Als de kolom *Limietsymbool* een '<'-teken bevat, dan wordt de meetwaarde door *Trendanalist* geïnterpreteerd als een gecensureerde waarde, die vervolgens voor verdere kwan-

titatieve verwerkingen standaard op de helft van de hoogste rapportagegrens wordt gezet (bij vrachten op nul), evenals eventuele niet-gecensureerde waarden die lager zijn dan de hoogste rapportagegrens. Zo wordt vermeden dat er trends worden gedetecteerd die louter het gevolg zijn van een veranderende rapportagegrens. De gebruiker kan er overigens via 'Gecensureerde waarden' en 'Gecensureerde vrachten' onder 'Opties' (zie § 4.6) ook voor kiezen dat een gecensureerde waarde op de helft van zijn eigen rapportagegrens wordt gezet. Als de kolom *Limietsymbool* ontbreekt in het csv-bestand, dan worden alle meetwaarden als niet gecensureerd beschouwd.

2.3.3 Vereiste formaat invoerbestand met extensie '.dia'

Dia-bestanden worden gebruikt voor het beschikbaar stellen van gegevens uit de database Donar, de centrale database voor de gegevens van de 'natte' Rijkswaterstaat.

De regels met codes die Trendanalist inleest van een dia-bestand.

Code	Voorbeeld	Omschrijving
Loc	Maassluis	Code meetlocatie, volledige naam
Loc;RD X	7675000	Rijksdriehoekstelsel X-coördinaat [m]
Loc;RD Y	43717000	Rijksdriehoekstelsel Y-coördinaat [m]
PAR	Naftaleen	Code parameter
dg	mg/kg	Code meeteenheid
HDH	dg	Hoedanigheid
BEM	VTIJD	Code typering
CPM	50	Compartiment
TYP	TN	Equidistante (TE) of non-equidistante reeks (TN)
TYD	20120118:1340;20121219;1110	Begin-en einddatum
[WRD]	<+.500E-01/0	Gecensureerde meetwaarde met kwaliteitscode 0

De parameternaam wordt bepaald door de codes: PAR, HDH, BEM en CPM. In het bovenstaande voorbeeld is de parameternaam: naftaleen [dg][VTIJD][50].

Verdere opmaakspecificaties

1) Elke meetreeks moet oplopend zijn gesorteerd op datum. Als een gesorteerde meetreeks twee of meer gelijke waarden bevat met dezelfde datum (en indien vermeld ook hetzelfde tijdstip), dan wordt de eerste van deze waarden meegenomen. Trendanalist sorteert de meetreeksen oplopend op meetlocatie, parameternaam, meeteenheid en datum.
2) Als de meeteenheid eindigt met ('uur', 'dag', 'week', 'maand', 'kwartaal' of 'jaar') dan interpreteert Trendanalist de meetwaarden als zijnde vrachten, een product van meetwaarden en debieten.
3) Als de meetwaarde een '<'-teken bevat, dan wordt de meetwaarde door *Trendanalist* geinterpreteerd als een gecensureerde waarde, die vervolgens voor verdere kwantitatieve verwerkingen standaard op de helft van de hoogste rapportagegrens wordt gezet (bij vrachten op nul), evenals eventuele niet-gecensureerde waarden die lager zijn dan de hoogste rapportagegrens. Zo wordt vermeden dat er trends worden gedetecteerd die louter het gevolg zijn van een veranderende rapportagegrens. De gebruiker kan er overigens via 'Gecensureerde waarden' en 'Gecensureerde vrachten' onder 'Opties' (zie § 4.6)

ook voor kiezen dat een gecensureerde waarde op de helft van zijn eigen rapportage-grens wordt gezet.

4) Is de kwalteitscode in een meetwaarde gelijk aan 99, dan wordt de meetwaarde als missend beoordeeld.
5) Trendanalist kan alleen meetreeksen verwerken waarbij de code TYP de waarde heeft TE (equidistante meetreeks) of TN (niet-equidistante meetreeks).

3 Het rekenhart van *Trendanalist*

Trendanalist is bedoeld om trendanalyse uit te voeren van meetreeksen, met onderscheid tussen trend*detectie* en trend*kwantificering*. Trenddetectie, uit te voeren aan de hand van een statistische toets op trend, levert een objectief antwoord op de vraag of er een trend bestaat (zie § 3.2), terwijl trendkwantificering een schatting oplevert van de grootte van een trend (zie § 3.6).

3.1 De inspiratiebronnen van *Trendanalist*

De werking van *Trendanalist* is gebaseerd op een synthese van twee eerder ontwikkelde procedures voor trendanalyse. De eerste is een procedure ontwikkeld in opdracht van de RIWA en neergelegd in het programma Trendpak [Baggelaar en Baggelaar, 1989 en 1991]. Dit betreft echter een MS-DOS-programma, dat een zeer omvangrijke en ook zeer specialistische pré- en postprocessing vergt. De tweede is een procedure voor trendanalyse van de lozing van verontreinigingen naar de Noordzee, ontwikkeld in opdracht van het Rijksinstituut voor Kust en Zee / RIKZ [Baggelaar, 1994]. Wij hebben de kernprincipes van beide procedures geprogrammeerd in een Matlab-omgeving, zodanig dat de interactie eenvoudig wordt en het programma ook in staat is tot een volledig automatische trendanalyse van grote aantallen meetreeksen.

3.2 Trenddetectie met *Trendanalist*

Voor wat betreft de *trenddetectie* kan *Trendanalist* voor elke afzonderlijke meetreeks statistisch toetsen of deze een *monotone* trend bevat. Een monotone trend is een verandering die doorgaans dezelfde kant opgaat, dat wil zeggen opwaarts of neerwaarts, ongeacht of dit lineair, stapsgewijs of in een andere specifieke vorm geschiedt. Het hypothetische *startpunt* van de trend kan vooraf door de gebruiker worden opgegeven. Als dat startpunt niet wordt opgegeven, zal standaard worden getoetst op trend vanaf het begin van de reeks. Er kan naar keuze *eenzijdig* of *tweezijdig* worden getoetst op trends. Eenzijdig toetsen is van toepassing als men specifiek wil toetsen op hetzij een dalende, hetzij een stijgende trend, terwijl tweezijdig toetsen van toepassing is als het teken van de trend er niet toe doet.

Maatwerk

Er bestaan velerlei statistische toetsen op monotone trend in meetreeksen. De bekendste daarvan is de lineaire regressietoets. Deze klassieke toets gaat uit van onderling onafhankelijke waarden, afkomstig uit een normale kansverdeling. Maar chronologisch gerangschikte waarden van milieuparameters zijn zelden onderling onafhankelijk te noemen, vanwege seizoensinvloeden en autocorrelatie. Verder blijken dergelijke waarden ook vaak een niet-symmetrische frequentieverdeling te vertonen [Hirsch et al. 1982, McLeod et al., 1991], doorgaans met een langere rechterstaart (positieve scheefheid). Aanvankelijk werd wel aanbevolen een dergelijke reeks om te zetten in een reeks van jaargemiddelden, omdat het middelen doorgaans leidt tot onderling onafhankelijke waarden, afkomstig uit een normale kansverdeling, zodat dan toch op trend kan worden getoetst met lineaire regressie (zie bijvoorbeeld [Sanders et al., 1983]). De laatste drie decaden zijn er echter verscheidene andere statistische toetsen op trend ontwikkeld, die wél rekening kunnen houden met de speciale karakteristieken van meetreeksen van milieuparameters. In verschillende simulatiestudies is al aangetoond dat het toepassen van deze toetsen op de oorspronkelijke reeks

doorgaans meer onderscheidend vermogen[5] oplevert dan het toepassen van de lineaire regressietoets op de reeks van jaargemiddelden [Baggelaar en Baggelaar, 1991; Harcum et al., 1992]. Gezien de huidige grote verscheidenheid aan statistische toetsen op trend, elk met specifieke mogelijkheden en onderliggende vooronderstellingen, is er veel expertise vereist om de meest geschikte toets te kunnen selecteren. Mede daarom wordt het gebruik van statistiek voor een evaluatie van veranderingen in het milieu wel net zo zeer een kunst als een wetenschap genoemd [Ward and Loftis, 1986].

Trendanalist bevat meerdere trendtoetsen, zodat het maatwerk kan leveren, met een toets die aansluit op de statistische karakteristieken van de meetreeks. Alleen op deze wijze kan optimaal rekening worden gehouden met het soort kansverdeling waar de waarden uit afkomstig zijn (normaal of niet-normaal) en met het al of niet optreden van seizoeneffecten en/of autocorrelatie. Immers, als bij trendanalyse van meerdere meetreeksen slechts één trendtoets wordt gehanteerd, zal bij een deel van de reeksen verlies aan onderscheidend vermogen optreden en zal bij een ander deel onterecht een trend worden gedetecteerd.

De automaat van *Trendanalist* zal voor elke meetreeks de trendtoets hanteren die het best aansluit op de karakteristieken van die reeks. De daartoe ter beschikking staande trendtoetsen zijn in te delen in twee groepen, namelijk:

1. de *lineaire regressietoets* en zijn uitbreidingen;
2. de *Mann-Kendall-toets* en zijn uitbreidingen.

Een toets uit de eerste groep is van toepassing als de waarden van de meetreeks afkomstig zijn uit een normale kansverdeling. Als dat niet het geval is, is een toets uit de tweede groep van toepassing. Voor beide groepen zijn de uitbreidingen bedoeld om ook rekening te kunnen houden met seizoenseffecten en/of autocorrelatie.

In het nu volgende worden de verschillende trendtoetsen toegelicht en wordt ook aangegeven hoe de automaat van *Trendanalist* tot de selectie van de juiste keuzetoets komt.

3.3 Lineaire regressietoets en zijn uitbreidingen

3.3.1 Lineaire regressietoets

Als de waarden van de beschouwde variabele afkomstig zijn uit een normale verdeling en geen seizoenseffecten en geen autocorrelatie vertonen, dan kan de lineaire regressietoets worden toegepast. Het daarbij gehanteerde model luidt als volgt:

$$Z_t = b_0 + b_1 \cdot X_t + e_t$$

met Z de waarden van de beschouwde variabele, b_0 het geschatte intercept, b_1 de geschatte lineaire helling, X de tijd [jaar, inclusief jaarfractie], t de tijdindex en tenslotte e het modelresidu. De toetsingsgrootheid T wordt berekend als:

$$T = \frac{b_1}{stft[b_1]}$$

[5] Het onderscheidend vermogen is de kans om een werkelijk bestaande trend te detecteren. Dit vermogen neemt toe met de grootte van de trend.

met *stft*[b_1] de standaardfout van de schatter van de lineaire helling. Als de modelresiduën (e_1...e_n) afkomstig zijn uit dezelfde normale kansverdeling en tevens onafhankelijk van elkaar zijn, mogen we met 100%·(1-α) betrouwbaarheid spreken van een trend, als geldt:

$$|T| > t_{(1-\alpha/2,n-2)} \text{ bij tweezijdig toetsen}$$
$$T > t_{(1-\alpha,n-2)} \text{ bij eenzijdig toetsen op een positieve trend}$$
$$T < t_{(\alpha,n-2)} \text{ bij eenzijdig toetsen op een negatieve trend}$$

met $t_{(1-\alpha/2,n-2)}$, $t_{(1-\alpha,n-2)}$ en $t_{(\alpha,n-2)}$ de Student-t-waarden met een eenzijdige onderschrijdingskans van respectievelijk 1-α/2, 1- α en α bij *n*-2 vrijheidsgraden (*n* is hier het aantal waarden in de reeks).

3.3.2 Lineaire regressietoets met verdiscontering seizoenseffecten en/of autocorrelatie

Als de waarden van de beschouwde variabele afkomstig zijn uit een normale verdeling en tevens seizoenseffecten en/of autocorrelatie vertonen, dan dient daarmee rekening te worden gehouden. Er moet dan worden getoetst op trend met uitgebreidere vormen van het lineaire regressiemodel.

Verdiscontering seizoenseffecten

Als er alleen sprake is van seizoenseffecten kan worden getoetst op trend met het volgende model:

$$Z_t = b_0 + b_1 \cdot X_t + \sum_{i=2}^{s} (\gamma_i \cdot I_i) + e_t$$

met *Z* de waarden van de beschouwde variabele, b_0 het geschatte intercept, b_1 de geschatte lineaire helling, *X* de tijd [jaar, inclusief jaarfractie], *t* de tijdsindex, *s* het aantal waarden per jaar (te interpreteren als het aantal seizoenen), *i* de seizoensindex, *γ* de seizoensinvloed, *I* de seizoensindicator (deze is 1 als X_t in seizoen *i* valt en anders 0) en tenslotte *e* het modelresidu.
Het eerste seizoen is niet opgenomen in dit uitgebreide model, zodat zijn (eventuele) effect zal zijn gevat in de schatting van het intercept (b_0). Een relevant seizoenseffect zal daardoor tot uiting komen in een statistisch significante invloed (γ_i) van minstens één van de andere seizoenen. De geschatte grootte van deze invloed vertegenwoordigt dan het verschil tussen het seizoenseffect van seizoen *i* en dat van het eerste seizoen. Om te kunnen toetsen op trend aan de hand van dit uitgebreide lineaire regressiemodel, moeten de modelresiduën (e_1...e_n) afkomstig zijn uit dezelfde normale kansverdeling en tevens onafhankelijk van elkaar zijn. Het toetsen gaat dan als beschreven in § 3.3.1, zij het dat het aantal vrijheidsgraden niet meer gelijk is aan *n*-2, maar aan *n*-*k*, waarbij *k* het aantal gehanteerde modelparameters is.

Verdiscontering autocorrelatie

Als er alleen sprake is van autocorrelatie kan worden getoetst op trend met het volgende model:

$$Z_t = b_0 + b_1 \cdot X_t + N_t$$
$$N_t = \phi_1 \cdot N_{t-1} + e_t$$

met *Z* de waarden van de beschouwde variabele, b_0 het geschatte intercept, b_1 de geschatte lineaire helling, *X* de tijd [jaar, inclusief jaarfractie], *t* de tijdsindex en tenslotte *N*, de ruis van het model. De autocorrelatie van de ruis wordt beschreven door het tweede deel van het model, met *e*

het modelresidu en ϕ_1 de 1e-orde autoregressieve modelparameter [Box and Jenkins, 1976]. Om te kunnen toetsen op trend aan de hand van dit uitgebreide lineaire regressiemodel, moeten de modelresiduën ($e_1...e_n$) afkomstig zijn uit dezelfde normale kansverdeling en tevens onafhankelijk van elkaar zijn. Het toetsen gaat dan als beschreven in § 3.3.1, zij het dat het aantal vrijheidsgraden niet meer gelijk is aan *n*-2, maar aan *n*-*k*, waarbij *k* het aantal gehanteerde modelparameters is.

Verdiscontering seizoenseffecten en autocorrelatie
Als er zowel sprake is van seizoenseffecten als van autocorrelatie kan worden getoetst op trend met het volgende model:

$$Z_t = b_0 + b_1 \cdot X_t + \sum_{i=2}^{s} (\gamma_i \cdot I_i) + N_t$$
$$N_t = \phi_1 \cdot N_{t-1} + e_t$$

met *Z* de waarden van de beschouwde variabele, b_0 het geschatte intercept, b_1 de geschatte lineaire helling, *X* de tijd [jaar, inclusief jaarfractie], *t* de tijdsindex, *s* het aantal waarden per jaar (te interpreteren als het aantal seizoenen), *i* de seizoensindex, *γ* de seizoensinvloed, *I* de seizoensindicator (deze is 1 als T_t in seizoen *i* valt en anders 0), *N* de ruis van het model, *e* het modelresidu en ϕ_1 de 1e-orde autoregressieve modelparameter [Box and Jenkins, 1976]. Een dergelijk model is al eerder voorgesteld door [Duin, 1993].
Om te kunnen toetsen op trend aan de hand van dit uitgebreide lineaire regressiemodel, moeten de modelresiduën ($e_1...e_n$) afkomstig zijn uit dezelfde normale kansverdeling en tevens onafhankelijk van elkaar zijn. Het toetsen gaat dan als beschreven in § 3.3.1, zij het dat het aantal vrijheidsgraden niet meer gelijk is aan *n*-2, maar aan *n*-*k*, waarbij *k* het aantal gehanteerde modelparameters is.

3.4 Mann-Kendall-toets en zijn uitbreidingen

3.4.1 Mann-Kendall-toets

De Mann-Kendall-toets op trend is verdelingsvrij en stelt dus geen voorwaarden aan het soort kansverdeling waar de waarden uit afkomstig zijn. De vooronderstellingen zijn dat de waarden nóch seizoeneffecten, nóch autocorrelatie vertonen. De toetsingsgrootheid *S* wordt berekend als [Mann, 1945, Kendall, 1938 en 1975]:

$$S = \sum_{t=1}^{n-1} \sum_{k=t+1}^{n} \mathrm{sgn}[Z_k - Z_t]$$

met *Z* de waarden van de beschouwde variabele, *t* en *k* waarden van de tijdsindex, *n* het aantal waarden en sgn[ ] de signum-functie, volgens:

$$\mathrm{sgn}[\theta] = 1 \quad \text{als} \quad \theta > 0$$
$$\mathrm{sgn}[\theta] = 0 \quad \text{als} \quad \theta = 0$$
$$\mathrm{sgn}[\theta] = -1 \quad \text{als} \quad \theta < 0$$

De nulhypothese luidt dat de waarden $Z_1...Z_n$ afkomstig zijn uit *n* onderling onafhankelijke en identieke kansverdelingen en dat er dus geen trend bestaat. Voor n ≥ 5 en $n \leq 10$ is er een tabel be-

schikbaar met kritische waarden voor *S* onder de nulhypothese. Voor *n* > 10 zal *S* onder de nulhypothese een asymptotisch normale verdeling volgen en kan de volgende benadering worden gehanteerd. Onder de nulhypothese geldt voor de verwachtingswaarde en de variantie van *S* respectievelijk:

$$\mathrm{E}[S]=0 \quad \text{en} \quad \mathrm{Var}[S]=\frac{n\cdot(n-1)\cdot(2n+5)}{18}$$

Hieruit kunnen we, mede met behulp van een continuïteitscorrectie, de gestandaardiseerde toetsingsgrootheid *T* berekenen, als:

$$T=\frac{S-1}{\sqrt{Var[S]}} \quad \text{als} \quad S>0$$
$$T=0 \quad \text{als} \quad S=0$$
$$T=\frac{S+1}{\sqrt{Var[S]}} \quad \text{als} \quad S<0$$

Onder de nulhypothese ('géén trend') volgt *T* een standaardnormale kansverdeling. We mogen met 100%·(1-α) betrouwbaarheid spreken van een trend, als geldt:

$$|T|>U_{(1-\alpha/2)} \text{ bij tweezijdig toetsen}$$
$$T>U_{(1-\alpha)} \text{ bij eenzijdig toetsen op een positieve trend}$$
$$T<U_{(\alpha)} \text{ bij eenzijdig toetsen op een negatieve trend}$$

met $U_{(1-\alpha/2)}$, $U_{(1-\alpha)}$ en $U_{(\alpha)}$ de waarde van de standaardnormale kansverdeling met een eenzijdige onderschrijdingskans van respectievelijk 1-α/2, 1-α en α.

Correctie voor gelijke waarden

Als er gelijke waarden[6] voorkomen, zal *Trendanalist* daarvoor corrigeren bij het berekenen van de variantie van *S*. Gelijke waarden kunnen bijvoorbeeld voorkomen als er meerdere gecensureerde waarden in een reeks voorkomen. *Trendanalist* zet elke gecensureerde waarde standaard namelijk op de helft van de hoogste rapportagegrens (bij vrachten op nul) en als er enkele gecensureerde waarden voorkomen ontstaan er dus enkele gelijke waarden.

3.4.2 Mann-Kendall-toets met verdiscontering van autocorrelatie

De Mann-Kendall-toets met verdiscontering van autocorrelatie is een verdere uitbreiding van de Mann-Kendall-toets voor het geval een reeks, waarvan de waarden niet afkomstig zijn uit een normale kansverdeling, ook autocorrelatie vertoont [Yue et al. , 2002]. De autocorrelatie in een reeks verandert de variantie van de Mann-Kendall toetsingsgrootheid *S.* Bij een positieve autocorrelatie neemt de kans toe dat de Mann-Kendall-toets een significante trend detecteert en bij een negatieve autocorrelatie neemt die kans af. Meestal is er sprake van een positieve autocorrelatie bij milieuparameters.

De toepassing van deze toets vereist zes stappen:

1. Bepaal de lineaire helling van de waarden van de beschouwde variabele.
2. Indien deze helling verschilt van nul, verwijder dan de trend uit de waarden.
3. Bepaal de eerste orde autocorrelatiecoëfficiënt van de waarden zonder de trend.

[6] In het Engels worden deze aangeduid als 'ties', of 'tied data'.

4. Verwijder het autoregressieve proces uit de waarden.
5. Voeg aan deze geprewhitened waarden de trend toe.
6. Pas de Mann-Kendall-toets toe.

Bij stap 1 wordt de helling door Trendanalist geschat met de Theilhelling (zie § 3.6.1). In stap 2 wordt de trend als volgt verwijderd:

$$Z_t' = Z_t - b_1 \cdot X_t$$

met Z' zijn de trendverwijderde waarden, *Z* de waarden van de beschouwde variabele, b_1 de geschatte lineaire Theilhelling, *X* de tijd [jaar, inclusief jaarfractie] en *t* de tijdsindex. In stap 3 wordt de autocorrelatiecoëfficiënt geschat van de trendverwijderde waarden met de Spearman-correlatiecoëfficiënt. De Spearman-correlatiecoëfficiënt is gebaseerd op rangnummers van de waarden en is daarmee een verdelingsvrije schatting van de autocorrelatiecoëfficiënt. In stap 4 wordt de autocorrelatie als volgt verwijderd:

$$Y_t' = Z_t' - r_1 \cdot Z_t'$$

met Y' zijn de prewhitened trendverwijderde waarden, *Z'* zijn de trendverwijderde tijdsindex of tijdsindexwaarden van de beschouwde variabele, r_1 de geschatte autocorrelatiecoëfficiënt en *t* de tijdsindex. In stap 5 wordt de trend, die in stap 2 was verwijderd, toegevoegd:

$$Y_t = Y_t' + b_1 \cdot X_t$$

met Y zijn de prewhitened waarden, Y' zijn de prewhitened trendverwijderde waarden van stap 4, b_1 de geschatte lineaire Theilhelling van stap 2, *X* de tijd [jaar, inclusief jaarfractie] en *t* de tijdsindex. In laatste stap wordt de Mann-Kendal-toets (zie § 3.4.1) toegepast op de prewhitened waarden Y van de oorspronkelijke waarden Z.

3.4.3 Seizoenale Mann-Kendall-toets

De seizoenale Mann-Kendall-toets is een multivariate uitbreiding van de Mann-Kendall-toets voor het geval dat een reeks, waarvan de waarden niet afkomstig zijn uit een normale kansverdeling, seizoenseffecten vertoont [Hirsch et al., 1982]. De nulhypothese van de Mann-Kendall-toets, dat de waarden $Z_1 \ldots Z_n$ afkomstig zijn uit *n* onderling onafhankelijke en identieke kansverdelingen, is dan immers te beperkt. Want door de seizoenseffecten zullen de verdelingen afhangen van de tijd van het jaar en dus niet meer identiek zijn. De multivariate uitbreiding houdt in dat voor elk afzonderlijk seizoen een toetsingsgrootheid S_g wordt berekend, volgens:

$$S_g = \sum_{i=1}^{n_g-1} \sum_{k=i+1}^{n_g} \mathrm{sgn}[Z_{kg} - Z_{ig}]$$

met *g* de index voor het seizoen, n_g het aantal jaren met een waarde in seizoen *g* en *i* en *k* indices voor het jaar. Vervolgens wordt door sommatie van alle S_g-waarden de toetsingsgrootheid *S** verkregen:

$$S^* = \sum_{g=1}^{s} S_g$$

met *s* het aantal seizoenen in een jaar. De nulhypothese luidt nu dat de waarden van de reeks afkomstig zijn uit *n* onderling onafhankelijke kansverdelingen, die voor een bepaald seizoen tevens identiek zijn. En de alternatieve hypothese luidt dat voor één of meer seizoenen de kansverdelingen niet identiek zijn, wat neerkomt op een monotone trend in één of meer seizoenen. Onder de nulhypothese geldt voor de verwachtingswaarde van *S**:

$$\mathrm{E}[S^*] = \sum_{g=1}^{s} \mathrm{E}[S_g] = 0$$

en voor de variantie van *S**:

$$\mathrm{Var}[S^*] = \sum_{g=1}^{s} \mathrm{Var}[S_g] + \sum_{g=1}^{s} \sum_{h \neq g}^{s} \mathrm{Cov}[S_g, S_h]$$

Vanwege de onderlinge onafhankelijkheid geldt Cov[S_g,S_h]=0, waardoor bovenstaande formule reduceert tot:

$$\mathrm{Var}[S^*] = \sum_{g=1}^{s} \mathrm{Var}[S_g] = \sum_{g=1}^{s} \frac{n_g \cdot (n_g - 1) \cdot (2n_g + 5)}{18}$$

Als het product van het aantal seizoenen (met beschikbare waarden) en het aantal jaren minstens 20 bedraagt, zal *S** onder de nulhypothese een asymptotisch normale verdeling volgen en kan de volgende benadering worden gehanteerd. Uit *S** kan dan, mede met behulp van een continuïteitscorrectie, de gestandaardiseerde toetsingsgrootheid *T* worden berekend, als:

$$T = \frac{S^* - 1}{\sqrt{Var[S^*]}} \quad \text{als} \quad S^* > 0$$

$$T = 0 \quad \text{als} \quad S^* = 0$$

$$T = \frac{S^* + 1}{\sqrt{Var[S^*]}} \quad \text{als} \quad S^* < 0$$

Onder de nulhypothese (‘géén trend’) volgt *T* een standaardnormale kansverdeling. We mogen met 100%·(1-α) betrouwbaarheid spreken van een trend, als geldt:

$$|T| > U_{(1-\alpha/2)} \text{ bij tweezijdig toetsen}$$
$$T > U_{(1-\alpha)} \text{ bij eenzijdig toetsen op een positieve trend}$$
$$T < U_{(\alpha)} \text{ bij eenzijdig toetsen op een negatieve trend}$$

met $U_{(p)}$ de waarde van de standaardnormale kansverdeling met eenzijdige onderschrijdingskans *p*.

Correcties voor ontbrekende waarden en voor gelijke waarden
Trendanalist zal zowel in het geval van ontbrekende waarden als in het geval van gelijke waarden specifieke correcties uitvoeren bij de berekeningen.

3.4.4 Seizoenale Mann-Kendall-toets met verdiscontering van autocorrelatie

De seizoenale Mann-Kendall-toets met verdiscontering van autocorrelatie is een verdere uitbreiding van de seizoenale Mann-Kendall-toets voor het geval dat een reeks, waarvan de waarden niet afkomstig zijn uit een normale kansverdeling, behalve seizoenseffecten ook autocorrelatie vertoont [Hirsch and Slack, 1984]. Vanwege de autocorrelatie geldt nu niet Cov[S_g,S_h]=0, zodat deze bij het schatten van de variantie van *S** wel degelijk een rol speelt, volgens:

$$\mathrm{Var}[S^*] = \sum_{g=1}^{s} \mathrm{Var}[S_g] + \sum_{g=1}^{s} \sum_{h \neq g}^{s} \mathrm{Cov}[S_g, S_h]$$

Als het product van het aantal seizoenen (met beschikbare waarden) en het aantal jaren minstens 20 bedraagt, zal *S** onder de nulhypothese een asymptotisch normale verdeling volgen en kan worden getoetst met de gestandaardiseerde toetsingsgrootheid *T*, zoals beschreven in § 3.4.2.

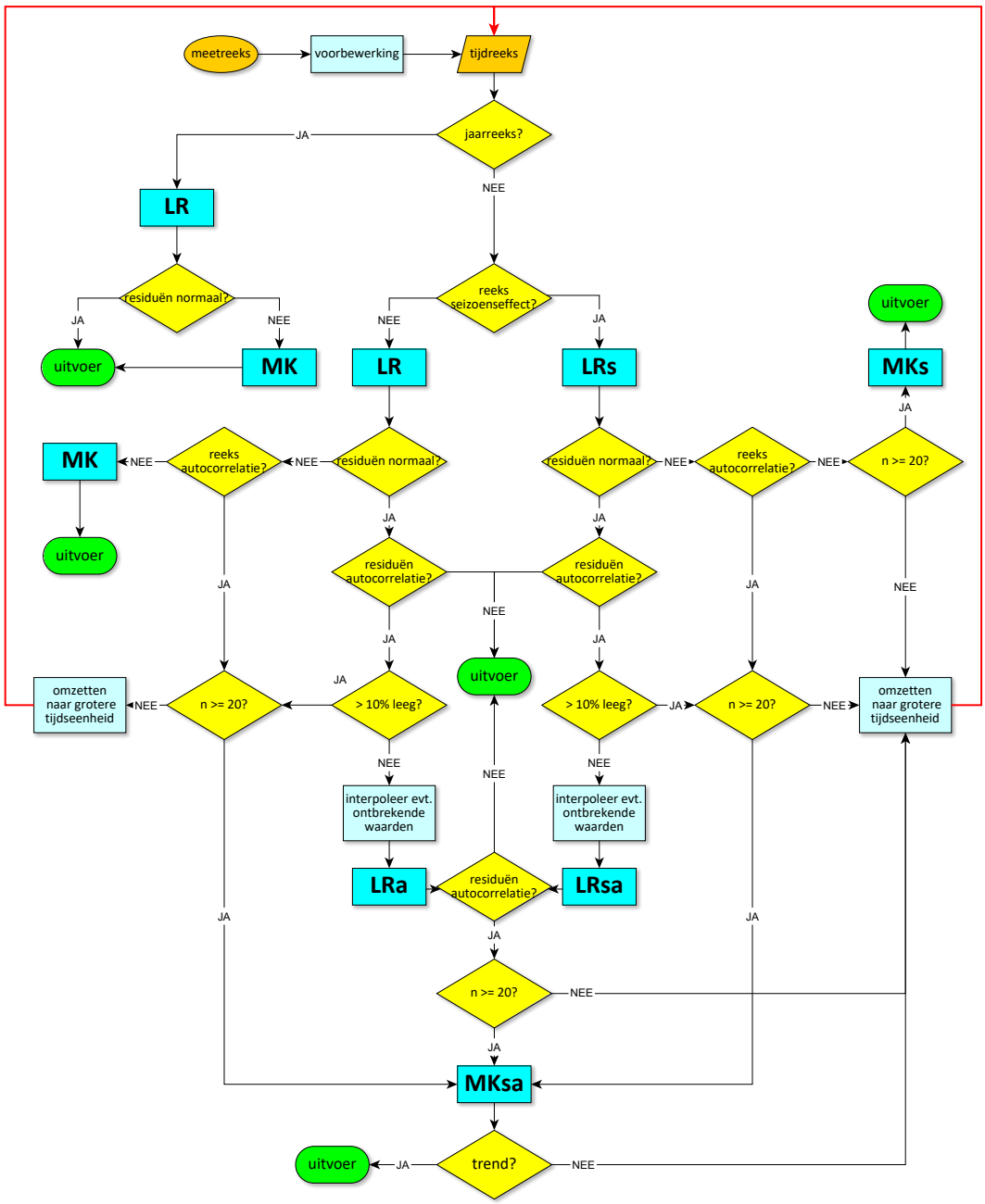
Om ervan uit te kunnen gaan dat het empirische significantieniveau in de buurt ligt van het nominale significantieniveau, dient het product van het aantal seizoenen (met beschikbare waarden) en het aantal jaren voldoende groot te zijn (zoals minstens 20 à 30).

3.5 Automatische selectieprocedure van *Trendanalist*

3.5.1 Stroomschema van de selectieprocedure

Om te bewerkstelligen dat voor elke meetreeks de meest geschikte trendtoets wordt gehanteerd, volgt de automaat van *Trendanalist* een selectieprocedure (zie figuur 3.1).

Figuur 3.1: De selectieprocedure die de automaat van Trendanalist voor elke meetreeks volgt.



De verschillende onderdelen van de selectieprocedure zijn onderscheiden door vormen en kleuren:

| oranje ellips: invoer | oranje parallellogram: bewerkte invoer | lichtblauwe rechthoek: actie | gele ruit: beslismoment | blauwe rechthoek: trendtoets | groene ellips: uitvoer |
Trendtoetsen: | **LR**: lineaire regressie | **MK**: Mann-Kendall-toets | **s**: met verdiscontering seizoenseffecten | **a**: met verdiscontering autocorrelatie | **sa**: met verdiscontering seizoenseffecten en autocorrelatie |

Hoofdlijnen van de selectieprocedure

Eerst wordt de meetreeks voorbewerkt en vervolgens omgezet tot een tijdreeks (§ 3.5.2) .
Als het een jaarreeks betreft wordt de lineaire regressietoets (LR) toegepast, maar het resultaat daarvan wordt alleen gepresenteerd als op basis van toetsing mag worden aangenomen dat de modelresiduën afkomstig zijn uit een normale kansverdeling (§ 3.5.4), anders wordt de Mann-Kendall-toets (MK) toegepast. Er wordt hierbij aangenomen dat autocorrelatie op jaarbasis geen of een verwaarloosbare rol speelt.
Als het geen jaarreeks betreft, wordt eerst getoetst of de reeks seizoenseffecten vertoont (§ 3.5.3). Als er geen sprake is van seizoenseffecten wordt de lineaire regressietoets (LR) toegepast, anders wordt de lineaire regressietoets met verdiscontering van seizoenseffecten (LRs) toegepast. In beide gevallen wordt het toetsresultaat alleen gepresenteerd als op basis van statistische toetsing mag worden aangenomen dat de modelresiduën afkomstig zijn uit een normale kansverdeling (§ 3.5.4) en tevens geen autocorrelatie vertonen (§ 3.5.5). Anders wordt één van de volgende paden gevolgd:

- Modelresiduën zijn niet normaal verdeeld en uit toetsing blijkt dat de tijdreeks geen autocorrelatie vertoont (§ 3.5.6) – Er wordt dan de bijpassende verdelingsvrije toets toegepast, namelijk de Mann-Kendall-toets (MK), of de Mann-Kendall-toets met verdiscontering van seizoenseffecten (MKs), zij het die laatste alleen als de tijdreeks minstens 20 waarden bevat, anders wordt deze omgezet naar een grotere tijdseenheid en wordt de selectieprocedure opnieuw doorlopen.
- Modelresiduën zijn niet normaal verdeeld en uit toetsing blijkt dat de tijdreeks autocorrelatie vertoont (§ 3.5.6) – Als de tijdreeks minstens 20 waarden bevat, wordt de Mann-Kendall-toets met verdiscontering van seizoenseffecten en autocorrelatie (MKsa) toegepast, anders wordt deze omgezet naar een grotere tijdseenheid en wordt de selectieprocedure opnieuw doorlopen.
- Modelresiduën zijn normaal verdeeld, maar ze vertonen autocorrelatie – Als de tijdreeks meer dan 10% lege tijdseenheden omvat en minstens 20 waarden bevat, wordt de Mann-Kendall-toets met verdiscontering van seizoenseffecten en autocorrelatie (MKsa) toegepast. Als de tijdreeks meer dan 10% lege tijdseenheden omvat en minder dan 20 waarden bevat, wordt deze omgezet naar een grotere tijdseenheid en wordt de selectieprocedure opnieuw doorlopen. Maar als de tijdreeks hooguit 10% lege tijdseenheden omvat, worden eventueel ontbrekende waarden geïnterpoleerd[7] en wordt de lineaire regressietoets met verdiscontering van autocorrelatie (LRa), of de lineaire regressietoets met verdiscontering van seizoenseffecten en autocorrelatie (LRsa) toegepast. Als uit toetsing blijkt dat de modelresiduën dan nóg autocorrelatie vertonen (§ 3.5.5), wordt de Mann-Kendall-toets met verdiscontering van seizoenseffecten en autocorrelatie (MKsa) toegepast, mits de tijdreeks minstens 20 waarden bevat, anders wordt deze omgezet naar een grotere tijdseenheid en wordt de selectieprocedure opnieuw doorlopen. Een resterende autocorrelatie kan immers zijn veroorzaakt door een niet-lineaire trend en de MKsa-toets kan daarmee overweg.

[7] Het interpoleren van ontbrekende waarden gebeurt met behulp van het voor die tijdreeks van toepassing zijnde lineaire regressiemodel. Dit wordt ook wel aangeduid als imputeren.

Voor een aantal tijdreeksen zal bij het doorlopen van de selectieprocedure de Mann-Kendall-toets met verdiscontering van seizoenseffecten en autocorrelatie (MKsa) de meest geschikte trendtoets blijken. Maar als die toets geen statistisch significante trend (of stijging of daling) detecteert, zal *Trendanalist* de tijdreeks omzetten naar de eerstvolgende grotere tijdseenheid en de selectieprocedure opnieuw doorlopen. Het is namelijk gebleken dat het onderscheidend vermogen van deze toets in bepaalde gevallen dermate gering is, dat het loont de tijdreeks om te zetten naar een grotere tijdseenheid, aangezien dan wellicht een andere, krachtigere trendtoets kan worden toegepast [Baggelaar en Baggelaar, 1991]. *Trendanalist* zal voor een meetreeks het resultaat presenteren van de laatste trendtoets die bij het doorlopen van de selectieprocedure is toegepast.

De onderdelen van de selectieprocedure worden in het volgende toegelicht.

3.5.2 Voorbewerking meetreeks

Meetreeksen van milieuvariabelen zijn zelden direct geschikt voor trendanalyse, vanwege uitschieters, tussentijdse veranderingen in meetfrequentie, ontbrekende waarden, of gecensureerde waarden. *Trendanalist* zal daarom eerst elke meetreeks controleren op dergelijke ongewenste karakteristieken en vervolgens zonodig aanpassingen aanbrengen. In het nu volgende is aangegeven hoe dat geschiedt. Eerst wordt echter toegelicht met welke tijdseenheden *Trendanalist* de trendanalyse uitvoert.

Gehanteerde tijdseenheid bij de trendanalyse

Trendanalist voert alleen trendanalyses uit op tijdreeksen met één van de zeven in tabel 3.1 vermelde tijdseenheden.

Tabel 3.1: De tijdseenheden waarop Trendanalist trendanalyse kan uitvoeren.

Tijdseenheid	Aantal waarden per jaar
Vier weken	13
Maand	12
Twee maanden	6
Kwartaal	4
Vier maanden	3
Zes maanden	2
Jaar	1

De redenen waarom geen trendanalyse wordt uitgevoerd op tijdreeksen met een kortere tijdseenheid dan vier weken zijn:

1. eventuele seizoenseffecten zijn dan moeilijker te herkennen;
2. de autocorrelatie kan dan zulke grote vormen aannemen, dat elke trendtoets daardoor onbruikbaar wordt.

Bij de voorbewerking wordt elke meetreeks omgezet in een tijdreeks met één van de in tabel 3.1 vermelde tijdseenheden. De tijdseenheid die het best past bij een meetreeks is de eerste tijdseenheid die een tijdreeks oplevert met minder dan 30% ontbrekende waarden, beginnend bij de tijdseenheid die het dichtst bij het meest voorkomende meetinterval in de meetreeks ligt. Als de tijd-

reeks 30% of meer ontbrekende waarden bevat, wordt een grotere tijdseenheid genomen, totdat de reeks wél minder dan 30% ontbrekende waarden bevat.
Een speciaal geval is de keuze tussen de tijdseenheid Vier weken en de tijdseenheid Maand, aangezien die zeer dicht bij elkaar liggen. De tijdseenheid Vier weken wordt pas verkozen boven de tijdseenheid Maand als dit geen groter percentage ontbrekende waarden oplevert. Zo wordt vermeden dat er bij het omzetten van een meetreeks onnodig ontbrekende waarden in de resulterende tijdreeks worden gecreëerd.

Het omzetten van een meetreeks naar een tijdreeks geschiedt door alle waarden van de meetreeks die binnen de tijdseenheid van de tijdreeks vallen te vervangen door hun *mediaan*. Dit principe wordt ook gehanteerd om een tijdreeks om te zetten naar een tijdreeks met een grotere tijdseenheid. Maar als de gebruiker wil dat hiervoor het rekenkundig gemiddelde wordt gehanteerd in plaats van de mediaan, kan dit worden aangepast (via 'Opties' en dan 'Middelingsmethode'). De aanpassing blijft echter alleen gelden voor de betreffende sessie.

Speciale criteria voor het omzetten naar een tijdreeks op jaarbasis
Als enige uitzondering mag bij de tijdseenheid Jaar wél 30% of meer van de jaren leeg zijn, op voorwaarde dat er wel minstens voor drie van de kalenderjaren een jaarwaarde is. De criteria die worden gehanteerd voor een jaarwaarde zijn:

a. als er in een kalenderjaar minstens twee meetwaarden voorkomen, dan is hun mediaan (of hun rekenkundig gemiddelde, als die optie door de gebruiker is geselecteerd) de jaarwaarde;
b. als er in een kalenderjaar maar één meetwaarde is, dan krijgt dat kalenderjaar geen jaarwaarde, tenzij er in geen van de andere kalenderjaren meer dan één meetwaarde voorkomt, want dan is de meetwaarde de jaarwaarde.

Op deze wijze wordt voorkomen dat er een tijdreeks ontstaat met zowel enkelvoudige meetwaarden als medianen (of gemiddelden) van meetwaarden. De statistische toetsen gaan er namelijk van uit dat alle waarden in de tijdreeks dezelfde nauwkeurigheid hebben en dat is niet het geval als er zowel meetwaarden als medianen (of rekenkundige gemiddelden) van meetwaarden in de tijdreeks voorkomen. Dit zal overigens ook al niet het geval zijn – zij het in mindere mate - als de jaren sterk verschillende aantallen meetwaarden bevatten, maar het is moeilijk daar passende criteria in het programma voor op te nemen. De gebruiker dient er daarom ook zelf op toe te zien dat er geen reeksen aan *Trendanalist* worden aangeboden met té sterk wisselende meetfrequenties.

Tijdstippen tijdreeks halverwege tijdseenheden
Bij het omzetten van een meetreeks naar een tijdreeks wordt elke tijdreekswaarde toegekend aan het tijdstip halverwege de betreffende tijdseenheid, waarbij de tijd wordt uitgedrukt als jaarfractie. Enkele voorbeelden ter verduidelijking:

- Bij omzetten van een meetreeks naar een tijdreeks op maandbasis, wordt bijvoorbeeld de tijdreekswaarde voor januari 2008 toegekend aan het tijdstip 2008,0416667, de tijdreekswaarde voor februari 2008 aan 2008,1250000, de tijdreekswaarde voor maart 2008 aan 2008,1666667, etcetera, met stappen van 0,0833333 (= 1/12).
- Bij omzetten van een meetreeks naar een tijdreeks op kwartaalbasis, wordt bijvoorbeeld de tijdreekswaarde voor het eerste kwartaal van 2008 toegekend aan het tijdstip 2008,125, de tijdreekswaarde voor het tweede kwartaal van 2008 aan 2008,375, de tijdreekswaarde voor het derde kwartaal van 2008,625, etcetera, met stappen van 0,250 (= 1/4).

Het is voor de gebruiker van belang dit te weten als hij/zij zelf ergens een trendlijn wil weergeven op basis van de met *Trendanalist* geschatte intercept en helling.

Voor een tijdreeks op jaarbasis gelden echter de volgende uitzonderingen op het bovenstaande.

1. Als er sprake is van een tijdreeks die is samengesteld uit een meetreeks met hooguit één meetwaarde per jaar, dan blijft het meettijdstip van een meetwaarde in de tijdreeks gehandhaafd. Dit voorkomt dat er een te grote vertekening van de trendlijn optreedt. Want als het meetinterval van een meetreeks met hooguit één meetwaarde per jaar soms sterk afwijkt van een jaar, treedt er vertekening op door het tijdreeksinterval vast te zetten op een jaar. Zo heeft de meetreeks met de meettijdstippen

20 januari 2007 - 16 november 2008 - 23 februari 2009 - 19 april 2010 – 21 maart 2011

   hooguit één meetwaarde per jaar, maar het meetinterval varieert van 3 maanden tot 21 maanden.
2. Als er sprake is van een tijdreeks die is samengesteld uit een meetreeks met meerdere meetwaarden per jaar, dan wordt de mediaan van de meetwaarden binnen een jaar toegekend aan het mediane meettijdstip van de betrokken meetwaarden. Dit voorkomt misverstanden bij het visueel beoordelen van een tijdreeksplot, want bij toekenning van een jaarwaarde aan het midden van het jaar is de visuele beoordeling lastiger.
   Als de gebruiker heeft aangegeven dat bij het omzetten van een meetreeks naar een tijdreeks per tijdseenheid niet de mediaan, maar het rekenkundig gemiddelde van de meetwaarden moet worden gehanteerd (via 'Opties' en dan 'Middelingsmethode'), dan wordt het gemiddelde van de meetwaarden binnen een jaar toegekend aan het gemiddelde meettijdstip van de betrokken meetwaarden.

Uitschieters

Uitschieters zijn meetwaarden die duidelijk afwijken van de andere meetwaarden. Ze kunnen zijn veroorzaakt door: 1. uitzonderlijke bemonsterings- of meetfouten en/of transcriptiefouten (fouten bij het schrijven en/of typen), of 2. een extreme situatie.
Voor trendanalyse is het van belang dat meetwaarden veroorzaakt door uitzonderlijke fouten niet meegenomen worden. Het vergt echter detailkennis van alle handelingen die tot een meetwaarde hebben geleid om objectief vast te kunnen stellen of er sprake is van een foute meetwaarde of van een meetwaarde die een extreme situatie weergeeft. Bij het ontbreken van deze detailkennis – en dit zal meer regel dan uitzondering zijn - is er dan ook geen enkele objectieve rechtvaardiging voor het verwijderen van uitschieters en moet dit achterwege blijven. *Trendanalist* zal daarom zelf geen uitschieters verwijderen, maar ze wel identificeren en vervolgens tonen in de meetreeksplot.

Het programma identificeert uitschieters door eerst de reeks van meetwaarden (Z_1, Z_2, ..., Z_n) om te zetten tot een reeks van gestudentiseerde afwijkingen (d_1, d_2, ..., d_n), volgens:

$$d_t = \frac{Z_i - \overline{Z}_{(-t)}}{s_{(-t)}}$$

met $\overline{Z}_{(-t)}$ en $s_{(-t)}$ het geschatte gemiddelde respectievelijk de geschatte standaardafwijking van de kansverdeling waar de meetwaarden uit afkomstig zijn, echter berekend *zonder* de betreffende meetwaarde Z_t. Het uitsluiten van de betreffende meetwaarde bij het berekenen van het gemiddelde en de standaardafwijking voorkomt dat deze kengetallen zodanig worden beïnvloed dat een uitschieter wordt gemaskeerd. Elke gestudentiseerde afwijking die groter is dan +3 of kleiner dan -

3 wordt door *Trendanalist* groen gemarkeerd als uitschieter in de meetreeksplot. De gebruiker kan deze uitschieter dan desgewenst interactief verwijderen, door deze aan te klikken in de meetreeksplot. Na het verwijderen van een meetwaarde zal het programma opnieuw berekenen of er uitschieters zijn (de kenmerken van de verzameling meetwaarden zijn dan immers veranderd). Een verwijderde meetwaarde wordt paars gemarkeerd in de meetreeksplot. Desgewenst kan deze waarde ook weer aan de meetreeks worden toegevoegd door hem opnieuw aan te klikken.

Gecensureerde waarden

Voor het omzetten van een meetreeks naar een tijdreeks worden alle gecensureerde waarden van een meetreeks standaard op de helft van de *hoogste* rapportagegrens van die reeks gezet (bij vrachten op nul). Ook niet-gecensureerde meetwaarden die lager zijn dan de hoogste rapportagegrens worden standaard op de helft van die grens gezet (tenzij de gebruiker die standaardoptie heeft uitgezet, zie ook verder) en zijn dan op te vatten als gecensureerde waarden. Deze aanpak kan informatieverlies en daarmee ook verlies aan onderscheidend vermogen geven, maar is nodig om het detecteren van kunstmatige trends te vermijden die louter zijn veroorzaakt door veranderingen van de rapportagegrens (zie bijvoorbeeld [Helsel and Hirsch, 1991]). Door deze voorbewerking kan het overigens voorkomen dat een tijdreeks meer gecensureerde waarden bevat dan de oorspronkelijke meetreeks.

Als voldaan wordt aan beide volgende criteria, dan wordt een aangepaste aanpak gevolgd:

1) de meetreeks bevat slechts één waarde die is gecensureerd ten opzichte van de hoogste rapportagegrens én
2) er komt minstens één niet-gecensureerde waarde voor, die tussen de hoogste rapportagegrens en de op één na hoogste rapportagegrens ligt, of die onder de hoogste rapportagegrens ligt als dat de enige rapportagegrens in de meetreeks is.

Alleen als wordt voldaan aan beide criteria dan wordt de ene waarde die is gecensureerd ten opzichte van de hoogste rapportagegrens verwijderd en worden vervolgens alle waarden (zowel de gecensureerde als de niet-gecensureerde) die onder de één na hoogste rapportagegrens liggen standaard op de helft van die één na hoogste rapportagegrens gezet, tenzij de meetreeks maar één rapportagegrens omvat, in welk geval er na de verwijdering van de ene gecensureerde waarde ten opzichte van die grens geen verdere censuur plaatsvindt. Dit voorkomt onnodig informatieverlies door een slechts eenmalig voorkomende relatief hoge rapportagegrens.

Als de gebruiker bovenstaande wil uitschakelen en wil dat elke gecensureerde waarde op de helft van zijn eigen rapportagegrens wordt gezet, kan dit via ‘Gecensureerde waarden’ en ‘Gecensureerde vrachten’ onder ‘Opties’ worden aangegeven. Er is dan echter wel een risico op het detecteren van kunstmatige trends, die louter zijn veroorzaakt door het tussentijds veranderen van de rapportagegrens.

We spreken na het omzetten van een meetreeks naar een tijdreeks van een gecensureerde tijdreekswaarde, als alle daaraan ten grondslag liggende meetreekswaarden gecensureerd zijn (hieronder vallen ook de aanvankelijk niet-gecensureerde meetwaarden onder de hoogste rapportagegrens, die bij de voorbewerking op de helft van die grens zijn gezet). Enkele voorbeelden:

- Een meetreeks met wekelijkse meetwaarden wordt omgezet naar een tijdreeks van vierwekelijkse waarden, door per periode van vier weken de mediaan van de daarbinnen vallende meetreekswaarden te nemen als tijdreekswaarde. Als in een bepaalde vierwekelijkse periode

de meetreekswaarden alle gecensureerd zijn, wordt de resulterende tijdreekswaarde geclassificeerd als zijnde gecensureerd. Maar als één of meer van deze meetreekswaarden niet gecensureerd is, wordt de resulterende tijdreekswaarde geclassificeerd als zijnde niet-gecensureerd.

- Een meetreeks met hooguit één meetwaarde per kwartaal wordt omgezet naar een tijdreeks van kwartaalwaarden, door per kwartaal de daarbinnen vallende meetreekswaarde te nemen als tijdreekswaarde. Als deze meetreekswaarde gecensureerd is, wordt ook de tijdreekswaarde geclassificeerd als zijnde gecensureerd.

Als na het omzetten van een meetreeks naar een tijdreeks de tijdreeks voor 80% of meer of bij vrachten 30% of meer bestaat uit gecensureerde tijdreekswaarden, dan is deze ongeschikt te achten voor trendanalyse. Deze reeks wordt na het uitvoeren van een automatische trendanalyse nog wel vermeld in het resultatenbestand, maar er wordt geen trendanalyse op uitgevoerd.

Verder adviseren wij de gebruiker om het resultaat van de trendanalyse niet mee te nemen als de tijdreeks meer dan 50% gecensureerde waarden bevat en er tevens geen statistisch significante trend is gedetecteerd. Dit is mede gebaseerd op bevindingen van [Alden et al., 2000], die met simulaties vaststelden dat de gebruikelijke toetsen op trend bij meer dan 50% gecensureerde waarden nog maar weinig onderscheidend vermogen hebben. Het niet detecteren van een trend heeft dan minder zeggingskracht dan gebruikelijk.

3.5.3 Vertoont de tijdreeks seizoenseffecten?

Als het geen jaarreeks betreft, zal *Trendanalist* nagaan of de tijdreeks seizoenseffecten vertoont, aan de hand van de Kruskal-Wallis-toets, het verdelingsvrije equivalent van één-weg-variantieanalyse.

Aangezien de Kruskal-Wallis-toets uitgaat van onafhankelijke waarden en beter zal werken op een trendloze reeks, zal *Trendanalist* de toets uitvoeren op een hulpreeks, die zo goed mogelijk aan die eisen voldoet. Deze hulpreeks wordt vervaardigd door:

1. de tijdreeks zo goed mogelijk van een eventuele trend te ontdoen, door van elke waarde de bijbehorende waarde van de trendlijn (met de Kendall-seizoenshelling) af te trekken (zie § 3.6.3);
2. als de voor trend gecorrigeerde tijdreeks meer dan één waarde per kwartaal bevat, deze om te zetten naar een reeks van kwartaalwaarden. Op kwartaalbasis zal er namelijk minder autocorrelatie optreden dan op vierwekelijkse, maandelijkse of tweemaandelijkse basis. Het omzetten geschiedt door alle waarden van de voor trend gecorrigeerde tijdreeks die binnen een bepaald kwartaal vallen te vervangen door hun *mediaan* (dus ongeacht wat er in 'Opties' is aangegeven als middelingsmethode).

De Kruskal-Wallis-toets hanteert dan het volgende model:

$$Z'_{ij} = \mu + \gamma_j + e_{ij}$$

met Z'_{ij} de waarde van de hulpreeks in jaar *i* (*i*=1...*r*) en seizoen *j* (*j*=1...*s*), μ het populatie-gemiddelde, γ het populatie-effect van seizoen *j* en e_{ij} de betreffende modelafwijking.
De nulhypothese ('géén seizoenseffecten') kan worden vertaald als:

$$H_0 : \gamma_1 = \gamma_2 = ... = \gamma_s$$

En de alternatieve hypothese luidt dat het populatie-effect van minstens één van de seizoenen ongelijk is aan dat van de andere seizoenen.

De procedure voor de toets kent dan de volgende stappen:

- rangschik de waarden Z'_{ij} van klein naar groot;
- ken deze waarden dan rangnummers R_{ij} toe (1...n, waarbij $n=r \cdot s$);
- bereken de variantie van de rangnummers, volgens:

$$s_R^2 = \frac{1}{n-1}\left[\sum_{i=1}^{r}\sum_{j=1}^{s} R_{ij}^2 - \frac{n \cdot (n+1)^2}{s}\right]$$

- bereken de toetsingsgrootheid T volgens:

$$T = \frac{1}{s_R^2}\left[\sum_{j=1}^{s} \frac{R_{+j}^2}{r} - \frac{n \cdot (n+1)^2}{s}\right]$$

met R_{+j} de som van alle rangnummers in seizoen j.

Omdat T onder H_0 een χ^2-verdeling zal volgen met s-1 vrijheidsgraden, kan met 95% betrouwbaarheid worden aangenomen dat de reeks seizoenseffecten vertoont als geldt:

$$T > \chi^2_{(95\%,s-1)}$$

met $\chi^2_{(95\%,s-1)}$ de χ^2-waarde met een éénzijdige onderschrijdingskans van 95% bij s-1 vrijheidsgraden.

3.5.4 Zijn de modelresiduën normaal verdeeld?

Na het toepassen van een lineair regressiemodel zal *Trendanalist* checken of de modelresiduën wel afkomstig zijn uit een normale kansverdeling. Daartoe wordt de Lilliefors-toets op normaliteit gehanteerd [Lilliefors, 1967 en 1969]. Dit is een aanpassing van de Kolmogorov-Smirnov-toets op normaliteit, voor díe gevallen waarin het gemiddelde en de standaardafwijking van de populatie moeten worden geschat uit de steekproefgegevens.

De toets berust op het vergelijken van de cumulatieve frequentieverdeling van de steekproef met die van de normale kansverdeling waar de steekproef volgens de nulhypothese uit afkomstig zou zijn. De cumulatieve frequentieverdeling van de steekproef ($F_s(e)$) wordt berekend door voor elk modelresidu (e_t, t=1,..., n) de empirische onderschrijdingskans te berekenen. Daartoe sorteren we eerst de modelresiduën van klein naar groot (e_r, r=1,..., n):

$$F_s(e_r) = \frac{r}{n}$$

Om de cumulatieve frequentieverdeling van de hypothetische normale kansverdeling ($F_p(e)$) te kunnen bepalen, moet eerst elk modelresidu (e_t, t=1,..., n) worden gestandaardiseerd, volgens:

$$u_t = \frac{e_t - \bar{e}}{s_e}$$

met $\bar{e}$ en s_e respectievelijk het gemiddelde en de standaardafwijking van de populatie modelresiduën, zoals geschat uit de steekproef. Vervolgens moet van elke u_t de eenzijdige onderschrijdingskans worden bepaald, zijnde de eenzijdige onderschrijdingskans van de waarde u_t van de standaardnormale kansverdeling.

De toetsingsgrootheid D van de Lilliefors-toets is vervolgens het maximale absolute verschil van beide cumulatieve frequentieverdelingen, volgens:

$$D = \max[|F_s(e) - F_p(e)|]$$

De kritische waarden voor D zijn voor verschillende steekproefgroottes en significantieniveaus getabelleerd. Er wordt getoetst op normaliteit met 95% betrouwbaarheid. Als *D* groter is dan de daarvoor geldende kritische waarde, kunnen we er niet van uitgaan dat de modelresiduën afkomstig zijn uit een normale kansverdeling. De automaat van *Trendanalist* zal dan één van de verdelingsvrije Mann-Kendall-toetsen toepassen.

3.5.5 Vertonen de modelresiduën autocorrelatie?

Als uit de Lilliefors-toets is gebleken dat de residuën van het lineaire regressiemodel afkomstig zijn uit een normale kansverdeling, zal de automaat van *Trendanalist* vervolgens nagaan of de modelresiduën wellicht autocorrelatie vertonen. Daartoe wordt de Portmanteau-toets op autocorrelatie gehanteerd. Deze baseert zich op een som van de gekwadrateerde autocorrelatiecoëfficiënten van de modelresiduën. De autocorrelatiecoëfficiënt is een maat voor de lineaire samenhang tussen waarden die zijn gescheiden door een bepaald tijdsinterval. Voor tijdsinterval *k* wordt deze maat geschat als [Box and Jenkins, 1976]:

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} e_t \cdot e_{t+k}}{\sum_{t=1}^{n} e_t^{\,2}}$$

met $\hat{\rho}_k$ de geschatte autocorrelatiecoëfficiënt voor tijdsinterval *k*, *e* het modelresidu, *t* de tijdsindex en *n* het aantal modelresiduën. Als de tijdreeks geen lege tijdseenheden omvat, wordt de toetsingsgrootheid Q_M van de Portmanteau-toets berekend als [Ljung and Box, 1978]:

$$Q_M = n \cdot (n+2) \cdot \sum_{k=1}^{L} \frac{\hat{\rho}_k^2}{n-k}$$

met *L* het aantal tijdsintervallen waarover de autocorrelatie wordt beschouwd (zie onder) en *n* het aantal modelresiduën. Onder de nulhypothese (‘géén autocorrelatie’) zal Q_M afkomstig zijn uit een χ^2-verdeling met *M* vrijheidsgraden, berekend als $M = L - p$, met *p* het aantal parameters van het ruismodel van het lineaire regressiemodel (in ons geval is *p* maximaal 1). We mogen daarom met 95% betrouwbaarheid aannemen dat de modelresiduën autocorrelatie vertonen als geldt:

$$Q_M > \chi^2_{(95\%,M)}$$

met $\chi^2_{(95\%,M)}$ de waarde van de χ^2-verdeling met een eenzijdige onderschrijdingskans van 95% bij *M* vrijheidsgraden.

Als de tijdreeks daarentegen wél lege tijdeenheden omvat, wordt een aangepaste toetsingsgrootheid van de Portmanteau-toets gehanteerd, die wordt berekend als [Stoffer and Toloi, 1992]:

$$Q_M^* = n^2 \cdot \sum_{k=1}^{L} \frac{C_{a,k} \cdot \hat{\rho}_{e,k}^2}{n-k}$$

met:

$$C_{a,k} = \frac{\sum_{t=1}^{n-k} a_t \cdot a_{t+k}}{n-k}$$

waarbij $a_t = 0$ als de waarde voor tijdstip *t* ontbreekt, anders is $a_t = 1$.
En met:

$$\hat{\rho}_{e,k} = \frac{\hat{C}_{z,k} \cdot C_{a,0}}{C_{a,k} \cdot \hat{C}_{z,0}}$$

en:

$$\hat{C}_{z,k} = \frac{\sum_{t=1}^{n-k} z_t \cdot z_{t+k}}{n}$$

waarin $z_t = a_t \cdot e_t$ en e_t het modelresidu voor tijdstip t.

Het aantal tijdsintervallen waarover de autocorrelatie wordt beschouwd (L) is in *Trendanalist* ingesteld op tweemaal het aantal tijdseenheden per jaar, echter met als minimum 5 en als maximum 15. Dus bij de tijdseenheden Vier weken en Maand is L 15 en bij Twee maanden is deze 12, bij Kwartaal is deze 8, bij Vier maanden is deze 6 en bij Zes maanden is deze 5. Er wordt aangenomen dat er bij de tijdseenheid Jaar geen storende autocorrelatie optreedt.

3.5.6 Vertoont de tijdreeks autocorrelatie?

Als uit de Lilliefors-toets is gebleken dat de modelresiduën van de lineaire regressietoets niet afkomstig zijn uit een normale kansverdeling, moet één van de verdelingsvrije Mann-Kendall-toetsen worden gebruikt. Om daarbij tot de geschikte selectie te kunnen komen, moet worden nagegaan of de tijdreeks autocorrelatie vertoont. Daartoe hanteert *Trendanalist* de runstoets, een verdelingsvrije toets. Aangezien het hierbij gaat om de autocorrelatie die resteert nadat de tijdreeks is ontdaan van een eventuele trend en eventuele seizoenseffecten, zal *Trendanalist* de toets uitvoeren op een hulpreeks, die zo goed mogelijk is ontdaan van die karakteristieken. Deze hulpreeks wordt vervaardigd door:

1. indien er geen sprake is van seizoenseffecten (volgens de aan het begin uitgevoerde Kruskal-Wallis-toets), de tijdreeks eerst zo goed mogelijk van een eventuele trend te ontdoen, door van elke waarde de bijbehorende waarde van de trendlijn (met de Theilhelling) af te trekken (zie § 3.6.1);
2. indien er sprake is van seizoenseffecten, de tijdreeks eerst zo goed mogelijk van een eventuele trend te ontdoen, door van elke waarde de bijbehorende waarde van de trendlijn (met de Kendall-seizoenshelling) af te trekken (zie § 3.6.2) en vervolgens zal *Trendanalist* van deze reeds voor trend gecorrigeerde reeks per seizoen de mediaan berekenen en vervolgens van elke reekswaarde de bijbehorende seizoensmediaan aftrekken.

We definiëren een ‘run’ als een aaneengesloten groep waarden die alle aan dezelfde kant van de mediaan van de reeks liggen. De toetsingsgrootheid van de runstoets is het aantal runs (r), met verwachtingswaarde:

$$\mathrm{E}[r] = 1 + \frac{2n_1 \cdot n_2}{n_1 + n_2}$$

en variantie:

$$\mathrm{Var}[r] = \frac{2n_1 \cdot n_2 \cdot (2n_1 \cdot n_2 - n_1 - n_2)}{(n_1 + n_2)^2 \cdot (n_1 + n_2 - 1)}$$

met n_1 het aantal waarden groter dan of gelijk aan de mediaan en n_2 het aantal waarden kleiner dan de mediaan. Als het aantal waarden groter is dan 50, kan er van worden uitgegaan dat r onder de nulhypothese (‘géén autocorrelatie’) een normale kansverdeling volgt. Uit r kan dan, mede met

behulp van een continuïteitscorrectie, de gestandaardiseerde toetsingsgrootheid *Tr* worden berekend, als:

$$T_r = \frac{r - 0{,}5 - \mathrm{E}[r]}{\sqrt{\mathrm{Var}[r]}} \quad \text{als } r - \mathrm{E}[r] > 0$$

$$T_r = \frac{r - \mathrm{E}[r]}{\sqrt{\mathrm{Var}[r]}} \quad \text{als } r - \mathrm{E}[r] = 0$$

$$T_r = \frac{r + 0{,}5 - \mathrm{E}[r]}{\sqrt{\mathrm{Var}[r]}} \quad \text{als } r - \mathrm{E}[r] < 0$$

Onder de nulhypothese zal T_r een standaardnormale kansverdeling volgen. We mogen daarom met 95% betrouwbaarheid aannemen dat de reeks autocorrelatie vertoont als voor de absolute waarde van T_r geldt:

$$|T_r| > U_{(97{,}5\%)}$$

met $U_{(97,5\%)}$ de waarde van de standaardnormale kansverdeling met een eenzijdige onderschrijdingskans van 97,5%.
Als het aantal waarden daarentegen niet groter is dan 50, wordt de tweezijdige overschrijdingskans bepaald van het aantal runs *r*, uitgaande van de binomiale kansverdeling. Als die tweezijdige overschrijdingskans kleiner is dan 5%, mogen we met 95% betrouwbaarheid aannemen dat de reeks autocorrelatie vertoont.

3.6 Trendkwantificering met *Trendanalist*

Voor wat betreft de *trendkwantificering* zal *Trendanalist* voor elke afzonderlijke tijdreeks de grootte van de trend schatten als de lineaire regressiehelling (in het geval van een normale kansverdeling), met de Theilhelling (in het geval van een niet-normale kansverdeling en geen seizoenseffecten), of met de Kendall-seizoenshelling (in het geval van een niet-normale kansverdeling en wel seizoenseffecten), ongeacht of er een statistisch significante trend in de tijdreeks is gedetecteerd. Aangezien ook bij de selectie van een trendtoets wordt afgegaan op kansverdeling, seizoenseffecten en autocorrelatie, zal *Trendanalist* altijd dezelfde combinaties van trendtoets en trendschatter hanteren. Deze combinaties zijn vermeld in tabel 3.2.

Tabel 3.2: De vaste combinaties van trendtoets en trendschatter die Trendanalist hanteert.

Trendtoets	Trendschatter
(Uitgebreide) lineaire regressie	Lineaire regressiehelling
Mann-Kendall	Theilhelling
Seizoenale Mann-Kendall	Kendall-seizoenshelling
Seizoenale Mann-Kendall met verdiscontering autocorrelatie	

3.6.1 Theilhelling

In het geval van een niet-normale kansverdeling en geen seizoenseffecten, wordt een trend door *Trendanalist* geschat als de Theilhelling b_{Th} [Theil, 1950, Sen, 1968]. Dit is de mediaan van alle mogelijke hellingen tussen de afzonderlijke waarden in de tijdreeks:

$$b_{Th} = \text{mediaan}\left[\frac{Z_t - Z_k}{X_t - X_k} \quad \text{voor alle } 1 \le k < t \le n\right]$$

met *Z* de waarden van de beschouwde variabele, *X* de tijd [jaar, inclusief jaarfractie], *t* en *k* waarden van de tijdsindex en *n* het aantal waarden in de reeks. Het betreft een zuivere hellingschatter– dat wil zeggen zonder systematische fout - en met een grotere precisie dan lineaire regressie bij niet-normaliteit [Hirsch et al., 1991].

3.6.2 Kendall-seizoenshelling

In het geval van een niet-normale kansverdeling en seizoenseffecten, wordt een trend door *Trendanalist* geschat als de Kendall-seizoenshelling [Sen, 1968; Hirsch et al., 1982], gedefinieerd als:

$$b_{Ks} = \text{mediaan}\left[\frac{Z_{tj} - Z_{kj}}{X_t - X_k} \quad \text{vooralle } 1 \le k < t \le n_j \text{ en voor } j = 1,\ldots,s\right]$$

met b_{Ks} de Kendall-seizoenshelling (eenheid/jaar) , *Z* de waarden van de beschouwde variabele, *X* de tijd [jaar], *t* en *k* waarden van de jaarindex, *j* de seizoensindex en n_j het aantal jaar met een tijdreekswaarde in seizoen *j*. Deze hellingen worden bepaald voor alle $1 \le k < t \le n_j$ en voor j=1,...s, met *s* het aantal tijdreekswaarden in een jaar (bij analyse op maandschaal is *s* bijvoorbeeld 12). Vervolgens wordt de Kendall-seizoenshelling berekend als de mediaan van álle hellingen.

Het is een niet-parametrische (of ook wel verdelingsvrije) schatter, vrij resistent tegen de invloed van extremen in de waarden, omdat de mediaan van alle individuele hellingen wordt gebruikt. Daarnaast wordt de schatter niet beïnvloed door seizoenseffecten, omdat de hellingen slechts worden bepaald voor waarden uit hetzelfde seizoen, die één of meer jaren uit elkaar liggen. Verder is de schatter zuiver – dat wil zeggen zonder systematische fout - en met een grotere precisie dan lineaire regressie bij het soort tijdreeksen dat kenmerkend mag worden geacht voor parameters van de waterkwaliteit, dat wil zeggen met seizoenseffecten, autocorrelatie en een scheve kansverdeling [Hirsch et al., 1982].

3.6.3 Trendlijn

Om de geschatte trend van een tijdreeks grafisch te kunnen weergeven in een trendlijn, is tevens een schatting van het intercept (b_0) nodig. Dit is het snijpunt van de trendlijn met de Y-as voor de tijdsindex t = 0.

Intercept in geval van normale kansverdeling

Als er sprake is van een normale kansverdeling, dan maakt de schatting van het intercept deel uit van de schatting van het lineaire regressiemodel. Als er daarbij ook sprake is van een seizoenseffect, dan wordt het geschatte intercept gecorrigeerd voor het daarin vervatte seizoenseffect (zie § 3.3.2).

Intercept in geval van geen normale kansverdeling

Als er geen sprake is van een normale kansverdeling, dan wordt het intercept verdelingsvrij geschat, volgens [Conover, 1980]:

$$b_0 = \text{mediaan}\left[Z_t \text{ voor } 1 \le t \le n\right] - b_1 \cdot \text{mediaan}\left[X_t \text{ voor } 1 \le t \le n\right]$$

met *Z* de waarden van de beschouwde variabele, *X* de tijd [jaar, inclusief jaarfractie], *t* de tijdsindex, *n* het aantal waarden in de reeks en b_1 de geschatte helling (dit is dan de Theilhelling of de Kendall-seizoenshelling).

De resulterende trendlijn

Met het geschatte intercept en de geschatte helling kan tenslotte de trendlijn worden berekend, volgens:

$$\hat{Z}_t = b_0 + b_1 \cdot X_t$$

4 De instelopties van *Trendanalist*

Het menu ‘Opties’ bevat de volgende mogelijkheden:

- Standaardmappen
- Middelingsmethode
- Alternatieve hypothese
- Betrouwbaarheid trendtoets
- Startpunt trend
- Gecensureerde waarden
- Gecensureerde vrachten
- Autocorrelatieplots
- Smooth
- Voorbewerken meetreeksen
- Criteria voor meetreeksen
- Overzicht opties
- Meldingen en uitvoer
- Grootte grafieken
- Font grafieken
- Taal (nl,en,du)

4.1 Standaardmappen

De gebruiker kan hier aangeven wat de standaardmappen (‘directories’) moeten zijn voor:

1. databestanden;
2. bestanden met numerieke resultaten;
3. bestanden met grafische resultaten;
4. tijdelijke bestanden.

4.2 Middelingsmethode

Om te kunnen toetsen op trend zal *Trendanalist* eerst bij de voorbewerking de meetreeks omzetten naar een tijdreeks met de best passende tijdseenheid. Het kan daarbij nodig zijn de meetwaarden binnen een bepaalde tijdseenheid samen te vatten tot een gemiddelde. Via de optie ‘Middelingsmethode’ kan de gebruiker aangeven of het rekenkundig gemiddelde moet zijn of de mediaan. *Trendanalist* zal standaard de mediaan hanteren, aangezien dat kengetal meer zeggingskracht heeft dan het gemiddelde bij de scheve kansverdelingen die doorgaans optreden bij milieuvariabelen. Verder wordt de mediaan niet of nauwelijks beïnvloed door uitschieters.

4.3 Alternatieve hypothese

Bij het toetsen op trend luidt de nulhypothese ‘er is geen trend’. Als alternatieve hypothese zal *Trendanalist* standaard hanteren ‘er is een trend’. Dit komt erop neer dat zowel op stijgingen als op dalingen wordt getoetst (tweezijdig toetsen). Maar de gebruiker kan via de optie ‘Alternatieve hypothese’ desgewenst aangeven of alleen op een stijging of alleen op een daling moet worden getoetst (eenzijdig toetsen).

Bij tweezijdig toetsen is het significantieniveau α (zie § 4.4) gelijk verdeeld over de onder- en bovenkant van de kansverdeling van de toetsingsgrootheid, terwijl deze zich bij eenzijdig toetsen slechts aan één kant van die kansverdeling bevindt.

4.4 Betrouwbaarheid trendtoets

De gebruiker kan bij deze optie aangeven met welke betrouwbaarheid moet worden getoetst op trend. De mogelijkheden zijn 95% (de standaard), 90% en 80%.
Bij het gebruiken van een statistische toets dient vóóraf de betrouwbaarheid te worden ingesteld, aangezien dat in feite de gevoeligheid van de toets regelt. Om dit te begrijpen moet een toets worden gezien als een instrument om objectief vast te kunnen stellen of een bepaalde nultoestand geldt (zoals: 'er is geen trend'). Aan elke toets zijn twee risico's verbonden, namelijk:

1. het onterecht concluderen dat de nultoestand niet geldt. Dit is het significantieniveau van de toets, aangeduid als α. Het wordt ook wel het risico van de 1^e soort genoemd. Hier is dat de kans om onterecht een trend te detecteren. De betrouwbaarheid van een uitspraak bij het toetsen is 1-α;
2. het onterecht concluderen dat de nultoestand geldt. Dit wordt aangegeven als β, of ook wel het risico van de 2^e soort. Hier is dat de kans om onterecht geen trend te detecteren. Deze kans neemt af met de grootte van de bestaande trend. De kans om een bestaande trend wél te detecteren (1-β) wordt aangeduid als het 'onderscheidend vermogen'. Dit vermogen neemt toe met de grootte van de bestaande trend.

Het is niet mogelijk beide risico's simultaan te beperken, want als α wordt verlaagd wordt β automatisch hoger. De keuze van de betrouwbaarheid 1-α is dan ook altijd min of meer een compromis. In de statistische praktijk is het algemeen gebruikelijk om te toetsen op trend met een betrouwbaarheid van 95% (tenzij er levens op het spel staan), wat inhoudt dat er dan slechts 5% kans bestaat om onterecht een trend te detecteren.

De binnen *Trendanalist* gehanteerde toetsen op normaliteit, autocorrelatie en seizoenseffecten hebben elk een betrouwbaarheid van 95%. Dit kan niet worden aangepast door de gebruiker.

4.5 Startpunt trend

Standaard zal *Trendanalist* elke reeks toetsen op een monotone trend die start bij het begin van de tijdreeks. De gebruiker kan via de optie 'Startpunt trend' echter ook een datum opgeven als startpunt van de trend, in het formaat dd/mm/jjjj (bijvoorbeeld 21/02/1990).

4.6 Gecensureerde waarden

Om trendanalyse uit te kunnen voeren, moet elke gecensureerde waarde – op te vatten als een semi-kwantitatieve waarde - worden vervangen door een kwantitatieve waarde.
De mogelijkheden zijn:

1. Op helft hoogste rapportagegrens - Standaard zal *Trendanalist* elke gecensureerde waarde van de meetreeks op de helft van de hoogste rapportagegrens zetten die in de meetreeks voorkomt. Ook niet-gecensureerde meetwaarden die lager zijn dan de hoogste rapportagegrens worden op de helft van die grens gezet. Deze aanpak kan informatieverlies geven, maar is nodig om kunstmatige trends door veranderingen van de rapportagegrens te vermijden (zie bijvoorbeeld [Helsel and Hirsch, 1991]). In het geval er slechts één waarde is gecensureerd ten opzichte van de hoogste rapportagegrens, wordt die waarde verwijderd en worden vervol-

gens alle waarden (zowel de gecensureerde als de niet-gecensureerde) die onder de één na hoogste rapportagegrens liggen op de helft van die één na hoogste rapportagegrens gezet. Dit voorkomt onnodig informatieverlies.

2. Op helft eigen rapportagegrens - Als de gebruiker wil dat elke gecensureerde waarde op de helft van zijn eigen rapportagegrens wordt gezet, moet dit via deze optie worden aangegeven. Er is dan echter wel een risico op het detecteren van kunstmatige trends die louter zijn veroorzaakt door het tussentijds veranderen van de rapportagegrens.

4.7 Gecensureerde vrachten

Trendanalist interpreteert meetwaarden als vrachten als de meeteenheid eindigt op de tijdseenheid ('uur', 'dag', 'week', 'maand', 'kwartaal' of 'jaar'). Voor de kwantitatieve waarde van gecensureerde vrachten zijn er drie mogelijkheden:

1. Op nul – Standaard zal *Trendanalist* elke gecensureerde meetwaarde van een vracht van de meetreeks op nul zetten die in de meetreeks voorkomt. Deze optie wordt toegepast bij vrachtberekeningen (product van meetwaarden en debieten) met gecensureerde meetwaarden bij ICBR[8]. Een product waarbij één van de variabelen gecensureerd is, kan extreem grote waarden geven. Ook dan is er een risico op het detecteren van kunstmatige trends die louter worden veroorzaakt door berekeningen met gecensureerde meetwaarden.
2. Op helft hoogste rapportagegrens – Als de gebruiker wil dat elke gecensureerde vrachtwaarde van de meetreeks op de helft van de hoogste rapportagegrens wordt gezet. De hoogste rapportagegrens bij een vrachtberekening is de hoogste waarde van het product van gecensureerde meetwaarde en debiet in de meetreeks. Ook niet-gecensureerde vrachtwaarden die lager zijn dan de hoogste rapportagegrens worden op de helft van die grens gezet. Deze aanpak kan informatieverlies geven, maar is nodig om kunstmatige trends door veranderingen van de rapportagegrens te vermijden (zie bijvoorbeeld [Helsel and Hirsch, 1991]). In het geval er slechts één vrachtwaarde is gecensureerd ten opzichte van de hoogste rapportagegrens, wordt die waarde verwijderd en worden vervolgens alle vrachtwaarden (zowel de gecensureerde als de niet-gecensureerde) die onder de één na hoogste rapportagegrens liggen op de helft van die één na hoogste rapportagegrens gezet. Dit voorkomt onnodig informatieverlies.
3. Op helft eigen rapportagegrens - Als de gebruiker wil dat elke gecensureerde vrachtwaarde op de helft van zijn eigen rapportagegrens wordt gezet, moet dit via deze optie worden aangegeven. Er is dan echter wel een risico op het detecteren van kunstmatige trends, die louter zijn veroorzaakt door het tussentijds veranderen van de rapportagegrens van de vrachtwaarde.

4.8 Autocorrelatieplots

Bij handmatig gebruik van *Trendanalist* kan de autocorrelatie van de tijdreeks of van de residuën van het (al dan niet uitgebreide) lineaire regressiemodel visueel worden beoordeeld op basis van: (1) een tijdreeksplot met daarin het voortschrijdend gemiddelde en (2) een autocorrelogram. De gebruiker kan zowel het tijdvenster van het voortschrijdend gemiddelde als de lengte van het au-

[8] ICBR, 2022. *Vermindering van microverontreinigingen in het Rijnstroomgebied – Monitoring en beoordelingssysteem.* Internationale Commissie ter Bescherming van de Rijn.

tocorrelogram opgeven. Het tijdvenster van het voortschrijdend gemiddelde is standaard 365 dagen en de lengte van de X-as van het autocorrelogram is standaard 25 tijdseenheden.

4.9 Smooth

Een trendplot bevat behalve een trendlijn ook een LOWESS-lijn (zie § 5.7). Deze laatste beschrijft een lokaal gemiddelde van de meetwaarden, op basis van lokale gewogen regressie. De gladheidsfactor ('smooth') f definieert het venster dat aan weerskanten van elk tijdstip bij de modellering wordt betrokken en bepaalt daarmee mede de gladheid van de resulterende lijn. Deze factor hebben we standaard op 0,5 gezet, aangezien die waarde in de praktijk doorgaans de bruikbaarste LOWESS-lijnen oplevert [Helsel and Hirsch, 1991]. De gebruiker kan de factor desgewenst echter aanpassen.

4.10 Voorbewerken meetreeksen

Voor het uitvoeren van trendanalyses dient elke meetreeks oplopend gesorteerd te zijn op datum. Trendanalist heeft de volgende twee mogelijkheden om dit te bewerkstelligen:

1. Meetreeksen te sorteren op basis van resp. meetlocatie, parameternaam, hoedanigheid, compartiment, eenheid en de datum (en eventueel tijdstip).
2. Dubbele metingen met gelijke datum (en eventueel tijdstip) met gelijke meetwaarde of verschillende meetwaarde te verwijderen door te kiezen voor de eerste van de dubbele meetwaarden.

Trendanalist zal een meetreeks oplopend sorteren op datum. De gebruiker kan er voor kiezen deze opties niet te gebruiken voor het controleren of elke meetreeks wel oplopend gesorteerd is op datum of om zelf dubbele meetwaarden te verwijderen.

4.11 Criteria meetreeksen

Trendanalist zal eerst van elke ingelezen meetreeks nagaan of deze de minimale informatieinhoud heeft om trendanalyse te kunnen uitvoeren. Die vereisten zijn:

1. meetwaarden over minstens 4,5 jaar (periode tussen start en eind meetwaarden);
2. minstens 5 meetwaarden, of anders minstens 3 verschillende jaarwaarden;
3. minder dan 80% gecensureerde meetwaarden, uitgezonderd vrachten (product meetwaarden en debiet) dan minder dan 30% gecensureerde waarden.

De gebruiker heeft de mogelijkheid om hardere criteria te stellen. Hardere criteria geven minder trends met mogelijk informatieverlies, maar kan de selectie geven van trends met de gewenste specificaties en betrouwbaarheid.

4.12 Overzicht opties

Hier zijn alle hiervoor beschreven instellingen samengevat in één venster.

4.13 Meldingen en uitvoer

Bij 'Meldingen en uitvoer' kunnen de onderstaande opties worden geactiveerd. Deze opties staan standaard uit.

- *Kenmerken ingelezen gegevens* – Na het inlezen worden in een apart scherm de algemene kenmerken van de ingelezen meetreeksen weergegeven. Dit betreft begin- en einddatum, aantal meetwaarden, aantal gecensureerde meetwaarden, gemiddelde, standaardafwijking,

minimum, maximum en aantal jaarwaarden (dit is het aantal jaar dat minstens 1 meetwaarde bevat). Verder is vermeld of alle ingelezen meetreeksen voldoen aan de drie daaraan gestelde criteria (deze zijn vermeld in § 2.2). Achter elk criterium is 'Ja' of 'Nee' vermeld. Een meetreeks die niet aan de criteria voor meetreeksen voldoet, kan niet worden geselecteerd voor de trendanalyse.

- *Ongeschikte reeksen* – Elke meetreeks die niet voldoet aan de criteria voor meetreeksen wordt vermeld in een apart scherm, evenals de criteria waaraan die reeks niet voldoet.
- *Parameterschattingen* – Bij de handmatige modus worden de geschatte modelparameters van het (al dan niet uitgebreide) lineaire regressiemodel vermeld in een apart scherm. Tevens worden hun schattingen van de voorlaatste iteratie vermeld, zodat kan worden beoordeeld of er voldoende convergentie is bereikt.
- Geen waarschuwingen – De waarschuwingen van *Trendanalist* kunnen bij de handmatige modus worden onderdrukt. Bij de automatische modus worden deze al standaard onderdrukt. De volgende waarschuwingen zijn mogelijk:
    - 'Geen trendanalyse mogelijk'
    - 'Meetreeks bevat te veel missende waarden'
    - 'Meetreeks is te kort voor trendanalyse'
    - 'Ongeschikte reeks'.
- Trendplots bewaren - Bij deze keuze worden bij het draaien van de automaat alle trendplots temporeel bewaard als emf-bestand, zodat ze na afloop desgewenst door de gebruiker opgeslagen kunnen worden via 'Opslaan grafieken' onder 'Bestand'.
- Ook ongeschikte reeksen in resultatenbestand – Reeksen die bij het inlezen niet bleken te voldoen aan de criteria worden hierdoor toch vermeld in het resultatenbestand.
- Controle opslaan resultatenbestand - De gebruiker wordt er hierdoor bij het beëindigen van *Trendanalist* op geattendeerd dat bepaalde resultaten nog niet bewaard zijn en daardoor verloren zullen gaan.

4.14 Grootte grafieken

De lijndikte en de tekengrootte van de grafieken kunnen naar keuze worden aangepast.

4.15 Font grafieken

Het font, de fontstijl en de fontgrootte van de grafieken kunnen naar keuze worden aangepast.

4.16 Taal

De taal binnen Trendanalist en die van de uitvoer kunnen aangepast worden. De gebruiker heeft de mogelijkheid om te kiezen tussen Nederlands, Engels en Duits.

5 Presentatiemogelijkheden

Trendanalist heeft meerdere presentatiemogelijkheden, die we hieronder toelichten.
De presentatiemogelijkheden zijn beschikbaar via het menu 'Verkennen'. Er zijn negen presentatiemogelijkheden, waarvan vier al direct na het inlezen beschikbaar zijn:

- *Meetreeksplots*
- *Jaarboxplots*
- *Meetdichtheidsmatrix*
- *Kengetallenmatrix*
- *Meetlocatiekaart*

De overige vijf presentatiemogelijkheden zijn pas beschikbaar na de trendanalyse:

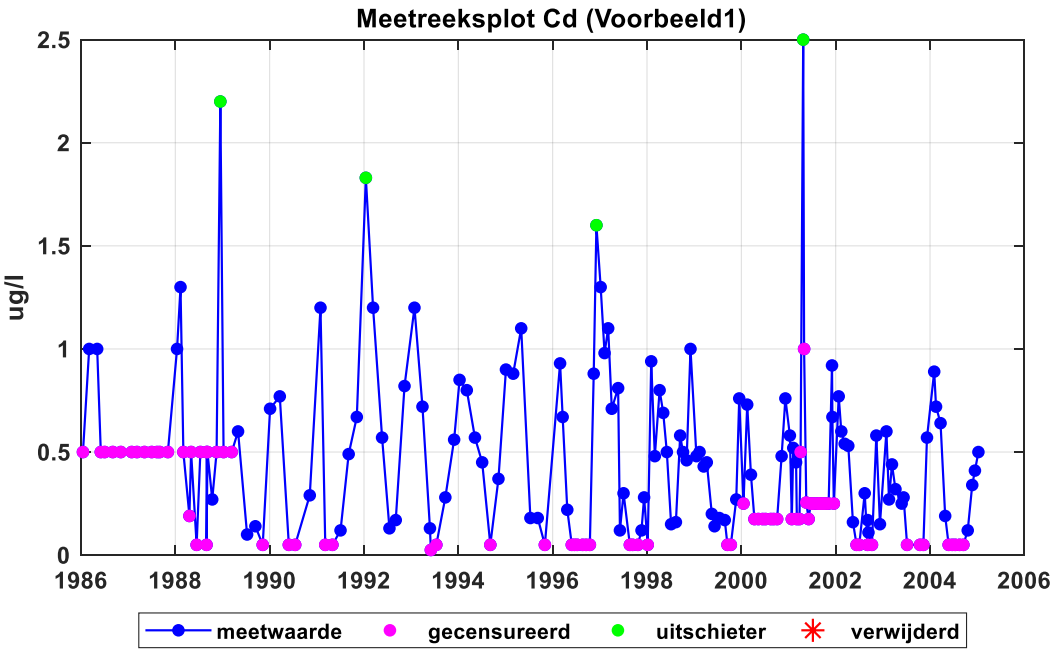
- *Bekijk sessie*
- *Trendplots*
- *Resultaten*
- *Trendpalet*
- *Trendkaart*

De ruimtelijke presentatiemogelijkheden *Meetlocatiekaart* en *Trendkaart* zijn overigens alleen beschikbaar als het invoerbestand ook coördinaten bevat. De negen bovengenoemde presentatiemogelijkheden worden toegelicht in de onderstaande paragrafen.

5.1 Meetreeksplots

Een meetreeksplot toont het verloop van de meetwaarden in de tijd, waarbij de meetwaarden zijn verbonden door een lijn (zie het voorbeeld in figuur 5.1).

Figuur 5.1: Voorbeeld van een meetreeksplot.

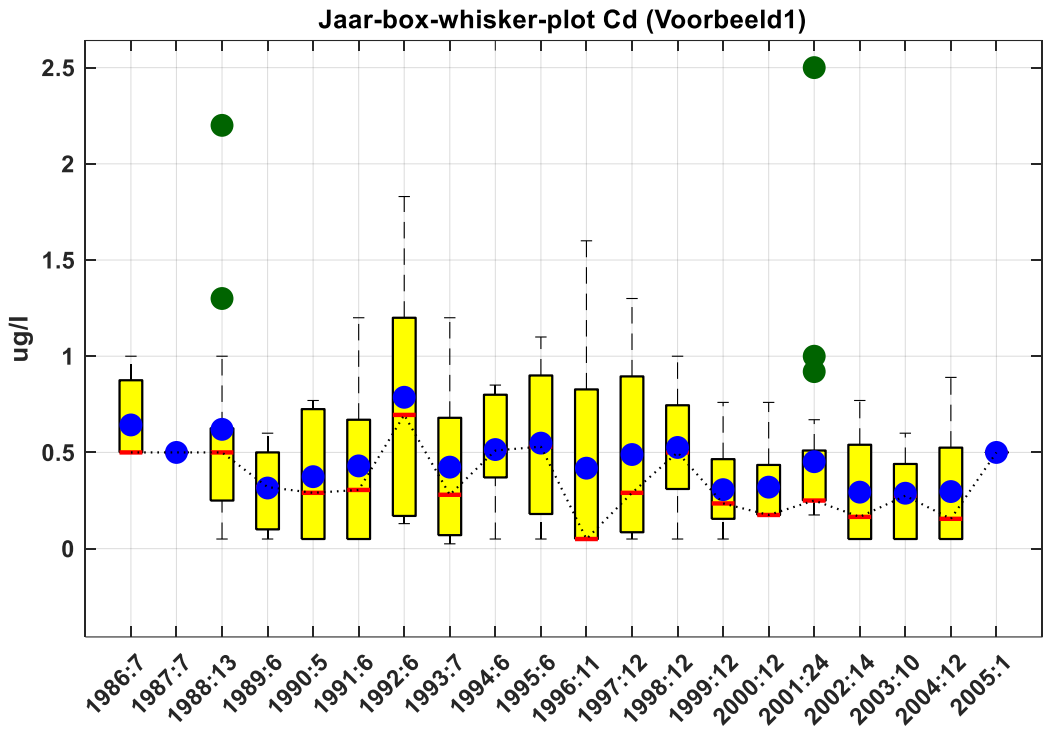


Als er sprake is van gecensureerde waarden, zijn deze in de meetreeksplot weergegeven op de helft van de eigen rapportagegrens, als een paars punt. Dit ongeacht wat de gebruiker heeft aangegeven bij 'Gecensureerde waarden' of 'Gecensureerde vrachten' onder 'Opties', want dat heeft alleen betrekking op hoe wordt omgegaan met gecensureerde waarden vanaf het omzetten van een meetreeks tot een tijdreeks.
Verder zijn gedetecteerde uitschieters in de meetreeksplot weergegeven met een groen punt.

5.2 Jaarboxplots

Een jaarboxplot toont het verloop van de meetwaarden in de tijd, waarbij jaarlijks de verdeling van de meetwaarden wordt weergegeven (zie het voorbeeld in figuur 5.2).

Figuur 5.2: Voorbeeld van een jaarboxplot.



De jaarboxplot toont de posities van de schattingen van de belangrijkste percentielen van de jaarlijkse kansverdeling waar de betreffende meetwaarden uit afkomstig zijn. Het middendeel, de 'box' (doos), loopt van het 25-percentiel naar het 75-percentiel, de blauwe punt in de box is het gemiddelde, terwijl het 50-percentiel, oftewel de mediaan, is aangegeven als een rode streep in de box. De 'whiskers' (snorharen) lopen van de box naar de uiteinden van de steekproef. Eventuele extreme waarden worden apart weergegeven boven de bovenste whisker of onder de onderste whisker als een groen bolletje. Die waarde ligt dan verder dan 1,5-maal de boxlengte vanaf de rand van de box.

5.3 Meetdichtheidsmatrix

De informatieinhoud van het ingelezen databestand kan duidelijk zichtbaar worden gemaakt met de meetdichtheidsmatrix. Deze geeft per combinatie van meetlocatie, parameter en jaar het aantal meetwaarden. Tabel 5.1 toont een voorbeeld van een dergelijke matrix.

Tabel 5.1: Voorbeeld van een meetdichtheidsmatrix.

Meetpunt	parameter	Eenheid	aantal	#gecensureerd	trendanalyse mogelijk?	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
G1	Ammonium	mg N/l	192	9	Ja	0	0	0	0	0	0	11	12	12	12
G1	Chloride	mg/l	201	0	Ja	0	0	0	0	0	0	11	12	12	12
G1	Fosfaat	mg P/l	201	3	Ja	0	0	0	0	0	0	11	12	12	12
G1	Kjeldahl stikstof	mg N/l	202	2	Ja	0	0	0	0	0	0	11	12	13	12
G1	Koper	ug/l	76	4	Ja	0	0	0	0	0	0	6	12	12	12
G1	Nitraat	mg N/l	201	0	Ja	0	0	0	0	0	0	11	12	12	12
G1	Nitriet	mg N/l	201	0	Ja	0	0	0	0	0	0	11	12	12	12
G1	Ntotaal	mg N/l	201	0	Ja	0	0	0	0	0	0	11	12	12	12
G1	Zink	ug/l	76	2	Ja	0	0	0	0	0	0	6	12	12	12
D2	Ammonium	mg N/l	252	47	Ja	3	6	6	6	12	11	12	12	11	12
D2	Chloride	mg/l	267	0	Ja	3	6	6	6	12	11	12	12	11	12
D2	Fosfaat	mg P/l	267	0	Ja	3	6	6	6	12	11	12	12	11	12
D2	Kjeldahl stikstof	mg N/l	266	6	Ja	3	6	6	6	12	11	12	12	11	12
D2	Koper	ug/l	147	37	Ja	3	6	6	6	6	7	6	6	11	12
D2	Nitraat	mg N/l	267	0	Ja	3	6	6	6	12	11	12	12	11	12
D2	Nitriet	mg N/l	266	1	Ja	3	6	6	6	12	11	12	12	11	12
D2	Ntotaal	mg N/l	265	0	Ja	3	6	6	6	12	11	12	12	11	12
D2	Zink	ug/l	147	84	Ja	3	6	6	6	6	7	6	6	11	12

De kolommen bevatten achtereenvolgens:

- de aanduiding (naam of code) van het meetpunt
- de parameternaam
- de meeteenheid
- het aantal meetwaarden
- het aantal gecensureerde waarden (cel is gekleurd als aantal > 0)
- het al of niet mogelijk zijn van trendanalyse, uitgaande van de in § 2.2 vermelde criteria voor de minimale informatieinhoud om trendanalyse te kunnen uitvoeren. Aangezien het mogelijk is dat er door het omzetten van een meetreeks naar een tijdreeks niet meer wordt voldaan aan één of meer van deze criteria, geeft een 'Ja' in deze kolom nog geen volledige garantie dat trendanalyse mogelijk is
- per kalenderjaar het aantal meetwaarden (cel is gekleurd als aantal > 0)

5.4 Kengetallenmatrix

De kengetallenmatrix vermeldt per combinatie van meetlocatie en parameter een aantal kengetallen. Dit zijn achtereenvolgens:

Eenheid	Meeteenheid
Aantal	Aantal meetwaarden
Min	Minimum van de meetwaarden
P5	Geschat 5-percentiel van de populatie meetwaarden
P25	Geschat 25-percentiel van de populatie meetwaarden
P50	Geschat 50-percentiel van de populatie meetwaarden
Gemidd	Gemiddelde van de meetwaarden
P75	Geschat 75-percentiel van de populatie meetwaarden
P95	Geschat 95-percentiel van de populatie meetwaarden
Max	Maximum van de meetwaarden
HKA	Halve kwartielafstand (robuuste spreidingsmaat), HKA=(P75-P25)/2
Stafw	Geschatte standaardafwijking van de populatie meetwaarden
%VC	Percentage coëfficiënt van variantie, *Stafw/Gemidd**100%
Scheefheid	Scheefheid van de meetwaarden

#nul	Aantal nulwaarden in de meetreeks
#gecens	Aantal gecensureerde waarden
#rap	Aantal rapportagegrenzen in de meetreeks
Min rap	Laagste rapportagegrens in de meetreeks
Max rap	Hoogste rapportagegrens in de meetreeks

5.5 Meetlocatiekaart

Als het invoerbestand ook de coördinaten van de meetlocaties bevat, kan de ligging van de meetlocaties worden weergegeven op een kaart (zie het voorbeeld in figuur 5.3). Daartoe wordt door *Trendanalist* het programma Google Earth aangeroepen. Als dit niet is geïnstalleerd op de betreffende computer, volgt er een melding.

Figuur 5.3: Voorbeeld van een meetlocatiekaart. Het betreft de ligging van de meetpunten van een Vlaams meetnet.

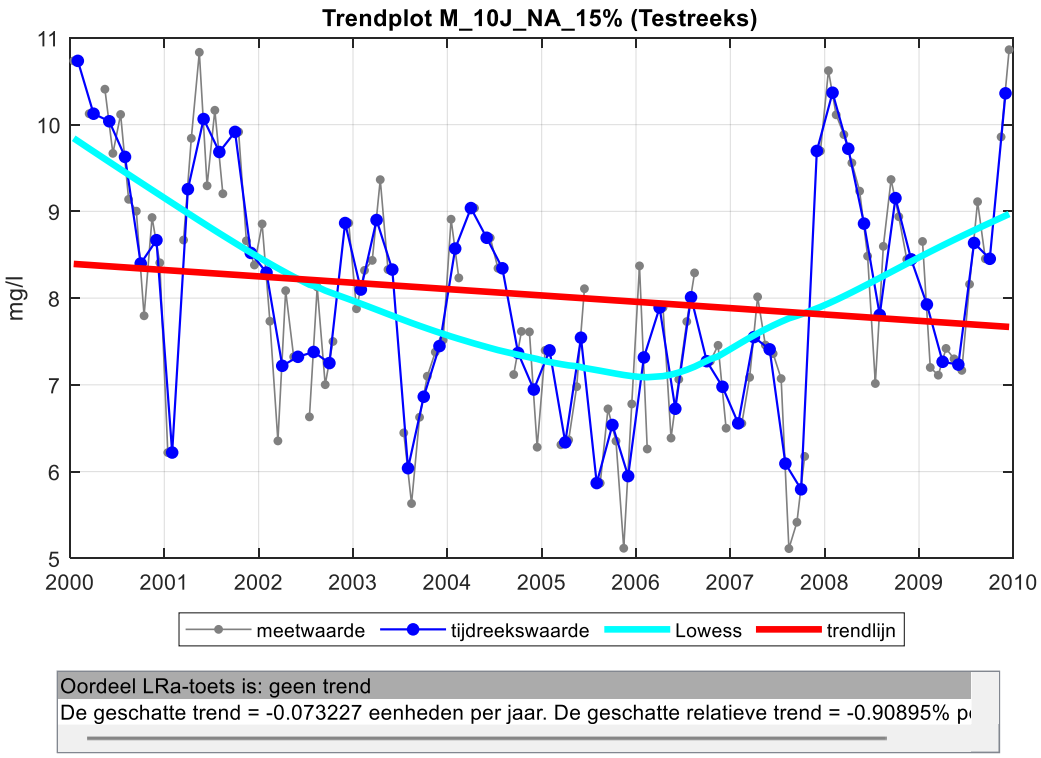
5.6 Bekijk sessie

Na het uitvoeren van een automatische of handmatige sessie kunnen de resultaten daarvan per parameter worden bekeken. Na de keuze van de parameter verschijnt een apart scherm met daarin alle resultaten vermeld. Deze zijn toegelicht in § 5.8.

5.7 Trendplots

Van elke geanalyseerde meetreeks kan een trendplot worden weergegeven. Deze toont de reeks, de LOWESS-lijn en de trendlijn (zie het voorbeeld in figuur 5.4). Meer op de achtergrond (in grijs) zijn tevens de oorspronkelijke meetreekswaarden weergegeven.

Figuur 5.4: Voorbeeld van een trendplot. De meetreeks bevat doorgaans één meetwaarde per maand, maar de automatische trendanalyse is uiteindelijk uitgevoerd met als tijdseenheid Twee maanden. De trendanalyse met als tijdseenheid Maand kwam namelijk uit op de MKsa-toets, maar die detecteerde geen statistisch significante trend, zodat is overgegaan naar een grotere tijdseenheid (zie de selectieprocedure in figuur 3.1).



De trendlijn kan alleen een lineaire trend beschrijven, terwijl de LOWESS-lijn een flexibelere beschrijving kan geven van het verloop van het gemiddelde van de meetreeks (zie onderstaand tekstkader). In bovenstaande figuur geeft de trendlijn nauwelijks een verandering aan over de beschouwde periode, terwijl de LOWESS-lijn nader specificeert dat er een daling was tot circa 2006, gevolgd door een stijging.
We wijzen erop dat de trendlijn een objectieve, toetsende basis heeft, terwijl de LOWESS-lijn slechts als explorerend kan worden gezien. Uitspraken over het verloop op basis van de LOWESS-lijn hebben daarmee ook niet meer dan een explorerend karakter.

Toelichting op de LOWESS-lijn

De LOWESS-lijn beschrijft een lokaal gemiddelde van de meetwaarden, op basis van lokale gewogen regressie [Cleveland, 1979]. LOWESS staat voor LOcally WEighted Scatterplot Smoothing. Het is een zeer flexibele en bruikbare manier om de lange-termijn veranderingen van een tijdreeks te tonen. Op iteratieve wijze wordt voor elk tijdstip de bijbehorende waarde van de LOWESS-lijn geschat met gewogen regressie, waarbij de gewichten zowel een functie zijn van de afstand tot het betreffende tijdstip, als van het modelresidu van de vorige modeliteratie. De gladheidsfactor f bepaalt het venster dat aan weerskanten van elk tijdstip bij de modellering wordt betrokken en bepaalt daarmee mede de gladheid van de resulterende lijn. Deze factor hebben we standaard op 0,5 gezet, aangezien die waarde in de praktijk doorgaans de bruikbaarste LOWESS-lijnen oplevert [Helsel and Hirsch, 1991].

5.8 Resultaten

De trendanalyseresultaten zijn vermeld in een uitgebreide teksttabel met per geanalyseerde meetreeks enkele karakteristieken en de resultaten van de analyse. Het uitvoerbestand krijgt standaard de naam van het invoerbestand, maar met de extensie '.tau' (dit is een afkorting van '**t**rend**a**nalist **u**itvoer') of met de extensie '.xlsx' (Excel).
Het tau-bestand is een ASCII-bestand en ook gemakkelijk leesbaar in Excel. Ga daartoe in Windows Explorer (of een vergelijkbaar programma) met de cursor op de bestandsnaam staan en klik vervolgens met de rechtermuisknop. Kies dan 'Open met' en 'Microsoft Excel'.

De kop van de uitvoertabel vermeldt de onderstaande algemene analysegegevens.

Analysedatum	Datum en uur van de start van de analyse.
Getoetst op	Hier is de gehanteerde alternatieve hypothese vermeld, met als mogelijkheden '**Trend**' (er is dan tweezijdig getoetst op een verandering, die dus zowel een stijging als een daling kan zijn), '**Stijging**', of '**Daling**' (in deze twee gevallen is er eenzijdig getoetst).
Gehanteerde betrouwbaarheid	Dit is **95%** (de standaardinstelling), **90%** of **80%**.
Startpunt trend	Dit is '**Reeksbegin**' (de standaardinstelling), of een door de gebruiker opgegeven datum.
Middelingsmethode	Dit is '**Mediaan**' (de standaardinstelling), of '**Gemiddelde**.'
Gecensureerde waarden	Dit is '**Op helft hoogste rapportagegrens**' (de standaardinstelling), of '**Op helft eigen rapportagegrens**'.
Maximaal percentage gecensureerde meetwaarden	Dit is **80%.**
Gecensureerde vrachten	Dit is **'Op nul (ICBR-methode)'**, of '**Op helft hoogste rapportagegrens**' (de standaardinstelling), of '**Op helft eigen rapportagegrens**'.
Maximaal percentage gecensureerde vrachten	*D*it is **30%.**
Versie	Datum en nummer van deze versie van *Trendanalist*.

Vervolgens worden voor elke geanalyseerde meetreeks afzonderlijk de volgende gegevens vermeld (een rij per meetreeks).

Meetpunt	Naam of code van het meetpunt.
Param	Naam of code van de geanalyseerde parameter.
Eenheid	Meeteenheid waarin de meetwaarden zijn uitgedrukt.
X-coord	X-coördinaat van het meetpunt [m]
Y-coord	Y-coördinaat van het meetpunt [m]
Start	Startdatum van de meetreeks.
Eind	Einddatum van de meetreeks.
#mw	Aantal meetwaarden van de ingelezen meetreeks.
#mw/jaar	Gemiddeld aantal meetwaarden van de meetreeks per jaar.
#gecens	Aantal gecensureerde meetwaarden.
%gecens	Percentage meetwaarden van de meetreeks dat gecensureerd is.
censfr	Fractie van rapportagegrens.
%maxcens	Maximaal percentage gecensureerde meetwaarden of vrachten.

#rap	Aantal rapportagegrenzen in de meetreeks.
Min	Minimum van de meetwaarden.
Med	Geschatte mediaan van de populatie van de meetwaarden.
MedICBR	Geschatte mediaan van de periode 2016 tot 2019 of de eerste drie jaar van de populatie van de meetwaarden.
Gemidd	Geschatte gemiddelde van de populatie van de meetwaarden.
Max	Maximum van de meetwaarden.
Stafw	Geschatte standaardafwijking van de populatie van de meetwaarden.
%VC	Geschatte variatiecoëfficiënt van de populatie van de meetwaarden, berekend als 100·*Stafw*/*Gemidd*.
Tijdseenh	Tijdseenheid van de geanalyseerde tijdreeks. Hier staat vermeld '**Ongeschikte reeks**' als de reeks niet voldoet aan de criteria voor trendanalyse.
Start	Startdatum van de tijdreeks.
Eind	Einddatum van de tijdreeks.
#tw	Aantal waarden van de tijdreeks.
%gevuld	Percentage tijdseenheden van de tijdreeks met een waarde.
#leeg	Aantal tijdseenheden van de tijdreeks met een ontbrekende waarde.
%leeg	Percentage tijdseenheden van de tijdreeks met een ontbrekende waarde.
#gecens	Aantal gecensureerde waarden van de tijdreeks.
%gecens	Percentage tijdreekswaarden van de tijdreeks dat gecensureerd is.
Min	Minimum van de tijdreekswaarden.
Med	Geschatte mediaan van de populatie van de tijdreekswaarden.
Gemidd	Geschatte gemiddelde van de populatie van de tijdreekswaarden.
Max	Maximum van de tijdreekswaarden.
Stafw	Geschatte standaardafwijking van de populatie van de tijdreekswaarden.
%VC	Geschatte variatiecoëfficiënt van de populatie van de tijdreekswaarden, berekend als 100·*Stafw*/*Gemidd*.
Toets	Gehanteerde trendtoets.
Intcpt	Geschatte intercept van de trendlijn (b_0), uitgedrukt in de meeteenheid (zie § 3.6.3). Als er is getoetst op trend met **LRs** of **LRsa**, dan is dit geschatte intercept gecorrigeerd voor het daarin vervatte seizoenseffect.
Trend/jaar	Grootte van de geschatte trend (b_1), uitgedrukt in meeteenheden per jaar (zie § 3.6).
Trend/med	Geschatte trend gedeeld door de mediaan van de tijdreekswaarden.
Rel-trend-ICBR	Geschatte trend gedeeld door de mediaan van de periode 2016 tot 2019 of de eerste drie jaar van de *meetreekswaarden*. Deze waarde is van belang voor het bepalen van de reductie van stoffen bij de ICBR-methode[9].
Signific?	Statistische significantie van de trend ('**Ja**', of '**Nee**').
Oordeel	Oordeel over de trend, stijging of daling. Dit oordeel luidt '**Geen trend/stijging/daling**' als de geschatte trend, stijging of daling statistisch niet significant is en anders

[9] ICBR, 2022. *Vermindering van microverontreinigingen in het Rijnstroomgebied – Monitoring en beoordelingssysteem.* Internationale Commissie ter Bescherming van de Rijn.

wordt het oordeel bepaald door de absolute waarde van de verhouding van trend en mediaan, volgens onderstaande tabel 5.2.

Tabel 5.2: Het oordeel over een trend, stijging of daling, hangt af van de statistische significantie en van de absolute waarde van de verhouding van trend en mediaan.

Trend significant?	|Trend/mediaan|	Oordeel
Nee	Irrelevant	Geen trend
Ja	<= 1%	Zeer kleine trend
	> 1% - <= 5%	Kleine trend
	> 5% - <= 10%	Matige trend
	> 10% - <= 25%	Grote trend
	>= 25%	Zeer grote trend

Stijging significant?	|Trend/mediaan|	Oordeel
Nee	Irrelevant	Geen stijging
Ja	<= 1%	Zeer kleine stijging
	> 1% - <= 5%	Kleine stijging
	> 5% - <= 10%	Matige stijging
	> 10% - <= 25%	Grote stijging
	>= 25%	Zeer grote stijging

Daling significant?	|Trend/mediaan|	Oordeel
Nee	Irrelevant	Geen daling
Ja	<= 1%	Zeer kleine daling
	> 1% - <= 5%	Kleine daling
	> 5% - <= 10%	Matige daling
	> 10% - <= 25%	Grote daling
	>= 25%	Zeer grote daling

5.9 Trendpalet

Het trendpalet is een Excel-uitvoerbestand dat elk van de in § 5.7 beschreven gegevens afzonderlijk presenteert in matrixvorm, zodat het betreffende gegeven voor alle combinaties van meetlocatie en parameter kan worden vergeleken. Elk gegeven is in matrixvorm gepresenteerd in een apart tabblad. Het laatste tabblad – *Rel trend oordeel* - is daarbij het belangrijkste (zie het voorbeeld in tabel 5.3).

Tabel 5.3: Voorbeeld van (een deel) van het trendpalet in het tabblad Rel trend oordeel.

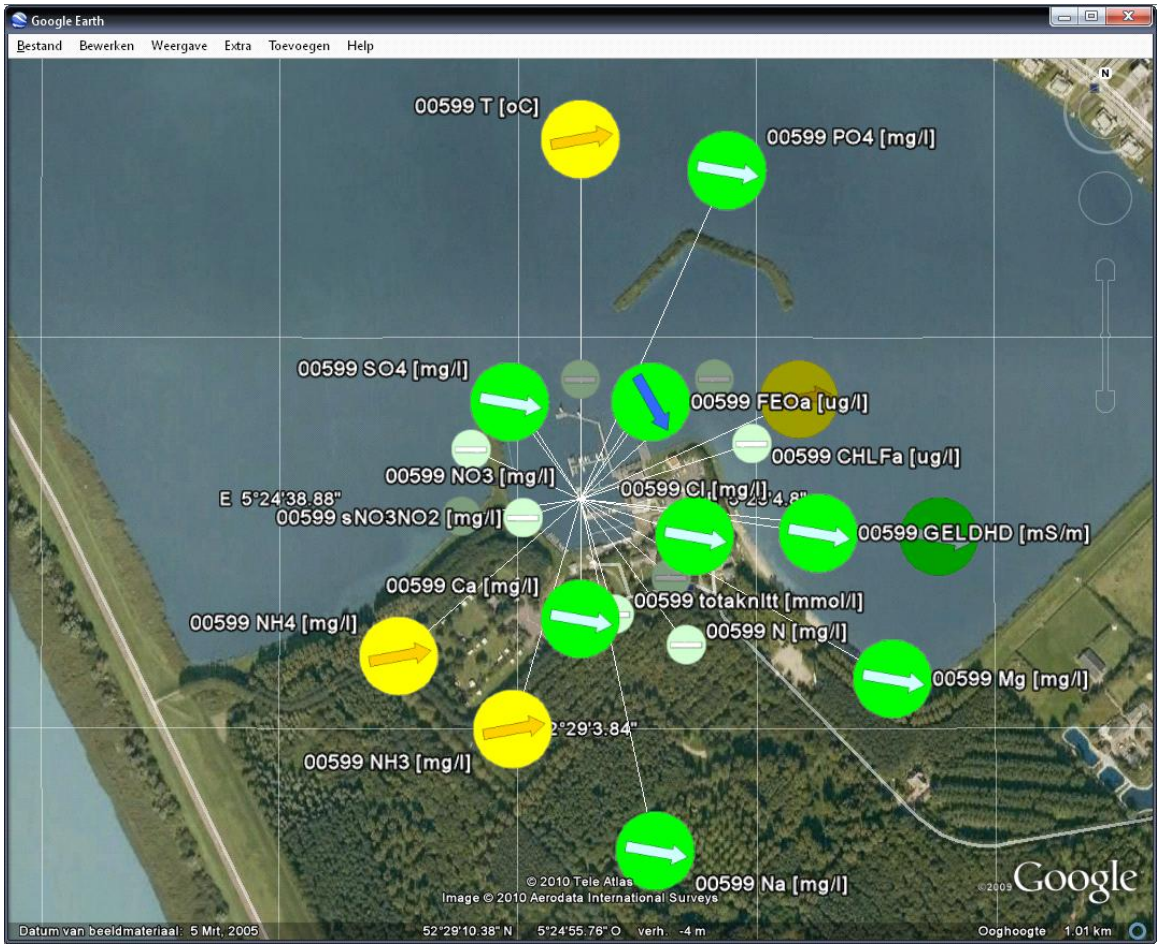
Parameter	Loc. A	Loc. B	Loc. C	Loc. D	Loc. E	Loc. F
ammonium	3,8%	Geen trend	Geen trend	10,8%	Geen trend	6,2%
cadmium	Geen trend	-17,7%	-17,8%	-8,1%	-12,5%	-17,4%
ijzer	Geen trend	-11,0%	Geen trend	12,8%	-10,0%	9,1%
koper	-3,8%	-11,2%	Geen trend	Geen trend	-9,3%	-12,7%
kwik	-21,1%	0,0%	0,0%	-11,3%	-0,1%	-16,5%
lood	Geen trend		-70,2%	-19,8%	-13,1%	-31,5%
nikkel	Geen trend	-3,0%	Geen trend	Geen trend	-3,6%	-1,3%
nitraat	Geen trend	Geen trend	Geen trend	-16,9%	-2,8%	-5,2%
nitriet	-2,5%	-25,0%	-3,1%	-16,1%	-25,0%	-2,0%
orthofosfaat	3,0%	Geen trend	Geen trend	3,1%	1,9%	2,2%
stikstof, kjeldahl	Geen trend	Geen trend	Geen trend	8,1%	Geen trend	6,4%
sulfaat	Geen trend	Geen trend	Geen trend	Geen trend	Geen trend	1,7%
totaal fosfaat	Geen trend	-4,1%	Geen trend	3,0%	Geen trend	4,3%
troebelingsgraad	Geen trend	Geen trend	Geen trend	9,1%	Geen trend	11,6%
zink	Geen trend	-26,7%	Geen trend	Geen trend	-25,0%	-25,0%
zuurgraad	0,1%	Geen trend	Geen trend	-0,2%	-0,3%	-0,2%
zuurstof	-0,9%	Geen trend	Geen trend	Geen trend	Geen trend	-6,3%

Elke cel in bovenstaand trendpalet geeft met een kleur de richting van de trend, per combinatie van meetlocatie en parameter. Een statistisch significante daling is groen gekleurd en een statistisch significante stijging is oranjerood gekleurd. Als er geen statistisch significante trend is gedetecteerd, dan is de cel geel gekleurd. Elke statistisch significante trend is relatief uitgedrukt als percentage verandering per jaar ten opzichte van de mediaan van de betreffende tijdreeks. Dit tabblad combineert de informatie in de tabbladen *Relatieve trend* en *Oordeel* (zij het alleen voor het oordeel 'Geen trend'), vandaar de naam *Rel trend oordeel*.
De standaardopmaak van de matrix is met de meetlocaties in kolom A en de parameters in rij 1. Maar als er meer dan 256 parameters zijn, staan de parameters in kolom A en de meetlocaties in rij 1 (zoals in tabel 5.3).

5.10 Trendkaart

Als het invoerbestand ook de coördinaten van de meetlocaties bevat, kunnen de resultaten van de trendanalyse worden weergegeven op een kaart. Daartoe wordt door *Trendanalist* het programma Google Earth aangeroepen. Figuur 5.5 geeft een voorbeeld van zo'n trendkaart.

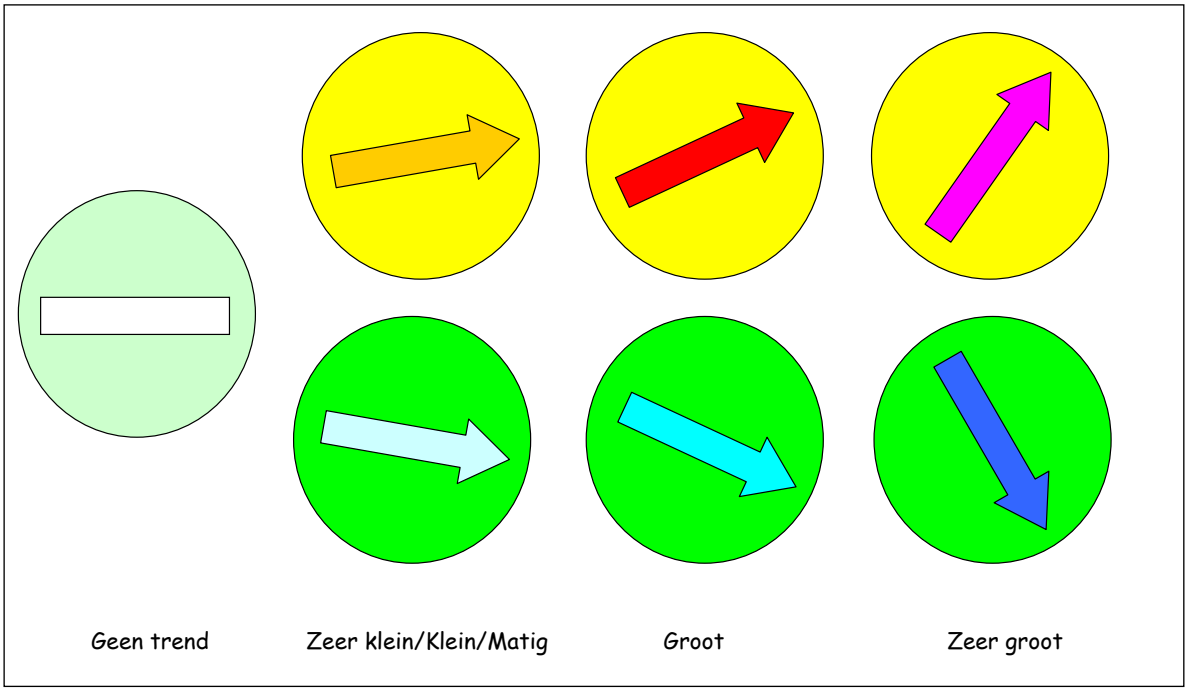
Figuur 5.5: Voorbeeld van een trendkaart. Het betreft hier de resultaten van één meetpunt. Oorspronkelijk liggen alle pictogrammen van een meetpunt over elkaar heen, maar als op het bovenste pictogram wordt geklikt, ontstaat een sterstructuur als hieronder.



In veel gevallen zal men per parameter een afzonderlijke trendkaart willen maken. De gebruiker moet dan voor de automatische trendanalyse alleen de meetreeksen van die parameter selecteren.

De oordelen over de trends worden op de kaart weergegeven met de pictogrammen van figuur 5.6.

Figuur 5.6: De pictogrammen die worden gebruikt om de oordelen over de trends weer te geven op de trendkaart. Zie tabel 5.2 voor de betekenissen van deze oordelen.



5.10.1.1 Geraadpleegde literatuur

Alden, R.W., Perry, E.S. and Lane, M.F. (2000). *A Comparison of Analytical Techniques for Determining Trends in Chesapeake Bay Water Quality Monitoring Program Data*. AMRL Technical Report # 3114. Applied Marine Research Laboratory, Norfolk, VA.

Baggelaar, P.K. en Baggelaar, D.H. (1989): *Project Trendbepaling. Rapportage 1e fase*. Kiwa-rapport SWO 89.335, augustus 1989, Nieuwegein.

Baggelaar, P.K. en Baggelaar, D.H. (1991): *Project Trendbepaling. Rapportage 2e fase*. Kiwa-rapport SWO 90.284, februari 1991, Nieuwegein.

Baggelaar, P.K. (1994): *A protocol for trend analysis of the input of pollutants to the North Sea*. Kiwa-rapport SWO 93.337, mei 1994, Nieuwegein.

Baggelaar, P.K. en Baggelaar, D.H. (1995): *Trends in de oppervlaktewaterkwaliteit van Rijn en Maas*. RIWA, Amsterdam en Kiwa, Nieuwegein, mei 1995.

Box, G.E.P. and Jenkins, J.M. (1976): *Time series analysis, forecasting and control*. Holden-Day, San Francisco.

Bradley, J.V. (1968): *Distribution-Free Statistical Tests*. Prentice Hall, Englewood Cliffs.

Cleveland, W.S. (1979): *Robust Locally Weighted Regression and Smoothing Scatterplots*. Journal of American Statistical Association 74, blz. 829 – 836.

Conover, W.J. (1980): *Practical nonparametric statistics*. John Wiley, New York.

Duin, R.N.M. (1993): *Trendanalyse van verschillende waterkwaliteitsvariabelen op 26 lokaties in de Noordzee over de periode 1975 t/m 1989*. Dienst Getijdewateren, Rijkswaterstaat (concept-rapport).

Harcum, J.B., Loftis, J.C. and Ward, R.C. (1992): *Selecting trend tests for water quality series with serial correlation and missing values*. Water Resources Bulletin, vol. 28, no. 3, blz. 469 – 478.

Helsel, D.R. and Hirsch, R.M. (1991): *Statistical Methods in Water Resources*. Studies in Environmental Science 49. Elsevier, Amsterdam, 510 blz.

Hirsch, R.M., Slack, J.R. and Smith, R.A. (1982): *Techniques of trend analysis for monthly water quality data*. Water Resources Research, vol. 18, no. 1, February 1982, blz. 107 – 121.

Hirsch, R.M. and Slack, J.R. (1984): *A nonparametric trend test for seasonal data with serial dependence*. Water Resources Research, vol. 20, no. 6, blz. 727 - 732.

Hirsch, R.M., Alexander, R.B. and Smith, R.A. (1991): *Selection of methods for the detection and estimation of trends in water quality*. Water Resources Research, vol. 27, no. 5, May 1991, blz. 803 – 813.

Kendall, M.G. (1938): *A new measure of rank correlation*. Biometrika, 30, 1938, blz. 81 - 93.

Kendall, M.G. (1975): *Rank Correlation Methods*. Charles Griffin, London, 1975.

Lilliefors, H.W. (1967): *On the Kolmogorov-Smirnov test for normality with mean and variance unknown*. Journal of the American Statistical Association, 62, blz. 399 – 402.

Lilliefors, H.W. (1969): *Correction to the paper: On the Kolmogorov-Smirnov test for normality with mean and variance unknown*. Journal of the American Statistical Association, 64, blz. 1702.

Ljung, G.M and Box, G.E.P (1978): *On a measure of lack of fit in time series models*. Biometrika, 65, blz. 297 – 303.

McLeod, A.I., Hipel, K.W. and Bodo, B.A. (1991): *Trend analysis methodology for water quality time series*. Environmetrics, 1991, 2(2), blz. 169 - 200.

Mann, H.B. (1945): *Non-parametric tests against trend*. Econometrica 13, blz. 245 - 259, 1945.

Sanders, T.G., Ward, R.C., Loftis, J.C., Steele, T.D., Adrian, D.D. amd Yevjevich, V. (1983): *Design of networks for monitoring water quality*. Water Resources Publications, Littleton, USA.

Sen, P.K. (1968): *Estimates of the regression coefficient based on Kendall's tau*. Journ. Am. Statist. Assoc., 63, blz. 1379 - 1389, 1968.

Stoffer, D.S. and Toloi, C.M.C. (1992): *A note on the Ljung-Box-Pierce portmanteau statistic with missing data*. Statistics & probability Letters 13 (1992), blz. 391 – 396.

Theil, H. (1950): *A rank-invariant method of linear and polynomial regression analysis, 1,2 and 3*. Ned. Akad. Wetensch. Proc., 53, blz. 386 - 392, 521 - 525 en 1397 - 1412.

Ward, R.C. and Loftis, J.C. (1986): *Establishing statistical design criteria for water quality monitoring systems: review and synthesis*. Water Resources Bulletin, vol. 22, no. 5, October 1986, blz. 759 - 767.

5.10.1.2 Bijlage – Verklaring van een aantal termen

Autocorrelatie: het verschijnsel dat opeenvolgende meetwaarden niet onafhankelijk van elkaar zijn. Als een milieugerelateerd proces met een hoge frequentie wordt waargenomen treedt doorgaans positieve autocorrelatie op, wat inhoudt dat opeenvolgende meetwaarden meer op elkaar lijken dan op verder in de tijd gelegen meetwaarden. Een aantal opeenvolgende meetwaarden zal daardoor meer de kenmerken weerspiegelen van een segment van de kansverdeling waar ze uit afkomstig zijn, dan van de gehele kansverdeling. Als hier geen rekening mee wordt gehouden, zal bijvoorbeeld de standaardafwijking van die kansverdeling worden onderschat.

Autocorrelatiecoëfficiënt: maat voor de lineaire samenhang tussen meetwaarden die zijn gescheiden door een bepaald tijdsinterval.

Autocorrelogram: grafiek die de autocorrelatiecoëfficiënt weergeeft als functie van het tijdsinterval tussen de meetwaarden.

Betrouwbaarheidsinterval: maat voor de precisie waarmee een bepaald kengetal, zoals een gemiddelde of een percentiel is geschat. Zo geeft het 95%-betrouwbaarheidsinterval van een geschat percentiel het interval aan waarbinnen het werkelijke percentiel zich ‘vast wel’ zal bevinden, namelijk in 95 van de 100 gevallen.

Correlatiecoëfficiënt: maat voor de lineaire samenhang tussen meetwaarden van twee verschijnselen (zoals de concentraties van twee parameters).

Gecensureerde meetwaarden: waarden zoals ‘< 1‘, waarbij de ‘1’ in dit voorbeeld staat voor de rapportagegrens. Bij een aantal analytisch chemische bepalingen wordt een meetwaarde onder een bepaald niveau slechts gecensureerd doorgegeven als zijn relatieve precisie door het betreffende laboratorium te laag wordt geacht.

Histogram: staafdiagram, waarin het aantal meetwaarden dat in een bepaalde grootte-klasse valt is uitgezet tegen de grootte-klasse.

Kansverdeling: functie die voor een bepaald verschijnsel (zoals de concentratie van een parameter) de kansen op realisatie geeft van alle mogelijke meetwaarden.

Lineair regressiemodel: statistisch model voor de lineaire relatie tussen twee verschijnselen (zoals de concentratie van een parameter en de tijd).

Meetinterval: de periode tussen twee opeenvolgende metingen.

Meetreeks: een chronologisch gerangschikte reeks meetwaarden van een bepaald verschijnsel, zoals de concentratie van een bepaalde stof, de grondwaterstandstand, of de afvoer.

Meetreeks- of tijdreeksplot: grafiek waarin de waarden van een reeks (meetwaarden, gemiddelden of voortschrijdende gemiddelden) zijn uitgezet tegen het meettijdstip.

Modelresidu: verschil tussen een gerealiseerde waarde van een bepaald verschijnsel (zoals de concentratie van een parameter) en de bijbehorende uitkomst van een model van dat verschijnsel.

Normale kansverdeling: veelvoorkomende symmetrische kansverdeling. Deze heeft een belvorm en is volledig te beschrijven aan de hand van het gemiddelde en de standaardafwijking.

Percentiel: kengetal van een kansverdeling. Het P-percentiel is de waarde die door P% van de kansverdeling wordt onderschreden en door (100-P)% wordt overschreden.

Standaardafwijking: kengetal van een kansverdeling. Het is een maat voor de spreiding in de meetwaarden.

Steekproef: een verzameling meetwaarden van een bepaald verschijnsel.

Transformeren: het wiskundig omzetten van meetwaarden, zoals het nemen van de logaritme, of machtsverheffen. Als meetwaarden niet afkomstig zijn uit een normale kansverdeling, dan lukt het soms om de meetwaarden zodanig te transformeren dat de getransformeerde waarden wél afkomstig zijn uit een normale kansverdeling.

Trendpalet: een Excelbestand waarin de verschillende soorten resultaten van de trendanalyse in afzonderlijke tabbladen in matrixvorm zijn gepresenteerd. Elke cel bevat het resultaat voor een bepaalde combinatie van meetlocatie en parameter.

Tijdreeks: chronologisch gerangschikte reeks equidistante waarden.

Uitschieters: meetwaarden die duidelijk afwijken van de andere meetwaarden.

Voortschrijdend gemiddelde: het (rekenkundig) gemiddelde van een aantal opeenvolgende meetwaarden. Naarmate er over meer opeenvolgende meetwaarden wordt gemiddeld, zal de reeks van het voortschrijdend gemiddelde minder fluctuatie vertonen.