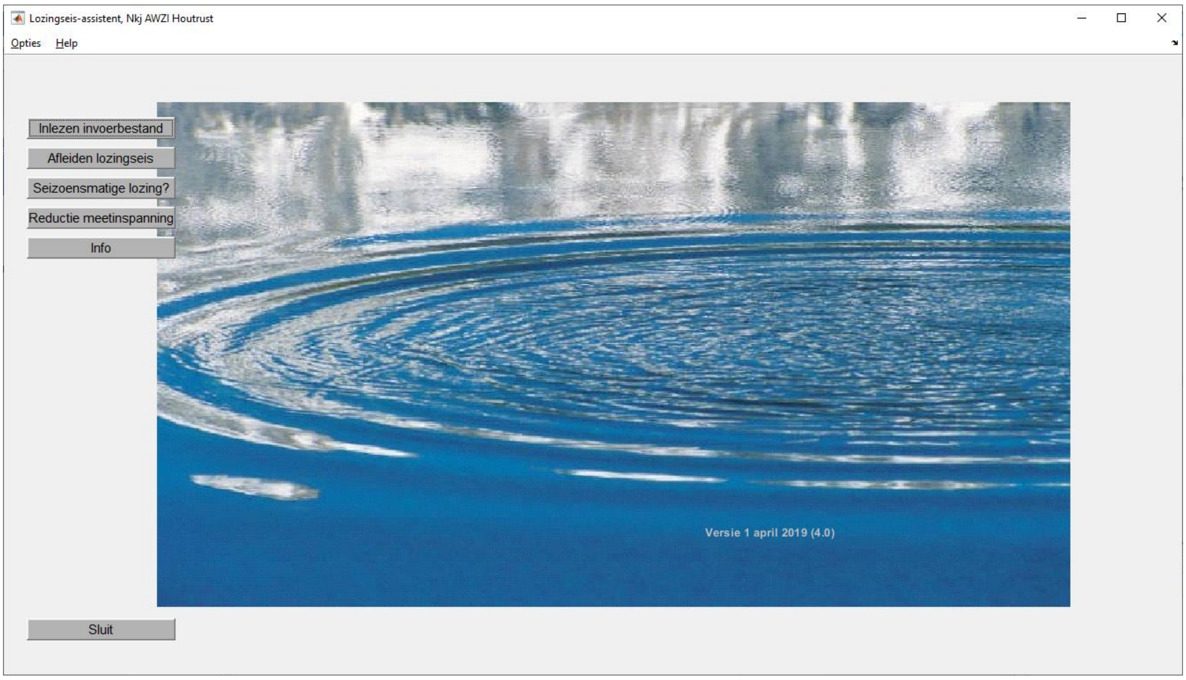
# Lozingseis-assistent

## *Gebruikershandleiding*




amo
Modellering en Optimalisatie

PB Icastat

22 juli 2024

# Lozingseis-assistent

## *Gebruikershandleiding*

Opdrachtgever: RIZA / CIW-4

Auteurs:
drs. Paul K. Baggelaar (PB Icastat Statistisch Adviesbureau)
ir. Eit C.J. van der Meulen (AMO, Adviesbureau Modellering en Optimalisatie)

1 juli 2024

Modellering en Optimalisatie

PB Icastat

amo@amo-nl.com
paul.baggelaar@planet.nl

## INHOUD

# 1 Inleiding

## 1.1 Achtergronden

Het programma Lozingseis-assistent is voornamelijk bedoeld voor Wvo-vergunningverleners en handhavers, opdat ze kunnen komen tot eenduidige, zo mogelijk uniforme, handhaafbare, onderscheidende en naleefbare lozingseisen. Maar uiteraard kunnen ook aanvragers van Wvo-vergunningen er gebruik van maken.
Het programma is ontwikkeld in opdracht van RIZA en de projectgroep Lozingseisen Wvo-vergunningen. Voor de theoretische overwegingen die ten grondslag hebben gelegen aan het programma verwijzen wij naar het rapport 'Statistische aspecten van lozingseisen' [Baggelaar, 2003]. Dit is in te zien via de 'Help'-knop van Lozingseis-assistent.

## 1.2 Mogelijkheden

Het programma Lozingseis-assistent kan vergunningverleners en handhavers ondersteunen bij de volgende werkzaamheden:

(1) het afleiden van lozingseisen voor meetwaarden en voor gemiddelden van 10 opeenvolgende meetwaarden (stap 8 van de nota 'Lozingseisen Wvo-vergunningen');
(2) het nagaan of er sprake is van een seizoensmatige lozing (stap 6 van de nota);
(3) het nagaan of er minder parameters kunnen worden gemeten door gebruik te maken van onderlinge relaties (stap 5 van de nota).

Zoals de naam al aangeeft is het programma bedoeld om te *assisteren* bij deze werkzaamheden. Om tot verantwoorde resultaten te kunnen komen, is naast dit programma, het gezonde verstand van de gebruiker onontbeerlijk. Zo is het bijvoorbeeld van groot belang dat de aan het programma verstrekte meetwaarden zo representatief mogelijk zijn voor de te vergunnen lozingssituatie (zie hiervoor § 2.3).

## 1.3 Leeswijzer

De eerste hoofdstukken van deze handleiding beschrijven de werking van de belangrijkste onderdelen van het programma Lozingseis-assistent, namelijk:

- hoofdstuk 2: het inlezen van het invoerbestand;
- hoofdstuk 3: het afleiden van lozingseisen op basis van de statistische karakteristieken van de ingelezen meetreeks;
- hoofdstuk 4: het nagaan of er sprake is van een seizoensmatige lozing;
- hoofdstuk 5: het nagaan of de relaties tussen parameters zodanig sterk zijn dat er minder parameters kunnen worden gemeten.

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

Deze handleiding heeft twee bijlagen. Bijlage 1 bevat een verklaring van enkele technische – veelal statistische - termen die in de handleiding worden gebruikt. En bijlage 2 geeft enkele oplossingen voor de gevallen dat er geen lozingseis voor gemiddelden is afgeleid.

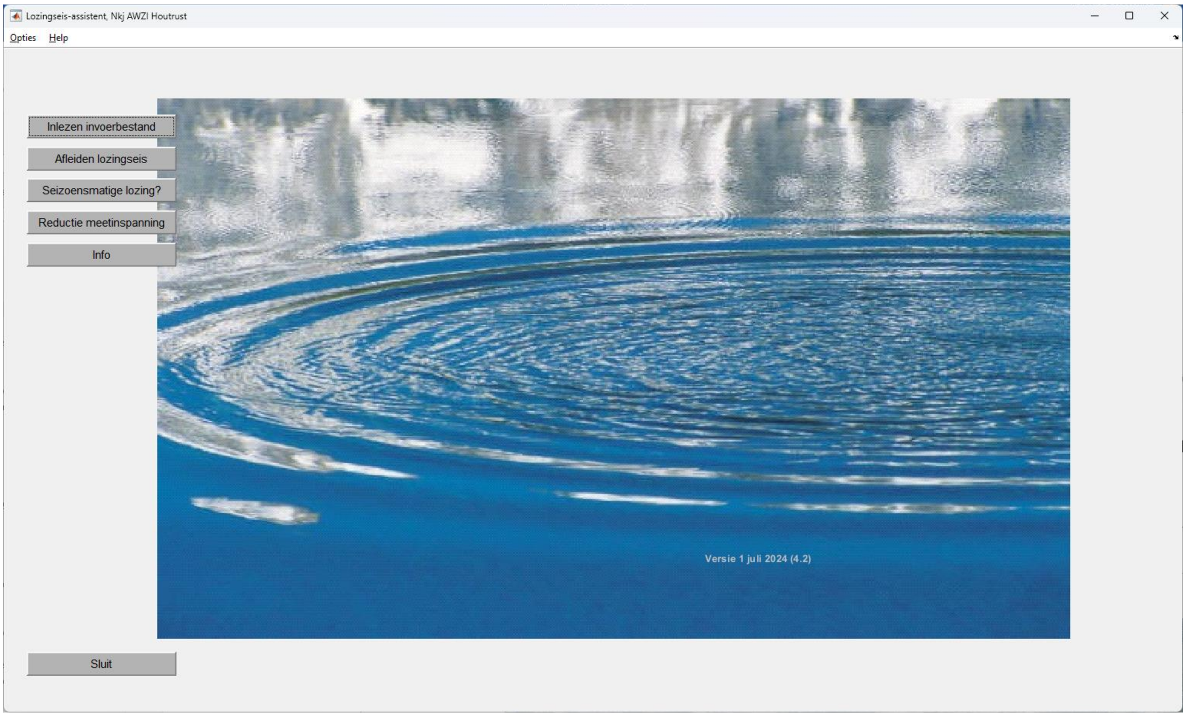
De handleiding richt zich op twee niveaus tot de gebruiker. Het eerste niveau beschrijft kort en bondig de bedoeling en werking van het programma, terwijl het tweede niveau – dit is in een kleiner lettertype geplaatst in de tekstkaders – meer technische (statistische) achtergronden geeft. Belangrijke waarschuwingen zijn geel gearceerd.

**Het wordt zeer sterk aanbevolen deze gebruikershandleiding goed te lezen (en ook bijlage 2), alvorens met het programma te starten !**

# 2 Inlezen invoerbestand

## 2.1 Bedoeling

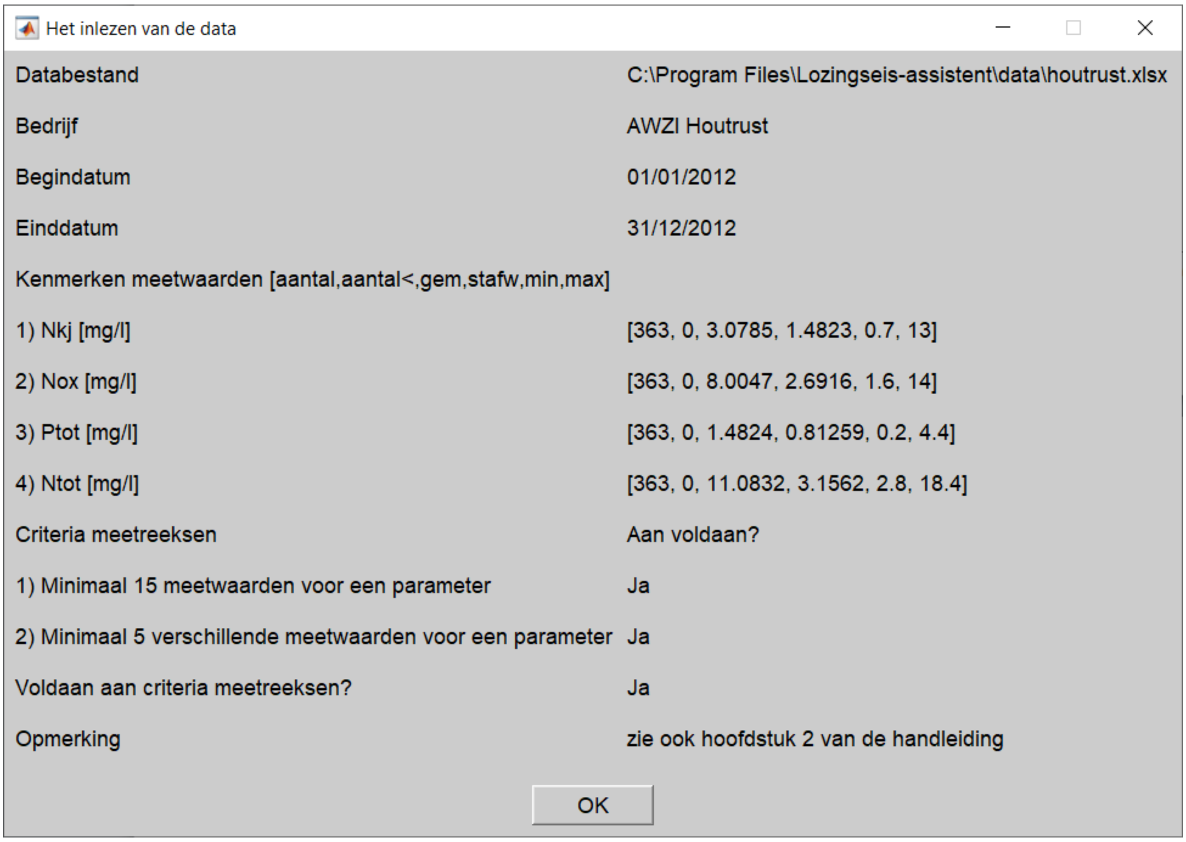
Als het programma is opgestart, verschijnt het hoofdscherm en is nog alleen de knop ‘Inlezen invoerbestand’ beschikbaar (zie onderstaande figuur). Door deze aan te klikken kan een bestand met meetreeksen van één of meer parameters worden ingelezen. In § 2.2 is aangegeven hoe een dergelijk bestand moet zijn opgebouwd.




Als het bestand is ingelezen, verschijnt er een schermpje dat de kenmerken van het ingelezene vermeldt. Dit betreft de bestandsnaam, de bedrijfsnaam, de begin- en einddatum van de meetreeks(en) en per ingelezen parameter het aantal meetwaarden, het aantal ‘<‘-waarden (zogenaamde gecensureerde waarden, zoals < 1 mg/l, zie ook bijlage 1), het gemiddelde, de standaardafwijking, het minimum en het maximum. Tenslotte vermeldt dit scherm of wordt voldaan aan de eisen die het programma stelt aan de meetreeks(en), namelijk:

(1) een meetreeks moet minimaal 15 meetwaarden bevatten;
(2) een meetreeks moet minimaal 5 verschillende meetwaarden bevatten (een vrijwel constante meetreeks kan dus niet worden geanalyseerd).

Als niet door alle meetreeksen wordt voldaan aan beide eisen, is een verdere analyse met het programma niet mogelijk. Van elke meetreeks die niet voldoet aan beide eisen zijn de kenmerken rood aangegeven in het scherm (zie onderstaand voorbeeld), zodat de gebruiker die reeks kan aanpassen of verwijderen. Een dergelijke handeling dient dan echter buiten het programma om te gebeuren.




Een meetreeks met nullen kan problemen opleveren. Vervang een gecensureerde meetwaarde (zoals < 1 mg/l) dus bij voorkeur **niet** door een nul. Dit is overigens ook al vanuit statistisch oogpunt sterk af te raden, want het vertekent het gemiddelde.

Als het ingelezene voldoet aan de eisen komen er meer knoppen beschikbaar en biedt het programma de gebruiker de volgende drie mogelijkheden:

(1) Afleiden lozingseis - het afleiden van een lozingseis voor een parameter, op basis van de statistische karakteristieken van zijn meetreeks (zie hoofdstuk 3);
(2) Seizoensmatige lozing - het nagaan of er voor een parameter sprake is van een seizoensmatige lozing (zie hoofdstuk 4);
(3) Reductie meetinspanning - het nagaan of de relaties tussen parameters zodanig sterk zijn dat er minder parameters kunnen worden gemeten (zie hoofdstuk 5).

## 2.2 Vereiste formaat invoerbestand

Lozingseis-assistent kan meetreeksen inlezen uit een tekstbestand of uit een Excel-bestand. De extensie van het tekstbestand mag hierbij echter alleen '.txt' of '.loz' zijn. Nieuw in versie 4 is dat de gegevens ook uitgelezen kunnen worden uit tabbladen van een Excel-bestand met de extensie '.xlsx'.

*Details van het vereiste formaat van het tekst-invoerbestand*

Het bestand mag meetreeksen van meerdere parameters bevatten. Als het bijvoorbeeld drie parameters bevat moet het formaat als volgt zijn:

1$^e$ regel: naam van het bedrijf
2$^e$ regel: naam parameter 1 ; naam parameter 2 ; naam parameter 3 ; etc.
3$^e$ regel: meeteenheid parameter 1 ; meeteenheid parameter 2 ; meeteenheid parameter 3 ; etc.
4$^e$ en volgende regels: datum tijd meetwaarde parameter 1 meetwaarde parameter 2 meetwaarde parameter 3 monstersoort

Als decimaal scheidingsteken van een meetwaarde wordt zowel de punt ('.'), als de komma (',') geaccepteerd. Scheidingstekens voor duizendtallen worden niet geaccepteerd.

Het scheidingsteken ';' (tussen parameternamen en tussen meeteenheden) mag ook worden voorafgegaan en/of gevolgd door één of meer spaties.

De meetreeks moet zijn gesorteerd op datum (oplopend). De datum heeft als formaat dag-maand-jaar, waarbij alleen nummers mogen worden gebruikt (dus voor januari een '1' en geen 'jan'). Het scheidingsteken mag een '/' zijn (voorbeeld: 26/1/2002), of een '-' (voorbeeld: 26-1-2002). Het jaartal mag met 4 cijfers zijn (voorbeeld: 1999), of met 2 cijfers (voorbeeld: 99). De tijd *hoeft niet* te zijn vermeld, aangezien de kleinste tijdseenheid een dag is. Als deze wel is vermeld, moet dit met het formaat uur:minuut, waarbij elk uit 1 of 2 cijfers bestaat (voorbeelden: 13:21 en 7:5).

Een meetwaarde kan ook worden voorafgegaan door een '<'-teken (zonder spatie). Dit wordt door het programma geïnterpreteerd als een gecensureerde waarde en vervolgens intern op de helft van de betreffende rapportagegrens gezet (bijvoorbeeld: < 1 mg/l wordt omgezet tot 0,5 mg/l).

Er mogen meetwaarden ontbreken, maar deze moeten dan wel zijn weergegeven als een asterisk ('*'). De meetreeksen van alle parameters in het bestand moeten even lang zijn.

De monstersoort moet als volgt zijn aangegeven:

- V24H als het een dagverzamelmonster betreft en
- S als het een steekmonster betreft.

Een bestand mag alleen dagverzamelmonsters of steekmonsters bevatten. Als ze samen in het bestand voorkomen volgt er een foutmelding. Zorg dus voor een homogeen bestand.

Het bestand mag in het deel boven de meetwaarden ook commentaarregels bevatten. De eerste kolom van zo'n regel moet echter worden aangegeven met het **%**-teken.

*Voorbeeld van de eerste regels van een tekst-invoerbestand met twee parameters*

```
% data zoals beschikbaar op 12 september 2005
Jansen Conservenindustrie
CZV ; N-Kj
mg/l ; mg/l
6-1-1999 0:00 285 5.3 V24H
7-1-1999 0:00 293  *   V24H
8-1-1999 0:00 276 4.8 V24H
```

### *Details van het vereiste formaat van het Excel-invoerbestand*

Het vereiste formaat van een tabblad in een Excel-bestand is direct vergelijkbaar met die van het tekstbestand, alleen bevatten nu de cellen in de eerste rij de naam van het bedrijf en de parameters en de cellen in de tweede rij de datum, de eenheden en eventueel het monstersoort. Als het Excel-bestand bijvoorbeeld drie parameters bevat moet het formaat als volgt zijn:

| |
|---|
| 1e rij: naam van het bedrijf, naam parameter 1, naam parameter 2, naam parameter 3, etc.<br>2e rij: 'datum', meeteenheid parameter 1, meeteenheid parameter 2, meeteenheid parameter 3, etc.<br>3e en volgende rijen: datum tijd meetwaarde parameter 1 meetwaarde parameter 2 meetwaarde parameter 3 monstersoort |

Bij meerdere tabbladen in het Excel-bestand geeft Lozingseis-assistent de lijst van de tabbladen waaruit een enkel tabblad gekozen kan worden.

*Voorbeeld van de eerste rijen van een Excel-invoerbestand met vier parameters*

| % AWZI Houtrust | | | | | |
|---|---|---|---|---|---|
| AWZI Houtrust | Nkj | Nox | Ptot | Ntot | |
| Datum | mg/l | mg/l | mg/l | mg/l | |
| 1-jan-12 | 5,7 | 5,5 | 1,4 | 11,2 | V24H |
| 2-jan-12 | 2,3 | 3,6 | 0,8 | 5,9 | V24H |
| 3-jan-12 | 5,2 | 5,7 | 1,2 | 10,9 | V24H |

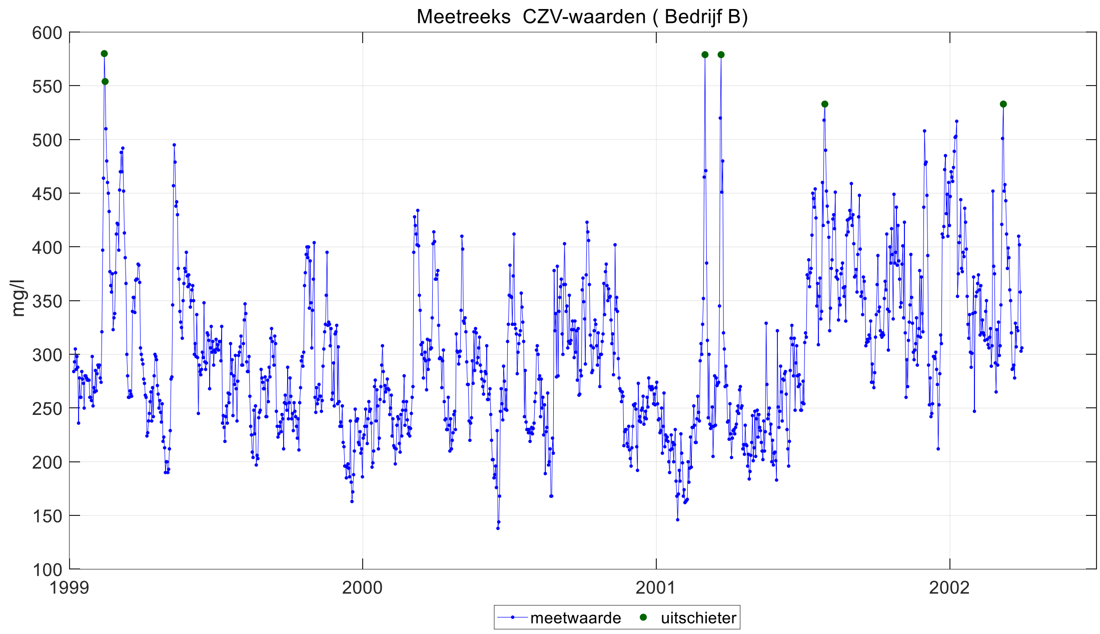
De subdirectory 'Data' van de directory waarin Lozingseis-assistent is geplaatst bevat enkele voorbeelden van bestanden. Deze zijn tevens bedoeld om ervaring met het programma te kunnen opdoen.

## 2.3 Zorg voor representatieve informatie!

Lozingseis-assistent dient zijn berekeningen alleen te baseren op de kenmerken van de meetwaarden bij de gebruikelijke, beheerste procesvoering. Anders zullen de berekende resultaten de plank volledig misslaan. Een lozingseis die is berekend met niet-representatieve meetwaarden zal immers ongeschikt zijn om er afwijkingen van de gebruikelijke, beheerste procesvoering mee te kunnen detecteren. Het is daarom zeer nodig er op aan te dringen en ook op toe te zien dat de aanvrager de vergunningverlener *voldoende* informatie verschaft die *representatief is voor de situatie waar de Wvo-vergunning voor dient te gelden*.

Helaas is niet aan te geven welk aantal meetwaarden en/of welke lengte van de meetreeks minimaal nodig is om een verantwoorde lozingseis af te kunnen leiden, doordat lozingsprocessen zoveel verschillende kenmerken en patronen kunnen vertonen. Als de concentratie van een geloosde parameter steeds volledig aselect rond een constant gemiddelde schommelt en ook geen uitschieters naar boven vertoont, kan bijvoorbeeld nog wel worden volstaan met 30 à 40 onafhankelijke meetwaarden van die concentratie, om een verantwoorde lozingseis af te kunnen leiden. Maar in andere gevallen zal het verloop van een concentratie een geleidelijke of stapsgewijze golfbeweging vertonen, of zelfs een trend. Het gemiddelde is dan niet constant (zie het voorbeeld in onderstaande figuur).

*Voorbeeld van een concentratieverloop met een golfbeweging. Als een lozingseis bij een dergelijk lozingsproces wordt afgeleid met té korte deelreeks, zal deze ongeschikt zijn voor toepassing.*




De golfbeweging kan veroorzaakt zijn door wijzigingen in de invoer van het proces (zoals een andere samenstelling van grondstoffen, of een influent met andere karakteristieken). Als dan een meetreeks wordt verstrekt die slechts een deel van deze golfbeweging vertegenwoordigt, zal het programma een verkeerde lozingseis afleiden. In sommige gevallen kan er zelfs minimaal een meetreeks van twee à drie jaar zijn vereist om tot een verantwoorde lozingseis te kunnen komen. Het oordeel over de representativiteit van een verstrekte meetreeks dient dan ook mede gebaseerd te worden op inzicht in:

- de fysische, chemische en/of biologische karakteristieken van het achterliggende proces en
- de gebruikelijkerwijs optredende veranderingen in de invoer van dat proces.

Om verder te waarborgen dat de meetreeks representatief is voor de situatie waar de Wvo-vergunning voor dient te gelden, dienen meetwaarden die fout zijn, of die zijn verkregen tijdens ongewone voorvallen, reeds door de aanvrager te zijn verwijderd, ongeacht of het een voorval betreft dat is voorzien, zoals onderhoud, of een voorval dat niet is voorzien, zoals een calamiteit. Om te voorkomen dat de vergunningverlener wordt opgezadeld met de opschoning van de meetwaarden – een klus die zeer arbeidsintensief kan zijn en zonder detailkennis van de lozing ook nauwelijks te objectiveren valt -, is het tevens raadzaam te eisen dat de aanvrager van elke uitschietende meetwaarde aantoont dat die nog de gebruikelijke, beheerste situatie vertegenwoordigt.

*Beperkte interne controle*

Het programma voert een beperkte controle uit om na te gaan of de informatieinhoud van de te analyseren meetreeks voldoende is om een verantwoorde lozingseis af te leiden. Er wordt namelijk nagegaan of de meetreeks kenmerken vertoont van één of andere grootschalige structuur, zoals een trend of een golfbeweging. Dit komt naar voren als een sterke autocorrelatie (zie het tekstkader), die slechts langzaam afneemt met het tijdsinterval tussen meetwaarden. Vervolgens bepaalt het programma of de reekslengte voldoende is om bij het berekenen van de lozingseis op verantwoorde wijze rekening te kunnen houden met die grootschalige structuur. Als dat niet het geval is, wordt er geen lozingseis berekend en wordt de reden toegelicht.

*Toelichting op autocorrelatie*

Autocorrelatie is 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. Dit leidt dan tot een te krappe lozingseis.

Deze interne controle geeft echter nog onvoldoende garantie dat berekende lozingseisen altijd bruikbaar zijn in de praktijk. Het programma wordt namelijk op een verkeerd been gezet als het een meetreeks krijgt aangeleverd, die alleen een min of meer stabiel deel van een grootschalige structuur vertegenwoordigt. Bij de interne controle wordt dan geen grootschalige structuur ontdekt en wordt een lozingseis afgeleid, die geen rekening houdt met de grote niveauschommelingen die buiten de geleverde meetreeks optreden. Het programma mag dus nooit een excuus worden om het gezond verstand uit te schakelen!

# 3 Afleiden lozingseis

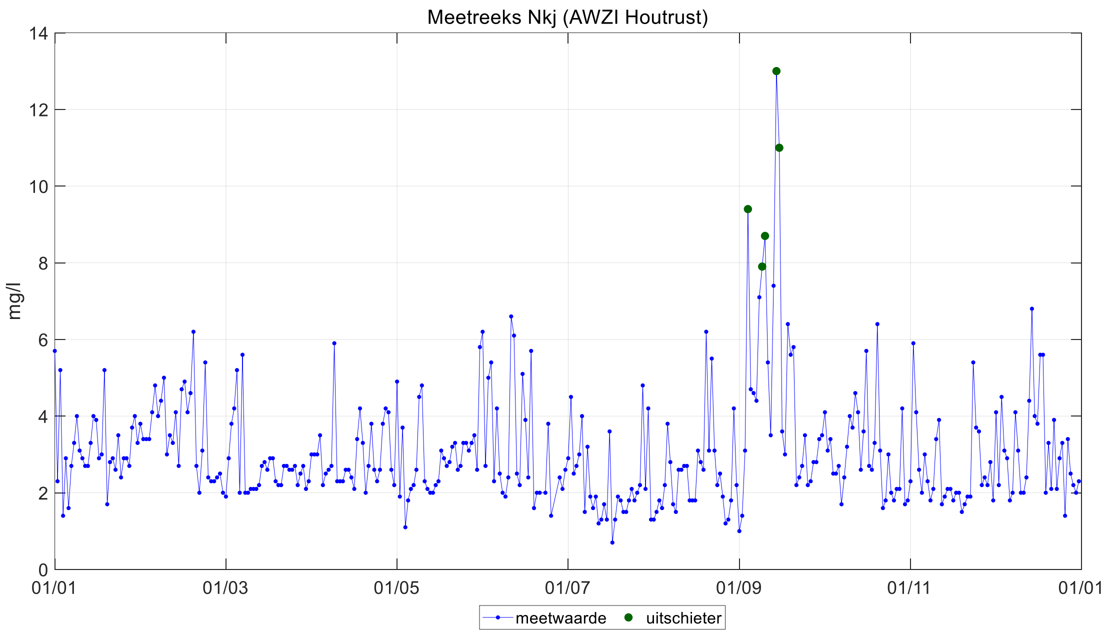
## 3.1 Bedoeling

Als een geldig invoerbestand is ingelezen, kan met behulp van het programma de lozingseis voor een parameter worden afgeleid. Daartoe moet de knop ‘Afleiden lozingseis’ worden aangeklikt. Doordat de lozingseis afhangt van de statistische kenmerken van de betreffende parameter, moeten eerst die kenmerken worden vastgesteld. Het programma zal de gebruiker hiertoe alle benodigde ondersteuning geven.

## 3.2 Verkenning te beschouwen meetreeks

Als het ingelezen invoerbestand slechts één meetreeks bevat, zal deze direct na het aanklikken van ‘Selectie parameter’ als tijdreeksplot worden getoond. Dit is een grafiek waarin de meetwaarden van de parameter zijn uitgezet tegen de tijd (zie onderstaande figuur). Maar als het invoerbestand meetreeksen van meerdere parameters bevat, dan verschijnt er eerst een schermpje waarop de gebruiker moet aangeven welke van de ingelezen parameters moet worden beschouwd. Daarna verschijnt de tijdreeksplot van de geselecteerde parameter.

*Voorbeeld van de tijdreeksplot die wordt getoond na het inlezen van een meetreeks.*




In de tijdreeksplot zijn de uitschieters - dit zijn de meetwaarden die volgens een bepaalde berekening duidelijk afwijken van de andere meetwaarden (zie § 3.2.1) - groen gemarkeerd.

### 3.2.1 Meetwaarden kunnen interactief worden verwijderd

Indien gewenst kan een meetwaarde interactief worden verwijderd, door deze in de tijdreeksplot aan te klikken. Hiervoor komen in ieder geval díe meetwaarden in aanmerking, die door een uitzonderlijke fout of een onbeheerste situatie zijn veroorzaakt. 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 tijdreeksplot. Desgewenst kan deze waarde ook weer aan de meetreeks worden toegevoegd door hem opnieuw aan te klikken.

*Toelichting op uitschieters*

Uitschieters zijn meetwaarden die duidelijk afwijken van de andere meetwaarden uit dezelfde steekproef. Ze kunnen zijn veroorzaakt door:

(1) uitzonderlijke bemonsterings- of meetfouten en/of transcriptiefouten (fouten bij het schrijven en/of typen);
(2) een onbeheerste situatie;
(3) een extreem geval van de gebruikelijke, beheerste situatie.

Voor onze toepassing is het van belang dat meetwaarden veroorzaakt door uitzonderlijke fouten (1) of onbeheerste situaties (2) niet meegenomen worden. Zonder detailkennis van de lozing valt het verwijderen van uitschieters echter nauwelijks te objectiveren. De aanvrager dient daarom slechts meetwaarden te verstrekken die representatief zijn voor de gebruikelijke, beheerste procesvoering. En voor elke nog resterende uitschieter dient de aanvrager bij voorkeur aan te tonen of te beargumenteren dat deze wél representatief is voor de gebruikelijke, beheerste procesvoering.

Het programma identificeert uitschieters door eerst de reeks van meetwaarden ($x_1$, $x_2$, ..., $x_n$) om te zetten tot een reeks van gestudentiseerde afwijkingen ($d_1$, $d_2$, ..., $d_n$), volgens:

$$d_i = \frac{x_i - \overline{x}_{(-i)}}{s_{(-i)}}$$

met $\overline{x}_{(-i)}$ en $s_{(-i)}$ het geschatte gemiddelde respectievelijk de geschatte standaardafwijking van de kansverdeling waar de meetwaarden uit afkomstig zijn, berekend *zonder* de betreffende meetwaarde $x_i$. 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 groen gemarkeerd als uitschieter in de tijdreeksplot.

## 3.3 Beschikbare knoppen

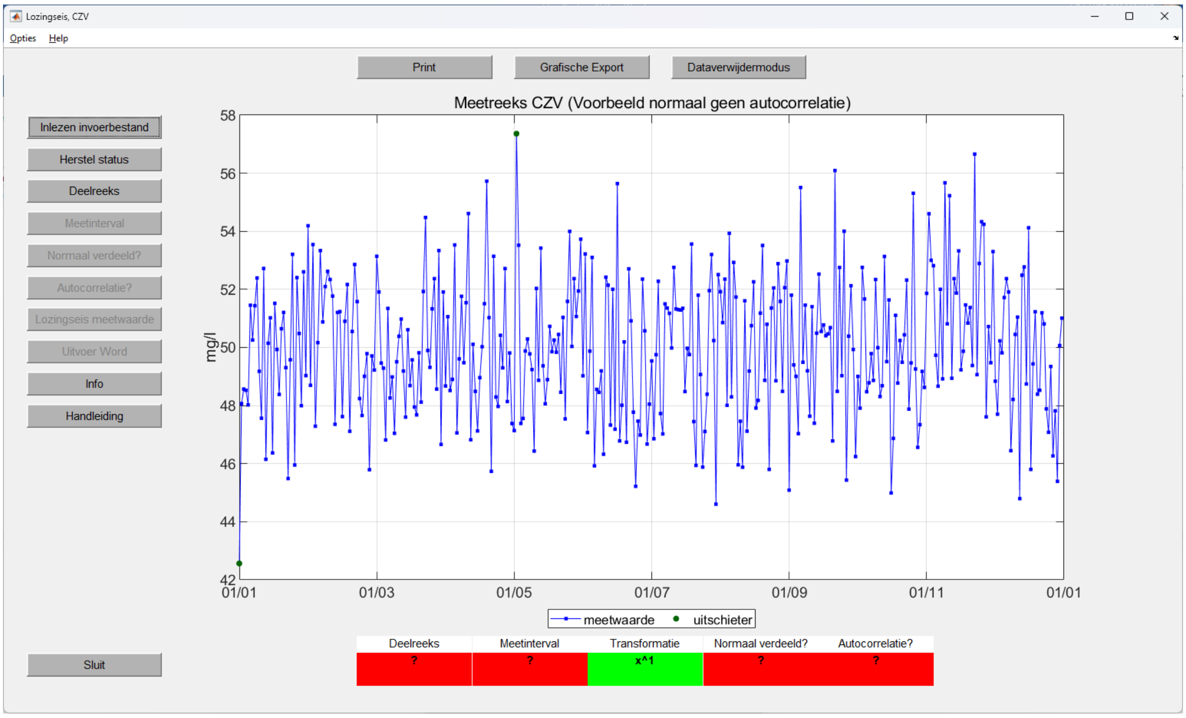
De volgende knoppen zijn verder beschikbaar voor dit onderdeel:

- Herstel status
- Deelreeks
- Meetinterval
- Normaal verdeeld?
- Autocorrelatie?
- Lozingseis
- Uitvoer naar Word
- Info
- Handleiding

**De statusbalk**

Om de gebruiker te ondersteunen bij het verstrekken van de informatie om de lozingseis af te kunnen leiden, wordt onderin het scherm een statusbalk weergegeven. Deze geeft per onderdeel aan welke informatie reeds beschikbaar is (groen veld) en welke nog moet worden ingevuld (rood veld). Om de lozingseis te kunnen berekenen (met de knop ‘Lozingseis’) dienen alle velden van deze statusbalk ingevuld - en dus groen- te zijn. Hiertoe moeten achtereenvolgens de knoppen ‘Deelreeks’, ‘Meetinterval’, ‘Normaal verdeeld?’ en ‘Autocorrelatie?’ worden aangeklikt. Ze zijn ook in deze volgorde geplaatst (van boven naar beneden).

*Scherm dat wordt getoond na de selectie van de parameter, met onderaan de statusbalk. De transformator staat bij aanvang op 1, wat wil zeggen dat er wordt begonnen zónder transformatie van de meetwaarden ($x^1$ is immers gelijk aan x).*



## 3.4 Herstel status

De knop 'Herstel status' is een herstelknop, die lopende een sessie kan worden aangeklikt, zodat opnieuw kan worden begonnen met de ingelezen meetreeks. Door deze aan te klikken, wordt de statusbalk weer in zijn oorspronkelijke staat teruggebracht en vormt de oorspronkelijk ingelezen meetreeks weer het uitgangspunt. Het is dus niet mogelijk om afzonderlijke stappen terug te gaan in het proces, men keert via deze knop terug naar het startpunt.

## 3.5 Deelreeks

Als de knop 'Deelreeks' wordt aangeklikt verschijnt er een schermpje dat de begin- en einddatum aangeeft van de geselecteerde meetreeks. Als de lozingseisen moeten worden afgeleid met alleen een bepaald deel van deze meetreeks, dan dient de gebruiker de begin- en/of einddatum in dit schermpje aan te passen. Als een deelreeks wordt geselecteerd die niet voldoet aan de criteria (minimaal 15 meetwaarden, waarvan minstens 5 verschillend zijn), zal het programma dat melden.

## 3.6 Meetinterval

Als de knop 'Meetinterval' wordt aangeklikt verschijnt een histogram van het meetinterval[1] van de meetwaarden in de beschikbare deelreeks. Dit histogram toont voor elk meetinterval hoe vaak dat in de meetreeks voorkomt. Daarbij is tevens een schermpje weergegeven dat reeds het meest voorkomende meetinterval aangeeft. Dat meetinterval zal worden gehanteerd bij het afleiden van de statistische kenmerken van deze deelreeks. Indien gewenst kan de gebruiker dit meetinterval in het schermpje aanpassen. Als een zodanig meetinterval wordt geselecteerd dat de deelreeks niet voldoet aan de criteria (minimaal 15 meetwaarden, waarvan minstens 5 verschillend zijn), zal het programma dat melden.

Na het kiezen van het meetinterval verschijnt weer de tijdreeksplot, maar ditmaal alleen van de meetwaarden voor het betreffende meetinterval. Als een meetwaarde ontbreekt, is de blauwe lijn onderbroken.

Als er verschillende meetintervallen voorkomen en de gebruiker wenst dat álle meetwaarden worden gehanteerd, dan moet hier '0' worden ingevuld. Er kan dan echter niet meer worden gecorrigeerd voor een eventuele autocorrelatie. Dit heeft als consequenties dat:

(1) de berekende lozingseis voor meetwaarden slechts bruikbaar kan worden geacht als de geanalyseerde meetreeks (min of meer) representatief is voor het hele bereik aan mogelijke meetwaarden bij de gebruikelijke, beheerste procesvoering;
(2) een lozingseis voor gemiddelden van 10 opeenvolgende meetwaarden niet berekend kan worden. Om die lozingseis toch af te kunnen leiden zal de gebruiker de sessie moeten herhalen, maar bij 'Meetinterval' moet dan een waarde groter dan 0 worden aangegeven (zie ook bijlage 2). Houd er wel rekening mee dat de afgeleide lozingseis voor gemiddelden dan alleen geldt voor gemiddelden van 10 opeenvolgende meetwaarden bij dát specifieke meetinterval.

[1] Het meetinterval is de periode tussen twee opeenvolgende metingen.

## 3.7 Normaal verdeeld?

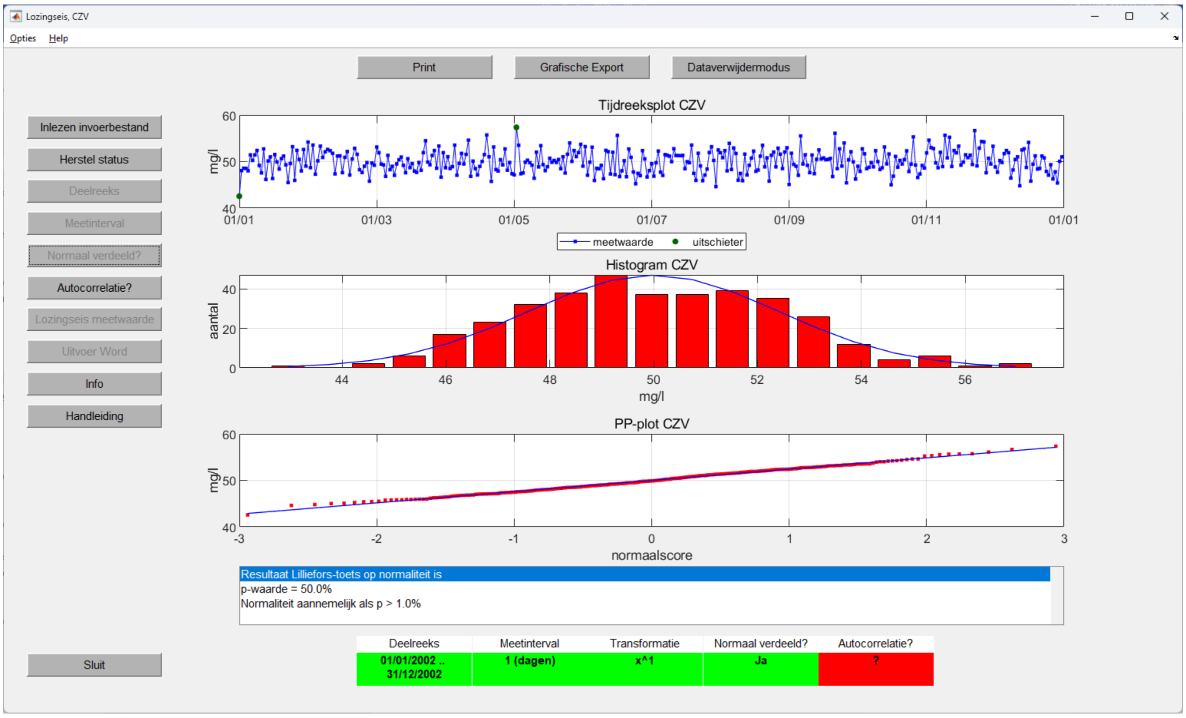
Het biedt voordelen als kan worden uitgegaan van een normale kansverdeling, aangezien dan een scherpere lozingseis kan worden gesteld dan wanneer dat níet het geval is. Dit komt door de vele theoretische kennis die beschikbaar is over de eigenschappen van die kansverdeling. Vandaar dat met het programma wordt nagegaan of er sprake is van een normale kansverdeling, zonodig na transformatie van de meetwaarden (§ 3.7.1).

Als de knop ‘Normaal verdeeld?’ wordt aangeklikt, verschijnen er vier hulpmiddelen om de gebruiker te assisteren bij het formuleren van zijn oordeel over het soort kansverdeling waar de meetwaarden uit afkomstig zijn. Dit zijn drie visuele hulpmiddelen (zie onderstaande figuur), aangevuld met het resultaat van een statistische toets:

(1) een tijdreeksplot van de meetwaarden;
(2) het histogram van de meetwaarden;
(3) de PP-plot van de meetwaarden;
(4) het resultaat van de Lilliefors-toets op normaliteit van de meetwaarden.

Tevens verschijnt er een schermpje dat de gebruiker vraagt of de meetwaarden al dan niet afkomstig zijn uit een normale kansverdeling.

*De visuele hulpmiddelen om een oordeel te kunnen vormen over het soort kansverdeling waar de meetwaarden uit afkomstig zijn. Op basis van de onderstaande grafieken heeft de gebruiker (terecht) geconcludeerd dat de betreffende meetwaarden afkomstig zijn uit een normale kansverdeling. Het statusveld bevat daardoor nu de tekst ‘Ja’ en is door het invullen ook groen geworden.*




ad (1) Tijdreeksplot

De tijdreeksplot is een grafiek waarin de meetwaarden zijn uitgezet tegen het meettijdstip.
Als de meetwaarden afkomstig zijn uit een normale kansverdeling, zullen zij min of meer symmetrisch rond hun gemiddelde liggen.

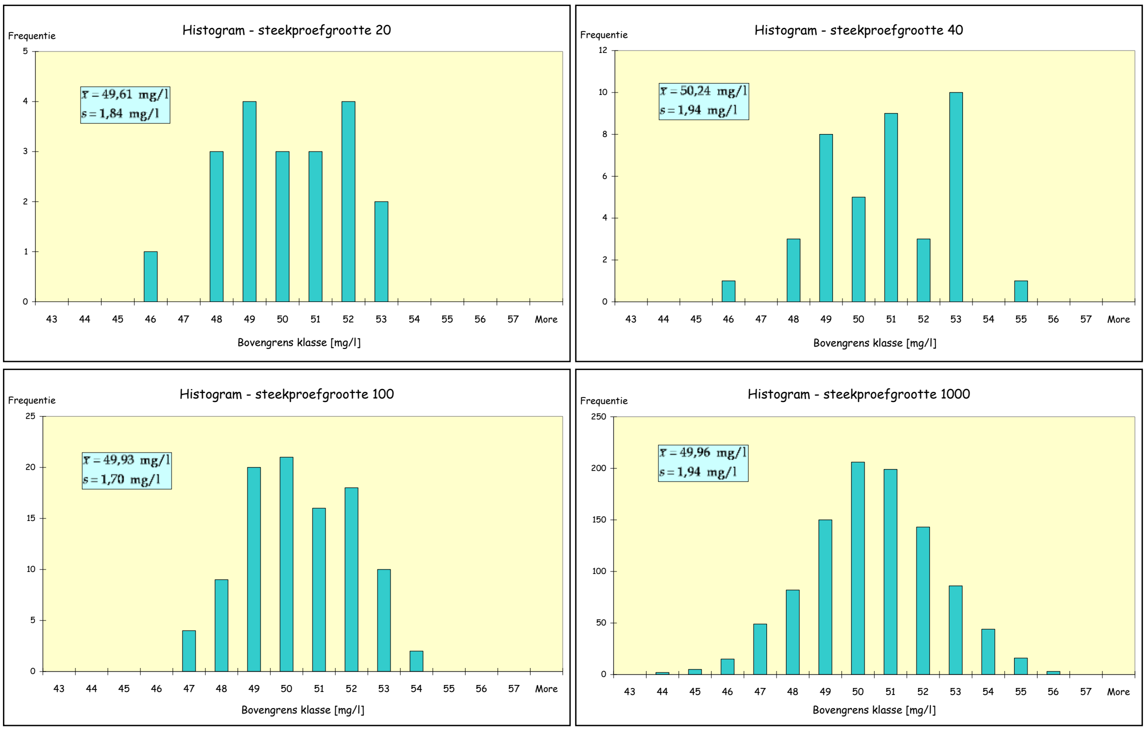
ad (2) Histogram van de meetwaarden

Het histogram toont voor elke grootte-klasse het aantal meetwaarden dat in die klasse valt. Tevens is met een blauwe lijn de vorm van de normale kansverdeling weergegeven die hetzelfde gemiddelde en standaardafwijking heeft als de meetwaarden. Als de reeks meer dan 100 meetwaarden bevat en deze afkomstig zijn uit een normale kansverdeling, dan zal deze lijn een goede beschrijving geven van het histogram.

*Over het beoordelen van histogrammen*

Bij het beoordelen van histogrammen van steekproeven dienen we ons te realiseren dat hun vorm sterk kan variëren met de steekproefgrootte. Dit verschijnsel is geïllustreerd in onderstaande figuur. Deze toont vier histogrammen van steekproeven van verschillende groottes uit dezelfde normaal verdeelde populatie.

*Histogrammen van vier steekproeven die aselect zijn genomen uit een populatie die een normale kansverdeling heeft, met gemiddelde 50 en standaardafwijking 2. De steekproeven verschillen slechts in het aantal meetwaarden (van linksboven naar rechtsonder respectievelijk 20, 40, 100 en 1000). Bij elk histogram zijn tevens de schattingen van het gemiddelde en de standaardafwijking vermeld. Uit deze figuren blijkt dat het met minder dan 100 meetwaarden lastig kan zijn om met een histogram vast te stellen of deze al dan niet afkomstig zijn uit een normale kansverdeling.*




ad (3) PP-plot

Een scherpere uitspraak over het voldoen aan een normale kansverdeling wordt mogelijk gemaakt door een visuele beoordeling van de PP-plot (percentiel-percentiel-plot). Als de meetwaarden afkomstig zijn uit een normale kansverdeling zullen ze namelijk min of meer op de weergegeven (blauwe) lijn moeten liggen. Let vooral ook op dat dit geldt voor de rechtsboven gelegen meetpunten, daar die de rechterstaart van de kansverdeling vertegenwoordigen. De lozingseis is namelijk een kengetal van die rechterstaart.

*Toelichting op de PP-plot*

De PP-plot is gebaseerd op het principe, dat als meetwaarden afkomstig zijn uit een normale kansverdeling, ze een lineaire relatie zullen vertonen met de waarden van de standaardnormale kansverdeling die overeenstemmende onderschrijdingskansen hebben. De onderschrijdingskans van een meetwaarde – ook wel aangeduid als zijn plot-positie – wordt berekend met de Cunnane-formule [Helsel and Hirsch, 1992]:

$$p_i = \frac{i - 0.4}{n + 0.2}$$

met $p_i$ de plotpositie van de $i$-de van de $n$ oplopend gerangschikte meetwaarden. Vervolgens wordt voor elke meetwaarde díe waarde van de standaardnormale kansverdeling afgeleid waarvan de onderschrijdingskans gelijk is aan de plotpositie van die meetwaarde (dit noemt men de normaalscore). Tenslotte worden de meetwaarden uitgezet tegen hun normaalscores. Daarbij is ook de lijn weergegeven die deze punten min of meer zullen volgen, als de meetwaarden afkomstig zijn uit een normale kansverdeling. De lijn gaat door het punt [0 ; gemiddelde van de meetwaarden] en zijn helling is gelijk aan de standaardafwijking van de meetwaarden. Het gemiddelde en de standaardafwijking zijn hierbij geschat als respectievelijk:

$$\bar{x} = \frac{\sum_{t=1}^{n} x_t}{n} \quad \text{en} \quad s = \sqrt{\frac{\sum_{t=1}^{n} (x_t - \bar{x})^2}{n-1}}$$

met $\bar{x}$ het geschatte gemiddelde, $s$ de geschatte standaardafwijking, $x_t$ de $t$-de van alle chronologisch gerangschikte meetwaarden en $n$ het aantal meetwaarden.

ad (4) Lilliefors-toets op normaliteit
Een aanvullende indicatie over het al of niet voldoen aan een normale kansverdeling wordt verkregen door hierop statistisch te toetsen met de Lilliefors-toets. Het resultaat van deze toets op normaliteit is de p-waarde, zijnde een maat voor de aannemelijkheid dat de meetwaarden afkomstig zijn uit een normale kansverdeling. Als de p-waarde kleiner is dan 1%, is het niet aannemelijk dat de meetwaarden afkomstig zijn uit een normale kansverdeling, mits er geen autocorrelatie optreedt in de meetreeks. Dit toetsresultaat moet echter niet zondermeer als belangrijker worden gezien dan de beoordeling van de drie visuele hulpmiddelen (zie hieronder).

*Het toetsresultaat moet niet per definitie als belangrijkste gelden*

Een waarschuwing is hier op zijn plaats, want de Lilliefors-toets gaat er van uit dat de meetreeks geen autocorrelatie vertoont (oftewel dat de meetwaarden onafhankelijk van elkaar zijn). Dat is echter zelden het geval bij meetreeksen van lozingsparameters, zodat het toetsresultaat niet zondermeer boven dat van de visuele beoordeling van de bovengenoemde grafieken mag worden gesteld.
Behalve de vooronderstelling van onafhankelijke meetwaarden, is een bijkomend bezwaar van het toetsen op normaliteit dat het onderscheidend vermogen gering is als er weinig onafhankelijke meetwaarden zijn, zoals minder dan circa 50. Afwijkingen van de getoetste kansverdeling kunnen dan dus niet snel worden gedetecteerd. Als er daarentegen veel meetwaarden zijn, is het onderscheidend vermogen dermate groot, dat ook al een praktisch gezien verwaarloosbare afwijking van de normale kansverdeling wordt gedetecteerd.

### 3.7.1 Transformatie-mogelijkheden

Als de gebruiker aangeeft dat de meetwaarden niet afkomstig zijn uit een normale kansverdeling, verschijnt er een schermpje met de vraag of de meetwaarden moeten worden getransformeerd. Als de gebruiker aangeeft dat er getransformeerd moet worden, berekent het programma díe transformator die de correlatiecoëfficiënt tussen de meetwaarden en hun normaalscores maximaliseert. Dit is dus de waarde die de punten in de PP-plot zoveel mogelijk een rechte lijn laat volgen. Na deze transformatie verschijnen weer de vier eerder toegelichte hulpmiddelen, die de gebruiker kunnen assisteren bij het formuleren van zijn oordeel over het soort kansverdeling waar de meetwaarden uit afkomstig zijn, namelijk:

(1) een tijdreeksplot van de getransformeerde meetwaarden;
(2) het histogram van de getransformeerde meetwaarden;
(3) de PP-plot van de getransformeerde meetwaarden;

(4) het resultaat van de Lilliefors-toets op normaliteit van de getransformeerde meetwaarden. Tevens verschijnt er een schermpje dat de gebruiker vraagt of de getransformeerde meetwaarden al dan niet afkomstig zijn uit een normale kansverdeling.

Als de gebruiker aangeeft dat ook de getransformeerde meetwaarden niet afkomstig zijn uit een normale kansverdeling, dan wordt de transformatie ongedaan gemaakt en gaat het programma weer verder met de oorspronkelijke meetwaarden. De consequentie dat er niet van kan worden uitgegaan dat de meetwaarden afkomstig zijn uit een normale kansverdeling, is dat er geen gesloten lozingseis (met verwaarloosbare, of minieme overschrijdingskans) kan worden afgeleid, maar slechts een open lozingseis (met niet-verwaarloosbare overschrijdingskans, zie § 3.9).

*Toelichting op het transformeren*

Veel niet-normale kansverdelingen kunnen wiskundig worden getransformeerd tot een normale kansverdeling, door de meetwaarden in een andere schaal uit te drukken, wat neerkomt op het inkrimpen of het uitrekken van de X-as. De daarvoor benodigde transformaties van de meetwaarden zijn doorgaans van de vorm $y=x^{\vartheta}$, met $y$ de getransformeerde meetwaarde, $x$ de oorspronkelijke meetwaarde en $\vartheta$ de transformator. Als $\vartheta$ nul is, wordt de logaritmische transformatie toegepast ($y=\log[x]$). Welke transformatie het meest geschikt is om symmetrie te bewerkstelligen, hangt af van de soort en de mate van scheefheid van de oorspronkelijke kansverdeling.

*Aanbevolen transformaties om symmetrie te bewerkstelligen voor verschillende soorten kansverdelingen (aangepast naar [Helsel and Hirsch, 1992]).*

| | negatief | | | | 0 | positief | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| scheefheid | ← | | | | | → | | | | | | |
| $\vartheta$ | **....** | **4** | **3** | **2** | **1** | **1/2** | **1/3** | **0** | **-1/2** | **-1** | **-2** | **...** |
| transformatie | | $x^4$ | $x^3$ | $x^2$ | $x$ | $\sqrt{x}$ | $x^{1/3}$ | $\log(x)$ | $-1/\sqrt{x}$ | $-1/x$ | $-1/x^2$ | |

<u>Toelichting</u>: de min-tekens bij de inverse transformaties dienen om de oorspronkelijke rangschikking van de meetwaarden te behouden.

*Wees zeer terughoudend met het transformeren en doe een visuele check op de lozingseis!*
Alhoewel het transformeren een elegante oplossing lijkt te bieden voor de gevallen dat de meetwaarden niet afkomstig zijn uit een normale kansverdeling, dienen we toch zeer terughoudend te zijn met het hanteren van deze mogelijkheid. Het transformatieproces introduceert namelijk grote onzekerheid bij het afleiden van de lozingseis. Het programma doet weliswaar enkele rekenkundige checks op een via transformeren afgeleide lozingseis, maar dat geeft nog onvoldoende garantie op een verantwoorde lozingseis. Doe daarom ook beslist zelf een check op een via transformeren afgeleide lozingseis, door visueel te beoordelen of deze wel past bij de karakteristieken van de meetreeks. Als dit tot twijfel leidt over de lozingseis, leid dan een open lozingseis af, door de sessie te herhalen en te kiezen voor een niet-normale kansverdeling, echter ditmaal zónder aansluitend te transformeren.

## 3.8 Autocorrelatie?

Als een meetreeks kleine meetintervallen heeft en/of een grootschalige structuur vertoont, zoals een golfbeweging, of een trend, kunnen de meetwaarden positieve autocorrelatie vertonen. Golfbewegingen of trends kunnen veroorzaakt zijn door veranderingen in de samenstelling van de grondstof, door veranderingen in de temperatuur van ingelaten koelwater of door veranderingen in het productieproces.
Als er positieve autocorrelatie optreedt wijken opeenvolgende meetwaarden doorgaans minder van elkaar af dan meetwaarden met een groter tijdsverschil, wat neerkomt op een soort naijling van informatie. Dit leidt tot een onderschatting van de standaardafwijking van de kansverdeling waar de meetwaarden uit afkomstig zijn, als we die op de gebruikelijke wijze berekenen. De meetwaarden in de meetreeks zijn dan namelijk minder gespreid dan verwacht zou mogen worden van een aselecte steekproef. Het is dus belangrijk om dit verschijnsel te kunnen onderkennen, zodat er vervolgens rekening mee kan worden gehouden bij het berekenen van de lozingseis, anders wordt de lozingseis te laag geschat.

Als de knop 'Autocorrelatie?' wordt aangeklikt en er is sprake van een normale verdeling van de meetwaarden (al of net getransformeerd) dan verschijnen er twee visuele hulpmiddelen, die de gebruiker kunnen assisteren bij het formuleren van zijn oordeel over het al of niet optreden van autocorrelatie in de meetreeks (zie onderstaande figuur), namelijk:

(1) een tijdreeksplot, met daarin ook weergegeven het voortschrijdend gemiddelde;
*(2)* het autocorrelogram van de meetreeks;
*(3)* toets op autocorrelatie.

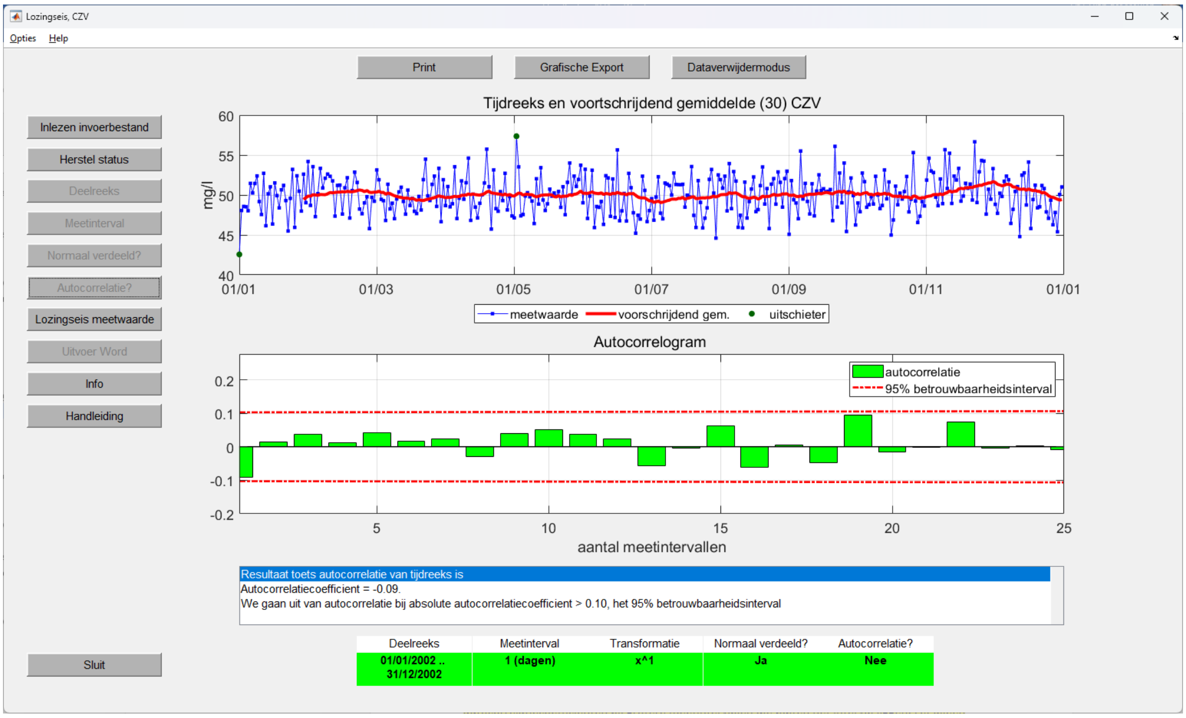
Tevens verschijnt er een schermpje dat de gebruiker vraagt of er al dan niet sprake is van autocorrelatie.

ad (1) Tijdreeksplot met voortschrijdend gemiddelde
Een eerste visuele beoordeling van eventuele autocorrelatie is mogelijk met een tijdreeksplot van de (al dan niet getransformeerde) meetreeks. Om een golfbeweging of een trend te kunnen onderkennen, wordt in deze figuur tevens het voortschrijdend gemiddelde over 30 (al dan niet getransformeerde) meetwaarden aangegeven, met een dikke rode lijn. Als dat voortschrijdend gemiddelde een golfbeweging of trend vertoont, is er duidelijk sprake van autocorrelatie. Het gehanteerde venster waarover het voortschrijdend gemiddelde wordt berekend staat standaard op 30, maar dit kan desgewenst handmatig worden aangepast via het menu 'Opties' (linksboven).

ad (2) Autocorrelogram van de meetreeks
Als de gebruiker een meetinterval groter dan 0 heeft opgegeven, kan tevens worden vastgesteld of er sprake is van autocorrelatie met het autocorrelogram. Dit toont voor elk mogelijk aantal meetintervallen (*l*) de geschatte autocorrelatiecoëfficiënt ( $\hat{\rho}_l$ ). Deze laatste is een maat voor de samenhang tussen meetwaarden die dat aantal meetintervallen uit elkaar liggen. Daarbij is ook het 95%-betrouwbaarheidsinterval van weergegeven van de geschatte autocorrelatiecoëfficiënten voor het geval dat de (eventueel getransformeerde) meetwaarden afkomstig zijn uit een normale verdeling. Dit interval wordt begrensd door de rode stippellijnen aan weerskanten van de nullijn. Deze lijnen lopen doorgaans niet evenwijdig aan de nullijn, daar het interval afhangt van het aantal meetwaarden waarmee een autocorrelatiecoëfficiënt is geschat en dat aantal neemt immers af als *l* toeneemt.

ad (3) Toets op autocorrelatie
Een aanvullende indicatie of er al of niet sprake is van autocorrelatie. Is er geen sprake van een normale verdeling voor de meetwaarden dan worden de resultaten van de runstoets (een verdelingsvrije toets) getoond, anders de toets op basis van de eerste autocorrelatiecoëfficiënt.

*Visuele hulpmiddelen om een oordeel te kunnen formuleren over het optreden van autocorrelatie in de meetreeks. Op basis van de onderstaande grafieken heeft de gebruiker (terecht) geconcludeerd dat in de betreffende meetreeks geen autocorrelatie optreedt. Het statusveld bevat daardoor nu de tekst 'Nee' en is door het invullen ook groen geworden.*




Als er geen autocorrelatie is en de (eventueel getransformeerde) meetwaarden afkomstig zijn uit een normale kansverdeling, dan zullen de autocorrelatiecoëfficiënten zich gemiddeld in 95 van de 100 gevallen binnen het 95%-betrouwbaarheidsinterval van de geschatte autocorrelatiecoëfficiënten bevinden. Als de eerste 3 à 5 autocorrelatiecoëfficiënten – deze bevinden zich aan de linkerkant van het autocorrelogram - zich binnen dat interval bevinden en niet alle positief zijn, dan kan er doorgaans al van worden uitgegaan dat er geen autocorrelatie is, ook al zijn er wellicht enkele autocorrelatiecoëfficiënten bij grotere meetintervallen die buiten het interval liggen. Let alleen wel op dat als de eerste autocorrelatiecoëfficiënten alle positief zijn, er meestal toch sprake is van autocorrelatie, ongeacht of ze al of niet binnen het 95%-betrouwbaarheidsinterval liggen.

Als de gebruiker na beoordeling van deze visuele hulpmiddelen aangeeft dat er sprake is van autocorrelatie, vraagt het programma tot en met welk meetinterval daarvan sprake is. Het betreffende schermpje bevat ook al een voorstel voor dat aantal, maar dit kan de gebruiker desgewenst aanpassen.

De breedte van de X-as van het autocorrelogram wordt door het programma zodanig ingesteld, dat zichtbaar is waar de eerste autocorrelatiecoëfficiënt negatief is. Minimaal worden echter altijd 25 autocorrelatiecoëfficiënten getoond.

*Toelichting op de autocorrelatiecoëfficiënt*

De autocorrelatiecoëfficiënt is een maat voor de samenhang tussen (al dan niet getransformeerde) meetwaarden die zijn verkregen met een tijdsinterval *l* en wordt geschat als [Box and Jenkins, 1976]:

$$\hat{\rho}_l = \frac{\sum_{t=1}^{n-l}(x_t - \overline{x})(x_{t+l} - \overline{x})}{\sum_{t=1}^{n}(x_t - \overline{x})^2}$$

met $\hat{\rho}_l$ de geschatte autocorrelatiecoëfficiënt voor tijdsinterval *l*, $\overline{x}$ de schatting van het gemiddelde van de kansverdeling waar de (al dan niet getransformeerde) meetwaarden uit afkomstig zijn, $x_t$ de *t*-de van de chronologisch gerangschikte (al dan niet getransformeerde) meetwaarden en *n* het aantal meetwaarden.

## 3.9 Lozingseis

Als er voldoende informatie beschikbaar is over de statistische eigenschappen van de deelreeks – de statusbalk is dan helemaal groen – kan door het aanklikken van 'lozingseis' de lozingseis worden afgeleid.

De lozingseis wordt zowel berekend voor individuele meetwaarden, als voor gemiddelden van 10 opeenvolgende meetwaarden[2]. De resultaten worden direct getoond in twee tijdreeksplots, de bovenste van de meetwaarden, de onderste van de gemiddelden van 10 opeenvolgende meetwaarden. De berekende lozingseis is daarbij weergegeven als een dikke rode lijn (zie onderstaande figuur). De in de tweede tijdreeksplot weergegeven gemiddelden zijn berekend over 10 opeenvolgende meetwaarden bij het opgegeven meetinterval. Als er meer dan 2 van de 10 meetwaarden ontbreken wordt het betreffende gemiddelde over 10 opeenvolgende meetwaarden echter niet getoond. Als er slechts 1 (of geen) gemiddelde kan worden getoond, wordt er geen lozingseis voor het gemiddelde berekend.

De berekende lozingseis voor meetwaarden kan zijn:

(1) een gesloten lozingseis, in de figuur aangegeven met een doorgetrokken dikke rode lijn. Deze heeft standaard een verwaarloosbare overschrijdingskans van 1/1.000 (oftewel 0,1%), tenzij de gebruiker via het menu 'Opties' (linksboven) heeft aangegeven dat deze overschrijdingskans 1/100 moet zijn (oftewel 1%). De gesloten lozingseis kan bij de handhaving als niet te overschrijden maximum worden gehanteerd;

(2) een open lozingseis, in de figuur aangegeven met een gestippelde dikke rode lijn – deze heeft een niet-verwaarloosbare, maar bekende overschrijdingkans, uitgaande van de procesvoering zoals die was gedurende de periode die de geanalyseerde meetreeks beslaat. Bij de handhaving dient dus rekening te worden gehouden met die overschrijdingskans.

De berekende lozingseis voor gemiddelden van 10 opeenvolgende meetwaarden is altijd een gesloten lozingseis. Maar als de gebruiker een meetinterval van '0' heeft opgegeven, wordt er geen lozingseis berekend voor gemiddelden van 10 opeenvolgende meetwaarden.

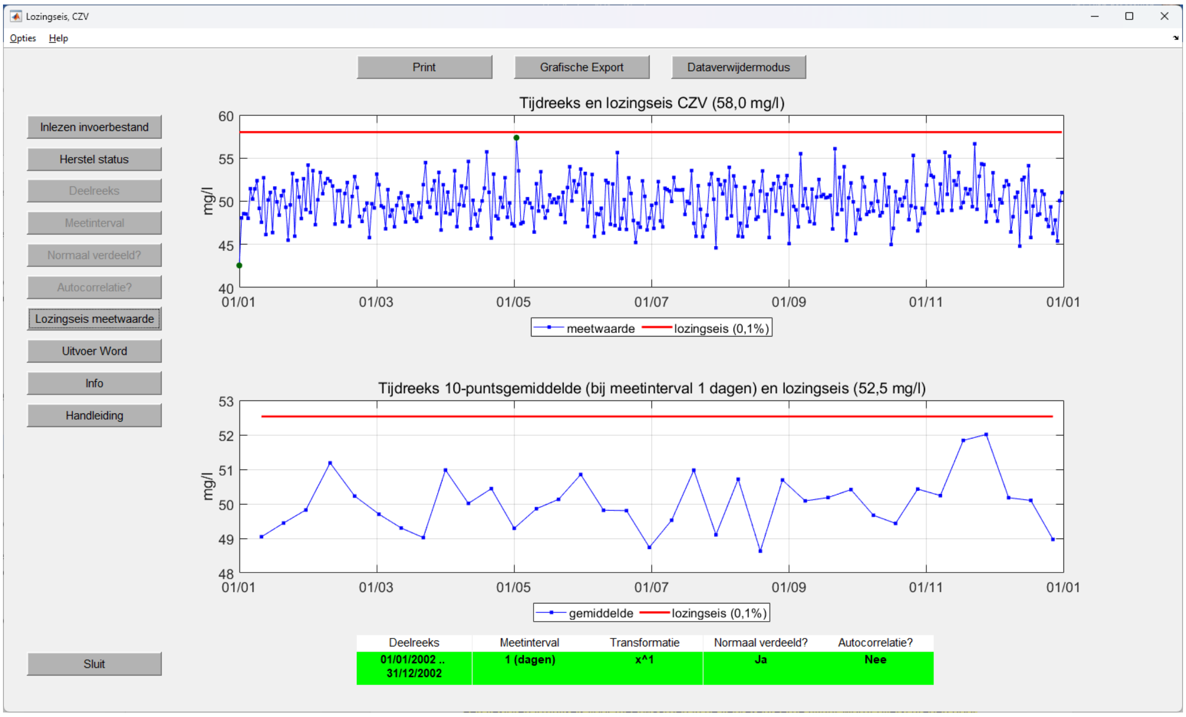
### *Lozingseis geldt alleen voor het monstertype van de ingevoerde meetwaarden*

Een afgeleide lozingseis geldt alleen voor het monstertype van de ingevoerde meetwaarden. Dus als een lozingseis is afgeleid met meetwaarden van dagverzamelmonsters, kan die niet worden gehanteerd voor steekmonsters (of andersom). Om een lozingseis voor meetwaarden van het ene monstertype te kunnen omzetten naar die van een ander type, zal er dus een vertaalslag moeten

[2] Het aantal van 10 is gekozen, omdat de kansverdeling van een gemiddelde van 10 meetwaarden beter zal naderen tot een normale kansverdeling. Dit aantal kan niet aangepast in het programma.

plaatsvinden. Zie hiervoor § 3.4 van het rapport ‘Statistische aspecten van lozingseisen’ [Baggelaar, 2003]. Dit rapport is in te zien via de ‘Help’-knop van Lozingseis-assistent.

*Voorbeeld van de grafische presentatie van de berekende lozingseisen. Boven: de tijdreeksplot van de meetwaarden en de daarvoor berekende lozingseis. Onder: de tijdreeksplot van de gemiddelden van 10 opeenvolgende meetwaarden en de daarvoor berekende lozingseis. Het betreft hier in beide gevallen gesloten lozingseisen. Deze zijn bij de handhaving te hanteren als waarden die niet over-schreden mogen worden.*




*Check de lozingseisen met deze figuur !*

Het is zeer sterk aan te raden de berekende lozingseisen visueel te checken met deze figuur. Als een gesloten lozingseis voor meetwaarden in de gepresenteerde figuur wordt overschreden door een percentage meetwaarden dat hoger is dan de gehanteerde overschrijdingskans (standaard 0,1%, of anders 1%), dan kan dat een aanwijzing zijn dat de gebruiker onterecht heeft aangegeven dat er sprake is van een normale kansverdeling. En als een gesloten lozingseis voor gemiddelden van 10 opeenvolgende meetwaarden in de gepresenteerde figuur wordt overschreden door één of meer van dergelijke gemiddelden, kan dat een aanwijzing zijn dat de gebruiker onterecht heeft aangegeven dat er geen sprake is van autocorrelatie.

*En wees vooral zeer kritisch als er is getransformeerd !*

En, zoals reeds is opgemerkt in § 3.7.1, wees vooral zeer kritisch als er via transformeren een lo-zingseis voor meetwaarden is afgeleid. Beoordeel daartoe visueel of deze wel past bij de karakte-ristieken van de meetreeks (is de lozingseis bijvoorbeeld niet veel te hoog?). Als dit tot twijfel leidt over de lozingseis, leid dan een open lozingseis af, door de sessie te herhalen en te kiezen voor een niet-normale kansverdeling, echter ditmaal zónder aansluitend te transformeren.

*Achtergronden lozingseis*
De lozingseis wordt door Lozingseis-assistent berekend aan de hand van de beschikbaar gestelde historische meetwaarden, waaruit wordt afgeleid binnen welke grens nieuwe waarden - dit zijn meetwaarden respectievelijk gemiddelden van 10 opeenvolgende meetwaarden - verwacht mogen worden, gegeven de gebruikelijke, beheerste procesvoering. Aan de hand van de ligging van nieuwe waarden ten opzichte van deze grens kan vervolgens worden vastgesteld of het lozingsproces al dan niet is gewijzigd.

De wijze waarop de lozingseis wordt berekend hangt af van: (a) de kansverdeling waar de waarden uit afkomstig zijn en (b) van het al of niet optreden van autocorrelatie in de meetreeks.

ad (a) Als kan worden uitgegaan van een normale kansverdeling, kan een scherpere lozingseis worden gesteld dan wanneer dat níet het geval is. Dit komt door de vele theoretische kennis die dan kan worden gehanteerd over de eigenschappen van die kansverdeling, als aanvulling op de informatie die reeds beschikbaar is in de meetreeks. Afhankelijk van de instelling berekent het programma in dit geval een lozingseis met een verwaarloosbare overschrijdingskans van 1/1.000, of een zeer geringe overschrijdingskans van 1/100. Deze kan worden gehanteerd als een lozingseis die niet overschreden mag worden door volgende waarden. Dit noemen we een *gesloten* lozingseis.

Als daarentegen níet kan worden uitgegaan van een normale kansverdeling, ook niet na transformatie, staat alleen de informatie uit de meetreeks ter beschikking. Om dan bijvoorbeeld een lozingseis met een overschrijdingskans van 1/1.000 te kunnen berekenen, dient de meetreeks minstens 4.500 onafhankelijke waarden te bevatten. Dat is uiteraard teveel gevraagd, zodat dan slechts een lozingseis kan worden afgeleid met een bekende en niet-verwaarloosbare overschrijdingskans. Dit noemen we een *open* lozingseis. Die overschrijdingskans is echter wel bekend, zodat daar bij de handhaving rekening mee kan worden gehouden. Een open lozingseis wordt aangegeven met een rode stippellijn. Zijn overschrijdingskans is vermeld in de uitvoer. Indien gewenst kan als aanvulling op deze open lozingseis ook een maximale eis worden opgenomen. Deze kan evenwel niet statistisch worden afgeleid, maar moet op andere overwegingen gebaseerd zijn, bijvoorbeeld een waarde die samenhangt met de acute toxiciteit.

ad (b) Het al of niet optreden van autocorrelatie in de meetreeks heeft ook grote invloed op de mogelijkheden van de lozingseisen. Als er autocorrelatie optreedt zal de beschikbare meetreeks namelijk de standaardafwijking van de achterliggende kansverdeling onderschatten, zodat daarvoor moet worden gecorrigeerd bij het opstellen van de lozingseis. Verder zal er bij de handhaving rekening mee moeten worden gehouden dat opeenvolgende meetwaarden doorgaans meer op elkaar zullen lijken dan op veel eerder of later verkregen meetwaarden. Waar het gaat om een open lozingseis (een lozingseis met een bekende en niet-verwaarloosbare overschrijdingskans), is het dan niet meer mogelijk om over een korte termijn te kunnen beoordelen of het aantal overschrijdingen van de lozingseis nog past bij wat verwacht zou mogen worden bij onveranderde procesvoering.

*Gebruikersoptie: overschrijdingskans van 1/1.000 of 1/100*
Bij het berekenen van een gesloten lozingseis zal het programma standaard een overschrijdingskans hanteren van 1/1.000 (0,1%). We mogen de lozingseis dan met 95% betrouwbaarheid opvatten als de waarde die met een kans van slechts 1/1.000 zal worden overschreden bij de gebruikelijke, beheerste procesvoering. Dit betekent dat bij dagelijks meten gemiddeld slechts eens in de drie jaar een overschrijding van de lozingseis zal plaatsvinden bij de gebruikelijke, beheerste procesvoering. Deze kans is dermate verwaarloosbaar, dat overschrijdingen van de lozingseis met grote zekerheid gezien kunnen worden als indicaties van veranderde lozingskenmerken.

Via het menu 'Opties' (aan te klikken op de menubalk) kan de gebruiker de overschrijdingskans voor de betreffende sessie desgewenst echter ook terugbrengen tot 1/100 (1%). Een berekende gesloten lozingseis is dan met 95% betrouwbaarheid op te vatten als de waarde die met een kans van 1/100 zal worden overschreden bij de gebruikelijke, beheerste procesvoering. Deze kans is weliswaar nog steeds gering, maar niet meer verwaarloosbaar. Het gebruik van deze instelling is dan ook doorgaans af te raden. Een mogelijke uitzondering kan worden gemaakt voor een parameter die met zeer lage frequentie wordt gemeten. Want als een parameter bijvoorbeeld 10 maal per jaar wordt gemeten, betekent een overschrijdingskans van 1/100 dat de lozingseis gemiddeld slechts eens in de 10 jaar zal worden overschreden bij de gebruikelijke, beheerste procesvoering.

Aangezien het gebruik van een overschrijdingskans van 1/100 maar in weinig gevallen opportuun zal zijn, zal het programma bij het opstarten altijd de overschrijdingskans van 1/1.000 hebben ingesteld, ook als de gebruiker die instelling bij een vorige sessie heeft veranderd. Zoals gezegd is dit echter per sessie gemakkelijk via 'Opties' te veranderen.

*De berekende lozingseis is een tolerantielimiet*

Lozingseis-assistent berekent de lozingseis als de bovengrens van het betrouwbaarheidsinterval van een geschat percentiel van de kansverdeling waar de waarden (meetwaarden danwel gemiddelden van 10 opeenvolgende meetwaarden) uit afkomstig zijn. De statistische benaming van die bovengrens is de tolerantielimiet$_{(\gamma,1-\alpha)}$, die wordt gekenmerkt door de volgende twee percentages:

(1) het percentiel, oftewel de dekkingsgraad, zijnde het percentage van alle nieuwe waarden dat de limiet *minstens* beoogt te begrenzen. Dit geven we aan als 100%·$\gamma$, met $0 < \gamma < 1$. De limiet zal dan dus door *hoogstens* 100%·(1-$\gamma$) van alle nieuwe waarden worden overschreden;
(2) de betrouwbaarheid, oftewel de mate van aannemelijkheid dat de limiet ook daadwerkelijk minstens de beoogde fractie van alle nieuwe waarden zal begrenzen. Dit geven we aan als 100%·(1-$\alpha$), met $0 < \alpha < 1$. Hierbij is $\alpha$ het (door ons geaccepteerde) risico dat de tolerantielimiet door meer dan 100%·(1-$\gamma$) van alle nieuwe waarden wordt overschreden, zónder dat er sprake is van een verandering van de lozingskenmerken. De onzekerheden ten gevolge van bemonsterings- en analysefouten en de steekproeffout zijn in dit betrouwbaarheidsinterval verdisconteerd.

De tolerantielimiet$_{(\gamma,1-\alpha)}$ kunnen we dus zien als de bovengrens van het 100%·(1-$\alpha$)-betrouwbaarheidsinterval van het geschatte 100·$\gamma$-percentiel van een kansverdeling.

**Voorbeeld**

We mogen bijvoorbeeld van het 99,9-percentiel van een kansverdeling verwachten dat 99,9% van de waarden daaronder zal vallen en 0,1% daarboven. Maar doordat we dat percentiel slechts kunnen schatten, moeten we daarbij ook zijn onzekerheid in acht nemen, uitgedrukt in het betrouwbaarheidsinterval. Zo zal het 95%-betrouwbaarheidsinterval van een geschat 99,9-percentiel in 95 van de 100 gevallen ook daadwerkelijk het 99,9-percentiel van de kansverdeling bevatten. De bovengrens van dat interval, aangeduid als de **tolerantielimiet$_{(99,9\%,\ 95\%)}$**, zal daarmee met 95% betrouwbaarheid door *minstens* 99,9% van de nieuwe waarden onderschreden worden en door *hoogstens* 0,1% van de nieuwe waarden overschreden worden. Als de overschrijding meer bedraagt, kan dat met 95% betrouwbaarheid worden opgevat als een signaal dat er geen sprake meer is van de gebruikelijke, beheerste procesvoering.

Door bij de handhaving na te gaan of nieuwe waarden onder of boven de tolerantielimiet liggen, kan dus worden nagegaan of er een verandering in het lozingsproces is opgetreden. De tolerantielimiet is daarmee een geschikt kengetal om als een *naleefbare* lozingseis te dienen.

*Toelichting op de berekende lozingseis voor meetwaarden*

Om een lozingseis voor meetwaarden te berekenen, onderscheidt Lozingseis-assistent de volgende gevallen:

(1) normale kansverdeling, zonder autocorrelatie;
(2) normale kansverdeling, met autocorrelatie;
(3) geen normale kansverdeling, zonder autocorrelatie;
(4) geen normale kansverdeling, met autocorrelatie.

(1) Normale kansverdeling, zonder autocorrelatie
Als we er van uit mogen gaan dat de (al dan niet getransformeerde) meetwaarden afkomstig zijn uit een normale kansverdeling en we geen rekening hoeven te houden met autocorrelatie, dan berekent het programma de tolerantielimiet$_{(100\%\cdot\gamma,\ 95\%)}$ van toekomstige (eventueel getransformeerde) meetwaarden als [Natrella, 1963]:

$$TL_{(100\%\cdot\gamma,95\%)} = \bar{x} + \frac{z_{(\gamma)} + \sqrt{\left(z_{(\gamma)}^2 - (1 - \frac{z_{(0,95)}^2}{2\cdot(n-1)})\cdot(z_{(\gamma)}^2 - \frac{z_{(0,95)}^2}{n})\right)}}{(1 - \frac{z_{(0,95)}^2}{2\cdot(n-1)})}\cdot s$$

met $\bar{x}$ en *s* de schattingen van het gemiddelde en de standaardafwijking van de kansverdeling waar de meetwaarden uit afkomstig zijn, *n* het aantal meetwaarden waarop die schattingen zijn gebaseerd, $z_{(\gamma)}$ het $100\cdot\gamma$-percentiel van de standaardnormale verdeling (standaard is dit het 99,9-percentiel en anders het 99-percentiel) en $z_{(0,95)}$ het 95-percentiel van de standaardnormale verdeling. De op deze wijze berekende tolerantielimiet is een gesloten lozingseis, dus met een verwaarloosbare (of minieme) overschrijdingskans.

Als we over oneindig veel meetwaarden beschikken ($n \rightarrow \infty$) en $\gamma = 0{,}999$ (de standaardinstelling), dan gaat bovenstaande formule over in:

$$TL_{(99,9\%,95\%)} = \bar{x} + z_{(0,999)}\cdot s = \bar{x} + 3{,}090\cdot s$$

Als de gebruiker daarentegen een overschrijdingskans van 1% heeft ingesteld ($\gamma = 0{,}990$), dan gaat bovenstaande formule bij oneindig veel meetwaarden over in:

$$TL_{(99\%,95\%)} = \bar{x} + z_{(0,990)}\cdot s = \bar{x} + 2{,}326\cdot s$$

Enige voorzichtigheid is wel op zijn plaats, daar we een extreme tolerantielimiet berekenen, namelijk diegene die slechts door 1 meetwaarde op 1.000 (of 100) onafhankelijke meetwaarden wordt overschreden. De geldigheid daarvan is namelijk gevoelig voor afwijkingen van de normale kansverdeling. Bij gerede twijfel aan het soort kansverdeling kan daarom beter voor een verdelingsvrije tolerantielimiet worden gekozen (zie verder), zij het dat die slechts dienst kan doen als open lozingseis.

(2) Normale kansverdeling, met autocorrelatie
Als we er van uit mogen gaan dat de (al dan niet getransformeerde) meetwaarden afkomstig zijn uit een normale kansverdeling en er sprake is van autocorrelatie, dan berekent het programma de tolerantielimiet$_{(100\%\cdot\gamma,\ 95\%)}$ van toekomstige (eventueel getransformeerde) meetwaarden als boven aangegeven, maar nu met *s* vervangen door *s**, de voor de autocorrelatie gecorrigeerde schatting van de standaardafwijking van de kansverdeling, volgens:

$$s^* = \sqrt{\frac{s^2}{1 - \frac{2}{n\cdot n-1}\cdot\sum_{l=1}^{n-1}((n-l)\cdot\hat{\rho}_l)}}$$

met $\hat{\rho}_l$ de geschatte autocorrelatiecoëfficiënt voor tijdsinterval *l*. Als alle autocorrelatiecoëfficiënten nul zijn, reduceert *s** weer tot *s*. Als de autocorrelatiecoëfficiënten positief zijn is *s** groter dan *s*. Een positieve autocorrelatie leidt dus tot een hogere tolerantielimiet.
De op deze wijze berekende tolerantielimiet is een gesloten lozingseis, dus met een verwaarloosbare (of minieme) overschrijdingskans.

Als de autocorrelatie relevant blijkt tot voorbij een tijdsinterval van *n*/4 (*n* is het aantal meetwaarden), door een slechts langzaam uitdempende autocorrelatie, dan zal het programma aangeven dat de lozingseis niet berekend kan worden. Er kunnen dan namelijk grote fouten ontstaan bij het schatten van de autocorrelatiecoëfficiënten. In feite is de meetreeks dan té kort ten opzichte van de daarin voorkomende grootschalige structuur.

(3) Niet-normale kansverdeling, zonder autocorrelatie

Als we níet uit kunnen gaan van een normale kansverdeling - ook niet na transformatie van de meetwaarden -, we geen rekening hoeven te houden met autocorrelatie en er voldoende meetwaarden beschikbaar zijn (zie onder), dan berekent het programma de tolerantielimiet$_{(100\%\cdot\gamma,\,95\%)}$ van toekomstige meetwaarden als [Gilbert, 1987]:

$$TL_{(100\%\cdot\gamma,\,95\%)} = x_{[u]} \quad \text{en} \quad u = \gamma\cdot(n+1) + z_{(0{,}95)}\cdot\sqrt{n\cdot\gamma\cdot(1-\gamma)}$$

met $x_{[u]}$ de meetwaarde die zich na rangschikking van klein naar groot op de $u^{e}$-positie bevindt (als $u$ geen geheel getal is wordt lineair geïnterpoleerd tussen de naastliggende meetwaarden), $n$ het aantal meetwaarden en $z_{(0,95)}$ het 95-percentiel van de standaardnormale verdeling. De op deze wijze berekende tolerantielimiet is een gesloten lozingseis, dus met een verwaarloosbare (of minieme) overschrijdingskans. Bij de standaardoverschrijdingskans van 0,1% ($\gamma$ = 0,999) is er sprake van voldoende meetwaarden bij $n$ > 4.500 en als de overschrijdingskans door de gebruiker is ingesteld op 1% ($\gamma$ = 0,990), dan is er sprake van voldoende meetwaarden bij $n$ > 450.

In de meeste gevallen zullen er echter onvoldoende meetwaarden beschikbaar zijn, zodat het dan niet mogelijk is de tolerantielimiet$_{(100\%\cdot\gamma,\,95\%)}$ te schatten. De reden is dat de berekende $u$ dan groter is dan $n$. In een dergelijk geval wordt de maximale meetwaarde ($x_{[n]}$) gehanteerd als tolerantielimiet, volgens:

$$TL_{(100\%\cdot\gamma,\,95\%)} = x_{[n]}$$

Dit betreft een open lozingseis, met een overschrijdingskans van 100%·(1-γ). De bij een betrouwbaarheid van 95% horende dekkingsgraad ($\gamma$) wordt hierbij vastgesteld aan de hand van de volgende relatie [Montgomery, 1991]:

$$\ln(\gamma) = \frac{\ln(0{,}05)}{n}$$

**Voorbeeld**

Stel dat we beschikken over 20 onafhankelijke meetwaarden van koper, verkregen bij lozing onder de gebruikelijke, beheerste procesvoering, over een periode van twee jaar. De meetwaarden blijken niet afkomstig uit een normale kansverdeling, zelfs ook niet na een transformatie. Als we het maximum van deze 20 meetwaarden ($x_{[20]}$) hanteren als tolerantielimiet en een betrouwbaarheid van 95% wensen ($\alpha$=5%), dan is de dekkingsgraad ($\gamma$) 86%. We mogen dus met 95% betrouwbaarheid verwachten dat deze tolerantielimiet door hoogstens 14% van de nieuwe meetwaarden zal worden overschreden, mits het lozingsproces niet wijzigt.

(4) Niet-normale kansverdeling, met autocorrelatie

Als we níet uit kunnen gaan van een normale kansverdeling - ook niet na transformatie van de meetwaarden - en er sprake is van autocorrelatie, dan berekent het programma een tolerantielimiet van toekomstige meetwaarden nog steeds zoals beschreven onder (3), zij het dat ditmaal niet wordt uitgegaan van het aantal beschikbare meetwaarden ($n$), maar van het aantal *onafhankelijke* meetwaarden. Het aantal onafhankelijke meetwaarden ($n$*) is een functie van het aantal beschikbare meetwaarden ($n$) en van het autocorrelogram van de meetreeks, volgens [Bayley and Hammersley, 1946]:

$$n^{*} = \left(\frac{1}{n} + \frac{2}{n^{2}}\cdot\sum_{l=1}^{n-1}\left((n-l)\cdot\hat{\rho}_{l}\right)\right)^{-1}$$

met $\hat{\rho}_{l}$ de geschatte autocorrelatiecoëfficiënt voor tijdsinterval $l$. Als alle autocorrelatiecoëfficiënten nul zijn, is $n$* gelijk aan $n$. Maar als de autocorrelatiecoëfficiënten positief zijn is $n$* kleiner dan $n$. Het betreft een open lozingseis, met een overschrijdingskans van 100%·(1-γ).

*Toelichting op de berekende lozingseis voor gemiddelden*
Behalve voor meetwaarden, stelt het programma ook een lozingseis op voor gemiddelden van 10 opeenvolgende meetwaarden.

Volgens de statistische theorie (de Centrale Limietstelling) zal een kansverdeling van gemiddelden meer naderen tot de normale kansverdeling, naarmate er meer meetwaarden worden gemiddeld, ongeacht de kansverdeling van de meetwaarden. Als de meetwaarden afkomstig zijn uit een zeer scheve kansverdeling moet er nog wel worden gemiddeld over veel meetwaarden (50 à 75) om dit te bewerkstelligen, maar naarmate de meetwaarden afkomstig zijn uit een symmetrischer kansverdeling, reduceert het benodigde aantal sterk. Doordat het aantal gemiddelden van 10 opeenvolgende meetwaarden hoogstens 10% kan bedragen van het aantal meetwaarden in de meetreeks, zijn er doorgaans te weinig van deze gemiddelden beschikbaar om vast te stellen uit welk soort kansverdeling die afkomstig zijn. Het programma gaat daarom uit van wat de gebruiker heeft ingevuld over de normaliteit van de *meetwaarden*.

Voor het formuleren van een lozingseis aan gemiddelden van 10 opeenvolgende meetwaarden onderscheidt dit programma alleen de volgende twee mogelijkheden:
(1) normale kansverdeling, zonder autocorrelatie;
(2) normale kansverdeling, met autocorrelatie.
Als de gebruiker heeft aangegeven dat de meetwaarden niet afkomstig zijn uit een normale kansverdeling, dan wordt er geen lozingseis afgeleid voor gemiddelden van 10 opeenvolgende meetwaarden. Een dergelijke lozingseis kan evenmin worden afgeleid als de gebruiker als meetinterval ‘0’ heeft opgegeven, daarmee aangevend dat álle meetwaarden moeten worden gehanteerd, ongeacht het meetinterval. Door het ontbreken van een vast meetinterval kan er dan namelijk niet meer worden vastgesteld óf er sprake is van autocorrelatie en zo ja, hoe groot die is, zodat er onvoldoende informatie is om de lozingseis voor gemiddelden van 10 opeenvolgende meetwaarden af te leiden. Bijlage 2 geeft echter oplossingen voor deze gevallen.

(1) Normale kansverdeling, zonder autocorrelatie
Als de meetwaarden bij het opgegeven meetinterval geen autocorrelatie vertonen, dan berekent het programma de tolerantielimiet$_{(100\%\cdot\gamma,\,95\%)}$ van toekomstige gemiddelden van 10 opeenvolgende (al dan niet getransformeerde) meetwaarden als:

$$TL_{(100\%\cdot\gamma,\,95\%)} = \bar{x} + \frac{z_{(\gamma)} + \sqrt{\left(z_{(\gamma)}^2 - (1-\frac{z_{(0,95)}^2}{2\cdot(n-1)})\cdot(z_{(\gamma)}^2 - \frac{z_{(0,95)}^2}{n})\right)}}{(1-\frac{z_{(0,95)}^2}{2\cdot(n-1)})}\cdot\frac{s}{\sqrt{10}}$$

met $\bar{x}$ en $s$ de schattingen van het gemiddelde en de standaardafwijking van de kansverdeling waar de (al dan niet getransformeerde) meetwaarden uit afkomstig zijn, $n$ het aantal meetwaarden waarop die schattingen zijn gebaseerd, $z_{(\gamma)}$ het 100·$\gamma$-percentiel van de standaardnormale verdeling (standaard is dit het 99,9-percentiel en anders het 99-percentiel) en $z_{(0,95)}$ het 95-percentiel van de standaardnormale verdeling. De op deze wijze berekende tolerantielimiet is een gesloten lozingseis, dus met een verwaarloosbare (of minieme) overschrijdingskans. Let op dat deze eis alleen geldt voor gemiddelden van 10 opeenvolgende (al dan niet getransformeerde) meetwaarden bij het opgegeven, of een groter meetinterval.

(2) Normale kansverdeling, met autocorrelatie
Als de meetwaarden bij het opgegeven meetinterval wél autocorrelatie vertonen, dan berekent het programma de tolerantielimiet$_{(100\%\cdot\gamma,\, 95\%)}$ van toekomstige gemiddelden van 10 opeenvolgende (al dan niet getransformeerde) meetwaarden als[3]:

$$TL_{(100\%\cdot\gamma,\,95\%)} = \bar{x} + \frac{z_{(\gamma)} + \sqrt{\left(z^2_{(\gamma)} - \left(1 - \frac{z^2_{(0,95)}}{2\cdot(n-1)}\right)\cdot\left(z^2_{(\gamma)} - \frac{z^2_{(0,95)}}{n}\right)\right)}}{\left(1 - \frac{z^2_{(0,95)}}{2\cdot(n-1)}\right)} \cdot \sqrt{\frac{s^2}{10\cdot\left(1 - \frac{2}{n\cdot n-1}\cdot\sum_{l=1}^{n-1}((n-l)\cdot\hat{\rho}_l)\right)} \cdot \left(1 + \frac{2}{10}\cdot\sum_{l=1}^{9}((10-l)\cdot\hat{\rho}_l)\right)}$$

met $\bar{x}$ en $s$ de schattingen van het gemiddelde en de standaardafwijking van de kansverdeling waar de (al dan niet getransformeerde) meetwaarden uit afkomstig zijn, $n$ het aantal meetwaarden waarop die schattingen zijn gebaseerd en $\hat{\rho}_l$ de geschatte autocorrelatiecoëfficiënt voor tijdsinterval *l*. Als alle autocorrelatiecoëfficiënten nul zijn, reduceert deze formule tot die vermeld onder (1). De op deze wijze berekende tolerantielimiet is een gesloten lozingseis, dus met een verwaarloosbare (of minieme) overschrijdingskans. Let op dat deze eis alleen geldt voor gemiddelden van 10 opeenvolgende (al dan niet getransformeerde) meetwaarden bij het opgegeven meetinterval.

Als de meetwaarden gedurende de sessie zijn getransformeerd en er vervolgens door de gebruiker is aangegeven dat de getransformeerde meetwaarden afkomstig zijn uit een normale kansverdeling, dan wordt deze tolerantielimiet berekend met de getransformeerde meetwaarden. De bij de uitvoer gepresenteerde lozingseis is echter weer teruggetransformeerd naar de meetschaal. In een dergelijk geval dient bij de handhaving een gemiddelde dus te worden berekend uit de getransformeerde meetwaarden en het dient vervolgens te worden teruggetransformeerd naar de meetschaal, alvorens het af te zetten tegen de lozingseis.

Als de autocorrelatie relevant blijkt tot voorbij een tijdsinterval van $n/4$ ($n$ is het aantal meetwaarden), door een slechts langzaam uitdempende autocorrelatie, dan zal het programma aangeven dat de lozingseis niet berekend kan worden. Er kunnen dan namelijk grote fouten ontstaan bij het schatten van de autocorrelatiecoëfficiënten. In feite is de meetreeks dan té kort ten opzichte van de daarin voorkomende grootschalige structuur.

## 3.10 Uitvoer naar Word

Als het programma de lozingseis heeft berekend, kan die met de knop ‘Info’ worden bekeken. Met de knop ‘Uitvoer’ worden de resultaten weggeschreven naar een Word-bestand. De resultaten omvatten niet alleen de berekende lozingseis, maar ook een onderbouwing van de wijze waarop deze is berekend. Dit uitvoerbestand is daardoor bijvoorbeeld geschikt om een berekende lozingseis te verantwoorden naar de aanvrager.

*Lozingseisen zijn niet afgerond*

Let op dat de afgeleide lozingseisen niet zijn afgerond en daardoor nog veel decimalen bevatten. De gebruiker dient deze dus zelf op redelijke wijze af te ronden.

Door met de knop ‘Handleiding’ de handleiding in Word te openen kan de gebruiker onderdelen kopiëren van de handleiding naar het uitvoerbestand om daarmee de rapportage van de resultaten verder te onderbouwen.

[3] Deze formule is uitgebreid ten opzichte van die in het rapport ‘Statistische aspecten van lozingseisen’ (mei 2003, Paul K. Baggelaar, Icastat Statistisch Adviesbureau).

### 3.10.1 Onderdelen uitvoerbestand

Het uitvoerbestand bevat de volgende onderdelen:

- ✓ datum
- ✓ bedrijfsnaam
- ✓ parameternaam [meeteenheid]
- ✓ soort monster (dagverzamelmonster of steekmonster)
- ✓ begin- en einddatum geselecteerde deelreeks
- ✓ verwijderde meetwaarden
- ✓ beschikbaar aantal meetwaarden deelreeks na verwijdering meetwaarden
- ✓ tijdreeksplot deelreeks
- ✓ histogram meetinterval
- ✓ gehanteerd meetinterval
- ✓ combinatie van:
    - o tijdreeksplot deelreeks (eventueel na transformatie)
    - o histogram deelreeks (eventueel na transformatie)
    - o PP-plot gehanteerde deelreeks (eventueel na transformatie)
- ✓ resultaat Lilliefors-toets deelreeks (eventueel na transformatie)
- ✓ oordeel gebruiker over normale kansverdeling meetwaarden
- ✓ de gehanteerde transformator van de meetwaarden
- ✓ tijdreeksplot en voortschrijdend gemiddelde deelreeks (eventueel na transformatie)
- ✓ autocorrelogram deelreeks (eventueel na transformatie)
- ✓ oordeel gebruiker over autocorrelatie
- ✓ combinatie van:
    - o tijdreeksplot meetwaarden en lozingseis
    - o tijdreeksplot gemiddelde van 10 opeenvolgende meetwaarden en lozingseis
- ✓ gehanteerde formule bij berekenen lozingseis meetwaarden (evenals de toelichting op de termen)
- ✓ lozingseis voor meetwaarden (en de gehanteerde overschrijdingskans)
- ✓ gehanteerde formule bij berekenen lozingseis voor gemiddelden van 10 opeenvolgende (eventueel getransformeerde) meetwaarden
- ✓ lozingseis voor gemiddelden van 10 opeenvolgende (eventueel getransformeerde) meetwaarden (en de gehanteerde overschrijdingskans).

De gebruiker heeft daarnaast nog de mogelijkheid onderdelen van de handleiding in het uitvoerbestand toe te voegen.

# 4 Seizoensmatige lozing?

## 4.1 Bedoeling

Als een geldig invoerbestand is ingelezen, kan met behulp van het programma worden nagegaan of er al dan niet sprake is van een seizoenspatroon bij het lozen (stap 6 van de nota 'Lozingseisen Wvo-vergunningen'). Er kan bijvoorbeeld sprake zijn van een campagnebedrijf of van seizoensinvloeden op de afvalwaterzuiveringsinstallatie (door de temperatuur). Daartoe moet de knop 'Seizoensmatige lozing?' worden aangeklikt. Het programma biedt dan zowel een visueel hulpmiddel als een statistische toets, om de gebruiker te assisteren bij het verkennen van de seizoensmatigheid.

Als er sprake is van een seizoensmatige lozing, met grote verschillen tussen de seizoenen, dan is het aan te bevelen om lozingseisen op te stellen voor de afzonderlijke seizoenen. Er dient in eerste instantie zoveel mogelijk vanuit theoretisch en praktisch inzicht te worden beredeneerd of er al dan niet sprake is van een seizoensmatige lozing. Maar in geval van twijfel kan dit ook met Lozingseis-assistent worden nagegaan met een beschikbare meetreeks. De gebruiker dient hiertoe dan echter wel zelf aan te geven van welke seizoenen er sprake zou kunnen zijn.

## 4.2 Verkenning te beschouwen meetreeks

Als het ingelezen invoerbestand slechts één meetreeks bevat, zal deze direct na het aanklikken van 'Selectie parameter' als tijdreeksplot worden getoond. Maar als het invoerbestand meetreeksen van meerdere parameters bevat, dan verschijnt er eerst een schermpje waarop de gebruiker moet aangeven welke van de parameters moet worden beschouwd. Na deze selectie verschijnt de tijdreeksplot van de parameter. Uitschieters zijn hierbij gemarkeerd.

Indien gewenst kan een meetwaarde interactief worden verwijderd, door deze in de tijdreeksplot aan te klikken. Hiervoor komen in ieder geval díe meetwaarden in aanmerking, die door een uitzonderlijke fout of een onbeheerste situatie zijn veroorzaakt. Na het verwijderen van een meetwaarde zal het programma opnieuw berekenen of er uitschieters resteren (de kenmerken van de verzameling meetwaarden zijn dan immers veranderd). Een verwijderde meetwaarde wordt paars gemarkeerd in de tijdreeksplot. Desgewenst kan deze ook weer aan de meetreeks worden toegevoegd door hem opnieuw aan te klikken.

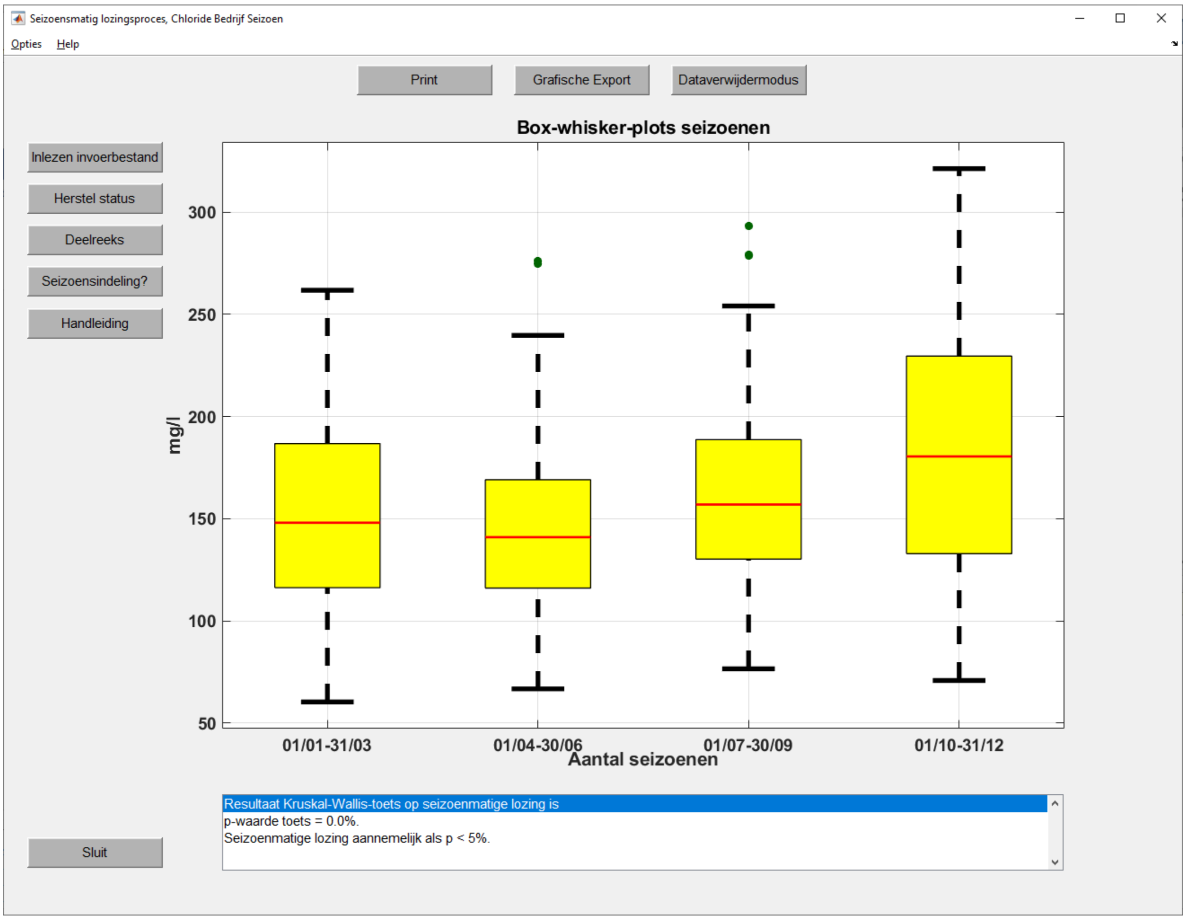
## 4.3 Verdere interactie

Door de knop 'Deelreeks' aan te klikken, verschijnt er een schermpje dat de begin- en einddatum aangeeft van de geselecteerde meetreeks. Als de exercitie moet worden uitgevoerd met een bepaald deel van deze meetreeks, dient de gebruiker de begin- en/of einddatum in dit schermpje aan te passen. Dit heeft overigens nog niets te maken met de seizoensindeling, want dat gebeurt pas in de volgende stap. Als een deelreeks wordt geselecteerd die niet voldoet aan de criteria (minimaal 15 meetwaarden, waarvan minstens 5 verschillend zijn), zal het programma dat melden.

Door vervolgens de knop 'Seizoensindeling?' aan te klikken, verschijnt eerst een schermpje dat vraagt naar het aantal seizoenen en vervolgens een schermpje dat vraagt naar de begindatum van elk seizoen. De seizoenen mogen ook over de jaarwisseling heen lopen, maar het programma gaat er wel vanuit dat de seizoenen samen precies een jaar beslaan.

Hierna volgt een resultaatscherm (zie hieronder), met zowel een visueel hulpmiddel (box-whiskerplots van de seizoenen), als het resultaat van een statistische toets (de Kruskal-Wallis-toets), om na te gaan of er sprake is van een seizoensmatige lozing.

*Voorbeeld van het resultaatscherm van het checken op een seizoensmatige lozing.*




## 4.4 Visueel beoordelen op seizoensmatige lozing

Om visueel te kunnen beoordelen of er sprake is van een seizoensmatige lozing, toont het programma de box-whisker-plots van de door de gebruiker gedefinieerde seizoenen naast elkaar. Eventuele verschillen in de achterliggende kansverdelingen zullen dan zichtbaar worden.

*Toelichting op de box-whisker-plot*

De box-whisker-plot[4] is een handzame manier om de kenmerken van een onderzochte populatie zo compact mogelijk grafisch samen te vatten. De plot toont de posities van de belangrijkste percentielen van de kansverdeling van de gemeten variabele. Het middendeel, de 'box' (doos), loopt van het 25-percentiel naar het 75-percentiel, terwijl het 50-percentiel, oftewel de mediaan, is aangegeven als een dikke streep in de box. De 'whiskers' (snorharen) lopen van de box naar de uiteinden van de verzameling. Extreme meetwaarden in de steekproef zijn weergegeven als een groene punt boven de whisker (de meetwaarde ligt dan meer 1,5 maal de boxlengte vanaf de box). Als er geen extreme meetwaarden zijn dan bevat de onderkant van de onderste whisker een groene punt.
We moeten ons realiseren dat de box-whisker-plot de percentielen toont van de kansverdeling van de gehele bemonsterde populatie, zoals die zijn geschat uit de meetwaarden van de steekproef. Dat is informatie die natuurlijk veel relevanter is dan de paar meetwaarden van de steekproef.

[4] Letterlijk te vertalen als 'doos-met-snorharen'-plot.

## 4.5 Statistisch toetsen op seizoensmatige lozing

Om een aanvullende indicatie te kunnen krijgen over een seizoensmatige lozing, presenteert het programma het resultaat van de Kruskal-Wallis-toets. Maar een waarschuwing is hier op zijn plaats, want deze toets gaat er van uit dat de meetreeks geen autocorrelatie vertoont (oftewel dat de meetwaarden onafhankelijk van elkaar zijn). Dat is echter zelden het geval bij meetreeksen van lozingsparameters, zodat het toetsresultaat enigszins vertekend kan zijn.

Het resultaat van de Kruskal-Wallis-toets is de p-waarde, zijnde een maat voor de aannemelijkheid dat er geen sprake is van een seizoensmatige lozing. Als de p-waarde kleiner is dan 5%, is het aannemelijk dat er sprake is van een seizoensmatige lozing, mits er geen autocorrelatie optreedt in de meetreeks.

*Toelichting op de Kruskal-Wallis-toets*

De Kruskal-Walis-toets is op te vatten als het verdelingsvrije equivalent van variantie-analyse met één factor. De toets is verantwoord toepasbaar, ongeacht het soort kansverdeling, mits de waarden onafhankelijk zijn. Het verlies aan efficiëntie ten opzichte van variantie-analyse is slechts gering, als de Kruskal-Wallis-toets wordt toegepast terwijl er tóch sprake is van normaliteit [Bradley, 1968].

De toets gaat echter uit van onafhankelijke waarden en kan dus een enigszins vertekend beeld opleveren als de meetreeks autocorrelatie vertoont.

Bij het toepassen van de Kruskal-Wallis-toets wordt uitgegaan van het volgende model:

$$x_{ij} = \mu + \gamma_j + e_{ij}$$

met $x_{ij}$ de waarde (hier is dat het maandgemiddelde) in jaar $i$ ($i$=1...$r$) en seizoen $j$ ($j$=1...$s$), $\mu$ het gemiddelde van de kansverdeling van alle mogelijke waarden, $\gamma_j$ het effect van seizoen $j$ en $e_{ij}$ het betreffende modelresidu. De te toetsen nulhypothese (er zijn geen seizoenseffecten) kan worden vertaald als:

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

En de alternatieve hypothese luidt dat het 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:

(1) rangschik de waarden $x_{ij}$ van klein naar groot;

(2) ken de waarden rangnummers $R_{ij}$ toe (1...$n$, waarbij $n=r{\cdot}s$);

(3) bereken de variantie van de rangnummers, volgens:

$$s_R^2 = \frac{1}{n-1}\left[\sum_{i=1}^{r}\sum_{j=1}^{s} R_{ij}^2 - \frac{n(n+1)^2}{4}\right]$$

(4) bereken de toetsingsgrootheid $T$ volgens:

$$T = \frac{1}{s_R^2}\left[\sum_{j=1}^{s}\frac{R_{+j}^2}{r} - \frac{n(n+1)^2}{4}\right]$$

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

(5) omdat $T$ onder de nulhypothese een $\chi^2$-verdeling zal volgen met ($s$-1) vrijheidsgraden, kan worden aangenomen dat de meetreeks seizoenseffecten vertoont als geldt:

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

met $\chi^2_{(1-\alpha,\, s-1)}$ het 100·(1-$\alpha$)-percentiel van de $\chi^2$-verdeling bij $s$-1 vrijheidsgraden (deze waarde is te vinden in statistische tabellen).

# 5 Reductie meetinspanning

## 5.1 Bedoeling

Als een geldig invoerbestand is ingelezen, kan met behulp van het programma worden nagegaan of de relatie tussen twee parameters voldoende sterk is om het meten van één van beide te laten vervallen (stap 5.2 van de nota 'Lozingseisen Wvo-vergunningen'). Daartoe moet de knop 'Reductie meetinspanning' worden aangeklikt. Voorwaarde is wel dat reeds de lozingseis is vastgesteld van de parameter die beoogd is om te laten vervallen uit het meetprogramma (hieronder aangegeven als de parameter Y).

*Gebruik van een model voor de relatie*

Als twee parameters (aan te geven als *Y* en *X*) een voldoende sterke relatie vertonen, kan er één vervallen uit het meetprogramma door gebruik te maken van hun relatie. De relevante uitkomst van dit programma-onderdeel is de *kritieke waarde* van *X*. Als een nieuwe meetwaarde van *X* boven die kritieke waarde ligt, is de kans dat de bijbehorende meetwaarde van *Y* boven de lozingseis voor *Y* ligt niet meer te verwaarlozen. De parameter *Y* dient dan alsnog te worden geanalyseerd in het betreffende monster (zie tabel 5.1). De lozingseis van *Y* dient voor deze toepassing echter wel reeds bekend te zijn.

*Toelichting op het gehanteerde model*

Om de kritieke waarde te kunnen berekenen, wordt de relatie tussen *Y* en *X* uitgedrukt in het volgende lineaire regressiemodel:

$$y_t = b_0 + b_1 \cdot x_t + e_t$$

met $y_t$ , respectievelijk $x_t$ de meetwaarde van *Y*, respectievelijk *X* op tijdstip *t*, $b_0$ en $b_1$ modelcoëfficiënten en $e_t$ het modelresidu voor tijdstip *t*. Uit elke nieuwe meetwaarde van *X* ($x_k$) kan met dit model vervolgens de daarbij behorende meetwaarde van *Y* ($y_k$) worden voorspeld, volgens:

$$\hat{y}_k = b_0 + b_1 \cdot x_k$$

met $\hat{y}_k$ de voorspelling van $y_k$. De kritieke waarde van *X* kan alleen worden berekend als de modelresiduën afkomstig zijn uit een normale kansverdeling. Als dat niet het geval blijkt, biedt het programma nog de mogelijkheid dit te bewerkstelligen door het transformeren van *Y* en/of *X*. Als echter ook na het transformeren de modelresiduën niet afkomstig zijn uit een normale kansverdeling zal het programma aangeven dat het geen kritieke waarde kan berekenen.

## 5.2 Verkenning te beschouwen meetreeks

Door 'Selectie parameters' aan te klikken verschijnt er een schermpje waarop de gebruiker moet aangeven van welke parameters de onderlinge relaties moeten worden verkend. Dit moeten er minstens twee zijn en kunnen er hoogstens acht zijn. Deelselecties kunnen worden gemaakt door behalve de muis ook de Ctrl-toets te gebruiken.

Vervolgens kunnen de tijdreeksplots van alle geselecteerde parameters worden verkend. Daartoe kan met het schuifje rechtsboven van de ene tijdreeksplot naar de andere worden gebladerd. In elke tijdreeksplot zijn de uitschieters groen gemarkeerd.

Indien gewenst kan een meetwaarde interactief worden verwijderd, door deze in de tijdreeksplot aan te klikken. Hiervoor komen in ieder geval díe meetwaarden in aanmerking, die door een uitzonderlijke fout of een onbeheerste situatie zijn veroorzaakt. Na het verwijderen van een meetwaarde zal het programma opnieuw berekenen of er uitschieters resteren (de kenmerken van de verzameling meetwaarden zijn dan immers veranderd). Een verwijderde meetwaarde wordt paars gemarkeerd in de tijdreeksplot. Deze kan ook weer aan de meetreeks worden toegevoegd door hem opnieuw aan te klikken.

## 5.3 Knoppen

De volgende knoppen zijn vervolgens beschikbaar voor dit onderdeel:

- Herstel status
- Deelreeksen
- Meetinterval
- X en Y
- Residu normaal?
- Residu autocorr?
- Kritieke waarde

**De statusbalk**

Om de gebruiker te ondersteunen bij het verstrekken van de informatie voor dit onderdeel, wordt onderin het scherm een statusbalk weergegeven. Deze geeft per onderdeel aan welke informatie reeds beschikbaar is (groen veld) en welke nog moet worden ingevuld (rood veld). Om de kritieke waarde te kunnen berekenen (knop 'Kritieke waarde') dienen alle velden van deze statusbalk ingevuld – en dus groen - te zijn.

*Beginstand van de statusbalk. De transformaties staan bij aanvang op 1, wat wil zeggen dat er wordt begonnen zónder transformatie van de meetwaarden ($x^1$ is immers gelijk aan x en $y^1$ is gelijk aan y).*

| Deelreeksen | Meetinterval | X en Y | Transformaties | Residu normaal? | Residu autocorr.? |
|---|---|---|---|---|---|
| ? | ? | ? | x^1 en y^1 | ? | ? |

## 5.4 Herstel status

De knop 'Herstel status' is een herstelknop, die lopende een sessie kan worden aangeklikt, zodat opnieuw kan worden begonnen met de ingelezen meetreeksen. Door deze aan te klikken, wordt de statusbalk weer in zijn oorspronkelijke staat teruggebracht en vormen de oorspronkelijk ingelezen meetreeksen weer het uitgangspunt.

## 5.5 Deelreeksen

Als de knop 'Deelreeksen' wordt aangeklikt, verschijnt een schermpje dat de begindatum en einddatum aangeeft van alle ingelezen meetreeksen (deze zijn immers even lang). Als de mogelijkheden van het reduceren van de meetinspanning moeten worden geëvalueerd met een bepaald deel van de verstrekte meetreeksen, dient de gebruiker de begin- en/of einddatum in dit schermpje aan te passen. Als na deze selectie niet wordt voldaan aan de criteria (minimaal 15 gepaarde meetwaarden, waarvan minstens 5 verschillend zijn), zal het programma dat melden.

## 5.6 Meetinterval

Als de knop 'Meetinterval' wordt aangeklikt verschijnt een histogram van het meetinterval van de meetwaarden in de beschikbare deelreeksen, evenals een schermpje dat reeds het meest voorkomende meetinterval aangeeft. Het in het schermpje aangegeven meetinterval zal worden gehanteerd bij het evalueren van de mogelijkheden van het reduceren van de meetinspanning. Indien gewenst kan de gebruiker dit meetinterval in het schermpje aanpassen. Als na deze selectie niet wordt voldaan aan de criteria (minimaal 15 gepaarde meetwaarden, waarvan minstens 5 verschillend zijn), zal het programma dat melden.

Als er verschillende meetintervallen voorkomen en de gebruiker wenst dat alle meetwaarden worden gehanteerd, moet '0' worden ingevuld. Er kan dan echter niet meer worden gecorrigeerd voor een eventuele autocorrelatie. Dit heeft als consequentie dat een berekende kritieke waarde voor meetwaarden van *X* slechts bruikbaar kan worden geacht als zowel de meetreeks van *X* als de meetreeks van *Y* representatief is voor het hele bereik aan mogelijke meetwaarden bij beheerste procesvoering (dit wordt ook als waarschuwing bij de uitvoer vermeld).

## 5.7 X en Y

Als de knop 'X en Y' wordt aangeklikt worden de relaties vastgesteld tussen de geselecteerde parameters. Voor elk paar parameters wordt de relatie gekwantificeerd in de Spearman-rangcorrelatiecoëfficiënt, zijnde een maat voor de samenhang in hun meetwaarden.

*Toelichting op de Spearman-rangcorrelatiecoëfficiënt*

De Spearman-rangcorrelatiecoëfficiënt is berekend als:

$$r_{Sp} = \frac{\sum_{t=1}^{n}(Rx_t \cdot Ry_t) - n \cdot \left(\frac{n+1}{2}\right)^2}{n \cdot (n^2 - 1)/12}$$

met $Rx_t$ het rangnummer van $x_t$ als de meetwaarden van *X* oplopend zijn gerangschikt, $Ry_t$ hetzelfde voor *Y* en *n* het aantal gepaarde meetwaarden van *X* en *Y*. Het is in feite de (gewone) correlatiecoëfficiënt van de rangnummers van *X* en *Y*.

De Spearman-rangcorrelatiecoëfficiënten worden weergegeven in een correlatiematrix, waarbij de sterke relaties ($r_{sp}$ > 0,50) zijn vetgedrukt en gearceerd (zie onder).

Door op een cel van de correlatiematrix te klikken, geeft de gebruiker aan van welke twee parameters de relatie verder moet worden uitgewerkt (voor bovenstaand voorbeeld ligt het voor de hand hiervoor de parameters Cl en $SO_4$ te kiezen). Let hierbij op het volgende:

(1) *Y* is de parameter die is beoogd is om minder te meten (de 'gecorreleerde parameter');
(2) *X* is de parameter die is beoogd om *Y* te voorspellen (de 'correlatieparameter');
(3) de lozingseis van *Y* dient reeds bekend te zijn.

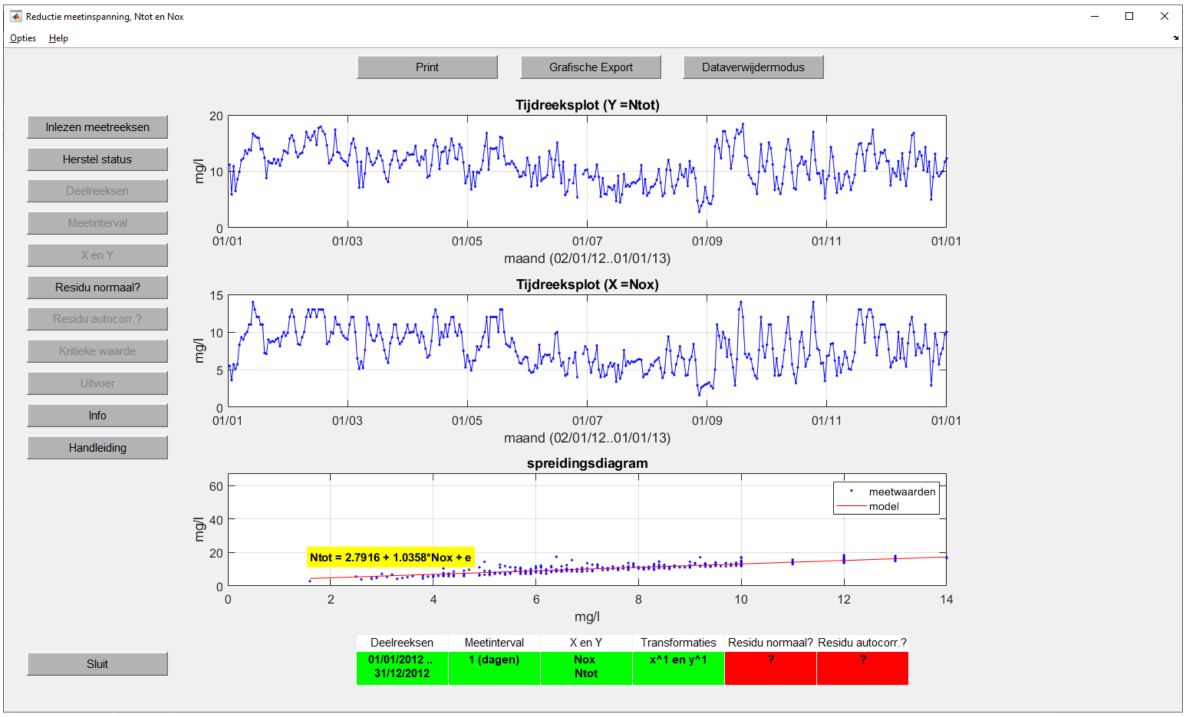
Het ligt voor de hand om als *Y* te kiezen de parameter die het duurst (of traagst) is om te meten.

Vervolgens vraagt het programma naar de lozingseis voor *Y* (het maakt hierbij niet uit of dit een gesloten of een open lozingseis is). Deze lozingseis moet dus al eerder zijn vastgesteld (bijvoorbeeld met het programma-onderdeel 'Afleiden lozingseis').

Het programma berekent vervolgens het lineaire regressiemodel en toont op een nieuw scherm de volgende drie grafieken (zie het voorbeeld in onderstaande figuur):

(1) een tijdreeksplot waarin de meetwaarden van *Y* zijn uitgezet tegen de tijd;
(2) een tijdreeksplot waarin de meetwaarden van *X* zijn uitgezet tegen de tijd;
(3) een spreidingsdiagram, waarin de meetwaarden van *Y* zijn uitgezet tegen die van *X*. Tevens is hierin de lijn weergegeven van het lineaire regressiemodel van hun relatie en is het model ook uitgeschreven.

*Tijdreeksplots van Y en X en hun spreidingsdiagram.*



## 5.8 Residu normaal?

Aangezien het voor de berekening van de kritieke waarde nodig is om vast te stellen of de modelresiduën al dan niet afkomstig zijn uit een normale kansverdeling, dient vervolgens de knop 'Residu normaal?' te worden aangeklikt. Er verschijnt dan een nieuw scherm, dat vier hulpmiddelen presenteert om de gebruiker te assisteren bij het formuleren van zijn oordeel over het soort kansverdeling waar de modelresiduën uit afkomstig zijn. Dit zijn drie visuele hulpmiddelen en het resultaat van een statistische toets (zie § 3.7 voor een toelichting):

(1) een tijdreeksplot van de modelresiduën;
(2) het histogram van de modelresiduën;
(3) de PP-plot van de modelresiduën;
(4) het resultaat van de Lilliefors-toets op normaliteit van de modelresiduën.

Tevens verschijnt er een schermpje dat de gebruiker vraagt of de modelresiduën al dan niet afkomstig zijn uit een normale kansverdeling.

*Als de modelresiduën niet afkomstig zijn uit een normale kansverdeling*

Als de gebruiker aangeeft dat de modelresiduën niet afkomstig zijn uit een normale kansverdeling, zal het programma zelf nagaan of normaliteit van de modelresiduën is te bewerkstelligen door *Y* en/of *X* te transformeren. Als criterium voor het al dan niet transformeren wordt het resultaat van de Lilliefors-toets op normaliteit (de p-waarde) gehanteerd. Als transformeren nodig wordt geacht, berekent het programma voor de betreffende parameter díe transformator die de correlatiecoëfficiënt tussen de meetwaarden en hun normaalscores maximaliseert. Dit is dus de waarde die de punten in de PP-plot zoveel mogelijk een rechte lijn laat volgen (zie § 3.7.1 voor een toelichting).

Als de bovenstaande procedure zowel voor *Y* als voor *X* is doorlopen en minstens één van beide is getransformeerd, berekent het programma een nieuw model en moet worden vastgesteld of de nieuwe modelresiduën al dan niet afkomstig zijn uit een normale kansverdeling. Daartoe verschijnt een nieuw scherm met:

(1) een tijdreeksplot van de modelresiduën;
(2) het histogram van de modelresiduën;
(3) de PP-plot van de modelresiduën;
(4) het resultaat van de Lilliefors-toets op normaliteit van de modelresiduën.

Tevens verschijnt er een schermpje dat de gebruiker vraagt of de modelresiduën al dan niet afkomstig zijn uit een normale kansverdeling. Als de gebruiker aangeeft dat dat niet het geval is, verschijnt er een melding dat de kritieke waarde alleen kan worden berekend als de modelresiduën afkomstig zijn uit een normale kansverdeling en stopt de sessie.

## 5.9 Residu autocorr?

Als door de bovenstaande interactie bekend is geworden dat de modelresiduën afkomstig zijn uit een normale kansverdeling en het opgegeven meetinterval groter is dan 0, moet nog worden vastgesteld of ze al dan niet autocorrelatie vertonen. Daartoe moet de knop 'Residu autocorr?' worden aangeklikt. Er verschijnt dan een nieuw scherm, met twee hulpmiddelen om de gebruiker te assisteren bij het formuleren van zijn oordeel over het al of niet optreden van autocorrelatie, namelijk (zie § 3.8 voor een toelichting):

(1) een tijdreeksplot van de modelresiduën, met tevens weergegeven hun voortschrijdend gemiddelde (over 30 modelresiduën);
(2) het autocorrelogram van de modelresiduën.

Tevens verschijnt er een klein scherm waarop de gebruiker kan aangeven of de modelresiduën al dan niet autocorrelatie vertonen. Als de gebruiker aangeeft dat er sprake is van autocorrelatie, vraagt het programma tot en met welk meetinterval daarvan sprake is. Het betreffende schermpje bevat ook al een voorstel voor dat aantal, maar dit kan de gebruiker desgewenst aanpassen.

## 5.10 Kritieke waarde

Als de statusbalk helemaal groen is, kan de knop 'Kritieke waarde' worden aangeklikt. Er verschijnt dan een nieuw scherm, met een spreidingsdiagram, waarin *Y* is uitgezet tegen *X*. Hierin zijn tevens weergegeven (zie onderstaande figuur):

- de lijn van het lineaire regressiemodel van hun relatie;
- de tolerantielimiet van *Y*, gegeven de meetwaarde van *X* (zie het tekstkader op de volgende pagina);
- de lozingseis voor *Y*;
- de verticale lijn door het snijpunt van de lozingseis voor *Y* en de tolerantielimiet van *Y*. Deze verticale lijn snijdt de X-as op de kritieke waarde van *X*;

De berekende kritieke waarde van *X* is tevens apart vermeld.

*De kritieke waarde van X wordt bepaald door het snijpunt van de lozingseis en de tolerantielimiet van Y.*

In bovenstaand voorbeeld is de kritieke waarde van *X* 61,8 mg/l. Als de meetwaarde van *X* meer bedraagt dan deze kritieke waarde, dient *Y* alsnog te worden geanalyseerd in het betreffende monster.

Als een berekende kritieke waarde groter is dan het maximum van de beschikbare meetwaarden van *X*, dan wordt de kritieke waarde op dat maximum gesteld. Het lineaire regressiemodel is namelijk afgeleid voor het bereik van de beschikbare meetwaarden van *X* en het is té speculatief om te veronderstellen dat het ook daarbuiten geldig is.

*Toelichting op het berekenen van de tolerantielimiet*

De berekening van de tolerantielimiet van *Y*, gegeven de meetwaarde $x_k$, houdt rekening met de kansverdeling en de autocorrelatie van de modelresiduën, zoals in onderstaande wordt toegelicht.

Het lineaire regressiemodel luidt:

$$y_t = b_0 + b_1 \cdot x_t + e_t$$

met $y_t$ en $x_t$ de meetwaarde van *Y*, respectievelijk *X* op tijdstip *t*, $b_0$ en $b_1$ modelcoëfficiënten en $e_t$ het modelresidu voor tijdstip *t*. Hiermee kan uit elke nieuwe meetwaarde van *X* ($x_k$) de daarbij behorende meetwaarde van *Y* ($y_k$) worden voorspeld, volgens:

$$\hat{y}_k = b_0 + b_1 \cdot x_k$$

met $\hat{y}_k$ de voorspelling van $y_k$. Om de tolerantielimiet van de meetwaarde van *Y* te berekenen, gegeven de meetwaarde $x_k$, onderscheidt Lozingseis-assistent de volgende twee gevallen:

(1) modelresiduën normaal verdeeld, zonder autocorrelatie;
(2) modelresiduën normaal verdeeld, met autocorrelatie.

(1) Modelresiduën normaal verdeeld, zonder autocorrelatie
Als de modelresiduën afkomstig zijn uit een normale kansverdeling en geen autocorrelatie vertonen, dan berekent het programma de tolerantielimiet$_{(100\%\cdot\gamma,\ 95\%)}$ van de meetwaarde van *Y*, gegeven de meetwaarde $x_k$, volgens:

$$TL_{k(100\%\cdot\gamma,95\%)} = \hat{y}_k + \frac{z_{(\gamma)} + \sqrt{\left(z_{(\gamma)}^2 - (1 - \frac{z_{(0,95)}^2}{2\cdot(n-1)})\cdot(z_{(\gamma)}^2 - \frac{z_{(0,95)}^2}{n})\right)}}{(1 - \frac{z_{(0,95)}^2}{2\cdot(n-1)})} \cdot s_e$$

met *n* het aantal gepaarde meetwaarden van *X* en *Y* waarmee het lineaire regressiemodel is geschat, $s_e$ de standaardafwijking van de modelresiduën, $z_{(\gamma)}$ het 100·$\gamma$-percentiel van de standaardnormale verdeling (standaard is dit het 99,9-percentiel en anders het 99-percentiel) en $z_{(0,95)}$ het 95-percentiel van de standaardnormale verdeling.

(2) Modelresiduën normaal verdeeld, met autocorrelatie
Als de modelresiduën wél autocorrelatie vertonen, zal $s_e$ geen zuivere schatting vormen van de standaardafwijking van de modelresiduën. Als de autocorrelatie niet te groot is, zal het programma daarvoor als volgt corrigeren:

$$s_e^* = \sqrt{\frac{s_e^2}{1 - \frac{2}{n \cdot n - 1} \cdot \sum_{l=1}^{n-1} ((n-l) \cdot \hat{\rho}_l)}}$$

met $\hat{\rho}_l$ de geschatte autocorrelatiecoëfficiënt van de modelresiduën voor tijdsinterval *l*.

## Geraadpleegde literatuur


ASTM (1984): ‘Annual Book of ASTM Standards Section 11, Water and Environmental Technology’. Vol. 11.01. Designation D4210-83, pp. 7 - 15. American Society for Testing and Materials.

Baggelaar, P.K. (2003): ‘Statistische aspecten van lozingseisen’. Icastat Statistisch Adviesbureau, mei 2003.

Barnett, V. and O’ Hagan, A. (1997): ‘Setting environmental standards: the statistical approach to handling uncertainty and variation’. Chapman and Hall, London.

Bayley, G.V. and Hammersley, J.M. (1946): ‘The effective number of independent observations in an autocorrelated time series’. Journal of the Royal Statistical Society, 8(1B), 1946, pp. 184 – 197.

Box, G.E.P. and Jenkins, G.M. (1976): ‘Time Series Analysis: Forecasting and Control’. Holden-Day, San Francisco.

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

Brockwell, P.J. and Davis, R.A. (1986): ‘Time Series: Theory and Methods’. Springer, Berlin.

Conover, W.L. (1980): ‘Practical Nonparametric Statistics’. John Wiley and Sons, New York, 493 pp.

Dodge, H.F. (1977): ‘Keep it Simple’. Journal of Quality Technology, 9, 3, pp. 102 (1977).

EPA (1980): ‘Upgrading Environmental Radiation Data’. EPA 520/1-80-012, Office of Radiation Programs, U.S. Environmental Protection Agency, Washington D.C.

Gibbons, R.D. (1994): ‘Statistical Methods for Groundwater Monitoring’. John Wiley and Sons, New York, 286 pp.

Gilbert, R.O. (1987): ‘Statistical Methods for Environmental Pollution Monitoring’. Van Nostrand Reinhold, New York, 320 pp.

Gilliom, R.J., Hirsch, R.M. and Gilroy, E.J. (1984): ‘Effect of censoring trace-level water-quality data on trend-detection capability’. Environmental Science and Technology, Vol. 18, 1984, pp. 530 - 535.

Gilliom, R.J. and Helsel, D.R. (1986): ‘Estimation of Distributional Parameters for Censored Trace Level Water Quality Data - 1 - Estimation Techniques’. Water Resources Research, Vol. 22, No. 2, February 1986, pp. 135 - 146.

Hald, A. (1952): ‘Statistical Theory with Engineering Applications’. John Wiley & Sons, New York, 783 pp.

Helsel D.R. and Gilliom, R.J. (1986): ‘Estimation of Distributional Parameters for Censored Trace Level Water Quality Data - 2 - Verification and Applications’. Water Resources Research, Vol. 22, No. 2, February 1986, pp. 147 - 155.

Helsel, D.R. and Cohn, T.A. (1988): ‘Estimation of Descriptive Statistics for Multiply Censored Water Quality Data’. Water Resources Research, Vol. 24, No. 12, December 1988, pp. 1997 - 2004.

Helsel, D.R. (1990): ‘Less than obvious - Statistical treatment of data below the detection limit’. Environmental Science and Technology, Vol. 24, No. 12, 1990, pp. 1766 - 1774.

Helsel, D.R. and Hirsch, R.M. (1992): ‘Statistical Methods in Water Resources’. Studies in Environmental Science 49, Elsevier, Amsterdam, 522 pp.

Klavers, H.C., De Vries, A., Bekkers, L. en Van Twuiver, H. (1992): ‘Vrachtberekeningsmethoden voor Maas en Rijn’. Rapport RIZA/DGW.

Matalas, N.C. and Langbein, W.B. (1962): ‘Information content of the mean’. Journal of Geophys. Research, 67(9), 1962, pp. 3441 - 3448.

Montgomery, D.C. (1991): ‘Introduction to Statistical Quality Control’. John Wiley & Sons, New York, second edition, 674 pp.

Natrella, M.G. (1963): ‘Experimental Statistics’. NBS Handbook 91, US Department of Commerce.

Porter, P.S., Ward, R.C. and Bell, H.F. (1988): ‘The detection limit - water quality monitoring data are plagued with levels of chemicals that are too low to be measured precisely’. Environmental Science and Technology, Vol. 22, No. 8, 1988, pp. 856 - 861.

Swaving, M. en De Vries, L. (2000): ‘Omgaan met waarden onder de detectiegrens’. Rapport E1680-01, CQM BV, Eindhoven, 7 november 2000.

Ward, R.C., Loftis, J.C. and McBride, G.B. (1990): ‘Design of Water Quality Monitoring Systems’. Van Nostrand Reinhold, New York, 231 pp.

Zhang, N.F. (1998): ‘A Statistical Control Chart for Stationary Process Data’. Technometrics, February 1998, Vol. 40, No. 1, pp. 24 – 38.

## Bijlage 1 – 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. Dit leidt dan tot een te krappe lozingseis.

**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 geloosde parameters).

**Gecensureerde meetwaarden**: waarden aangeduid als een '< '-teken, gevolgd door een bepaalde concentratie, die staat voor de rapportagegrens (bijvoorbeeld '< 1 mg/l'). Bij een aantal analytisch chemische bepalingen wordt een meetwaarde onder een bepaald niveau gecensureerd, als zijn relatieve precisie door het betreffende laboratorium te laag wordt geacht.

**Gesloten lozingseis**: een lozingseis die niet overschreden mag worden. Deze kan worden berekend voor meetwaarden als die, eventueel na transformatie, afkomstig zijn uit een normale kansverdeling. En deze kan worden berekend voor gemiddelden van 10 (eventueel getransformeerde) meetwaarden, als die gemiddelden afkomstig zijn uit een normale kansverdeling.

**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 geloosde parameter) de kansen op realisatie geeft van alle mogelijke meetwaarden.

**Lineair regressiemodel**: model voor de lineaire relatie tussen twee verschijnselen (zoals de concentraties van twee gelooosde parameters).

**Meetinterval**: de periode tussen twee opeenvolgende metingen.

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

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

**Open lozingseis**: een lozingseis die in een bepaald bekend percentage van de gevallen overschreden mag worden. Als de meetwaarden, ook na een eventuele transformatie, niet afkomstig zijn uit een normale kansverdeling, kan alleen een open lozingseis worden berekend. Daar wordt door het programma dan echter ook het toegestane percentage overschrijdingen bij vermeld.

**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.

**Tolerantielimiet$_{(\gamma,1-\alpha)}$**: de bovengrens van het 100%·(1-$\alpha$)-betrouwbaarheidsinterval van het geschatte 100·$\gamma$-percentiel van een kansverdeling.

**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. Dit heeft als voordeel dat er dan scherpere kansuitspraken kunnen worden gedaan, bijvoorbeeld over toekomstige meetwaarden.

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

**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.

## Bijlage 2 – Oplossingen als er geen lozingseis voor gemiddelden kan worden afgeleid

Bij het afleiden van een lozingseis zal het programma in bepaalde gevallen aangeven dat er geen lozingseis voor gemiddelden van 10 opeenvolgende meetwaarden kan worden berekend. Soms is er dan echter toch nog een oplossing beschikbaar om tot een dergelijke lozingseis te kunnen komen. Die oplossingen zijn hieronder beschreven.

**Geval 1**. *Niet-normale kansverdeling van de meetwaarden*
Als de gebruiker heeft aangegeven dat de (eventueel getransformeerde) meetwaarden niet afkomstig zijn uit een normale kansverdeling, dan wordt er geen lozingseis afgeleid voor gemiddelden van 10 opeenvolgende meetwaarden. Het programma moet er bij het afleiden van een dergelijke lozingseis namelijk van uit kunnen gaan dat de gemiddelden van 10 opeenvolgende (eventueel getransformeerde) meetwaarden afkomstig zijn uit een normale kansverdeling. Het aantal beschikbare gemiddelden is doorgaans echter te gering – namelijk hooguit 10% van het aantal meetwaarden in de meetreeks -, om empirisch te kunnen vaststellen of ze al dan niet afkomstig zijn uit een normale kansverdeling. Lozingseis-assistent moet daarom afgaan op wat de gebruiker aangeeft over het al dan niet normaal verdeeld zijn van de (eventueel getransformeerde) *meetwaarden*. Maar het is zeer wel mogelijk dat gemiddelden, die zijn berekend uit 10 opeenvolgende meetwaarden uit een niet-normale kansverdeling, zelf wél afkomstig zijn uit een normale kansverdeling. Volgens de statistische theorie (de Centrale Limietstelling) zal een kansverdeling van gemiddelden immers meer naderen tot de normale kansverdeling, naarmate er meer meetwaarden worden gemiddeld, ongeacht de kansverdeling van de meetwaarden. Als de gebruiker veronderstelt dat hier sprake van kan zijn, dan kan met de volgende stappen toch nog een lozingseis voor gemiddelden van 10 opeenvolgende meetwaarden worden afgeleid:

1. herhaal de sessie voor het afleiden van de lozingseis (via 'Herstel status');
2. kies bij 'Normaal verdeeld?' voor een normale kansverdeling van de meetwaarden (ook al is daar dus geen sprake van);
3. geef bij 'Autocorrelatie?' aan of daar al of niet sprake van is;
4. klik op 'Lozingseis' om de lozingseis af te leiden;
5. in de onderste tijdreeksplot is nu wél een lozingseis weergegeven voor gemiddelden van 10 meetwaarden. Check visueel of deze lozingseis past bij de karakteristieken van de weergegeven tijdreeks van gemiddelden;
6. als deze check geen twijfel oproept, neem dan de bij 'Uitvoer' of 'Info' vermelde lozingseis voor gemiddelden van 10 opeenvolgende meetwaarden over. Sla hierbij echter geen acht op de daar vermelde lozingseis voor meetwaarden, aangezien die uitgaat van een normale kansverdeling van de meetwaarden (dat gaat hier immers niet op).

De praktijk heeft geleerd dat deze aanpak in veel gevallen toch nog een bruikbare lozingseis op zal leveren voor gemiddelden van 10 opeenvolgende meetwaarden. Zeker als de meetwaarden niet afkomstig zijn uit een al te scheve kansverdeling, zal dit een vrij robuuste benadering blijken.

**Geval 2**. *Meetinterval '0'*
Een lozingseis voor gemiddelden van 10 opeenvolgende meetwaarden kan evenmin worden afgeleid als de gebruiker als meetinterval '0' heeft opgegeven. Daarmee heeft de gebruiker impliciet aangegeven dat de meetreeks (min of meer ) representatief mag worden geacht voor het bereik aan mogelijke meetwaarden bij de gebruikelijke, beheerste procesvoering. Voor wat betreft het afleiden van de lozingseis voor *meetwaarden* heeft dit als voordeel dat er dan geen rekening meer hoeft te worden gehouden met een eventuele autocorrelatie. Verder mogen nu álle meetwaarden bij de afleiding worden gehanteerd, ongeacht het meetinterval. Maar het nadeel is dat er geen lozingseis voor gemiddelden van 10 opeenvolgende meetwaarden kan worden afgeleid. Zonder specificatie van het meetinterval kan er immers niet meer worden vastgesteld óf er sprake is van

autocorrelatie en zo ja, hoe groot die is. Die informatie is echter onontbeerlijk bij het afleiden van de lozingseis voor gemiddelden van 10 opeenvolgende meetwaarden. Om die lozingseis toch af te kunnen leiden zal de gebruiker de sessie moeten herhalen, maar bij ‘Meetinterval’ moet dan een waarde groter dan 0 worden aangegeven. Houd er wel rekening mee dat de afgeleide lozingseis voor gemiddelden dan alleen geldt voor gemiddelden van 10 opeenvolgende meetwaarden bij dát specifieke meetinterval.