



KWR 2015.004 | April 2015

Kwaliteitsborging grondwaterstands- en stijghoogtegegevens

Systematiek en
methodiek voor
datakwaliteits-
controle (QC)

Kwaliteitsborging grondwaterstands- en stijghoogtegegevens

Systematiek en methodiek voor datakwaliteitscontrole (QC)

KWR 2015.004 | April 2015

Projectmanager

Martin de Haan

Kwaliteitsborger(s)

Willem Jan Zaadnoordijk, Flip Witte

Opdrachtgever(s)

Eric Castenmiller (Provincie Limburg) en Henny Kempen (Provincie Gelderland), namens de gezamenlijke provincies

Stuurgroep en begeleidingsgroep

Zie bijlagen A en B, onder voorzitterschap van Henny Kempen, Janco van Gelderen en Jan Meijles (Provincies Gelderland, Utrecht en Zuid-Holland)

Met dank aan

Joost Gooijer en Thomas de Meij (Provincie Overijssel) voor gebruik van gezamenlijk ontwikkeld materiaal. De stuurgroep en provinciale meetnetbeheerders voor de vele nuttige discussies en inbreng.

Auteur(s)

Jos von Asmuth, Frans van Geer

Jaar van publicatie
2015

Meer informatie

T
E

PO Box 1072
3430 BB Nieuwegein
The Netherlands

T +31 (0)30 60 69 511
F +31 (0)30 60 61 165
E info@kwrwater.nl
I www.kwrwater.nl

TNO innovation
for life

KWR Watercycle
Research
Institute

KWR | April 2015 © KWR

Alle rechten voorbehouden.

Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen, of enig andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

Inhoud

Inhoud	3
1 Inleiding	5
1.1 Aanleiding en achtergrond	5
1.2 Doelstelling	5
1.3 Organisatie	6
1.4 Leeswijzer	6
2 Bestaande methoden en richtlijnen	7
2.1 Inleiding	7
2.2 Kwaliteit van producten in het algemeen	7
2.3 Kwaliteit van (grond)waterdata in het bijzonder	9
2.4 Gerelateerde initiatieven en onderwerpen	10
3 Onderwerp en afbakening	13
3.1 Inleiding	13
3.2 Kwaliteitscontrole: wanneer en waarop?	13
3.3 Data: wat wel en wat niet?	15
4 Systematiek en terminologie	17
4.1 Inleiding	17
4.2 Data-integriteit (administratieve controle)	17
4.3 Consistentie (relationele controle)	18
4.4 Plausibiliteit (statistische controle)	19
4.5 Controle en correctie van sensordata (controlemeting)	19
4.6 Nauwkeurigheid (kwantificering van)	19
5 Data-integriteit (administratieve controle)	21
5.1 Inleiding	21
5.2 Methoden en criteria	22
5.3 Geautomatiseerde controle	26
5.4 Visuele controle	26
6 Consistentie (relationele controle)	27
6.1 Inleiding	27
6.2 Methoden en criteria	27
6.3 Geautomatiseerde controle	28
6.4 Visuele controle	28
7 Plausibiliteit (statistische controle)	31
7.1 Inleiding	31
7.2 Methoden en criteria	31
7.3 Geautomatiseerde controle	34
7.4 Visuele controle	34

8	Controle en correctie van sensordata (controlemeting)	37
8.1	Inleiding	37
8.2	Methoden en criteria	39
8.3	Geautomatiseerde controle	45
8.4	Visuele controle	45
9	Implementatie en aanbevelingen	47
9.1	Inleiding	47
9.2	Implementatie, facilitatie en borging van het gebruik	47
9.3	Aanbevelingen	49
	Literatuur	51
	Bijlage A: Data-integriteitscontrole	53
	Bijlage B: Consistentiecontrole	54
	Bijlage C: Begeleidingsgroepleden	55
	Bijlage D: Stuurgroepleden	56

1 Inleiding

1.1 Aanleiding en achtergrond

De afgelopen periode is het inzicht gegroeid dat de kwaliteit en betrouwbaarheid van grondwaterstands- en stijghoogtegegevens van wezenlijk belang zijn voor het gebruik ervan. Het is hierbij niet alleen belangrijk dat de kwaliteit van de gegevens ‘gemiddeld gezien hoog’ is. Minstens zo belangrijk is dat de kwaliteit ‘bekend en van een gegarandeerd niveau’ is, ofwel dat de kwaliteit geborgd is volgens gestandaardiseerde, eenduidige en heldere procedures en methoden. Het is wellicht vanuit die optiek dat de Kaderrichtlijn Water (KRW) het toepassen van een gestandaardiseerd protocol voor datakwaliteitscontrole verplicht stelt.

Het platform meetnetbeheerders grondwaterkwantiteit van de gezamenlijke provincies, heeft daarom het realiseren van een dergelijk protocol ter hand genomen, met KWR en TNO als samenwerkingspartners en opdrachtnemers. In dat kader zijn inmiddels een tweetal rapporten verschenen:

- **Von Asmuth, J.R. en F. C. Van Geer (2013)** Kwaliteitsborging grondwaterstands- en stijghoogtegegevens: op weg naar een landelijke standaard; rapportnr. KWR 2013.027, KWR Watercycle Research Institute / TNO, Nieuwegein / Utrecht.
- **Leunk, I. (2014)** Kwaliteitsborging grondwaterstands- en stijghoogtegegevens: Validatiepilot, analyse van bestaande data; rapportnr. KWR 2014.059, KWR Watercycle Research Institute, Nieuwegein.

In het eerste rapport is een aanzet gegeven voor een landelijke standaard, met daarin o.a. een inventarisatie en overzicht van methoden voor geautomatiseerde controle van de plausibiliteit van meetgegevens. In het tweede rapport zijn de fouten en afwijkingen die in de praktijk optreden in grondwaterstandsgegevens, en de relatieve grootte en belang daarvan, vervolgens beter in beeld gebracht.

1.2 Doelstelling

Er is binnen en buiten de gezamenlijke provincies en de KRW-monitoring een duidelijke behoefte aan een standaard en richtlijnen voor kwaliteitscontrole (zie ook Sorensen en Butcher, 2011; Von Asmuth en Van Geer, 2013; Broekhuis e.a., 2014). Het onderwerp kwaliteit, kwaliteitsborging en kwaliteitscontrole van (grondwater)data is echter complex en veelomvattend, waardoor er naast veel verschillende meningen en werkwijzen ook veel spraakverwarring over is.

De doelstelling van dit rapport is in de eerste plaats het bieden van een systematisch kader, om verschillende aspecten van datakwaliteit van elkaar te kunnen onderscheiden en ordenen. Het geeft daarnaast een overzicht, definities en achtergronden bij verschillende methoden van datakwaliteitscontrole. Het rapport is vervaardigd ten behoeve van het besluitvormingsproces rond de totstandkoming van het genoemde protocol voor datakwaliteitscontrole van de gezamenlijke provincies. Het protocolrapport zal tegelijkertijd verschijnen, en bevat de eigenlijke werkwijzen, controles en criteria die in de praktijk gehanteerd zullen worden door de meetnetbeheerders van de provincies. Dit rapport fungeert daarmee ook als naslagwerk, opdat de in het protocol gehanteerde termen, werkwijzen en richtlijnen helder en eenduidig gedefinieerd zijn.

In dit rapport zijn de opgedane ervaringen en resultaten van de eerste twee projecten meegenomen. Er is bovendien geput uit ervaringen van derden en andere projecten (bijv. Goering, 2009; De Meij en Von Asmuth, 2011; Walvoort e.a., 2013; Van Geer, 2013; Leunk en von Asmuth, 2013; Broekhuis e.a., 2014; Von Asmuth en Leunk, 2014). Alhoewel het de Kaderrichtlijn Water is die het hanteren van een protocol voorschrijft, is het wenselijk dat het

protocol daarnaast ook door andere partijen en voor andere meetlocaties gebruikt kan worden, inclusief die van andere waterpartners zoals waterschappen, waterleidingbedrijven, terreinbeherende instanties en gemeentes.

1.3 Organisatie

Omdat toepassing van de hier beschreven systematiek en methodiek in de praktijk berust op het draagvlak daarvoor, is gekozen voor een zorgvuldige sturing, begeleiding en terugkoppeling binnen de provincies, maar ook daarbuiten. De begeleidingsstructuur van het project voorziet in een stuurgroep en begeleidingsgroep. De stuurgroep bestaat uit de volgende personen die er zitting in hebben vanuit verschillende sectoren:

- Henny Kempen (Provincie Gelderland)
- Janco van Gelderen (Provincie Utrecht)
- Jan Meijles (Provincie Zuid-Holland)
- Jean Hacking (Provincie Limburg)
- Jolanda Bauwens, Marja Segers (Provincie Brabant)
- Lester Reiniers, Nanko de Boorder (Provincie Noord-Holland)
- Erik Simmelink (TNO / BRO)
- Gert van den Houten (Waterschap Rijn en IJssel)
- Ate Oosterhof (Waterleidingbedrijf Vitens)
- Cor Beets (Staatsbosbeheer)

Het project is daarnaast begeleid door het voltallige platform meetnetbeheerders, onderdeel kwantiteit, van de provincies (zie bijlage C). De uitvoering berust bij KWR en TNO gezamenlijk, in de personen van Jos von Asmuth (KWR) en Frans van Geer (TNO). De inhoud van dit rapport en het protocol zullen bovendien teruggekoppeld worden met de verschillende waterpartners, tijdens een nog te organiseren bijeenkomst.

1.4 Leeswijzer

De opzet van dit rapport is als volgt. Allereerst wordt ingegaan op bestaande richtlijnen en methoden m.b.t. kwaliteitsborging, welke we daarvan hier volgen en welke niet. Daarna wordt het onderwerp en de reikwijdte van het protocol nader uitgewerkt en toegelicht, omdat het bijvoorbeeld veel uitmaakt of je de kwaliteit van historische meetreeksen of nieuw ingewonnen gegevens controleert. Vervolgens wordt de hier gebruikte systematiek en terminologie uitgewerkt en toegelicht, waarna op elk van de vier basisonderdelen ervan (data-integriteit, consistentie, plausibiliteit en controle van sensordata) in meer detail wordt ingegaan in afzonderlijke hoofdstukken. Op de onderdelen data-integriteit en consistentie zijn daarbij meerdere kleinere datacontroles onderscheiden (die verregaand geautomatiseerd kunnen worden) waarvan overzichten zijn te vinden in bijlagen A en B. In hoofdstuk 9 wordt vervolgens de implementatie van het protocol en bijbehorende controles besproken. De hoofdstukken 1 t/m 9 bevatten dus kortgezegd een beschrijving van de achtergronden van het protocol en de systematiek en verschillende onderdelen ervan. Ze beschrijven daarmee ook het proces dat doorlopen is, en onderbouwen de gezamenlijk gemaakte keuzes.

Het eigenlijke protocol, tenslotte, zal tegelijkertijd als apart rapport verschijnen, en bevat de eigenlijke, in de praktijk te hanteren werkwijzen, controles en criteria.

2 Bestaande methoden en richtlijnen

2.1 Inleiding

Aandacht voor de kwaliteit van de (grond)watergegevens, op de schaal waarop daar inmiddels aandacht voor is, is een relatief recent fenomeen (alhoewel TNO eertijds eigen kwaliteitsprocedures had voor de controle van handmatige invoer in DINO). De aandacht voor kwaliteit (en problemen daarin) is toegenomen met het toegenomen gebruik van de gegevens, maar ook met de toegenomen gebruiks-, visualisatie- en analysemogelijkheden van de gegevens.

Daarnaast speelt een rol dat het meetproces zelf in toenemende mate niet meer handmatig, maar geautomatiseerd met sensoren en dataloggers wordt verricht. Dit zorgt er aan de ene kant voor dat er minder tijd in het meten zelf gaat zitten, maar het verwerkingsproces van de gegevens is aan de andere kant wel complexer geworden. Sensoren vertonen daarnaast niet alleen andersoortige maar ook complexere en deels systematische fouten en afwijkingen (McLaughlin en Cohen, 2011; Von Asmuth, 2011; Post en Von Asmuth, 2013). De hoeveelheid data is door sensorisering bovendien flink toegenomen, waardoor handmatige controle steeds minder goed uitvoerbaar is.



Figuur 1: Voorbeelden van wetenschappelijke boeken en tijdschriften rond het onderwerp kwaliteitsborging en -controle.

Ter onderbouwing van de keuzes die in het protocol gemaakt dienen te worden, en om te profiteren van de kennis en ervaringen die buiten de watersector zijn opgedaan, wordt allereerst ingegaan op kwaliteitsborging van producten in het algemeen. Bij industriële productieprocessen is kwaliteitscontrole dusdanig algemeen en belangrijk dat er een eigen wetenschappelijke discipline voor is ontstaan, met bijbehorende boeken en tijdschriften (Figuur 1). Paragraaf 2.3 behandelt vervolgens de reeds beschikbare richtlijnen en procedures die betrekking op de kwaliteitsborging van (grond)watergegevens en de systematiek en terminologie die daar gehanteerd wordt.

2.2 Kwaliteit van producten in het algemeen

2.2.1 Inleiding

Buiten de wereld van de (grondwater)monitoring is aandacht voor kwaliteit en kwaliteitsborging al langer vanzelfsprekend. De kwaliteit van een product is wellicht hét middel voor fabrikanten en marketeers om zich te onderscheiden van de concurrentie. Een

gegarandeerde en vaste kwaliteit is daarbij een groot goed, en een mogelijke aanleiding om een serie producten terug te roepen uit de winkel en te vernietigen. Om te profiteren van de kennis en ervaring die daar is opgedaan gaan we allereerst in op de kwaliteitsborging van producten en productieprocessen in het algemeen. Zo bezien zijn grondwatergegevens eigenlijk ook gewoon producten uit de monitoringsfabriek van de provincies en haar waterpartners.....

2.2.2 Quality Assurance (QA) en Quality Control (QC)

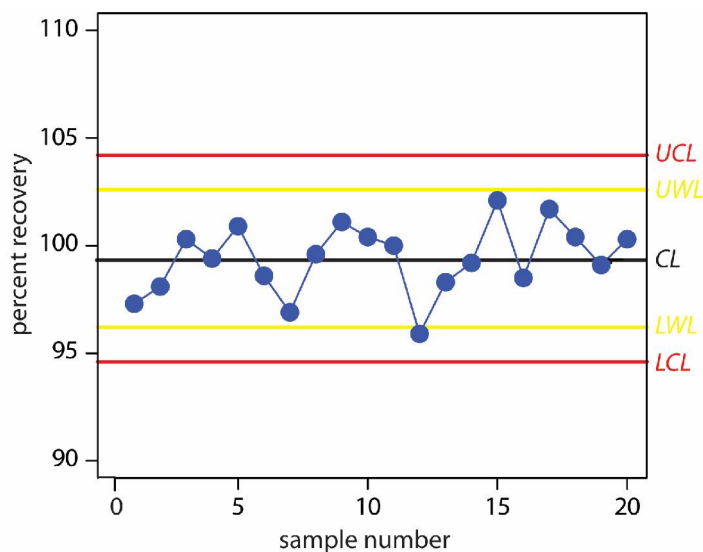
Het voert logischerwijze te ver om in het kader van dit project de beschikbare kennis over kwaliteitssystemen en kwaliteitsborging en -controle door te nemen en samen te vatten. Naast de constatering dat kwaliteitscontrole meer gemeengoed is buiten de waterwereld dan daarbinnen, is hier ook van belang dat er in het algemeen een strikt onderscheid wordt gemaakt tussen *Quality Assurance* (QA) aan de ene kant en *Quality Control* (QC) aan de andere. *Quality Assurance* handelt daarbij over alles wat met preventie te maken heeft, vertaald in termen van (grond)watermonitoring bijvoorbeeld het opstellen van handboeken, normen en richtlijnen over de inrichting van de meetopstelling, het te gebruiken meetinstrumentarium en het meet- en verwerkingsproces. Het handboek dat de Stowa op heeft laten stellen over het meten van grondwaterstanden in peilbuizen (Bouma e.a., 2012) is hier een voorbeeld van. *Quality Assurance* valt als zodanig buiten de scope van dit project. In hoofdstuk 9 is de aanbeveling opgenomen om ook op dit gebied een richtlijn vast te stellen. Dit protocol gaat dus over het *Quality Control* deel, waarbij het in de praktijk toetsen en detecteren van 'producten' of metingen die niet voldoen aan de eerder gespecificeerde kwaliteitseisen centraal staat.



Figuur 2: Illustratie van het verschil tussen, en samengaan van, Quality Assurance (preventie) en Quality Control (detectie) (bron: www.dialog.com.au)

2.2.3 Statistical Quality Control en Quality Control Charts

Naast het hierboven beschreven onderscheid tussen *Quality Assurance* en *Quality Control* is van belang dat statistiek niet alleen bij grondwater (waar we eerder de term 'plausibiliteit' gebruikt hebben) maar ook in het algemeen een belangrijke rol speelt bij kwaliteitscontrole. Voor het controleren van producten of productieprocessen is het immers het meten van bepaalde eigenschappen daarvan nodig, die een bepaalde variatie zullen vertonen omdat de werkelijkheid nu eenmaal niet exact of volledig te sturen is. Toetsing van de vraag of de (gemeten) eigenschappen van



Figuur 3: Voorbeeld van een Quality Control Chart met Upper en Lower Warning en Control Limits (bron: chemwiki.ucdavis.edu)

een product voldoen aan ofwel vooraf gedefinieerde kwaliteitseisen, ofwel te sterk afwijken van de gemiddelde of doeleigenschappen, kan in principe volledig automatisch gebeuren. Bij kwaliteitscontrole wordt echter veelvuldig gebruikt gemaakt van zogenaamde *Quality Control Charts* waarin het verloop van het meetproces gevisualiseerd wordt en dus ook handmatig en visueel beoordeeld kan worden (Figuur 3). Kennelijk is ook buiten de hydrologie de stand van zaken dat visuele en expertoordelen aanvullende waarde hebben bovenop automatische beoordeling en toetsing.

2.3 Kwaliteit van (grond)waterdata in het bijzonder

2.3.1 Grondwaterkwaliteitsdata (TNO / DINO)

Richtlijnen voor de kwaliteitsborging van (grond)waterdata zijn door verschillende instanties al eerder en meer of minder uitgebreid en direct ter hand genomen en gedocumenteerd. We lichten de beschikbare richtlijnen en procedures hieronder toe, met de daar gehanteerde terminologie en systematiek en de overeenkomsten en verschillen daartussen. Allereerst is er vanuit de DINO-groep van TNO een methodiek ontwikkeld en procedure opgesteld voor de kwaliteitsborging van grondwaterkwaliteitsgegevens (Van der Meij en Van der Meer, 2010). In deze methodiek is de kwaliteitscontrole of Quality Control (QC) gesplitst in een drietal stappen, die opeenvolgend genummerd zijn als QC1, QC2, en QC3. De eerste twee controlestappen hebben een administratief karakter (QC1 en 2), de derde is meer chemische georiënteerd (QC3). Van der Meij & Van der Meer (2010) omschrijven ze als volgt:

1. QC1: controleren van het format van de aangeleverde gegevens;
2. QC2: controleren van de volledigheid van informatie van de aangeboden gegevens;
3. QC3: controleren van de 'plausibiliteit' van de aangeleverde gegevens en het toekennen van een kwaliteitslabel.

2.3.2 Grondwaterkwaliteitsdata (RIVM)

Vanuit het RIVM is in het kader van het project 'Van Peilbuis tot KRW portal' een alternatieve methodiek ontwikkeld en procedure opgesteld voor het borgen van de kwaliteit van grondwaterkwaliteitsgegevens. Deze procedure is beschikbaar als losse 'conceptnotitie', en is niet in definitieve vorm verschenen (Van Vliet, niet gepubliceerd). De kwaliteitscontrole wordt ook hier onderverdeeld in een aantal genummerde stappen, die qua afbakening en omschrijving echter verschillen van (Van der Meij en Van der Meer, 2010). Van Vliet omschrijft ze als volgt:

1. QC0: eisen aan veldwerk, monsterbehandeling en analyse (de laatste conform iso 17025)
2. QC1: controleren van het dataformaat ('formaat check')
3. QC2: controleren van de plausibiliteit ('plausibiliteitscheck')
4. QC3: 'reeks analyse'

In de latere omschrijving van onderdeel QC2 of de plausibiliteitscheck staat dat in die stap 'niet gekeken wordt naar de waarschijnlijkheid van de meetwaarde zelf'. Er wordt in stap QC2 wel gekeken naar gegevens die elkaar tegenspreken en die dus met zekerheid fout zijn. De term 'plausibiliteitscheck' lijkt dus de lading niet te dekken, en kan verwarring opleveren.

2.3.3 Oppervlaktewaterdata (Rijkswaterstaat)

Rijkswaterstaat heeft, als verantwoordelijke partij voor het monitoren van de Rijkswateren ofwel grotere oppervlaktewaterlichamen, richtlijnen opgesteld voor de monitoring daarvan die eveneens voortvloeien uit de verplichtingen vanuit de Kaderrichtlijn Water (Van Herpen e.a., 2009; Faber e.a., 2011). Anders dan in de vorige paragrafen beschreven richtlijnen richt Rijkswaterstaat zich niet zozeer op het bewaken van het meet- en dataverwerkingsproces, maar vooral op het gebruik van de gegevens vanuit de KRW-doelstelling: het bepalen van de huidige toestand van waterlichamen en beoordelen van lange termijn trends daarbij. De

toestand en trendparameters worden daarbij getoetst aan normen die zijn vastgelegd in officiële regelingen zoals het BKMW (Besluit kwaliteitseisen en monitoring water, 2009) of de KRW-doelendatabase (KRW-portaal voorheen kwrdoelen.nl). Voor het uitvoeren van de toetsing bestaan toetsinstrumentaria zoals iBever, QBWAT en de KRW-integratiemodule (Aquokit).

De kwaliteitsborging van meetgegevens komt als zodanig niet expliciet aan de orde in de richtlijnen van Rijkswaterstaat. Kwaliteit valt hier onder de noemer 'Statistische berekening precisie en betrouwbaarheid' waarmee de precisie en betrouwbaarheid van de classificatie of eindoordeel en toetsing worden bedoeld. Het onderwerp plausibiliteit komt terug in deze richtlijnen als het detecteren van uitbijters, maar dan om het effect van de uitbijters te bepalen op de uiteindelijke toestandsbeoordeling en toetsing.

2.3.4 Grond- en oppervlaktewaterdata (EU / Kaderrichtlijn Water)

In (Von Asmuth en Van Geer, 2013) zijn de richtlijnen en daarbij gebruikte termen in Guidance Documents 7 en 15 van de Kaderrichtlijn Water (European Communities, 2003; European Communities, 2007) al uitgebreider aan de orde geweest. We herhalen de belangrijkste punten hier nog eens kort, omdat de KRW-verplichtingen leidend zijn voor dit protocol. De Guidance Documents schrijven niet zozeer voor hoe maar dat kwaliteitsborging uitgevoerd dient te worden, en bevatten als zodanig zelf geen concrete procedures of protocollen. Er worden wel een aantal (ambitieuze) randvoorwaarden gesteld en resultaten opgelegd met betrekking tot de kwaliteit en kwaliteitsborging, die hebben geleid tot het opstellen van dit protocol:

- Er dient een kwaliteitsborgingsprocedure gebruikt te worden om fouten in de monitoringsgegevens te beperken
- Het hele monitoringsproces dient volgens (inter)nationale standaarden verricht te worden, om een vergelijkbare kwaliteit en vergelijkbaarheid te garanderen
- Controle op de consistentie van data is verplicht, en de regels die gehanteerd worden bij het beheer en de interpretatie van gegevens dienen helder gedocumenteerd te zijn.
- Monitoringsgegevens dienen gerapporteerd te worden met een bijbehorende schatting van de grootte en waarschijnlijkheid van de fouten en onzekerheden die ze bevatten.

2.4 Gerelateerde initiatieven en onderwerpen

Uit bovenbeschreven inventarisatie blijkt dat er op dit moment geen geschikte concrete, formele richtlijnen en methoden zijn die voldoende toegespitst zijn op kwantitatieve hydrologische monitoring, terwijl de bestaande systematiek voor (grond)waterkwaliteitsgegevens onvoldoende eenduidig is en verwarring kan opleveren (zie voor nadere onderbouwing ook hoofdstuk 4). Ook buiten de directe opdrachtgevers en het directe kader van dit project wordt deze behoefte duidelijk gevoeld (zie bijv. ook Sorensen en Butcher, 2011; Bouma e.a., 2012; Broekhuis e.a., 2014) en zijn verschillende initiatieven op dit terrein genomen. Datakwaliteit en datakwaliteitscontrole heeft daarnaast een niet te onderschatten relatie met de in den lande gebruikte dataverwerking-, databeheer- en data-analysesystemen, en sensoren en dataloggers die de ruwe meetgegevens inwinnen en vastleggen. Al met al zijn de belangrijkste projecten, initiatieven en onderwerpen die een directe relatie hebben met datakwaliteit de volgende:

- **Komst en inrichting van de BRO** – Vanwege zowel het centrale als verplichtende karakter van de BRO en de verregaande automatisering van de data-uitwisseling daarmee, heeft de lopende realisatie van de BRO een grote

impact op alles wat met ondergronddata te maken heeft. Alhoewel veel nog uitgewerkt moet worden zal de BRO waarschijnlijk 1) eigen eisen stellen aan de interne data-integriteit van de BRO, 2) via synchronisatie uitwisselfouten en inconsistenties tussen databases voorkómen en 3) aanpassingen vergen aan lokale databeheersystemen, waar aanpassingen t.b.v. dit protocol op mee kunnen liften. In paragraaf 3.2.1 wordt in meer detail ingegaan op de data-uitwisseling met de BRO, en de relatie van dit protocol daarmee.

- **Gebruik van lokale databeheersystemen** – verschillende bronhouders zijn bezig met de overstap op, inrichting en/of verbetering van lokale beheersystemen voor hydrologische data als WISKI, DAWACO, FEWS en Menyanthes.
- **Software voor datakwaliteitscontrole** – naast al langer bestaande systemen die *outlier*-detectiefunctie bevatten zoals de Dataprofeet en FEWS, zijn er inmiddels ook verschillende systemen die zich (mede) richten op (druk)sensordata, zoals Menyanthes, ArtDiver en de screeningtool van Eijkelkamp. De verschillende softwaresystemen hanteren logischerwijze elk (eigen) interne methoden en criteria bij de controle.
- **Hercontrole en herziening van data uit het verleden** – alhoewel er in het kader van dit project geen overzicht van gemaakt kon worden, is wel duidelijk dat verschillende bronhouders bezig zijn om hun ‘historische’ of bestaande data, waarvan de kwaliteit in het verleden minder stringent bewaakt is, opnieuw te controleren en te herzien. Dit speelt in verkennende zin bij de provincies (Leunk, 2014), in uitvoerende zin bij Vitens en anderen (zie bijv. Leunk & Von Asmuth, 2013; Von Asmuth & Leunk, 2014).
- **Kwaliteit van meetopstelling, meetproces en meetinstrumenten** – naast het controleren van de kwaliteit van data, is van belang deze te kennen en/of bij de bron (meetopstelling, sensoren) te verbeteren, waarvoor verschillende initiatieven hebben gelopen of nog lopen. In dit kader is het opstellen van richtlijnen relevant, waartoe de STOWA is overgegaan (Bouma e.a., 2012) en de provincies nog zullen doen, maar ook onderzoek naar het gedrag van fouten en afwijkingen (Von Asmuth, 2011, De Meij & Von Asmuth, 2011, Knotters e.a., 2013) en verbeteringen die fabrikanten doorvoeren aan hun sensoren.

Tijdens dit project heeft overleg plaatsgevonden met de BRO (Erik Simmelink en Henco Kuiphof) over afstemming van de consequenties en het ‘datamodel’ van dit protocol en dat van de BRO. Vlak voor het verschijnen van dit rapport is het data- of domeinmodel van de BRO van onderdeel ‘Grondwatermonitoringput’ gereed gekomen. In de data-integriteits- en consistentiecontroles in bijlagen A en B is een verwijzing opgenomen naar de gebruikte term in het BRO-domeinmodel, voor de overige onderdelen (die betrekking hebben op de metingen en het meetproces) dient dat te gebeuren na het gereedkomen daarvan. Andersom is het attribuut ‘drukdot’ (aanwezig bij monitoring van artesisch grondwater) uit dit protocol overgenomen in het domeinmodel van de BRO. Een één-op-één vertaling van gebruikte attributen bleek niet mogelijk, bijvoorbeeld omdat het domeinmodel van de BRO anders is dan dat van DINO en beide verschillend omgaan met de historie van gegevens.

Er is daarnaast overleg geweest met Eijkelkamp (John de Geest, Leon van Hamersveld) rond hun initiatief vanuit het Nationaal Watertraineeship (Broekhuis e.a., 2014) en over afstemming met en gebruik van de methoden die de door hen en Alterra (Martin Knotters, Dennis Walvoort) ontwikkelde screeningtool gebruikt. Uit (Broekhuis e.a., 2014) komt naar voren dat niet alleen bij de provincies maar ook bij de andere waterpartners een sterke behoefte bestaat aan methoden en richtlijnen voor datakwaliteitscontrole, en dat men

daarom uit ziet naar de inhoud van dit protocol. De inhoudelijke conclusies omtrent de methodiek van *outlier*-detectie zijn verwerkt in hoofdstuk 7, en vergen deels nog verder onderzoek. Beide partijen beamen daarnaast dat een cursus rond het thema datakwaliteitscontrole aan te bevelen is, om de implementatie en verspreiding van de ontwikkelde en beschikbare kennis te bevorderen, en stellen hun eigen expertise en ervaring daarvoor graag ter beschikking.

3 Onderwerp en afbakening

3.1 Inleiding

In de hierna volgende hoofdstukken worden de verschillende onderdelen van kwaliteitscontrole nader uitgewerkt en besproken. Het doel van de tekst is en was mede het bieden van een raamwerk voor en faciliteren van de besluitvorming over welke toetsen op welke manier in het QC-protocol opgenomen en uitgevoerd dienen te worden. De eerste stap in het opstellen van een dergelijk raamwerk is het bepalen en afbakenen van het onderwerp en de reikwijdte ervan. Afbakening is eveneens nodig om te komen tot een zinvol, helder en concreet toepasbaar protocol. In ons geval betekent dit dat een antwoord op de volgende vragen geformuleerd moet worden:

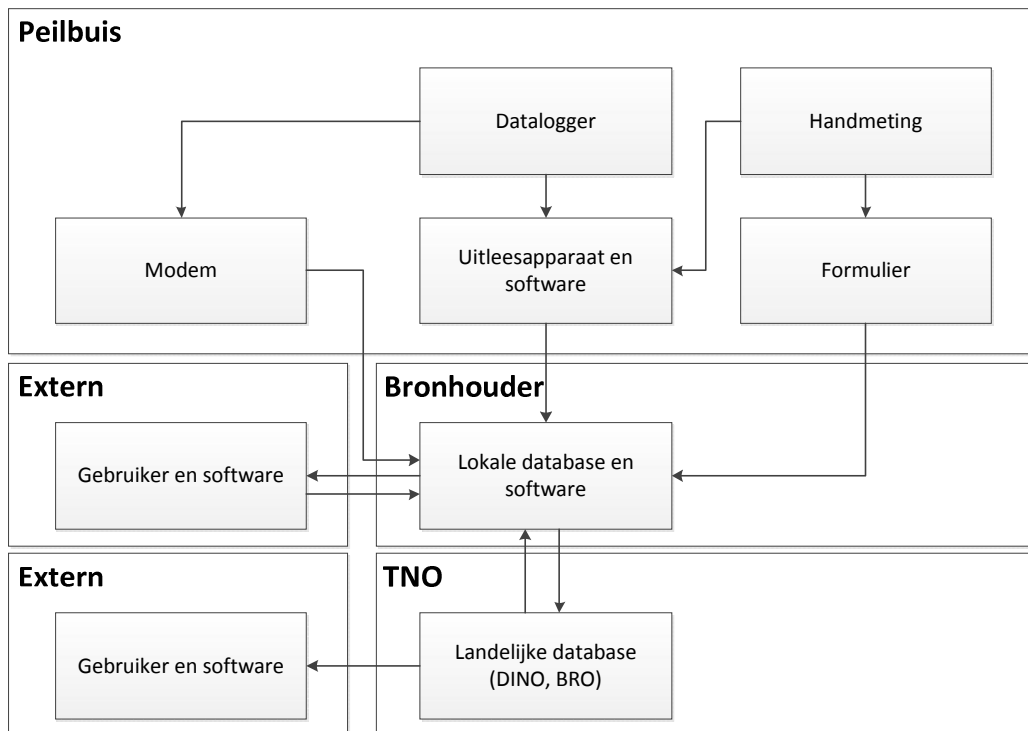
- Kwaliteitscontrole:
 - Op welke stappen in de datastroom richt het protocol zich? Op alle stappen, of op een specifiek systeem en aanleverproces (bijv. DINO en de BRO)?
 - Beperkt het protocol zich tot nieuw ingewonnen data, of vallen ook de data en situatie uit het verleden daaronder?
 - En hoe definiëren we datakwaliteit precies?
- Data:
 - Op welke gegevens heeft het protocol wel, en op welke geen betrekking?

De bovenstaande vragen en afbakening zullen uitgewerkt en toegelicht worden in de volgende paragrafen.

3.2 Kwaliteitscontrole: wanneer en waarop?

3.2.1 Data-uitwisseling van Bron tot BRO, en verder

Zoals in de vorige paragraaf gesteld is, is het de vraag of dit protocol alleen in dient te gaan op de kwaliteitscontrole van data op het moment dat ze aangeleverd worden aan de landelijke database DINO of de BRO. Dit is wel het uitgangspunt geweest in 'Kwaliteitsborging van meetgegevens van het grondwater in DINO' (Van der Meij en Van der Meer, 2010), zoals de titel al aangeeft. Op het moment van schrijven is de BRO echter nog in wording, zodat niet vaststaat welke gegevens daar wel en niet in zullen worden opgenomen. Een concept-datamodel voor grondwaterstanden en peilbuizen is eerder besproken door een expertgroep die daarvoor door TNO in het leven is geroepen, maar zal naar verwachting dit jaar nog niet definitief worden. Zowel door bronhouders als door externe partijen worden naast de BRO bovendien verschillende andere computersystemen en -programma's gebruikt, waar gerelateerde en aanpalende informatie in opgeslagen is (bijv. informatie over beheer en onderhoud, gebruik, kwaliteitscontrole, analyse, modellering, beïnvloeding, interpretatie, etc.). De BRO is daarnaast het eindstation van de gegevens, terwijl fouten en problemen bij elke (uitwisselings)stap in het monitoringsproces kunnen ontstaan (zie Figuur 4). De BRO, tenslotte, is uitsluitend bedoeld voor data die door of in naam van overheidsorganisaties is ingewonnen.



Figuur 4: Stroomschema van grondwaterstandsgegevens vanaf de peilbuis in het veld, via het kantoor van de bronhouder naar de landelijke database bij TNO. Gebruikers betrekken hun gegevens zowel van bronhouders als van DINO en de BRO.

Om bovenstaande redenen kiezen we als onderwerp van het protocol niet alleen voor de data in of data-aanlevering aan DINO of de BRO. Het protocol geldt voor de hele datastroom tijdens het monitoringsproces, en alle databestanden en databases die er bij betrokken zijn.

3.2.2 Data en dataverkeer in verleden, heden en toekomst

De manier waarop grondwatergegevens verzameld, verwerkt en opgeslagen worden was en is sterk in ontwikkeling. Voor de kwaliteit en dus de kwaliteitscontrole daarvan heeft dit aanzienlijke gevolgen:

- **Verleden** – in het verleden was handmatige meting, verwerking en invoer de manier om grondwatergegevens te verzamelen. Omdat vergissen menselijk is en de mogelijkheden voor het hanteren van gedetailleerde, systematische procedures en exacte definities daarbij beperkt zijn, zijn dergelijke gegevens doorspekt met administratieve en communicatieve fouten (zie bijv. Knotters e.a., 2008; Jansen e.a., 2013).
- **Heden** – op het moment van schrijven bestaan handmatige en geautomatiseerde meting, verwerking en communicatie naast elkaar (waarbij geautomatiseerde meting inmiddels de overhand heeft). Er worden bovendien verschillende computersystemen, software en databases voor, naast en na elkaar gehanteerd. Onvoldoende aandacht voor de consequenties van invoering van sensoren voor de monitoring heeft datakwaliteitsproblemen en een achterstand in de verwerking daarvan opgeleverd (zie bijv. Von Asmuth, 2011; Leunk en von Asmuth, 2013; Leunk, 2014).
- **Toekomst** – het lijkt het logisch (en dit protocol dient daaraan bij te dragen) dat de datastromen in de toekomst sterker geüniformeerd, gespecificeerd en gecontroleerd zullen zijn. De huidige problemen en activiteiten die zich voordoen op het gebied van data-integriteit, data-uitwisseling, datamodel- en

dataformatdefinities zullen daarbij waarschijnlijk grotendeels opgelost en naar de achtergrond verdrongen zijn. Een belangrijke vraag is echter of er in de toekomst één centraal (BRO) systeem zal of dient te zijn, of dat de toekomst ligt in een netwerk van onderling met elkaar communicerende systemen. De verantwoordelijkheid voor de data en datakwaliteit, en ook het uitvoeren van de kwaliteitscontrole, zal naar alle waarschijnlijkheid (grotendeels) decentraal en bij de bronhouder blijven liggen.

3.3 Data: wat wel en wat niet?

3.3.1 Kwaliteitsborging en interpretatie als doelstelling

In eerste instantie zouden we onder 'de' te controleren data alle informatie kunnen scharen die ontstaat tijdens of betrekking heeft op het monitoringsproces, voor zover deze relevant en in praktisch te hanteren vorm opgeslagen is. De relevantie van informatie wordt daarbij ingegeven door het doel. (Meta)gegevens kunnen relevant zijn voor:

1. Administratief en organisatorisch kader van het grondwatermeetnet
2. Beheer en onderhoud van het meetnet
3. Het inwinnen van de meetgegevens
4. Beoordeling van de kwaliteit van de gegevens
5. Gebruik en hydrologische interpretatie van de gegevens

Vanuit de doelstelling van dit protocol kunnen we ons tot de kwaliteit van grondwaterstands- en stijghoogtemetingen, ofwel de omkaderde doelstellingen 4) en 5) beperken. Dat wil overigens niet zeggen dat gegevens die primair vanuit een andere doelstelling verzameld zijn, bijvoorbeeld voor beheer en onderhoud, niet ook relevant zijn voor kwaliteitscontrole en hydrologisch gebruik. In iets ruimere zin is een goede datakwaliteit nodig voor het gebruik en de interpretatie van de gegevens, en bovendien gaat het beoordelen van afwijkende patronen naadloos over in het interpreteren daarvan. De constatering dat er meer doelstellingen zijn en dus ook meer relevante gegevens dan hier behandeld worden, is overigens wel relevant, bijvoorbeeld vanuit het oogpunt van data-uitwisseling.

3.3.2 Transparantie en reproduceerbaarheid als uitgangspunten

Naast het formuleren van doelstellingen zijn er ook middelen nodig om die te bereiken. Hoe beoordelen we de kwaliteit van grondwaterstandsreeksen, en wat is of welke gegevens zijn daarvoor nodig? Het thema datakwaliteit gaat over fouten en afwijkingen, en fouten ontstaan doorgaans niet spontaan maar tijdens een (meet- of verwerkings)proces. Om te kunnen beoordelen of een gegeven fout is of niet, is het van belang dat het meet- en verwerkingsproces a) transparant en b) reproduceerbaar is. Transparantie houdt in dat helder en overzichtelijk is wat de meetsituatie en het meet- en verwerkingsproces precies is geweest. Transparantie heeft daarmee ook met visualisatie te maken, mede waarom ook bij de kwaliteitscontrole van productieprocessen (zie paragraaf 2.2) zogenaamde 'Quality Control Charts' een belangrijke hulpmiddel blijken te zijn. Reproduceerbaarheid sluit daarbij aan en betekent dat niet alleen duidelijk is wat het verwerkingsproces precies is geweest, maar dat subjectieve onderdelen daarvan indien nodig anders beoordeeld en gewijzigd kunnen worden.

3.3.3 Classificatie van 'objecten' of gegevenstypen in het protocol

Vanuit bovenstaande doelstellingen en uitgangspunten kunnen we het onderwerp van dit protocol afbakenen. Het is daarnaast belangrijk om te constateren dat er rond de monitoring meer gegevens verzameld worden dan de grondwaterstands- en stijghoogtemeetreeksen alleen. De gegevens beschrijven in technische termen ook verschillende typen 'objecten' of 'entiteiten', ook wel aangeduid als 'klassen'. We classificeren de verschillende soorten gegevens in dit protocol dus in een aantal klassen, die in onderstaande tabel zijn

weergegeven. De kolommen 'kwaliteitscontrole' en 'interpretatie' in die tabel geven daarbij aan of de desbetreffende data van belang is voor de kwaliteitscontrole, de hydrologische interpretatie, of beide.

Tabel 1: Klassen of gegevenstypen in het protocol, met omschrijving en indicatie van de doelstelling van waaruit ze zijn opgenomen

Klasse	Omschrijving	Kwaliteitscontrole	Interpretatie
Administratie	Coderingen en/of namen die de meetlocatie uniek identificeren	Ja	Ja
Processen	Hydrologische of antropogene processen die leiden tot bijzonder gedrag	Ja	Ja
Peilbuiscluster	Informatie over de situering en afwerking van een cluster van peilbuizen	Ja	Ja
Peilbuis	Informatie over de constructie en situering van de individuele peilbuizen en hun filters	Ja	Ja
Meetinstrument	Informatie over het gebruikte meetinstrument en de situering daarvan	Ja	
Dataverwerking	Informatie over hoe en door wie de gegevens beoordeeld zijn, en wat er mee gebeurd is	Ja	
Locatiehistorie	Informatie over de geschiedenis van peilbuis en meetinstrument	Ja	Ja
Handmeting	De handmatige metingen van het waterpeil in een peilbuis en daaraan gerelateerde informatie	Ja	Ja
Loggermeting	Metingen van waterpeil of -druk met een datalogger in een peilbuis, en daaraan gerelateerde informatie	Ja	Ja
Controlemeting	De (handmatige) metingen die erop gericht zijn dataloggermetingen te controleren en daaraan gerelateerde informatie	Ja	

4 Systematiek en terminologie

4.1 Inleiding

Uit de in hoofdstuk 2 beschreven inventarisatie van beschikbare richtlijnen en methoden concluderen we het volgende:

- De richtlijnen voor kwaliteitsborging van grondwaterkwaliteitsgegevens die door TNO / DINO en het RIVM zijn opgesteld verschillen van elkaar, ook al hanteren ze grotendeels dezelfde benadering en codes voor de controlestappen (QC-1, QC-2 en QC-3). Gebruik van een dergelijke korte code kan voordelen bieden wanneer de stappen helder afgebakend en gedefinieerd zijn, maar wanneer dat niet zo is bieden ze onvoldoende houvast om verwarring en misverstanden te voorkomen. De richtlijnen van Rijkswaterstaat voor oppervlaktewatermonitoring verschillen sterk van die van TNO en het RIVM. Deze richten zich primair op de toestands- en trendbepaling voor de KRW, en dekken de kwaliteitsborging van het meet- en dataverwerkingsproces niet volledig.
- De kennis en ervaring rond kwaliteitscontrole van Industriële producten en productieprocessen in het algemeen biedt wel inzicht en houvast, maar geen geschikte concrete richtlijnen en methoden die voldoende zijn toegespitst op (grondwater)monitoring. De *Guidance Documents* van de KRW stellen wel voorwaarden aan de te gebruiken procedures en methoden, maar bevatten zelf ook geen specifieke procedure.

Om bovenstaande redenen stellen we voor om de terminologie en systematiek die gehanteerd wordt in (Von Asmuth en Van Geer, 2013) over te nemen in het op te stellen protocol, aangevuld met een tweetal aspecten die eerder niet aan de orde gesteld zijn (data-integriteit en nauwkeurigheid). De in het protocol te hanteren kwaliteitscontroleonderdelen zijn daarmee:

QC-nummer	QC-term	QC-type
1	Data-integriteit	Administratieve controle
2	Consistentie	Relationele controle
3	Plausibiliteit	Statistische controle
4	Controle en correctie van sensordata	Controlemeting
5	Nauwkeurigheid (kwantificering van)	Berekening

In de volgende paragrafen zullen we bovenstaande onderdelen kort definiëren en onderling afbakenen. Per onderdeel volgt daarna een apart hoofdstuk, waarin de uitwerking van de verschillende toetsen, stappen en gerelateerde aspecten aan bod zullen komen.

4.2 Data-integriteit (administratieve controle)

In (Von Asmuth en Van Geer, 2013) is het probleem van de kwaliteitsborging of kwaliteitscontrole opgepakt en beschreven vanaf het punt dat de te controleren gegevens geïnterpreteerd, gekoppeld en integraal zijn geïmporteerd en opgeslagen in de database en/of het systeem dat voor de kwaliteitscontrole ingezet wordt. Om algemeen toepasbaar te

zijn dient dit protocol echter bij het begin te beginnen: de databestanden die aangeleverd en/of beschikbaar zijn (net als het DINO- en RIVM-protocol). We definiëren hier het onderdeel 'data-integriteit' als de vraag of de relevante gegevens integer en integraal bewaard zijn gebleven. In ons geval zijn dat de:

- Meetreeksen (data)
- Meetopstelling en andere dataklassen in dit protocol (metadata)

De data en/of databestanden zijn in dat geval:

- Uniek, indien eenmalig
- Volledig, indien verplicht
- Eenduidig, qua:
 - Dataformat
 - Betekenis of definitie
 - Status
 - Datamodel en relaties tussen de verschillende tabellen en/of bestanden

Gegevens die een relatie met elkaar hebben en daarbij onderling in tegenspraak zijn, zou je in het dagelijks spraakgebruik ook als niet eenduidig kunnen betitelen. Op dit punt gaat data-integriteit in feite over in consistentie, maar elkaar tegensprekende gegevens scharen we hier onder de noemer consistentie. Het is daarbij natuurlijk de vraag wat wel tot 'de' gegevens behoort en wat niet (zie ook paragraaf 3.3). In de praktijk blijkt een gebrekkige 'data-integriteit' in ieder geval een belangrijk probleem voor de interpretatie en het gebruik van grondwatermeetgegevens te kunnen vormen (zie bijv. Leunk en von Asmuth, 2013).

4.3 Consistentie (relationele controle)

Bij gegevens die door een computersysteem geïnterpreteerd en geïmporteerd zijn ligt als het goed is de definitie en het zogenaamde datamodel vast. We definiëren hier het onderdeel 'consistentie' als de vraag of de relevante gegevens in dat computersysteem, niet logisch in tegenspraak met elkaar of fysisch onmogelijk zijn. Ook hier betreft dat de gegevens over of van de:

- Meetreeksen
- Meetopstelling

Inconsistenties kunnen daarbij optreden tussen de verschillende gegevens van de:

- Meetreeksen zelf
- Meetopstelling zelf
- Meetreeksen en meetopstelling onderling

De eigenlijke toetsing vindt plaats op de relatie die de gegevens met elkaar hebben. Het gaat daarbij om:

- Chronologische relaties
- Ruimtelijke relaties
- Administratieve relaties

Naast gegevens die foutief zijn of elkaar tegenspreken kunnen zich ook situaties voordoen in de meetopstelling, waarbij de eigenlijke meting juist is maar niet representatief voor het proces dat gemonitord wordt. Voorbeelden daarvan zijn metingen gedurende droogval van de sensor of ijsvorming in de peilbuis.

4.4 Plausibiliteit (statistische controle)

Bij gegevens die onderling in tegenspraak dan wel logisch of fysisch onmogelijk zijn is de conclusie relatief duidelijk: minstens (!) één van de onderling vergeleken gegevens is dan fout. Naast eenvoudigweg foute gegevens zijn er ook gegevens die meer of minder waarschijnlijk fout zijn. We definiëren hier het onderdeel 'plausibiliteit' als de vraag hoe waarschijnlijk het is dat gegevens toe te schrijven zijn aan een:

- Daadwerkelijke gebeurtenis
- Meet- of verwerkingsfout

gezien:

- Andere metingen in de tijd
- Andere metingen in de ruimte
- Metingen van andere parameters
- Systeem- en meetnetkennis

4.5 Controle en correctie van sensordata (controlemeting)

De controle en correctie van sensorgegevens is niet per definitie een losstaande stap die pas volgt na de hiervoor beschreven toetsen. Natuurlijk kunnen ook fouten en afwijkingen van sensoren leiden tot niet-consistente en/of niet-plausibele meetwaarden, zodat de in de hoofdstukken 6 en 7 behandelde toetsen ook voor sensordata van belang zijn. We bakenen dit onderdeel daarom af tot toetsing aan de hand van controlemetingen, die fouten op kunnen sporen die te klein zijn om met andere toetsen te constateren. We definiëren hier het onderdeel 'controle en correctie van sensordata' als de vraag of sensormetingen behept zijn met:

- Systematische fouten en afwijkingen
- Toevallige fouten en afwijkingen

Bij constatering daarvan volgt de vraag of, wanneer en hoe:

- Systematische afwijkingen gecorrigeerd dienen te worden
- Geconstateerde afwijkingen leiden tot het afkeuren van de metingen en/of het meetinstrument

Onder controlemeting valt hier meer dan alleen handmatige controle van het waterpeil. Controlemeting kan bestaan uit:

- Voor eigen luchtdrukmetingen: betrouwbare nabijgelegen luchtdrukmetingen, zoals beschikbaar via het KNMI (zie ook De Meij en Von Asmuth, 2011).
- Voor waterdruk- of waterpeilmetingen: handmatige controlemetingen van het waterpeil
- Voor de klok van de datalogger: controle aan de hand van de klok van het uitleesapparaat en/of een horloge

Naast controlemetingen brengt het gebruik van (druk)sensoren ook andere specifieke kwesties met zich mee, zoals droogval van de sensor en fouten bij de omrekening van druk naar stand. Toetsing op metingen op of onder het sensornulpunt is in feite consistentietoetsing, maar behandelen we net als de omrekening toch hieronder.

4.6 Nauwkeurigheid (kwantificering van)

De doelstelling van de kwaliteitscontroleonderdelen, die in de vorige paragrafen in grote lijnen beschreven zijn, is kort gezegd het optimaliseren van de kwaliteit van de (geaccepteerde of als voldoende betrouwbaar gelabelde) meetgegevens. Wat de kwaliteit of

nauwkeurigheid van de definitieve gegevens precies is, is daarmee nog niet geïnventariseerd of gekwantificeerd, terwijl dat bijvoorbeeld wel een vereiste is van de Kaderrichtlijn Water (European Communities, 2003; European Communities, 2007).

Een goede en goed onderbouwde kwantificering van de nauwkeurigheid van de definitieve gegevens is geen eenvoudige opgave. Conform de projectmatige afspraken valt dit onderdeel buitende scope van het onderhavige onderzoek en rapport. Daarbuiten vindt er weliswaar weldegelijk onderzoek plaats naar nauwkeurigheidskwantificering. We verwijzen hier allereerst naar een onderzoek waarin de nauwkeurigheid van handmatige metingen onder de loep genomen is (Knotters e.a., 2013). Er loopt inmiddels een tweede project in opdracht van de provincie Overijssel, waarin de nauwkeurigheid van de druksensordata van de provincie onderzocht en gekwantificeerd zal worden en de bijbehorende methodiek daarvoor wordt ontwikkeld.

We definiëren hier het onderdeel 'kwantificering van de nauwkeurigheid' als de vraag wat de nauwkeurigheid is van de definitieve meetgegevens, gegeven de daaraan ten grondslag liggende inwinnings-, omreken-, verwerkings- en controlestappen, instrumenten en methoden.

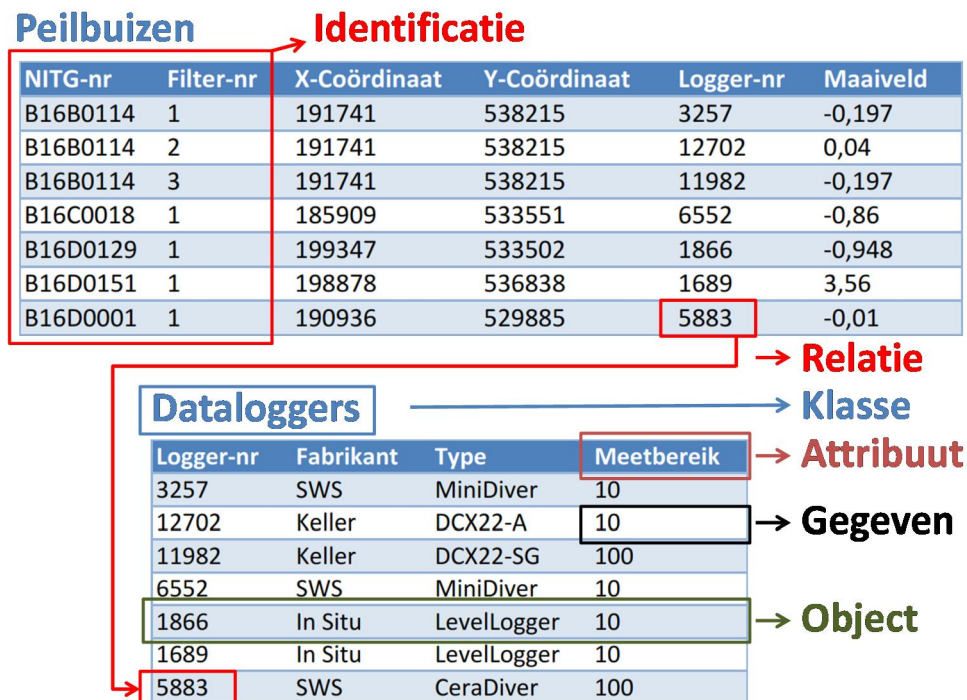
5 Data-integriteit (administratieve controle)

5.1 Inleiding

5.1.1 De tabel als basis van het data- en denkmodel

Conform de definitie in hoofdstuk 0 verstaan we onder het begrip 'data-integriteit' de vraag of de data a) uniek (eenmalige gegevens), b) volledig (verplichte gegevens) en c) eenduidig (qua definitie van format en datamodel) zijn. Alvorens op die vragen en mogelijke toetsen daarvoor in te gaan lichten we hier allereerst een aantal basisbegrippen en termen toe. De doelgroep van dit rapport en het op te stellen protocol is immers in principe breed: van veldmedewerkers en databeheerders, via hydrologen, onderzoekers en ICT-ers tot aan bestuurders en beleidsmaker. De inhoud van het rapport dient dus voor zoveel mogelijk partijen toegankelijk te zijn. Aan de andere kant dient een protocol helder te zijn en eenduidig geïnterpreteerd te worden, reden waarom we de begrippen en richtlijnen hier toch zo exact en technisch-inhoudelijk mogelijk definiëren.

We kiezen de 'tabel' met bijbehorende elementen als basis om verschillende aspecten van, en mogelijke toetsing op, data-integriteit toe te lichten. Met nadruk dient daarbij gesteld te worden dat 'tabel' hier bedoeld wordt als denkmodel: de eigenlijke gegevens worden lang niet altijd in de vorm van een tabel opgeslagen of uitgewisseld. Desalniettemin zijn elementen van de tabel doorgaans ook in andere opslagvormen te herkennen. Binnen de ICT wordt onderscheid gemaakt in een zogenaamd 'conceptueel datamodel' (definitie van objecten of entiteiten en relaties op een hoger abstractieniveau), logisch datamodel (definitie



Figuur 5: Een tabel met peilbuizen en één met dataloggers, ter illustratie van de verschillende elementen en gebruikte termen.

van tabellen, attributen en relaties daartussen) en fysiek datamodel (datadefinities in de eigenlijke database) (zie bijv. <http://nl.wikipedia.org/wiki/datamodel>). Om niet te diep op technische details in te gaan gebruiken we hier slechts het begrip tabel. We lichten de daarbij gehanteerde begrippen en termen hieronder kort toe:

- **Tabel** – daadwerkelijke tabel in een spreadsheet of (relationele) database, gegevens in één of meer afzonderlijke bestanden, of klasse in een object-georiënteerd raamwerk
- **Klasse** – of objecttype in een object-georiënteerd raamwerk, entiteittype in een conceptueel datamodel
- **Object** – rij in een spreadsheet, record in een database of object in een object-georiënteerd raamwerk, entiteit in een conceptueel datamodel
- **Attribuut** – kolomnaam of ‘header’ in een spreadsheet of tekstbestand, veld in een database, of eigenschap van een object
- **Attribuutdefinitie** – betekenis van en/of toelichting op de attributen, bijv. eenheid en referentie van hoogtegegevens.
- **Gegeven** – cel in een tabel of spreadsheet, getal of tekst in een bestand, waarde van een attribuut
- **Metadata** – informatie over een gegeven, bijvoorbeeld hoe en wanneer het is bepaald
- **Identificatie** – uniek nummer of sleutel in een database, naam van een object, of (combinatie van) attributen die een object uniek identificeren
- **Relatie** – koppeling tussen objecten of gegevens in verschillende tabellen of bestanden
- **Datamodel** – wijze waarop data in een informatiesysteem gestructureerd zijn, met definitie van relaties (conceptueel model), tabellen en attributen (logisch model) en data(base)definities (fysiek model).

5.2 Methoden en criteria

5.2.1 Uniceit (eenmalige gegevens)

Bepaalde gegevens zijn van nature uniek of eenmalig. De stijghoogte kan bijvoorbeeld op één bepaald tijdstip en plaats in de ruimte slechts één waarde hebben, een individu of individueel object is per definitie uniek en dient eenmalige en met een unieke identificatie of sleutel opgeslagen te zijn. Bij dit soort gegevens kunnen er doublures en versieconflicten optreden, en wel:

- **Conflicten binnen een dataset**
- **Conflicten tussen datasets**

Wanneer bekend is welke gegevens uniek dienen te zijn, kan eenvoudig op doublures en versieconflicten binnen een dataset getoetst worden. Conflicten tussen datasets, zoals die bijvoorbeeld tussen verschillende databases als DINO en Dawaco bestaan (Leunk en von Asmuth, 2013), zijn moeilijker te voorkomen omdat de toetsing en synchronisatie (nog) niet geautomatiseerd kunnen gebeuren. Van gegevens die uniek dienen te zijn, maar waarvan op

verschillende plaatsen verschillende versies bestaan, is de status vaak niet eenvoudig te achterhalen.

5.2.2 Volledigheid (verplichte gegevens)

Om te kunnen toetsen of een bepaalde dataset volledig is, bijvoorbeeld tijdens een importproces, dient allereerst gedefinieerd te worden bij welke gegevens dat een vereiste is. Het is, met andere woorden, de vraag welke gegevens verplicht zijn. Voor de overige toetsen is daarnaast van belang welke gegevens relevant zijn, of hoe we het onderwerp van het protocol precies afbakenen. De vraag welke dataklassen of typen gegevens we onderscheiden en welke daarmee dus relevant zijn is aan de orde gekomen in paragraaf 3.3.3.

Per dataklasse dient gekozen te worden welke gegevens of attributen daarvan verplicht zijn en welke niet. In het algemeen lijkt het logisch om die attributen verplicht te stellen die dusdanig van belang zijn dat zonder deze kwaliteitsborging of interpretatie niet goed mogelijk is. 'Belang' en 'goed' zijn echter subjectieve oordelen. Naast dit criterium is het daarom ook goed om de gevolgen van verplichtstelling in dat oordeel te betrekken. De gevolgen van verplichtstelling van historische gegevens verschilt daarbij van die van nieuw in te winnen gegevens:

- **Historische gegevens** – verplichtstelling leidt hier in principe tot het afkeuren van onvolledige gegevens. Hierbij dient weer onderscheid gemaakt te worden in:
 - **Nog bestaande peilbuizen** – waar alsnog verschillende eigenschappen van opgemeten kunnen worden
 - **Niet meer bestaande peilbuizen** - waarvan de eigenschappen voorgoed verloren zijn
- **Nieuw in te winnen gegevens** - verplichtstelling leidt hier, zeker in combinatie met het voor overheden verplichtende karakter van de BRO, tot verhoging van de kwaliteit van de toekomstige data

Op historische gegevens kan verplichtstelling van volledigheid dus een negatief effect hebben. Het afkeuren van gegevens is daarbij in feite voor historische gegevens ook onnodig. Softwarematig gezien is het probleem goed te ondervangen door gegevens niet af te keuren maar te filteren, afhankelijk van het gebruiksdoel en persoonlijke voorkeuren van de gebruiker

We stellen voor om de verplichtstelling te beperken tot nieuw in te winnen gegevens. Een overzicht van welke gegevens we verplicht en welke we relevant achten is te vinden in bijlage A.

5.2.3 Eenduidigheid (definitie van uitwisselformat en datamodel)

5.2.3.1 *Expliciete en open standaarden*

De vraag of het datamodel en de dataformats in een dataset eenduidig en expliciet gedefinieerd zijn, is iets wat de meeste gebruikers van grondwatergegevens waarschijnlijk niet als bewust en dagelijks probleem ervaren. Het is in principe ook aan de gezamenlijke software- en databaseontwikkelaars om eenduidige datamodellen en uitwisselingsformats te definiëren, en aan databeheerders om de eenduidigheid van de data te bewaken en garanderen.

Toch leiden gesloten, gebrekkige en/of wisselende datamodel- en dataformatdefinities tot problemen en ongemakken voor gebruikers in de praktijk, die wellicht door velen niet als zodanig herkend worden. Misverstanden over bijvoorbeeld de eenheid van gegevens heeft meer dan eens tot miljoenen kostende schade geleid (zie ook paragraaf 5.2.3.3). Kleiner maar telkens terugkerend ongemak betreft het haperen van inleesroutines, handmatige en gebrekkige data-uitwisseling tussen verschillende programma's. Des te meer handwerk er nodig is bij data-uitwisseling, des te meer fouten zullen er gemaakt worden.

```
PUT CONSTRUCTIE;BUIS CONSTRUCTIE;X COORDINAAT (m);Y CO
P;B;56685;426020;-30;6-5-1987;18/09/1996;-18;-8;-17;
P;B;56700;426000;-29;6-5-1987;18/09/1996;-19;-10;-17;
P;B;56700;425960;-35;6-5-1987;14/09/1995;-18;-3;-16;
P;B;56700;425960;-39;14/09/1995;;38;77;-180;-197;23;
;P;B;56740;425880;-34;N;6-5-1987;14/09/1995;-46;-12;-;
;P;B;56740;425880;-37;N;14/09/1995;;26;63;-139;-204;2;
;P;B;56740;425880;-34;N;6-5-1987;14/09/1995;-43;-9;-7;
;P;B;56740;425880;-37;N;14/09/1995;;24;61;-706;-721;7;
P;B;56785;425790;-38;6-5-1987;14/09/1995;-13;-5;-17;
P;B;56785;425790;-38;14/09/1995;;21;59;-179;-194;21;
P;B;56825;425805;-41;6-5-1987;18/09/1996;-13;-2;-17;
P;B;56900;425820;-19;6-5-1987;23/02/1993;-5;-6;-15;
P;B;56900;425820;-20;23/02/1993;;-25;-5;-15;-225;20;
P;B;56880;425820;-15;6-5-1987;23/02/1993;-9;6;-144;-;
P;B;56880;425820;-16;23/02/1993;;-21;-5;-136;-186;10;
;P;B;56700;425260;3;N;6-5-1987;14/09/1995;-3;-6;-138;-;
```

Figuur 6: Uitsnede uit een csv-bestand voor uitwisseling van stijghoogtegegevens, met een niet eenduidige en intern wisselende notatie of datumformat.

Het gebruik van open standaarden voor datamodellen en uitwisselformats is vanuit deze optiek sterk aan te bevelen, en het is uiteindelijk mede aan gebruikers en beleidsmakers om dat voor elkaar te krijgen.

5.2.3.2 Datauitwisseling en uitwisselformat

Een logisch moment om gegevens te controleren en interpreteren is bij de import of uitwisseling ervan. Data-uitwisseling kan in principe in twee verschillende basisformats:

- **Tekst** - in dat geval bestaan alle gegevens uit ASCII-of Unicode-tekenen. Een voordeel van tekst-formats is dat de gegevens in een willekeurige tekst editor geopend en 'handmatig' bekeken en geïnterpreteerd kunnen worden. Een nadeel is dat ook een inleesroutine de tekst bij import moet interpreteren en omzetten in verschillende gegevenstypen (bijv. tekst, getallen, datums, etc.), wat vraagt om de definitie van bijv. decimaaltekenen en datumformats. De grotere 'openheid' van tekstbestanden gaat gepaard met een grotere foutengevoeligheid als nadelig effect. Een intermediaire oplossing daarvoor is het gebruik van tekst-bestanden met een vast en eenduidig format zoals XML.
- **Binair** - in dat geval bestaan alle gegevens uit bits en bytes. Een voordeel van het gebruik van binaire formats is dat deze doorgaans eenduidiger gedefinieerd zijn en daarmee minder foutengevoelig. Een nadeel is dat de gegevens niet geïmporteerd kunnen worden zolang de formatdefinitie onbekend is. Dit zorgt ervoor dat softwareleveranciers de bestanden en data van hun software gesloten en voor zichzelf kunnen houden, om concurrentie tegen te gaan.

Naast verschillende uitwisselformats zijn er ook verschillende technische oplossingen om data uit te wisselen, zoals het gebruik van webservices en databasekoppelingen.

5.2.3.3 Dataopslag en datamodel

5.2.3.3.1 Attributen en relaties

Bij attribuutdefinitie gaat het net als bij metadata om informatie op een hoger niveau. Het zijn beiden gegevens die eigenschappen van andere gegevens beschrijven. We gebruiken hier de term attribuutdefinitie naast metadata, omdat die laatste term een veel ruimere betekenis heeft. Informatie op een hoger niveau vraagt in feite ook om bewustzijn en communicatie op een hoger niveau. Gegeven de waan van de dag waarin het uitwisselen van de gegevens zelf nog vaak handwerk is en lang niet altijd vlekkeloos gaat, is dat wellicht vragen om problemen (ook hier conform de wet van Murphy). Een mogelijke oorzaak van problemen met datadefinities is dat ze vaak niet expliciet gemaakt worden, en niet meegegeven worden met de eigenlijke data. In veel gevallen spreekt de definitie ook bijna (!) vanzelf: bij het attribuut 'X-Coördinaat' uit Figuur 5 zal het wel om coördinaten volgens het Rijksdriehoeksstelsel in meters gaan.

Het hierboven gebruikte woordje 'bijna' brengt echter niet onbelangrijke risico's met zich mee. Het niet expliciet aan de orde stellen van de vraag 'wat een bepaald waarde precies voorstelt' leidt te makkelijk tot misverstanden, getuige bijvoorbeeld de vele 'sprongen' in de grondwatermeetreeksen in DINO. De ervaring leert inmiddels dat fouten in datadefinities en metadata grote gevolgen kunnen hebben, niet alleen bij grondwatergegevens (zie bijv. Leunk, 2014), Von Asmuth en Van Geer, 2013) maar ook daarbuiten (zie bijv. 'Some Famous Unit Conversion Errors', beschikbaar op <http://spacemath.gsfc.nasa.gov/weekly/6Page53.pdf>).

Definitie van de relaties tussen verschillende databestanden of tabellen is vooral van belang wanneer de gegevens niet of niet integraal zijn ondergebracht in een adequaat databasesysteem waarin het datamodel en relaties vastgelegd en liefst open en expliciet gedefinieerd zijn. In dergelijke gevallen is het van belang dat niet alleen de verschillende gegevens en bestanden zelf bewaard blijven, maar ook de relaties daartussen (zodat bijvoorbeeld te traceren blijft welke dataloggerbestanden bij welke peilbuis horen).

5.2.3.3.2 Metadata, oorsprong, betekenis, interpretatie en status van data

Naast de engere en meer technisch gedefinieerde term 'attribuutdefinitie' is het van belang om binnen het datamodel voldoende informatie of metadata op te nemen om de oorsprong en status van verschillende (basis)gegevens vast te leggen en expliciet te maken. Deze metadata is logischerwijze zelf ook vastgelegd in tabellen met attributen (en bijbehorende definitie). Het beschrijven van de oorsprong van gegevens is bijvoorbeeld van belang wanneer de statistische eigenschappen daarvan (bijv. de nauwkeurigheid) sterk variabel zijn. Een goed voorbeeld daarvan zijn coördinaten, die soms ruwweg van een kaart afgelezen zijn, maar soms ook nauwkeurig ingemeten zijn. Het is in dit voorbeeld ook van belang om de status van de gegevens vast te leggen, omdat nieuwe coördinaten eenvoudigweg nauwkeuriger kunnen zijn dan de oude, en die in principe dus kunnen vervangen. Nieuwe coördinaten kunnen echter ook te maken hebben met een daadwerkelijke verplaatsing van de peilbuis, wat een heel andere betekenis heeft.

Voorbeelden van het vastleggen en expliciet maken van interpretatie en status zijn het opnemen van opmerkingenvelden of kwaliteitslabels bij metingen. Een lastiger op te lossen probleem met betrekking tot de status van gegevens zijn de eerder beschreven versieconflicten die binnen en vooral tussen verschillende datasets kunnen bestaan. Dat vraagt om eenduidige keuzes op welke plaats mutaties in de data gedaan worden, welke dataset dus de leidende en enige juiste is, en de organisatie en bewaking daarvan (zolang de databasesystemen niet onderling gekoppeld zijn en geautomatiseerd gesynchroniseerd worden).

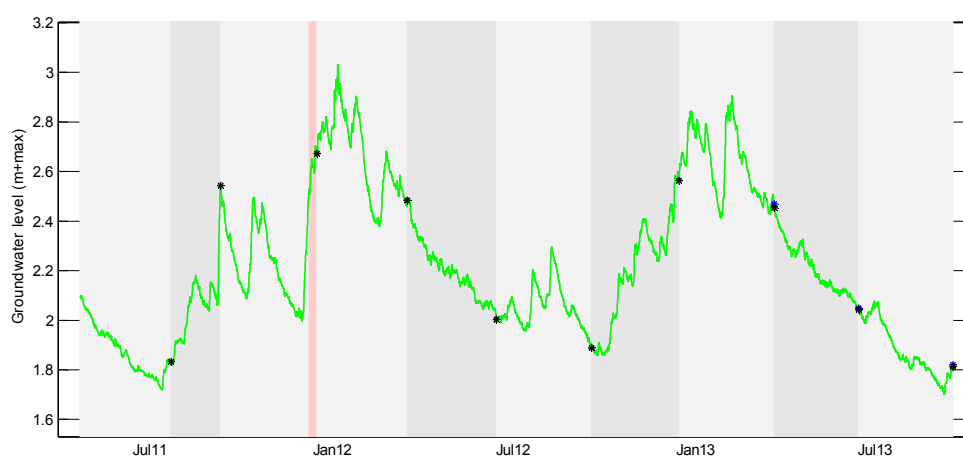
5.3 Geautomatiseerde controle

Data-integriteit is in sterke mate een technisch probleem, en vraagt daarom ook vooral om technische en geautomatiseerde oplossingen. De hierboven beschreven punten betekenen dus vooral werk voor de verzamelde informatietechnici die zich met software voor verwerking en beheer van (grond)waterdata bezig houden. Over de technische achtergronden zouden databeheerders, hydrologen en eindgebruikers zich niet al te veel zorgen moeten maken, maar het is anderzijds wel goed dat zij zich bewust zijn van de problematiek omdat ze er wel op andere manieren mee te maken of een rol in hebben:

- **Sturing van softwareontwikkeling** - Dit gaat niet vanzelf. Individuele softwareleveranciers hebben niet perse baat bij openheid en een goede koppeling van systemen, omdat gesloten systemen ook een gesloten markt met zich meebrengen. In open concurrentie en data-uitwisseling zijn voor eindgebruikers echter de efficiëntste en kwalitatief beste oplossingen te vinden.
- **Eenduidigheid van het werkproces** - Zolang dat niet via automatisering afgedwongen of opgelost is, kunnen databeheerders zelf zorgen voor een eenduidig werkproces en daarmee eenduidige data, bijvoorbeeld door data niet in meerdere systemen tegelijk te muteren maar te kiezen voor één bron-database (al dan niet lokaal) van waaruit de data verspreid worden.
- **Toename van de hoeveelheid metadata** - eenduidigheid vraagt om een grotere aanlevering en aanrijking van de data, waarbij bijvoorbeeld door databeheerders in meer detail aangegeven zal moeten worden hoe bepaalde data precies ingewonnen en tot stand zijn gekomen.

5.4 Visuele controle

Alhoewel in de validatiepilot (Leunk, 2014) en elders geconcludeerd is dat visualisatie en visuele toetsing net zo belangrijk is als geautomatiseerde toetsing, zijn de mogelijkheden daarvoor in het geval van data-integriteit beperkt. Visualisatie van data-integriteitsproblemen met kleuren, bijvoorbeeld bij overlappende meetreeksen of meerdere meetwaarden op eenzelfde tijdstip (Figuur 7) of kleuring van data met integriteitsproblemen in een tabel, is daarbij de belangrijkste optie.



Figuur 7: Voorbeeld van visualisatie van data-integriteitsproblemen. De grafiek toont een druksensorreeks (groen) met handmatige controlemetingen (zwarte punten) en uitleesrondes (afwisselende lichte- en middelgrijze achtergrond). Er is sprake van overlap en doublures in de meetreeks, waar voor gewaarschuwd wordt door voor die periode een rode achtergrond te gebruiken.

6 Consistentie (relationele controle)

6.1 Inleiding

6.1.1 Afbakening t.o.v. data-integriteit

Conform de definitie in hoofdstuk 0 verstaan we onder het begrip 'consistentie' de vraag of de gegevens over of van de meetreeksen en meetopstelling niet in tegenspraak met elkaar of fysisch onmogelijk zijn. We maken daarbij onderscheid tussen 'data-integriteit' en 'consistentie', terwijl het verschil tussen die twee wellicht niet meteen vanzelf spreekt. In het dagelijkse spraakgebruik kun je bijvoorbeeld de versieconflicten die we hier onder 'data-integriteit' scharen best 'niet-consistent' noemen. Versieconflicten verschillen echter van inconsistenties doordat het bij de eerste om dezelfde gegevens gaat die in principe identiek zouden moeten zijn, terwijl het bij inconsistenties gaat om verschillende gegevens die logisch of fysisch een bepaalde relatie met elkaar zouden moeten hebben. Het gaat bij consistentie-toetsen dus om toetsing op relaties. Bij zowel data-integriteit als consistentie speelt daarnaast het datamodel, en zogenaamde redundancies daarin, een grote rol. Versieconflicten kunnen voorkomen worden door in het datamodel eenmalige gegevens als zodanig te classificeren en gegevens dus per definitie niet dubbel op te slaan. Inconsistenties kunnen voorkomen worden door een juiste en scherpe keuze van attributen in het datamodel, zodat zoveel mogelijk slechts niet-redundante gegevens opgeslagen worden. Een goed voorbeeld hiervan is de opslag van de bovenkant en de onderkant van een peilbuisfilter. Wanneer beiden ten opzichte van NAP of een willekeurig ander referentiepunt worden opgeslagen kunnen beide gegevens inconsistent met elkaar zijn (de onderkant kan dan abusievelijk hoger zijn dan de bovenkant). Wanneer naast de bovenkant van het peilbuisfilter de lengte daarvan als positief getal wordt opgenomen in het datamodel, zijn er geen inconsistenties tussen beide mogelijk (maar andere typen fouten natuurlijk nog wel).

6.2 Methoden en criteria

6.2.1 Relationele operatoren

Toetsing op inconsistenties tussen gerelateerde gegevens is technisch gezien eenvoudig, omdat daarvoor een aantal standaardoperatoren c.q. zogenaamde relationele operatoren beschikbaar zijn in iedere programmeertaal. We zullen deze operatoren ook gebruiken in het protocol, en lichten ze hieronder toe:

=	gelijk aan
≠	niet gelijk aan
<	kleiner dan
≤	kleiner dan of gelijk aan
≥	groter dan of gelijk aan
>	groter dan

Bij administratieve of niet-kwantitatieve relaties kan ook verzamelingstheorie en de specifieke operatoren daarbinnen relevant zijn (bijv. \in , is element van). De uitvoering van een individuele consistentietoets in een willekeurig ICT-systeem vraagt doorgaans niet meer dan één regel programmacode. Het afhandelen van de toetsresultaten en de consequenties daarvan brengt meer werk met zich mee, maar het grootste deel daarvan hoeft niet voor iedere individuele consistentietoets apart te gebeuren.

6.2.2 Toetsbare relaties

Voor stijghoogte- en andere monitoringsgegevens geldt dat ze op verschillende manieren een relatie met elkaar kunnen hebben. Het is van belang daarbij onderscheid te maken tussen:

- **Chronologische relaties** – voorbeelden hiervan zijn de chronologische volgorde van de meetreeks, het feit dat een individuele drukopnemer niet op het zelfde moment in meerdere peilbuizen voor kan komen, of een meetreeks die niet gestart kan zijn vóórdat de peilbuis geplaatst is.
- **Ruimtelijke relaties** – hier kunnen we het eerdere voorbeeld van de bovenkant en de onderkant van een peilbuisfilter hergebruiken
- **Administratieve relaties** – een voorbeeld zijn de verschillende typen dataloggers die een individuele fabrikant daarvan in zijn collectie heeft

Naast het onderscheid in tijd en ruimte is het verhelderend om de eerder onderscheidde gegevens- of objecttypen ook hier uit elkaar te houden. De gegevens van een individueel object kunnen elkaar daarbij tegenspreken, maar ook die tussen verschillende objecten onderling. In praktijk van de grondwatermonitoring (zie ook Leunk, 2014) is het zinvol om te toetsten op inconsistenties in de:

- **Relaties binnen de meetreeksen** – daarbij gaat het om de chronologie, inclusief het feit dat metingen niet in de toekomst gedaan kunnen worden
- **Relaties binnen de meetopstelling** – hier gaat het zowel om de chronologie als om de ruimtelijke relaties in de constructie
- **Relaties tussen meetgegevens en meetopstelling** – daarbij gaat het om chronologie, maar vooral ook over de beperkingen die de meetopstelling oplegt aan de meetwaarden, zoals het feit dat er geen stijghoogtes onder de onderkant van het filter gemeten kunnen worden

6.3 Geautomatiseerde controle

6.3.1 Harde en zachte toetsing

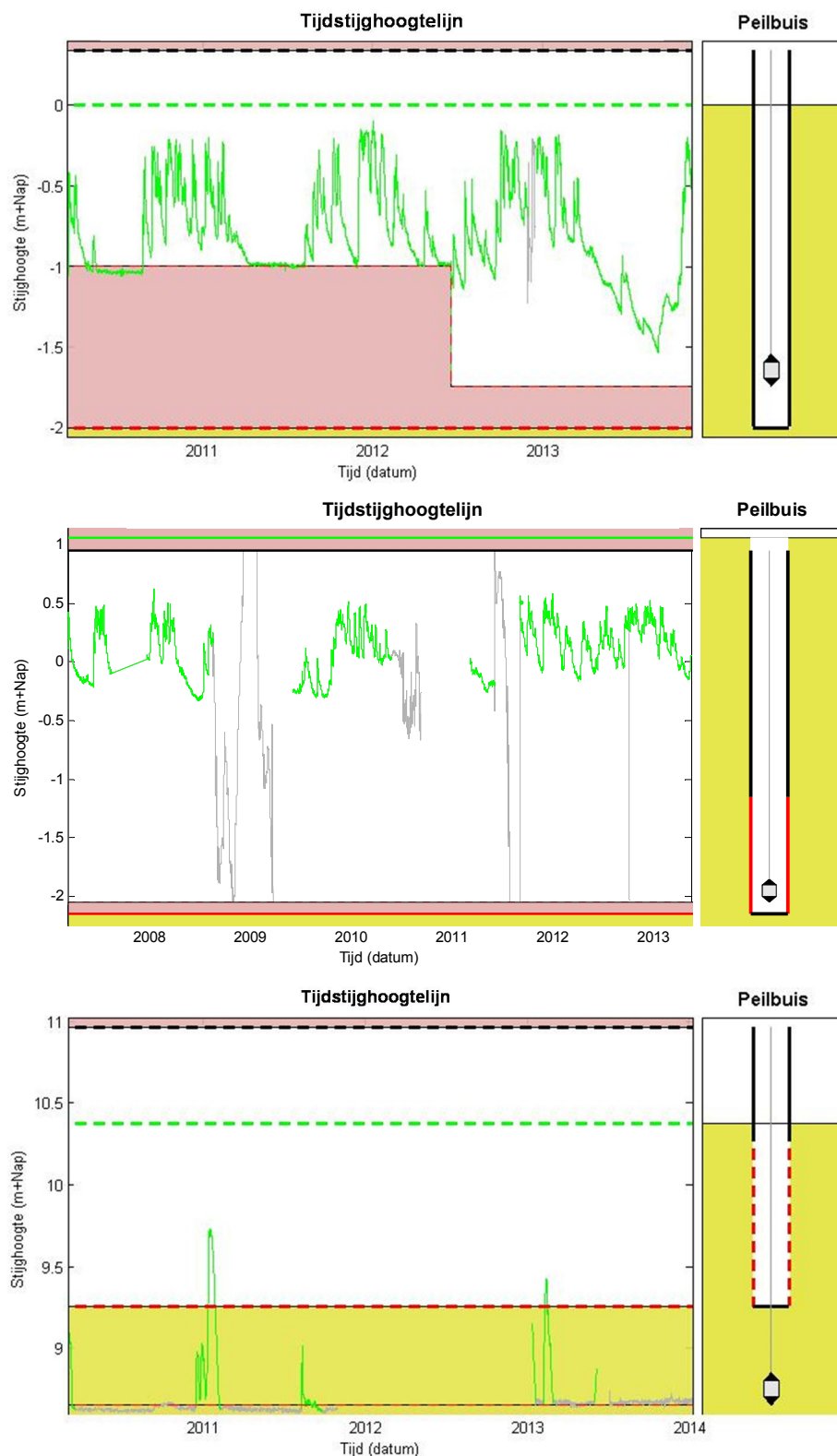
Een complicerende factor bij geautomatiseerde toetsing is dat de werkelijkheid minder hard, eenvoudig en exact is dan het op het eerste gezicht vaak lijkt. Dat is wellicht een van de redenen dat in de harde werkelijkheid van fysische modellen zachtere stochastische methoden, schattingen, optimalisatie en methoden als *fuzzy logica of data-assimilatie* zo'n grote vlucht hebben genomen.

Ook bij inconsistentietoetsing, die op het eerst gezicht hard lijkt, komen bij nadere beschouwing toch ook zachte aspecten naar voren. Zo is het een hard gegeven dat druksensoren geen waterdruk meer kunnen meten zodra de druksensor drooggevallen is, maar het nulpunt van de sensor is vanwege bias, drift en meetonzekerheden niet exact bekend. Een ander voorbeeld is toetsing op eventuele invoerfouten bij de te gebruiken waterdichtheid voor de omrekening van waterdruk naar waterstand. Ook daarbij bestaat er geen harde grens tussen welke waarden wel en welke niet fysisch mogelijk zijn, en bestaat een overgangsgebied waarbij waarden meer of minder plausibel zijn.

6.4 Visuele controle

Consistentie leent zich uitstekend voor visuele toetsing, omdat de ruimtelijke en chronologische relaties die er zijn goed gevisualiseerd kunnen worden in grafieken en tekeningen. Het is hierbij belangrijk op te merken dat de meetreeks en de meetopstelling

beiden zowel een dynamische / chronologische component hebben als een ruimtelijke. De meetopstelling kan daardoor zowel in als naast de meetreeks getoond worden om de begrenzingen die ze oplegt aan de meetreeks ook grafisch te illustreren. Buiten dit protocol om is in samenwerking met Joost Gooijer en Thomas de Meij, medewerkers van de provincie Overijssel, veel ervaring opgedaan met visuele beoordeling en toetsing. In dat kader zijn verschillende visualisaties ontwikkeld die zijn opgenomen in het programma Menyanthes (Von Asmuth e.a., 2012), waar we hier dankbaar gebruik van maken. Figuur 8 toont een aantal meetreeksen en peilbuizen uit de praktijk, waarin verschillende meer en minder grote inconsistenties duidelijk naar voren komen.



Figuur 8: Visualisatievoorbeelden van meer en minder grote consistentieproblemen. De grafieken tonen de gecorrigeerde druksensorreeks (groen), de originele druksensorreeks (donkergrijs), het nulpunt van de druksensor en onderkant filter (rode stippellijn), het maaiveld (groene stippellijn en groene vlakken) en de bovenkant van de peilbuis (zwarte stippellijn). In de grafieken is te zien:

Bovenste grafiek: Droogval van de druksensorreeks. In 2012 is de datalogger dieper ingehangen.

Middelste grafiek: Druksensorreeks met inconsistenties vanwege haperingen en defecten. De gaten zijn ontstaan doordat defecte drukopnemers niet altijd direct zijn vervangen.

Onderste grafiek: Inconsistentie tussen inhangdiepte en filterstelling. De reeks vertoont droogval en gaten.

7 Plausibiliteit (statistische controle)

7.1 Inleiding

Conform de definitie in hoofdstuk 0 verstaan we onder het begrip 'plausibiliteit' de vraag hoe waarschijnlijk het is dat gegevens toe te schrijven zijn aan ofwel een daadwerkelijke gebeurtenis, ofwel een meet- of verwerkingsfout. Voor het beantwoorden van die vraag zijn andere metingen in tijd en ruimte, naast metingen van andere parameters en kennis van het grondwatersysteem- en meetnet van groot belang. In (Von Asmuth en Van Geer, 2013) zijn de mogelijkheden geïnventariseerd (mede op basis van Van Geer, 2012) om gegevens geautomatiseerd te controleren aan de hand van hun statistische kenmerken en relaties met andere gegevens. Uit deze inventarisatie resulteerde een overzicht van toetsen die gebruik maken van respectievelijk het of de:

1. Verdeling van de meetwaarden
2. Regimecurve
3. Autocorrelatie (opeenvolging)
4. Tijdreeksmodellen zonder externe input
5. Tijdreeksmodellen met externe input
6. Verschil met (een) andere stijghoogtereeks(en) in de tijd
7. Verschil met andere stijghoogtereeksen in ruimte en tijd

Voor meer informatie over de verschillende toetsen, die hier niet verder in detail worden behandeld, verwijzen we naar de genoemde rapporten. Een kenmerk dat al deze methoden gemeenschappelijk hebben is dat ze (al dan niet een tijdreeks van) verwachte waarnemingen opleveren, waar de daadwerkelijke waarnemingen mee vergeleken kunnen worden. Dit resulteert in een reeks verschillen, waarbij het verschil c.q. de afstand tot de verwachtingswaarde een maat is voor hoe onverwacht, of omgekeerd hoe plausibel, de waarneming is.

In die zin reduceren de toetstypen 2 t/m 7 uit bovenstaande lijst uiteindelijk allemaal tot toetstypen 1 (de detectie van *outliers*), met als verschil dat de andere toetsen kennis bevatten over welke waarde verwacht mag worden, waarvan de complexiteit min of meer toeneemt met het oplopen van de nummering. In het algemeen wordt het onderscheidend vermogen van de toets groter naar mate er meer kennis wordt toegevoegd, wat echter ook een nadeel kan hebben. Met het toevoegen van kennis neemt het aantal aannamen dat daarbij gedaan wordt ook toe. Bij type 1 wordt de toets daarbij uitgevoerd op een reeks meetwaarden, terwijl het bij de overige om een reeks verschillen gaat (alhoewel het onderscheid daartussen vervalft als je de gemiddelde of mediane meetwaarde als verwachtingswaarde ziet). In (Von Asmuth en Van Geer, 2013) is relatief kort aandacht besteed aan de manier waarop je uit een reeks meetwaarden of verschillen geautomatiseerd de uitbijters kunt verwijderen. Mede omdat in (Leunk, 2014) bleek dat dat niet zomaar recht-toe-recht-aan is, en omdat het ook in (Walvoort e.a., 2013) een belangrijk onderwerp is gebleken, besteden we daar in dit rapport meer aandacht aan.

7.2 Methoden en criteria

7.2.1 Bezwaren tegen *outlier*-verwijdering

Bij het behandelen van methoden en criteria voor *outlier*-detectie hoort allereerst een waarschuwing: er kleven niet te onderschatten risico's aan het verwijderen van metingen, alleen vanwege het feit dat ze een extreme of afwijkende waarde hebben. Extreme waarden kunnen veroorzaakt worden door extreme omstandigheden, en daarmee feitelijk juist en

bovendien beleidsmatig zeer relevant zijn. Deze waarschuwing is zeker zo belangrijk bij beschouwingen over geautomatiseerde toepassing zoals hier het geval is. Automatisering van *outlier*-detectie leidt er toe dat *outlier*-verwijdering sneller en eenvoudiger wordt, met als risico dat dit ondoordacht en onterecht gebeurt. We citeren hier ter onderbouwing allereerst <http://en.wikipedia.org/wiki/Outlier>:

'Deletion of outlier data is a controversial practice frowned on by many scientists and science instructors; while mathematical criteria provide an objective and quantitative method for data rejection, they do not make the practice more scientifically or methodologically sound....'

Belangrijke aan te voeren bezwaren tegen het verwijderen van *outliers* zijn:

- **Extremen zijn niet onverwacht** – iedere dataset heeft immers meer of minder grote extremen. Een te beperkt voorkomen van extremen kan zelfs een aanwijzing zijn dat de data gemanipuleerd zijn.
- **Afwijkend is niet per definitie onmogelijk** - en onverwacht is niet per definitie fout. Interpreteer je dit wel zo, dan bedrijf je een cirkelredenering.
- **Op extremen moet je zuinig zijn** - ze zijn doorgaans van groter belang dan andere waarnemingen, juist omdat ze weinig voorkomen en bijv. vanwege aan extremen gerelateerde schade. Extremen kunnen bovendien wijzen op het veranderen van het systeem.

7.2.2 Methoden van *outlier*-detectie

Statistische controle en *outlier*-detectie wordt, zoals ook al in hoofdstuk 2 aan de orde kwam, in allerlei disciplines gebruikt. Er bestaat helaas geen eenduidige of harde definitie van het begrip *outlier*, maar er zijn veel methoden ontwikkeld die het probleem benaderen en daarbij de eerder genoemde bezwaren pogen te ondervangen. Voor de vuist weg zijn de verschillende methoden onder te verdelen in methoden die toetsen met behulp van:

1. **Gemiddelde en standaarddeviatie (gevoelig voor *outliers*)** – De afstand tot het gemiddelde, geschaald met de standaarddeviatie, wordt hierbij gebruikt als maat voor de afwijkendheid. Voor het kwantificeren van de kans dat een waarneming tot de verdeling behoort of een 'objectief' criterium te definiëren wanneer dat niet zo is, is een aanname over die verdeling nodig (die dan getoetst moet worden), net als voor het daarbij meewegen van het aantal waarnemingen dat beschikbaar is. Bij scheve verdelingen dient de afwijking in negatieve zin anders gewogen te worden dan die in positieve zin.
2. **Iteratieve benaderingen (minder gevoelig voor *outliers*)** – Een variant op bovenstaande aanpak is het iteratief detecteren van één *outlier* per keer, waarna de karakteristieken van de verdeling herberekend worden en opnieuw op de aanwezigheid van *outliers* getest wordt. Een bekende variant hiervan is de zogenaamde test van Grubbs (Grubbs, 1950). Deze aanpak zorgt er o.a. voor dat aanwezige *outliers* de berekende karakteristieken minder sterk verstoren.
3. **Kwartielen, afstanden of dichtheden (ongevoelig voor *outliers*)** - Het gebruik van kwartielen (waaronder de mediaan als 2^e kwartiel), afstand tussen of dichtheid van waarnemingen heeft als voordeel dat dit soort maten (in tegenstelling tot gemiddelde en standaarddeviatie) ongevoelig of robuust zijn voor de waarde van individuele *outliers*. Dit heeft als voordeel dat *outliers* niet eerst uit een referentiedataset verwijderd hoeven te worden, om andere waarnemingen als *outlier* te kunnen herkennen. In boxwhisker- plots wordt dit principe bijv. toegepast om

extremen en outliers te visualiseren, samen met de kwartielen zelf. De o.a. door (Walvoort e.a., 2013) gebruikte Hampel identifier (Hampel e.a., 1986) gebruikt bijv. de mediane en absolute mediane afstand daarvan als maat. Ook bij dergelijke methoden dienen echter aannamen gedaan te worden om te bepalen vanaf welke waarde(n) een waarneming (in meer of mindere mate) als outlier gezien wordt. Een ander voorbeeld is toetsing aan de hand van k-nearest neighbors (Altman, 1992).

4. **Gekozen verdelingsfuncties (ongevoelig voor aantal waarnemingen)** – omdat extremen zoals gezegd zelf niet onverwacht zijn, kan het probleem geherformuleerd worden in de vraag of er grotere of meer extremen zijn dan op grond van de aangenomen verdeling en het aantal waarnemingen verwacht mag worden. Door bijv. Chauvenet is een methode ontwikkeld waarin dit principe wordt toegepast (Chauvenet, 1960). Het voordeel van een dergelijk aanpak is dat het aantal gedetecteerde outliers (onder aanname van de verdelingsfunctie) niet afhangt van het aantal waarnemingen in de dataset die getoetst wordt. Wordt het aantal niet meegeewogen, dan neemt de kans op het classificeren van werkelijke extremen als outliers toe met de grootte van de dataset.

Naaste bovengenoemde methoden wordt ook de zogenaamde *leverage* of invloed gebruikt bij *outlier*-verwijdering, die aangeeft welke invloed een bepaalde waarneming heeft op een berekende statistiek of de uitkomsten van een (regressie)model. Twijfel aan de juistheid van een dergelijke meting kan dan een gegronde reden zijn om deze waarneming weg te laten bij de berekeningen (zie bijv. Cook en Weisberg, 1982). Omgekeerd kan een grote invloed op de uitkomsten ook aangeven dat een waarneming een *outlier* is. Bij gebruik van leverage komt echter de toepassing op de eerste plaats, en niet zozeer de metingen zelf.

7.2.3 Gebruik van statistische- en/of fysische kennis bij *outlier*-detectie

Alhoewel o.a. in (Walvoort e.a., 2013) praktische bezwaren worden aangevoerd tegen gebruik van meer complexe kennis (bijv. in de vorm van tijdreeksmodellen) bij toetsing en *outlier*-detectie, zijn er ook argumenten die juist voor het gebruik van statistische- en/of fysische kennis daarbij pleiten. We citeren hier wederom allereerst

<http://en.wikipedia.org/wiki/Outlier>:

'Rejection of outliers is more acceptable in areas of practice where the underlying model of the process being measured and the usual distribution of measurement error are confidently known.'

Alhoewel deze stelling niet nader toegelicht wordt, is ze intuïtief wel te begrijpen vanuit de eerder genoemde bezwaren. Belangrijke aan te voeren argumenten voor het incorporeren van kennis bij toetsing zijn aldus:

- (*extremen zijn niet onverwacht*) - Door de optredende extremen zoveel mogelijk te verklaren vanuit bestaande kennis en verwachting, is de *outlier*-detectie beter terug te brengen tot detectie van waarnemingen die echt onverwacht zijn
- (*afwijkend is niet onmogelijk*) - Door fysische kennis te gebruiken, al is het in elementaire vorm, verschuift de toetsing van statistisch afwijkend meer in de gewenste richting van fysisch niet-plausibel.
- (*op extremen moet je zuinig zijn*) - Behoud van de echte extremen is beter geborgd door kennis over wanneer extremen optreden zoveel mogelijk mee te nemen bij de toetsing (hoge standen komen bijv. vaker voor in de winter of na grote neerslagsommen, zie gebruik van de regimecurve of tijdreeksanalyse).

- **Praktische voordelen:**

- De efficiëntie van de toets neemt toe naarmate de gegevens beter verklaard worden door het model. Ook de fout die je maakt door abusievelijk een echt extreem te verwijderen wordt kleiner. Daartegenover staat dat met toenemende incorporering van kennis vaak ook het aantal aannamen groeit, wat kan leiden tot abusievelijke verwijdering als een aanname onterecht is.
- Scheve en grillige verdelingen bemoeilijken *outlier*-detectie. Des te sterker de scheefheid en het gedrag van de meetwaarden al door het model verklaard worden, des te groter is de kans op 'vriendelijke' verdelingen van de verschilreeks of residuen.

7.3 Geautomatiseerde controle

Automatisering van *outlier*-verwijdering stuit op principiële bezwaren, en bij ondoordacht gebruik of tijdsdruk kan verregaande automatisering van *outlier*-detectie daar makkelijk toe leiden (in ieder geval met de huidige kennis). Geen van de nu beschikbare methoden wint daarbij duidelijk van andere en is universeel toepasbaar. Terwijl de uitkomsten van de eerder beschreven data-integriteits- en consistentietoetsen eenvoudig en recht-toe-recht-aan zijn, is het controleren en eventueel verwijderen van waarnemingen op basis van afwijkend gedrag geen eenduidig te definiëren of objectiveren opdracht.

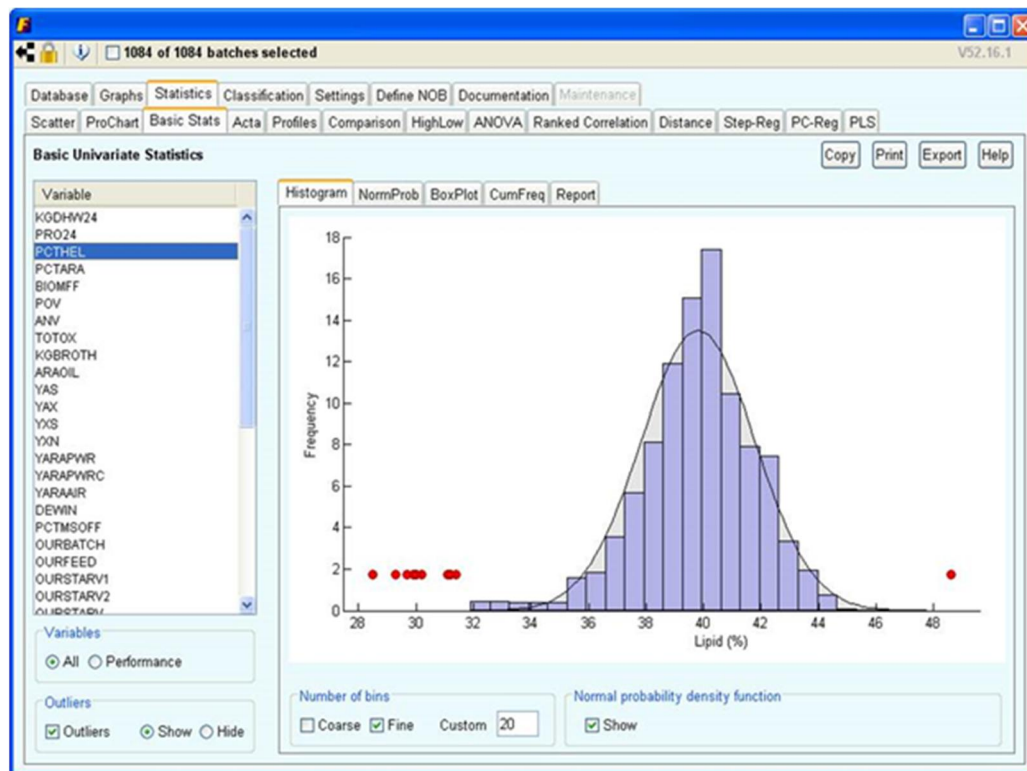
Er zijn wel aanbevelingen en mogelijkheden om de bruikbaarheid van de techniek te verhogen. Het combineren van toetsen biedt perspectief om verschillende typen fouten beter op te kunnen sporen. Nader onderzoek is nodig naar het gebruik van statistische- en/of fysische kennis (heeft principiële voordelen), subtielere, evt. gecombineerde methoden voor het opsporen van kleinere fouten, en naar het controleren van historische gegevens. Er is inmiddels onderzoek in gang gezet in opdracht van de provincie Overijssel naar het gebruik van scheve verdelingen en/of meer complexe kennis bij *outlier*-detectie.

Statistische controle van meetreeksen dient dus met verstand van meetnet en hydrologie uitgevoerd en beoordeeld te worden. Mede om deze reden verdient het aanbeveling (zie ook hoofdstuk 9) om een cursus of workshop rond datakwaliteitscontrole te organiseren, voor en zo mogelijk in overleg met databeheerders en hydrologen van geïnteresseerde partijen.

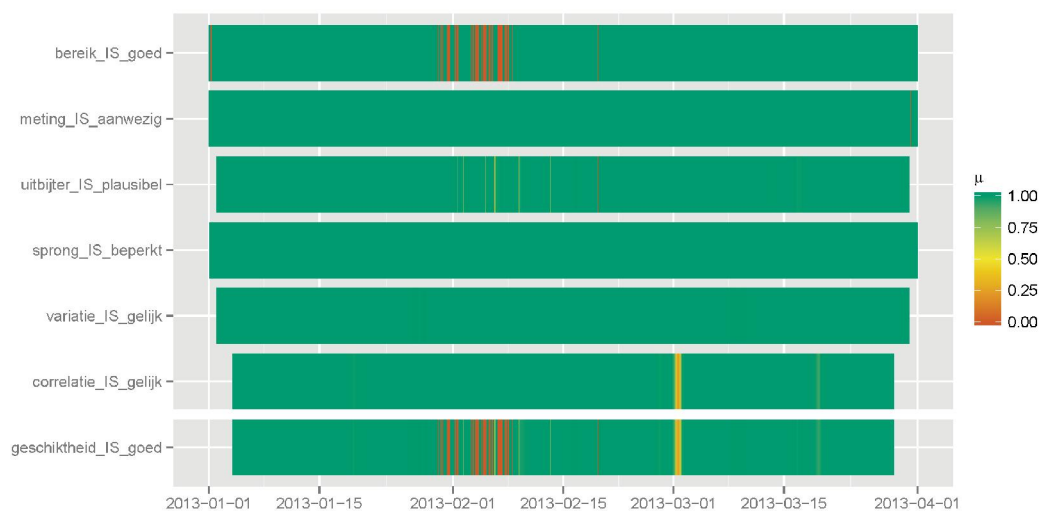
7.4 Visuele controle

Omdat de resultaten van een plausibiliteitscontrole voor meerdere uitleg vatbaar zijn, en dit met verstand van zaken dient te gebeuren, is een goede visualisatie van deze gegevens ook hier zeker zo belangrijk. Zoals genoemd in paragraaf 7.2.2 is een zogenaamde boxwhisker-plot een goede manier om een verdeling met de extremen en *outliers* daarvan te visualiseren. Indien daar ruimte voor is, geeft visualisatie van het volledige histogram en de gedetecteerde *outliers*, naast eventueel de gefitte verdelingsfunctie, nog meer informatie en mogelijkheden om het optreden van *outliers* visueel te beoordelen (zie bijv. Figuur 9).

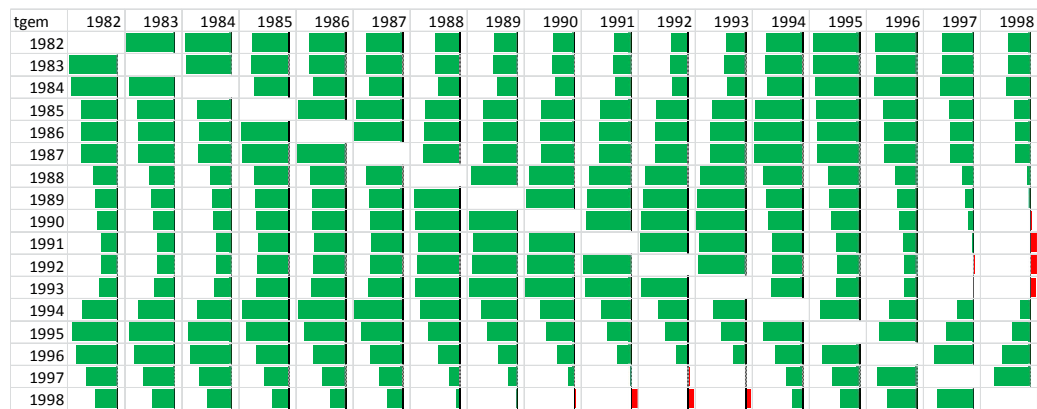
Wanneer meerdere toetsen worden uitgevoerd op dezelfde reeks, dient deze informatie samengevat te worden om ze in samenhang te kunnen beoordelen. Een mooi voorbeeld daarvan is te vinden in Figuur 10. Figuur 11 toont daarnaast een voorbeeld van visualisatie van de resultaten van een toets waarbij verschillende perioden (jaren) binnen een stijghoogtereeks zijn vergeleken.



Figuur 9: Screenshot van een applicatie van een AEX-genoteerd bedrijf uit de voedselindustrie, met visualisatie van histogram, outliers en de gekozen (normale) verdelingsfunctie.



Figuur 10: Samenvatting van de (meerwaardige) resultaten van verschillende controles op eenzelfde hoogfrequente druksensorreeks (bron: Walvoort e.a., 2013).



Figuur 11: Visualisatie van de resultaten van een t-toets in een matrix, waarbij verschillende perioden (jaren) binnen een stijghoogtereeks zijn vergeleken (reeks B31H065101; tijdvak =9 jaar, $p=0.01$ (bron: van Geer, 2013)).

8 Controle en correctie van sensordata (controlemeting)

8.1 Inleiding

8.1.1 Systematische en toevallige fouten en afwijkingen

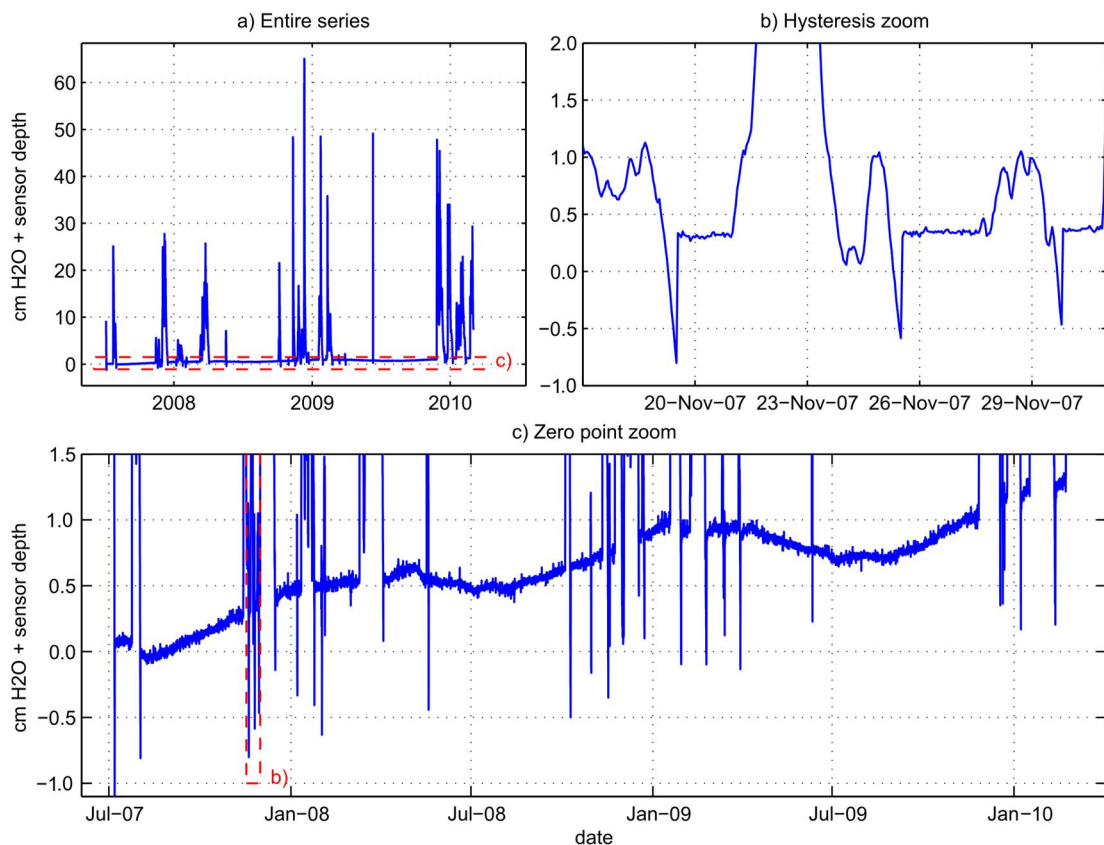
Conform de definitie in hoofdstuk 0 verstaan we onder 'controle en correctie van sensordata' de vraag of sensormetingen behept zijn met systematische en/of toevallige fouten en afwijkingen, en wanneer en hoe de metingen gecorrigeerd of afgekeurd dienen te worden. We beperken ons daarbij tot het gebruik van controle- of dubbelmetingen, en enkele specifieke controles die het gebruik van druksensoren met zich meebrengen, zoals droogval van de sensor en fouten bij de omrekening van druk naar stand.

Voor het opsporen van fouten en afwijkingen in sensorgegevens is van belang dat de eigenschappen daarvan in belangrijke mate verschillen van die van fouten bij handmatige meting en verwerking (zie ook McLaughlin en Cohen, 2011; Sorensen en Butcher, 2011; Von Asmuth, 2011; Post en Von Asmuth, 2013). De afwijkingen van sensoren zijn in de regel relatief klein ten opzichte van bijv. fouten in de metadata of fouten die gemaakt worden tijdens het verwerkingsproces, maar ze zijn desondanks vaak wel belangrijk (Von Asmuth en Van Geer, 2013; Leunk, 2014). Afwijkingen van sensoren volgen daarnaast vaak een dusdanig systematisch of vast patroon dat nieuwe metingen niet sterk afwijken van de karakteristieken van de periode daarvoor, waardoor bijv. de in hoofdstuk 7 beschreven plausibiliteitstoetsen ze niet goed op kunnen sporen. In een eerder onderzoek naar de kwaliteit van druksensorreeksen (Von Asmuth, 2011) zijn de verschillende typen fouten en afwijkingen van (druk)sensoren behandeld. De belangrijkste daarvan zijn:

1. **Bias en Drift** – deze beschrijven de verschuiving van het nulpunt van de sensor (Bias), die vaak langzaam oploopt in de tijd (Drift). De drift blijkt in veel gevallen ongeveer lineair te verlopen (zie bijv. Figuur 12), maar kan ook plots veel groter worden. Dit laatste scharen we onder 'defect van de sensor'.
2. **Temperatuurafhankelijkheid** - afwijkingen die samenhangen met de temperatuur, die invloed heeft op elektrische weerstandsmetingen in de sensor.
3. **Haperingen en outliers** – incidentele of kortstondige afwijkingen. Een oorzaak die hiervoor genoemd wordt is de aanwezigheid van condens c.q. waterdruppeltjes op het meetmembraan.
4. **Defect of uitval van de sensor** – afwijkingen die plots sterk toenemen of dataloggers die anderszins disfunctioneren.

In mindere mate zijn van belang:

5. **Ruis** - random variatie van de gemeten waarde rond de werkelijke waarde.
6. **Hysteresis** – Verschil in de relatie tussen druk en indrukking van het membraan bij oplopende en afnemende druk.



Figuur 12: Verschillende 'zooms' op een stijghoogtereeks van een fysisch gecompenseerde drukopnemer die periodiek droogvalt, waarin verschillende typen afwijkingen zichtbaar worden. a) grafiek van de hele reeks, met weergave van de aslimieten uit figuur c. b) zoom-in op vier gebeurtenissen waarop het waterniveau tot onder de onderkant van de sensor rijkt, waarbij fouten ten gevolge van hysteresis zichtbaar worden (resultierend in een negatieve waterkolom wanneer het peil dalend is); c) zoom op het nulpunt van de sensor, waarbij temperatuurafwijking (de seizoensfluctuatie in het nulpunt), drift (de geleidelijke toename van het nulpunt in de tijd), ruis (de random variatie rond het nulpunt) en hysteresis zichtbaar worden (bron: Post en Von Asmuth, 2013).

Er kunnen naast de hierboven beschreven fouten en afwijkingen in de meetwaarden ook fouten en afwijkingen optreden in de geregistreerde tijdstippen vanwege het voor- of achterlopen van de klok van de datalogger, en/of onjuist gebruik van zomer- en wintertijd.

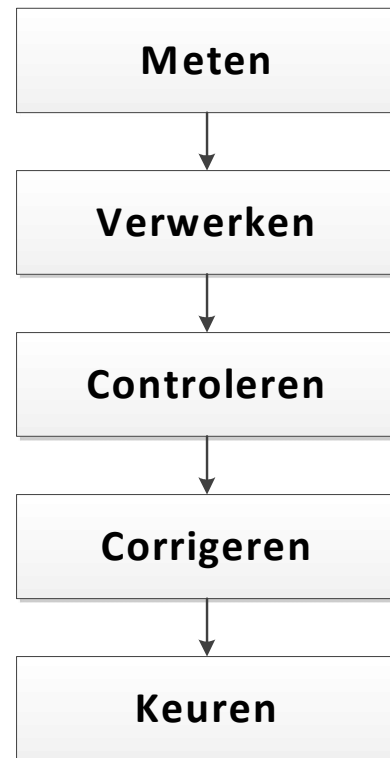
Het is voor het controleren en corrigeren van sensordata belangrijk om op een iets hoger niveau onderscheid te maken tussen:

- **Systematische afwijkingen** – hieronder vallen met name nulpuntsverschuivingen zoals bias, drift en temperatuursafhankelijkheid. Wanneer de grootte en het gedrag van systematische afwijkingen ingeschat kunnen worden, kunnen deze gecorrigeerd worden. Correctie en monitoring van de nulpuntsverschuiving komt daarbij neer op softwarematige ijking van het instrument.
- **Toevallige afwijkingen** – zoals haperingen, outliers en defecten. Afhankelijk van de grootte van de afwijking of fout is de enige optie het afkeuren van de data en/of het instrument.

8.1.2 Stappen in het meetproces

Terwijl de andere onderdelen van dit protocol generiek zijn en bijv. net zo goed op historische data toegepast kunnen worden, richt een onderwerp als 'controle en correctie van sensordata' zich vooral op het actuele monitoringsproces en het inwinnen van nieuwe data (alhoewel het ook voor hercontrole van bestaande sensordata opgaat). Dit protocol maakt daarom onderscheid tussen de verschillende stappen in het werkproces van inrichting van de meetlocaties tot definitieve meetreeks, omwille van de eenvoud en praktische toepasbaarheid ervan. We onderscheiden daarbij de:

- **Metten** – en de specificaties daarvan (bijv. frequentie)
- **Verwerken** – stappen die nodig zijn om tot de meetreeks te komen (bijv. omrekening naar waterkolom)
- **Controleren** – het vergelijken van metingen (bijv. verschil tussen hand- en loggermeting)
- **Corrigeren** – bijstellen van systematische afwijkingen (bijv. van de loggerklok)
- **Keuren** – goed- of afkeuren van data en/of logger, wanneer de fout te groot is



Figuur 13: Stappen in het proces van meting tot definitieve meetreeks.

Naast een onderverdeling in controles en richtlijnen per stap kent het protocol ook een onderverdeling in datatype (luchtdruk, waterdruk of -hoogte, tijdsregistratie), waarover hieronder meer.

8.2 Methoden en criteria

8.2.1 Specifieke kwesties en toetsen voor druksensoren

Naast de fouten en afwijkingen die (digitale) sensoren in het algemeen kunnen vertonen, zijn er zoals gezegd ook een aantal kwesties en gegevens die specifiek zijn voor het gebruik van druksensoren. Een specifiek punt voor het gebruik van absolute druksensoren is bijvoorbeeld dat metingen van de luchtdruk nodig zijn om metingen van de totale druk om te kunnen rekenen naar de hoogte van de waterkolom boven het sensornulpunt. Omdat de luchtdrukmeting een geheel op zichzelf staande en eigen type meting betreft besteden we daar apart aandacht aan in de volgende paragraaf. Hier behandelen we in het kort het:

- **Omrekenen van luchtdrukreeksen** – wanneer de hoogteligging van de luchtdruksensor significant verschilt van het waterpeil in de peilbuis dienen de luchtdrukgegevens hiernaartoe omgerekend te worden. KNMI-luchtdrukgegevens worden standaard omgerekend naar NAP niveau, in het KNMI-handboek is de daartoe te gebruiken formule te vinden (KNMI, 2000).
- **Omrekenen van druk naar waterkolom** – de waterdruk die resteert na luchtdrukcompensatie dient omgerekend te worden naar de hoogte van de waterkolom boven het sensornulpunt. Om dit te kunnen dienen zowel de dichtheid van het water als de gravitatieconstante opgegeven te worden (waarbij die waarden soms 'ingebakken' zitten in de uitleessoftware, maar soms ook niet, zie ook van

Asmuth, 2001). Om invoer- en interpretatiefouten hierbij te kunnen controleren dienen de gebruikte parameters opgeslagen te worden. **N.B.:** Daarnaast is voor het berekenen van de stijghoogte de dichtheid van de hele waterkolom in de peilbuis van belang, waar verschillen in op kunnen treden, bijvoorbeeld door lekken in de peilbuis (zie ook Post en Von Asmuth, 2013). Alhoewel controles daarop wel mogelijk zijn, laten we dit punt hier verder buiten beschouwing.

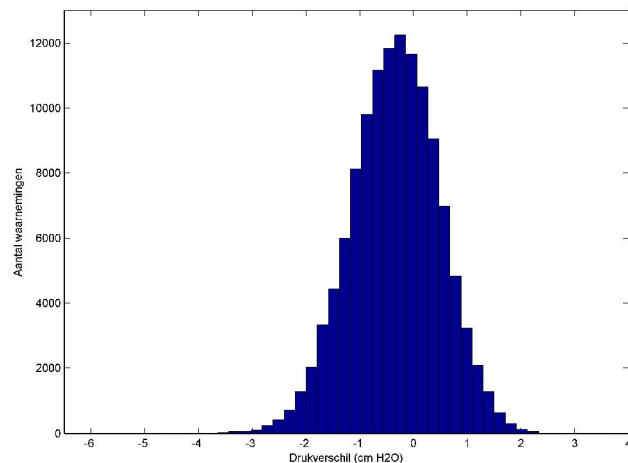
- **Droogval van de sensor** – wanneer het waterpeil in een peilbuis onder het sensornulpunt zakt, registreert deze alleen nog de luchtdruk. In dergelijke gevallen dienen de metingen niet meer als stijghoogte, maar als ‘droge sensor’ gelabeld te worden **N.B.:** dit is niet hetzelfde als het huidige ‘droog’ label in DINO, dat aangeeft dat de stijghoogte onder de onderkant van het filter ligt.

8.2.2 Luchtdrukmetingen

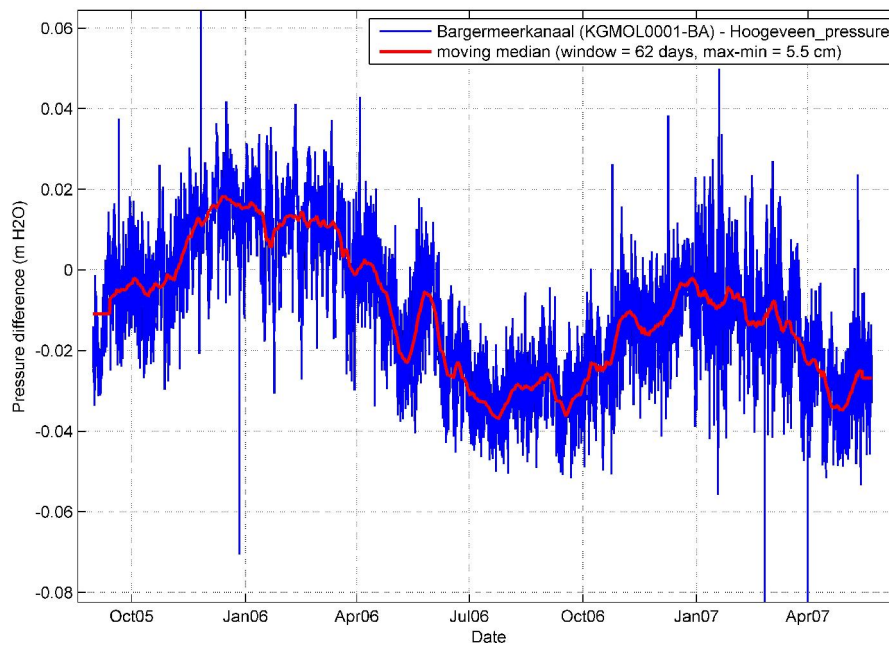
8.2.2.1 Lokale versus KNMI-luchtdrukmetingen

Voor de fouten en afwijkingen die optreden bij de luchtdrukcompensatie van druksensormetingen is van belang dat luchtdrukgegevens ofwel a) kunnen worden ingewonnen m.b.v. eigen, lokale (lucht)druksensoren, ofwel b) gedownload kunnen worden van de KNMI-site. De overwegingen die hierbij spelen en tegen elkaar afgewogen moeten worden zijn:

1. **Kwaliteitsverschil** - Het KNMI gebruikt digitale barometers met een kwalitatief hoogwaardiger type druksensoren dan bij grondwatermonitoring gebruikelijk is (zie ook Von Asmuth, 2011, <http://www.knmi.nl/klimatologie/metadatas/stationslijst.html>). De zorg voor de kwaliteit van de data en meetopstelling kan in dat geval worden overgelaten aan het KNMI. Omdat de luchtdrukgegevens van het KNMI standaard omgerekend worden naar NAP-niveau, dienen deze nog wel teruggerekend worden naar de hoogte van het waterpeil in de peilbuis.
2. **Afstandsverschil** - Het KNMI meet logischerwijze op haar meetstations en niet ter plaatste van de individuele peilbuis. Het verschil in luchtdruk tussen de twee meetlocaties neemt toe met de afstand daartussen, waardoor lokale luchtdrukmetingen een nauwkeuriger resultaat opleveren wanneer de kwaliteit daarvan goed is. Ter illustratie toont Figuur 14 een histogram van het verschil in luchtdruk (in cm H₂O) tussen twee nabijgelegen KNMI-stations (Maastricht en Eindhoven) in de periode 1-1-2010 tot 1-1-2014. Het gemiddelde absolute verschil bedraagt 0.7 cm, het maximale absolute verschil 6.3 cm. Hierbij is van belang dat de relatieve drukopnemers van de firma Keller standaard een eigen



Figuur 14: Histogram van het verschil in luchtdruk (in cm H₂O) tussen twee naburige KNMI-stations (Maastricht en Eindhoven) in de periode 1-1-2010 tot 1-1-2014. Het gemiddelde absolute verschil bedraagt 0.7 cm, het maximale absolute verschil 6.3 cm).



Figuur 15: Verschil tussen een lokale luchtdrukreeks en die van KNMI-station Hoogveen. Een lopende mediaanfilter (in rood) kan gebruikt worden om systematische verschillen over een langere periode te berekenen en te corrigeren (bron: De Meij & Von Asmuth, 2011).

lokale luchtdruksensor bevatten (Keller BV, 2010), die geen extra kosten met zich mee brengt. Daarnaast vergemakkelijkt nauwkeurige luchtdrukcompensatie de interpretatie van fouten en afwijkingen in waterdrukmetingen. Bij gebruik van eigen luchtdrukmetingen heeft een afzonderlijke kwaliteitscontrole en correctie van deze gegevens echter de voorkeur (zie volgende paragraaf).

3. **Tijdsverschil** - Voor controle in het veld is van belang dat de KNMI-luchtdrukmetingen niet instantaan beschikbaar zijn. De metingen van de vorige dag komen ineens ter beschikking via internet, maar die gegevens zijn doorgaans bij de start van een veldwerkdag nog niet beschikbaar. Er is dus een tijdsverschil tussen de actuele luchtdruk en de laatst beschikbare KNMI-meting dat doorgaans 32 tot 40 uur bedraagt. De verschillen die hierdoor veroorzaakt worden kunnen bij onstabiel weer oplopen tot +/- 20 cm waterkolom, wat KNMI-gegevens ongeschikt maakt voor controle van actuele metingen in het veld. Waterschap Noorderzijlvest is om die reden bezig een Field Wizard te laten ontwikkelen die actuele en lokale luchtdrukcompensatie mogelijk maakt m.b.v. een USB-barosensor.

8.2.2.2 Controle en correctie van lokale luchtdrukmetingen

Het controleren en corrigeren van zelf verzamelde luchtdrukgegevens is relatief eenvoudig en kan grotendeels geautomatiseerd worden, ook hier vanwege de vrije beschikbaarheid van luchtdrukgegevens van het KNMI. In 'Het weer, een lessenserie over meteorologie' (KNMI, 2012) adviseert het KNMI scholieren al om eigen barometergegevens te iken aan die van het instituut en het iken van goedkopere aan betrouwbaarder sensoren wordt ook geadviseerd en toegepast door fabrikanten van digitale barometers (zie bijv.

<http://www.paroscientific.com/technel740-2.htm> of de barometers van Vaisala waarbij meerdere sensoren elkaars nulpunt bewaken (Anonymous, 2009). Vanwege het verschil in afstand tussen het KNMI-station en de eigen, lokale luchtdruksensoren kunnen afwijkingen in het nulpunt van die laatste niet zomaar gecorrigeerd worden door middel van directe vergelijking. Ook bij de controle en correctie van luchtdrukgegevens is het zinvol onderscheid te maken tussen:

- **Systematische afwijkingen** – zoals bias, drift en temperatuurafhankelijkheid. Door gebruik te maken van een lopend mediaanfilter zoals beschreven in (De Meij en Von Asmuth, 2011), zie ook Figuur 15) kunnen systematische verschillen in het nulpunt over een langere periode berekend en gecorrigeerd worden, waarmee de lokale variatie in tijd en ruimte behouden blijft. Een dergelijk filter is vanwege het gebruik van een mediaan robuust voor outliers
- **Toevallige afwijkingen** – zoals haperingen, outliers en defecten. Waarnemingen die meer verschillen van de omringende KNMI-stations dan dat deze van elkaar verschillen vallen buiten de regionale gradiënten en zijn daardoor minder plausibel.

Het is ook hier van belang dat de originele gegevens bewaard blijven, naast de correcties die gedaan zijn.

8.2.3 Waterdruk- en waterpeilmetingen

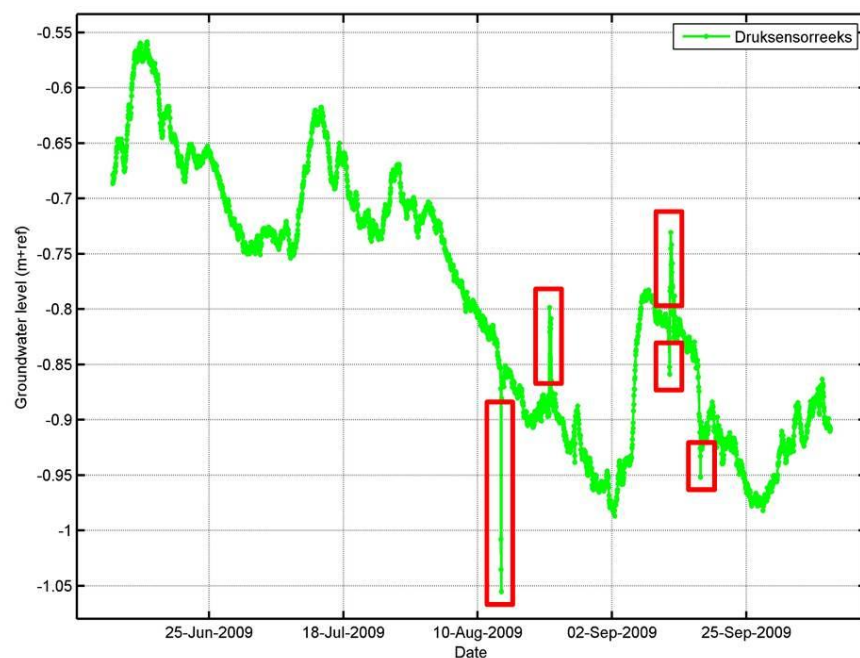
8.2.3.1 Gebruik van handmatige controlemetingen

Zoals reeds genoemd zijn de afwijkingen van sensoren in de regel relatief klein en volgen daarnaast vaak een dusdanig systematisch of vast patroon dat handmatige controlemetingen nodig zijn om deze op te sporen en waar nodig te corrigeren. Naast de wijze van controleren en corrigeren zijn van belang de:

- **Verschilberekening** – omdat het tijdstip van de controle- en sensormetingen vaak niet identiek zijn (wel als realtime meting mogelijk is, zoals bij Keller-loggers) , dient er hierbij a) een criterium te zijn voor wat het maximale tijdsverschil is en b) dient het tijdsverschil overbrugt te worden met bijvoorbeeld lineaire interpolatie
- **Frequentie van de handmeting** – die bepaalt hoe snel en goed fouten en afwijkingen op te sporen zijn
- **Frequentie van de sensormeting** – die het maximale tijdsverschil tussen controle- en sensormeting bepaalt
- **Tijdstip en snelheid van de kwaliteitscontrole** – die bepaalt hoe snel en adequaat problemen hersteld kunnen worden. Bij controle in het veld kan de uitlezing en handmeting zo nodig nog eens overgedaan worden, en/of defecte apparatuur direct worden vervangen.

Bij de controle en correctie van waterdrukgegevens is het zinvol onderscheid te maken tussen:

- **Systematische afwijkingen** – zoals bias en drift (de temperatuur van grondwater varieert beduidend minder dan die van de lucht). Systematische verschillen in het nulpunt kunnen ook hier liefst over een langere periode berekend en gecorrigeerd worden door de verschillen tussen sensor- en handmeting over een langere periode te beoordelen.
- **Toevallige afwijkingen** – zoals haperingen, *outliers* en defecten. De frequentie van handmetingen is doorgaans onvoldoende om toevallige afwijkingen alleen op basis daarvan op te sporen. Daarnaast kan ook wanneer een toevallige afwijking wel samenvalt met een controlemeting, daarmee nog niet de vraag beantwoord worden vanaf wanneer bijv. een ontstaan defect is opgetreden. Toevallige afwijkingen verschillen doorgaans echter qua waarde en gedrag van de naburige sensormeetreeks, zodat de sensorreeks zelf behulpzaam is hierbij. Voor



Figuur 16: Druksensorreeks met schijnbaar toevallige, kortstondige afwijkingen in zowel positieve als negatieve richting.

kortstondige afwijkingen (zie bijv. Figuur 16) kunnen hierbij ook de in het vorige hoofdstuk beschreven plausibiliteitstoetsen behulpzaam zijn.

8.2.4 Tijdsregistratie

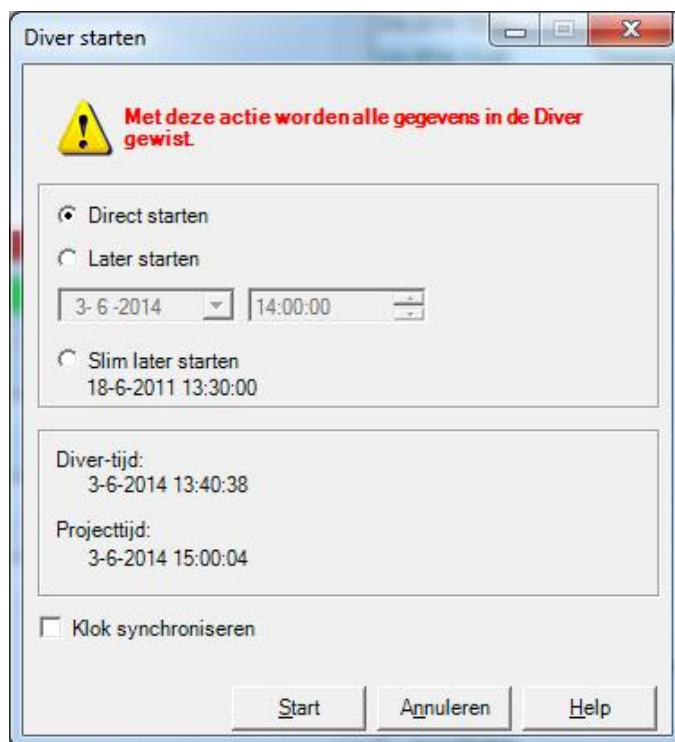
Naast afwijkingen in de meetwaarde kunnen sensoren ook afwijkingen in de tijdsregistratie of interne klok vertonen. Een fenomeen dat specifiek is voor het gebruik van absolute druksensoren is dat de klok van de waterdruksensor al dan niet in toenemende mate afwijkt van de klok of tijdsregistratie van de luchtdrukmetingen, wat de fout die ontstaat bij de luchtdrukcompensatie doet toenemen. Omdat door verschillen in tijdstip zowel te hoge als te lage luchtdrukken voorkomen kan dit resulteren in een aanzienlijke 'ruis' op de uiteindelijke meetreeks (Figuur 18). Het is dus belangrijk om ook de tijdregistratie van software, data en datalogger te controleren op:

1. **Correcte omgang met de tijdseenheid** - Het ligt voor de hand om data in een computersysteem op te slaan in een standaard tijdseenheid (wintertijd of UTC+1 voor Nederland), en daarbij niet de overschakeling van zomer- en wintertijd te volgen, bijv. omdat er anders verschillende waarnemingen op hetzelfde tijdstip mogelijk zijn bij het terugzetten van de klok. Gebruikers en veldwaarnemers gaan aan de andere kant wel uit van zomer- en wintertijd, waardoor voor hen juist de zomertijd voor de hand ligt, bijv. bij het registreren van een handwaarneming in de zomer. Verwarring ligt hierdoor altijd op de loer. Deze is alleen op te lossen door te zorgen dat programmeurs en computersystemen voor elke databron (incl. handmatige invoer) nagaan wat daarvan de tijdseenheid is en deze telkens netjes te converteren naar UTC+1.
2. **Verschil met de klok van de (veld)computer** - Met de uitleessoftware van gangbare dataloggers is het niet eenvoudig om de opgetreden en eventueel gesynchroniseerde tijdsverschillen te registreren en bewaren, wat veel problemen op dit gebied zou kunnen oplossen. Bij Diver-Office is het tijdsverschil alleen zichtbaar in de gebruikersinterface bij het herstarten van de Diver (zie Figuur 17, met hier een significant tijdsverschil van 1 uur en 20 minuten), en is verder niet

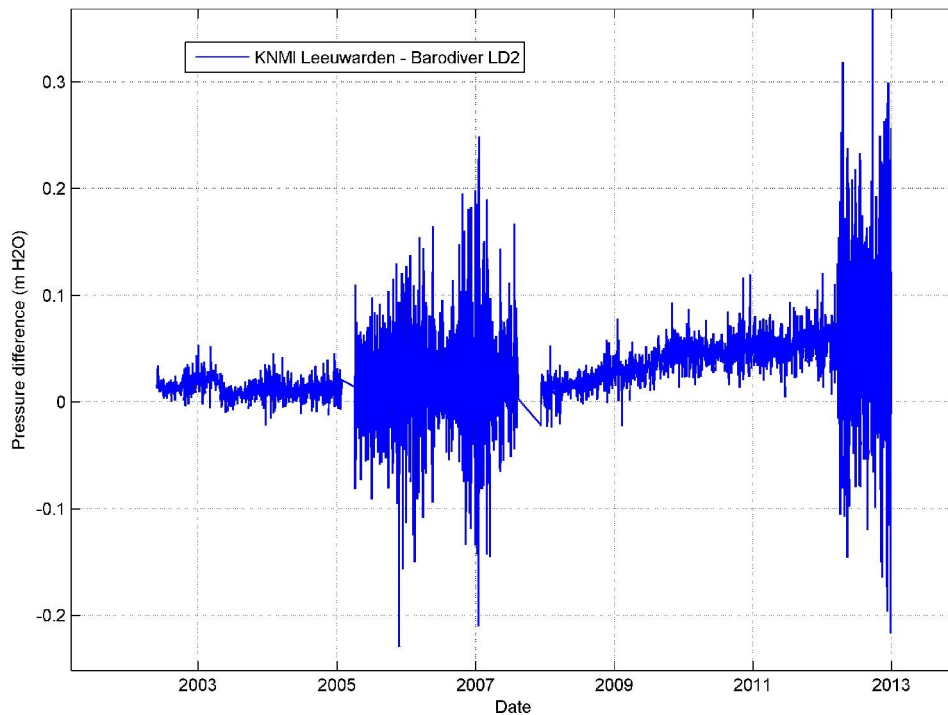
terug te vinden in de uitvoerdata (MON-files). In de Keller-reader is het tijdsverschil standaard zichtbaar bij het uitlezen, en wordt daarnaast ook weggeschreven naar een LOG-file. Deze wordt echter makkelijk overschreven door andere modules van de Keller-software en niet vaak bewaard. Mede hierdoor is er geen goed inzicht in hoe groot de problemen zijn met de tijdsregistratie in de praktijk, en wordt de tijdsregistratie niet standaard gecontroleerd en zeker niet gecorrigeerd.

Ook voor fouten in de tijdsregistratie geldt dat deze niet per definitie alleen bij de datalogger hoeven te liggen. Uitleessoftware van dataloggers zoals Diver-Office en de LoggerDCX-reader van Keller bieden de mogelijkheid om de klok van de logger te synchroniseren met die van de veldcomputer (Figuur 17). Bij voorbeelden uit de praktijk waar grotere fouten in de tijdsregistratie zijn aangetroffen (zoals in Figuur 18) bleek juist de klok van de veldcomputer verkeerd te lopen.

Omdat dataloggers doorgaans op vaste tijdstippen meetwaarden vastleggen (bijv. op de hele uren), zijn afwijkingen en correcties (synchronisaties) van de klok doorgaans niet zichtbaar in de data. Tijdsverschillen die groter zijn dan de meetfrequentie kunnen wel opgespoord worden, door het tijdstip van het aanmaken van de uitvoerbestanden te vergelijken met die van de laatste waarneming of door het ontbreken van waarnemingen na synchronisatie. Het tijdsverschil zelf is hiermee niet exact te reconstrueren. Geautomatiseerde registratie en controle van het tijdsverschil is op dit moment alleen mogelijk bij Keller-loggers, door de LOG-file van de reader te bewaren en in te lezen, of door deze handmatig af te lezen en vast te leggen bij gebruik van Divers. Het verdient aanbeveling dat de verschillende fabrikanten van dataloggers hun software zodanig aanpassen dat tijdsverschillen en eventuele synchronisaties voortaan vastgelegd en uitgevoerd worden in hun databestanden.



Figuur 17: Screenshot uit Diver Office waarin het verschil tussen de klok van de logger en die van de (veld)computer zichtbaar is (hier 1 uur en 20 minuten), met optie om de loggerklok bij te stellen c.q. te synchroniseren.



Figuur 18: Verschil tussen lokaal gemeten luchtdruk (LD2) en luchtdruk van KNMI station Leeuwarden. Een fout in de tijdsregistratie van de lokale sensor van +/- 24 uur veroorzaakt hier 'ruis' of verschillen van meer dan 20 cm H₂O (bron: Leunk & Von Asmuth, 2013).

8.3 Geautomatiseerde controle

Terwijl luchtdrukmetingen geautomatiseerd gecontroleerd en gecorrigeerd kunnen worden is dat bij waterdrukgegevens (in ieder geval met de huidige kennis) niet goed mogelijk omdat:

- de frequentie van controlemetingen minder groot is
- de oorzaak van fouten en afwijkingen niet alleen bij de sensor gezocht moet worden, maar ook bij de controlemeting en/of bij berekeningsstappen en de daarbij gebruikte metadata.

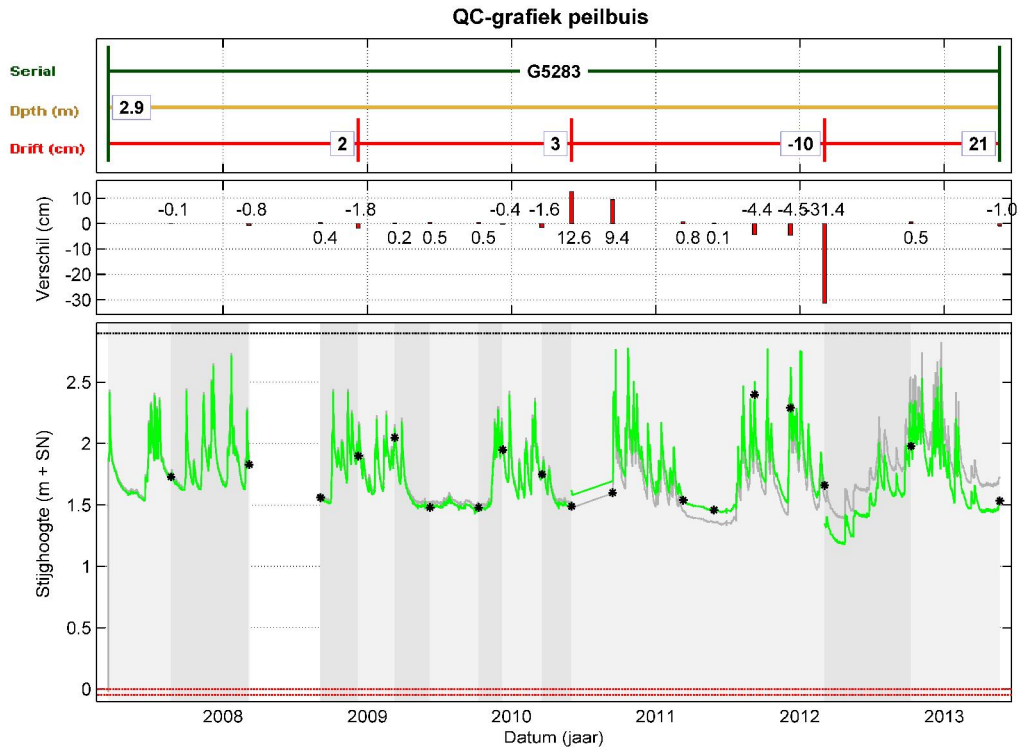
De controle en eventuele correctie van waterdrukreeksen dient dus met kennis van zaken beoordeeld en uitgevoerd te worden!

8.4 Visuele controle

Omdat de data bij de controle en correctie van waterdrukreeksen voor meerdere uitleg vatbaar zijn en dit met verstand van zaken dient te gebeuren, is een goede visualisatie van deze gegevens nog belangrijker dan het altijd al is. Om onderscheid te kunnen maken tussen:

- Systematische en toevallige fouten en afwijkingen
- Fouten en afwijkingen in de sensormetingen en de controlemetingen
- Fouten in de bewerkingsstappen en/of in de metadata

dienen de genoemde gegevens zoveel mogelijk in samenhang gevisualiseerd te worden. Een voorbeeld waar dit is toegepast en waar de gegevens tegelijkertijd gecorrigeerd zijn voor de systematische nulpuntsverschuiving is te vinden in Figuur 19.



Figuur 19: Een 'Quality Control' grafiek met gecontroleerde en gecorrigeerde stijghoogtereeks, als voorbeeld van een reeks waar behoorlijke driftcorrecties op nodig waren. In deze figuur is het volgende te zien:

Onderste grafiek - Voor drift gecorrigeerde druksensorreeks (groen), originele (hergecompenseerde) druksensorreeks (donkergrijs), handmatige controlemetingen (zwarte punten), nulpunt van de druksensor en onderkant filter (rood), uitleesrondes (afwisselende lichte- en middelgrijze achtergrond)
Middelste grafiek - Verschil tussen gecorrigeerde druksensor- en handmatige controlemetingen (in cm)
Bovenste grafiek - Historie van toegepaste druksensoren, met boven het serienummer, daaronder de inhangdiepte, en onder de (geschatte) instrumentdrift.

N.B.: Het gat in de meetreeks is ontstaan doordat defecte drukopnemers niet telkens direct zijn vervangen.

9 Implementatie en aanbevelingen

9.1 Inleiding

Hoe helder en doorwrocht een protocol ook is, als het niet toegepast wordt en verankerd is in de praktijk was de moeite van het opstellen ervan tevergeefs. Om duurzaam gebruik en verankering te verzekeren dient na realisatie van het eigenlijke protocol eveneens gerealiseerd te worden:

- **Draagvlak** – door het bestaan, de opzet en het belang van het protocol te delen met een brede groep belanghebbenden
- **Kennis** - bij de gebruikers van het protocol over de achtergronden en inhoud ervan en hoe het toe te passen
- **Hulpmiddelen** – voor diezelfde gebruikers, zodat zij het protocol eenvoudig en eenduidig toe kunnen passen
- **Regelgeving** - om toepassing van het protocol niet alleen te stimuleren maar waar nodig ook te garanderen

We lichten de beoogde wijze van realisatie van deze punten toe in de volgende paragrafen.

9.2 Implementatie, facilitatie en borging van het gebruik

9.2.1 Communicatie en afstemming met belanghebbenden

Om inbreng van en draagvlak onder belanghebbenden te verzekeren is gekozen voor een begeleidingsstructuur met stuurgroep en bredere begeleidingsgroep met bronhouders en belanghebbenden uit verschillende sectoren en experts. Dit rapport en het eigenlijke protocol zullen gedeeld worden met belanghebbenden via een:

- **Symposium** - een bijeenkomst met de hele begeleidingsgroep, voor het delen van de behaalde resultaten en overwegingen daarbij, en feedback en discussie over vervolgstappen

9.2.2 Verspreiden van kennis onder toepassers

Uit de discussies in de begeleidingsgroep en stuurgroep, maar ook uit de inhoud van dit rapport, komt als consensus naar voren dat er voor het correct beoordelen van de kwaliteit van meetgegevens veel kennis nodig is. Deze kennis betreft enerzijds de bijzonderheden en historie van de meetpunten, anderzijds de hydrologie en hydrologische dynamiek ter plaatse, als ook het gedrag van (fouten van) sensoren en instrumenten. Voor een deel is deze kennis (naast dat ze aanwezig is bij verschillende personen) vastgelegd in dit en andere verschenen rapporten, maar de ervaring leert dat het beschikbaar stellen van schriftelijke kennis alleen vaak onvoldoende effectief is. Om het delen van de beschikbare en ontwikkelde kennis onder de beoogde toepassers van het protocol (databeheerders en hydrologen bij bronhouders en gelieerde partijen) te bevorderen, bevelen we het organiseren van een specifieke cursus of workshop op dit gebied aan.

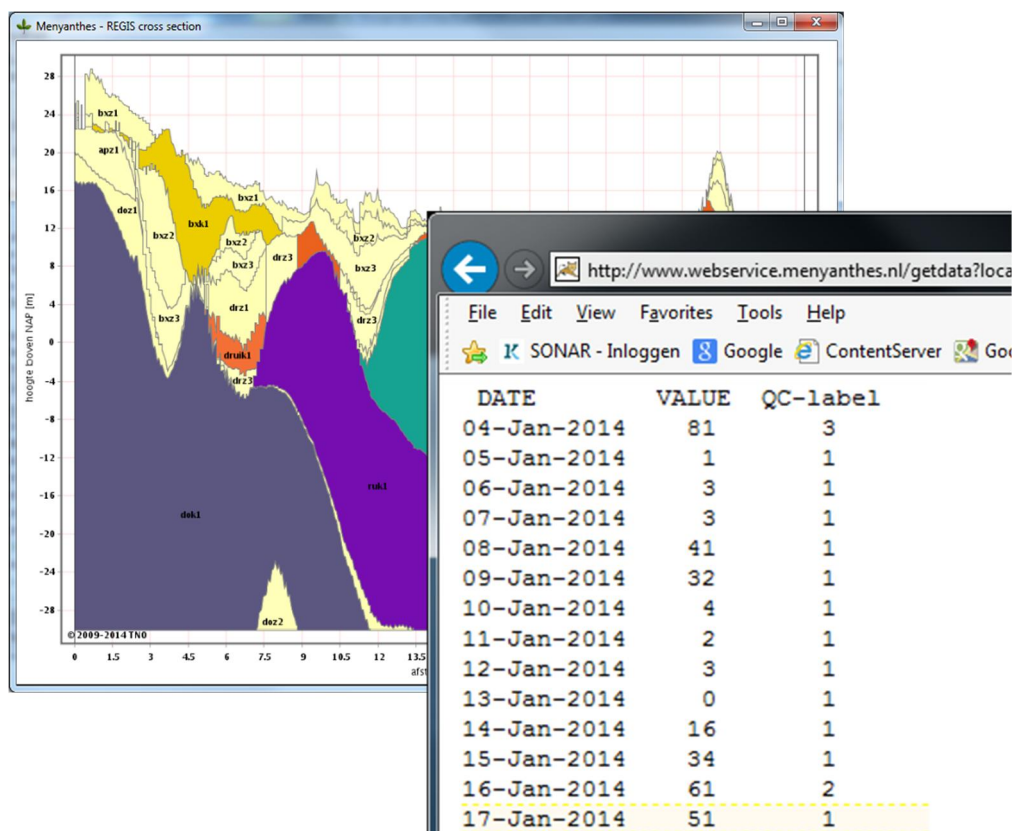
9.2.3 Facilitatie via (softwarematige) hulpmiddelen

Het uitgangspunt en de voorwaarde is dat de databeheerders bij de verschillende bronhouders en partijen elk hun eigen databases en software gebruiken voor de gegevensinwinning en verwerking, en dat ook zullen en moeten kunnen blijven doen. Desalniettemin zijn er ook softwarematige hulpmiddelen nodig om dit protocol eenvoudig en eenduidig toe te kunnen passen, die deels nog gerealiseerd moeten worden of niet vrij beschikbaar zijn. Realisatie van softwarematige hulpmiddelen t.b.v. dit protocol kan als:

- **Centrale, open webservice of component** – die eenvoudig aangeroepen en gebruikt kan worden door (al dan niet lokale) databases en software
- **Lokale (doorgaans gesloten) software** – ofwel inbouw in de verschillende thans gebruikte (commerciële) softwarepakketten afzonderlijk

Realisatie van een webservice of component (*component-based software engineering*) heeft als voordeel dat de functionaliteit hergebruikt kan worden, en toepassing van dit protocol niet noodzakelijkerwijs voor elke individuele partij kosten met zich meebrengt. Centrale implementatie is daarnaast eenduidig (er is maar één versie). Implementatie als webservice ligt mede voor de hand gezien de rol die webservices in de BRO spelen. De praktijk is echter nu dat er veel verschillende gesloten en/of lokale systemen zijn, zowel voor databeheer als datakwaliteitscontrole.

De verdere realisatie van de BRO heeft hoe dan ook een belangrijke relatie met het protocol, reden waarom hier bijzondere aandacht besteed is aan de BRO. De BRO zal naar verwachting zelf eisen stellen aan de interne data-integriteit, waarmee delen van dat onderdeel door de



Figuur 20: Voorbeelden van eenvoudige webservices voor uitwisseling van figuren (REGIS webservice van TNO) en data en eventueel kwaliteitslabels (t.b.v. het voorbeeld aangepaste webservice van KWR).

BRO afgedekt zullen worden. Ook de automatisering van de data-uitwisseling met de BRO via webservices zal de hoeveelheid fouten daarbij drastisch reduceren. Desalniettemin zullen integriteitsproblemen in de lokale data fouten kunnen blijven veroorzaken. Alhoewel het datamodel van de BRO voor grondwaterstands- en stijghoogtemetingen zelf nog niet gerealiseerd is, zal het wellicht mogelijk gemaakt worden om daarbij kwaliteitslabels op te nemen.

9.2.4 Borging van het gebruik

Het stimuleren en faciliteren van de toepassing van het protocol alleen is gezien vanuit bijvoorbeeld de Kaderrichtlijn Water (KRW) onvoldoende. De KRW vereist dat de verzamelde gegevens niet naar eigen inzicht en behoefte gecontroleerd worden, maar standaard en volgens dezelfde, helder gedefinieerde regels.

Het protocol zal in definitieve vorm ter vaststelling worden voorgelegd aan de Landelijke Werkgroep Grondwater en/of de ambtelijke adviescommissie duurzame ruimtelijke ontwikkeling en waterbeheer (AAC DROW) van het Interprovinciaal Overleg (IPO). Er zal bovendien een verkenning plaatsvinden naar de opties om het te verankeren in beleid of wetgeving, bijvoorbeeld door het op te nemen als bijlage in het Draaiboek Monitoring, dat onderdeel uitmaakt van het BKMW (Besluit kwaliteitseisen en monitoring water, 2009).

Het verdient overweging om dit protocol in een vervolproces de vorm en status van een NEN-norm te geven.

9.3 Aanbevelingen

Naast aanbevelingen die betrekking hebben op de implementatie van het protocol en die hierboven aan de orde zijn geweest, volgen uit het doorlopen proces ook de volgende, gerelateerde aanbevelingen:

- **Kwaliteitsvoorschrift (QA)** – Alhoewel er wel handboeken beschikbaar zijn met richtlijnen voor het inrichten van peilbuizen en verrichten van metingen (bijv. het handboek van de STOWA (Bouma e.a., 2012) en het kennisdocument van de waterleidingbedrijven (Leunk, 2011)) heeft geen daarvan een formele of wettelijke status. Het is aan te bevelen dat de provincies onderzoeken en vaststellen welke richtlijnen zij volgen bij hun (KRW) monitoring.
- **Geautomatiseerde plausibiliteitscontrole** – Uit de validatiepilot (Leunk, 2014) bleek dat er aandacht nodig is voor het verbeteren van de effectiviteit van geautomatiseerde plausibiliteitscontrole of *outlier*-detectie. Zoals in hoofdstuk 7 behandeld is, kan geautomatiseerde *outlier*-detectie echter ook op principiële bezwaren stuiten. Nader onderzoek naar de effectiviteit en legitimiteit van de techniek verdient aanbeveling, vanwege het gemak, de uniformiteit en de mogelijkheid om daarmee grotere datahoeveelheden te verwerken. Een onderzoek in die richting is in gang gezet door de provincie Overijssel.
- **Kwaliteitscontrole en –verbetering van bestaande data** – De in dit rapport beschreven systematiek en methodiek zijn in beginsel generiek, maar met name het onderdeel sensordata en de aanleiding vanuit de KRW maken dat de toepassing ervan vooral nieuw ingewonnen data zal zijn. Voor data geldt echter net als voor historische bebouwing of kunst dat de waarde ervan gerelateerd is aan leeftijd en onvervangbaarheid. Het verdient dus aanbeveling om ook de kwaliteitscontrole en –verbetering van historische of bestaande data ter hand te nemen (wat verschillende individuele bronhouders al doen), om de kwaliteit, bruikbaarheid en waarde van de data in DINO (en straks die met IMBRO/A status in de BRO) verder te verbeteren.

- **Verdere afstemming met de BRO en andere softwaresystemen** – Vanwege de nog in gang zijnde realisatie van de BRO was alleen afstemming met BRO onderdeel ‘Grondwatermonitoringput’ mogelijk. Afstemming met de nog komende volgende onderdelen van de BRO en andere (lokale) softwaresystemen is aan te bevelen.
- **Monitoren van brak of zout grondwater** – hieraan zijn specifieke kwesties verbonden die betrekking hebben op zowel de meetopstelling, de verwerking van gemeten drukken tot waterpeilen als de vertaling van waterpeilen naar stijghoogtes. Deze zijn in het kader van dit project niet meegenomen, maar het verdient aan te bevelen dat alsnog te doen.
- **Kwantificering van de nauwkeurigheid** – de KRW stelt niet alleen als voorschrift dat de kwaliteit van verzamelde gegevens gecontroleerd wordt (waartoe dit protocol dient), maar ook dat deze onderzocht, gekwantificeerd en gerapporteerd wordt. Voor een deel zijn er projecten en initiatieven geweest op dit gebied (zie bijv. Von Asmuth, 2011, Knotters e.a., 2013), voor een ander deel is onderzoek in gang gezet (bijv. onderzoek naar de nauwkeurigheid van sensordata van Keller door de provincie Overijssel). Voor andere onderdelen is hier nog niet in voorzien, en het verdient aanbeveling om ook dat te doen.

Literatuur

- Altman, N.S. (1992)** An introduction to kernel and nearest-neighbor nonparametric regression; in: *The American Statistician*, jrg 46, nr 3, pag 175-185. doi:10.1080/00031305.1992.10475879.
- Bouma, J., M. Maasbommel en I. Schuurman (2012)** Handboek meten van grondwaterstanden in peilbuizen; rapportnr. 2012-50 STOWA, Amersfoort.
- Broekhuis, R., I. Van Erp, D. Lue, B. Van Loon, M. Van de Merwe, R. Veldman en L. Van Hamersveld (2014)** Toekomst perspectief 'screeningtool' Royal Eijkkelkamp; Royal Eijkkelkamp, Nationaal Watertraineeship, Giesbeek.
- Chauvenet, W. (1960)** A Manual of Spherical and Practical Astronomy V. II.; Reprint of 1891. 5th ed. Dover, N.Y.
- Cook, R.D. en S. Weisberg (1982)** Residuals and Influence in Regression.; Chapman and Hall., New York.
- De Meij, T. en J. R. Von Asmuth (2011)** Correctie van eigen luchtdrukmetingen is noodzakelijk; in: *H2O*, jrg 4, pag 29-32.
- European Communities (2003)** Common Implementation Strategy For The Water Framework Directive (2000/60/EC), Monitoring under the Water Framework Directive; WFD CIS Guidance Document No. 7, Working Group 2.7 - Monitoring, Office for Official Publications of the European Communities, Luxembourg.
- European Communities (2007)** Common Implementation Strategy For The Water Framework Directive (2000/60/EC), Guidance on Groundwater Monitoring; WFD CIS Guidance Document No. 15, Office for Official Publications of the European Communities, Luxembourg.
- Faber, W., D. Wielakker, A. Bak, J. L. Spier en C. Smulders (2011)** Richtlijn KRW Monitoring Oppervlaktewater en Protocol Toetsen & Beoordelen; Eindrapport vastgesteld in het Directeuren Water Overleg (DWO), Rijkswaterstaat, Ministerie van Infrastructuur en Milieu, Lelystad.
- Goering, T. (2009)** Groundwater level data processing and validation, Standard Operating Procedure; SOP-5230, Waste and Environmental Service Division, Los Alamos National Laboratory,
- Grubbs, F.E. (1950)** Sample criteria for testing outlying observations; in: *The Annals of Mathematical Statistics*, jrg doi: 10.1214/aoms/1177729885, nr 1, pag 27-58.
- Hampel, F.R., E. M. Ronchetti, P. J. Rousseeuw en W. A. Stahei (1986)** Robust Statistics: The Approach Based on Influence Functions.; Wiley, New York.
- Jansen, A.J.M., J. R. Von Asmuth, P. J. T. Van Bakel, E. Brouwer, R. J. Ketelaar en R. L. Terhürne (2013)** De Engbertsdijksvenen: advies van de Commissie van Deskundigen; Kennisnetwerk OBN, Driebergen.
- Keller BV (2010)** Maak kennis met de nieuwe KELLER DCX niveau-, overstort- en neerslag dataloggers; Brochure KELLER Meettechniek BV, Reeuwijk, Nederland.
- KNMI (2000)** Handboek Waarnemingen; Koninklijk Nederlands Meteorologisch Instituut, De Bilt.
- KNMI (2012)** Het weer, een lessenserie over meteorologie; Koninklijk Nederlands Meteorologisch Instituut, De Bilt.
- Knotters, M., T. De Meij en M. Pleijter (2013)** Nauwkeurigheid van handmatig gemeten grondwaterstanden en stijghoogtes. Verslag van een veldexperiment.; Alterra Wageningen UR, Wageningen.
- Knotters, M., S. P. J. Van Delft, H. E. Keizer-Vlek, J. R. Von Asmuth, P. C. Jansen, F. P. Sival en C. E. Van 't Klooster (2008)** Evaluatie monitoring

- Deurnese Peel en Mariapeel. Kwantificering van effecten van maatregelen en advies over het monitoringplan; Alterra-Document2, Alterra, Wageningen.
- Leunk, I. (2011)** Kennisdocument Putten(velden), Ontwerp, aanleg en exploitatie van pomp- en waarnemingsputten; rapportnr. KWR 2012.014, KWR, Nieuwegein
- Leunk, I. (2014)** Kwaliteitsborging grondwaterstands- en stijghoogtegegevens. Validatiepilot; analyse van bestaande data.; rapportnr. KWR 2014.059, KWR Watercycle Research Institute, Nieuwegein.
- Leunk, I. en J. R. von Asmuth (2013)** Hervalidatie drukopnemerdata Noardburgum; rapportnr. KWR 2013.054, KWR Watercycle Research Institute, Nieuwegein.
- McLaughlin, D.L. en M. J. Cohen (2011)** Thermal artifacts in measurements of fine-scale water level variation; in: *Water Resources Research*, jrg W09601, nr 47, doi:10.1029/2010WR010288.
- Post, V.E.A. en J. R. Von Asmuth (2013)** Hydraulic head measurements: New technologies, classic pitfalls; in: *Hydrogeology journal*, DOI 10.1007/s10040-013-0969-0.
- Sorensen, J.P.R. en A. S. Butcher (2011)** Water Level Monitoring Pressure Transducers-A Need for Industry-Wide Standards; in: *Ground Water Monitoring & Remediation*, doi: 10.1111/j.1745-6592.2011.01346.x, pag 1-7.
- Van der Meij, J.L. en G. J. Van der Meer (2010)** Kwaliteitsborging van meetgegevens van het grondwater in DINO; rapport nr. TNO-034-UT-2010-00635, TNO Bouw en Ondergrond, Utrecht.
- Van Geer, F.C. (2012)** Validatie en plausibiliteit stijghoogte observaties; TNO, Utrecht.
- Van Geer, F.C. (2013)** Eenvoudige toetsing en visualisatie van veranderingen in het karakter van een stijghoogtereeks; in: *Stromingen*, jrg 19, nr 3&4, pag 21-34.
- Van Herpen, F., O. van Tongeren, R. Knoben, P. Baggelaar en W. van Loon (2009)** Quickscan precisie en betrouwbaarheid KRW-monitoringsprogramma's; rapport 9V0539/R00002/902795/AH/DenB in opdracht van Rijkswaterstaat Waterdienst, Royal Haskoning.,
- Van Vliet, M. (niet gepubliceerd)** Validatie van grondwaterkwaliteitsgegevens, bijlage 1 bij rapportage project 1 fase 1 Programma van Eisen vptp; RIVM, Bilthoven.
- Von Asmuth, J.R. (2011)** Over de kwaliteit, frequentie en validatie van druksensorreeksen; Rapportnr. KWR 2010.001, KWR Watercycle Research Institute, Nieuwegein.
- Von Asmuth, J.R., C. Maas, M. Knotters, M. F. P. Bierkens, M. Bakker, T. N. Olsthoorn, D. G. Cirkel, I. Leunk, F. Schaars en D. C. Von Asmuth (2012)** Software for hydrogeologic time series analysis, interfacing data with physical insight; in: *Environmental Modelling & Software*, jrg 38, pag 178-190, <http://dx.doi.org/10.1016/j.envsoft.2012.06.003>.
- Von Asmuth, J.R. en F. C. Van Geer (2013)** Kwaliteitsborging grondwaterstands- en stijghoogtegegevens: op weg naar een landelijke standaard; rapportnr. KWR 2013.027, KWR Watercycle Research Institute / TNO, Nieuwegein / Utrecht.
- Von Asmuth, J.R. en I. Leunk (2014)** Hydrologisch meetnet Electraboezem 2e schil, controle en verbetering van de datakwaliteit; Rapportnr KWR 2014.053, KWR Watercycle Research Institute, Nieuwegein.
- Walvoort, D., M. Knotters en T. Hoogland (2013)** Een tool voor controle van hoogfrequente grondwaterstandsreeksen; in: *Stromingen*, jrg 19, nr 3&4 , pag 49-61.

Bijlage A: Data-integriteitscontrole

Klasse	Attribuut	BRO-term	Data-integriteit		
			Format	Verplicht	Uniek
Peilbuiscluster		Grondwatermonitoringput			
(Administratie)	BRO-ID	BRO-ID	numeriek	v	v
	NITG-code	(naar keuze:) Historische naam	tekst	v	v
			geldig kaartblad		
	OLGA-code	(naar keuze:) Historische naam	tekst		v
	Bronhouder-code	Object-ID bronhouder	tekst	v	
	Naam bronhouder	Bronhouder	tekst	v	
	Doelstelling / beleidskader	Kader aanlevering	vaste opties	v	
	Afwerking	Beschermconstructie	vaste opties	v	
(Historie)	Startdatum	Datum ingericht	datum+tijdstip	v	
	Einddatum	Datum opgeruimd	datum+tijdstip	v	
(Processen)	Inundatie	-	boolean		
	Bodemdaling	-	boolean		
	Overige	-	tekst		
Peilbuis		Monitoringbuis			
(Administratie)	Filternummer	Buisnummer	integer	v	
	Plaatsing door (persoon)	-	tekst		
	Plaatsing door (instantie)	-	tekst		
(Historie)	Startdatum	<<dynamisch>>	datum+tijdstip	v	
	Einddatum	<<dynamisch>>	datum+tijdstip	v	
(Coördinaten)	XY-coördinaat	Locatie	numeriek	v	i.c.m. filter
	Maaiveldshoogte	Positie maaiveld	numeriek	v	
	Bovenkant buis	Stijgbuis: positie bovenkant	numeriek	v	
	Bovenkant filter	Filter: positie bovenkant	numeriek	v	i.c.m. XY
	Onderkant filter	Filter: positie onderkant	numeriek	v	i.c.m. XY
	Onderkant buis (Zandvang)	Zandvang: positie onderkant	numeriek	v	
(Constructie)	Drukdpod / Artesisch (ja/nee)	Drukdpod	boolean	v	
	Diameter buis	Interne diameter	numeriek		
	Materiaal buis	Materiaal	tekst		
	Afdichting boorgat	-	tekst		
	Omstorting	-	tekst		
Logger / sensor		?			
	Logger-ID / Serienummer	?	afh. van fabrikant	v	i.c.m. datum
	Startdatum (inhangen)	?	datum+tijdstip	v	
	Einddatum (verwijderen)	?	datum+tijdstip	v	
	Inhanddiepte logger	?	numeriek	v	
	Instrument type	?	vaste opties	v	
	Batterijstatus	?	numeriek		
	Fabricagedatum	?	datum		
	Fabrikant	?	tekst	v	
Loggermeting		?			
	Uitgelezen door (persoon)	?	tekst		
	Uitgelezen door (instantie)	?	tekst		
	Loggermeting waarde	?	numeriek	v	
	Loggermeting tijd	?	datum+tijdstip	v	v
	Compensatiestatus	?	boolean	v	
	Eigen luchtdrukreeks	?	tijdreeks	v, tenzij KNMI reeks	
	KNMI luchtdrukstation	?	vaste opties	v, tenzij eigen reeks	
	Waterdichtheid	?	numeriek	v	
	Gravitatieconstante	?	numeriek	v	
Handmeting		?			
(al dan niet ter controle)	Meting door (persoon)	?	tekst	v	
	Meting door (instantie)	?	tekst	v	
	Meetinstrument	?	vaste opties	v	
	Handmeting waarde	?	numeriek	v	
	Handmeting tijd	?	datum+tijdstip	v	v
Dataverwerking		?			
	Persoon	?	tekst	v	
	Instantie	?	tekst	v	
	QC methodiek	?	vaste opties	v	
	Aantal controles		54	39	12
	Totaal aantal controles		105	(Niet allemaal uniek)	

Bijlage B: Consistentiecontrole

Klasse	Attribuut	BRO-term	Consistentie			
			<	<=	>=	>
Peilbuiscluster		Grondwatermonitoringput				
(Administratie)	BRO-ID	BRO-ID				
	NITG-code	(naar keuze:) Historische naam				
	OLGA-code	(naar keuze:) Historische naam				
	Bronhouder-code	Object-ID bronhouder				
	Naam bronhouder	Bronhouder				
	Doelstelling / beleidskader	Kader aanlevering				
	Afwerking	Beschermconstructie				
(Historie)	Startdatum	Datum ingericht				
	Einddatum	Datum opgeruimd				
(Processen)	Inundatie	-				
	Bodemdaling	-				
	Overige	-				
Peilbuis		Monitoringbuis				
(Administratie)	Filternummer	Buisnummer			1	
	Plaatsing door (persoon)	-				
	Plaatsing door (instantie)	-				
(Historie)	Startdatum	<<dynamisch>>			Vorige startdatum	
	Einddatum	<<dynamisch>>		Volgende startdatum		
(Coördinaten)	XY-coördinaat	Locatie	max (XY) Nederland Vorige XY + X meter			min(XY) Nederland Vorige XY - X meter
	Maaiveldshoogte	Positie maaiveld	AHN2 + 50 cm			AHN2 - 50 cm
	Bovenkant buis	Stijgbuis: positie bovenkant				Onderkant filter
	Bovenkant filter	Filter: positie bovenkant	Bovenkant buis Vorige BF+X cm			Bovenkant filter Onderkant filter
	Onderkant filter	Filter: positie onderkant	Bovenkant filter		Onderkant buis	Vorige BF - X cm
	Onderkant buis (Zandvang)	Zandvang: positie onderkant			Onderkant filter	
(Constructie)	Drukdp / Artesisch (ja /nee)	Drukdp				
	Diameter buis	Interne diameter				0
	Materiaal buis	Materiaal				
	Afdichting boorgat	-				
	Omstorting	-				
Logger / sensor		?				
	Logger-ID / Serienummer	?				
	Startdatum (inhangen)	?	Einddatum buis		Startdatum buis	
					Einddatum logger elders	
	Einddatum (verwijderen)	?		Startdatum logger elders		
	Inhangdiepte logger	?	Bovenkant buis			Onderkant buis
	Instrument type	?				
	Batterijstatus	?				0
	Fabricagedatum	?	Startdatum			
	Fabrikant	?				
Loggermeting		?				
	Uitgelezen door (persoon)	?				
	Uitgelezen door (instantie)	?				
	Loggermeting waarde	?		Bovenkant buis (tenzij drukdp of inundatie)	Inhangdiepte logger	
	Loggermeting tijd	?		Verwijderdatum logger	Inhangdatum logger	
	Compensatiestatus	?				
	Eigen luchtdrukreeks	?				
	KNMI luchtdrukstation	?				
	Waterdichtheid	?		Minimum	Maximum	
	Gravatieconstante	?		Minimum	Maximum	
Handmeting		?				
(al dan niet ter controle)	Meting door (persoon)	?				
	Meting door (instantie)	?				
	Meetinstrument	?				
	Handmeting waarde	?		Bovenkant buis (tenzij drukdp of inundatie)	Onderkant filter	
	Handmeting tijd	?		Einddatum	Startdatum	
Dataverwerking		?				
	Persoon	?				
	Instantie	?				
	QC methodiek	?				
	Aantal controles		9	9	11	10
	Totaal aantal controles		39	(niet allemaal uniek)		

Bijlage C: Begeleidingsgroepleden

Dit rapport is tot stand gekomen onder begeleiding van het platform meetnetbeheerders van de gezamenlijke provincies (onderdeel grondwaterkwantiteit), waarin zitting hebben:

- Jan Meijles (Provincie Zuid-Holland)
- Henny Kempen (Provincie Gelderland)
- Janco van Gelderen (Provincie Utrecht)
- Jolanda Bauwens (Provincie Brabant)
- Marja Segers (Provincie Brabant)
- Jean Hacking (Provincie Limburg)
- Lester Reiniers (Provincie Noord-Holland)
- Nanko de Boorder (Provincie Noord-Holland)
- Johan Wortelboer (Provincie Noord-Holland)
- Geert Jan Steenbergen (Provincie Groningen)
- Bert Luinge (Provincie Drenthe)
- Anne Venema (Provincie Friesland)
- Thomas de Meij (Provincie Overijssel)
- Ronnie Hollebrandse (Provincie Zeeland)
- Bart Hamer (OFGV Flevoland)
- Jack van Velthuisen (OMWB Noord Brabant)
- Erik Simmelink (TNO)
- Ton Ebbing (Vitens)

Bijlage D: Stuurgroepleden

De volgende personen uit verschillende sectoren hadden zitting in de stuurgroep:

- Henny Kempen (Provincie Gelderland)
- Janco van Gelderen (Provincie Utrecht)
- Jan Meijles (Provincie Zuid-Holland)
- Jolanda Bauwens, Marja Segers (Provincie Brabant)
- Lester Reiniers, Nanko de Boorder (Provincie Noord-Holland)
- Erik Simmelink (TNO)
- Gert van den Houten (Waterschap Rijn en IJssel)
- Ate Oosterhof (Waterleidingbedrijf Vitens)
- Cor Beets (Staatsbosbeheer)