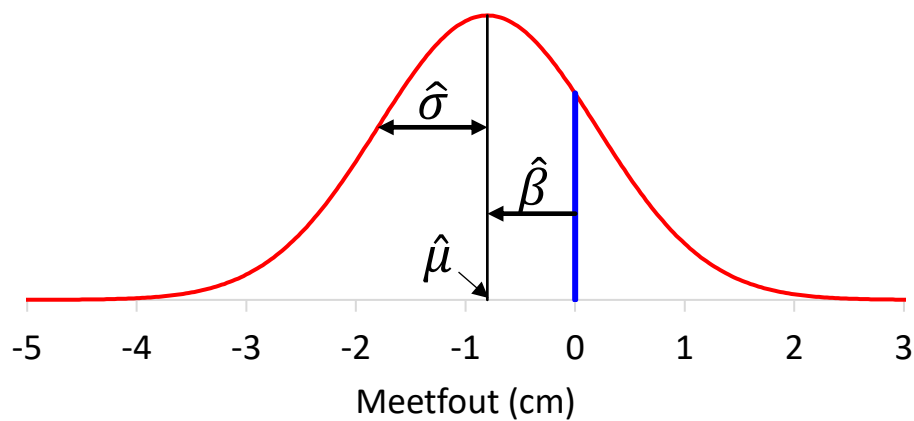


Kenmerken van de meetfout van gevalideerde loggermetingen van de grondwaterstand van het meetnet van de Provincie Overijssel

Kansverdeling meetfout



Kenmerken van de meetfout van gevalideerde loggermetingen van de grondwaterstand van het meetnet van de Provincie Overijssel

Aanvrager:
Provincie Overijssel

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

Met medewerking van:
ing. René Buijs (*Hydro-ecologisch Onderzoek & Advies*)

13 december 2021



AMO (Adviesbureau Modelling en Optimalisatie)
Deldenerstraat 81
7551 AC Hengelo
Tel.: 074-7850124 / 06-24865321
Email : amo@amo-nl.com

INHOUD

SAMENVATTING.....	2
1 INLEIDING.....	3
1.1 ACHTERGRONDEN	3
1.2 VRAAGSTELLING VAN DEZE STUDIE	4
1.3 HOOFDLIJNEN VAN DE AANPAK.....	4
1.4 OVER DIT RAPPORT	4
2 KENMERKEN MEETFOUT PROTOCOL PROVINCIE OVERIJSEL	5
2.1 THEORETISCH SCHATTEN KENMERKEN MEETFOUT.....	5
2.2 EMPIRISCH SCHATTEN KENMERKEN MEETFOUT VÓÓR VALIDATIE	5
2.3 EMPIRISCH SCHATTEN KENMERKEN MEETFOUT NÁ VALIDATIE	7
3 OPZET STEEKPROEF HANDMETINGEN.....	8
3.1 KEUZE STEEKPROEFGROOTTE	8
3.2 SELECTEREN VAN PEILFILTERS VOOR DE STEEKPROEF	9
3.3 OVERDRACHT GEGEVENS EN INSTRUCTIES AAN UITVOERDER STEEKPROEF	9
4 STATISTISCHE VERWERKING STEEKPROEFRESULTATEN	11
4.1 AFLEIDEN VERSCHILWAARDEN L1-SH EN L0-SH	11
4.2 VERDELINGEN VERSCHILWAARDEN L1-SH EN L0-SH	12
4.3 KENMERKEN VERSCHILWAARDEN L1-SH EN L0-SH.....	14
4.4 AFLEIDEN KENMERKEN MEETFOUT GEVALIDEERDE LOGGERMETING.....	16
4.5 JUISTHEID EN PRECISIE VAN EEN GEVALIDEERDE LOGGERMETING	20
4.6 CONCLUSIES	21
AANGEHAALDE LITERATUUR.....	22
BIJLAGE: LIJST VERSCHILWAARDEN L1-SH EN L0-SH	23

Samenvatting

Op verzoek van de Provincie Overijssel zijn kenmerken geschat van de meetfout van gevalideerde loggermetingen van de grondwaterstand van het Provinciaal meetnet, op basis van een steekproef van willekeurige handmetingen die niet zijn gebruikt bij het valideren van de loggermetingen.

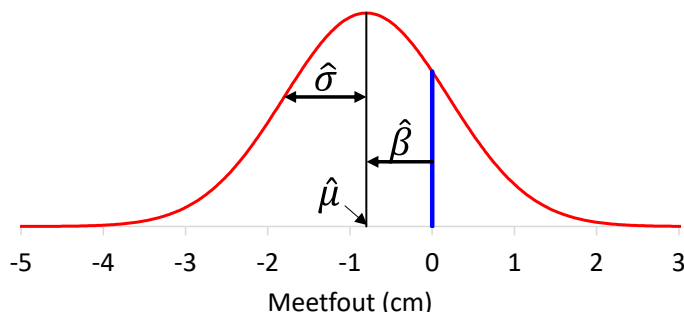
Het betrof een steekproef van 200 handmetingen van de grondwaterstand, die geheim zijn gehouden voor de Provincie Overijssel en dus niet konden worden gebruikt om loggermetingen te valideren. Deze handmetingen zijn later vergeleken met de daaraan gepaard te achten gevalideerde loggermetingen. Gemiddelde en standaardafwijking van de resulterende set verschilwaarden zijn vervolgens gecorrigeerd voor gemiddelde en standaardafwijking van de meetfout van een handmeting (zoals met een proefopzet geschat in een andere studie).

Op basis van deze studie geschat gemiddelde en standaardafwijking van de meetfout van een gevalideerde loggermeting (blauw), inclusief de onder- en bovengrenzen van het 95%-betrouwbaarheidsinterval van die schattingen.

Kenmerken meetfout	Gemiddelde	Standaardafw
Schatting (cm)	-0,81	0,96
Ondergrens 95%-b.i. (cm)	-1,05	0,33
Bovengrens 95%-b.i. (cm)	-0,57	1,49

Weergave van de in deze studie geschatte dimensies van de kansverdeling van de meetfout van een door de Provincie Overijssel gevalideerde loggermeting van de grondwaterstand.

Kansverdeling meetfout



Toelichting

$\hat{\mu}$: geschatte gemiddelde meetfout (hier -0,8 cm)

$\hat{\beta}$: geschatte systematische meetfout (hier -0,8 cm, omdat $\hat{\beta} = \hat{\mu} - 0$)

$\hat{\sigma}$: geschatte standaardafwijking van de meetfout (hier 1,0 cm)

Uit deze verkenning bleek het volgende over de kansverdeling van de meetfout van een door de Provincie Overijssel gevalideerde loggermeting van de grondwaterstand:

- de gemiddelde meetfout is geschat op -0,8 cm, maar gegeven de onzekerheden kan deze met 95% betrouwbaarheid ergens in het interval van -1,0 tot -0,6 cm liggen. Er is daarmee met 95% betrouwbaarheid sprake van een statistisch significante negatieve systematische meetfout;
- de standaardafwijking van de meetfout is geschat op 1,0 cm, maar gegeven de onzekerheden kan deze met 95% betrouwbaarheid ergens in het interval van 0,3 tot 1,5 cm liggen. Dit houdt in dat het niet aannemelijk is dat er vaak toevallige meetfouten groter dan +3 cm of kleiner dan -3 cm optreden (bovenop de bovengenoemde systematische meetfout).

1 Inleiding

1.1 Achtergronden

Provincie Overijssel beheert een meetnet van ongeveer 500 peilbuizen om toestand en ontwikkeling van de grondwaterstand en stijghoogte¹ in de provincie te kunnen bepalen. Oorspronkelijk werden de metingen handmatig uitgevoerd met een peilklokje aan een meetlint, op min of meer halfmaandelijke basis (op of rond de 14^e en de 28^e). Maar vanaf eind 2003 is het meetnet geleidelijk voorzien van automatische drukopnemers (in het nu volgende aangeduid als *loggers*), die de grondwaterstand eens per dag of per uur meten. Deze automatisering heeft als voordelen dat er mankracht wordt bespaard en dat er makkelijker hoogfrequent kan worden gemeten. Maar een nadeel is dat het meten complexer en daarmee foutgevoeliger is geworden. Er is immers geen sprake meer van een eenvoudige handmeting van de grondwaterstand met een peilklokje aan een meetlint, maar van het simultaan met twee drukopnemers meten van de luchtdruk en het totaal van luchtdruk en waterdruk. Uit het verschil van deze drukmetingen wordt dan door de logger de grondwaterstand ten opzichte van een bepaald referentieniveau afgeleid.

De praktijk heeft geleerd dat het automatisch meten van de grondwaterstand met een drukopnemer regelmatig valideren² vergt, om te voorkomen dat de meetfouten te groot worden (zie bijvoorbeeld [Von Asmuth, 2010, Post and Von Asmuth, 2013, Pleijter et al., 2015, Von Asmuth en Van Geer, 2015] en [Bagelaar en Van der Meulen, 2019]).

Protocol Provincie Overijssel voor meten grondwaterstand met logger

De Provincie Overijssel hanteert al vrijwel vanaf het begin een protocol, waarbij de automatische metingen van een logger regelmatig - doorgaans circa halfjaarlijks - worden gevalideerd op basis van vergelijking met een controlehandmeting, indien nodig (bij relevante afwijking) gevolgd door correctie [Gooijer et al., 2018]. Tot 2010 werd dat protocol ondersteund door de loggersoftware en daarna zijn enkele onderdelen overgenomen door de module HydroData van het programma *Menyanthes* [Von Asmuth et al., 2010].

In een recente studie zijn kenmerken geschat van de meetfout van dat protocol van de Provincie Overijssel [Bagelaar en Van der Meulen, 2019]. De resultaten daarvan zijn samengevat in hoofdstuk 2. Bij die studie kon echter nog geen standaardafwijking worden geschat van de meetfout van gevalideerde loggermetingen, omdat daarvoor handmetingen nodig zijn die niet zijn gebruikt bij het valideren. Provincie Overijssel heeft ons daarom verzocht een steekproef van willekeurige handmetingen te verzamelen en daarmee de kenmerken van de meetfout van gevalideerde loggermetingen te schatten. Dit rapport beschrijft aanpak en bevindingen van die studie.

¹ Voor de leesbaarheid wordt in het nu volgende verder geen onderscheid gemaakt tussen *grondwaterstand* en *stijghoogte* en worden beide aangeduid als *grondwaterstand*.

² Onder *valideren* van metingen wordt in strikte zin verstaan controleren of wordt voldaan aan eisen (zoals nauwkeurigheid). Maar omdat in het meetproces van de Provincie Overijssel het valideren van metingen bij gebleken relevante onnauwkeurigheid gepaard gaat met corrigeren, zullen we onder gevalideerde metingen impliciet ook verstaan gevalideerde en zonodig gecorrigeerde metingen. Het Protocol voor Datakwaliteitscontrole [Von Asmuth, 2018] stelt overigens de aanduiding ‘controleren en/of corrigeren’ voor als alternatief voor ‘valideren’.

1.2 Vraagstelling van deze studie

De hoofdvraag van deze studie is kenmerken te schatten van de meetfout van gevalideerde loggermetingen van de grondwaterstand van het meetnet van de Provincie Overijssel, op basis van een steekproef van willekeurige handmetingen die niet zijn gebruikt bij het valideren van de loggermetingen.

1.3 Hoofdpijnen van de aanpak

Om de kenmerken van de meetfout van gevalideerde loggermetingen te kunnen schatten is een steekproef verzameld van 200 handmetingen van de grondwaterstand, die geheim zijn gehouden voor de Provincie Overijssel en dus niet konden worden gebruikt om loggermetingen te valideren. Deze handmetingen zijn later vergeleken met de daaraan gepaard te achten gevalideerde loggermetingen. Gemiddelde en standaardafwijking van de resulterende set verschilwaarden zijn vervolgens gecorrigeerd voor gemiddelde en standaardafwijking van de meetfout van een handmeting.

Het ontwerpen van de steekproef en het statistisch verwerken van zijn gegevens zijn uitgevoerd door Eit van der Meulen (AMO) en Paul Baggelaar (PB Icastat). De 200 handmetingen zijn uitgevoerd door René Buijs (*Hydro-ecologisch Onderzoek & Advies*), in de periode 23 april 2019 t/m 6 mei 2020.

1.4 Over dit rapport

Na deze inleiding beschrijft hoofdstuk 2 resultaten van de voorgaande studie [Baggelaar en Van der Meulen, 2019], waarin op verschillende wijzen kenmerken zijn geschat van de meetfout van de grondwaterstand zoals met een logger gemeten door de Provincie Overijssel. Hoofdstuk 3 beschrijft de opzet van onze steekproef van 200 handmetingen van de grondwaterstand. En hoofdstuk 4 beschrijft aanpak en bevindingen van het statistisch verwerken van die steekproef tot schattingen van de kenmerken van de meetfout van gevalideerde loggermetingen. Het hoofddeel van dit rapport sluit af met de alfabetische lijst van de aangehaalde literatuur.

2 Kenmerken meetfout protocol Provincie Overijssel

Dit hoofdstuk beschrijft de relevantste resultaten van de voorgaande studie [Bagelaar en Van der Meulen, 2019], waarin op verschillende wijzen kenmerken zijn geschat van de meetfout van het protocol dat de Provincie Overijssel hanteert bij het meten van de grondwaterstand met een logger. Dit zal helpen om de resultaten van de steekproef beter te kunnen plaatsen.

In de betreffende studie zijn die kenmerken zowel *theoretisch* geschat, aan de hand van technische specificaties van de gebruikte logger, als *empirisch*, uit verschillen tussen ruwe loggermetingen en bijbehorende handmetingen en uit verschillen tussen gevalideerde loggermetingen en bijbehorende handmetingen. Deze onderwerpen worden hieronder kort beschreven in § 2.1, 2.2 en 2.3.

Codering van het soort meting

Bij onze codering van het soort meting staat C voor controlemeting, L voor loggermeting, H voor handmeting, 0 voor ruwe meting en 1 voor gevalideerde (en zonodig gecorrigeerde) meting. De controleloggermeting (CLO) wordt ook wel aangeduid als de online-loggermeting.

2.1 Theoretisch schatten kenmerken meetfout

Aan de hand van technische specificaties van de door de Provincie gebruikte logger (Keller DCX-22 AA) is langs theoretische weg de standaardafwijking van zijn meetfout geschat, voor de situatie dat de logger nog geen kwaliteitsverslechtering vertoont. Deze blijkt dan tussen 0,8 en 1,3 cm waterhoogte te kunnen liggen, afhankelijk van de interpretatie van die technische specificaties.

2.2 Empirisch schatten kenmerken meetfout vóór validatie

De kenmerken van de meetfout van een ruwe loggermeting (dus vóór validatie en zonodige correctie) zijn empirisch geschat uit 3.665 verschillen tussen ruwe loggermetingen (LO) en de daaraan gepaarde controle-handmetingen (CH0), die zijn vastgesteld bij het circa halfjaarlijks uitlezen van een logger volgens het meetprotocol van de Provincie (zie tabel 2.1). Het gemiddelde verschil van LO en CH0 blijkt -0,39 cm en de standaardafwijking van het verschil blijkt 3,63 cm. Na verdiscontering van kenmerken van de meetfout van CH0³ wordt het gemiddelde van de ruwe meetfout van de logger geschat op 0,11 cm en de standaardafwijking op 3,53 cm.

Deze kengetallen hebben betrekking op een verzameling loggers die heterogeen is voor wat betreft de kwaliteit van de loggermeting, omdat ze zijn afgeleid uit verschilwaarden van een groot aantal loggers, die verschillende leeftijden hebben en die ook onder verschillende omstandigheden functioneren, zodat ze ook zullen verschillen in kwaliteit van de ruwe loggermeting. De oudere loggers bijvoorbeeld zullen in veel gevallen drift vertonen, wat de standaardafwijking van de verschilwaarde LO-CH0 zal hebben vergroot. Dit alles beperkt de bruikbaarheid van de afgeleide kengetallen.

³ Voor de meetfout van een handmeting is uitgegaan van de resultaten van een uitgebreide experimentele proefopzet met 16 waarnemers en 16 waterdichte peilbuizen. Voor het handmeten met een elektronisch peilapparaat met geluidssignaal aan een meetlint (de handmeetmethode die ook door de Provincie Overijssel wordt gebruikt), kwam de geraamde gemiddelde meetfout uit op 0,50 cm en de geraamde standaardafwijking op 0,84 cm [Knotters, De Meij en Pleijter, 2013].

Tabel 2.1. Kenmerken van de meetfout van een ruwe loggermeting (ϵ_{L0} , blauw gekleurd). Deze zijn afgeleid uit het combineren van: i) de kenmerken van 3.665 verschillen tussen ruwe loggermetingen en de daaraan gepaarde (controle)handmetingen (L0-CH0) en ii) de kenmerken van de meetfout van een (controle)handmeting (ϵ_{CH0}).

Alle loggers Kengetal	L0-CH0 ¹⁾	ϵ_{CH0} ²⁾	ϵ_{L0}
Gemiddelde (cm)	-0,39	0,50	0,11 ³⁾
Standaardafwijking (cm)	3,63	0,84	3,53 ⁴⁾

L0 : ruwe loggermeting CH0 : controlehandmeting	
ϵ_{L0} resp. ϵ_{CH0}	meetfout van L0 resp. CH0
1) empirische raming [Baggelaar en Van der Meulen, 2019]	
2) empirische raming [Knotters et al., 2013]	
3) volgt uit toepassing formule	$\bar{x}_{L0-CH0} = \bar{e}_{L0} - \bar{e}_{CH0}$
4) volgt uit toepassing formule	$S_{L0-CH0}^2 = S_{e,L0}^2 + S_{e,CH0}^2$
\bar{x}_{L0-CH0}	gemiddelde van L0-CH0
\bar{e}_{L0} resp. \bar{e}_{CH0}	gemiddelde meetfout van L0 resp. CH0
S_{L0-CH0}^2	variantie van L0-CH0
$S_{e,L0}^2$ resp. $S_{e,CH0}^2$	variantie meetfout van L0 resp. CH0

De kengetallen zullen bruikbaar worden als we ze uitsplitsen naar homogeneren deelverzamelingen loggers. Zo wordt de standaardafwijking van de ruwe meetfout van een logger van leeftijd 0 jaar geschat op minder dan 1,40 cm en het gemiddelde op 0,56 cm (zie tabel 2.2). Deze standaardafwijking is al duidelijk lager dan die van de heterogene verzameling loggers (3,53 cm, zie tabel 2.1 hierboven), wat zal komen doordat er bij leeftijd 0 jaar nog geen sprake zal zijn van sterke drift. De schatting van een standaardafwijking van minder dan 1,40 cm sluit overigens aan bij de theoretisch afgeleide schatting die uitkwam tussen 0,8 en 1,3 cm (zie § 2.1). Daarom kunnen deze kengetallen bijvoorbeeld dienen als uitgangspunt voor correctiecriteria van ruwe loggermetingen, in plaats van de subjectieve criteria die nu doorgaans gebruikt worden.

Tabel 2.2. Kenmerken van de meetfout van een ruwe meting van een logger van leeftijd 0 jaar (ϵ_{L0} , blauw gekleurd). Deze zijn afgeleid uit het combineren van: i) de kenmerken van 773 verschillen tussen ruwe metingen van loggers met leeftijd 0 jaar en de daaraan gepaarde (controle)handmetingen (L0-CH0) en ii) de kenmerken van de meetfout van een (controle)handmeting (ϵ_{CH0}). De standaardafwijking van L0-CH0 betreft een overschatting, zodat de hier afgeleide standaardafwijking van ϵ_{L0} dat ook zal zijn.

Loggers met leeftijd 0 jaar Kengetal	L0-CH0 ¹⁾	ϵ_{CH0} ²⁾	ϵ_{L0}
Gemiddelde (cm)	0,06	0,50	0,56 ³⁾
Standaardafwijking (cm)	<1,63	0,84	<1,40 ⁴⁾

L0 : ruwe loggermeting (loggerleeftijd 0 jaar) CH0 : controlehandmeting	
ϵ_{L0} resp. ϵ_{CH0}	meetfout van L0 resp. CH0
1) empirische raming [Baggelaar en Van der Meulen, 2019]	
2) empirische raming [Knotters et al., 2013]	
3) volgt uit toepassing formule	$\bar{x}_{L0-CH0} = \bar{e}_{L0} - \bar{e}_{CH0}$
4) volgt uit toepassing formule	$S_{L0-CH0}^2 = S_{e,L0}^2 + S_{e,CH0}^2$
\bar{x}_{L0-CH0}	gemiddelde van L0-CH0
\bar{e}_{L0} resp. \bar{e}_{CH0}	gemiddelde meetfout L0 resp. CH0
S_{L0-CH0}^2	variantie van L0-CH0
$S_{e,L0}^2$ resp. $S_{e,CH0}^2$	variantie meetfout L0 resp. CH0

2.3 Empirisch schatten kenmerken meetfout ná validatie

De kenmerken van de meetfout van een logger ná validatie en zonodige correctie zijn afgeleid uit de verschillen tussen gevalideerde loggermetingen (L1) en de daaraan gepaarde controlehandmetingen (CH0). Het gemiddelde verschil van L1 en CH0 blijkt 0,01 cm en de standaardafwijking van het verschil blijkt 0,94 cm (tabel 2.3). Na verdiscontering van de meetfout van CH0 wordt het gemiddelde van de meetfout van de logger ná validatie en zonodige correctie geschat op 0,51 cm, zodat er blijkbaar ook na validatie nog sprake is van enige overschatting. De standaardafwijking van de meetfout ná validatie en zonodige correctie wordt geschat op *minstens* 0,42 cm. Het is niet mogelijk deze laatste nauwkeuriger te herleiden dan deze ondergrens, doordat de meetfout van L1 (ϵ_{L1}) door de validatie (met zonodig correctie) gecorreleerd is geraakt aan de meetfout van CH0 (ϵ_{CH0}), zodat we ook de covariantie van de meetfouten van L1 en CH0 nodig hebben bij de berekening. Die covariantie is echter niet bekend.

Tabel 2.3. Kenmerken van de meetfout van een loggermeting ná validatie en zonodige correctie (ϵ_{L1} , blauw gekleurd). Deze zijn afgeleid uit het combineren van: i) de kenmerken van 3.665 verschillen tussen gevalideerde loggermetingen en de daaraan gepaarde (controle)handmetingen (L0-CH0) en ii) de kenmerken van de meetfout van een (controle)handmeting (ϵ_{CH0}).

Alle loggers			
Kengetal	L1-CH0 ¹⁾	ϵ_{CH0} ²⁾	ϵ_{L1}
Gemiddelde (cm)	0,01	0,50	0,51 ³⁾
Standaardafwijking (cm)	0,94	0,84	>0,42 ⁴⁾

L1 : gevalideerde loggermeting CH0 : controlehandmeting	
ϵ_{L1} resp. ϵ_{CH0}	meetfout van L1 resp. CH0
1) empirische raming [Baggelaar en Van der Meulen, 2019]	
2) empirische raming [Knotters et al., 2013]	
3) volgt uit formule	$\bar{x}_{L1-CH0} = \bar{e}_{L1} - \bar{e}_{CH0}$
4) volgt uit formule	$s_{L1-CH0}^2 = s_{e,L1}^2 + s_{e,CH0}^2 - 2 \cdot Cov(e_{L1}, e_{CH0})$
\bar{x}_{L1-CH0}	gemiddelde van L1-CH0
\bar{e}_{L1} resp. \bar{e}_{CH0}	gemiddelde meetfout L1 resp. CH0
s_{L1-CH0}^2	variantie van L1-CH0
$s_{e,L1}^2$ resp. $s_{e,CH0}^2$	variantie meetfout L1 resp. CH0
$Cov(e_{L1}, e_{CH0})$	covariantie meetfout L1 en CH0

De voorgaande studie leverde dus nog geen bruikbare standaardafwijking op van de meetfout van gevalideerde loggermetingen (ϵ_{L1}). Op verzoek van de Provincie Overijssel hebben wij deze vervolgstudie uitgevoerd om die standaardafwijking toch te kunnen schatten. Deze studie is gebaseerd op het verzamelen van een steekproef willekeurige handmetingen buiten de uitleesmomenten, die daarna ook niet zijn gebruikt om loggermetingen te valideren.

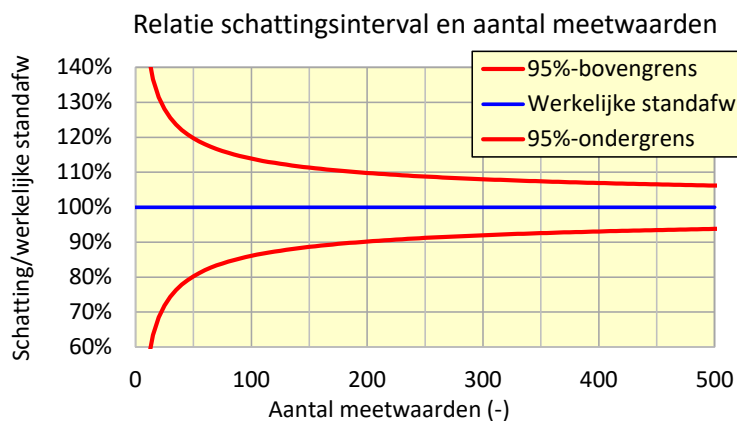
3 Opzet steekproef handmetingen

Om de kenmerken van de meetfout van gevalideerde loggermetingen (L1) te kunnen schatten, dient van elk van een groot aantal van deze metingen het verschil te worden bepaald met een daaraan gepaarde (steekproef)handmeting (SH), die niet is gebruikt bij het valideren. De kenmerken van de verschilwaarden L1-SH moeten vervolgens nog worden gecorrigeerd voor de kenmerken van de meetfout van de handmetingen (ϵ_{SH}). In dit hoofdstuk is beschreven hoe de daarvoor benodigde handmetingen zijn verzameld met een steekproef.

3.1 Keuze steekproefgrootte

De steekproefgrootte is zodanig gekozen dat de standaardafwijking van de verschilwaarden L1-SH voldoende nauwkeurig kan worden geraamd. Om een goede onderbouwing van die keuze mogelijk te kunnen maken is de relatie afgeleid tussen de steekproefgrootte (en daarmee ook de kosten) en de uit de steekproef volgende precisie van de standaardafwijking (figuur 3.1).

Figuur 3.1. Relatie tussen het 95%-schattingsinterval van de standaardafwijking en het aantal waarden waar de schatting op gebaseerd is. Deze relatie gaat er van uit dat de waarden afkomstig zijn uit een normale kansverdeling en aselekt zijn verkregen.



Uit figuur 3.1 blijkt dat er bij 100 aselekt verkregen verschilwaarden L1-SH die afkomstig zijn uit een normale kansverdeling een kans van 95% is dat de schatting van de standaardafwijking van het verschil L1-SH niet meer dan circa 14% afwijkt van de werkelijke standaardafwijking van L1-SH. Bij 150 verschilwaarden zal deze niet meer dan circa 11% afwijken en bij 200 verschilwaarden niet meer dan circa 10%. De precisiewinst per extra steekproefeenheid is bij meer dan circa 150 verschilwaarden dus nog maar gering. Op basis hiervan is in samenspraak met de Provincie Overijssel gekozen voor een steekproef van 200 handmetingen van de grondwaterstand.⁴

De steekproef is vervolgens gespreid uitgevoerd over een jaar, van 23 april 2019 t/m 6 mei 2020, om vertekening door seizoensinvloeden zoveel mogelijk te minimaliseren. Er is namelijk sprake van een (beperkte) seizoensinvloed op het verschil van loggermeting en handmetingen [Baggelaar en Van der Meulen, 2019].

⁴ Bij de uitvoering bleek dat de verzameling verschilwaarden door enkele uitschieters niet voldoet aan de normale kansverdeling (zie hoofdstuk 4), zodat er minder garantie is op de hier beoogde precisie.

3.2 Selecteren van peilfilters voor de steekproef

Het selecteren van peilfilters voor de steekproef omvatte de volgende stappen.

1. Door de Provincie is een lijst overlegd met metadata van 548 peilfilters van de te onderzoeken populatie (peilbuis, filter, coördinaten, filterstelling, maaiveldhoogte, meetpunthoogte, logger-nummer en loggerdiepte). Deze lijst omvat alleen peilfilters die worden beheerd door de Provincie, die nog in gebruik zijn en die een logger bevatten. Het betreft alleen Keller-loggers van het type DCX-22 AA. De Provincie heeft daarbij bepaalde peilfilters bij voorbaat uitgesloten, bijvoorbeeld omdat bekend is dat daar door een afwijkende meetopstelling minder betrouwbare metingen worden verricht, of omdat ze moeilijk bereikbaar zijn.
2. Op basis van de overlegde lijst van 548 peilfilters – het steekproefkader - hebben wij de 200 peilfilters van de steekproef geselecteerd volgens het verlotingsprincipe, zodanig dat elk dezelfde kans had om te worden geselecteerd. Er is geen stratificatie toegepast bij het selecteren. Stratificatie naar één of meer factoren kan de nauwkeurigheid van de steekproef bij dezelfde steekproefgrootte vergroten, mits elk van die factoren een voldoende grote invloed heeft op de standaardafwijking van de steekproefvariabele (hier is dat L1-SH).

3.3 Overdracht gegevens en instructies aan uitvoerder steekproef

De lijst met 548 peilfilters en recente metadata daarvan zijn door ons beschikbaar gesteld aan de uitvoerder van de steekproef (René Buijs). Dit betrof een Excelbestand met de 200 geselecteerde peilfilters en een Excelbestand met de 348 overige peilfilters. Ook zijn meegegeven links naar topografische kaarten en situatiefoto's met de peilbuizen (zie figuur 3.2).

Figuur 3.2. Topografische kaart met locatie van een peilbuis en een bijbehorende situatiefoto.



Verder is de instructie meegegeven dat de overige 348 peilfilters als reserve moeten fungeren voor de gevallen dat er bij één van de 200 geselecteerde peilfilters geen hand- of loggermeting uitgevoerd kan worden. In zo'n geval dient te worden gekozen voor een ander peilfilter van de betreffende peilbuis en pas als dat niet beschikbaar is, of als daar geen meting kan worden uitgevoerd, dient te worden gekozen voor een ander peilfilter van de reservelijst.

Geheimhouden geselecteerde peilfilters

De Provincie Overijssel mag niet te weten komen welke peilfilters zijn geselecteerd, om te vermijden dat het validatieproces daardoor beïnvloed kan worden. Verder zullen wij er op toezien dat de steekproefopzet ook na afronding van het onderzoek vertrouwelijk blijft, zodat deze ook in de toekomst kan worden gebruikt. Er kunnen immers in een later stadium meetreeksen opnieuw worden gecorrigeerd op basis van het langjarige beeld van de verschillen met handmetingen, of door ontwikkeling van nieuwe technieken.

Het aansluiten van een logger op een veld-laptop en het vervolgens uitlezen van loggerdata heeft volgens de meetnetbeheerder geen invloed op het meetproces van de logger. Om te waarborgen dat ook de handmeting geen invloed kan hebben op het meetproces, is de uitvoerder van de steekproef geïnstrueerd om binnen een tijdsvenster van 5 minuten rondom een automatische loggermeting geen handmeting uit te voeren. Hiertoe moet deze bij aankomst de veld-laptop aansluiten op de logger, om zo te kunnen zien wanneer de laatste loggermeting is geweest en de eerstvolgende komt. Verder is de veld-laptop vóór elke interactie synchroon gezet met de loggertijd, omdat de loggers het hele jaar op wintertijd blijven staan.

De steekproefhandmeting is uitgevoerd met een elektronisch peilapparaat met geluidssignaal aan een meetlint. Deze handmeetmethode wordt ook door de Provincie Overijssel gebruikt, namelijk voor de controlehandmetingen bij het uitlezen van de loggers.

De eerste meting van de steekproef is uitgevoerd op 23 april 2019 en de laatste op 6 mei 2020. De steekproef omvatte 200 peilfilters, waarbij 187 van de initiële lijst en 13 van de reservelijst. Dit leverde 199 steekproefhandmetingen (SH) op⁵ en 200 steekproefloggermetingen (SL).

De steekproefuitkomsten zijn ons door de uitvoerder aangeleverd in een Excel-bestand, waarbij deze de metagegevens van peilbuis, peilfilter en logger heeft aangevuld met:

- gegevens over de bij het veldbezoek operationele loggerreeks, namelijk:
 - het tijdstip van de start van die reeks. Dit is het eerste meettijdstip na het uitlezen van de voorgaande reeks door de Provincie;
 - het tijdstip van de laatste loggermeting vóórdat de uitvoerder van de steekproef zijn metingen heeft uitgevoerd (t_v);
 - het tijdstip van de eerste loggermeting nádat de uitvoerder van de steekproef zijn metingen heeft uitgevoerd (t_n);
- gegevens over de steekproefmetingen:
 - de steekproefhandmeting (SH, in cm onder meetpunt) en het tijdstip daarvan (volgens de veld-laptop);
 - de steekproefloggermeting (SL, in cm onder meetpunt) en het tijdstip daarvan (volgens de logger);
- eventuele bijzonderheden.

⁵ Door droogval ontbrak een steekproefhandmeting.

4 Statistische verwerking steekproefresultaten

Dit hoofdstuk beschrijft hoe de kenmerken van de meetfout van gevalideerde loggermetingen (L1) zijn geschat, door van elk van een groot aantal van deze metingen het verschil te bepalen met een daaraan gepaarde (steekproef)handmeting (SH), die niet is gebruikt bij het valideren. De kenmerken van de verschilwaarden L1-SH zijn vervolgens gecorrigeerd voor de kenmerken van de meetfout van de handmetingen (ε_{SH}).

4.1 Afleiden verschilwaarden L1-SH en L0-SH

Om de bij de steekproef verzamelde handmetingen van de grondwaterstand (SH) te kunnen vergelijken met daaraan gepaard te achten gevalideerde loggermetingen (L1), heeft de Provincie ons op 7 november 2020, een half jaar na de laatste handmeting van de steekproef, zowel reeksen van ruwe loggermetingen (L0) als van gevalideerde loggermetingen (L1) beschikbaar gesteld van alle 613 peilfilters die oorspronkelijk in aanmerking kwamen als steekproefkader voor de steekproef en waaruit uiteindelijk 200 peilfilters zijn geselecteerd voor de steekproef.

Dit stelde ons in staat de verzameling verschillen L0-SH en de verzameling verschillen L1-SH als volgt te bepalen:

- L0-SH is bepaald als de verzameling van m verschillen $L0_i-SH_i$ ($i = 1, 2, \dots, m$), waarbij $L0_i$ de ruwe loggermeting is die het kortst vóór of ná de steekproefhandmeting SH_i is genomen, echter op voorwaarde dat de twee metingen niet meer dan een uur uit elkaar liggen;
- L1-SH is bepaald als de verzameling van n verschillen $L1_i-SH_i$ ($i = 1, 2, \dots, n$), waarbij $L1_i$ de gevalideerde (en daarbij zonodig gecorrigeerde) versie is van de ruwe loggermeting $L0_i$ die het kortst vóór of ná de steekproefhandmeting SH_i is genomen, echter op voorwaarde dat de twee metingen niet meer dan een uur uit elkaar liggen.

Voor 8 peilfilters bleek het niet mogelijk om bruikbare verschilwaarden L1-SH te bepalen en evenmin om bruikbare verschilwaarden L0-SH te bepalen, met de volgende redenen:

- bij één peilfilter bleek het niet mogelijk een handmeting te doen door droogval;
- bij twee peilfilters bleek uit het bij de steekproef genoteerde loggernummer dat de logger niet overeenstemde met degene die daar volgens de later geleverde reeksen van ruwe en gevalideerde loggermetingen aanwezig zou moeten zijn;
- van vijf peilfilters bleken er geen ruwe en gevalideerde loggermetingen beschikbaar van minder dan een uur vóór of ná de steekproefhandmetingen (de gerealiseerde hiaten lagen tussen 2 en 375 dagen).

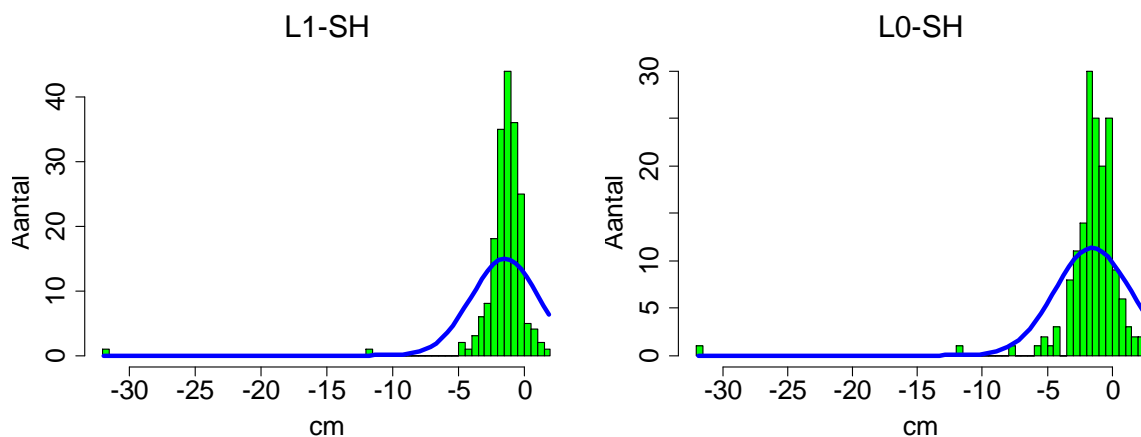
Daar bovenop bleek het voor 27 peilfilters niet mogelijk bruikbare verschilwaarden L0-SH te bepalen, omdat de ons geleverde reeksen van ruwe loggermetingen (L0) meer hiaten bleken te bevatten dan de ons geleverde reeksen van gevalideerde loggermetingen (L1).

Daarmee bleek het uiteindelijk voor 192 peilfilters mogelijk bruikbare verschilwaarden L1-SH te bepalen en voor 165 peilfilters om bruikbare verschilwaarden L0-SH te bepalen. Deze verzamelingen verschilwaarden zijn vermeld in de bijlage.

4.2 Verdelingen verschilwaarden L1-SH en L0-SH

Figuur 4.1 toont het histogram van de 192 verschilwaarden L1-SH en dat van de 165 verschilwaarden L0-SH, met daaronder een aantal kenmerken van de twee verzamelingen.

Figuur 4.1. Histogram van de 192 verschilwaarden L1-SH (links) en dat van de 165 verschilwaarden L0-SH (rechts), met daaronder een aantal kenmerken van de twee verzamelingen. De blauwe lijn geeft de theoretische vorm aan van een normale kansverdeling met het gemiddelde en de standaardafwijking van de betreffende verzameling verschilwaarden.



Kenmerk	L1-SH	L0-SH
Aantal (-)	192	165
Minimum (cm)	-31,90	-31,90
5-percentiel (cm)	-3,30	-4,07
Mediaan (cm)	-1,20	-1,34
Gemiddelde (cm)	-1,47	-1,52
95-percentiel (cm)	0,10	0,77
Maximum (cm)	1,90	2,32
Standaardafw (cm)	2,55	2,90

Toelichting

L0: ruwe loggermeting van de grondwaterstand

L1: gevalideerde en zonodig gecorrigeerde loggermeting van de grondwaterstand

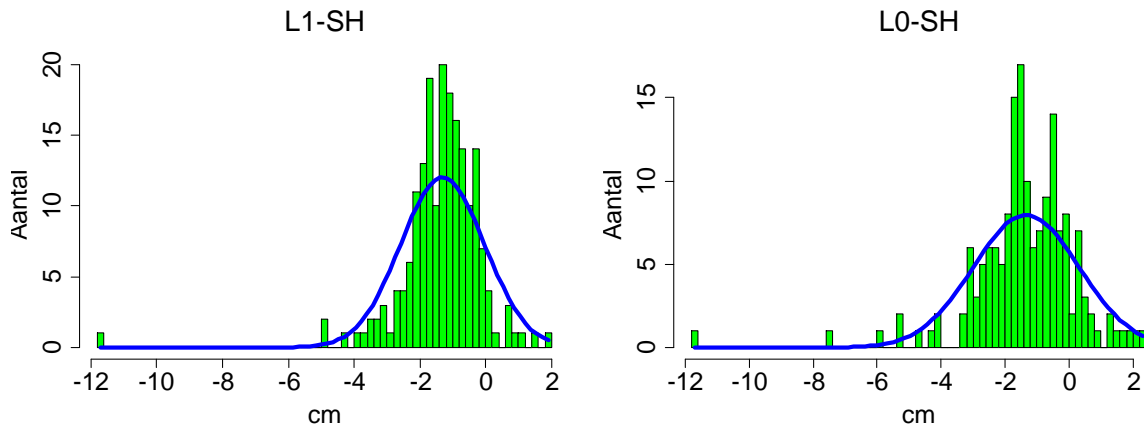
SH: steekproefhandmeting

Uit figuur 4.1 blijkt dat de twee verzamelingen nauwelijks verschillen in gemiddelde (-1,47 versus -1,52 cm) en dat L1-SH een kleinere standaardafwijking heeft dan L0-SH (2,55 versus 2,90 cm). Dit laatste mag ook verwacht worden, als effect van het valideren en corrigeren. Verder blijken beide verzamelingen sterk scheef naar links te zijn, door enkele negatieve uitschieters.

Voor beide verzamelingen betreft steekproefnummer 198 veruit de grootste uitschieter, met een waarde van -31,90 cm. Deze extreme waarde betreft met grote waarschijnlijkheid een transcriptiefout van de uitvoerder bij het registreren van de steekproefhandmeting SH. Die handmeting is namelijk doorgegeven als 114,0 cm onder meetpunt, maar de uitvoerder gaf bij navraag aan dat het getal in zijn handgeschreven veldnotitie bij nader inzien meer lijkt op 144,0 dan op 114,0. Als we daarvoor corrigeren komt de verschilwaarde uit op 1,90 cm in plaats van -31,90 cm (zowel voor L1₁₉₈-SH₁₉₈, als voor L0₁₉₈-SH₁₉₈). Dit is op te vatten als een gecorrigeerde schatting en in het nu volgende gaan we alleen nog uit van de verzamelingen verschilwaarden L1-SH en L0-SH die ontstaan ná deze correctie. Deze verzamelingen zijn in de bijlage aangeduid als respectievelijk L1-SHc en L0-SHc.

Figuur 4.2 toont het histogram van de 192 verschilwaarden L1-SH en dat van de 165 verschilwaarden L0-SH, beide na correctie van SH₁₉₈, met daaronder een aantal kenmerken van de twee verzamelingen.

Figuur 4.2. Histogram van de 192 verschilwaarden L1-SH (links) en dat van de 165 verschilwaarden L0-SH (rechts), beide na correctie van SH₁₉₈, met daaronder een aantal kenmerken van de twee verzamelingen. De blauwe lijn geeft de theoretische vorm aan van een normale kansverdeling met het gemiddelde en de standaardafwijking van de betreffende verzameling verschilwaarden.



Kenmerk	L1-SH	L0-SH
Aantal (-)	192	165
Minimum (cm)	-11,70	-11,66
5-percentiel (cm)	-3,19	-3,90
Mediaan (cm)	-1,20	-1,34
Gemiddelde (cm)	-1,31	-1,34
95-percentiel (cm)	0,10	0,77
Maximum (cm)	1,90	2,32
Standaardafw (cm)	1,28	1,66

Toelichting

L0: ruwe loggermeting van de grondwaterstand

L1: gevalideerde en zonodig gecorrigeerde loggermeting van de grondwaterstand

SH: steekproefhandmeting (hier na correctie van SH₁₉₈)

Door de correctie van SH₁₉₈ zijn beide gemiddelden enigszins vergroot (met 0,16 en 0,18 cm), maar de twee standaardafwijkingen zijn relatief sterk verlaagd (met 1,27 en 1,24 cm). Verder blijkt dat beide verzamelingen verschilwaarden ook ná de correctie nog scheef naar links zijn. Bij toetsen op normaliteit met de Shapiro-Wilk-toets wordt voor beide verzamelingen de nulhypothese van normaliteit verworpen bij een gehanteerd significantieniveau van 5% (in beide gevallen is de p-waarde < 0,001).

Deze schattingen van gemiddelde en standaardafwijking van L0-SH (-1,34 en 1,66 cm) wijken af van de schattingen (-0,39 en 3,63 cm), zoals afgeleid uit 3.665 verschillen tussen ruwe loggermetingen en daaraan gepaarde (controle)handmetingen (L0-CH0) bij de voorgaande studie [Baggelaar en Van der Meulen, 2019]. Het is echter niet uit te sluiten dat die verzameling van 3.665 verschillen nog diverse foute handmetingen bevatte, omdat er nauwelijks objectieve checks mogelijk waren.

4.3 Kenmerken verschilwaarden L1-SH en L0-SH

Aangezien de twee verzamelingen verschilwaarden L1-SH en L0-SH scheef zijn en bij toetsen niet voldoen aan normaliteit, zijn de standaardmethoden om de 95%-betrouwbaarheidsintervallen van de kengetallen gemiddelde en standaardafwijking af te leiden niet geschikt. Daarom hebben we die afgeleid met bootstrappen, zoals hieronder toegelicht.

Bootstrappen

Bootstrappen is een statistische techniek, gebaseerd op het herhaald met teruglegging trekken van meetwaarden uit een beschikbare steekproef en daar vervolgens kengetallen uit afleiden. Hier houdt dat in dat voor afzonderlijk de verzameling met 192 verschilwaarden L1-SH en de verzameling met 165 verschilwaarden L0-SH 100.000 maal een substeekproef is getrokken die even groot is als de steekproef waaruit is getrokken. Door het terugleggen hebben die substeekproeven doorgaans een enigszins andere samenstelling dan de steekproef, met minder unieke waarden. Vervolgens is per substeekproef zowel gemiddelde als standaardafwijking bepaald, wat voor beide verzamelingen verschilwaarden zowel 100.000 gemiddelden opleverde – aangeduid als de bootstrapverdeling van het gemiddelde - als 100.000 standaardafwijkingen - aangeduid als de bootstrapverdeling van de standaardafwijking. Het gemiddelde is vervolgens geschat als het gemiddelde van de bootstrapverdeling van het gemiddelde en onder- en bovengrens van zijn 95%-betrouwbaarheidsinterval zijn geschat als het 2,5-percentiel en het 97,5-percentiel van die bootstrapverdeling. De standaardafwijking is geschat als het gemiddelde van de bootstrapverdeling van de standaardafwijking en onder- en bovengrens van zijn 95%-betrouwbaarheidsinterval zijn geschat als het 2,5-percentiel en het 97,5-percentiel van die bootstrapverdeling. De betrouwbaarheidsintervallen van het geschatte gemiddelde en de geschatte standaardafwijking zijn daarmee dus geschat zonder dat een vooronderstelling over normaliteit nodig was.

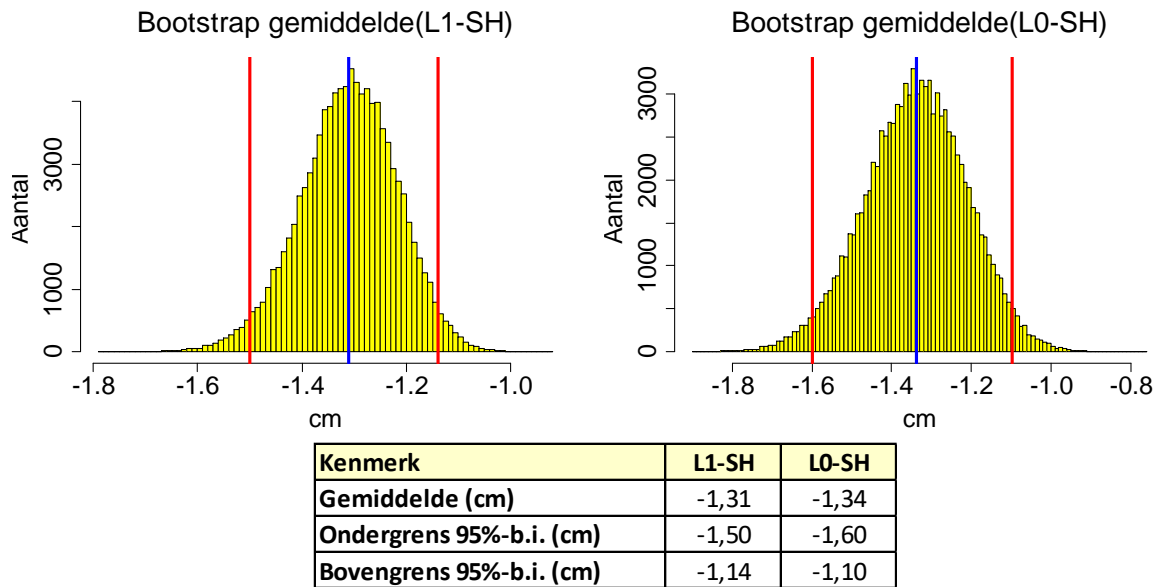
Het principe achter bootstrappen is dat een statistisch verantwoorde steekproef kan worden opgevat als een mini-populatie, die kenmerken heeft van de moederpopulatie waar deze uit is getrokken. Die kenmerken kunnen met analyses van de herhaald met teruglegging verkregen substeekproeven in beeld worden gebracht. Deze techniek kan vooral uitkomst bieden als er sprake is van een sterk scheve kansverdeling (zoals door uitschieters) en de steekproef nog niet groot genoeg blijkt om na middeling uit te mogen gaan van een normale kansverdeling van het gemiddelde.⁶ Zie bijvoorbeeld [Efron, 1979] en [Chernick, 2007] voor toelichtingen op bootstrappen.

⁶ Volgens de Centrale Limietstelling zal de kansverdeling van een gemiddelde de normale kansverdeling meer benaderen naarmate over meer meetwaarden wordt gemiddeld, ongeacht het soort kansverdeling van de meetwaarden.

Resultaten van het bootstrappen

Figuur 4.3 toont de histogrammen van de 100.000 met bootstrappen geschatte gemiddelden van L1-SH (links) en L0-SH (rechts), met daaronder een aantal kenmerken van de twee verzamelingen.

Figuur 4.3. Histogrammen van de 100.000 met bootstrappen geschatte gemiddelden van L1-SH (links) en L0-SH (rechts), met daaronder een aantal kenmerken van de twee verzamelingen. De verticale blauwe lijn geeft het gemiddelde aan van de 100.000 geschatte gemiddelden en de twee verticale rode lijnen het 2,5-percentiel en het 97,5-percentiel (hier op te vatten als de grenzen van het 95%-betrouwbaarheidsinterval van het gemiddelde).



Toelichting

L0: ruwe loggermeting van de grondwaterstand

L1: gevalideerde en zonodig gecorrigeerde loggermeting van de grondwaterstand

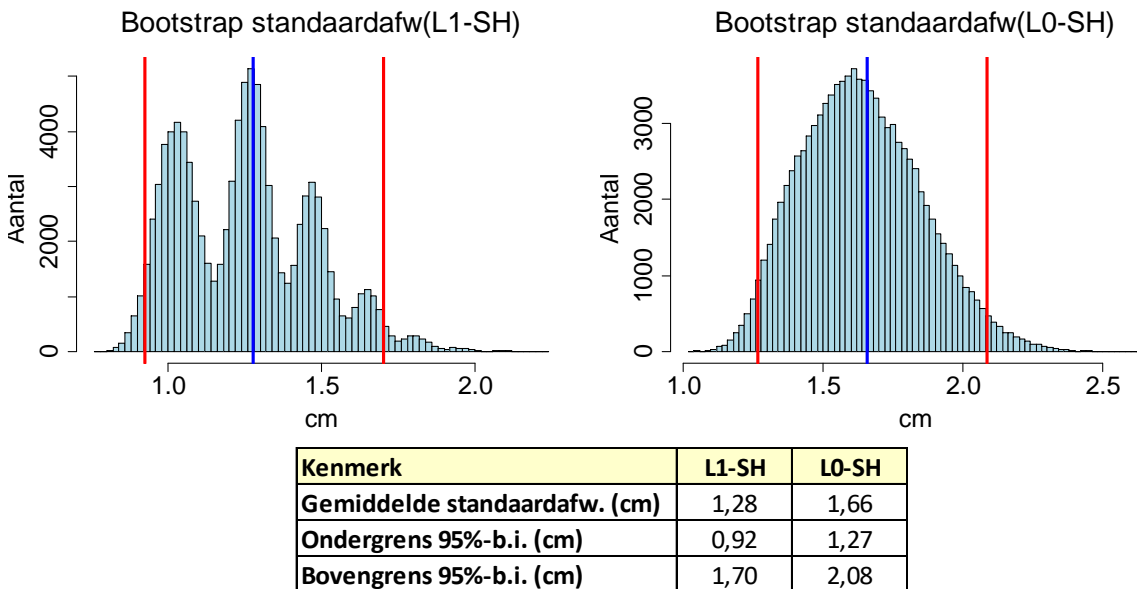
SH: steekproefhandmeting (hier na correctie van SH₁₉₈)

Voor L1-SH is het gemiddelde geschat op -1,31 cm, met 95%-betrouwbaarheidsinterval dat loopt van -1,50 tot -1,14 cm. Voor L0-SH is het gemiddelde geschat op -1,34 cm met 95%-betrouwbaarheidsinterval dat loopt van -1,60 tot -1,10 cm. Voor beide soorten verschilwaarden is er dus sprake van een gemiddeld verschil tussen loggermeting en handmeting dat statistisch significant afwijkt van nul.

Voor beide verzamelingen stemt het met bootstrappen geschatte gemiddelde goed overeen met het gemiddelde van de oorspronkelijke steekproef (vermeld in figuur 4.2), zodat er geen onzuiverheidscorrectie hoeft te worden toegepast bij het schatten van het betrouwbaarheidsinterval (zie daarvoor [Efron, 1987]).

Figuur 4.4 toont de histogrammen van de 100.000 met bootstrappen geschatte standaardafwijkingen van L1-SH (links) en L0-SH (rechts), met daaronder een aantal kenmerken van de twee verzamelingen.

Figuur 4.4. Histogrammen van de 100.000 met bootstrappen geschatte standaardafwijkingen van L1-SH (links) en L0-SH (rechts), met daaronder een aantal kenmerken van de twee verzamelingen. De verticale blauwe lijn geeft het gemiddelde aan van de 100.000 geschatte standaardafwijkingen en de twee verticale rode lijnen het 2,5-percentiel en het 97,5-percentiel (hier op te vatten als de grenzen van het 95%-betrouwbaarheidsinterval van de standaardafwijking).



Toelichting

L0: ruwe loggermeting van de grondwaterstand

L1: gevalideerde en zonnodig gecorrigeerde loggermeting van de grondwaterstand

SH: steekproefhandmeting (hier na correctie van SH₁₉₈)

Voor L1-SH is de standaardafwijking geschat op 1,28 cm, met 95%-betrouwbaarheidsinterval dat loopt van 0,92 tot 1,70 cm. Voor L0-SH is de standaardafwijking geschat op 1,66 cm met 95%-betrouwbaarheidsinterval dat loopt van 1,27 tot 2,08 cm.

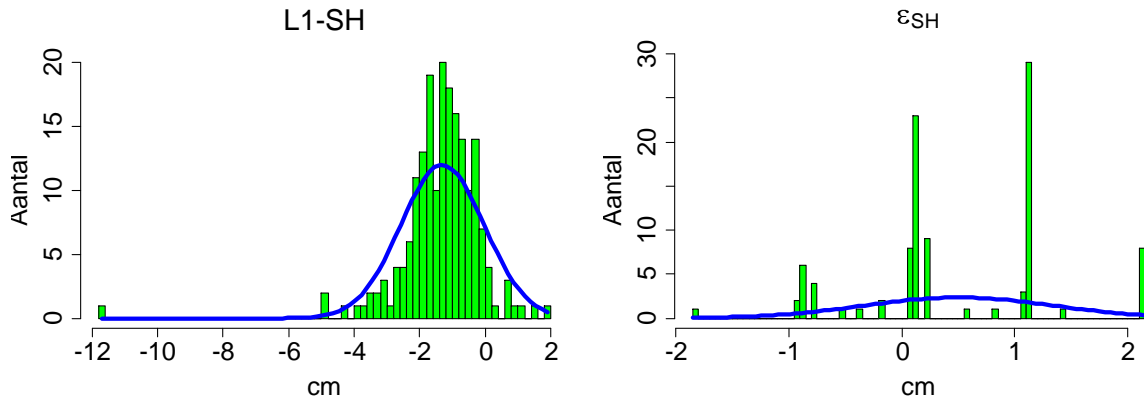
Voor beide verzamelingen stemt de met bootstrappen geschatte standaardafwijking goed overeen met de standaardafwijking van de oorspronkelijke steekproef (vermeld in figuur 4.2).

4.4 Afleiden kenmerken meetfout gevalideerde loggermeting

In het voorgaande zijn de kenmerken afgeleid van het verschil tussen een gevalideerde loggermeting (L1) en de daarbij horende steekproefhandmeting (SH). Hieruit kunnen we de kenmerken afleiden van de meetfout van een gevalideerde loggermeting, door te corrigeren voor de kenmerken van de meetfout van een handmeting (ϵ_{SH}).

Figuur 4.5 toont het histogram van de 192 bij deze studie verzamelde verschilwaarden L1-SH (na correctie, links) en de 100 meetfouten van steekproefhandmetingen (ϵ_{SH} , rechts), zoals verzameld bij een proefopzet [Knotters et al., 2013]. Met de kenmerken van die twee verzamelingen zijn de kenmerken afgeleid van de meetfout van L1, zoals vermeld en toegelicht in tabel 4.1.

Figuur 4.5. Histogram van de 192 bij deze studie verzamelde verschilwaarden L1-SH (na correctie, links) en de 100 meetfouten van steekproefhandmetingen (ϵ_{SH} , rechts), zoals verzameld bij een proefopzet⁷. De blauwe lijn geeft de theoretische vorm aan van een normale kansverdeling met het gemiddelde en de standaardafwijking van de betreffende verzameling. Onder de histogrammen een aantal kenmerken van de twee verzamelingen.



Kenmerk	L1_SH	ϵ_{SH}
Aantal (-)	192	100
Minimum (cm)	-11,70	-1,85
5-percentiel (cm)	-3,19	-0,85
Mediaan (cm)	-1,20	0,25
Gemiddelde (cm)	-1,31	0,50
95-percentiel (cm)	0,10	2,15
Maximum (cm)	1,90	2,15
Standaardafw (cm)	1,28	0,84

Tabel 4.1. Kenmerken van de meetfout van een gevalideerde loggermeting (ϵ_{L1} , blauw gekleurd). Deze zijn afgeleid uit het combineren van: i) de kenmerken van 192 verschillen tussen gevalideerde loggermetingen en daaraan gepaarde steekproefhandmetingen (L1-SH) en ii) de kenmerken van de meetfout van een (steekproef)handmeting (ϵ_{SH}).

Kengetal	L1-SH ¹⁾	ϵ_{SH} ²⁾	ϵ_{L1}
Gemiddelde (cm)	-1,31	0,50	-0,81 ³⁾
Standaardafwijking (cm)	1,28	0,84	0,96 ⁴⁾

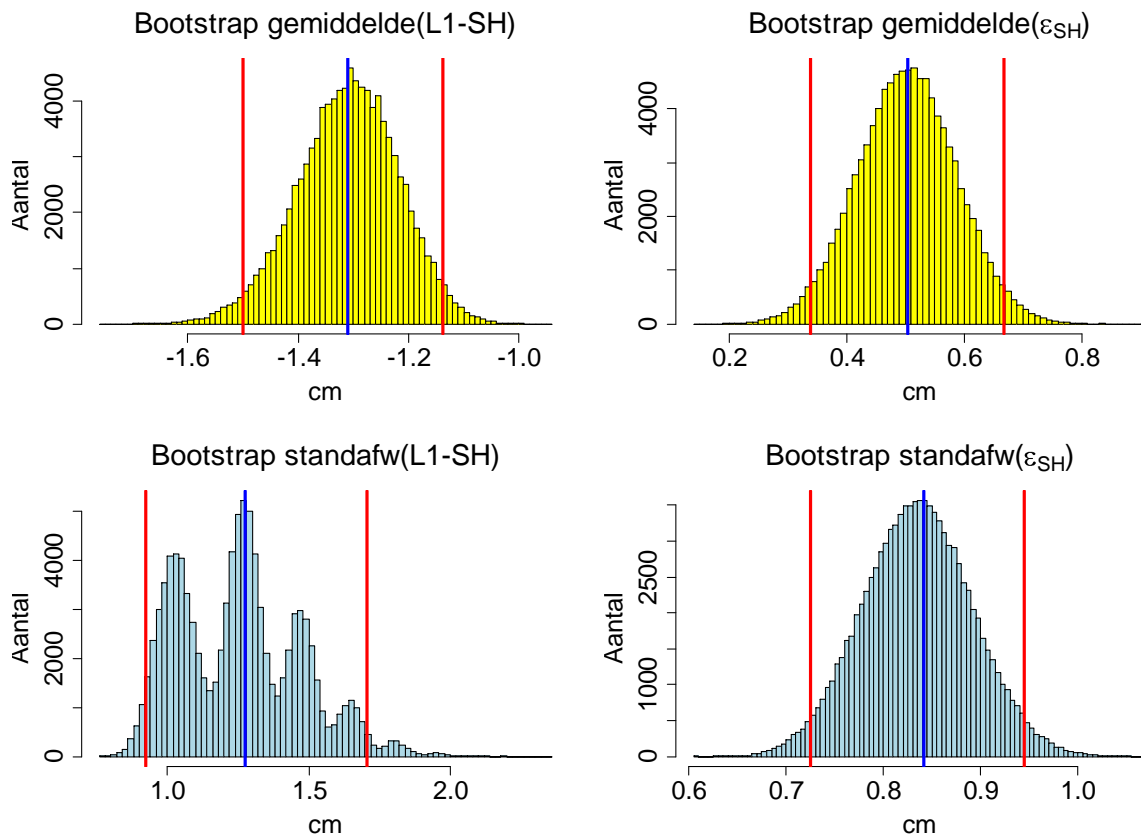
L1 : gevalideerde loggermeting SH : steekproefhandmeting	
ϵ_{L1} resp. ϵ_{SH} meetfout van L1 resp. SH	
1) empirische raming (zie figuur 4.3 en figuur 4.4)	
2) empirische raming [Knotters et al., 2013]	
3) volgt uit toepassing formule	$\bar{x}_{L1-SH} = \bar{e}_{L1} - \bar{e}_{SH}$
4) volgt uit toepassing formule	$s_{L1-SH}^2 = s_{e,L1}^2 + s_{e,SH}^2$
\bar{x}_{L1-SH}	gemiddelde van L1-SH
\bar{e}_{L1} resp. \bar{e}_{SH}	gemiddelde meetfout L1 resp. SH
s_{L1-SH}^2	variantie van L1-SH
$s_{e,L1}^2$ resp. $s_{e,SH}^2$	variantie meetfout L1 resp. SH

Volgens deze berekening bedraagt de gemiddelde meetfout van een gevalideerde loggermeting -0,81 cm en de standaardafwijking 0,96 cm. Dit betreft echter schattingen die gepaard zullen gaan met onzekerheden, door de onzekerheden van gemiddelde en standaardafwijking van zowel L1-SH

⁷ Het betrof een uitgebreide experimentele proefopzet met 16 waarnemers en 16 waterdichte peilbuizen [Knotters, De Meij en Pleijter, 2013]. We zijn uitgegaan van de 100 daarbij geconstateerde meetfouten bij handmeten met een elektronisch peilapparaat met geluidssignaal aan een meetlint. Dat is namelijk de handmeetmethode die ook door de Provincie Overijssel wordt gebruikt.

als SH. De onzekerheden van gemiddelde en standaardafwijking zijn voor beide verzamelingen in beeld gebracht met bootstrappen (zie figuur 4.6).

Figuur 4.6. Histogrammen van de 100.000 met bootstrappen geschatte gemiddelden van L1-SH (linksboven) en ϵ_{SH} (rechtsboven) en van de 100.000 met bootstrappen geschatte standaardafwijkingen van L1-SH (linksonder) en ϵ_{SH} (rechtsonder), met daaronder relevante kenmerken van de twee verzamelingen. De verticale blauwe lijn geeft het gemiddelde aan van de 100.000 schattingen en de twee verticale rode lijnen het 2,5-percentiel en het 97,5-percentiel (hier op te vatten als de grenzen van het 95%-betrouwbaarheidsinterval van de gemiddelde schatting).



Kenmerken gemiddelde	L1-SH	ϵ_{SH}
Gemiddelde (cm)	-1,31	0,50
Ondergrens 95%-b.i. (cm)	-1,50	0,34
Bovengrens 95%-b.i. (cm)	-1,14	0,67

Kenmerken standaardafwijking	L1-SH	ϵ_{SH}
Gemiddelde standaardafw. (cm)	1,28	0,84
Ondergrens 95%-b.i. (cm)	0,92	0,72
Bovengrens 95%-b.i. (cm)	1,70	0,95

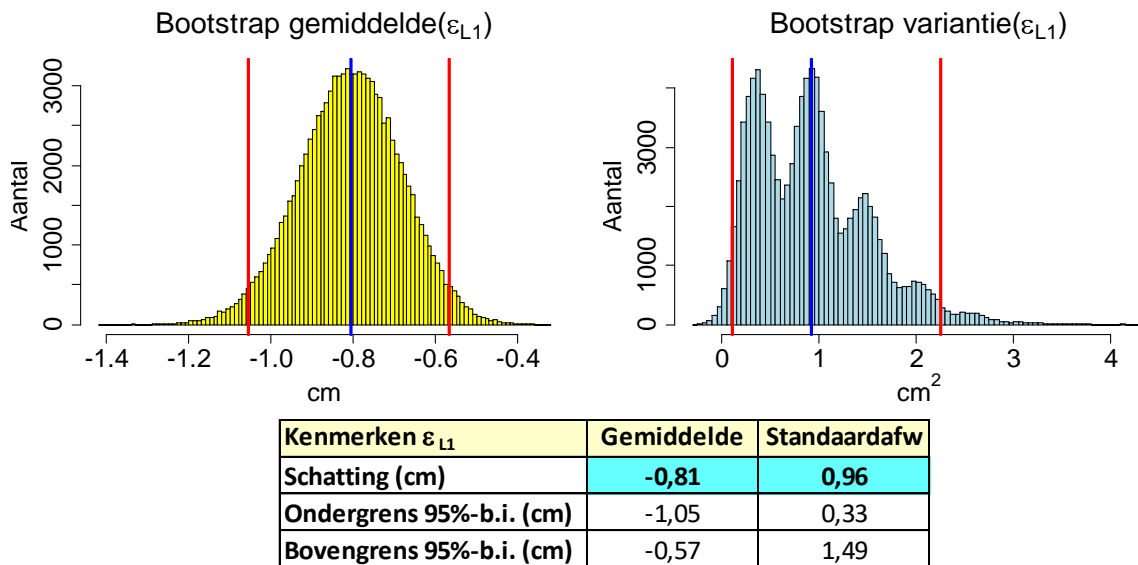
Met deze bootstrapverdelingen van gemiddelde en standaardafwijking van L1-SH en ϵ_{SH} konden vervolgens de bootstrapverdelingen van gemiddelde en variantie van ϵ_{L1} (de meetfout van L1) worden afgeleid:

- de bootstrapverdeling van het gemiddelde van ϵ_{L1} is afgeleid door voor elk van 100.000 combinaties van een gebootstrapt gemiddelde van L1-SH en een gebootstrapt gemiddelde van ϵ_{SH} een bijbehorende schatting van het gemiddelde van ϵ_{L1} te bepalen met formule 3 (vermeld in het tekstvak onder tabel 4.1);
- de bootstrapverdeling van de variantie van ϵ_{L1} is afgeleid door voor elk van 100.000 combinaties van een gebootstrapte standaardafwijking van L1-SH en een gebootstrapte standaardaf-

wijking van ε_{SH} een bijbehorende schatting van de variantie van ε_{L1} te bepalen met formule 4 (vermeld in het tekstvak onder tabel 4.1).

De twee resulterende bootstrapverdelingen zijn weergegeven in figuur 4.7 en daaronder zijn zowel voor het geschatte gemiddelde als de geschatte standaardafwijking van ε_{L1} de grenzen van het 95%-betrouwbaarheidsinterval vermeld.

Figuur 4.7. Histogrammen van de 100.000 met bootstrappen geschatte gemiddelden van ε_{L1} (links) en varianties van ε_{L1} (rechts). De verticale blauwe lijn geeft het gemiddelde aan van de 100.000 schattingen en de twee verticale rode lijnen het 2,5-percentiel en het 97,5-percentiel (hier op te vatten als de grenzen van het 95%-betrouwbaarheidsinterval van de gemiddelde schatting). Daaronder de daaruit afgeleide schattingen van gemiddelde en standaardafwijking van ε_{L1} , evenals de bijbehorende 95%-betrouwbaarheidsintervallen.



Met deze verkenning is de gemiddelde meetfout van een gevalideerde loggermeting geschat op -0,81 cm en de standaardafwijking op 0,96 cm. Maar als we rekening houden met de schattingsonzekerheden zal met 95% betrouwbaarheid de gemiddelde meetfout van een gevalideerde loggermeting ergens in het interval van -1,05 tot -0,57 cm liggen. Aangezien dit interval alleen negatieve waarden omvat, is er met 95% betrouwbaarheid sprake van een statistisch significante negatieve systematische fout. En eveneens rekening houdende met de schattingsonzekerheden zal met 95% betrouwbaarheid de standaardafwijking van de meetfout van een gevalideerde loggermeting ergens in het interval van 0,33 tot 1,49 cm liggen.

Bovenstaande resultaten kunnen worden ook aangeduid als de (on)juistheid en de precisie van een gevalideerde loggermeting, die samen zijn nauwkeurigheid bepalen. Dit is uitgewerkt in de volgende paragraaf (§ 4.5).

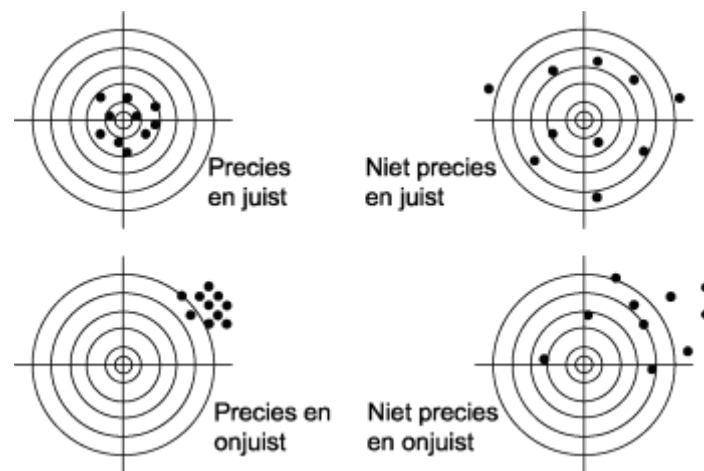
4.5 Juistheid en precisie van een gevalideerde loggermeting

De in figuur 4.7 vermelde resultaten kunnen ook worden aangeduid als de (on)juistheid en de precisie van een gevalideerde loggermeting, die samen zijn nauwkeurigheid bepalen:

- Juistheid - Een meetmethode kan worden aangeduid als juist als deze geen systematische fout vertoont. De gemiddelde meetfout bedraagt dan nul. Als de gemiddelde meetfout afwijkt van nul kan de meetmethode worden aangeduid als onjuist.
- Precisie – De precisie van een meetmethode is gerelateerd aan de spreiding van de meetfout. Deze kan bijvoorbeeld worden uitgedrukt in de standaardafwijking van de meetfout.

In figuur 4.8 is aangegeven hoe de juistheid en de precisie samen de nauwkeurigheid van een meetmethode bepalen.⁸

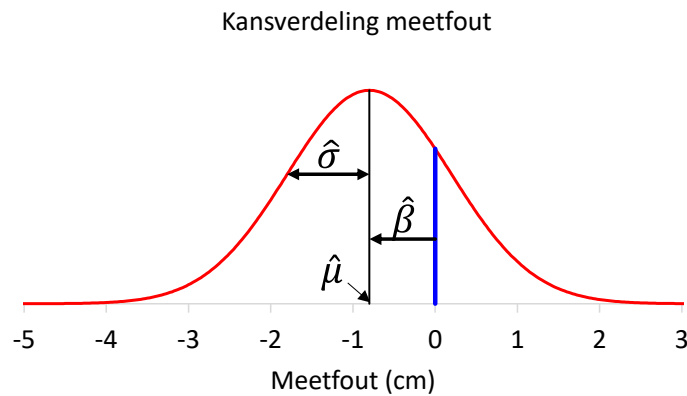
Figuur 4.8. Samenhang tussen de begrippen juistheid, precisie en nauwkeurigheid. Linksboven is de nauwkeurigheid het grootst en rechtsonder het kleinst.



Afgaande op de in figuur 4.7 vermelde resultaten is een door de Provincie Overijssel gevalideerde loggermeting van de grondwaterstand onjuist, met een systematische fout (β) die is geschat als -0,8 cm en statistisch significant negatief is (bij een gehanteerde betrouwbaarheid van 95%). De precisie van de meting (σ) is geschat als 1 cm. Figuur 4.9 toont de kansverdeling van de meetfout, met deze geschatte dimensies.

⁸ In de statistische en metrologische literatuur zijn ook andere definities van *nauwkeurigheid* (Engels: *accuracy*) te vinden. Het is daarom zaak altijd eenduidig in een rapport te definiëren welke gehanteerd wordt.

Figuur 4.9. De in deze studie geschatte dimensies van de kansverdeling van de meetfout van een door de Provincie Overijssel gevalideerde loggermeting van de grondwaterstand.



Toelichting

$\hat{\mu}$: geschatte gemiddelde meetfout (hier -0,8 cm)

$\hat{\beta}$: geschatte systematische meetfout (hier -0,8 cm, omdat $\hat{\beta} = \hat{\mu} - 0$)

$\hat{\sigma}$: geschatte standaardafwijking van de meetfout (hier 1,0 cm)

Kenmerken ε_{L1}	Gemiddelde	Standaardafw
Schatting (cm)	-0,81	0,96
Ondergrens 95%-b.i. (cm)	-1,05	0,33
Bovengrens 95%-b.i. (cm)	-0,57	1,49

4.6 Conclusies

Uit deze verkenning bleek het volgende over de kansverdeling van de meetfout van een door de Provincie Overijssel gevalideerde loggermeting van de grondwaterstand:

- de gemiddelde meetfout is geschat op -0,8 cm, maar gegeven de onzekerheden kan deze met 95% betrouwbaarheid ergens in het interval van -1,0 tot -0,6 cm liggen. Er is daarmee met 95% betrouwbaarheid sprake van een statistisch significante negatieve systematische meetfout;
- de standaardafwijking van de meetfout is geschat op 1,0 cm, maar gegeven de onzekerheden kan deze met 95% betrouwbaarheid ergens in het interval van 0,3 tot 1,5 cm liggen. Dit houdt in dat het niet aannemelijk is dat er vaak toevallige meetfouten groter dan +3 cm of kleiner dan -3 cm optreden (bovenop de bovengenoemde systematische fout).

Aangehaalde literatuur

Baggelaar, P.K. en Van der Meulen, E.C.J. (2019): *Kenmerken meetfout bij automatisch meten grondwaterstand door Provincie Overijssel*. AMO en Icastat, 14 mei 2019, Hengelo, 76 blz.

Chernick, M. (2007). *Bootstrap methods: A guide for researchers and practitioners*. Wiley, New York, 400 blz.

Efron, B. (1979). *Bootstrap methods: Another look at the jackknife*. The Annals of Statistics. 7 (1): 1–26. doi:10.1214/aos/1176344552.

Efron, B. (1987). *Better Bootstrap Confidence Intervals*. Journal of the American Statistical Association, Vol. 82, No. 397. 82 (397): 171–185. doi:10.2307/2289144.

Gooijer, J. en De Meij, T. (2018): *Werkwijze meten grondwaterstanden Provincie Overijssel*. Interne notitie Provincie Overijssel, 3 november 2013, aangepast 12 juni 2018, 8 blz.

Knotters, M., De Meij, Th. en Pleijter, M. (2013): *Nauwkeurigheid van handmatig gemeten grondwaterstanden en stijghoogtes - Verslag van een veldexperiment*. Alterra Wageningen UR, rapport 2476, december 2013, 30 blz.

Pleijter, M., Van Hamersveld, L. en Knotters, M. (2015): *Systematische fouten in metingen van grondwaterstanden met drukopnemers: verslag van een data-analyse*. Alterra Wageningen UR, rapport 2666, September 2015, 23 blz.

Post, V.E.A. and Von Asmuth, J.R. (2013): *Review: Hydraulic head measurements—new technologies, classic pitfalls*. Hydrogeology Journal, 21(4), June 2013, pp. 737 - 750.

Von Asmuth, J.R., Maas, K. Knotters, M. en Leunk, I. (2010): *Handleiding Menyanthes*. KWR, Nieuwegein, 52 blz.

Von Asmuth, J.R. (2010). *Over de kwaliteit, frequentie en validatie van druksensorreeksen (t.b.v. een optimale meetfrequentie en verwerkingsprotocol voor verdrogingsmeetnet Overijssel)*. Nieuwegein, november 2010, 44 blz.

Von Asmuth, J.R. en Van Geer, F.C. (2015): *Kwaliteitsborging grondwaterstands- en stijghoogtegegevens: systematiek en methodiek voor datakwaliteitscontrole (QC)*. KWR-rapport 2015.004, april 2015, 56 blz.

Bijlage: Lijst verschilwaarden L1-SH en L0-SH

Toelichting

nr : steekproefnummer

L0 : ruwe loggermeting van de grondwaterstand (cm NAP)

L1 : gevalideerde en zonodig gecorrigeerde loggermeting van de grondwaterstand (cm NAP)

SH : steekproefhandmeting (cm NAP)

SHc: steekproefhandmeting na correctie (cm NAP)

nr	L1-SH	L0-SH	L1-SHc	L0-SHc
1	-1,600	-1,608	-1,600	-1,608
2	-2,700	-7,535	-2,700	-7,535
3	-0,800	0,403	-0,800	0,403
4	-1,600	0,436	-1,600	0,436
5	-1,300	-1,287	-1,300	-1,287
6	-1,600	-1,734	-1,600	-1,734
7	-0,900	-3,100	-0,900	-3,100
8	-2,500	-2,861	-2,500	-2,861
9	-1,100	-1,882	-1,100	-1,882
10	-1,800	-2,899	-1,800	-2,899
11	-1,700	-1,657	-1,700	-1,657
12	-0,300	-0,846	-0,300	-0,846
13	-1,200	-0,654	-1,200	-0,654
14	-2,200	-3,184	-2,200	-3,184
15	-2,100	-1,641	-2,100	-1,641
16	-1,300	-2,481	-1,300	-2,481
17	-0,700	-0,575	-0,700	-0,575
18	-1,100	-1,395	-1,100	-1,395
19	-0,800	-1,602	-0,800	-1,602
20	-1,300	-0,429	-1,300	-0,429
21	-1,400	-1,367	-1,400	-1,367
22	-2,500	-2,534	-2,500	-2,534
23	-1,600	-1,598	-1,600	-1,598
24	-0,400	0,581	-0,400	0,581
26	-3,600	-2,586	-3,600	-2,586
27	-0,400	-0,107	-0,400	-0,107
28	-0,500	0,290	-0,500	0,290
29	1,200	0,572	1,200	0,572
30	-2,100	-1,460	-2,100	-1,460
31	-3,900	-1,344	-3,900	-1,344
32	-2,200	-5,172	-2,200	-5,172
33	-0,300	-1,229	-0,300	-1,229
34	-0,200	-0,161	-0,200	-0,161
35	-0,700	-1,175	-0,700	-1,175
36	-2,000	-2,259	-2,000	-2,259
37	-1,800	-2,956	-1,800	-2,956
39	-2,400	-2,412	-2,400	-2,412
40	-1,200	-1,434	-1,200	-1,434
41	-1,400	-1,373	-1,400	-1,373
42				
43	-1,200	-1,547	-1,200	-1,547
44	-1,300	-0,112	-1,300	-0,112
45	-0,900	-0,552	-0,900	-0,552

nr	L1-SH	L0-SH	L1-SHc	L0-SHc
46	-0,700	-0,444	-0,700	-0,444
47	-0,100	-2,568	-0,100	-2,568
48	-0,800	-0,437	-0,800	-0,437
49	-0,100	-0,327	-0,100	-0,327
50	-0,800	-4,239	-0,800	-4,239
51	-0,800	-0,409	-0,800	-0,409
52	-0,200	-1,623	-0,200	-1,623
53	-0,600	-0,286	-0,600	-0,286
54	-2,700	-2,717	-2,700	-2,717
55	-3,100	-3,126	-3,100	-3,126
56	-1,100	-2,110	-1,100	-2,110
57	-0,500	-0,715	-0,500	-0,715
58	-1,800	-2,717	-1,800	-2,717
59	-1,600	-1,552	-1,600	-1,552
62	0,200	-0,091	0,200	-0,091
63	-0,700	-1,336	-0,700	-1,336
64	-2,000	-4,037	-2,000	-4,037
65	-2,100	-2,089	-2,100	-2,089
66	-0,900	-1,374	-0,900	-1,374
67				
68				
69	-1,900	-1,893	-1,900	-1,893
70	-1,700	-1,739	-1,700	-1,739
71	-1,300	0,226	-1,300	0,226
72	0,100	0,582	0,100	0,582
73	0,100	0,729	0,100	0,729
74	1,000	0,952	1,000	0,952
75	-0,300	1,680	-0,300	1,680
76	-1,300	-2,209	-1,300	-2,209
77	-2,900	-2,947	-2,900	-2,947
78	-1,100	-2,290	-1,100	-2,290
79	-1,100	-1,413	-1,100	-1,413
80	-1,800	-1,813	-1,800	-1,813
81	-0,500	-0,334	-0,500	-0,334
82	-1,300	-1,713	-1,300	-1,713
83	-0,300	-0,009	-0,300	-0,009
85	-0,700	2,318	-0,700	2,318
86	-2,600	-2,069	-2,600	-2,069
87	-1,000	-1,959	-1,000	-1,959
88	-0,100	-1,218	-0,100	-1,218
90	-1,000	-1,498	-1,000	-1,498
91	-3,400	-1,400	-3,400	-1,400
92	-1,500	-0,066	-1,500	-0,066
93	-2,000	-1,965	-2,000	-1,965
94	-1,400	-0,353	-1,400	-0,353
95	-1,400	-0,990	-1,400	-0,990
96	-0,900	-3,142	-0,900	-3,142
97	-1,600	-0,416	-1,600	-0,416
98	-0,900	-0,493	-0,900	-0,493
99	-0,600	0,360	-0,600	0,360
100	-3,300	-3,327	-3,300	-3,327
101	1,900	2,136	1,900	2,136
102	-1,900	-1,096	-1,900	-1,096

nr	L1-SH	L0-SH	L1-SHc	L0-SHc
103	-1,100	-0,883	-1,100	-0,883
104	-1,200	-1,475	-1,200	-1,475
105	-4,800	-2,426	-4,800	-2,426
106				
107	-3,000	-3,017	-3,000	-3,017
108	-0,800	-0,529	-0,800	-0,529
109	0,700	0,355	0,700	0,355
110	-1,900	-5,856	-1,900	-5,856
111	-0,900	-1,426	-0,900	-1,426
112	-1,700	-2,153	-1,700	-2,153
113				
114	-0,200	-0,091	-0,200	-0,091
115	0,000	-0,045	0,000	-0,045
116	-2,100	-1,091	-2,100	-1,091
117	-2,300	-1,534	-2,300	-1,534
118	-0,700	-0,689	-0,700	-0,689
119	-0,100	-0,353	-0,100	-0,353
120	-1,200	-0,518	-1,200	-0,518
121	-0,700	-1,176	-0,700	-1,176
122	-1,000	-1,043	-1,000	-1,043
123	1,500	1,495	1,500	1,495
124	-1,900	-5,311	-1,900	-5,311
125	-1,500	-1,458	-1,500	-1,458
126	0,300	-0,722	0,300	-0,722
127	-0,600	-0,603	-0,600	-0,603
129	-11,700	-11,656	-11,700	-11,656
130	-0,400	-0,391	-0,400	-0,391
131	-1,800	-1,814	-1,800	-1,814
132	-0,200	-0,167	-0,200	-0,167
133	-1,100	-0,859	-1,100	-0,859
134	-0,900	-0,877	-0,900	-0,877
136	-2,500	-1,710	-2,500	-1,710
137	-0,200	-1,943	-0,200	-1,943
138	-4,900		-4,900	
139	-3,100	-3,086	-3,100	-3,086
140	-1,700	-1,725	-1,700	-1,725
141	-1,500	-0,560	-1,500	-0,560
142	-1,600	0,163	-1,600	0,163
143	-0,500	-0,801	-0,500	-0,801
144	-1,000	-1,647	-1,000	-1,647
146	-3,300	-0,765	-3,300	-0,765
147	-4,300	-2,267	-4,300	-2,267
148	-1,600		-1,600	
149	-0,600	0,366	-0,600	0,366
150	-2,000	-2,429	-2,000	-2,429
151				
152	-2,600	-4,084	-2,600	-4,084
153	-1,500	-1,533	-1,500	-1,533
154	-1,600		-1,600	
155	-1,400		-1,400	
156	-1,200		-1,200	
157	-1,300	-1,848	-1,300	-1,848
159	-0,900		-0,900	

nr	L1-SH	L0-SH	L1-SHc	L0-SHc
160	-1,000		-1,000	
161	-0,500		-0,500	
162	-0,300		-0,300	
163	-0,100		-0,100	
164	-2,100		-2,100	
165	-1,200		-1,200	
166	-1,100		-1,100	
167	-1,200		-1,200	
168	-2,000		-2,000	
169	-1,400	-1,834	-1,400	-1,834
170	-1,700	-1,725	-1,700	-1,725
171	-1,000		-1,000	
172	-0,700		-0,700	
173				
174	-1,700		-1,700	
175	-1,200		-1,200	
176	-2,300		-2,300	
177	-0,500	1,987	-0,500	1,987
178	-1,000	-1,004	-1,000	-1,004
179	-2,200	-1,206	-2,200	-1,206
180	-1,000	-0,850	-1,000	-0,850
181	-1,600	-2,603	-1,600	-2,603
182	-1,000		-1,000	
183	-3,500		-3,500	
184	-1,800	0,030	-1,800	0,030
185	-0,400	-0,393	-0,400	-0,393
186	-1,000	-2,450	-1,000	-2,450
188	-1,700	-1,532	-1,700	-1,532
189	-2,300	-4,694	-2,300	-4,694
190	-0,600	1,412	-0,600	1,412
191				
192	0,100	-0,675	0,100	-0,675
194	-0,200	0,439	-0,200	0,439
196	-1,200	-1,166	-1,200	-1,166
197	-1,600	-1,558	-1,600	-1,558
198	-31,900	-31,904	-1,900	-1,904
199	-0,200		-0,200	
200	-1,200		-1,200	
324	-0,800	-0,549	-0,800	-0,549
326	0,800	1,395	0,800	1,395
524	-0,300	-0,370	-0,300	-0,370
529	-0,800	-0,186	-0,800	-0,186
530	-1,900		-1,900	
531	-2,100	-2,235	-2,100	-2,235
533	-1,900		-1,900	
536	-1,700	-1,357	-1,700	-1,357
537	0,800	0,783	0,800	0,783
538	-0,300	-0,251	-0,300	-0,251
540	-0,600	-0,977	-0,600	-0,977
541	-1,200	-1,191	-1,200	-1,191
547	-0,100		-0,100	