Notitie

Ministerie van Verkeer en Waterstaat
Directoraat-Generaal Rijkswaterstaat
Meetkundige Dienst

Aan
Monitoring Maas projectgroep

Van
Ardis Bollweg
Marc Crombaghs
Regine Brügelmann
Erik de Min

Doorkiesnummer
015 - 269 1480

Datum
4 juni 2001

Bijlage(n)
-

Onderwerp
Precisie

Wat is precisie?
Deze ogenschijnlijk simpele vraag blijkt in de praktijk de aanleiding te zijn voor interessante discussies. De verschillende interpretaties van precisie leiden vaak tot ongewenste situaties.

> *Intermezzo:*
> *Veel mensen praten over nauwkeurigheid in plaats van precisie. Dit verschilt nogal per vakgebied. Met precisie bedoelen we meestal iets met een standaardafwijking of 95%-kansgebied. In de geodesie wordt met nauwkeurigheid zowel precisie als betrouwbaarheid bedoeld, en betrouwbaarheid is de kans dat je een bepaalde fout kunt vinden. In de GIS-wereld wordt in het engels met 'precision' de toevallige fout bedoeld, en met 'áccuracy'(=nauwkeurigheid) de systematische fout. Om verdere misverstanden te voorkomen hebben we het hier verder over precisie.*
> *De term 'kwaliteit' tenslotte, is een koepelbegrip, waarbij behalve precisie en betrouwbaarheid, ook actualiteit, beschikbaarheid, consistentie, integriteit en volledigheid een rol (kunnen) spelen (zie het MD-kwaliteitsbegrippenhandboek).*

Een voorbeeld van verschillende interpretaties vormt het project Monitoring Maas, waarbij verschillende partijen precisie-eisen hebben gespecificeerd, waarmee meetplannen opgesteld zullen worden. Een bloemlezing:

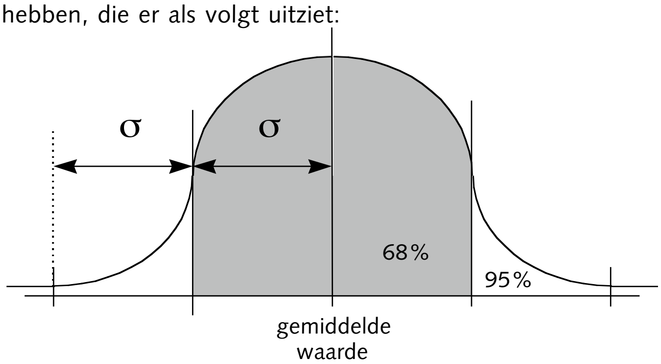
- cm ± 1
- m^3/s ± 3-5%
- enkele cm's tot dm
- < 10 cm / 100 m
- c-factor 5
- ± factor 2 (relatief)
- 5 stenen bij elkaar
- ng/l
- vracht in ton per jaar
- aanwezigheid / afwezigheid

Postadres: Postbus 5023, 2600 GA Delft
Bezoekadres: Kanaalweg 3b, 2628 EB Delft
Telefoon (015) 269 11 11
Telefax 015 - 2618962
E-mail a.e.bollweg@mdi.rws.minvenw.nl

Deze specificaties zijn vaak onvolledig en multi-interpretabel. Dit wordt in deze notitie uitgelegd aan de hand van een voorbeeld.

Stel: een beleidsafdeling zegt dat men de bodemligging met een precisie van cm ± 10 wil weten. Deze eis zegt wel íets over de amplitude (de grootte) van de fout die toe wordt gestaan, maar lang niet alles. Stel dat de bodemligging is gegeven in gridcellen van 1 m x 1 m voor een stuk Maas van 10 m x 100 m. Totaal zijn er dan dus 1000 waardes gegeven. Moeten die allemaal minder dan 10 cm van de werkelijkheid afwijken? Of is het niet zo erg als er soms een foutje van 20 cm voorkomt? En hoe vaak mag dat dan gebeuren? En een fout van 50 cm?

Om met dit soort vragen om te kunnen gaan, is het handig om met kansverdelingen te werken. Een kansverdeling geeft weer hoe vaak een meetwaarde voor een bepaalde grootheid (b.v. een afstand) te verwachten is als deze grootheid heel vaak gemeten wordt. Het vaakst zal dus de "ware" (of gemiddelde) waarde gemeten worden maar heel vaak ook niet. Een kansverdeling geeft dus een beeld van de spreiding in waarnemingen (zoals lodingen) op grond van meetfouten, of in daaruit afgeleide grootheden (zoals bodemligging). Uit ervaring is gebleken dat veel variabelen (bij benadering) een 'normale kansverdeling' hebben, die er als volgt uitziet:

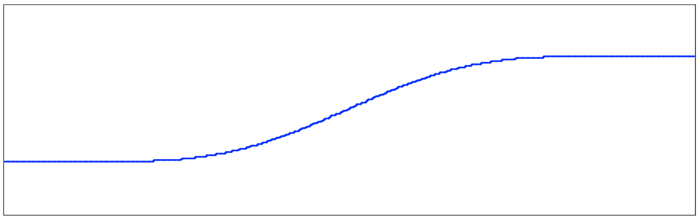


De spreiding wordt normaal gesproken gekwantificeerd met de afstand tussen de gemiddelde waarde μ en de waarde bij de buigpunten van zo'n kansverdeling. Dit wordt de standaardafwijking σ genoemd. Voor een normale verdeling geldt dat 68% van de waardes binnen het interval $[\mu-1\sigma\ ;\ \mu+1\sigma]$ valt, en dat 95% van de waardes binnen het interval $[\mu-2\sigma\ ;\ \mu+2\sigma]$ valt, zie ook onderstaande tabel.

$[\mu-1\sigma\ ;\ \mu+1\sigma]$	68%
$[\mu-2\sigma\ ;\ \mu+2\sigma]$	95%
$[\mu-3\sigma\ ;\ \mu+3\sigma]$	99%
$[\mu-4\sigma\ ;\ \mu+4\sigma]$	100%

Terug nu naar het voorbeeld. Als de beleidsafdeling zegt dat men de bodemligging met een precisie van een standaardafwijking van 10 cm wil weten, dan kan er al meer gezegd worden over wat nu eigenlijk de bedoeling is: van de 1000 waardes mogen er 320 meer dan 10 cm afwijken van de werkelijkheid. Daarvan mogen er 50 zelfs meer dan 20 cm afwijken, en 10 mogen er tussen de 30 en 40 cm afwijken. Deze precisie-specificatie wordt ook vaak genoteerd als “20 cm (2σ)”, of als “20 cm met een 95% betrouwbaarheidsinterval“.

Maar daarmee is de kous nog niet af. De eis van σ = 10 cm zegt namelijk helemaal niets over de ruimtelijke verdeling van de fout die wordt toegestaan. Mogen die 50 waardes met een afwijking van meer dan 20 cm allemaal bij elkaar zitten in een gebiedje van 8 m x 8 m? Onderstaand een voorbeeldje dat het verschil nog eens laat zien.



De beide soorten meetfouten, de random fout (groen) en de systematische fout (rood) hebben allebei dezelfde standaardafwijking (=amplitude), maar een heel verschillend ruimtelijk karakter. De ene is random per meetpuntje en de daadwerkelijk optredende fout is per meetpunt niet te voorspellen, alleen dat de grootte binnen de normale verdeling zal liggen. De andere fout is systematisch van karakter, hij heeft voor groepen meetpunten die dicht bij elkaar liggen (min of meer) dezelfde grootte, pas een heel eind verderop is de grootte van de fout anders en ongecorreleerd.
Binnen een gebied van bepaalde omvang, b.v. 10 m x 10 m, is de systematische fout (min of meer) constant maar per 10 m x 10 m gebied ‘random’. Als we dus in een groter

gebied naar de systematische fouten van alle deze 10 m x 10 m gebieden kijken dan zijn deze weer normaal verdeeld.

In het middelste plaatje is uit de groene lijn veel beter het profiel van de rivier terug te vinden dan uit de rode. Dat komt omdat je op het oog de random-ruis aardig kunt wegdenken. Met filtertechnieken kun je dat ook echt wegfilteren (wegmiddelen). De rode lijn lijkt heel goed, omdat er geen ruis op voorkomt, dus je hebt de neiging om ook snel te denken dat hij zonder fouten is. Echter, de fouten hebben hier een heel ander karakter, ze zijn systematisch van aard, dus ze zijn voor een groter gebied, of voor een grote groep gemeten punten, hetzelfde. Pas voor een gebiedje verderop heeft deze fout weer een toevallige, niet te voorspellen waarde uit de normale verdeling. Dit soort verschillende fouten komen we steeds meer tegen doordat veel inwinsystemen (zoals bijvoorbeeld laseraltimetrie en multibeam) uit verschillende sensoren zijn samengesteld. Elk van deze sensoren heeft een eigen foutengedrag. En de ene sensor geeft per milli-seconde 'random'-fouten, terwijl de ander dat per 10 seconden doet, waarin meestal al een groot gebied is gemeten waarvoor deze fout dan systematisch is (terwijl het eigenlijke foutenkarakter per 10 seconden wel 'random' is).

Bij een standaardafwijking hoort dus in veel praktische situaties ook een gebiedsgrootte. Voor het voorbeeld zou de precisie-specificatie er dan als volgt uit kunnen zien:

Losse waardes	$\sigma = 20$ cm
5 m x 5 m	$\sigma = 5$ cm
10 m x 100 m	$\sigma = 2$ cm

De betekenis is als volgt: individuele bodemliggings-waardes mogen een standaardafwijking van 20 cm hebben. Dit wordt ook wel de puntruis genoemd. Dat betekent dus dat 5% van de waardes een fout van meer dan 40 cm mag hebben. De standaardafwijking van de gemiddelde waarde voor een gebied van 5 m x 5 m mag echter maar 5 cm zijn, en die van een gebied van 10 m x 100 m zelfs maar 2 cm.

Hieruit is af te leiden dat de beleidsafdeling het belangrijk vindt dat de gemiddelde bodemligging voor het totale gebied precies wordt vastgesteld, en dat fouten in de losse meetpunten helemaal niet zo erg zijn. Deze beleidsafdeling is bijvoorbeeld geïnteresseerd in baggervolumes: het gaat om het volume van het totale gebied; fouten in de losse meetpunten zijn niet zo belangrijk.

Als men de vaardiepte in kaart gebracht had willen brengen, dan was de eis voor de precisie van de losse meetpunten een stuk strenger geweest. De eis van $\sigma = 20$ cm betekent namelijk dat er fouten tot wel 80 cm voor kunnen komen in de bodemliggings-waarde, en dat kan voldoende zijn om een binnenvaartschip aan de grond te laten lopen. In dit geval is de precisie van de bodemligging van grotere gebieden niet zo belangrijk, zodat de tabel er bijvoorbeeld als volgt komt uit te zien:

Losse waardes	$\sigma = 6$ cm
5 m x 5 m	$\sigma = 6$ cm

10 m x 100 m	σ = 6 cm

In plaats van, of naast een ruimtelijke verdeling, kan ook een verdeling naar tijdseenheid worden gemaakt.

Op dezelfde manier kan voor verschillende inwintechnieken (verschillende apparaten, of verschillende manieren om een meetnet op te zetten) ook in een tabelletje worden beschreven welke soort fouten er optreden en met welke amplitude (standaardafwijking).
De totale fout kan berekend worden door alle optredende deelfouten kwadratisch bij elkaar op te tellen en daaruit de wortel te trekken. Voorbeeld: 12,7_ = (10_ + 6_ + 5_).

	Inwinscenario 1	Inwinscenario 2	Inwinscenario 3
Puntruis	σ = 10 cm	σ = 10 cm	σ = 5 cm
10 m x 10 m	σ = 6 cm	σ = 4 cm	σ = 6 cm
100 m x 100 m	σ = 5 cm	σ = 2 cm	σ = 5 cm
Totaal	σ = 12,7 cm	σ = 11,0 cm	σ = 9,3 cm
Kosten per ha	1000 hfl	2000 hfl	1200 hfl

In plaats van in een tabelvorm zijn er ook wiskundige technieken beschikbaar om een gecombineerde precisiebeschrijving te maken. Dit is de fouten-covariantiefunctie (ook bekend als variogram). Hiermee kan ook gemakkelijk écht worden gerekend.

Conclusie
Bij het opstellen van precisie-specificaties is het geven van één getalletje vaak niet voldoende. Er hoort vermeld te worden wat dit getalletje voorstelt (een standaardafwijking, een maximale waarde), en er moet onderscheid gemaakt worden naar ruimtelijke (of temporele) schaal. Pas dan kan de inwintechniek op het gewenste resultaat (de gewenste precisie) worden afgestemd.