

Foutenanalyse in waterbalansstudies

M.F.P. Bierkens

sc-dlo



32/44b(460) 2e ex

Foutenanalyse in waterbalansstudies

**BIBLIOTHEEK
STARINGGEBOUW**

M.F.P. Bierkens

Rapport 460

DLO-Staring Centrum, Wageningen, 1996

- 4 JULI 1996

13- 926974

REFERAAT

M.F.P. Bierkens, 1996. *Foutenanalyse in waterbalansstudies*. Wageningen, DLO-Staring Centrum. Rapport 460. 54 blz.; 2 tab.; 19 ref.

De berekening van de waterbalans geeft meestal een restterm, waaraan de kwel/wegzijging of waterinlaat wordt gelijkgesteld. Dit rapport behandelt de interpretatie van deze restterm, ervan uitgaande dat die het gevolg is van fouten bij de berekening van de waterbalansen. Een algemeen foutenmodel wordt geïntroduceerd waarmee kan worden getoetst of de restterm het gevolg is van een systematische fout in de waterbalansberekeningen, meetfouten of onverklaarde temporele en ruimtelijke variatie. Basistechnieken van foutenrekening worden behandeld om de nauwkeurigheid van de waterbalansen te bepalen. Verbetering van de nauwkeurigheid van de minst nauwkeurige term vergroot de nauwkeurigheid van de totale waterbalans het meest.

Trefwoorden: foutenrekening, hydrologie

ISSN 0927-4499

Tevens verschenen als onderdeel syllabus PHLO-cursus "Omgaan met de Waterbalans"

©1996 DLO-Staring Centrum, Instituut voor Onderzoek van het Landelijk Gebied (SC-DLO)
Postbus 125, 6700 AC Wageningen.
Tel.: (0317) 474200; fax: (0317) 424812; e-mail: postkamer@sc.dlo.nl

Niets uit deze uitgave mag worden veeleelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van DLO-Staring Centrum.

DLO-Staring Centrum aanvaardt geen aansprakelijkheid voor eventuele schade voortvloeiend uit het gebruik van de resultaten van dit onderzoek of de toepassing van de adviezen.

Inhoud

	blz.
Woord vooraf	7
Samenvatting	9
1 Inleiding	11
2 Beschrijving van het algemene foutenmodel	13
2.1 Restterm en schattingsfout	13
2.2 Stochastische variabelen en verwachtingswaarden	14
2.3 Stochastische analyse van de schatter R^*	16
3 Praktische toepassing van het foutenmodel	19
3.1 Het belang van de normale verdeling	19
3.2 Toetsen op een systematische afwijking	19
3.3 De onzekerheid van de waterbalans	20
3.4 R^* als schatter voor een onbekende balansterm	21
4 Foutenberekening van individuele termen	23
4.1 Meetfouten	23
4.2 Berekeningen	24
4.3 Tijdsintegratie	27
4.4 Ruimtelijke integratie	31
4.5 Toepassing op balanstermen	35
5 Foutenanalyse, meetnetoptimalisatie en monitoring	37
5.1 Het monitoren van de kwaliteit van een berekende waterbalans	37
5.2 Het monitoren van veranderingen in termen van de waterbalans	38
6 De stoffenbalans	41
7 Toepassing op waterbalans Polder Oudendijk	43
7.1 Inleiding	43
7.2 Het foutenmodel	44
7.3 Berekening van de varianties van de termen	45
7.4 Interpretatie van de restterm	47
7.5 Berekening van de kwel uit de restterm	49
8 Conclusies	51
Literatuur	53

Tabellen

1 Gegevens om de winterbalans van Polder Oudendijk te berekenen	43
2 Winterbalans Polder Oudendijk (in mm)	44

Woord vooraf

Als onderdeel van de cursus "Omgaan met de Waterbalans" van PHLO, gehouden op 9 en 10 november 1995, werd de docenten gevraagd om een onderdeel te schrijven voor een syllabus. Mijn onderdeel betrof de foutenanalyse. Mijn doel was om een zo algemeen mogelijk raamwerk te geven om met fouten en onzekerheid in waterbalansberekeningen om te gaan. Het resultaat is een algemeen foutenmodel voor de waterbalans als geheel en een beschrijving van een aantal basistechnieken van foutenrekening en onzekerheidsanalyse, om de schattingsvarianties van de individuele termen van de waterbalans te berekenen. Het geheel is een raamwerk waarbinnen de meetnetoptimalisatie voor de monitoring van de waterbalans gemakkelijk kan worden ingepast. Omdat de gepresenteerde methoden ook van belang kunnen zijn voor hydrologen en waterbeheerders die niet aan deze cursus hebben deelgenomen, is besloten mijn bijdrage aan de syllabus in aangepaste versie in een DLO-Staring Centrum rapport te vatten. Dit is gebeurt onder de vlag van het WDT-onderzoeksprogramma 228 van DLO "Patronen en variabiliteit van bodem en grondwater".

Mijn dank gaat uit naar ir. B. van der Veer van het Hoogheemraadschap van Rijnland te Leiden voor zijn toestemming om het praktische rekenvoorbeeld van de Polder Oudendijk te mogen gebruiken. Ik bedank ing. M. Knotters, ir. J.M.P.M. Peerboom en drs. ing. J.W.J. van der Gaast (DLO-Staring Centrum) en drs. J.H. Oude Voshaar (DLO-Groep Landbouw-Wiskunde) voor het leveren van waardevol commentaar op de concept-tekst.

Samenvatting

Bij de berekening van de termen van een waterbalans treden systematische en toevallige fouten op. Wanneer alle termen afzonderlijk worden berekend is er daarom sprake van een *sluitpost* of *restterm*. Ook wordt vaak een onbekende term van de waterbalans zoals kwel/wegzijging of waterinlaat gelijk gesteld aan deze restterm. De aldus berekende onbekende term kan hierdoor echter waarden vertonen die sterk afwijken van die men redelijkerwijs kan verwachten. Doel van dit rapport is om een handreiking te geven bij het interpreteren van de restterm. Hierbij wordt ervan uitgegaan dat de restterm het gevolg is van fouten of onzekerheden in de berekening van de waterbalans.

De hoofdstukken 2 t/m 6 van het rapport vormen het theoretisch gedeelte. In hoofdstuk 2 is een algemeen foutenmodel gepresenteerd voor de restterm. Met dit foutenmodel is het mogelijk om te toetsen of de restterm het gevolg is van een systematische fout in de waterbalansberekeningen of slechts het gevolg is van toevalligheden, zoals meetfouten (hoofdstuk 3). In het eerste geval dienen de berekeningsmethoden te worden verbeterd, en in het tweede geval is het beter om preciezer, frequenter of op meer plaatsen te meten. In het geval dat de restterm dienst doet als een schatting voor een onbekende balansterm, zoals kwel/wegzijging, kan worden getoetst of de geschatte waarde significant afwijkt van een fysisch redelijke waarde.

Fouten in waterbalansberekeningen kunnen het gevolg zijn van meetfouten, de berekening van afgeleide grootheden, tijdsintegratie en ruimte-integratie. In hoofdstuk 4 zijn deze foutenbronnen uitgebreid besproken en een aantal basistechnieken van foutenrekening is behandeld waarmee men de nauwkeurigheid van elk van de termen van de waterbalans kan bepalen. Verbetering van de nauwkeurigheid van de minst nauwkeurige term, via het preciezer, frequenter of op meer plaatsen meten en berekenen, levert dan de grootste winst in nauwkeurigheid op voor de totale waterbalans.

Bij het berekenen van afgeleide grootheden, zoals bijvoorbeeld de berekening van stuwaflow uit waterhoogten, worden fouten bij de invoer (i.c. waterhoogten) vertaald naar fouten in de uitvoer (i.c. de afvoer). Met behulp van de foutenvoortplanting (Heuvelink, 1993) kan worden nagegaan wat de fouten in de uitvoer zijn, gegeven de fouten in de invoer. Er zijn verschillende technieken van foutenvoortplanting voorhanden, zoals Taylor-methoden en Monte-Carlo simulatie. De keuze van de juiste methode hangt af van de relatie tussen de invoer- en de uitvoervariabelen. In paragraaf 4.2 is ingegaan op de keuze van de juiste techniek van foutenvoortplanting.

Een waterbalans wordt altijd opgesteld voor een bepaald beheersgebied en voor een bepaalde tijdsperiode (bijvoorbeeld een jaar). Dit betekent dat grootheden die gemeten of berekend zijn op een bepaald tijdstip of voor een bepaalde lokatie moeten worden gemiddeld over tijd en ruimte. Bij dit middelen ontstaan ook fouten omdat de waarden van de grootheden tussen de meettijdstippen of meetlokaties niet bekend

zijn. In de paragrafen 4.2 en 4.3 zijn twee methodieken besproken waarmee de groottes van de fouten die ontstaan bij respectievelijk tijds- en ruimtemiddeling kunnen worden geschat. De eerste methodiek, de steekproefbenadering, kan worden toegepast wanneer de meettijdstoppen of de meetlokaties aselekt zijn gekozen. Is dit niet het geval, dan kan de tweede methodiek, de stochastische-functiebenadering, worden toegepast.

Het foutenmodel en de beschouwde basistechnieken van foutenrekening zijn zeer algemeen. Elke hydrologische berekeningsmethode en elk deterministisch model past binnen dit raamwerk. Het algemene foutenmodel en de technieken van foutenrekening vormen samen een blauwdruk voor het uitvoeren van meetnetoptimalisaties ten behoeve van het monitoren van trends in de individuele balanstermen en het monitoren van de kwaliteit van de waterbalans als geheel. In hoofdstuk 5 is aangegeven hoe het algemene foutenmodel en de foutenrekening kan worden gebruikt bij monitoring en meetnetoptimalisatie van de waterbalans.

De technieken die in hoofdstukken 2 t/m 5 zijn besproken kunnen direct worden toegepast voor de foutenanalyse van de stoffenbalans, voor zover deze beperkt blijft tot conservatieve stoffen. In hoofdstuk 6 is hiervan een voorbeeld gegeven.

Ter illustratie is een praktische toepassing van de foutenanalyse gegeven in hoofdstuk 7. De toepassing betreft een foutenanalyse van de winterwaterbalans van Polder Oudendijk (Hoogheemraadschap Rijnland). De foutenanalyse toont aan dat er sprake is van een systematische fout in de berekening van de winterwaterbalans van Polder Oudendijk. Analyse van de individuele termen maakt duidelijk dat de grootste fouten optreden bij de berekening van de kwelterm. De toekomstige meetinspanning dient zich dus vooral te richten op het nauwkeuriger bepalen van de kwel in het gebied.

1 Inleiding

Waterbalansstudies spelen een belangrijk rol in het waterbeheer. Zij verschaffen inzicht in de grootte van de verschillende componenten van de hydrologische kringloop en de variatie ervan over de seizoenen. In het dagelijks waterbeheer verschaft de waterbalans inzicht in de grootte van een tijdelijk tekort of overschot aan water. Ook vormt de waterbalans een belangrijk hulpmiddel bij het evalueren van effecten van ingrepen in het beheersgebied.

Aangezien de verschillende termen van de waterbalans afzonderlijk worden gemeten of berekend, kan niet worden gegarandeerd dat de waterbalans kloppend is. Er zal water overblijven of tekort zijn: $\text{in} - \text{uit} - \text{bergingsverandering} \neq 0$. Er is dus sprake van een *sluitpost* of *restterm* in de waterbalans. De restterm is het gevolg van fouten die gemaakt worden bij de berekening van de termen van de waterbalans. Omdat op basis van waterbalansstudies vaak belangrijke beslissingen op het gebied van het waterbeheer worden genomen, is het van belang te weten hoe groot deze fouten dan wel zijn. Anders gezegd: het is van belang om te weten hoe nauwkeurig we de verschillende termen van de waterbalans kennen. Deze nauwkeurigheid kan dan worden meegewogen bij het nemen beheersbeslissingen en het evalueren van de effecten van beheersmaatregelen. Het belang van een maat voor de betrouwbaarheid van waterbalansstudies wordt steeds vaker onderkend. Dit blijkt ondermeer uit recente waterbalansstudies waarin de foutenanalyse een belangrijk onderdeel vormde van het onderzoek (Van Bakel en De Kruijff, 1994; Van Der Gaast en Peerboom, 1996). De foutenanalyses in deze studies zijn toegespitst op het specifieke onderzoeksprobleem en het beschouwde onderzoeksgebied. Het doel van dit rapport is om een algemene methodiek te presenteren waarmee foutenanalyses van waterbalansstudies kunnen worden uitgevoerd. De methodiek kan dienen als een raamwerk waarbinnen elke berekeningsmethode of deterministisch model kan worden toegepast.

In waterbalansstudies kan op twee wijzen met de restterm worden omgegaan. De eerste manier is om deze niet apart te berekenen, maar hem te verdisconteren in bergingsveranderingen van oppervlaktewater, grondwater of bodemvocht of moeilijk te meten termen als kwel, welke dan zelf als sluitpost worden berekend. De tweede manier is om de restterm expliciet te berekenen, na afzonderlijke berekening van alle termen van de waterbalans. De tweede methode is verstandiger, want bij de eerste methode worden fouten in de berekening van de waterbalans – we zullen later zien welke fouten hier worden bedoeld – alleen toegeschreven aan fouten in bepaalde termen van de waterbalans, terwijl deze net zo goed het gevolg kunnen zijn van fouten in andere termen. Een ander nadeel van de eerste methode is dat we bij een dergelijke berekening geen enkel idee hebben hoe goed we voor een bepaalde periode de waterbalans als geheel kennen. Door de restterm te berekenen hebben we daar wel een idee van. Des te groter de restterm, des te groter de fout in de waterbalans. In dit rapport gaan we er daarom voornamelijk van uit dat er volgens de tweede methode met de restterm wordt omgegaan. We zullen zien dat het mogelijk is om te berekenen hoe nauwkeurig we die restterm kennen en te toetsen of deze het gevolg

is van onzekerheid in de waterbalansberekeningen of dat deze meer systematisch van aard is. Omdat in de praktijk de restterm toch vaak gebruikt wordt als een schatting van een onbekende balansterm, zal ook kort worden ingegaan hoe in dat geval met de onzekerheid van de restterm kan worden omgegaan.

In hoofdstuk 2 wordt eerst een algemeen foutenmodel voor de totale waterbalans opgesteld. Vervolgens wordt in hoofdstuk 3 het belang van de normale verdeling besproken. Getoond wordt hoe kan worden getoetst of de restterm systematisch of toevallig is. Ook wordt besproken hoe de onzekerheid in de restterm kan worden geïnterpreteerd als deze wordt gebruikt als schatting voor een onbekende balansterm. Vervolgens wordt behandeld hoe het foutenmodel kan worden gebruikt om tot een verbetering van de waterbalansberekeningen te komen. Hoofdstuk 4 geeft een beschrijving van een aantal technieken om de grootte van de fout in de individuele termen van de waterbalans te berekenen. In hoofdstuk 5 wordt kort ingegaan op de centrale rol die een foutenanalyse kan spelen bij meetnetoptimalisatie in monitoring systemen. Hoofdstuk 6 toont dat het algemene foutenmodel en de beschreven technieken van foutenrekening direct kunnen worden toegepast op de stoffenbalans van conservatieve stoffen. In hoofdstuk 7 wordt een praktische toepassing van de methodiek gegeven door middel van een foutenanalyse van de winterwaterbalans van Polder Oudendijk (Hoogheemraadschap Rijnland). Hoofdstuk 8 besluit het rapport met een opsomming van de belangrijkste conclusies.

2 Beschrijving van het algemene foutenmodel

2.1 Restterm en schattingsfout

We nemen als voorbeeld de volgende waterbalans, waarbij de termen zijn uitgedrukt als totale hoeveelheid water in een bepaalde periode (dag, week, maand, jaar) per oppervlakte-eenheid. De eenheid van de termen is dus in m:

$$r = p + k - e - d - \Delta o - \Delta g - \Delta b \quad (1)$$

met

p	: neerslag
k	: kwel (>0) of infiltratie (<0)
e	: evapotranspiratie
d	: gebiedsafvoer
Δo	: bergingsverandering in oppervlaktewater
Δg	: bergingsverandering in grondwater
Δb	: bergingsverandering in bodemvocht
r	: restterm

We gaan dus uit van een restterm die een maat is voor de fout in de totale waterbalans: als de restterm r ongelijk is aan 0 moet er sprake zijn van een onregelmatigheid in de waterbalans voor de beschouwde periode. Dit betekent dat er iets fout is met de definities of berekeningen van de individuele termen of dat er één of meerdere termen ten onrechte niet in de waterbalans voorkomen of er juist ten onrechte in zijn opgenomen.

In de praktijk echter zijn de termen van de waterbalans niet exact bekend. Zij worden immers berekend met methoden en modellen waarvoor benaderende veronderstellingen zijn gedaan. Ook kan er sprake zijn van meetfouten of onzekere schattingen, omdat nu eenmaal niet overal en altijd gemeten kan worden. De waterbalans van vergelijking (1) wordt dus in de praktijk benaderd door een waterbalans bestaande uit geschatte termen:

$$R^* = P^* + K^* - E^* - D^* - \Delta O^* - \Delta G^* - \Delta B^* \quad (2)$$

De asterisks geven aan dat het gaat om "schatters" van de termen van de waterbalans. Waarom hoofdletters worden gebruikt wordt uitgelegd in de volgende paragraaf. Zoals gezegd zijn de geschatte termen niet vrij van fouten. Dit heeft tot gevolg dat ook de restterm niet foutloos wordt berekend. De fout die gemaakt wordt, laten we die R' noemen, is het verschil tussen de schatter van de restterm en de werkelijke waarde ervan: $R' = R^* - r$. Andersom kunnen we de schatter van de restterm uitdrukken als:

$$R^* = r + R' \quad (3)$$

We zien dus dat de geschatte restterm bestaat uit een feitelijke restterm r en een schattingsfout R' . De feitelijke restterm kan beschouwd worden als een systematische fout in de waterbalans en wordt veroorzaakt doordat de berekeningsmethoden van de verschillende termen van (1) systematische fouten bevatten of omdat één of meerdere termen ten onrechte niet in de waterbalans voorkomen of er juist ten onrechte in opgenomen zijn. De systematische fout in de waterbalans kan alleen worden weggewerkt door een verandering van de berekeningsmethoden of door een betere calibratie van de meettechnieken. De schattingsfout daarentegen is een toevallige fout. In tegenstelling tot een systematische fout levert een toevallige fout, bij herhaalde berekeningen met andere sets metingen uit dezelfde periode telkens andere waarden op. Toevallige fouten zijn het gevolg van een complex van veelal snel veranderende factoren, zoals onverklaarde ruimtelijke en temporele variatie en onbekende verstoringen van de metingen, en zijn moeilijk te beheersen. Zij kunnen veelal alleen worden gereduceerd door vaker en preciezer te meten.

2.2 Stochastische variabelen en verwachtingswaarden

In de foutenrekening wordt de toevallige fout beschouwd als een uitkomst van een kansexperiment (denk aan het gooien van een dobbelsteen). Dit betekent dat de schattingsfout R' en dus ook de schatter R^* zogenaamde *stochastische variabelen* zijn: We weten niet welke waarde R^* precies heeft, maar we weten wel wat voor waarden R^* kan aannemen en met welke kans we deze waarden kunnen verwachten. De functie $f(r^*)$ die aangeeft met welke kans een bepaalde waarde r^* van R^* kan voorkomen wordt de *kansdichtheidsfunctie* genoemd. De bekendste kansdichtheidsfunctie is de belvormige Gauss-curve. Een stochastische variabele met een dergelijke kansdichtheidsfunctie noemt men ook wel normaal verdeeld. Hier dient opgemerkt te worden dat de stochastische variabele meestal met een hoofdletter wordt geschreven (zoals R^*) terwijl de kleine letter (zoals r^*) dan staat voor één feitelijke uitkomst van R^* (in dit geval: R^* is de schatter en r^* de schatting). Dit verschil lijkt wat academisch maar kan in bepaalde gevallen verwarring voorkomen. Hier wordt het onderscheid gemaakt omdat het in de statistische literatuur gebruikelijk is (zie o.a. Papoulis, 1986). De asterisks in vergelijking (2) geven dus aan dat het hier gaat om schatters van de termen van de waterbalans en de hoofdletters geven aan dat deze schatters stochastische variabelen zijn; er is dus sprake van schattingsfouten.

Als een variabele X een stochastische variabele is, dan heeft deze twee belangrijke maten: de verwachtingswaarde en de variantie.

De *verwachtingswaarde* $E[X]$ geeft het gemiddelde van de stochastische variabele weer en wordt berekend uit de kansdichtheidsfunctie als:

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx \quad (4)$$

De *variantie* $\text{var}(X)$ is een maat voor de breedte van het waardenbereik van de stochastische variabele:

$$\text{var}(X) = \int_{-\infty}^{\infty} (x - E[X])^2 f(x) dx \quad (5)$$

Als symbool voor de variantie wordt vaak de Griekse letter sigma gebruikt: $\text{var}(X) = \sigma_x^2$. In plaats van de variantie gebruikt men als spreidingsmaat ook wel de *standaardafwijking* σ_x , welke wordt verkregen door de wortel uit de variantie te nemen.

Om met de verwachting en de variantie te rekenen is het van belang een aantal regels te kennen. In het volgende zijn X en Y stochastische variabelen en α en β constanten. De volgende eigenschappen gelden voor één variabele:

$$E[\alpha] = \alpha \quad (6)$$

$$E[\alpha X] = \alpha E[X] \quad (7)$$

$$\text{var}(\alpha X) = \alpha^2 \text{var}(X) \quad (8)$$

Uit de definities van verwachting (4) en variantie (5) kunnen we afleiden dat we de variantie ook kunnen uitdrukken in termen van verwachtingswaarden:

$$\text{var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2 \quad (9)$$

Voor meer variabelen en constanten vinden we:

$$E[\alpha + \beta] = \alpha + \beta \quad (10)$$

$$E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y] \quad (11)$$

$$\text{var}(\alpha X + \beta Y) = \alpha^2 \text{var}(X) + \beta^2 \text{var}(Y) + 2\alpha\beta \text{cov}(X, Y) \quad (12)$$

waarbij $\text{cov}(X, Y)$ staat voor de *covariantie* tussen X en Y en wordt gegeven door:

$$\text{cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E[X])(y - E[Y]) f(x, y) dx dy \quad (13)$$

In deze vergelijking staat $f(x, y)$ voor een kansdichtheidsfunctie van twee variabelen. Analooeg aan (9) kan de covariantie ook geschreven worden in termen van verwachtingswaarden:

$$\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y] \quad (14)$$

Bovenstaande regels vormen de basis voor het rekenen met stochastische variabelen. Als we aannemen dat de restterm wordt beschreven door vergelijking (3), dan vinden we voor de verwachting:

$$E[R^*] = r + E[R'] \quad (15)$$

In het algemeen wensen we alleen schatters van r waarvoor de schattingsfout gemiddeld 0 is, dus als we onze berekeningen voor een groot aantal perioden of voor verschillende sets metingen zouden herhalen, dan zou het gemiddelde van al de daarbij horende schattingsfouten 0 moeten zijn. In termen van verwachting betekent dit dat $E[R'] = 0$. Als dat het geval is dan noemen we R^* een zuivere schatter. De verwachting wordt dan:

$$E[R^*] = r \quad (16)$$

De variantie van R^* volgt uit vergelijking (9):

$$\text{var}(R^*) = E[(R^* - E[R^*])^2] = E[(R^* - r)^2] = E[R'^2] = \text{var}(R') \quad (17)$$

We zien in (16) en (17) dat de verwachting van de schatter gelijk is aan de feitelijke restterm, terwijl de variantie van de schatter gelijk is aan de variantie van de schattingsfout.

2.3 Stochastische analyse van de schatter R^*

Wanneer we de schatter van de restterm als een stochastische variabele beschouwen willen we ook weten wat de verwachting en variantie is van deze variabele. Dus als zoals in dit geval de restterm wordt opgevat als een fout in de waterbalans is de verwachting gelijk is aan de systematische fout in de waterbalans, terwijl de variantie een maat is voor de fout die we maken bij het schatten van r . Om deze te berekenen kunnen we in het algemeen geen gebruik maken van vergelijkingen (4) en (5) omdat de kansdichtheidsfunctie $f(r^*)$ meestal niet van tevoren bekend is. We kunnen ze echter wel afleiden uit de verwachtingen, varianties en covarianties tussen de verschillende termen in vergelijking (2). Om het wat algemener te maken gaan we uit van een algemene lineaire balansvergelijking van de volgende vorm:

$$r = \sum_{i=1}^n \alpha_i z_i \quad (18)$$

Hier is r de restterm van een waterbalans die berekend wordt uit n termen van de waterbalans z_i , $i = 1, \dots, n$. De α_i zijn gewichten. In geval van een gewone waterbalans zoals vergelijking (1) nemen de α_i maar twee waarden aan: 1 en -1. Elk van de termen in de waterbalans is niet exact bekend maar wordt geschat:

$$R^* = \sum_{i=1}^n \alpha_i Z_i^* \quad (19)$$

Net als R^* zijn de Z_i^* opgebouwd uit een gemiddelde waarde $E[Z_i^*]$ en een toevallige

variatie rond dit gemiddelde Z_i' . De variatie rond dit gemiddelde Z_i' is dan de toevallige fout die wordt gemaakt bij het schatten van z_i . Een andere interpretatie van Z_i' is dat het de onzekere fluctuatie van z_i is. Dan is $\text{var}(Z_i^*)$ een maat voor de onzekerheid over de werkelijke waarde van z_i .

We veronderstellen nu dat we deze termen kunnen beschrijven met de volgende parameters ($i, j = 1, \dots, n$):

- $E[Z_i^*]$, de verwachting van de Z_i^* ;
- $\text{var}(Z_i^*) = \sigma_i^2$ de variantie van de Z_i^* (met σ_i de standaardafwijking);
- ρ_{ij} , de correlatiecoëfficiënt tussen Z_i^* en Z_j^* ; deze kan berekend worden uit de covarianties en de standaardafwijkingen van Z_i^* en Z_j^* :

$$\rho_{ij} = \frac{\text{cov}(Z_i^*, Z_j^*)}{\sigma_i \sigma_j} \quad (20)$$

Als bovenstaande statistische parameters bekend zijn voor alle waterbalanstermen dan kunnen hieruit de verwachting en de variantie van de schatter van de restterm worden berekend volgens:

$$E[R^*] = \sum_{i=1}^n \alpha_i E[Z_i^*] \quad (21)$$

$$\text{var}(R^*) = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \sigma_i \sigma_j \rho_{ij} \quad (22)$$

Vergelijkingen (21) en (22) zijn een algemenere vorm zijn van respectievelijk (11) en (12), maar nu voor een willekeurig aantal stochastische variabelen. Uit vergelijkingen (18) en (21) is te zien dat, om te zorgen dat R^* een zuivere schatter is ($E[R^*] = r$), het voldoende is om te eisen dat alle balanstermen zuiver worden geschat: $E[Z_i^*] = z_i$, $i = 1, \dots, n$. We moeten hierbij twee dingen opmerken:

- een term z_i bestaat uit een feitelijke hoeveelheid water + een eventuele systematische afwijking die bijdraagt tot de werkelijke waarde van r . Met een zuiver schatter van z_i wordt dus ook een schatter bedoeld die de term + zijn systematische afwijking zuiver schat;
- we kunnen weliswaar garanderen (of aannemen) dat de schatters van alle termen zuiver zijn en dat dus ook r zuiver wordt geschat, echter de echte waarden van de z_i en r komen we natuurlijk nooit te weten vanwege de toevallige schattingsfout.

Bij het toepassen van bovenstaande vergelijkingen op vergelijking (2) nemen we vooralsnog aan dat de schattingsfouten van de individuele termen onafhankelijk zijn. Dit betekent dat de correlatiecoëfficiënten tussen de schattingsfouten van de verschillende termen gelijk zijn aan 0. Dus alle termen in (22) voor $i \neq j$ worden gelijk aan 0. We houden alleen de termen voor $i=j$ over waarvoor geldt dat $\rho_{ii}=1$ (een stochastische variabele is perfect gecorreleerd met zichzelf).

Samenvattend, als we veronderstellen dat de termen van de waterbalans (1) zuiver worden geschat en dat schattingsfouten van de individuele termen onafhankelijk zijn, volgt het volgende algemene foutenmodel:

de restterm van de waterbalans wordt zuiver geschat door:

$$R^* = P^* + K^* - E^* - D^* - \Delta O^* - \Delta G^* - \Delta B^* \quad (23)$$

de variantie van de schattingsfout wordt gegeven door:

$$\text{var}(R^*) = \sigma_p^2 + \sigma_k^2 + \sigma_e^2 + \sigma_d^2 + \sigma_{\Delta o}^2 + \sigma_{\Delta g}^2 + \sigma_{\Delta b}^2 \quad (24)$$

waarbij $\sigma_p^2 = \text{var}(P^*)$, $\sigma_k^2 = \text{var}(K^*)$, etc.

Vergelijkingen (23) en (24) laten zien dat als de schattingen en de schattingsvarianties van de afzonderlijke termen van de waterbalans bekend zijn, we de restterm kunnen schatten en de bijbehorende schattingsvariantie kunnen berekenen. Zoals gezegd is de schattingsvariantie een maat voor de grootte van toevallige fout of, anders beschouwd, een maat voor de onzekerheid in de totale waterbalans.

3 Praktische toepassing van het foutenmodel

Als we in staat zijn om de restterm te schatten en de bijbehorende schattingsvariantie te berekenen kunnen deze een praktische rol spelen bij het waterbeheer. We zullen achtereenvolgens bekijken welke conclusies kunnen worden verbonden aan de berekende R^* en de variantie $\text{var}(R^*)$. Allereerst wordt echter bekeken welke kansdichtheidsfunctie we bij R^* kunnen verwachten.

3.1 Het belang van de normale verdeling

In vergelijking (2) zien we dat de stochastische variabele R^* in feite een som is van andere stochastische variabelen: de individuele termen van de waterbalans. Deze stochastische variabelen zijn veelal zelf weer sommaties van stochastische variabelen. Bijvoorbeeld de totale afvoer van een jaar kan geschat zijn uit een som van 365 dagafvoeren. De *centrale limietstelling* is een belangrijke regel in de statistiek die stelt dat de som van een groot aantal onafhankelijke stochastische variabelen met een willekeurige maar identieke kansdichtheidsfunctie bij benadering normaal verdeeld is. De schatter R^* voldoet weliswaar niet aan de eis dat de variabelen die gesommeerd worden dezelfde kansdichtheidsfunctie hebben en de individuele termen in de sommaties zijn niet allen onafhankelijk, maar uit de statistiek is eveneens bekend dat als het aantal variabelen dat gesommeerd wordt maar groot genoeg is de benadering van normaliteit ook in dat geval op gaat; voor de precieze definities wordt verwezen naar Papoulis (1986).

We kunnen dus verwachten dat de schatter van de restterm R^* bij benadering normaal verdeeld is. Deze veronderstelling gaat beter op wanneer de waterbalans over langere perioden wordt berekend. Het voordeel van normaal verdeelde variabelen is dat de bijbehorende kansdichtheids-functie de belvormige Gauss-curve is die kan worden beschreven met twee parameters: de verwachting en de variantie. De verwachting van R^* is gelijk aan r en is dus niet precies bekend. R^* is echter wel een zuivere schatter van r . Als we nu via (23) R^* en via (24) $\text{var}(R^*)$ kunnen berekenen, kennen we bij benadering ook de kansdichtheidsfunctie van R^* . We kunnen dan ook berekenen wat de kans is dat r tussen bepaalde grenzen zal liggen. Met andere woorden, we kunnen nu een betrouwbaarheidsinterval voor r opstellen.

3.2 Toetsen op een systematische afwijking

Zoals gesteld is het te verwachten dat R^* normaal verdeeld is. Het berekenen van R^* en $\text{var}(R^*) = \sigma_r^2$ is dus voldoende om te berekenen wat de kans is dat r tussen bepaalde grenzen ligt. Zo geldt dan dat r met een kans van 95% ligt tussen de grenzen $R^* - 2\sigma_r$ en $R^* + 2\sigma_r$. We noemen het interval $[R^* - \sigma_r, R^* + \sigma_r]$ ook wel het 95%-betrouwbaarheids-interval van r . We kunnen dus stellen dat, gegeven de

gebruikte berekeningsmethoden en de kwaliteit (= precisie, hoeveelheid en frequentie) van de gemeten variabelen, de feitelijke waarde van de restterm r met 95% zekerheid zal liggen in het interval $[R^* - 2\sigma_r, R^* + 2\sigma_r]$. Het 95%-betrouwbaarheidsinterval wordt veel gebruikt als toetsingscriterium. Een belangrijke toets is te kijken of de berekende R^* het gevolg is van toevallige schattingsfouten of dat er een systematische fout wordt gemaakt in de definitie of berekening van onze waterbalans. Dit komt neer op het toetsen of r significant van 0 afwijkt. In statistische termen:

H_0 : nul-hypothese: $r = 0$, d.w.z. geen systematische afwijking in de waterbalans

H_1 : alternatieve hypothese: $r \neq 0$, d.w.z. een systematische afwijking wordt gemaakt

Voor elke balansperiode kunnen R^* en $\text{var}(R^*) = \sigma_r^2$ worden berekend. De nul-hypothese wordt geaccepteerd als geldt dat $R^* \in [-2\sigma_r, 2\sigma_r]$, dus als de waarde van de schatter niet verder dan twee standaardafwijkingen van 0 af ligt. De kans dat ten onrechte wordt geconcludeerd dat er een systematische afwijking is in de berekende waterbalans is nu kleiner dan 5%.

Als de 0-hypothese niet wordt geaccepteerd dan is er wellicht met de feitelijke berekening van de waterbalans voor die periode wat aan de hand en is dit een reden om de berekening van de termen wat beter te bekijken of zelfs te herhalen. Eventueel moet uitgekeken worden naar betere berekeningsmethoden. De resultaten van een foutenberekening kunnen dus worden gebruikt als een soort waarschuwingsinstrument voor de berekende waterbalans.

3.3 De onzekerheid van de waterbalans

De waarde van $\text{var}(R^*)$ zegt iets over de grootte van de schattingsfout in R^* en dus over de onzekerheid van de berekende waterbalans: Als er een systematische fout is in de berekening van de waterbalans willen we die ook graag identificeren. Het is niet wenselijk dat systematische fouten in onze berekeningen ten onrechte worden toegeschreven aan toevallige fouten. Als op de in paragraaf 3.2 beschreven wijze wordt getoetst, dan betekent dit dat een systematische afwijking die ten onrechte niet wordt opgemerkt een grootte van maximaal $2\sigma_r$ kan hebben; anders gesteld: *de kleinste significante systematische afwijking is gelijk aan $2\sigma_r$* . De onzekerheid van de waterbalans is dus gelijk aan $2\sigma_r$.

Uit vergelijking (24) is te zien dat wanneer de schattingsfouten in de individuele termen van de waterbalans onafhankelijk zijn we in staat zijn te bepalen welke van de individuele termen het meest bijdraagt aan de onzekerheid over r . Door verbeteringen te richten op deze termen is de grootste verkleining in de onzekerheid van de totale waterbalans te verwachten. Zoals we in het volgende hoofdstuk zullen zien, hangen de schattingsvarianties van de meeste individuele termen sterk samen met de precisie van de metingen, de hoeveelheid meetpunten en de meetfrequentie.

Door deze op te voeren bij de meest onzekere termen kan de onzekerheid van de totale waterbalans het meest verkleind worden. Het gevolg van een verkleining van de onzekerheid $2\sigma_r$ van de geschatte waterbalans (verbeterde nauwkeurigheid) is dat kleinere systematische afwijkingen worden ontdekt en de waarschuwingsbel hierdoor eerder gaat rinkelen. Dit is logisch omdat bij meer en preciezer meten ook van de berekening van de waterbalans meer verwacht mag worden en dus minder grote fouten dienen te worden getolereerd.

3.4 K^* als schatter voor een onbekende balansterm

In paragrafen 3.2 en 3.3 zijn we er vanuit gegaan dat de restterm een maat is voor de fout in de gehele waterbalans. Vaak worden in de praktijk echter niet alle balanstermen apart berekend, maar wordt de restterm gebruikt om een onbekende balansterm zoals kwel/wegzijging of waterinlaat te berekenen. Als we bijvoorbeeld kwel/wegzijging berekenen dan wordt het foutenmodel:

de kwel/wegzijging wordt geschat door:

$$K^* = -P^* + E^* + D^* + \Delta O^* + \Delta G^* + \Delta B^* \quad (25)$$

de variantie van de schattingsfout wordt gegeven door:

$$\text{var}(K^*) = \sigma_p^2 + \sigma_e^2 + \sigma_d^2 + \sigma_{\Delta o}^2 + \sigma_{\Delta g}^2 + \sigma_{\Delta b}^2 \quad (26)$$

Wanneer vergelijking (25) alle balanstermen bevat en elke balansterm wordt door de schatters in (25) zuiver geschat (dit is een strengere aanname dan die gedaan wordt voor het foutenmodel (23); zie paragraaf 3.2), dan is K^* ook een zuivere schatter van k , de kwel/wegzijging. De waarde $\sigma_k^2 = \text{var}(K^*)$ is dan een maat voor de grootte van de schattingsfout van k . Net als voor r kunnen we nu k een betrouwbaarheidsinterval opstellen $[K^* - 2\sigma_k, K^* + 2\sigma_k]$.

Meestal heeft men wel enig idee hoe groot de onbekende term in de waterbalans moet zijn. Men kan bijvoorbeeld uit de resultaten van een pompproef en metingen van stijghoogten van het grondwater een schatting maken van de grootte van de kwel/wegzijging. Men kan dan toetsen of de uit de waterbalans berekende kwel/wegzijging significant afwijkt van de verwachte waarde. Ligt de verwachte waarde in het betrouwbaarheidsinterval $[K^* - 2\sigma_k, K^* + 2\sigma_k]$ dan is er geen significante afwijking en kan het verschil het gevolg zijn van (toevallige) schattingsfouten in de overige termen van de waterbalans. Is er wel een significant verschil dan kan er sprake zijn van een systematische fout. Of sommige van de termen van de waterbalans worden systematisch fout berekend of de waarde die men op grond van andere bronnen verwachtte voor de kwel/wegzijging klopt niet. In ieder geval zal verder onderzoek nodig zijn naar de kwel/wegzijging zelf of de berekeningswijze van sommige van de overige termen van de waterbalans.

De waarde $2\sigma_k$ is een maat voor de onzekerheid in de kwel/wegzijging. Men kan eisen stellen aan de maximale grootte van $2\sigma_k$ om de waterbalansstudie als voldoende nauwkeurig te beschouwen. De onzekerheid $2\sigma_k$ dient dan tevens als een geaggregeerde maat voor de onzekerheid van de waterbalans als geheel. Aan de varianties van de individuele termen in (25) kan men zien welke term het meest bijdraagt tot deze onzekerheid. Eventuele verbeteringen van de nauwkeurigheid van de waterbalansstudie kunnen dan het best worden bereikt door deze term nauwkeuriger te schatten. Dit komt neer op het verbeteren van de meetstrategie die aan de schatting van deze term ten grondslag ligt; i.c. het opvoeren van precisie van de metingen, de hoeveelheid meetpunten en de meetfrequentie.

4 Foutenberekening van individuele termen

Dit hoofdstuk behandelt een aantal technieken waarmee de verwachtingen en varianties van de schatters van de individuele termen, die nodig zijn in vergelijkingen (23) en (24) of (25) en (26), kunnen worden berekend. Om te kunnen analyseren hoe de onzekerheid in de berekening van de termen in vergelijking (1) binnensluipt en hoe eventuele fouten zich voortplanten moeten we het berekeningsproces in een aantal stappen onderverdelen. In de meest uitgebreide vorm bestaat de berekening van een term van de waterbalans uit de volgende stappen:

- 1 *Metingen*. De eerste fout ontstaat bij het meten van variabelen. Zoals gezegd kunnen dit systematische en toevallige fouten zijn; de systematische fout draagt bij tot de werkelijke waarde van r , terwijl de toevallige fout terug te vinden is in de variantie van R^* .
- 2 *Berekeningen*. Vervolgens wordt deze fout doorgegeven naar eventuele afgeleide variabelen die uit de gemeten variabelen worden berekend. Hierbij ontstaan geen nieuwe fouten, doch fouten in de invoervariabelen worden (versterkt of verzwakt) omgezet in fouten in de berekende uitvoer. Dit verschijnsel wordt vaak aangeduid met foutenvoortplanting.
- 3 *Tijdsintegratie*. Veel termen van de waterbalans worden op een beperkt aantal tijdstippen gemeten, waarna met dit beperkte aantal metingen een tijdsgemiddelde of cumulatieve hoeveelheid (over de balansperiode) van deze variabele moet worden geschat. Hierbij wordt geen rekening gehouden met de onbekende variatie tussen de meetpunten, zodat dit een extra fout oplevert.
- 4 *Ruimte-integratie of ruimtelijke middeling*. Analoog aan het tijdsgemiddelde moet voor een aantal variabelen (denk aan neerslag of verdamping) een gebiedsgemiddelde worden berekend. Ook dit levert een fout op.

We hebben dus te maken met meetfouten, die versterkt of verzwakt kunnen worden door berekeningen, en fouten ten gevolge van onbekende ruimtelijke en temporele variabiliteit. In de volgende paragrafen zal op al deze fouten worden ingegaan en worden besproken hoe de gemiddelden en varianties kunnen worden (door)berekend.

4.1 Meetfouten

Meetfouten moeten vaak experimenteel worden bepaald, bijvoorbeeld door in een korte tijd (in een tijdsbestek waarin de variabele niet verandert) een groot aantal metingen te doen door verschillende personen. Bij veel meetapparatuur wordt de precisie van metingen door de fabrikant gegarandeerd. Deze precisie is equivalent met de standaardafwijking van de meetfout.

4.2 Berekeningen

Vaak wordt een aantal variabelen gemeten en andere variabelen hieruit berekend. Men kan bijvoorbeeld denken aan het berekenen van de open-waterverdamping (met de Penman formule) uit de netto straling, watertemperatuur, luchttemperatuur en relatieve luchtvochtigheid of het berekenen van de afvoer boven een stuw uit de overstorthoogte, of de bodemvochthoeveelheid uit een bodemfysisch model met onzekere parameters. Als de invoervariabelen van een berekening stochastisch zijn (door bijvoorbeeld een onbekende meetfout) zijn de resultaten van een berekening dit ook. We kunnen dan de verwachting en de variantie van de uitvoervariabele berekenen als de verwachtingen en varianties van de invoervariabelen en hun onderlinge correlatiecoëfficiënten bekend zijn. Het algemene probleem is het volgende (naar Heuvelink et al., 1989):

We beschouwen een functie

$$Y = g(Z_1, Z_2, \dots, Z_n) \quad (27)$$

van de stochastische variabelen $Z_1, Z_2, \dots, Z_n$. We nemen aan dat we verwachtingen $E[Z_i]$ en varianties $\text{var}(Z_i) = \sigma_i^2$ van de Z_i ($i=1, \dots, n$) kennen, alsmede de correlatiecoëfficiënten ρ_{ij} tussen Z_i en Z_j voor alle mogelijke combinaties van $i, j=1, \dots, n$. Omdat de Z_i stochastische variabelen zijn geldt dit ook voor Y . De vraag is nu: Wat is de verwachting $E[Y]$ en de variantie $\text{var}(Y)$ van Y ? De methode die we het beste kunnen gebruiken om deze vraag te beantwoorden hangt af van de vorm van de functie $g(Z_1, Z_2, \dots, Z_n)$. We kunnen de volgende gevallen onderscheiden:

- de functie $g()$ is lineair;
- de functie $g()$ is bekend, continu en differentieerbaar in alle Z_i ; σ_i^2 klein;
- de functie $g()$ heeft een willekeurige vorm; σ_i^2 hebben willekeurige waarden.

De functie $g()$ is lineair

In dat geval heeft $g()$ de vorm van vergelijking (18) en geven vergelijkingen (21) en (22) de oplossingen voor $E[Y]$ en $\text{var}(Y)$. Veel voorkomende gevallen zijn (α en β constanten):

(a) simpele lineaire relatie:

$$Y = \alpha + \beta Z_1 \rightarrow E[Y] = \alpha + \beta E[Z_1], \quad \text{var}(Y) = \beta^2 \sigma_1^2$$

(b) verschil van twee variabelen (bijvoorbeeld grondwaterstand op twee tijdstippen):

$$Y = Z_1 - Z_2 \rightarrow E[Y] = E[Z_1] - E[Z_2], \quad \text{var}(Y) = \sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}$$

(c) optellen van n onafhankelijke variabelen Z_i met dezelfde variantie σ_z^2

$$Y = \sum_{i=1}^n Z_i \rightarrow E[Y] = \sum_{i=1}^n E[Z_i], \quad \text{var}(Y) = n\sigma_z^2$$

(d) optellen van n onafhankelijke variabelen Z_i met proportionele fout $\sigma_{zi} = E[Z_i]\sigma_z$

$$Y = \sum_{i=1}^n Z_i \rightarrow E[Y] = \sum_{i=1}^n E[Z_i], \quad \text{var}(Y) = \sigma_z^2 \sum_{i=1}^n E[Z_i]^2$$

De functie $g()$ is bekend, continu en differentieerbaar in alle Z_i ; σ_i^2 klein

Als $g(Z_1, Z_2, \dots, Z_n)$ niet lineair is, maar het is wel een bekende (= expliciete) en continue (= ononderbroken) functie van de Z_i die bovendien differentieerbaar is, dan kunnen we de zogenaamde Taylorreeksontwikkeling gebruiken om $g(Z_1, Z_2, \dots, Z_n)$ te lineariseren. Deze linearisatie gebeurt dan in de buurt van de verwachtingen $E[Z_1], E[Z_2], \dots, E[Z_n]$. Een Taylorreeksontwikkeling bestaat uit het benaderen van een functie in termen van lineaire combinaties van zijn eerste, tweede, derde, vierde etc. afgeleiden. De afgeleiden worden dan geëvalueerd voor $E[Z_1], E[Z_2], \dots, E[Z_n]$. Meestal worden alleen de termen met de lagere (eerste, tweede) orde afgeleiden meegenomen. Omdat de functie daarna lineair is in $Z_1, Z_2, \dots, Z_n$ kunnen vervolgens de regels van de lineaire kansrekening (vergelijkingen 21 en 22) worden toegepast. De Taylorreeksontwikkeling van $g(Z_1, Z_2, \dots, Z_n)$ is een goede benadering als de $Z_1, Z_2, \dots, Z_n$ niet te ver van hun verwachtingswaarden afwijken, dus als de varianties σ_i^2 ($i=1, \dots, n$) niet te groot zijn.

Om de Taylor-vergelijkingen wat korter te kunnen uitdrukken wordt de volgende vector gedefinieerd: $[E[Z_1], E[Z_2], \dots, E[Z_n]]^T = \mu$. Dus als we μ gebruiken dan worden de verwachtingswaarden $E[Z_1], E[Z_2], \dots, E[Z_n]$ bedoeld. Voor de verwachting van Y vinden we volgens een tweede-orde-Taylorbenadering (Heuvelink et al., 1989):

$$E[Y] = g(\mu) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \left\{ \sigma_i \sigma_j \rho_{ij} \left(\frac{\partial^2 g}{\partial z_i \partial z_j} (\mu) \right) \right\} \quad (28)$$

De variantie vinden we via een eerste-orde-Taylorbenadering (Heuvelink et al., 1989):

$$\text{var}(Y) = \sum_{i=1}^n \sum_{j=1}^n \left\{ \sigma_i \sigma_j \rho_{ij} \left(\frac{\partial g}{\partial z_i} (\mu) \right) \left(\frac{\partial g}{\partial z_j} (\mu) \right) \right\} \quad (29)$$

Hogere-orde-benaderingen zijn ook mogelijk, maar zijn in het algemeen niet praktisch omdat ze berusten op de veronderstelling dat alle Z_i samen normaal verdeeld zijn.

Als voorbeeld van een toepassing: De afvoerintensiteit Q [$L^3 T^{-1}$] over een volkomen overlaat wordt gegeven door de vergelijking

$$Q = BCH^{3/2} \quad (30)$$

waarbij B de breedte van de overlaat is [L], C de (gecalibreerde) afvoerconstante [$L^{1/2} T^{-1}$] en H de overstorthoogte [L]. We weten dat de overstorthoogte H onzeker is tengevolge van fouten in de waterhoogtemeting. Ook de afvoerconstante C is onzeker omdat deze in een waterloopkundig laboratorium is bepaald onder omstandigheden die iets anders zijn dan in het veld. De afvoer Q_i is dus ook onzeker en we willen weten wat de verwachting en variantie van Q_i zijn. Uit het waterloopkundig laboratorium is bekend dat C gemiddeld gelijk is aan $E[C]$ en een

variantie σ_C^2 heeft. Verder weten we van de fabrikant van het apparaat waarmee we de waterhoogte meten dat als we een overstorthoogte h_m meten we een fout maken die gemiddeld nul is met variantie σ_H^2 . Omdat de fout gemiddeld nul is de meting h_m de beste beschikbare schatting van de onbekende waarde van de verwachte overstorthoogte $E[H]$; de Taylor-ontwikkeling zal dus plaats moeten vinden rond $E[C]$ en h_m . Verder kunnen we aannemen dat C en H ongecorreleerd zijn. Door toepassing van vergelijkingen (28) en (29) op (30) vinden we voor de verwachting en de variantie van $Q(t)$ op tijdstip t met gemeten waterhoogte $h_m(t)$:

$$E[Q(t)] = BE[C]h_m^{3/2}(t) + (3/8)\sigma_H^2 BE[C]h_m^{-1/2}(t) \quad (31)$$

$$\text{var}(Q(t)) = B^2\sigma_C^2 h_m^3(t) + (9/4)\sigma_H^2 B^2(E[C])^2 h_m(t) \quad (32)$$

We kunnen in vervolgberekeningen $E[Q(t)]$ nemen als de 'gemeten' afvoerintensiteit Q_t en $\text{var}(Q(t))$ is dan de variantie van de 'meetfout'.

De functie $g()$ heeft een willekeurige vorm; σ_i^2 hebben willekeurige waarden

De meest algemene techniek van de analyse van foutenvoortplanting is de Monte Carlo-simulatie. Bij Monte Carlo-simulatie worden trekkingen $\{z_1, z_2, \dots, z_n\}$ gedaan uit de gezamenlijke kansdichtheidsfunctie (ook wel "kansverdeling" genoemd) van $Z_1, Z_2, \dots, Z_n$. Zo'n trekking is eigenlijk een kansexperiment, zoals het gooien van een dobbelsteen: de stochastische variabele N is het aantal ogen, en het feitelijke aantal ogen n na een worp is een trekking uit de kansverdeling van N ; door de dobbelsteen een zeer groot aantal malen te rollen kunnen we schatten wat de kans is dat $n=5$ door te tellen hoeveel maal de dobbelsteen op 5 ogen blijft liggen en dit te delen door het aantal keer dat is geworpen. Dit kansexperiment is triviaal omdat de kans op $n=5$ gelijk is aan $1/6$. We kunnen dit experiment echter ook toepassen op ingewikkelder gevallen. Bijvoorbeeld in het geval van de stochastische variabele $Y = g(Z_1, Z_2, \dots, Z_n)$. Als we de gezamenlijke kansdichtheidsfunctie $f(Z_1, Z_2, \dots, Z_n)$ kennen, dan kunnen we hier een groot aantal (bijv. m) trekkingen $\{z_1, z_2, \dots, z_n\}_k$, $k=1, \dots, m$ uit genereren. Vervolgens kunnen we de functie $g()$ voor elke van deze trekkingen evalueren. We verkrijgen dan m verschillende uitkomsten $y_k = g(\{z_1, z_2, \dots, z_n\}_k)$, $k=1, \dots, m$. Als m groot genoeg is kunnen we uit de y_k de verwachting en variantie van Y schatten of, bij zeer grote m , zelfs de kansdichtheidsfunctie van Y .

Monte Carlo-simulatie heeft grote voordelen.

- De methode is toepasbaar op elke willekeurige vorm van $g(Z_1, Z_2, \dots, Z_n)$. Zo kan $g(Z_1, Z_2, \dots, Z_n)$ een niet-differentieerbare of discontinuë functie zijn of zelf een niet expliciete functie zoals de oplossing van een differentiaalvergelijking. In het laatste geval kan men denken aan een grondwatermodel met onzekere doorlatendheden of een bodemfysisch model, zoals SWATRE (Belmans et al., 1983) met onzekere Van Genuchten parameters.
- De methode is toepasbaar bij elke willekeurige grootte van de varianties σ_i^2 van Z_i .

De Monte Carlo-methode heeft ook nadelen.

- De rekentijden kunnen lang zijn. Vaak zijn vele trekkingen $\{z_1, z_2, \dots, z_n\}_k$ nodig om $E[Y]$, $\text{var}(Y)$ en zeker de kansdichtheidsfunctie $f(Y)$ betrouwbaar te schatten (vaak is het niet duidelijk hoeveel trekkingen voldoende is). Zeker als

$g(Z_1, Z_2, \dots, Z_n)$ een oplossing is van een differentiaalvergelijking kan dit leiden tot lange rekentijden op zelfs grote computers. Natuurlijk verdwijnt dit bezwaar meer en meer met de verdergaande technologische ontwikkelingen.

- Om trekkingen te doen uit de kansdichtheidsfunctie van $Z_1, Z_2, \dots, Z_n$ moet men veelal aannemen dat deze Gaussisch is, en dus dat $Z_1, Z_2, \dots, Z_n$ gezamenlijk normaal verdeeld zijn. In dat geval zijn er een groot aantal verschillende algoritmen om trekkingen te genereren. Veel gebruikte algoritmen zijn de "LU-decompositie" (Press et al., 1988) en de "sequentiële simulatie" (Deutsch & Journal, 1992) waarbij van de Z_i dezelfde statistische parameters bekend moeten zijn als voor de Taylormethoden.

Uit het bovenstaande is duidelijk dat:

- wanneer de functie $g(Z_1, Z_2, \dots, Z_n)$ lineair is de varianties van Y direct berekend kunnen worden
- wanneer $g(Z_1, Z_2, \dots, Z_n)$ een niet-lineaire expliciet functie is die in ieder geval rond μ continue en differentieerbaar is en wanneer de varianties σ_i^2 klein zijn, Taylormethoden kunnen worden gebruikt;
- slechts als aan bovenstaande voorwaarden niet wordt voldaan Monte Carlo-methoden moeten worden toegepast.

Voor een uitgebreide afweging tussen de verschillende methoden van foutendoorrekening wordt verwezen naar Heuvelink (1993).

4.3 Tijdsintegratie

Bepaalde variabelen worden in de balansperiode op min of meer gelijke tijdsintervallen bemeten, waarna er de cumulatieve hoeveelheid voor de gehele balansperiode mee wordt geschat. Men kan hierbij denken aan het één maal per dag meten van een afvoerintensiteit [$L^3 T^{-1}$] en het hiermee berekenen van de totale jaarafvoer per oppervlakte-eenheid [$L^3 L^{-2}$]. Stel dat we een balansperiode hebben van lengte T en de afvoerintensiteit q_i wordt gemeten op de tijdstippen t_i , $0 \leq t_1 \leq t_2 \leq \dots \leq t_n \leq T$. We nemen we aan dat het beheersgebied een bekende oppervlakte A [L^2] heeft en dat de jaarafvoer per oppervlakte-eenheid d wordt berekend met de volgende ongewogen som:

$$d = \frac{T}{A n} \sum_{i=1}^n q_i \quad (33)$$

Vergelijking (33) wordt dan gebruikt om de d in de balansvergelijking (1) te berekenen. Op de tijdstippen t_i meten we echter de afvoerintensiteiten met een meetfout waarvan de variantie gelijk is aan $\sigma_{Q_i}^2$. We gaan daarom uit van stochastische afvoerintensiteiten Q_i . De meting Q_i en variantie $\sigma_{Q_i}^2$ kunnen dan bijvoorbeeld met vergelijkingen (31) en (32) zijn berekend. De schatter D^* van de totale afvoerhoeveelheid d (in m per oppervlakte-eenheid) voor het beheersgebied is ook een stochastische variabele, omdat:

- de afvoermetingen zijn behept met een meetfout;
- er een bijkomende fout ontstaat omdat de totale gebiedsafvoer wordt geschat aan

de hand van een beperkt aantal metingen. Hierdoor wordt de temporele variatie van de afvoerintensiteit tussen de metingen verwaarloosd of geschat.

De stochasticiteit van D^* is dus het gevolg van meetfouten en temporele variatie. De schatter van d wordt:

$$D^* = \frac{T}{A n} \sum_{i=1}^n Q_i \quad (34)$$

Voor de berekening van de variantie van D^* zijn er twee benaderingen: steekproefmethoden en stochastische-functiemethoden.

Steekproefmethoden

Steekproefmethoden kunnen gebruikt worden als het mogelijk is om de metingen *willekeurig in de tijd* te nemen. Dit houdt in dat het beoogde aantal metingen n willekeurig (door trekkingen uit een uniforme verdeling) over een eindig interval worden verdeeld. De tijdstippen t_i , $0 \leq t_1 \leq t_2 \leq \dots \leq t_n \leq T$ zijn hierdoor stochastische variabelen. De n metingen zijn dan een steekproef uit de populatie van alle afvoeren in het interval $[0, T]$. De onbekende constante d wordt dan zuiver geschat met vergelijking (34) en de schattingsvariantie wordt berekend door de volgende stappen (Cochran, 1977):

(a) bereken de steekproefvariantie S_q^2 (met D^* uit 34):

$$S_q^2 = \frac{1}{(n-1)} \sum_{i=1}^n (Q_i - [AD^*/T])^2 \quad (35)$$

(b) bereken $\text{var}(D^*)$ met de vergelijking:

$$\text{var}(D^*) = \frac{T^2 S_q^2}{A^2 n} \quad (36)$$

Bij het berekenen van de variantie hoeft niet apart rekening gehouden te worden met de meetfouten in de Q_i omdat die automatisch worden meegerekend door hun bijdragen in de steekproefvariantie S_q^2 .

De bovenstaande opzet heet een "enkelvoudige aselechte steekproef". Er zijn ook steekproefopzetten te bedenken waarbij de balansperiode kan worden verdeeld in tijdvakken (bijvoorbeeld in seizoenen) en de metingen op willekeurige tijdstippen binnen elk tijdvak worden genomen. Men noemt dit een "gestratificeerde aselechte steekproef". Hiermee is het mogelijk de variantie $\text{var}(D^*)$ te verlagen (zie Cochran, 1977).

Stochastische-functiemethoden

Meestal zijn de metingen op *voorbedachte tijdstippen* genomen, omdat dit het uitvoeren van een meetprogramma vergemakkelijkt. Vaak is het zelfs zo dat gemeten wordt met vaste tijdsintervallen tussen de metingen. Er is dan sprake van een tijdreeks. Als de tijdstippen waarop gemeten wordt niet willekeurig in de tijd genomen zijn kunnen de steekproefmethoden niet worden gebruikt. Een alternatieve methode die wel kan worden toegepast is de stochastische-functiemethode. Deze houdt in dat de onbekende variatie tussen de meetpunten wordt gemodelleerd door een continue stochastisch functie (als deze alleen een functie van de tijd is ook wel "stochastisch proces" genoemd). Zonder hier te diep op in te gaan kunnen we stellen dat de onbekende waarden van de afvoerintensiteit tussen de metingen worden gezien als in de tijd gecorreleerde stochastische variabelen¹. Het standaardwerk waarin stochastische-functiemethoden, ook wel "geostatistiek" genoemd, uitgebreid worden beschreven is van Journel & Huijbregts (1978). Eenvoudige korte introducties kunnen worden gevonden in de Marsily (1986; hfst 11) en Bierkens (1994; hfst 2). Een uitstekend leerboek met voorbeelden is geschreven door Isaaks & Srivastava (1989). Brus & De Gruijter (1993) gaan op uitgebreide wijze in op het verschil tussen de steekproefmethoden en de stochastische-functiemethoden.

Het berekenen van $\text{var}(D^*)$ gaat dan als volgt (n is nu het aantal metingen op regelmatige afstand in de tijd in de balansperiode):

(a) Schat de volgende variantie met behulp van de vergelijking (met D^* uit 34):

$$S_q^2 = \frac{1}{(n-1)} \sum_{i=1}^n (Q_i - [AD^*/T])^2 \quad (37)$$

(b) Schat de correlatiefunctie $\rho(h)$:

De correlatiefunctie $\rho(h)$ geeft de correlatie weer tussen de stochastische variabelen $Q(t)$ en $Q(t+h)$ die op een afstand h op de tijdas van elkaar liggen. De correlatiefunctie kan bij n contante meetintervallen $<0, t_1]$, $<t_1, t_2]$, ..., $<t_{n-1}, T]$ geschat worden voor de discrete afstanden $h_k = t_{i+k} - t_i$:

$$\rho(h_k) = \frac{1}{S_q^2(n-k-1)} \sum_{i=1}^{n-k} (Q_i - [AD^*/T])(Q_{i+k} - [AD^*/T]) \quad (38)$$

De correlatiefunctie die op discrete tijdstippen is geschat wordt ook wel correlellogram genoemd. In het geval dat er geen sprake is van constante meetintervallen kan het correlellogram ook worden geschat. Hoe dit gaat wordt in de volgende paragraaf uitgelegd. Om zinvolle (i.c. positieve) waarden voor $\text{var}(D^*)$

¹ Dus in plaats van een constante d proberen we nu een trekking (ook wel realisatie genoemd) van een stochastische variabele D te schatten (n.b. om het verschil tussen een constante en een realisatie van een stochastische variabele aan te geven wordt in het laatste geval in plaats van de term "schatten" vaak de term "voorspellen" gebruikt). Dit verschil heeft twee belangrijke consequenties: 1) als een schatter D^* zuiver is betekent dit nu dat $E[D^*] = E[D]$; 2) de schattingsvariantie heeft nu betrekking op $(D^* - D)$ en bij een zuivere schatter geldt dus dat we in het foutenmodel van de waterbalans eigenlijk substitueren: $\text{var}(Z_i^*) \equiv E[(Z_i^* - Z_i)^2]$.

te berekenen kunnen we voor de correlatiefunctie niet elke willekeurige functie nemen. Er zijn slechts een beperkt aantal geoorloofde functies. Journal & Huijbregts (1978) geven een overzicht van een groot aantal geoorloofde functies.

Voor de correlatiefunctie in de tijd wordt vaak het exponentiële model genomen:

$$\rho(h) = e^{-(|h|/I)} \quad (39)$$

De parameter I noemt men ook wel de correlatieschaal of karakteristieke tijd. Hoe groter I hoe groter het tijdsverschil waarover twee waarden van $Q(t)$ nog gecorreleerd blijven. Bij een exponentiële correlatiefunctie geldt dat als h groter is dan $3I$ we kunnen stellen dat $Q(t)$ en $Q(t+h)$ ongecorreleerd zijn. Voordat $\text{var}(D^*)$ verder berekend kan worden moet nu eerst vergelijking (39) gefit worden door de punten van het met (38) geschatte correlogram. Dit levert een geschatte waarde van de correlatielengte op.

(c) Bereken de relatieve bijdrage van de meetfout aan de schattingsvariantie: Dit is eigenlijk de variantie van de gemiddelde meetfout en wordt berekend als:

$$\sigma_{Qm}^2 = \frac{1}{n^2} \sum_{i=1}^n \sigma_{Qi}^2 \quad (40)$$

(d) Bereken de volgende twee integralen:

$$\rho(T,t) = \frac{1}{T} \int_0^T \rho(t-\tau) d\tau \quad (41)$$

$$\rho(T,T) = \frac{1}{T^2} \int_0^T \int_0^T \rho(\tau_1-\tau_2) d\tau_1 d\tau_2 \quad (42)$$

(e) Bereken tenslotte de schattingsvariantie met de volgende vergelijking:

$$\text{var}(D^*) \equiv \text{var}(D^* - D) =$$

$$\frac{T^2 (S_q^2 - \sigma_{Qm}^2)}{A^2} \left\{ \rho(T,T) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \rho(t_i - t_j) - \frac{2}{n} \sum_{i=1}^n \rho(T, t_i) \right\} + \sigma_{Qm}^2 \quad (43)$$

Vergelijking (34) geeft een ongewogen gemiddelde als schatter voor d . Binnen de stochastische-functiebenadering behoeft dit echter niet de beste schatter te zijn. Een gewogen lineaire schatter, in dezelfde vorm als vergelijking (18), kan kleinere schattingsvarianties opleveren. De gewichten worden dan gevonden door de variantie van de schattingsfout te minimaliseren. Deze vorm van schatten noemt men "kriging", of in het bijzonder "blokkriging" omdat het gaat om het schatten van het gemiddelde van een interval (Journal & Huijbregts, 1978). Als de metingen echter regelmatig over het interval $[0, T]$ zijn verdeeld, krijgen ze allen ongeveer het gewicht $1/n$ en is er geen verschil tussen bovenstaande vergelijkingen en blokkriging. Als er sprake is van een trend ($E[Q(t)] \neq \text{constante}$) – men kan bijvoorbeeld denken aan een

seizoensfluctuatie – dan moet die eerst uit de gemeten Q_i worden verwijderd: $Q_i' = Q_i - E[Q(t_i)]$. Vergelijkingen (37) t/m (43) moeten dan op Q_i' worden toegepast. Het probleem is dat men de trend exact moet kennen. Is de trend onbekend dan bestaan er speciale krigingmethoden waarin de trend automatisch wordt geschat en de onzekerheid over de trend wordt verdisconteerd in de schattingsvariantie (Journel & Huijbregts, 1978).

Als we de twee methoden met elkaar vergelijken is het duidelijk dat de steekproefmethode veel gemakkelijker is toe te passen. Dit is met name het geval wanneer we maar weinig metingen in de tijd mogen nemen. In zo'n geval is het moeilijk om een correlatiefunctie of een eventueel aanwezige trend te schatten. De traditionele manier van meten in de tijd gaat vaak uit van vaste, veelal equidistante, meettijdstippen: het willekeurig meten in de tijd wordt toch als enigszins tegennatuurlijk ervaren. Als de situatie echter zodanig is dat meettijdstippen nog gekozen mogen worden en we maar weinig metingen in de tijd mogen nemen, verdient het zeker de aanbeveling om te kijken of steekproefmethoden niet efficiënter zijn dan het meten met een vaste frequentie.

4.4 Ruimtelijke integratie

De uitmiddeling van variabelen die op een beperkt aantal punten zijn gemeten over het beheersgebied is een belangrijke stap in waterbalansberekeningen. Een aantal voorbeelden zijn: het berekenen van de gebiedsneerslag uit metingen op een aantal punten; het berekenen van de gebiedsverdamping uit verdampingsberekeningen op een aantal punten; het berekenen van de bergingsverandering van het bodemvocht en de grondwaterhoeveelheid uit berekende grondwaterstanden en bodemvochtgehalten met een bodemfysisch model zoals SWATRE (Belmans et al., 1983). In al deze gevallen moeten uit puntwaarden gebiedsgemiddelden worden geschat en bij een foutenanalyse dus ook de variantie van de schattingsfout.

Als voorbeeld zullen we de berekening van de bergingsverandering van het grondwater nemen. In een beheersgebied met een oppervlakte A is op n plaatsen aan het begin en aan het eind van de balansperiode de grondwaterstand gemeten. Uit de porositeit en het verschil in grondwaterstand kan dan op elke meetlocatie de bergingsverandering Δg [L] worden geschat. We hebben dus n metingen van de bergingsverandering, die we ΔG_i ($i=1,...,n$) noemen, en welke niet foutloos zijn gemeten maar met een meetfout waarvan de varianties, respectievelijk $\sigma_{\Delta G_i}^2$ ($i=1,...,n$), bekend zijn. Het doel is nu om uit deze metingen de gebiedsgemiddelde bergingsverandering (= totale bergingsverandering per oppervlakte-eenheid [L]) uit vergelijking (1) te schatten en de bijbehorende schattingsvariantie $\text{var}(P^*)$. Net als in het geval van de tijdsintegratie kunnen we hier weer op twee manieren te werk gaan.

Steekproefmethoden

Vaak kan de grondwaterstand maar op een beperkt aantal punten worden gemeten. Als dit het geval is en we het meetnet nog moeten inrichten verdient het de aanbeveling om de lokaties van de grondwaterstandsbuizen willekeurig te kiezen. Hierbij worden een x -coördinaat en een y -coördinaat uit een uniforme verdeling getrokken en punt (x,y) de toekomstige positie van een grondwaterstandsbuis, mits dit natuurlijk binnen het beheersgebied valt. Is dit niet het geval dan volgt een nieuwe trekking. Dit wordt herhaald totdat n willekeurige posities zijn getrokken. Op deze posities kan dan de bergingsverandering worden gemeten voor de balansperiode. Het schatten van de gebiedsgemiddelde bergingsverandering is dan zeer eenvoudig:

$$\Delta G^* = \frac{1}{n} \sum_{i=1}^n \Delta G_i \quad (44)$$

De variantie van de schattingsfout wordt dan berekend als:

$$\text{var}(\Delta G^*) = \frac{1}{(n-1)} \sum_{i=1}^n (\Delta G_i - \Delta G^*)^2 \quad (45)$$

Wanneer het beheersgebied uit verschillende peilvakken of verschillende grondsoorten bestaat is het waarschijnlijk dat de bergingsveranderingen van plaats tot plaats sterk verschillen. Het is dan aan te bevelen een "gestratificeerde aselechte steekproef" uit voeren. In dat geval wordt het beheersgebied verdeeld in subgebieden met min of meer dezelfde geologie of hetzelfde peilbeheer en worden voor elk subgebied de punten willekeurig toegedeeld. Net als bij het berekenen van het tijdsgemiddelde (zie § 4.3) is hiermee een aanzienlijke reductie van de schattingsvariantie te verwachten (zie Cochran, 1977).

Stochastische-functiemethoden

Wanneer de posities van de meetpunten niet willekeurig zijn – in dit voorbeeld: wanneer de grondwaterstandsbuizen bewust op bepaalde punten zijn neergezet – kunnen de steekproefmethoden niet worden gebruikt. Als we de schattingsvariantie willen berekenen zijn we gedwongen om een stochastische-functiemethode te gebruiken. In tegenstelling tot metingen in de tijd, zijn de metingen in de ruimte slechts zelden regelmatig verdeeld. De meest nauwkeurige lineaire schatters zijn in dat geval gewogen gemiddelden, waarbij de gewichten een minimale schattingsvariantie garanderen. We kunnen dan het beste "blokkriking" toepassen. We nemen aan dat $\Delta G(\mathbf{x})$ een stochastische functie is, met $\mathbf{x} = (x,y)$ een plaatsvector die de positie in het gebied A aangeeft. De coördinaten van de meetpunten ΔG_i zijn bekend: $\mathbf{x}_i = (x_i, y_i)$. De berekeningsstappen, die horen bij "ordinary blockriking" (Journel & Huijbregts) gaan dan als volgt:

(a) Bereken het semivariogram:

Het semivariogram geeft net als het correlellogram een indruk van de afstand (in ruimte of tijd) waarover een stochastische functie, zoals $\Delta G(\mathbf{x})$, gecorreleerd is. Het semivariogram is gedefinieerd als:

$$\gamma(\mathbf{h}) = \frac{1}{2} E[(\Delta G(\mathbf{x}) - \Delta G(\mathbf{x} + \mathbf{h}))^2] \quad (46)$$

Het semivariogram is dus gelijk aan de helft van de verwachting van het kwadratisch verschil van twee punten die op een *vectoriële* afstand $\mathbf{h} = (h_x, h_y)$ van elkaar liggen, waarbij h_x en h_y respectievelijk de afstanden tussen de punten in de x - en de y -richting voorstellen. Het variogram kan geschat worden uit verspreid liggende punten als (Journel & Huijbregts, 1978):

$$\gamma^*(\mathbf{h}) = \frac{1}{2N(h)} \sum_{k=1}^{N(h)} [\Delta G_i(\mathbf{x}_k) - \Delta G_j(\mathbf{x}_k + \mathbf{h} \pm \Delta \mathbf{h})]^2 \quad (47)$$

We middelen dus de kwadratische verschillen van alle puntenparen ΔG_i en ΔG_j die op een afstand $\mathbf{h} \pm \Delta \mathbf{h}$ van elkaar liggen. $N(h)$ is het aantal puntenparen dat op deze afstand van elkaar ligt.

Voor het berekenen van positieve schattingsvarianties moet een geoorloofde functie door het met (47) berekende experimentele semivariogram worden aangepast. Geoorloofde functies worden gegeven in alle boeken over geostatistiek, waaronder dat van Journel & Huijbregts (1978). De exponentiële functie is er één van en heeft de volgende vorm:

$$\gamma(\mathbf{h}) = C(1 - e^{-(|\mathbf{h}|/I)}) \quad (48)$$

waarbij $|\mathbf{h}| = (h_x^2 + h_y^2)^{1/2}$, de *scalaire* afstand tussen twee punten in het vlak. Ook hier geldt dat I de correlatielengte is. Te zien is dat in de functie is aangenomen dat de correlatielengte in alle richtingen gelijk is. Men kan ook functies aanpassen waarbij de correlatielengtes in de x - en de y -richting verschillend zijn.

Voor het schatten van het experimentele semivariogram (47) behoeven de punten niet op vaste afstanden van elkaar te liggen. Als de meetfrequentie in de tijd niet constant is kan het semivariogram worden gebruikt om de correlatiefunctie te achterhalen. Daarvoor kan eerst het experimentele semivariogram in de tijd met (47) worden geschat. Vervolgens kan de functie van het exponentiële variogram (48) worden gefit. De temporele correlatiefunctie volgt dan uit:

$$\rho(h) = 1 - [\gamma(h)/C] \quad (49)$$

(b) Bereken de volgende twee integralen:

$$\gamma(A, \mathbf{x}) = \frac{1}{A} \int_A \gamma(\mathbf{x} - \xi) d\xi \quad (50)$$

$$\gamma(A, A) = \frac{1}{A^2} \int_A \int_A \gamma(\xi_1 - \xi_2) d\xi_1 d\xi_2 \quad (51)$$

waarbij een enkele integraal van de integratievariabele ξ over A zelf staat voor een dubbele integraal over A : één in de x -richting en één in de y -richting.

(c) Bereken de onbekende gewichten λ_i en de constante μ :

De gewichten λ_i worden gebruikt in de gewogen som $\Delta G^* = \sum_i \lambda_i \Delta G_i$ en de constante μ (de zogenaamde "Lagrange multiplier") is toegevoegd om te zorgen dat ΔG^* een zuivere schatter is. Bij kriging worden de gewichten λ_i zo berekend dat ΔG^* de lineaire schatter is met de kleinste schattingsvariantie. Dit is gegarandeerd als de λ_i worden opgelost uit het volgende stelsel van $n+1$ vergelijkingen (de $\mathbf{x}_i$, $i = 1, \dots, n$, zijn de lokaties van metingen):

$$\begin{aligned} \sum_{j=1}^n \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) - \lambda_i \sigma_{\Delta G_i}^2 + \mu &= \gamma(A, \mathbf{x}_i) & i = 1, \dots, n \\ \sum_{i=1}^n \lambda_i &= 1 \end{aligned} \quad (52)$$

We zien dat bij het berekenen van de gewichten ook rekening wordt gehouden met de meetfout.

(d) Bereken de schatting en de schattingsvariantie:

Als de gewichten zijn berekend zijn kunnen we tenslotte de gebiedsgemiddelde bergingsverandering schatten en de bijbehorende schattingsvariantie berekenen.

$$\Delta G^* = \sum_{i=1}^n \lambda_i \Delta G_i \quad (53)$$

$$\text{var}(\Delta G^*) \equiv \text{var}(\Delta G^* - \Delta G) = \sum_{i=1}^n \lambda_i \gamma(A, \mathbf{x}_i) + \mu - \gamma(A, A) \quad (54)$$

Ook voor kriging geldt dat het apart schatten van de grondwater-bergingsverandering en de bijbehorende varianties voor min of meer homogene subgebieden de schattingsvariantie van het totale gebiedsgemiddelde kan verlagen (Stein et al., 1988). Als de grondwaterstand goed gecorreleerd is met bijvoorbeeld de maaiveldshoogte, dan kunnen hoogtecijfers ook direct gebruikt worden als hulpinformatie (Te Riele et al., 1995).

Een vergelijking van (47) t/m (54) met (44) en (45) laat zien dat steekproefmethoden voor ruimtelijke gemiddelden veel eenvoudiger zijn dan de krigingmethode. Echter, er bestaat nu goedkope maar zeer professionele software waarmee allerlei soorten kriging betrekkelijk eenvoudig en snel kunnen worden uitgevoerd. Een goed voorbeeld hiervan is het geostatistisch pakket GSLIB van Deutsch & Journel (1992). Dit pakket, dat bestaat uit een set Fortran-77-programma's (source code) plus een instructieboek, kost slechts een paar honderd gulden.

Het grootste probleem bij het toepassen van de stochastische-functiemethoden voor het berekenen van ruimtelijke gemiddelden in het waterbeheer ligt in vinden van het semivariogram; dit kan niet geschat worden uit het handjevol waarnemingen dat doorgaans in het beheersgebied aanwezig is. Het verdient dus bijna altijd de aanbeveling om steekproefmethoden te gebruiken als het waarnemingsnet voor ruimtelijke gemiddelden nog moet worden ingericht. Uitzonderingen hierop vormen de regenmeters die op zeer specifieke plaatsen moeten worden neergezet. Als het

meetnet reeds is ingericht en de metingen worden gedaan op gericht gekozen lokaties, dan moet men wel een stochastische-functiemethode gebruiken. In het geval dat men te weinig metingen heeft om een variogram te schatten kan men proberen een variogram te gebruiken van een andere studie (zie bijvoorbeeld Witter et al., 1991). Als dit ook niet lukt dan zal men een traditionele methode zoals de Thiessenmethode voor bijvoorbeeld de gebiedsneerslag moeten toepassen; bij deze methode kunnen we echter niet berekenen hoe groot de bijbehorende schattingsfout is.

4.5 Toepassing op balanstermen

In deze paragraaf zullen we laten zien dat het mogelijk is om met de technieken die in § 4.1 t/m § 4.4 zijn beschreven elk van de termen van de waterbalans (1) te schatten en de bijbehorende schattingsvariantie te berekenen. Het kan zijn dat berekeningen in de praktijk heel anders verlopen dan in het volgende wordt voorgesteld. Hier wordt slechts aangegeven hoe het zou kunnen (met verwijzingen naar § 4.1 t/m § 4.4 voor de verschillende onderdelen). Het moet echter in principe mogelijk zijn elk type berekening van een waterbalansterm op te splitsen in een aantal elementaire stappen die met de technieken van § 4.1 t/m § 4.4 kunnen worden uitgewerkt.

Neerslag: r

Neerslag wordt niet gemeten als een intensiteit maar direct als een geïntegreerde hoeveelheid voor een bepaald tijdsinterval. Als de neerslag bijvoorbeeld wordt gemeten in dagtotalen met een variantie van de fout voor dag k gelijk aan σ_k^2 (§ 4.1), dan is de variantie van het jaartotaal voor meetpunt $\mathbf{x}_i$ te berekenen als $\sum_{k=1,365} \sigma_k^2$ en hoeven we niet verder de technieken van tijdsintegratie (§ 4.3) toe te passen. Voor elk meetpunt kunnen we zo de meetfout van het jaartotaal berekenen en uit de meting en de meetfout een schatting maken van het ruimtelijk gemiddelde P^* [L] en de bijbehorende schattingsvariantie (§ 4.4).

Kwel: k

Op bepaalde punten $\mathbf{x}_i$ in het beheersgebied worden op bepaalde tijdstippen t_k grondwaterstanden bemeten (diep $h_d(\mathbf{x}_i, t_k)$ en ondiep $h_o(\mathbf{x}_i, t_k)$), alsmede de verticale weerstand $c(\mathbf{x}_i)$ van de ondergrond. De geschatte kwelsterkte [LT^{-1}] op deze punten en tijdstippen wordt dan gelijk aan $Q_{kwel}(\mathbf{x}_i, t_k) = [h_d(\mathbf{x}_i, t_k) - h_o(\mathbf{x}_i, t_k)]/c(\mathbf{x}_i)$. Als de varianties van de meetfouten (§ 4.1) van $h_d(\mathbf{x}_i, t_k)$, $h_o(\mathbf{x}_i, t_k)$ en $c(\mathbf{x}_i)$ bekend zijn kan hieruit via de foutendoorrekening (§ 4.2) de verwachte kwel (= meting) en de variantie (= variantie meetfout) op tijdstip t_k voor elk meetpunt worden berekend. Vervolgens volgt voor elk meetpunt uit de tijdsintegratie (§ 4.3) de cumulatieve kwelhoeveelheid [L] en de bijbehorende variantie van de fout. Hieruit kan dan vervolgens het ruimtelijke gemiddelde worden geschat (K^* [L]) en de bijbehorende schattingsvariantie (§ 4.4).

Evapotranspiratie: e

Het berekenen van de gebiedsgemiddelde cumulatieve evapotranspiratie volgt dezelfde weg als beschreven voor de kwel: meting invoervariabelen (§ 4.1), berekening intensiteiten [LT^{-1}] op een aantal punten (§ 4.2), tijdsintegratie op die punten [L] (§ 4.3), ruimte-integratie [L] (§ 4.4). De intensiteit van de evapotranspiratie plus de bijbehorende foutenvariantie op één tijdstip voor een meetpunt kan dan worden berekend uit de meetfouten in variabelen als netto-straling, relatieve luchtvochtigheid, bodemvochtgehalte etc. via bijvoorbeeld de formules van Makkink of via een gewasmodel als WOFOST (van Diepen et al, 1989) in combinatie met de technieken uit (§ 4.2). Het verdient de aanbeveling om bij de ruimtelijke interpolatie van de cumulatieve evapotranspiratie te stratificeren naar gewas en bodemsoort.

Gebiedsafvoer: d

Het schatten van de cumulatieve gebiedsafvoer [L] hebben we reeds uitgebreid behandeld in § 4.2 en § 4.3. De stappen: meten invoervariabelen (§ 4.1), berekening afvoerintensiteiten (§ 4.2) en tijdsintegratie (§ 4.3).

Bergingsveranderingen: Δo , Δg en Δb

De bergingsverandering in het grondwater op een punt kan berekend worden door het verschil van het gemeten grondwaterpeil aan het eind en aan het begin van de balansperiode te nemen (§ 4.1) en dit te vermenigvuldigen met de porositeit van de ondergrond (zie voorbeeld § 4.4). Veranderingen in de berging van het bodemvocht kunnen op een punt bijvoorbeeld worden berekend met een eenvoudig bodembalansmodel in combinatie met een grondwaterstandsmeting. Bij onzekere invoerparameters kunnen de verwachte waarden (= meting) en de varianties (= variantie meetfout) van de bergingsveranderingen worden berekend met de technieken uit (§ 4.2). De gebiedsgemiddelde bergingsverandering volgt uit ruimtelijke middeling van de berekening op een aantal punten (§ 4.4). Het is hierbij aan te bevelen hulpinformatie te gebruiken. Bij de vochtberging valt te denken aan een stratificatie naar bodemtype (Stein et al., 1988). Bij de ruimtelijke integratie van de verandering in grondwaterberging kan de maaiveldshoogte als nuttige hulpinformatie dienen (Te Riele et al., 1995). Bergingsveranderingen in het oppervlaktewater kunnen berekend worden uit peilverschillen en de dimensies van waterlopen, kanalen, rivieren en meren. De schattingsvarianties volgen uit de meetfouten in de peilen en de dimensies (§ 4.1) en de technieken die gegeven zijn in § 4.2.

5 Foutenanalyse, meetnetoptimalisatie en monitoring

Grofweg kan men 'monitoring' beschrijven als het in de tijd volgen van variabelen om na te gaan of de grootte van deze variabelen zich ontwikkelt in overeenstemming met vooraf gestelde beleidsdoelen. Om de mate van overeenstemming van de gemonitorde variabelen te toetsen moet men een indruk hebben hoe nauwkeurig de variabelen bekend zijn. Hieruit volgt het "monitoring vraagstuk": Welke variabelen moet men meten, hoe precies, hoe vaak en waar moet men deze meten, om een te monitoren variabele met een bepaalde nauwkeurigheid te kunnen volgen tegen minimale kosten. Het oplossen van het monitoring vraagstuk heet ook wel 'meetnetoptimalisatie'. Bij meetnetoptimalisatie moet dus een gewenste nauwkeurigheid (= maximale onzekerheid) worden afgewogen tegen gewenste kosten. De foutenanalyse-technieken die in dit rapport staan beschreven vormen in feite een blauwdruk voor het optimaliseren van meetnetten.

5.1 Het monitoren van de kwaliteit van een berekende waterbalans

In paragraaf 3.3 werd gesteld dat de kleinste systematische afwijking in de waterbalans die kan worden opgemerkt gelijk is aan $2\sigma_r$. We noemen dit de onzekerheid van de waterbalans en hoe kleiner deze waarde is des te beter is de kwaliteit van de berekende waterbalans als geheel. In het geval dat de restterm dient om een onbekende term uit de waterbalans te schatten geldt de onzekerheid van deze term in de volgende beschouwing ook als een geaggregeerde maat voor de onzekerheid van de waterbalans als geheel. Uit de gepresenteerde technieken in hoofdstuk 4 is duidelijk dat we de onzekerheid kunnen verkleinen door:

- nauwkeuriger te meten;
- vaker te meten (soms betekent dit ook vaker berekeningen uit te voeren: zie § 4.5);
- op meer plaatsen te meten (soms betekent dit ook op meerdere plaatsen berekeningen uit te voeren: zie § 4.5).

Met de gepresenteerde technieken is het mogelijk, met behulp van reeds beschikbare informatie, de verandering van de onzekerheid $2\sigma_r$ te berekenen bij veranderingen in precisie, hoeveelheid en frequentie van de metingen. We hebben dus de beschikking over een "foutenmodel". De meet- en berekeningskosten nemen toe naarmate we $2\sigma_r$ kleiner eisen. Het "kostenmodel" geeft nu aan hoe financiële, materiële (en eventueel maatschappelijke) kosten gerelateerd zijn aan de kwaliteit (precisie) en kwantiteit (hoeveelheid en frequentie) van de metingen. Door het foutenmodel en het kostenmodel te combineren verkrijgen we een criterium voor het optimaliseren van het meetnetwerk ten behoeve van de waterbalans. Zo kunnen we ons bijvoorbeeld afvragen: Welke variabelen moet men meten, hoe precies, hoe vaak en waar moet men deze meten, om een systematische fout van tenminste 50 mm te kunnen opmerken tegen minimale kosten? Een dergelijke optimalisatie zal zich met name richten op de termen van de waterbalans die de grootste bijdrage aan $\text{var}(R^*)$

leveren. Het resultaat van een dergelijke optimalisatie is dat als voor een bepaald jaar de restterm groter is dan 50 mm de berekening moet worden overgedaan, bepaalde termen moeten worden toegevoegd of verwijderd of er moet gezocht gaan worden naar verbeterde berekeningsmethoden voor bepaalde termen in de waterbalans.

5.2 Het monitoren van veranderingen in termen van de waterbalans

De methoden die beschreven zijn in de vorige hoofdstukken kunnen ook gebruikt worden om te toetsen of er sprake is van een trend in één van de termen van de waterbalans. Stel dat men wil toetsen of de gebiedsafvoer is veranderd. Hiervoor zijn de totale gebiedsafvoeren in periode 1 (D_1^*) en periode 2 (D_2^*) geschat, alsmede de bijbehorende schattingsvarianties σ_{d1}^2 en σ_{d2}^2 berekend. We toetsen nu:

$$H_0 : d_1 = d_2 \text{ (de afvoeren in perioden 1 en 2 zijn gelijk)}$$

$$H_1 : d_1 \neq d_2 \text{ (de afvoeren in perioden 1 en 2 zijn niet gelijk)}$$

Om dit te toetsen berekenen we eerst de gemiddelde variantie: $\sigma_d^2 = \frac{1}{2}(\sigma_{d1}^2 + \sigma_{d2}^2)$. We kunnen nu met 95% zekerheid stellen dat de afvoer in periode 1 anders is dan in periode 2 als geldt dat het verschil ($D_1^* - D_2^*$) groter is dan $2\sigma_d$ of kleiner dan $-2\sigma_d$. Op deze wijze kunnen we dus toetsen of trends significant zijn. Meetnetoptimalisatie voor trenddetectie zal zich met name moeten richten op het kleinst te detecteren verschil: $2\sigma_d$. Met de technieken die in hoofdstuk 4 zijn gepresenteerd en informatie uit reeds beschikbare metingen kunnen we ook een inschatting te maken van de verandering van de onzekerheid $2\sigma_d$ bij veranderingen in precisie, hoeveelheid en frequentie van de metingen. De meetnetoptimalisatie ten behoeve van trenddetectie komt dan bijvoorbeeld neer op het beantwoorden van de volgende vraag: Welke variabelen moet men meten, hoe precies, hoe vaak en waar moet men deze meten, om een verschil van tenminste 20 mm tussen perioden te kunnen aantonen tegen minimale kosten. Meetnetoptimalisatie voor trenddetectie behoeft zich niet alleen te richten op gebiedsgemiddelden. We kunnen ons ook richten op het kunnen detecteren van een minimaal verschil in een bepaalde balansvariable op één meetpunt. In dat geval wordt alleen de meetfrequentie en eventueel de precisie van de metingen geoptimaliseerd.

Meetnetoptimalisatie voor de monitoring van de waterbalans zou in zijn meest uitgebreide vorm het volgende inhouden:

- optimalisatie van de meetprecisie en meetfrequentie voor elke afzonderlijke term van de waterbalans op elk afzonderlijk meetpunt om een bepaald minimaal verschil tussen balansperioden te kunnen aantonen;
- optimalisatie van het aantal meetpunten voor elke variabele om een bepaald minimaal verschil van ruimtelijk gemiddelde balanst termen tussen balansperioden te kunnen aantonen;
- eventuele verdere verhoging van de meetprecisie, de meetfrequentie en het aantal meetpunten om een bepaald minimaal systematisch verschil in de restterm tussen balansperioden te kunnen detecteren.

Afhankelijk van het doel van de monitoring kunnen alleen bepaalde stappen van deze meest uitgebreide vorm worden uitgevoerd (zie bijvoorbeeld Van Geer en Gieske, 1995). Verder geldt dat het doel van één (of twee) van bovenstaande stappen kan zijn bewerkstelligd zonder dat aan het doel van de andere stappen wordt voldaan.

6 De stoffenbalans

In het waterbeheer is ook de stoffenbalans van belang. De technieken die in deze bijdrage zijn gepresenteerd kunnen direct worden toegepast op de stoffenbalans, voor zover het conservatieve stoffen betreft. In dat geval worden de waterhoeveelheden in vergelijking (1) vervangen door vrachten [M] of vrachten per oppervlakte-eenheid [ML^{-2}]. De toepassing van het algemene foutenmodel bij de meetnetoptimalisatie ten behoeve van de monitoring van vrachten van individuele termen en de nauwkeurigheid van de totale stoffenbalans is identiek aan die voor de waterbalans (hoofdstuk 5). Bij het schatten van de vrachten voor de individuele termen en de bijbehorende schattingsvarianties komt er de extra meting bij van de concentratie (plus een meetfout) en een extra vermenigvuldiging van de concentratie met waterhoeveelheid of intensiteit.

Als voorbeeld zullen we bekijken hoe de chloridevracht via kwel kan worden geschat, alsmede de bijbehorende schattingsvariantie. Op zekere punten $\mathbf{x}_i$ in het beheersgebied worden op zekere tijdstippen t_i grondwaterstanden bemeten (diep $h_d(\mathbf{x}_i, t_i)$ en ondiep $h_o(\mathbf{x}_i, t_i)$) en de concentratie chloride van het diepere grondwater $C_{ch}(\mathbf{x}_i, t_i)$ [ML^{-3}]. Ook de verticale weerstand $c(\mathbf{x}_i)$ [T] van de ondergrond is geschat. De geschatte vracht chloride per tijdseenheid en per oppervlakte-eenheid [$\text{MT}^{-1}\text{L}^{-2}$] op deze punten wordt dan gelijk aan $V_{kwel}(\mathbf{x}_i, t_i) = C_{ch}(\mathbf{x}_i, t_i)[h_d(\mathbf{x}_i, t_i) - h_o(\mathbf{x}_i, t_i)]/c(\mathbf{x}_i)$. Als de varianties van de meetfouten (§ 4.1) van $h_d(\mathbf{x}_i, t_i)$, $h_o(\mathbf{x}_i, t_i)$, $C_{ch}(\mathbf{x}_i, t_i)$ en $c(\mathbf{x}_i)$ bekend zijn kan hieruit via de foutendoorrekening (§ 4.2) de verwachte $V_{kwel}(\mathbf{x}_i, t_i)$ (= meting voor dat tijdstip) en de variantie (= variantie meetfout) worden berekend. Vervolgens volgt voor elk meetpunt $\mathbf{x}_i$ uit de tijdsintegratie (§ 4.3) de totale vracht per oppervlakte-eenheid voor de balansperiode [ML^{-2}] en de bijbehorende variantie van de fout. Hieruit kan dan vervolgens de ruimtelijk gemiddelde vracht per oppervlakte-eenheid worden geschat [ML^{-2}] en de bijbehorende schattingsvariantie (§ 4.4).

7 Toepassing op waterbalans Polder Oudendijk

7.1 Inleiding

Om de theorie van de foutenanalyse te illustreren wordt hier een kort voorbeeld uitgewerkt. Het voorbeeld komt uit de PHLO-cursus "Omgaan met de Waterbalans" en is onderdeel van een praktische oefening gegeven door ir. B. van der Veer van het Hoogheemraadschap van Rijnland te Leiden. De volgende beschrijving is letterlijk uit de opgave overgenomen:

Polder Oudendijk ligt in het gebied van het Hoogheemraadschap van Rijnland. Het is een polder met een agrarisch karakter met een oppervlakte van 384 ha. De neerslag en verdampingscijfers voor 1993 en 1994 zijn weergegeven in tabel 1. Het neerslagoverschot wordt afgevoerd via een gemaal, dat het overtollige water uit slaat op het boezemwater. Het gemaal van de polder heeft een capaciteit van 2502 m³/uur. De draaiuren van het gemaal worden geregistreerd met behulp van het energieverbruik. De verdeling van de maaluren over de maanden is niet bekend, maar van een vergelijkbare polder wel (tabel 1). Bij watertekort wordt water ingelaten via een tiental inlaten. De inlaten worden door de boeren bediend. In de polder ligt de R.W.Z.I. van Woubrugge, die loost op een van de vaarten. De gemiddelde lozing van effluent bedroeg in 1993 1036 m³/dag en in 1994 956 m³/dag. Volgens een kwelkaart ligt de kwel tussen de 0,5 en 1 mm/dag.

Tabel 1 Gegevens om de winterbalans van Polder Oudendijk te berekenen

Maand	Neerslag (mm/mnd)	Verdamping (mm/mnd)	Maaluren (h/mnd)	R.W.Z.I. (m ³ /d)
Okt 1993 ³¹	88	25	169	1036
Nov 1993 ³⁰	61	9	194	1036
Dec 1993 ³¹	161	5	272	1036
Jan 1994 ³¹	90	9	196	956
Feb 1994 ²⁸	25	15	127	956
Mrt 1994 ³¹	81	34	159	956

De eerste vraag was om de waterbalans van het hydrologisch winterhalfjaar op te stellen, dus van 1 oktober 1993 t/m 31 maart 1994. Hierbij wordt aangenomen dat bergingsveranderingen te verwaarlozen zijn en dat er in de winter geen waterinlaat plaatsvindt. De resultaten staan in tabel 2. We zien dat er een restterm (sluitpost) van 135 mm overblijft. Men kan een aantal verklaringen trachten te vinden voor deze restterm: de aanname "geen inlaat in de winter" klopt niet; de kwel is te laag genomen; de uitlaat klopt niet omdat voor de verdeling van de maaluren over de maanden de cijfers van een andere polder zijn gebruikt; de aanname "geen bergingsverandering" klopt niet. Deze verklaringen betreffen alleen systematische

fouten. Men kan zich echter de vraag stellen of de restterm niet het gevolg is van toevallige fouten bij de berekening en de meting van de verschillende termen. De actie bij een systematische fout is dat de berekening anders moet, bijvoorbeeld met meer termen of met andere aannamen. Bij een toevallige fout kan de restterm net zo goed het geval zijn van de grote onzekerheid in één of meer van de termen en kan de restterm het volgend jaar weer een veel kleinere of grotere waarde hebben of zelfs aan de output kant zitten. Om dit te onderzoeken gebruiken we de foutenanalyse.

Tabel 2 Winterbalans Polder Oudendijk (in mm)

IN		UIT	
Neerslag	506	Verdamping	97
Inlaat	0	Uitlaat	728
R.W.Z.I.	48		
Kwel	137		
Totaal	691	Totaal	825
Sluitpost:	134		

7.2 Het foutenmodel

Als we de winter-waterbalans voor Polder Oudendijk uitdrukken in termen van vergelijking (1) dan ziet deze er als volgt uit (alles in mm):

$$r = p + s + k - e - d \tag{55}$$

met

- p : neerslag
- s : aanvoer uit R.W.Z.I.
- k : kwel (>0) of infiltratie (<0)
- e : verdamping
- d : gebiedsafvoer

Bij de foutenrekening worden bovenstaande termen geschat zodat ze stochastisch (d.w.z. kansvariabelen) worden. Dus we hebben voor de schatter van de restterm van de waterbalans (zie verg. 23):

$$R^* = P^* + S^* + K^* - E^* - D^* \tag{56}$$

De variantie van de schattingsfout wordt gegeven door:

$$\text{var}(R^*) = \sigma_r^2 = \sigma_p^2 + \sigma_s^2 + \sigma_k^2 + \sigma_e^2 + \sigma_d^2 \tag{57}$$

Het betrouwbaarheidsinterval van de restterm is dan gelijk aan $[R^* - 2\sigma_r, R^* + 2\sigma_r]$. Als 0 dus niet in dit interval ligt dan is de waarde van de restterm significant en kunnen we stellen dat er een systematische fout zit in de waterbalans. Uit tabel 2 volgt dat $R^* = 134$ mm. Om het betrouwbaarheidsinterval te berekenen moeten we vervolgens σ_r^2 berekenen en dus de varianties van alle termen in de winterwaterbalans van Polder Oudendijk. Er moet gezegd worden dat men bij het berekenen van de waterbalans met een foutenanalyse eigenlijk wat extra informatie nodig heeft. Naast informatie over de verschillende termen zelf zal ook informatie moeten worden verzameld over hun onzekerheid. Dit vereist meestal wat extra metingen. In dit voorbeeld zal ik het dus moeten doen met minder informatie dan men nodig heeft voor een uitgebreide fouten-analyse. Ik zal dus waar nodig wat aannamen (geïnformeerde schattingen) over de onzekerheid moeten maken.

7.3 Berekening van de varianties van de termen

In het volgende wordt getracht de varianties van de schatters van de verschillende termen van de bovenstaande waterbalans te achterhalen.

Neerslag

Bekend is dat in regenmeters fouten van 10% per neerslagperiode kunnen optreden. Op basis hiervan lijkt een fout van 5% in de totale neerslaghoeveelheid op maandbasis een redelijke schatting. Als we aannemen dat de standaardafwijking een goede maat is voor de fout dan geldt voor deze standaardafwijking:

$$\sigma_{pi} \approx 0,05 \cdot P_i^* \quad (58)$$

waarin P_i^* en σ_{pi} respectievelijk de geschatte hoeveelheid neerslag voor maand i en de standaardafwijking van de bijbehorende schattingsfout zijn. Als we aannemen dat de schattingsfouten voor afzonderlijke maanden onafhankelijk zijn (een redelijke aanname) dan kunnen we de schattingsfout van totale hoeveelheid in n maanden berekenen als:

$$\sigma_p^2 = \sum_{i=1}^n \sigma_{pi}^2 = \sum_{i=1}^n (0,05)^2 (P_i^*)^2 \quad (59)$$

Voor het voorbeeld betekent evaluatie van vergelijking (5) voor de winterperiode (maanden okt 93 t/m mrt 1994) een waarde van $\sigma_p^2 \approx 132$ mm². Stel nu dat in het gebied de neerslag gemeten wordt met drie regenmeters (op aselekt gekozen posities) en dat alle variatie in maandtotalen binnen het gebied het gevolg is van meetfouten. Dan geldt: $\sigma_p^2 \approx 132/3 = 44$ mm².

R.W.Z.I.

Allereerst nemen we aan dat de variaties in dagdebiet (of fouten in gemeten dagdebiet) van de R.W.Z.I. onafhankelijk zijn. In dat geval geldt de volgende relatie tussen de variantie in dagdebiet en de variantie in het jaargemiddelde dagdebiet (de laatste is gegeven in de tabel 1):

$$\text{var}(\text{dag}) = \text{var}(\text{jaar}) \cdot 365 \quad (60)$$

en geldt voor het winterhalfjaar (182 dagen):

$$\text{var}(\text{winter}) = \text{var}(\text{dag})/182 = 2 \cdot \text{var}(\text{jaar}) \quad (61)$$

De variantie van het jaargemiddelde debiet kan worden geschat als:

$$\text{var}(\text{jaar}) = [(1036-956)/1]^2 = 6400 \text{ m}^6/\text{d}^2 \quad (62)$$

en dus volgt voor de variantie van het halfjaargemiddelde: $\text{var}(\text{winter}) = 12\,800 \text{ m}^6/\text{d}^2$. Bij berekening van de R.W.Z.I. aanvoer in mm moet het debiet vermenigvuldigd worden met het aantal dagen T en gedeeld door het oppervlak van het gebied A zodat we uiteindelijk krijgen:

$$\sigma_s^2 = \frac{T^2 \text{var}(\text{halfjaar})}{A^2} = \frac{(182)^2 \cdot 12800}{(3840)^2} \approx 29 \text{ mm}^2 \quad (63)$$

Kwel

Volgens de kwelkaart varieert de kwel tussen de 0,5 en 1,0 mm/d. De waterbalans in tabel 2 is gebaseerd op 0,75 mm/d. Op grond van de kwelkaart nemen we aan dat de jaargemiddelde kwelsterkte op een punt kan variëren tussen de 0,5 en 1,0 mm/d. Het is waarschijnlijk dat het bereik van de gebiedsgemiddelde jaargemiddelde kwelsterkte kleiner is, maar omdat we hierover geen informatie hebben moeten we het "worst case" scenario beschouwen en dus voor de gebiedsgemiddelde jaargemiddelde kwelsterkte met hetzelfde bereik werken als voor de jaargemiddelde kwelsterkte op een punt. Als $[0,5;1,0]$ het 95% interval is van de verdeling in gebiedsgemiddelde jaargemiddelde kwelsterkte en deze is normaal verdeeld dan geldt voor de standaardafwijking van de gebiedsgemiddelde jaargemiddelde kwelsterkte $\sigma = (1,0-0,5)/4 = 0,125 \text{ mm/d}$. Omdat we verder niets weten dan het bereik van de kwelsterkte is de schattingsvariantie van de jaarsom kwel (mm) gelijk aan de populatievariantie van jaarsommen kwel: $\sigma_k^2 = \sigma_{182}^2 = 0,125^2 \cdot 182^2 = 518 \text{ mm}^2$.

Uitlaat

De verdeling van het aantal maaluren is afgeleid van een ander gemaal. Stel dat per maand het aantal maaluren 5% kan afwijken van de cijfers zoals die gegeven zijn in de tabel. We hebben dan een probleem vergelijkbaar met dat van de neerslag:

$$\sigma_{hi} \approx 0,05 \cdot H_i^* \quad (64)$$

waarin H_i^* en σ_{hi} respectievelijk de geschatte hoeveelheid maaluren voor maand i en de standaardafwijking van de bijbehorende schattingsfout zijn. Dan geldt voor de schattingsfout van de totale hoeveelheid maaluren in n maanden (onder aanname van onafhankelijkheid van schattingsfouten van verschillende maanden):

$$\sigma_h^2 = \sum_{i=1}^n \sigma_{hi}^2 = \sum_{i=1}^n (0,05)^2 (H_i^*)^2 \quad (65)$$

Hieruit vinden we voor de variantie van de totale hoeveelheid uitgeslagen water:

$$\sigma_d^2 = \frac{\sigma_h^2 (\text{capaciteit})^2}{A^2} = \frac{550 \cdot (5202)^2}{(3840)^2} \approx 233 \text{ mm}^2 \quad (66)$$

Verdamping

Stel dat de verdamping per etmaal wordt berekend (in mm) met een standaardafwijking van de meetfout van $\sigma_m = 0,05$ mm. Omdat de verdamping veelal berekend wordt uit een groot aantal andere variabelen en velen hiervan over de tijd gemiddelde variabelen zijn, is het zeer aannemelijk dat de fouten in de verdampingsberekeningen per etmaal statistisch afhankelijk zijn. Stel dat de correlatielengte I (uit vergelijking 39) van de meetfout 10 dagen is. De variantie van de fout op de totale verdamping is dan $(\rho(t_i, t_j))$ uit vergelijking 39) met $I=10$, $n = 182$:

$$\sigma_e^2 = (0,05)^2 \sum_{i=1}^n \sum_{j=1}^n \rho(t_i, t_j) \approx 9 \text{ mm}^2 \quad (67)$$

Restterm

De variantie van de restterm volgt uit vergelijking (3):

$$\sigma_r^2 = 44 + 29 + 518 + 233 + 9 = 833 \text{ mm}^2 \quad (68)$$

7.4 Interpretatie van de restterm

De totale variantie van de restterm is gelijk aan 833 mm^2 . Dit betekent dat het betrouwbaarheidsinterval gelijk is aan $[R^* - 2\sigma_r, R^* + 2\sigma_r] = [134 - 2 \cdot 29, 134 + 2 \cdot 29] = [76, 192] \text{ mm}$. De waarde "0" ligt niet in dit betrouwbaarheidsinterval zodat we kunnen aannemen dat er een systematische fout zit in de waterbalans of in de onderliggende aannamen ervan. De onzekerheid van de waterbalans is echter zeer groot. Een restterm van 76 mm is al niet meer significant van 0 te onderscheiden en is dus de maximale toevallige fout die met deze berekening kan worden gemaakt.

Duidelijk is ook dat het grootste probleem ligt in de kwelberekening. Omdat geen metingen gedaan zijn, zijn we gedwongen te werken met een grove schatting van de gebiedsgemiddelde kwelsterkte die constant wordt verondersteld in de tijd. Omdat voor het gehele gebied en voor het gehele halfjaar met één kwelsterkte wordt gewerkt heeft een kleine systematische fout in deze kwelsterkte grote gevolgen voor de systematische fout in de waterbalans. Ook een toevallige fout in de kwelsterkte werkt

sterk door zoals blijkt uit de grote variantie van de schattingsfout van de totale kwelhoeveelheid. Het ligt dan ook voor de hand de kwelsterkte op een aantal punten in het gebied te schatten uit grondwaterstandsmetingen op meerdere diepten en schattingen van de verticale weerstanden van de deklaag.

De vraag is dan op hoeveel plaatsen we moeten meten om een met een bepaalde betrouwbaarheid de kwel te schatten. We gaan uit van dezelfde populatievariantie als voorheen, $\sigma_{182}^2 = 0,125^2 \cdot 182^2 = 518 \text{ mm}^2$. Stel dat we n grondwaterstandsbuizen op aselechte lokaties zouden zetten (met een diep en een ondiep filter) en daar de totale kwel voor de winterperiode uit zouden berekenen. Als de gebiedsgemiddelde kwel voor de winter (mm) berekend wordt uit deze n buizen dan kunnen we de volgende steekproefvariantie verwachten:

$$S_k^2 = [n/(n-1)] \cdot \sigma_{182}^2 \quad (69)$$

En variantie van de schatter van het gemiddelde σ_k^2 wordt geschat door:

$$\sigma_k^2 = S_k^2/n = \sigma_{182}^2/(n-1) \quad (70)$$

We kunnen hieruit zien dat we tenminste drie waarnemingspunten moeten hebben om een grotere nauwkeurigheid te garanderen. Bij drie waarnemingspunten vinden we $\sigma_k^2 = 259 \text{ mm}^2$. De onzekerheid van de restterm is dan gelijk aan $2\sigma_r = 48$. Om te concluderen dat de restterm niet significant afwijkt van 0 moet deze dus liggen in het interval $[-48, 48] \text{ mm}$. De geschatte gemiddelde kwel moet dan liggen in het interval $[223, 319] \text{ mm}$. Ligt de kwel die geschat wordt uit het gemiddelde van de drie waarnemingspunten niet in dit interval dan volgt de conclusie dat er een systematische fout zit in één of meer van de overige termen van de waterbalans. Bij meer waarnemingen zal deze conclusie eerder volgen. Dit is logisch, omdat men bij meer waarnemingen meer mag verwachten van de kwaliteit van de waterbalansberekeningen

We zien uit dit voorbeeld dat het met de juiste voorinformatie in principe mogelijk is om de schattingsvarianties van alle termen van de waterbalans te berekenen. Naast het toetsen op systematische afwijkingen van de waterbalans en het berekenen van de nauwkeurigheid van de waterbalans als geheel, kunnen we individuele termen voor verschillende jaren met elkaar vergelijken en toetsen of er significante trends in bepaalde termen van de waterbalans zitten. Ook kunnen we op deze wijze achterhalen wat, waar, hoeveel en hoe precies gemeten moet worden om een trend in een zekere balansterm met een voorgeschreven nauwkeurigheid op te merken. Foutenanalyse van de waterbalans is dus ook een instrument voor monitoring en meetnetoptimalisatie.

7.5 Berekening van de kwel uit de restterm

Tenslotte volgt nog een voorbeeld waarbij dat de winterwaterbalans wordt opgesteld om uit de restterm de kwel te schatten. De schatting van de kwel wordt dan $825 - 554 = 271$ mm (1.49 mm/d). De variantie van de geschatte kwel wordt dan gegeven door:

$$\text{var}(K^*) = \sigma_k^2 = \sigma_p^2 + \sigma_s^2 + \sigma_e^2 + \sigma_d^2 = 315 \text{ mm}^2 \quad (71)$$

Dit levert een betrouwbaarheidsinterval op van $[R^* - 2\sigma_r, R^* + 2\sigma_r] = [271 - 2 \cdot 18, 271 + 2 \cdot 18] = [235, 307]$. Zelfs de bovengrens van de kwelsterke volgens de kwelkaart (182 mm) ligt niet in dit betrouwbaarheidsinterval. Er is dus sprake van een systematische fout. Of de berekende kwel uit de waterbalans klopt niet door systematische fouten in één of meer van de balanstermen en/of de waarden op de kwelkaart zijn onjuist. De enige manier om hier achter te komen is de kwelsterkte proberen te schatten op een aantal punten in het gebied (zie hoofdstuk 4). Stel dat we uitgaan van eenzelfde populatievariantie als berekend uit vergelijking (18). Bij twee aselekt gekozen meetlocaties betekent dit dat wanneer het gemiddelde van beide metingen niet tussen de 235 en 307 mm ligt we zullen concluderen te concluderen dat er in één of meer van de overige termen van de waterbalans een systematische fout zit.

8 Conclusies

De sluitpost of restterm van de waterbalans geeft een indicatie van fouten die gemaakt zijn bij de waterbalansberekeningen. Door de restterm te schatten en de bijbehorende schattingsvariantie te berekenen kan men toetsen of de fout in de waterbalans systematisch is of dat er sprake is van een toevallige afwijking. Als voor een bepaalde periode de afwijking in de waterbalans als systematisch wordt aangemerkt dan is dat een teken dat de berekeningen van de waterbalans termen moeten worden herhaald of verbeterd of dat bepaalde termen ten onrechte niet of ten onrechte wel in de waterbalans zijn opgenomen.

De schattingsvariantie van de restterm is een maat voor de minimale systematische fout in de waterbalans die nog kan worden ontdekt. Het is dus een maat voor de onzekerheid in de waterbalansberekeningen. Deze onzekerheid kan worden teruggebracht door de individuele termen van de waterbalans nauwkeuriger te schatten, met name de termen die het meest bijdragen tot de onzekerheid in de restterm.

Vaak worden in de praktijk niet alle balans termen apart berekend, maar wordt de restterm gebruikt om een onbekende balans term zoals kwel/wegzijging of waterinlaat te berekenen. Men kan dan toetsen of de uit de waterbalans berekende onbekende balans term significant afwijkt een of andere verwachte waarde. Is er een significant verschil dan kan er sprake zijn van een systematische fout. Of sommige van de termen van de waterbalans worden systematisch fout berekend of de waarde die men op grond van andere bronnen verwachtte voor deze term is onjuist. De schattingsvariantie van een zo berekende onbekende term is dan tevens een geaggregeerde maat voor de onzekerheid van de waterbalans als geheel.

De schattingen en varianties van de schattingsfouten van de termen van de waterbalans kunnen worden berekend met technieken van foutenanalyse die betrekking hebben op een viertal stappen in het berekeningsproces:

- het meten van grootheden;
- het berekenen van afgeleide grootheden;
- tijdsintegratie;
- ruimtelijke integratie.

Bij tijdsintegratie en ruimtelijke integratie is er de keuze tussen de steekproefbenadering en de stochastische-functiebenadering. Steekproefmethoden kunnen alleen gebruikt worden als de metingen willekeurig in de tijd of in de ruimte zijn genomen. Daar staat tegenover dat de berekeningen veel eenvoudiger zijn en dat ze al toepasbaar zijn bij een gering aantal metingen.

De voorbeelden in dit rapport tonen aan dat de foutenanalyse niet zomaar een toevoeging is op de normale waterbalansberekeningen. Als de waterbalans als geheel of bepaalde termen van de waterbalans gemonitord moeten worden om beleidsmaatregelen te ondersteunen of het effect ervan te toetsen, dan dient de foutenanalyse zelfs een centrale rol te spelen: een belangrijke stap bij het opzetten

van een monitoring systeem is namelijk de meetnetoptimalisatie, en de laatste is niet mogelijk zonder een behoorlijke foutenanalyse.

De technieken die in dit rapport zijn gepresenteerd kunnen direct worden toegepast voor de foutenanalyse van de stoffenbalans, voor zover deze beperkt blijft tot conservatieve stoffen.

Het algemene foutenmodel en de technieken van foutenrekening zijn toegepast voor een foutenanalyse van de winterwaterbalans van Polder Oudendijk (Hoogheemraadschap Rijnland). De foutenanalyse toont aan dat er sprake is van een systematische fout in de berekening van de winterwaterbalans van Polder Oudendijk. Analyse van de individuele termen maakt duidelijk dat de grootste fouten optreden bij de berekening van de kwelterm. De toekomstige meetinspanning dient zich dus vooral te richten op het nauwkeuriger bepalen van de kwel in het gebied.

Literatuur

Bakel, P.J.T. van, en J. de Kruijff, 1994. *Plan voor de inrichting van een meetnet in de waterschappen Noordlik Westergoa, Tuskenwaed en Ie*. Deventer, Tauw Civiel en Bouw, Rapport 941341.wp1.

Belmans, C., J.G. Wesseling en R.A. Feddes, 1983. 'Simulation model of the water balance of a cropped soil: SWATRE'. *Journal of Hydrology* 63: 271–286.

Bierkens, M.F.P., 1994. *Complex confining layers; a stochastic analysis of hydraulic properties at various scales*. Utrecht, Proefschrift Universiteit Utrecht, Nederlandse Geografische Studies 184.

Brus, D.J. en J.J. de Gruijter, 1993. 'Design-based versus model-based estimates of spatial means: theory and application in environmental soil science'. *Environmetrics* 4(2): 123-152.

Cochran, W.G., 1977. *Sampling Techniques*, 3rd edition. Wiley, New York.

Deutsch, C.V. en A.G. Journel, 1992. *GSLIB, Geostatistical software library and user's guide*. Oxford University Press, New York.

Diepen, C.A. van, J. Wolf, H. van Keulen en C. Rappoldt, 1989. 'WOFOST: a simulation model of crop production' *Soil Use and Management* 5: 16-24.

Gaast, J.W.J. en J.M.P.H. Peerboom, 1996. *Effecten van de sanering van gasbronnen in en rond de Beemster op de nutriënten- en chloridebelasting van het oppervlaktewater*. Wageningen, Staring Centrum DLO, Rapport 411.

Geer, F.C. van en J.M.J. Gieske, 1995. *Standaard meetprotocol verdroging; richtlijnen voor meetnetontwerp en analyse van de meetgegevens*. Delft, TNO Grondwater en Geo-Energie, rapport OS 95-07A.

Heuvelink, G.B.M., 1993. *Error propagation in quantitative spatial modelling; applications in geographical information systems*. Utrecht, Proefschrift Universiteit Utrecht, Nederlandse Geografische Studies 163.

Heuvelink, G.B.M., P.A. Burrough en A. Stein, 1989. 'Error propagation in spatial modelling with GIS'. *International Journal of Geographical Information Systems* 3, pp. 303-322.

Isaaks, E.H. en R.M. Srivastava, 1989. *An introduction to applied geostatistics*. Oxford University Press.

Journel, A.G. en C.J. Huijbregts, 1978. *Mining geostatistics*. London, Academic Press.

Marsily, G. de, 1986. *Quantitative hydrogeology; groundwater hydrology for engineers*. Orlando, Academic Press.

Papoulis, A., 1986. *Probability, random variables and stochastic processes*, 2nd edition. New York, McGraw-Hill.

Press, W.H., B.P. Flannery, S.A. Teukolsky en W.T. Vetterling, 1988. *Numerical Recipes in Fortran*. Cambridge, Cambridge University Press.

Riele, W.J.M. te, E.P. Querner en A.B. Pomper, 1995. *Geostatistische interpolatie van grondwaterstanden met behulp van geografische informatie en resultaten van een regionaal stromingsmodel*. Wageningen, Staring Centrum DLO, rapport 414.

Stein, A., M. Hoogerwerf en J. Bouma, 1988. 'Use of soil-map delineations to improve (co-)kriging of point data on moisture deficits'. *Geoderma* 43: 163-177.

Witter, J.V., A.C. van Eijnsbergen en M.A.J. van Montfort, 1991. 'Covariantiestructuur van kaarten'. In: J.C. Hooghart (redactie). *Ruimtelijke statistiek in bodem en water. Verslag van de lezingendag in de Reehorst te Ede op 24 januari 1991*. Delft, Rapporten en Nota's no.24, CHO-TNO.