

**stichting
mathematisch
centrum**



AFDELING MATHEMATISCHE STATISTIEK

SD 110/77

NOVEMBER

P. GROENEBOOM, J. HEMELRIJK & F.J.A. OVERWEEL

EERSTE ANALYSE VAN DE NAUWKEURIGHEID VAN LODINGEN
VERRICHT DOOR VERSCHILLENDE SCHEPEN OP DEZELFDE
PLAATSEN op 2 EN 3 MAART 1976

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O).

Eerste analyse van de nauwkeurigheid van ladingen verricht door verschillende schepen op dezelfde plaatsen op 2 en 3 maart 1976

door

P. Groeneboom, J. Hemelrijk & F.J.A. Overweel

SAMENVATTING

De resultaten van een onderzoek naar de nauwkeurigheid van ladingen zijn met behulp van een variantie-analyse model met stochastische termen geanalyseerd. De schattingen van de parameters in dit model zijn met behulp van twee verschillende methoden verkregen: de momentenmethode en de methode van de grootste aannemelijkheid. De twee methoden leverden hetzelfde algemene beeld op en in het bijzonder een schatting van ongeveer $0,7 \text{ dm}^2$ voor de variantie van de systematische afwijkingen van de schepen ten opzichte van elkaar.

TREFWOORDEN: *gemengde modellen*

INHOUD

1. Inleiding.....	1
2. Model.....	1
3. Kwadratenschatters volgens de momenten methode.....	3
4. Berekening van de schattingen.....	7
5. Toetsingen.....	10
6. Methode van de grootste aannemelijkheid.....	12
7. Conclusies.....	21
8. Literatuur.....	22

1. INLEIDING

Bij de lodingen in de Waddenzee worden verschillende schepen gebruikt, die metingen verrichten in series, varende over bepaalde trajekten. Daar deze metingen individueel veel minder nauwkeurig zijn dan het resultaat, dat men uiteindelijk wenst te bereiken is het van groot belang te beschikken over een bruikbaar wiskundig model. Dit model moet enerzijds nog eenvoudig genoeg zijn om statistische methoden toe te passen, doch anderzijds niet zoveel vereenvoudigende veronderstellingen bevatten, dat het zich te ver van de werkelijkheid verwijderd om nog betrouwbare resultaten te geven.

In dit rapport wordt op grond van de waarnemingen, waar in de titel op wordt gedomd, een dergelijk model voorgesteld en voor zover de huidige waarnemingen en beschikbare statistische technieken dit toelaten, worden daarin schattingen berekend en toetsingen uitgevoerd. Hierbij zijn twee methoden gevolgd, een eenvoudige overzichtelijke en een verfijndere, maar minder overzichtelijke, waarvan de resultaten met elkaar worden vergeleken.

Bij eerste lezing krijgt men een indruk van de resultaten door de paragrafen 1,2 en 7 te lezen, benevens de tabellen 1,2,3,4 en 6.

2. MODEL

Als model hebben we gekozen:

$$(2.1) \quad Y_{ijk} = \mu + \xi_i + \tau_j + w_{jk} + e_{ijk} \quad \begin{array}{l} i = 1, \dots, n \text{ meetpunten} \\ j = 1, \dots, p \text{ schepen} \\ k = 1, \dots, q \text{ herhalingen} \end{array}$$

We veronderstellen τ_j , w_{jk} en e_{ijk} stochastisch onafhankelijk. Om goed te onderscheiden tussen stochastische grootheden enerzijds en parameters en verrichte waarnemingen anderzijds worden stochastische grootheden aangegeven door onderstreepte symbolen. De betekenis van de symbolen is de volgende:

Y_{ijk} is de dieptemeting van het j -de schip, op het i -de punt in het

- trajekt, tijdens de k-de vaart over dat trajekt.
- μ is een algemeen gemiddelde van alle dieptepeilingen.
- ξ_i is een afwijking van dat algemeen gemiddelde voor het i-de punt, of ook:
- $\mu + \xi_i$ is de diepte voor het i-de punt.
- τ_j is een stochastische grootheid, die het effect van het schip aan- geeft. Deze grootheid is stochastisch genomen, omdat bij andere waarnemingsseries andere schepen zijn ingeschakeld. Op deze wijze kunnen we ons het schip-effekt denken als te zijn getrokken uit een bepaalde verdeling. We veronderstellen dat τ_j een normale verdeling met verwachting 0 en variantie T^2 heeft. (Notatie $\tau_j \sim N(0, T^2)$).
- w_{jk} is de systematische fout in de k-de vaart van schip j tengevolge van bijvoorbeeld koersafwijking of getijverschillen. We veronder- stellen $w_{jk} \sim N(0, W_j^2)$.
- e_{ijk} is een algemene storingsterm. Ook hier veronderstellen we norma- liteit: $e_{ijk} \sim N(0, F_j^2)$.

De aannames van normaliteit zijn nodig voor een deel van de uit te voeren toetsen, niet voor de schattingen van §3, maar wel voor de afleiding van die van §6. Het zal duidelijk zijn, dat de τ_j een grotere invloed heeft dan w_{jk} en e_{ijk} op de uiteindelijke onnauwkeurigheid. Immers het aantal vaarten (lanes dat gevaren wordt) is groot, zeg k_j en voor elke vaart wordt een nieuwe trekking uit de $(0, W_j^2)$ -verdeling gedaan, zodat de variantie van de gemiddelde systematische fout in de vaarten $\frac{W_j^2}{k_j}$ is. Het aantal meetpunten is nog veel groter, zeg n_j , zodat de variantie van de gemiddelde systema- tische fout in de punten $\frac{F_j^2}{n_j}$ is. De variantie van de gemiddelde systema- tische fout van de τ_j 's blijft vrij aanzienlijk, daar er slechts een beperkt aantal schepen is gebruikt. Hoewel het mogelijk is dat deze τ_j in de tijd verandert, is daar nu nog niets over bekend en gaan we er voorlopig vanuit, dat de τ_j de grootste invloed heeft op de uiteindelijke onnauwkeurigheid.

3. KWADRATENSCHATTERS VOLGENS DE MOMENTEN METHODE

3.1 Schatters voor ieder der schepen afzonderlijk.

In deze paragraaf worden formules afgeleid, waarmee schatters voor de parameters F_j^2 , W_j^2 en T^2 worden verkregen. Deze formules zijn ontleend aan de theorie van de gemengde lineaire modellen, wat betekent, dat in het model naast deterministische ook meerdere stochastische termen voorkomen.

Vooraf nog even een notatie: we geven het gemiddelde over een bepaalde index aan met behulp van een punt, bijvoorbeeld: $\underline{Y}_{.jk} = \frac{1}{n} \sum_{i=1}^n \underline{Y}_{ijk}$.

De schatters, die afgeleid worden zijn:

$$(3.1) \quad \hat{F}_j^2 = \frac{1}{(q-1)(n-1)} \sum_{i=1}^n \sum_{k=1}^q (\underline{Y}_{ijk} - \underline{Y}_{ij.} - \underline{Y}_{.jk} + \underline{Y}_{.j.})^2$$

$$(3.2) \quad \hat{W}_j^2 = \frac{1}{n(q-1)} \sum_{i=1}^n \sum_{k=1}^q (\underline{Y}_{ijk} - \underline{Y}_{ij.})^2 - \hat{F}_j^2$$

$$(3.3) \quad \hat{T}^2 = \frac{1}{p-1} \sum_{j=1}^p (\underline{Y}_{.j.} - \underline{Y}_{...})^2 - \frac{1}{npq} \sum_{j=1}^p \hat{F}_j^2 - \frac{1}{pq} \sum_{j=1}^p \hat{W}_j^2$$

Afleiding:

LEMMA. Zijn $\underline{x}_1, \dots, \underline{x}_m$ stochastisch onafhankelijk met

$$E \underline{x}_h = 0 \text{ en } \sigma^2(\underline{x}_h) = E \underline{x}_h^2 = \sigma^2, \text{ dan is}$$

$$E \sum_{h=1}^m (\underline{x}_h - \underline{x}_{..})^2 = (m-1) \sigma^2 \text{ en}$$

$$E (\underline{x}_h - \underline{x}_{..})^2 = \frac{m-1}{m} \sigma^2.$$

BEWIJS.

$$\begin{aligned} \sum_{h=1}^m (\underline{x}_h - \underline{x}_{..})^2 &= \sum_{h=1}^m \{ \underline{x}_h^2 - 2 \underline{x}_h \underline{x}_{..} + \underline{x}_{..}^2 \} = \\ &= \sum_{h=1}^m \underline{x}_h^2 - \frac{2}{m} \left(\sum_{h=1}^m \underline{x}_h \right) \cdot \left(\sum_{h=1}^m \underline{x}_h \right) + m \cdot \left(\frac{1}{m} \sum_{h=1}^m \underline{x}_h \right)^2 = \end{aligned}$$

$$= \sum_{h=1}^m \underline{x}_h^2 - m \underline{x}_\cdot^2$$

$$E \sum_{h=1}^m (\underline{x}_h - \underline{x}_\cdot)^2 = m E \underline{x}_h^2 - m E \underline{x}_\cdot^2 = m \sigma^2 - m \frac{\sigma^2}{m} = (m-1) \sigma^2 = (m-1) E \underline{x}_h^2.$$

De tweede bewering in het lemma is een gevolg van het feit dat $E(\underline{x}_h - \underline{x}_\cdot)^2$ voor elke h dezelfde is:

Schatting voor F_j^2

Om F_j^2 te schatten beschouwen we voor $j = 1, \dots, p$

$$E \sum_{i=1}^n \sum_{k=1}^q (\underline{y}_{ijk} - \underline{y}_{ij\cdot} - \underline{y}_{\cdot jk} + \underline{y}_{\cdot j\cdot})^2 = \quad (\text{na substitutie van 2.1})$$

$$= E \sum_{i=1}^n \sum_{k=1}^q (\underline{e}_{ijk} - \underline{e}_{ij\cdot} - \underline{e}_{\cdot jk} + \underline{e}_{\cdot j\cdot})^2 =$$

$$= \sum_{k=1}^q (n-1) E (\underline{e}_{ijk} - \underline{e}_{ij\cdot})^2 = (n-1)(q-1) F_j^2$$

zodat een zuivere schatter voor F_j^2 is

$$\hat{F}_j^2 = \frac{1}{(n-1)(q-1)} \sum_{i=1}^n \sum_{k=1}^q (\underline{y}_{ijk} - \underline{y}_{ij\cdot} - \underline{y}_{\cdot jk} + \underline{y}_{\cdot j\cdot})^2.$$

Schatting voor W_j^2

Teneinde een schatter voor W_j^2 af te leiden beschouwen we voor $j = 1, \dots, p$

$$(3.4) \quad E \sum_{i=1}^n \sum_{k=1}^q (\underline{y}_{ijk} - \underline{y}_{ij\cdot})^2 = E \sum_{i=1}^n \sum_{k=1}^q (\underline{w}_{jk} - \underline{w}_{j\cdot} + \underline{e}_{ijk} - \underline{e}_{ij\cdot})^2 =$$

$$= E \sum_{i=1}^n \sum_{k=1}^q \{ (\underline{w}_{jk} - \underline{w}_{j\cdot})^2 + (\underline{e}_{ijk} - \underline{e}_{ij\cdot})^2 - 2(\underline{w}_{jk} - \underline{w}_{j\cdot})(\underline{e}_{ijk} - \underline{e}_{ij\cdot}) \} =$$

$$= E \sum_{i=1}^n \sum_{k=1}^q \{ (\underline{w}_{jk} - \underline{w}_{j\cdot})^2 + (\underline{e}_{ijk} - \underline{e}_{ij\cdot})^2 \},$$

immers daar $\underline{w}_{jk}$ en $\underline{e}_{ijk}$ stochastisch onafhankelijk zijn en verwachting 0 hebben is $E(\underline{w}_{jk} - \underline{w}_{j\cdot})(\underline{e}_{ijk} - \underline{e}_{ij\cdot}) = 0$. We bekijken de termen van (3.4) apart.

$$(3.5) \quad E \sum_{i=1}^n \sum_{k=1}^q (\underline{w}_{jk} - \underline{w}_{j\cdot})^2 = \sum_{i=1}^n (q-1) W_j^2 = n(q-1) W_j^2$$

$$(3.6) \quad E \sum_{i=1}^n \sum_{k=1}^q (\underline{e}_{ijk} - \underline{e}_{ij.})^2 = \sum_{i=1}^n (q-1) F_j^2 = n(q-1) F_j^2$$

dus

$$E \sum_{i=1}^n \sum_{k=1}^q (\underline{y}_{ijk} - \underline{y}_{ij.})^2 = n(q-1) W_j^2 + n(q-1) F_j^2$$

zodat een zuivere schatter voor W_j^2 wordt

$$\hat{W}_j^2 = \frac{1}{n(q-1)} \sum_{i=1}^n \sum_{k=1}^q (\underline{y}_{ijk} - \underline{y}_{ij.})^2 - \hat{F}_j^2.$$

Schatter voor T^2

We beschouwen de volgende verwachting van de som van kwadraten om T^2 te schatten

$$(3.7) \quad E \sum_{j=1}^p (\underline{y}_{.j.} - \underline{y}_{...})^2 = E \sum_{j=1}^p (\underline{\tau}_j - \underline{\tau}_{..} + \underline{w}_{j.} - \underline{w}_{..} + \underline{e}_{.j.} - \underline{e}_{...})^2 =$$

$$E \sum_{j=1}^p \{ (\underline{\tau}_j - \underline{\tau}_{..})^2 + (\underline{w}_{j.} - \underline{w}_{..})^2 + (\underline{e}_{.j.} - \underline{e}_{...})^2 \}$$

daar $\underline{w}_{jk}$, $\underline{e}_{ijk}$ en $\underline{\tau}_j$ stochastisch onafhankelijk zijn met verwachting 0.
We bekijken de drie sommen van kwadraten apart:

$$E \sum_{j=1}^p (\underline{\tau}_j - \underline{\tau}_{..})^2 = (p-1) T^2$$

$$E \sum_{j=1}^p (\underline{w}_{j.} - \underline{w}_{..})^2 = \frac{p-1}{p} \sum_{j=1}^p \sigma^2(\underline{w}_{j.}) = \frac{p-1}{pq} \sum_{j=1}^p W_j^2$$

$$E \sum_{j=1}^p (\underline{e}_{.j.} - \underline{e}_{...})^2 = \frac{p-1}{p} \sum_{j=1}^p \sigma^2(\underline{e}_{.j.}) = \frac{p-1}{npq} \sum_{j=1}^p F_j^2$$

dus

$$E \sum_{j=1}^p (\underline{y}_{.j.} - \underline{y}_{...})^2 = (p-1) T^2 + \frac{p-1}{pq} \sum_{j=1}^p W_j^2 + \frac{p-1}{npq} \sum_{j=1}^p F_j^2$$

een zuivere schatter voor T^2 is dus

$$\hat{T}^2 = \frac{1}{p-1} \sum_{j=1}^p (\underline{y}_{.j.} - \underline{y}_{...})^2 - \frac{1}{pq} \sum_{j=1}^p \hat{W}_j^2 - \frac{1}{npq} \sum_{j=1}^p \hat{F}_j^2.$$

3.2 Gewogen gemiddelde schatters

Stel we hebben voor de verschillende schepen de varianties F_j^2 , W_j^2 en T^2 geschat. We hebben dit gedaan voor 4 series waarnemingen met verschillende n en q , maar gelijke p (= aantal schepen). Deze n en q zijn n_1, n_2, n_3 en n_4 en q_1, q_2, q_3 en q_4 . Definieer nu voor elk van die waarnemingsseries de volgende sommen van kwadraten:

$$\begin{aligned} S_{\ell} &= \sum_{i=1}^{n_{\ell}} \sum_{k=1}^{q_{\ell}} (\underline{Y}_{ijk} - \underline{Y}_{ij.} - \underline{Y}_{.jk} + \underline{Y}_{.j.})^2 \\ S'_{\ell} &= \sum_{i=1}^{n_{\ell}} \sum_{k=1}^{q_{\ell}} (\underline{Y}_{ijk} - \underline{Y}_{ij.})^2 \\ S''_{\ell} &= \sum_{i=1}^{n_{\ell}} \sum_{k=1}^{q_{\ell}} (\underline{Y}_{.jk} - \underline{Y}_{.j.})^2 \end{aligned} \quad \ell = 1, 2, 3, 4.$$

We noemen de schatting voor F_j^2 op grond van waarnemingsserie ℓ $\hat{F}_{j\ell}^2$ en op analoge wijze zijn $\hat{W}_{j\ell}^2$ en $\hat{T}_{\ell}^2$ gedefinieerd.

Dan geldt voor elke $\ell = 1, \dots, 4$ en $j = 1, \dots, p$ (zie (3.1))

$$(q_{\ell}-1)(n_{\ell}-1) \hat{F}_{j\ell}^2 = S_{\ell}, \text{ sommeren over } \ell:$$

$$\sum_{\ell=1}^4 \hat{F}_{j\ell}^2 (q_{\ell}-1)(n_{\ell}-1) = \sum_{\ell=1}^4 S_{\ell}$$

de gewogen gemiddelde schatter van F_j^2 is dan

$$\hat{F}_j^2 = \frac{\sum_{\ell=1}^4 S_{\ell}}{\sum_{\ell=1}^4 (q_{\ell}-1)(n_{\ell}-1)}.$$

Op dezelfde wijze volgt uit (3.2)

$$\begin{aligned} n_{\ell} (q_{\ell}-1) \cdot (\hat{F}_{j\ell}^2 + \hat{W}_{j\ell}^2) &= S'_{\ell} \\ \sum_{\ell=1}^4 (\hat{F}_{j\ell}^2 + \hat{W}_{j\ell}^2) n_{\ell} (q_{\ell}-1) &= \sum_{\ell=1}^4 S'_{\ell}. \end{aligned}$$

De gewogen gemiddelde schatter voor W_j^2 wordt:

$$\hat{\underline{W}}_j^2 = \frac{\sum_{\ell=1}^4 \underline{S}'_{\ell}}{\sum_{\ell=1}^4 n_{\ell}(q_{\ell}-1)} - \hat{\underline{F}}_j^2.$$

Uit formule (3.3) volgt:

$$(p-1)\hat{\underline{T}}_{\ell}^2 + \frac{p-1}{n_{\ell}pq_{\ell}} \sum_{j=1}^p \hat{\underline{F}}_{j\ell}^2 + \frac{p-1}{pq_{\ell}} \sum_{j=1}^p \hat{\underline{W}}_{j\ell}^2 = \underline{S}_{\ell}''$$

$$\sum_{\ell=1}^4 (p-1) \hat{\underline{T}}_{\ell}^2 = \sum_{\ell=1}^4 \underline{S}_{\ell}'' - \sum_{\ell=1}^4 \left(\frac{p-1}{n_{\ell}pq_{\ell}} \sum_{j=1}^p \hat{\underline{F}}_{j\ell}^2 \right) - \sum_{\ell=1}^4 \left(\frac{p-1}{pq_{\ell}} \sum_{j=1}^p \hat{\underline{W}}_{j\ell}^2 \right)$$

dus de gewogen gemiddelde schatter voor T^2 wordt:

$$\hat{\underline{T}}^2 = \frac{1}{4(p-1)} \sum_{\ell=1}^4 \left[\underline{S}_{\ell}'' - \left(\frac{p-1}{n_{\ell}pq_{\ell}} \cdot \sum_{j=1}^p \hat{\underline{F}}_{j\ell}^2 \right) - \left(\frac{p-1}{pq_{\ell}} \sum_{j=1}^p \hat{\underline{W}}_{j\ell}^2 \right) \right].$$

4. BEREKENING VAN DE SCHATTINGEN

Bij het uitvoeren van de berekeningen hebben we vier volledige schema's uit het materiaal gehaald, we duiden ze aan met blok 1 t/m blok 4. Een volledig schema betekent in ons geval, dat er per schip per meetpunt evenveel herhalingen van de meting zijn.

blok 1 - inlood 4 - punten 835,00 t/m 840,00 en

$n = 26$ $p = 5$ $q = 5$ 840,75 t/m 841,75

blok 2 - inlood 6 - punten 835,00 t/m 844,50

$n = 37$ $p = 5$ $q = 2$

blok 3 - inlood 4 - punten 820,50 t/m 834,75

$n = 57$ $p = 5$ $q = 4$

blok 4 - inlood 7 - punten 835,00 t/m 840,50

$n = 16$ $p = 5$ $q = 3$ 841,50 t/m 843,00.

In de blokken 1,2 en 3 zijn drie tussenpunten gemeten tussen twee opvolgende geheelwaardige punten, dus bijvoorbeeld: 835,00; 835,25; 835,50; 835,75; etc., bij het laatste blok telkens één tussenpunt, dus: 835,00; 835,50; 836,00 etc.

Uit blok 1 en blok 4 is een aantal punten weggelaten, omdat het profiel van de bodem daar erg steil is, zodat er onnauwkeurigheid als gevolg van de steilte geïntroduceerd wordt. Om een indruk te geven van het cijfermateriaal volgt hier een aantal tabellen. De eerste bevat $\underline{Y}_{.jk} - \underline{Y}_{...} = (\underline{\tau}_j - \underline{\tau}_{.}) + (\underline{w}_{jk} - \underline{w}_{..}) - (\underline{e}_{.jk} - \underline{e}_{...})$. De tweede tabel geeft een indruk van de spreiding van het schipeffect; hierin staan de waarden van $\underline{Y}_{.j.} - \underline{Y}_{...} = (\underline{\tau}_j - \underline{\tau}_{.}) + (\underline{w}_{j.} - \underline{w}_{..}) + (\underline{e}_{.j.} - \underline{e}_{...})$. In tabel 3 vatten we de blokken van de tabel 2 samen, door elke waarde per blok gewicht $n_{\ell} \times q_{\ell}$ (aantal punten $\times$ aantal herhalingen van blok ℓ te geven en te delen door het totaal).

Tabel 1 $\underline{Y}_{.jk} - \underline{Y}_{...} = \underline{\tau}_j - \underline{\tau}_{.} + \underline{w}_{jk} - \underline{w}_{..} + \underline{e}_{.jk} - \underline{e}_{...}$

	k	Blauwe Slenk	Zwarte Haan	Breesem	Falga	A. Vierlingh
blok 1	1	-2,45	1,35	-0,22	0,28	-1,30
	2	-0,18	0,62	0,20	1,01	-0,45
	3	-2,65	0,20	0,28	-0,65	-0,22
	4	-2,03	0,12	1,59	-0,11	0,16
	5	-2,22	1,51	1,78	0,47	2,93
blok 2	1	-0,80	-0,77	0,80	1,34	0,80
	2	-0,58	-0,88	0,04	1,61	-1,55
blok 3	1	-1,65	-0,23	0,02	1,54	-3,18
	2	-1,33	0,40	0,21	1,25	-0,58
	3	-0,74	0,12	0,07	1,35	0,25
	4	-1,14	0,30	0,93	0,95	1,49
blok 4	1	-1,31	0,32	0,82	0,50	-0,50
	2	-0,93	-0,06	0,32	0,50	0,82
	3	-1,81	0,57	1,13	1,13	-1,50

Tabel 2 $Y_{.j.} - Y_{...} = \tau_j - \tau_{.} + w_{j.} - w_{..} + e_{.j.} - e_{...}$

blok 1	-1,91	0,76	0,72	0,20	0,22
blok 2	-0,69	-0,83	0,42	1,47	-0,38
blok 3	-1,22	0,15	0,30	1,27	-0,50
blok 4	-1,35	0,28	0,75	0,71	-0,39

Tabel 3 Gewogen gemiddelden van de waarden uit tabel 2

-1,34	0,18	0,48	0,96	-0,28
-------	------	------	------	-------

Opmerking: de eenheid van de getallen in de tabellen 1,2 en 3 is dm.
 We gebruiken de in §3 ontwikkelde formules om F_j^2 , W_j^2 en T^2 te schatten.
 De resultaten zijn in dm^2 .

Tabel 4 Schattingen van de varianties F_j^2 , W_j^2 en T^2 per blok

blok 1	Blauwe Slenk	Zwarte Haan	Breesem	Falga	A.Vierlingh
F_j^2	2,62	5,90	2,73	1,20	12,01
W_j^2	0,88	0,19	0,70	0,34	2,12
T^2	1,00				
blok 2					
F_j^2	1,48	1,77	4,32	2,41	5,14
W_j^2	-0,02	-0,04	0,17	-0,03	2,63
T^2	0,60				
blok 3					
F_j^2	2,64	3,36	3,89	2,73	6,29
W_j^2	0,10	0,02	0,12	0,01	3,78
T^2	0,65				
blok 4					
F_j^2	0,66	1,74	1,40	0,26	3,08
W_j^2	0,15	-0,01	0,08	0,11	1,15
T^2	0,65				

Opmerking: negatieve schattingen van de varianties kunnen ontstaan, doordat (zie formule (3.2)), het kan gebeuren, dat

$$\frac{1}{n(q-1)} \sum_{i=1}^n \sum_{k=1}^q (Y_{ijk} - Y_{ij.})^2 < \hat{F}_j^2.$$

De beste schatter is dan 0, maar voor de berekening van de gecombineerde schattingen in tabel 5 zijn de negatieve waarden gehandhaafd en gewogen meegeteld. Negatieve waarden komen dan niet meer voor.

Tabel 5 gewogen gemiddelde variantie schattingen

	Blauwe Slenk	Zwarte Haan	Breesem	Falga	A.Vierlingh
F_j^2	2,33	3,80	3,36	2,01	7,59
W_j^2	0,33	0,07	0,29	0,11	2,92
T^2	0,73				

5. TOETSINGEN

5.1 Toetsing van gelijkheid van $F_1^2, \dots, F_5^2$

Om een beter inzicht te krijgen in de waarde en betekenis van de verkregen schattingen, voeren we een aantal toetsingen uit. Laten we eerst nagaan in hoeverre het nodig is aan te nemen, dat de variantie van e_{ijk} per schip verschilt; daartoe toetsen wij

$$H_0: F_1^2 = F_2^2 = F_3^2 = F_4^2 = F_5^2.$$

Als toetsingsgrootte hiervoor nemen wij

$$\underline{V} = \max_{1 \leq j \leq p} \hat{F}_j^2 / \sum_{j=1}^p \hat{F}_j^2.$$

Kritieke waarden voor $\alpha = 0,05$ en $\alpha = 0,01$ zijn getabelleerd in tabel 31a van Biometrika Tables for Statisticians [4]. De toets wordt gewoonlijk naar zijn ontwerper W.G. Cochran genoemd. Voor onze vier blokken krijgen we

V	kritieke waarden			
	$v = (n-1)(q-1)$	$\alpha = 0,05$	$\alpha = 0,01$	
12,01/24,46 = 0,49	100	0,26	0,28	*)
5,14/15,12 = 0,34	36	0,31	0,34	
6,29/18,91 = 0,33	168	0,24	0,25	*)
3,08/ 7,14 = 0,43	30	0,31	0,35	*)

*) de kritieke waarden voor $v = 100$, 168 en 30 zijn niet getabelleerd en verkregen door interpolatie.

Blijkbaar is ongelijkheid van F_j^2 wel degelijk het geval.

5.2 Toetsing van gelijkheid van $W_1^2, \dots, W_5^2$

Om na te gaan in hoeverre het nodig is aan te nemen dat de varianties van $\underline{w}_{jk}$ per schip verschillen is de te toetsen nulhypothese

$$H_0: W_1^2 = W_2^2 = W_3^2 = W_4^2 = W_5^2.$$

We kunnen niet dezelfde toets gebruiken, omdat de schatters van de W_j^2 bestaan uit de verschillen van twee sommen van kwadraten. Daarom nemen we onze toevlucht tot de toets van Friedman, ook wel de methode van m rangschikken genoemd. Een beschrijving van de methode staat in de Jonge [3]. We nummeren de schattingen van W_j^2 per blok naar opklimmende grootte. Deze rangnummers brengen we voor de 4 blokken onder in de volgende tabel en we tellen kolomsgewijs op.

	schip 1	schip 2	schip 3	schip 4	schip 5
blok 1	4	1	3	2	5
blok 2	3	1	4	2	5
blok 3	3	2	4	1	5
blok 4	4	1	2	3	5
totaal	14	5	13	8	20

De bijbehorende toetsingsgrootte is:

$$K = \sum_{j=1}^5 (S_j - S_{\cdot})^2, \text{ waarbij } S_j \text{ de kolomtotalen zijn.}$$

$$K = (14-12)^2 + (5-12)^2 + (13-12)^2 + (8-12)^2 + (20-12)^2 =$$

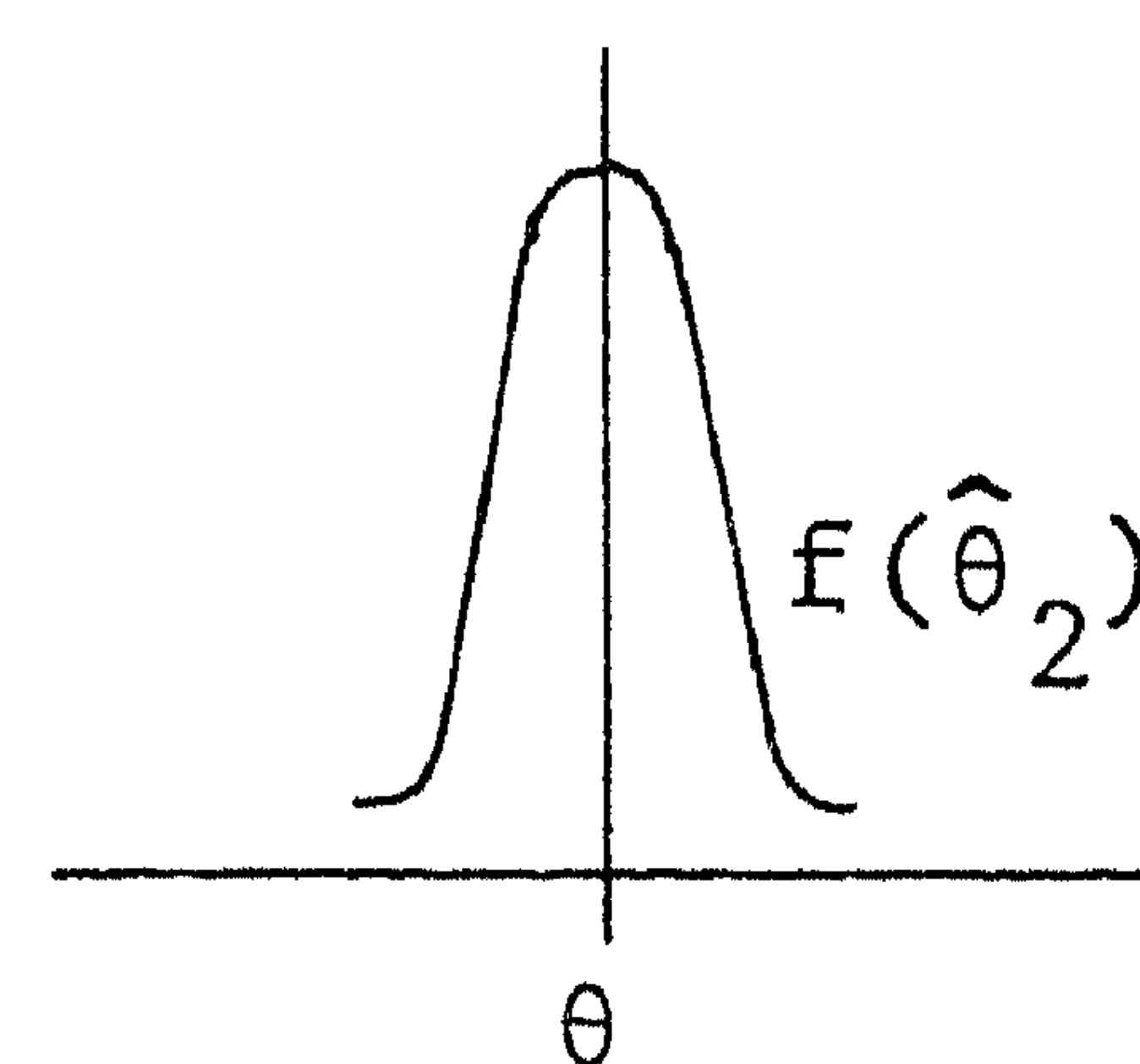
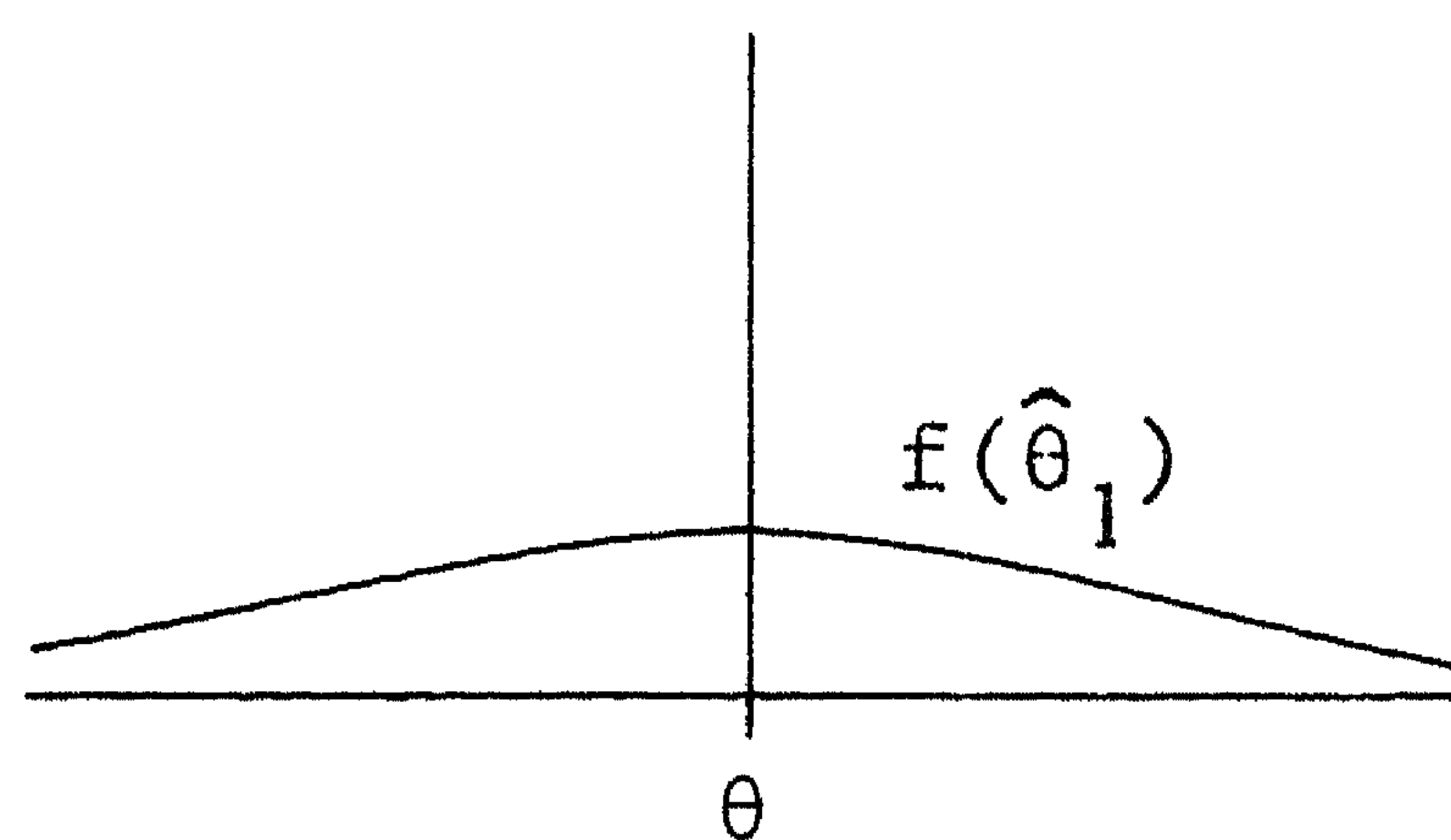
$$4 + 49 + 1 + 16 + 64 = 134$$

deze waarde overtreft de kritieke waarde voor $\alpha = 0,05$, die 89 is, verre, zodat we de nulhypothese moeten verwerpen. De aanname dat de W_j^2 per schip verschillen kan dus ook niet vermeden worden.

6. METHODE VAN DE GROOTSTE AANNEMELIJKHEID

6.1 Inleiding

In de vorige paragrafen werden de resultaten van de voor onzuiverheid gecorrigeerde "momentenmethode" bij het schatten van parameters beschreven. Dit betekent dat men bij het schatten van een parameter θ zoekt naar een schatter $\hat{\theta}$ waarvan men (onder de modelassumpties) weet dat $E\hat{\theta} = \theta$, dat wil zeggen men zoekt naar *zuivere* schatters. Hierbij moet echter worden opgemerkt dat zo'n zuivere schatter niet uniek bepaald is; er zijn meestal zeer veel zuivere schatters te bedenken, waaronder goede en slechte. Vergelijk de volgende situaties:



$\hat{\theta}_1$ en $\hat{\theta}_2$ zijn allebei zuivere schatters van een parameter θ , maar $\hat{\theta}_1$ heeft een dichtheid $f(\hat{\theta}_1)$ met tamelijk dikke staarten, terwijl de dichtheid $f(\hat{\theta}_2)$ horend bij de schatter $\hat{\theta}_2$ om de parameter θ geconcentreerd ligt. Intuïtief kan men de situatie als volgt beschrijven: hoewel $\hat{\theta}_1$ en $\hat{\theta}_2$

beide "gemiddeld" bij herhaalde experimenten een juiste schatting van de parameter θ zullen geven, zal $\hat{\theta}_2$ meestal vlak bij θ liggen terwijl $\hat{\theta}_1$ regelmatig ver uit de buurt van θ terecht zal komen.

Minstens zo belangrijk als het criterium van zuiverheid lijkt daarom het criterium van de *nauwkeurigheid* van een schatter. Een schatter $\hat{\theta}$ wordt een *nauwkeurige* schatter van θ genoemd als $E(\hat{\theta} - \theta)^2$ "klein" is. In bovenstaand voorbeeld is duidelijk $E(\hat{\theta}_2 - \theta)^2 < E(\hat{\theta}_1 - \theta)^2$.

Een methode die in het algemeen nauwkeurige schatters oplevert is de methode van de grootste aannemelijkheid. De zo verkregen "maximum likelihood" schatters zijn niet noodzakelijk zuiver, maar meestal wel asymptotisch raak, dat wil zeggen: bij toenemende steekproefomvang zal de schatter met tot 1 naderende kans willekeurig dicht bij de geschatte parameterwaarde komen te liggen.

6.2 Toepassingen van de methode van de grootste aannemelijkheid op het gemengde lineaire model

Om de methode van de grootste aannemelijkheid te kunnen toepassen, gaan we het model eerst nog in een andere vorm schrijven. Het model is

$$Y_{ijk} = \mu + \xi_i + \tau_j + w_{jk} + e_{ijk},$$

waarbij

$$i = 1, \dots, n; j = 1, \dots, p \text{ en } k = 1, \dots, q$$

$$\tau_j \sim N(0, T^2), w_{jk} \sim N(0, W_j^2) \text{ en } e_{ijk} \sim N(0, F_j^2).$$

We hanteren dezelfde onafhankelijkheidsveronderstellingen als in paragraaf 2 en bovendien gebruiken we voor de afleidingen van de meest aannemelijke schatters de normaliteitsassumpties. Zij nu

$$\underline{Y} = (\underline{Y}_{111}, \dots, \underline{Y}_{11q}, \dots, \underline{Y}_{1p1}, \dots, \underline{Y}_{1pq}, \dots, \underline{Y}_{np1}, \dots, \underline{Y}_{npq})'$$

dat wil zeggen $\underline{Y}$ is de uitgerolde kolomvector van waarnemingen waarbij de

de laatste index het snelst loopt en de eerste index het langzaamst.

Zij

$\underline{\tau} = (\tau_1, \dots, \tau_p)'$ de vector van schipeffecten

$\underline{w}_j = (w_{j1}, \dots, w_{jq})'$ de vector van "vaarteffecten" van schip j , $j = 1, \dots, p$

$\underline{e}_j = (e_{1j1}, \dots, e_{1jq}, \dots, e_{nj1}, \dots, e_{njq})'$ de vector van toevallige fouten van schip j , $j = 1, \dots, p$ en

$\theta = (\theta_1, \dots, \theta_n)'$ de vector van werkelijke dieptes op het traject (d.w.z. $\theta_i = \mu + \xi_i$):

Zij verder $A \otimes B$ het zg. Kronecker produkt van de matrices A en B , d.w.z. als

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \quad \text{en} \quad B = \begin{pmatrix} b_{11} & \dots & b_{1s} \\ \vdots & & \vdots \\ b_{r1} & \dots & b_{rs} \end{pmatrix},$$

dan is

$$A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{pmatrix}.$$

Dus $A \otimes B$ is een $mr \times ns$ matrix met bijv. op de eerste rij de getallen $a_{11}b_{11}, a_{11}b_{12}, \dots, a_{11}b_{1s}, \dots, a_{1n}b_{11}, \dots, a_{1n}b_{1s}$. Laat tenslotte de vectoren J_m , d_j en de matrix I_m gedefinieerd zijn door

$$J_m = \underbrace{(1, 1, \dots, 1)'}_m, \quad d_j = \underbrace{(0, 0, \dots, 0, 1, 0, \dots, 0)'}_p$$

en

$$I_m = \underbrace{\begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}}_m = m \times m \text{ identiteitsmatrix.}$$

Dan kan het model als volgt geschreven worden

$$(6.1) \quad \underline{Y} = X\theta + U_1 \underline{\tau} + \sum_{j=1}^p U_{2j} \underline{w}_j + \sum_{j=1}^p U_{3j} \underline{e}_j,$$

met

$$\begin{aligned} X &= I_n \otimes J_p \otimes J_q \\ U_1 &= J_n \otimes I_p \otimes J_q \\ U_{2j} &= J_n \otimes d_j d_j' \otimes I_q, \quad j = 1, \dots, p \\ U_{3j} &= I_n \otimes d_j d_j' \otimes I_q, \quad j = 1, \dots, p \end{aligned}$$

Aan de schrijfwijze (6.1) van het model kunnen we nu onmiddellijk zien dat de waarnemingsvector $\underline{Y}$ een meerdimensionale normale verdeling heeft met verwachting $\theta \otimes J_{pq}$ en covariantiematrix $\sigma^2 H$ met

$$(6.2) \quad \sigma^2 H \stackrel{\text{def}}{=} \sigma^2 \left(\gamma_1 U_1 U_1' + \sum_{j=1}^p \gamma_{2j} U_{2j} U_{2j}' + \sum_{j=1}^p \gamma_{3j} U_{3j} U_{3j}' \right),$$

waarbij $\gamma_1 = T^2/\sigma^2$, $\gamma_{2j} = W_j^2/\sigma^2$ en $\gamma_{3j} = F_j^2/\sigma^2$ (σ^2 is een hulpgrootheid ten opzichte waarvan de andere varianties door de quotiënten γ_1 , γ_{2j} en γ_{3j} bepaald zijn). Merk op dat hoewel in het model de τ_j 's, w_{jk} 's en e_{ijk} 's onderling onafhankelijk zijn de $\underline{Y}_{ijk}$'s in het geheel niet onafhankelijk zijn, maar een covariantiestructuur vertonen die gespecificeerd wordt door (6.2).

De dichtheid en tevens aannemelijkheidsfunctie van een realisatie Y van de stochast $\underline{Y}$ gedefiniëerd door (6.1) is dus:

$$L(Y) = (2\pi)^{-\frac{1}{2}npq} \sigma^{-npq} |H|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (Y-X\theta)' H^{-1} (Y-X\theta) \right\}$$

($|H|$ is de determinant van H). Stel $\lambda(Y) = \log L(Y)$. (Vector-)differentiatie van $\lambda(Y)$ naar θ en σ^2 geeft

$$\frac{\delta \lambda(Y)}{\delta \theta} = \frac{1}{\sigma^2} \{X' H^{-1} Y - (X' H^{-1} X) \theta\}$$

$$\frac{\delta \lambda(Y)}{\delta \sigma^2} = -\frac{npq}{2\sigma^2} + \frac{1}{2\sigma^4} (Y-X\theta)' H^{-1} (Y-X\theta).$$

Nul stellen van deze afgeleides geeft de meest aannemelijke schatters

$$(6.3) \quad \hat{\underline{\theta}} = (X' H^{-1} X)^{-1} X' H^{-1} \underline{Y}$$

en

$$(6.4) \quad \hat{\sigma}^2 = \frac{1}{npq} (\underline{Y} - \underline{X}\hat{\theta})' H^{-1} (\underline{Y} - \underline{X}\hat{\theta}) =$$

$$\frac{1}{npq} \{ \underline{Y}' H^{-1} \underline{Y} - (\underline{X}' H^{-1} \underline{Y})' (\underline{X}' H^{-1} \underline{X})^{-1} \underline{X}' H^{-1} \underline{Y} \}.$$

Als $\gamma_1, \gamma_{21}, \dots, \gamma_{2p}$ en $\gamma_{31}, \dots, \gamma_{3p}$ bekend zouden zijn, zouden $\hat{\theta}$ en $\hat{\sigma}^2$ gedefiniëerd in (6.3) en (6.4) tevens de *gewogen kleinste kwadratenschatters* zijn. Als schattingen voor de dieptes krijgen we

$$(6.5) \quad \hat{\theta}_i = \frac{1}{\gamma_3} \sum_{j=1}^p \frac{1}{\gamma_{3j}} (\underline{Y}_{ij} - \underline{Y}_{.j}) + \frac{1}{\delta_3} \sum_{j=1}^p \frac{1}{\delta_{3j}} \underline{Y}_{.j},$$

waarbij

$$\gamma_3 = \sum_{j=1}^p \frac{1}{\gamma_{3j}}, \quad \delta_{3j} = nq\gamma_1 + n\gamma_{2j} + \gamma_{3j}$$

en

$$\delta_3 = \sum_{j=1}^p \frac{1}{nq\gamma_1 + n\gamma_{2j} + \gamma_{3j}}.$$

Het onmiddellijke praktische gevolg van de weging met de gewichten γ_{3j} en δ_{3j} is, dat een schip met grote geschatte onnauwkeurigheid minder invloed heeft bij de schatting van de werkelijke diepte op een bepaald meetpunt. We gaan hier in 6.3 nader op in voor het schip "Vierlingh" bij de besprekingen van de uitkomsten van deze methode.

Uitwerking van (6.4) geeft

$$(6.6) \quad \hat{\sigma}^2 = \sum_{j=1}^p \left\{ \sum_{i=1}^n \sum_{k=1}^q \frac{1}{\gamma_{3j}} (\underline{Y}_{ijk} - \hat{\theta}_i)^2 - \frac{n^2}{n\gamma_{2j} + \gamma_{3j}} \sum_{k=1}^q (\underline{Y}_{.jk} - \hat{\mu})^2 \right.$$

$$\left. - \frac{(nq)^2 \gamma_1}{(n\gamma_{2j} + \gamma_{3j}) \delta_{3j}} (\underline{Y}_{.j} - \hat{\mu})^2 \right\}$$

waarbij $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_i$ de schatter van de gemiddelde diepte is.

Tenslotte vinden we de meest aannemelijke schatters van $\gamma_1, \gamma_{21}, \dots, \gamma_{2p}$ en $\gamma_{31}, \dots, \gamma_{3p}$ weer door de partiële afgeleiden van $\log L(Y)$ naar deze

parameters nul te stellen. Dit geeft de vergelijkingen

$$(6.7) \quad \text{spoor } (U_1' H^{-1} U_1) = \frac{1}{\sigma^2} (Y - X\theta)' H^{-1} U_1 U_1' H^{-1} (Y - X\theta)$$

$$(6.8) \quad \text{spoor } (U_{ij}' H^{-1} U_{ij}) = \frac{1}{\sigma^2} (Y - X\theta)' H^{-1} U_{ij} U_{ij}' H^{-1} (Y - X\theta)$$

$$i = 2, 3; j = 1, \dots, p$$

(spoor (A) = som van de diagonaalelementen van A).

We krijgen dus een stelsel (niet-lineaire) vergelijkingen met als onbekenden de parameters $\theta_1, \dots, \theta_n, \sigma^2, \gamma_1, \gamma_{21}, \dots, \gamma_{2p}$ en $\gamma_{31}, \dots, \gamma_{3p}$. Deze vergelijkingen zijn door middel van een iteratieve methode met als startwaarden $\gamma_1 = \gamma_{21} \dots \gamma_{2p} = \gamma_{31} \dots \gamma_{3p} = 1$ opgelost.* Deze methode wordt in principe beschreven in Hartley en Rao (1967), sectie 5 en Harville (1977), sectie 6.1, hoewel in beide artikelen een iets ander type model dan het onze gehanteerd wordt.

6.3 Resultaten van de methode van de grootste aannemelijkheid

De meest aannemelijke schattingen van de varianties F_j^2 , W_j^2 en T^2 , $j = 1, \dots, p$ worden gegeven in tabel 6 (in dm^2).

*) Wij zijn veel dank verschuldigd aan de heer R. van der Horst, die een computerprogramma voor deze berekeningen geschreven heeft.

Tabel 6 Schattingen van de varianties F_j^2 , W_j^2 en T^2

blok 1	Blauwe Slenk	Zwarte Haan	Breesem	Falga	A. Vierlingh
F_j^2	3,05	6,58	2,77	1,38	17,40
W_j^2	1,08	0,16	0,68	0,32	1,69
T^2	0,73				
blok 2	Blauwe Slenk	Zwarte Haan	Breesem	Falga	A. Vierlingh
F_j^2	1,09	1,76	2,91	2,25	4,16
W_j^2	0,0008	0,0004	0,18	0,001	1,74
T^2	0,77				
blok 3	Blauwe Slenk	Zwarte Haan	Breesem	Falga	A. Vierlingh
F_j^2	3,09	3,85	3,70	2,97	11,26
W_j^2	0,09	0,009	0,12	0,01	3,21
T^2	0,70				
blok 4	Blauwe Slenk	Zwarte Haan	Breesem	Falga	A. Vierlingh
F_j^2	0,77	2,17	1,92	1,09	3,73
W_j^2	0,17	0,002	0,05	0,06	0,92
T^2	0,58				

Het is duidelijk, dat het algemene beeld van tabel 4 wordt bevestigd. Als we allereerst kijken naar de belangrijkste schatting nl. T^2 , dan krijgen we voor de 4 blokken respectievelijk de schattingen 0,73, 0,77, 0,70 en 0,58, terwijl de schattingen volgens de momentenmethode respectievelijk 1,00, 0,60, 0,65 en 0,65 waren met als gewogen gemiddelde over de 4 blokken 0,73. De opmerkingen in paragraaf 6.1 over de grotere nauwkeurigheid van de meest aannemelijke schatters (t.o.v. schatters van de momentenmethode) lijkt hier dus enigzins geïllustreerd te worden: het verschil tussen de grootste en kleinste schatting is bij de momentenmethode gelijk aan 0,35 terwijl dit verschil bij de methode van de grootste aannemelijkheid slechts 0,19 is.

Als we vervolgens de meest aannemelijke schattingen van W_j^2 bekijken zien we dat het beeld van tabel 4 weer bevestigd wordt. In het bijzonder krijgen we steeds op de plaatsen waar in tabel 4 negatieve schattingen

voorkomen, schattingen die dicht bij nul liggen (eerste twee decimalen achter de komma zijn nullen).

Tenslotte vinden we bij de schattingen van F_j^2 de grootste afwijkingen van tabel 4. Als we bijvoorbeeld de schatting van F_5^2 in het eerste blok bekijken (schip A. Vierlingh), dan geeft de momentenmethode een schatting van $12,01 \text{ dm}^2$, maar de methode van de grootste aannemelijkheid een schatting van $17,40 \text{ dm}^2$. Dit verschil wordt veroorzaakt door het feit dat in de loop van de iteratieve berekening van de meest aannemelijke schattingen van de varianties de aanvankelijke schattingen van de dieptes $\theta_i = \mu + \xi_i$ veranderen. Het schip A. Vierlingh krijgt in de loop van het iteratieve proces steeds minder gewicht in de schatting van de dieptes (vanwege de hoge schattingen van de onnauwkeurigheid van dit schip). Dit heeft tot gevolg dat de afwijkingen van de A. Vierlingh in de loop van de berekening nog geaccntueerd worden.

Ter illustratie vergelijk we de resultaten van de eerste en de laatste (33^{e}) iteratie bij blok 1.

Tabel 7. Vergelijking van de schattingen van de varianties F_j^2 , W_j^2 en T^2 bij de 1^{e} en 33^{e} iteratie voor blok 1.

1^{e} iteratie	Blauwe Slenk	Zwarte Haan	Breesem	Falga	A. Vierlingh
F_j^2	3,74	5,90	2,82	2,57	13,96
W_j^2	1,90	1,43	1,66	1,31	3,06
T^2	1,63				
33^{e} iteratie	Blauwe Slenk	Zwarte Haan	Breesem	Falga	A. Vierlingh
F_j^2	3,05	6,58	2,77	1,38	17,40
W_j^2	1,08	0,16	0,68	0,32	1,69
T^2	0,73				

De schatting van F_5^2 is in de loop van het iteratieproces opgeklommen van 13,96 tot 17,40. Ook is in deze vergelijkende tabel te zien hoe de schattingen van de varianties, die aan het begin van het proces allemaal gelijk genomen werden, in de loop van het proces steeds meer gaan verschillen (na de 33^{e} iteratie werd gestopt omdat van de 32^{e} naar de 33^{e} iteratie de

schattingen gelijk gebleven waren in de 4^e decimaal; het criterium was: stoppen als de verschillen kleiner zijn dan $\frac{1}{2} \cdot 10^{-4}$).

We vergelijken vervolgens de schattingen van de dieptes bij de eerste en laatste iteratie voor blok 1.

Tabel 8. Schattingen van de dieptes bij de 1^e en 33^e iteratie voor blok 1

1 ^e iteratie	33 ^e iteratie
1 = 132,52	1 = 133,88
2 = 136,28	2 = 137,24
3 = 140,28	3 = 140,98
4 = 143,56	4 = 144,14
5 = 146,44	5 = 146,94
6 = 149,16	6 = 149,53
7 = 150,44	7 = 150,31
8 = 151,60	8 = 151,72
9 = 152,80	9 = 153,35
10 = 151,88	10 = 151,65
11 = 149,60	11 = 149,40
12 = 148,88	12 = 148,93
13 = 144,96	13 = 145,04
14 = 144,60	14 = 144,17
15 = 141,16	15 = 140,99
16 = 136,88	16 = 135,84
17 = 120,24	17 = 119,64
18 = 118,20	18 = 117,76
19 = 116,32	19 = 115,29
20 = 114,20	20 = 113,71
21 = 112,96	21 = 113,18
22 = 39,36	22 = 37,86
23 = 33,96	23 = 34,30
24 = 31,28	24 = 31,56
25 = 29,96	25 = 30,44
26 = 28,28	26 = 28,86

Uit de tabel blijkt dat er soms aanzienlijke wijzigingen optreden. Als we nu bijvoorbeeld nader bekijken wat er bij de meetpunten no. 1 en no. 22 aan de hand was, dan vinden we voor de A. Vierlingh hier respectievelijk de metingen 133, 120, 126, 127, 130 en 36, 61, 36, 35, 59. Het is duidelijk dat in de 33^e iteratie de lage waarden van de metingen van de A. Vierlingh bij meetpunt 1 en de zeer hoge waarden gemeten door dit schip bij meetpunt 22 minder invloed hebben op de schatting van de dieptes.

De methode levert dus tevens een soort detectiesysteem voor extreem uitvallende waarnemingen (zg. "uitbijters") op.

7. CONCLUSIES

7.1 Het *algemene beeld* dat uit beide analyses naar voren komt is, dat er duidelijke systematische verschillen tussen de schepen zijn. Deze vallen te onderscheiden in:

- a) verschil in nauwkeurigheid: in het bijzonder is de A. Vierlingh veel onnauwkeuriger te werk gegaan dan de andere schepen (zie tabel 4).
- b) verschillen in gemiddelde: hier springt duidelijk de Blauwe Slenk eruit (zie tabel 1,2 en 3).

7.2 De *overeenstemming tussen beide analysemethoden* is goed. Het lijkt, na de vergelijking tussen deze twee methoden geen twijfel, dat de tweede (grootste aannemelijkheid) - in de toekomst toe te passen zonder verdeling in blokken - superieur is. Voor dit doel is een computerprogramma in ontwikkeling, dat hier voor volledige blokken reeds is gebruikt.

7.3 De *grootste bron van onnauwkeurigheid* in het uiteindelijke resultaat zit, zoals in §2 reeds is opgemerkt, in de T^2 . Houden wij rekening met $T^2 = 0,7 \text{ dm}^2$, dan is de invloed hiervan op de gemiddelde gemeten diepte, onafhankelijk van het totale aantal meetpunten, bij gebruik van 5 schepen, die ieder evenveel metingen verrichten, gelijk aan $0,70/5 = 0,14 \text{ dm}^2$; dit is een standaardafwijking van $\sqrt{0,14} = 0,37 \text{ dm}$. Deze is te groot om vertrouwen te hebben in een gevonden verschil van 0,5 dm tussen twee series metingen. Nader onderzoek van het gevolg van het gedrag van de τ_j in de loop van de tijd is nodig om hierin (eventueel) verbetering te kunnen brengen.

7.4 *Persistentie van de τ_j* . Het huidige waarnemingsmateriaal strekt zich slechts uit over twee dagen. Indien, op grond van latere meervoudige waarnemingen zou blijken, dat de afwijking die nu voor de Blauwe Slenk gevonden is persisteert, dan moet het mogelijk zijn de oorzaak daarvan op te sporen. Indien de afwijking niet persistent is, is de oorzaak, die tot het nu geconstateerde verschil heeft geleid, wellicht niet meer te achterhalen.

Dit behoeft dan echter niet te gelden voor latere eventueel te vinden systematische verschillen. Voor de opsporing van de oorzaken dienen dan speciaal daartoe te ontwerpen experimenten te worden uitgevoerd. De mate van persistentie van de r_j 's is van essentiële betekenis voor de uiteindelijke berekening van de nauwkeurigheid en deze dient dus mede op grond van verdere duplobepalingen te worden onderzocht. Deze conclusie sluit geheel aan bij de beschouwingen op bladzijde 2 en 3 van het rapport "Foutenbeschouwing aanzandings en erosieonderzoek" van de Studiedienst Hoorn van Rijkswaterstaat.

7.5 *Verdere mogelijke punten van analyse* van het huidige waarnemingsmateriaal zijn:

- a) het nagaan van de invloed van "diep" en "ondiep", benevens de hier achterwege gelaten analyse van steile stukken.
- b) onderzoek naar systematische verschillen van viertallen en drietallen schepen na verwijdering van de Blauwe Slenk resp. A. Vierlingh.
- c) toetsing van $W_j^2 = 0$ voor sommige schepen.

Deze verdere analyse lijkt echter minder belangrijk dan voortzetting van het onderzoek op met nieuwe metingen uitgebreid waarnemingsmateriaal.

8. LITERATUUR

- 1 HARTLEY, H.O. & J.N.K. RAO (1967), *Maximum-likelihood estimation for the mixed analysis of variance model*, Biometrika 54, 93-108.
- 2 HARVILLE, D.A. (1977), *Maximum-likelihood Approaches to Variance Component Estimation and to Related Problems*, Journal of the American Statistical Association 72, 320-340. (Vermeldt veel literatuur).
- 3 JONGE DE H. (1963), *Inleiding tot de medische statistiek*, deel 1, (2^e druk, Nederlands Instituut voor Praeventieve Geneeskunde, Leiden).
- 4 PEARSON, E.S. & H.O. HARTLEY (1970), *Biometrika Tables for Statisticians*, vol. 1 (3rd. ed., Cambridge University Press, Cambridge).