

KONINKLIJK NEDERLANDS
METEOROLOGISCH INSTITUUT

De Bilt

Verslagen

V - 273

P.J. Rijkoort

De zeldzaamheid van het aantal dagen met grote
windsnelheden in de Waddenzee gedurende het tijd-
vak juni-september 1974

De Bilt, 1976

Publikationsnummer: K.N.M.I. V-273 (Stat.Bur.)

U.D.C.: 551.553.6 :
519.2

De zeldzaamheid van het aantal dagen met grote windsnelheden in de Waddenzee gedurende het tijdvak juni-september 1974.

Dr. P.J. Rijkoort

Inleiding.

Door een baggermaatschappij kon in de periode juni-september 1974 bij werkzaamheden in de Waddenzee op een vrij groot aantal dagen niet gewerkt worden wegens te hoge windsnelheden. De vraag is nu of het aantal dagen met grote windsnelheden zo groot was dat van onvoorziene omstandigheden kan worden gesproken.

Op grond van een vergelijking van waargenomen windsnelheden van het station Terschelling in de betrokken periode met een frequentie verdeling van uurwaarden van de windsnelheid van hetzelfde station over het tijdvak 1958-1967, werd door de baggermaatschappij geconcludeerd dat in 1974 inderdaad bijzonder veel hoge windsnelheden voorkwamen zodat i.v.m. de werkzaamheden van onvoorziene omstandigheden kan worden gesproken.

Door Rijkswaterstaat arrondissement Friesland West werd getracht een statistisch verantwoorde kansschatting te geven voor de zeldzaamheid van de waargenomen frequentie van grote windsnelheden in 1974. Er dient dan uiteraard nog te worden vastgesteld wat onder onvoorziene omstandigheden moet worden verstaan. Door RWS is voorgesteld dit vage begrip aan de 5% significante drempel te koppelen. De berekening van de overschrijdingskans van de waargenomen frequentie van grote windsnelheden in 1974 leverde volgens het rapport van RWS 6½% zodat dus niet van onvoorziene omstandigheden kan worden gesproken.

Aangezien de gebruikte methode om de overschrijdingskans te schatten een grote mate van onzekerheid bevat werd door RWS aan het K.N.M.I. verzocht na te gaan in hoeverre de gebruikte methode correct is en zo mogelijk te pogen met meer adequate gegevens het door RWS verkregen resultaat te verifiëren.

In het onderhavig rapport wordt dit probleem behandeld, waarbij in het eerste hoofdstuk de methode en het resultaat van de berekening in het kort worden samengevat, terwijl in de volgende hoofdstukken,

voor lezers die meer in detail willen weten op welke wijze het resultaat is verkregen, de gebruikte methode uitvoeriger wordt behandeld.

1. Samenvatting en resultaten.

Alvorens over te gaan tot de samenvatting van de in dit rapport toegepaste methode wordt eerst in het kort weergegeven op welke wijze de berekening volgens het RWS rapport is uitgevoerd.

Volgens opgave van de baggermaatschappij was, tijdens de 63 dagen die voor de werkzaamheden in aanmerking kwamen, 22% van de tijd onwerkbaar. In de frequentie verdeling van de windsnelheden over juni-september 1974 van Terschelling komt 22% overeen met 21 kts of meer. 10%. Uitgaande van deze 10% kans is vervolgens de kans berekend dat in 63 dagen gedurende 22% van de tijd de windsnelheid groter dan 21 kts is. Hierbij is nog rekening gehouden met het feit dat het voorkomen van windsnelheden boven een gegeven grens op opeenvolgende dagen niet onderling onafhankelijk is.

In principe is deze methode van berekening wel correct; het bezwaar is echter dat er een grote mate van onzekerheid bestaat in de nauwkeurigheid van de einduitkomst van de berekening, vooral in verband met de reductie methode om met de persistentie rekening te houden.

Hoewel 21 kts een geschikte drempelwaarde is om de onwerkbaarheid aan te geven is voor het K.N.M.I. onderzoek een ander criterium gebruikt (zie hfd. 2). Per dag is n.l. de maximum uurwaarde (u_M) van de windsnelheid tussen 7 en 18 uur bepaald. Voor u_M kan een drempelwaarde L worden bepaald zodanig dat voor werkbare dagen geldt $u_M < L$ en voor volledig onwerkbare dagen $u_M > L$. Voor de toepassing van dit criterium is de in het RWS rapport gebruikte reeks gegevens van Terschelling uitgebreid tot het tijdvak 1952 - 1973, terwijl daarnaast nog gegevens van Kornwerderzand, maar alleen over het tijdvak 1963 - 1973, zijn gebruikt.

De waarde van L blijkt nog iets te kunnen variëren n.l. voor Terschelling tussen 23 en 24 kts en voor Kornwerderzand 22 en 23 kts. Er moet hier nog even op worden gewezen dat het criterium niet volledig omkeerbaar is d.w.z.: zowel in geval $u_M < L$ geldt als wanneer $u_M > L$ juist

is kan de betrokken dag gedeeltelijk onwerkbaar zijn.

Het aantal dagen per jaar, beperkt tot juni-september, waarvoor $u_M > L$, genoteerd als Y_j is gebruikt als karakteristieke grootheid om de mate van onwerkbaarheid in de periode juni-september van het betrokken jaar aan te geven. De index j duidt daarbij het jaar aan.

Het probleem is nu van Y_{74} de overschrijdingskans aan te geven op basis van de Y_j gegevens van de jaren vóór 1974. De oplossing wordt gelukkig vergemakkelijkt doordat Y_j zich als een normaal verdeelde grootheid blijkt te gedragen. Van de Y_j waarden vóór 1974 worden gemiddelde Y en standaard deviatie s_y berekend en hiermede kan uit een tabel voor de normale verdeling de overschrijdingskans voor Y_{74} worden gevonden. Het resultaat van de berekeningen blijkt nu vrij sterk afhankelijk van de keuze van tijdvak en station (zie tabel 5); de overschrijdingskans P varieert tussen 1 en 20% en derhalve is het niet zonder meer mogelijk een uitspraak te doen ten aanzien van de vraag of voor 1974 van "onvoorziene omstandigheden" wat betreft dagen met grote windsnelheden kan worden gesproken d.w.z. of P al dan niet kleiner is dan 5%.

De grote variatie in de P waarden is natuurlijk het gevolg van het feit dat het aantal Y_j waarden waaruit Y en s_y is berekend zeer klein is. Men zou eigenlijk P moeten berekenen met behulp van μ en σ , de "ware" of universum waarden van Y en s_y . Omdat deze laatsten van μ en σ afwijken zal P ook van de "ware" P afwijken. Het is eigenlijk nodig om een marge te bepalen rond de berekende P zodanig dat deze marge met b.v. 95% betrouwbaarheid de ware P omvat. Men loopt echter het risico dat de marge ook de 5% omvat zodat dan nog geen uitspraak mogelijk is.

Het probleem kan echter ook op andere wijze worden geformuleerd n.l. als een beslissingsprobleem i.p.v. als een schattingsprobleem. Daartoe worden twee hypothesen ingevoerd n.l. $H_1: P < 5\%$ en $H_2: P > 5\%$ en er wordt vervolgens getracht een criterium te vinden op grond waarvan beslist kan worden welke hypothese kan worden aanvaard.

Voor de oplossing van het probleem wordt nu eerst de grootheid $k_0 = (Y_j - Y) / s_y$ ingevoerd. Er zijn dan twee waarden \tilde{k} en \check{k} te bepalen (zie hfd 4) met de eigenschap dat H_1 kan worden aanvaard als $k_0 > \tilde{k}$ en

H_2 als $k_0 < \check{k}$. Als geldt: $\check{k} < k_0 < \tilde{k}$ dan is er geen beslissing mogelijk. Deze uitspraken gelden met een gegeven mate van betrouwbaarheid (b.v. 95%).

De waarden \check{k} en \tilde{k} zijn alleen afhankelijk van het aantal Y_j 's waaruit Y en s_y zijn berekend. Het resultaat van de toepassing van deze methode is te vinden in tabel 6. Daaruit blijkt dat in geen der gevallen de hypothese $P < 5\%$ kan worden aanvaard. In twee gevallen kan wel de hypothese $P > 5\%$ worden aanvaard n.l. Terschelling 1952-'73 voor $u_M = 24$ en Kornwerderzand voor $u_M = 22$. In de overige gevallen is er geen uitspraak mogelijk. Uiteraard is dit laatste niet in strijd met het aanvaarden van de hypothese $P > 5\%$ in de eerder genoemde gevallen, zodat men tenslotte als eindresultaat kan stellen dat $P > 5\%$, mits men bedenkt dat deze uitspraak geen absolute uitspraak is maar slechts geldt behoudens een kleine kans op onjuist zijn. Voor een enkelvoudige toepassing van de methode was deze kans 0,05. Als twee onafhankelijke toepassingen elkaar bevestigen is deze kans uiteraard veel kleiner. Hoe het met deze kans gesteld is in de in dit rapport aanwezige situatie van zes min of meer afhankelijke gevallen is zeer moeilijk te bepalen maar het is wel plausibel dat de kans kleiner dan 5% is.

Volledigheidshalve is in hfd 5 het materiaal nog op een andere wijze bewerkt. De dagen met $u_M > L$ kunnen n.l. verdeeld worden in z.g. runs van diverse lengten l_r waarbij onder een run wordt verstaan een aantal opeenvolgende dagen met $u_M > L$ voorafgegaan en gevolgd door een dag met $u_M < L$. Als n_{1r} het aantal runs van lengte l_r is geldt: $Y = \sum l_r \cdot n_{1r}$. De verdelingen van de aantallen runs over de diverse lengten kan men theoretisch berekenen waarbij men desgewenst een zekere mate van persistentie in de opeenvolgende waarden van u_M kan invoeren. Dit laatste blijkt inderdaad nodig te zijn; theoretische verdelingen zonder persistentie passen in het geheel niet bij de waargenomen verdelingen (zie de tabellen 8).

Men kan ook trachten de verdeling van de runlengten te gebruiken om te onderzoeken of een bepaald jaar sterk afwijkt van de overigen. Dit blijkt voor 1974 niet het geval te zijn, hetgeen dus in feite in overeenstemming is met het reeds gevonden resultaat dat $P > 5\%$.

Tenslotte valt nog het volgende op te merken: De vrij grote verschillen die in de schattingen van P voorkomen kunnen ook samenhangen met het feit dat er in de waarden van u_M van jaar tot jaar vrij grote verschillen zijn, waardoor er in een bepaald tijdvak "toevallig" wel vrij weinig grote u_M voorkomen. Dit is blijkbaar in de periode 1958 - 1967 het geval. De berekeningen op basis van deze periode geven dan ook vrij kleine P waarden. Er is nog afzonderlijk getoetst of deze periode wat betreft de u_M waarden verschilt van de tijdvakken 1952 - 1957 en 1968 - 1973. Er blijkt geen significant verschil te zijn (zie hfd. 4)

2. criterium voor onwerkbare dagen wegens teveel wind.

Er is een lijst beschikbaar waarop, voor de werkdagen die in principe voor baggeractiviteiten in het Waddengebied in aanmerking kwamen, is aangegeven in hoeverre de betrokken dagen onwerkbaar waren wegens teveel wind. De onwerkbaarheid was uitgedrukt in eenheden van $\frac{1}{4}$ van het tijdvak 7 - 18 uur.

Tabel 1

Onwerkbare dagen in het Waddengebied wegens te veel wind.

Datum	fr	u	u_M	d_{21}
1 - 7 - 1974	1	26,0	30	11
2 - 7	$\frac{1}{2}$	13,5	18	0
3 - 7	$\frac{1}{2}$	26,8	40	11
4 - 7	1	19,4	24	8
5 - 7	$\frac{3}{4}$	20,6	28	7
11 - 7	1	23,5	25	10
29 - 7	$\frac{1}{2}$	20,4	22	9
31 - 7	$\frac{1}{4}$	19,2	26	8
1 - 8	$\frac{3}{4}$	17,0	24	5
12 - 8	1	22,5	31	9
13 - 8	$\frac{1}{4}$	7,2	9	0
3 - 9	1	34,2	38	12
4 - 9	$\frac{3}{4}$	26,3	34	12
5 - 9	$\frac{3}{4}$	22,7	34	7
6 - 9	1	19,9	40	6
23 - 9	1	31,4	33	12
24 - 9	$\frac{3}{4}$	19,9	40	6
25 - 9	$\frac{1}{2}$	15,8	23	2
26 - 9	$\frac{1}{2}$	12,6	14	0

De werkdagen waar het om gaat zijn 63 dagen in de perioden 17 juni - 11 juli en 29 juli - 26 september. In tabel 1 zijn de data genoteerd van de dagen die geheel of gedeeltelijk onwerkbaar waren, waarbij verder is aangegeven: de fractie onwerkbare tijd (fr), de gemiddelde windsnelheid (\bar{u}), de maximum waarde van u (u_M) en het aantal uren met snelheden ≥ 21 kts (d_{21}), allen betrekking hebbende op het tijdvak 7 - 18 uur. De grens van 21 kts is overgenomen uit het RWS rapport, terwijl gemiddelde en maximum snelheid daarnaast de meest voor de hand liggende grootheden zijn die in aanmerking komen om als criterium te dienen.

De windsnelheidsgegevens zijn eveneens voor de werkbare dagen bepaald maar hier niet vermeld.

In fig. 1^{a,b,c}, zijn voor de drie criteria stipendiagrammen geconstrueerd waarin de betrokken grootheid tegen fr is uitgezet. Het blijkt dat men geen van de criteria de onwerkbaarheid van de betrokken dag volledig en éénduidig is te bepalen. Wel zijn drempelwaarden aan te geven zodanig dat enerzijds alle werkbare dagen een windcriteriumwaarde beneden de drempelwaarde hebben en anderzijds alle volledig onwerkbare dagen een windcriteriumwaarde boven de drempelwaarde. Er is daarbij soms nog enige speling in de drempelwaarde; er zijn daarom steeds twee drempelwaarden gekozen waarmede tevens het effect van enige variatie in de drempelwaarde kan worden bekeken.

In tabel 2 is een overzicht gegeven van de aantallen onwerkbare resp. werkbare dagen in afhankelijkheid van onder- resp. overschrijding van de drempelwaarde.

Tabel 2

Onwerkbaar weer criterium, windsnelheid van Terschelling.

		aantal onwerkbare dagen als	aantal werkbaar dagen als		aantal onwerkbare dagen als	aantal werkbaar dagen als
w	L	w<L	w>L	L	w<L	w>L
\bar{u}	21	10	0	16	4	9
u_M	24	5	0	23	4	3
d_{21}	7	6	0	5	3	3

Op grond van deze resultaten kan men zonder meer vaststellen dat \bar{u} het minst geschikt is als criterium, terwijl u_M en d_{21} vrijwel even bruikbaar zijn.

Voor het onderhavige onderzoek zal de keuze op u_M worden bepaald en daarbij zullen de berekeningen zowel voor $L = 23$ kts als voor $L = 24$ kts worden uitgevoerd om na te gaan wat het effect is van zo'n klein verschil.

3. De aantallen dagen met $u_M \geq L$ op grond van de windsnelheid van Terschelling en Kornwerderzand.

De uitgangsggegevens voor de berekening zullen de aantallen dagen met grote windsnelheden over de periode juni - september per jaar zijn. In tabel 3^{a, b} zijn voor de jaren 1952 - 1974 ¹⁾ en voor het station Terschelling deze aantallen vermeld. In verband met mogelijke afhankelijkheid in het optreden van dagen met grote windsnelheden zijn niet alleen de totale aantallen dagen met $u_M \geq 23$ resp 24 kts gegeven (Y), maar ook de aantallen z.g. runs van opeenvolgende dagen met $u_M \geq 23$ resp. 24 kts; tevens is het totale aantal runs (X) vermeld.

Als men de waarde van Y voor 1974 met die van de jaren 1958-1967 vergelijkt dan is het inderdaad zo dat 1974 meer dagen met grote

¹⁾ 1961 is buiten beschouwing gelaten wegens ontbrekende waarnemingen.

windsnelheden heeft dan in de jaren 1958 - 1967 zijn voorgekomen n.l. 35 tegenover 26 als maximum in de 10 jaren resp. 28 tegenover 25. In de jaren voor 1958 resp. na 1967 komen echter wel aantallen voor die gelijk aan of zelfs groter dan die van 1974 zijn. Het tijdvak 1958 - 1967 is blijkbaar toevallig een tijdvak met vrij weinig grote windsnelheden.

Tabel 3^a

Terschelling; runlengten en aantallen dagen met $u_M \geq 23$ kts	
r	1952 53 54 55 56 57 58 59 60 62 63 64 65 66 67 68 69 70 71 72 73 74
1	11 9 6 9 5 8 12 4 7 7 11 13 11 6 2 6 8 8 9 11 5 14
2	4 6 4 - 7 4 3 2 3 4 4 2 1 - 1 2 1 8 3 - 1 3
3	2 4 5 - - 2 1 1 1 - 1 1 1 4 - 2 1 1 2 1 2 -
4	- - 1 - 1 - - - 4 - 1 - 2 - - - - 2 - - - 1
5	1 - - - - - 1 - 1 1 - 1 - - 1 1 - - - - 1
6	- - 1 - - - - 1 - 1 - - - - - - - - - 1
7	- - - - - 1 - - - - - - - - - - - - - -
8	- - - - - - - - - - - - - - - - 1 - - - - -
X	18 19 17 9 13 15 17 8 13 13 17 17 15 10 4 11 11 19 14 12 8 20
Y	30 33 39 9 23 29 26 17 25 26 26 25 24 18 9 21 21 35 21 14 13 35

Tabel 3^b

Terschelling; runlengten en aantallen dagen met $u_M \geq 24$ kts.	
r	1952 53 54 55 56 57 58 59 60 62 63 64 65 66 67 68 69 70 71 72 73 74
1	9 10 6 9 5 8 11 5 5 6 11 13 12 6 2 5 9 6 10 9 2 10
2	5 4 4 - 7 4 3 2 3 3 5 1 2 - 1 2 2 6 2 - 1 2
3	3 2 5 - - 2 1 2 2 - 1 1 1 4 - 2 2 3 1 1 - -
4	- - 1 - 1 - - - - - - - 1 - - - - - - - 2
5	- 1 - - - - 1 - 1 1 - 1 - - 1 1 - - - - -
6	- - 1 - - - - - - 1 - - - - - - - - - -
7	- - - - - 1 - - - - - - - - - - - - - - -
X	17 17 17 9 13 15 16 9 11 11 17 16 16 10 4 10 13 15 13 10 3 15
Y	28 29 39 9 23 29 25 15 22 23 24 23 23 18 9 20 19 27 17 12 4 28

Om een statistische beoordeling van deze getallen te kunnen geven, is de gemakkelijkste oplossing te vinden als men er van zou kunnen uitgaan dat de waarden Y een onderling onafhankelijke reeks normaal verdeelde grootheden vormen. Dit blijkt inderdaad een aanvaardbare hypothese te zijn. In de eerste plaats bleken de 21 waarden Y over de periode 1952 - 1973 autocorrelatiecoëfficiënten $+0,15$ resp. $+0,14$ (voor $Y \geq 23$ resp. 24 kts) te bezitten. Dit is bij 19 paren (19 wegens het ontbreken van 1961) niet significant van 0 afwijkend. De waarden van Y kunnen dus als onderling onafhankelijke grootheden worden beschouwd. Ten tweede levert toepassing van de toets van Shapiro-Wilks (zie o.a. Kruizinga W.R. 74-7) als uitspraak dat er geen reden is de hypothese dat Y normaal verdeeld is te verwerpen. In het geval van 23 kts is de overschrijdingskans van de Shapiro-Wilks toetsingsgrootte zelfs zeer groot (85%); in het geval van 24 kts veel kleiner n.l. 8%, maar dit is toch altijd nog voldoende ver boven de 5% grens om voor beide gevallen de normaliteitshypothese aanvaardbaar te achten. ¹⁾ In de figuren 2a en 2b zijn op normaal waarschijnlijkheidspapier de numerieke cumulatieve verdelingen van Y uitgezet. Tevens zijn de aangepaste rechten en een 95% betrouwbaarheidsband aangegeven. De aanpassing van de rechte is verkregen op basis van gemiddelde en standaard-deviatie. De 95% betrouwbaarheidsband werd bepaald op basis van de normale benadering voor de verdeling van gemiddelde en standaard-deviatie (zie b.v. Dey W.R. 72-9).

Op grond van de berekende waarden van \bar{Y} en s_Y voor een gegeven steekproef van een aantal jaren kan voor Y van 1974 (Y_{74}) de overschrijdingskans worden bepaald. Hiertoe wordt $k = (Y_{74} - \bar{Y})/s_Y$ berekend, waarna voor deze k in een tabel van de standaard normale verdeling de overschrijdingskans kan worden opgezocht.

De berekening is uitgevoerd met \bar{Y} en s_Y van de volledige reeks 1952 - 1973, van de deelreeks 1958 - 1967 ter vergelijking met het resultaat van RWS en van de deelreeks 1963 - 1973 ter vergelijking met een reeks van Kornwerderzand die slechts over die periode beschikbaar was. Deze laatste reeks is bij het onderzoek betrokken teneinde bezwaren van een te eenzijdige keuze van een station te ontgaan. Analooq aan hetgeen voor Terschelling is gedaan, is in fig. 3 de relatie van de windsnelheid (beperkt tot u_M) van Kornwerderzand met de onwerkbare dagen

1) Dit grote verschil in overschrijdingskans is het gevolg van het feit dat dat in 1974 "toevallig" vrij veel waarden $u_M = 23$ voorkwamen, terwijl dit aantal veel kleiner was.

weergegeven. Het blijkt dat de limietwaarden hier op 22 of 23 kts gesteld kunnen worden. De tabellen 4a, b geven de runs en de totale aantallen dagen met $u_M \geq 22$ resp. 23 kts.

Tabel 4a

Kornwerderzand; runlengten en aantallen dagen met $u_M \geq 22$ kts.

r	1963	'64	'65	'66	'67	'68	'69	'70	'71	'72	'73	'74
1	12	13	10	5	8	7	7	7	5	7	9	9
2	4	3	4	-	1	2	2	5	1	5	-	4
3	-	-	1	3	-	3	1	-	4	1	2	-
4	-	-	3	-	1	-	-	2	-	-	1	1
5	-	2	-	-	1	1	-	-	-	-	-	1
6	-	-	-	-	-	-	-	-	1	-	-	1
7	-	-	-	-	-	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-	-	-	-	-
9	-	-	-	-	-	1	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	1	-	-	-
X	16	18	18	18	11	14	10	14	12	13	12	16
Y	20	29	33	24	19	34	14	25	35	20	19	32

Tabel 4b

Kornwerderzand; runlengten en aantallen dagen met $u_M \geq 23$ kts.

r	1963	'64	'65	'66	'67	'68	'69	'70	'71	'72	'73	'74
1	9	9	8	14	7	5	6	7	4	12	7	8
2	3	-	6	2	2	2	2	4	2	1	1	4
3	-	-	1	-	-	3	-	1	2	-	2	-
4	-	1	1	-	1	-	-	1	-	-	-	1
5	-	1	-	-	-	1	-	-	-	-	-	1
6	-	-	-	-	-	-	-	-	1	-	-	1
7	-	-	-	-	-	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-	-	-	-	-
9	-	-	-	-	-	1	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	1	-	-	-
X	12	11	16	16	10	12	8	13	10	13	10	15
Y	15	18	27	18	15	32	10	22	30	14	15	31

In tabel 5 zijn vervolgens voor de diverse combinaties van periode en limietwaarde de overschrijdingskansen P samengevat. Tevens werden \bar{Y} , s_Y en k vermeld.

Tabel 5

Overschrijdingskansen P(%) voor Y_{74}

Terschelling

	$u_M \geq 24$ kts $Y_{74} = 28$				$u \geq 23$ kts $Y_{74} = 35$			
	\bar{Y}	s_Y	k	P	\bar{Y}	s_Y	k	P
1958-'67	19,9	5,16	1,57	5,8	21,8	5,90	2,24	1,2
1952-'73	20,9	8,07	0,88	19,9	23,0	7,95	1,51	6,6
1963-'73	17,8	6,97	1,46	7,2	20,6	7,15	2,01	2,2
Kornwerderzand								
	$u_M \geq 23$ kts $Y_{74} = 31$				$u_M \geq 22$ kts $Y_{74} = 32$			
	\bar{Y}	s_Y	k	P	\bar{Y}	s_Y	k	P
1963-'73	19,6	7,18	1,58	5,7	24,7	7,10	1,03	15,0

Uit deze tabel blijkt dat er vrij grote verschillen bestaan tussen de schattingen van de gezochte kans, zowel tussen die voor de diverse tijdvakken, als bij variatie van de limietwaarde voor u_M . Daarbij is het zo dat voor Terschelling de kansen kleiner zijn bij kleinere limietwaarde, terwijl voor Kornwerderzand het juist andersom is. Dit wijst er eigenlijk al wel op dat de onnauwkeurigheid van de schattingen vrij groot moet zijn zodat de verschillen mogelijk niet significant zijn; hierop zal in de volgende paragraaf nader worden ingegaan. Dat bij Terschelling voor alle drie perioden P voor $u_M \geq 23$ kleiner is dan voor $u_M \geq 24$ zegt in dit opzicht natuurlijk niets omdat de perioden niet onderling onafhankelijk zijn.

Op te merken valt nog dat de door RWS voor 1958 - 1967 berekende kans van 6,5% vrij aardig overeenkomt met de P waarde voor $u_M \geq 24$ voor dezelfde periode.

4. De onnauwkeurigheid van de kansschattingen.

In het voorgaande is gebleken dat de schattingen voor de overschrijdingskans P van het in 1974 waargenomen aantal dagen met grote windsnelheden Y_{74} sterk uiteenlopen. Dit zal wel hoofdzakelijk een gevolg zijn van het feit dat de reeksen waarop de berekening berust zeer klein zijn, zodat de berekende waarden \bar{Y} en s_Y (hier verder korthedshalve met m_0 en s_0 aangegeven) zeer onnauwkeurige schattingen zijn van de "ware" waarden μ en σ van het universum van alle mogelijke Y waarden. Men zou de waarde willen kennen van de overschrijdingskans van Y_{74} in het (μ, σ) universum of wel van $K = (Y_{74} - \mu) / \sigma$ in een standaard normale verdeling, maar men heeft echter alleen de kans P_{k_0} waarin $k_0 = (Y_{74} - m) / s$, in plaats van de gewenste kans P_K .

Nu zou men kunnen onderstellen dat het mogelijk moet zijn een marge rond P_{k_0} aan te geven zodat met een zekere betrouwbaarheid gesteld kan worden dat deze marge P_K omvat. In principe kan dat wel maar het bezwaar van een dergelijke marge is echter dat men niet altijd de vraag kan beantwoorden of de gezochte kans kleiner (of groter) dan een bepaalde limietkans ξ is, n.l. als deze limietkanswaarde ook in de marge ligt.

Het is daarom beter het probleem te formuleren als een beslissingsprobleem waarbij het er om gaat of een keuze gemaakt kan worden tussen de hypothesen $H_1 : P < \epsilon$ en $H_2 : P > \epsilon$.

Zo gesteld is het probleem op te lossen (zie appendix). Het resultaat is, in algemene vorm, dat men twee k-waarden verkrijgt n.l. \tilde{k}_ϵ en \tilde{k}_ϵ waarvoor geldt dat als $k_0 > \tilde{k}_\epsilon$ men de hypothese H_1 kan aanvaarden en als $k_0 < \tilde{k}_\epsilon$ de hypothese H_2 acceptabel is. Beide keuzen met een risico van hoogstens $(1-\gamma)$ kans op een onjuiste beslissing. Als geldt: $\tilde{k}_\epsilon < k_0 < \tilde{k}_\epsilon$ is geen beslissing mogelijk, althans niet met hoogstens $1-\gamma$ kans op een foute beslissing.

Voor toepassing in dit verslag is voor $\gamma=0,95$ gekozen en voor $\epsilon=0,05$. In tabel 6 zijn, voor de verschillende reeksen en criteria, de k-waarden uit tabel 5 overgenomen, daarnaast zijn \tilde{k}_ϵ en \tilde{k}_ϵ vermeld.

Tabel 6

Vertrouwenslimieten voor k.

		Terschelling		Kornwenderzand		$\tilde{k}_{5\%}$	$\tilde{k}_{5\%}$
		L=24	L=23	L=23	L=22		
1958-'67	9	1,57	2,24			3,032	0,990
1952-'73	21	<u>0,88</u>	1,51			2,371	1,185
1963-'73	11	1,46	2,01	1,58	<u>1,03</u>	2,815	1,042

Het blijkt dus dat in geen der gevallen de hypothese $P < 5\%$ kan worden aanvaard (op basis van 95% betrouwbaarheid). In twee gevallen (onderstreept) kan men wel de hypothese $P > 5\%$ aanvaarden. In de overige gevallen is er geen uitspraak mogelijk, maar dit is niet in strijd met de hypothese $P > 5\%$, zodat tenslotte deze laatste hypothese zonder bezwaar kan worden aanvaard.

Het feit dat tenslotte één hypothese voor alle reeksen is te aanvaarden, kan er op wijzen dat de jaren 1958 - 1967, hoewel dit een tijdvak met vrij weinig grote windsnelheden is, toch met de jaren er voor en er na een homogene reeks vormen. De resultaten van tabel 6 zijn natuurlijk in statistische zin niet geschikt om deze conclusie te recht-

vaardigen mede omdat de reeksen elkaar overlappen.

Een eenvoudige methode hiervoor is de parameter vrije toets van Kruskal en Wallis (zie de Jonge 1963 of Conover 1971). Deze zal hier alleen op 21 Y waarden van tabel 3_a worden toegepast. De periode 1952 - '73 wordt verdeeld in de drie tijdvakken 1952 - '57, 1958 - '67 (zonder 1961) en 1968 - '73 waarvan de lengten dus resp. zijn 6, 9 en 6 jaren. Aan de 21 Y waarden worden de rangnummers 1 - 21 toegekend volgens hun afnemende grootte. Voor de drie groepen worden de sommen van de rangnummers in die groepen bepaald. Onder de nulhypothese dat de drie groepen onderling onafhankelijke steekproeven uit een zelfde universum zijn, is de verwachtingswaarde van de rangnummersommen $\frac{1}{2}m_i(n+1)$ waarin m_i het aantal elementen in de eerste groep is en n het totale aantal elementen. Naarmate de waargenomen rangnummersommen sterker van deze theoretische waarden afwijken zal er meer niveau verschil zijn tussen de verschillende groepen. Als toetsingsgrootheid is door Kruskal en Wallis ingevoerd: $H = (12/n(n+1))(\sum s_i^2/m_i - 3(n+1))$ Deze H is bij benadering verdeeld als χ^2 met $v = k-1$ vrijheidsgraden waarbij k het aantal groepen is. In bovenstaand geval blijkt H de waarde 2,88 te verkrijgen terwijl $v = 2$, waarbij een overschrijdingskans 0,22 behoort. Er is dus geen reden om de nulhypothese dat de drie steekproeven uit één universum komen te verwerpen.

5. Runs van dagen met grote windsnelheden.

In de tabellen 3 en 4 zijn naast de totalen aantallen dagen met $u_M \geq 22, 23$ of 24 kts ook de aantallen z.g. "runs" van verschillende lengten vermeld.

Een statistisch verantwoorde beoordeling van de vraag of 1974 in aantallen runs van diverse lengten afwijkt van de vroegere jaren is moeilijk te geven. Een globale beschouwing geeft echter niet de indruk dat 1974 wat betreft de aantallen runs en de verdeling daarvan erg bijzonder is. De grootste runlengte (lr) die voorkomt is 6 (in alle vier gevallen). Deze zelfde waarde en nog grotere komen in andere jaren ook voor. Bij $u_M \geq 23$ kts voor Terschelling b.v. in 5 van de 21 jaren voor 1974. Het aantal "losse" dagen in 1974 is alleen in het ge-

Het blijkt dat er alleen voor Terschelling $u_M \geq 23$ en de periode 1958-'67 een significante afwijking is tussen de waarde voor 1974 en de "theoretische" waarde.

Voor de toepassing van de tweede methode wordt eerst het runlengte kansmodel besproken.

Men kan modellen maken met en zonder persistentie. Een model zonder persistentie is zeer eenvoudig; in dit geval is er alleen sprake van de kans p dat een willekeurige dag aan het criterium voldoet (genoteerd als: C-dag). Deze kans is te schatten als het quotient van het totale aantal C-dagen (Y uit de tabellen 3 en 4) en het totale aantal dagen N . Dus: $p = Y/N$. De kans op een run van lengte k is de kans op k C-dagen achter voorafgegaan en gevolgd door een niet C-dag dus: $(1-p)p^k(1-p)$. Het aantal runs van lengte k is derhalve:

$$x_k = N(1-p)^2 p^k = N(1-Y/N)^2 (Y/N)^k = (N-Y)^2 Y^2 / N^{k+1}.$$

Het totaal aantal runs is dus:

$$X = N(1-p)^2 \sum p^k = N(1-p)^2 p / (1-p) = Np(1-p) = Y(N-Y)/N.$$

Het model met persistentie is iets ingewikkelder. Naast de kans dat een willekeurige dag geen C-dag is n.l. $1-p$, moet voor het geval dat een bepaalde dag wel een C-dag is onderscheid worden gemaakt tussen de kans p_0 dat een dag een C-dag is als de vorige dag geen C-dag is en de kans p_1 dat een dag een C-dag is als de vorige dag wel een C-dag is. Een verdere onderscheiding in kansen al naar het aantal voorafgaande dagen dat ook C-dagen zijn is mogelijk maar zal hier niet worden beschouwd.

De kans p_0 is te schatten als het aantal runs gedeeld door het aantal niet C-dagen, immers de eerste dagen van de runs zijn de enige C-dagen die door niet C-dagen worden voorafgegaan.

$$\text{Dus: } p_0 = X/N(N-Y)$$

Verder zijn de C-dagen die door een C-dag worden voorafgegaan alle dagen van de runs behalve de eerste dag van iedere run zodat:

$$p_1 = X^2(Y-X)^{k-1} Y^{-k}.$$

De kans op een runs van de lengte k is nu: $(1-p)p_0 p_1^{k-1}(1-p_1)$ en het aantal runs van lengte k dus: $x_k = N(1-p)p_0 p_1^{k-1}(1-p_1) = X^2(Y-X)^{k-1} Y^{-k}$.

Er moet in dit geval gelden: $\sum x_k = X$, hetgeen als men het doorrekent inderdaad blijkt te kloppen.

In tabel 8 zijn voor de diverse perioden en criteria de verdelingen van de runlengten gegeven met de aantallen volgens de theoretische verdelingen volgens beide methoden (I; zonder persistentie en II: met persistentie).

Tabel 8^a
Terschelling $u_M \geq 23$

	1958 - '67			1952 - '73			1963 - '73		
	W	I	II	W	I	II	W	I	II
1	73	132,3	66,3	168	318,4	162,0	90	156,7	83,9
2	20	23,6	27,7	60	60,2	68,3	23	26,5	32,9
3	10	4,2	11,6	32	11,4	28,8	16	4,48	12,9
4	4	0,75	4,85	8	2,15	12,1	5	0,76	5,06
5	5	0,13	2,03	7	0,41	5,11	3	0,13	1,98
6	2	0,02	0,85	3	0,08	2,15	-	0,022	0,78
7	-	-	0,36	1	0,01	0,91	-	0,0036	0,30
8	-	-	0,15	1	0,002	0,38	1	0,0006	0,12
X	114			280			138		
Y	196			484			227		
P		0,178			0,189			0,169	
P ₀			0,126			0,135			0,124
P ₁			0,418			0,421			0,392
P(χ^2)			0,05			0,11			0,04

Tabel 8b
Terschelling $u_M \geq 24$

	1958 - '67			1952 - '73			1963 - '73		
	W	I	II	W	I	II	W	I	II
1	71	126,7	66,5	159	301,0	156,7	85	142,9	82,3
2	20	21,0	26,3	57	51,5	63,0	22	20,9	29,0
3	12	3,48	10,4	33	8,79	25,3	16	3,05	10,2
4	1	0,58	4,12	3	1,50	10,2	1	0,46	3,59
5	5	0,10	1,63	7	0,26	4,09	3	0,065	1,26
6	1	0,015	0,64	2	0,044	1,64	-	0,009	0,44
7	-		0,255	1	0,008	0,66	-		0,17
8	-		0,101	-		0,27	-		0,055
X	110			262			127		
Y	182			438			196		
P		0,166			0,171			0,146	
P ₀			0,120			0,129			0,111
P ₁			0,396			0,402			0,352
P(χ^2)			0,17			0,05			0,02

Tabel 8c
Kornwerderzand 1963 - '73

	$u_M \geq 22$			$u_M \geq 23$		
	W	I	II	W	I	II
1	100	172,9	89,5	88	152,1	79,4
2	27	35,0	38,2	25	24,5	31,3
3	15	7,10	16,3	9	3,94	12,3
4	7	1,44	6,94	4	0,63	4,84
5	4	0,29	2,96	2	0,10	1,95
6	1	0,059	1,26	1	0,016	0,75
7	-	0,012	0,54	-	0,0026	0,30
8	-	0,002	0,230	-	0,0004	0,116
9	1	0,0004	0,098	1	0,00006	0,046
10	1	0,0001	0,042	1	0,00001	0,018
X	156			131		
Y	272			216		
P		0,203			0,161	
P ₀			0,146			0,116
P ₁			0,426			0,394
P(χ^2)			0,03			0,07

Het is zonder meer duidelijk dat het model zonder persistentie niet voldoet; het model met eerste orde persistentie daarentegen geeft een vrij redelijke overeenstemming, hoewel de χ^2 toetsing (de overschrijdingskansen zijn in de tabel vermeld) in 3 van de 8 gevallen een kleine significante afwijking geeft.

Uitgaande van de gevonden waarden van p , p_0 en p_1 kan men nu voor een willekeurig jaar de "theoretische" aantallen runs van diverse lengten berekenen. De resultaten van 1974 kunnen hiermede worden vergeleken. Beperking tot twee klassen is weer nodig. Tabel 9 geeft, analoog aan tabel 7, de resultaten van de χ^2 toetsing.

Tabel 9

		Terschelling						Kornwerderzand				
		$u_M \geq 23$			$u_M \geq 24$			$u_M \geq 22$		$u_M \geq 23$		
	1974	'58 '67	'52 '73	'63 '73	'74	'58 '67	'52 '73	'63 '73	'74	'63 '73	'74	'63 '73
lr=1	14	7,35	7,73	7,64	10	7,37	7,80	7,49	9	8,15	8	7,20
lr=2	6	5,29	5,63	4,93	5	4,84	5,25	4,07	7	6,05	7	4,67
χ^2		6,11	5,11	5,53		0,94	0,63	1,05		0,24		1,25
P		0,04	0,07	0,06		>0,6	>0,6	0,52		>0,6		0,5

Deze resultaten kloppen vrijwel volledig met die van tabel 7; ook nu is de periode 1958/67 voor het geval $u \geq 23$ van Terschelling het enige dat een significante afwijking voor 1974 geeft, met dien verstande dat volgens tabel 9 de significante veel minder sterk is dan volgens tabel 8.

Appendix

Gegeven: x is een grootheid die aan een normale verdeling $N(\mu, \sigma)$ met gemiddelde μ en standaard deviatie σ voldoet. De distributieve resp. cumulative vorm van de verdelingsfunctie (d.f. resp. c.f.) wordt als $f_{\mu, \sigma}(x)$ resp $F_{\mu, \sigma}(x) = \int_{-\infty}^x f_{\mu, \sigma}(\xi) d\xi$ genoteerd.

Probleem: Als een steekproef $S_n(m_0, s_0)$ uit $N(\mu, \sigma)$ is gegeven waarin het aantal elementen van S is n en m_0 en s_0 resp gemiddelde en standaard deviatie van S zijn, en verder nog een element x^* uit $N(\mu, \sigma)$ kan dan een uitspraak gedaan worden over de vraag of voor de overschrijdingskans van $P(x^*)$ van x^* geldt: $P(x^*) < \epsilon$?

Oplossing: Er is een x , die met x_ϵ wordt aangegeven, zodanig dat $P(x_\epsilon) = 1 - F_{\mu, \sigma}(x_\epsilon) = \epsilon$ (zie fig. 4.1). Deze x_ϵ is ook te schrijven als $\mu + \kappa_\epsilon \sigma$. De grootheid $\kappa = (x - \mu) / \sigma$ volgt een standaard normale verdeling $N(0, 1)$ met d.f. $f_{0, 1}(\kappa)$ en c.f. $F_{0, 1}(\kappa)$ kortheidshalve $f(\kappa)$ en $F(\kappa)$ (zie fig. 2.1).

Uit het universum $N(\mu, \sigma)$ kan een universum van steekproeven $S_n(m_i, s_i)$ worden samengesteld. De steekproef $S_n(m_0, s_0)$ is een exemplaar uit dit universum (zie fig. 4.2).

Er wordt nu een grootheid k ingevoerd. Als k constant is vormen de grootheden $x_i = m_i + k s_i$ een universum met d.f. en c.f. die resp met $f_k(x)$ en $F_k(x)$ zullen worden aangegeven.

Twee waarden \tilde{k}_ϵ en \check{k}_ϵ worden nu bepaald door de eisen:

$$P(m_i + \tilde{k}_\epsilon s_i \geq \mu + \kappa_\epsilon \sigma) = \delta$$

en

$$P(m_i + \check{k}_\epsilon s_i \leq \mu + \kappa_\epsilon \sigma) = \gamma$$

waarbij γ vrij groot b.v. 0,95.

Voor deze eisen is ook te schrijven:

$$P(\tilde{k}_\epsilon \sqrt{n} \geq ((\mu - m_i) \sqrt{n} + \delta \sigma) / s_i) = \delta$$

en

$$P(\check{k}_\epsilon \sqrt{n} < ((\mu - m_i) \sqrt{n} + \delta \sigma) / s_i) = \gamma \quad \text{met } \delta = \kappa_\epsilon \sqrt{n}$$

De grootheid $((\mu - m_i) \sqrt{n} + \delta \sigma) / s_i$ wordt met t' (δ) aangeduid.

De verdelingsfunctie hiervan is theoretisch te berekenen; het is de z.g. niet-centrale t-verdeling met δ als parameter en $\nu = n-1$ vrijheidsgraden. Voor $\delta = 0$ gaat de verdeling in de bekende Student verdeling over. De verdeling is voor een beperkt aantal waarden

van δ en n getabelleerd. Het is mogelijk om bij gegeven n , δ en γ de waarden \tilde{k}_ϵ en \check{k}_ϵ te vinden, hetzij rechtstreeks hetzij door interpolatie.

In fig. 4.3 en 4.4 zijn de functies $f_{\tilde{k}_\epsilon}(x)$ resp $f_{\check{k}_\epsilon}(x)$ in beeld gebracht.

Het voorgaande kan nu als volgt worden geïnterpreteerd:

Neemt men een willekeurige steekproef $S(m_i, s_i)$ en berekend men de waarde $x_i = m_i + \tilde{k}_\epsilon s_i$ dan is er 100% kans dat de overschrijdingkans $P(x_i)$ van x_i kleiner is dan ϵ . Evenzo geldt: $P(x_j) > \epsilon$ met 100% waarschijnlijkheid, als $x_j = m_i + \check{k}_\epsilon s_i$.

De gegeven waarde x^* (b.v. x_p^* resp x_q^* in fig. 4.1) correspondeert met een κ^* (κ_p^* resp. κ_q^* in fig. 5.1). In feite is het zo dat in fig. 4.1 de ligging van de d.f. $\mathcal{Y}_{\kappa, \sigma}(x)$ niet bekend is maar de positie van x^* wel, terwijl in fig. 5.1 de ligging van de d.f. $\mathcal{Y}_{0,1}(x)$ wel bekend is maar de positie van κ^* niet.

Uit x^* , m_0 en s_0 wordt $k_0 = (x^* - m_0)/s_0$ berekend. Deze k_0 is een exemplaar uit een universum van $k_i = (x^* - m_i)/s_i$ waarden. De k_i 's voldoen aan een verdelingsfunctie waarvan de cumulative vorm als $G_{x^*}(k)$ wordt genoteerd. Dus:

$$G_{x^*}(k) = P((x^* - m_i)/s_i < k)$$

In fig. 5.2 zijn een aantal van de cumulative verdelingen $G_{x^*}(k)$ getekend. Een bijzonder geval is $G_{x_\epsilon}(k)$; hiervoor geldt:

$$G_{x_\epsilon}(\tilde{k}_\epsilon) = P((x_\epsilon - m_i)/s_i < \tilde{k}_\epsilon) = P(m_i + \tilde{k}_\epsilon s_i \geq x_\epsilon) = \delta$$

en

$$G_{x_\epsilon}(\check{k}_\epsilon) = P((x_\epsilon - m_i)/s_i < \check{k}_\epsilon) = P(m_i + \check{k}_\epsilon s_i \geq x_\epsilon) =$$

$$1 - P(m_i + \check{k}_\epsilon s_i < x_\epsilon) = 1 - \delta$$

Deze verdeling is in fig. 2.2 getekend, tevens zijn op de k -as de waarden \tilde{k}_ϵ , κ_ϵ en \check{k}_ϵ aangegeven. Voor de functies $G_{x^*}(k)$ geldt nu dat ze rechts van $G_{x_\epsilon}(k)$ liggen als $x^* > x_\epsilon$ en links van $G_{x_\epsilon}(k)$ als $x^* < x_\epsilon$.

Ten aanzien van de overschrijdingskans van k_0 is nu uit fig. 5.2 het volgende schema af te leiden:

	$x^* < x_\epsilon$	$x^* > x_\epsilon$
$\tilde{k}_\epsilon < k$	$P(k_0) < 1 - \delta$	$0 \dots 1$
$\check{k}_\epsilon < k_0 < \tilde{k}_\epsilon$	$P(k_0) < \delta$	$P(k_0) > 1 - \delta$
$k_0 < \check{k}_\epsilon$	$0 \dots 1$	$P(k_0) > \delta$

Op grond van dit schema kan nu voor de keuze tussen de hypothesen $x^* < x_\epsilon$ en $x^* > x_\epsilon$ een beslissingsregel worden vastgesteld n.l.: Als $k_0 > \tilde{k}_\epsilon$

aanvaardt men de hypothese $x^* < x_e$ en als $k < \check{k}_e$ de hypothese $x^* > x_e$.
In het geval $\check{k}_e < k_0 < \tilde{k}_e$ is geen beslissing mogelijk.

De vraag is nu tenslotte: in hoeverre kan de aanvaarde hypothese onjuist zijn? Dus: Hoe groot zijn $P(k > \tilde{k}_e | x^* < x_e)$ en $P(k < \check{k}_e | x^* > x_e)$? Dit volgt direct uit fig. 5.2 n.l.:

$$P(k > \tilde{k}_e | x^* < x_e) < 1 - \gamma$$

evenzo: $P(k < \check{k}_e | x^* > x_e) < 1 - \gamma$

Men kan dus stellen: Op grond van de voorgestelde regel kan een keuze tussen de hypothesen $x^* < x_e$ gemaakt worden waarbij men hoogstens $1 - \gamma$ kans heeft op een foute uitspraak. Uiteraard is het daarbij zo dat het risico op een foute beslissing kleiner is naarmate het verschil tussen k_0 en \tilde{k}_e resp k_0 en \check{k}_e groter is.

In fig. 6 tenslotte het verloop van \tilde{k}_e en \check{k}_e als functie van n getekend voor 1,5 en 10%.

Literatuur

- C.E.P. Brooks and N. Carruthers: Handbook of Statistical Methods in
Meteorology. Her. Maj. Off. London 1953
- Dr. L.J.L. Dey: Over de zeldzaamheid van het optreden
van een droge winter gevolgd door een
droge lente, en van een droge herfst +
winter gevolgd door een droge lente +
zomer. K.N.M.I. W.R. 72-9
- W.J. Conover: Practical Nonparametric Statistics.
Wiley, London 1971
- N.L. Johnson and S. Kotz: Distributions in Statistics.
Continuous univariate distributions III
Houghton Mifflin Co 1970
- H. de Jonge: Inleiding tot de Medische Statistiek I
Inst. voor Prev. Geneesk. Leiden
1963
- Drs. S. Kruizinga: Aanpassingstoetsen K.N.M.I. W.R. 74-7
- D.B. Owen: Handbook of Statistical Tables.
Addison Wesley - Mass. U.S.A. 1962
- F. Ynsen ing.tha: Onderzoek naar abnormale windsnelheden in
het tijdvak Juni t/m september 1974.
Rijkswaterstaat Arr. Friesl. W.
rapport 75.3

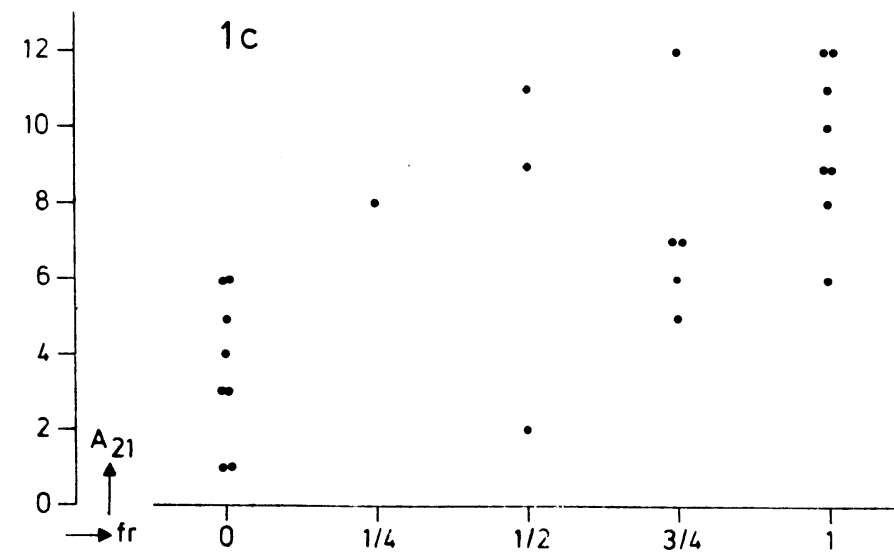
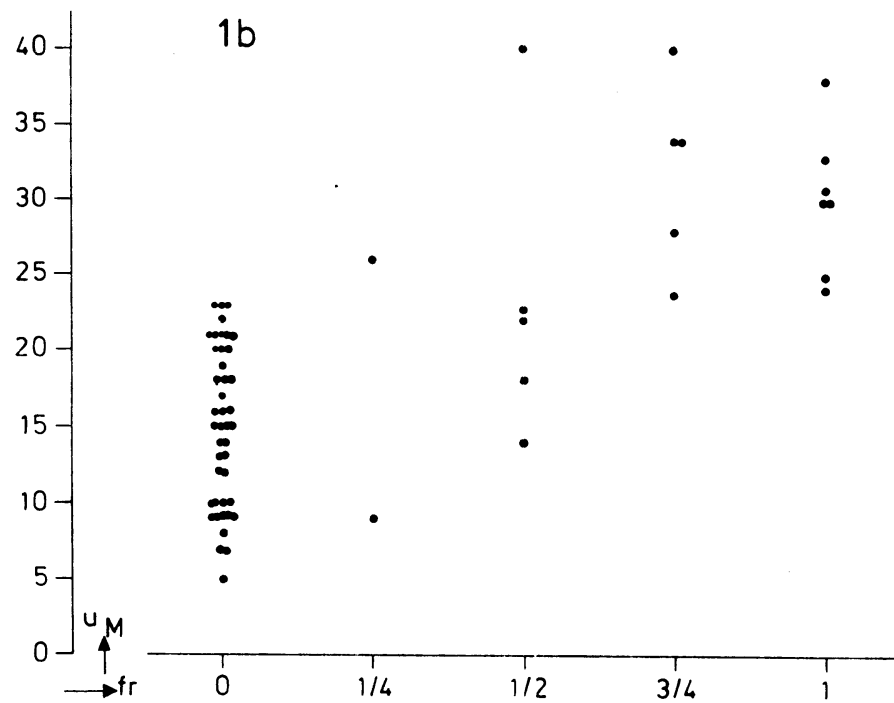
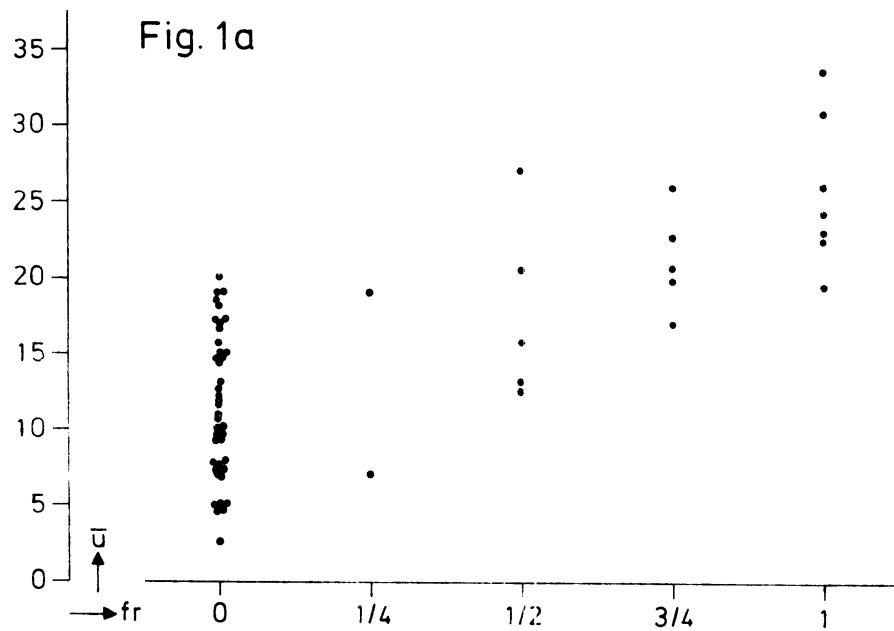
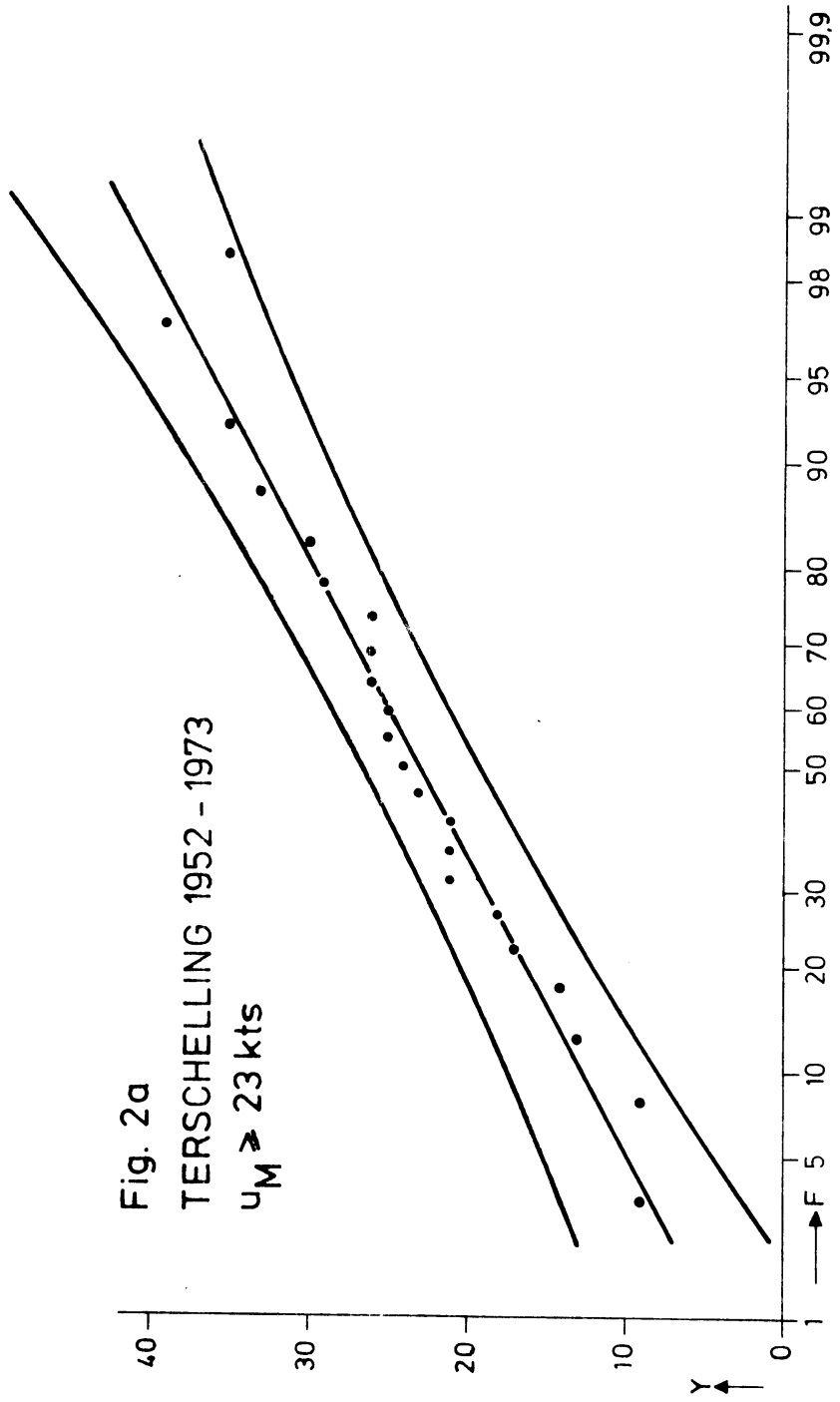


Fig. 2a
TERSCHELLING 1952 - 1973
 $u_M \geq 23$ kts



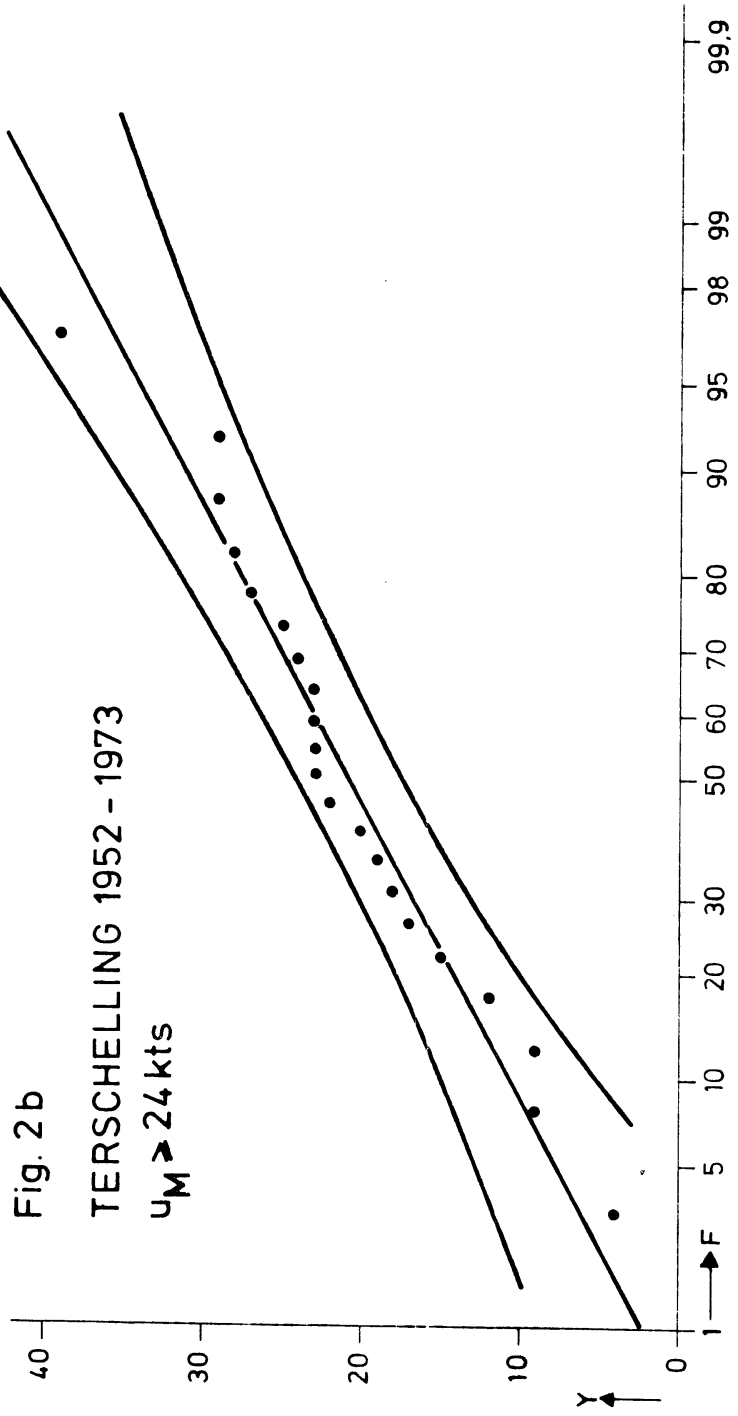


Fig. 3 KORWERDERZAND

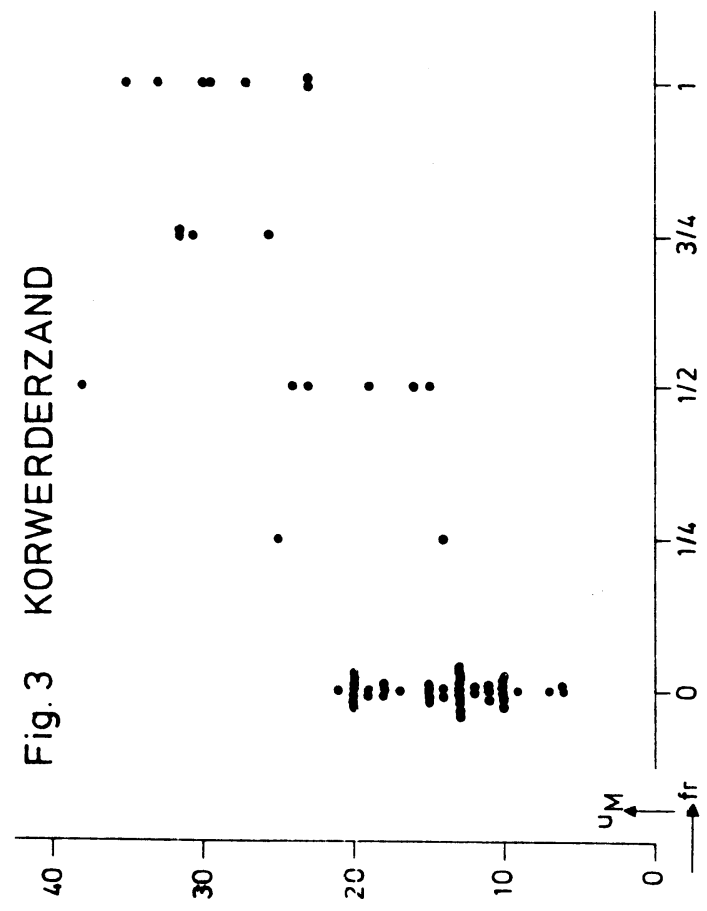
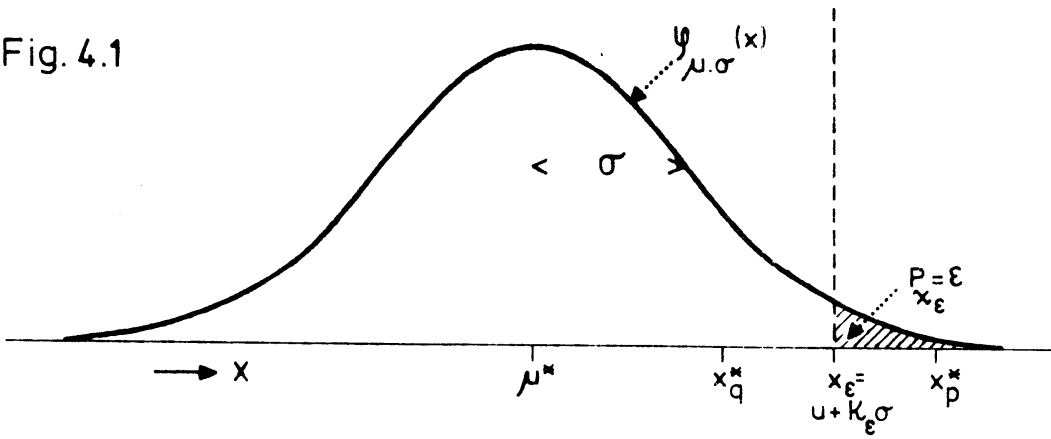
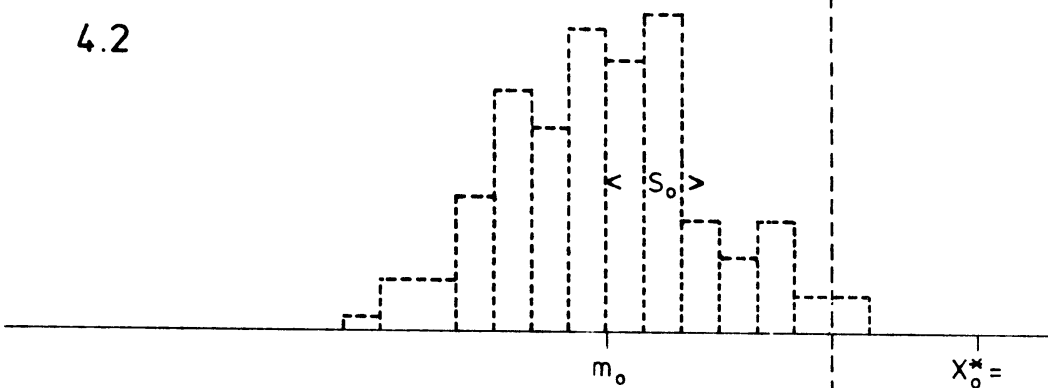


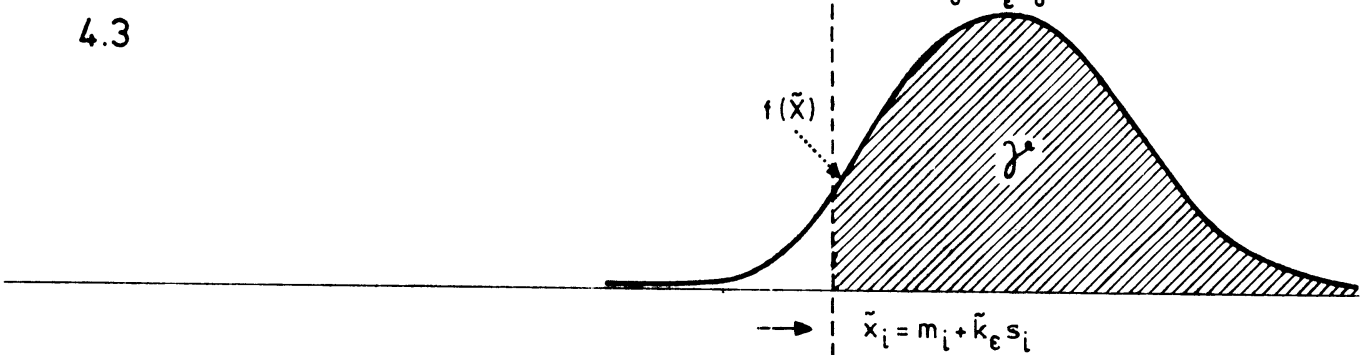
Fig. 4.1



4.2



4.3



4.4

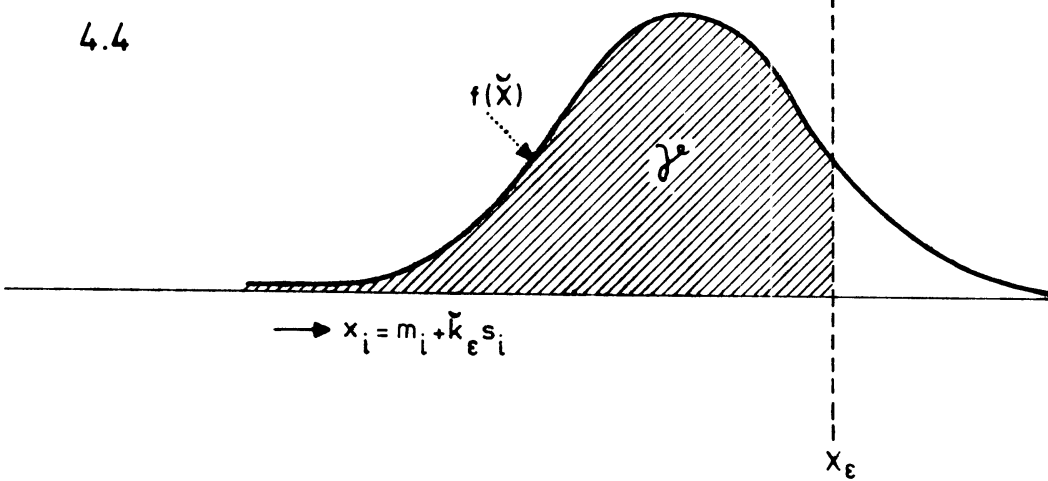
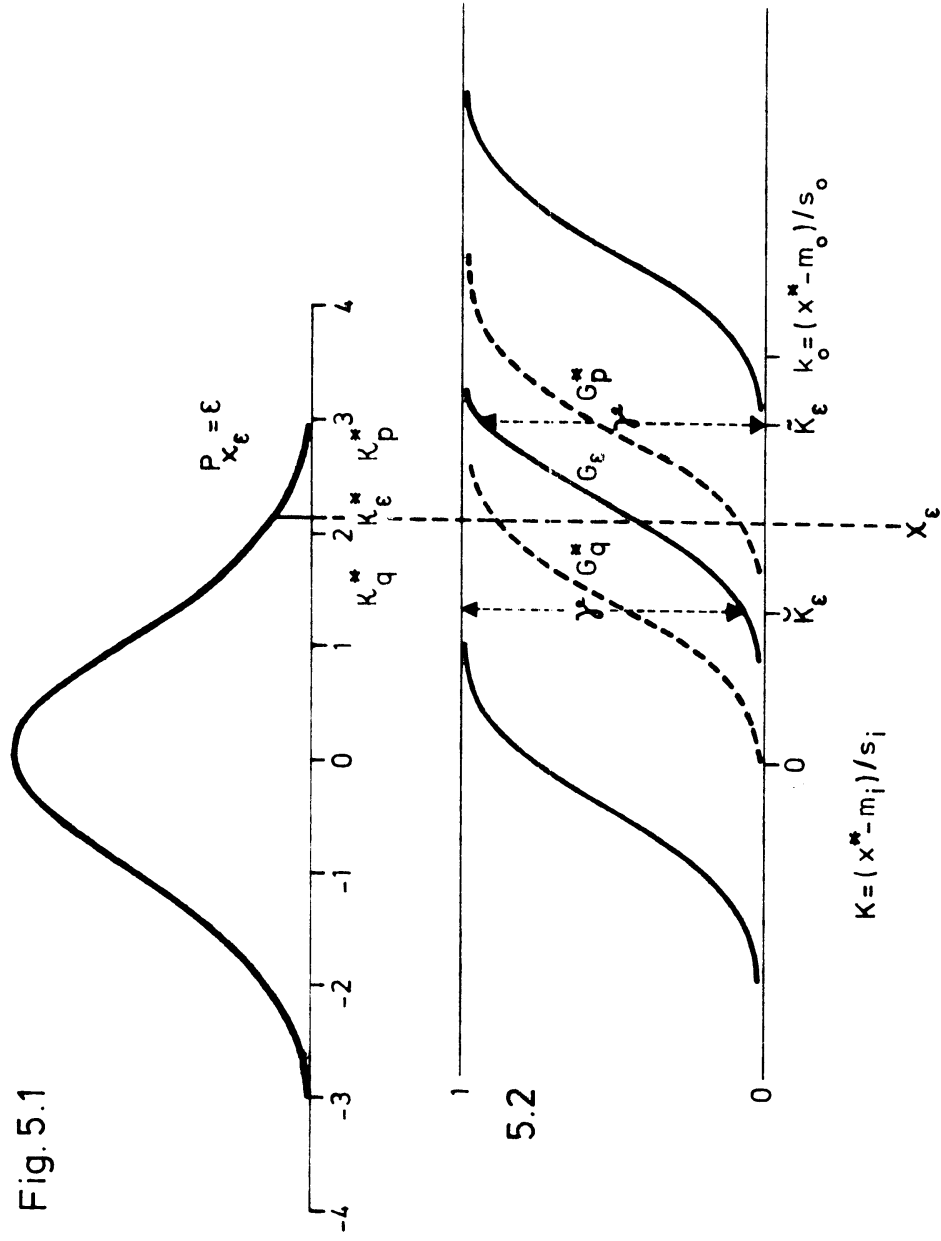


Fig. 5.1



5.2

Limieten voor individuele kansen uit een normaal verdeeld Universum op basis van een steekproef van n elementen.

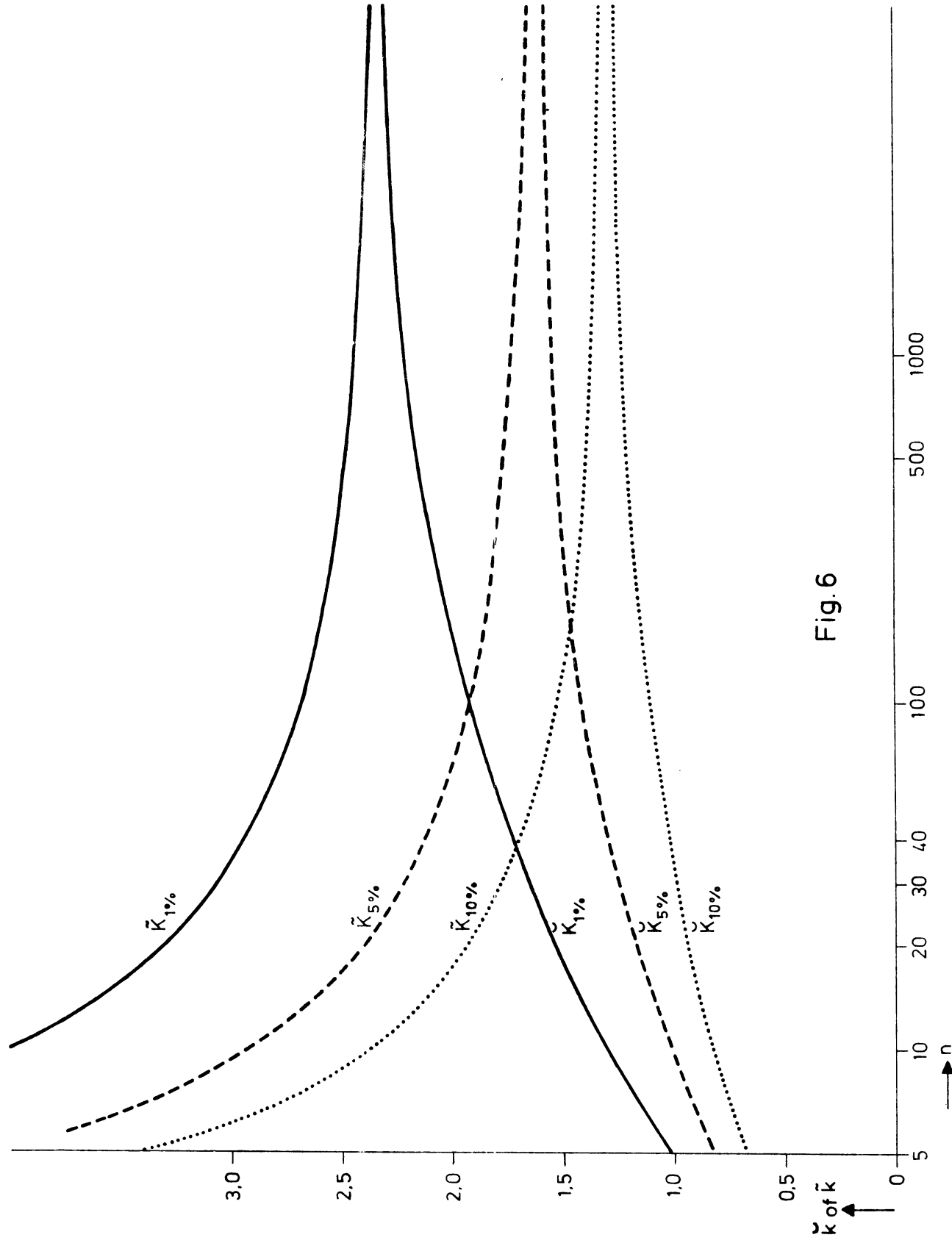


Fig.6