

31 AUG. 1965

KONINKLIJK NEDERLANDS  
METEOROLOGISCH INSTITUUT

Verslagen V-168  
(R III-294-1965)

"Een mogelijkheid tot een mathematische  
voorstelling van cumulatieve frequentie-  
verdelingen van k-daagse neerslagsommen,  
waarbij k kleiner of gelijk aan 30 is"

door

551.577.36 :  
519.2

Dr. H.J. de Boer

De Bilt, 1965

Kon. Ned. Meteor. Inst.  
De Bilt

## Inhoud

	blz.
1. Inleiding.	1
2. Momenten ten opzichte van het gemiddelde van een reeks getallen met persistentie.	2
3. Het verband tussen de momenten t.o.v. het gemiddelde van een reeks gepersisteerde getallen en de constanten van een Polya-verdeling.	5
4. Aanpassing van de momentenformules aan de momenten van de waargenomen cumulatieve frequentieverdelingen van k-daagse neerslag-sommen van juli te Hoofddorp, waarbij $k \leq 30$ .	6
5. Berekening van de numerieke waarden van de constanten van een Polya-verdeling voor verschillende k-waarden.	10
6. Goedheid van aanpassing van de waargenomen overlappende cumula-tieve frequentieverdelingen door Polya-verdelingen.	12
Samenvatting.	17
Literatuur.	18

1. Inleiding

In enkele publicaties, zie [1], [2] en [3], is aangetoond, dat cumulatieve frequentieverdelingen van k-daagse neerslagsommen op elk willekeurig station in Nederland op mathematische en voldoende nauwkeurige wijze kunnen worden voorgesteld door cumulatieve Poisson-verdelingen, mits k groter of gelijk aan 30 is. Deze laatste voorwaarde houdt in feite in, dat deze k-daagse neerslagsommen geen persistentie vertonen.

Als een cumulatieve frequentieverdeling van k-daagse neerslagsommen met overschrijdende kansen wordt voorgesteld door  $W(\geq s)$  dan wordt het verband tussen  $W$  en een cumulatieve Poisson-verdeling gegeven door de uitdrukking:

$$W(\geq s) = \sum_{j=s}^{\infty} e^{-\alpha} \frac{\alpha^j}{j!} \dots\dots\dots (1)$$

In betrekking (1) betekent de Poisson-constante  $\alpha$  het gemiddeld aantal neerslaghoeveelheden q in mm in het betreffende k-daagse tijdvak, zodat dan geldt

$q_k \alpha_k = \bar{h}_k$ .  $\bar{h}_k$  is de gemiddelde hoeveelheid neerslag in mm in het k-daagse tijdvak. Uit het tweede moment van de Poisson-verdeling kan nog een betrekking worden afgeleid  $q_k \bar{h}_k = \sigma_k^2$ . In deze betrekking is  $\sigma_k^2$  de variantie.

Worden van een cumulatieve frequentieverdeling  $W$  de grootheden  $\bar{h}_k$  en  $\sigma_k^2$  berekend, dan kan met behulp van (1)  $W$  door een cumulatieve Poisson-verdeling worden voorgesteld met de Poisson-constante  $\alpha_k$  en de Poisson-eenheid  $q_k$  in mm. Wil men de Poisson-verdeling op Gausz-papier uitzetten, dan wordt de overschrijdingskans voor een neerslaghoeveelheid van q mm uitgezet bij  $1/2 q$  mm, de kans op  $2 q$  en meer bij  $3/2 q$  mm, de kans op  $3 q$  mm en meer bij  $5/2 q$  mm, enz.

Voor meerdere details omtrent deze voorstellingswijze mag vooral naar [3] worden verwezen.

In de genoemde publikaties is voornamelijk over de voorstelling van cumulatieve frequentieverdelingen van k-daagse neerslagsommen medegedeeld, waarbij  $k \gg 30$ . Slechts in [1] is een poging gedaan om de verdelingen voor  $k \leq 30$  door cumulatieve Poisson-verdelingen voor te stellen. Dit bleek niet goed te gelukken. Vervolgens is eveneens gepoogd om de neerslagverdelingen voor  $k \leq 30$  voor te stellen door Poisson-verdelingen met persistentie, de z.g. Polya-verdelingen. Ook deze poging slaagde niet [1, p.50], hoewel dit verwondering wekte.

Daarom zal nu worden gepoogd om cumulatieve frequentieverdelingen van k-daagse neerslagsommen, waarbij  $k \leq 30$ , toch door cumulatieve Polya-verdelingen voor te stellen, maar dan na eerst enkele verbanden tussen eigenschappen van de waargenomen verdelingen en de constanten van een Polya-verdeling te hebben opgespoord.

2. Momenten ten opzichte van het gemiddelde van een reeks getallen met persistentie

Stel dat een reeks getallen  $u_i$ , waarbij  $i$  de waarden 1, 2, 3, ..... doorloopt, de eigenschappen heeft, dat  $\bar{u}_i = \mu$  en  $(u_i - \mu)(u_{i+s} - \mu) = 0$  voor  $s \neq 0$ , terwijl  $s$  een geheel getal is.

Nu wordt uit de niet-persistente  $u_i$ -reeks een reeks  ${}_1v_i$  met persistentie gevormd door het functioneel verband:

$${}_1v_i = \frac{\sum_{l=-m}^m r_l u_{i+l}}{\sum_{l=-m}^m r_l} \dots\dots\dots (2)$$

waarbij dus het aantal  $r_l$ -waarden  $2m+1$  bedraagt en  $r_0$  de waarde 1 heeft. Daarmede worden de termen van de  $u_i$ -reeks stuk voor stuk met behulp van telkens  $2m+1$  omliggende  $u_i$ -getallen gladgestreken.

Het functionele verband houdt in dat  ${}_1\bar{v}_i = \bar{u}_i = \mu$

Nu kan gemakkelijk de standaardafwijking van  ${}_1v_i$  worden uitgerekend:

$$\sigma_{{}_1v_i}^2 = \frac{\sum_{l=-m}^m r_l^2}{\left(\sum_{l=-m}^m r_l\right)^2} \sigma_{u_i}^2 \dots\dots\dots (3)$$

Daar de  ${}_1v_i$ -reeks gepersisteerd is kan de eerste orde autocorrelatiecoëfficiënt gemakkelijk worden berekend:

$$f_1 = \frac{(v_i - \mu)(v_{i+1} - \mu)}{\sigma_{{}_1v_i}^2} \dots\dots\dots (4)$$

Verder kan men schrijven:

$$\overline{(v_i - \mu)(v_{i+1} - \mu)} = \frac{\sum_{l=-m+1}^m r_l r_{l+1}}{\left(\sum_{l=-m}^m r_l\right)^2} \sigma_{u_i}^2 \dots\dots\dots (5)$$

en dus wordt met (3), (4) en (5)

$$f_1 = \frac{\sum_{l=-m+1}^m r_l r_{l-1}}{\sum_{l=-m}^m r_l^2} \dots\dots\dots (6)$$

Op dezelfde wijze kan worden afgeleid dat de s-de orde autocorrelatiecoëfficiënt  $f_s$  het volgende verband heeft met de  $r_1$ -coëfficiënten:

$$f_s = \frac{\sum_{-m+s}^m r_1 r_{1-s}}{\sum_{-m}^m r_1^2} \dots\dots\dots (7)$$

Als van de autocorrelatiefunctie s termen bekend zijn, d.w.z.  $f_0, f_1, f_2, f_3, \dots, f_s$ , en als ook nog s zo gekozen kan worden, dat s de waarden 0, 1, 2,  $\dots, 2m$  aanneemt, dan staan  $2m$  vergelijkingen met  $2m$  onbekenden  $r_{-m}, r_{-m+1}, r_{-m+2}, \dots, r_m$  tot onze beschikking ( $r_0 \equiv 1$ ). De  $2m$  onbekenden kunnen worden opgelost en het verband tussen de gepersisteerde en de niet-gepersisteerde reeks is bekend.

Daar  ${}_1v_i$  een ééndaagse neerslagsom kan voorstellen zal

$${}_k v_i = {}_1 v_i + {}_1 v_{i+1} + \dots + {}_1 v_{i+k-1} \dots\dots\dots (8)$$

een k-daagse neerslagsom kunnen voorstellen. Dan is eenvoudig af te leiden dat

$${}_k \bar{v}_i = \overline{({}_1 v_i + {}_1 v_{i+1} + \dots + {}_1 v_{i+k-1})} = k {}_1 \bar{v}_i = k \mu \dots (9)$$

De grootheid  ${}_k v_i$  kan met behulp van (2) ook als een functie van termen van de niet-gepersisteerde reeks worden uitgedrukt:

$${}_k v_i = \frac{\sum_{-m}^m r_1 (u_{i+1} + u_{i+1+1} + u_{i+1+2} + \dots + u_{i+1+k-1})}{\sum_{-m}^m r_1} \dots\dots\dots (10)$$

Dan kan men eveneens gemakkelijk de variantie van  ${}_k v_i$  berekenen:

$$\sigma_{k v_i}^2 = \frac{k \sum_{-m}^m r_1^2 + 2(k-1) \sum_{-m}^{m-1} r_1 r_{1+1} + 2(k-2) \sum_{-m}^{m-2} r_1 r_{1+2} + \dots + 2 \sum_{-m}^{m-k+1} r_1 r_{1+k-1}}{\left( \sum_{-m}^m r_1 \right)^2} \sigma_{u_i}^2 \dots\dots\dots (11)$$

Met behulp van de betrekkingen (3), (4) en (5) kan formule (11) omgewerkt worden tot de eenvoudige gedaante:

$$\sigma_{k v_i}^2 = \left\{ k+2(k-1) f_1 + 2(k-2) f_2 + 2(k-3) f_3 + \dots + 2 f_{k-1} \right\} \sigma_{1 v_i}^2 \quad (12)$$

Uitdrukking (12) geldt natuurlijk alleen, als de grootheden  $k v_i$  onafhankelijk van elkaar zijn. De persistentie, welke in (12) tot uitdrukking wordt gebracht, zit in de  $k$  grootheden  $1 v_i$ , waaruit  $k v_i$  is opgebouwd.

Hoe de persistentie er in zit of wel hoe de aard van de persistentie is, doet niet ter zake; immers de lineaire combinatie van niet-gepersisteerde grootheden, waaruit elke gepersisteerde grootheid bestaat, is niet meer in formule (12) terug te vinden. (12) bezit dus algemene geldigheid en kan daarom tevens op andere manieren worden afgeleid. Het is dan ook een reeds bekende formule.

Dezelfde methode kan nu worden gevolgd om het derde moment ten opzichte van het gemiddelde van een reeks gepersisteerde grootheden te bepalen. Dit derde moment t.o.v. het gemiddelde kan dan als volgt worden geschreven:

$$K_{k v_i}^3 = \left\{ k+3(k-1) \frac{\sum_{l=-m}^{m-1} r_l \cdot r_{l+1} (r_l + r_{l+1})}{\sum_{-m}^m r_l^3} + 3(k-2) \frac{\sum_{-m}^{m-2} r_l \cdot r_{l+2} (r_l + 2r_{l+1} + r_{l+2})}{\sum_{-m}^m r_l^3} \right. \\ \left. + 3(k-3) \frac{\sum_{-m}^{m-3} r_l \cdot r_{l+3} (r_l + 2r_{l+1} + 2r_{l+2} + r_{l+3})}{\sum_{-m}^m r_l^3} \right. \\ \left. + 3(k-4) \frac{\sum_{-m}^{m-4} r_l \cdot r_{l+4} (r_l + 2r_{l+1} + 2r_{l+2} + 2r_{l+3} + r_{l+4})}{\sum_{-m}^m r_l^3} + \dots \right\} K_{v_i}^3 \quad (13)$$

Bij de berekening is verondersteld, dat

$$(u_i - \mu)(u_{i+s} - \mu)(u_{i+p} - \mu) = 0 \quad \text{voor } s = 0 \text{ en } p = 0$$

en

$$(u_i - \mu)^2 (u_{i+s} - \mu) = 0 \quad \text{voor } s \neq 0$$

Zijn nu op een of andere wijze de  $r_l$ -waarden bekend of berekend, dan is formule (13) een handige uitdrukking voor het gebruik tezamen met formule (12). Zijn de

$r_1$ -waarden niet bekend, dan kan worden geschreven:

$$K_{k v_i}^3 = \left\{ k+3a_1 (k-1) + 3a_2(k-2) + 3a_3(k-3) + \dots \right\} K_{1 v_i}^3 \dots \dots \dots (13a)$$

De constanten  $a_1, a_2, a_3$ , enz. moeten dan uit het materiaal worden bepaald; dit kost niet zoveel rekenarbeid. Immers  $a_1$  kan uit  $K_{k v_i}^3$  en  $K_{1 v_i}^3$  worden bepaald,  $a_2$  uit  $K_{k v_i}^3$  en  $K_{3 v_i}^3$  na invulling van de waarde van  $a_1$ ,  $a_3$  uit  $K_{k v_i}^3$  en  $K_{1 v_i}^3$  na invulling van de waarde van  $a_1$  en die van  $a_2$ , enz.

Zijn de voorwaarden, waaronder formule (13) is afgeleid, niet vervuld, dan zal in het algemeen moeten worden gesteld:

$$K_{k v_i}^3 = (a + b k + c k^2 + d k^3) K_{1 v_i}^3 \dots \dots \dots (14)$$

De constanten  $a, b, c$  en  $d$  moeten dan op een of andere wijze uit het materiaal worden bepaald.

3. Het verband tussen de momenten ten opzichte van het gemiddelde van een reeks gepersisteerde getallen en de constanten van een Polya-verdeling

De Polya-verdeling is evenals de Poisson-verdeling een discrete verdeling. Er bestaat een sterk verband tussen beide soorten verdelingen; de Polya-verdeling is namelijk een Poisson-verdeling met persistentie. Het  $j$ -de punt van een distributieve Polya-verdeling is bepaald door de uitdrukking (zie [5]):

$$\frac{\alpha (\alpha+d)(\alpha+2d) \dots \dots \dots \{ \alpha + (j-1)d \}}{j! (1+d)^{\alpha/d+j}} \dots \dots \dots (15)$$

In uitdrukking (15) is  $\alpha$  het gemiddelde aantal gebeurtenissen in het beschouwde tijdvak van een zekere lengte, terwijl  $d$  de persistentiefactor wordt genoemd. Een Polya-verdeling heeft zijn eigen eenhedenstelsel. Een Polya-eenheid is, uitgedrukt in een ons bekende eenheid, gelijk aan  $q$ . Wanneer dit nu wordt betrokken op  $k$ -daagse neerslagsommen, dan is  $\alpha_k$  het gemiddelde aantal Polya-neerslageenheden in  $k$  dagen. Eén Polya-neerslageenheid is  $q_k$  mm.  $d_k$  is een maat voor de persistentie in het  $k$ -daagse tijdvak. Het eerste moment van een Polya-verdeling is

$$q_k \alpha_k = \bar{h}_k = \mu_k \dots \dots \dots (16)$$

Het tweede moment ten opzichte van het gemiddelde  $\bar{h}_k$  is

$$\sigma_{k i}^2 = q_k^2 \alpha_k (1+d_k) = \sigma_k^2 \dots \dots \dots (17)$$

en het derde moment ten opzichte van het gemiddelde  $\bar{h}_k$  is

$$K_{k i}^3 = q_k^3 \alpha_k (1+d_k)(1+2d_k) = K_k^3 \dots \dots \dots (18)$$

Aan de rechterzijde van de betrekkingen (16), (17) en (18) is een eenvoudiger notatie voor de momenten ingevoerd. Uit de drie genoemde betrekkingen zijn de Polya-constanten  $\alpha_k$  en  $d_k$  en de Polya-eenheid  $q_k$  te bepalen. Zo vindt men:

$$\alpha_k = \frac{\sigma_k^2 \mu_k^2}{2\sigma_k^4 - K_k^3 \mu_k}; \quad q_k = \frac{2\sigma_k^4 - K_k^3 \mu_k}{\sigma_k^2 \mu_k}; \quad d_k = \frac{K_k^3 \mu_k - \sigma_k^4}{2\sigma_k^4 - K_k^3 \mu_k} \dots \dots (19)$$

In (19) is

$$\mu_k = k \mu_1,$$

terwijl  $\sigma_k^2$  en  $K_k^3$  reeds bekende functies van  $k$  zijn en gegeven worden door (12) en (13a). In bepaalde gevallen worden beide laatstgenoemde grootheden gegeven door (12) en (14).

4. Aanpassing van de momentenformules aan de momenten van de waargenomen cumulatieve frequentieverdelingen van k-daagse neerslagsommen van juli te Hoofddorp, waarbij  $k \leq 30$

De in de vorige paragraaf afgeleide formules voor het berekenen van momenten van een reeks getallen met persistentie zullen nu worden toegepast op de momenten berekend uit de cumulatieve frequentieverdelingen van k-daagse neerslagsommen van de maand juli te Hoofddorp;  $k$  is hierbij kleiner of gelijk aan 30.

Uit een maand juli kunnen  $(31-k+1)$  k-daagse neerslagsommen worden gemaakt; deze  $(31-k+1)$  k-daagse sommen zijn overlappend geformeerd. In één k-daagse som zit de van dag tot dag persistentie verborgen; in de overlappende k-daagse sommen vertonen de k-daagse sommen ten opzichte van elkaar een kunstmatige door de overlapping aangebrachte persistentie.

In het frequentieboek Hoofddorp bevinden zich alleen frequentieverdelingen van



overlappende k-daagse sommen. Teneinde een eventueel effect van het overlappen in de frequentieverdelingen en dus in de berekening van de momenten te vermijden, zijn voor de maand juli ook k-daagse neerslagsommen, die geen overlapping vertoonden, tot frequentieverdelingen verwerkt. Zowel uit de frequentieverdelingen van de overlappende als uit die van de niet-overlappende k-daagse neerslagsommen zijn voor verschillende waarden van k de grootheden  $\mu_k$ ,  $\sigma_k^2$  en  $K_k^3$  berekend; de berekende waarden zijn in tabel I verzameld.

Tabel I

k	overlappend			niet-overlappend	
	$\mu_k$	$\sigma_k^2$	$K_k^3$	$\sigma_k^3$	$K_k^3$
1	2,2	21,39	336,8	21,39	336,8
2	4,2	49,45	956,6	52,57	1048,1
3	6,3	82,23	1611,5	84,53	1640,1
4	-	-	-	117,46	-
5	10,7	150,63	2890,8	153,66	2972
6	-	-	-	184,45	-
7	15,2	225,50	4158,4	216,57	3625
10	21,2	364,57	7773,4	340,57	7061
15	32,0	614,50	14404	483	(7844)
30	67,7	1397,05	32.165	1397	32.165

Uit tabel I blijkt voor k = 2, 3, 4 en 5 de waarden van de niet-overlappende  $\sigma_k^2$ -waarden iets groter zijn dan die van de overlappende, terwijl van k = 7 tot en met k = 15 het omgekeerde het geval is. Voor k = 30 zijn beide  $\sigma_k^2$ -waarden natuurlijk even groot. Bij het tweede moment zijn dus de verschillen tussen overlappende en niet-overlappende k-daagse neerslagsommen niet groot. In principe zal er dus voor berekening van momenten wel geen verschil tussen beide soorten k-daagse neerslagsommen bestaan. De bestaande verschillen zullen dus aan het steekproefeffect zijn toe te schrijven. Als dit zo is, dan is het even vanzelfsprekend dat na berekening van hogere momenten grotere verschillen tussen de overeenkomstige momenten van beide soorten neerslagsommen zullen blijken op te kunnen treden. Zo blijkt het steekproefeffect reeds op  $K_k^3$  een grotere invloed te hebben dan op  $\sigma_k^2$ .  
 Uit de tweede kolom van tabel I kan met behulp van formule (20) worden berekend:

$$\mu_k = 2,1849 k. \dots\dots\dots (21)$$

Uit de derde en vijfde kolom van tabel I kan met behulp van formule (12) worden berekend:

$$\sigma_k^2 = \sigma_1^2 \left\{ k + 0,234(k-1) + (0,234)^2 (k-2) + (0,234)^3 (k-3) + (0,234)^4 (k-4) \right\} \quad (22)$$

of

$$\sigma_k^2 = 21,39 \left\{ k + 0,234(k-1) + 0,0548(k-2) + 0,0128(k-3) + 0,0030(k-4) \right\} \quad (22a)$$

In tabel II zijn voor waarden van  $k = 1, 2, 3, 4, 5, 6, 7, 10, 15$  en  $30$  de waarden van  $\mu_k$  na berekening volgens formule (21) en de waarden van  $\sigma_k^2$  na berekening volgens formule (22a) in de tweede en derde kolom ingevuld. Het blijkt nu, dat de waarden van  $\mu_k$  in tabel II en die in tabel I slechts zeer weinig verschillen. Hetzelfde kan worden gezegd van de waarden van  $\sigma_k^2$ , behalve misschien voor  $k = 15$  en  $k = 30$ . Maar de berekende waarde van  $\sigma_{15}^2$  is nog groter dan de waargenomen  $\sigma_{15}^2$  bij niet-overlappende  $k$ -daagse neerslagsommen, maar kleiner dan die bij overlappende sommen. Daar voor  $k=30$  alleen niet-overlappende sommen zijn te gebruiken, kan niet worden beoordeeld of de berekende  $\sigma_{30}^2$  tussen de beide soorten waargenomen  $\sigma_{30}^2$  ligt. Er kan nog opgemerkt worden dat formule (22) een aanpassing van  $\sigma_k^2$  voorstelt met een lineaire persistentie, n.l.  $f_i = f_1^i$ . De aanpassing kan niet slecht worden, terwijl door de genoemde voorwaarden slechts één waarde, n.l.  $f_1$ , kan

Tabel II

k	$\mu_k$	niet-overlappend	
		$\sigma_k^2$	$\mu_k^3$
1	2,2	21,39	336,8
2	4,4	52,79	956,6
3	6,6	86,23	1.611,5
4	8,7	120,82	
5	10,9	155,18	3.743,6
6	13,1	189,53	
7	15,3	223,89	5.876
10	21,8	326,95	9.074
15	32,8	499	14.404
30	65,5	1.014	30.395

worden gekozen.

Indien nu de derde momenten van beide soorten verdelingen worden beschouwd, dan blijkt uit de vierde en zesde kolom van tabel I dat er verschillen van 3 tot 10% tussen beide soorten kunnen bestaan.

Blijkbaar moeten dergelijke verschillen tussen de derde momenten gewoon geacht worden. Voor  $k = 15$  is het verschil zeer veel groter; dit is echter een gevolg van het feit dat voor het niet-overlappende geval slechts de 15-daagse som van de eerste helft van de maand is gebruikt. Daarom is de betreffende waarde tussen haakjes gezet.

Ter berekening van  $K_k^3$  als  $K_1^3$  bekend is, kunnen de coëfficiënten  $a_1$ ,  $a_2$  en  $a_3$  met formule (13a) berekend worden uit de waargenomen overlappenden  $K_k^3$ -waarden voor  $k = 2, 3$  en  $15$ . De waarden van  $a_1$ ,  $a_2$  en  $a_3$  zijn dan respectievelijk  $0,28009$ ;  $0,03474$  en  $0,02069$ . Hiermede kunnen  $K_k^3$ -waarden worden berekend en de berekende waarden zijn ingevuld in tabel II. De overeenstemming tussen waarneming en berekening is nog niet groot, maar kan ongetwijfeld worden vergroot door een vierde term in te voeren met een coëfficiënt  $a_4$  in formule (13a).

Een andere wijze van aanpassing aan de momenten als functies van  $k$  is er één, welke gedeeltelijk gedekt wordt door de theorie gegeven in paragraaf 2. We stellen:

$$\sigma_k^2 = \sigma_1^2 (a_1 + a_2 k + a_3 k^2) \quad (23)$$

en

$$K_k^3 = K_1^3 (b_1 + b_2 k + b_3 k^2 + b_4 k^3) \quad (24)$$

Met behulp van de methode van de kleinste kwadraten, waarbij de waarden van de momenten van de overlappende sommen voor  $k = 2, 3, 5, 7, 10, 15$  en  $30$  als aanpassingsmateriaal worden genomen, vindt men dan:

$$\sigma_k^2 = 21,39 (-0,48237 + 1,45699k + 0,025381 k^2) \quad (23a)$$

$$K_k^3 = 336,8 (-0,46312 + 1,36731k + 0,10660k^2 - 0,00146k^3) \quad (24a)$$

In tabel III zijn de resultaten van deze aanpassingen verzameld en wel in kolom 3 en kolom 5. Op het oog lijkt de aanpassing in tabel III beter met de waarnemingen overeen te stemmen dan de aanpassingen in tabel II. Hoewel de aanpassingen in tabel II geheel volgens de theorie zijn en die van tabel III alleen voor  $K_k^3$ , ligt het wel in de lijn om de  $\sigma_k^2$  volgens formule (12) en de  $K_k^3$  volgens formule (14) aan te passen, daar de meeste verdelingen - zeker die

Tabel III

k	$\sigma_k^2(\text{wrg.})$	$\sigma_k^2(\text{ber.})$	$K_k^3(\text{wrg.})$	$K_k^3(\text{ber.})$
1	21,39	21,39	336,8	336,8
2	49,45	54,18	956,6	904,7
3	82,23	88,06	1.611,5	1.535,4
5	150,63	159,51	2.890,8	2.982,6
7	225,50	234,44	4.158,4	4.658,0
10	364,57	355,56	7.773,4	7.547,4
15	614,50	579,31	14.404,-	13.170,-
30	1.397,05	1.413,24	32.165,-	32.692,-

voor kleine k-waarden - niet symmetrisch zijn. Er is ook een mogelijkheid dat voor andere maanden dan juli een aanpassing met de formules (12) en (13a) op het oog betere resultaten oplevert. Dit zal nog moeten worden onderzocht.

5. Berekening van de numerieke waarden van de constanten van een Polya-verdeling voor verschillende k-waarden.

Met behulp van de drie formules (19) zijn de Polya-constanten  $\alpha_k$  en  $d_k$  en de Polya-eenheid  $q_k$  berekend uit de grootheden  $\mu_k$ ,  $\sigma_k^2$  en  $K_k^3$  van de frequentieverdelingen van de overlappende en van de niet-overlappende k-daagse neerslag-sommen. De waarden van de berekende  $\alpha_k$ ,  $d_k$  en  $q_k$  zijn voor beide soorten sommen in tabel IV verenigd.

Tabel IV

k	overlappend			niet-overlappend		
	$\alpha$	d	q	$\alpha$	d	q
1	0,5727	1,5310	3,8414	0,5727	1,5310	3,8414
2	1,0496	1,8010	4,0014	1,1590	2,1103	3,8189
3	0,9681	1,0058	6,5073	1,0878	1,0917	6,0951
4	-	-	-	1,1945	0,8674	7,2568
5	1,1937	0,5705	8,9638	1,3044	0,6413	8,4719
6	-	-	-	1,2607	0,3280	10,4961
7	1,3535	0,3210	11,2305	1,2640	0,2035	11,8954
10	1,6219	0,3156	13,0714	2,0738	0,4818	10,5523
15	2,1383	0,2832	14,9650	1,9372	0,0178	15,6434
30	3,7101	0,1309	18,2473	3,7101	0,1309	18,2473

De waarden van de Polya-grootheden van beide soorten zijn blijkbaar hier geen gladde functies van  $k$ ; echter wel kan worden geconstateerd dat de overlappende soort gladder is dan de niet-overlappende. Welke functies  $\alpha_k$ ,  $q_k$  en  $d_k$  zijn van  $k$ , kan worden bepaald met formule (19).

Deze formule geeft de drie grootheden als functies van  $\mu_k$ ,  $\sigma_k^2$  en  $K_k^3$ , terwijl formule (21)  $\mu_k$ , formule (23a)  $\sigma_k^2$  en formule (24a)  $K_k^3$  als functie van  $k$  laat zien. Hierdoor worden alle drie grootheden quotiënten, waarvan zowel teller als noemer vierde graads functies van  $k$  zijn. Indien men ziet naar de meer theoretische formules voor  $\mu_k$ ,  $\sigma_k^2$  en  $K_k^3$  van symmetrische frequentieverdelingen, n.l. formules (11), (12) en (13a), dan blijkt  $\alpha_k$  een lineaire functie van  $k$  te worden. Zo kan met de methode van de kleinste kwadraten uit de tweede kolom van tabel IV worden gevonden:

$$d_k = 0,64520 + 0,10163 k \quad (25)$$

Dan wordt automatisch:

$$q_k = 21,64715 - \frac{13,96674}{0,10163k + 0,64520} \quad (26)$$

Met dezelfde argumentatie wordt  $d_k$  een quotiënt van twee tweede graads functies in  $k$ . Zo kan  $d_k$  worden voorgesteld:

$$d_k = a + \frac{b}{k} + \frac{c}{k^2} \quad (27a)$$

Indien nu de constanten in (27a) worden bepaald met behulp van de methode van de kleinste kwadraten uit de waargenomen  $d_k$ -waarden in de derde kolom van tabel IV, dan vindt men

$$d_k = -0,15398 + \frac{4,97589}{k} - \frac{3,15560}{k^2} \quad (27)$$

Met behulp van (25), (26) en (27) worden  $\alpha_k$ ,  $q_k$  en  $d_k$  berekend en de gevonden waarden worden in tabel V verzameld naast de waargenomen waarden van de drie grootheden, welke bepaald waren uit de frequentieverdelingen van de overlappende  $k$ -daagse neerslagsommen.

De berekende  $\alpha_k$ - en  $q_k$ -waarden verschillen maar weinig van de waargenomen overeenkomstige waarden, behalve die voor  $k = 1$  en  $k = 2$ . Voor deze twee waarden van  $k$  zijn de waargenomen waarden van  $\alpha$  en  $q$  verre van glad vergeleken met de die voor de overige  $k$ -waarden.

Tabel V

k	$\alpha_k$	$\alpha_k$	$q_k$	$q_k$	$d_k$	$d_k$
	wrg. overl.	berek.	wrg. overl.	berek.	wrg. overl.	berek.
1	0,5727	0,7468	3,8414	2,9450	1,5310	1,6663
2	1,0496	0,8485	4,0014	5,1866	1,8010	1,5451
3	0,9681	0,9501	6,5073	6,9469	1,0058	1,1540
5	1,1937	1,1534	8,9638	9,5380	0,5705	0,7150
7	1,3535	1,3566	11,2305	11,3518	0,3210	0,4925
10	1,6219	1,6615	13,0714	13,2410	0,3156	0,3120
15	2,1383	2,1697	14,9650	15,2100	0,2832	0,1637
30	3,7101	3,6941	18,2473	17,8663	0,1309	0,0837

De waargenomen  $d_k$ -waarden zijn niet alleen voor  $k = 1$  en  $k = 2$  maar ook voor  $k = 7, 10$  en  $15$  verre van glad.

De vermoedelijke oorzaak van dit niet gladde verloop met  $k$  van de waargenomen waarden van  $\alpha_k$ ,  $q_k$  en  $d_k$  ligt in het feit, dat deze drie grootheden uit de eerste drie momenten van de waargenomen verdelingen zijn berekend. De fout in de berekening van het tweede moment neemt met het kwadraat toe ten opzichte van het eerste moment en de fout in het derde moment neemt toe met de derde macht ten opzichte van het eerste moment.

In plaats van de Polya-constanten en de Polya-eenheid met behulp van de momenten van de waargenomen verdelingen te berekenen verdient het misschien aanbeveling de berekening van de Polya-constanten uit te voeren met behulp van de methode van de maximum likelihood en de Polya-eenheid te bepalen met formule (16).

#### 6. Goedheid van aanpassing van de waargenomen cumulatieve frequentieverdelingen door Polya-verdelingen

Thans zullen de cumulatieve frequentieverdelingen van de  $k$ -daagse neerslag-sommen voor juli te Hoofddorp voor  $k = 1, 2, 5$  en  $30$  getoetst worden met de overeenkomstige Polya-verdelingen, waarvan de constanten met behulp van de momentenmethode uit de genoemde waargenomen verdelingen zijn bepaald geworden. De frequentieverdeling voor  $k = 1$  in het frequentieboek voor Hoofddorp is gemaakt van 2572 ééndaagse neerslagsommen. Nu zijn de op eenvolgende ééndaagse sommen niet onafhankelijk van elkaar ten gevolge van het verschijnsel van persistentie. Door de persistentie wordt het aantal van elkaar onafhankelijke

ééndaagse sommen, het effectieve aantal =  $N_{\text{eff}}$ , kleiner dan het waargenomen aantal. Voor  $k = 2$  en groter blijken er twee soorten persistentie te zijn, nl. de reeds genoemde interne persistentie en de kunstmatige persistentie, welke is aangebracht door de  $k$ -daagse neerslagsommen in een maand overlappend te nemen. Levert [4] heeft de effectieve aantallen uitgerekend; waarbij  $k = 30$  niet-overlappend is:

k	$N/N_{\text{eff}}$
1	1,63
2	2,58
5	5,24
30	1,02

Hoe moet nu een aangepaste Polya-verdeling worden uitgezet. Met behulp van formule (15) wordt de distributieve Polya-verdeling berekend door achtereenvolgens  $j = 1, 2, 3$  enz. in te vullen. Van de distributieve verdeling wordt door optelling de cumulatieve verdeling verkregen. Op lineair Gausz-papier wordt het eerste punt van de cumulatieve verdeling uitgezet bij  $1/2 q$  mm, het tweede punt bij  $1 1/2 q$  mm, enz. (zie [1]). Zo wordt, hetgeen in figuur 1 is uitgezet, in tabel VI bijeengebracht. Naast de theoretische Polya-verdeling is in figuur 1 tevens de waargenomen verdeling uitgezet.

Nu zal de theoretische verdeling met behulp van de  $\chi^2$ -toets op de goede aanpassing aan de waargenomen verdeling worden getoetst.

Tabel VI

j	distr. kans	cum. kans overschrijdend	q	mm neerslag
1	0,70651	0,70651	$\frac{1}{2}$	1,9207
2	0,15986	0,86637	$1\frac{1}{2}$	5,7621
3	0,06644	0,93281	$2\frac{1}{2}$	9,6035
4	0,03180	0,96461	$3\frac{1}{2}$	13,4449
5	0,01623	0,98084	$4\frac{1}{2}$	17,2863
6	0,00859	0,98943	$5\frac{1}{2}$	21,1277
7	0,00465	0,99408	$6\frac{1}{2}$	24,9691
8	0,00256	0,99664	$7\frac{1}{2}$	28,8105
9	0,00143	0,99807	$8\frac{1}{2}$	32,6519
10	0,00080	0,99887	$9\frac{1}{2}$	36,4933
11	0,00045	0,99932	$10\frac{1}{2}$	40,3347

In tabel VII staat het voorbeeld van de uitgevoerde toetsing. We beginnen bovenaan bij  $\bar{W} = 100\%$  om te zien vanaf hoeveel mm de theoretische verdeling voldoende aan de waargenomen verdeling is aangepast. Bij een dergelijke toetsing moet steeds worden gezorgd dat in elke klasse 6 of meer punten aanwezig zijn. Dat kan met de theoretische verdeling erg gemakkelijk dan moet de tweede kolom ( $1-\bar{W}$ ) telkens bij 0,2333% omhoog gaan. Dit levert grote praktische moeilijkheden op bij het aflezen van fig.1 en dus het invullen van kolom 1 en bij interpoleren

Tabel VII

neerslagin- terval in mm	$1-\bar{W}$ in %	t	w	$\frac{(w-t)^2}{t}$	$\frac{)}^2}{1,63}$	$\sum \chi^2/1,63$	v	P( $\chi^2$ )
∞ - 31	0,245	6,3	6	0,01	0,01	0,01		
31 - 26	0,51	6,8	6	0,09	0,06	0,07		
26 - 25	0,59	2,1	6	7,72	4,73	4,80		
25 - 22	0,94	9,0	6	1,50	0,92	5,72		
22 - 21	1,14	5,1	5	-	-	5,72	1	
21 - 20	1,32	4,7	7	1,13	0,69	6,41	2	
20 - 19	1,52	5,1	6	0,16	0,10	6,51	3	
19 - 18	1,75	5,9	12	6,31	3,87	10,38	4	
18 - 17	2,05	7,7	10	0,69	0,42	10,80	5	0,055
17 - 16	2,45	10,3	8	0,51	0,31	11,11	6	0,084
16 - 15	2,75	7,8	9	0,18	0,11	11,22	7	0,13
15 - 14	3,20	11,5	12	0,02	0,01	11,23	8	0,19
14 - 13	3,80	15,5	13	0,40	0,25	11,48	9	0,24
13 - 12	4,50	18,0	18	0	0	11,48	10	0,32
12 - 11	5,35	21,8	24	0,22	0,13	11,61	11	0,40
11 - 10	6,25	23,2	35	6,00	3,68	15,29	12	0,23
10 - 9,5	6,90	16,7	22	1,68	1,03	16,32	13	0,23
9,5- 9	7,50	15,2	12	0,67	0,41	16,73	14	0,27
9 - 8,5	8,05	14,4	14	0,01	0,01	16,74	15	0,34
8,5- 8	8,80	19,2	14	1,41	0,87	17,61	16	0,35
8 - 7,5	9,60	20,6	16	1,03	0,63	18,24	17	0,37
7,5- 7	10,50	23,2	22	0,06	0,04	18,28	18	0,43
7 - 6,5	11,45	24,4	19	1,20	0,74	19,02	19	0,45
6,5- 6	12,75	33,5	30	0,37	0,23	19,25	20	0,51
6 - 5,5	14,05	33,4	21	4,60	2,82	22,07	21	0,41
5,5- 5	15,2	29,6	33	0,39	0,24	22,31	22	0,45



neerslagin- terval in mm	$1-\bar{W}$ in %	t	w	$\frac{(w-t)^2}{t}$ = $\chi^2$	$\chi^2/1,63$	$\sum \chi^2/1,63$	$\sqrt{\quad}$	$P(\chi^2)$
5 - 4,5	17,0	46,3	28	7,23	4,44	26,75	23	0,29
4,5- 4	19,1	54,0	32	8,96	5,50	32,25	24	0,125
4 - 3,5	21,3	56,6	43	3,27	2,01	34,26	25	0,10
3,5- 3	23,8	64,3	64	-	-	34,26	26	0,125
3 - 2,5	26,0	56,5	50	0,75	0,46	34,72	27	0,165
2,5- 2	28,7	69,5	59	1,59	0,98	35,70	28	0,17
2 - 1,5	32,5	97,7	81	2,85	1,75	37,45	29	0,145
1,5- 1,0	37,1	118,3	98	3,48	2,14	39,59	30	0,115
1,0- 0,8	38,6	38,6	43	0,50	0,31	39,90	31	0,14
0,8- 0,6	40,3	43,7	40	0,31	0,19	40,09	32	0,15
0,6- 0,4	43,7	87,4	70	3,86	2,37	42,46	33	0,12
0,4- 0,2	46,4	69,5	98	11,69	7,17	49,63	34	0,04
0,2- 0	100	1378,6	1480	7,46	4,58	54,21	35	0,019

ten behoeve van het bepalen van het aantal punten van de waargenomen verdeling in dezelfde klasse. Daarom keren we, hoewel dit niet gebruikelijk is, het procédé om. We kiezen n.l. voor de w-kolom een aantal millimeterklassen bij elkaar tot dat we 6 punten hebben en als in één millimeterklasse of in een halve millimeterklasse meer dan 6 punten zijn, zoals in de meeste, dan wordt die millimeter of die halve millimeter in de eerste kolom als klasse genomen. Zo blijkt uit de eerste regel van tabel VII dat het zes malen is voortgekomen (w) dat een aantal van 31 of meer millimeters neerslag is opgetekend in de maand juli te Hoofddorp. Op de theoretische verdeling lees ik bij 31 mm af:  $1 - \bar{W} = 0,245\%$ . Nu is  $0,245\%$  van de 2572 gemeten ééndaagsneerslaghoeveelheden 6,3; d.w.z. in de klasse van  $\infty - 31$  mm wijst de theoretische verdeling 6,3 punten aan. In de klasse van 31 - 26 mm bleek wederom  $w = 6$ . Bij 26 mm lees ik op de theoretische kromme af:  $1 - \bar{W} = 0,51$ . De klasse 31 - 26 mm heeft dus een theoretische kans van voorkomen van  $0,265\%$ . Dit geeft  $t = 2572 \times 0,00265 = 6,8$ . Uit de waarden van t en w worden de waarden van de grootheid  $(w-t)^2/t$  berekend. Deze grootheid ook  $\chi^2$  genoemd heeft de dimensie van een aantal waarnemingen. Het effectieve aantal waarnemingen ten gevolge van persistentie bij ééndaagse neerslagsommen is gelijk aan het aantal waarnemingen gedeeld door 1,63. Zo vindt men in de kolom naast die van  $\chi^2$  de grootheid  $\chi^2/1,63$ . Tenslotte wordt de grootheid  $\chi^2/1,63$  in de daaropvolgende kolom gesommeerd. Het aantal vrijheidsgraden  $\sqrt{\quad}$  van de  $\chi^2$ -sommen is gelijk aan het

aantal klassen minus één verminderd met het aantal constanten in de theoretische verdeling. Daar het aantal constanten in de Polya-verdeling drie bedraagt, begint het aantal vrijheidsgraden  $\nu$  eerst bij de klasse van 22 - 21 mm te tellen. Tenslotte wordt in de laatste kolom de overschrijdingskans  $P(\chi^2)$  van de genoemde  $\chi^2$ -som bij dat aantal vrijheidsgraden uit een grafiek afgelezen. Zolang  $P$  groter of gelijk aan 0,05 is, mag worden gezegd, dat de theoretische verdeling aan de waargenomen dito voldoende nauwkeurig is aangepast.

Hier mag dus worden geconstateerd dat voor  $k = 1$  de aangepaste Polya-verdeling voldoende nauwkeurig is voor hoeveelheden neerslag groter dan 0,4 mm.

Op geheel dezelfde wijze blijkt voor  $k = 2$  de Polya-verdeling met  $\alpha_2 = 1,0496$ ;  $q_2 = 4,0014$  en  $d_2 = 1,8010$  voldoende nauwkeurig aan de waargenomen verdeling te zijn aangepast voor  $h$  ( $h =$  hoeveelheid neerslag) is groter of gelijk aan 0,2 mm.

Voor  $k = 5$  en  $k = 30$  blijkt het, dat de respectieve Polya-verdelingen goede aanpassingen aan de waargenomen verdelingen zijn.

Er moet wel aan worden gedacht dat voor  $k = 2$  de  $\chi^2$ -bijdragen ten gevolge van de persistentie door 2,58 moeten worden gedeeld; voor  $k = 5$  door 5,24 en voor  $k = 30$  door 1,02.

Uit deze resultaten mag worden besloten, dat cumulatieve Polya-verdelingen redelijk goede aanpassingen kunnen zijn van cumulatieve frequentieverdeling van  $k$ -daagse neerslagsommen, waarbij  $k \leq 30$ .

Alvorens te beëindigen zijn nog enkele opmerkingen te maken.

1. Het onderzoek is feitelijk nog verre van afgesloten. Immers we weten ten naaste bij hoe  $\alpha_k$ ,  $q_k$  en  $d_k$  veranderen met  $k$ . Misschien wordt dit verband duidelijker, wanneer de genoemde constanten met behulp van de methode van de maximum likelihood worden bepaald in plaats van met de momentenmethode; het verband behoeft niet duidelijker te worden daar in vele gevallen met beide methoden hetzelfde antwoord wordt verkregen. Op het eerste gezicht lijkt ons de kans op een minder onnauwkeurig antwoord met de maximum likelihood-methode groot.
2. Bij de aanpassing van cumulatieve frequentieverdelingen van  $k$ -daagse neerslagsommen in Nederland voor  $k \gg 30$  met cumulatieve Poisson-verdelingen bleek  $\alpha_k$  een lineaire functie van  $k$  te zijn en onafhankelijk van de jaarlijkse gang; de Poisson-eenheid  $q_k$  bleek alleen afhankelijk te zijn van de jaarlijkse gang. Het zou aangenaam zijn, indien dit feit zich ook manifesteerde voor  $k \leq 30$  bij Polya-verdelingen. Hoewel ook hier  $\alpha_k$  een lineaire functie van  $k$  blijkt te zijn en ook geldt dat  $q_k \alpha_k = \mu_k$  zijn hier  $\alpha_k$  en  $\mu_k$  helaas niet dezelfde lineaire functie van  $k$ . Het zal dus nog moeten

worden uitgezocht òf en h e  $\alpha_k$  en  $q_k$  beide afhankelijk zijn van de jaarlijkse gang.

3. Bij het in grafiek brengen van een Polya-verdeling wordt het eerste punt ( $j = 1$ ) uitgezet bij  $1/2 q$  mm; het punt  $j = 2$  bij  $1 1/2 q$  mm, enz.

Het is nu de vraag, hoe de kans op 0 mm neerslag en meer moet worden bepaald. We behelpen ons op de volgende wijze.

Het  $j$ -de punt van een distributieve Polya-verdeling wordt gegeven door de uitdrukking

$$\frac{\alpha (\alpha + d)(\alpha + 2d) \dots \{\alpha + (j - 1)d\}}{j! (1+d)^j + \alpha / d}$$

Dit kan ook geschreven worden als

$$\frac{\Gamma(\frac{\alpha}{d} + j) \times d^j}{\Gamma(\frac{\alpha}{d}) \Gamma(j + 1) \times (1 + d)^j (1 + d)^{\alpha/d}} \quad (28)$$

Waar nu in het algemeen cumulatief het  $j$ -de punt wordt uitgezet bij  $(j - 1/2) q$  mm, zal dus 0 mm moeten worden uitgezet bij  $j = 1/2$  ingevuld in betrekking (28). Helaas brengt dit geen uitkomst voor het geval  $k = 1$ . In hoeverre de hier aangegeven weg juist is, is ons onbekend.

#### Samenvatting

Er wordt aangetoond, dat de cumulatieve frequentieverdelingen van  $k$ -daagse neerslagsommen van de maand juli te Hoofddorp, waarbij  $k$  kleiner of gelijk aan 30 is, redelijk goed kunnen worden voorgesteld door een cumulatieve Polya-verdeling. De constanten van de Polya-verdeling worden bepaald met de momentenmethode. Op theoretische en op praktische wijze wordt bepaald welke functie van  $k$  de momenten ten opzichte van het gemiddelde van een reeks getallen met persistentie zijn. Daardoor kan worden nagegaan, welk verband de constanten van een Polya-verdeling met  $k$  hebben.

Literatuur

- 1 Boer, H.J. de; "De cumulatieve frequentieverdelingen van de k-daagse neerslagsommen van Hoofddorp", K.N.M.I. WR 56-004, De Bilt (1956).
- 2 Boer, H.J. de: "De cumulatieve frequentieverdelingen van de k-daagse neerslagsommen van Winterswijk", K.N.M.I. WR 57-007, De Bilt (1957).
- 3 Boer, H.J. de: "On the cumulative frequency distributions of k-day period amounts of precipitation for any station in the Netherlands, while  $k \geq 30$ ", Archiv F.Met.Geoph. u. Bioklim. B, Bd 9 (1958), pp. 244-253.
- 4 Levert, C. : "De persistentie van de dagelijkse hoeveelheid neerslag", K.N.M.I. WR 55-004, De Bilt (1955), pp. 34-38.
- 5 Brooks, C.E.P. and Carruthers, N.: "Handbook of Statistical Methods in Meteorology", p. 315.

# HOOFFDORP JULI

Cum. freq. verd. 1 - daagse neerslagsommen

$\alpha = 0.5727$   
 $d = 1.5310$   
 $q = 3.8414 \text{ mm}$

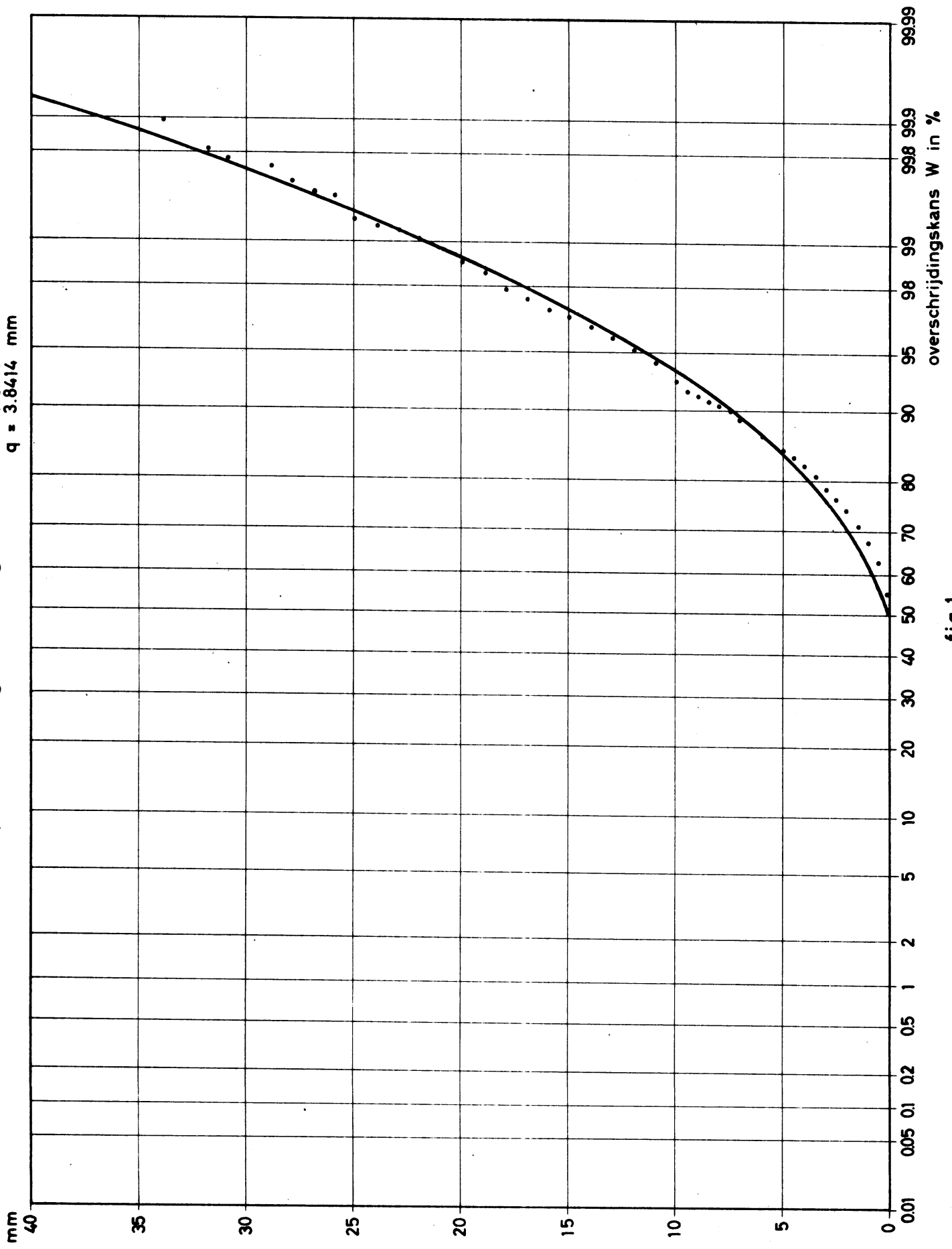


fig 1