

KONINKLIJK NEDERLANDS  
METEOROLOGISCH INSTITUUT

Wetenschappelijk Rapport W.R. 55-006 (R III-155)

Dr C. Levert

Statistische kritiek op het artikel van E.G. Bowen

The influence of meteoritic dust on rainfall

De Bilt, 1955

**All Rights Reserved.**

**Nadruk zonder toestemming van het K.N.M.I. is verboden.**

Corrigenda

bij W.R. 55-006 (RIII-155 - 1955)

Statistische kritiek op het artikel van E.G. Bowen  
The influence of meteoritic dust on rainfall

Dr C. Levert

- in noot <sup>1)</sup> blz. 8, onder "voorbeeld". Lees  $(\frac{1}{6})^{10} = 1.7 \cdot 10^{-8}$   
en lees "één op de  $6 \cdot 10^7$  keren"
- in noot <sup>1)</sup> blz. 21, voeg toe: zie ook "Fluctuations of long-period accumulations  
of daily rainfall amounts". J. Neumann; [8]

De volgende verbeteringen zijn bijna alle nodig vanwege het feit, dat bij  $y = 1$   
uit de gestandaardiseerde normale verdeling een overschrijdingskans 0.139 afge-  
lezen werd, i.p.v. 0.159.

- blz. 28, onder "mogelijkheid 1",  
achter I lees  $K_b = 0,159$ ;  $p = 0,0398$ ;  $p^2 = 0,00158$ ;  
 $P_i = 0,00158$   
achter II lees  $p^2 = 0,000032$ ;  $P_i = 0,000032$   
r. 10 en 9 v.o. lees  $31 \times 0,00158 = 0,049$  en  $31 \times 0,000032 = 0,001$
- blz. 29, r. 7, 8, 9 v.b. lees  $5 \times 0,00158 = 0,00790$ ;  $5 \times 0,000032 = 0,00016$ ;  
 $31 \times 0,0079 = 0,24$ ;  $31 \times 0,00016 = 0,005$   
r. 18, 19 v.b. lees  $p = 0,159^2 = 0,0254$ ;  $P_i = 0,0254^2 = 0,00064$ ;  
 $P = 31 \times 0,00064 = 0,020$   
r. 23 en 24 v.b. lees  $P = 5 \times 0,020 = 0,10$  en  $P = 5 \times 8.4 \cdot 10^{-6} = 4.2 \cdot 10^{-5}$   
r. 7 v.o. lees  $K_b = 0,159$   
r. 6 v.o. lees  $p = 0,159 \cdot 0,159^2 = 0,004$ ;  $P_i = 0,004^2 = 1.6 \cdot 10^{-5}$   
r. 5 v.o. lees  $P = 31 \times 1.6 \cdot 10^{-5} = 5 \cdot 10^{-4}$ ;  
r. 4 v.o. lees  $\frac{p}{P} = 0,159 \times 0,0228^2 = 8.3 \cdot 10^{-5}$ ;  $P_i = (8.3 \cdot 10^{-5})^2 =$   
 $6.9 \cdot 10^{-9}$ ;  $P = 31 \times 6.9 \cdot 10^{-9} = 2.1 \cdot 10^{-7}$   
r. 1 v.o. lees  $P = 5 \times 5 \cdot 10^{-4} = 2.5 \cdot 10^{-3}$  en  $P = 5 \times 2.1 \cdot 10^{-7} = 1.05 \cdot 10^{-6}$
- blz. 30 r. 10 v.b. lees 0,159
- blz. 31 r. 8 v.b. lees  $K_b = 0,159$  i.p.v. 0,139  
r. 9 v.b. lees  $K_c = 0,159$  i.p.v. 0,139  
r. 10 v.b. lees 0,0398 i.p.v. 0,0348; 0,0254 i.p.v. 0,0194;  
0,0040 i.p.v. 0,0027;  $8.3 \cdot 10^{-5}$  i.p.v.  $3.2 \cdot 10^{-3}$ .

blz. 31, r. 11 v.b. lees 0,00158 i.p.v. 0,00121; 6.4 i.p.v. 3.8;  
1.6.10<sup>-5</sup> i.p.v. 7.3.10<sup>-6</sup>; 7.10<sup>-9</sup> i.p.v. 10<sup>-5</sup>

r. 12 v.b. lees 0,049 i.p.v. 0,037; 0,020 i.p.v. 0,012;  
5 i.p.v. 2.2; 2.10<sup>-7</sup> i.p.v. 3.1.10<sup>-4</sup>

r. 13 v.b. lees 0,24 i.p.v. 0,186; 0.10 i.p.v. 0,060;  
2.5 i.p.v. 1,1; 10<sup>-6</sup> i.p.v. 1.6.10<sup>-4</sup>.

blz. 37 onder Literatuur [8] Australian Journal of Physics 7 522 1954.

C.L.

Dr C. Levert

Statistische kritiek op het artikel van E.G. Bowen

The influence of meteoritic dust on rainfall

| blz. |     | <u>Inhoud</u>  |
|------|-----|--|
| 2    | 0   | Inleiding  |
| 2    | 1   | Korte inhoud van het artikel van Bowen   |
| 6    | 2   | Eigen statistische kritiek   |
| 19   | 3   | Kritiek van anderen  |
| 19   | 3.1 | Kritiek van W.C. Swinbank  |
| 20   | 3.2 | Kritiek van D.F. Martyn  |
| 21   | 3.3 | Kritiek van J. Neumann   |
| 23   | 3.4 | Kritiek in Annual Report 1953/54 Commonwealth scientific and industrial<br>research organization section of meteorological physics |
| 24   | 4   | Bevestiging van Bowen's bewering.  |
| 24   | 4.1 | Artikeltje van M.B. Oliver and V.J. Oliver   |
| 24   | 4.2 | Artikeltje van O.N. Dhar   |
| 25   | 5   | Enige citaten uit "Der Jahresgang der Niederschlagswahrscheinlichkeit<br>auf dem Somblick", F. Steinhauser                         |
| 26   | 6   | Behandeling van het probleem aan Nederlands materiaal  |
| 28   | 7   | Formulering van een nieuw probleem i.v.m. Bowen's onderzoek  |
| 32   | 8   | Addendum: een en ander over enkele statistische toetsen  |
| 37   | 9   | Literatuur   |
| 38   | 10  | Summary  |

0

## Inleiding

Dit rapport houdt zich voornamelijk bezig met de statistische aspecten van de fundering van een door E.G. Bowen uitgesproken stelling, dat meteorieten stof een belangrijke factor zou zijn bij het stimuleren van regenval [1]. Deze stelling vloeit voort uit een studie van de statistiek van de dagelijkse regenval op een aantal Australische stations; de statiek vertoont eigenaardigheden, die niet of zeer moeilijk met "gewone" meteorologische processen te begrijpen zouden zijn.

Ofschoon - althans voor zover wij Bowen in zijn artikel kunnen volgen - de statistische aanpak van het probleem niet tenvolle ontbreekt, wordt nergens een werkelijk afdoende statistische toets toegepast en mist de gehele beschouwing de hoogste noodzakelijke zelf-kritiek.

Het probleem vroeg onze aandacht in het bijzonder om de statistische zijde. Tegelijk bleek onze kritische studie van nut te zijn ook bij eigen werk, waarover verderop nog gesproken zal worden.

Terwijl wij het artikel van Bowen onder handen namen, verschenen van de hand van W.C. Swinbank en D.F. Martijn al spoedig in hetzelfde tijdschrift twee zeer kritische commentaren, waarvan de inhoud in 3. kort zal worden weergegeven. J. Neumann (Israël), met wie ik mijn eigen statistische bezwaren te Rome, tijdens het U.G.G.I.-Congres 1954, besprak, publiceerde inmiddels eveneens eigen kritiek. Verder halen wij nog enkele bezwaren aan uit het jaarrapport 1953/54 van de Commonwealth Scientific and Industrial Research Organization Section of Meteorological Physics, terwijl wij van M.B. Oliver and V.J. Oliver een - overigens weinig zeggende - bevestiging van Bowen's theorie citeren. Dat eigenaardigheden analoog aan die, welke Bowen tot zijn gedurfde stelling brachten, meer de aandacht trokken, laten wij zien door enige passages uit "Der Jahresgang der Niederschlagswahrscheinlichkeit auf dem Somblick" door F. Steinhauser aan te halen. In 6. behandelen wij Nederlands materiaal terwijl wij in 7. proberen te komen tot de formulering van een probleem waarvan de oplossing het "beste" antwoord op de aanpak van Bowen zou kunnen vormen. Tot slot vertellen wij in 8. een en ander over enkele statistische toetsen, die bij de kritische beschouwing van Bowen's artikel ter sprake komen.

1

## Korte inhoud van het artikel van Bowen

In zijn inleidingen vertelt Bowen, dat zeer veel aandacht in de meteorologische literatuur besteed is aan periodiciteiten in het weer en

en aan singulariteiten <sup>1)</sup> in temperatuur of druk. Een analyse van de dagelijkse regenvalcijfers werd tot nu toe niet dikwijls ter hand genomen. Hij ontdekte in verschillende delen van de wereld een verrassend gedrag van deze getallen en probeerde hiervoor een verklaring te vinden.

Het is een feit (E.C. <sup>2)</sup>: is dat zo? staat dit "feit" zo statistisch zeker vast?), dat de hoeveelheden neerslag per weer of maand voor één station zelden afwijkingen van de gemiddelde waarde vertonen, die groter zijn dan louter statistische gronden verwacht mogen worden. Maar als men de veeljarige dagsommen beschouwt (d.i. de totale hoeveelheid neerslag op alle 1 Jan.-dagen; die op alle 2 Jan.-dagen, enz., het hele jaar rond) dan is het "apparent that in some localities there is a marked tendency for heavy falls of rain to occur on certain days rather than on other, and for this pattern to be repeated year after year".

We kunnen deze stelling de quintessence van het artikel van Bowen noemen. De schrijver illustreert zijn artikel met een 4-tal figuren, die wij hier zo goed mogelijk weergeven (fig. 1, 2, 3, 4). De directe gegevens staan ons niet ter beschikking, doch dit doet er niet veel toe.

In elk der figuren 1a en 1b ziet men dus  $31 + 5 = 36$  punten (1 Jan. t/m 5 Feb.) In beide stellen zij de over 43 jaren gesommeerde daghoeveelheden H voor, in fig. 1a: 1859 - 1901, in fig. 1b: 1902 - 1944. Wij hebben opzettelijk zelf het gemiddelde van beide perioden gemaakt, zie fig. 1c.

Bowen wijst op de algemene toeneming van H door de maand, overeenkomstig de jaarlijkse gang. Opvallend zijn de pieken op 12 Jan. en 22 Jan. in fig. 1a en <sup>op</sup> 12, 13, 23 en 31 Jan. in fig. 1b. In fig. 1a bedraagt de hoogte van elk der twee pieken ongeveer het dubbele van het gemiddelde omstreeks de datum van de betreffende piek, terwijl het verschil tussen de piekwaarde en het gemiddelde ongeveer vier keren "the standard deviation of the remainder" is.

Verder wijst Bowen vooral op de "near coincidence in time" van de eerste twee pieken (12 en 13; 22 en 23 Jan.). Dit was zulk een onverwacht verschijnsel, dat het stimuleerde tot een onderzoek van de regen-cijfers van andere stations op corresponderende breedten. Toen ontdekte Bowen, dat vele stations, over een uitgestrekt gebied, pieken van een dergelijke grootte op bijna dezelfde dagen vertoonden (7 stations op het Zuidelijk halfrond over  $180^{\circ}$  lengte van Zuid Afrika tot Nieuw Zeeland;

---

1) Onderstrepingen door mij.

2) E.C.: eigen commentaar.

basistijdvakken van ongeveer 50 jaren). Men vindt een en ander in tabel 1.

- (d) De data, waarop de veeljarige som het gemiddelde met 50% overtreft, worden speciaal genoemd.

Enkele stations op de Britse Eilanden tonen aan, dat er een

- (e) kleinere variabiliteit is dan op de Australische stations. Pieken, die niet  
(f) als toevallig beschouwd mogen worden, komen er niet voor. Bij onderzoek bleek, dat op die stations, waarop toch pieken leken voor te komen, deze  
(g) het gevolg waren van een betrekkelijk klein aantal dagen met een exceptioneel zware regen en niet van een grotere frequentie van regen op die dagen. Om dit duidelijker te laten uitkomen tekende Bowen fig. 2. Verticaal is uitgezet het aantal keren, dat elk der data 1 Jan. t/m 5 Feb. in de periode 1919 - 1949 een "heavy-rain day" was (dagsom van meer dan  $2\frac{1}{2}$  inch = 61 mm) (E.C. wolk station? Een station of meerdere stations tezamen?).

- (h) De curve vertoont een maximum op 2 Januari, welke datum wij niet als piek in de figuren 1a en 1b ontmoeten. De toppen bij 11 (13), 23 Januari en 1 Feb. liggen precies of zo goed als precies op de piekdagen der figuren 1a en 1b. Zie ook het overzicht in tabel 1.

In figuur 3 brengt Bowen al deze piek-data bij elkaar om een

- (i) "distribution of rainfall peaks" te verkrijgen. Hij constateert "a remarkable grouping of the peaks" ..... "around January 13, 23 and 31 with a spread of  $\pm 2$  days about these dates".

Bovendien stelt hij vast: "similar results are obtained in other months of the year" / E.C.: jammer, dat deze ook niet in kleine tabellen samengenomen werden /.

Deze eigenaardigheden aldus geconstateerd hebbende (en laten wij onderstellen dat zij statistisch significant zijn), rijst de vraag of dit verschijnsel toegeschreven kan worden aan klimatologische factoren alleen of aan deze tezamen met andere invloeden.

Bowen merkt dan op, dat de jaren, waarin zware regens bijdroegen tot de piek van 14 Januari, bijv. te Alice Springs, geheel andere zijn dan die, waarin te Christchurch de piek op dezelfde dag ontstond. Daar komt bij, dat het moeilijk te begrijpen is, dat door klimatologische omstandigheden zware regens op het zuidelijk halfrond op ongeveer dezelfde data zouden vallen als op het noordelijk halfrond. Van het grote aantal buitenlandse verschijnselen, die zich jaar voor jaar op dezelfde data afspelen, komt alleen dat der meteorieten-zwermen in aanmerking.

- (j) Bowen merkt verder op, dat de meteor-deeltjes, die men met het blote oog ergens aan de hemel zien kan (gemiddeld aantal 10 à 12 per uur) meer of minder toevallig over de wereldruimte verdeeld zijn.



(k) Nu en dan echter passeert de aarde door brede meteoren-stromen, die een elliptische baan om de zon volgen en kan het aantal zichtbare meteoren toenemen tot wel 50 à 100 per uur. De waarnemers op het noordelijk halfrond kennen tien nachtelijke stromen van deze soort. Zij komen elk jaar op dezelfde data voor en op hetzelfde uur. De data der meteoren-stromen, die het dichtst bij die der regenpieken liggen zijn:  
Geminiden: 13 - 14 Dec.; Ursiden: 22 Dec. en Quadrantiden: 3 Jan.  
Het lijkt wel of de regenpieken ongeveer een maand na de meteorenzwermen vallen. Bowen weet wel, dat er ook meteoren-stromen zijn gedurende Mei, Juni en Juli, maar deze volgen elkaar zo snel op, dat zij haast niet te onderscheiden zijn. Tenslotte meent Bowen dus te mogen concluderen, dat er op zekere stations en op zekere tijden in het jaar een grote waarschijnlijkheid bestaat voor een regenpiek 29 à 30 dagen, nadat de aarde in een hoofd meteoren-stroom is binnengetreden.

Daarna tracht Bowen het fysische proces, volgens hetwelk deze meteorieten-zwermen de wereldregenval zouden beïnvloeden, te beschrijven. Natuurlijk moet men numerieke gegevens omtrent de gemiddelde doorsnede der deeltjes en hun bewegingssnelheid kennen. Tenslotte concludeert Bowen, dat meteorieten-stof in voldoende hoeveelheden aanwezig is om de regenval in de lage atmosferische niveaus te beïnvloeden en dat het tijdstip van binnenvallen zo gelegen is, dat de waargenomen tijdsafstand tussen de meteorieten-zwermen en de regenvalpieken verklaard kan worden.

Deze fysische finesses in Bowen's artikel laten wij echter buiten beschouwing.

Uit zijn afsluitende paragraaf citeren wij nog.

- a) De hoeveelheid neerslag wordt beïnvloed door de voorziening met regen-vormende kernen in de atmosfeer. Wanneer kernen van aardse oorsprong afwezig zijn, kan kennis omtrent een ophanden zijnd binnenvallen van meteorietenstof van belang zijn bij overwegingen of extremen regenhoeveelheden onder gegeven klimatologische omstandigheden te verwachten zijn.
- b) Het onderzoek kan resultaten afwerpen die van betekenis zijn voor het vraagstuk van het maken van regen. Wanneer zij bevestigd worden, betekent een en ander, dat de aanwezigheid van betrekkelijk kleine concentraties van deeltjes, die uit de bovenste lagen van de atmosfeer binnenvallen, tot gevolg hebben kan, dat de regenval op bepaalde dagen ongeveer de dubbele is van die op vorige dagen. Men

mag misschien concluderen, althans voor zekere gedeelten van de aarde, dat de potentiële toeneming in de regenval, bewerkstelligd volgens kunstmatige methodes, wellicht de tot nu toe geschatte 10 à 15% overtreft. (E.C.: bedoelt de auteur, dat door menselijk toedoen - als wij voor die "stof" zorgen - de hoeveelheid neerslag de dubbele van die op de vorige dag gemaakt kan worden? Maar hoeveel "stof" dan? En hoe deze in de atmosfeer te verspreiden?).

2

## Onze statistische kritiek

### 2.1 Range-toets en "constellatie-toets"

Bowen beweert, dat het verschil tussen de betreffende piekwaarde en de trend-waarde op de datum van de piek ongeveer vier keren "the standard deviation of the remainder" bedraagt. Wij vragen ons af, of dit juist is. Het lijkt ons zeer veel. En waarom "the remainder".

We hebben getracht Bowen na te rekenen en moesten hiervoor de 43-jarige dagsommen, zo goed mogelijk, van zijn prentje fig. 1a aflezen. Dit lukte voor 34 van de 36 dagen (1 Jan. t/m 5 Feb.; de waarden op 19 en 21 Feb. konden wij niet aflezen). Deze 34 43-j.dagsommen hebben een gemiddelde  $g = 4.81$  inch en een standaarddeviatie  $s = 3.39$  inch <sup>1)</sup>.

---

1) a) Eigenlijk behoort dit a.v. te gebeuren: welke standaarddeviatie  $\sigma_1$ , hebben de 1 Jan.-hoeveelheden? Welke  $\sigma_2$  de 2 Jan.-hoeveelheden, t/m welke  $\sigma_{36}$  de 5 Feb.-hoeveelheden? Onderzoek of deze 36  $\sigma$ 's significant verschillen. Zo neen, neem het gemiddelde  $\bar{\sigma}$ . Bijgevolg spreiden de 43-jarige 1 Jan.-dagsommen met een standaard deviatie  $= \bar{\sigma}\sqrt{43}$ ; idem de 43 j. 2 Jan.-dagsommen, etc. Al deze 36 frequentieverdelingen zijn zo goed als normaal (omdat 43 een voldoende groot getal is). Welke universumwaarden  $\mu_1, \mu_2, \dots, \mu_{36}$  hebben de gemiddelden van deze 36 freq. verdelingen? Is er een trend van  $\mu_1 \rightarrow \mu_{36}$ ? Wellicht wel (jaarl. gang). Zo ja, dan de beste trend berekenen. Het getal 3.38 in de tekst is een magere schatting van deze  $\bar{\sigma}\sqrt{43}$ . "Mager"? Omdat (zie ons rapport W.R. 55-004) de standaard deviatie der 36 43 j.-dagsommen (ook als wij elk dezer 36 getallen eerst verminderen met de bijbehorende trendwaarde  $\mu$ ) niet identiek gelijk is (ook niet als wij de verwachtingswaarde beschouwen) met bedoelde  $\bar{\sigma}\sqrt{43}$  en wel vanwege de interne correlatie. In genoemd rapport werd bewezen dat de verwachtingswaarde van de variantie tussen de 31 daghoeveelheden in bijv. Januari niet identiek is met de datumvariantie  $\sigma^2$  (ook al denken wij deze datum-onafhankelijk), maar gelijke  $\sigma^2 \left\{ 1 - \frac{1}{31} \sum_{i=1}^{31} \rho_i / (1 - \rho_i) \right\}$  is, wanneer de interne correlatie-coëfficiënt van de orde  $i$  ( $= 1, 2, \dots$ ) gegeven wordt door  $\rho_i = \rho_1^i$ . Men ziet, dat de factor  $\left\{ \dots \right\} = 1$  als  $\rho = 0$ ; echter  $< 1$  als  $\rho > 0$ , doch nog ong. 0.9 als  $\rho_1 = 0.6$  - zeer groot -.

(Zie vervolg op blz. 7)

Als wij de bewuste pieken (12 en 22 Januari met resp. 13.2 en 11.7 inch) niet meenemen, wordt het gemiddelde der overige 32 stuks 43 j.dagsommen 4.65 inch, met een standaarddeviatie 2.84 inch. Men ziet hoe sterk de standaarddeviatie verandert met het al of niet meenemen der twee pieken. Het lijkt ons niet geoorloofd reeds a priori deze pieken buiten beschouwing te laten, zoals Bowen blijkbaar doet. Nu loopt de trend (op het oog getrokken; men zou haar ook kunnen berekenen, maar dat zou zeer weinig veranderen aan onze conclusies) omstreeks 12 Jan. bij 4 en omstreeks 22 Jan. bij 5 inch. Zodat wij moeten beschouwen:

$$\text{op 12 Jan. het quotiënt } q = \frac{13.2 - 4.0}{3.38} = 2.72 \text{ (doch Bowen: } \frac{13.2 - 4.0}{2.84} =$$

3.24; is dat nu het getal 4, waarvan hij spreekt?) en

$$\text{op 22 Jan. } q = \frac{11.7 - 5.0}{3.38} = 1.98 \text{ (doch Bowen: } \frac{11.7 - 5.0}{2.84} = 2.36)$$

Wij zien, dat Bowen tot geflatteerd hoge q's komt. Wat nu met deze q's te doen? Welke toets te gebruiken om te onderzoeken of deze q's "boventoevallig" groot zijn?

Ons lijken hier de twee volgende toetsen voor de hand te liggen:

### 2.1.1

#### Range-toets

De range of de spreidingsbreedte, d.i. het verschil tussen de kleinste en grootste waarde in een steekproef, gehoorzaamt aan een waarschijnlijkheidsverdeling, die volkomen bekend is als de steekproef (N-elementen) genomen wordt uit een normale verdeling met bekende  $\mu$  en  $\sigma$ . Bij Bowen is in fig. 1 a de spreidingsbreedte  $R = 13.2 - 0.6 = 12.6$ . Wij moeten beschouwen  $\frac{R}{s} = \frac{12.6}{3.38} = 3.73$  (Bowen zou beschouwen:  $\frac{12.6}{2.84} = 4.44$ )

---

(Vervolg noot 1) van blz. 6)

- b) Men kan zich afvragen of niet eerst elk der 34 43 j.-dagsommen met de trend-waarde moet worden verminderd? Ja, echter welke trend? Hoe de trend-curve te trekken? Lineair, kwadratisch? Wij trokken de rechte door 4.0 inch op 12 Jan. en 5.0 inch op 22 Jan. Na correctie voor de trend hebben de 34 punten (waaronder de piekwaarden zelf) een standaarddeviatie 2.42 inch, zodat de piekwaarden quotiënten opleveren:  $(3.2 - 4.0) : 2.42 = 3.8$  resp.  $(11.7 - 5.0) : 2.42 = 2.8$ . Als we bij de berekening van de standaarddeviatie de twee pieken niet meenemen, dan is de standaarddeviatie: 1.42 ( $\ll 2.42$ ) en zijn de twee quotiënten  $(13.2 - 4.0) : 1.42 = 6.5$  en  $(11.7 - 5.0) : 1.42 = 4.8$ . De laatste quotiënten liggen zo veel boven Bowen's vier, dat wij denken, dat Bowen bij de berekening van de standaarddeviatie niet eerst voor de trend gecorrigeerd heeft.

Nu is aan een zeer essentiële voorwaarde niet voldaan, t.w. de  $N = 36$  elementen in deze steekproef mogen niet gepersisteerd (ge-correleerd) zijn. Anders gezegd: vóór wij de tabel voor de toetsing van spreidingsbreedten mogen gebruiken, moeten wij  $N$  door het equivalente effectieve aantal  $N_e$  onafhankelijke elementen vervangen. Zeker zal  $N_e < N$  zijn, doch hoeveel kleiner? Dit hangt met de sterkte van de persistentie samen (waarover wij uit Bowen's artikel niets te weten komen), dus met de mate waarin de hoeveelheid neerslag van heden, met die van  $i$  dagen ( $i = 1, 2, \dots$ ) later gecorreleerd is. Daarom proberen wij maar wat; als  $N_e = 15$  is, is de kans dat  $R/s$  een waarde 3.73 zal bereiken of overschrijden gelijk (Bowen: 0.187). Wanneer  $N_e = 20$  is, is diezelfde kans 0.476 (Bowen: 0.277).

Hoe dichter  $N_e$  bij  $N$  ligt (d.i. hoe geringer de persistentie), hoe meer komt bedoelde kans boven de drempel 0.05 uit, d.w.z. hoe dringender mogen wij concluderen, dat de spreidingsbreedte door louter toeval zo groot is, ofwel, dat er niets bijzonders achter de piek (de spreidingsbreedte hangt immers met de piek samen) schuilt. In het bovenstaande geval: bij zware persistentie ( $N = 36$  en  $N_e = 15$ ) en zo gunstig mogelijk rekenende (door een te kleine  $s$  te nemen, door namelijk de pieken bij de berekening van  $s$  niet mee te nemen) komen wij tot de laagste overschrijdingskans; maar die zelfs ligt ver boven 0.05 ( $18.7 > 5$ ). Conclusie: volstrekt geen significantie <sup>1)</sup>.

1)

Men kan redeneren, dat aldus slechts beweerd is, dat de piek als toevallig beschouwd kan worden, doch dat niet "bewezen" zou zijn, dat ze louter toevallig is, d.w.z., dat de piek niet toch aan een bepaalde aanwijsbare oorzaak kan worden toegeschreven. Men kan zelfs redeneren, dat de mogelijkheid, om het meteorietenstof als oorzaak te beschouwen nog altijd open blijft, en wel, in de grond van de zaak, omdat we niet meer dan een waarschijnlijkheidsoordeel uitspreken, dat direct met die 5%-drempel samenhangt. Inderdaad, de statisticus pretendeert dan ook niet meer dan waarschijnlijkheidsuitspraken te doen. Doch hij gebruikt de woorden "waarschijnlijk", "mogelijk" enz. niet zonder meer, maar tezamen met numerieke kansen. Het gaat niet om de "mogelijkheid" - zie boven - , maar om de grootte van die mogelijkheid? 1%? 80%? 1‰? Bovendien wat is "bewijzen" eigenlijk? Wat is "toeval" eigenlijk? Het "toeval" is identiek met een constellatie van fysische oorzaken, zo talrijk, dat zij tezamen consequenties hebben, die aan de toevalswetten der kansrekening gehoorzamen. Hoe "bewijst" men, dat A causaal met B "samenhangt"? Door een redenering alleen? Als men ook de uitkomsten van metingen wil laten meespreken in het antwoord moet men weer statistische toetsen e.d. toepassen.

Voorbeeld: Als iemand maar één keer werpen mag met 10 dobbelstenen en hij krijgt daarbij 10 zessen, wat zal hij dan besluiten? 1° Allemaal "eerlijk" of "zuiver"? 2° Tenminste één "onzuiver"? Dat alle van 10 dobbelstenen met 6 bovenkomen heeft een kans  $(1/6)^{10} = 0,006$ . Hij kan dus rustig het 2° besluiten, want in één op de 167 keren, dat hij zulk een conclusie trekt, zou die conclusie toch met de waarheid in strijd zijn.

(Zie vervolg op blz. 9)

2.1.2

"Constellatie-toets"

Een piek kan gedefinieerd worden als een element  $\geq \mu + 1.98\sigma$ , met voldoende grote  $q$ . Laten wij in het oog nemen de  $q = 2.72$  resp.  $1.98$  (12 en 22 Jan.) De tabel van de normale waarschijnlijkheidsverdeling levert een overschrijdingskans  $0.0033$  bij  $q = 2.72$  en  $0.0238$  bij  $1.98$ . Vraag: bereken de kans  $P(\geq 2)$  op 2 of meer elementen per 36 stuks, die  $1.98$  of meer keren  $\sigma$  hoger dan  $\mu$  liggen. Gedacht is aan het nemen van een aselechte steekproef van  $N = 36$  stuks elementen uit een normale verdeling

(1)  $\mu, \sigma$ . Antwoord: de kans op  $i$  successen in  $N_e$  experimenten, als de succeskans per experiment  $0.0238$  is, is:

$$P(i) = \binom{N_e}{i} 0.0238^i (1 - 0.0238)^{N_e - i}$$

Dus  $P(\geq 2) = 1 - (P_0 + P_1)$ . Weer komen wij voor de moeilijkheid: hoe groot is  $N_e$ ?

Bij zware persistentie is het effectieve aantal dagen bijvoorbeeld  $N_e = 15$  ( $N = 36$ ) en is  $P(\geq 2) = 1 - (P_0 + P_1) = 1 - (0.697 + 0.255) = 0.048$ ; bij minder zware persistentie:  $N_e = 20$  en  $P(\geq 2) = 1 - (0.618 + 0.295) = 0.087$  en bij vrijwel geen persistentie:  $N_e = 30$  en  $P(\geq 2) = 1 - (0.486 + 0.356) = 0.158$ . Als persistentie volledig afwezig is, is  $N_e = N = 36$  en  $P(\geq 2) = 1 - (0.420 + 0.358) = 0.222$ . Conclusie: als er geen persistentie is (en het artikel geeft de indruk dat Bowen met persistentie geen rekening houdt) is er 22 % kans op het krijgen van twee pieken wanneer men geheel willekeurig 36 elementen neemt uit een normaal verdeeld universum. (een piek gedefinieerd als  $\geq \mu + 1.98\sigma$ ) Is er wel

---

(Vervolg noot 1) van blz. 8)

Voor hen, die het "statistisch denken" niet gewend zijn, voegen wij nog het volgende toe: men stelt de z.g. "nulhypothese"  $H_0$ ; deze is hier, dat zowel de 43 jarige 1-, als de 43 j. 2-, als de 43 j. 3-, enz. Jan. dagsom 1859-1901 een element is uit een en dezelfde normale verdeling. Men voert een "toetsingsgrootheid"  $T$  in, d.w.z. een grootheid, die een functie is van de elementen van de steekproef; bij ons is  $T$  de "spreidingsbreedte"  $R$  in de steekproef der 36 elementen, t.w. de 43 j. 1 Jan.-som, de 43 j. 2 Jan.-som, etc. Voor een andere groep van 43 jaren zal  $R$  een andere waarde hebben; dus is  $R$  een stochastische variabele. Van deze  $T$  moet de verdelingsfunctie, als  $H_0$  waar is, bekend zijn.  $T$  doorloopt een gebied  $\Omega$ ; hierin zonderen wij een kritieke zone af, zodanig, dat de kans, dat  $T$  tot  $\omega$  behoort, als  $H_0$  waar is, gelijk  $\gamma$  is. Deze  $\gamma$  heet de "onbetrouwbaarheids-drempel". Meestal neemt men  $\gamma = 0.05$ . In de steekproef berekent men  $T_s$ ; als  $T_s$  tot  $\omega$  behoort verwerpt men  $H_0$  (bij ons viel  $T$  buiten  $\omega$  en  $H_0$  werd niet verworpen). Zulk een uitspraak heeft geen absolute zin. Ook al is  $H_0$  juist, dan is er een kans  $\gamma$ , dat  $T$  in valt, zodat men dan  $H_0$  onjuist verwerpt. Dit is een fout van de eerste soort, die men klein (dus 0.05) wenst. Voor verdere finesses zij verwezen naar de uitgebreide literatuur (zie o.a. K.N.M.I.-rapport R III 120 1953).

persistentie dan neemt deze kans af met toenemende persistentie, doch hij is eerst 5 % (de onbetrouwbaarheidsdrempel) bij een persistentiecoëfficiënt van ongeveer 0,4 (zeer grote persistentie!), waarbij het schijnbare aantal  $N = 36$  equivalent is met een effectief aantal  $N_e = 15$ .

Met de  $q$ -waarden van Bowen, t.w. 3.24 en 2.36 werkende, moeten wij, met de in de tabel voor de gestandaardiseerde normale verdeling bij 2.36 afgelezen kans 0.0091, berekenen  $\binom{N_e}{i} 0,0091^i (1 - 0,0091)^{N_e - i}$ , voor  $i = 0$  en  $1$  en voor  $N_e = 15$  20 30 en 36. Er komt: bij  $N_e = 15$  een  $P(\geq 2) = 0,0081$ ; bij  $N_e = 20$   $P = 0,0137$ ; bij  $N_e = 30$   $P = 0,0310$ ; bij  $N_e = 36$   $P = 0,0163$ . Natuurlijk liggen deze  $P$ 's heneden de overeenkomstige voor het geval dat van de betere  $q = 1,98$  uitgegaan wordt. Bowen zou hierdoor foutief tot significantie kunnen besluiten. Op drie zaken dient nog de aandacht gevestigd te worden:

a) Wat is hier die normale verdeling? Antw.: de verdeling der 43 jarige dagsommen. Dit vanwege de centrale limietstelling, welke zegt, dat een lineaire combinatie van  $n$  grootheden, die elk aan een eigen waarschijnlijkheidsverdeling gehoorzamen, een kansverdeling heeft, welke meer en meer tot een normale nadert, naarmate  $n$  toeneemt en dit ongeacht de preciese vorm der initiaalverdelingen. In ons geval is  $n = 43$ , terwijl de initiaalverdeling die der dagsommen neerslag is, welke - gelijk bekend is - beslist geen normale is. 1)

b) Wij hebben in ons rapport R III 118, (1953) bewezen, dat het met  $N = 36$  equivalente effectieve aantal  $N_e$  gelijk is aan 36 : w met  $w = 1 + 2 \left[ \frac{35}{36} \rho_1 + \frac{34}{36} \rho_2 + \dots + \frac{1}{36} \rho_{35} \right]$ , waarin  $\rho_1, \rho_2, \dots$  de persistentiecoëfficiënten van de 1ste, 2de ..... orde zijn. Dikwijls is

---

1) Toelichting: ook al is de frequentieverdeling  $\varphi(h)dh$  der dagsommen neerslag  $h$  zeer scheef - in Nederland is een goede benadering  $\textcircled{H} (\geq h) = \int_0^h \varphi(k)dk = \exp.[-a\{(h+b)^c - b^c\}]$  met constanten  $a, b$  en  $c$  - toch is de frequentieverdeling  $\varphi(H)dH$  bijv. der 1 Jan.-dagsommen, maar gesommeerd over 43 jaren, zo goed als normaal, d.w.z.  $\varphi(H)dH = (\sigma\sqrt{2\pi})^{-1} \exp.[-(H-\bar{H})^2/2\sigma^2]$  met  $\bar{H} = \text{gemiddelde} = \int_0^H H\varphi dH = 43 \bar{h}$ , als  $\bar{h} = \text{gem. 1 Jan. dagsom} = \int_0^h h\varphi dh$  (de  $\sigma$  hangt samen met de  $a, b, c$  en 43)

in goede benadering  $\rho_i = \rho_1^i$  (zie rapport W.R. 55 - 004; 1955). In dat geval laat zich  $w$  omwerken tot

$$1 + \frac{2}{N} \left[ -N + \frac{N - (N+1) \rho_1 + \rho_1^{N+1}}{(1 - \rho_1)^2} \right]. \text{ Als } \rho_1^N \ll 1 \text{ geldt}$$

$$(3) \quad w \approx 1 + \frac{2}{N} \left[ -N + \frac{N - (N+1) \rho_1}{(1 - \rho_1)^2} \right] \approx 1 + 2 \frac{\rho_1 - \rho_1^2}{(1 - \rho_1)^2} = \frac{1 + \rho_1}{1 - \rho_1}$$

Het volgende staatje geeft enig inzicht in het verband tussen  $\rho_1$ ,  $w$  en  $N_e = 36 : w$

| $\rho_1$ | $w$ | $N_e$ |
|----------|-----|-------|
| 1.0      | -   | -     |
| 0.6      | 4   | 9     |
| 0.5      | 3   | 12    |
| 0.4      | 2.3 | 15.5  |
| 0.3      | 1.9 | 19.4  |
| 0.2      | 1.5 | 24    |
| 0.1      | 1.2 | 35    |
| 0.0      | 1.0 | 36    |

c) Wij zijn niet geheel correct geweest bij de berekening van  $s$ , als variantie tussen de 36 43 j.-dagsommen. In ons rapport W.R. 55 - 004; 1955 toonden wij aan, dat de gemiddelde waarde van de variantie tussen de 36 43 j.-dagsommen geschreven kan worden in de vorm

$$(4) \quad \overline{t^2} = 43 \sigma^2 \left[ 1 - \frac{2}{35 \cdot 36} (35 P_1 + 34 P_2 + \dots + P_{35}) \right]. \text{ Let wel:}$$

$P_1$  = universumwaarde van de correlatiecoëfficiënt tussen de 43 j.-dagsommen op 1 en 2 Jan. Deze is gelijk gedacht aan die voor 2 en 3 Jan.

en die voor 3 en 4 Jan. etc.  $P_2$  = idem, doch voor 1 op 3 Jan. enz.;

$P_3$  = idem, doch voor 1 op 4 Jan. enz.

Wij bewezen in genoemd rapport:  $P_1 = \rho_1, P_2 = \rho_2, \dots$ , met  $\rho_i$  als persistentiecoëfficiënt (autocorrelatiecoëfficiënt) van de  $i$ -de orde (d.i. tussen de hoeveelheid op dag  $k$  en die op dag  $k + i$ ). De  $\sigma$  is de datumvariantie, d.i. de variantie in alle 1 Jan.-dagsommen, die gelijk gedacht is aan die in alle 2 Jan.-dagsommen. etc. Wij kennen de  $\rho$ 's niet voor Sydney, maar laten wij eens onderstellen, dat de persistentie zwaar

is en wel  $\rho_1 = 0.50$  en  $\rho_i = \rho_1^i$ . Welke  $t^2$ ? Hier op basis van twee 43 j.-periodes:  $t^2 = \frac{1}{2}(3.38^2 + 3.05^2) = 10.4$ . De onder 2.1.2 bedoelde standaarddeviatie  $\sigma^2$  mag niet worden vereenzelvigd met de in de formule voor  $t^2$  bedoelde  $\sigma$ . Het verband is  $\sigma^2 = \sigma \sqrt{43}$ . Wij behoren dus te rekenen niet met  $t^2$  maar met  $43 \sigma^2$ , zijnde een factor  $1 - \frac{2}{35.35}$

( $35 \rho_1 + \dots$ ) (dus  $< 1$ ) maal zo klein <sup>1)</sup>. Hoe klein? Er komt met  $\rho_1 = 0.50$   $\rho_2 = 0.25$   $\rho_3 = 0.12$   $\dots$   $1 - \frac{2}{35.36}$  ( $35 \rho_1 + \dots \rho_{35}$ )  $\approx 1 - 0.042 = 0.96$ . Wij werken dus met een 4 % te kleine variantie of een 2 % te kleine standaarddeviatie.

Met afnemende  $\rho_1$  (afnemende persistentie) komt de factor  $\left[ \frac{N}{N-1} \right]$  in (4) dichter bij 1 ofwel is er minder relatief verschil tussen  $t^2$  en  $43 \sigma^2$ . Doordat wij rekenden met  $t^2$  en niet met  $43 \sigma^2$  ( $> t^2$ ) komen wij "gemakkelijker" tot een significantie van de piek. Overigens is dit effect van zeer weinig betekenis, omdat  $\left[ \frac{N}{N-1} \right] \approx 1$ .

Nota bene: als  $t^2 \approx 43 \sigma^2$  volgt uit  $t^2 = 10.33$  inch<sup>2</sup> een  $\sigma \approx 12.9$  mm. Bij ons (Hoofddorp)  $\sigma = 3 \text{ à } 4$  mm <sup>2)</sup>. Duidelijk ligt de gemiddelde dagsom te Sydney in Januari op een hoger niveau dan te Hoofddorp (28 tegen 1.7 mm) en de datumvariantie is vele malen groter dan die te Hoofddorp (12.9 en 3 à 4 mm).

Voor de volledigheid passen wij al het bovenuitgevoerde ook op fig. 1b (de andere 43 j.-periode) toe.

Er zijn 4 pieken (12 Jan.: 11.2; 13 Jan.: 11.4; 23 Jan.: 8.7; 31 Jan.: 10.0; gemiddelde der 36 dagen: 4.50; inches). De variantie tussen alle 36 dagsommen is 3.05 inch; tussen de 32 dagsommen (4 pieken niet meegenomen): 2.33 inch. De trend is vrijwel horizontaal; hoogte geschat: 4.50 inch. De q-waarden op de piekdagen zijn derhalve  $q = (11.2 - 4.50) : 3.05 = 2.20$ ;  $q_3 = (8.7 - 4.50) : 3.05 = 1.38$  en  $q_4 = (10.0 - 4.50) : 3.05 = 1.80$ . Bowen, rekenende met de standaarddeviatie der niet-piekdagen,

1) Voor de grote N en kleine  $\rho_1$  (mits  $\rho_1^N \ll 1$ ) is  $1 - \frac{2}{N(N-1)} \left[ N \rho_1 + (N-1) \rho_2 + \dots + \rho_N \right]$ , mits  $\rho_i = \rho_1^i$ , ongeveer gelijk aan  $1 - \frac{2}{N} \frac{\rho_1}{1 - \rho_1}$ ; als  $\rho_1 = \frac{1}{2}$  en N = 36 is dat  $1 - \frac{1}{18} = 1 - 0.056 = 0.94$  (in tekst 0.96)

2) Zie Rapport W.R. 55 - 004; 1955.

$q_2 = (11.4 - 4.50) : 3.05 = 2.26$ ;



zou komen tot resp. 2.88 2.97 1.80 en 2.36. De range is  $R = 11.4 - 0.6 = 10.8$  en dus  $R : s = 10.8 : 3.05 = 3.54$  (Bowen:  $10.8 : 2.33 = 4.64$ ). Bij zware persistentie is  $N_e = 15$  en  $P (\geq R) = 0.455$  en bij minder zware persistentie  $P (\geq R) = 0.587$ . (Bowen: resp.: 0.072 en 0.114) Conclusie: ook in de tweede 43-jarige reeks en bij zware persistentie en rekenende met een te kleine  $s$  (door de  $s$  op de niet-piekdagen te baseren; d.w.z. onder omstandigheden, die het vinden van significantie begunstigen) komt er een overschrijdingskans ver boven 5 %, of wel de spreidingsbreedte mag als toevallig beschouwd worden en dus mag de aan deze spreidingsbreedte schuldige piek aan het toeval toegeschreven worden.

Wij kunnen ook naar de kans op een constellatie van 36 waarden met daarin 4 pieken. (alle  $1.38 \sigma$  of meer boven  $\mu$ ) vragen, gelijk onder 2.1.2 voor 1899 - 1901 gebeurde. Dan is  $P (\geq 4) = 1 - (P_0 + P_1 + P_2 + P_3)$  met  $P_i = \binom{N_e}{i} 0.084^i 0.916^{N_e - i}$ , waarin 0.084 afgelezen wordt (behorende bij  $\mu + 1.38 \sigma$ ) uit de Gauss-tabel. Er komt met  $N_e = 15$   $K = 0.032$ ;  $N_e = 20$   $K = 0.099$ ;  $N_e = 30$   $K = 0.242$ . Conclusie: een frequentie van 4 pieken (gedefinieerd als boven) op  $N = 36$  waarden, is zeker niet "boventoevallig", tenzij de persistentie zeer sterk zou zijn ( $N_e \approx 15$  ofwel  $\rho \approx 0.4$ ). Deze constellatie is wel boventoevallig wanneer men de kans berekent op basis van Bowen's  $q = 1.80$ , die wij te groot achten. Zie schema.

## 2.2 Coïncidentie-toets

Toegegeven moet worden, dat onze kritiek tot nu toe gericht is op de beschouwing bij elk der twee prentjes fig. 1a en 1b afzonderlijk. Wij menen aangetoond te hebben (Neumann gaat aldus ook te werk), dat de constellatie in elk der gekozen perioden geheel toevallig ontstaan kan zijn en menen op grond daarvan gerechtigd te zijn geen bijzondere betekenis te hechten aan "zekere" coïncidentie van pieken. Opzettelijk zeggen we: "zekere", want ook Bowen spreekt van "near coincidence in time" Één (of misschien twee) dag (dagen) verschil in de piekdatum acht hij blijkbaar niet erg. Maar dan wordt de kans op een louter toevallige "coïncidentie" al direct groter! Bovendien, wat wil Bowen onder een piek verstaan? Wanneer men de definitie niet zwaar stelt ( $> \mu + k \sigma$ , met  $k = 1$  of 0.5 of 0.2 etc.), dan zal men zeker meer "coïncidenties" van "pieken" tussen de 2 prentjes vinden, maar ook meer "niet-coïncidenties". In verband daarmee is het vreemd, dat Bowen, verwijzend naar zijn resultaten bij stations elders (in tabel 1 samengenomen), opeens beschouwt die dagen, waarop de hoeveelheid neerslag het gemiddelde met (maar) 50 % of meer

overtreft. In de figuren 1 bedroeg de afwijking toch meer dan 100 % ! Men kan natuurlijk vragen naar een meer directe coincidentie-toets. Zulk een toets moet men zelf ontwikkelen. Wij kennen er geen, die gebruiksklaar ligt. Geen der auteurs, wier artikelen wij verderop noemen, deed het zo. Trouwens de formulering zelf van het probleem brengt nog moeilijkheden. De behandeling daarvan stellen wij liever uit tot 7.

Tenslotte hebben wij nog de in 2.1 genoemde berekening in bijgaand schema overzichtelijk bijeen genomen.

Sydney

RANGE - toets

| 43 j. 1859 - 1901 |             |            |       |                            |       |                  |       |      |      |
|-------------------|-------------|------------|-------|----------------------------|-------|------------------|-------|------|------|
| Goed              |             |            |       | Bowen                      |       |                  |       |      |      |
| $g = 4.81$        |             | $s = 3.38$ |       | $\hat{g} = g(36-2) = 4.65$ |       | $\hat{s} = 2.84$ |       |      |      |
| Pieken            |             |            |       | Pieken                     |       |                  |       |      |      |
|                   | 12 J.       |            | 22 J. |                            | 12 J. |                  | 22 J. |      |      |
| waarde            | 13.2        |            | 11.7  |                            | 13.2  |                  | 11.7  |      |      |
| trend             | 4           |            | 5     |                            | 4     |                  | 4     |      |      |
| q =               | 2.72        |            | 1.98  |                            | 3.24  |                  | 2.36  |      |      |
| RANGE = 12.6      |             |            |       |                            |       |                  |       |      |      |
| $R/s = 3.73$      |             |            |       | 4.14                       |       |                  |       |      |      |
| $N_e = 15$        | $P = 0.345$ |            |       | 0.187                      |       |                  |       |      |      |
| 20                | 0.476       |            |       | 0.277                      |       |                  |       |      |      |
| 43 j. 1902 - 1944 |             |            |       |                            |       |                  |       |      |      |
| Goed              |             |            |       | Bowen                      |       |                  |       |      |      |
| $g = 4.50$        |             | $s = 3.05$ |       | $\hat{g} = 3.77(36-4)$     |       | $\hat{s} = 2.33$ |       |      |      |
| Pieken            |             |            |       | Pieken                     |       |                  |       |      |      |
|                   | 12J.        | 13J.       | 23J.  | 31J.                       |       | 12J.             | 13J.  | 23J. | 31J. |
| piek              | 11.2        | 11.4       | 8.7   | 10.0                       |       | 11.2             | 11.4  | 8.7  | 10.0 |
| trend             | 4.5         | 4.5        | 4.5   | 4.5                        |       | 4.5              | 4.5   | 4.5  | 4.5  |
| q =               | 2.20        | 2.26       | 1.38  | 1.80                       |       | 2.88             | 2.97  | 1.80 | 2.36 |
| RANGE = 10.8      |             |            |       |                            |       |                  |       |      |      |
| $R/s = 3.54$      |             |            |       |                            | 4.67  |                  |       |      |      |
| $N_e = 15$        | $P = 0.455$ |            |       |                            | 0.072 |                  |       |      |      |
| 20                | 0.587       |            |       |                            | 0.114 |                  |       |      |      |

CONSTELLATIE - toets

| 1859 - 1901                                    |  |
|--|--|
| Goed   | Bowen                                    |
| $q = 1.98 \rightarrow K = 0.0238$              | $q = 2.36 \rightarrow K = 0.0091$        |
| 2 pieken op 36 stuks                           |  |
| $P(\geq 2) = 1 - P_0 - P_1$                    |  |
| $P_i = \binom{N_e}{i} 0.0238^i 0.9762^{N_e-i}$ | $\binom{N_e}{i} 0.0091^i 0.9909^{N_e-i}$ |
| $N_e = 15$ $P(\geq 2) = 0.048$                 | 0.0081                                   |
| 20 = 0.087                                     | 0.0137                                   |
| 30 = 0.158                                     | 0.0310                                   |
| 36 = 0.222                                     | 0.0163                                   |
| 1902 - 1944                                    |  |
| Goed   | Bowen                                    |
| $q = 1.38 \rightarrow K = 0.084$               | $q = 1.80 \rightarrow K = 0.036$         |
| 4 pieken op 36 stuks                           |  |
| $P(\geq 4) = 1 - P_0 - P_1 - P_2 - P_3$        | $P(\geq 4) = 1 - P_0 - P_1 - P_2 - P_3$  |
| $P_i = \binom{N_e}{i} 0.084^i 0.916^{N_e-i}$   | $\binom{N_e}{i} 0.036^i 0.964^{N_e-i}$   |
| $N_e = 15$ $P(\geq 4) = 0.032$                 | 0.0002                                   |
| 20 = 0.099                                     | 0.0051                                   |
| 30 = 0.242                                     | 0.0402                                   |

2.3 De  $\chi^2$  - toets

Zie fig. 2. Bowen schijnt speciaal te willen wijzen op het maximum op 2 Jan., dat hij niet vindt in fig. 1. Hier willen wij, statistisch, dieper op ingaan. In de eerste plaats zijn in de figuren 1 en 2 verschillende grootheden voorgesteld n.l.: in fig. 1 de over vele jaren gesommeerde daghoeveelheden per datum en in fig. 2 de kans, dat deze datum een daghoeveelheid boven een gegeven drempel toont <sup>1)</sup>.

1)

Dit is niet helemaal waar, maar daar komt de figuur toch op neer. Als er bijv. 20 keren op 8 Jan. een dagsom  $\geq 2\frac{1}{2}$  inch viel, in 40 jaren, dan komt dat overeen met een relatieve frequentie van  $20 : 40 = 0.50$ . (Intussen is deze frequentie zó groot, dat de figuur welhaast onmogelijk op één enkel station betrekking kan hebben; maar op hoevele dan wel? Bowen vertelt het niet)

Deze grootheden behoeven niet strak gecorreleerd te zijn. In de tweede plaats vragen we ons af; zijn de pieken in fig. 2 statistisch significant? Dit kan met de  $\chi^2$  - toets onderzocht worden. Echter: is er wel of geen trend en zo ja, welke? Als er geen trend is luidt de nulhypothese, dat er een gelijkmatige verdeling van de a's (zie fig.) is over de data <sup>1)</sup>. Als beste schatting voor het gemiddelde  $\mu$  van deze verdeling vinden wij het rekenkundig gemiddelde van alle a's, die wij uit de figuur aflezen:  $\bar{a} = 14.3$ . Dan is  $\chi^2 = \left[ \sum_1^{36} (a_i - \bar{a})^2 \right] : \bar{a} = 63.5$

bij  $36 - 1 = 35$  vrijheidsgraden. Deze  $\chi^2$  is ongeveer normaal verdeeld rondom  $n = 36$  als gemiddelde en met  $2n = 72$  als variantie, zodat  $(63.5 - 36) : \sqrt{72} = 2.07$ ; de overschrijdingskans  $P \approx 4\%$ . <sup>o</sup>. significantie. M.a.w. de nulhypothese wordt verworpen, wij mogen derhalve niet aannemen, dat er geen trend (jaarlijkse gang) zou zijn. Eigenlijk konden wij dat op het oog ook al zien.

We nemen dus wel een trend en corrigeren daarmee eerst de a's. Welke trend? Wij willen niet op het probleem "hoe een trend numeriek te benaderen" ingaan, doch aannemen dat zij lineair is en hebben de trendlijn op het oog getrokken, zie de rechte. Nu wordt

$$\chi^2 = \sum_1^{36} \left[ (a_i - a_{T_i})^2 : a_{T_i} \right], \text{ met } a_{T_i} = \text{trendwaarde op datum } i.$$

Er komt  $\chi^2 = 39.4$  bij  $36 - 2 = 34$  graden van vrijheid; en een overschrijdingskans  $P \approx 67\%$ . In elk geval  $P \gg 5\%$  ofwel er is hoegenaamd geen reden om de pieken "boventoevallig" (d.i. significant) te achten <sup>2)</sup>

1) Dat zou betekenen, dat elke datum dezelfde kans heeft een hoeveelheid  $\geq 2\frac{1}{2}$  inch te ontvangen (over een oneindig groot materiaal zouden de punten in fig. 2 dus alle even hoog (en wel a) boven de tijdas moeten liggen). Als de kans, dat datum i een dagsom  $\geq 2.5$  inch heeft gelijk p gezet wordt (p onafhankelijk van i), dan wordt de kans dat deze datum n zulke dagsommen toont in 40 jaren gegeven door  $K(n) = \binom{40}{n} p^n (1-p)^{40-n}$ ; het gemiddeld te verwachten aantal  $\mu = 40p$ .

2) Ter verduidelijking: als er geen trend is (d.i. als de nulhypothese  $H_0$  luidt:  $\bar{a}_1 = \bar{a}_2 = \dots \bar{a}_{31} = \dots \bar{a}_{36}$ ) dan kan de in fig. 2 getekende constellatie van a's ontstaan, doch de kans op zulk een constellatie of een nog onwaarschijnliker (als  $H_0$  juist is), is zeer klein (0,04). Aangezien  $0.04 < 0.05$  (onbetrouwbaarheidsdrempel) verwerpen wij de nulhypothese. Wat betekent dit verwerpen? Eigenlijk niet meer dan, dan niet alle  $\bar{a}_i$ 's ( $i = 1, 2, \dots, 36$ ) gelijk zijn. Wat betekent dit weer? Men zou kunnen zeggen: alle  $\bar{a}_i$ 's gelijk (=  $\bar{a}$ ) behalve bijv.  $\bar{a}_2$  en  $\bar{a}_{10}$  (pieken). Deze twee namelijk veel groter dan  $\bar{a}$ . Hoeveel groter? Dit is moeilijk te zeggen. Zodra wij dit aannemen wordt  $\chi^2$  kleiner en neemt de overschrijdingskans tot boven 0.05 toe en wordt deze hypothese niet verworpen. Deze hypothese onderstelt dus dat

(Zie vervolg op blz. 17)

Natuurlijk hangt P er mee samen hoe we de trend leggen doch de ware ("beste") trend, die natuurlijk niet perse lineair behoeft te zijn, zal zeker dicht bij de getekende liggen. Aldus blijkt hoe sterk de overschrijdingskans samenhangt met het al of niet stellen van een jaarlijkse gang.

#### 2.4 Andere kritische opmerkingen

Hiervoor verwijzen wij naar de kantletters a, b ..... in 1.

Bij a "this pattern to be repeated year for year". Dit is ons niet duidelijk. Het kan toch niet letterlijk zo bedoeld zijn? Bedoelt Bowen dat in ieder jaar op 12 Januari (veel) meer viel dan op omringende dagen? Als dat waar zou zijn, zou de datum een zeer merkwaardig klimatologisch gedrag openbaren en zouden wij wel aan de significantie van de piek moeten geloven. Bovendien: de statistische toetsen hebben juist betekenis in die gevallen, waarin het niet al "direct duidelijk" is, dat zeker verschijnsel niet louter toevallig kan zijn. Juist dan is het uitermate zaak een numerieke maatstaf voor de scheiding van toeval en niet-toeval aan te leggen.

De punten b, c en d werden inmiddels reeds in onze tekst besproken.

Bij c Vermoedelijk zal Bowen hier bedoelen, dat de datum- $\sigma$  op Britse stations kleiner is dan op Australische stations. Wij stelden zelf reeds Hoofddorp tegenover Sydney, met resp.  $\sigma = 3 \text{ à } 4 \text{ mm}$  en  $13 \text{ mm}$ . Een en ander klopt dus.

Bij f Hoe toetste Bowen dit?

Bij g Inderdaad: waarom vertelt Bowen hierover niet ook iets ten aanzien van de piekdata te Sydney? Waren er maar enkele dagen onder de 43 12 Jan.-dagen met zóveel mm neerslag, dat daardoor deze 12 Jan. als een

---

(Vervolg noot 2) van blz. 16)

2 en 10 Jan. piekdata zijn, d.w.z. zó dikwijls zulke grote dagsommen ontvangen dat hun  $\bar{a}_2$  en  $\bar{a}_{10}$  veel groter zijn dan de overige, onderling gelijke,  $\bar{a}_1$ 's. Toch wel een gekunstelde hypothese! Meer voor de hand ligt de andere alternatieve hypothese, dat de  $\bar{a}_1$  geleidelijk daalt van 1 Jan. naar 5 Feb. (voor deze daling probeerden wij een rechte). Mocht óók nu de  $\chi^2$ -som tot een  $P < 0.05$  leiden, dan zou er reden geweest zijn om te onderstellen, dat de bijzonder hoog gelegen punten op 2 en 10 Jan. zo hoog liggen, doordat deze data door oorzaken anders dan die, welke aan de overige punten schuldig zijn, opvallend dikwijls grote dagsommen ontvingen. Zo is de situatie echter niet. De punten boven 2 en 10 Jan. zouden dan veel hoger gelegen moeten hebben!

piek in fig. 1a uit de bus kwam? Of blijkt de situatie niet zo eenvoudig? Het is typisch, dat Swinbank het probleem van die zijde aanpakt en juist aldus tegen Bowen's conclusie in oppositie komt.

Bij i Is deze pieken-groepering zo merkwaardig? Vooral nu Bowen het coïncideren zo weinig scherp stelt.

Tenslotte nog enige kritiek.

1) Is er selectie-effect? Wij zouden zeggen van wel. Het wil ons voorkomen, dat Bowen naar de meest markante pieken in het ganse jaar gekeken heeft en deze op hun realiteit getoetst. Men zou daarentegen alle pieken moeten toetsen. Wanneer dat niet simultaan geschieden kan (omdat er een jaarlijkse gang is), dan zou men eerst de jaarlijkse gang moeten elimineren en, zo men daar tegen op ziet, het jaar in een aantal segmenten verdelen, in elk waarvan de jaarlijkse gang mag worden verwaarloosd, ten-einde in elk daarvan de pieken te onderzoeken. Iets dergelijks deden wij aan Nederlands materiaal.

2) Zouden wij de Bowense aanpak a.v. kunnen verbeteren?

De data der meteorietenzwermen liggen niet volkomen vast. Zo lezen wij in de Sterrengids (Ned. Ver. v. Weer- en Sterrenkunde) 1955, dat de Geminiden-meteorenvlaag haar grootste heftigheid (vijftig meteoren/ uur) omstreeks 12 December heeft, doch overigens gedurende een 10-tal dagen rondom deze datum verschijnt. Men zou nu in elk jaar op de  $n^{\text{de}}$  dag na de datum  $d_M$  van de maximum intensiteit de hoeveelheid neerslag kunnen beschouwen. Dit aldus te verstaan: men zoekt de grootste dagsom  $h_M$  op in het tijdvak van bijv. 20 tot 40 dagen na  $d_M$ . Zij deze  $h_M$  gelegen  $n_M$  dagen na  $d_M$ . Dit voor vele jaren, zodat wij een groot aantal paren  $h_M, n_M$  krijgen, met een  $\bar{h}_M$  en  $\bar{n}_M$ . Liggen de afzonderlijke  $n_M$ 's veel uiteen? Nu ook te nemen het ene jaar de hoeveelheid  $h_{10}$  op de 10<sup>de</sup> dag na  $d_M$ ; een ander jaar  $h_{43}$  op de 43<sup>ste</sup> dag na  $d_M$ ; dan weer  $h_{19}$  op de 19<sup>de</sup> dag, enz. enz., dit geheel van het toeval afhankelijk makende. Vervolgens onderzoeken of  $\bar{h}_M \gg$  gemiddelde der willekeurige  $h_i$ 's. Dit verantwoord statistisch.

Een andere vraag: doet de waarde van de bovenbedoelde max. intensiteit van de meteorenvlaag er niet toe? Is er misschien een drempelwaarde, die eerst overschreden moet worden voor er een effect is?

3

Kritiek van anderen

3.1 Kritiek van W.C. Swinbank

[2]

Is het gedrag der regengegevens werkelijk zó uitermate verrassend, dat het een zeer bijzondere verklaring vraagt en juist die, welke Bowen biedt?

Swinbank merkt op, dat door toevoeging van de periode 1902 - '44 (pieken 12, 13, 23, 31 Jan.) aan 1859 - 1901 (pieken 12 en 22 Jan.) de piek 31 Jan. verdwijnt. Dit kan ik niet met hem eens zijn, zie onze figuur 1c, met als hoogste 43 j.-dagsommen (gemiddelde der 2 43 j.-perioden), in afnemende grootte, 12.2 inch (12 Jan.), 11.7 (22 Jan.), 11.4 (13 Jan.) en 10.0 (31 Jan.). Hier dus kritiek op de kritiek van Swinbank. Terecht merkt Swinbank op, dat Bowen beweert, dat het bijzondere patroon zich elk jaar herhaalt (a), hetgeen hij in twijfel trekt (dat doen wij ook). Swinbank argumenteert a.v.: bij onderzoek blijkt de piek het gevolg van enkele dagen van zware regen te zijn en niet van een grotere frequentie van regen juist op die bijzondere dagen (E.C. echter behoeven "hoeveelheid" en "frequentie" niet strak gecorreleerd te zijn).

Verder heeft Swinbank er bezwaren tegen dat de significantie van de piekwaarde bestudeerd wordt <sup>door</sup> eerst haar verschil met de trendwaarde omtrent de piekdatum te berekenen en dan dit verschil uit te drukken in de standaarddeviatie der remaining fluctuations from the mean". Swinbank vindt dat wij hier niet gebruik mogen maken van de normale verdeling (Bowen doet dat immers), omdat de neerslag zwaar anormaal verdeeld is. Deze kritiek lijkt ons onjuist. Het is waar, dat de daghoeveelheden (uit Januari bijv.) zeer scheef verdeeld zijn (nulwaarden het meest frequent), maar de n-jarige dagsommen zijn wel "bijna" normaal verdeeld, hoe groter n, hoe beter (en bij Bowen is  $n = 43$ ). En juist van deze verdeling wordt bij de toetsing gebruikt gemaakt (zie onze berekeningen).

Voorts heeft Swinbank bezwaren tegen Bowen's conclusie, dat pieken van dezelfde grootte als te Sydney op vele stations voorkomen "on nearly the same days", waarbij Bowen op het prentje fig. 3 wijst. De ene groep heeft een breedte van 6 dagen, de andere twee groepen zijn 5 dagen breed en dat bij afstanden van ongeveer 9 dagen. Met zulk een interval en zulk een spreiding zullen vele stations wel in te passen zijn in het patroon, ofschoon hun pieken zeer verschillende oorzaken kunnen hebben.

Verder merkt Swinbank op: in fig. 1a heeft de 12 Jan. piek een hoogte van 13.2 inch. Een dagsom van 10 inches is te Sydney zeker niet onbekend. Mocht de 13.2 zulk een 10 of meer hoeveelheid bevatten, dan

maakt het al of niet meenemen van het jaar met deze bijzonder grote daghoeveelheid zeer veel verschil (E.C.: maar ook op de  $s$ , die vrij sterk met de pieken samenhangt; de significantie hangt samen én met de piekhoogte, in absolute zin, én met de  $s$ , n.l. met het verschil tussen de piek- en de trendwaarde, uitgedrukt in  $s$ ).

Zou men de periode 1902 - 1944 hebben uitgebreid tot en met 1951 dan zou men mee hebben moeten nemen 2.24 inch op 18 Jan. 1951, waardoor er een piek (18 J.) tussen 13 en 23 J. bijgekomen zou zijn.

Verder merkt Swindbank op, dat zware depressies zich voortbewegen en aldus regenhoeveelheden brengen, groot genoeg om hoofdpieken te bewerkstelligen op verschillende plaatsen op opeenvolgende dagen over een interval van verscheidene dagen. Maar dan verliest zulk een patroon als in fig. 3 zijn statistische zin.

Swindbank concludeert tenslotte (wij kunnen hier niet zijn ganse artikel citeren): "the inference is doubtful, and the case for synchronicity of peak dates must remain open until figure 3 has been supplemented by the inclusion of many more stations".

### 3.2 Kritiek van D.F. Martyn

[3]

Martyn vertelt, dat Bowen niet het bestaan van wereldomvattende verschijnselen, zoals hij beschrijft, werkelijk bewezen heeft. Het blijkt mogelijk de regenvalpieken, die Bowen selecteert, toe te schrijven aan 1 of 2 lokale depressies; ook is het mogelijk pieken te bewerkstelligen op iedere willekeurige datum. Martyn vestigt de aandacht op 13 Januari (zie fig. 1b), als piek in de periode 1902 - 1944 (te Sydney). Deze piek-waarde (11.4 inch) is tevens de grootste 43 jarige dagsom voor onverschillig welke datum in Januari. Het gemiddelde der 31 43 jarige dagsommen in Januari is 4.50 inch. Dus heeft 13 Jan. meer dan het dubbele van het Januari-gemiddelde. Bowen spreekt bovendien van "an excess of rain repeated year after year". Echter is dat wel waar? Op 13 Jan. 1911 viel te Sydney 7.08 inch.

Dus eerst een piek van 11.4 inch tegenover 4.70 inch, gemiddeld over de gehele Jan.-maand (43 jarige dagsommen); nu (1911 niet meenemende)  $11.4 - 7.08 = 4.32$  inch, tegenover  $g_{42}$  inch, gemiddeld over alle Jan.-dagen (42 j.-dagsommen) Martyn lijkt ons bij  $g_{42}$  weer aan 4.70 te denken. Eigenlijk verandert ook deze 4.70, echter vermoedelijk zeer weinig (Martyn geeft geen getallen) Als de som der 31 Jan.-dagsommen in 1911 gelijk  $S$  is, is  $g_{42} = 4.70 - \frac{S}{31}$ . Één der 31 dagsommen is 7.08 inch (13 Jan.). Eerst  $11.4 \gg 4.70$ ; nu  $4.32 < g_{42}$  en toch heet 13 Jan. een "piek-datum". Zo betrekkelijk is dus het begrip piek!



Wij citeren Martyn: "It seems difficult indeed to derive any worthwhile conclusion from data so variable en so limited, and impossible to attach significance to daily peaks which, contrary to the quoted statement, do not derive from a repeated yearly pattern, but are built of one major fall."

Als men wil kan men piekdata maken, door de juiste basisperiode te kiezen en jaren met grote dagsommen al of niet mee te nemen. Dit laatste doet Martyn dan ook zeer nadrukkelijk. Hij beschouwt de regengegevens voor Australië gedurende Jan. 1911. Voor elke dag (tussen 8 en 14 Jan.) werd dat station gezocht, waar meer dan 35 keren het gemiddelde viel. Deze 35 is zo groot, dat het zo goed als zeker is, dat op elk dier stations op die bewuste dag een piek te voorschijn komt als men er een prentje analoog aan dat voor Sydney (fig. 1) voor zou tekenen (mits men maar zorgt dat het jaar 1911 deel uitmaakt van de beschouwde periode). Er zouden aldus piekdata komen: 8 Jan. (West Leichhard), 9 Jan. (Cloncurry), ..... 13 Jan. (Sydney), 14 Jan. (Eden en Walhalla). Als men deze stations op de kaart markeert kan men de depressie a.h.w. volgen. Inderdaad blijkt een en dezelfde depressie aldus getrokken te zijn vanuit Upper Western Queensland op 8 Jan., naar het Z.O. naar Eden op 14 Jan. in het Zuiden. Doch dan zijn deze piekdata (6 dagen uit elkaar liggende; let op de breedte van de piekdata-groepering in fig. 3) louter via de depressie gelieerd en geenszins door tussenkomst van meteorietenstof e.d.

Verder maakt Martyn een opmerking bij Bowen's figuur 3, analoog aan de onze. Heeft het prentje betrekking op één station of op vele tezamen? Zo het laatste het geval is, is een dag een "natte dag" als er een zware regen viel ergens in Engeland. Maar dan moet men bij toetsingen met statistische toetsen dubbel oppassen!

### 3.3 Kritiek van Neumann 1)

Neumann merkt op, dat men alvorens Bowen's hypothese te aannemen, moet onderzoeken of de pieken + troggen (dalen) louter toevallig over de tijdas verspreid liggen. Ondersteld is als nulhypothese, dat men N elementen aselekt trekt uit een normaal verdeeld universum en deze als tijdreeks noteert. Kendall heeft bewezen, dat het gemiddeld aantal "turning points" in zulk een N-reeks gelijk  $E(p) = \frac{2}{3}(N - 2)$ .

---

1) Resultaten, die Neumann ons te Rome tijdens het 10<sup>de</sup> U.G.G.I. Congres (1954) mededeelde.

Van een turning-point is sprake óf als  $x_{i-1} < x_i > x_{i+1}$  (maximum) óf als  $x_{i-1} > x_i < x_{i+1}$  (minimum). Een fase is de afstand tussen een max. (min.) en eerstvolgend min. (max.).

Een tweede toets betreft de kansverdeling van de lengte  $d$  van een fase. Kendall bewees: de kans op een lengte  $d$  (bij een steekproef van  $N$  - termen) is

$$P = \frac{6 (d^2 + 3d + 1) (N - d + 2)}{(d + 3) ! (2N - 7)}$$

Toepassing op de Sydney-reeksen, die Bowen op het oog heeft:

|                                     |             |             |
|-------------------------------------|-------------|-------------|
| a)                                  | 1859 - 1901 | 1902 - 1944 |
| waargenomen aantal "turning points" | 22          | 23          |
| $E(P)$ , met $N = 36$               | 22.3        | 22.3        |

Dus ligt  $E(P)$  vlak bij 22 of 23.

|                       |          |             |       |             |       |
|-----------------------|----------|-------------|-------|-------------|-------|
| b)                    |          | 1859 - 1901 |       | 1902 - 1944 |       |
| fase lengte $d$ dagen | kans $P$ | W           | T     | W           | T     |
| 1                     | 0.63     | 13          | 13.23 | 12          | 13.86 |
| 2                     | 0.27     | 7           | 5.67  | 9           | 5.94  |
| $\geq 3$              | 0.10     | 1           | 2.10  | 1           | 2.20  |
| som                   |          | 21          | 21.00 | 22          | 22.00 |

De overeenkomst tussen  $W$  = waargenomen aantal fasen en  $T$  = theoretische aantal is goed. De  $\chi^2$  - toets behoeft niet toegepast te worden. Resultaat: beide toetsen leren, dat de twee reeksen een toevalskarakter hebben. De pieken en dalen schijnen toevalsresultaten te zijn ..... en dus moet in de coincidentie van pieken het spel van het toeval gezien worden.

Enige toelichting: de nulhypothese luidt, dat ondersteld is, dat wij een aselechte steekproef van  $N$  elementen uit een normaal verdeeld universum genomen hebben. Zulk een nulhypothese is eigenlijk op zijn minst genomen tweeledig, wanneer wij letten op de qualificaties "aselect" en "normaal verdeeld universum". Het tweede heeft te maken met de vraag of

het getal 43 al zo groot is, dat de over 43 jaren gesommeerde dagsommen zo goed als normaal verdeeld zijn (terwijl de dagsommen dit zelf beslist niet zijn). Onze ervaring doet ons vermoeden, dat deze onderstelling wel juist is, maar van een "aselect-heid" is zeker geen sprake. Vermoedelijk gaat Neumann van de gedachte uit, dat tussen de over 43 jaren getotaliseerde 31 Januari-dagsommen geen autocorrelatie meer is, maar dat heeft hij o.i. toch mis (zie hierover vooral ons rapport W.R. 55 - 004).

In verband daarmee doet o.i. Neumann niet goed om  $N = 36$  te substitueren; het moet zijn  $N_p < 36$ , doch hoeveel? Wij weten het niet, omdat Bowen geen getallen omtrent de persistentie vermeldt.

Overigens voelen wij niet veel voor deze "turning-point" toets, omdat hij maar van weinig informatie gebruik gemaakt. Het gaat meer om de grootten der pieken dan om het verschijnsel van pieken zelf.

Een ander bezwaar onzerzijds tegen Neumann's critiek is deze, dat ons niet zo zeer het al- of niet-toevalskarakter van elk der twee figuren 1 (a en b) interesseert maar meer dat van de coincidentie der pieken. Voor een toetsing daarvan hebben wij aan deze twee Kendall-toetsen niet veel. Hoe dan wel?.

N.B. Kendall heeft afgeleid, dat in een reeks met autocorrelatie de gemiddelde afstand  $\bar{a}$  tussen twee successieve maxima gegeven wordt door  $\bar{a} = 2\pi / bg \cos b$  met  $b = -\frac{1}{2} + \frac{\rho_1 - \rho_2}{2(1 - \rho_1)}$ , met  $\rho_1 =$  pers.

coëff. van 1ste orde;  $\rho_2$ : 2de orde. Dikwijls is in goede benadering

$\rho_2 = \rho_1^2$  en dus  $b = -\frac{1}{2} + \frac{1}{2}\rho_1$ . Wij willen deze uitdrukking toepassen

op de reeks der 43 j.-dagsommen fig. 1a. Wij tellen 11 maxima en  $\bar{a} = 3.3$  dag; dus  $bg \cos b = \frac{2}{4.1} = 1.53 \text{ rad.} = 87^\circ 36'$ ; dus  $b = \cos 87^\circ 36' =$

0,042 en  $\rho_1 = 0.32$  en  $\rho_2 = 0.10$ . Dit komt ons zeer groot voor. De tweede reeks levert toevallig ook  $\bar{a} = 3.3$ .

Terecht merkt Brooks op, dat een berekening van  $\rho$  uit het gemiddelde interval tussen maxima meestal af te raden is, aangezien de aldus berekende  $\rho$  zeer gevoelig is voor kleine veranderingen in het gem. interval. Zo zou voor 10 maxima zijn:  $\rho_1 = 0.65$  en  $\rho_2 = 0.43$

3.4 In het Annual Report for the year 1953/54 Commonwealth Scientific and Industrial Research Organization, Section of Meteorological Physics, wordt gezegd: "Een schatting werd gemaakt van de invloed van meteorietenstof op de regenval, met de conclusie dat op basis van de huidige kennis een significante invloed onwaarschijnlijk bleek". In het rapport wordt niet verteld hoe men tot de conclusie kwam.

4

Bevestiging van Bowen's theorie ?

4.1 Artikeltje van M.B. Oliver and V.J. Oliver [4]

In dit artikel zeggen de auteurs, dat zij op grond van een onderzoek in tropisch Afrika de hypothese van Bowen bevestigd vinden, hoewel hun de ware fysische verklaring van het proces nog duister is. Hoe dit onderzoek plaats vond en hoe deze bevestiging bedoeld is vertellen zij niet.

4.2 "Meteoric dust and rainfall" by O.N. Dhar [5]

In dit artikel vertelt Dhar, dat hij de opvatting van Bowen bevestigd vindt aan de regenstatistieken van Pamban en Madras, aan de oostzijde van Voor Indië. Het artikel vertoont twee prentjes 1. de 53 j.-dagsommen te Pamban voor elk der dagen 1 Nov. - 27 Jan. 2. de 53 j.-dagsommen te Madras (1 Nov. - 27 Jan.). Periode: 1900 - 1952.

Figuur 1 vertoont (allicht) een grillig beeld; vele pieken en dalen, met een duidelijke jaarlijkse gang: omstreeks 1 Nov. bedraagt de 53 j.-dagsom ongeveer 20 en op 27 Jan. ongeveer 2 inches. Langs de bovenrand van beide figuren worden aangegeven de data van de belangrijkste meteorieten-zwermen: Orioniden (20 - 23 Oct.), Tauriden (3 - 10 Nov.), Leoniden (16 - 17 Nov.), Germiniden (13 - 14 Nov.) en de Ursiden (22 Dec.). Sommige regenvalpieken (te Pamban is de grootste ong. 27 inches, op 8 December; te Madras: ong. 28 inches, op 21 November) lijken 30 + 2 dagen te komen ná de inval der meteorieten. .... maar er zijn ook "tussenpieken", die niet 30 dagen na zekere meteorietenzwermen gedateerd zijn. Dhar "verklaart" de laatste pieken door te zeggen, dat zij samenhangen met sporadische hevige regens bij cyclonen en depressies, die juist omtrent 26 - 28 Januari zeer frequent zijn. Dhar concludeert tenslotte tot een associatie met de meteorieten.

E.C.: Omdat iedere statistische aampak ontbreekt en andere beschouwingen dit gemis evenmin goedmaken, lijkt ons de waarde van het artikel niet groot.

5

Enige citaten uit "Der Jahresgang der Niederschlagswahrscheinlichkeit auf dem Sonnblick, 3106 m" von F. Steinhauser, in het 48ste "Jahresbericht des Sonnblick-Vereines für das Jahr 1950".

Op basis van een 60 jarig materiaal (1891 - 1950) werd voor Sonnblick per dag de relatieve neerslagwaarschijnlijkheid (in procenten) berekend. De jaarkromme geeft natuurlijk een zeer grillig beeld te zien; duidelijk is echter, dat omstreeks half Juni de neerslagkans ongeveer 75 à 80 %, omstreeks half October ongeveer 40 % bedraagt. (toelichting: als er onder de 60 2 Jan.-dagen 36 stuks waren met neerslag -d.i.  $\geq 0.1$  mm -, is de relatieve neerslagwaarschijnlijkheid  $(36 : 60) 100 = 60 \%$ ). Daarna werden voortlopende 5 dagen-gemiddelden berekend volgens  $w_i^* = \frac{1}{5} (w_{i-2} + w_{i-1} + w_i + w_{i+1} + w_{i+2})$ ,  $i = 1, 2, \dots, 365$ . Ook daarna vertoont de jaarcurve grillige veranderingen (fig. 1). Hetzelfde werd gedaan op basis van 1891 - 1920 (30 j.)

De schrijver merk met nadruk op, dat in bepaalde gedeelten van het jaar de regenkanen ver boven of ver beneden het algemeen gemiddelde liggen en vraagt zich af of deze anomalieën slechts toevallig zijn, d.w.z. bij een groter materiaal wel zullen verdwijnen, of dat er bijzondere fysieke oorzaken te noemen zouden zijn voor deze pieken en dalen (Wij zijn aldus midden in het onderwerp van ons rapport terecht gekomen). In het laatste geval zou deze kennis van enig gewicht voor de voorspelling van regen kunnen zijn. Echter "Eine überzeugende Entscheidung über die Realität dieser "Singularitäten" oder Anomalien im Jahresgang ist schwer zu bringen". Wel lijkt het de auteur gewenst de periode in tweeën te delen en dan te zien of de singulariteiten in beide deelreeksen op dezelfde plaats voorkomen. Zo ja, dan zullen "diese Singularitäten einen hohen Grad von Wahrscheinlichkeit haben". Toch voelt Steinhauser zich niet zeker: "Unabhängig davon, ob mit einer solchen Unterteilung schon ein "Beweis" als erbracht angesehen werden kann, ist es jedenfalls zu sehen, wie weit oder in welchen Abschnitten die Teilkurven einen gleichen oder zumindest ähnlichen Verlauf nehmen". Waarom vertelt de schrijver niet nader wat hij als "bewijs" definiëren wil? Hij constateert dan "in kleineren Details eine beträchtliche Ähnlichkeit", tussen de figuren 2 en 3, die nog groter wordt als men van de kleinste details afziet en meer op de tendensen let, waarbij verschuivingen voor een dag toegelaten worden, hetgeen bij 30 jarige reeksen wel geoorloofd is. Hier komt Steinhauser op glad ijs. Wat is nu wel en wat niet statistisch geoorloofd en waarom nog wel bij 30 j.-reeksen, niet bij 15 j.-reeksen en zeker wel bij 100j.-reeksen? Als het

betoog voor ons interessant begint te worden, maakt Steinhauser ons niet veel wijzer. Zijn singulariteiten-onderzoek aan de Sonnblick-regengegevens heeft dan ook voor de statisticus weinig waarde.

6

Behandeling van Nederlands materiaal

Wij lieten machinaal met de geponste daghoeveelheden van Hoofd-dorp (1867 - 1953; enkele jaren vielen uit) de 81-jarige dagsommen be-rekenen. In fig. 4 ziet men de gemiddelden. In fig. 5 ziet men de 81-jarige dagsommen voor 1 Januari t/m 10 Februari. In de fig. 6 en 7 hebben wij voor gedeelten uit het jaar de 31- en 48 j.-dagsommen in grafiek gebracht (2 deelreeksen, die geen gemeenschappelijke jaren hebben). De vier krommen in fig. 6 <sup>en 7</sup> bekijkend, zien wij maar voor één flinke top (27 Aug.) een preciese coïncidentie. Natuurlijk zijn er meer coïncidenties van pieken als wij tussen de piekdata 0, 1 of 2 dagen verschil toelaten, doch dan komen zowel coïnciderende toppen als dalen voor.

Wij hebben getracht iets omtrent de mogelijke realiteit van de pieken in elk der 12 maanden te zeggen. Daarbij zijn nodig: de spreidings-breedte; het effectieve aantal dagen per maand en de standaarddeviatie.

In de volgende tabel zijn de resultaten tezamen genomen voor die maanden, waarin de spreidingsbreedte misschien boventoevallig groot is. Hierbij is  $s$  = datumstandaarddeviatie, d.i. de standaarddeviatie tussen de hoeveelheden op bijv. 1 Jan.; deze is binnen de maand Jan. constant gedacht. Verder is de variantie van de 81-jarige dagsom  $s\sqrt{81}$ . Voorts is  $N_{eff}$  berekend m.b.v. de benadering  $N_e = N \frac{1 - \rho_1}{1 + \rho_1}$ , waarbij de persisten-tiecoëfficiënt tussen de daghoeveelheden is  $\rho_i = \rho_1^i$ ;  $R$  = range (spreidings-breedte). De significantie-drempel is 0.05 (de waarde van  $s$  is overgenomen uit ons rapport W.R. 55 - 004, 1955)

| 1    | 2    | 3    | 4               | 5                                    |     |     |     | 6               |         | 7               |         | 8               |         |
|------|------|------|-----------------|--------------------------------------|-----|-----|-----|-----------------|---------|-----------------|---------|-----------------|---------|
|      |      |      |                 | de range is signif. indien $\geq$ mm |     |     |     | $\rho_1 = 0.15$ |         | $\rho_1 = 0.25$ |         | $\rho_1 = 0.40$ |         |
|      | R mm | s mm | $s\sqrt{81}$ mm | $N_e = 20$                           | 15  | 10  | 5   | $N_e$           | sign. R | $N_e$           | sign. R | $N_e$           | sign. R |
| Jan. | 156  | 3.7  | 33.3            | 167                                  | 160 | 149 | 128 | 23              | -       | 19              | -       | 13              | +       |
| Mrt. | 120  | 2.5  | 22.5            | 113                                  | 108 | 101 | 86  | 23              | ?       | 19              | +       | 13              | +       |
| Mei  | 173  | 3.8  | 34.2            | 172                                  | 164 | 153 | 132 | 23              | +       | 19              | +       | 13              | +       |
| Aug. | 239  | 4.8  | 43.2            | 217                                  | 208 | 193 | 166 | 23              | +       | 19              | +       | 13              | +       |
| Sep. | 180  | 4.5  | 40.5            | 203                                  | 194 | 181 | 164 | 22              | -       | 18              | -       | 12              | -       |

De werkelijke  $\int_1$  is ongeveer 0.25 (zie ons rapport W.R. 55 - 004; 1955), maar deze waarde ligt door allerlei oorzaken weinig precies vast. Misschien is ze met 95 % zekerheid gelegen tussen 0.35 en 0.15. Men ziet, dat de uitkomst, wat de significantie van R betreft, met de  $\int_1$ -waarde samenhangt. Misschien zelfs moeten wij de waarde van  $\int_1$  van de maand afhankelijk stellen. In genoemd rapport zetten wij uiteen hoe moeilijk dit probleem is. In ieder geval mogen wij het aantal maanden, waarin de range boventoevallig groot is (misschien 1 à 3 op de 12), niet verontrustend groot noemen, m.a.w. de conclusie, dat wij (voorlopig) de aan deze ranges schuldige pieken als toevallig mogen beschouwen, lijkt ons gepermitteerd. 1)

Natuurlijk valt de 27 Aug.-piek erg op; de 81 j.-som is 398 mm, terwijl 26 Aug. levert 199 en 28 Aug. levert 237 mm. Wij vragen ons allicht af hoe 27 Aug. aan zulk een piek komt. Aan een paar exceptioneel grote dagsommen? De grootste 27 Aug.-hoeveelheid was: 48 mm (1867); die op 27 Aug.: 22 mm (1877) en op 28 Aug.: 25 mm (1935). De 26ste Aug. telde 1 dagsom  $\geq$  20 mm; 27 Aug.: 8 stuks (48; 36; 34; 28; 25; 24; 21 mm) en 28 Aug.: 1 stuks. Nu een andere piek, bijv. die van 2 Jan., met 248 mm; 1 Jan.: 161 en 3 Jan.: 154 mm. Op 1 Jan.: grootste dagsom: 16 mm; op 2 Jan.: 29 en op 3 Jan.: 16 mm. Op 1 Jan. 5 keren een dagsom  $\geq$  10 mm; op 2 Jan.: 7 keren en op 3 Jan.: 2 keren.

Men zou bijna gaan geloven aan significante verschillen in regenkansen op naburige dagen. Een streng statistisch onderzoek zou vereist zijn. Het is zeer waarschijnlijk, dat een basisperiode van 81 jaren te klein zou blijken voor betrouwbare statistische uitspraken in deze zaken.

---

1) De betrouwbaarheidsdrempel 0.05 stellende, is de kans  $K(3; 12)$  op 12 experimenten met daaronder 3 of meer successen (succeskans per experiment = 5%):  $1 - \sum_{i=0}^2 \binom{12}{i} 0.05^i 0.95^{12-i} = 0.02$  (terwijl  $K(2; 12) = 0.12$ ). Deze kans is inderdaad zeer klein. Wat nu te concluderen? De pieken zijn dus toch reëel? Of er is een verwaarlozing gemaakt, die achteraf niet geoorloofd blijkt? Welke? Wij voelen meer voor de tweede mogelijkheid. Immers wij hebben de jaarlijkse gang binnen een maand verwaarloosd. Een spreidingsbreedte wordt allicht groter door een sterke jaarlijkse gang binnen de maand. Hiervoor is wel te corrigeren, maar wij zagen af van het enorme werk. Wij zouden dan namelijk iedere dagsom van ieder jaar moeten verminderen met de trendwaarde (= jaarlijkse gang) ter plaatse (de berekening van deze trend is een probleem op zichzelf). Daarna dient dit (pos. of neg.) verschil  $\Delta$  gedeeld te worden door de standaarddeviatie ter plaatse (deze is ook al datum-afhankelijk, d.w.z. ook de trend in de st.dev. zou berekend moeten worden); q. Deze pos. en neg. q's (ong. 84x365 stuks) leveren een nieuwe figuur, analoog fig. 4. Wij hebben in feite aldus gestandaardiseerd in de hoop de jaarlijkse gang te elimineren. Nu kunnen wij weer de ranges gaan onderzoeken, per maand, per halfjaar, per jaar, enz.

7 Formulering van een nieuw probleem i.v.m. Bowens' onderzoek

7.1 Gegeven zij een normaal verdeeld universum  $(\mu; \sigma)$ , waaruit wij op aselechte wijze 31 elementen nemen, die in deze volgorde als een tijdreeks genoteerd worden:  $x_1, x_2, \dots, x_{31}$ . Wij nemen een bepaalde plaats  $i$  in het oog en wensen daar een "piek". Echter hoe is de definitie van een "piek"? Er zijn al direct tenminste 3 mogelijke definities. Zij kans  $K(x < a) = K_a$ ;

Mogelijkheid 1

$$x_{i-1} < a; x_i > b; x_{i+1} < a \quad 1)$$

De kans op 31 elementen, zodanig, dat op plaats  $i$  een piek komt, terwijl voor de overige  $x$ -waarden geen voorwaarden gesteld worden, is  $P = K_a^2 K_b$ . Hierbij is de kans  $K(x < a) = K_a$  en  $K(x > b) = K_b$ . Eigenlijk stelt deze  $p$  de kans op een constellatie van 31 elementen voor, met tenminste 1 piek, waarvan er één op plaats  $i$  moet liggen. Wanneer men  $A$  keren zulk een steekproef doet en dus  $A$  "prentjes" krijgt, dan is de kans, dat daaronder  $B$  prentjes voorkomen, die elk de piek op plaats  $i$  hebben ("coïncidentie op  $i$ ") :  $P_i = \binom{A}{B} p^B (1-p)^{A-B}$ .

Bij Bowen is  $A = 2$ ;  $B = 2$ . Numerieke voorbeelden:

I  $a = \mu; b = \mu + \sigma; K_a = \frac{1}{2}; K_b = 0.139; p = 0.03475; p^2 = 0.00121;$   
 $P_i = \binom{2}{2} 0.03475^2 (1 - 0.03475)^0 = 0.0012$

II  $a = \mu; b = \mu + 2\sigma; K_a = \frac{1}{2}; K_b = 0.0228; p = 0.0057;$   
 $p^2 = 0.000065; P_i = \binom{2}{2} 0.0057^2 (1 - 0.0057)^0 = 0.000065$

Wanneer de waarde  $i$  onverschillig is, is de coïncidentie-kans "B in A" gelijk  $P = 31 P_i$ ; in I dus  $31 \times 0,0012 = 0,038$ ; in II:  $31 \times 0,000065 = 0,001$ .

1)

Men zou ook alleen de eis  $x_i > b$  kunnen stellen,  $b$  voldoende groot en ongeacht de waarden van  $x_{i-1}$  en  $x_{i+1}$ . Maar kijken wij bij de definitie van een piek niet ook naar de "buren"? Intussen ligt dit geval, zo men er veel voor gevoelt, in onze notatie  $x_{i-1} < a; x_i > b; x_{i+1} < a$  opgesloten, als men  $a = \infty$  substitueert (en  $K_a = 1$ ). Natuurlijk kunnen er ook twee pieken naast elkaar liggen:  $x_{i-1} < a; x_i > b; x_{i+1} > b; x_{i+2} < a$ . Wij doen hier maar voorstellen; voor nadere uitwerking ontbrak de tijd.



Nu kunnen wij ook nog, zoals Fowen doet, de coincidentie minder scherp stellen, d.w.z. van een coincidentie spreken als de data der pieken 0,1 of 2 dagen verschillen. In het ene prentje de piek op  $i$ , in het andere op  $i - 2$  of  $i - 1$  of  $i$  of  $i + 1$  of  $i + 2$ .

In zulk een geval is  $p = K_a^2 K_b$  en voor het tweede prentje 5 p. Zijn er slechts twee prentjes gemaakt, dan wordt coincidentiekans (in de nieuwe betekenis) :  $P_i = 5p^2$ . Dus in geval I:  $5 \times 0,0012 = 0,006$ ; II:  $5 \times 0,000065 = 0,000325$ . Vervolgens laten wij  $i$  zelf onverschillig; dan I:  $P = 31 \times 0,006 = 0,186$  en II:  $P = 31 \times 0,000325 = 0,010$ . Men ziet hoe groot deze kansen thans zijn. De eerste komt ver boven de gebruikelijke significantie-drempel (5 %) uit.

### Mogelijkheid 2

$$x_i - x_{i-1} > c; x_i - x_{i+1} > c$$

Wij weten, dat de verschillen  $\Delta = x_p - x_{p-1}$  ( $p = 1, 2, \dots$ ) een normale verdeling volgen met gemiddelde  $\mu_\Delta = 0$  en  $\sigma_\Delta = \sigma\sqrt{2}$  (let wel: de  $x$ 'en zijn niet gecorreleerd). Wij beschouwen weer twee gevallen

I:  $c = \mu_\Delta + \sigma_\Delta$  en II:  $c = \mu_\Delta + 2\sigma_\Delta$ .

In geval I is  $p = 0,139^2 = 0,0194$ ; de coincidentiekans (2 prentjes):  $P_i = 0,0194^2 = 0,000376$  en onverschillig de  $i$ :  $P = 31 \times 0,000376 = 0,012$ . In geval II is  $p = 0,0228^2 = 0,00052$ ; coïnc. kans (2 prentjes):  $P_i = 0,00052^2 = 27,1 \cdot 10^{-8}$  en onverschillig de  $i$ :  $P = 31 \times 27,1 \cdot 10^{-8} = 8,4 \cdot 10^{-6}$

Nu kunnen wij ook nog de coincidentie soepeler definiëren, als boven, en in I:  $P = 5 \times 0,012 = 0,060$  en in II:  $P = 5 \times 8,4 \cdot 10^{-6} = 4,2 \cdot 10^{-4}$ .

### Mogelijkheid 3

$$x_i - x_{i-1} > c; x_i - x_{i+1} > c \text{ en } x_i > b.$$

Wij nemen, als voorbeeld, voor  $b$  alleen  $b = \mu + \sigma$  ( $K_b = 0,139$ ) en weer 2 gevallen voor  $c$  ( $\mu_\Delta + \sigma_\Delta$  resp.  $\mu_\Delta + 2\sigma_\Delta$ ).

In geval I:  $p = 0,139$ .  $0,139^2 = 0,0027$ ;  $P_i = 0,0027^2 = 0,00000726$ ;  $P = 31 \times 0,00000726 = 2,25 \cdot 10^{-4}$ . In geval II:  $p = 0,139$ .  $0,0228^2 = 3,17 \cdot 10^{-3}$ ;  $P_i = (3,17 \cdot 10^{-3})^2 = 10^{-5}$ ; en  $P = 31 \times 10^{-5} = 3,1 \cdot 10^{-4}$ .

Nu ook nog de coincidentie soepeler definiëren en er komt, in I:  $P = 5 \times 2,25 \cdot 10^{-4} = 1,1 \cdot 10^{-3}$  en in II:  $P = 5 \times 3,1 \cdot 10^{-4} = 1,6 \cdot 10^{-4}$ .

7.2

Boven werd ondersteld, dat persistentie afwezig is. Hoe als er wel autocorrelatie is? De kans op  $x_j < \mu$  (als voorbeeld) is dan niet meer persé  $\frac{1}{2}$ , doch afhankelijk van de waarden van  $x_{j-1}$ ,  $x_{j-2}$  enz. De aanpak wordt dan moeilijk.

Men kan misschien verder komen met de tweede definitie, onder mogelijkheid 2. De  $\Delta$ 's volgen nu namelijk een normale verdeling met  $\mu_{\Delta} = 0$  en  $\sigma_{\Delta} = \sigma \sqrt{2(1-r)}$ , als  $r$  de c.c. is. Wil men vasthouden aan de definitie van de piek in de vorm  $x_i - x_i > \sigma \sqrt{2}$  (idem  $x_i - x_{i+1}$ ), dan moet men bedenken, dat  $\sigma \sqrt{2} = \frac{1}{\sqrt{1-r}} \{ \sigma \sqrt{2(1-r)} \}$ . In geval I is  $p$  dan niet meer 0.139, maar kleiner, aangezien  $1/\sqrt{1-r} > 1$ . Wij vullen een aannemelijke  $r$  in, t.w. 0.25 en dus  $1/\sqrt{1-r} = 1.156$ . In geval I:  $p = 0.123^2 = 0.0152$ ;  $P_i = 0.0152^2 = 0.000233$  en  $P = 31 \times 0.000233 = 0.007$ . In geval II:  $p = 0.0104^2 = 1.085 \cdot 10^{-4}$ ;  $P_i = 1.085^2 \cdot 10^{-8} = 1.18 \cdot 10^{-8}$  en  $P = 31 \times 0.18 \cdot 10^{-8} = 3.7 \cdot 10^{-7}$ .

Nu ook nog de coïncidentie soepeler definiëren: in I:  $P = 5 \cdot 0.007 = 0.035$  en in II:  $P = 5 = 5 \times 3.7 \cdot 10^{-8} = 1.8 \cdot 10^{-7}$ . Wij zien duidelijk de coïncidentie-kans afnemen ( $0.060 \rightarrow 0.035$ , in I en  $4.2 \cdot 10^{-4} \rightarrow 1.8 \cdot 10^{-6}$ , in II) als er persistentie ( $r = 0.25$ ) is, als wij de piek even scherp stellen.

Wij hebben een en ander in het volgende schema overzichtelijk samengenomen (blz. 31)

Er zullen nog wel andere piekdefinities zijn. Ook kan men de piek zelf een breedte van bijv. 2 of 3 dagen geven. In zulk een geval zouden analoog aan de mogelijkheid 1,  $x_{i-2} < a$   $x_{i-1} > b$   $x_i > b$   $x_{i+1} > b$   $x_{i+2} < a$ , kunnen zijn. Bovendien kan men een coïncidentie van meer dan één piek eisen. Enz.

De tijd ontbrak om al deze mogelijkheden verder door te rekenen.

Berekening van coïncidentie-kansen

| 1   |                     | 2  |                                       | 3                                     |                                       |
|---|---------------------|--|---------------------------------------|---------------------------------------|---------------------------------------|
| $x_i > b$   |                     |  |                                       | $x_i > b$                             |                                       |
| $x_{i-1} < a;$  | $x_{i+1} < a$       | $x_i - x_{i-1} > c;$                         | $x_i - x_{i+1} > c$                   | $x_i - x_{i-1} > c;$                  | $x_i - x_{i+1} > c$                   |
| $a = \mu$   | $a = \mu$           | $c = \frac{\sigma}{\Delta} = \sigma\sqrt{2}$ | $c = 2\sigma\sqrt{2}$                 | $c = \sigma\sqrt{2}$                  | $2\sigma\sqrt{2}$                     |
| $b = \mu + \sigma$  | $b = \mu + 2\sigma$ |  |                                       | $b = \mu + \sigma$                    | $\mu + \sigma$                        |
| $K_a = 0.5$   | 0.5                 | $K_c = 0.139$                                | 0.0228                                | 0.139                                 | 0.139                                 |
| $K_b = 0.139$   | 0.0228              | $K_c^2 = 0.0194$                             | 0.00052                               | $K_b K_c^2 = 0.0027$                  | $3.2 \cdot 10^{-3}$                   |
| $p = K_a^2 K_b = 0.0348$  | 0.0057              | $3.8 \cdot 10^{-4}$                          | $2.7 \cdot 10^{-7}$                   | $7.3 \cdot 10^{-6}$                   | $10^{-5}$                             |
| $P_i = p^2 = 0.00121$   | $6.5 \cdot 10^{-5}$ | 0.012  | $8.4 \cdot 10^{-6}$                   | $2.2 \cdot 10^{-4}$                   | $3.1 \cdot 10^{-4}$                   |
| $P = 0.037$   | 0.002               | <u>0.060</u>                                 | <u><math>4.2 \cdot 10^{-5}</math></u> | <u><math>1.1 \cdot 10^{-3}</math></u> | <u><math>1.6 \cdot 10^{-4}</math></u> |
| $(P) = \underline{0.186}$                                       | <u>0.010</u>        |  |                                       |                                       |                                       |
| Persistentie $r = 0.25$   |                     |  |                                       |                                       |                                       |
| $c = \frac{\sigma^*}{\sqrt{1-r}}$                               |                     |  |                                       |                                       |                                       |
| $\sigma^* = \sigma \sqrt{2(1-r)}$                               |                     |  |                                       |                                       |                                       |
| $K_c = 0.123$ $0.0104$  |                     |  |                                       |                                       |                                       |
| $p = K_c^2 = 0.0152$ $1.08 \cdot 10^{-4}$                       |                     |  |                                       |                                       |                                       |
| $P_i = p^2 = 0.00023$ $1.2 \cdot 10^{-8}$                       |                     |  |                                       |                                       |                                       |
| $P = 0.007$ $3.7 \cdot 10^{-7}$                                 |                     |  |                                       |                                       |                                       |
| $(P) = \underline{0.035}$ <u><math>1.8 \cdot 10^{-6}</math></u> |                     |  |                                       |                                       |                                       |

Addendum

Een en ander over enige range-toetsen.

8.1

Gesteld er worden  $n$  elementen op aselechte wijze uit een normaal verdeeld universum, waarvan het gemiddelde  $\mu$  en de standaarddeviatie  $\sigma$  zijn, genomen en in volgorde van grootte genoteerd:  $x_1 \dots x_n$ . Het verschil  $x_n - x_1$  heet de "range" of spreidingsbreedte in de steekproef. Beschouw  $w = (x_n - x_1) : \sigma$ . Er bestaat geen simpele uitdrukking voor  $f_n(w)$  dw

van  $w$ , d.i. de waarschijnlijkheidsverdeling van de stochastische variabele  $w$ . Pearson [6] heeft  $\int_0^W f_n(w) dw$  voor speciale waarden  $W$  van  $w$  berekend en wel voor  $n = 1, 2, 3 \dots 20$  en voor  $W = 0.00; 0.05; 0.10; \dots 7.25$ . Deze  $P$  geeft aan de kans, dat de spreidingsbreedte in een randomsteekproef van  $n$  elementen ten hoogste  $W$  keren de standaarddeviatie van het normaal verdeelde universum bedraagt. Men bedenke, dat  $f_n(w) dw$  scheef is, d.w.z. de gemiddelde  $\bar{w}_n \neq$  de meest waarschijnlijke  $\hat{w}_n$ .

Tippett maakte tabellen voor de  $\bar{w}_n$  en de  $\sigma_w =$  standaarddeviatie in  $f_n(w) dw$  voor allerlei  $n$ -waarden. Wanneer  $n$  toeneemt (laten we zeggen tot boven 20) neemt ook het gevaar toe, dat de tabel van Pearson ( $P$  als functie van  $W$ ) misleidend is in de praktijk, aangezien  $f_n(w)$  meer en meer gevoelig is voor kleine afwijkingen van de normaliteit in de staarten van de populatieverdeling.

In de praktijk komt er bovendien de moeilijkheid bij, dat men  $\sigma$  (meestal) niet kent; men moet zich behelpen met de standaarddeviatie  $s$  in de steekproef zelf, zijnde de beste schatting van  $\sigma$ . Mogelijk verschilt  $s$  veel van  $\sigma$ .

De tabel van Tippett (de "mean range is samples of size  $n$  taken from normal population, given in terms of standard deviation") is opgesteld voor  $n = 1, 2, 3, 4 \dots 1000$ . Het bezwaar is o.i., dat men er niet zoveel aan heeft. Voorbeeld:  $n = 10; (x_n - x_1) : s = 3.50$ ; Tippett's tabel levert  $\overline{(x_n - x_1)} : s = 3.08$ ; nu is  $3.50 > 3.08$ . Maar wat nu? Is de range "boventoevallig" groot? Met de Tippett-tabel komt men niet verder; wel met de Pearson-tabel, die leert  $P_{10}(W \geq 3.50) = 0.282$ ;  $P_{10}(W \geq 4.00) = 0.127$  en  $P_{10}(W \geq 4.50) = 0.047$ . Antwoord: Een  $W \geq 3.50$  bij  $n = 10$  heeft een toevalskans van 28% ( $\gg 5\%$ ).

Als men dus in een steekproef van 10 onafhankelijke elementen een spreidingsbreedte meet, die 3.50 keren de standaarddeviatie in deze steekproef bedraagt, behoeft men de nulhypothese, inhoudende, dat de steekproef genomen werd uit een normaal verdeeld universum, zeker niet op grond van deze grote 3.5 te verwerpen.

8.2

Er is nog een range-toets, die wij bij Brooks [7] aantreffen. Brooks houdt een zeer eigenaardige redenering: gegeven een normaal verdeeld universum  $\mu$ ;  $\sigma$ . De (positieve en negatieve) verschillen  $\Delta$  tussen de elementen in willekeurig uit deze populatie genomen tweetallen volgen een normale verdeling met  $\mu_2 = 0$  en met  $\sigma_2 = \sigma\sqrt{2}$ . Neem eens  $N = 7$  elementen uit de initiaal-populatie. Dan zijn er  $N(N-1) = 42$  verschillen  $\Delta$  (bij elke  $\Delta$  is een even grote, doch tegengestelde). De tabel bij de gestandaardiseerde normale verdeling leert, dat er een kans  $1/42$  is op een pos.  $\Delta \geq 1.97\sigma\sqrt{2} = 2.80\sigma$  of een kans  $2/42$  op een  $|\Delta| \geq 2.80\sigma$ . Anders gezegd: in  $42/2 = 21$  willekeurig (dus onafhankelijk zijnde) uit het  $\Delta$ -universum gekozen  $\Delta$ 's (die in doorsnede voor de helft positief, voor de helft negatief zijn) is er, gemiddeld genomen, één  $|\Delta| \geq 2.80\sigma$ . Nu zijn er onder de 7 elementen in een steekproef van 7 stuks  $\frac{42}{2} = 21$  positieve  $\Delta$ 's. Een van deze is de grootste; deze  $\Delta$  stelt de "spreidingsbreedte" voor. Dus concludeert Brooks: per steekproef van 7 elementen komt, gemiddeld beschouwd, één positieve  $\Delta \geq 2.80\sigma$  voor. Deze  $2.80\sigma$  noemt hij de "gemiddeld te verwachten range R". Aldus zijn de getallen in zijn tabel berekend. Hier volgt een gedeelte van deze tabel

| N  | R/ $\sigma$ | N  | R/ $\sigma$ | N   | R/ $\sigma$ |
|----|-------------|----|-------------|-----|-------------|
| 7  | 2.80        | 18 | 3.85        | 50  | 4.74        |
| 8  | 2.97        | 20 | 3.95        | 60  | 4.88        |
| 9  | 3.11        | 25 | 4.15        | 70  | 5.00        |
| 10 | 3.24        | 30 | 4.31        | 80  | 5.10        |
| 12 | 3.43        | 35 | 4.44        | 100 | 5.25        |
| 14 | 3.59        | 40 | 4.55        | 200 | 5.71        |
| 16 | 3.73        | 45 | 4.65        | 300 | 5.99        |

Hoewel Brooks' redenering niet volkomen streng is, ligt de curve W tegen N vlak bij de kromme van Tippett (gemiddelde spreidingsbreedte). Zie fig. 8. Wij hebben ook de punten  $W = \sqrt{N}$  aangebracht. Er is namelijk een vuistregel, welke zegt dat, voor  $N < 10$ ,  $\bar{R}/\sigma \approx \sqrt{N}$ .

Intussen is duidelijk, dat men met dit recept weer niet veel kan beginnen. Het vormt eigenlijk geen toets. Brooks rekent slechts de gemiddeld te verwachten spreidingsbreedte uit; maar als wij een andere waarde krijgen, wat is dan haar overschrijdingskans?

Er valt trouwens nog meer op te merken: zij  $n = N(N-1)$ . Een steekproef van  $N$  elementen noemt Brooks blijkbaar "equivalent" met een steekproef van  $n = N(N-1)$  onafhankelijke  $\Delta$ 's (pos. en neg.)<sup>1)</sup>. Geef aan een  $\Delta$  de naam "succes". Deze  $g = g(n)$  en wel gedefinieerd via  $\frac{1}{n}$  m.b.v. de gestandaardiseerde normale verdeling. Maak er het volgende probleem van: wat is de kans op  $i$  successen in  $n$  proefnemingen, als de succes-kans per proefneming  $p = \frac{1}{n}$  is? Het gemiddeld aantal successen per  $n$  proefnemingen is dan precies 1, gelijk wij het willen.

$K(i) = \binom{n}{i} \left(\frac{1}{n}\right)^i \left(1 - \frac{1}{n}\right)^{n-i}$ . Kans op 0 successen (op  $n$  proefnemingen) is  $K_0 = \left(1 - \frac{1}{n}\right)^n \approx e^{-1} = 0.36$  als  $n \gg 1$ ; en  $K_1 = K_0 \left(1 - \frac{1}{n}\right)^{-1}$ , iets kleiner dan  $K_0$ . En  $K_2 = \frac{1}{2} K_1$ ;  $K_3 = K_2 \frac{1 - 2/n}{1 - 1/n} \frac{1}{3} \approx \frac{1}{3} K_2$  als  $n \gg 1$ ; en

$K_4 \approx \frac{1}{4} K_3$  enz.

Wat ons hierbij opvalt, is, dat de kans op een steekproef van  $N$  elementen, waarin geen enkele  $\Delta$  de waarde  $g(n)$  bereikt of overschrijdt (dus een spreidingsbreedte  $< g(n)$ ) ongeveer even groot is (en wel ongeveer 0.36) als wanneer geëist wordt dat er precies één zulk een  $\Delta$  is. De kans op een steekproef met daarin tenminste één  $\Delta \geq g(n)$  is  $1 - 0.36 = 0.64$ . Men vraagt zich af of  $g(n)$  dan wel de goede  $n$ -functie is; liever zouden wij een  $n$ -functie wensen zódanig dat  $K(i \geq 1) \ll K(i=0)$  of liever nog  $K(i \geq 1) \ll 0.05$ . Wanneer wij dat eens proberen moeten wij  $p$  niet meer  $\frac{1}{n} = 1 : N(N-1)$  nemen, als boven, doch aanvankelijk onbekend beschouwen. Het gaat er dan om, dat  $K(i \geq 1) = 1 - K_0 = 1 - (1-p)^n = 0.05$ . Zodra  $n$  bekend is, is  $p$  bekend. Bijv.  $N = 10$ , dus  $n = 90$  en  $p = 0.0006$  (terwijl  $\frac{1}{n} = 0.0011$ ). De Gausz-tabel geeft 4.87 bij  $p = 0.0006$ , d.w.z.  $4.87 \sqrt{2} = 6.89$ , terwijl de redenering van Brooks zou leiden tot  $3.24$  ( $< 6.89$ ). Onder woorden: er is een kans

1) Heeft Brooks gelijk? Bij bijv.  $N = 4$  getallen bijv. 1.0 2.3 4.2 5.8 kunnen we  $3 \times 4 = 12$   $\Delta$ 's onderscheiden, n.l.  $1.0 - 2.3 = -1.3$  (en  $2.3 - 1.0 = 1.3$ );  $1.0 - 4.2 = -3.2$  (en  $4.2 - 1.0 = 3.2$ );  $1.0 - 5.8 = -4.8$  (en  $5.8 - 1.0 = 4.8$  de spreidingsbreedte);  $2.3 - 4.2 = -1.9$  (en  $4.2 - 2.3 = 1.9$ );  $2.3 - 5.8 = -3.5$  (en  $5.8 - 2.3 = 3.0$ ) en  $4.2 - 5.8 = -1.6$  (en  $5.8 - 4.2 = 1.6$ ) Zijn deze  $\Delta$ 's onafhankelijk? Natuurlijk niet. Als wij de  $\Delta$ 's noemen  $\Delta_1, \Delta_2 \dots$

$\dots \Delta_{\frac{1}{2}(n-1)n}$  en  $\Delta_1^* = -\Delta_1 \dots \Delta_{\frac{1}{2}(n-1)n}^* = -\Delta_{\frac{1}{2}(n-1)n}$ , dan is altijd  $\sum_{i=1}^{\frac{1}{2}(n-1)n} (\Delta_i + \Delta_i^*) = 0$ .

van 5 % op een steekproef van 10 elementen (uit een normaal verdeeld universum) met een range  $\geq 6.89 \sigma$ . Toch ligt deze 6.89 veel hoger dan de Pearson-waarde 4.45 (zie fig. 8), hetgeen er op duidt, dat er iets mis is aan onze redenering, maar dan ook aan die van Brooks. Vermoedelijk ligt dit in het feit, dat, ook al zijn de N elementen onafhankelijk, de N (N - 1)  $\Delta$ -waarden het nimmer zijn.

8.3

Brooks geeft nog een methode aan, volgens welke men kan onderzoeken of de gevonden spreidingsbreedte door louter toeval zo groot kan zijn. De methode houdt direct verband met wat wij onder 2.1.2 de "constellatie-toets" noemden. Hij redeneert a.v.: stel er zijn N metingen x, y gedaan, tussen welke punten door men een gladde kromme i zou willen trekken. Deze y's leveren een standaarddeviatie  $s_y$ . Hij legt nu aan weerszijden een curve parallellel aan de (eigenlijk nog niet gevonden) curve c, op zulk een afstand D, dat, gemiddeld genomen, twee van deze n punten op of buiten de parallelcurven komen te liggen. Deze afstand D (uitgedrukt in  $s_y$ -eenheden) wordt opgezocht uit de tabel voor de gestandaardiseerde normale verdeling bij een overschrijdingskans  $2/n$ . Letterlijk: "..... one of the plotted points should lie on or near to each of these outer curves .....". Waarom "on or near"? Bedoelt Brooks "near" letterlijk? Bovendien wat is "near"? En hoe gaat de toetsing nu verder? Hier geeft Brooks geen antwoord op. Naar onze mening behoort men a.v. te redeneren: de kans op een constellatie van bijv. 10 punten, met daaronder geen enkel punt, dat "on or near" de op bovenbeschreven wijze geconstrueerde parallelcurven ligt, is  $K(0) = \binom{10}{0} \left(\frac{2}{10}\right)^0 \left(1 - \frac{2}{10}\right)^{10} = \left(1 - \frac{1}{5}\right)^{10} \approx (e^{-1})^2 =$

0.14 (exact 0.11). Deze 14 % is nog behoorlijk groot. Een ander uiterste: stel er liggen 3 of meer der 10 punten "on or near the outer curves". Welke is de kans op zulk een constellatie?  $K(\geq 3) = 1 - K(0) - K(1) - K(2) = 0.29$  (exact 0.32). Toch niet bepaald een kleine kans. Wij berekenen deze kansen om te ontdekken waarom Brooks wil hebben, dat er maar één punt op of nabij de parallelcurven ligt. Dit brengt ons ertoe de vraag te stellen: voor welk aantal (x) punten "on or near" de bandgrenzen (geconstrueerd op basis  $2/N$ ) liggende, heeft de constellatie een kans van 0.05? Dus los x op uit

$$K(\geq x) = 1 - K_0 - K_1 - \dots - K_{x-1} < 0.05 \text{ met } K_i = \binom{N}{i} \left(\frac{2}{N}\right)^i \left(1 - \frac{2}{N}\right)^{N-i}.$$

Voorbeeld:  $N = 10$   $p = 0.2$  (halve bandbreedte  $0.84 s_y$ ). Er komt:  $x = 5$ . Het zal misschien verbazen, dat pas bij 5 op de 10 punten op of buiten de op basis van  $2/10$  geconstrueerde parallelcurven een constellatie aanwezig is, die uiterst zeldzaam (kans 0.05) genoemd zou mogen worden. (terwijl Brooks van 2 op de 10 punten spreekt).

Conclusie: ook in deze vorm lijkt ons de toets van Brooks niet veel waard. Hoe deze te wijzigen? Men zal aanvankelijk de kans  $P$ , die door Brooks  $2/10$  genomen worde ( $N = 10$ ), onbekend moeten laten en zo moeten kiezen, dat er op een constellatie met  $10.p$  of meer dan  $10p$  punten op of buiten de met deze  $p$  geconstrueerde parallelcurven een kans van 0.05 is. D.w.z.  $K (\geq 10.p) = 0.05$ , voor  $N = 10$ . Wat is  $p$ ? Een schatting: 0.005, want  $K (\geq 10.0.0.5) = K (\geq 1) = 1 - K_0 = 1 - 0.95 = 0.05$ .

De gezochte  $p^*$  is dus ong. 0.005 en de halve bandbreedte (opgezocht bij 0.005 in de tabel van de gestandaardiseerde normale verdeling) is  $2.6 \sigma$  (i.p.v.  $0.84 \sigma$ , volgens Brooks' redenering op basis van  $p = 2/10$ ). Interpretatie: bij een bandbreedte van  $2.6 \sigma$  heeft een constellatie van 10 punten met 2 of meer punten op of buiten de parallelkrommen een kans van 0.05.

Het is duidelijk, dat in dit alles de trendfunctie (de gladde kromme tussen de punten door) bekend moet zijn; hoe zou men anders de band kunnen trekken? Het is denkbaar, dat men deze trend vrij goed op het oog schatten kan. Als men daarna de band tekent, kan men uittellen hoeveel punten buiten de band liggen. Wanneer er dat zovele zijn, dat de ganse constellatie zeer onwaarschijnlijk wordt (d.w.z. op grond van toeval een kans heeft kleiner dan 5 %), dan "is er iets mis", d.i. dan zullen sommige van die "outer points" "echte" pieken zijn (mogelijk zijn er ook "echte" dalen!).

Wij zijn hiermede zo uitvoerig geweest om duidelijk te maken dat men met de door Brooks in zijn handboek genoemde toetsen op zijn qui-vive moet zijn, hoe interessant de in zijn boek aangesneden kwesties ook mogen zijn. Al te vaak vraagt men zich af, wat Brooks nu eigenlijk zeggen wil. Juist als het er op aankomt, laat hij de lezer in het onzekere. Voorbeeld: hij tekent in fig. 59, pg. 259, de gemiddelde dagelijkse temperatuur per dag van de maand Januari, te Greenwich (50 jarige gemiddelden). Natuurlijk liggen de punten zeer grillig. "Boventoevallig" grillig? Men wil er graag een gladde kromme door heen leggen, welke vermoedelijk omstreeks half Januari zijn laagste punt heeft. Nu moet men eerst  $N = 31$  door  $N_e$  vervangen. Zeker  $N_e < N = 31$ , vanwege de persistentie.



Brooks zegt  $N_e = 10$ . Eerst wordt de range getoetst; deze is  $3.07^\circ\text{F}$ . Zie tabelletje onder 6.2. Uitkomst: deze range kan zo breed zijn louter door toeval. Dus zijn wij gerechtigd een gladde kromme tussen de punten door te trekken. Doch hoe? Daarna legt Brooks de band op basis van  $p = 2/10$ . Als dan de band getrokken is telt hij één punt (van het effectieve aantal 10), dat "on or near the outer curves" ligt. Conclusie: "the distribution of the crosses in relation to these outer curves confirms that the central curve is a reasonable estimate of the trend of temperature in January".

Het ligt voor de hand dit procédé analoog ook toe te passen op de punten in onze fig. 4. Zie bijv. naar de Augustusmaand. De punten liggen zeer grillig. Wij moeten dan eerst onderzoeken of wij gerechtigd zijn tussen deze 31 punten een gladde kromme te leggen (te contrueren), d.w.z. wij moeten onderzoeken of sommige der "piekachtig"-aandoende punten louter door toeval zo hoog liggen. Dit onderzoek vindt weer aan de spreidingsbreedte plaats. Als deze reeds mét trend (want er is zeker trend) niet significant is, dan zeker ook niet, indien voor trend gecorrigeerd wordt. Vervolgens (gesteld, dat wij tot significantie-afwezigheid concluderen) moeten wij die "gladde kromme" leggen. Hoe? Via een vereffeningsprocédé? Welk? Indien de kromme de ware goed benadert en wij leggen twee parallelcurven, zoals wij boven beschreven volgens het door ons verbeterde procédé, dan mogen er ook niet meer dan 2 punten op of buiten zulk een parallelcurve liggen.

9

#### Literatuur

- [1] Australian Journal of Physics 6 490 1953
- [2] Australian Journal of Physics 7 354 1954
- [3] Australian Journal of Physics 7 358 1954
- [4] Bull. Amer. Met. Soc. 35 90 1954
- [5] Indian J. Met. and Geoph. 5 356 1954
- [6] Biometrika 32 301 1941/42
- [7] Handbook of Statistical Methods in Meteorology, 1953, pg. 43

S U M M A R Y

A statistical study of the article of E.G. Bowen, entitled  
"The influence of meteoritic dust on rainfall"  
(Austr. J. of Phys. 6 490 1953)

This report consists of 9 chapters.

In chapter 0 the reason is stated why it was necessary to study the above mentioned article in a statistical way as critically as possible. The results of our study proved to be important for our own investigations, namely the analysis of the annual course of the daily rainfall and the calculation of the frequency of the total rainfall in periods of  $k$  consecutive days ( $k = 1, 2, 3 \dots$ ), for each of the 24 stations in the country.

In chapter 1 we summarize the main results of Bowen's investigation.

In chapter 2 we deal with our statistical remarks and observations with respect to the way in which Bowen tackles his problem, which is one of singularities.

Chapter 3 discusses the comments of D.F. Martyn and W.C. Swinbank (Australia) and of J. Neumann (Israel). Moreover we cite some comment from the Annual Report 1953/54 of the Commonwealth Scientific and Industrial Research Organization Section of Meteorological Physics. All these articles intend to demonstrate that Bowen was not forced to state that meteoritic dust should influence the rainfall. It is not necessary (necessary in a statistical meaning) to take into account factors different from the ordinary climatological or meteorological ones.

In chapter 4 we mention a confirmation of Bowen's hypothesis, but in our opinion the confirmation is of little importance.

Chapter 5 treats some passages of an article "Der Jahresgang der Niederschlagswahrscheinlichkeit auf dem Somblick" by F. Steinhauser. This article deals with singularities (peaks) on specified days in the annual course of the daily-rainfall frequency.

In chapter 6 we examine dutch material, namely the daily rainfalls totalized over 81 years; station Hoofddorp. We test statistically the many peaks in the annual course of daily rainfalls while paying attention to the persistence between the daily falls.

In chapter 7 we formulate a new problem of coincidence. The solution of this problem should become the conclusive answer to the statistical question which was not treated in the correct way neither by Bowen himself nor by the other investigators, mentioned above. Due to lack of time, we do not solve this problem fully, but we indicate the way how to continue. Of course the persistence ought to be considered.

In chapter 8 we treat the tests we used in our report. The following details are of importance.

In the first place we cite Bowen: "..... If the daily rainfall figures are examined, however, it is apparent that in some localities there is a marked tendency for heavy falls of rain to occur on certain days rather than on others, and for this pattern to be repeated year after year. For purposes of illustration the total rainfall of Sydney for January and the first five days of February for the period from 1859 to 1901 is given in fig. 1a. It shows a general increase throughout the month, conforming to the seasonal pattern and two exceptional peaks occurring respectively on January 12 and January 22. The magnitude of each peak is approximately twice the mean and their departure from the mean about four times the standard deviation of the remainder. The corresponding curve for the period from 1902 to 1944 is given in figure 1b and it seen that two similar peaks occur on January 12 - 13 and January 23, while another appears on January 31. It will be shown later that this third peak is a significant one.

The near coincidence in time of the first two peaks was unexpected enough to stimulate an examination of the rainfall figures of other stations in corresponding latitudes to see whether they too showed any unusual characteristics. The surprising discovery was made that many stations over a wide area tended to show peaks of similar magnitude on nearly the same days. Figures for periods of approximately 50 years are available for seven stations in the southern hemisphere extending over 180° in longitude from South Africa to New Zealand. These are listed in top of figure 3, together with the dates on which the rainfall exceeds the mean by 50 percent." So far Bowen himself. (underlining by us).

Bowen regards these strange matters as significant and states that it is impossible to understand the facts by climatological factors. There must be an extraterrestrial cause: meteoritic dust. We are not going to discuss his physical reasoning, but we only criticize the way in which he tests the results of his rainfall analysis as to the significance; his way cannot be correct. It is impossible to repeat our computations

al over again. The main point is that we prove by means of the so called range test and constellation test that each of the figures 1a and 1b is a random result. It is curious that Bowen has computed the variance in the 36 totalized daily falls (figures 1a and 1b) only on basis of the non-peak-days. In this way he finds a variance which is too small and so it is more probable that the same peaks are statistically real.

Next we think that also the coincidence and non coincidence of peaks (and coincidence of troughs) is caused by chance. Stress is laid on the fact that these tests should be applied carefully in consequence of the persistence between the 43 years-total rainfalls of consecutive days. We believe that Bowen has neglected this fact; we do not know the numerical value of this persistence for Sydney; so we substituted some reasonable values. In the same way the constellation of the points in figure 2 proved to be a random result if we apply the  $\chi^2$ -test with respect to the hypothesis that the straight line 2 is a good approximation of the annual course. In this figure the ordinate indicates the number of times that the day, indicated by the abscis, received more than  $2\frac{1}{2}$  inch. The period is 1919 - 1949. We think that Bowen has totalized several stations on the British Isles. He does not mention these stations.

Swinbank (Austr. J. of Phys. 7 354 1954) says that many peaks are caused by only a few days of large rainfalls. Daily falls of 10 or more inches are not seldom. In this way it may occur that a specified day (for instance 12 January) is not a peak-date if all years are considered with exception of the year in which this date received an extremely large fall.

Martyn (Austr. J. of Phys. 7 358 1954) states that the same cyclone, wandering over Eastern Australia, has caused the peak of 8 January in the 43 years figure of West Leichhard and the peak of 9 January at Cloncurry, and so on finally the peak of 13 Jan. in figure 1a of Sydney. In this way all these peaks are correlated by means of one depression and not by means of the meteoritic dust.

Neumann applies the turning point test. A turning point (t.p.) is a relative maximum or minimum. He counts the number of t.p. in figures 1a and 1b and compares with the mean number if a random sample of 36 elements is taken from a normal population and plotted as points in a

figure. Kendall has derived the probability distribution for this number of t.p. Neumann also examines the frequency of the phases of various lengths and he compares this frequency with the theoretical one, caused by chance (the phase is the distance between two consecutive t.p.) Finally he concludes that both figures may be considered as random results. So he does not regard the peaks as statistically real (oral communication)

In figure 4 the daily rainfalls in the dutch station Hoofddorp, averaged over 81 years have been plotted. Attention is drawn to the fact that the points lie very irregularly notwithstanding the fact that the period is fairly large. Next we examined the range in each month. Of course there is an annual course, but probably this course may be neglected within a month being a fairly small interval of the year. See table pg 26.  $R$  = range in mm;  $s$  = data standard deviation for instance between the 81 1 January-falls; this  $s$  is considered equally to the 2 Jan.-s and 3 Jan.-s and so on up to and including the 31 Jan.-s. In this way each month has its own  $s$ -value. Of course the standard deviation of the 1 Jan.-falls, totalized over 81 years, is  $s \sqrt{81}$ . Since we do not know the persistence,  $\sqrt{\quad}$  in the series of 36 consecutive mean daily rainfall amounts  $\sqrt{\quad}$ , we substituted some values for the effective number  $\sqrt{\quad}$  Of course  $N_e = 5$  indicates a very strong, nearly impossible persistence. Column 5 shows the threshold values of significant ranges for these  $N_e$ -values. In 6, 7 and 8 we mention the values of  $N_e$  if the persistence coefficient is  $\rho = 0.15, 0.25$  or  $0.40$ . Each significant range is indicated by a + sign. In this way we treated the month January, March, May, August and September (the other months certainly do not have a significant range). It is obvious that the range depends on the peak value and the lowest trough.

In 7 we formulate the following interesting problem. Let a normal universum  $(\mu; \sigma)$  be given. We draw a random sample of 31 elements and plot these elements as a time series  $x_1, x_2 \dots x_{31}$ . What is the probability  $p$  of getting a peak in a fixed place  $i$ ? The definition of a peak may be given in many ways. We introduce three possibilities:

- (1)  $x_{i-1} < a; x_i > b; x_{i+1} < a$       (2)  $x_i - x_{i-1} > c; x_i - x_{i+1} > c$   
and (3)  $x_i - x_{i-1} > c; x_i > b; x_i - x_{i+1} > c$ .

Now  $p = K_a^2 \cdot K_b$  if  $K_a$  = probability of  $x < a$

$K_b$  = probability of  $x > b$ .

---

$\sqrt{\quad}$  (the apparent number is  $N = 36$ ) : 20, 15, 10, 5.

In such a constellation there must be at least one peak and one of these peaks must be situated in the given place  $i$ . Now we do this  $A$  times ( $A$  figures). Then the probability that on  $B$  figures ( $0 \leq B \leq A$ ) there is a peak on  $i$ , is given by

$$P_i = \binom{A}{B} p^B (1-p)^{A-B}.$$

Bowen used:  $A = 2$ ;  $B = 2$ . Next we choose  $i = 1$  or  $2$  or  $3 \dots$  or  $31$ . So  $P = 31P_i$ . Next the definition of a coincidence of peaks is made less exact, in this way: on the first figure the peak must be situated at place  $i$  and on the second one on  $i-2$   $i-1$   $i$   $i+1$  or  $i+2$ . So  $(P) = 5P$ .

These  $(P)$  we have calculated for  $3 \times 2 = 6$  cases, see the scheme pg. 31 ( $a = \mu$ ;  $b = \mu + \sigma$  or  $\mu + 2\sigma$ ;  $c = \sigma\sqrt{2}$  or  $2\sigma\sqrt{2}$ ).

Using the definition (2) we state that the difference  $\Delta$   $x_p - x_{p-1}$  ( $p = 1, 2 \dots$ ) satisfies a normal distribution with  $\mu_\Delta = 0$  and  $\sigma_\Delta = \sigma\sqrt{2}$ . If there exists persistence this  $\sigma_\Delta$  should be replaced by  $\sigma_\Delta^* = \sigma\sqrt{2(1-\rho)}$ ,  $\rho$  being the persistence coefficient of the first order. In this way it is possible to modify the computations starting with definition (2). We have substituted  $\rho = 0.25$ . The probability of a "near" coincidence of peaks when drawing two figures becomes 0.035. The peak is defined as  $\Delta_1 = x_i - x_{i-1} \geq \sigma_\Delta$  with  $\Delta_2 = x_i - x_{i+1} \geq \sigma_\Delta$ .

Next we should compute these probabilities also for peaks which have a breadth of for instance 2 or 3 days and for the coincidence of more than one peak, etc., etc. This is not done in this report.

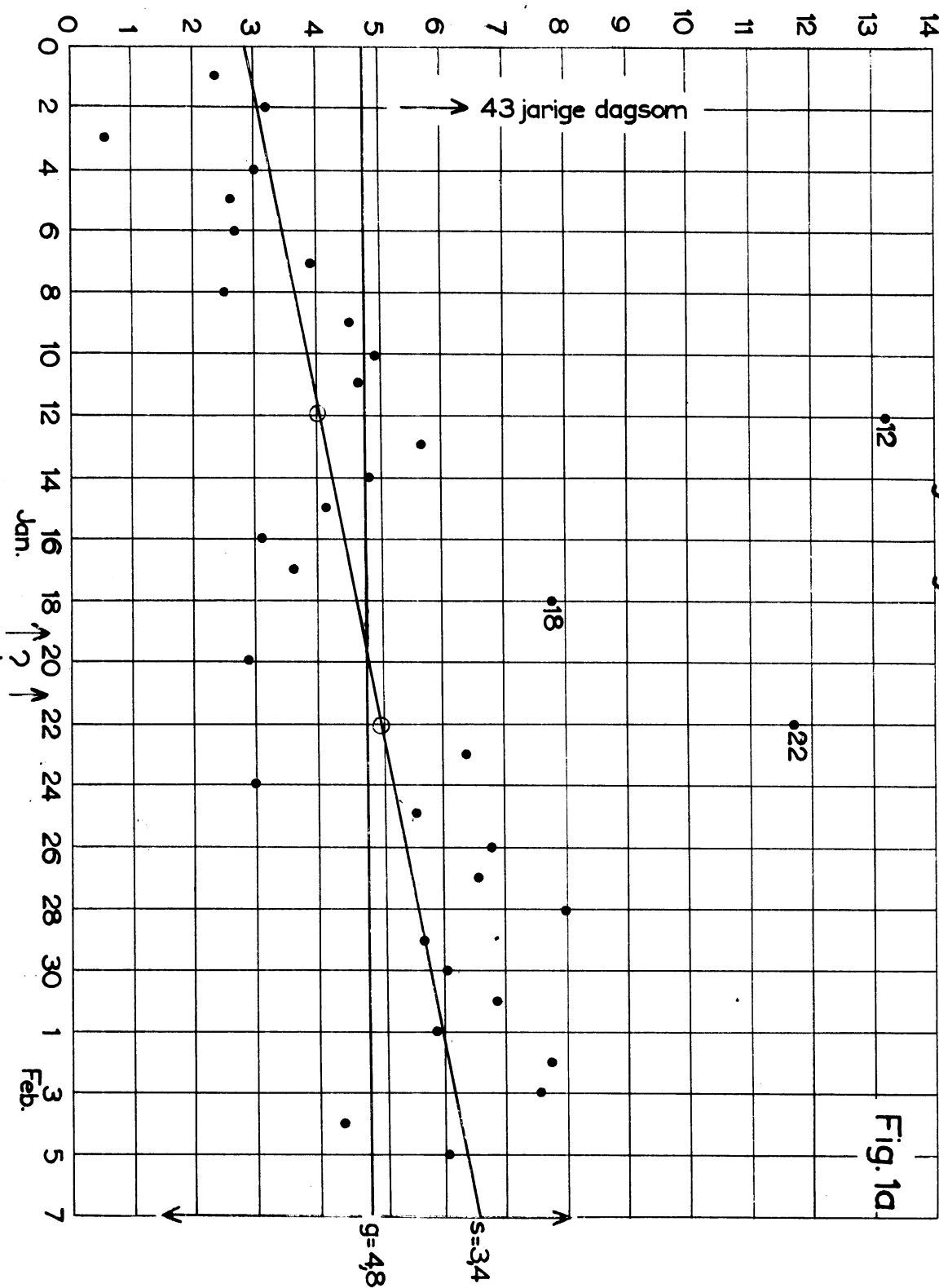
In chapter 8 the range tests, developed by Pearson - Tippett and by Brooks (Handbook of Statistical Methods in Meteorology pg. 43 1953) are discussed. We show that the Brooks's test cannot be correct (and propose some modifications), but if we do not ask for exact results this simple range test may still be used. We should prefer the test of Pearson and the corresponding table. (See fig. 8). The table gives the probability of getting a range  $R \geq W \cdot \sigma$  when drawing a random sample of  $N$  elements from a normal population ( $\mu; \sigma$ ):

$N = 2, 3, 4, \dots, 20;$   
 $W = 0.00, 0.05, 0.10 \dots, 7.25.$

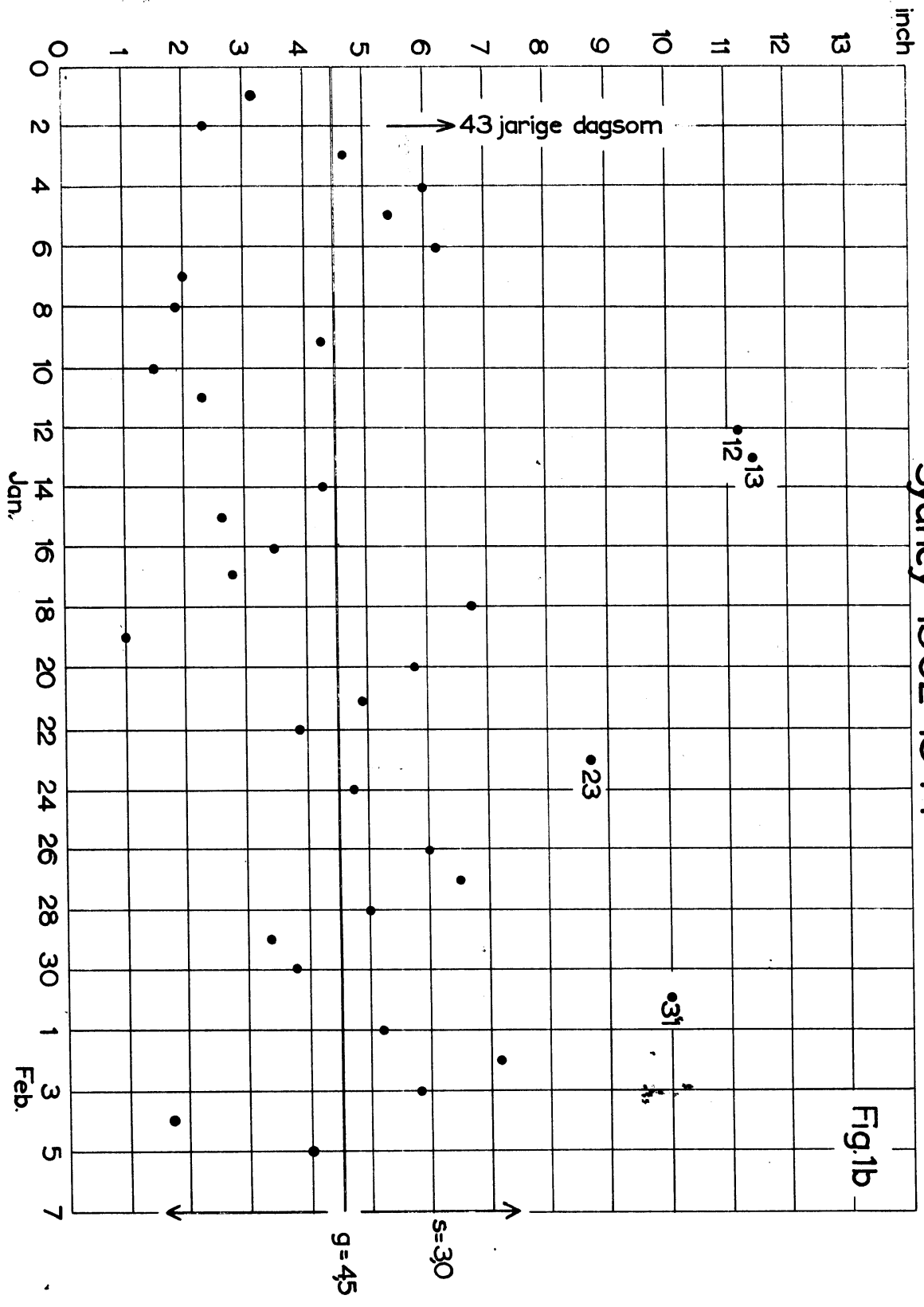
inch

# Sydney 1859-1901

Fig. 1a

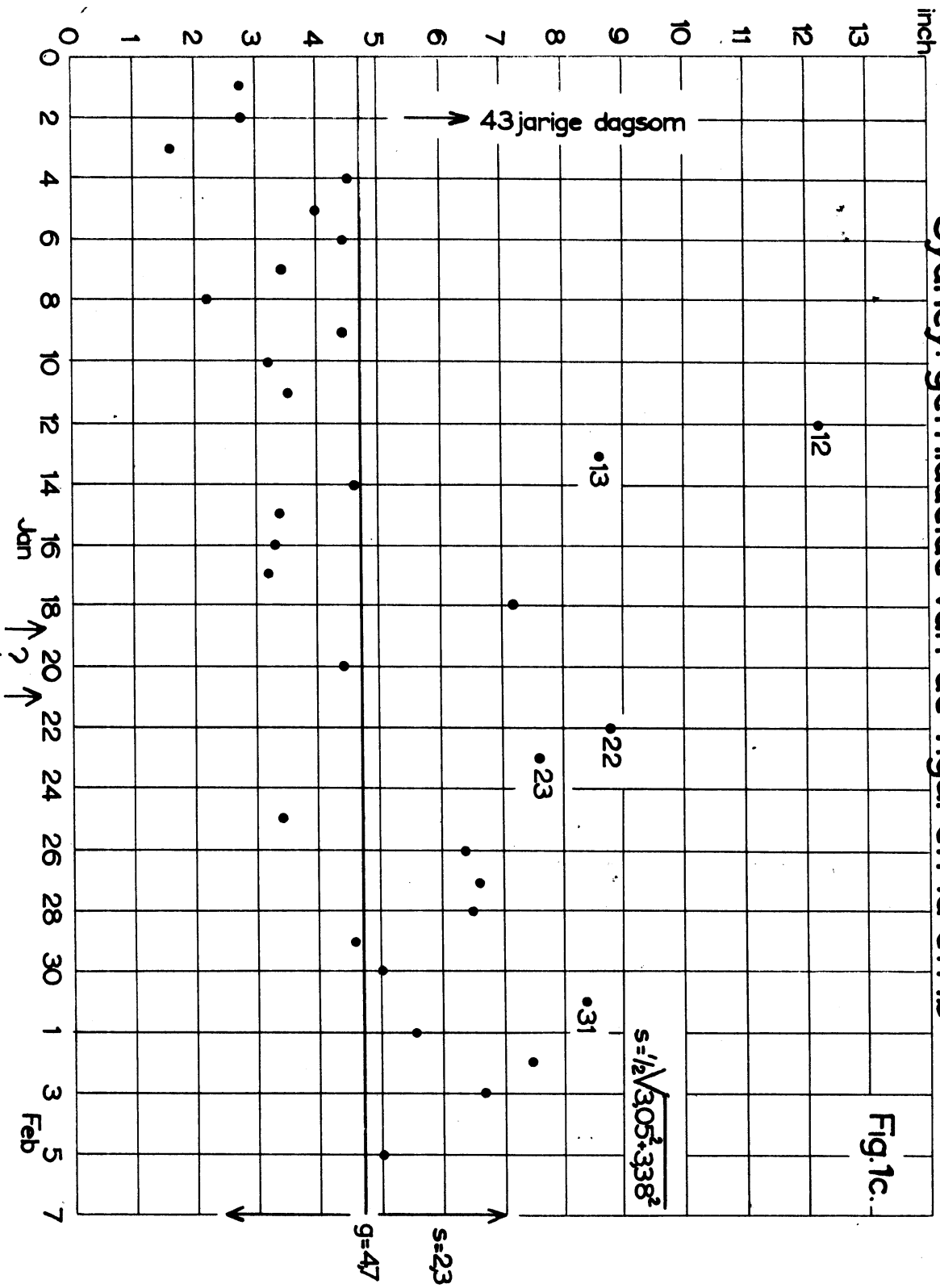


# Sydney 1902-1944





# Sydney: gemiddelde van de figuren 1a en 1b



# Britse Eilanden 1919-1949

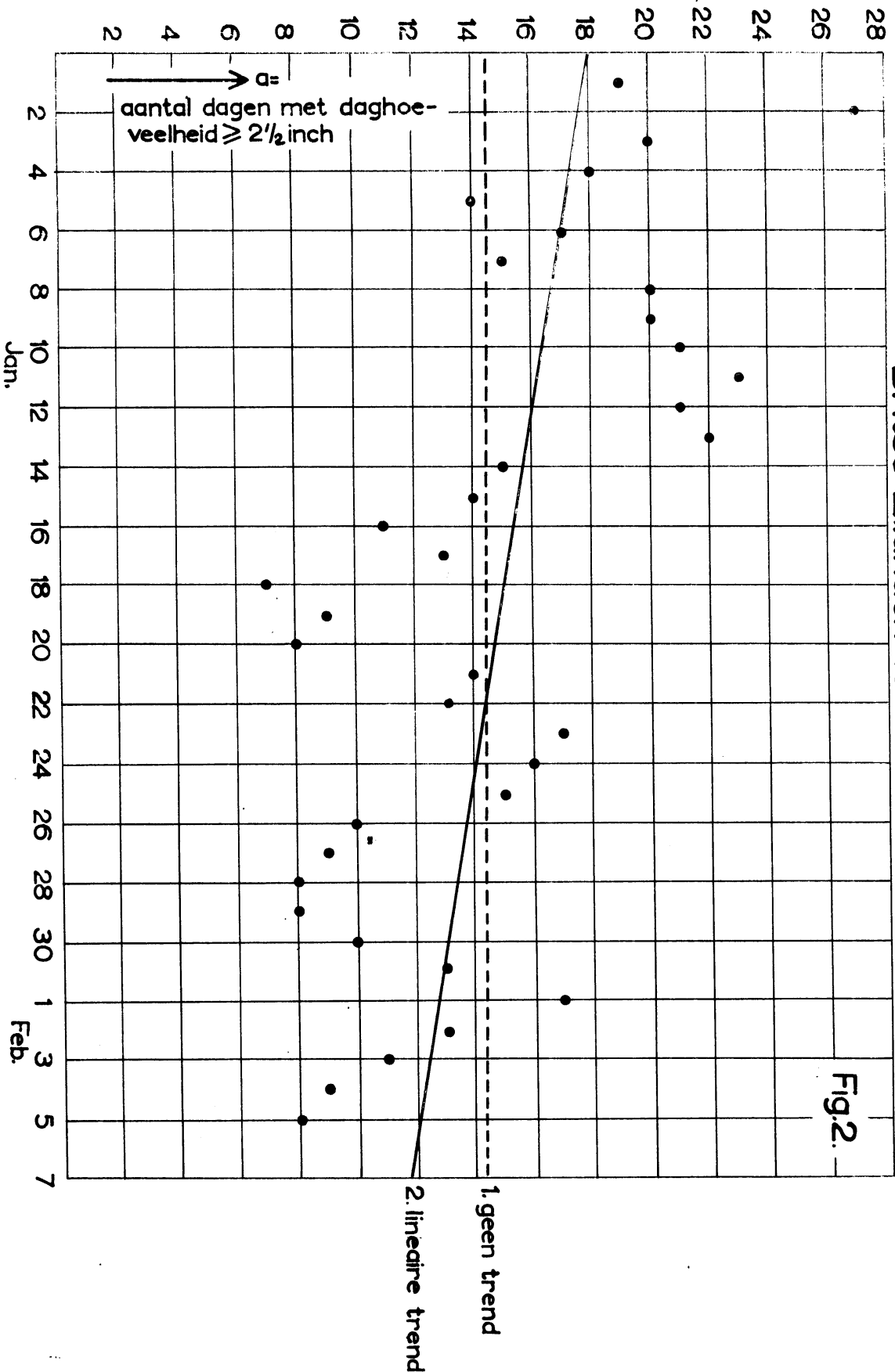
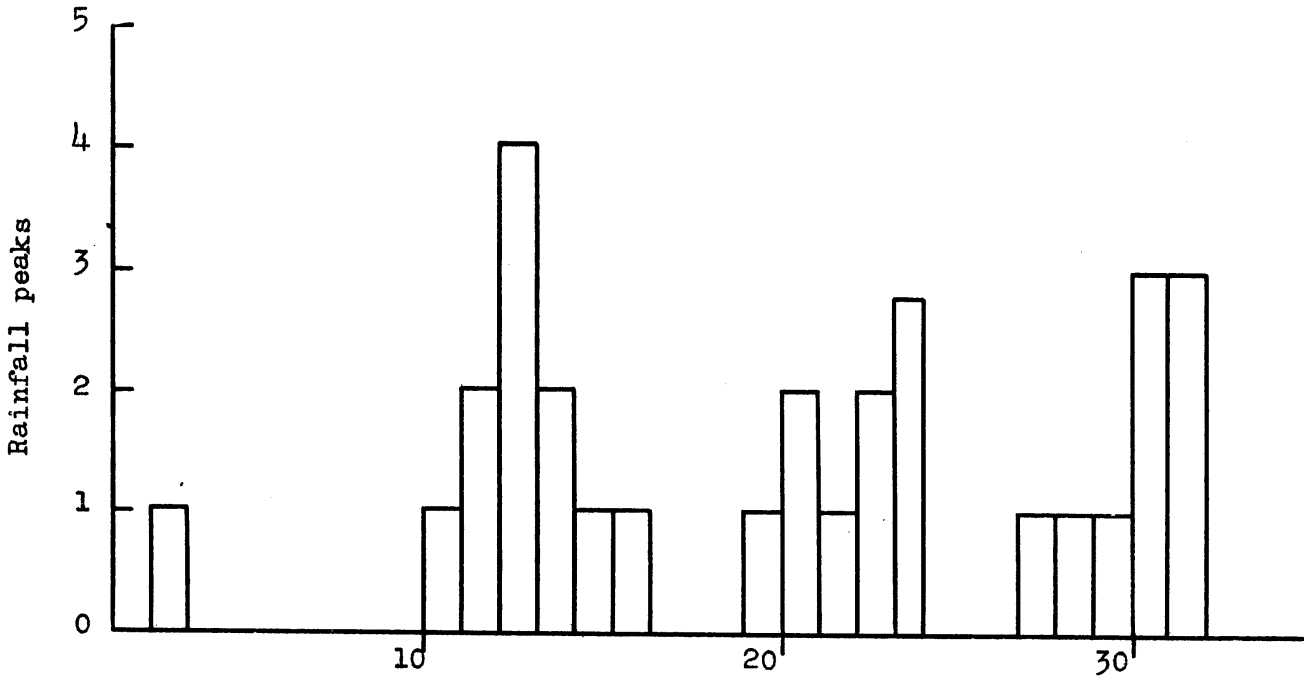


Fig.2.

# Fig 3

## Dates of rainfall peaks

| Place            | Period  | Dates on which Peaks of Rain Occur |              |         |                     |
|------------------|---------|------------------------------------|--------------|---------|---------------------|
|                  |         | Jan. 2                             | Jan. 11 & 13 | Jan. 23 | Jan. 28 & 31        |
| Durban .. ..     | 1900-53 |                                    | Jan. 12 & 16 | Jan. 22 | Jan. 30             |
| Perth .. ..      | 1907-52 |                                    | 15           | 20 & 24 | 31                  |
| Alice Springs .. | 1900-52 |                                    | 14           | 21      | 28 & 31             |
| Sydney .. ..     | 1900-49 |                                    | 12 & 13      | 23      | 31                  |
| Brisbane .. ..   | 1900-49 |                                    | 13           | 24      | Feb. 1              |
| Auckland .. ..   | 1900-53 |                                    | 13           | 21      | Jan. 29 &<br>Feb. 1 |
| Christchurch ..  | 1905-53 |                                    | 14           | 24      | -                   |
| Great Britain .. | 1919-49 | Jan. 2                             | 11 & 13      | 23      | Feb. 1              |



January

Fig. 3. - Distribution of rainfall peaks in January for Durban, Perth, Alice Springs, Brisbane, Sydney, Auckland, Christchurch, and the British Isles.

Hoofddorp gem. daghoeveelheid neerslag  
(gem. over 81 jaren)

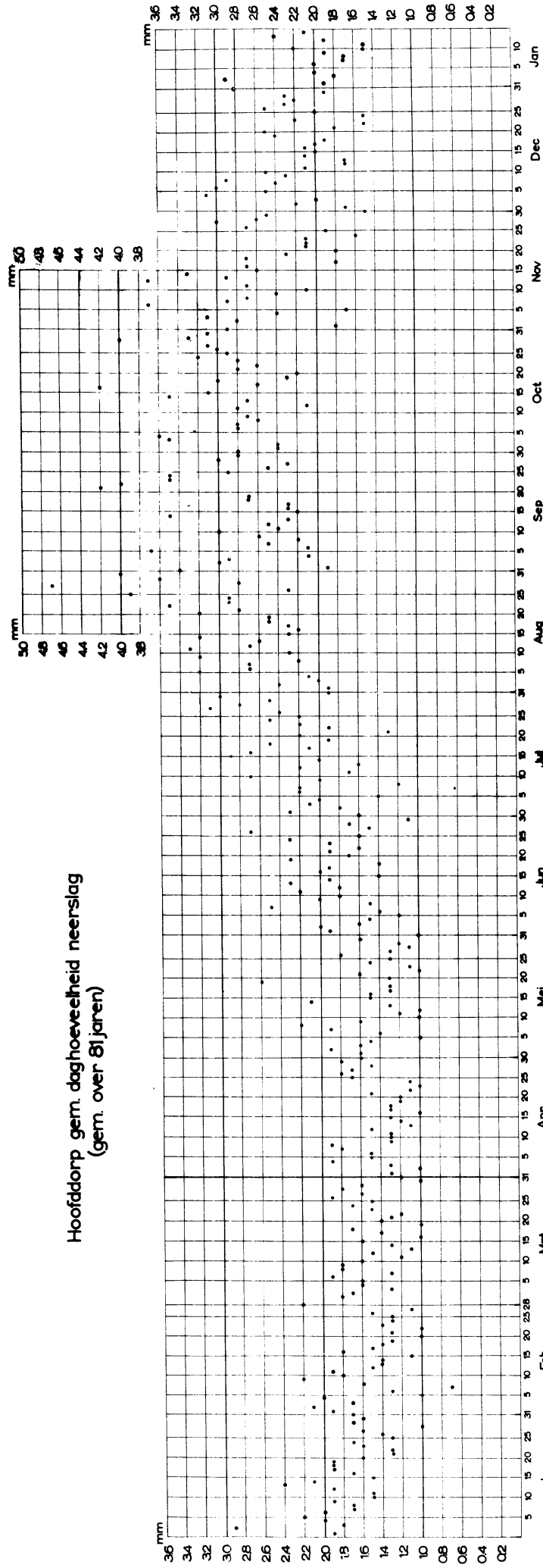


Fig. 4

# 81j dagsom Hoofddorp 1867-1953

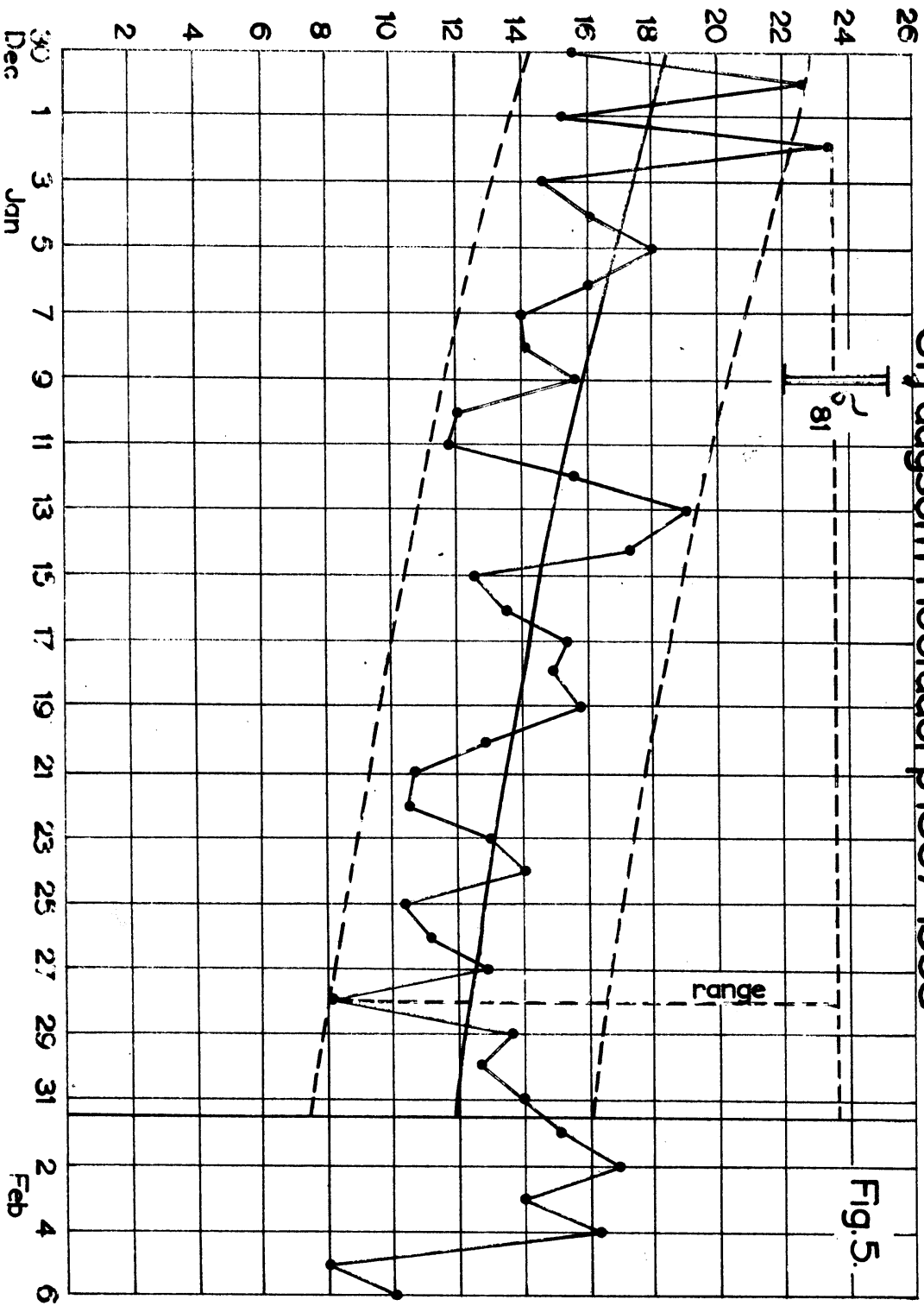
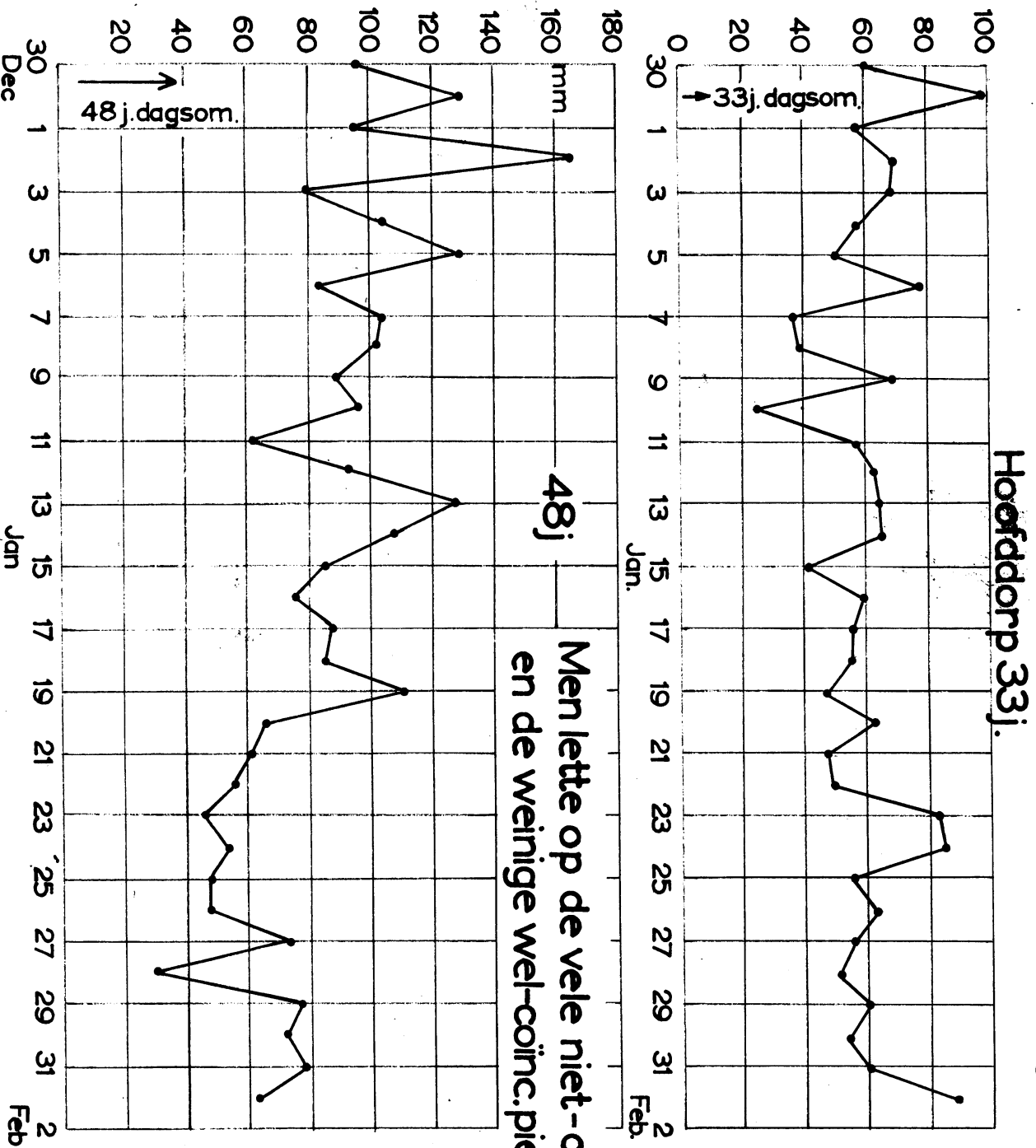
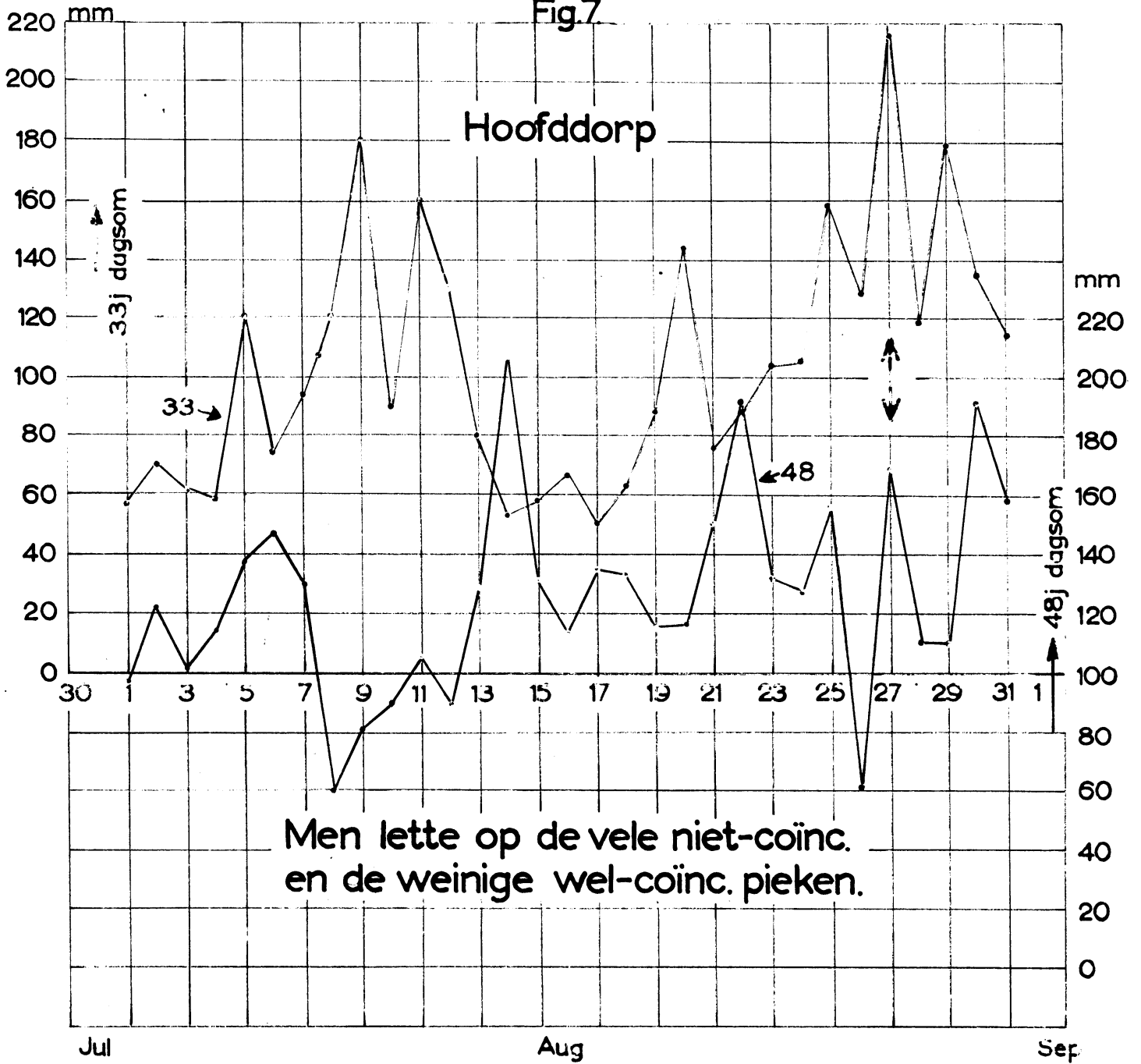


Fig.6.



Men lette op de vele niet-coïnc.  
en de weinige wel-coïnc.pieken.

Fig.7



# Range-Toets

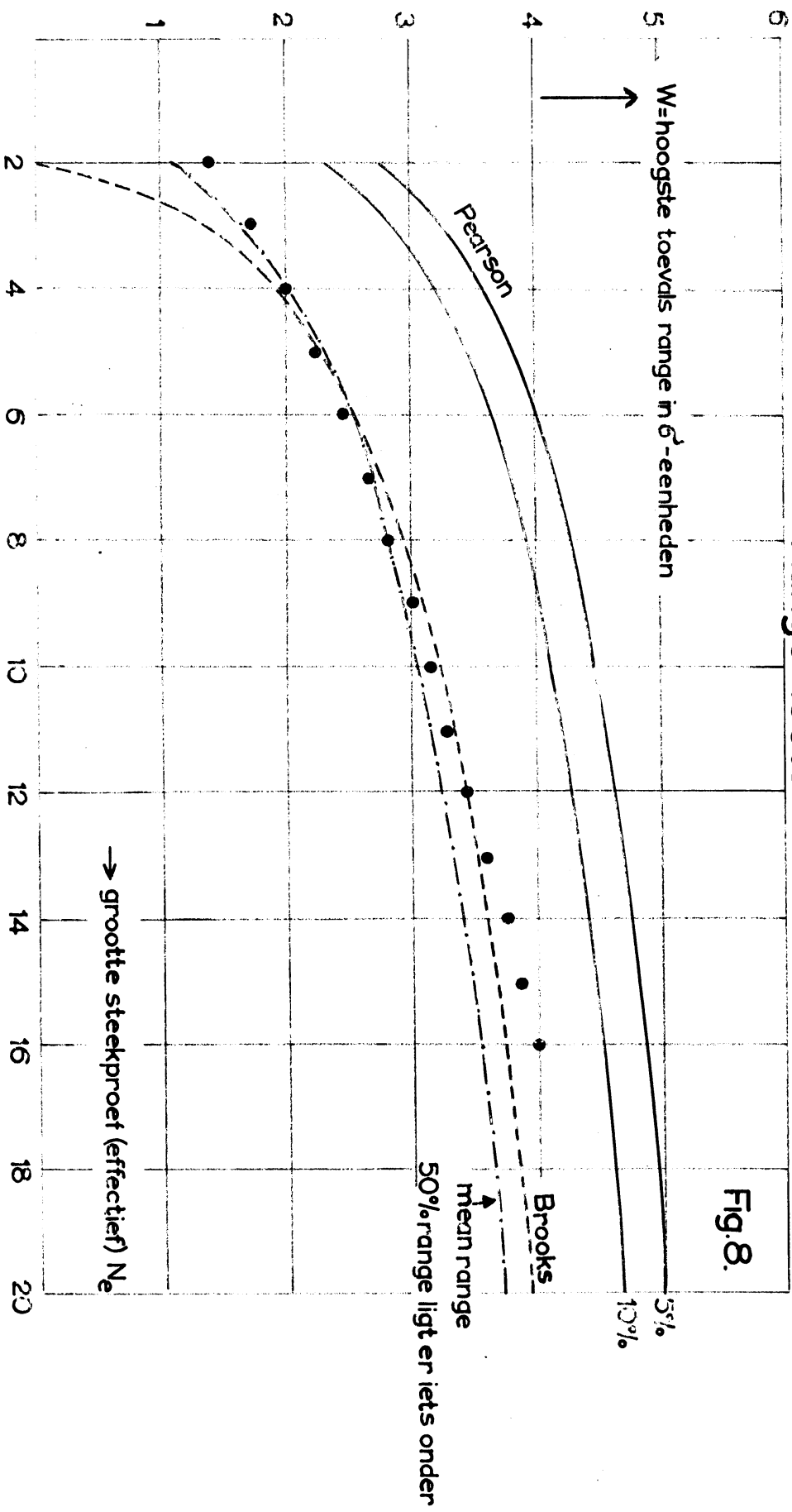


Fig. 8.