

- 1 JULI 1957

KONINKLIJK NEDERLANDS  
METEOROLOGISCH INSTITUUT

Verslagen V-3  
(R III 191-1957)

Afd.: Klimatologie en  
Landbouwmeteorologie.

Bespreking van enige voordrachten gehouden tijdens het 4de  
INTERNATIONAAL CONGRES VOOR ALPINE METEOROLOGIE  
te Chamonix, 17-20 september 1956

door

551.585.7

Dr. C. Levert

Inleiding.

In dit rapport worden twee der 47 voordrachten uitgebreid in beschouwing genomen (de eerste speciaal op verzoek van Prof. Berlage), terwijl aan enkele andere een korte bespreking wordt gewijd.

De bedoeling is commentaar en -waar nodig- kritiek te leveren, o.m. op verzoek van de auteurs zelf, maar verder om vast te leggen op welke punten wij bedacht moeten zijn in het geval, dat wij in ons land analoge onderzoekingen zouden gaan ondernemen.

Tenslotte is ook mijn eigen voordracht opgenomen.

Deze voordrachten zijn:

1. L. Martel: Utilisation des données pluviométriques pour la prévision des apports dans les réservoirs de haute montagne.
2. L. Serra: La représentation des facteurs climatologiques par des courbes d'égalé probabilité.
3. M. Bossolasco : Densité de la neige et température de l'air.
4. M. de Quervain: Modell eines Tagbogenschreibers.

Kon. Ned. Meteor. Inst.  
De Bijl.

1. L. Martel.

Utilisation des données pluviométriques pour la prévision des apports dans les réservoirs de haute montagne.

A. Inhoud van de voordracht.

De "Service de la production hydraulique d'E.D.F.", die belast is met de exploitatie der hydroëlectrische installaties, houdt zich sedert de nationalisatie bezig met het vraagstuk der verwachting der wateraanvoeren in die bassins, die de voornaamste installaties voeden, vooral teneinde een betere beheersing der energiereserves, aanwezig in de door de grote stuwdammen opgeslagen watermassa's, mogelijk te maken.

De eerste verwachting met betrekking tot de alpine-stromen werd in 1949 uitgegeven. In 1956 heeft de dienst 460 verwachtingen gemaakt met betrekking tot 25 bassins, waarvan er 210 slaan op 11 bassins in de Alpen (vóór sept.).

Het gaat allereerst om het verband tussen de te verwachten aanvoeren enerzijds en een combinatie van elementen, die de hydrologische situatie van het bassin ten tijde van het uitbrengen van de verwachting karakteriseren, anderzijds.

Deze elementen zijn essentieel:

1. de in of rondom het bassin gemeten hoeveelheden neerslag,
2. de aanvoeren in het bassin, voorafgaande aan het tijdstip, waarop de verwachting gemaakt wordt.

Daarna behandelt Martel de keuze der regenstations; hoe lang de reeksen waren; de ruimtelijke verdeling; vestiging van nieuwe stations; de betrouwbaarheid der metingen.

Het net telt thans 281 stations (in 1949 : 184) in het gebied der Franse Alpen en de Jura. Al deze stations meten vloeibare en vaste neerslag (sneeuw); 160 ervan bovendien de temperatuur; enige bezitten ook een zelfregistrerende regenmeter, zonnenschijnautograaf, verdampingsmeter, vochtigheidsmeter, enz.

In deze berg-bassins, merendeels boven 1000 à 1500 m hoogte gelegen, komt de neerslag hoofdzakelijk als sneeuw neer, (bijna) uitsluitend als zodanig gedurende de 9 tot 11 maanden oktober (november) t/m juli (augustus).

Te beginnen in maart of april begint de vaste laag, die in de loop van de herfst en winter gevormd werd, steeds sneller te smelten en voedt daarbij ten eerste de vloeibare oppervlakte- of

diepgelegen reserves en vervolgens de beken en rivieren zelf tot einde juli of augustus (al naar de streek).

Martel's studie richtte zich tot nu toe op twee zaken:

- a. hoe kunnen wij het best de grootte van die vaste voorraad op het tijdstip, waarop wij de verwachting willen formuleren, representeren? en
- b. welk verband is er tussen de waarde van de zoeven bedoelde grootte en de globale wateraanvoeren in de smeltings-periode?

Natuurlijk zijn er meer factoren in het spel, die eigenlijk óók beschouwd zouden moeten worden.

- a. wat de maanden betreft, voorafgaande aan het ogenblik, waarop de verwachting gemaakt zal worden:
  1. welke verdeling over de voorbije herfst en winter had de neerslag (sneeuw)?
  2. welke was de bodemtoestand op het ogenblik van vorming van het sneeuwdek?
  3. welk verloop had de temperatuur op de verschillende hoogten gedurende de winter?
- b. wat de maanden betreft, waarvoor de verwachting gelden moet (de maanden, waarin de sneeuw smelt dus):
  1. welke zullen de hoeveelheden regen zijn in deze maanden en hoe zullen zij over de ruimte en tijd verdeeld zijn?
  2. hoe zal het temperatuursverloop worden?
  3. hoeveel zonneschijn zal er zijn?
  4. welke verdamping zal er zijn?

Gegevens omtrent  $a_1, a_2, a_3$  zou men tot zijn beschikking kunnen hebben; over  $b_1, b_2, b_3, b_4$  is geen zekerheid (men zou ook hiervoor verwachtingen moeten maken).

Goede sneeuwmetingen (z.g. sonderingen) vinden op tamelijk uitgebreide schaal én op grote hoogte eerst plaats sedert 1948 en kunnen daardoor niet goed in de statistische berekeningen betrokken worden. Martel overweegt dus óf te beschouwen het rekenkundige of gewogen gemiddelde  $P_1$  der hoeveelheden neerslag (hoofdzakelijk sneeuw) sedert 1 oktober of 1 november op de verschillende (dus vrijwel slechts in de lage regionen gelegen) stations van het bassin of in de naaste omgeving óf het verschil tussen deze (doch dan verbeterde) hoeveelheid neerslag en de hoeveelheid water, die in de loop van dezelfde periode, uit het bassin wegstroomde ( $P_w - U$ ).

Aangezien de stations natuurlijk merendeels in de bewoonde oorden gelegen zijn, d.w.z. op lage hoogten, representeren hun neerslaghoeveelheden niet de ware hoeveelheid over het gehele bassin, maar zijn meestal te klein. Men moet daarvoor corrigeren. Hoe? Daartoe wordt over vele hydrologische jaren, het gemiddelde der op deze (grotendeels op lage hoogten gelegen) stations gemeten hoeveelheden neerslag  $\bar{P}_L$  vergeleken met het gemiddelde  $\bar{U}$  der uitstromingen van het bassin (meestal blijkt  $\bar{U} > \bar{P}_L$ ), het laatste nog zo mogelijk vermeerderd met de gemiddelde verdamping  $\bar{V}$  (dus ook meestal  $\bar{P}_L < \bar{U} + \bar{V}$ ). Men kan nu rekenen of met bedoelde  $\bar{P}_L$  of met het verschil  $\bar{P}_w - U$ . Maar hoe weet (meet) men de ware  $\bar{P}_w$ ? Hiervoor beschouwt Martel de gemiddelde toestand in een hydrologisch jaar; daarvoor mag men aannemen, dat  $\bar{P}_w$  de som is van  $U$  en  $V$ ; ook kent men de  $\bar{P}_L$  op basis der referentie-stations. Hij berekent dan per bassin de factor  $K = (U + V) : \bar{P}_L$ ; altijd  $K > 1$ .

Bijv. a. bassin Bissorte; 10 referentiestationen, op gemiddelde hoogte 1300 m; gemiddelde hoogte van bassin is 2600 m;  $K = 1.5$ .

b. bassin des Sept-Laux; 12 referentiestationen, op gem. hoogte 900 m; gemiddelde hoogte van bassin is 2400 m;  $K = 1.9$ .

---

Zie Commentaar 1 en 6

---

Daarna vertelt Martel, dat hij ten eerste slechts stations met basisreeksen van 20 à 25 jaren koos, doch uit deze weer slechts die, waarvoor de correlatiecoëfficiënt tussen hun neerslaghoeveelheid en de uitstroming van het bestudeerde bassin groot is; hij noemt zulk een station voldoende representatief voor de hydrologische gesteldheid van het bassin.

---

Zie Commentaar 2

---

Vervolgens de verdere statistische procedure:

Als voorbeeld behandelt Martel het Bissorte-bassin; voor ieder 5-maanden-tijdvak 1 okt. t/m 28 febr. der jaren 1936/37 t/m 1953 t/m 54 (18 jaren) werd  $S = 1.5 \bar{P}_{10.2} - \bar{U}_{10.2}$  berekend voor het bassin (de 10 slaat op okt.; de 2 op febr.). Deze  $S$  (die representatief is voor de natuurlijke reserve op de eerste maart van ieder jaar) is de coördinaat  $x$ . Voor elk der 18 6-maanden-tijdvakken mrt. t/m aug. wordt de uitstroming berekend (eenheid is 1 mm water over gehele hellende oppervlak van bassin): de  $y$ -coördinaat;  $\bar{x} = 455$  en  $\bar{y} = 899$  mm.;  $s_x = 222$  en  $s_y = 221$  mm.

Dit geeft 18 punten in een cartesisch stelsel op coördinaten  $x - \bar{x}$  en  $y - \bar{y}$  (fig. 1). Elk punt stelt een jaar voor. De punten schijnen op een lineair verband tussen  $y$  en  $x$  te wijzen. Er valt niet veel te toetsen aan een materiaal van maar 18 getallenparen. Daarom waagt Martel de hypothesen, dat:

1. bij iedere waarde van  $x$  de  $y$  een normale verdeling volgt met een standaarddeviatie, die van  $x$  niet afhangt, en
2. dat het gemiddelde  $\bar{y}(x)$  lineair met  $x$  (de regressie-rechte) samenhangt. Deze rechte wordt op de klassieke wijze berekend:

$$b_{y,x} = 0.82; r = 0.83; \sqrt{1-r^2} = 0.56. \therefore y = 0.82x + 515$$

Vervolgens worden twee rechten  $\sqrt{\quad}$  getrokken evenwijdig aan de regressierechte en wel op gelijke ( $\parallel$   $y$ -as gemeten) afstanden, t.w.  $1.28 \cdot s_y^1 = 1.28 \cdot s_y \sqrt{1-r^2} = 159$  mm. Deze 1.28 hangt samen met de wens de rechten zó te leggen, dat binnen de door deze 2 rechten bepaalde ruimte 0,8 allen punten zullen liggen. In de gestandaardiseerde normale verdeling heeft de waarde  $t = 1.28$  de overschrijdingskans  $\frac{1}{2} (1-0,8) = 0.1$  hetgeen betekent, dat de kans op  $-1.28 < t < +1.28$  gelijk 0.8 is. *rechten 3' en 3''*

---

Zie Commentaar 3

---

Martel voegt toe: op 1 maart 1956 bedroeg  $x - \bar{x} \cong -145$  mm. Via de rechte (die gebaseerd is op de punten 1937 t/m 1954) is dan (ik verbeter hier tevens een rekenfout van M.)  $y - \bar{y} = -125$  of  $y = 764$ , zodat de 0.1- en 0.9-punten zijn  $764 \pm 159$ , d.i. 605 en 919 mm. Met een kans 0.8 zal dus de afvoer over 1 mrt. t/m 31 aug. 1956 bedragen 605 à 919 mm. Men mat: 714. Fraaie uitkomst!

Martel weet, dat men niet al te blij moet zijn met dit mooie resultaat; er zijn zo veel factoren, die niet beschouwd werden.

Vervolgens berekent hij, terwille van het gebruik in de praktijk, die waarden voor  $y$  (over 1 mrt. t/m 31 aug. 1956) die met een kans 0.9 0.8 0.5 0.2 en 0.1 zullen worden overtroffen en wel met behulp van de op 28 febr. 1956 aanwezige gegevens.

Maar ook worden verwachtingen gemaakt voor het tijdvak 1 mrt. t/m 30 jun., 1 mrt. t/m 31 juli en 1 mrt. t/m 30 sept.

In dezelfde grafiek tekent hij voor elk dezer 4 tijdvakken (van 4, 5, 6, 7 mnd.) het 18-jarig-gemiddelde der uitstroming, zie fig. 2, *de stippen* •

---

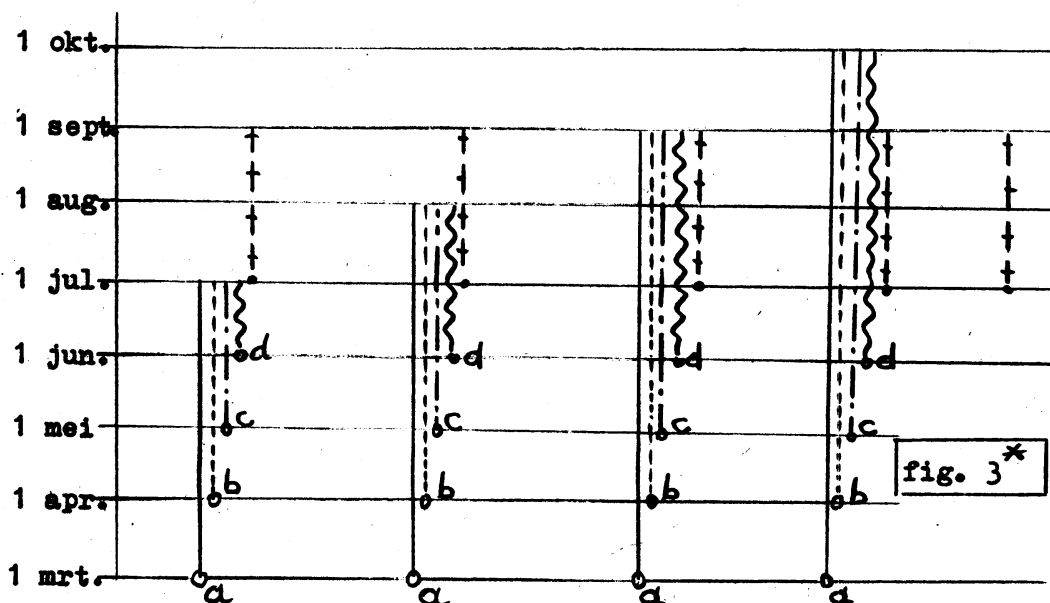
Zie Commentaar 4

---

Men maakt nog meer verwachtingen. Om precies te zijn:

- a. op 1 mrt. worden verwachtingen gemaakt voor de tijdvakken mrt. t/m juni (4 mnd.), mrt. t/m jul. (5), mrt. t/m aug. (6) en mrt. t/m sept. (7).
  - b. op 1 apr. voor de tijdvakken apr. t/m jun. (3 mnd.), apr. t/m jul. (4), apr. t/m aug. (5) en apr. t/m sept. (6).
- In de grafiek fig. 3 wordt dan die verwachte hoeveelheid over apr. t/m aug. vermeerderd met de gemeten hoeveelheid over mrt. en dan boven 31 aug. d.i. 1 sept. uitgezet (segment 2). enz.
- c.
  - d.
  - e.
  - f. op 1 aug. voor de tijdvakken aug. (1 mnd.) en aug. + sept. (2 mnd.)

In fig. 3 heb ik alleen a, b, c, d getekend.



Deze uitvoerige verwachtingen gelden voor 11 bassins in de Alpen. Het is aardig om te zien hoe het "met een kans 0.8 te verwachten" gebied kleiner wordt, naarmate er meer gegevens beschikbaar zijn en daardoor het tijdvak, waarvoor de verwachting opgesteld wordt, korter wordt (zie mijn commentaar 5). Dit wordt goed geïllustreerd door fig. 3, weer geldende voor het Bissorte-Bassin, dat alle in 1956 gemaakte verwachtingen in beeld brengt.

Toelichtingen: boven 1 jul. ziet men 4 vlak bij elkaar (eigenlijk langs elkaar vallende) getekende rechte segmenten, genummerd 1, 2, 3, 4. Zij stellen voor de 0.8-kans verwachtingsintervallen voor tijdvakken, die eindigen met 1 juli en beginnen resp. 1 mrt. (1), 1 apr. (2), 1 mei (3) en 1 jun. (4). Men lette op de ordinaten van de eindpunten van bijv. segment (1); ze liggen bij ongeveer 640 en

470 mm. D.i. op 1 mrt. werd de verwachting gemaakt, dat de uitstroming over mrt. t/m juni met een kans 0.8 zou liggen tussen 470 en 640 mm; op 1 april werd een verwachting voor de uitstroming over apr. t/m jul. gemaakt, die, vermeerderd met de gemeten hoeveelheid over maart, een som tussen 480 en 600 mm levert. (segment 2 boven 1 juli).

Men ziet dat boven zo goed als elk der data 1 jul., 1 aug., 1 sept., 1 okt., 1 nov. de segmenten 1, 2, 3, 4, 5 in deze volgorde in lengte afnemen (d.i. de verwachtingen betrouwbaarder worden). Ook zijn met dikke stippen aangebracht de werkelijk gemeten afvoeren over mrt. t/m jun., mrt. t/m jul. en mrt. t/m aug. Deze 3 verwachtingen mogen geslaagd genoemd worden.

Verder zijn nog drie gestippelde krommen getekend, de kromme van 0.1 kans-, die van de gemiddelde en die van de 0.9 kanswaarden, waarmee het volgende bedoeld wordt. Beschouw bijv. alle 18 uitstromingen (18 jaren immers) over de tijdvakken mrt. t/m aug. Deze hebben een gemiddelde (zie in fig. 3: ong. 960 mm), terwijl men ook schatten kan, op basis van deze 18 getallen, welke waarde in het universum dezer getallen met een kans 0.1 resp. 0.9 zouden worden overschreden (zie fig.: 1280 resp. 690 mm). De kennis van deze getallen (690 en 1280) zou ook reeds een verwachting mogelijk maken (een zuiver klimatologische verwachting) doch natuurlijk behoort gebruikmaking van meteorologische en hydrologische gegevens een meer betrouwbare voorspelling mogelijk te maken en inderdaad zien we het gebied 690 tot 1280 mm zich vernauwen tot 680 tot 1010 mm (als de verwachting op 1 mrt. geformuleerd wordt) of 660 tot 950 mm (als de verwachting op 1 apr. gemaakt wordt) etc.

---

Zie Commentaar 7.3

---

Tenslotte eindigt Martel met nog eens het essentiële doel te noemen. Wij citeren gedeeltelijk: "Le but essentiel de nos prévisions est de permettre de fixer dans quelles limites il est raisonnable d'utiliser les apports aux réservoirs pendant la période de fusion si l'on veut garder une probabilité de remplissage élevé et ainsi d'établir un programme de marche rationnel des usines qui leur sont associées".

Men gaat voort de methodiek te verbeteren en op meer bassins toe te passen.

B. Eigen commentaar.

1. Wat is een hydrologisch jaar?

Een hydrologisch jaar ligt zodanig, dat de watervoorraad in de grond aan het begin en aan het einde van dit jaar dezelfde is. In het algemeen kan men de totale hoeveelheid neerslag  $P$ , die over een willekeurig gelegen jaar op het bassin viel, schrijven als de som van de totale uitstroming  $U$  (er is misschien uit- en instroming, waarbij de eerste overweegt, bedoeld wordt dan de algebraïsche som), de verdamping  $V$  en de verandering van de voorraad water in het bassin  $W$ , dus  $P = U + V + W$ . Voor een hydrologisch jaar is  $W = 0$ . Blijkbaar kan men  $U$  "goed" meten (rivieren, beken, etc.), doch hoe is het met de meting van  $V$  gesteld? Is deze betrouwbaar? Hoe wordt ze verricht? Zijn het schattingen? Het is jammer, dat Martel geen getallen ter illustratie toevoegt.

Algemener (bijv. voor een Hollandse polder):

$$P = V - W_i + W_u + \Delta W_g + \Delta W_s = V + (W_u - W_i) + (\Delta W_s + \Delta W_g)$$

$P$  = neerslag;  $W_i$  = ingelaten en  $W_u$  = uitgelaten water;  $\Delta W_g$  = verandering van grondwaterhoeveelheid.  $\Delta W_s$  = verandering in de hoeveelheid water in sloten, plassen etc. Blijkbaar geldt volgens Martel voor een hydrologisch jaar:  $\Delta W_s + \Delta W_g = 0$ , terwijl hij  $U = W_u - W_i$  stelt, zodat  $P = V + U$ .

2. Representativiteit; het selecteren van correlatiecoëfficiënten.

Welke definitie houdt Martel voor "representativiteit" aan? Wij zouden geneigd zijn te zeggen, dat de stations, waar neerslag gemeten wordt (tezamen  $P_L$ ), des te meer representatief zijn (d.w.z. ten aanzien van de neerslaghoeveelheid) voor het gehele bassin ( $P_B$ ) hoe "meer functioneel" het verband tussen  $P_L$  en  $P_B$  is. Als men jaar voor jaar én  $P_L$  én  $P_B$  zou weten, zou men ook jaar voor jaar het quotiënt  $k = P_B : P_L > 1$  kennen. Hoe nauwer deze  $k$ -waarden rondom hun gemiddelde (waarvoor Martel voor het Bissorte-bassin  $3/2$  aanvaardt) spreiden, hoe meer zijn dan o.i. die stations (die  $P_L$  leveren) representatief voor het gehele bassin. Doch aldus handelt Martel niet, want hij berekent de correlatiecoëfficiënt tussen enerzijds de regenval  $P_L$  op elk station over de 6 maanden-periode september t/m februari en anderzijds de totale afvoer  $U$  uit het ganse bassin over het direct volgende 6 maanden-tijdvak van maart t/m augustus. Dit geeft zovele correlatiecoëfficiënten  $r$  (elk op 18 paren gebaseerd)



als er regenstations liggen in en rondom het bassin; gerangschikt in volgorde van afnemende  $r_1 \geq r_2 \geq r_3 \dots \geq r_{10} \geq r_{11} \geq r_{12} \dots \geq r_n$ . Blijkbaar neemt Martel hiervan slechts de (eerste) 10, d.w.z.  $r_1, r_2 \dots r_{10}$ , waardoor 10 bepaalde stations aangewezen worden. Met deze 10 werkt hij verder. Hebben wij hier met het beruchte statistische selectieëffect te doen?

Wij willen deze procedure generaliseren. Gesteld  $a$  zowel als  $b$  is gecorreleerd met een derde grootte  $m$ ;  $\rho(a, m) = \rho_1$ ;  $\rho(b, m) = \rho_2$ ; verder is  $a$  met  $b$  gecorreleerd volgens  $\rho(a, b) = \rho_{12}$ .

Welke is de correlatiecoëfficiënt  $\rho$  tussen  $\frac{1}{2}(a + b)$  (of  $a + b$ ) en  $m$ ?

$$\text{Antw.: } \rho(a + b, m) = \frac{\sum (a + b)m - \sum (a + b)\sum m}{\sigma_m \left\{ \sum (a + b)^2 - \frac{(\sum (a + b))^2}{n} \right\}^{1/2}} =$$

$$\frac{(\sum am - \sum a\sum m) + (\sum bm - \sum b\sum m)}{\sigma_m \left\{ (\sum a^2 - \frac{(\sum a)^2}{n}) + (\sum b^2 - \frac{(\sum b)^2}{n}) + 2(\sum ab - \frac{\sum a\sum b}{n}) \right\}^{1/2}}$$

$$\frac{\rho_1 \sigma_1 + \rho_2 \sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho_{12}\sigma_1\sigma_2}} \quad \text{als } \sigma_1^2 = \text{variantie van } a; \sigma_2^2 \text{ van } b.$$

$$\text{Onderstel } \sigma_1 = \sigma_2, \text{ dan is } \rho = \frac{\rho_1 + \rho_2}{\sqrt{2(1 + \rho_{12})}} = \frac{\frac{1}{2}(\rho_1 + \rho_2)\sqrt{2}}{\sqrt{1 + \rho_{12}}}$$

Hoe ligt  $\rho$  t.o.v.  $\rho_1$  en  $\rho_2$ ? Nu denken wij aan het concrete geval;  $a$  = regenval op station A;  $b$  die op station B;  $m$  = afvoer. Allicht ligt  $\rho_{12}$  dicht bij één, terwijl in het algemeen  $\rho_1 \neq \rho_2$  (stel  $\rho_1 < \rho_2$ ). Aangezien  $0 < \rho_{12} < 1$ , is  $\rho_1 < \frac{1}{2}(\rho_1 + \rho_2)$ . Zeker ligt derhalve  $\rho$  boven  $\rho_1$ , doch ook boven  $\rho_2$ ? Dit hangt af van de waarden van  $s = \frac{\frac{1}{2}(\rho_1 + \rho_2)\sqrt{2}}{\sqrt{1 + \rho_{12}}}$  en  $\rho_{12}$ . Immers de  $\rho$  ligt óók boven  $\rho_2$  als  $\frac{\frac{1}{2}(\rho_1 + \rho_2)\sqrt{2}}{\sqrt{1 + \rho_{12}}} > \rho_2$  d.i. als  $s > \frac{1.414\sqrt{1 + \rho_{12}}}{2} - 1$ .

Enige getallen ter oriëntatie: als  $\rho_{12} = 0.8$ , dan moet  $0.9 < s < 1$  gelden, opdat  $\rho > \rho_1$  en  $\rho > \rho_2$  zal zijn; als  $\rho_{12} = 0.5$ , dan moet gelden  $0.73 < s < 1$  opdat  $\rho > \rho_1$  en  $\rho > \rho_2$ . Conclusie: Door van het verband tussen  $a$  en  $m$  over te stappen op dat tussen  $a + b$  en  $m$  gaan we altijd vooruit, d.i. komen wij tot een  $\rho > \rho_1$  en het kán zijn, dat óók  $\rho > \rho_2$  is (dan gaan we optimaal vooruit). Analoge waarheden zullen gelden als men overgaat van de correlatie tussen  $a$  en  $m$  op die tussen  $a + b + c + \dots$  en  $m$ . Als de stations A, B, C .... strak gecorreleerd zijn (d.w.z. de correlatiecoëfficiënten

$\rho_{ij}$  liggen vlak bij één) dan zullen de correlatiecoëfficiënten  $\rho_i$  (tussen A en m, tussen B en m, etc.) zeer weinig van elkaar moeten verschillen, opdat de correlatiecoëfficiënt tussen a + b + c + ... en m groter is dan de grootste der  $\rho_i$ 's. In dit geval heeft men een optimaal gunstig gevolg van een combinatie van stations.

Overzichtstabelletje (men moet altijd zorgen, dat  $\rho_1 \rho_2 - \sqrt{(1 - \rho_1^2)(1 - \rho_2^2)} \leq \rho_{12} \leq \rho_1 \rho_2 + \sqrt{(1 - \rho_1^2)(1 - \rho_2^2)}$  !)

$\rho_1$	$\rho_2$	$s = \rho_1 \rho_2$	$\rho_{12}$	$\rho$	$\rho > \rho_2$ als
0.4	0.6	0.67	0.96	0.51	$0.98 < s < 1$
0.4	0.6	0.67	0.8	0.53	$0.90 < s < 1$
0.4	0.6	0.67	0.5	0.58	$0.73 < s < 1$
0.4	0.6	0.67	0.3	0.62	$0.61 < s < 1$
0.4	0.6	0.67	0.0	0.71	$0.41 < s < 1$
met $\rho_1 = 0.4$ en $\rho_2 = 0.6$ is $-0.49 \leq \rho_{12} \leq 0.97$					

Dus  
altijd  $\rho > \rho_1 = 0.4$   
en ook  
 $\rho > \rho_2 = 0.6$   
indien  
 $\rho_{12} \geq 0.39$

3. Hoe betrouwbaar is deze  $r = 0.83$ ?

De significantie-tabel leert, dat de in een aselechte steekproef van 18 paren, genomen uit een ongecorreleerd normaal verdeeld universum, berekende correlatiecoëfficiënt met een kans 0.05 boven 0.50 gelegen is, met kans 0.01 boven 0.62 en met een kans, veel kleiner dan 0.01, boven 0.83. Als er reden is om te onderstellen, dat de x zowel als de y in het onderzoek van Martel normaal verdeeld is (alleen dan mag men de gebruikelijke significantie-tabel voor correlatiecoëfficiënten raadplegen), dan is de realiteit van het verband "zeker" (in statistische zin). Martel vertelt dit er niet bij. Het lijkt me zeer waarschijnlijk, dat  $\hat{u}$  en  $\hat{P}$  normaal verdeeld zijn, mits ze maar beide betrekking hebben op perioden van voldoende lengte (bijv. niet een dag), zoals een maand en meer. En inderdaad is dit het geval. En als  $\hat{u}$  en  $\hat{P}$  normaal verdeeld is, dan is  $3/2 P_L - U = S$  het ook.

Wij denken hierbij ook aan de waarschuwende woorden, die A.M. Grant uitsprekt in zijn artikel: "The application of correlation and regression to forecasting" in Bur. Met. Study No. 7, Melbourne; jan. 1956.

4. Het volume water, dat moet weggepompt worden.

Uit een, hier niet opgenomen, figuur lezen wij af:  $46 \text{ Mm}^3$  bij een overschrijdingskans 0.1, Hoezo 46? Uit fig. 2 lezen wij:

919 mm heeft een overschrijdingskans 0.1. Het bassinoppervlak bedraagt 48.4 km<sup>2</sup>, zodat 919 mm overeenkomt met 919 x 48.4 10<sup>3</sup>m<sup>3</sup> = 44 Mm<sup>3</sup> water, waarbij 1 Mm<sup>3</sup> = 10<sup>6</sup>m<sup>3</sup>. Klopt ongeveer.

5. Grotere betrouwbaarheid als er meer gegevens zijn.

Ik zou dit nog anders hebben willen illustreren: waarom niet de 5 verwachtingen voor steeds hetzelfde 3 maanden-tijdvak bijv. 1 juli t/m 31 sept., doch opgesteld begin maart, begin april, begin mei, begin juni, begin juli, tegenover elkaar gesteld? Juist dan zou de invloed van meer kennis, noodzakelijk bij de verwachting, moeten blijken in een markant smaller worden der betrouwbaarheidsbanden. Het komt me voor, dat Martel dit niet deed.

6. Moeten wij U<sub>2</sub> correleren met P<sub>L</sub> of met kP<sub>L</sub> - U<sub>1</sub>?

Is Martel erg vooruitgegaan door i.p.v. P<sub>L</sub> met U<sub>2</sub> de (3/2)P<sub>L</sub> - U<sub>1</sub> met U<sub>2</sub> te correleren? Daarbij stelt P<sub>L</sub> voor de hoeveelheid neerslag over alle stations over het tijdvak sept. t/m febr., U<sub>1</sub> de totale hoeveelheid uitstroming gedurende ditzelfde tijdvak en U<sub>2</sub> de totale uitstroming over het aansluitende tijdvak mrt. t/m aug.

Wij werken uit a.v.:

$$\rho(U_2; \frac{3}{2}P_L - U_1) = \frac{E U_2 (\frac{3}{2}P_L - U_1) - E U_2 E (\frac{3}{2}P_L - U_1)}{\sigma_2 \sqrt{E (\frac{3}{2}P_L - U_1)^2 - E^2 (\frac{3}{2}P_L - U_1)^2}} = \frac{\frac{3}{2}\rho_2\sigma_P + \rho_{12}\sigma_1}{\sqrt{\frac{3}{4}\sigma_P^2 + 2\frac{3}{2}\rho_1\sigma_1\sigma_2 + \sigma_1^2}}$$

Hierin is  $\sigma_1^2$  = variantie in U<sub>1</sub>-waarden;  $\sigma_2^2$  idem in U<sub>2</sub>-waarden;  $\sigma_P^2$  idem in P-waarden;  $\rho_1$  = correlatiecoëff. tussen U<sub>1</sub> en P<sub>L</sub>;  $\rho_2$  idem tussen U<sub>2</sub> en P<sub>L</sub>;  $\rho_{12}$  idem tussen U<sub>1</sub> en U<sub>2</sub>.  
Laten we de factor 3/2 door K vervangen,  $\sigma_1$ ;  $\sigma_P$  door  $q$  en  $q/K$  door m. Dan is

$$\rho = \frac{\rho_2 + m\rho_{12}}{(1 + 2\rho_1 m + m^2)^{1/2}} \quad \text{Zal } \rho > \rho_2 \text{ zijn?}$$

Daarvoor onderzoeken wij het tegendeel:

$$\frac{\rho_2 + m\rho_{12}}{(1 + 2\rho_1 m + m^2)^{1/2}} < \rho_2 \quad (\text{te bedenken, dat alle correlatiecoëfficiënten positief zijn}).$$

$$\text{Dit geldt als } \rho_{12} < \rho_2 \left( \frac{-1 + \sqrt{1 + 2\rho_1 m + m^2}}{m} \right) = f \rho_2:$$

Aangezien  $0 < \rho_1 < 1$  en  $m > 0$ , is  $f < 1$ . Derhalve is  $\rho < \rho_2$  als  $\rho_{12} < f\rho_2$ .

Helaas beschikken wij niet over getallenwaarden. Het zou prettig geweest zijn als Martel wat meer van  $\sigma_1$ ,  $\sigma_2$ ,  $\rho_1$ ,  $\rho_2$  etc. verteld zou hebben. Hoe dicht  $\rho_1$  bij 1 gelegen is (hoe meer daardoor  $f$  bij 1) en hoe kleiner  $\rho_{12}$  (wij vermoeden, dat  $\rho_{12}$  wel dicht bij nul gelegen zal zijn) hoe waarschijnlijk is het, dat  $\rho_{12} < f \rho_2$  geldt. Als dit inderdaad het geval mocht zijn, had Martel beter met de correlatie tussen  $P_L$  en  $U_2$  kunnen werken i.p.v. met die tussen  $(3/2)P - U_1$  en  $U_2$ .

Nu kunnen wij ook de invloed van  $K$  nagaan. Bij gelijkblijvende  $\rho_2$ ,  $\rho_{12}$  en  $g$ , neemt  $m$  met toenemende  $K$  af (naar 0) en komt  $\rho$  op de  $n$  duur bij  $\rho_2$ .

Samenvattende:

1. het komt me voor, dat Martel er niet op vooruitgegaan is door niet met de correlatie tussen  $U_2$  en  $P_L$  (c.c.  $\rho_2$ ), maar met die tussen  $U_2$  en  $(3/2)P_L - U$ , (c.c.  $\rho$ ) te werken (want het lijkt ons, dat  $\rho < \rho_2$ )
2. de waarde  $3/2$  voor  $K$  (in het geval van Bissorte-bassin) moet goed geargumenteerd zijn, want de waarde van  $\rho$  kan daarmee sterk samenhangen.

7. Wat beogen wij met een regressie-analyse?

Wanneer is het gebruik van de regressierechte wel en niet zinvol?

7.0 Het gaat hier om het gebruik van de regressierechte bij forecasting. Er zijn hierbij 3 belangrijke vragen :

Vraag 1: Gegeven de waarde van  $x_i$ , hoe groot is dan het betrouwbaarheidsinterval voor de gemiddeld bij deze  $x_i$  behorende  $y$ -waarde?

Vraag 2: Binnen welk betrouwbaarheidgebied ligt de onbekende regressie rechte ( $\eta = \alpha + \beta x$ ) met een gegeven betrouwbaarheid?

Andere formulering: binnen welk gebied zullen bijv. 95% der rechten liggen, die men verkrijgt als men herhaaldelijk een  $n$ -tal punten met dezelfde  $x$ -waarden beschouwt?

Vraag 3. Binnen welk gebied mag men, met een voorgeschreven zekerheid, de  $y$ -waarde verwachten die kan behoren bij een  $n + 1$  ste  $x$ -waarde ( $x_{n+1}$ )?

Opmerking: Vooral als  $n =$  aantal paren  $x, y$  klein is (hier slechts 18) heeft het zin op deze vragen een exact antwoord te geven.

7.1 Antw. 1. Wij moeten, evenals Martel doet, eerst enige onderstellingen invoeren, anders is geen enkele vraag te beantwoorden. Zij zijn:

1. Bij iedere x-waarde is de bijbehorende y-waarde normaal verdeeld rondom een gemiddelde  $\bar{y}$ . Deze normale verdelingen hebben alle dezelfde standaarddeviatie  $\sigma$ .
2. Deze  $\bar{y}$  hangt lineair met x samen, d.i.  $\eta \equiv \bar{y} = \alpha + \beta x$ . De coëfficiënten  $\alpha$  en  $\beta$ , alsmede  $\sigma$ , kennen wij niet. <sup>1)</sup>

Neem nu een bepaalde x in het oog, bijv. die bij het jaar 1941 (fig. 2 leert  $x_{15} = 455 + 230 = 685$  mm; deze x heeft rangnummer 15 in de 18 in volgorde van toeneming genoteerde x-waarden).

De beste schatting van het universum-gemiddelde  $\eta_{15}$  in de waarschijnlijkheidsverdeling van alle bij deze  $x_{15}$  behorende y-waarden is dan  $\hat{y}_{15} = \hat{\alpha} + \hat{\beta} x_{15}$ , indien we voor de volgens de methode der kleinste kwadraten berekende beste regressierechte schrijven:  $y = \hat{\alpha} + \hat{\beta} x$ . Hoe betrouwbaar is deze  $\hat{y}_{15}$ ? D.w.z. hoe goed benadert deze speciale  $\hat{y}_{15}$  de ware  $\eta_{15} = \alpha + \beta x_{15}$  of nog anders uitgedrukt: hoe nauw liggen, indien men vele keren n paren x, y meten zou (wel alle met dezelfde  $x_{15}$ ), de verschillende  $\hat{y}_{15}$ -waarden rondom de (onbekende)  $\eta_{15}$ -waarde? Welke waarschijnlijkheidsverdeling volgt deze  $\hat{y}_{15}$ ? Men kan bewijzen, dat deze verdeling heeft een gemiddelde  $\bar{\hat{y}}_{15} = \eta_{15}$  en een variantie

$$\sigma^2(\hat{y}_{15}) = \sigma^2 \left[ \frac{1}{18} + \frac{(x_{15} - \bar{x})^2}{\sum_{i=1}^{18} (x_i - \bar{x})^2} \right]$$

Hierin is  $\bar{x} = \sum_{i=1}^{18} x_i / 18$

en  $\sigma^2$  de reeds genoemde (ook onbekende) variantie. Het valt op, dat deze  $\sigma^2(\hat{y})$  samenhangt met het rangnummer van x en daardoor tevens slechts gedefinieerd is voor alle x-waarden, die deel uitmaken van de steekproef.

Voor de  $\sigma^2$  moeten wij een beste schatting maken; deze is  $\hat{\sigma}^2 = \sum_{i=1}^{18} (y_i - \hat{y}_i)^2 : (18-2)$  (let wel: 18-2 en niet 18-1).

De beste schatting voor  $\sigma(\hat{y}_{15})$  wordt derhalve

$$\hat{\sigma}(\hat{y}_{15}) = \left[ \frac{\sum (y_i - \hat{y}_i)^2}{18-2} \left\{ \frac{1}{18} + \frac{(x_{15} - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right\} \right]^{1/2}$$

$$= \hat{\sigma} \sqrt{f} \quad , \text{ met } f \geq \frac{1}{18}.$$

Het bovenstaande geldt voor elk der 18 x-waarden, d.w.z.

---

noot 1) Eigenlijk is er nog een drietal onderstellingen, t.w. zowel x als y is foutloos gemeten en de x is verdeeld volgens een bepaald type verdeling. Tot dit type behoort ook de normale verdeling.

niet voor  $x_{15}$  alleen. Men kan bewijzen, dat het quotiënt  $(\hat{y}_{15} - \eta_{15}) : \hat{\sigma}(\hat{y}_{15})$  een t-verdeling volgt met  $n-2 = 18-2 = 16$  graden van vrijheid (welke verdeling voor zo weinig graden van vrijheid vrij slecht door een normale benaderd wordt). Ik heb voor elk der 18  $x_i$ -waarden de waarde van  $\hat{\sigma}(\hat{y}_i)$  berekend. Deze is rondom het midden der  $x$ -waarden het kleinst, t.w. voor die  $x_i$ , welke het dichtst bij  $\bar{x}$  ligt, en voor  $x_1$  resp.  $x_{18}$  het grootst. Wij hebben iedere waarde  $y_i = \hat{\alpha} + \hat{\beta}x_i$  ( $i = 1, 2 \dots 18$ ) vermeerderd en verminderd met  $t_{16} \cdot \hat{\sigma}(\hat{y}_i)$ ; deze  $t_{16} = 2.12$  stelt die waarde van  $|t|$  voor (in de t-verdeling op  $18-2 = 16$  g.v.v.), die overschreden wordt met kans 0.05 (te vinden in een t-tabel; als  $n$ , hier 18, toeneemt, nadert  $t_n$  tot 1.98, gemakshalve meestal afgerond tot 2.00). Aldus ontstaat de licht gearceerde band, begrensd door hyperbooltakken, rondom de regressierechte. Deze band is, ondanks de kleine waarde van  $n$ , zo nauw door de kleinheid van  $\hat{\sigma}$  en deze  $\hat{\sigma}$  is weer zo klein, doordat de 18 punten weinig sterk spreiden rondom de rechte (hetgeen weer tevens in een grote waarde van de correlatiecoëfficiënt weerspiegeld wordt).

Een opmerking: Martel berekent  $\hat{\sigma}$  uit  $s_y \sqrt{1-r^2} = 221 \sqrt{1-0.83^2} = 124$ . Helemaal juist is dit niet. De theorie leidt af  $\hat{\sigma}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{(n-2)}$  en men kan dit omvormen tot  $(1-r^2) \cdot \sum (y_i - \bar{y})^2 / (n-2)$

Echter is  $s_y^2 = \sum (y_i - \bar{y})^2 / (n-1)$  en niet  $\sum / (n-2)$

Voor veel punten (grote  $n$ ) is het bovenstaande van zeer weinig betekenis. In dit geval komt Martel tot 124 mm, doch wij vinden 128 mm.

De betekenis van dit alles is deze: de op grond van de gemeten  $x_{19} = 310$  mm (de totale hoeveelheid  $1\frac{1}{2}$  P - U over de maanden okt. 1956 t/m febr. 1957) gemiddeld te verwachten uitstroming, totaal over de maanden maart t/m augustus, ligt met 0.95 zekerheid tussen rond 750 en 780 mm. Bij deze  $x = 310$  mm behoort namelijk een  $\hat{y} = 764$  mm, af te lezen van de regressierechte, en een 2.15. 128.

$$\sqrt{\frac{1}{18} + \frac{(310 - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = 15 \text{ mm, zodat de gezochte grenzen zijn } 764 - 15 \text{ en } 764 + 15 \text{ mm.}$$

7.2 Antw. 2. Wat deze vraag betreft willen we kort zijn. Weer zetten wij vanuit ieder der 18 punten  $x_i$ ,  $y_i$  ( $= \hat{\alpha} + \hat{\beta} x_i$ ) op de regressierechte naar boven en naar beneden eenzelfde segment uit, dat met  $x_i$  samenhangt, n.l.

$$\hat{\varepsilon}_i = \hat{\sigma} \sqrt{2 F_{2;18-2}} \sqrt{\frac{1}{18} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$$

Men ziet, dat  $\hat{\varepsilon}_i$  voor een factor  $\sqrt{2 F_{2;18-2}}$  :  $t_{18-2}$  zo groot is als  $\hat{\sigma}(\hat{y}_i)$ . Dit is een factor  $> 1$ ;  $F_{2;18-2}$  stelt voor die F-waarde, welke in de F-P-tabel bij 2 en 18-2 g.v.v. met een kans 0.05 overschreden wordt.

Enige getallenvoorbeelden: bij  $n = 10$  is  $\sqrt{2 F_{2;8}} = 3.00$  en  $t_8 = 2.31$ ; bij  $n = 20$  is  $\sqrt{2 F_{2;18}} = 2.68$  en  $t_{18} = 2.10$ ; bij  $n = 120$  is  $\sqrt{2 F_{2;118}} = 2.49$  en  $t_{118} = 1.98$ ; als  $n \rightarrow \infty$  dan  $\sqrt{2 F_{2;n}} \rightarrow 2.46$  en  $t_n \rightarrow 1.98$ . Met toenemende  $n$  daalt  $\sqrt{2 F_{2;n-2}}$  } asymptotisch naar 1.24.  $t_{n-2}$

7.3 Antw. 3. Toelichting: de 18 paren  $x_i, y_i$  leverden een beste rechte  $y = \hat{\alpha} + \hat{\beta} x$ . Er wordt een 19de  $x$ -waarde  $x_{19}$  gemeten (die kleiner dan de kleinste of groter dan de grootste der 18  $x$ -waarden mag zijn). Welke kan de bijbehorende  $y$ -waarde zijn? De regressierechte levert als beste schatting van het gemiddelde van dit  $y$ -universum:  $\hat{y}_{19} = \hat{\alpha} + \hat{\beta} x_{19}$ .

Gemeten wordt zekere  $y_{19}$ . Aan welke kansverdeling behoort het verschil  $d = y_{19} - \hat{y}_{19}$ ? Men kan bewijzen, dat

$$E d = 0 \text{ en } \sigma_d^2 = \sigma^2 \left[ 1 + \frac{1}{18} + \frac{(x_{19} - \bar{x})^2}{\sum_{i=1}^{18} (x_i - \bar{x})^2} \right]$$

De beste schatting van  $\sigma^2$  is  $\hat{\sigma}^2 = \frac{\sum_{i=1}^{18} (y_i - \hat{y}_i)^2}{(18-2)}$

Dan is de beste schatting van  $\sigma_d$

$$\hat{\sigma}_d = \hat{\sigma} \sqrt{1 + \frac{1}{18} + \frac{(x_{19} - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = \hat{\sigma} \sqrt{g} \quad \text{met } g \geq 1 + \frac{1}{18}$$

Verder volgt het quotiënt  $d/\hat{\sigma}_d$  een t-verdeling met 18 - 2 g.v.v. Vergelijkende met  $\hat{\sigma}(\hat{y}_{19})$  zien we, dat de factor  $f$  vervangen is door  $g = f + 1$ . We zien  $\hat{\sigma}_d^2 = \{\hat{\sigma}(\hat{y})\}^2 + \hat{\sigma}^2$

De totale variantie van de forecast is blijkbaar opgebouwd uit twee varianties, de ene: de "error variance" van de schattingen rondom de ware regressie, de andere: de steekproef-onnauwkeurigheid van deze regressie zelf. Vooral

als  $n =$  aantal paren klein is en als bovendien de punten sterk rondom de rechte spreiden, zal men  $\{\hat{\sigma}(\hat{y})\}^2$  niet tegenover  $\hat{\sigma}^2$  mogen verwaarlozen. Meestal let men alleen op  $\hat{\sigma}^2$ , zoals Martel doet, maar in zijn geval is inderdaad  $\hat{\sigma}(\hat{y}) \ll \hat{\sigma}$ .

Met deze  $\hat{\sigma}_d$ , berekend voor elk der  $n = 18$  x-waarden, hebben wij de uitdrukking  $2.12 \hat{\sigma}_d$  berekend. Daarmede hebben wij de twee 0.95-hyperbooltakken rondom de regressierechte getekend. De hyperbooltakken zijn zeer flauw gebogen. 4 en 4"

Resultaat: gegeven de  $x_{19} = 310$  mm, mag men daarom de y-waarde met 0.95 zekerheid verwachten tussen  $\hat{y}_{19} - 2.12 \hat{\sigma}_d = 438$  en  $\hat{y}_{19} + 2.12 \hat{\sigma}_d = 1050$  mm;  $\hat{y}_{19} = 764$  mm is de waarde bij  $x_{19}$ , afgelezen van de regressierechte.

De  $\hat{\sigma}_d = 133$  mm wordt berekend via de genoemde formule voor  $\hat{\sigma}_d$ , door in te vullen  $\hat{\sigma} = 128$  mm,  $x_{19} = 310$  mm,  $\bar{x} = 455$  mm en  $\sum_7^8 (x_i - \bar{x})^2 = 836400$  mm<sup>2</sup>.

Martel voelde meer voor de 0.80-grenzen. De t-verdeling levert bij  $18-2 = 16$  g.v.v. en bij 0.10 een  $t_{16} = 1.33$ , terwijl Martel 1.28 gebruikt (kennelijk bij  $\infty$  g.v.v., d.w.z. bij zeer grote  $n$ ). Bovendien moet hij niet rechten // regressierechte trekken, doch hyperbooltakken, ofschoon deze in dit geval zeer vlak blijken te zijn, maar vooral wanneer men een x zou meten ver beneden de kleinste of ver boven de grootste x-waarde is het beter de hyperbooltakken te gebruiken.

Wij kunnen hieraan nog een vraag verbinden: verbeteren wij de forecast m.b.v. de regressierechte? Martel houdt zich maar even bezig met deze vraag.

Wat zullen wij onder "verbeteren" verstaan?

Zonder een regressierechte te kennen, kunnen wij altijd beweren, dat de y een waarschijnlijkheidsverdeling volgen zal (en wel een normale) rondom zeker onbekend gemiddelde  $\mathcal{E}(y)$  (hier geschat als  $\bar{y} = 899$  mm) en met zekere onbekende standaarddeviatie  $\chi$  (hier geschat als  $c = S_y = 221$  mm). Welke ook de waarde van  $x_{19}$  moge zijn (de 19de x-waarde, nadat 18 x-waarden, die voor de regressie gebruikt werden, "achter de rug" zijn), altijd is waar, dat met een zekerheid 0.95 de onbekende y-waarde zal liggen tussen  $\bar{y} - 2.12 c = 424$  en  $\bar{y} + 2.12 c = 1375$  mm. Hierbij is 2.12 die waarde van  $|t|$ , die in de t-verdeling op  $18 - 2 = 16$  g.v.v. met een kans



0.05 overtroffen wordt (d.w.z. een kans 0.95 op  $-2.12 < t < + 2.12$ . Dit is een zuiver klimatologische voorspelling.

Met behulp van de regressierechte kan worden gezegd, weer met 0.95 zekerheid, dat de komende y-waarde gelegen is tussen  $\hat{y}_{19} - 2.12 \hat{\sigma}_d = 438$  en  $\hat{y}_{19} + 2.12 \hat{\sigma}_d = 1050$  mm. Het gebied 438 tot 1050 mm is (nauwer) dan het gebied 424 tot 1375 mm, waarin reeds een voordeel van een forecast d.m.v. de lineaire regressie gezien kan worden. Belangrijk is voornamelijk dat men in een situatie verkeert waarin  $\hat{\sigma}_d < c$ . Dit is hier inderdaad het geval, met  $\hat{\sigma}_d = 133$  en  $c = 221$  mm. Wij kunnen de vraag stellen, wat is in het algemeen nodig, opdat  $\hat{\sigma}_d < c$ ? Daarvoor schrijven wij uit:

$$c = \frac{\sum (y_i - \bar{y})^2}{n-1}; \hat{\sigma}_d^2 = \frac{\sum (y_i - \bar{y})^2}{n-1} (1-r^2) \frac{n-1}{n-2} \left[ \frac{n+1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right]$$

c heeft natuurlijk niets te maken met  $x_{n+1}$ , echter  $\hat{\sigma}_d$  wel; het gaat om de factor  $m =$

$$\hat{\sigma}_d / c = (1-r^2) \frac{n-1}{n-2} \left[ \frac{n+1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right]$$

Wanneer is  $m < 1$ ? Stel  $q = (x_{n+1} - \bar{x})^2 : \sum (x_i - \bar{x})^2$ ;  $q \geq 0$ ;  $q = 0$  als  $x_{n+1} = \bar{x}$ ;  $q$  neemt toe naarmate  $x_{n+1}$  meer van  $\bar{x}$  verschilt (dus naar de "einden" van de regressierechte). Voor de  $x_{19} = 310$  mm is  $q = 0.025$ .

Voor  $m < 1$  is nodig  $r^2 > 1 - \frac{n(n-2)}{n^2(q+1) - nq - 1}$  of  $|r| > \sqrt{1 - \frac{n(n-2)}{n^2(q+1) - nq - 1}}$

Teneinde een indruk te krijgen maken wij het volgende tabelletje

		$\sqrt{1 - \frac{n(n-2)}{n^2(q+1) - nq - 1}} \xrightarrow{n \rightarrow \infty} \sqrt{\frac{q}{q+1}}$			
$q \backslash n$		10	20	50	$\infty$
0		0.436	0.314	0.200	0.000
0.1		0.509	0.442	0.352	0.301
0.5		0.666	0.624	0.597	0.576
1		0.760	0.734	0.719	0.707
5		0.925	0.919	0.916	0.912

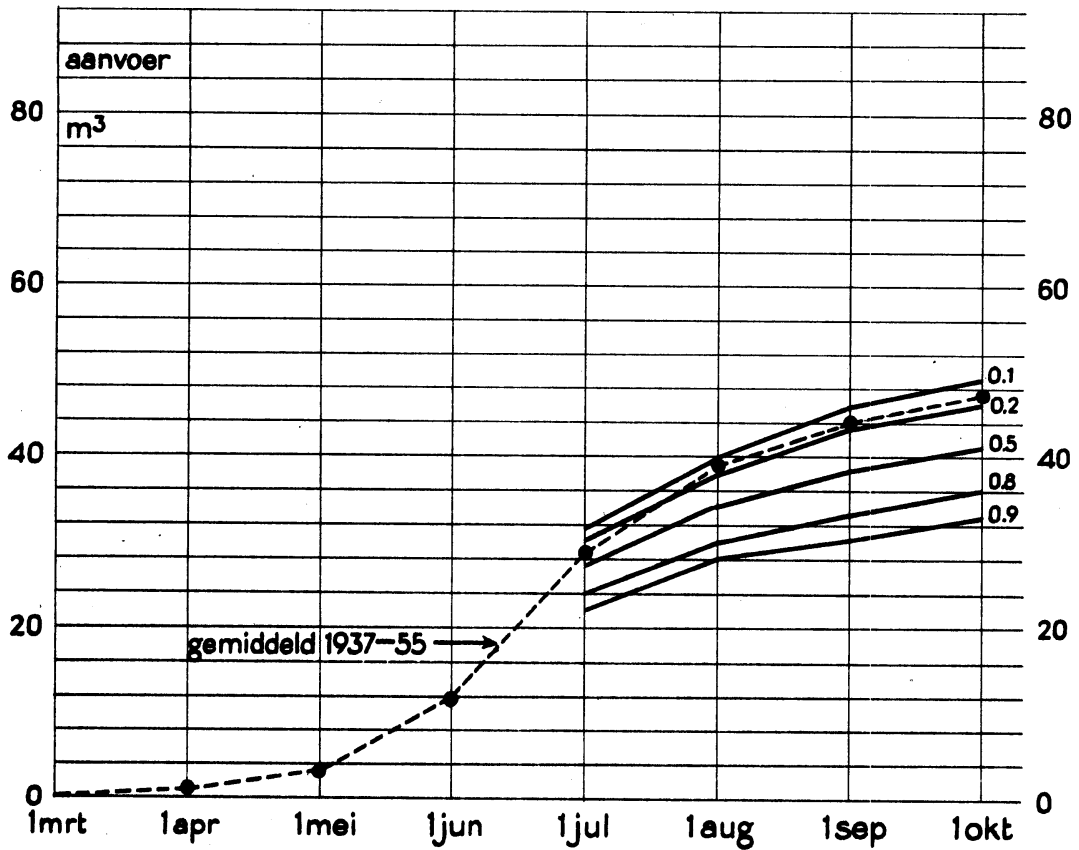
In het onderhavige geval is  $n = 18$  en  $q \approx 0.025$ . Voor dit geval is de voorspelling via de regressierechte beter dan de klimatologische ("beter" in de zin van  $\hat{\sigma}_d < c$ ) als  $|r| > 0.37$  (interpoleren uit tabelletje). Er werd gemeten 0.83. Zolang in Martel's geval de nieuwe  $x_{19}$  binnen het traject der gebruikte 18 x-waarden blijft of er niet ver buiten komt, blijft steeds de voorspelling via de

regressierechte beter dan de gewone klimatologische. Maar gesteld, dat  $x_{19}$  geweest zou zijn 2455 mm  $\gg \bar{x} = 455$  (misschien een dwaze onderstelling?), dan zou  $\hat{\sigma}_d$  geweest zijn 4.8. In dat geval zou wel  $\hat{\sigma}_d > c$  geweest zijn (want  $r = 0.83$  en uit het tabelletje volgt ongeveer 0.90). Kortom: voor "zeer ver" van  $\bar{x}$  verwijderde  $x$ -waarden zal het zeker gebeuren dat  $\hat{\sigma}_d > c$ ; dan heeft het weinig zin de regressierechte te gebruiken. Wat "zeer ver" is hangt van het geval af; hoe nauwer de correlatie (d.i. hoe groter de  $r$  bij de bij de lineaire regressie gebruikte paren) en hoe groter de  $n$ , hoe verder is dit "zeer ver", d.w.z. hoe groter is de mogelijkheid de regressierechte zinvol bij de voorspelling te gebruiken.

Men drukt dit wel aldus uit:

Men mag de verwachting pas betrouwbaar noemen als de gemiddelde fout bij deze verwachting d.m.v. de regressierechte beneden de klimatologische fout blijft.



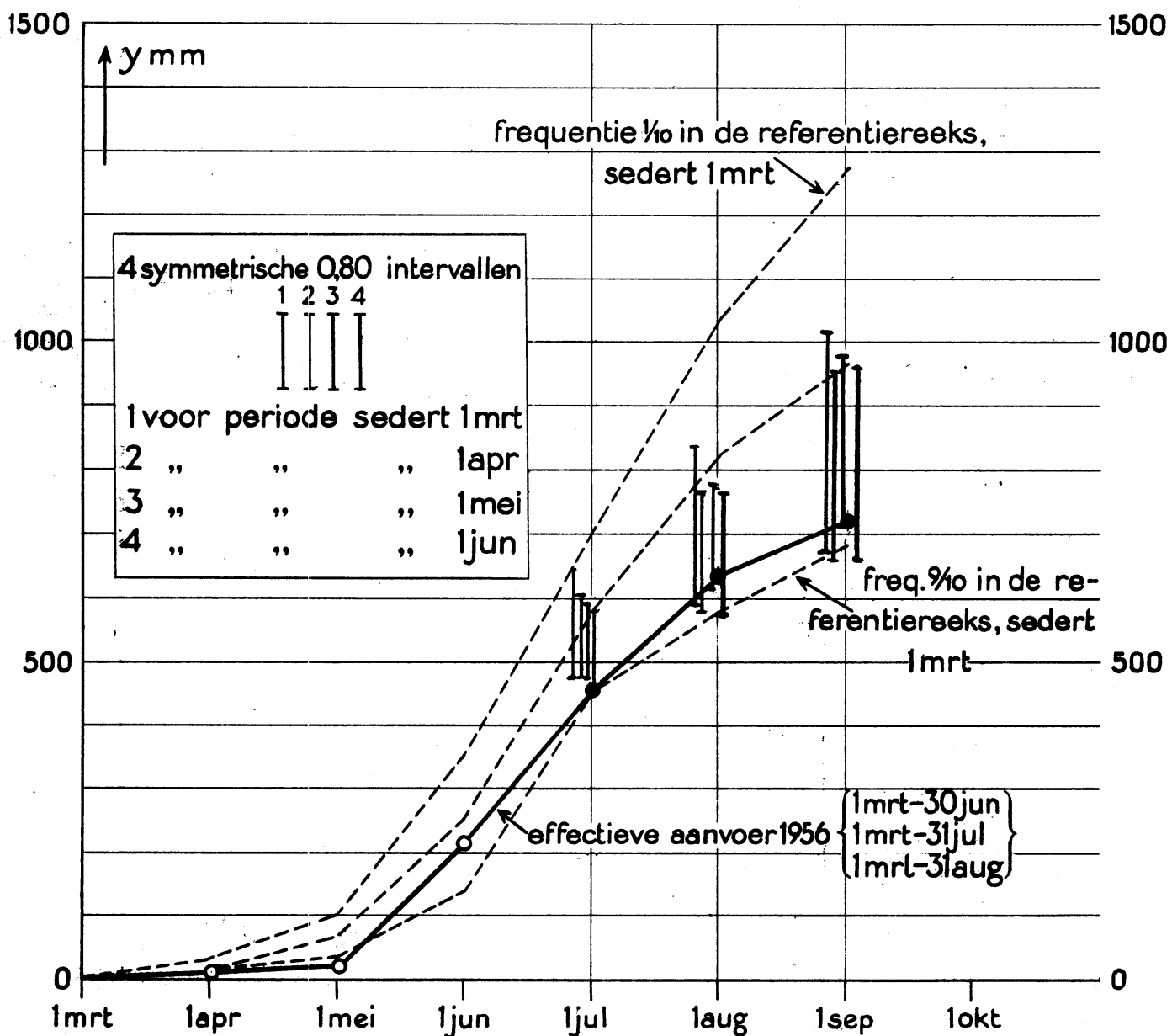


Bassin: BISSORTE

Verwachtingen, gemaakt op 1-3-1956.

De verwachtingen betreffen de aanvoeren, die aan het einde van elk der maanden juni, juli, aug. en sept. met kansen 0.1 0.2 0.5 0.8 0.9 bereikt of overschreden worden.

fig. 2



Bassin: BISSORTE

Vergelijking tussen verwachting en werkelijkheid

fig. 3

2. L. Serra.

La représentation des facteurs climatologiques par des courbes d'équale probabilité.

A. Inhoud van de voordracht.

a. Men representeert nog dikwijls een verzameling metingen door één enkel karakteristiek geval, waarvoor in de regel het rekenkundig gemiddelde gekozen wordt. Echter kan dit alleen de verzameling nimmer volledig karakteriseren. Er is behoefte aan een "indice de dispersion", "spreidingsindex".

De klimatologen werken dikwijls alleen met gemiddelden, waarmee zij onder meer de verschillende klimaten klassificeren.

In de "Résumés mensuels du temps" van de meeste nationale meteorologische diensten vindt men tabellen met afwijkingen ten opzichte van het gemiddelde en soms ook kaartjes met isokrommen van gelijke verhouding tot het gemiddelde (bij de neerslaghoeveelheden in het bijzonder). Deze kaartjes en tabellen hebben zeker hun verdiensten:

- a) zij elimineren de invloed van de opstelling der regenmeters en de invloed van de opstellingshoogte,
- b) zij zijn gemakkelijk te lezen.

Helaas leiden ze ook tot foutieve interpretaties.

b. Beschouw bijv. twee verschillende stations, waar men gedurende een lange reeks jaren tot dezelfde gemiddelde jaarsom of maandsom neerslag gekomen is, echter met verschillende "variabiliteiten" (Men zou de variabiliteit kunnen karakteriseren door de verhouding tussen de gemeten kleinste en grootste jaarsom, danwel maandsom, gedurende de gehele basisperiode).

Als nu bijv. over oktober 1955, dezelfde hoeveelheid neerslag op de twee stations gemeten is, zal op het kaartje met de krommen van gelijke verhouding tot het veeljarige gemiddelde, in niets het verschil tussen de twee stations blijken. En toch is er verschil, want de bedoelde hoeveelheid kan zeldzaam groot zijn voor het ene stations en zeer gewoon voor het andere.

Een meteorologische studie behoort zich niet te beperken tot de blote opsomming der waarnemingsresultaten. Het is niet voldoende een verschijnsel te beschrijven, men moet het ook proberen te begrijpen en verklaren.

c. Gedacht wordt ook aan de operaties om kunstmatige regen te verwekken. Men brengt wolken tot uitregening, zo mogelijk vele keren per maand, en tekent na afloop van de maand voor de ganse streek, boven een klein deel waarvan de regenoperaties plaatsvonden, een kaartje met krommen van gelijke verhouding tot het gemiddelde. Als men dan juist boven het bewerkte gebied een kromme met zeer hoge parameter krijgt, zal men graag deze uitkomst als een succes van de poging om kunstmatig regen te maken willen zien. Zo mag men echter niet handelen, men moet veeleer krommen van gelijke waarschijnlijkheid schetsen over het gehele gebied.

d. Het simpelste procédé is er een van rangschikking in volgorde van toeneming. Stel op het station A is reeds 20 jaren gemeten. Het jaar 1955 (het 21ste jaar) gaf  $a_{21}$  mm. Rangschik dan de 21 jaarsommen tot  $a_1^1, a_2^1, \dots, a_{21}^1$ , waarin genoemde  $a_{21}^1$  bijv. plaats 8 inneemt ( $a_8^1 \equiv a_{21}^1$ ). Het relatieve rangnummer is dan  $r_a = 8/21$ . Doe dit ook voor station B, met basisreeks van 35 jaren, waarover 1955 de hoeveelheid  $b_{36}$  viel. In de reeks gerangschikte b-waarden neme  $b_{36}$  de plaats 10 in. Het relatieve rangnummer is dan  $r_b = 10/36$ . Als we dit voor alle stations doen, kunnen wij op de relatieve rangnummers isokrommen baseren.

Een en ander komt neer op een grafische schatting van de kans, dat de jaarsom bereikt of overschreden wordt.

Deze kans kan precieser becijferd worden indien de voor de frequentie verdeling van alle verleden gegevens een analytische uitdrukking gevonden kan worden, die de berekening van waarschijnlijkheden mogelijk maakt.

e. Het schema, geleverd door de normale verdeling, blijkt niet te gelden voor geïsoleerde regens. De statisticus beschikt over middelen, die een scheve verdeling in een verdeling volgens Gauss overvoeren.

f. Als men de totale hoeveelheden neerslag beschouwt, gevallen over een tijdvak van tenminste een maand en als deze nog niet aan een normale verdeling gehoorzamen, dan toch wel de wortel uit de hoeveelheid (Sommige auteurs verkiezen  $\sqrt[3]{}$ ).

g. Bij wijze van voorbeeld wordt het gebied der franse Alpen beschouwd en wel de winter 1954 op 1955 (december + januari +

februari + maart) (regen + sneeuw), omdat de hoeveelheden neerslag in deze winter buitengewoon groot waren.

Er werden twee kaarten getekend:

1. De kaart met krommen van gelijke verhouding  $p$  tot het gemiddelde. De referentie-periode moet voor alle stations in het gebied hetzelfde zijn (d.w.z. geheel dezelfde jaren). Genomen werd: 1935 t/m 1954 (20 j.) Stations met kortere basistijdvakken deden niet mee en van stations met langere basisperioden werden slechts de genoemde 20 jaren beschouwd. *fig. 1*

2. De kaart met krommen van gelijke overschrijdingskans. *fig. 2*. Bij deze kaart werd a.v. gehandeld: beschouw station A; de 20 wintersommen mogen heten  $H_1, H_2, H_3 \dots H_{20}$ , met gemiddelde  $G_a$ . De wortelwaarden  $\sqrt{H_1}, \sqrt{H_2} \dots \sqrt{H_{20}}$  hebben een gemiddelde  $m_a$  en een standaarddeviatie  $s_a$ . Ondersteld wordt dat deze  $\sqrt{H_1}$  een normale verdeling volgt. Zij  $q_a = (\sqrt{H_a} - m_a) : s_a$ , als  $H_a \equiv H_{21} =$  winterhoeveelheid 1954/55 te A. Deze  $q_a$  heet de gestandaardiseerde normaal verdeelde variabele, ~~waarvoor~~ uitvoerige tabellen de overschrijdingskans geven. Bijv. als  $q_a = 2$ , dan is de overschrijdings  $P(q_a) = 0.163$ ; bij  $q_a = 3$ , dan  $P(q_a) = 0.041$  etc.

h. Zie de twee figuren 1 en 2.

De krommen 1/500, 1/1000, 1/2000 zijn slechts curiositeitshalve toegevoegd. Men lette op het gebied, het hooggelegen dal van de Isère, rechts boven (fig. 1) met grote  $p$ . Men mag niet denken aan meetfouten (juist in dit gebied worden de regen- en sneeuwmeters dikwijls gecontroleerd), derhalve moet men naar zeer speciale oorzaken zoeken.

Een gedetailleerde studie der dagelijkse meteorologische situaties boven de Alpen leidde niet tot een "natuurlijke oorzaak", waardoor juist "la haute Isère" veel meer neerslag ontving dan de naburige dalen... dus invloed van de mens, hetgeen thans zeer waarschijnlijk was, want men had juist boven dit gebied en in deze winter verschillende keren kunstmatig regen gemaakt.

i. Als men fig. 1 met fig. 2 vergelijkt, zou men, althans op het eerste gezicht, tot een gelijke ligging der isokrommen willen besluiten. In werkelijkheid zijn de twee configuraties zeer verschillend.

De krommen  $P$  zijn eigenlijk tevens krommen van gelijke  $q$ . Een kromme met gegeven waarde van  $p$  (fig. 1) is tevens een isokromme  $P$  uit fig. 2 (d.w.z. langs de gehele kromme is de waarde van  $q$  hetzelfde)



als althans langs de kromme de waarde van  $s/m$  hetzelfde is. Is dit het geval? Neen, zie nl. fig. 3 (a en b), waarin voor 10 der stations uit de franse Alpen voor bedoelde 20 jaren voor elk der 12 maanden de waarde van  $s/m$  berekend werd (vervolgens werden uit de hand vloeiende krommen getrokken). Voor eenzelfde maand (bijv. december) en ten aanzien van de 20 jaren 1935 t/m 1954, blijkt de  $s/m$  van plaats tot plaats sterk te variëren (over enige tientallen kilometers soms met een factor 2).

j. Slot: Deze weinige voorbeelden brengen ons ertoe de grote betekenis van krommen van gelijke kans onder het oog te brengen. Wij willen niet voorstellen hen de welbekende en zeker nuttige krommen van gelijke percentuele afwijking t.o.v. het gemiddelde te doen vervangen. Ons voorstel is slechts een waarschuwend woord te laten horen tegen het te grote en soms uitsluitende gebruik van het gemiddelde.

#### B. Eigen commentaar en kritiek.

ad c. Er is vrijveel literatuur over de juiste statistische methode om te onderzoeken in hoeverre de resultaten van zulke proefnemingen niet puur toevallig zijn. Ik denk o.a. aan 1) het lijvige rapport, prepared by I. Lanmuir, uitgegeven door het Gen. Electric Res. Lab., onder de naam "Final Report, Project Cirrus; part II: Analysis of the effects of periodic seeding of the atmosphere with silver iodide; report No. RL - 785; May 1953) 2) het artikel: Random tests for the effects of cloud seeding", by J. Putter, verschenen in een uitgave van Hakirya, Tel Aviv (Israël), March 1956.

ad f. Een probleem, waarmee zich vele statistici bezighouden.

Zeer onlangs nog J. Neumann uit Israël, met wie ik hierover correspondeerde, naar aanleiding van een door hem samen met S. Kotz, in bovenbedoelde uitgave van Hakirya gepubliceerd artikel: "Some Pearsonianlike frequency functions capable of a modified normalization" (Zie mijn W.R. 57-002(III-198).

ad g. Om welke stations gaat het en vooralom hoevele? Hoe dicht is het net?

Wij willen op de laatste vraag dieper ingaan.

$G_a = \left( \sum_1^{20} H_i \right) / 20$  en  $m_a = \left( \sum_1^{20} \sqrt{H_i} \right) / 20$ ; te bedenken, dat  $m_a \neq \sqrt{G_a}$  in het algemeen, terwijl Serra de gelijkheid onderstelt (doch

dit niet vermeldt) Immers: beschouw een kromme met gegeven waarde van  $p$  in fig. 1, bijv.  $p_0$ . Dus voor station A (op deze kromme gelegen) is  $H_a = p_0 G_a$ , als  $H_a$  de hoeveelheid in de 21ste winter voorstelt. En voor station B, op dezelfde kromme gelegen, eveneens  $H_b = p_0 G_b$  etc. Voor A geldt p.d.  $q_a = (\sqrt{H_a} - m_a) : s_a = (\sqrt{p_0 G_a} - m_a) : s_a$ . Indien  $m_a = \sqrt{G_a}$ , dan is  $q_a = \sqrt{G_a} (\sqrt{p_0} - 1) : s_a = (\sqrt{p_0} - 1) : c_a$ , als  $c_a = s_a : \sqrt{G_a} = s_a : m_a$

Inderdaad: én als voor alle stations langs de kromme  $p = p_0$  het quotiënt  $s_a : \sqrt{G_a}$  dezelfde waarde heeft én als  $m_a = \sqrt{G_a}$  en  $m_b = \sqrt{G_b}$  enz., dan is deze curve tevens een kromme  $q = \text{constant}$ , t.w.  $(\sqrt{p_0} - 1) : c$ , dus ook een kromme  $P = \text{constant}$ .

Derhalve 2 voorwaarden! Hoe zwaar weegt de voorwaarde  $m \approx \sqrt{G}$ ? D.w.z. wanneer is  $\sum \sqrt{H_i} \approx \sqrt{\sum H_i} \cdot \sqrt{n}$ ? Men voelt wel aan, dat het antwoord zeer sterk samenhangt met de verhouding  $\sigma/\mu$  zelf, als  $\sigma = \text{universumwaarde van } s \text{ en } \mu = \text{universumwaarde van } G$ .

Een schatting a.v.:

Er geldt algemeen:

$$E F(x) = F_0(x) + \frac{1}{2} \left( \frac{\partial^2 F}{\partial x^2} \right)_0 \sigma^2, \text{ met } F_0(x) = F(\xi);$$

$\sigma^2 = \text{variantie van } x; \xi = E x = \text{gemiddelde van } x$ .

Hier  $F(x) = \sqrt{x}$ , zodat  $E(\sqrt{x}) = \sqrt{\xi} + \frac{\sigma^2}{8\xi\sqrt{\xi}} = \sqrt{\xi} + \frac{1}{8} c \frac{\sigma}{\sqrt{\xi}} = (1 + \frac{1}{8} c^2) \sqrt{\xi}$ , met  $c = \text{coëffi. v. variantie} = \frac{\sigma}{\mu}$ .

Aangezien  $c < 0.6$  is  $\frac{1}{8} c^2 < 0,045$  en  $E \sqrt{x} \approx \sqrt{\xi}$

Serra mag dus de onderstelling  $\sum \sqrt{H_i} = \sqrt{\sum H_i} \cdot \sqrt{n}$  maken.

Ik heb kaartje 1 herleid tot een kaartje 1<sup>st</sup> door te onderstellen, dat over het gehele franse Alpen-gebied  $\sigma/\mu$  dezelfde waarde zou hebben, terwijl aan de voorwaarde  $m \approx G$  per station voldoende goed voldaan zou worden.

Aangezien Serra geen tabel met de  $\sigma/\mu$  (beter is:  $s/\bar{x}$ )-waarden geeft, heb ik de figuur III gebruikt om voor elk der 10 stations zo goed als het ging de waarden van  $s/\bar{x}$  voor elk der 4 maanden dec, jan, feb en mrt af te lezen. Zie het volgende tabelletje.

waarden van  $s/\bar{x}$ , afgelezen uit fig. 3

		dec	jan	feb	mrt	gemiddeld
S <sub>1</sub>	Chamonix	0.26	0.24	0.38	0.44	0.32
S <sub>2</sub>	St. Foix	0.31	0.34	0.39	0.40	0.36
S <sub>3</sub>	Pralagnan	0.32	0.23	0.37	0.53	0.41
S <sub>4</sub>	Montsapey	0.34	0.37	0.42	0.50	0.41
S <sub>5</sub>	Jarrier	0.44	0.28	0.42	0.56	0.43
S <sub>6</sub>	Nevache	0.38	0.36	0.50	0.41	0.41
S <sub>7</sub>	Orchières	0.36	0.40	0.48	0.43	0.42
S <sub>8</sub>	Pelvoux	0.33	0.43	0.58	0.50	0.46
S <sub>9</sub>	Barcelonnette	0.42	0.40	0.59	0.47	0.47
S <sub>10</sub>	Valdeblore	0.44	0.59	0.63	0.53	0.55
	gem.	0.36	0.37	0.48	0.47	algeheel gem.: 0.42

De 10 stations leveren een gemiddelde  $(s/\bar{x}) = 0.42$ . Dan correspondeert kromme  $p = 2$  met kromme  $q = (1/0.42) (\sqrt{2}-1) = 0.983$  of  $P = 0.163$ . Evenzo  $p = 2.25 \rightarrow q = 1.19$  en  $P = 0.117$ ;  $p = 2\frac{1}{2} \rightarrow q = 1.38$  en  $P = 0.084$ ;  $p = 2\frac{3}{4} \rightarrow q = 1.565$  en  $P = 0.0585$  en  $p = 3 \rightarrow q = 1.738$  en  $P = 0.0412$ . Deze P-waarden heb ik naast de p-waarden genoteerd in fig. 1.\* Wij vergelijken vervolgens met fig. 2 (overgenomen van Serra). Veel overeenkomst in de ligging der krommen is niet te ontdekken. Wat zegt dit? Serra oordeelt dit zeer begrijpelijk, omdat wij zijn uitgegaan van een onderstelling t.w., dat over het gehele gebied de  $\sigma/\mu$ -waarden (thans wèl de universum-waarden) gelijk zouden zijn, aan welke onderstelling z.i. volgens fig. 4 (uit zijn toegezonden artikel overgenomen) geenszins voldaan wordt. Nu valt dit nog te bezien. Alles berust maar op 20 jaren! Terecht werd in de discussie door iemand de opmerking gemaakt, dat de krommen in fig. 4 wel erg grillig zijn, zo grillig, dat men aan de realiteit ervan gaat twijfelen. Spelen niet juist hier statistische (toevals-) fluctuaties een rol? Wij willen trachten hierop in te gaan. Allereerst wat nader op de grootheid  $c = s/\bar{x}$ . Gegeven een normale verdeling  $(\mu; \sigma)$ . De variantiecoëfficiënt  $\gamma$  is gedefinieerd als  $\gamma = \sigma/\mu$ . Iedere aselechte steekproef van  $n$  elementen uit dit universum levert een gemiddelde  $\bar{x}$ , een standaarddeviatie  $s$  en een variantiecoëfficiënt  $c = s/\bar{x}$ , die alle drie statistisch verdeeld zijn. De  $\bar{x}$  volgt exact een normale verdeling  $(\mu; \sigma/\sqrt{n})$ ; de  $s$  is bij benadering normaal verdeeld; hoe groter  $n$ , hoe beter de benadering; met gemiddelde  $\sigma$  en standaarddeviatie  $\sigma/\sqrt{2(n-1)}$ . Tenslotte de  $c$ ; deze volgt een scheve verdeling, doch de scheefheid neemt af en de verdeling gaat steeds minder van

een normale verschillen, naarmate  $n$  groter is;  $\mu_c = \gamma$  en  $\sigma_c^2 \approx \frac{(1 + 2\gamma^2)}{2(n-1)} \gamma^2$  Wij zien:  $\sigma_c/\gamma = \sqrt{(1 + 2\gamma^2)/2(n-1)}$ .

De kleinste onder alle 10 x 4  $s/\bar{x}$ -waarden is 0.23. (Pralognan; jan.). Aangezien onze enige schatting van de  $\sigma/\mu$ -waarde deze 0.23 is, is de beste schatting van  $\sigma_c$  ook  $0.23 \sqrt{\frac{1 + 2 \cdot 0.23^2}{2 \times 19}} = 0.23 \times 0.17 = 0.04$ . Zodat (als reeds voor  $n = 20$  de  $c$  een normale verdeling volgt) met een betrouwbaarheid 95% de  $\sigma/\mu$  tussen  $0.23 - 2(0.04)$  en  $0.23 + 2(0.04)$ , d.i. tussen 0.15 en 0.31 gelegen is.

Nu ook voor de grootste variantiecoëfficiënt in het tabelletje, t.w. 0.63 (Valdeblore; feb.). Hier is  $0.63 \sqrt{\frac{1 + 2 \cdot 0.63^2}{2 \times 19}} = 0.63 \times 0.217 = 0.137$ , zodat nu  $\sigma/\mu$  gelegen is tussen  $0.63 - 2(0.14)$  en  $0.63 + 2(0.14)$ , d.i. tussen 0.35 en 0.91. Het ziet er dus wel naar uit (nl. omdat het interval 0.15 tot 0.31 geheel buiten het interval 0.35 tot 0.91 gelegen is), dat er tussen enige van de 10 x 4  $\sigma/\mu$ -waarden, reeds op basis van de 10 x 4 steekproefwaarden, een statistisch verzekerd verschil bestaat, doch we zien hoe breed deze intervallen zijn (0.16 op 0.23, d.i. 70%, resp. 0.28 op 0.63, d.i. 44%) juist door de kleinheid van  $n$  (hier  $n = 20$ ). Door ditzelfde feit staat de realiteit van de jaarlijkse gang van de  $\sigma/\mu$ -waarde per maand voor ieder station statistisch niet vast, zodat wij voor elk station afzonderlijk de 4  $s/\bar{x}$ -waarden als een aselechte greep van 4 elementen uit het universum der  $s/\bar{x}$ -waarden mogen beschouwen, hetgeen wij verderop inderdaad gaan doen. Hoe het vermoeden, dat er een reële verandering van  $\sigma/\mu$  van plaats tot plaats bestaat, te toetsen? Zoals zo dikwijls, kunnen wij verschillende nulhypothesen stellen. De keuze ervan wordt onder meer bepaald door de reeds aanwezige kennis omtrent de situatie, over welke kennis wij in dit geval niet beschikken (Serra waarschijnlijk wel). Wij zouden de volgende nulhypothesen kunnen stellen:

a).  $H_0 \equiv$  alle stations hebben identieke, normaal verdeelde, waarschijnlijkheidsverdelingen der wortels uit de maandsommen neerslag (dec, jan, feb en mrt). Dit normaal verdeeld zijn zal niet te veel geëist zijn (als wij Serra mogen geloven). De nulhypothese heeft dan eigenlijk alleen betrekking op het identiek gelijk zijn. Er is a.h.w. 40 keren (10 stations; 4 maanden) een greep van 20 elementen gedaan, in welke grepen de 40  $s/\bar{x}$ -waarden: 0.26; 0.24 0.38 .... 0.53 gemeten werden, die een standaarddeviatie 0.08 bezitten. Onder de nulhypothese volgen deze  $s/\bar{x}$ -waarden een, zo goed als normale, verdeling (als  $n = 20$  voldoende groot is) rondom een gemiddelde en met een standaarddeviatie  $\sqrt{\frac{1 + 2\gamma^2}{2 \times 19}}$ ; de beste schatting voor  $\gamma$  is

het gemiddelde der 40 waarden 0.26 0.24 .... 0.53, zijnde 0.42, zodat de beste schatting van die theoretische standaarddeviatie is:  $0.42 \sqrt{\frac{1 + 2 \cdot 0.42^2}{38}} = 0.10$ . Aangezien 0.08 en 0.10 statistisch niet verschillen (het is niet nodig dit nog met de F-toets te toetsen) behoeven wij de  $H_0$  niet te verwerpen.

Serra zal misschien oordelen, dat deze  $H_0$  tóch moet verworpen worden, doch dan omdat hij meer weet dan in zijn voordracht vermeld wordt, bijv. omdat hij met zekerheid weet dat er een plaatselijke variatie is in de gemiddelde maandsom. Zo ja, dan heeft het al direct geen zin om bovengenoemde  $H_0$  te stellen.

b)  $H_0 \equiv$  de 10 universa, waaruit elk der rijen van 4 getallen (zie weer het tabelletje) komen, hebben eenzelfde gemiddelde, d.w.z. er is geen verandering van plaats tot plaats in de universumwaarde van  $s/\bar{x}$  (dus van  $\sigma/\mu$ ) per maand. Deze  $H_0$  laat zich goed toetsen met de gegeneraliseerde toets van Wilcoxon (d.i. H-toets). Wij geven dan rangnummers aan elk der  $10 \times 4 = 40$  getallen (1 t/m 40) en geven aan twee gelijken het gemiddelde rangnummer, enz. Daarna per rij de som  $S_i$  ( $i = 1, 2 \dots 10$ ). Vervolgens  $\frac{1}{2} \cdot 4 \cdot (40 + 1) = 82$  en per rij  $S_i - 82$ , en  $(S_i - 82)^2$ . Er komt

rangnummers				$S_i$	$(S_i - 82)^2$
3	2	$14\frac{1}{2}$	$20\frac{1}{2}$	40	1764
5	$8\frac{1}{2}$	16	18	$47\frac{1}{2}$	1190
6	1	$12\frac{1}{2}$	$34\frac{1}{2}$	54	784
$8\frac{1}{2}$	$12\frac{1}{2}$	23	32	76	64
$27\frac{1}{2}$	4	23	36	$90\frac{1}{2}$	56
$14\frac{1}{2}$	$10\frac{1}{2}$	32	$20\frac{1}{2}$	$77\frac{1}{2}$	20
$16\frac{1}{2}$	18	30	$25\frac{1}{2}$	84	4
7	$25\frac{1}{2}$	37	32	$101\frac{1}{2}$	342
23	18	$38\frac{1}{2}$	29	$108\frac{1}{2}$	702
$27\frac{1}{2}$	$38\frac{1}{2}$	40	$34\frac{1}{2}$	$140\frac{1}{2}$	3422
som					8449

Vervolgens maken wij gebruik van een nomogram (zie "Simple Approximations and nomograms for two ranking tests" by P.J. Rijkooort and M.E. Wise, Proceedings Kon.Ned.Akad.v.Wet., A.56, No.3) en er komt een overschrijdingskans van ongeveer 0.03 d.i. kleiner dan de gebruikelijke significantie-drempel 0.05 (strikt genomen mag het nomogram niet gebruikt worden als er gelijken zijn).

Conclusie:  $H_0$  moet verworpen worden, dus zeer waarschijnlijk is er een plaatselijke variatie in de  $\sigma/\mu$ -waarde (maar de realiteit van de

jaarlijkse gang, waar Serra óók de aandacht op vestigt, is niet bewezen), en heeft Serra dus gelijk.

Nog enige opmerkingen:

a). Waarom niet, waar mogelijk, van sommige stations álle meetjaren gebruikt? Voor die stations (met lange reeksen) zouden dan én G, én s, én m én s/m veel betrouwbaarder zijn dan voor de overige. Zo heeft men stevige referentie-waarden.

Jammer, dat de auteur per sé voor alle stations dezelfde basisperiode van helaas slechts 20 jaren heeft willen beschouwen. De grote invloed van het steekproefeffect maakt dan het significantstellen van regionale of temporale verschillen zeer moeilijk, terwijl voorts te veel over de ongelijkheid  $m \neq \sqrt{G}$  heen gelopen werd.

b). Ook in ons land wordt na afloop van iedere maand op een kaartje naast ieder regen-, hoofd- en termijnstation de totale hoeveelheid neerslag over deze maand genoteerd (ong. 300 stations). In het "Maandelijks Overzicht der Weersgesteldheid" wordt bovendien een kaartje getekend, waarop zijn onderscheiden de gebiedsdelen met een maandsom tussen 10 en 20 mm, 20 en 30 mm enz. (verschillend gearceerd), met parameters aangepast aan het geval in kwestie. Tevens wordt het land in streken verdeeld. Per streek wordt het gemiddelde berekend en voor dit gemiddelde wordt de procentuele afwijking t.o.v. het langjarig gemiddelde van dit streekgemiddelde berekend. Dit geeft zowel percentages als er streken gemaakt werden. Deze worden genoemd in de tekst. Met genoemd kaartje en deze tekst moet de lezer het dus doen. Wat doen? Hij zal ondermeer willen weten, waar in het land veel neerslag viel ("veel" in verhouding tot de hoeveelheden elders), doch ook of dit veel was voor dat gebied zelf (nu in de zin van zeldzaam). Een zeer groot positief percentage-getal (in de boven beschreven zin) zal wel tevens op een grote zeldzaamheid wijzen, doch zekerheid daaromtrent hebben wij niet, tenzij wij enige wijzigingen in de procedure aanbrenge.

1. In de eerste plaats lijkt mij een isohyeten-kaartje illustratiever dan een kaartje met verschillend gearceerde gebieden; dat zou ook meer in overeenstemming zijn met het temperatuurkaartje, waarop isokrommen van gelijke gemiddelde overdagtemperatuur getekend zijn.

2). Bedoeld stukje tekst zou kunnen worden weggelaten als een kaartje met krommen van gelijke relatieve afwijking t.o.v. de normale waarde (doch van op alle of zeer vele stations gebaseerd) getekend en toegevoegd zou worden.

3. Een stapje verder zou zijn een kaartje met iso-overschrijdingskans-krommen. Het zou nuttig zijn te onderzoeken of overal in het land de maandsommen zelf reeds normaal verdeeld zijn, en zo niet, misschien wel de wortelvormen (zie Serra) of misschien de logs. Normalisering lijkt wel geboden, want ná normalisering is het zeer gemakkelijk via de relatieve afwijking t.o.v. de normale waarde, de overschrijdingskans te berekenen (slechts aflezen uit een tabel voor de gestandaardiseerde normale verdeling).

Met behulp van de "bekende" 24 stations met basisreeksen 60 à 110 jaren, waarvan alle frequentieverdelingen van k-daagse sommen ( $k = 1, 2, 3, 5, 10, 30, 60, 90, 180, 360, 720, 1080, 1800$ ) gemaakt zijn of worden, moet het niet moeilijk zijn, zekerheid te verkrijgen over deze normaliseringsmogelijkheid. Misschien zal blijken, dat de variatiecoëfficiënten van de genormaliseerde maandsommen wel over het gehele land gelijk zijn (in tegenstelling tot Serra's conclusie), hetgeen de berekeningen zou vergemakkelijken.

Mijn voorstel is dus: publiceer iedere maand de volgende twee kaartjes.

- 1) een isohyetenkaartje, gebaseerd op de maandsommen zelf (mm).
- 2) een kaartje met isokrommen van overschrijdingskansen (%).

Voorbeeld: voor station S geldt 30 mm en 80%.

Interpretatie: er viel 30 mm en de kans op  $\leq 30$  mm is 20%, d.i. zulk een maand komt 1 keer in gem.  $100 : 20 = 5$  jaren voor. Een ander station  $S^1$  geeft toevallig ook 30 mm, doch 50%, d.w.z. hier komt een maand met  $\leq 30$  mm (of zo men wil  $\geq 30$  mm) 1 keer in gem.  $100 : 50 = 2$  jaren voor. Een derde station  $S^{11}$  geeft 75 mm en 5%. Hier is de frequentie van een maandsom  $\geq 75$  mm blijkbaar 1 keer in gem.  $100 : 5 = 20$  jaren. Derhalve had  $S^{11}$  een  $2\frac{1}{2}$  keer zo grote maandsom als  $S^1$ , en was deze  $20 : 2 = 10$  keer "zo zeldzaam". Op kaartje 2) kan men zeer snel zien, waar de hoeveelheden normaal waren, waar zeldzaam laag, waar zeldzaam hoog, mits deze begrippen maar opgehangen worden aan overschrijdingskansen. Het vergelijken van een kaartje over bijv. februari 1955 met dat over bijv. juli 1955 wordt nu ook zinvoller.



isokrommen p van gelijke verhouding tot  
het gemiddelde; dec'54-mrt'55; 1935-'54

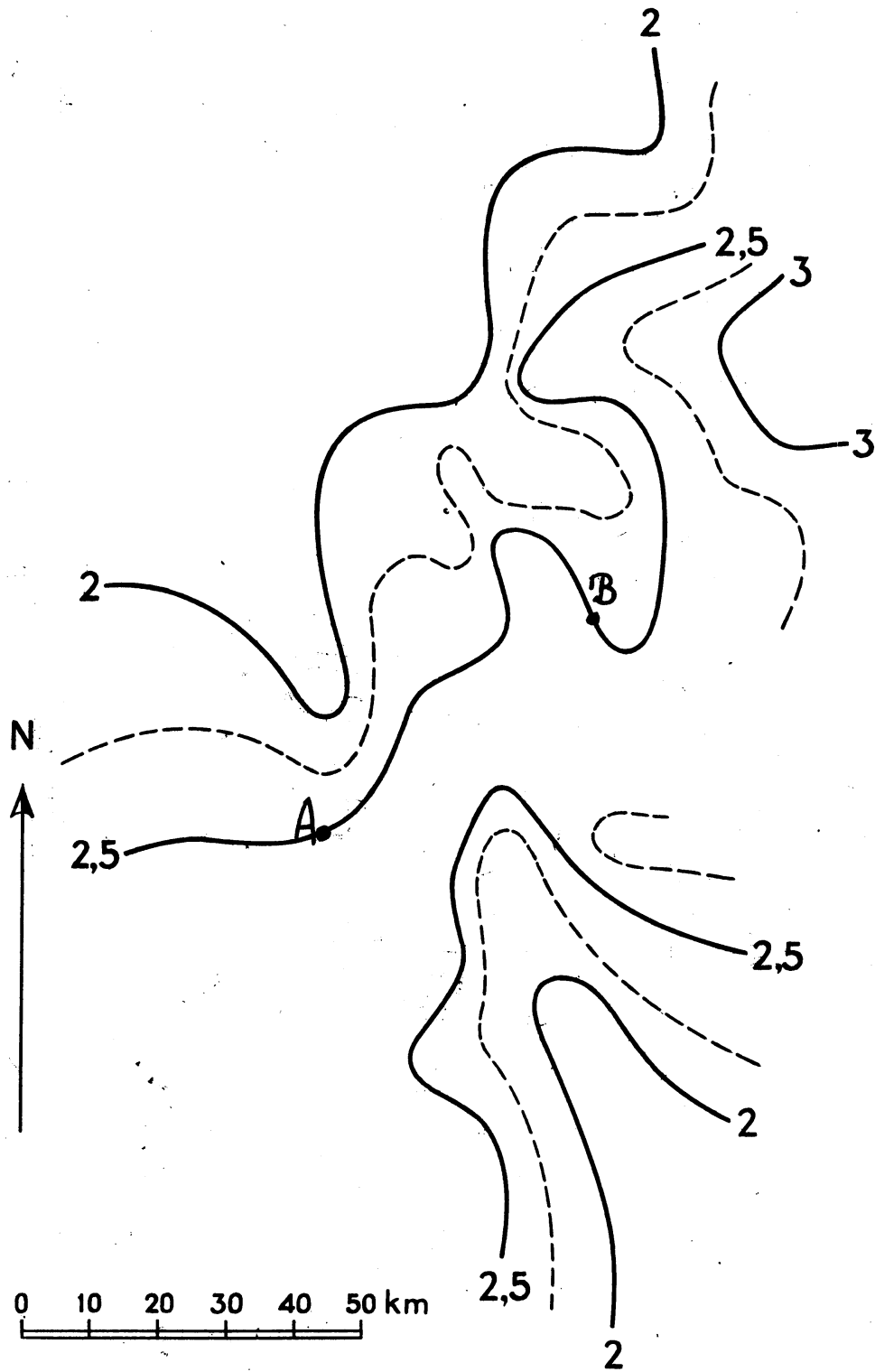


fig. 1





Meer  
van Genève

isokrommen P van gelijke overschrij-  
dingskans; dec'54 - mrt'55 ; 1935-'54

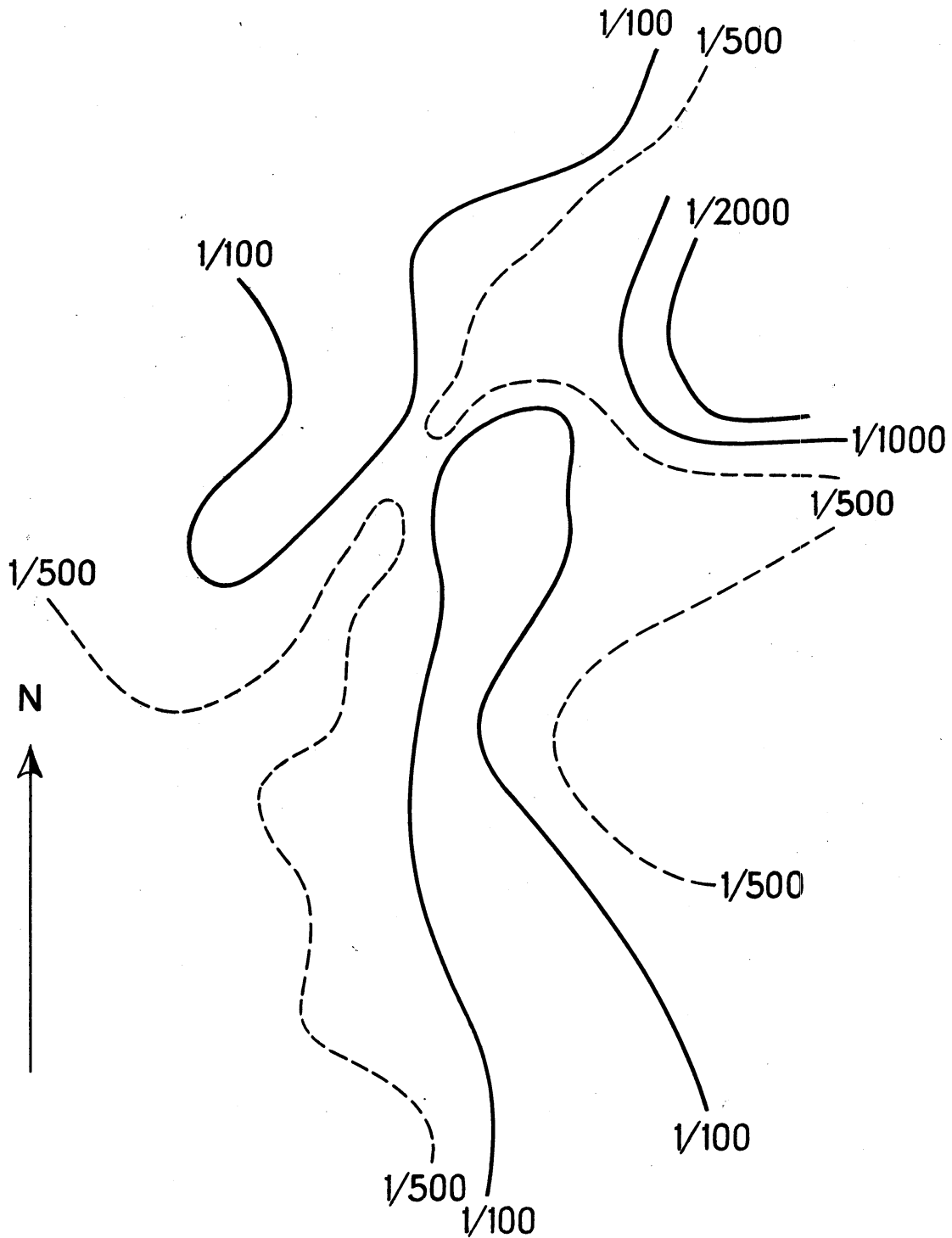


fig. 2

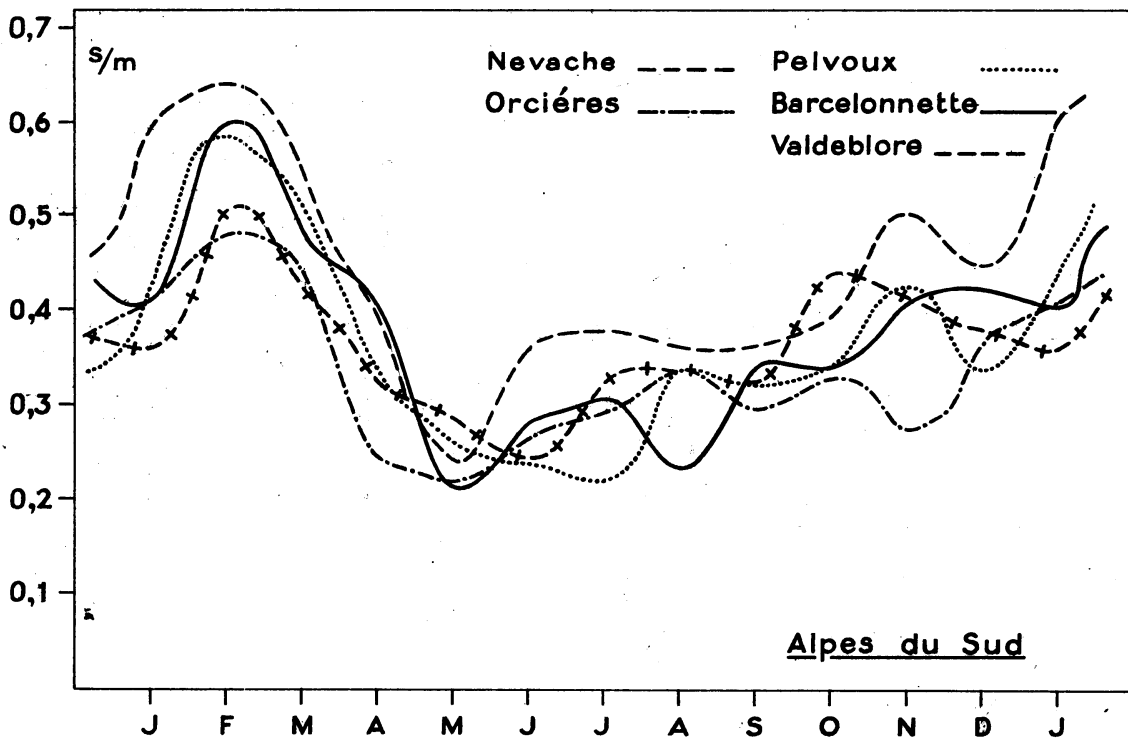
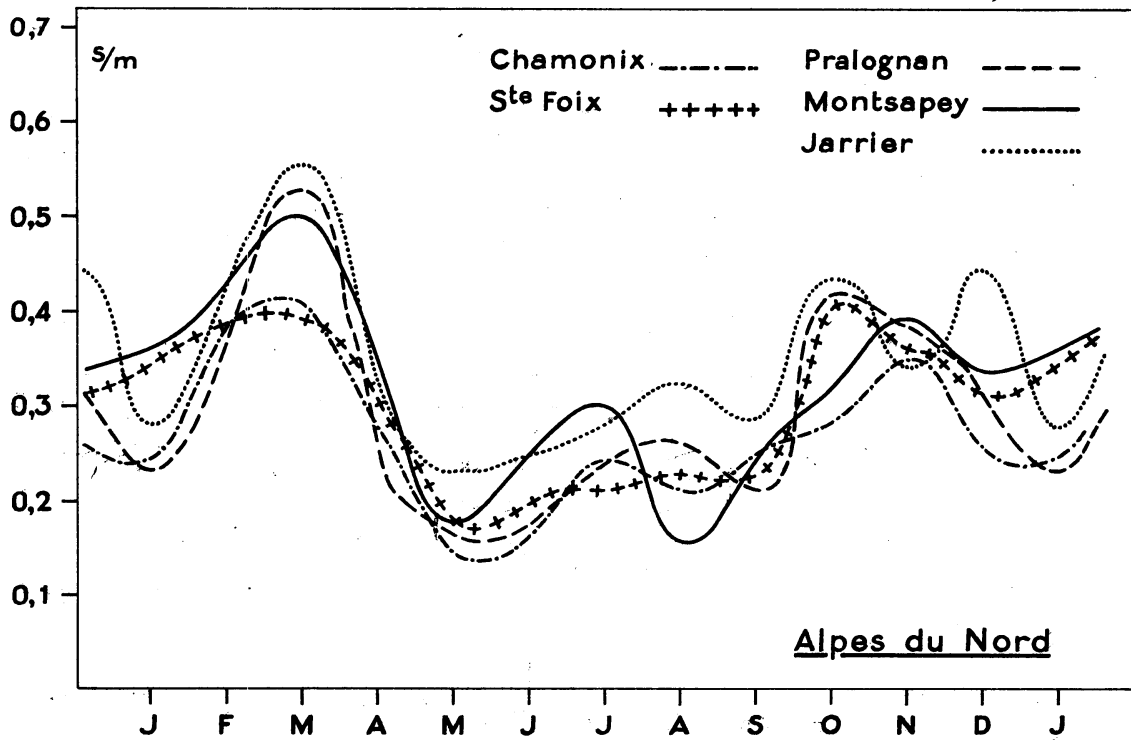


fig. 3



3. M. Bossolasco.

Densité de la neige et température de l'air.

De fysische eigenschappen van een sneeuwlaag hangen samen én met meteorologische omstandigheden gedurende de sneeuwvui, én met meteorologische factoren nadat de sneeuw gevallen is. Bossolasco deed verslag van de resultaten van metingen aan 53 geselecteerde sneeuwvuien, die plaats gevonden hadden bij temperaturen onder  $0^{\circ}\text{C}$ , terwijl wind afwezig was, en iedere sneeuwval tenminste 10 cm sneeuw geleverd moest hebben (d.i. een sneeuwduur van gewoonlijk meer dan 6 uren). De metingen werden verricht te Davos, Weissfluhjoch, 2600 m hoog. Gemeten werd de schijnbare sneewdichtheid<sup>p</sup> van het sneeuwdek (wisselende tussen 0.05 en 0.20 gr./cm<sup>3</sup>) en de temperatuur t. Deze temperatuur was het gemiddelde van minimum en maximum, gemeten aan de grond, op de dag, waarop de sneeuwvui viel of het overdaggemiddelde (waarnemingen om  $7\frac{1}{2}$ ,  $13\frac{1}{2}$ ,  $21\frac{1}{2}$  uur).

Als men<sup>p</sup> verticaal uitzet en t horizontaal, krijgt men 53 punten, die rondom  $11^{\circ}\text{C}$  het dichtst bij elkaar liggen én het laagst, hetgeen beduidt, dat de meeste sneeuwvuien vallen bij ong.  $-11^{\circ}\text{C}$ , terwijl daarbij de dichtheid het kleinst is ( $\sim 0.05$ ). Daarbij is  $12\frac{1}{2}$  cm sneeuw equivalent met 1 cm water; losse sneewpakking. Deze uitkomst m.b.t. de hoogste frequentie is niet vreemd, want metingen van sneeuwvui-frequenties op de Rigi, 1787 m hoog, over de jaren 1938-1949, deden een maximale frequentie zien ook bij  $-11^{\circ}\text{C}$ , op grond van 614 sneeuwvuien; op de Sántis, 2500 m, 1938-1952, 1870 sneeuwvuien eveneens; evenzo op de Weissfluhjoch, 2067 m, 1936-1951, 914 sneeuwvuien; op de Zugspitze, 2962 m, 1922-1934, 1757 sneeuwvuien en op de Sonnblick, 3106 m, 1938-1952 met 2309 sneeuwvuien. In alle 5 gevallen een relatieve frequentie van 6 à 11% bij temperaturen om  $-11^{\circ}\text{C}$ , tegen ongeveer 1% bij temperaturen van  $0^{\circ}\text{C}$  of onder  $-22^{\circ}\text{C}$ .

Het aardige is, dat men uit laboratoriumproeven leerde, dat de kristalkernen, die tot sneewvorming leiden, vooral bij  $-10$  à  $-11^{\circ}\text{C}$  ontstaan, en dat verschillende kristalvormen het meest frequent zijn bij verschillende (negatieve) temperaturen. De stervorm bij ongeveer  $-11^{\circ}\text{C}$ , waarbij de sneewpakking het meest los is en de schijnbare dichtheid het kleinst.

Bossolasco wilde temidden der 53 punten een curve tekenen, waarvan de vorm meer intuïtief aangevoeld werd dan door de punten zelf bewezen. In de discussie bleek van enige scepsis t.a.v. zijn handelwijze. Bovendien trachtte Bossolasco met behulp van zijn

uitkomsten begrijpelijk te maken, waardoor op verschillende hoogten in de bergen de hoeveelheden sneeuw zo verschillen. Zelfs trachtte hij een verband te zien met de vorming van gletsjers via sneeuw.

Ik wees de spreker op een artikel "Dependence of snowfall on temperature" by L. Gold and G. Williams, in Nature 177, june 1956, waarin de schrijvers betwijfelen of de temperatuur gemeten aan het aardoppervlak wel voldoende representatief is voor die in het sneeuwvormingsniveau.

Literatuur.

1. Mason and Ludlam: "Reports on Progress in Physics" 14 147 1951.
2. Rau: Geofisica pura e applicata 26 75 1953.
3. Bossolasco: Geofisica pura e applicata 28 208 1954
4. Bossolasco: Nature 174 362 1954
5. Ooge: Nature 175 465 1955
6. Gold and Williams: Nature 177 june 1956

4. M. de Quervain.

Modell eines Tagbogenschreibers.

Het doel van het in de titel genoemde instrument is om voor een willekeurig punt in het veld de door de horizon bepaalde optimaal mogelijke bezonningsomstandigheden te weten te komen.

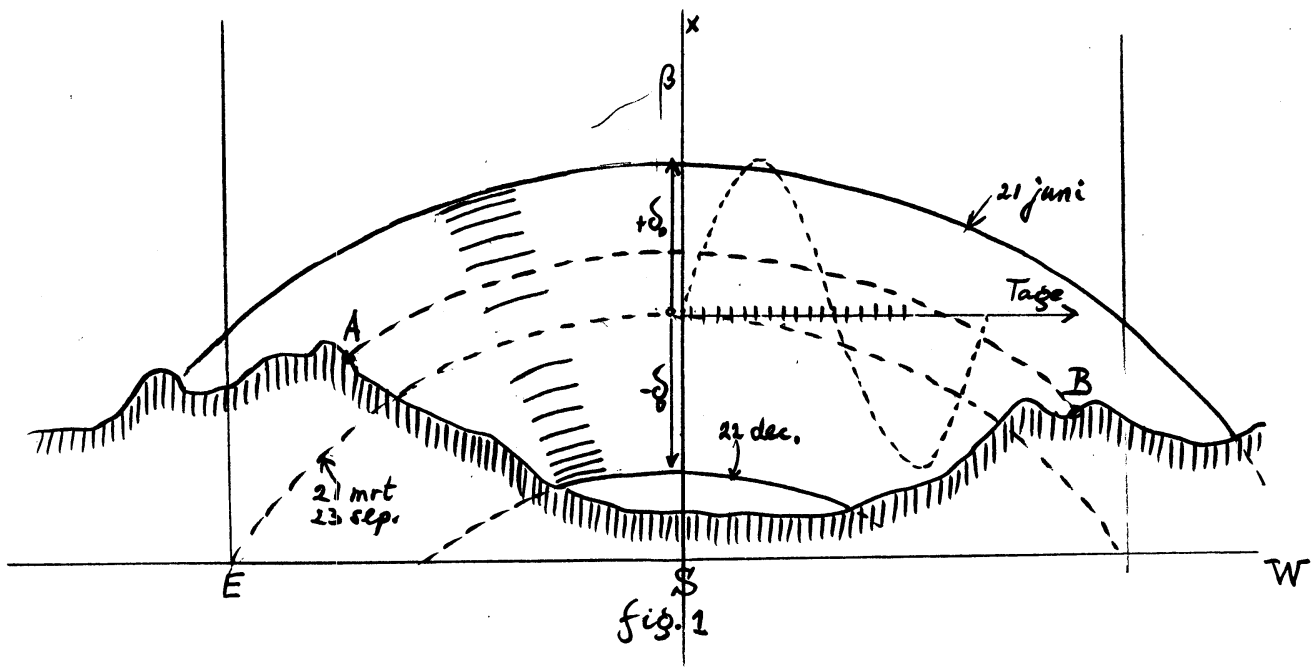
Men wil daartoe weten:

- a) voor een willekeurige dag in het jaar het tijdstip van zonsopkomst en -ondergang en daarmede de dagbooglengte.
- b) voor gekozen perioden, bijv. dekaden, maanden enz. het aantal uren van de theoretisch maximaal mogelijke zonnenschijnduur.

Deze getallen zijn, zoals bekend, van belang voor de klimaatsbeschrijving van plaatsen, in het bijzonder voor een vergelijking tussen de werkelijke zonnenschijnduur en de theoretisch maximaal mogelijke. Afgezien van de lokaalklimatologische en hydrologische onderzoekingen is de meting van de dagbooglengte ook van betekenis in samenhang met de beoordeling van bouwplaatsen (woningen, ziekenhuizen, etc.)

De meting van de dagbooglengte op zichzelf is een reeds lang opgelost probleem. Met de ontwikkeling van een nieuw instrument werd beoogd, het omvangrijke meetwerk, vooral als men het gehele land, vele stations, in kaart brengen wil, tot een minimum te reduceren.

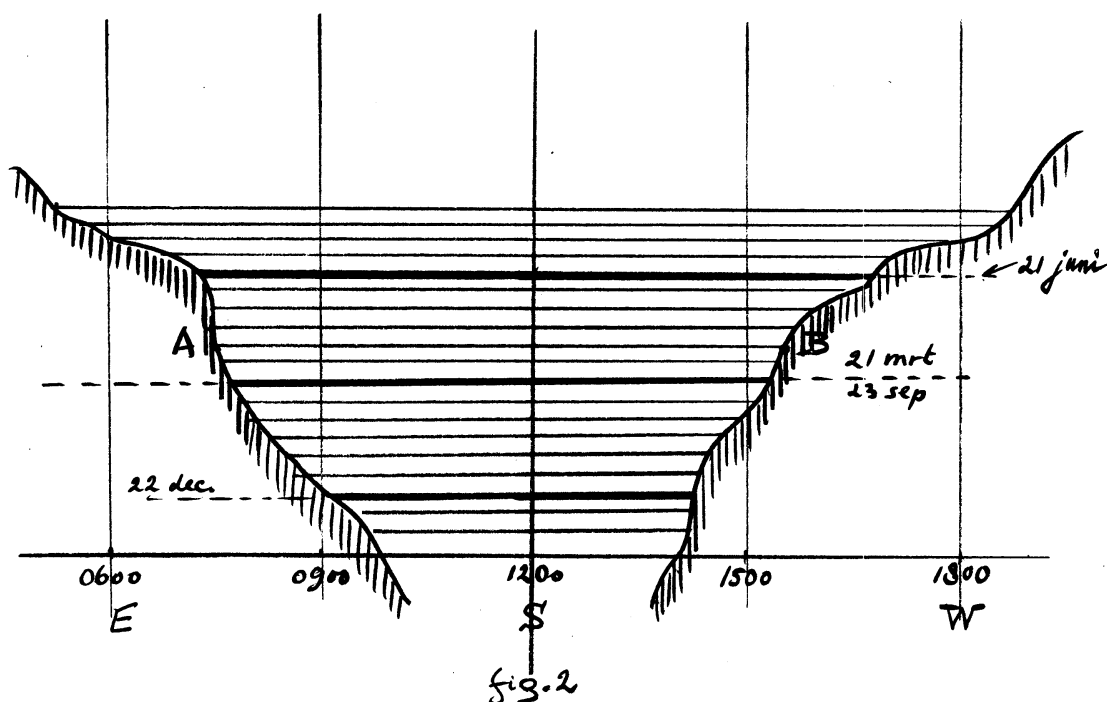
Van een bepaald standpunt uitgezien ziet de horizon, afgekweld in een plat vlak, er bijv. als volgt uit:



Alle zonnebanen tezamen verschijnen als een band met een breedte gelijk aan de dubbele absolute maximale declinatie  $\delta_0$  van de zon (die onafhankelijk is van de N.B.  $\varphi$  van de plaats). Men lette op de boog van 22 juni, die om 12 uur plaatselijke ware tijd een hoogte heeft van  $(90 - \varphi) + \delta_0$  ( $= 61\frac{1}{2}^\circ$  als  $\varphi = 52^\circ$ ;  $\delta_0 = 23\frac{1}{2}^\circ$ ); verder de boog van 21 maart én 23 september met een maximale hoogte van  $90 - \varphi$  en de laagste boog, die van 22 december,  $90 - \varphi - \delta_0$ . De maximaal mogelijke zonnenschijnduur op een willekeurige dag, bijv. 12 april, d.i. tevens 1 september, komt dan overeen met het gedeelte van de dagboog van die dag tussen de punten A en B. Het oppervlak van het zichtbare gedeelte van de band van alle dagbogen representeert de maximaal mogelijke zonnenschijnduur van een jaar, echter in vertekende vorm, want men kan niet planimetreren. De successieve dagbogen hebben n.l. in de tekening een wijdere ligging rondom 21 mrt en 23 sept dan rondom 22 juni en 22 december.

Teneinde wèl te kunnen planimetreren bedacht de Quervain een instrument, waarbij de zonnebanen equidistante rechte lijnen, evenwijdig aan de tijdas (met arithmetische schaal), worden. In dit geval is de maximale zonnenschijnduur per dag rechtevenredig met de lengte ervan in centimeters papier op het diagram. De uurlijnen worden equidistante rechten, alle  $\perp$  horizontale as (en de hoogtelijnen zullen zich krommen, doch deze neemt de Quervain niet in beschouwing).

Aangezien ik de Quervain beloofde de technische finesses niet te publiceren alvorens zijn patentaanvraag geregeld was, bespreek ik deze niet. Het diagram bij zijn instrument ziet er dan uit a.v.:





Het instrument kan op iedere breedte ingesteld worden. Men volgt door een kijker met de kruisdraden de contouren van de horizon (bergen, gebouwen, enz.) en onderwijl worden deze automatisch op het diagram getekend. Dit gebeurt in hoogstens 15 minuten.

In ons land wordt dit met een theodoliet gedaan en wel punt voor punt der contouren. Nadat op een nieuw punt is ingesteld, wordt azimuth en hoogte afgelezen. Deze worden overgebracht op een diagram (fig. 1) der zonnebanen, waarna uit de hand een vloeiende kromme getekend wordt. Het ~~geen~~ met de Quervain's instrument een kwartier kost, vraagt volgens onze methode een kleine dag werk, vooral als de horizon grillig van vorm is.

De Bilt, mei 1957.