

22 juli 1960

KONINKLIJK NEDERLANDS  
METEOROLOGISCH INSTITUUT

Wetenschappelijk Rapport W.R. 60-4 (RIII 251-1960)

Dr. C. Levert

Twee statistische modellen van persistentie  
in klimatologische tijdreeksen

(Two statistical models of persistency  
in climatological time series)

De Bilt - 1960

Kon. Ned. Meteor. Inst.  
De Bilt

All Rights Reserved.

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

Dr. C. Levert

Twee statistische modellen van persistentie  
in klimatologische tijdreeksen

INHOUD

0. Inleiding
  1. Begrensde en onbegrensde persistentie
  2. Model A
    - 2.1 Definitie
    - 2.2 Autocorrelatie-coëfficiënten  $\rho_k$
    - 2.3 Restvarianties in lineaire relaties in de  $y$ -reeks
    - 2.4 Effectieve aantallen
    - 2.5 De reeks van de eerste orde differenties  $\Delta$
  3. Model B
    - 3.1 Definitie
    - 3.2 Autocorrelatie-coëfficiënten  $\rho_k$
    - 3.3 Restvarianties in lineaire relaties in de  $y$ -reeks
    - 3.4 Effectieve aantallen
    - 3.5 De reeks van de eerste orde differenties  $\Delta$
  4. Toepassingen
    - 4.1 Het kiezen tussen de twee modellen A en B d.m.v. de F-toets
    - 4.2 Het kiezen tussen de twee modellen A en B via berekening van het repetitiegetal  $a$ 
      - 4.2.1 Via het equivalente herhalingsgetal  $q$
      - 4.2.2 De toets van Wald-Wolfowitz
      - 4.2.3 Via de  $r_k$ -waarden
    - 4.3 Schatting van de autocorrelatie-coëfficiënt  $\rho_k$
    - 4.4 Het kiezen tussen de twee modellen A en B via de autocorrelatie-coëfficiënt  $\gamma_2$  in de reeks der eerste orde differenties
  5. Persistentie-vrije reeksen
  6. Stationaire reeksen
- APPENDIX
- A. Model A met niet-zuivere persistentie
  - B. Model B met persistentie in de  $u$ 's
  - C. Kunstmatige reeksen met persistentie
    - C.1 Positieve persistentie
      - C.1.1 Een reeks van twee alternatieven
      - C.1.2 Een reeks van waarden van een continu verdeelde grootheid
    - C.2 Negatieve persistentie
      - C.2.1 Een reeks van twee alternatieven
      - C.2.2 Een reeks van waarden van een continu verdeelde grootheid
7. Summary



## 0. Inleiding

Dit rapport is ontstaan naar aanleiding van een uitvoerige gedachtenwisseling met Dr. Sneyers, Kon. Meteorologisch Instituut van België, Ukkel. In de appendix zijn nog enkele beschouwingen, los hiervan opgenomen.

## 1. Begrensde en onbegrensde persistentie

De meeste meteorologische grootheden (luchtdruk, temperatuur, neerslag, enz.) blijken in de tijdreeksen van hun waarden niet persistentie-vrij<sup>1)</sup> te zijn, anders gezegd: autocorrelatie te vertonen. In sommige statistische studies heeft het weinig zin zich hierover zorgen te maken, in andere daarentegen wel, in het bijzonder als men de afhankelijkheid tussen opeenvolgende waarden in getallen wil en moet uitdrukken, ten einde deze weer in andere berekeningen te kunnen gebruiken. Dan is het noodzakelijk om verschillende statistische "modellen" voor persistentie op te stellen en aan het empirische materiaal, indien dit althans voldoende groot is, te toetsen. Het spreekt vanzelf, dat men bovendien het liefst die modellen zal kiezen, zo mogelijk, welke bedoelde berekeningen eenvoudig houden.

Er zijn reeds vele modellen voorgesteld en geregeld ziet men auteurs weer nieuwe suggereren. In dit rapport worden er slechts twee besproken. Het eerste, gemakshalve model A geheten, wordt gekenmerkt door een autocorrelatie, waarvan de autocorrelatie-coëfficiënt van de orde  $k$ , d.i.  $\rho_k$ , met toenemende  $k$  asymptotisch naar nul gaat. Het tweede model, B, echter bezit begrensde persistentie; alle  $\rho_k \equiv 0$  voor  $k \geq k_0$ , met  $k_0$  eindig. Wij willen deze twee modellen en hun consequenties in verschillende statistische kwesties, uitvoerig bespreken.

## 2. Model A

### 2.1 Definitie

Zij een tijdreeks ....  $y_i, y_{i+1}, y_{i+2} \dots$  opgebouwd als volgt. Elke volgende term  $y_i$  wordt verkregen door de vorige  $y_{i-1}$  met een vaste factor  $f$  te vermenigvuldigen en het product met een toevalsvariabele  $\epsilon_i$

---

1) Voor de definitie van "persistentie-vrij" en van "stationair" zie § 5 en 6.

te vermeerderen. Zulk een proces is een voorbeeld van een Markoff-keten<sup>1)</sup> van het eerste type, waarbij de waarde, die elk element aanneemt, slechts afhankelijk is van de waarde van het direct daaraan voorafgaande element

$$y_i = f \cdot y_{i-1} + \varepsilon_i ; \quad i = 1, 2 \dots \quad (1)$$

Ondersteld zij, dat  $E\varepsilon_i = 0$  en  $\sigma^2(\varepsilon_i) = \sigma_\varepsilon^2$  voor iedere  $i$ , hetgeen (2) betekent, dat op iedere plaats  $i$  in de reeks  $\varepsilon_i$  gedacht kan worden aselekt gekozen te zijn uit een universum met gegeven gemiddelde nul en gegeven variantie  $\sigma_\varepsilon^2$ .

Voorts zij:

$$E\varepsilon_i \cdot \varepsilon_j = 0 \quad \text{voor iedere } i \neq j \quad \text{en} \quad (3)$$

$$E\varepsilon_i \cdot y_j = 0 \quad \text{voor elke } i, j, \quad , \quad \text{hetgeen} \quad (4)$$

ongeveer betekent, dat noch tussen  $\varepsilon_i$  en  $\varepsilon_j$ , noch tussen  $\varepsilon_i$  en  $y_j$  enig verband bestaat.

Over de waarde van  $f$  wordt voorlopig nog niets ondersteld. Het zal echter blijken, dat het verstandig is om  $0 < |f| < 1$  te postuleren. Immers:

$$y_N = f^N y_0 + f^{N-1} \varepsilon_1 + f^{N-2} \varepsilon_2 + \dots + f \varepsilon_{N-1} + \varepsilon_N , \quad (5)$$

als onze reeks met  $y_0$  begint.

Alleen als  $0 < |f| < 1$ , dan is  $0 < y_N < y_0 + \sum_1^N \varepsilon_i$ , waaruit (6)

$$0 < \lim \frac{1}{N} y_N < \lim \frac{1}{N} y_0 + \lim \frac{1}{N} \sum_1^N \varepsilon_i \quad \text{als } N \rightarrow \infty .$$

Vanwege (2) is dan  $\lim_{N \rightarrow \infty} y_N/N = 0$ , wat ook  $y_0$  moge zijn. Verder (7)

is

$$\sum_1^N y_i = f \sum_0^{N-1} y_i + \sum_1^N \varepsilon_i = f \{ \sum_1^N y_i + y_0 - y_N \} + \sum_1^N \varepsilon_i , \quad \text{waaruit}$$

---

1) Uit "A dictionary of statistical terms" by M.C. Kendall and W.R. Buckland: Een Markoff's proces is een stochastisch proces, zodanig dat de voorwaardelijke kansverdeling voor de toestand op ieder toekomstig ogenblik, gegeven de tegenwoordige toestand, niet wordt beïnvloed door enige additionele kennis van de geschiedenis van het stelsel. Een Markoff-keten is een proces  $[x_t]$ , waarbij de tijdparameter discontinu is.

$$(1 - f) \frac{1}{N} \sum_1^N y_i = \frac{1}{N} f(y_0 - y_N) + \frac{1}{N} \sum_1^N \varepsilon_i ; \text{ met (7) is dan}$$

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_1^N y_i \equiv \varepsilon_y = 0 \quad (8)$$

Anders geformuleerd: elke  $y_i$  kan worden beschouwd als de afwijking t.o.v. een gemiddelde waarde. Beschouw bijv. de overdagtemperaturen der dagen  $i = 1, 2, 3 \dots 365, 366, \dots 730, 731 \dots$ ; een  $y_i$  stelt dan voor de op dag  $i$  gemeten waarde, verminderd met het veeljarige gemiddelde van die dag, en verder nog gedeeld door de standaarddeviatie der waarden van die datum (de laatste toevoeging vanwege de onderstelling dat  $\sigma_\varepsilon^2$  onafhankelijk van  $i$  is).

Zij dus voortaan  $0 < f < 1$ . Wij zullen aantonen, dat  $f$  de autocorrelatie-coëfficiënt van de eerste orde in de  $y$ -reeks is.

## 2.2 Autocorrelatie-coëfficiënten $\rho_k$

Per definitie is  $\rho_k$ , de autocorrelatie-coëfficiënt van de orde  $k$  in de  $y$ -reeks,

$$\rho_k \equiv \frac{\varepsilon_{y_i \cdot y_{i+k}} - \varepsilon_{y_i} \cdot \varepsilon_{y_{i+k}}}{\sigma_{y_i} \cdot \sigma_{y_{i+k}}} , \text{ waarin p.d.} \quad (9)$$

$$\varepsilon_{y_i \cdot y_{i+k}} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N y_i \cdot y_{i+k} \quad \text{en} \quad (10)$$

$$\sigma_{y_i}^2 = \lim_{N \rightarrow \infty} \frac{1}{N-1} \left[ \sum_{j=i}^{N+i-1} y_j^2 - \left\{ \sum_{j=i}^{N+i-1} y_j \right\}^2 : N \right]$$

Men kan bewijzen, dat  $\sigma_{y_i} \equiv \sigma_y$ , voor elke  $i$  en wel  $\sigma_y = \sigma_\varepsilon : \sqrt{1 - \rho^2}$ .

Vanwege (8), (4) en (1) is:

$$\rho_k = \frac{\varepsilon_{y_i \cdot y_{i+k}}}{\sigma_y^2} = \frac{\varepsilon_{y_i} (f \cdot y_{i+k-1} + \varepsilon_{i+k})}{\sigma_y^2} = \frac{f \varepsilon_{y_i \cdot y_{i+k-1}} + \varepsilon_{y_i} \cdot \varepsilon_{i+k}}{\sigma_y^2} =$$

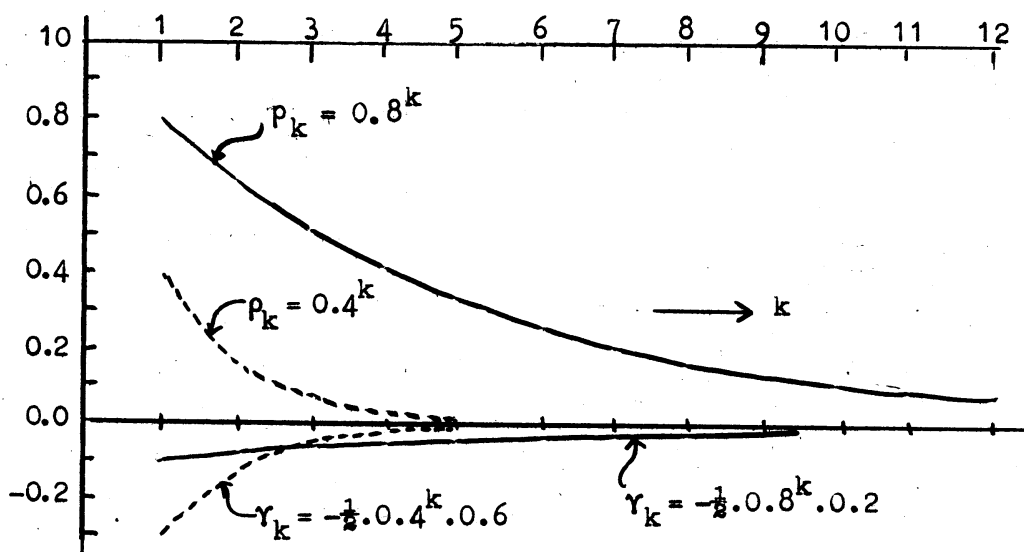
$$f \frac{\varepsilon_{y_i \cdot y_{i+k-1}}}{\sigma_y^2} \equiv f \rho_{k-1} . \quad \text{Derhalve:}$$

$$\rho_k = f \rho_{k-1} = f^2 \rho_{k-2} = \dots = f^{k-1} \rho_1 = f^k \cdot \rho_0, \text{ met } \rho_0 \equiv \frac{\varepsilon_{y_i \cdot y_i}}{\sigma_y^2} \equiv \frac{\sigma_y^2}{\sigma_y^2} = 1.$$

Aldus blijkt:  $f \equiv \rho_1$ . Laten wij gemakshalve de index 1 in  $\rho_1$  weg, dan is

$$\boxed{\rho_k = \rho^k} \quad (11)$$

Deze persistentie heet p.d. "zuiver" (voor de niet-zuivere, zie appendix). De autocorrelatie-coëfficiënten dalen met toenemende orde als een meetkundige reeks en worden dus nimmer exact nul. De persistentie heet onbegrensd. (Het kan wel, dat het niet mogelijk is om alle  $\rho_k$  van een zekere waarde van  $k$  af, bijv. met  $k \geq 4$ , van nul te onderscheiden, doch dit is een praktisch aspect; in het bovenstaande is de zaak principieel gesteld.)



Verder zagen wij reeds  $\sigma_\varepsilon^2 = (1 - \rho^2) \sigma_y^2$ . (12)

Noem  $\beta = \sigma_\varepsilon : \sigma_y$ , dan is (13)

$$\boxed{\beta^2 + \rho^2 = 1} \quad (14)$$

- 1) Als wij stellen  $y_i = \rho y_{i-1} + \tau y_{i-2} + \varepsilon_i$ , dan laat zich gemakkelijk bewijzen,  $\rho_k = \frac{E y_i \cdot y_{i+k}}{\sigma_y^2}$  stellende, dat  $\rho_k = \rho \cdot \rho_{k-1} + \tau \rho_{k-2}$  (in ons model is  $\rho_k = \rho \cdot \rho_{k-1}$ )  
 Uitwerkende:  $\rho_k = \rho \rho_{k-1} + \tau \rho_{k-2} = \rho^2 \rho_{k-2} + 2\rho \cdot \tau \rho_{k-3} + \tau^2 \rho_{k-4} = \rho^3 \rho_{k-3} + 3\rho^2 \tau \rho_{k-4} + 3\rho \cdot \tau^2 \rho_{k-5} + \tau^3 \rho_{k-6}$  enz. Het is moeilijk een plezierige algemene uitdrukking voor  $\rho_k$  in  $\rho$ ,  $\tau$  en  $k$  te vinden. Er komt  $\rho_0 \equiv 1$ ;  $\rho_1 = \frac{\rho}{1-\tau}$  (als  $\tau = 0$ , dan  $\rho_1 = \rho$ );  $\rho_2 = \frac{\rho}{1-\tau} + \tau$ ;  $\rho_3 = \frac{\rho^3}{1-\tau} + \rho\tau(1 + \frac{1}{1-\tau})$ ;  $\rho_4 = \frac{\rho^4}{1-\tau} + \tau(\rho^2 + \tau + \frac{2\rho^2}{1-\tau})$  enz. Elke  $\rho_k \rightarrow \rho^k$  als  $\tau \rightarrow 0$ . Nú hebben wij te doen met een soort van Markoff-keten, type II. Lees M.G. Kendall "On autoregressive time series" Biom. 33 105 1943/46.



Commentaar: naarmate de toevalsvariabele  $\varepsilon$  "minder betekent" (d.i.  $\beta \rightarrow 0$ ) gaat  $\rho \rightarrow 1$ , want dan gaat de stochastische afhankelijkheid in een functionele (lineaire) over, nl. ongeveer  $y_i = \rho y_{i-1}$ .

### 2.3 Restvarianties in lineaire relaties in de y-reeks

De lineaire relatie tussen  $y_i$  en de  $k^{\text{de}}$  voorafgaande term  $y_{i-k}$  kan geschreven worden als:

$$y_i = \rho_k y_{i-k} + v_i, \text{ met } \rho_k = \rho^k \text{ en een} \quad (15)$$

$$\text{restvariantie } \sigma_y^2 = (1 - \rho_k^2) \sigma_y^2 = \frac{1 - \rho^{2k}}{1 - \rho^2} \sigma_\varepsilon^2 \quad (15a)$$

Met toenemende  $k$  en vaste  $\rho$  (d.i. met toenemende afstand tussen de gelieerde  $y$ 's) neemt deze restvariantie toe tot  $\frac{1}{1 - \rho^2} \sigma_\varepsilon^2 = \sigma_y^2$ . Dit is zeer begrijpelijk:  $y_i$  en  $y_{i-k}$  zijn asymptotisch onafhankelijk met  $k \rightarrow \infty$ .

### 2.4 Effectieve aantallen

Beschouw in de  $y$ -reeks een lopende groep van  $n$  successieve  $y$ 's:  $y_k, y_{k+1}, \dots, y_{k+n-1}$  met  $k = 1, 2, 3, \dots$ . Het groepsgemiddelde zij  $\bar{y}_k^n$ . Het universum der  $\bar{y}^n$ -waarden heeft een gemiddelde nul en een variantie

$$\sigma^2(\bar{y}^n) = \sigma_y^2 / n_e, \quad (16)$$

waarin  $n_e = n/\omega =$  effectieve aantal, met  $n =$  nominale aantal en  $\omega =$  equivalente herhalingsgetal van Bartels.

Men kan bewijzen dat  $\omega$  met de autocorrelatie-coëfficiënten  $\rho_k$  der initiaalreeks als volgt samenhangt:

$$\omega = 1 + 2 \left[ \frac{n-1}{n} \rho_1 + \frac{n-2}{n} \rho_2 + \dots + \frac{1}{n} \rho_{n-1} \right]. \quad (16a)$$

Als elke  $\rho_k \geq 0$ , dan is zeker  $\omega > 1$  en  $n_e < n$ .

Substitueer  $\rho_k = \rho^k$  in (16a) en er komt

$$\omega = \frac{1+\rho}{1-\rho} - 2 \frac{\rho}{n} \frac{1-\rho^n}{(1-\rho)^2} \quad (17)$$

Indien  $n \gg 2\rho \frac{1-\rho^n}{1-\rho^2}$ , dan: hoe kleiner  $\rho$  en groter  $n$ , hoe dichter ligt  $\omega$  bij  $(1+\rho) : (1-\rho)$  en  $n_e$  bij  $n \frac{1-\rho}{1+\rho}$ .

Met toenemende  $\rho$  neemt  $n_e$  af.

2.5 De reeks van de eerste orde-differenties  $\Delta$

Zij  $\Delta_i \equiv y_i - y_{i-1}$ ,  $i = 1, 2 \dots$ ; deze  $\Delta$ 's heten de eerste orde (18) differenties.

De nieuwe reeks is  $\Delta_1, \Delta_2, \Delta_3 \dots$ . Ook deze reeks bezit autocorrelatie met coëfficiënten  $\gamma_k$  ( $k = 1, 2 \dots$ ). Wij willen  $\gamma_k$  in  $\gamma_1$  en  $k$  uitdrukken.

Algemeen geldt p.d.

$$\gamma_k = \frac{\epsilon \Delta_i \cdot \Delta_{i+k} - \epsilon \Delta_i \cdot \epsilon \Delta_{i+k}}{\sigma_\Delta^2} \quad (19)$$

Uitwerkende

$$\gamma_k = \frac{-\rho_{k-1} + 2\rho_k - \rho_{k+1}}{2(1-\rho)} \quad (20)$$

Aangezien  $\rho_k$  een functie van  $k$  is (in model A:  $\rho_k = \rho^k$ ) is ook  $\gamma_k$  een functie van  $k$ .

Wij merken op, dat het teken van  $\gamma_k$  niet vaststaat, ook al is  $\rho > 0$ .

Substitueer  $\rho_k = \rho^k$ . Er komt

$$\gamma_k = -\frac{1}{2} \rho^{k-1} (1-\rho) \quad , k = 1, 2 \dots \quad (21)$$

$\gamma_k$  is negatief voor elke  $k$  (natuurlijk is  $0 \leq \rho < 1$  gedacht)

Als de initiaalreeks persistentie-vrij is ( $\rho = 0$ ), dan is  $\gamma_1 = -\frac{1}{2}$ ;  $\gamma_a \equiv 0$ ,  $a \geq 2$ .

Ook de (negatieve)  $\gamma_k$ 's volgen een meetkundige reeks (mits  $\rho > 0$ ).

Er bestaat een lineaire regressie tussen  $\Delta_i$  en de  $k^{\text{de}}$  voorgaande  $\Delta_{i-k}$ .

$$\Delta_i = \gamma_k \cdot \Delta_{i-k} + \tau_i \quad , \text{ waarbij als restvariantie}$$

$$\sigma^2(\tau) = (1 - \gamma_k^2) \sigma_{\Delta}^2 = 2 \frac{1 - \frac{1}{2} \rho^{2(k-1)} (1 - \rho)^2}{1 + \rho} \sigma_{\Delta}^2, \text{ waarin} \quad (22)$$

$$\begin{aligned} \sum \sigma_{\Delta}^2 &= 2(1 - \rho) \sigma_y^2 \\ \text{h} \sigma_y^2 &= \frac{1}{1 - \rho^2} \sigma_{\varepsilon}^2 \end{aligned}$$

Met toenemende  $k$  ( $\Delta_{i-k}$  en  $\Delta_i$  verder verwijderd) en vaste  $\rho$  neemt deze  $\sigma^2(\tau)$  asymptotisch toe tot  $\frac{2\sigma_{\varepsilon}^2}{1+\rho} = 2(1 - \rho) \sigma_y^2$ .

Met afnemende  $\rho$  ( $\rightarrow 0$ ) en vaste  $k$ , gaat de restvariantie naar  $2\sigma_{\varepsilon}^2 = 2\sigma_y^2$

Wij beschouwen nu ook in de  $\Delta$ -reeks een lopende groep van  $n$  successieve  $\Delta$ 's:  $\Delta_k, \Delta_{k+1} \dots \Delta_{k+n-1}$ , met een gemiddelde  $\bar{\Delta}_{(k)}^n$  ( $k = 1, 2, 3 \dots$ ). Het universum der  $\bar{\Delta}^n$  waarden heeft zekere variantie en wel  $\sigma^2(\Delta) / n_e$ , waarin  $n_e =$  effectieve aantal  $= n/\omega$ , waarin  $\omega$  gedefinieerd is door:

$$\omega = 1 + 2 \left[ \frac{n-1}{n} \gamma_1 + \frac{n-2}{n} \gamma_2 + \dots + \frac{1}{n} \gamma_{n-1} \right], \text{ met } \gamma_k \text{ zie (21)}. \quad (23)$$

Uitwerkende:

$$\omega = \frac{1 - \rho^n}{1 - \rho} \text{ en } \boxed{n_e = \frac{1 - \rho}{1 - \rho^n} n^2}, \text{ zodat } n < n_e < n^2 \text{ voor } 1 > \rho > 0 \quad (23a)$$

Men bedenke: thans is  $n_e > n$ , louter vanwege het feit, dat alle autocorrelatie-coëfficiënten in de  $\Delta$ -reeks negatief zijn. Voor  $\rho = 0$  (de initiaal  $y$ -reeks persistentie-vrij) staat er (voor de  $\Delta$ -reeks)  $n_e = n^2$ . Op dit opvallende resultaat wees o.a. v.d. Bijl in zijn dissertatie "Toepassing van statistische methoden in de klimatologie", 1952, p. 45.

Opmerking: wij hadden met dit aspect te maken, toen wij in onze studie van de interdiurne variabiliteit bij de 260 in de eerste dekaden van 26 januari-maanden gelegen dagelijkse minimumtemperaturen de 260  $\Delta$ -waarden (interdiurne veranderingen) berekenden en daarvan de standaarddeviatie. Deze variantie heeft niet een aantal graden van vrijheid gelijk aan 260 (of 259), maar  $26 \cdot 10_e$  als  $10_e$  het effectieve aantal is, dat in de plaats van het nominale 10 gezet moet worden. Bijv. is voor  $\rho = 0.7$  de  $10_e = 31$ ;  $10 < 31 < 10^2$ . Evenzo voor bijv. juli. Wil men vervolgens onderzoeken of het verschil tussen de  $s_{\Delta}^2$  van januari en die van juli significant is (jaarlijkse gang), dan is

kennis van de aantallen graden van vrijheid, waarop de twee steekproef-varianties berusten, nodig. Hoe groter deze zijn (d.i. bij kleinere  $\rho$ ) hoe "vlugger" wordt reeds een klein verschil tussen de gemeten  $s_{\Delta}^2$ -waarden voor significant verklaard. Een en ander is mathematisch niet volkomen correct. Zie Appendix in W.R. 60-2.

### 3. Model B

#### 3.1 Definitie

Laat de reeks  $x_1, x_2, x_3 \dots$  bestaan uit onderling even grote groepjes van gelijke waarden, bijv.  $u_1, u_1, u_1, u_2, u_2, u_2, u_3, u_3, u_3, u_4, u_4, u_4 \dots$ . In dit voorbeeld is  $x_1 = x_2 = x_3 = u_1$ ;  $x_4 = x_5 = x_6 = u_2 \neq u_1$ ;  $x_7 = x_8 = x_9 = u_3 \neq u_2$  en  $\neq u_1$  enz. Het "repetitie-getal"  $a = 3$ .

Laten deze  $u_i$ 's komen uit een kansverdeling met  $\mathcal{E}u = 0$  en variantie  $\sigma_u^2$ . Algemener geformuleerd:

$$x_{a \cdot j+k} = x_{a \cdot j+1} = u_{j+1} \quad \text{met } j = 0, 1, 2 \dots; k = 1, 2 \dots a; a \text{ een } (24)$$

gegeven positief geheel getal.

Op deze reeks wordt een toevalsvariabele  $\varepsilon$  gesuperponeerd, d.i. p.d.  $y_i = x_i + \varepsilon_i$  met inachtneming van (24) (25)

Verder zij ondersteld:  $\mathcal{E}\varepsilon_i = 0$ ;  $\sigma^2(\varepsilon_i) = \sigma_\varepsilon^2$  voor elke  $i$ ;  $\mathcal{E}u_i = 0$ ;  $\sigma^2(u_i) = \sigma_u^2$ ;  $\mathcal{E}\varepsilon_i \cdot \varepsilon_j = 0$  voor elke  $i \neq j$ ;  $\mathcal{E}\varepsilon_i \cdot u_j = 0$  voor elke  $i$  en  $j$ . (25a)

Bovendien  $\mathcal{E}u_i \cdot u_j = 0$  voor elke  $i \neq j$ , hetgeen impliceert, dat de  $u$ 's onafhankelijk van elkaar zijn. Weer zij:  $\beta = \sigma_\varepsilon : \sigma_u$ .

In het appendix proberen wij na te gaan welke de consequenties zijn van een zuivere persistentie in de  $u$ -reeks, bijv. beschreven door een autocorrelatie-coëfficiënt  $\omega_k = \omega^k$ , d.w.z.  $\mathcal{E}u_i \cdot u_{i+k} = \omega^k \sigma_u^2$ .

N.B. Voor  $a = 1$  bestaat iedere  $y_i$  uit de som van twee onafhankelijke stochastische variabelen. Dan volgt  $y_i$  een kansverdeling met gemiddelde nul (omdat  $\mathcal{E}u = \mathcal{E}\varepsilon = 0$ ) en met variantie  $\sigma_u^2 + \sigma_\varepsilon^2 = (1 + \beta^2)\sigma_u^2$ , welke ook deze verdelingen mogen zijn. Men mag model A echter niet zien als model B voor  $a = 1$ .

### 3.2 Autocorrelatie-coëfficiënten

Gesteld  $a \geq 2$  (voor  $a = 1$  is alles triviaal). Per definitie is

$$\rho_k = \frac{\varepsilon_{y_i} \cdot y_{i+k}}{\sigma_y^2}; \quad \text{N.B. } \varepsilon_{y_i} = \varepsilon_{y_{i+k}} = 0$$

De teller staat voor:  $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_1^N (x_i + \varepsilon_i)(x_{i+k} + \varepsilon_{i+k}) =$

$$\lim_{N \rightarrow \infty} \frac{1}{N} [ \underbrace{x_1 x_{1+k} + x_2 x_{2+k} + \dots + \dots + \dots}_{\substack{u_1^2 + u_1^2 + \dots \\ \text{a-k termen}}} + \underbrace{\dots + \dots + \dots}_{\substack{u_2^2 + u_2^2 + \dots \\ \text{a-k termen}}} ] = (a-k) \frac{\sigma_u^2}{a} .$$

De noemer  $\sigma_y^2 = \sigma_u^2 + \sigma_\varepsilon^2$ . Bijgevolg

$$\rho_k = \frac{(a-k) \sigma_u^2}{a(\sigma_u^2 + \sigma_\varepsilon^2)} = \frac{a-k}{a(1 + \beta^2)} . \quad (26)$$

Als  $k = 1$  is  $\rho_1 = \frac{a-1}{a(1 + \beta^2)}$ , voortaan  $\rho$  genoemd. Derhalve

$$\boxed{\rho_k = \frac{a-k}{a-1} \rho} ; \quad k = 1, 2 \dots a-1 \quad \boxed{\rho_k = 0} \quad \text{als } k \geq a . \quad (27)$$

Thans dalen de autocorrelatie-coëfficiënten  $\rho_k$  met toenemende  $k$  volgens een rekenkundige reeks (zulke noemt v.d Bijl, louter bij wijze van voorbeeld, op p. 107 van zijn proefschrift, zonder te vertellen bij welke persistentiemodellen de autocorrelatie-coëfficiënten zich aldus kunnen gedragen).

De persistentie is begrensd. Het aantal ongelijk nul zijnde autocorrelatie-coëfficiënten wordt door het herhalingsgetal  $a$  bepaald. Het is het aspect der herhaling, hetwelk de autocorrelatie bewerkstelligt.

### 3.3 Restvarianties in lineaire relaties in de y-reeks

De lineaire relatie tussen  $y_i$  en de  $k^{\text{de}}$  voorafgaande term  $y_{i-k}$  kan beschreven worden door:

$$y_i = \rho_k y_{i-k} + v_i \quad \text{met } \rho_k, \text{ zie (27).}$$

De restvariantie  $\sigma^2(\nu)$  is als

$$1 < k \leq a-1 \quad \sigma_{\nu}^2 = (1-\rho_k^2) \sigma_y^2 = \frac{a(a-1)}{a(1-\rho) - 1} \left\{ 1 - \left( \frac{a-k}{a-1} \right)^2 \rho^2 \right\} \sigma_{\varepsilon}^2 \quad (28)$$

en als

$$k \geq a \quad \sigma_{\nu}^2 = \frac{a(a-1)}{a(1-\rho) - 1} \sigma_{\varepsilon}^2 \quad (29)$$

Enig commentaar:

Met toenemende  $k$  (afstand tussen de twee  $y$ 's) neemt  $\sigma_{\nu}^2$  toe tot die waarde, d.i.  $\sigma_y^2 = \frac{a(a-1)}{a(1-\rho) - 1} \sigma_{\varepsilon}^2$ , die bereikt wordt als de  $y$ 's zover van elkaar liggen (d.i.  $k \geq a$ ), dat ze niet meer afhankelijk zijn. Met vaste  $k$  en afnemende  $\rho$  ( $\rightarrow 0$ ), neemt de restvariantie toe tot  $a \sigma_{\varepsilon}^2$ . De  $\rho$  neemt af, bij gegeven  $a$ , als  $\beta$  toeneemt, d.w.z. naarmate meer en meer de  $\sigma_u$  door  $\sigma_{\varepsilon}$  wordt overtroffen, d.w.z. naarmate de spreiding in de gesuperponeerde toevalsvariable groter is dan die der grond (d.i.  $u$ -)waarden.

### 3.4 Effectieve aantallen

Beschouw weer het gemiddelde van een lopend groepje van de successieve  $y$ 's. Het universum dezer gemiddelden heeft een variantie  $\sigma_y^2/n_e$ , met  $n_e = n/\omega$  en  $\omega$ , zie (17).

Substitueer (27) in  $\omega$  en er komt:

$$\text{als } n < a \quad \omega = 1 + (n-1)\rho - \frac{(n-1)(n-2)}{3(a-1)} \rho^2 > 1 \quad \text{en} \quad (30)$$

$$\text{als } n > a \quad \omega = 1 + a\rho - \frac{a(a+1)\rho}{3n} > 1 \quad \text{. Voor } n = a \text{ zijn de } \omega\text{'s} \quad (31)$$

gelijk.

De  $\omega$  is thans gekarakteriseerd door 3 grootheden:  $n$ ,  $\rho$  en  $a$ .

(in model A door 2 grootheden:  $n$  en  $\rho$ ).

Hoe groter  $n$  t.o.v.  $a$ , hoe meer nadert  $\omega$ , groeiende, tot  $1+a\rho$  en  $n_e$  tot  $n_e^* = \frac{n}{1+a\rho} < n$ .

Hoe groter  $\rho$  en hoe groter  $a$ , hoe kleiner  $n_e^*$ .

Laat ons hier, ten aanzien van het facet der effectieve aantallen, model A met model B vergelijken.

Noem  $n_A = n \frac{1-\rho}{1+\rho}$  en  $n_B = \frac{n}{1+a\rho}$ , d.z. de effectieve aantallen voor grote n. (32)

Duidelijk is  $n_B < n_A$ , zodra  $\rho < 1 - \frac{2}{a}$  (model B eist reeds  $\rho < 1 - \frac{1}{a}$ , zie (26)).

Ter toelichting:

als  $a = 2$  dan is  $n_B < n_A$  als  $\rho < 1/3 = 0.33$  of  $\beta > \frac{1}{2}\sqrt{2} = \sqrt{\frac{1}{2}}$   
als  $a = 3$  " "  $< 2/4 = 0.50$  of  $\beta > \frac{1}{3}\sqrt{3} = \sqrt{\frac{1}{3}}$   
als  $a = 4$  " "  $< 3/5 = 0.60$  of  $\beta > \frac{1}{4}\sqrt{4} = \sqrt{\frac{1}{4}}$

enz.

Daarentegen is  $n_B > n_A$  zodra  $1 - \frac{2}{a} < \rho < 1 - \frac{1}{a}$ .

Het blijkt, dat  $n_A$  en  $n_B$  veel kunnen verschillen al naar gelang model A danwel model B gekozen wordt (als het materiaal  $\rho$  levert).  
Neem bijv.  $n = 30$ ; een maand dagelijkse maximumtemperaturen, die zwaar gepersisteerd zijn. Met welk effectieve aantal is 30 equivalent? Gesteld  $\rho = 0.7$  en  $a = 9$  (zie tabel 3), dan is, met de benadering werkende,  $n_A = 30 \frac{0.7}{1.7} = 5.3$  en  $n_B = \frac{30}{1 + 9 \times 0.7} = 18.4 \gg 5.3$ . Kennis van deze effectieve aantallen is vereist, wanneer wij willen onderzoeken of de variantie der dagelijkse maximumtemperaturen van bijv. juli significant verschilt van die in november of wanneer wij willen weten of er tussen twee stations voor de maand juli verschil is t.a.v. de varianties  $\sigma_\Delta^2$ .

### 3.5 De reeks van de eerste orde differenties $\Delta$

Zij weer  $\Delta_i = y_i - y_{i+1}$ . Beschouw de reeks  $\Delta_1, \Delta_2, \dots$ .  
Gezocht de autocorrelatie-coëfficiënt  $\gamma_k$  in deze  $\Delta$ -reeks. Algemeen geldt weer

$$\gamma_k = \frac{-\rho_{k-1} + 2\rho_k - \rho_{k+1}}{2(1-\rho)}$$

Substitueer  $\rho_k$  uit (27) hierin. Er komt (als  $a \geq 2$ ):

$$\gamma_1 = - \frac{\beta^2}{2(\beta^2 + \frac{1}{a})} = - \frac{a-1-\rho}{2(a-1)(1-\rho)} ; \quad (33)$$

$$\gamma_a = \gamma_3 = \dots = \gamma_{a-1} \equiv 0$$

$$\gamma_a = - \frac{\rho}{2(a-1)(1-\rho)}$$

$$\gamma_k \equiv 0 \text{ als } k \geq a-1$$

alle zijn  $\leq 0$

Men vindt deze formules gemakkelijk door ze successievelijk te berekenen voor  $a = 2, 3, 4 \dots$

Model B leidt derhalve tot  $\gamma_1 \neq 0, \gamma_a \neq 0$ , terwijl de  $\gamma$ 's tussen deze twee en die voorbij  $\gamma_a$  identiek nul zijn. Voorts is zeker niet per se  $|\gamma_4| < |\gamma_1|$ . Het teken  $<$  geldt als  $\rho < \frac{a-1}{a+1}$ , d.i. als  $\beta > 1/\sqrt{a}$ .

Voorbeeld:

$$a = 4; \beta = 0.1; \rho_1 = \frac{3}{4(1 + \beta^2)} = 0.744; \text{ hier } \beta < 1/\sqrt{a}.$$

Tabel 1

k =	1	2	3	4	5	6
$\rho \approx$	0.75	0.50	0.25	0	0	enz.
$\gamma \approx$	-0.02	0	0	-0.50	0	enz.

Hoe groter het repetitiegetal  $a$ , d.i. hoe kleiner  $\frac{1}{a}$  t.o.v.  $\beta^2$ , hoe dichter ligt de altijd negatieve  $\gamma_1$  bij  $-\frac{1}{2}$ , welke waarde exact bereikt wordt wanneer de initiaalreeks vrij van autocorrelatie is. Deze uitkomst is zeer plausibel, want de initiaalreeks bestaat dan uit zeer lange stukken, in elk waarvan op een constante "ondergrond" een toevalsvariabele gesuperponeerd wordt.

Ook nu vragen wij naar het effectieve aantal  $n_e$ , behorende bij een lopende groep van  $n$  successieve  $\Delta$ 's.

$n_e = n/\omega$ . Substitueer (33) in (23). Er komt:



als  $a \leq n$   $\omega = \frac{1}{n(1-\rho)}$  , met  $n_e = n^2(1-\rho)$  (bevat a niet) (34)

als  $a \geq n$   $\omega = \frac{1 - \rho \frac{a-n}{a-1}}{n(1-\rho)}$  , met  $n_e = n^2 \frac{1-\rho}{1 - \rho \frac{a-n}{a-1}} > n^2(1-\rho)$  . (35)

I Voor  $n \geq a$  is in model B  $(n_e)_B = (1-\rho)n^2$  en in model A:  $(n_e)_A = \frac{1-\rho}{1-\rho^A} n^2$ . Altijd  $(n_e)_A > (n_e)_B$ .

Het verschil wordt kleiner met kleinere  $\rho$  en grotere  $n$ .

II Voor  $n < a$  is in model B  $n_e = \frac{1-\rho}{1 - \rho \frac{a-n}{a-1}} n^2$  en in model A:

$\frac{1-\rho}{1-\rho^n} n^2$ .

Nu  $(n_e)_A > (n_e)_B$  als  $\rho > \sqrt[n-1]{\frac{a-n}{a-1}}$  en  $(n_e)_A < (n_e)_B$  als  $\rho < \sqrt[n-1]{\frac{a-n}{a-1}}$ .

Getallenvoorbeeld.

Tabel 2

n = 10; $\rho = 0.8$				
model A	model B		$\sqrt[n-1]{\frac{a-n}{a-1}}$	
	a = 12	a = 20	a = 12	a = 20
nominale n = 10	10	10	0.828	0.934
effectieve $n_e = 223$	234	556		
nominale n = 15	15	15	-	0.909
effectieve $n_e = 47$	45	45		

4. Toepassingen

4.1 Het kiezen tussen de twee modellen A en B d.m.v. de F-toets

Gesteld met heeft de beschikking over  $\nu + 1$  groepen van  $n$  succesieve  $y$ -waarden. De gemiddelden daarvan zijn  $\bar{y}_1^n, \bar{y}_2^n \dots \bar{y}_{\nu+1}^n$ . Gesteld deze gemiddelden zijn onafhankelijk ( de groepen mogen derhalve geen overlapping hebben). De variantie der  $\bar{y}^n$ -waarden zij  $s^2(\bar{y}^n)$ . Het univarium der  $\bar{y}^n$ -waarden:  $\sigma^2(\bar{y}^n)$ , zie (16). De  $n_e$  daarin kan berekend worden òf via model A òf via model B, door de reeds bekend onderstelde

$\rho$ -waarden te substitueren òf in (17) òf in (30), (31). Er is een stelling: als  $z$  normaal verdeeld is rondom een gemiddelde  $\mu_z$  met variantie  $\sigma_z^2$ , dan heeft de uitspraak, dat de variantie  $s_z^2$  in een steekproef van  $\nu + 1$  elementen, aselekt gekozen uit deze populatie, gelegen is tussen  $(\chi_1^2/\nu) \sigma_z^2$  en  $(\chi_2^2/\nu) \sigma_z^2$  en betrouwbaarheid 0.95, indien  $\chi_1^2$  (resp.  $\chi_2^2$ ) de met een kans 0.025 (resp. 0.975) in de  $\chi^2$ -verdeling op  $\nu$  graden van vrijheid overschreden waarde voorstelt. Wij kunnen ook schrijven

$$P \left[ \chi_1^2/\nu < F < \chi_2^2/\nu \mid \mu_z; \sigma_z \right] = 0.95, \text{ als } F = s^2/\sigma^2. \quad (36)$$

Zo beschouwde Sneyers  $\nu + 1 = 30$  maanden januari. Elke maand bevat  $n = 31$  daggemiddelden van de temperatuur (door hem gedefinieerd als het gemiddelde van maximum en minimum), die een gemiddelde  $\bar{y}_i$  leveren ( $i = 1, 2, \dots, 30$ ). De variantie zij  $s^2(\bar{y})$ . De maandgemiddelden  $\bar{y}$  zijn zeker onafhankelijk, aangezien een invloed van januari op januari uitgesloten of zeer onbelangrijk geacht mag worden. Met kennis van  $\rho$  (en  $a$ ) is  $\sigma^2(\bar{y})$  te berekenen (wanneer men de variantie in de  $30 \times 31$   $y$ -waarden als  $\sigma_y^2$  mag opvatten), voor model A zowel als voor model B. Aldus is ook  $F_A$  en  $F_B$  bekend;  $\nu$  is hier 29. Het bovengenoemde betrouwbaarheidsinterval is dan eveneens bekend. Het bleek, dat zowel  $F_A$  als  $F_B$  erin lag. Conclusie: deze beschouwingswijze leidt tot verwerping, nòch van model A, nòch van model B.

(N.B. een bezwaar vind ik hier, dat Sneyers  $\rho$  en  $a$  uit hetzelfde materiaal berekende als waarvoor hij het bovenstaande met  $F$  verricht; bovendien zijn deze  $\rho$  en  $a$  slechts schattingen).

#### 4.2 Het kiezen tussen de twee modellen via de berekening van het repetitiegetal $a$

Er staan verschillende wegen open om tot  $a$  te komen. Wij behandelen hier een drietal daarvan.

##### 4.2.1 Via het equivalente herhalingsgetal $\omega$

Formule (16) levert  $\omega = \sigma_y^2 / n \cdot \sigma^2(\bar{y})$ .

Het materiaal levert  $\sigma_y^2$  en  $\sigma^2(\bar{y})$  (eigenlijk  $s^2$ -waarden).

Hetzelfde materiaal levert ook  $\rho$ , welke gesubstitueerd kan worden in (30), (31). Het resultaat is dan een vierkantsvergelijking in  $a$ . Aldus ging Sneyers te werk voor de gemiddelde dagtemperaturen  $y$  te Ukkel, 1901-1930. Zie de  $a$ -waarden in rij 2 van tabel 3.

Tabel 3

		jan	feb	mrt	apr	mei	jun	jul	aug	sep	okt	nov	dec
1	$\rho$	0.85	0.89	0.78	0.73	0.76	0.67	0.71	0.66	0.75	0.80	0.84	0.85
2	$a$	10.1	13.9	7.7	6.5	3.8	7.2	14.0	9.2	6.6	11.1	7.0	8.4
3	$a'$	10.8	10.6	9.4	6.0	5.9	6.8	8.6	8.4	9.0	8.3	8.8	8.5
4	$\frac{a'-1}{a'}$	0.01	0.90	0.89	0.83	0.83	0.85	0.88	0.88	0.89	0.88	0.89	0.89
5	$a^*$	10	16	11	5	5	7	14	7	10	7	6	11
6	$\frac{a^*-1}{a^*}$	0.90	0.94	0.91	0.80	0.80	0.86	0.93	0.86	0.90	0.86	0.83	0.91

Model B moet al direct verworpen worden als mocht blijken, dat  $\rho > \frac{a-1}{a}$  is (zie 27). De twaalf  $a$ -waarden (rij 2) spreiden grillig, daarom vereffende Sneyers met  $a_i' = \frac{1}{3}(a_{i-1} + a_i + a_{i+1})$ . Het blijkt daarna dat iedere  $\rho < \frac{a'-1}{a'}$  is, zodat model B mogelijk is.

#### 4.2.2 De toets van Wald-Wolfowitz

Wil men parameter vrij handelen, dan ligt de toets van Wald-Wolfowitz voor de hand. Hiertoe moet elke  $y$  door een + of - teken vervangen worden, want de toets beschouwt runs (iteraties, reeksen) van plussen of minnen. Het ligt voor de hand  $y +$  te noemen als  $y$  boven het veeljarige datumgemiddelde gelegen is en anders -. Beschouw bijv. alle  $10 \times 31$  dagelijkse maximumtemperaturen  $y$  der maanden januari uit de periode 1948-1957 (Sneyers). Na ze door tekens vervangen te hebben worden zij samengenomen in een tabel.

Tabel 4

$y_1$	$y_{k+1}$	$y_{k+2}$	$y_{k+3}$	.....
$y_2$	$y_{k+2}$	$y_{k+3}$	$y_{k+4}$	.....
$\vdots$				
$\vdots$				
$y_k$	$y_{2k}$	$y_{3k}$	$y_{4k}$	.....

Met de runtoets van W.W. wordt onderzocht of the nulhypothese  $H_0$ , dat deze  $k$  rijen gelijktijdig regelloos zijn, al of niet verworpen moet worden. Men begint bijv. met  $k = 3$ . Wanneer  $H_0$  verworpen moet

worden, wordt overgegaan op  $k = 4$ . Mocht  $H_0$  opnieuw verworpen worden, dan naar  $k = 5$  enz., tot de kleinste  $k$ , zegge  $m$ , waarbij  $H_0$  niet meer verworpen moet worden. Deze  $m$  stelt dan  $a$  voor. Zie de  $a^*$ -waarden in tabel 3. Ook nu blijkt in elk der 12 maanden, dat  $\rho < \frac{a^* - 1}{a^*}$ , zodat model B mogelijk is.

N.B. deze toets is vrij grof, zo werpt men veel informatie weg door van de waarden der  $y$ 's op hun tekens (t.o.v.  $\bar{y}$ ) over te gaan. Bovendien lijkt mij de procedure om tot  $m$  te komen verre van fraai. De grilligheid van de jaarlijkse gang van  $a^*$  zal wel daarmee samenhangen.

#### 4.2.3 Via de $r_k$ -waarden

Het verdient naar mijn mening de voorkeur de aan het materiaal ontleende  $r$ -waarden,  $r_1, r_2, r_3 \dots$  (schattingen van  $\rho_1, \rho_2, \dots$ ) te gebruiken om via de formule 27 tot  $a$  te komen (in een eigen studie koos ik model A en paste ik  $r_1, r_2, \dots$  aan de formule  $\rho_k = \rho^k$ , met  $\rho$  als onbekende, aan). De berekeningen zijn helaas verre van eenvoudig. Het gaat om de beste  $a$  en  $\rho$ , gedefinieerd als die waarden, welke

$$S = \sum_1^{a-1} (r_k - \rho_k)^2 \text{ minimaliseren, met } \rho_k = \rho \frac{a-k}{a-1}. \quad (37)$$

De oplossing van  $\partial S / \partial \rho = 0$  en  $\partial S / \partial a = 0$  luidt

$$\begin{cases} a \sum_1^{a-1} r_k - \sum_1^{a-1} k r_k = \frac{1}{6} a \rho (2a-1) \text{ en} \\ \sum_2^{a-1} (k-1) r_k = \frac{1}{3} \rho a (a-2) . \end{cases}$$

$$\text{Invoerende } S_a \equiv \sum_1^{a-1} r_k \text{ en } S_a^* \equiv \sum_1^{a-1} k r_k \text{ staat er ook} \quad (38)$$

$$\left. \begin{cases} a S_a - S_a^* = \frac{1}{6} a \rho (2a-1) \\ S_a^* - S_a = \frac{1}{3} a \rho (a-2) \end{cases} \right\} \text{ onbekenden: } a \text{ en } \rho$$

$$\text{De eenvoudigste oplossing lijkt me } \frac{S_a^* - S_a}{a S_a - S_a^*} = 2 \frac{a-2}{2a-1} \quad (38a)$$

Men probeer verscheidene  $a$ -waarden, berekene bij elke  $a$ -waarde  $S_a$  en  $S_a^*$ , alsmede de breuk  $\frac{S_a^* - S_a}{a S_a - S_a^*}$  en  $2(a-2) : (2a-1)$ , aldus approximatief tot de gezochte  $a$  komende.

Mocht  $a$  groot zijn, zegge 10 of meer, dan zijn blijkbaar ook  $r_k$ 's met  $k \leq 10$  nodig. Als de vergelijkingen tot negatieve of imaginaire  $a$ 's leiden, mag men model B voor onmogelijk houden.

#### 4.3 Schatting van de autocorrelatie-coëfficiënten $\rho_k$

Deze schatting maakte Sneyers precies zoals wij het deden; bij ons onderzoek van de Interdiurne Variabiliteit, zie W.R. 60-2 § 5,

via  $\rho_k = 1 - \frac{1}{2} \frac{\sigma^2(k\Delta)}{2}$  als  $k\Delta_i = t_i - t_{i-k} = k^{\text{de}}$  orde differentie.

#### 4.4 Het kiezen tussen de twee modellen A en B via de autocorrelatie-coëfficiënt $\gamma_2$ (orde 2) in de reeks der eerste orde differenties

In § 3.5 werd bewezen, dat voor elk model B met  $a \geq 2$  in de  $\Delta$ -reeks de  $\gamma_2 \equiv 0$ , zie (33), terwijl, zie (21), in model A de  $\gamma_2$  slechts dan nul is, als de initiaalreeks der  $y$ 's autocorrelatie afwezig is. Nu is de  $y$ -reeks ongetwijfeld niet-regelloos en dus zou een  $\gamma_2 = 0$  slechts te rijmen zijn met model B, als men althans onderstellen mag, dat het repetitiegetal  $\geq 2$  (het heeft trouwens weinig zin van een repetitiegetal 1 te spreken; bovendien zou in het geval  $a = 1$  de  $y$ -reeks regelloos zijn). Het heeft derhalve zin om  $c_2$ , schatting van  $\gamma_2$ , te berekenen in de  $\Delta$ -reeks en dan te toetsen of deze  $c_2$  significant van nul verschilt. Sneyers verreed deze berekening door weer slechts de tekens der  $\Delta$ 's te beschouwen ( $\bar{\Delta}$  is nu nul) en de Wald-Wolfowitz-toets toe te passen op de reeksen  $\Delta_1, \Delta_3, \Delta_5, \dots$  en  $\Delta_2, \Delta_4, \dots$ . De conclusie was, dat model B moest worden verworpen. Kiezende slechts tussen de modellen A en B, verkoos Sneyers tenslotte het eerste, beheerst door  $\rho_k = \rho^k$ , waarmee hij een bewijs leverde voor de juistheid van het vermoeden, dat wij in ons onderzoek m.b.t. de interdiurne variabiliteit (W.R. 60-2) en daarvoor m.b.t. een persistentie-onderzoek inzake de dagelijkse neerslagsommen (W.R. 55-004) slechts intuïtief hadden. Vanzelfsprekend is hiermede nog niet gezegd, dat nu model A het beste is onder de vele modellen, die ook nog voor verificatie in aanmerking zouden kunnen komen. Doch model A is zo eenvoudig dat het om deze reden waard is hiermede voortaan te werken.

5. Persistentie-vrije reeksen

5.1 Hier volgt de zeer algemeen gehouden definitie van het begrip persistentie, ontleend aan de Vocabulaire Météorologique Internationale, Paris 1957 (Edition provisoire), uitgegeven door de W.M.O.-Commission de Bibliographie et des Publications. Onder Tendence à la persistence du temps - Erhaltungstendenz des Wetters - Tendency for persistence of weather wordt verstaan: Phénomène constitué par le fait que les valeurs d'un grand nombre d'éléments ou/et de conditions météorologiques considérés comme séries statistiques temporelles, montrent une corrélation positive entre les valeurs successives.

5.2 Het begrip persistentie-vrij vereist een nauwkeuriger definitie: Een reeks zou persistentie-vrij genoemd kunnen worden, wanneer de kans, dat het volgende element in de reeks één van een aantal elkaar onderling uitsluitende alternatieven zal zijn (in het geval er sprake is van een discontinu verdeelde grootheid), of een gegeven waarde zal bereiken of overschrijden (bij een continue verdeelde variabele), in generlij opzicht afhankelijk is van de door voorafgaande elementen aangenomen waarden.

5.3 Een oneindige reeks van een eindig aantal karakteristieken heet een "regelloos kollektief" (irregular collectix), random-reeks, als aan de volgende wetten gehoorzaamd wordt:

- a) de relatieve frequentie van een gegeven karakteristiek in een n-tal aansluitende termen tendeert naar een limiet bij  $n \rightarrow \infty$ , welke limiet niet afhankelijk van de ligging van de eerste term van deze n termen;
- b) iedere oneindige onderreeks bezit dezelfde eigenschap als de moederreeks.

Toelichting: een onderreeks  $y_0, y_1, y_2 \dots$  bestaat uit termen, die volgens een gegeven voorschrift uit moederreeks  $x_0, x_1, x_2 \dots$  gekozen zijn.

- Bijv.: a)  $y_k = x_{n+j_k}$ , met  $j_k = \frac{1}{2}k(1+k)$ ;  $k = 0, 1, 2 \dots$   
d.i.  $x_n, x_{n+1}, x_{n+3}, x_{n+6}, x_{n+10}, x_{n+15}, x_{n+21}$  enz.
- b)  $y_k = x_{kn}$ , dus  $x_0, x_{2n}, x_{3n}$  enz.
- c)  $y_k = x_{n^k}$ , dus  $x_1, x_n, x_{n^2}, x_{n^3}$  enz.

Enkele voorbeelden.

- a) Ik werp met 10 zuivere dobbelstenen tegelijk en tel bij elke worp het totale aantal bovenliggende ogen. De som moge  $y$  heten. Aldus ontstaat de reeks  $y_1, y_2, y_3 \dots$ . De 51 verschillende  $y$ -waarden,  $10 \leq y \leq 60$  ( $y$  geheel), zijn niet gelijkwaarschijnlijk, en op iedere plaats  $m$  in de reeks is de kans op een gegeven  $y$ -waarde dezelfde d.i. van  $m$  onafhankelijk.
- b) De volgende reeks van + en - tekens is beslist niet persistentie-vrij -+-+-----+++++-----+ enz. Gestart werd met een -; gevolgd door 2 plussen; gevolgd door 1+2=3 minnen; gevolgd door 1+2+3=6 plussen; daarna 1+2+3+6=12 minnen; daarna 1+2+3+6+12=24 plussen enz.
- c) De volgende reeks samengesteld uit de cijfers 0, 1, 2 .... 9 is weer wel volkomen regelloos:  
4 0 4 9 5 4 9 7 1 1 4 8 6 5 2 6 1 1 4 1 enz.

$y_1, y_2,$

$y_{20}$

Het zijn de cijfers, die boven lagen bij het werpen met een 10 kantige dobbelsteen (cijfers 0, 1, 2 .... 9 op de zijvlakken).

Men merke het verschil tussen de b)- en c)-reeks; beide zijn persistentie-vrij = regelloos, doch in b) weet ik te voren het cijfer op plaats  $i$ , maar in c) kan dat 0, 1, 2 .... 9 zijn, met kansen  $1/10$ .

6. Stationaire reeksen

In het genoemde woordenboek lezen wij:

"A stochastic process  $x_t$  is said to be strictly stationary if the multivariate distribution of  $x_{t_1+h}, x_{t_2+h} \dots x_{t_n+h}$  is independent of  $h$  for any finite set of parameters  $t_1+h, \dots, t_n+h, t_1, t_2 \dots t_n$ .

The process is said to be stationary in the wide sense if the mean and the variance are independent of  $t$ ".

De definitie is nogal ingewikkeld. Ter toelichting: zij  $h_i$  de totale hoeveelheid neerslag op station  $S$  in het jaar  $i$ . De reeks  $h_1, h_2 \dots$  heet stationair als bijv. de kans, dat van een viertal jaren de jaarsommen gelegen zullen zijn in de intervallen 50 à 100, 0 à 25, 300 à 500, > 800 mm, voor ieder viertal jaren, hoe ook gelegen (al of niet successief) dezelfde is en als dit eveneens het geval is voor

ieder n-tal jaren en alle combinaties van intervallen, anders gezegd de kans op een willekeurige voorgeschreven samenstelling van een niet noodzakelijk uit direct aansluitende termen bestaande greep is, ongeacht de grootte van de greep, onafhankelijk van de plaats in de reeks.

Wij simplificeren deze definitie doorgaans tot: de verdelingsdichtheid moet onafhankelijk van de plaats in de reeks zijn, d.i.

$$F[\underline{x}_i = a] \equiv F[\underline{x}_j = a] \text{ voor elke } i, j, a.$$

Is de y-reeks, model A, stationair?

Er kwam  $y_N = \rho^N y_0 + \rho^{N-1} \varepsilon_1 + \dots + \rho \varepsilon_{N-1} + \varepsilon_N$ . Houd de N vast en herhaal het voorgeschreven procédé vele keren, steeds met dezelfde  $y_0$  beginnende (bijv. nul). Dan volgt  $y_N$  een kansverdeling en wel een normale als  $\varepsilon$  een normale verdeling volgt en zo het laatste niet het geval is, dan toch in goede benadering, welke benadering des te beter is naarmate N groter is. De  $y_N$ -verdeling heeft een gemiddelde  $E y_N = \rho^N y_0$ , want elke  $E \varepsilon_i = 0$  en een variantie  $\sigma^2(y_N) = (1 + \rho + \rho^2 + \dots + \rho^{N-1}) \rho \sigma_\varepsilon^2 = \frac{1 - \rho^{N-2}}{1 - \rho} \sigma_\varepsilon^2$ , want  $\sigma^2(\varepsilon_i) = \sigma_\varepsilon^2$ , terwijl ook nog (3) geldt. Dus  $\sigma^2(y_N) \rightarrow \frac{\sigma_\varepsilon^2}{1 - \rho}$  als  $N \rightarrow \infty$ . Al heel gauw is de reeks ongeveer stationair.

-----



A P P E N D I X

A. Model A met "niet zuivere persistentie"

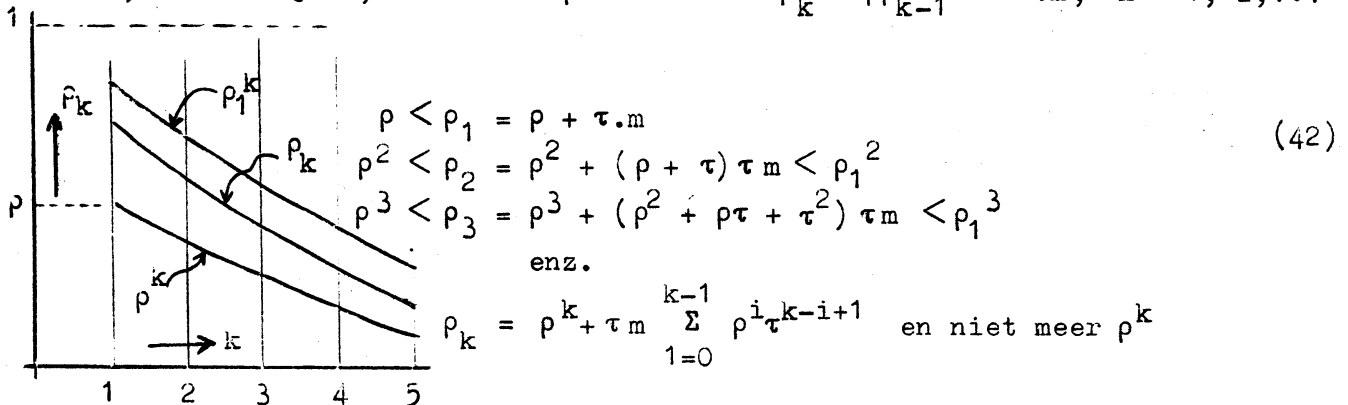
Formule (11) is het resultaat o.m. van onderstelling (4). Wij willen deze onderstelling laten vallen en vervangen door

$$\varepsilon y_i \cdot \varepsilon_{i+k} = \tau^k \sigma_y \sigma_\varepsilon, \text{ met } 0 < \tau < 1; k \rightarrow \infty, \tau^k \rightarrow 0. \quad (40)$$

Er komt

$$\rho_k = \frac{\varepsilon y_i \cdot y_{i+k}}{\sigma_y^2} = \rho \rho_{k-1} + \tau^k \frac{\sigma_\varepsilon}{\sigma_y}. \quad (41)$$

Thans is  $\sigma_\varepsilon / \sigma_y \neq \sqrt{1-\rho^2}$ , want uit  $y_i = \rho y_{i-1} + \varepsilon_i$  volgt nu  $\sigma^2(y_i) = \rho^2 \sigma^2(y_{i-1}) + \sigma_\varepsilon^2 - 2\rho \cdot \tau \cdot \sigma_y \cdot \sigma_\varepsilon$ . Noem  $m = \sigma_\varepsilon / \sigma_y$ , dan is  $m^2 - 2\rho\tau m - (1-\rho^2) = 0$ , waaruit  $m = \rho\tau \pm \sqrt{\rho^2\tau^2 + (1-\rho^2)}$ . Alleen de + wortel heeft zin, deze hete verder m. Zodra  $\tau = 0$ , is  $m = \sqrt{1-\rho^2}$  maar nu,  $\tau > 0$  zijnde, is  $m > \sqrt{1-\rho^2}$ . Er komt:  $\rho_k = \rho \rho_{k-1} + \tau^k \cdot m$ ;  $k = 1, 2, \dots$



In het bovenstaande werd  $\tau$  positief gedacht, omdat dit ons het meest plausibel leek indien er althans een  $\tau$  zou bestaan. Bij beschouwing van de realiteit van een  $\tau$ , en in het bijzonder wat het teken ervan betreft, kunnen fysische overwegingen een indicatie geven. Wanneer bijv. achter de  $\rho$  een opeenvolging van weertypen (luchtmassa's) gedacht zou mogen worden, dan zou de  $\tau$  te doen kunnen hebben met de overgang van de ene weertypenserie op de andere, welke vermoedelijk niet random zal plaatsvinden. Het is niet de bedoeling in dit rapport nader op deze kwestie in te gaan.

B. Model B met persistentie in de u's

Wij laten nu in (2) de <sup>onderstelling</sup>  $\epsilon u_i \dots u_j = 0, i \neq j$ , vallen en vervangen haar door  $\epsilon u_i \cdot u_{i+k} = \omega_k \sigma_u^2 = \omega^k \sigma_u^2$ , met  $0 < \omega < 1$ . Weer vragen wij naar de  $\rho_k$  in de y-reeks (25).

De zaak is veel ingewikkelder geworden. Hier volgen enkele gevallen, geheel doorgerekend, alle met  $a = 5$ .

$k = 3$ , d.i.  $k < a$ .

$$u_1 u_1 u_1 u_1 u_1 \left| u_2 u_2 u_2 u_2 u_2 \right| u_3 u_3 u_3 u_3 u_3 \quad \text{enz.}$$

$$+ \epsilon_1 \epsilon_2 \epsilon_3 \epsilon_4 \epsilon_5 \left| \epsilon_6 \epsilon_7 \epsilon_8 \epsilon_9 \epsilon_{10} \right| \quad \text{enz.}$$

Algemeen:

$$\rho_k = \frac{\epsilon y_i \cdot y_{i+k}}{\sigma_y^2} \quad \text{In de teller staat:}$$

$$\epsilon y_i \cdot y_{i+3} = \frac{1}{N} \sum_{N \rightarrow \infty} y_i y_{i+3} = \frac{1}{N} [(u_1 + \epsilon_1)(u_1 + \epsilon_2) + (u_1 + \epsilon_2)(u_1 + \epsilon_5) + (u_1 + \epsilon_3)(u_2 + \epsilon_6) + (u_1 + \epsilon_4)(u_2 + \epsilon_7) + (u_1 + \epsilon_5)(u_2 + \epsilon_8) + \dots \text{enz.}] = \frac{1}{N} [2(u_1^2 + u_2^2 + \dots) + 3(u_1 u_2 + u_2 u_3 + \dots) + u \cdot \epsilon \text{-termen}]; \rho_k \frac{2\sigma_u^2 + 3\omega\sigma_u^2}{5(\sigma_u^2 + \sigma_\epsilon^2)}$$

Algemeen:  $\frac{(a-k)\sigma_u^2 + k\omega\sigma_u^2}{a(\sigma_u^2 + \sigma_\epsilon^2)} = \frac{a - k(1-\omega)}{a - (1-\omega)}$ , mits  $1 \leq k < a$

$k = 5$  (nu  $k = a$ )

Teller in  $\rho_5$  wordt  $\frac{1}{N} (u_1 + \epsilon_1)(u_2 + \epsilon_6) + (u_1 + \epsilon_2)(u_2 + \epsilon_7) + \dots = \frac{5u_1 u_2 + 5u_2 u_3 + \dots}{N}; \rho_k \rightarrow \frac{5\omega\sigma_u^2}{5(\sigma_u^2 + \sigma_\epsilon^2)}$ , algemeen  $\frac{a\omega}{a - (1-\omega)}$

$k = 6$ , d.i.  $a < k < 2a$ . Dan bevat de teller van  $\rho_k$

$$\frac{1}{N} [(u_1 + \epsilon_1)(u_2 + \epsilon_7) + \dots] = \frac{1}{N} [4u_1 u_2 + 4u_2 u_3 + \dots]; \rho_k \frac{4\omega}{5(\sigma_u^2 + \sigma_\epsilon^2)}$$

algemeen  $\frac{(2a-k)\omega}{a - (1-\omega)}$

$k = 11$ , nu  $2a < k < 3a$ . De  $\rho_k$  heeft in de teller  $\frac{1}{N} [uu_1 u_2 u_3 + uu_2 u_3 u_4 + \dots + u \cdot \epsilon \text{ termen}]$  en  $\rho_k \frac{4\omega^2}{5(\sigma_u^2 + \sigma_\epsilon^2)}$ , algemeen  $\frac{3a - k}{a - (1-\omega)} \omega^2 \rho$  enz.

Samenvattend:

$$\begin{aligned} k < a & \quad \rho_k = \frac{a - k(1-\omega)}{a - (1-\omega)} \rho & (43) \\ a \leq k < 2a & \quad \rho_k = \frac{2a - k}{a - (1-\omega)} \omega \rho \\ 2a \leq k < 3a & \quad \rho_k = \frac{3a - k}{a - (1-\omega)} \omega^2 \rho \end{aligned}$$

enz.

Substitueert men  $\omega = 0$ , dan zien wij weer  $\rho_k = \frac{a-k}{a-1} \rho$ .

Conclusie: zodra tussen de u's (d.i. van groepje tot groepje) autocorrelatie in gevoerd wordt (fysisch lijkt me dit wenselijk) gaat de persistentie principieel onbegrensd worden.

tabel 5

		ρ-waarden														
		orde 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\omega = 0.2$		0.80	65	49	34	19	15	11	8	4	3.8	3.0	2.3	1.5	0.8	enz.
$\omega = 0$		0.80	60	40	20	0	0	0	0	0	0	0	0	0	0	enz.

Opmerking. Het geval  $a = 1$  wordt in de literatuur druk besproken. De y-reeks is:  $u_1 + \varepsilon_1$   $u_2 + \varepsilon_2$   $u_3 + \varepsilon_3$  ...., terwijl de u-reeks autocorrelatie bezit, beheerst door coëfficiënten  $\omega_k$ ,  $k = 1, 2, \dots$ . In de y-reeks is dan autocorrelatie met coëfficiënten  $\rho_k = \omega_k \cdot f$ , met  $f = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2} < 1$ . Superpositie van een toevalsvariabele op een gegeven tijdreeks (meetonnauwkeurigheden bijv.: noise, ruis, in het algemeen) doet de autocorrelatiecoëfficiënten met een factor f kleiner worden, met verdere gevolgen daarvan (zoals het moeilijker vinden van periodiciteiten, trends e.d.)

## C Kunstmatige reeksen met persistentie

Hoe kan men zich reeksen maken met een voorgeschreven persistentie? Hier volgen enige methodes.

### C1. Positieve persistentie

#### C1.1 Een reeks van twee alternatieven

Een vaas zij gevuld met N ballen; zij zijn identiek, behoudens het feit, dat er  $N_+$  het + teken en  $N_-$  het -teken dragen;  $N = N_+ + N_-$ . Trek een bal. Zij deze +. Werp nu  $\Delta + 1$  + ballen terug ( $\Delta$  bijv. 5). Schud de vaas en trek weer. Misschien is de bal weer +. Werp dan opnieuw  $\Delta + 1$  + ballen terug. Trek. Zij de bal nu -. Werp nu  $\Delta + 1$  - ballen terug, etc. Noteer de tekens  $++-+---+--+$  enz. De reeks bestaat uit twee

alternatieven. Door ons voorschrift van trekken en surplus - terugwerpen heeft de reeks autocorrelatie. Aan dit model is reeds veel gerekend door Eggenberger, Polya, Schumann e.a. Gesteld er zijn  $n$  trekkingen verricht, waarvan  $n_+$  en  $n_-$  ballen. De  $(n+1)$ ste bal kan + of - zijn (gevallen a en b). De kans  $k$ , dat de  $(n+2)$ de bal + is hangt daarmee samen. In geval a is  $k_a(+)$  =  $\frac{f_+ + (n+1)d}{1 + (n+1)d}$ , met  $f_+ = N_+/N$  en  $d = \Delta/N$  in geval b is

$$k_b(+1) = \frac{f_+ + n d}{1 + (n+1)d} .$$

Het is gemakkelijk te bewijzen, dat  $k_a > k_b$ , onverschillig  $f_+$ ,  $n_+$  en  $d$ .

Het procédé bevredigt m.i. niet, want men krijgt er niet de ware, d.i. natuurlijke persistentie mee. Immers, gesteld er zijn  $n$  trekkingen verricht, deze leveren gemiddeld  $nf_+$  plus- en  $nf_-$  minballen. Onderscheid nu twee gevallen

a) de  $n+1$ ste bal is +. Werp dus  $\Delta + 1$  plusballen terug, schud en trek de  $n+2$ de bal. De kans, dat deze weer + is, is gemiddeld  $\mathcal{E}f_a(+)$  =  $\frac{f_+ + (f_+ n+1)d}{1 + (n+1)d} > f_+$ .

b) de  $n+1$ ste bal is -. Werp dan  $\Delta+1$  minballen terug, schud en trek de  $n+2$ de bal. De kans, dat deze + is, is gemiddeld  $\mathcal{E}f_b(+)$  =  $\frac{f_+ + f_+ n.d}{1 + (n+1)d} < f_+$ .

Met  $n \rightarrow \infty$ , daalt  $\mathcal{E}f_a(+)$  naar  $f_+$  en stijgt  $\mathcal{E}f_b(+)$  naar  $f_+$ .

Dit is m.i. niet in overeenstemming met de persistentie in een natuurlijke reeks van twee alternatieven (bijv. regen-, droge dag). Waar men zich ook moge bevinden in de reeks, d.i. hoe  $n$  ook zij, overal heeft  $f_a(+|+)$  een vaste waarde, evenals bijv.  $f_b(+|-)$ . Wel heeft het model de gewenste eigenschap, dat de gemiddelde kans op een +, nadat direct al  $a$  plussen voorafgingen, met  $a$  toeneemt, zie

$$f = \frac{f_+ + (a + nf_+)d}{1 + (a+n)d} .$$

Een reeks van twee alternatieven (+, - tekens), die in dit opzicht meer bevredigt, namelijk een Markoff-ketting van het eerste type, wordt a.v. verkregen.

Er zijn gegeven twee vazen; de ene is gevuld met  $N_+$  plus- en  $N_-$  minballen, fracties  $p$  en  $q = 1-p$ ; de andere met  $N_+'$  en  $N_-'$ , fracties

$p'$  en  $q'$ . Het voorschrift luidt: start met de eerste vaas. Trek een bal, noteer het teken en werp de bal terug. Blijf uit dezelfde vaas trekken zolang er geen tekenverwisseling is, doch ga op de andere vaas over zodra het teken wel verwisselt. De reeks tekens zou kunnen zijn  $++\text{---}+ \text{---}++\text{---}++++\text{---}$  enz. Met  $x$ ' zijn de tekens gemerkt, die uit de tweede (accent) vaas stammen. Door dit procédé is de kans op een +, nadat een + voorafging, waar men zich ook moge bevinden in de reeks, altijd  $p$ ; geschreven  $P(+|+) = p$ ; die op een - na een + is  $q = 1-p$ , d.i.  $p(-|+) = q$ ; evenzo  $p(+|-) = p'$  en  $p(-|-) = q' = 1-p'$  en dat is precies wat per definitie het geval behoort te zijn in een Markoff-keten I. Talrijke statistische kwesties, zoals de kans op het verkrijgen van  $a$  plussen in een reeks van  $n$ -trekkingen of de kans op  $b$  runs van bijv. drie plussen etc. (hier niet behandeld) hangen samen met de determinantwaarde  $\Delta$  van  $\begin{vmatrix} p & q \\ p' & q' \end{vmatrix}$ . Zodra  $\Delta > 0$  is er positieve persistentie. Voor  $\Delta = 0$  is de reeks random.

### C1.2 Een reeks van waarden van één continu verdeelde grootheid

C1.2.1 Het procédé  $y_i = \rho y_{i-1} + \varepsilon_i$ , met  $0 < \rho < 1$  en bijv.  $\varepsilon_1 \equiv N(0; 1)$ . Dit werd in dit rapport uitvoerig beschreven.

C1.2.2 Laat een  $\varepsilon$ -reeks  $\varepsilon_1, \varepsilon_2, \varepsilon_3 \dots$  bestaan uit aselect uit een standaard normale verdeling  $N(0; 1)$  getrokken getallen. Laat een groepje van  $n$  successieve  $\varepsilon$ -waarden voortschuiven en beschouw het groepsgemiddelde  $\bar{\varepsilon}_k^n = \frac{1}{n} \sum_{i=k}^{k+n-1} \varepsilon_i$ ;  $k = 1, 2, 3 \dots$ . In deze  $\bar{\varepsilon}$ -reeks van  $\bar{\varepsilon}$  waarden is positieve autocorrelatie louter en alleen vanwege het feit, dat twee successieve  $\bar{\varepsilon}$ -waarden berusten op groepen, die een overlapping  $n-1$  hebben. De autocorrelatie-coëfficiënten zijn (hier niet bewezen)  $\rho_1 = \frac{n-1}{n}$ ;  $\rho_2 = \frac{n-2}{n} \dots \rho_n = \frac{1}{n}$  en  $\rho_k \equiv 0$ , voor  $k \geq n+1$ . Zij dalen als een rekenkundige reeks. De variantie van  $y$  is 1, het gemiddelde is nul, de verdeling is eveneens normaal.

C1.2.3 Het  $u, \varepsilon$ -procédé, in dit rapport als model B beschreven, levert ook als in een rekenkundige reeks dalende  $\rho_k$ 's, althans als de  $u$ 's niet gecorreleerd zijn. Zou men bijv. de  $u$ 's aselect steken uit een  $N(0; \sigma_u)$ -verdeling en de  $\varepsilon$ 's uit een  $N(0; \sigma_\varepsilon)$ -verdeling, zodat  $\beta = \sigma_\varepsilon / \sigma_u$  vastligt, en  $a = 4$  als repetitiegetal kiezen, dan heeft de  $y (= u + \varepsilon)$ -reeks autocorrelatie-coëfficiënten  $\rho_1 = \frac{3}{4(1 + \beta^2)}$ ;  $\rho_2 \equiv \frac{3}{4} \rho_1$ ;  $\rho_3 = \frac{2}{4} \rho_1$ ;  $\rho_4 = \frac{1}{4} \rho_1$ ;  $\rho_k \equiv 0$ ,  $k \geq 5$ .

Men zou voor de u-reeks kunnen nemen de onder C1.2.1 geconstrueerde y-reeks, waardoor de u-reeks positieve autocorrelatie zal hebben, beheerst door  $\rho_k = \rho^k$ . Daarna zou men weer het u- $\varepsilon$ -procédé kunnen volgen. Dan ontstaat er een reeks met onbegrensde autocorrelatie.

## C2. Negatieve persistentie<sup>1)</sup>

### C2.1 Een reeks van twee alternatieven

Zie het spel met de vaas met ballen onder C1.1. Werp nu echter de getrokken bal terug, doch voeg aan de vaas bovendien  $\Delta$  ballen toe van het teken, tegengesteld aan het getrokken. Hierdoor wordt, nadat een + bal getrokken werd, de kans op weer een + trekking verkleind. Wij werken dit hier niet verder uit.

### C2.2 Een reeks van waarden van een continu verdeelde grootheid

C2.2.1 Het procédé  $y_i = \rho y_{i-1} + \varepsilon_i$ , doch  $-1 < \rho < 0$ , met bijv.  $\varepsilon_i \equiv N(0; 1)$ . De y-reeks zal zuivere persistentie bezitten, want  $\rho_k = \rho^k$ , doch de tekens van  $\rho_1, \rho_2, \rho_3 \dots$  zijn om en om plus en min.

C2.2.2 Hoe verkrijgen wij zuivere persistentie met enkel negatieve autocorrelatie-coëfficiënten, die als een meetkundige reeks  $|\rho_k| = |\rho_1|^k$  dalen? Wij denken allicht aan de reeks der eerste orde differenties  $\Delta_i = y_i - y_{i-1}$ ,  $i = 1, 2, 3 \dots$ . Als de y-reeks regelloos is (elke  $\rho_k \equiv 0$ ), heeft de  $\Delta$ -reeks negatieve persistentie<sup>1)</sup>, beheerst door  $\gamma_1 = -\frac{1}{2}$  en  $\gamma_k \equiv 0$ ,  $k \geq 2$ . Als de y-reeks positieve autocorrelatie heeft, die  $\rho_k = \rho^k$  volgt, dan heeft de  $\Delta$ -reeks negatieve autocorrelatie, beschreven door  $\gamma_k = -\frac{1}{2} \rho^{k-1} (1-\rho)$ . Elke  $\gamma_k < 0$ ; de  $\gamma_k$ 's dalen meetkundig, doch  $|\gamma_k| \neq |\gamma_1|^k$ .

Op de gestelde vraag weten wij het antwoord (nog) niet.

---

1) Het woord deugt niet: contradictio interminis. Beter is het te spreken van positieve autocorrelatie  $\equiv$  persistentie; negatieve autocorrelatie of non-persistentie  $\equiv$  apersistentie.

SUMMARY

Two statistical models of persistency  
in climatological time series

---

0.1 In this report, written on the base of exchanges of view with my Belgian colleague Dr. Sneyers (Royal Meteorological Institute, Ukkel), the mathematical and statistical aspects of two highly different models of persistency are discussed. In model A the persistency is unlimited, that is the autocorrelation-coefficients  $\rho_k$  of orders  $k$  ( $= 1, 2, 3 \dots$ ) decrease geometrically. In model B these coefficients decrease arithmetically; consequently the persistency is limited.

2. Model A

The time series  $y_1, y_2 \dots$  is defined according to (1) under the conditions (2), (3) and (4). Only if  $0 < |f| < 1$ , then  $f \equiv \rho_1$ . In this case it is easily shown that  $\rho_k = \rho^k$ . This persistency is called "pure". Denoting  $\sigma_\varepsilon/\sigma_y$  by  $\beta$ , we obtain (14). Mind the conditions  $\varepsilon y_i = 0$  and  $\varepsilon y_i = \sigma_y^2 = \sigma_\varepsilon^2 : (1 - \rho^2)$  for each  $i$ . The  $y$ -series turns out to be asymptotically stationary (definition c). Each  $y_i$  is stochastically linearly related with each  $y_{i-k}$ , see (15). The restvariance is (15a).

Next moving groups of  $n$  successive  $y$ 's are considered, with group means  $\bar{y}^{-n}$ 's. The variance is (16), with effective number  $n_e$  and  $n_e = n : \omega$ , with  $\omega$ , see (16a). Substituting  $\rho_k = \rho^k$ ,  $\omega$  becomes (17); which approximately is  $n \frac{1 - \rho}{1 + \rho}$ , for large  $n$ , small  $\rho$ . Next the series of differences  $\Delta_i = y_i - y_{i-1}$  is considered. The autocorrelation-coefficients may be called  $\gamma_k$  ( $k = 1, 2 \dots$ ); see (20).

With  $\rho_k = \rho^k$  (pure persistency in the initial  $y$ -series) (21) follows for  $\gamma_k$ . All coefficients are negative (in case  $\rho = 0$   $\gamma_1 = -\frac{1}{2}$ ;  $\gamma_2 = \gamma_3 = \dots \equiv 0$ ); they decrease geometrically, but  $\gamma_k \neq \gamma_1^k$ .

Again a moving group of  $n$  successive values is considered, but now in the  $\Delta$  series. The effective number  $n_e$  follows (24), provided that  $\rho_k = \rho^k$ . Mind  $n < n_e < n^2$ , for  $1 > \rho > 0$ . In consequence of a negative autocorrelation the inequality  $n_e > n$  holds instead of  $n_e < n$ .

3. Model B

Consider, for instance, a series  $u_1 u_1 u_1 u_2 u_2 u_2 u_3 u_3 u_3 \dots$ . The "repetition number"  $a = 3$ . Let  $\mathcal{E}u_i = 0$  and  $\mathcal{E}u_i^2 = \sigma_u^2$ , but  $\mathcal{E}u_i u_j = 0$ .

A new series is constructed by  $y_i = x_i + \mathcal{E}_i$ , with  $x_1 = x_2 = x_3 = u_1$ ;  $x_4 = x_5 = x_6 = u_2$ , etc. See the conditions (25a).

This  $y$  series shows autocorrelation. The autocorrelation-coefficients are  $\rho_k$ , see (26), (26a), (27). They decrease arithmetically, that is, they are positive for  $k \leq a$  and zero for  $k > a$ . This persistency is limited. Mind  $\rho < 1 - \frac{1}{a}$ . It is the repetition which causes the autocorrelation. For  $a = 1$  each  $\rho_k = 0$ . Again each  $y_i$  is stochastically linearly related with each  $y_{i-k}$ ;  $y_i = \rho_k y_{i-k} + v_i$ . The restvariance  $\sigma^2(v_i)$  is given by (28) and (29).

Next the effective number in a moving group of  $n$  successive  $y$  values is calculated,  $n_e = n/\omega$ ; for  $\omega$  see (30) and (31).

For sufficiently large  $n$ , in model A  $n_e = n \frac{1 - \rho}{1 + \rho}$  and in model B  $n_e = \frac{n}{1 + a\rho}$ . Of course in both models  $n_e < n$ , but  $n_B < n_A$ , provided that  $\rho < 1 - \frac{2}{a}$  and  $n_B > n_A$ , if  $\rho > 1 - \frac{2}{a}$ . These differences may be statistically important in significance investigations.

Then again the  $\Delta$  series is considered. The autocorrelation-coefficients  $\gamma_k$  obey law (33). Mind:

$$\gamma_1 \neq \frac{1}{a}; \gamma_2 = \gamma_3 = \dots = \gamma_{a-1} \equiv 0; \gamma_a \neq 0; \gamma_{a+1} = \gamma_{a+2} = \dots \equiv 0. (a \geq 2)$$

The effective number  $n_e$ , which statistically equals the nominal number of a moving group of  $n$  successive  $\Delta$  values, is given by (34) and (35).

4. Applications

4.1 Choosing between models A and B by means of the F-test

Let  $\bar{y}_1^n, \bar{y}_2^n, \dots, \bar{y}_{\nu+1}^n$  be the independent means of  $\nu + 1$  groups of  $n$  successive  $y$  values (in the initial series). The variance in the population of  $\bar{y}$  values  $\sigma^2(\bar{y}^n)$ , see (16), is estimated as  $s^2(\bar{y}^n)$ ,

The effective number  $n_e$  can be computed in two ways, either in model A or in model B, see (17) and (30), (31), if  $\rho_k$  is known.



Denoting  $F = s^2/\sigma^2$  (for  $\bar{y}^n$ ), we obtain

$P[\chi_1^2/\nu < F < \chi_2^2/\nu] = 0.95$ , provided that  $y$  follows a normal distribution, wherein  $\chi_1(\chi_2^2)$  is exceeded with 0.025 (0.975) probability in the  $\chi^2$  distribution for  $\nu + 1$  degrees of freedom.

If  $F$  is computed in both model A and model B, it proves to lie within the mentioned reliability interval in both cases.

Consequently neither A nor B should be rejected.

#### 4.2 Choosing between the two models by means of the repetition number $a$

The number  $a$  can be computed in three ways.

Firstly, along (16). Solve for  $\omega$ , substitute the computed values for  $\sigma_y^2$  and  $\sigma^2(\bar{y})$ , for instance, when considering a series of January months ( $n = 31$ ) each day gives a mean temperature  $y$ ; the month yields a mean  $\bar{y}^{31}$ . From (30), (31) results a quadratic equation in  $a$ . See the values in row 2, smoothed in row 3 of table 1.

Secondly, the run test of Wald-Wolfowitz can be applied. Each  $y$  is substituted by  $+(-)$  if it lies above (below) the general mean value. Consider, for instance, all  $10 \times 31$  daily maximum temperatures  $y$  of 10 months of January and write these as in table 2. The null hypothesis  $H_0$  is: the  $k$  rows are simultaneously random. Start, for instance, with  $k = 3$ . If  $H_0$  is rejected, take  $k = 4$  and so on. If  $m$  is the first value of  $k$  for which  $H_0$  is not rejected, then  $a = m$ . These  $a^*$  values are mentioned in row 5 of table 1. Neither on the base of the  $a$ - nor on the base of the  $a^*$  values it is necessary to reject model B, because each of the twelve  $p$ -values is smaller than  $(a - 1) : a$ .

Thirdly, it is possible (but it was not done in this way) to compute the value of  $a$  with the aid of the computed  $r_k$  values and expression (27) for  $p_k$ . The equations in  $a$  and  $p$  become 38, resulting in the complicated equation (38a) for  $a$ .

This can be solved only approximately.

#### 4.3 Choosing between the two models by means of $\gamma_2$ in the $\Delta$ series

It was shown, see (33), that for model B in the  $\Delta$  series  $\gamma_2 \equiv 0$  for all  $a$  values  $\geq 2$ , but in model A  $\gamma_2 \neq 0$  (since the initial series is not random). Therefore test  $c_2$ , an estimation of  $\gamma_2$ , against zero

and see whether  $\gamma_2$  may be zero or not. In the first case model B should be rejected. This was done in our investigation of the interdiurnal variability as to the daily minimum and maximum temperature at three stations in the Netherlands, relating to a 30 years period. Of course this does not mean that model A is the statistically best one of all possible models of persistency which furnish a  $\gamma_2 \neq 0$  in the  $\Delta$  series, but it is sufficiently good and simple.

5. Persistency free series

A series may be called persistency-free if the probability that the next element will be one of a finite set of alternatives (the variable being distributed discretely) or that the next element will equal or exceed a given value (in case the variable is distributed continuously) is independent of the foregoing elements.

6. Stationary series

A simplification of the exact definition mentioned in Kendall and Buckland's "Dictionary of statistical terms", may be: "A stochastic process  $\{x_t\}$  is said to be stationary if the density probability  $F[\underline{x}_i \geq a]$  is constant, that is independent of the place  $i$  in the series (in case the variable is distributed continuously)".

A P P E N D I X

A. Model A with not purely persistent series

Replacing supposition (4) by  $\varepsilon y_i \cdot \varepsilon_{i+k} = t^k \sigma_y \cdot \sigma_\varepsilon$ , with  $0 < t < 1$ , we obtain a  $\rho_k$  according to (41). For  $\rho_1, \rho_2 \dots$  see (42). Mind  $\rho_1 > \rho$ ;  $\rho^2 < \rho_2 < \rho_1^2$ ;  $\rho^3 < \rho_3 < \rho_1^3$  etc. Consequently all  $\rho_k$ 's have been enlarged.

B. Model B with autocorrelation in the u-series

Replace supposition (25) by  $\varepsilon u_i \cdot u_j = \omega^k \sigma_u^2$ , with  $0 < \omega < 1$ . Next let us recompute the  $\rho_k$  in the  $y$  series, resulting in (43). For  $\omega = 0$  the expression (27) returns. As soon as  $\omega \neq 0$  the persistency in the  $y$  series becomes unlimited.

C. Artificial series with persistency

The statistician wants to be able to study the mathematical aspects of time series by means of artificial series with autocorrelation. Hence the question arises how to produce series with prescribed autocorrelation. Some methods to achieve this are suggested here.

C1. Positive persistency

C1.1 A series of two alternatives

A vase contains  $N$  balls. These are identical, but  $N_+$  bear a + sign and  $N_-$  a - sign;  $N = N_+ + N_-$ . Draw a ball. Let the sign be +. Now reject the ball, but also add  $\Delta$  plus balls to the vase. Next draw a second time. Let this again be +. Reject this ball, and again add  $\Delta$  plus balls to the vase. Draw a third time. Let this be -. Now reject this ball, plus  $\Delta$  minus balls and draw a fourth time, and so on. In this way a series  $++-+---+---+ \dots$  may result. Eggenberger, Polya, Schumann a.o. have worked mathematically with this model, in which the probability of a + is enlarged whenever the foregoing sign too was a +, however, the series is not a Markoff chain I. Such a series is obtained in the following way.

Let be given two vases  $V$  and  $V^1$ ; the first one contains  $N_+$  plus-balls and  $N_-$  minus-balls, the second one  $N_+^1, N_-^1$ ; fractions  $p, q; p^1, q^1$ . Draw a ball from vase  $V$ . Suppose it is +. Reject the ball in vase  $V$  and remain drawing balls from  $V$  as long as the sign does not change. As soon as the sign changes ( $+ \rightarrow -$  or  $- \rightarrow +$ ) chose the other vase and so on. Now a series of +, - signs results in which the probability that the next element will be + only depends on the sign of the foregoing element and the same for -; there are 4 probabilities, written  $p(+/+)=p; p(-/-)=q; p(+/-)=p^1; q(-/-)=q^1$ . This is per definition a Markoff chain I.

C1.2 A series with values of a variable, which is distributed continuously

C1.2.1 The procedure  $y_i = \rho y_{i-1} + \epsilon_i$ , the so-called model A;  $0 < \rho < 1$ ;  
 $\rho_k = \rho^k$ .

C1.2.2 Start from a random series  $\epsilon_1, \epsilon_2, \epsilon_3 \dots$

Let a group of  $n$  successive elements move through the series and consider the mean values in these groups. The series of these

mean values shows positive autocorrelation with  $\rho_1 = \frac{n-1}{n}$ ;  
 $\rho_2 = \frac{n-2}{n}$  .....  $\rho_n = \frac{1}{n}$ ;  $\rho_k \equiv 0, k \geq n+1$ . These coefficients  
decrease arithmetically.

C1.2.3 The  $u, \varepsilon$ -procedure, treated as model B.

Draw, for instance, the  $u_i$ 's from a  $N(0; \sigma_u)$  population and the  
 $\varepsilon_i$ 's from a  $N(0; \sigma_\varepsilon)$  universum. Take  $a = 4$ ; then the  $y$ -series  
shows autocorrelation-coefficients  $\rho_1 = \frac{3}{4(1+\beta^2)}$ ;  $\rho_2 = \frac{3}{4} \rho_1$ ;

$\rho_3 = \frac{2}{4} \rho_1$ ;  $\rho_4 = \frac{1}{4} \rho_1$ ;  $\rho_k \equiv 0, k \geq 5$ , with  $\beta = \sigma_\varepsilon / \sigma_u$ .

They decrease arithmetically.

C2. Negative autocorrelation (alternating tendency)

C2.1 A series with two alternatives

Proceed as in C1.1, but if a plus-ball is drawn, then reject this  
plus-ball, and also add  $\Delta$  minus-balls to the vase. If a minus-ball were  
drawn, reject this minus-ball, plus  $\Delta$  plus-balls.

C2.2 A series with values of a variable which is distributed continuously

C2.2.1 Follow the procedure  $y_i = \rho y_{i-1} + \varepsilon_i$ , model A, but now  $-1 < \rho < 0$ .

C2.2.2 Let be given a random series  $y$ ; each  $\rho_k \equiv 0$ . Form the  $\Delta$  series.

Then  $\gamma_1 = -\frac{1}{2}$ ;  $\gamma_k \equiv 0, k \geq 2$ . Now start from a  $y$  series with pure  
persistency,  $\rho_k = \rho^k$ , constructed according to C1.2.1.

Then in the  $\Delta$  series  $\gamma_k = -\frac{1}{2} \rho^{k-1} (1-\rho)$ ; each  $\gamma_k < 0$ , but  
 $|\gamma_k| \neq |\gamma_1|^k$ .

How to construct a series with negative autocorrelation-coeffi-  
cients  $\gamma_k$ , for which  $|\gamma_k| = |\gamma_1|^k$  does hold, is a question  
which I have not been able to answer.

-----