KONINKLIJK NEDERLANDS
METEOROLOGISCH INSTITUUT

Verslagen V-65
(R III-247-1960)

Het elimineren van de relatieve inhomogeniteit van twee klimatologische reeksen

door

Dr C. Levert

Inhoud.

519.2

Het elimineren van de relatieve inhomogeniteit van twee klimatologische reeksen.

0. Inleiding

In dit rapport wordt een aspect besproken, dat enigszins analoog is met dat, behandeld in het Wetenschappelijk Rapport (W.R. 58-3) van Rijkoort, getiteld: „Homogeniteit, betrouwbaarheid en reductie van klimatologische reeksen" (zie daarin 2.3 „Reductie van afzonderlijke maandwaarden" pg.27). Men zou het probleem aldus kunnen formuleren: „wat moet men doen met slechte metingen? Wegwerpen? Corrigeren en zo ja, hoe?

1. Definities

a. Homogeen: een stochastisch proces $\{x_i\}$, d.i. een reeks x_1, x_2 ..... heet stationnair of homogeen als, voor gegeven n, de vorm van de multipele verdelingsdichtheid $F(x_{i_1}; x_{i_2}; \ldots x_{i_n})$ dezelfde is voor iedere combinatie van i_1, i_2 .... i_n en als dit het geval is voor elke n.
Deze definitie nemen wij over uit het „Dictionary of statistical terms" door Kendall and Buckland.

b. Relatief homogeen: twee synchrone reeksen x_1, x_2 ....; y_1, y_2 .... heten relatief homogeen als er transformaties $x^* = f(x)$ en $y^* = \varsigma(y)$ bestaan van zodanige aard, dat de verschilreeks $\Delta_i = x_i^* - y_i^*$ homogeen is.

c. Betrouwbaar. Een reeks x_1, x_2 .... heet betrouwbaar als in de reeks der meetfouten ε_i de verdelingsfunctie $F_i(\varepsilon_i)$ onafhankelijk van i is, d.w.z. als de ε-reeks stationnair of homogeen is.

Toelichting

Bij a. Bijv. x_i = jaarsom neerslag op station S in jaar i. Zij de kans, dat van een 4-tal jaren de jaarsommen gelegen zijn in de intervallen 50 à 100, <25, >300, 800 - 1200 mm, voor elk viertal jaren, hoe ook gelegen, al dan niet chronologisch successievelijk, dezelfde en laat dit waar zijn voor elk n-tal jaren en alle combinaties van intervallen. Zulk een reeks heet homogeen. Een engere definitie is: de reeks x_1, ... heet homogeen als de verdelingsfunctie $F_i(x_i)$ onafhankelijk van i is (zie bij Rijkoort, def.D, blz. 1). Anders gezegd: de kansverdeling verandert niet met de tijd (afwezigheid van trend, periodieke schommelingen, etc.).

Bij b. De verschilreeks $\Delta_i = x_i^* - y_i^*$ heet homogeen, als $F_i(\Delta_i)$ onafhankelijk van i is. In de praktijk drukken wij dit dikwijls statistisch minder fraai uit als: „Δ_i is quasi-constant". Dit betekent zo ongeveer: de Δ_i's bewegen zich gemiddeld om eenzelfde niveau en spreiden daaromheen, de gehele reeks door, met een vaste spreiding. Overigens impliceren de voorwaarden E Δ_i = vast en $\sigma^2(\Delta_i)$ = vast geenszins, dat $F_i(\Delta_i)$ onafhankelijk is van i, tenzij de verdeling van Δ een normale mocht zijn. Wij zouden een reeks Δ_i's, die minder spreiden rondom een constant gemiddelde (dat nul kan zijn), „meer quasi-constant" kunnen noemen.

Bij c. Zij w_i de ware, onbekende, waarde, gemeten als x_i, dan heet $\varepsilon_i = w_i - x_i$ de meetfout. De x-reeks heet betrouwbaar, als de nauwkeurigheid waarmede de w gemeten wordt (bijv. gekarakteriseerd door σ_ε) niet met de plaats in de reeks (d.i. dikwijls de tijd) verandert.

Een verandering in de opstelling der instrumenten; een wijziging in zekere voorschriften van behandeling der gegevens, kan een discontinuïteit in de betrouwbaarheid veroorzaken.

2. De V- en de Q-methodes

2.1 De V-, d.i. verschillenmethode.

Def.: De synchrone reeksen $x_1, x_2 \ldots\ldots x_m$; $y_1, y_2 \ldots\ldots y_m$ heten relatief homogeen, als het verschil $v_i \equiv y_i - x_i$ quasi-constant is. Dit is de definitie, die wij in de meeste leerboeken ontmoeten. Wel is waar vermindert men dan dikwijls eerst elke x_i met $\bar{x}$ en elke y_i met $\bar{y}$, doch dit is niet essentieel; als v_i quasi-constant is, dan is $(y_i - \bar{y}) - (x_i - \bar{x})$ het ook en omgekeerd.

Gegeneraliseerde def.: de twee reeksen heten r.h. als het verschil $v_i \equiv y_i - (kx_i + \ell)$ quasi-constant is, waarbij de k en ℓ zodanig gekozen zijn, dat de som $\sum_1^m \left[y_1 - (kx_i + \ell)\right]^2$ minimaal (zegge $\sum_{min.}$) is. Indien x_i en y_i stochastisch lineair verbonden zijn, houdt dit in, dat bij gegeven x_i de synchrone y_i vast ligt, behoudens een toevalsvariabele ε, d.i. $y_i = kx_i + \ell + \varepsilon_i$, waarbij $E\,\varepsilon_i = 0$ en $\sigma_\varepsilon^2 = (1-\rho^2)\,\sigma_y^2$, als ρ de correlatiecoëfficiënt is. Door de berekening van k en ℓ, via de methode der kleinste kwadraten, is $\sum_1^m \left[y_i - (kx_i + \ell)\right] = 0$

en $s_\varepsilon^2 = (1-r^2) s_y^2$, als r de steekproefschatting van ρ is.

(1) Er geldt:

$$k = \sum_1^m (x_i-\bar{x})(y_i-\bar{y}) \Big/ \sum_1^m (x_i-\bar{x})^2 ; \quad \ell = \bar{y} - k\bar{x} \ ; \quad r = \sum(x_i-\bar{x})(y_i-\bar{y}) \Big/ \sqrt{\sum(x_i-\bar{x})^2 \sum(y_i-\bar{y})}$$

Men weet, dat ten opzichte van iedere andere rechte, gekarakteriseerd door een ander stel k_o, ℓ_o, geldt

$$\sum\left[y_i-(k_o x_i+\ell_o)\right]^2 > \sum_{min.} \quad \text{en} \quad \sum\left[y_i-(k_o x_i + \ell_o)\right] \lesseqgtr 0.$$

Zo is $\sum(y_i-x_i)^2 > \sum_{min.}$ Wij hebben derhalve door de kleinste kwadraten-rechte (i.p.v. de 45^o rechte) te beschouwen een gegeneraliseerd verschil gekregen, dat minder spreidt (rondom een gemiddelde nul), d.w.z. „meer quasi-constant" is.

<u>Het minder gegeneraliseerde verschil</u> (waaraan wij verderop toch de voorkeur gaan geven) is $v_i \equiv y_i - k^1 x_i$, waarin k^1 zodanig gekozen is, dat $\sum(y_i - k^1 x_i)^2$ minimaal is. Dan geldt

(2) $k^1 = \sum x_i y_i \Big/ \sum x_i^2$. Echter is $\sum(y_i - k^1 x_i)^2 > \sum_{min.}$ en $\sum(y_i - k^1 x_i) \lesseqgtr 0$

2.2 <u>De Q-, d.i. quotiënten-methode.</u>

<u>Def.</u>: De synchrone reeksen $x_1, x_2 \ldots\ldots x_m$; $y_1, y_2 \ldots\ldots y_m$ heten relatief homogeen als het quotiënt $q_i \equiv y_i/x_i$ quasi constant is.
Deze definitie ontmoet men meestal in de leerboeken. Sommigen delen eerst x_i resp. y_i door $\bar{x}$ resp. $\bar{y}$, doch dit is weer niet essentiëel. Als y_i/x_i q.c. is, dan is $(y_i : \bar{y}) / (x_i : \bar{x})$ het ook en omgekeerd.

<u>Gegeneraliseerde def.</u>: De reeksen heten relatief homogeen als het gegeneraliseerde quotiënt $q_i \equiv y_i / \ell x_i^k$ quasi constant is (van de V- naar de Q-methode gaande, vermenigvuldigen → machtverheffen en optellen → vermenigvuldigen). Daarbij zijn k en ℓ zodanig gekozen, dat $\sum q_i^2$ minimaal is. Zij volgen dus als oplossing der twee vergelijkingen

(3) $\frac{\partial}{\partial k} \sum q_i^2 = 0$ en $\frac{\partial}{\partial \ell} \sum q_i^2 = 0$. De wortels zijn uitermate gecompliceerde uitdrukkingen in de x_i en y_i.

Men maakt zich gewoonlijk van de moeilijkheden af door op de logarithmen $X_i = \log x_i$ en $Y_i = \log y_i$ over te stappen en dan k^1 en ℓ^1 zodanig te kiezen, dat $\sum\left[Y_i - (k^1 X_i + \ell^1)\right]^2$ minimaal is en dan maar, gemakshalve, te besluiten, dat $k = k^1$ en $\ell = 10^{\ell^1}$, zonder na te gaan welke de waarde is van de benadering. Hoe dan ook, de benadering zal wel zodanig zijn, dat toch $\sum(y_i/\ell x_i^k)^2$ kleiner is dan $\sum(y_i/x_i)^2$, zodat $q_i \equiv y_i/\ell' x_i^{k'}$ „beter quasi-constant" is dan $q_i = y_i/x_i$, hoewel het niet zeker is, dat voor deze k' en ℓ' de kleinste kwadratensom bereikt werd. Dan is

$$k^1 = \sum(X_i - \bar{X})(Y_i - \bar{Y}) \Big/ \sum(X_i - \bar{X})^2;\ \ell^1 = \bar{Y} - k^1\bar{X}\ ;\ r(X,Y) =$$

(4) $$\frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\left[\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2\right]^{1/2}}$$

Bovendien wordt dikwijls, ook weer gemakshalve, ondersteld, dat $r(X,Y) \cong r(x,y)$, d.w.z. de c.c. der logarithmen is ongeveer gelijk aan die der waarnemingen zelf. Wij zien, dat $\ell = 10^{\ell^1} = 1$ als $\ell^1 = 0$, d.i. als $\bar{Y} = k^1\bar{X}$ en dit is zeker zo, als $\overline{\log x} = \overline{\log y} = 0$. Nu is $\overline{\log x} \neq \log \bar{x}$ en $\overline{\log y} \neq \log \bar{y}$, doch àls de = tekens in voldoende goede benadering zouden gelden (verderop hierover meer), dan zou zulks inhouden $\bar{x} = \bar{y} = 1$, hetgeen impliceert, dat wij verstandig doen niet de x_i en y_i te beschouwen, maar de quotiënten $\tilde{x}_i \equiv x_i/\bar{x}$ en $\tilde{y}_i \equiv y_i/\bar{y}$ (en dit zou volkomen analoog zijn aan $x_i - \bar{x}$ resp. $y_i - \bar{y}$ in de V-methode). In dát geval zou de quasi-constante de simpele vorm $\tilde{y}_i/\tilde{x}_i^{\,k}$ hebben.

<u>Het minder gegeneraliseerde quotiënt</u> (waaraan wij tenslotte toch de voorkeur geven) luidt vanzelfsprekend $q_i \equiv y_i / x_i^{k''}$, waarin k'' gekozen gedacht is, zodanig, dat $\sum q_i^2$ minimaal is. Men neemt dan niet de exacte k'' volgende uit

(5) $$\frac{d}{dk''}\left(\sum q^2\right) = 0, \text{ doch de benadering } k'' = \sum X_i Y_i \Big/ \sum X_i^2$$

2.3 <u>Hoe ver reikt de dualiteit tussen de V- en Q-methode?</u>

Wij maken in het volgende gebruik van twee (hier niet bewezen) stellingen:

Zij $w = f(v)$; afgeleiden $f^{(p)} = \frac{\partial f}{\partial v^p}$; $\mu \equiv Ev$; $\mu_2 \equiv \sigma^2 = E(v - \mu)^2$; $\mu_i \equiv E(v - \mu)^i$. Dan geldt:

(6) $$E(w) = \underbrace{f + f^{(2)}\frac{\mu_2}{2!}} + f^{(3)}\frac{\mu_3}{3!} + \ldots \quad \text{en}$$

(7) $$\sigma^2(w) = \underbrace{\mu_2\left[f^{(1)}\right]^2} + (\mu_4 - \mu_2)(f^{(2)} : 2!)^2 + (\mu_6 - \mu_2^3)(f^{(3)} : 3!)^2 + \ldots$$
$$2 f^{(1)}\left[f^{(2)}(\mu_3 : 2!) + f^{(3)}(\mu_4 : 3!) + \ldots\right] + \text{enz.}$$

Daarin zijn $f^{(i)}$ gedacht voor het argument μ.

Gewoonlijk is het geoorloofd slechts met de $\sim\!\sim\!\sim$ gedeelten te rekenen. Dan geldt:

(8) $$E(\ln w) \cong \ln\mu - \tfrac{1}{2}(\sigma/\mu)^2 \text{ en } \sigma^2(\ln W) \cong (\sigma/\mu)^2$$

Wij zeiden reeds, ten aanzien van X_i en Y_i, dat de methode der kleinste kwadraten impliceert dat

$$\sum\left[Y_i - (kX_i + \ell'')\right] = 0 \text{ en } s^2\left[Y_i - (kX_i + \ell')\right] = (1-r^2)\,s^2(Y) = s_R^2$$

eveneens

$$E\left[Y_i - (\alpha X_i + \beta)\right] = 0 \text{ en } \sigma^2\left[Y_i - (\alpha X_i + \beta)\right] = \sigma_R^2$$

waarbij k en ℓ' beste schattingen der populatiewaarden α resp. β zijn. Substituerende $q_i = y_i \,/\, \ell a_i^k$, met $\ell = 10^{\ell'}$, staat er $E(\log q_i) = 0$ en $\sigma^2(\log q_i) = \sigma_R^2$.

De twee bovengenoemde stellingen (6 en 7) gebruikende, komt er

(9) $\ln \mu_q \cong \frac{1}{2}(\sigma_q/\mu_q)^2$ en $\sigma^2(\ln q) \cong (\sigma_q/\mu_q)^2$

waaruit

(10)
$$\begin{cases} 2\mu_q^2 \ln \mu_q \cong \sigma_q^2 \\ \ln \mu_q \cong \frac{1}{2}.2{,}30^2\,\sigma_R^2 \\ \sigma_q \cong 2{,}30\,\sigma_R \exp.(\frac{1}{2}.2{,}30^2\,\sigma_R^2) \\ \sigma_q/\mu_q \cong 2{,}30\,\sigma_R \end{cases}$$

met $2{,}30 = {}^{e}\log 10$.

Daarin lezen wij o.a.: hoe strakker het lineaire stochastische verband tussen $X_1 = \log x_i$ en $Y_i = \log y_i$ (d.w.z. hoe dichter de ρ bij 1 en hoe kleiner σ_R^2), hoe kleiner ook μ_q, σ_q en σ_q/μ_q . Anders gezegd: hoe „meer quasiconstant" is de q_i en dit is juist wat wij nastreven met de overgang van het simpele quotiënt y_i/x_i op het minder eenvoudige: $y_i/\ell x_i^k$

Rijkoort heeft er in zijn rapport op pg.15 op gewezen, dat men in principe de quotiëntenmethode in de verschillenmethode kan omzetten door van de grootheden de logarithmen te nemen. Dat zulks niet volkomen waar is willen we laten zien door stap voor stap de dualiteit te onderzoeken.

Verschilmethode	Quotiëntmethode

Gegeven zijn twee synchrone reeksen x_i, y_i ; i = 1, 2 .... n. De universumgemiddelden zullen zijn $Ex \equiv \mu_x$, $E_y = \mu_y$; varianties σ_x^2, σ_y^2 ; in de steekproef $\bar{x}$, $\bar{y}$, s_x^2, s_y^2.

Definieer

$\tilde{x}_i = x_i - \bar{x}$; $\tilde{y}_i = y_i - \bar{y}$

Kies die k, waarvoor

$\sum_1^m (\tilde{y}_i - k\tilde{x}_i)^2$ minimaal is. Dit is

voor $k = \dfrac{\sum \tilde{x}_i \tilde{y}_i}{\sum \tilde{x}_i^2}$. Dan is $\sum_1^m (\tilde{y}_i - k\tilde{x}_i) = 0$

en

$s^2(\tilde{y}_i - k\tilde{x}_i) = s_y^2(1-r^2)$ als

$$r = \frac{\sum \tilde{x}_i \cdot \tilde{y}_i}{\sqrt{\sum \tilde{x}_i^2 \sum \tilde{y}_i^2}}$$

Definieer analoog

$\tilde{x}_i = x_i/\bar{x}$ en $\tilde{y}_i = y_i/\bar{y}$

Analoog: kies die k, waarvoor $S = \sum_1^m (\tilde{y}_i/\tilde{x}_i^k)^2$ minimaal is. De k wordt gevonden uit $\frac{dS}{dk} = 0$, d.i. $\sum_1^m \left(\frac{\tilde{y}_i}{\tilde{x}_i^k}\right)^2 \ln x_i = 0$. Een eenvoudige expliciete uitdrukking voor k is er niet. De dualiteit zou fraai zijn als voor deze k tevens $\sum_1^m (\tilde{y}_i/\tilde{x}_i^k)$ gelijk aan 1 zou zijn, doch noch de ongelijkheid, noch de gelijkheid kan men bewijzen. Overigens lost men doorgaans de k op een andere, niet exacte, wijze op, n.l. via de logarithmen $\tilde{X}_i \equiv \log \tilde{x}_i$ en $\tilde{Y}_i \equiv \log \tilde{y}_i$.

Dan wordt $\sum \left[\tilde{Y}_i - (k\tilde{X}_i + \ell')\right]^2$ geminimaliseerd ten aanzien van k en ℓ'.

Dan is

$$k = \frac{\sum (\tilde{X}_i - \bar{\tilde{X}})(\tilde{Y}_i - \bar{\tilde{Y}})}{\sqrt{\sum (\tilde{X}_i - \bar{\tilde{X}})^2 \sum (\tilde{Y}_i - \bar{\tilde{Y}})^2}} \quad \text{en}$$

$\ell' = \bar{\tilde{Y}} - k\bar{\tilde{X}}$ terwijl $\sum \left[\tilde{Y}_i - (k\tilde{X}_i + \ell')\right] = 0$

Nu is $\tilde{X} \equiv \log \tilde{x} = \log x - \log \bar{x}$ en idem y. Dus $\bar{\tilde{X}} = \overline{\log x} - \log \bar{x} \neq 0$ en idem y. Derhalve

$$k = \frac{\sum (\log x_i - \overline{\log x})(\log y_i - \overline{\log y})}{\sqrt{\sum (\log x_i - \overline{\log x})^2 \sum (\log y_i - \overline{\log y})^2}}$$

met $\ell' = (\overline{\log y} - \log \bar{y}) - k(\overline{\log x} - \log \bar{x})$. Gebruikmakende van de stelling over $\sigma^2(\ln \ldots)$, is dan $\ell' \cong -\frac{0.43}{2}\left\{(\sigma_y/\mu_y)^2 - k(\sigma_x/\mu_x)^2\right\}$.

Wij interesseren ons voor het geval $\ell' = 0$. Daartoe moet

$k \simeq (\sigma_y/\mu_y)^2 : (\sigma_x/\mu_x)^2$ gelden. Ook is in de lineaire stochastische relatie $k = \rho \frac{\sigma(Y)}{\sigma(X)} \simeq \rho \frac{\sigma_y/\mu_y}{\sigma_x/\mu_x}$. Maar dan zou $\rho \simeq \frac{\sigma_y/\mu_y}{\sigma_x/\mu_x}$ moeten zijn, met als steekproefschatting $\frac{s_y/\bar{y}}{s_x/\bar{x}}$, terwijl eigenlijk

$$r = \frac{\sum(\tilde{X}-\bar{\tilde{X}})(\tilde{Y}-\bar{\tilde{Y}})}{\left[\sum(\tilde{X}-\bar{\tilde{X}})^2 \sum(\tilde{Y}-\bar{\tilde{Y}})^2\right]^{1/2}}$$

Het is zeer de vraag of dit geldt. Noodzakelijk is de gelijkheid geenszins.

Men zou ook $\sum(\tilde{Y}_i - k^*\tilde{X}_i)^2$ kunnen minimaliseren. De som is minimaal voor $k^* =$

$$\frac{\sum \tilde{Y}_i\tilde{X}_i}{\sum \tilde{X}_i^{\,2}} = \frac{\sum(\log x_i-\log \bar{x})(\log y_i-\log \bar{y})}{\sum(\log x_i-\log \bar{x})^2},$$

doch nu is niet tevens $\sum\left[\tilde{Y}_i - k^*\tilde{X}_i\right] = 0$, want zou de gelijkheid gelden, dan zou k^* ook gelijk aan $\frac{\overline{\log y} - \log \bar{y}}{\overline{\log x} - \log \bar{x}}$ zijn en dat is niet te bewijzen.

Ondertussen: $\boxed{\text{deze } k^* \neq k}$!

Men zou graag zien, terwille van de dualiteit, dat voor deze k^* zou gelden $E\left[\tilde{y}_i/(\tilde{x}_i)^k\right] = 1$ of $\ln E\left[\;\right] = 0$. Via de meermalen genoemde stelling luidt de vraag dan: $E(\ln q_i) = \frac{1}{2}(\sigma_q/\mu_q)^2$? Het is helaas niet te zeggen of de gelijkheid geldt. Indien echter in voldoend goede benadering $E(\log z) \simeq \log(Ez)$ geldt, dan is in bovenstaande beschouwing automatisch $\ell' = 0$ en $10^{\ell'} = \ell = 1$ en bijgevolg ook $E(\tilde{Y}_i - k\tilde{X}_i) = 0$, d.i. $E(\log \tilde{y}_i - k \log \tilde{x}_i) = 0$ d.i. $E\log(\tilde{y}_i/\tilde{x}_i^{\,k}) = 0$ en ook $\log(E\,\tilde{y}_i/\tilde{x}_i^{\,k}) = 0$ of $E\,\tilde{y}_i/\tilde{x}_i^{\,k} = 1$. Dàn is de dualiteit wel volkomen.

Conclusie: De dualiteit tussen de twee methoden is, exact beschouwd, niet volkomen.

3. Wanneer gaan Q- en V-methode in elkaar over?

3.1 Beschouw weer twee synchrone reeksen: x_i en y_i. Zij p.d. $x_i = x_0 + \xi_i$; $y_i = y_0 + \eta_i$; x_0 en y_0 constanten. Bijgevolg: $\lg x_i = \lg x_0 + \xi_i/x_0$, mits $\xi_i \ll x_0$ en $\lg y_i = \lg y_0 + \eta_i/y_0$, mits $\eta_i \ll y_0$. Dus

$\lg(y_i/x_i) = \lg(y_0/x_0) + (\frac{\eta_i}{y_0} - \frac{\xi_i}{x_0})$. Stel nu $x_0 = y_0$, dan is

$\lg y_i/x_i = \frac{1}{x_0}(\eta_i - \xi_i)$.

Aangezien $\xi_i \ll x_0$ en $\eta_i \ll y_0 = x_0$, is

$\eta_i - \xi_i \ll x_0$ en

$$\begin{cases} y_i/x_i \simeq 1 + \frac{1}{x_0}(\eta_i - \xi_i) \quad ; \text{ ook is} \\ y_i - x_i = \eta_i - \xi_i \end{cases} \tag{11}$$

Hierin lezen wij, dat het verschil en het quotiënt <u>gelijktijdig</u> quasi-constant zijn, mits $x_0 = y_0$ en $\xi_i , \eta_i \ll x_0$.

Voorbeelden:

3.1.2 <u>Temperatuurreeks</u>; $x_0 = y_0 = 273^oK$; $x\ ^oC \equiv 273 + x^oK$. Nu is zeker elke ξ_i en $\eta_i \ll 273$. Beide methoden (V- en Q-) zijn toepasbaar, maar men zal de verschillenmethode kiezen om de eenvoud.

3.1.3 <u>Dagsommen neerslag</u>.

De dagsommen zijn eenzijdig begrensd, met nul als kleinste waarde. De verschillen $y_i - x_i$ niet. Als men $x_0 = y_0 = 0$ stelt, komt men in moeilijkheden. Het blijkt niet mogelijk te bewijzen, dat uit de quasiconstantheid van $y_i - x_i$ ook die van y_i/x_i volgt, welk quotiënt tussen 0 en ∞ wisselt (als bij afspraak voor 0/0 de 1 genomen wordt). Hier is het niet mogelijk èn $x_0 = y_0$ èn $x_i, y_i \ll x_0$ te bereiken.

3.1.4 <u>Jaarsommen neerslag</u>.

Wel is waar is $x = 0$ nu ook mogelijk, echter zeer onwaarschijnlijk (wellicht nog nooit voorgekomen in ons land). Beschouw twee stations S_1 en S_2. Kies bijv. $x_0 = y_0 = 750$ mm, waarna het zeker is, dat $\xi_i , \eta_i \ll 750$ (het teken $\ll$ nader te definiëren), zodat thans zowel de V- als de Q-methode van toepassing is.

3.2 <u>Specialisatie van 3.1 , door $x_0 \equiv \mu_x$ en $y_0 \equiv \mu_y$ te kiezen.</u>

Zij $x_i = \mu_x + \tilde{x}_i$ en $y_i = \mu_y + \tilde{y}_i$, waarna

$\lg x_i \cong \lg \mu_x + \frac{\tilde{x}_i}{\mu_x}$ en $\lg y_i \cong \lg \mu_y + \frac{\tilde{y}_i}{\mu_y}$, als $\tilde{x}_i \ll \mu_x$ en $\tilde{y}_i \ll \mu_y$.

(12) Kies k zodanig, dat $k \lg \mu_x = \lg \mu_y$, d.i. $k = \lg \mu_y : \lg \mu_x$.

Gevolg:

(13) $$\lg \left(\frac{y_i}{x_i^k}\right) \cong \frac{1}{\mu_x^k} (\tilde{y}_i - k^* \tilde{x}_i) \text{ met } k^* = \frac{\mu_y \lg \mu_x}{\mu_x \lg \mu_x} = k\mu_x^{k-1}$$

(14) waaruit volgt: $k \cong \frac{E \lg y}{E \lg x} \cong E\left(\frac{\lg y}{\lg x}\right)$ en $k^* \cong \frac{Ey \lg y}{Ex \lg x} \cong E \frac{\lg y^y}{\lg x^x}$

waaruit volgt:

(15) $$y_i/x_i^k \cong 1 + \mu_x^{-k} \left[\tilde{y}_i - k^* \tilde{x}_i\right] \text{ , mits } \tilde{y}_i - k^* \tilde{x}_i \ll \mu_y .$$

Daarin leest men:

$\tilde{y}_i - k^* \tilde{x}_i$ (een gegeneraliseerd verschil) is gelijktijdig quasiconstant met y_i/x_i^k (een gegeneraliseerd quotiënt), als $\tilde{x}_i \ll \mu_x$; $\tilde{y}_i \ll \mu_y$; $\tilde{y}_i - k^* \tilde{x}_i \ll \mu_y$, met

$$\mu_y = \mu_x^k \text{ en } \underline{k^* = k\mu_x^{k-1}}$$

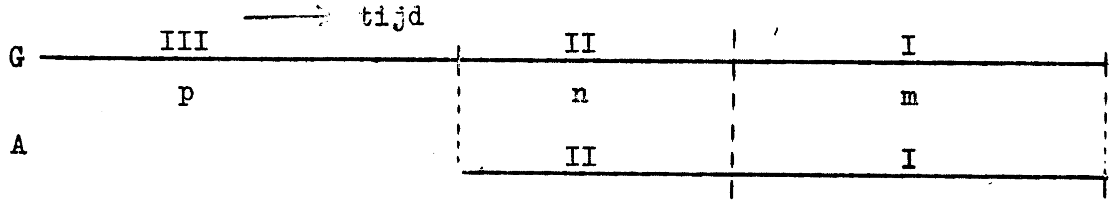
Er staat niet, dat de beste $\hat{k}$, waarvoor $y_i/x_i^{\hat{k}}$ quasiconstant is (d.w.z. deze breuk een minimale kwadratensom levert) juist gelijk aan $\lg \mu_y / \lg \mu_x$ zou zijn. Evenmin staat er, dat de beste $\hat{k}^*$, waarvoor $\tilde{y}_i - \hat{k}^* \tilde{x}_i$ quasiconstant zou zijn (d.i. dit verschil een minimale kwadratensom bezorgt), met de $\hat{k}$ zou samenhangen, zoals in het bovenstaande de k^* met k samenhangt.

Mocht dit in benadering wel het geval zijn, dan zou, indien volgens de Q-methode gewerkt wordt, de tijdrovende berekening van k (lineaire regressie op gelogarithmiseerde metingen) kunnen worden omzeild door direct $k = \lg \mu_y / \lg \mu_x$, d.w.z. in de steekproef $k = \lg \bar{y} / \lg \bar{x}$, te berekenen. Voelt men meer voor de V-methode, dan kieze men $k^* = k\,(\bar{x})^{k-1}$.

N.B. k en k^* zijn alleen dan principieel gelijk (en wel 1) als $\mu_x = \mu_y$. Dan zijn wij teruggekeerd tot de "gewone" quotiënten- en verschillen-methode.

4. Het homogeniseren.

Gesteld er doet zich de volgende situatie voor.



G stelt het station voor, waarop gedurende tijdvak I (m jaren) de jaarsommen correct gemeten werden: g_{1i} ; $i = 1,2 \ldots m$. Gedurende de tijdvakken II en III (n en p jaren) werd minder correct gemeten: g_{2i} , $i = 1,2 \ldots n$ en g_{3i} , $i = 1,2 \ldots p$.
A stelt het referentiestation voor, waar de metingen voortdurend even correct geschiedden: a_{1i} ; $i = 1,2 \ldots m$ en a_{2i} ; $i = 1,2 \ldots n$.
Anders gezegd: de meetreeks g heeft een discontinuïteit in de betrouwbaarheid op de overgang van II naar I. Er wordt niet beweerd (althans nù nog niet), dat de a-reeks homogeen is.
Het bovenstaande impliceert, dat de g- en a-reeksen niet relatief homogeen zijn. De vraag rijst, àls wij de metingen g_{2i} en g_{3i} niet zonder meer willen wegwerpen, hoe zullen wij ze dan kunnen behandelen, zodanig, dat na de behandeling een reeks g^1_{3i}, g^1_{2i}, g_{1i} verschijnt, welke wel relatief homogeen is vergeleken met de a-reeks?
Een bijkomstige, doch zeker niet minder belangrijke vraag is:
Kunnen wij een criterium opstellen, dat beslist over het al of niet zinvol zijn van zulk een behandeling?

4.1 De Q-methode, toegepast op de waarnemingen zelf.

Beschouw de m synchrone paren g_{1i}, a_{1i} in deelvak I en de n synchrone paren g_{2i}, a_{2i} in II. Kies de methode berustend op
(16) het "minder gegeneraliseerde quotiënt" $q_i = g_i/a_i^k$ en bereken zowel in I als in II de beste k, d.i. resp. k_1 en k_2. Zoals gezegd in 2.2 , is een benadering
(17) $k_1 = \sum_1^m A_{1i} G_{1i} \,/\, \sum_1^m A_{1i}^2$ en analoog k_2 . Daarbij $A = \lg a$; $G = \lg g$.
Er is dan een steekproef van m q_{1i}-waarden en een tweede van n q_{2i}-waarden, met gemiddelden $\bar{q}_1$, $\bar{q}_2$ en varianties λ_1^2 en λ_2^2 .
Men interessert zich in het kader van het onderzoek der relatieve homogeniteit in het bijzonder voor de vraag of $\bar{q}_1$ en $\bar{q}_2$ significant verschillen. Mocht ter beantwoording van deze vraag de t-toets gekozen worden, dan zal onderzocht moeten worden of de twee universa (uit de ene komen de q_1's; uit de andere de q_2's) normaal verdeeld

zijn en zal eveneens de statistische gelijkheid van s_1^2 en s_2^2 onderzocht moeten worden.
Gesteld de uitkomst luidt, dat $\bar{q}_1$ en $\bar{q}_2$ statistisch ongelijk zijn. De vraag rijst dan: hoe de g_{2i}'s te "behandelen" tot g_{2i}^1's, zodanig dat daarna wèl relatieve homogeniteit bereikt is? Exact beschouwd zou men voor de paren g'_{2i}, a_{2i} de nieuwe beste k_2^1 moeten berekenen en daarna verifiëren of $\bar{q}_2^1$ nu wel statistisch gelijk is aan $\bar{q}_1$. Wij willen dit vereenvoudigen door te eisen, dat ons doel bereikt is zodra

$$\overline{(g^1_{2i}/ a_{2i}^{k_2})}^{n} = \overline{(g_{1i} / a_{1i}^{k_1})}^{m}$$

; $\bar{j}^n$ duidt op het gemiddelde van n j-waarden.

Er blijken twee "behandelingen" tot dit doel te leiden.

(18) de "correctie" $g_{2i}^{*} = g_{2i} \dfrac{\hat{q}_1}{\hat{q}_2}$

(19) de schatting $g_{2i}^{**} = \hat{q}_1 . a_{2i}^{k_2}$.

als $\hat{q}_1 = \overline{\left(\dfrac{g_{1i}}{a_{1i}^{k_1}}\right)}^{m}$ en $\hat{q}_2 = \overline{\left(\dfrac{g_{2i}}{a_{2i}^{k_2}}\right)}^{n}$

Vervolgens het bewijs van de relatieve homogeniteit:

$$\overline{(g_2^{*}/ a_2^{k_2})}^{n} = \frac{\hat{q}_1}{\hat{q}_2} \overline{(g_2 / a_2^{k_2})}^{n} = \frac{\hat{q}_1}{\hat{q}_2} . \hat{q}_2 = \hat{q}_1 \quad \text{q.e.d. en}$$

$$\overline{(g_2^{**}/a_2^{k_2})}^{n} = \overline{(q_1 . a_2^{k_2}/a_2^{k_2})}^{n} = \hat{q}_1 \quad \text{q.e.d.}$$

De vraag rijst onmiddellijk: welke van de twee behandelingen zullen wij kiezen? Is de eerste beter dan de andere, en wat zal "beter" heten? Bij de keuze kunnen wij aanhaken aan de kwestie van de "kleinste foutenvariantie", waarover in § 6.

In deelvak III kunnen wij vanzelfsprekend alleen de correctie formule gebruiken, aangezien het referentiestation A geen parallelmetingen a_{3i} leverde. Er is echter geen $\hat{q}_3$, zodat wij in $g_{3i}^{*} = g_{3i} . \hat{q}_1/\hat{q}_3$ de $\hat{q}_3$ noodgedwongen door $\hat{q}_2$ vervangen, hetgeen "des te minder erg" is (dit eigenlijk ook statistisch te vertalen) naarmate elk der deelreeksen II en III langer is en naarmate de quasiconstante "meer constant" is.
Hier blijkt weer, dat het nut heeft het simpelste quotiënt g_i/a_i te generaliseren tot g_i/a_i^k of zelfs $g_i/\ell a_i^k$, hetwelk meer constant is. Dat we het eerste van deze twee kiezen (ofschoon onder beste aanpassing der constanten het tweede een kleinere kwadratensom heeft) heeft zijn reden hierin, dat wij (louter bij afspraak) de

quasi-constantheid bestuderen aan het gemiddelde $\bar{q}$. Dit is eigenlijk geen noodzaak, immers, de exacte definitie volgende, komt het erop aan te onderzoeken of de kansverdeling in het universum, waaruit de II-steekproef van q's komt, volkomen identiek is aan die van het universum, uit welke de I-steekproef van q's stamt; zulk een onderzoek omvat méér dan verifiëren of de populatiegemiddelden dezelfde zijn. Toch willen wij ons tot het laatste beperken. Dit doet de keuze op g_i/a_i^k vallen (d.i. $\ell = 1$); in de V-methode wordt dit nog duidelijker. Deze volgt nu.

4.2 <u>De V-methode, toegepast op de logarithmen.</u>

Zij weer $G_i = \lg g_i$ en $A_i = \lg a_i$.
Zij $\hat{k}_1$ de k-waarde, waarvoor in I de som $\sum (G_i - k_1 A_i)^2$ minimum is en idem k_2 in II. Zoals gezegd:

(20) $\hat{k}_1 = \sum_1^m A_{1i} G_{1i} / \sum_1^m A_{1i}^2$ en idem $\hat{k}_2$ in II. In het geval

van een discontinuïteit in de betrouwbaarheid der g-reeks (dan ook in de G-reeks) bij overgang van II naar I, zullen $\hat{k}_1$ en $\hat{k}_2$ significant verschillen, evenals de gemiddelden

$\overline{V_1}$ en $\overline{V_2}$, met $\overline{V_1} = \overline{G_{1i} - \hat{k}_1 A_{1i}}^m$ en idem $\overline{V_2}$.

N.B.: Wanneer wij gewerkt zouden hebben met G_{1i} tegenover $k_1 A_{1i} + \ell_1$ (index 2 in II) dan zou automatisch
èn $\overline{G_{1i} - (k_1 A_{1i} + \ell_1)}^m = 0$ èn $\overline{G_{2i} - (k_2 A_{2i} + \ell_2)}^n = 0$ geweest zijn en zou een verificatie of $\overline{V_1}^m$ significant van $\overline{V_2}^n$ verschilt vanzelfsprekend zonder zin geworden zijn en zou het significantieonderzoek op de statistische gelijkheid der varianties gericht moeten worden (met verdere consequenties t.a.v. behandelingen van G_{2i}). Dit willen wij hier opzettelijk vermijden.
Zie overigens de appendix

Gevraagd opnieuw hoe G_{2i} te behandelen opdat er relatieve homogeniteit bereikt wordt. Antw.: vervang elke G_{2i} door

(21) $\boxed{G_{2i}^* = G_{2i} + \overline{V_1}^m - \overline{V_2}^n}$ want dan is

$$\overline{G_{2i}^* - k_2 A_{2i}}^n = \overline{G_{2i} - k_2 A_{2i} + \overline{V_1}^m - \overline{V_2}^n}^n = \overline{V_1}^m \qquad \text{q.e.d.}$$

Verder volgt g_{2i}^* uit $G_{2i}^* = \lg g_{2i}^*$

(22) Wij kunnen schrijven $\boxed{g_{2i}^{*} = f\, g_{2i}}$ met $f \simeq \dfrac{\overline{(g_{1i}/a_{1i}^{k_1})}^{m}}{\overline{(g_{2i}/a_{2i}^{k_2})}^{n}}$

Hoe beter de benadering $\overline{\lg x}^{m} \simeq \lg \overline{x}^{m}$ geldt, hoe beter geldt de bovenstaande uitdrukking voor f.
Hiermee zijn wij, zeer begrijpelijk overigens, bijna tot de correctie-formule uit 4.1 teruggekeerd.
Minder benaderend zou zijn

(23) $f \simeq \dfrac{\overline{q_1}^{m}}{\overline{q_2}^{n}}\, e^{-\frac{1}{2}(t_1^2 - t_2^2)}$, waarin $q_1 = g_{1i}/a_{1i}^{k_1}$ en analoog q_2, alsmede $t_1 = s_1/\overline{q}_1$ als $\overline{q}_1$ resp. s_1 het gemiddelde en de variantie der m q_{1i} - waarden voorstellen. Idem t_2.

Hoe minder t_1 en t_2 verschillen, hoe meer geldt de benadering (22).

5. De "zuiverheid" van correctie en schatting.

5.0 In het volgende wordt gewerkt met de logarithmen der metingen. Laat de n metingen G_{2i} in deelvak II "niet correct" zijn; d.i. zij wijken meer van de onbekende ware waarde W_{2i} ($= \lg w_{2i}$) af dan louter het gevolg van meetfouten alleen kan zijn (bijv. staat ten westen van de regenmeter een hoge boom).
De fout heet $G_{2i} - W_{2i}$. Deze zal, gemiddeld genomen, niet nul zijn. Voor de correctie G_{2i}^{*}, zowel als voor de schatting G_{2i}^{**}, zullen wij toch in de eerste plaats willen eisen, dat, zowel $G_{2i}^{*} - W_{2i}$ en $G_{2i}^{**} - W_{2i}$, gemiddeld over de n metingen nul zou leveren. Daarbij is

(24) $G_{2i}^{*} = G_{2i} + \overline{V_1}^{m} - \overline{V_2}^{n}$, zie (21) en

(25) $G_{2i}^{**} = k_2 A_{2i} + \overline{V_1}^{m}$

5.1 Wij werken uit:

$E(G_{2i}^{*} - W_{2i}) = E(G_{2i} - W_{2i}) + E(\overline{V_1}^{m}) - E(\overline{V_2}^{n})$ Noem $Eg_{1i} \equiv Ew_{1i} = \mu_1$;

$Eg_{2i} \equiv \mu_{g_2}$; $Ew_{2i} \equiv \mu_{w_2}$; $Ea_{1i} \equiv \mu_{a_1}$; $Ea_{2i} \equiv \mu_{a_2}$.

N.B. In I zijn Eg_{1i} en Ew_{1i} gelijk, omdat de metingen correct zijn (zoveel betekenende als: zou men in staat zijn eenzelfde ware waarde talloos vele keren te meten, dan zou het gemiddelde van de verzameling meetresultaten met toenemend aantal metingen naar de ware waarde tenderen). In II juist niet. Het is bovendien niet zeker,

dat μ_{w1} gelijk aan μ_{w2} is, evenals μ_{a1} gelijk aan μ_{a2}, omdat trend in de reeks nog niet uitgesloten werd.
Gesteld de benadering $EG_{2i} = E\lg g_{2i} \cong \lg Eg_{2i} = \lg \mu_{g2}$ etc. is geoorloofd. Dan staat er

$$E(G_{2i}^{*} - W_{2i}) \cong \lg\mu_{g2} - \lg\mu_{w2} + \lg\mu_{g1} - k_1\lg\mu_{a1} - \lg\mu_{g2} + k_2\lg\mu_{a2} =$$

(26) $$\lg\left(\mu_{g1}/\mu_{a_1}^{k_1}\right) - \lg\left(\mu_{g2}/\mu_{a_2}^{k_2}\right)$$

De relatieve homogeniteit der reeksen $(I+II)_G$ en $(I+II)_A$ impliceert

(27) $$E\left(\frac{g_{1i}}{a_{1i}^{k_1}}\right) = E\left(\frac{g_{2i}^{1}}{a_{2i}^{k_2}}\right),$$ als

g_{2i}^{1} het meetresultaat voorstelt bij een wèl correcte meting.
$Eg_{2i}^{1} = Ew_{2i} = \mu_{w2}$, doch $Eg_{2i} \neq Eg_{2i}^{1}$
Als nu ook nog bij benadering geldt:

(28) $$E\left(\frac{g_{1i}}{a_{1i}^{k_1}}\right) \cong \frac{E(g_{1i})}{(E\,a_{1i})^{k_1}} = \frac{\mu_{g_1}}{\mu_{a_1}^{k_1}}$$ en idem voor II

dan is (26) inderdaad nul en is de "zuiverheid" der schatting G_{2i}^{*} bewezen.

5.2 $$E(G_{2i}^{**} - W_{2i}) = E\left[k_2A_{2i} + \overline{V_1}^{m} - W_{2i}\right] = k_2EA_{2i} + E\overline{G_{1i}}^{m} - k_1\overline{EA_{1i}}^{n}$$
$$- EW_{2i} = k_2\lg\mu_{a_2} + \lg\mu_{g_1} - k_1\lg\mu_{a_1} - \lg\mu_{w_2} = \lg\frac{\mu_{g1}}{\mu_{a_1}^{k_1}} - \lg\frac{\mu_{w2}}{\mu_{a_2}^{k_2}}$$

$\cong 0$. Derhalve is zowel G_{2i}^{*} als G_{2i}^{**} zo goed als een zuivere schatting in de statistische betekenis. Bijgevolg heeft het weinig zin bij de keuze tussen schatting en correctie op de kwestie van de zuiverheid te letten. Blijft over de foutenvariantie te beschouwen.

6. Foutenvariantie.

6.0 Een schatting willen wij behalve "zuiver" ook "meest doeltreffend". Als t een schatting is van τ, dan heet t een zuivere schatting als $Et = \tau$. Een zuivere schatting heet meest doeltreffend, als $E(t-\tau)^2 \equiv \sigma^2(t-\tau)$ kleiner is dan de waarden, die deze varianties aannemen bij andere schatters van τ.
Wij willen nu de foutenvarianties bij g_{2i}^{*} en g_{2i}^{**} berekenen; de

"fout" is p.d. $g_{2i} - w_{2i}$ resp. $g_{2i}^{**} - w_{2i}$. Wij doen dit weer aan de logarithmen.

6.1 De foutenvariantie van de incorrecte meting.

Zij de beste lineair stochastische relatie tussen de incorrecte meting G_{2i} en A_{2i}, alsmede die tussen W_{2i} en A_{2i} :

(29) $G_{2i} = \hat{k}_{2g} A_{2i} + V_{g2i}$ en $W_{2i} = \hat{k}_{2w} A_{2i} + V_{w2i}$

Daarbij is streng formeel $k_{2g} = \sum A_{2i} G_{2i} / \sum A^2_{2i}$ en $\hat{k}_{2w} = \sum A_{2i} W_{2i} / \sum A_{2i}^2$ (Wij hebben in 4.2 de $\overline{V_{g2i}}^n$ met $\overline{V}_2$ aangeduid).

Er komt

$$\sigma^2(G_{2i} - W_{2i}) = \sigma^2\left\{A_{2i}(\hat{k}_{2g} - \hat{k}_{2w}) + (V_{g2i} - V_{w2i})\right\}.$$

(30) Wij willen nu óók nog, gemakshalve, $\hat{k}_{2g} = \hat{k}_{2w}$ onderstellen.

Dan is

(31) $$\boxed{\sigma^2(G_{2i} - W_{2i}) = \sigma^2(V_{g2i}) + \sigma^2(V_{w2i}) - 2\rho\sigma(V_{g2i})\sigma(V_{w2i})}\ ;\ \text{genoemd } \sigma^2$$

Hierin is ρ de correlatiecoëfficiënt tussen $V_{g2i} = G_{2i} - \hat{k}_2 A_{2i}$ en $V_{w2i} = W_{2i} - \hat{k}_2 A_{2i}$. Deze is zeker niet nul; hoe minder incorrect de meting g_{2i}, hoe dichter ρ bij 1, lijkt ons.

De ρ heeft dus iets te maken met de „slechtheid" der metingen.

6.2 De foutenvariantie voor de correctie.

$$\sigma^2(G_{2i}^{*} - W_{2i}) = \sigma^2\left\{G_{2i} - W_{2i} + \overline{G_{1i} - k_1 A_{1i}}^m - \overline{G_{2i} - k_2 A_{2i}}^n\right\} =$$

$$\sigma^2\left\{V_{g2i} - V_{w2i} + \overline{V_{g1}}^m - \overline{V_{g2}}^n\right\} =$$

(32) $$\boxed{(1 - \frac{1}{\sqrt{n}})^2 \sigma^2(V_{g2i}) + \sigma^2(V_{w2i}) + \frac{1}{m}\sigma^2(V_{g1i}) - 2\rho(1 - \frac{1}{\sqrt{n}})\sigma(V_{g2i}).\sigma(V_{w2i})}\ ;\ \text{genoemd } \sigma_{*}^2$$

Bij de afleiding is gebruik gemaakt van een stelling, dat de correlatiecoëfficiënt tussen z_1 en $\frac{1}{n}(z_1 + z_2 + \dots z_n)$ gelijk aan $1/\sqrt{n}$ is, als althans de z_i's ongecorreleerd zijn.

6.3 De foutenvariantie voor de schatting.

$$\sigma^2(G_{2i}^{**} - W_{2i}) = \sigma^2\left\{k_2 A_{2i} + \overline{V_1}^m - W_{2i}\right\} = \sigma^2\left\{\overline{V_1}^m - V_{w2i}\right\} =$$

(33) $$\boxed{\sigma^2(V_{w2i}) + \frac{1}{m}\sigma^2(V_{w1i})}\ ;\ \text{genoemd } \sigma_{**}^2$$

Eerst nu is het nuttig afwezigheid van trend te onderstellen.

Dus is $\sigma(V_{w1i}) = \sigma(V_{w2i})$, Bovendien geldt

$\sigma^2(V_{g1i}) \equiv \sigma^2(V_{w1i})$. Daardoor worden (32) en (33) gemakkelijker vergelijkbaar.

6.4 Vergelijking der foutenvarianties.

Zie de uitdrukking voor σ^2, σ_*^2 en σ_{**}^2 in 6.1 6.2 6.3

Het gaat om de kleinste dezer drie. Schrijf ter vereenvoudiging

$\sigma^2(V_{g2i}) \equiv \sigma_g^2$; $\sigma^2(V_{w1i}) = \sigma^2(V_{w2i}) = \sigma_w^2$.

Dan blijkt

(34) $\sigma_*^2 < \sigma_{**}^2$ als $\rho > \frac{1}{2}(1 - \frac{1}{\sqrt{n}})\, T_g$; het rechter lid bevat de m niet.

(35) $\sigma_*^2 < \sigma^2$ als $\rho < \frac{2\sqrt{n} - 1}{2n} F - \frac{1}{2}\frac{\sqrt{n}}{m}(T_g T_w)^{-1}$

(36) $\sigma_{**}^2 < \sigma^2$ als $\rho < \frac{1}{2}\left[F - (mT_g T_w)^{-1}\right]$; het rechter lid bevat de n niet.

Hierin is $T_g = \sigma_{g2}/\sigma_{g1}$; $T_w = \sigma_w/\sigma_{g1}$; $F = T_g/T_w = \sigma_{g2}/\sigma_w$.

De correctieformule is de "beste", d.i. $\sigma_* < \sigma_{**} < \sigma$,

als

(37) $$\boxed{\frac{1}{2}(1 - \frac{1}{\sqrt{n}})T_g < \rho < \frac{1}{2}\left[F - (mT_g T_w)^{-1}\right]}$$

De schattingsformule is de "beste", d.i. $\sigma_{**} < \sigma_* < \sigma$, als

(38) $$\boxed{\text{én } \rho < \frac{1}{2}(1 - \frac{1}{\sqrt{n}})T_g \text{ én } \rho < \frac{2\sqrt{n}-1}{2n}F - \frac{1}{2}\frac{\sqrt{n}}{m}(T_g T_w)^{-1}}$$

Er kan zich ook de situatie voordoen, dat het noch zinvol is om de correctieformule te gebruiken, noch de schattingsformule.

Dat is het geval als én $\sigma_* > \sigma$ én $\sigma_{**} > \sigma$, d.i. als

(39) $$\boxed{\text{én } \rho > \frac{1}{2}\left[F - m(T_g T_w)^{-1}\right] \text{ én } \rho > \frac{2\sqrt{n}-1}{2n}F - \frac{1}{2}\frac{\sqrt{n}}{m}(T_g T_w)^{-1}}$$

In dat geval kunnen wij de incorrecte metingen beter laten zoals ze zijn.

Het is helaas moeilijk deze gecompliceerde uitdrukkingen te overzien.

6.5 Wat doet de afstand $\overline{GA}$ tussen de stations G en A ertoe?

Wordt de ρ van (31) uitgeschreven, dan volgt

$$\rho = \frac{\rho_{gw}\sigma_g\sigma_w - k\rho_{ga}\sigma_g\sigma_a - k\rho_{aw}\sigma_a\sigma_w + k^2\sigma_a^2}{\sqrt{(\sigma_g^2 - k^2\sigma_a^2)(\sigma_w^2 - k^2\sigma_a^2)}}$$

als ρ_{gw} de c.c. tussen G en W, ρ_{aw} de c.c. tussen A en W, ρ_{ga} de c.c. tussen G en A voorstelt.

Laat $\overline{GA}$ groter worden (station A verder). Dan zullen de σ_g, σ_w niet veranderen, daar ze alleen met de g-reeks te doen hebben.

Wel de ρ_{aw}, ρ_{gw}, σ_a en k, doch laten wij wèl onderstellen dat σ_a constant is. De ρ_{aw} en ρ_{gw} zullen wel afnemen, (minder correlatie tussen de stations), daarmee k eveneens (op den duur $k = 0$). Wat doet de gehele breuk ρ ? Het valt niet gemakkelijk te zeggen, hoewel $\rho(\overline{GA} \to \infty)$ naar ρ_{gw} moet tenderen. Wanneer het juist zou zijn dat ρ met toenemende $\overline{GA}$ toeneemt, zal op grote afstand de correctieformule het winnen, op kleine afstand de schattingsformule.

Nadere bestudering.

Ga uit van een situatie $\sigma_{*}^{2} < \sigma_{**}^{2} < \sigma^{2}$, zie (37). Laat $\overline{GA}$ toenemen; houd daarbij T_g en T_w vast, hetgeen kan. De ρ zal toenemen én op zekere afstand tussen G en A boven het rechterlid in (37) uitgroeien. Dan wordt $\sigma_{**}^{2} > \sigma^{2}$ en vervalt daarom al de schattingsformule. Hoe met g^{*}?
Voorlopig is $\sigma_{*}^{2} < \sigma^{2}$ omdat de gelijkheid in (35) geldt, zodat corrigeren het meest zinvol is. Totdat $\overline{GA}$ zodanig groot is, dat ook de ongelijkheid in (35) niet meer geldt en, zie (33), noch schatten, noch corrigeren zinvol is. Aangezien de functie $\rho = \rho(\overline{GA})$ onbekend is, is dit alles zeer moeilijk in formule te brengen. Toch zou hierin een criterium t.a.v. de afstand inzake de vraag "al of niet met de incorrecte metingen werken" gelegen kunnen zijn.

7. Appendix.

7.1 Een opmerking bij 4.2

Men kan natuurlijk ook die $\hat{k}_1$ en $\hat{l}_1$-waarden berekenen, waarvoor de som $\sum_1^m \left[G_{1i} - (\hat{k}_1 A_{1i} + \hat{l}_1)\right]^2$ minimaal is en dan toch verder werken met het gemiddelde $\frac{1}{m} \sum_1^m (G_{1i} - \hat{k}_1 A_{1i})$, hetwelk, in tegenstelling tot $\frac{1}{m} \sum_1^m \left[G_{1i} - (\hat{k}_1 A_{1i} + \hat{l}_1)\right]$, nu niet identiek nul is. Zoals bekend, geldt:

$$\hat{k}_1 = \sum (A_{1i} - \overline{A}_{1i})(G_{1i} - \overline{G}_{1i}) \Big/ \sqrt{\sum (A_{1i} - \overline{A}_{1i})^2 \ \sum (G_{1i} - \overline{G}_{1i})^2} \text{ en } \hat{l}_1 = \overline{G}_{1i} - \hat{k}_1 \overline{A}_{1i} .$$

Dit ook doende voor deelvak II, kan men daarna onderzoeken of deze twee gemiddelden significant verschillen. Bij een niet-parametervrije toets (bijv. de Student-toets) zullen dan zeker ook twee varianties $\frac{1}{m-1} \sum_1^m (G_{1i} - \hat{k}_1 A_{1i})^2$ en $\frac{1}{n-1} \sum_1^n (G_{2i} - \hat{k}_2 A_{2i})^2$ beschouwd moeten worden, die nu, zoals reeds gezegd werd, groter dan $\frac{1}{m-1} \sum_1^m \left[G_{1i} - (\hat{k}_1 A_{1i} + \hat{l}_1)\right]^2$ resp. $\frac{1}{n-1} \sum_1^n \left[G_{2i} - (\hat{k}_2 A_{2i} + \hat{l}_2)\right]^2$ zijn. Toch lijkt me dit geen elegante handelwijze, men haalt immers twee procedures door elkaar.

7.2 <u>De onbepaaldheid van het probleem.</u>

De redenen waarom de gegeven beschouwingen ons niet volledig bevredigen, liggen in de ongedefinieerdheid van het begrip <u>"quasiconstante"</u>. Men kan vaag zeggen: "een grootheid, die waarden aanneemt, welke weinig rondom het gemiddelde schommelen, heet een quasiconstante". In § 1 verscherpten wij dit door te eisen, dat de spreiding en het gemiddelde overal in de reeks dezelfde zijn. Een stap verder: de grootheid moet een normale verdeling volgen: $N(\mu_i; \sigma_i)$, met μ_i en σ_i voor elke i dezelfde. Algemener: de quasiconstante grootheid moet een van i onafhankelijke verdeling F_i volgen en <u>eigenlijk zouden wij moeten verifiëren of aan deze eis voldaan wordt!</u>

Zodra terwille van een toets gewenst is, dat deze quasi-constante grootheid in kwestie een normale verdeling volgt, staan wij voor consequenties als deze:

a) Als de synchrone x_i- en y_i-reeksen relatief homogeen zijn in die zin, dat het verschil $x_i - y_i$ een van i onafhankelijke normale verdeling volgt (waaruit geenszins persé volgt, dat én x_i én y_i normaal verdeeld is), kan niet het quotiënt x_i/y_i tegelijkertijd normaal verdeeld zijn. Het omgekeerde geldt evenmin.

b) Als x_i en y_i gecorreleerd binormaal verdeeld zijn en zelfs zodanig, dat de normale verdeling, die door het verschil $y_i-(\alpha x_i + \beta)$ (waarin $\alpha = \rho \sigma_y / \sigma_x$ en $\beta = \mu_y - \alpha . \mu_x$) dientengevolge gevolgd wordt, van i onafhankelijk is, dan zijn $X_i = \ln x_i$ en $Y_i = \ln y_i$ zeker niet binormaal verdeeld en zijn er geen constanten k en ℓ te vinden, waardoor $Y_i-(kX_i + \ell)$ normaal verdeeld is en zelfs onafhankelijk van i.

Dikwijls zijn de gegevens te schaars om statistisch zeker te kunnen vaststellen dat het een waar is en het andere niet. De transformaties $x^* = f(x)$ en $y^* = g(y)$, waarvan in de meest algemene definitie van "relatieve homogeniteit" sprake is (zie onder 1.6) zijn dientengevolge meestal niet éénduidig vast te stellen (tenzij er zeer duidelijke fysische aanwijzingen zijn, doch dit is zelden het geval), anders gezegd: er zullen meestal verschillende zulke transformaties zijn, die de reeks $x_i^* - y_i^*$ homogeniseren. Wil men tussen hen kiezen, dan geeft men natuurlijk de voorkeur aan de eenvoudigste.

7.3 Andere grootheden.

De voorgaande beschouwingen zijn niet slechts van toepassing op neerslagreeksen, doch geheel analoog ook op klimatologische reeksen van temperatuur, zonneschijn, luchtdruk, enz. Nu eens zijn er redenen om de Q-methode (meestal in de niet gegeneraliseerde vorm) toe te passen, dan weer verdient de V-methode (meestal eveneens in de niet gegeneraliseerde vorm) de voorkeur. Overigens mag men dikwijls de Q-methode vervangen door de V-methode, mits toegepast op de logarithmen.

Hoe een en ander is voor discreet verdeelde grootheden (bijv. het aantal regendagen per maand, het aantal zonnige dagen per decade, enz.) zou nader bestudeerd moeten worden.

8. Numerieke voorbeelden.

Tabellen

Het is via de stationsbeschrijving bekend, dat te Groningen (G) vóór het einde van 1905 de regenmetingen slechter zijn dan die er na; anders gezegd: de gehele meetreeks bestaat uit twee niet verenigbare gedeelten: 1840-1905 ($p+n=66$ jaren) en 1906-1953 ($m =$ 48 jaren).

Het dichtstbijgelegen regenstation, dat én over een zeer lange reeks goede metingen beschikt én eveneens geponst is, is Leeuwarden (A), 1876-1953. Het plan rees Leeuwarden als referentiestation te gebruiken, om daarmede Groningen te "corrigeren", zowel wat het gedeelte III ($p = 37$ jaren), dat geen parallelmetingen te A heeft, als wat het gedeelte II($n=30$ jaren) betreft. Met "corrigeren" wordt dan bedoeld: "echt corrigeren" in III en òf gebruikmaken van de correctieformule, òf van de schattingsformule in II. Wij zouden dit alles hebben willen doen aan de synchrone dagsommen, doch kozen, louter ter oriëntering, de maandsommen januari en juli. Dit kwam neer op

A. januari: II: $n = 26$ maanden G en A; I:$m=48$ maanden G en A.
B. juli : II: $n = 29$ " G en A; I:$m=47$ " G en A.

In de tabellen 1 en 2 geven wij een globaal overzicht der resultaten. Daarbij is $\bar{a}^m$ het gemiddelde van m maandsommen; idem $\bar{g}^m$ enz.
$A = \log a$ enz. $\hat{g}$ staat voor minimumkwadraten-schatting.
t = de Student-statistic in de t-toets van Student.
s_Δ = de standaarddeviatie van het verschil van twee onafhankelijke steekproefgemiddelden.

Enig commentaar.

De A-reeks schijnt zonder trend te zijn, want de toets inzake de nulhypothese H_0, dat de twee steekproeven, de ene van 29 elementen (29 juli-sommen) en de andere van 47 elementen (47 juli-sommen) uit identieke normale verdelingen komen, leidt niet tot verwerping van H_0. Via grafiekjes op lineair gauszisch papier bleek ons, louter visueel, dat de populatieverdelingen normaal kunnen zijn. Hetzelfde voor de januarisommen te A. Nu is deze toets eigenlijk niet een trend toets in de volledige zin van het woord, maar wij oordeelden dit voldoende.

Aan de Minimumkwadratenmethode is inhaerent, dat in tabel 1 de regressie onder $E\underset{\sim}{\xi} = ka + \ell$ de kleinste restvariantie moet opleveren.

Wij zien: in juli, in I: 27.3^2 is de kleinste van 28.4^2; 27.3^2 en 29.6^2 en in II: 24.1^2 de "kleinste" van 24.1^2, 24.1^2 en 26.2^2; in januari: in I: 10.6^2 de kleinste van 10.7^2; 10.6^2 en 12.4^2 en in II: 7.8^2 de kleinste van 10.0^2; 7.8^2 en 7.9^2. Het analoge doet zich voor in tabel 2 onder de A,G-regressies.

Het is geen noodzaak, dat zich ditzelfde dan ook voordoet onder de "vertalingen". Met "vertaling" wordt bedoeld: vinden wij bijv. $EG = 0.591\ A + 0.760$ (zie juli; II), dan schrijven wij ook $E\mathrm{g} = 5.76\ a^{0.591}$, maar wij zeiden reeds, dat het niet per se deze speciale waarden zijn, 5.76 en 0.591, die $\sum\left[g/\ell a^{k}\right]^2$ minimaal maken. Daaraan zal het wel moeten worden toegeschreven (doch bovendien is er nog het steekproefeffect; steekproeven van 26 à 48 zijn nog niet bijzonder groot te noemen bij zulk een sterke variërende grootheid als de maandsom neerslag), dat wij in de 4 gevallen (jan.I en II; jul.I en II) zien:

0.374^2 is de kleinste van 0.560^2, 0.374^2 en 0.560^2 (jul.II) en 0.363^2 de kleinste van 0.440^2, 0.363^2 en 0.418^2 (jul.I), doch 0.184^2 is niet de kleinste van 0.171^2, 0.184^2 en 0.210^2 (jan.II) en 0.155^2 niet van 0.164^2, 0.155^2 en 0.151^2 (jan.I).

Corrigeren of schatten?

Wij dienen tenslotte in G II te "handelen"

in juli

òf volgens de correctieformule (18): $g^{*}_{2i} = g_{2i}\dfrac{\hat{q}_1}{\hat{q}_2} = g_{2i}\dfrac{1.085}{1.143} = \boxed{0.95 g_{2i}}$ $(1.03\ g_{2i})$

òf volgens de schattingsformule (19): $g^{**}_{2i} = \hat{q}_1 . a_{2i}^{k_2} = \boxed{1.08\ a_{2i}}$ $(1.18\ a_{2i})$

, waarin $k_2 = 0.9998 \approx 1$

in jan.

òf volgens: $g^{*}_{2i} = g_{2i}\dfrac{\hat{q}_1}{\hat{q}_2} = g_{2i}\dfrac{0.985}{1.012} = \boxed{0.96\ g_{2i}}$ $(1.29\ g_{2i})$

òf volgens: $g^{**}_{2i} = q_1 . a_{2i}^{k_2} = \boxed{0.98\ a_{2i}^{0.945}}$ $(1.05\ a_{2i})$

Tussen haakjes zijn de uitdrukkingen geplaatst, werkende met de simpelste quotiënten, t.w. $q = g/a$, zoals in de meeste leerboeken geschiedt: $1.030 = 1.178/1.143$; $1.29 = 1.054/0.817$. Opvallend zijn de verschillen tussen 0.95 en 1.03 (jul) en tussen 0.96 en 1.29 (jan).

Nu nog de moeilijke keuze tussen de drie alternatieven: metingen onveranderd laten, corrigeren, schatten? Zie de criteria (37),(38) en (39).

In de ontwikkelde formules treden helaas onbekende grootheden op en toch moeten wij numerieke waarden substitueren, alvorens de criteria toegepast kunnen worden. Wij mogen waarschijnlijk stellen, dat "minder correcte metingen" een grotere $\sigma(V_{g2i})$ impliceert, zodat, hoe slechter de metingen zijn, hoe groter T_g zal zijn; T_w vast denkende, impliceert dit een grotere F. Louter ter illustratie proberen wij eens:

A) $F = 1$ (d.i. $T_g = T_w$) B) $F = 1.5$ (d.i. $T_g = 1\frac{1}{2}\,T_w$) en C) $F = 2$ (d.i. $T_g = 2\,T_w$) en $m = 48$, $n = 26$, $T_g = \sigma(g_2)/\sigma(g_1) = 0.1187/0.1077 = 1.10$ (jan.)

De ρ is onbekend; hij stelt voor de c.c. tussen

$V_{g2i} \equiv G_{2i} - k_2A_{2i} = G_{2i} - 0.945\,A_{2i}$ en $V_{w2i} \equiv W_{2i} - 0.945A_{2i}$

($i = 1, 2 \ldots 26$).

Van de ρ zouden wij een schatting hebben kunnen maken, als gedurende enige jaren te Groningen ook goede metingen verricht zouden zijn en wel op een zo nabij gelegen plaats, dat deze goede metingen de ware waarden op de plek, waar de slechte metingen plaatsvinden, voldoende goed representeren. Wij voelen ook aan, dat ρ en T_g niet vrij van elkaar gekozen zullen mogen worden. Afneming van T_g zal wel toeneming van ρ impliceren.

In bijgaande grafische voorstelling hebben wij ρ laten veranderen van 0 tot 1 en successieve intervallen opgezocht, waarin de ρ willekeurig mag veranderen en toch de grootte-volgorde van σ, σ_* en σ_{**} dezelfde blijft.

In het traject 0.28 à 0.33 bijv. is in geval A $\sigma_{**} < \sigma < \sigma_*$ (dus schattingsformule gebruiken), in geval B idem en in geval C $\sigma_* < \sigma < \sigma_{**}$ (dus correctieformule gebruiken).

De grafiek laat zien o.m., dat <u>voor kleine ρ</u>, zegge < 0.3, <u>meestentijds de schattingsformule het van de correctieformule "wint" en dat</u> het <u>zinvol</u> is überhaupt <u>de metingen g_{2i} te "behandelen"</u>. Niet zinvol is de behandeling voor grote ρ's; dan zijn blijkbaar de metingen niet slecht genoeg. Er zitten helaas onverklaarbare aspecten in de grafiek, zoals én $\sigma_* < \sigma_{**}$ én $\sigma_* > \sigma$ én $\sigma_{**} < \sigma$, aan welke ongelijkheden niet tegelijk voldaan kan worden.

TABEL I

waarnemingen zelf

juli-maanden	II : n = 29		I : m = 47			
mm	mm	mm		mm	mm	
gem.	st.dev.	rest-st.dev.	gem.	st.dev.	rest-st.dev.	Toetsingen, met opmerkingen
$\bar{a} = 76.9$ $\bar{g} = 77.1$ $\overline{g-a} = 0.2$	$s_a = 37.4$ $s_g = 34.9$ $s_{g-a} = 27.1$		73.1 81.4 7.3	35.0 39.0 28.4		Trend? 73.1 tegen 76.9. Student-toets $s_\Delta = 8.7$ en $t = (76.9-73.1):8.7 = 0.44 \ll 2$. Geen trend. Ook niet in G-reeks.
$\hat{g} \equiv Eg = a$ $\hat{g} \equiv Eg = 0.672a + 25.5$ $\hat{g} \equiv Eg = 0.945a$	$\overline{g-\hat{g}} = 0.2$ $\overline{g-\hat{g}} = 0$ $\overline{g-\hat{g}} = 4.5$	(rest) = 24.1 24.1 26.2	a 0.791a + 22.5 1.046a	7.3 $\equiv 0$ 3.9	28.4 27.3 29.6	Relatief homogeen? 7.3 tegen 0.2; t-toets: $s_\Delta = 6.8$; $t = (7.3-0.2):6.8 = 1.0 < 2$; signifikant relatief homogeen. Var. 24.1^2 en 28.4^2 verschillen niet. Varianties 27.3^2 en 24.1^2 verschillen niet signifikant, want $F = 27.3^2/24.1^2 = 1.28$ en $F(5\%) \cong 1.9$
$r(a,g) = 0.720$			0.710			
januari-maanden	II : n = 26		II : m = 40			
mm	mm	mm	mm	mm	mm	
$\bar{a} = 54.5$ $\bar{g} = 44.3$ $\overline{g-a} = -10.2$	$s_a = 26.8$ $s_g = 22.0$ $s_{g-a} = 10.0$		58.8 62.1 3.3	27.0 29.4 10.7		Trend? 58.8 tegen 54.5; $s_\Delta = 6.6$ $\therefore t = 0.7 \ll 2$. In A geen trend; ook 62.1 tegen 44.3; $s_\Delta = 7.7$ $\therefore t = 2.3 > 2$. Wel trend, echter vanwege breuk in homogeniteit. Relatief homogeen? 3.3 tegen −10.2; $s_\Delta = 2.54$ $\therefore t = 5.3 \gg 2$. Signifikante breuk in relatieve homogeniteit; Var. 10.0^2 en 10.7^2 verschillen niet.
$\hat{g} \equiv Eg = a$ $\hat{g} \equiv Eg = 0.767a + 2.5$ $\hat{g} \equiv Eg = 0.804a$	$\overline{g-\hat{g}} = -10.2$ $\equiv 0$ $= 0.5$	(rest) = 10.0 7.8 7.9	a 1.015a + 2.5 1.049a	3.3 $\equiv 0$ 0.5	10.7 10.6 12.4	Var. 7.8^2 en 10.6^2 verschillen bijna sign., want $F = 10.6^2:7.8^2 = 1.85$ en $F(5\%) \cong 1.9$ 0.5 en 0.5 verschillen natuurlijk niet signifikant.
$r(a,g) = 0.935$			$r(a,g) = 0.932$			

logarithmen der waarnemingen

juli-maanden	II : n = 29		I : m = 47			
gem.	st.dev.	rest-st.dev	gem.	st.dev.	rest st.dev.	Toetsingen, met opmerkingen
$\bar{A}$ = 1.822	s_A = 0.261		1.806	0.243		
$\bar{G}$ = 1.837	s_G = 0.223		1.849	0.236		
$\overline{G-A}$ = 0.015	s_{G-A} = 0.193		0.043	0.159		
$\hat{G} \equiv EG = A$	$\overline{G-\hat{G}}$ = 0.015	s(rest) = 0.193	A	0.043	0.159	Relatief homogeen? Vergelijk 0.015 met 0.043. s_Δ = 0.100 ∴ t = 0.26 ≪ 2; signifikant relatief homogeen.
$\hat{G} \equiv \mathcal{E}G = 0.591\,A + 0.760$	= 0	= 0.161	0.758A + 0.482	= 0	0.160	Relatief homogeen? Vergelijk 0.015 met 0.006; s_Δ = 0.137 ∴ t = (0.015 − 0.006) : 0.137 = 0.07 ≪ 2; signifikant relatief homogeen.
$\hat{G} \equiv \mathcal{E}G = 0.9998A \approx A$	= 0.015	= 0.193	1.020A	0.006	0.178	
k_2 = 0.9998	r (G,A) = 0.690 ≠ r(g,a) = 0.720		r(G,A) = 0.730 ≠ r(g,a) = 0.710			
	V E R T A A L D T O T					
$\hat{g} \equiv \mathcal{E}g = a$	$\overline{g/a}$ = 1.143 ($\frac{\bar{g}}{\bar{a}}$ = 1.013)	s = 0.560	a	1.178 (1.100)	0.440	Verschilt 1.143 van 1.178? s_Δ = 0.364; t = 0.1 ≪ 2. Geen signifikant verschil.
$\hat{g} \equiv \mathcal{E}g = 5.76a^{0.591}$	$\overline{g/5.76a^{0.591}}$ = 1.063	= 0.374	$3.034a^{0.758}$	1.052	0.363	Tussen 1.063 en 1.052 geen signifikant verschil.
$\hat{g} \equiv \mathcal{E}g = a^{0.9998}$	$\hat{q}_2 = \overline{g/a^{0.9998}}$ = 1.143	= 0.560	$a^{1.020}$	$\hat{q}_1$ = 1.085	0.410	Tussen 1.063 en 1.052 geen signifikant verschil. Evenmin tussen 1.143 en 1.085.

januari maanden	II n = 26		m = 48			
$\bar{A}$ = 1.0844	s_A = 0.224		1.723	0.206		Verschil tussen −0.0970 en 0.0178? s_Δ = 0.60; t = (0.0178 + 0.0970) : 0.60 = 0.19 ≪ 2. Verschil niet signifikant.
$\bar{G}$ = 1.5875	s_G = 0.246		1.745	0.222		Varianties 0.092^2 en 0.068^2 verschillend? F = 1.83; F (5%) = 1.83 Suggestief signifikant.
$\overline{G-A}$ = −0.0970	s_{G-A} = 0.092		0.0178	0.068		
$\hat{G} \equiv \mathcal{E}G = A$	$\overline{G-\hat{G}}$ = −0.0970	s(rest) = 0.092	A	0.0178	0.068	Verschillen −0.0045 en −0.0070? s_Δ = 0.0272; t = 0.1 ≪ 2. Geen signifikant verschil
$\hat{G} \equiv \mathcal{E}G = 1.021A - 0.0624$	= 0	0.092	1.026A − 0.024	= 0	0.068	
$\hat{G} \equiv \mathcal{E}G = 0.945\,A$	= −0.0045	0.119	1.017A	−0.0070	0.108	
k_2 = 0.945	r(G,A) = 0.928 ≠ r(g,a) = 0.935		r(G,A) = 0.952 ≠ r(g,a) = 0.932			
	V E R T A A L D T O T					
$\hat{g} \equiv \mathcal{E}g = a$	$\overline{(g/a)}$ = 0.817 ($\frac{\bar{g}}{\bar{a}}$ = 0.812)	s = 0.171	a	1.054 (1.059)	0.164	Verschil tussen $\overline{(g/a)}$ = 0.817 in II en 1.054 in I? t ≅ 6. Zeer signifikant. Echter volgen de g/a-waarden een norm. verd.? Verschil tussen $s^2 = 0.184^2$ en 0.155^2 (II en I)? Neen, want F = $0.184^2/0.155^2$ = 1.41 en F(5%) = 1.75.
$\hat{g} \equiv \mathcal{E}g = 0.866 \cdot a^{1.021}$	$\overline{g/0.866 \cdot a^{1.021}}$ = 0.870	= 0.184	$0.95a^{1.026}$	1.001	0.155	Verschil tussen 0.870 (II) en 1.001 (I)? s_Δ = 0.040 en t ≅ 3 > 2. Zeer signifikant. Verschil tussen 0.210^2 en 0.151^2? Ja, want F = $0.210^2 : 0.151^2$ = 1.93 en F(5%) = 1.75.
$\hat{g} \equiv \mathcal{E}g = a^{0.945}$	$\hat{q}_2 = \overline{g/a^{0.945}}$ = 1.012	= 0.210	$a^{1.017}$	$\hat{q}_1$ = 0.985	0.151	Verschil tussen 1.012 en 0.985? Varianties poolen. Verschil blijkt niet sign.

ρ		$\sigma_*\ \sigma_{**}$	$\sigma\ \sigma_{**}$	$\sigma\ \sigma_*$		Wat te doen?
0.0–1.0	A	>	<	<	$\sigma_{**}<\sigma_*<\sigma$	schatten of corrige
	B	>	<	<	idem	idem
	C	>	<	<	idem	idem
	A	>	>	<	$\sigma_{**}<\sigma<\sigma_*$	schatten
	B	>	<	<	$\sigma_{**}<\sigma_*<\sigma$	schatten of corr
	C	>	<	<	idem	idem
	A	>	>	<	$\sigma_{**}<\sigma<\sigma_*$	schatten
	B	>	>	<	idem	idem
	C	>	<	<	$\sigma_{**}<\sigma_*<\sigma$	schatten of corr
	A	>	>	<	$\sigma_{**}<\sigma<\sigma_*$	schatten
	B	>	>	<	idem	idem
	C	<	<	<	$\sigma_*<\sigma_{**}<\sigma$	corr. of schatte
	A	>	>	<	$\sigma_{**}<\sigma<\sigma_*$	schatten
	B	>	>	<	idem	idem
	C	<	>	<	kan niet	??
	A	>	>	<	$\sigma_{**}<\sigma<\sigma_*$	schatten
	B	<	>	<	kan niet	??
	C	<	>	<	kan niet	??
	A	<	>	<	kan niet	??
	B	<	>	<	kan niet	??
	C	<	>	<	kan niet	??
	A	<	>	>	$\sigma<\sigma_*<\sigma_{**}$	metingen zelf
	B	<	>	<	kan niet	??
	C	<	>	<	kan niet	??
	A	<	>	>	kan niet	??
	B	<	>	>	kan niet	??
	C	<	>	<	kan niet	??
	A	<	>	>	$\sigma<\sigma_*<\sigma_{**}$	metingen zelf
	B	<	>	>	idem	idem

ρ: 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

januari

m = 48 in I
n = 26 in II
$T_g = 1.10$

Van A→B→C de ρ kleiner, d.i. de metingen g_{2i} slechter, d.i. $F = 1 \to 1\frac{1}{2} \to 2$, als $F = T_g/T_w$ en

$T_g = \sigma_{g2}/\sigma_{g1}$ en $T_w = \sigma'_{w2}/\sigma_{g1}$

Dit speciale voorbeeld laat zien:
als betrekkelijk slechte metingen een $\rho < \frac{1}{2}$ impliceren, dan lijkt veelal schatten het beste

Typisch, dat in dit voorbeeld weinig de correctieformule geprefereerd moet worden.

Hoe minder slecht de metingen, hoe minder zinvol én corr. én schatting.

Dan metingen zelf gebruiken.