KONINKLIJK NEDERLANDS
METEOROLOGISCH INSTITUUT

Verslagen V-39
(R III-223-1958)

De mogelijkheid van tijdsbesparing bij de berekening van maand- en decade- gemiddelden uit uurwaarnemingen

door

Drs. P.J. Rijkoort

0. Inleiding.

De vraag, die in dit rapport beantwoord zal worden, luidt:

Is het noodzakelijk om bij de berekening van maand- of decade- gemiddelden uit uurwaarnemingen alle uurwaarnemingen te gebruiken, of is het misschien voldoende slechts een gedeelte te benutten door de waarnemingen die een zeker aantal (p) uren uiteen liggen te gebruiken?

Het is duidelijk, dat we de vraag slechts dan kunnen beantwoorden en de waarden van p aangeven, als vastgesteld is welke onnauwkeurigheid in de bepaling van het gemiddelde getolereerd kan worden. De waarde van p zal dan tenslotte afhangen van de standaarddeviatie van de uurwaarnemingen zelf en van de waarden van de autocorrelatiecoëfficiënten.

Als we de beschouwde grootheid met x_i aangeven en het totale aantal uurwaarden met N, dan kunnen we met $\bar{x}^N = \frac{1}{N} \sum_{i=1}^{N} x_i$ het gemiddelde over alle N uurwaarden aanduiden en met $\bar{x}^M = \frac{1}{M} \sum_{i=1}^{M} x_{i.p.}$ (met $M = N/p$) het gemiddelde over het beperkte aantal uurwaarden. De nauwkeurigheid van $\bar{x}^M$ t.o.v. $\bar{x}^N$ is gedefinieerd door de variantie van het verschil, n.l.: $\sigma_v^2 \equiv \sigma^2(\bar{x}^N - \bar{x}^M)$

1. Afleiding van de formule voor de variantie σ_v^2.

Om een eenvoudige berekening mogelijk te maken, moeten we onderstellen, dat de reeks uurwaarnemingen z.g. zuiver persistent is d.w.z. $\rho_k = \rho_1{}^k$. Hier in is k de correlatiecoëfficiënt tussen de temperatuur op het uur j en op het uur j + k(zie W.R. 55-004 (R III-150) van (I) Dr. C. Levert). Deze onderstelling is althans voor kleine waarden van k in eerste instantie wel juist; bij grotere waarden van k zal, bij aanwezigheid van een dagelijkse gang, deze onderstelling echter niet meer opgaan. Het is dus gewenst de dagelijkse gang te elimineren. Deze eliminatie zal in de volgende paragraaf nader besproken worden.

Na de eliminatie van de dagelijkse gang, waardoor aan

$$E\, \underline{x}_i = c \qquad \text{(II)}$$

voor elke i is voldaan, onderstellen we tenslotte nog:

$$\sigma_{x_i} = \sigma_{x_j} \; (= \sigma_x) \text{ voor alle i en j.} \qquad \text{(III)}$$

De berekening van σ_v^2 verloopt als volgt:

$$\sigma_v^2 = E(\bar{x}^N - \bar{x}^M)^2 = \frac{1}{N^2} E\left\{(\sum^N x_i)^2 + p^2(\sum^M x_{ip})^2 - 2p \sum^N x_i \sum^M x_{ip}\right\} =$$

$$\frac{1}{N^2} E\left\{\sum^N x_i^2 + \sum_{i \neq j}^N x_i x_j + p^2 \sum^M x_{ip}^2 + p^2 \sum_{ip \neq jp}^M x_{ip} x_{jp} - 2p \sum_{j=1}^M \sum_{i=1}^N x_i x_{jp}\right. \qquad (1)$$

De verwachtingswaarden van de vijf sommen berekenen we achtereenvolgens als volgt:

$$E \sum^N x_i^2 = N E x_i^2 = N \sigma_x^2 \qquad (2)$$

$$E \sum^N x_i x_j = 2 \sum_{q=1}^N (N-q) E x_i x_{i+q} = 2 \sum_{q=1}^N (N-q) \rho^q \sigma x^2 \qquad (3)$$

$$E \sum^M x_{ip}^2 = \sum^M x_{ip}^2 = M E x_{ip}^2 = M \sigma_x^2 \qquad (4)$$

$$E \sum_{i+j}^M x_{ip} x_{jp} = 2 \sum_{q=1}^{N/p} (\frac{N}{p} - q) \rho^{pq} \sigma_x^2 \qquad (5)$$

en tenslotte:

$$E \sum_{i=1}^N \sum_{j=1}^M x_i x_{jp} = \sigma_x^2 \left\{ 1 - N/p + (2N/p-1)(1+\rho + \dots + \rho^{p-1}) + \rho^p \right.$$
$$+ (2N/p-3)(\rho^h + \dots + \rho^{2p-1}) + \rho^{2p}$$
$$\left. + \text{— — — — — —} \right\}$$

$$= \sigma_x^2 \left\{ 1 - N/p + \sum_{q=1}^N \rho^{qp} + \sum_{S=0}^{N/p-1} (2\,N/p-1-2\,S) \sum_{q=0}^{p-1} \rho^{Sp+q} \right. \qquad (6)$$

Nu is $\sum_{q=\alpha}^B q\rho^q = \frac{\rho^x - \rho^{B+1}}{1-\rho}$ en

$$\sum_{q=\alpha}^B q\,\rho^q = \frac{\alpha\rho^x - (\alpha-1)\,\rho^{\beta+1} - (\beta+1)\rho^{\beta+1} + \beta\rho^{\beta+2}}{(1-\rho)^2}$$

In onze formule is β = N of M en deze waarde is in het algemeen groot, zodat $\rho^{\beta+1}$ enz. te verwaarlozen is t.o.v. ρ^α, omdat α = 1 of 0 is.

We vinden achtereenvolgens:

$$E \sum_{i \neq j}^{N} x_i \, x_j = \left(2N \sum_{q=1}^{N} \rho^{q} - 2 \sum_{q=1}^{N} q \, \rho^{q}\right) \sigma_x^{\,2} = \left(\frac{2N\rho}{1-\rho} - \frac{2\rho}{(1-\rho)^2}\right) \sigma_x^{\,2} \tag{7}$$

$$E \sum^{M} x_{ip} \, x_{jp} = \left(2\frac{N}{p} \; \frac{\rho^p}{1-\rho^p} - \frac{2\rho^p}{(1-\rho^p)^2}\right) \sigma_x^{\,2} \tag{8}$$

en

$$E \sum_{i=1}^{N} \sum_{j=1}^{N/p} x_i \, x_{jp} = \left\{1 - N/p + \frac{\rho^p}{1-\rho^p} + \left(2\frac{N}{p} - 1\right) \sum_{q=0} \rho^{q} - 2 \sum_{S=1} \rho^{Sp} \sum_{q=0} \rho^{q}\right\} \sigma_x^{\,2} =$$

$$\left\{1 - N/p + \frac{\rho^p}{1-\rho^p} + (2\,N/p - 1)\frac{1}{1-\rho} - 2\frac{1}{1-\rho} \cdot \frac{\rho^p}{1-\rho^p}\right\} \sigma_x^{\,2} = \left(\frac{1+\rho}{1-\rho} \, \frac{N}{p} - \frac{\rho + \rho^p}{(1-\rho)(1-\rho^p)}\right) \sigma_x^{\,2} \tag{9}$$

Substitutie van (2), (7), (4), (8) en (9) in (1) geeft:

$$\sigma_v^{\,2} = \sigma^2(\bar{x}^N - \bar{x}^M) = \frac{\sigma_x^2}{N^2} \left\{ N\left(1 + p - \frac{2(1+\rho)}{1-\rho}\right) + \frac{2N\rho}{1-\rho} - \frac{2\rho}{(1-\rho)^2} + \frac{2pN\rho^p}{1-\rho^p} - \frac{2p^2\rho^p}{(1-\rho^p)^2} + \frac{2p(\rho+\rho^p)}{(1-\rho)(1-\rho^p)} \right\} =$$

$$\frac{\sigma_x^2}{N^2} \left\{ N\left(1 + p - \frac{2(1+\rho)}{1-\rho} + \frac{2\rho}{1-\rho} + \frac{2p\rho^h}{1-\rho^h}\right) - \left(\frac{2\rho}{(1-\rho)^2} + \frac{2p^2\rho^p}{(1-\rho^p)^2} - \frac{2p(\rho+\rho^p)}{(1-\rho)(1-\rho^p)}\right) \right\} =$$

$$\frac{\sigma_x^2}{N} \left\{ \left(p\,\frac{1+\rho^h}{1-\rho^h} - \frac{1+\rho}{1-\rho}\right) - \frac{2}{N}\left(\frac{\rho}{(1-\rho)^2} - p\frac{\rho + \rho^p}{(1-\rho)(1-\rho^p)} + p^2 \frac{\rho^p}{(1-\rho^p)^2}\right) \right\} = \frac{\sigma_x^2}{N} A(p,\rho,N) \tag{10}$$

of

$$A\,(p,\rho,N) = \frac{N\,\sigma_v^{\,2}}{\sigma_x^{\,2}} \tag{10*}$$

In fig. 1 is de functie A (p,ρ,N) in grafiek weergegeven voor N=720 en N = 240. Dit zijn resp. de gevallen van maand- en decade- gemiddelden uit uurwaarnemingen met de maand op 30 en de decade op 10 dagen gesteld.

Als $\sigma_x^{\,2}$ en ρ bekend zijn er voor $\sigma_v^{\,2}$ een zekere waarde gëeist wordt, waardoor A vastligt, kan uit fig. 1 afgelezen worden welke waarde voor p genomen moet worden.

3. <u>De eliminatie van de dagelijkse gang.</u>

Nemen we aan, dat de dagelijkse gang gemiddeld over de gehele maand of decade is te schrijven als

$$\sum_{n=1} a_n \cos\left(\frac{2\,n\,\pi}{T}\,t + \varphi_n\right)$$

(T is de lengte van de dag)

dan is de beschouwde meteorologische grootheid θ te schrijven als

$$\underline{\theta} = \underline{x} + \sum_{n=1}^{k} a_n \cos\left(\frac{2\,n\,\pi}{T}\,t + \varphi_n\right) = \underline{x} + \alpha(t) \text{ met b.v. } t=1,2,3 \ldots \text{ en } T=24.$$

Voor $\underline{x}$ kunnen we onderstellen, dat aan de in 2 gestelde voorwaarden (I) en (II) is voldaan. In de praktijk zullen we voor $\alpha(t)$ bij T=24 en t=1, ..., 24 de gemiddelden per uurvak over maand resp. decade gebruiken.

De ρ en σ^2 moeten dus niet voor de grootheid θ zelf worden berekend, maar voor de grootheid $x = \theta - \alpha$.

Voor $\underline{x}$ wordt nu de waarde van p bepaald. Opdat tenslotte $\bar{\theta}^M = \bar{x}^M$ zal zijn, moet $\bar{\alpha}^M = 0$ zijn.

In 't algemeen is dit, als t = ip en i=1, ... , M, wel het geval. Immers

$$\bar{\alpha}^M = \frac{1}{M}\sum_{i=1}^{M} \alpha(i.p.) = \frac{1}{M}\sum_{n=1}^{k} a_n \left\{ \cos\varphi_n \sum_{i=1}^{M} \cos\frac{2n\pi p}{T}\,i - \sin\varphi_n \sum_{i=1}^{M} \sin\frac{2n\pi h}{T}\,i \right\} =$$

$$= \frac{1}{M}\sum_{n=1}^{k} \left\{ a_n \cos\varphi_n \cdot \frac{\sin\left(\frac{n\pi p}{T} + \frac{2n\,N\pi}{T}\right) - \sin\frac{n\pi p}{T}}{2\sin\frac{n\,\pi\,p}{T}} - a_n \sin\varphi_n \frac{\cos\frac{n\pi n}{T} - \cos\left(\frac{n\pi h}{T} + \frac{2nN\pi}{T}\right)}{2\sin\frac{n\,\pi\,n}{T}} \right\}$$

Daar we steeds een hele maand of decade gebruiken is N zeker een veelvoud van T en dus

$$\sin\left(\frac{n\pi h}{T} + \frac{2nN\pi}{T}\right) = \sin\frac{n\pi h}{T}\,,\ \text{resp. } \cos\left(\frac{n\pi h}{T} + \frac{2nN\pi}{T}\right) = \cos\frac{n\pi h}{T}$$

Dus geldt inderdaad $\bar{\alpha}^M = 0$, mits $\sin\frac{n\pi p}{T} \neq 0$.

Derhalve moet $p \neq C\,\frac{T}{N}$ zijn.

Bij de temperatuur bijvoorbeeld, die zoals bekend voldoende goed met de 1e en 2e harmonische kan worden benaderd, waarbij dus k = 1 en 2 is, mag p dus niet de waarde 12, 24, ... aannemen.

4. <u>Toepassingen</u>

<u>1e voorbeeld.</u>

Temperatuurgegevens van de eerste decade van juni 1956 te Castricum.

De berekende waarde S_x^2 en r van σ_x^2 en ρ zijn $S_x^2 = 4,27$ en r = 0,91.

We zullen voor σ_v^2 twee waarden kiezen n.l. 0,00625 en 0,0025. In het eerste geval is $\sigma_v = 0,025$ d.w.z. dat $\bar{x}^M$ met 95% kans hoogstens $\pm$ 0,05°C van $\bar{x}^N$ afwijkt, terwijl in het tweede geval $\bar{x}^M$ met 95% kans hoogstens $\pm$ 0,1°C afwijkt. De waarden van $A(p,\rho,N)$ zijn resp. $\frac{240 \times 0,000625}{4,27} = 0,035$ en $\frac{240 \times 0,0025}{4,27} = 0,14$.

Uit fig. 1 kunnen we de bijbehorende waarden van p aflezen. We vinden bij $\sigma_v = 0,025$ voor p bijna 2 en p = 3 voor $\sigma_v = 0,05$.

Ter vergelijking zijn in de tabel 1 de direct berekende waarden van de gemiddelden $\bar{x}^N$ en $\bar{x}^M$ aangegeven, voor p = 2,3,4,6 en 8 en voor alle drie decaden van juni 1956. Nemen we aan, dat $\sigma_x^2 = 4,27$ en $\rho = 0,91$ ook voor de decaden II en III geldt, dan kunnen we de uitkomsten vergelijken met wat we op grond van de theorie kunnen verwachten. Dit is natuurlijk slechts een globale vergelijking die geen bewijskracht heeft om de juistheid van de theoretische afleiding aan te geven.

Tabel 1

Decade	I		II		III	
	$\bar{x}$	$\bar{x}^N - \bar{x}^M$	$\bar{x}$	$\bar{x}^N - \bar{x}^M$	$\bar{x}$	$\bar{x}^N - \bar{x}^M$
N	12.93		11.29		12.23	
p=2	12.98	+ 0.05	11.28	- 0.01	12.26	+ 0.03
3	12.97-	+ 0.04	11.28	- 0.01	12.28	+ 0.05
4	13.04	+ 0.11	11.30	+ 0.01	12.33	+ 0.10
6	12.84	- 0.11	11.18	- 0.11	12.35	+ 0.12
8	13.32	+ 0.39	11.34	+ 0.05	12.39	+ 0.16

We zien dus reeds bij p = 2 een geval waar de gestelde "toelaatbare" afwijking 0.05 bereikt is. Bij de keuze van 0.1 als 95% grens zien we bij p = 4 een overschrijding van deze grens.

We kunnen deze resultaten in ieder geval wel in overeenstemming met de theorie achten.

In het geval we de maand in zijn geheel beschouwen en daarbij onderstellen, dat ook nu $\rho = 0.91$ en $\sigma_v^2 = 4,27$ geldt, dan moeten de A waarden resp. 0.10 en 0.42 zijn, waarbij volgens figuur 1 p = 2 à 3 en p = 5 behoort. Voor de maand juni 1956 te Castricum berekenen we:

Tabel 2

	$\bar{x}$	$\bar{x}^N - \bar{x}^M$
	12.14	
p=2	12.17-	+ 0.03
3	12.18	+ 0.04
4	12.23	+ 0.09
6	12.12	- 0.02
8	12.35	+ 0.21

2^e voorbeeld

De dampspanningswaarden van december 1957 te Castricum

In dit geval is van een dagelijkse gang niets te merken. We werken dus met de uurwaarden zelf.

Uit de 1e decade blijkt de autocorrelatiecoëfficiënt r = 0.975 te zijn. De standaarddeviatie is $S_x^2 = 3.41$. Nemen we weer $\sigma_v^2 = 0.000625$ en $\sigma_v^2 = 0.0025$ als nauwkeurigheids eis, dan is A resp. 0.044 en 0.176. Voor p vinden we dan resp. 3 en 6.

Tabel 3 geeft de resultaten van de directe bepaling van $\bar{x}^N$:

Tabel 3

Decade	I		II		III	
	$\bar{x}$	$\bar{x}^N - \bar{x}^M$	$\bar{x}$	$\bar{x}^N - \bar{x}^M$	$\bar{x}$	$\bar{x}^N - \bar{x}^M$
	5.45		4.59		6.33	
p=2	5.44	- 0.01	4.59	0.00	6.33	0.00
3	5.42	- 0.03	4.60	+ 0.01	6.31	- 0.02
4	5.40	- 0.05	4.61	+ 0.02	6.32	- 0.01
6	5.36	- 0.09	4.66	+ 0.07	6.33	0.00
8	5.32	- 0.13	4.64	+ 0.05	6.23	- 0.10

Voor de maandcijfers neemt A de waarden 0.132 resp. 0.528 aan. Hierbij wordt p volgens de grafiek 5 resp. ca. 10.

Tabel 4 geeft de direct berekende waarden

Tabel 4

	$\bar{x}$	$\bar{x}^N - \bar{x}^M$
	5.48	
p=2	5.48	0.00
3	5.47	- 0.01
4	5.47	- 0.01
6	5.48	0.00
8	5.42	- 0.06

Ook hier zien we, dat de resultaten in overeenstemming met de theorie geacht kunnen worden.

5. Samenvatting

De bepaling van het getal p, dat aangeeft de duur van het tijdvak (in uren) tussen 2 opvolgende waarnemingen, die gebruikt moeten worden om een voldoend betrouwbaar maand- of decade gemiddelde te berekenen, kan worden uitgevoerd met behulp van grafiek 1. Als ingangen van deze grafiek worden gebruikt de autocorrelatiecoëfficiënt ρ en de grootheid $A = \frac{N \sigma_v^2}{\sigma_x^2}$. Hierin is σ_v de van tevoren geeiste nauwkeurigheid, waarmede het "beperkt" berekende gemiddelde t.o.v. het juiste gemiddelde wordt bepaald.

Hebben we met een grootheid te maken die een duidelijke dagelijkse gang bezit dan moet de bepaling van σ_x^2 en ρ niet geschieden voor de grootheid zelf, maar voor die, welke na eliminatie van de dagelijkse gang wordt verkregen;in de praktijk wordt dit gedaan door de uurwaarden te verminderen met het gemiddelde van het beschouwde uurvak over de gehele maand resp. decade.

Enkele voorbeelden lichten de theorie toe. Uit deze voorbeelden blijkt, dat in de praktijk bij temperatuurwaarnemingen de waarde van p ongeveer 3 à 4 kan zijn en bij dampspanningswaarnemingen 6 à 10.

$$A(p,\rho,N) = \left(p\frac{1+\rho^p}{1-\rho^p} - \frac{1+\rho}{1-\rho}\right) - \frac{2}{N}\left(\frac{\rho}{(1-\rho)^2} - p\frac{\rho+\rho^p}{(1-\rho)(1-\rho^p)} + p^2\frac{\rho^p}{(1-\rho^p)^2}\right)$$

Fig. 1.

——— N=720
– – – – N=240