On the time dependence of smoothed variables

By C. L. GODSKE, *Geofysiska Institutet, Bergen, Norway*

(Manuscript received January 11, 1966)

ABSTRACT

In Meteorology one often concentrates the study not on instantaneous values of a variable, but on smoothed values, like daily, pentadic and monthly averages. When studying the time dependency of instantaneous and smoothed variables, the problem arises of fitting statistical models to observed series of data. The present paper considers which models must be applied to smoothed data when models of moving average type or autoregressive models have been introduced for the instantaneous variables. A result of the simple formulae derived is to give a warning against use of simple models for the smoothed data.

1. Introduction

The study of time series is becoming of increasing importance in geophysics. Different simple mathematical models have been introduced and applied, among others, to instantaneous values of meteorological variables. However, in many climatological investigations one applies time averages (say diurnal, pentadic, monthly) and also space averages and combined space-time averages. The question then arises about the time dependence of such smoothed variables when the instantaneous variables are described by simple time series models (moving averages, autoregressive schemes, etc.). This problem is of some interest for medium range or long range prediction problems. The simple formulae derived below, perhaps well known among statisticians, show that one has to be very careful when introducing time series models for smoothed meteorological variables. Even an extremely primitive "smoothed model" will often, for the instantaneous variable, lead to a very complicated time dependency—which may even be devoid of any simple physical meaning.

The following terminology will be systematically adopted:

ε_t denotes a "random" time series; for each integer value of t (the time unit depends on the problem) we have by definition:

$$E[\varepsilon_t]=0, \quad E[\varepsilon_t^2]=\varepsilon^2, \qquad (1.1)$$

where E, as usual, denotes expectation. Moreover, for all integer positive and negative, values of j we assume

$$E[\varepsilon_t \varepsilon_{t-j}]=0, \quad j \neq 0. \qquad (1.2)$$

x_t will be denoted the "primary" time series (the "x-series"), to be defined in different ways in the following sections. Analogously $y_{n,t}$, $z_{n,T}$, v_t denote "secondary time series" derived from x_t as shown below.

$y_{n,t}$ is defined as the "backward" moving n-average of x_t, i.e. by the equation

$$y_{n,t}=\frac{1}{n}\sum_{i=0}^{n-1} x_{t-i}. \qquad (1.3)$$

The time unit for $y_{n,t}$ is the same as for x_t. It is of interest to consider also time steps of length n for the y-series. We therefore introduce the notation

$$z_{n,T}=y_{n,nt}=\frac{1}{n}\sum_{i=0}^{n-1} x_{nt-i}. \qquad (1.4)$$

With the terminology used by J. Nordø (1959), $y_{n,t}$ will be called *overlapping means* and $z_{n,T}$ *successive means*.

In particular we note the formula

$$\varrho(z_{n,T}, z_{n,T-1})=\varrho(y_{n,nt}, y_{n,nt-n}) \qquad (1.5)$$

where ϱ denotes the correlation coefficient.

Finally we introduce the time series for the "day-to-day variation" v_t defined by the equation

$$v_t=x_t-x_{t-1} \qquad (1.6)$$

Introducing for simplification

$$\varrho_{x,i} = \varrho(x_t, x_{t-i}), \tag{1.7}$$

we have:

$$\sigma_v^2 = 2\sigma_x^2(1 - \varrho_{x,1}), \tag{1.8}$$

and

$$\varrho(v_t, v_{t-i}) = \tfrac{1}{2}(1 - \varrho_{x,1})^{-1}(2\varrho_{x,i} - \varrho_{x,i-1} - \varrho x, i+1). \tag{1.9}$$

The main object of this paper is the study of the mathematical properties of the time series for $y_{n,t}$, $z_{n,T}$, and v_t when simple models have been chosen for the x-series.

2. The case of random x-series:

Assuming

$$x_t = \varepsilon_t \tag{2.1}$$

we find:

$$y_{n,t} = \frac{1}{n}\sum_{i=0}^{n-1} \varepsilon_{t-i}, \tag{2.2}$$

and:

$$z_{n,T} = y_{n,nt} = \frac{1}{n}\sum_{i=0}^{n-1} \varepsilon_{nt-i}. \tag{2.3}$$

Thus the z-series is of the same random type as the x-series. The y-series, on the other hand, obey a simple scheme of moving averages of "fashion length n", according to a terminology introduced by the author (1965, p. 61). We note in particular the values of the autocorrelation coefficients

$$\varrho(y_{n,t}, y_{n,t-j}) = \begin{cases} \dfrac{n-|j|}{n} & \text{for } |j| \leqslant n \\ 0 & \text{for } |j| > n \end{cases} \tag{2.4}$$

Thus, considering directly the y_n-series, the best single predictor for $y_{n,t}$ is $y_{n,t-1}$. Since for $k \geqslant 2$

$$\varrho(y_{n,t}, y_{n,t-k}; y_{n,t-1}) = \frac{\varrho(y_{n,t}, y_{n,t-k}) - \varrho(y_{n,t}, y_{n,t-1})\varrho(y_{n,t-1}, y_{n,t-k})}{\sqrt{[1-\varrho^2(y_{n,t}, y_{n,t-1})][1-\varrho^2(y_{n,t-1}, y_{n,t-k})]}}, \tag{2.5}$$

we find, by combining with (4)

$$\varrho(y_{n,t}, y_{n,t-k}; y_{n,t-1}) = \begin{cases} -\sqrt{\dfrac{k-1}{(2n-1)(2n+1-k)}}, & \text{for } k \leqslant n \\ 0, & \text{for } k > n. \end{cases} \tag{2.6}$$

The best "additional predictor" is consequently $y_{n,t-n}$, corresponding to the regression equation

$$Y_{n,t} = \frac{1}{n+1}(ny_{n,t-1} - y_{n,t-n}). \tag{2.7}$$

The relative residual variance becomes

$$1 - R^2 = \frac{2}{n+1}, \quad \text{i.e.} \quad R^2 = \frac{n-1}{n+1} \tag{2.8}$$

We will not enter into the discussion of the generalization of this formula to more predictors, but only compare (7)–(8), describing the best "two-term autopredictability" of $y_{n,t}$, with the best predictor of $y_{n,t}$ when the *primary x-series are known*, i.e. the variable

$$y_{n-1,t-1} = \frac{1}{n}\sum_{i=1}^{n-1} x_{t-i}, \tag{2.9}$$

with

$$\varrho^* = \varrho(y_{n,t}, y_{n-1,t-1}) = \sqrt{\frac{n-1}{n}}. \tag{2.10}$$

Comparing with (8) we find

$$\varrho^{*2} - R^2 = \frac{n-1}{n(n+1)}, \tag{2.11}$$

a difference in residual variance of the order n^{-1}.

The prognosis based on $y_{n-1,t-1}$ corresponds to the *maximum* or *total predictability* inherent in our model.

For pentadic means, $n = 5$, formula (9) becomes

$$y_{n,t} = \tfrac{5}{6}y_{n,t-1} - \tfrac{1}{6}y_{n,t-5}, \quad \text{with} \quad R^2 = \tfrac{2}{3}, \tag{2.9'}$$

whereas (11) gives

$$\varrho^{*2} - R^2 = \frac{4}{30} = 0.1333. \tag{2.11'}$$

For the v-series 1 (6) we find by using (1)

$$\varrho(v_t, v_{t-1}) = -0.5, \quad \varrho(v_t, v_{t-j}) = 0, \quad j > 1. \tag{2.12}$$

Thus, a special type of "negative persistence" or "periodicity" will be observed in the interdiurnal variation.

It is, of course, impossible to pass in an unambigous way from the z_n-series to the x-series owing to the much smaller "information" contained in the former. x_t can be expressed by $y_{n,t}$-values by formulae of the types

$$x_t = n(y_{n,t} - y_{n,t-1}) + x_{t-n} = n(y_{n,t} - y_{n,t-1} + y_{n,t-n} - y_{n,t-n-1}) + x_{t-2n} \quad (2.13)$$

showing that a passage from the y_n-series to the x-series is not simple. In particular we conclude from the above discussion that if a moving n-average $y_{n,t}$ obeys a scheme of type (2) but with *weights different from unity*, the primary x-series must be expected to be of quite complicated nature.

3. Moving-average x-series

If x_t is a moving average of fashion length q

$$x_t = \sum_{j=0}^{q-1} a_j \varepsilon_{t-j}, \quad (3.1)$$

we have for $y_{n,t}$

$$n y_{n,t} = \sum_{j=0}^{q-1} \sum_{i=0}^{n-1} a_j \varepsilon_{t-i-j} = \sum_{i=0}^{m-2} \left(\sum_{j=0}^{i} a_j \right) \varepsilon_{t-i} + \sum_{i=m-1}^{M-1} \left(\sum_{j=0}^{m} a_j' \right) \varepsilon_{t-i} + \sum_{i=M}^{n+q-2} \left(\sum_{j=0}^{n+q-2-i} a_{q-1-j} \right) \varepsilon_{t-i} \quad (3.2\,\text{a})$$

with

$$m = \min(n, q), \; M = \max(n, q);$$

$$a_j' = a_j \text{ for } n > q, \; a_j' = a_{i-j} \text{ for } n < q. \quad (3.2\text{b})$$

Thus $y_{n,t}$ obeys a moving average scheme of fashion length $n + q - 1$ so that $y_{n,t}$ and $y_{n,t-1}$ have more common terms than x_t and x_{t-1} —and are, as was to be expected (if all $a_i > 0$) —more strongly correlated. Using the connection 1(4) between $y_{n,t}$ and $z_{n,T}$, we see that $z_{n,T}$ and $z_{n,T-1}$ have $n + q - 1 - n = q - 1$ terms of type ε in common. Although $z_{n,T}$ is not formally given by an ordinary moving average scheme of time step n, it will behave like such a scheme of fashion length $[(n + q - 2)/n]$, where as usual $[x]$ denotes the highest integer contained in x.

Finally we have for v_t defined by 1(6)

$$v_t = a_0 \varepsilon_t + \sum_{i=1}^{q-1} (a_i - a_{i-1}) \varepsilon_{t-i} - a_{q-1} \varepsilon_{t-q}, \quad (3.3)$$

showing the interdiurnal variation to be presented by a moving average scheme of fashion length $q + 1$. For $a_i = \text{constant} = a_0$ this scheme is of the simple form

$$v_t = a_0(\varepsilon_t - \varepsilon_{t-q}) \quad (3.3')$$

giving $\varrho(v_t, v_{t-i}) \neq 0$ only when $i = q$.

As illustrations let us briefly consider two cases of pentadic means

(**a**) $n = 5$, $q = 3$, $a_i = 1$, so that $\varrho_{x,1} = \frac{2}{3}$, $\varrho_{x,2} = \frac{1}{3}$, $\varrho_{x,i} = 0$ for $i > 2$.

For $y_{n,t}$ we find

$$5 y_{n,t} = \varepsilon_t + 2\varepsilon_{t-1} + 3(\varepsilon_{t-2} + \varepsilon_{t-3} + \varepsilon_{t-4}) + 2\varepsilon_{t-5} + \varepsilon_{t-6}. \quad (3.2')$$

The autocorrelations for y_n are given by the table

i	1	2	3	4
$\varrho(y_{n,t}, y_{n,t-i})$	0.9189	0.7297	0.4865	0.2703
i	5	6	>6	
$\varrho(y_{n,t}, y_{n,t-i})$	0.1081	0.0270	0	

Moreover, we have $\varrho(y_{n,t}, y_{n-1,t-1}) = 0.9631$—showing how much (also in this case) is lost in predictability by using $y_{n,t-1}$ instead of $y_{n-1,t-1}$ as predictor.

(*b*) $n = 5$, $q = 7$, $a_i = 1$. In this case there is a much stronger correlation in the x-series, since we have

$$\varrho_{x,i} = \frac{7 - i}{7} \quad \text{for} \quad i \leq 7; \; \varrho_{x,t} = 0 \quad \text{for} \quad i > 7.$$

For $y_{n,t}$ we have

$$5 y_{n,t} = \varepsilon_t + 2\varepsilon_{t-1} + 3\varepsilon_{t-2} + 4\varepsilon_{t-3} + 5 \sum_{i=4}^{6} \varepsilon_{t-i} + 4\varepsilon_{t-7} + 3\varepsilon_{t-8} + 2\varepsilon_{t-9} + \varepsilon_{t-10}.$$

The correlations in the y_n-series are given by the following table:

i	1	2	3	4
$\varrho(y_{n,t}, y_{n,t-i})$	0.9630	0.8667	0.6963	0.5630
i	5	6	7	8
$\varrho(y_{n,t}, y_{n,t-i})$	0.4000	0.2593	0.1481	0.0741
i	9	10	>10	
$\varrho(y_{n,t}, y_{n,t-i})$	0.0296	0.0074	0	

Finally we have

$$\varrho(y_{n,t}, y_{n-1}, y_{n-1,t-1}) = 09871.$$

For $z_{n,T}$ we have

$\varrho(z_{n,T}, z_{n,T-1}) = 0.5630$, $\varrho(z_{n,T}, z_{nT-2}) = 0.0296$, $\varrho(z_{n,T}, z_{n,T-i}) = 0$ for $i > 2$, corresponding to a considerable time dependence between pentadic means.

4. Markovian autoregressive x-series

Let us assume

$$x_t = b_1 x_{t-1} + \varepsilon_t; \; |b_1| < 1, \sigma_x^2 = 1, \quad \text{for } i > 0. \qquad (4.1)$$

$$E[x_{t-i}\varepsilon_t] = 0$$

Then we have, using the notation 1(7)

$$E(\varepsilon_t^2] = 1 - b_1^2, \varrho_{x,i} = b_1^i. \qquad (4.1')$$

$$E[x_t \varepsilon_{t-i}] = b_1^i(1 - b_1^2).$$

From 1(3) we find

$$y_{n,t} = b_1 y_{n,t-1} + \frac{1}{n}\sum_{i=0}^{n-1} \varepsilon_{t-i}. \qquad (4.2)$$

Consequently the y-series are of "mixed" type with Markovian autoregressive part and fashion length n. Elementary, but quite laborious computations give

$$\sigma_y^2 = \sigma_z^2 = n^{-1} + \frac{2b_1}{n^2(1-b_1)^2}[n(1-b_1) - 1 + b_1^n]$$

$$= n^{-1}(1 + \phi(b_1)), \qquad (4.3)$$

and

$$\text{cov}\,(y_{n,t}, y_{n,t-1}) = \frac{n-1}{n^2} + \frac{b_1}{n^2(1-b_1)^2} \times [2(n-1) - 2nb_1 + b_1^{n-1} + b_1^{n+1}] \qquad (4.4)$$

Combining (3) and (4) we find an expression for the correlation coefficient $\varrho(y_{n,t}, y_{n,t-1})$. The corresponding higher coefficients are quite complicated. However, we note that

$$\frac{\varrho_{y,n+i+1}}{\varrho_{y,n+i}} = \frac{\varrho(y_{n,t}, y_{n,t-n-i-1})}{\varrho(y_{n,t}, y_{n,t-n-i})} = b_1 \text{ for } i > 0 \quad (4.5)$$

We further have the following formula, useful for prediction problems

$$\text{cov}\,(y_{n,t}, x_{t-i}) = n^{-1} \times \left[1 + \frac{b_1}{1-b_1}(2 - b_1^i - b_1^{n-1-i})\right]. \qquad (4.6)$$

Hence, in particular

$$\text{cov}\,(y_{n,t}, x_{t-i}) = \text{cov}\,(y_{n,t}, x_{t-n+1+i}), \qquad (4.6')$$

the interpretation of which is self-evident.

The best prediction of $y_{n,t}$ by x_{t-i} is obtained by giving to i its "central" value, i.e. $(n-1)/2$ for n odd and $(n/2) - 1$ for n even.

Finally we observe that a good i-step predictor is $y_{n-i,t-i}$, for which we have for $i \leqslant n-2$

$$\text{cov}\,(y_{n,t} y_{n-i,t-i}) = \frac{1+b_1}{(1-b_1)n} - \frac{b_1(1+b_1^i)(1-b_1^{n-i})}{n(n-i)(1-b_1)^2}. \qquad (4.7)$$

For $i = n-1$ the covariance follows by putting $i = n-1$ in (6).

Since we have

$$\sum_{j=0}^{n-1}\sum_{i=0}^{n-1} b_1^i \varepsilon_{nt-i-j} = \sum_{i=0}^{n-1} \frac{1-b_1^{i+1}}{1-b_1}\varepsilon_{nt-i} + \sum_{i=0}^{n-2} b_1^{i+1} \times \frac{1-b_1^{n-i-1}}{1-b_1}\varepsilon_{nt-n-i} = n\varepsilon_T^* + n\varepsilon_T^{**},$$

we may write $z_{n,T}$ in the form

$$z_{n,T} = b_1^n z_{n,T-1} + \varepsilon_T^* + \varepsilon_T^{**} \qquad (4.8)$$

We note that ε_T^* is composed by $\varepsilon_{nt}, \varepsilon_{nt-1}, \cdots, \varepsilon_{nt-n+1}$, ε_T^{**} by $\varepsilon_{nt-n}, \varepsilon_{nt-n-1}, \cdots, \varepsilon_{nt-2n+2}$, and that $\varepsilon_T^{**} \neq \varepsilon_{T-1}^*$. Thus the z-series are of mixed type, although not expressed in the "classical" form. Moreover we have

$$\text{cov}\,(z_{n,T}, z_{n,T-i}) = \frac{b_1^{(i-1)n+1}}{n^2}\left(\frac{1-b_1^n}{1-b_1}\right)^2 \quad (4.9)$$

Combining with (3) gives $\varrho(z_{n,T}, z_{n,T-i})$. We have in analogy with (5)

$$\frac{\varrho_{z,i+1}}{\varrho_{z,i}} = \frac{\varrho(z_{n,T}, z_{n,T-i-1})}{\varrho(z_{n,T}, z_{n,T-i})} = b_1^n \quad \text{for} \quad i > 0. \quad (4.10)$$

The above formulae (3), (5), and (9) have been applied by J. Nordø (1959) to clear up some misunderstandings with respect to often used procedures in long-range weather prediction.

The simplest prognostic formula when only the z-series are known is that connecting $z_{n,T}$ with $z_{n,T-1}$, i.e. the formula

$$z_{n,T} = \varrho(z_{n,T}, z_{n,T-1})\, z_{n,T-1} = \varrho_{z,1} z_{n,T-1} \quad (4.11)$$

with a relative error of $1 - \varrho_{z,1}^2$.

The formula for partial correlations

$$\varrho_{z,i;\,1} = \varrho(z_{n,T}, z_{n,T-i}; z_{n,T-1})$$
$$= \frac{\varrho(z_{n,T}, z_{n,T-i}) - \varrho(z_{n,T}, z_{n,T-1})\varrho(z_{n,T-1} z_{n,T-i})}{\sqrt{[1-\varrho^2(z_{n,T}, z_{n,T-1})][1-\varrho^2(z_{n,T-1}, z_{n,T-i})]}}$$

where $i > 1$, becomes by using (8) and the stationary properties:

$$\varrho_{z,i;\,1} = \frac{\varrho_{z,i} - \varrho_{z,1}\varrho_{z,i-1}}{\sqrt{(1-\varrho_{z,1}^2)(1-\varrho_{z,i-1}^2)}} = \frac{\varrho_{z,1}}{\sqrt{(1-\varrho_{z,1}^2)}}$$
$$\times \frac{b_1^{ni} - \varrho_{z,1} b_1^{n(i-1)}}{\sqrt{1-\varrho_{z,1}^2\, b_1^{2n(i-1)}}}. \quad (4.12)$$

For $n \geqslant 5$ and practically all values of b_1 we may put:

$$\varrho_{z,i;\,1} = \frac{\varrho_{z,1}}{\sqrt{(1-\varrho_{z,i}^2)}} b_1^{(n-i)i} (b_1^i - \varrho_{z,1})$$
$$\times (1 + \tfrac{1}{2}\varrho_{z,1}^2 b_1^{2(n-1)i}),$$

from which we conclude that $z_{n,T-2}$ is the best additional predictor at least when b_1 is not very near to unity.

When also the x-series (1) are known, the best prediction formulae for $z_{n,T}$ are found by using x_{nt-in}. We have

$$\text{cov}\,(z_{n,T}, x_{nt-n-in}) = n^{-1} b_1^{1+in} \frac{1-b_1^n}{1-b_1}$$
$$= b_1^{in}\, \text{cov}\,(z_{n,T}, x_{nt-n}), \quad (4.13)$$

and corresponding formulae exist for the correlation coefficient. The greatest value, obtained for $i = 0$ and giving the total predictability of $z_{n,T}$ is considerably greater than $\varrho(z_{n,T}, z_{n,T-1})$ as is seen by comparing with (9) for $i = 1$.

The prediction formula can be written, by using (3)

$$z_{n,T}^* = \sigma_z^2 \varrho(z_{n,T}, x_{nt-n}) x_{nt-n}$$
$$= \frac{b_1(1-b_1^n)}{(1-b_1)(1+\phi(b_1))} x_{nt-n} \quad (4.14)$$

Finally we have for the day-to-day variation according to 1(6)

$$v_t = b_1 v_t + \varepsilon_t - \varepsilon_{t-1}, \quad (4.15)$$

so that the v-series are of mixed type with fashion length 2 and tradition length unity. Moreover, 1(7) and 1(8) give

$$\sigma_v^2 = 2(1-b_1), \varrho(v_t, v_{t-i}) = -\tfrac{1}{2}(1-b_1) b_1^{i-1}$$
$$= b_1^{i-1} \varrho(v_t, v_{t-1}). \quad (4.16)$$

As examples of the preceding formulae we have computed corresponding residual variances for pentadic and monthly averages.

Fig. 1 corresponds to pentadic averages. The continuous curves labeled 1,2,3,4 give $1 - \varrho^2 (y_{it}, y_{i-5,\,t-5})$ corresponding to (7). The dot-dashed line describes the one step prediction of $y_{5,t}$ by $y_{5,t-1}$—which consequently does not utilize the total predictability inherent in the model. The dashed lines correspond to prediction of $y_{5,t}$ by x_{t-i} as described by formula (6). The diagram also represents the 5-step predictabity characterizing the successive means; the continuous curves I, II, ..., V corresponding to prediction of $z_{n,T}$ by x_{nt-ni}, the dot-dashed to prediction of $z_{n,T}$ by $z_{n,T-i}$—i.e. to those types of prediction often used in long-range prognostic problems. The diagram shows clearly that this latter method does not utilize the total predictability inherent in the model.

Fig. 2 gives a similar representation for monthly means, restricted to b_1-values between 0.5 and 1.0, in order to present a clear way the predictability of $z_{n,T}$. For $n = 30$, the prediction of $y_{30,t}$ by $y_{30,t-1}$ is practically identical to that of $y_{30,t}$ by $y_{29,t-1}$, not presented in the diagram.

The general conclusions from the above dia-

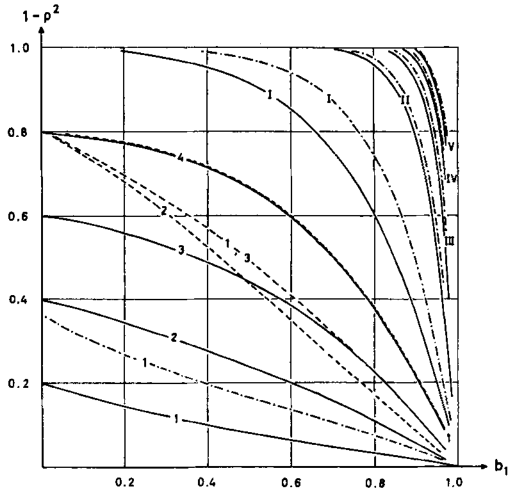

FIG. 1

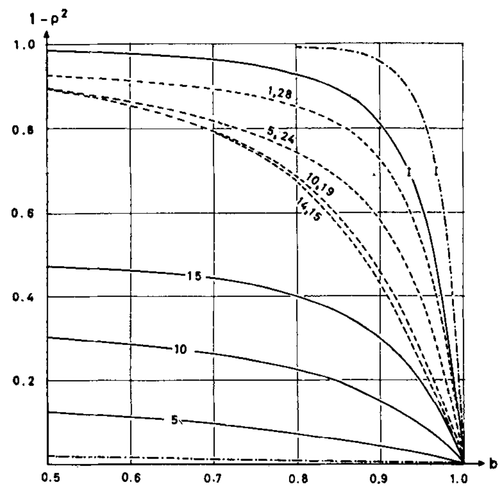

FIG. 2

FIG. 1. Residual variances corresponding to different predictions of pentadic means.
FIG. 2. Residual variances corresponding to different predictions of monthly means.

grams correspond to those already drawn in the important paper by J. NORDØ (1959), where also more details can be found.

5. General autoregressive x-series

Studies by J. NORDØ (1959), C. L. GODSKE (1962, 1965), J. H. KNUDSEN (1966, not yet published), and others have shown that time series in meteorology in general are not Markovian. Thus it will be of interest to consider the y_n- and z_n-series also when the x-series have tradition lengths greater than one. At present we restrict our study to the simple case when the x-series have the form

$$x_t = b_1 x_{t-1} + b_2 x_{t-2} + \varepsilon_t. \qquad (5.1)$$

For the y_n-series we easily find

$$y_{n,t} = b_1 y_{n,t-1} + b_2 y_{n,t-2} + n^{-1} \sum_{i=0}^{n-1} \varepsilon_{t-i}, \qquad (5.2)$$

i.e. a mixed series with the same autoregressive part as the x-series. This result is seen to hold also for the general autoregressive x-series.

The z_n-series are not of the ordinary mixed type of finite tradition length, as is easily seen even in the simple case $n = 2$, for which we have

$$z_{n,T} = (b_1^2 + b_2) z_{n,T-1} + b_1^2 b_2 \sum_{i=0}^{\infty} b_2^i z_{n,T-2-i}$$

$$+ 0.5(\varepsilon_t + \varepsilon_{t-1}) + 0.5 b_1 \sum_{i=0}^{\infty} b_2^i (\varepsilon_{t-i-1} + \varepsilon_{t-i-2}). \qquad (5.3)$$

Consequently, *no z_n-series of mixed type with finite tradition length can correspond to non-Markovian autoregressive x-series with finite tradition length.*

Finally we have for the v-series

$$v_t = b_1 v_{t-1} + b_2 v_{t-2} + \varepsilon_t - \varepsilon_{t-1}, \qquad (5.4)$$

so that this series are of mixed type with fashion length 2 and the autoregressive component the same as for the x-series—the latter result is valid also when the x-series obeys a general autoregressive scheme.

An interesting generalization is encountered when "errors" are superimposed on the time series. In this particular case it is possible that prediction of smoothed values are more successful when based on smoothed than when based on "perturbed" values. This problem will be discussed later as will also the case when the x-series are of "mixed" autoregressive-moving average type.

6. Average of quantities defined by a "chain of series"

The generalization of our study to "chains of time series" (see GODSKE, 1962) may be of some importance for meteorological variables having a diurnal variation. We will at present consider the simplest Markovian series of period 2:

$$\left.\begin{aligned} x_{2t} &= b_{01}x_{2t-1} + \varepsilon_{2t}, \\ x_{2t-1} &= b_{11}x_{2t-2} + \varepsilon_{2t-1}, \end{aligned}\right\} \qquad (6.1)$$

where we have simplified by assuming $\sigma^2_{x_{2t}}$ or $\sigma^2(x_{2t}) = \sigma^2_{x_{2t}-1}$ or $\sigma^2(x_{2t-1}) = 1$. The y_n-series have also period 2, and seem somewhat artificial from the meteorological point of view. If x_{2t} denotes, say the temperature at 19^h, x_{2t-1} at 01^h, $z_{2,T} = \frac{1}{2}(x_{2t} + x_{2t-1})$ is a rough expression for the average daily temperature for "day t". For $z_{2,T}$ we have

$$z_{2,T} = b_{01}b_{11}z_{2,T-1} + \varepsilon_{2t} + (1 + b_{01})\varepsilon_{2t-1} + b_{11}\varepsilon_{2t-2}, \qquad (6.2)$$

so that the z_2-series is of *mixed type*. Observing (see GODSKE, 1962, p. 190)

$$\left.\begin{aligned} \varrho(x_{2t}, x_{2t-2n}) &= (x_{2t-1}, x_{2t-1-2n}) = (b_{01}b_{11})^n, \\ \varrho(x_{2t}, x_{2t-2n-1}) &= b_{01}(b_{01}b_{11})^n, \\ \varrho(x_{2t-1}, x_{2t-2n-2}) &= b_{11}(b_{01}b_{11})^n \end{aligned}\right\} \qquad (6.3)$$

we further find

$$E[z^2_{n,T}] = 0.5(1 + b_{01}),$$

$$\text{cov}\,[z_{n,T}, z_{n,T-j}] = 0.25b_{11}(1 + b_{01})^2(b_{01}b_{11})_{j-1},\ j > 0$$

so that

$$\varrho(z_{n,T}, z_{n,T-j}) = 0.5b_{11}(1 + b_{01})(b_{01}b_{11})^{j-1}. \qquad (6.5)$$

We note that

$$\frac{\varrho(z_{n,T}, z_{n,T-j-1})}{\varrho(z_{n,T}, z_{n,T-j})} = b_{01}b_{11} \quad \text{for} \quad j > 0,$$

corresponding to the fashion length 2 for the z_n-series.

Generalizing these studies to n-chains (say with $n = 24$), we conclude that it may be reasonable to apply to daily average temperatures not Markovian or higher order autoregressive schemes, but time series of mixed type.

7. Averages for variables obeying a cross-correlation scheme

Let us assume the variables x_{I} and x_{II} to obey the simple scheme

$$\left.\begin{aligned} x_{\text{I},t} &= a_{\text{I}}x_{\text{I},t-1} + b_{\text{I}}x_{\text{II},t-1} + \varepsilon_{\text{I}}, \\ x_{\text{II},t} &= a_{\text{II}}x_{\text{I},t-1} + b_{\text{II}}x_{\text{II},t-1} + \varepsilon_{\text{II}}, \end{aligned}\right\} \qquad (7.1)$$

where ε_{I} and ε_{II} satisfy conditions 1(1).

For $y_{2,t} = \frac{1}{2}(x_{\text{I},t} + x_{\text{II},t})$ we then have

$$y_{2,t} = \tfrac{1}{2}(a_{\text{I}} + a_{\text{II}})x_{\text{I},t-1} + \tfrac{1}{2}(b_{\text{I}} + b_{\text{II}})x_{\text{II},t-1} + \tfrac{1}{2}(\varepsilon_{\text{I}} + \varepsilon_{\text{II}}). \qquad (7.2)$$

Only when

$$a_{\text{I}} + a_{\text{II}} = b_{\text{I}} + b_{\text{II}} = 2a, \qquad (7.3)$$

can $y_{2,t}$ be expressed exclusively by means of $y_{2,t-1}$

$$y_{2,t} = ay_{2,t-1} + \tfrac{1}{2}(\varepsilon_{\text{I}} + \varepsilon_{\text{II}}) \qquad (7.4)$$

corresponding to a simple Markovian time dependence. If we put

$$a_{\text{I}} + a_{\text{II}} = 2a + 2b,\ b_{\text{I}} + b_{\text{II}} = 2a - 2b, \qquad (7.5)$$

we may write in the general case

$$y_{2,t} = ay_{2,t-1} + b(x_{\text{I},t-1} - x_{\text{II},t-1}) + \tfrac{1}{2}(\varepsilon_{\text{I}} + \varepsilon_{\text{II}}), \qquad (7.6)$$

so that the average $y_{2,t}$ does not only depend on earlier values of y_2, but also of earlier values of the "amplitude" $x_{\text{I}} - x_{\text{II}}$.

Generalizations to non-Markovian cross-correlation schemes may be taken up later. At present we only note that the model (1) and its non-Markovian generalization, for n variables, may be of some interest if we want to study time dependence between *space averages*. As follows from the above simple examples these dependences may be expected to be quite complicated.

REFERENCES

GODSKE, C. L., Contributions to Statistical Meteorology I. Geophysica Norvegica, 24, pp. 161–210, Oslo, 1962.

GODSKE, C. L., Statistics of Meteorological Variables. Final Report, Contract AF 61 (052)-416. Bergen, 1965.

NORDØ, J., Expected Skill of Long-Range Forecast when derived from daily Forecasts and past Weather Data. Scientific Report. Det norske meteorologiske institutt. Oslo, 1959.

О ВРЕМЕННОЙ ЗАВИСИМОСТИ СГЛАЖЕННЫХ ПЕРЕМЕННЫХ

В метеорологии часто изучаются не мгновенные, а сглаженные значения переменных, осредненные, например, за день, за десять дней и за месяц. Когда изучается временная зависимость мгновенных и сглаженных переменных, возникает проблема применимости статистических моделей к полученным рядам наблюдений. В данной статье рассматривается вопрос, какие модели следует применять к сглаженным данным, если мгновенные переменные обрабатывались с помощью модели типа движущихся средних или авторегрессивной модели. Полученная в результате формула служит предостережением относительно возможности использования простых моделей сглаженных величин.