CONTRIBUTION TO STATISTICAL METEOROLOGY

I. On the application of non stationary time series to the study of Norwegian air temperatures

BY C. L. GODSKE

FREMLAGT I VIDENSKAPS-AKADEMIETS MØTE DEN 9DE FEBRUAR 1962

Summary. The aim of statistical climatology is to arrive at mathematical models giving an adequate description of the statistical properties of the atmospheric variables. Proceeding from simpler to more complicated studies, we may start with the following problems:

1. The construction of a mathematical model, say a distribution function, describing the statistical properties of "isolated" atmospheric variables for a given time and locality.

2. The construction of a model describing the statistical properties of the time variation of the atmospheric variables characteristic of a fixed locality, e.g. by autocorrelation analysis.

3. The construction of a model describing, at a given time, the statistical properties of the space variability of the variables (studies of representativeness).

4. The construction of a model taking into consideration the statistical relationship between different atmospheric variables measured at a given time and locality.

Combining 1—4 we arrive at the more difficult, but extremely important, problems characteristic of *synoptical climatology and statistical weather prediction.*

The present paper is a contribution to problem 2. In Chapter 1 we present a summary of the statistical analyses of air temperatures from Oslo and Bergen, taking especially into consideration the time persistency in the data; the time unit of the correlograms is partly chosen as one day, partly as one hour. Stationary time series can, with some caution, be used for representing the day to day persistency. However, the correlograms with one hour as time unit are of non-stationary type, owing to the daily variation of temperature. In Chapter 2 the stationary time series are generalized to what we have called a *chain of time series,* some simple properties of which have been summarily studied. Finally, Chapter 3 gives some simple numerical examples, making probable that chains of stationary time series may be useful mathematical models for the study of daily variations of quantitative atmospheric variables like the air temperature.

CHAPTER 1

SOME STATISTICAL PROPERTIES OF THE BERGEN AND OSLO AIR TEMPERATURES

1. Introduction. The thermograms for the air temperature at the Norwegian meteorological stations are only partially utilized in climatology, owing to the laborious work connected with the reading off of hourly values. For some few stations a longer series of carefully checked hourly temperatures T_{01}, T_{02}, ... T_{24} exist, thus for Bergen 1904—30 and for Oslo 1901—30. These data have been punched under the supervision of TH. WERNER JOHANNESSEN, chief of the Climatological division of the Norwegian Meteorological Institutte. Correlation computations have been carried through by K. FLØISAND and E. RAMM on the IBM 650 digital computer at the University of Bergen, and correlograms have been drawn by M. GODSKE. For working out the numerical examples in Chapter 3 the assistance of B. GRUNG and J. H. KNUDSEN has been valuable; the final drawing of the diagrams is due to E. BOLSTAD and S. O. SIVERTSEN. The investigation has been sponsored by the Geophysical Research Directórate, Air Force Cambridge Research Center under Contract AF61 (052)—416.

Among the computations hitherto performed we will only mention those which are of interest for the following time series studies, namely:

1. Computation of monthly temperature averages for each hour of observation: $\overline{T_{01}} = \frac{1}{n} \sum T_{01}$, $\overline{T_{02}}$, ... $\overline{T_{24}}$ for Bergen. The sum is extended to all days of the months and to a certain number of years.
2. Computation of corresponding "between-days" standard deviations:
$$s_{01} = \sqrt{\frac{1}{n} \sum (T_{01} - \overline{T_{01}})^2}, \; s_{02}, \; \ldots\ldots \; s_{24}.$$
3. Computation of corresponding serial correlation coefficients with time lags, 1, 2, ... 8 days for all hours of observation and all months.
4. Computation, for Bergen and Oslo, of serial correlation coefficients with time lags 1, 2, 3, ... 72 hours for the 8 hours of observation 01^h, 04^h, ... 22^h, and all months.

The aim of these computations has, among others, been to furnish data enabling us to decide how far the concept of stationary time series is applicable to the study of daily and seasonal variations in air temperatures, and which generalizations of the series are necessary if stationary series cannot be utilized.

2. Daily and seasonal variations in the "between-days" variability of Bergen air temperatures. The standard deviations "between days" of the Bergen temperatures have been computed, for each month, separately for the 12 odd years 1907, 1909, . . . 1929, for the 13 even years 1906, 1908, . . . 1930, and for the 25 years 1906—30. It may be quite difficult, owing to persistency and non-normality in the temperature distribution, to judge the statistical significance and reliability of diurnal and seasonal variations if we only consider results from the complete 25 years period. However, features which appear in clear form both in the 12 odd years and in the 13 even years series, can with some confidence be considered as statistically significant. (The division of the complete series into an odd and an even one has also been chosen so as to eliminate changes, if any, in climate; in fact the two series thus selected may possibly be considered as random samples from the same "statistical universe").

Fig. 1 shows, for the 25 years period, the daily and seasonal variation in s_i, starting with January at the top of the diagram. In order to give some indication of the "time stability" of the variations without overcrowding the diagram, we have drawn the curves for January, April, July, and October also for the 12 years (dotted curves) and for the 13 years (dashed curves). The following features appear, and are confirmed by the curves for the 12 odd and the 13 even years for the other months:

In October – February the between-days variability of the temperature has a flat minimum between 13^h and 15^h, and only slight variations between 22^h and 10^h. The daily amplitude is greatest in January—February. In April—August a maximum exists about 13^h—16^h; the values of s_i are practically constant between 22^h and 07^h. In March the 25 years period — as well as the 12 years and 13 years period — shows a morning maximum in s_i. A very flat secondary maximum at about 16^h is found for the 25 years period; this is more pronounced for the 12 years period but does not exist in the 13 years period. In September the 25 years period shows minima at 09—10^h and at 19—20^h, with maxima at 07^h and 14—15^h; the same features are also evident in the odd and the even series.

Although the physical meaning of the arithmetic mean of 24 standard deviations for a given month is more than doubtful, we have represented this quantity, for 12, 13, and 25 years, in fig. 2. Its seasonal variation is very marked: maxima occur in April—May and December—January, minima in February—March and August—September. The agreement between the odd and the even periods indicates the statistical significance of the variations. We note in particular the strong change from September to October, corresponding to almost 50 per cent increase in the between-days temperature variance.

In a stationary time series

1 (1) $$\ldots, x(t-1), x(t), x(t+1). \ldots$$

where x, for each value of t, is a stochastic variable, the standard deviation, σ_x, of x must be independent of time. Then σ_x estimated from between-days variations by time averaging over a month (or the same month for many years) keeping the hour of

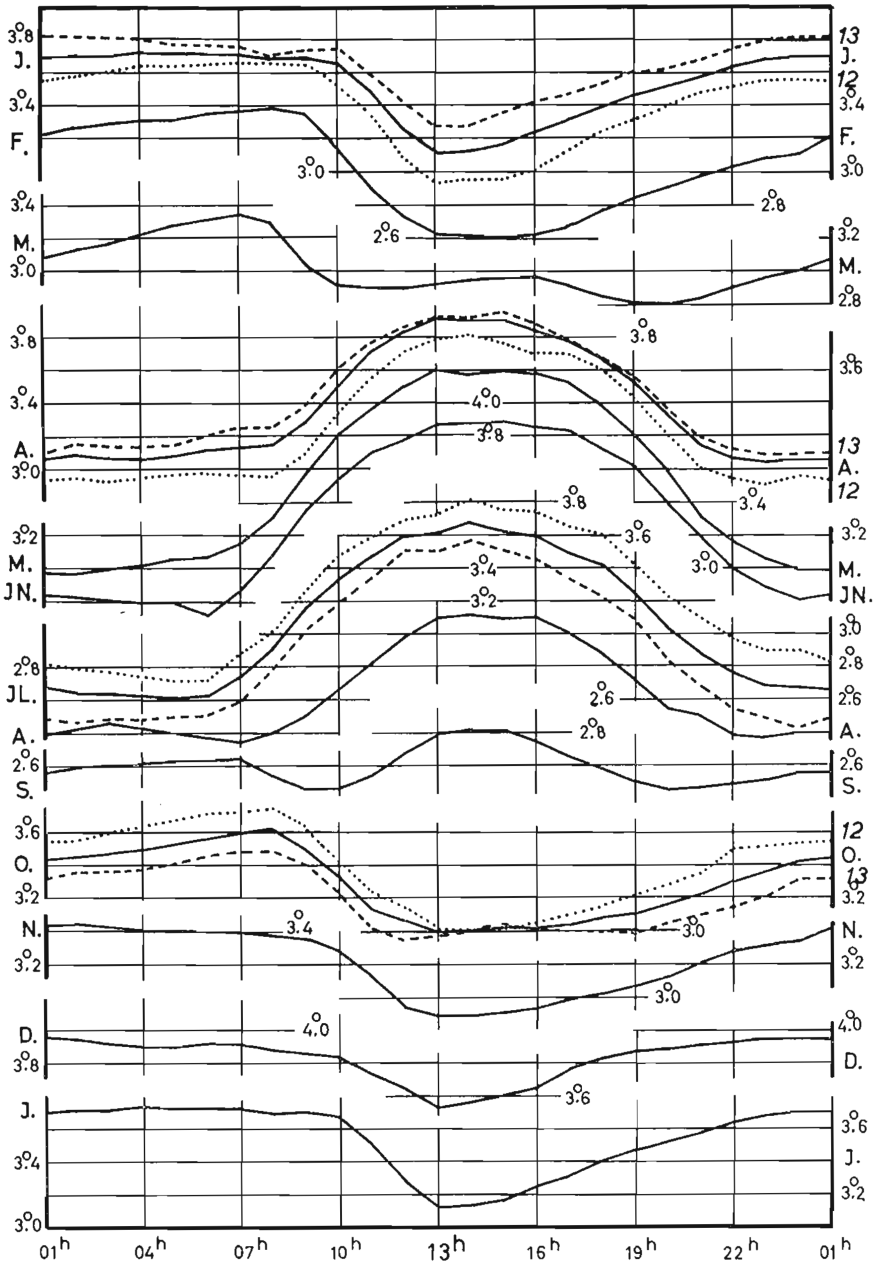

Fig. 1. Daily and seasonal variations of the between-days variability, s_i, for the Bergen air temperature during the period 1906—30. Dotted and dashed curves represent the variations for the odd 12 years period and the even 13 years period respectively for 4 selected months.

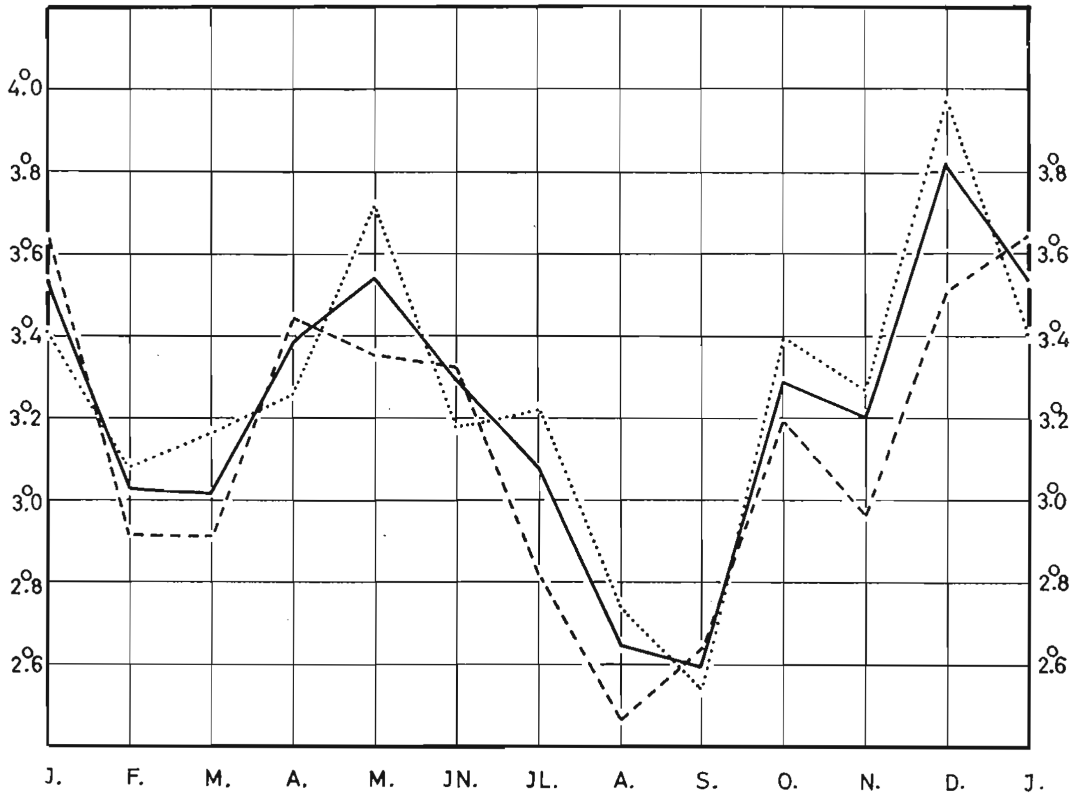

Fig. 2. Seasonal variation for Bergen 1906—30 of the "average" between-days temperature variability as measured by the standard deviation s_i, together with corresponding curves for the odd 12 years period (dotted) and the even 13 years period (dashed).

observation constant, should show only slight and insignificant daily and seasonal variations. Fig. 1 shows that this is not the case for the Bergen air temperature T.

Thus we are justified in concluding that both the series of the type:

1 (2) $$\ldots T(t - 1^h),\ T(t),\ T(t + 1^h),\ T(t + 2^h) \ldots, \qquad (h = \text{hour})$$

and the series of the type:

1 (3) $$\ldots T(t - 1^d),\ T(t),\ T(t + 1^d),\ T(t + 2^d) \ldots, \qquad (d = \text{day})$$

(where T denotes the Bergen air temperatures) are non-stationary.

It might, however, be possible that the time series obtained by dividing the temperature anomaly $T_i - \overline{T}_i$ by the corresponding standard deviation were stationary. Then the correlation coefficient $r\,[x(t),\ x(t + \tau)]$, where $x = \dfrac{T - \overline{T}}{\sigma_T}$, should be independent of t, only depending on τ. In order to obtain more information about the character of the series (2) and (3), we consequently have to consider the serial correlations corresponding to these series.

3. Correlograms for the Bergen temperatures with time lags 1, 2, . . . 8 days. Serial correlation coefficients with 1, 2, . . .8 days lag have been computed for Bergen, for each hour of observation and each month for the above defined groups of 12, 13, and 25 years; in all 864 curves have been drawn. For Oslo analogous curves have later been drawn for 3 periods of 10 years and for the 30 years period 1901—30; the discussion of the Oslo curves will not be attempted in the present paper. Some few of the Bergen curves, all referring to the 25 years 1906—30, have been presented in fig. 3, namely the curves for the main hours of observation 01^h, 07^h, 13^h, 22^h, and for the months January, April, July, October. To the left we have given 4 small diagrams for the main hours of observation starting with 01^h at the bottom; to the right the same data have been presented, but the small diagrams refer to the selected months, starting with January at the bottom.

Fig. 3 left shows, among others, that for lags 1—4 days for all hours of observation the October correlations are the highest; for 13^h, April also shows high values. Thus the "persistence predictability" 1—4 days ahead of a Bergen air temperature is greatest in October, a result whose reliability is made probable by considering also the 12 years and 13 years periods. Fig. 3 right shows that the predictability in January is practically the same for all hours of observation; also in July the difference is small with regard to predictions some few hours ahead. For October, and for lags 1—6 for April, however, the persistence predictability P_i of the different hours of observation is characterized by the following inequalities: $P_{07} \leq P_{01} < P_{13} < P_{19}$. In order to obtain some quantitative information about the difference, we present in table 1 the percentual residual variance for the 12, 13, and 25 years groups, for 1 day and 4 days persistency; decreasing values of the presented figures correspond to increasing predictability. The differences in the predictability are not great, but show a systematic character and are probably real.

Table 1.1 *Percentual residual variance for "persistency prediction" 1 and 4 days ahead.*

Hour of observation		07^h	01^h	13^h	19^h
1 day	April 12	55	53	47	44
	13	53	51	43	40
	25	52	51	44	41
	October 12	46	47	48	40
	13	54	52	45	39
	25	50	49	46	39
4 days	April 12	93	90	90	87
	13	95	95	89	89
	25	93	92	88	87
	October 12	88	88	87	87
	13	84	82	85	76
	25	86	86	85	82

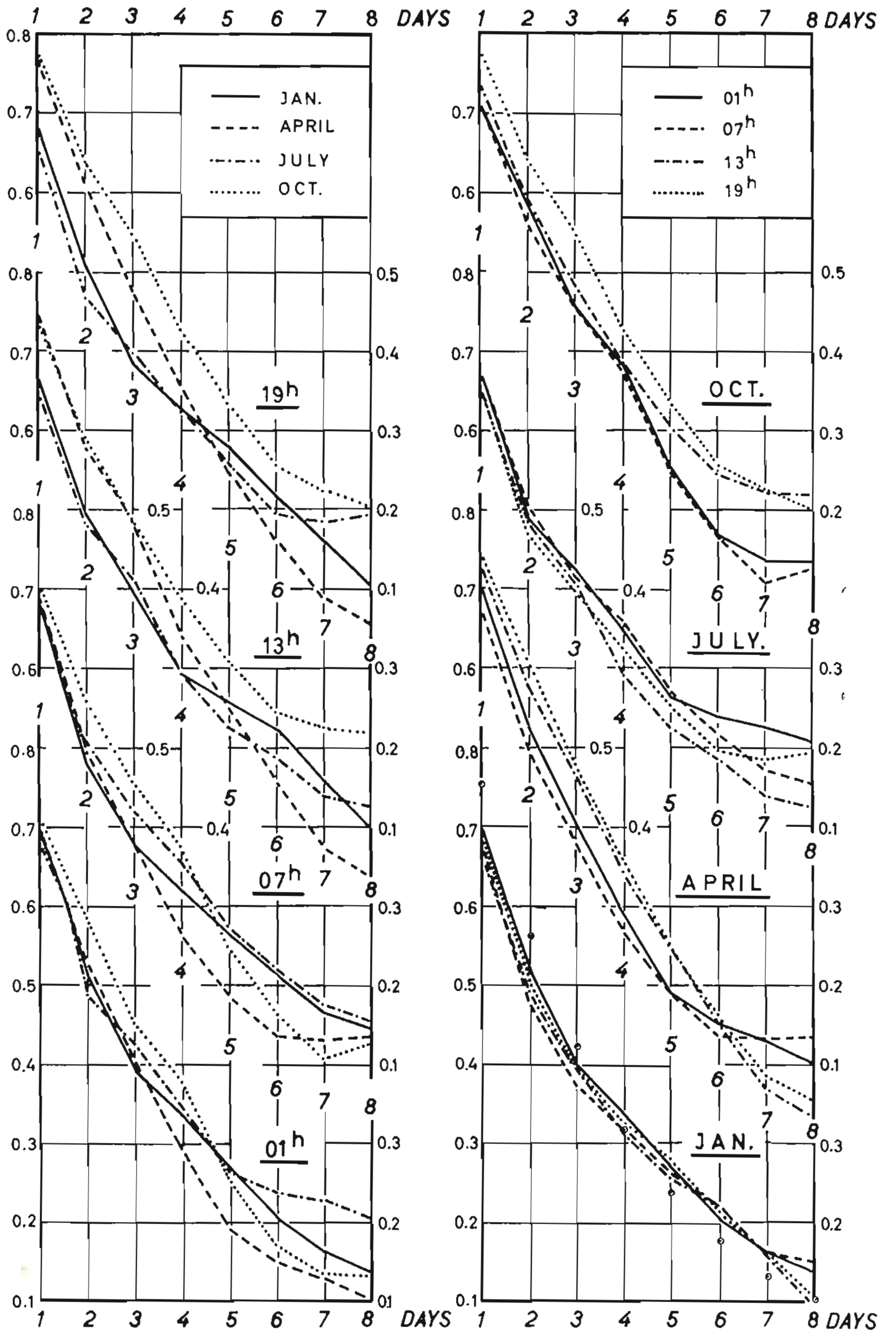

Fig. 3. Correlograms for the Bergen air temperatures 1906–30, 01^h, 07^h, 13^h, 19^h for January, April, July, October.

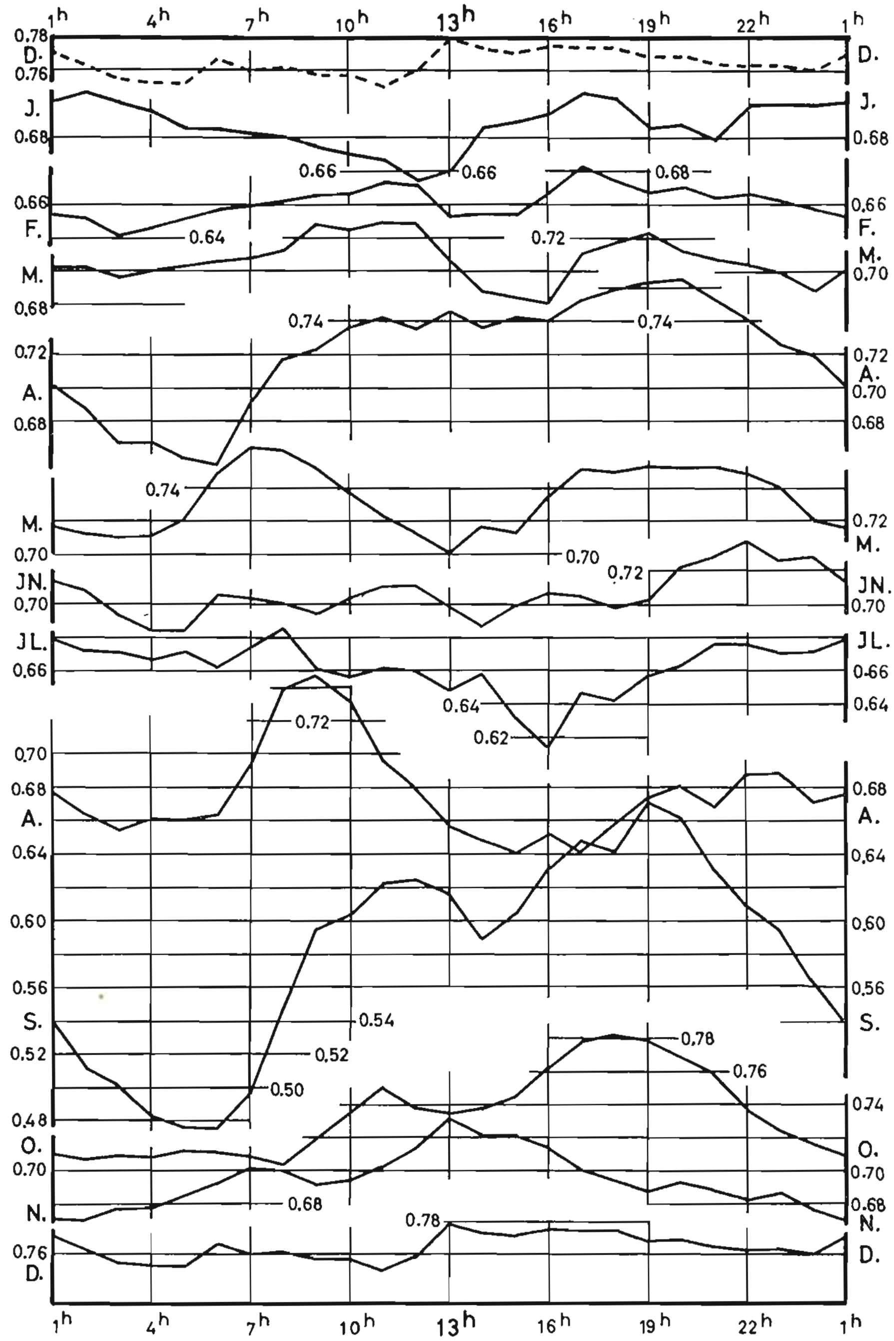

Fig. 4. The daily and seasonal variations of the Bergen temperature autocorrelation coefficients 1906—03 for time lag one day.

In the January curve (bottom, right) we have added small circles corresponding to an "average" exponential decrease of the correlation coefficient r_i. The decrease observed is for small i-values more rapid, for greater i more slow than the exponential. This indicates that not all previous information about $T(t)$ is contained in $T(t-1)$; the temperature time series, at least in January, cannot be considered as Markovian. A systematic treatment of curves of the type fig. 3 will be taken up later when mathematical models will be fitted to the empirical date (see for instance J. NORDØ, 1959).

Fig. 4 (analogous to fig. 1) gives, as a supplement, for different hours of observation and months the value of $r[T(t), T(t-1^d)]$, but only for the 25 years period; similar curves have, however, also been drawn for the 12 years and 13 years period. In January a noon minimum and a midnight maximum is indicated in the 12, 13, and 25 years periods — and a slight indication of a secondary maximum at 17^h and minimum at 21^h. December gives curves of quite other types, so also does February. If, therefore, the daily variations in r shown in these three months are significant, there must exist a rapid change in the character of the persistency during the winter months — probably associated with similar changes in the frequency of different weather types.

March is not too unlike February, with a pronounced afternoon and a flat night minimum, the corresponding maxima occurring at 11^h and 19^h. In April the morning minimum at 06^h is very marked for all 3 groups (12, 13, 25 years), whereas the afternoon minimum is clearly indicated only in the 13 years period. The highest value of r is found at 19^h. The highest r-values (0.765) and the lowest (0.655) correspond to residual variances of 41.5 and 57.0 per cent, showing that a considerable difference exists in one day persistency between morning and evening.

In May both maxima are pronounced, as in March, but the first maximum is as early as 07^h. The minimum at 13^h is very weak in the 12 years period, but extremely strong in the 13 years period. The June curve shows only slight variations, which may be of random character although the three minima at 04^h, 08^h, and 14^h and the maxima at 11^h and 22^h appear both in the 12 years and the 13 years periods. For July the odd and the even periods show quite different features, but the minimum at 16^h and the morning maximum seem to be real. In August, these two extremes are much more pronounced, moreover a minimum appears in the early morning and a maximum in late afternoon, so that May and August show strong similarities. The September curves show a very deep morning minimum and a high maximum at 19^h, the secondary extremes being at 12^h and 14^h. Although the average r-value for the 12 years and the 13 years period (0.535 and 0.617 respectively) are quite different (see fig. 5), the daily variation of r is surprisingly similar for the two series, so that at least the *form* of the September curve must be considered as securely established. For the 25 years period the maximum in r (0.671) corresponds to a residual variance of 55.0 per cent, whereas the corresponding value for the minimum (0.475) is 77.5 per cent, showing the very low one-day predictability of the September morning temperatures.

In October the level of predictability is markedly higher than in September; the deep night minimum has flattened out; both maxima and the intermediary noon

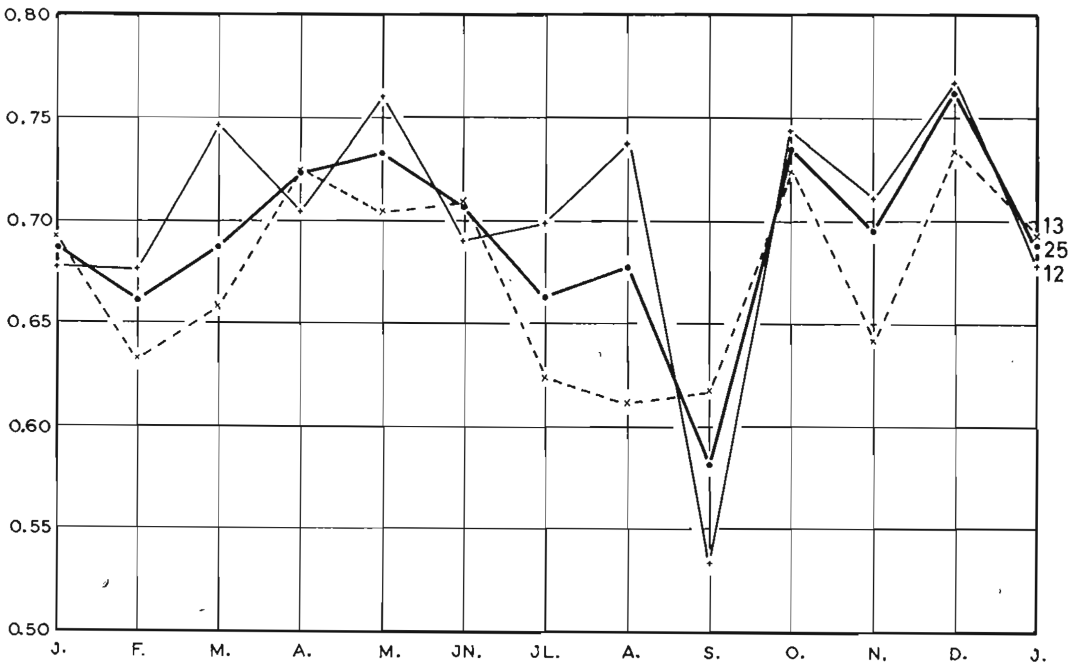

Fig. 5. Seasonal variation of the "average" autocorrelation coefficient for the Bergen temperatures 1906—30 (time lag one day).

minimum exist in the 12 years and 13 years periods. November shows a radical change with the main maximum at noon and a main minimum at midnight. In December, like January, the daily variation of r is very small, with a high level of r. The highest of all r-values in the 25 years series is found for October 18^h, namely 0.782, corresponding to a residual variance of 38.5 per cent — to be compared with the value 77.5 per cent found for this quantity at the minimum in September.

In order to compare the general levels of "one day persistency predictability" for the different months we have computed the arithmetic mean for the 24 hours of observation for each month and presented them in fig. 5 for 12, 13, and 25 years. A pronounced minimum exists in September, a maximum in December; January, April, June, and October show very "stable" conditions, with practically the same values for the 12 and the 13 years periods; the greatest scattering is found in August. As in fig. 2 we note the great change from September to October, corresponding to an increase of 50 per cent in r^2.

Summing up we may draw the following conclusion with regard to the autocorrelation coefficient (lag one day) for the Bergen temperatures:

1. The value of r shows a seasonal variation, so that only short series (say a month) ought to be used when computing serial correlations; in order to obtain sufficient data such small series from different years must be pooled.

2. The value of r shows in spring and autumn typical daily variations, the greatest predictability existing some hours after sunrise and about 19^h-22^h. In winter and summer these variations are much smaller, but still some of them may possess statistical significance.

In a later paper we will discuss the analogous Oslo curves and possibly also data from other meteorological stations in order to find the "region of validity" for the above results. The time stability, revealed by a comparison between the 12 years and the 13 years series, makes probable that 10 or 15 years may be sufficient, when trying to extend the investigation to other stations. However, our results indicate that it is recommendable to perform all computations separately for each month and decide about the pooling of different months only after careful inspection of the data (for Bergen, it might be permissible to pool February and March, and June and July).

4. Correlograms for the Oslo and Bergen temperatures with time lags 1, 2, . . . 72 hours. The correlograms $r\ (T_j,\ T_{j-i})$, $j = 01^h,\ 04^h,\ \ldots 22^h$; $i = 1,\ 2,\ \ldots 72^h$, for Bergen have been computed for five 5-years periods, namely 1906—10, 1911—15, 1916—20, 1921—25, and 1926—30, and for the "total" period of 25 years. Since 5 years curves gave a too bad approximation to the 25 years curve, we applied later 10, 15, and 30 years periods for the analogous computations for Oslo, and eliminated also most of the climatical variations by basing all curves on the following 5 years periods:

a: 1901, 1907, . . . 1925;	*b:* 1902, 1908, . . . 1926;
c: 1903, 1909, . . . 1927;	*d:* 1904, 1910, . . . 1928;
e: 1905, 1911, . . . 1929;	*f:* 1906, 1912, . . . 1930.

The three 10 years periods were the following: $a + d$, $c + e$, $b + f$, whereas the 15 years periods were $b + c + e$, and $a + d + f$. Since correlograms thus were computed for 8 hours of observation, for all months, and for 6 different groups, for both Oslo and Bergen, 1152 curves have been drawn, each of 72 points.

As an introduction to our provisional study of these curves, we present in fig. 6 some few of the more complicated ones, namely curves for Oslo, 07^h in April for the 10 years periods and for the period 1901—30. As shown by the exponential curve $e^{-\alpha t}$, where $e^{-72\alpha} = 0.05,\ 0.10,\ 0.15,\ \ldots\ 0.60$ (light continuous lines), we again see (compare fig. 3) that even the regular decrease in r for 1, 2, and 3 days cannot be approximated by an exponential curve (therefore, not much would be gained by using logarithmic paper).

The curves do not present an "undisturbed" regular decrease of r, but a series of regularly distributed "relative" maxima and minima — "relative" when referred to some decreasing "average" curve. Two main problems now exist: to find the position (i.e. the i-values) of the relative extremes and to estimate the reliability of this position.

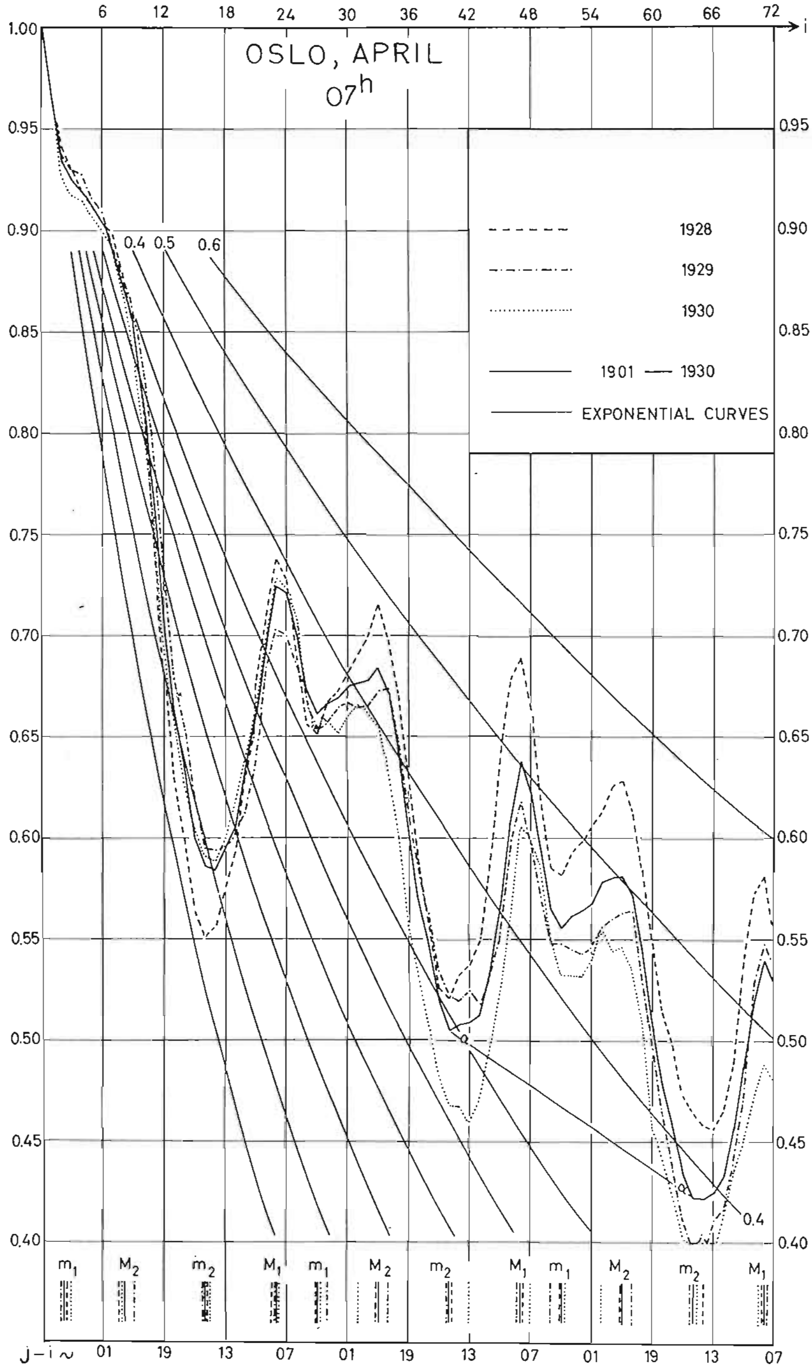

Fig. 6. Correlograms for the Oslo air temperatures, April 07^h, 10-years periods and 30-years period; time unit one hour.

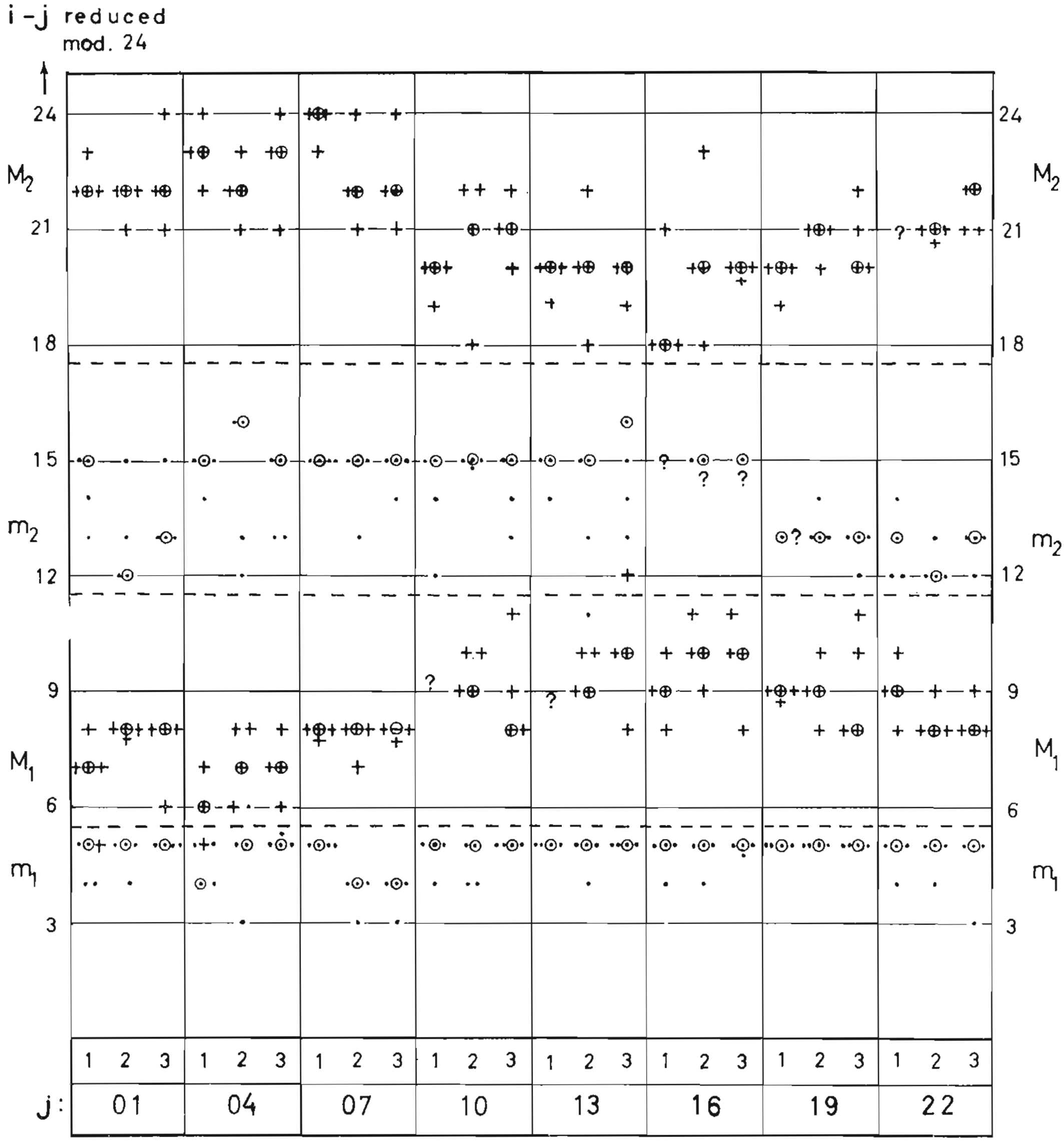

Fig. 7. Times of occurrence, measured by $i-j$ corresponding to the relative extremes in the correlograms for T_j ($j = 01^h, 04^h, \ldots 22^h$). 10-years periods and the period 1901—30 have been used; corresponding to $i_{max} = 72$ we have, in each period, 3 different determinations of the hours of extremes.

The solution of the first problem ought to be based on well defined "undisturbed" curves. Since the definition of such curves may be somewhat arbitrary, we will apply simple visual methods, for the most pronounced extremes based upon "common tangents" (see line segments $a-a$ in fig. 6). For the relative extremes for small i-values the exponential curves may also be of some value (still the extremes may be quite difficult to fix accurately to the left of our diagram). As an example consider the 4th relative minimum of the 30 years curve; its position is given by $i = 40$ corresponding to $j - i = 7 - 40 = - 33$ which, modulo 24, is equal to 15. The "$j - i$" value, reduced modulo 24, is given at the bottom of our diagram and marked by m_2; in a similar way the other extremes m_1, M_1, M_2, have been marked in the diagram.

The values of $j - i$ corresponding to a given extreme show a high degree of stability from one 10 years period to another, and can therefore reasonably be assumed as being statistically significant. The accordance between the 3 curves for 10 years is, on the whole, for all diagrams of the type represented in fig. 6, surprisingly good. As a consequence we conclude that a *10-years period is sufficient for obtaining the general form of the correlograms having time unit 1 hour.* (even 5-years curves, as computed for Bergen, reveal most of the characteristic features).

A cursory glance at the non-represented April curves for $j = 01^h, 04^h, 10^h \ldots, 22^h$ (see also fig. 8) makes probable that relative minima occur *not for constant values of* i, but — approximately — *for constant values of* $j - i$. We have consequently determined these values for the extremes of the 8 April curves ($01^h, 04^h, \ldots 22^h$) for Oslo and presented them in fig. 7. Dots represent the position of the 10-years minima, crosses refer to the analogous maxima, whereas dots and crosses surrounded by a small circle refer to 30-years extremes. In case of doubt, interrogation marks have been introduced. Each extreme is represented thrice (see figures 1, 2, 3 at the bottom of the diagram) corresponding to i-values between 1 and 24, between 25 and 48, and between 49 and 72 hours. We may note that (as follows from non-represented diagrams) m_1 is much more pronounced than m_2 for $j = 10^h, 13^h, 16^h, 19^h$, whereas the opposite is the case for $01^h, 04^h, 07^h$. For $j = 22^h$ both minima are about equal in intensity. For $22^h, 01^h, 04^h, 07^h, 10^h$ the maxima M_1 and M_2 are about equally strong, for $13^h, 16^h, 19^h$, M_1 is almost non-existent, but M_2 is very pronounced.

The minimum m_1 shows a remarkable "stability in time" (least so for 04^h and 07^h when it has a small intensity and its position may be difficult to determine). We may safely conclude that $j - i$ for this minimum is independent of j. The minimum m_2 and the maximum M_1 and M_2 show a slight dependency on j; the study of the statistical significance of this dependency will be taken up later. Thus our general conclusion is: At 05^h and $14^h - 15^h$ the information rendered by the observed temperature about the future temperature is abnormally small; we may also say that the *prediction value* of T is small and the *noise level* high at 05^h and $14^h - 15^h$, whereas the prediction value is great and the noise level low at the hours corresponding to the maxima M_1 and M_2.

As a supplement to our provisional study of the 07^h April curves for Oslo, we have in figs. 8—11 presented for 8 hours of observation the 25-years curves for Bergen (continuous lines) and the 30-years curves for Oslo (dot-dashed lines) for 4 selected months, January, April, July, and October (the curves for 5-years and 10-years periods are used in the discussion when we have some doubt about the statistical significance of the time position of the extremes). In the discussion we also refer to corresponding, not represented, curves for the other months.

Let us first, in connection with figures 6 and 7, consider the diagram for April (fig. 8). On our curves the times corresponding to the Oslo extremes, namely 05^h, 15^h, 18^h, and $20-23^h$ have been marked by small dashed vertical line segments. For the Bergen diagrams we have identified the same maxima and minima as for Oslo; they appear often a little later than in Oslo (partly, at least, owing to the longitude difference of 6°), and are, on the whole, less marked (since the Bergen climate is more oceanic than that of Oslo).

In January (fig. 9), with a very small daily temperature variation, the only feature which seems to be statistically significant in the Oslo curves is a weak minimum at 09^h; the Bergen curves are highly variable from one 5-years period to another; perhaps a minimum at 12^h-14^h may possibly be real (compare also the minimum in the January curve of fig. 4)

In February, Oslo, the 14^h minimum is quite pronounced, a secondary minimum being found at $7-8^h$; two rather diffuse maxima occur at about 10^h and 20^h. For Bergen there is a pronounced variability between the 5-years periods (thus r_{32} for 01^h is equal to 0.2 for 1906—10, and equal to 0.7 for 1921—25!) but the general form of the curves is the same for all groups, and the distribution of maxima and minima is one the whole as for Oslo. Similar conditions seem to obtain in March, but the night minimum is a little earlier than in February. Thus there is no difficulty in identifying for February—March the extremes m_1, m_2, M_1, and M_2 studied in detail for April. In May a maximum is indicated for Oslo at 07^h-09^h and a little later in Bergen; the minimum at 14^h-16^h is less pronounced than the night minimum at 04^h-05^h. Finally we have a maximum at 19^h-21^h. Also in June the 04^h minimum for Oslo and the 05^h minimum for Bergen are very marked; the maximum at 8^n-9^h, (somewhat variable in time for different values of j), the secondary minimum at 13^h-16^h, and the maximum at around 20^h are less pronounced.

The correlograms of July, fig. 10 are similar to those of June, with the Bergen extremes slightly later than the Oslo ones (minima at 04^h-05^h and 15^h-18^h, maxima at 19^h-20^h and 07^h-08^h). In August and September the night minimum is very pronounced, especially so for Oslo; also the day minima for $j = 2$ and 1 are well developed. In August the night minimum occurs at 05^h-06^h, in September at $05-07^h$. The afternoon minimum is similar to that of July; the evening maximum occurs at 20^h-21^h in August and at 18^h-19^h in September, the morning maximum at 07^h-09^h in August and at 08^h-09^h in September.

The October curves (fig. 11) are less wave-shaped, but still the 4 extremes are clearly visible: minima at 06^h-08^h and 14^h, maxima at 08^h-11^h and 18^h-19^h. In November the only reliable extreme is the minimum at 14^h-16^h, which is very weak for Bergen; in December a very weak minimum can be identified for Oslo.

The seasonal variation in the occurrences of the Oslo extremes is summarized in fig. 12, where also the time of the sunrise is indicated. m_1 shows a typical seasonal variation; its time of occurrence coincides fairly well with the minimum temperature. Analogously, m_2 occurs around the time of the temperature maximum. *Thus the noise level is highest at the times of the daily temperature extremes.*

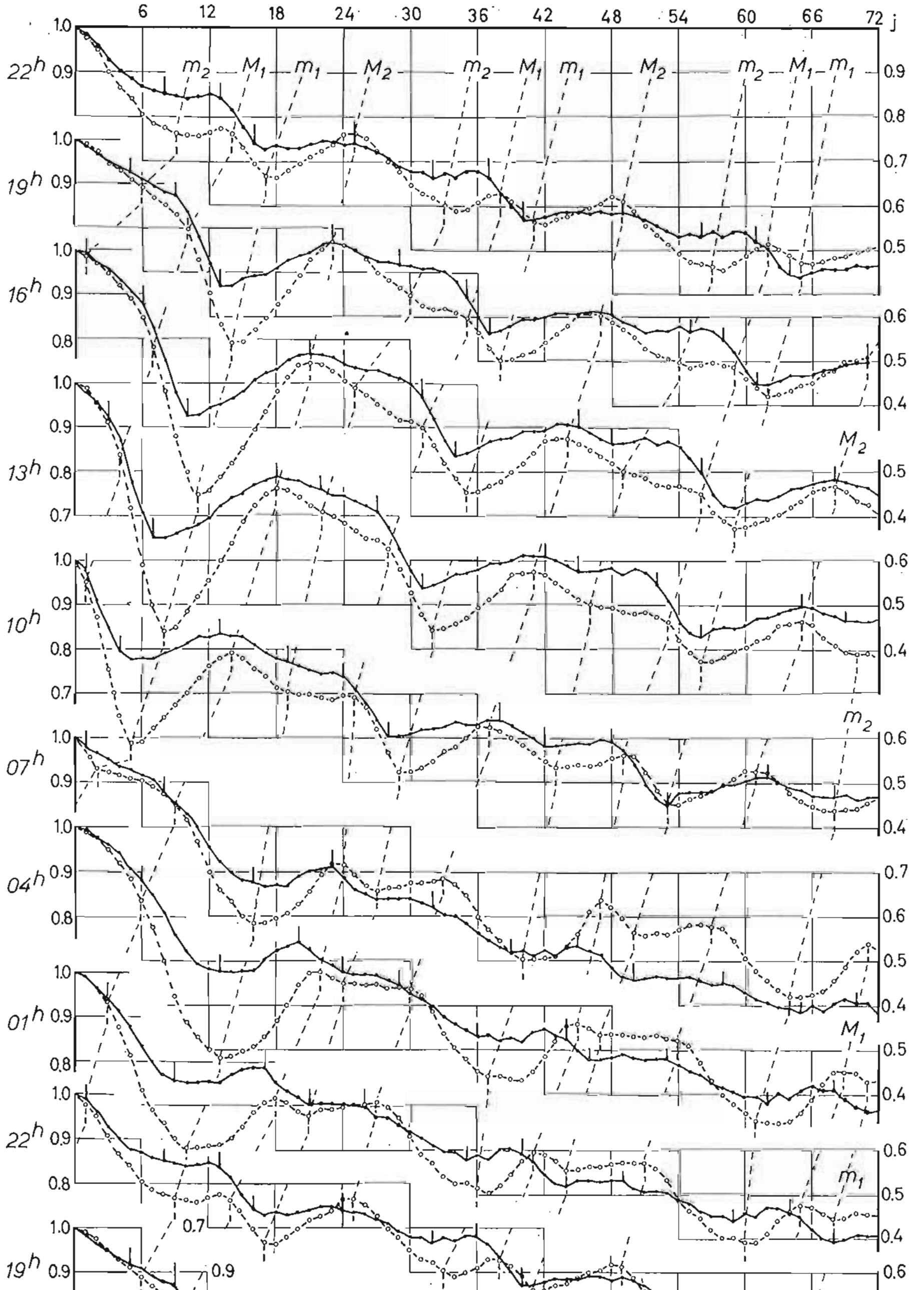

Fig. 8. Correlograms for the April air temperature in Bergen (25 years, continuous lines) and Oslo (30 years, dashed lines); time unit one hour.

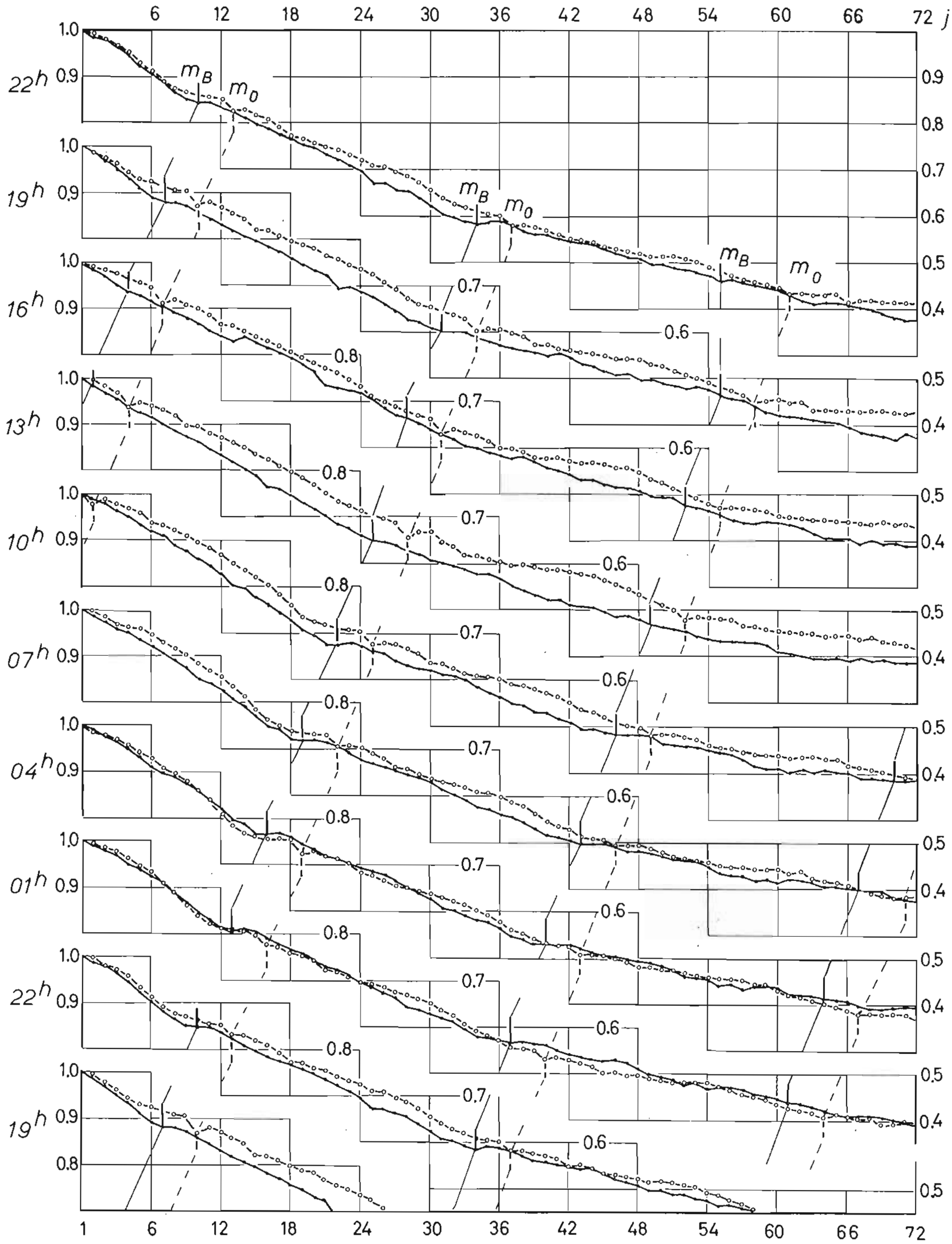

Fig. 9. Correlograms for the January air temperatures in Bergen and Oslo.

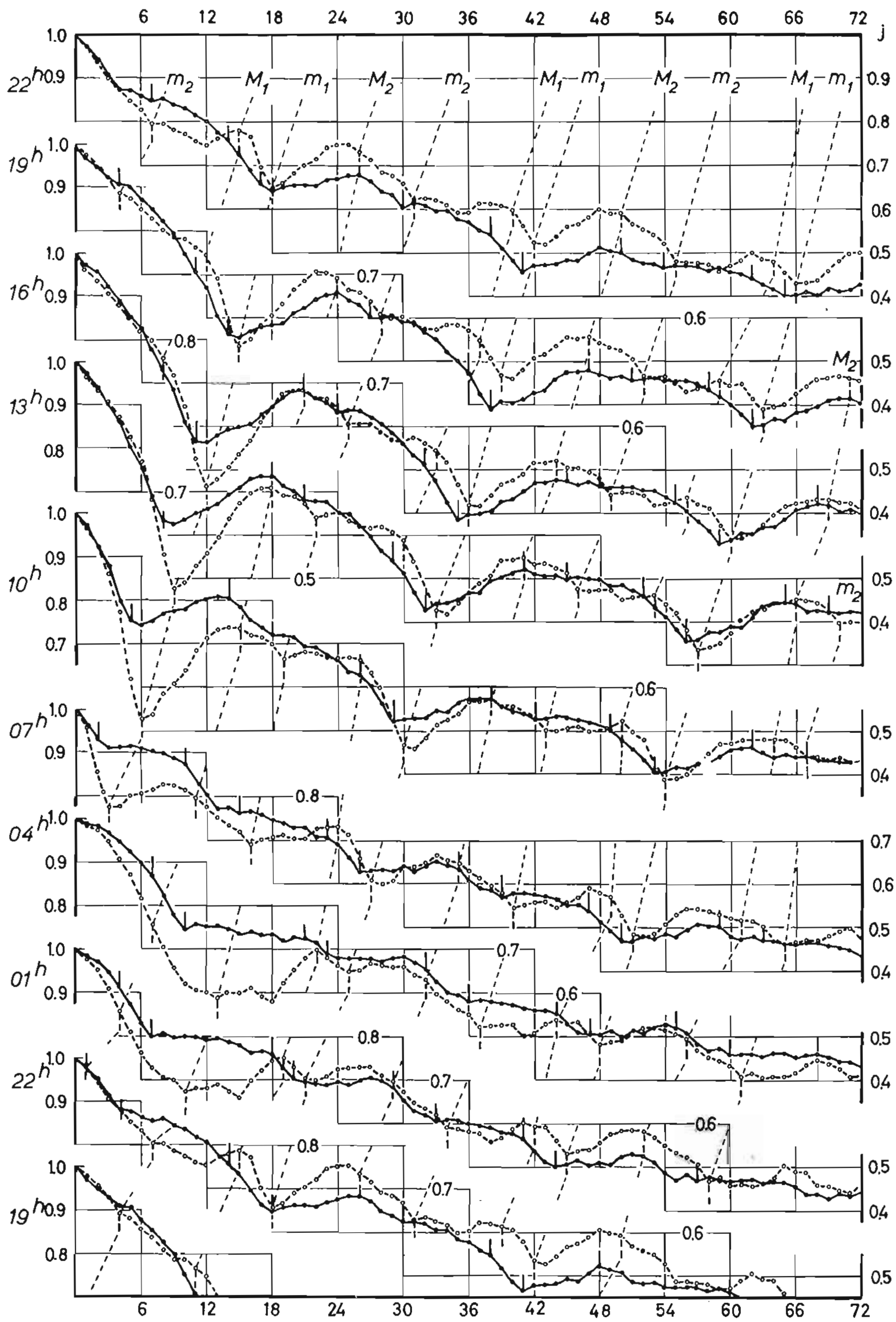

Fig. 10. Correlograms for the July air temperatures in Bergen and Oslo.

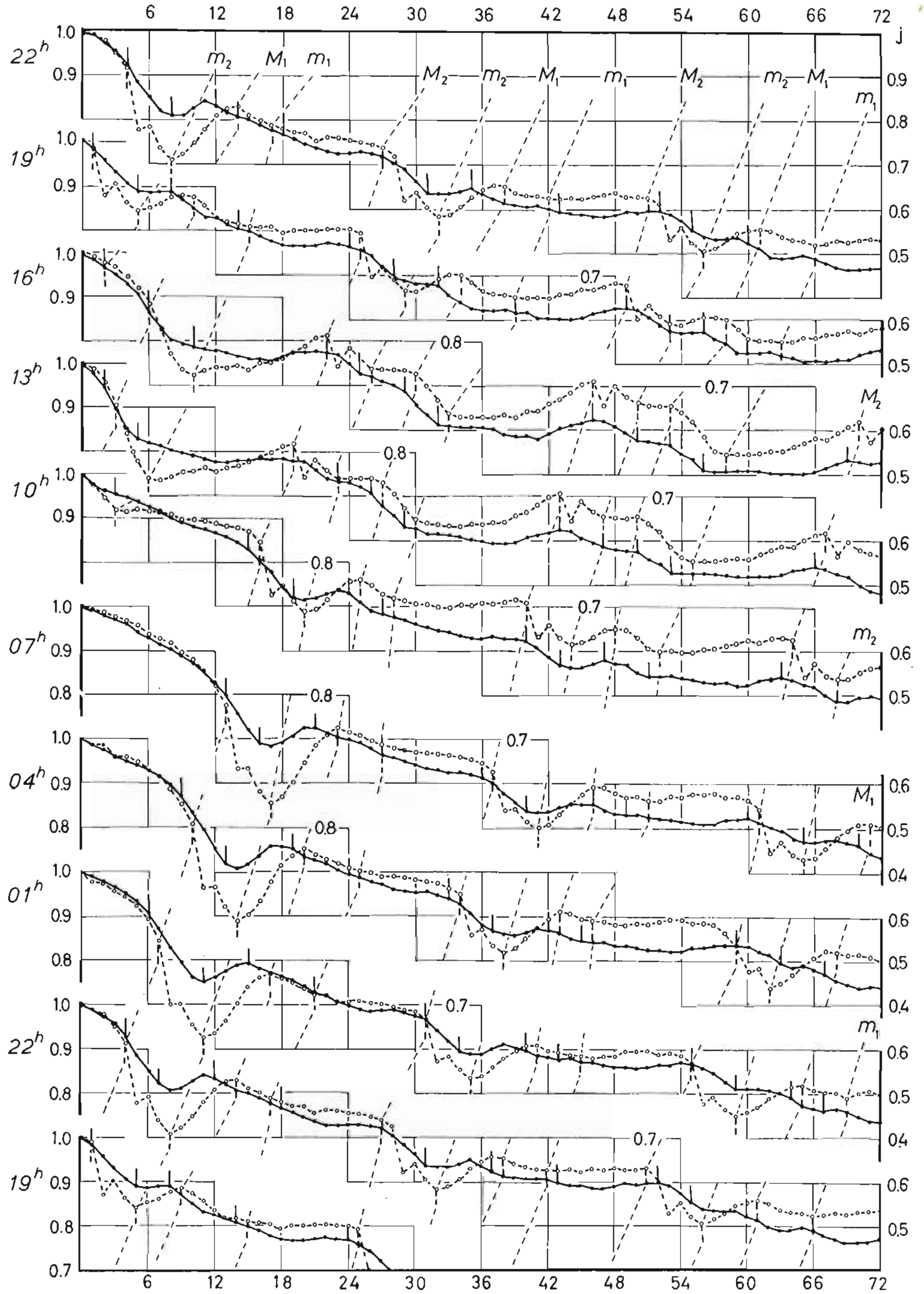

Fig. 11. Correlograms for the October air temperatures in Bergen and Oslo.

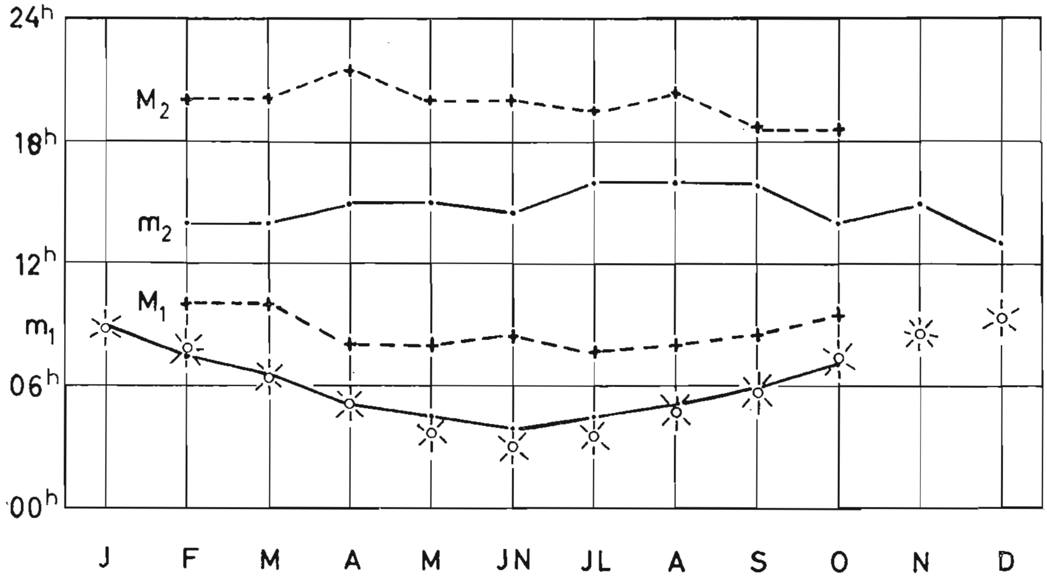

Fig. 12. Seasonal variation in the hours corresponding to relative minima or maxima in the temperature correlograms (or to high and low temperature noise levels.)

Fig. 13 finally, gives another representation of our computed correlation coefficients, namely for Oslo, April the values of $r(\dot{T}_j, T_{j+i})$ for $t_j = 05^h, 09^h, 14^h$ and 20^h corresponding roughly to the relative maxima and minima discussed above. The diagrams thus give the *prediction value* of T_j for i hours ahead. Since the original tables were computed for the main hours of observation, the coefficients in fig. 13 can only be given by 3 hours intervals. The diagram clearly shows the difference between the extremes in the prediction value; they are very strong for short range predictions (8—16 hours ahead) but less marked for 24 hours, as is also shown in fig. 4. Moreover, as shown by the scales at the bottom of the diagram, T_{05} (corresponding to m_1) has an extremely small prediction value for the hours 13^h—16^h corresponding to m_2; analogously, T_{14} (or m_2) has a small prediction value for T_{05} (or m_1). The fact that the April temperature maximum gives small information about the temperature minimum and vice versa is, of course, not surprising. In fact, in spring an increase in cloudiness may have small influence upon the daily average temperature, increase the minimum temperature, and decrease the maximum temperature. For M_1 (corresponding to 09^h) the "weakest points" in the predictions occur at around 14^h—16^h (near m_2) and 04^h (near m_1), the detailed determination being difficult since only 3 hours values are available; maxima occur near 20^h and 08^h. For $M_2(20^h)$ the extremes occur at almost the same time. Thus there seems to be a strong "mutual information" between M_1 and M_2 whereas m_1 and m_2 have small prediction value for all future hours.

It is reasonable to assume that the hours with great noise, or a low prediction value, also will have a smaller predictability than hours with great prediction value. This

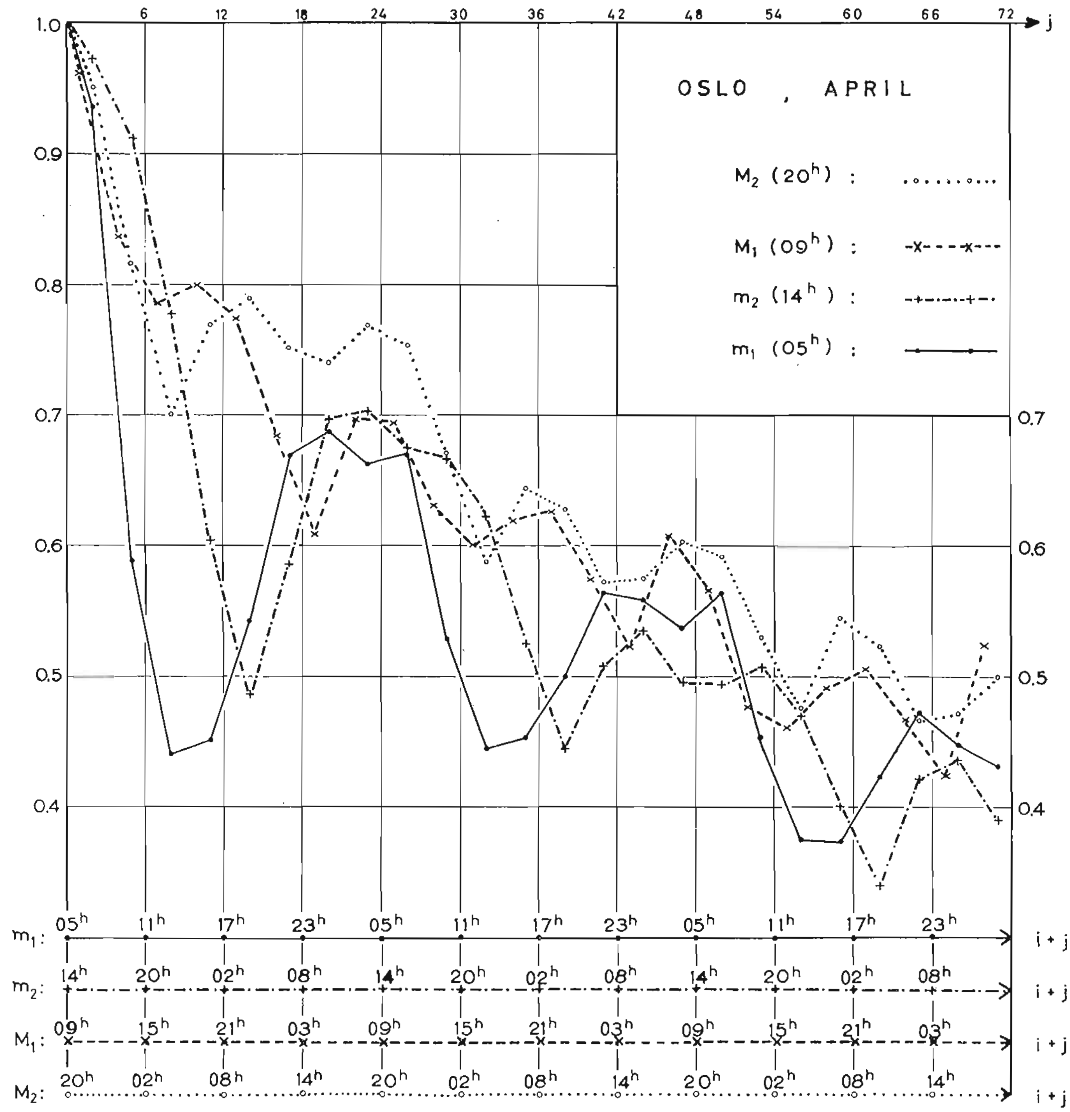

Fig. 13. Diagram giving the prediction value of T_{05}, T_{14}, T_{09}, and T_{20} for one future temperature.

is to a certain extent confirmed by fig. 3 and fig. 4, and shown in clearer form in fig. 14, which is analogous to fig. 6, but refers to those among the main hours of observation which can represent m_1, m_2, M_1, and M_2. The curves for T_{04} and T_{13} dip down much lower than the curves for T_{10} and T_{19}; moreover, the relative maximum for T_{04} is very low. We note that none of the curves in figs. 13 and 14 lie for all values of $j - i$ below or above the others, the reason being that the correlation coefficients are not only decreased by noise in the predictand, but also, and perhaps even more, by noise in the predictor.

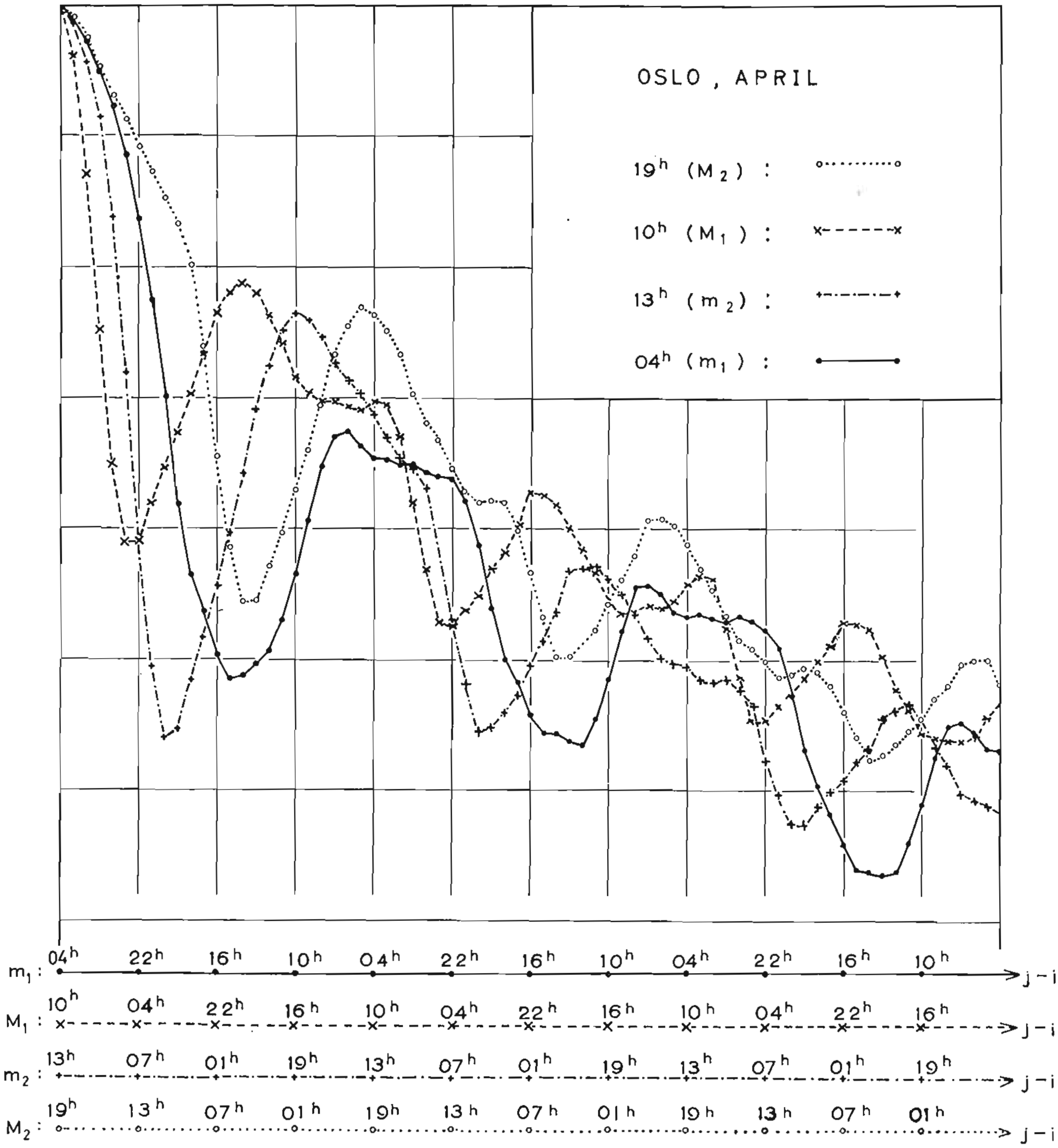

Fig. 14. Diagram giving the predictability of T_{04}, T_{13}, T_{10}, and T_{19} by one past temperature.

Summing up the results presented in figs. 6—14 we conclude that the temperature correlograms with time unit 1 hour need as a mathematical model a generalization of the stationary time series hitherto discussed. Such a model will be presented in the following chapter and applied summarily to the 30 years temperature data of Oslo in Chapter 3. However, a general and detailed discussion of the model and its meteorological interpretation and application will not be attempted in the present paper.

CHAPTER 2

A GENERALIZATION OF THE STATIONARY TIME SERIES

5. Definition of a mathematical model for a chain of stationary time series. Let us consider a stochastic variable y_t with expectation $\bar{y}_t$ and standard deviation s_t, defined at time t. Assuming t to vary discontinuously with steps equal to 1, we have thus defined a *stochastic process*. Generalizing the linear autoregressive scheme applied for stationary time series (see for instance, M. G. Kendall, p. 414), we assume successive values of y to be connected through the following formula:

2(1) $$s_{mt-j}^{-1}\,(y_{mt-j} - \bar{y}_{mt-j}) = \sum_{i=1}^{n_j} a_{j,i}\, s_{mt-j-i}^{-1}\,(y_{mt-j-i} - \bar{y}_{mt-j-i}) + \varepsilon_{mt-j},$$

where $$j = 0, 1, 2, \ldots m - 1.$$

ε_j is a stochastic variable (whose distribution may be assumed to be normal); $a_{j,i}$ are constants (satisfying certain conditions), and i, j, n, and m are integers.

Let us further suppose the stochastic variables to be characterized by the following equations:

2(2) $$\bar{\varepsilon}_t = 0 \text{ for all values of } t,$$

2(3) $$\bar{y}_{t_1} = \bar{y}_{t_2},\ s_{t_1}^2 = \overline{(y_{t_1} - \bar{y}_{t_1})^2} = s_{t_2}^2,\ \overline{\varepsilon_{t_1}^2} = \overline{\varepsilon_{t_2}^2} \text{ when } (t_2 - t_1)m^{-1} \text{ is an integer.}$$

2(4a) $$\overline{\varepsilon_t y_{t-k}} = 0,$$
2(4b) $$\overline{\varepsilon_t\, \varepsilon_{t-k}} = 0,$$
for all integer k.

The bars denote expectations, averages which can be defined as "phase averages" over different realizations of the stochastic processes or, more practically, as time averages estimated by adding the variables at k times mt, $(m + 1)t$, ... $(m + k)t$ and dividing by k, supposed to be sufficiently great.

Moreover, we assume for the constants:

2(5) $$a_{j_1,i} = a_{j_2,i}, \text{ when } (j_2 - j_1)m^{-1} \text{ is integer.}$$

Thus the erratic terms ε_t are uncorrelated with each other and with all preceding values of y, and a displacement in time equal to m units will not affect the statistical properties of the series. Under the assumptions (2)—(5) formulae (1) define a *chain of stationary time series*. Each of the series (1) will be denoted as a *primitive series* of the chain; m is called the *order* of the chain.

Writing down (1) for $j = 0,1, \ldots m - 1, m, \ldots$ we find by a process of successive eliminations m series with time step m:

$$2(6) \qquad s_{mt-j}^{-1}\,(y_{mt-j} - \bar{y}_{mt-j}) = \sum_{i=1}^{\infty} A_{j,i}\,(y_{mt-j-mi} - \bar{y}_{mt-j-mi}) + \varepsilon'_{mt-j},$$

where $j = 0, 1, 2, \ldots m-1.$

The series (6) will be denoted as "*m-series*" in the subsequent considerations. We assume the series to converge as will probably be the case in problems which have a clear physical interpretation. The erratic term in (6) may be written:

$$2(7) \qquad \varepsilon_t' = L(\varepsilon_t, \varepsilon_{t-1} \ldots \varepsilon_{t-m-1} \ldots),$$

where L denotes a linear function. Thus, using assumption (2), we find:

$$2(8) \qquad \overline{\varepsilon_t'} = 0.$$

In general ε'_{mt-j} will contain ε-terms which in time are prior to $y_{mt-j-mi}$ and thus contribute to this value. Consequently, we have:

$$2(9) \qquad \overline{\varepsilon'_{mt-j}\, y_{mt-j-mi}} \neq 0.$$

Moreover, ε'_{mt-j} and $\varepsilon'_{mt-j-mi}$ will in general have certain ε-terms in common, so that:

$$2(10) \qquad \overline{\varepsilon'_{mt-j}\, \varepsilon'_{mt-j-mi}} \neq 0.$$

Taking into consideration (3) and (5), we conclude that the m series represented by (6) are stationary time series for time lags m, $2m$, ..., but not of that type (with independent error terms) for which the "classical" theory, due to YULE, WOLD, and others, is valid.

The computations leading from (1) to (6) are elementary, but fairly complicated except when n and m are small, and lead to a *unique system of m-series*. The opposite problem, to find the m primitive series (1) when the m-series (6) are given, will in general be very complicated from the mathematical point of view. Since all m-series in general contain an infinity of terms, their coefficients must satisfy an infinite number of equations if n_j in the primitive series shall be finite. Moreover, it can be shown (see p. 193) that if series of type (6) can be considered as m-series, to one system of m-series there will correspond an infinity of primitive series of type (1), associated with different strengths of "coupling".

The quantities $a_{j,i}, y_{mt-j}$ and s_{mt-j} ($i = 1, 2, \ldots n_j, j = 0, 1, \ldots m-1$) give a complete definition of our chain of series. They must satisfy the following inequality, obtained by squaring (1), taking the expectations, using (4) and (5), and observing that $\overline{\varepsilon^2_{mt-j}} \geq 0$:

$$2(11) \qquad 1 \geq \sum_{i=1}^{n_j} \sum_{h=1}^{n_j} a_{j,i}\, a_{j,h}\, r\,(y_{mt-j-i}, y_{mt-j-h}).$$

The matrix

$$\{a_{j,i}\}$$

(dimension $m \times \max(n_j)$) will be denoted as the *matrix of tradition.* The number of rows corresponds to the order of the chain, whereas the number of columns will be called the *maximum length of the tradition; n_j* will be denoted as the *individual length of tradition* for row j.

Introducing a new matrix $a_{j,k}$ where $a_{j,k} = a_{j,k-j}$ we may write equation (1) in the ordinary form

2(12) $$\boldsymbol{y}_{mt} = \{a_{j,k}\} \boldsymbol{y}'_{mt-1}$$

where $\boldsymbol{y}_{mt}$ and $\boldsymbol{y}'_{mt-1}$ are the row vector $(y_{mt}, y_{my-1}, \ldots y_{mt-j+1})$ and the column vector $[y_{mt-1}, \ldots y_{mt-\max(j+n_j)}]$ respectively. However, the matrix $\{a_{j,k}\}$ is of a special form, having zeros below the diagonal starting at the upper left corner. Moreover, displacing the starting time in equations (1), we may essentially change the appearance of the corresponding a-matrix, whose order depends on the maximum of $j + n_j$ (as is easily seen in the simple chain treated in section 4). Therefore, we will not make use of the matrix notation (12), but concentrate our attention upon the coefficients $a_{j,i}$, an interpretation of which will be given in the following section. Moreover, we will "normalize" our variables introducing instead of y_t the quantity x_t defined by:

2(13) $$x_t = s_t^{-1} (y_t - \bar{y}_t).$$

Thus $s_{x_t} = 1$ for all values of t. The formulae characterizing the chain of stationary time series for the normalized variable x follow from (1) — (10) by substituting x, 0, 1 for $y, \bar{y}, s$ respectively.

A simple generalization of our model (1) would follow if we consider non-equidistant time steps $\tau_1, \tau_2, \ldots \tau_m$, so that

$$\sum_1^m \tau_j = m.$$

The mathematical developments will be practically the same as when $\tau_j = 1$; an application of the generalized model is given in example 10.

When applying the model (1) to, say, the time persistency of air temperature, the coefficients $a_{j,i}$ must be estimated from empirical data. The correlograms can then be used, giving estimates of the theoretical autocorrelation coefficients, the knowledge of which makes possible the determination of the regression coefficients $a_{j,i}$. We will not, at present, discuss the general formulae connecting regression and correlation coefficients; some few simple examples are given in sections 7—10. Moreover, a discussion of the difficult problems of "goodness of fit" of the theoretical models will not be attempted in the present paper.

7. Tradition interpretation of a chain of stationary time series. Let us consider a "normalized" property x, partly determined by "former values", partly by "noise". For simplification we first consider the "*Markov*" type chain (to be treated in more detail in section 3), for which $n_j = 1$.

If $m = 2$ we can illustrate the series by assuming the quantity x to represent a tradition following the "male line"; we assume that a man receives all tradition via his mother, whereas a woman obtains the tradition via her husband (see fig. 15d), x_m, x_{m-2}, x_{m-4}, ... thus refer to the man of generation no. 1, 2, 3 .. (reckoned backwards in time), x_{m-1}, x_{m-3}, .. to the woman of generation no. 2, 3,... No tradition passes in this case directly from father, grandparents, etc. to the son and daughter-in-law, only the mother is able to transfer tradition. If $a_{0,1} = 0$, the tradition is broken because the mother transmits to her son nothing of that tradition which she received from her husband; if $a_{1,1} = 0$ the tradition is broken because the wife cannot obtain from her husband any part of the tradition he received from his mother. In these cases the tradition can be characterized as "intermittent"; each man acts as a "source", each woman as a "sink" of tradition when $a_{0,1} = 0$, whereas the opposite is the case when $a_{1,1} = 0$.

Let us next assume, considering as before $n_j = 1$, that each generation is statistically different from the parent generation, but similar to the generation of grandparents, with respect to their sense of tradition, so that $m = 4$.

Counting the generations backward in time, we then have:

2(14)
$x_m, x_{m-4}, \ldots$ correspond to man of first, third, ... generation,
$x_{m-1}, x_{m-5}, \ldots$ correspond to woman of second, fourth, ... generation,
$x_{m-2}, x_{m-6}, \ldots$ correspond to man of second, fourth,... generation,
$x_{m-3}, x_{m-7}, \ldots$ correspond to woman of third, fifth,... generation.

Also in this case the tradition can be broken. If, say $a_{0,1} = 0$, the tradition is broken because, in every second generation, the woman is absolutely without any interest in her husband's family tradition.

As our next, non-Markovian, example, let us assume $m = 2$, $n_j = 2$, a case illustrated in fig. 15f. The son may then be assumed to receive tradition via mother and father, his wife via husband and mother-in-law. We may note that $a_{0,1} = 0$ ("mother contributes nothing to the male line tradition") does not mean that there is no correlation between mother and son (both are influenced by the son's father); in fact, discussing time series we have to be very careful in not confounding zero order and partial correlations and regressions.

$m = 2$, $n_0 = 3$, $n_1 = 2$ would correspond to a direct influence also of the grandmother upon her grandson; if also $n_1 = 3$, the father would transmit tradition directly to his daughter-in-law, etc.

Finally, considering as in (14) a periodicity in the generations we have for $m = 4$, $n_j = 4$:

son influenced directly from parents and grandparents,
wife » » » the husband, his parents, and his grandmother.

In the following sections, when discussing different types of chains and "sources" and "sinks" of information, the notion of tradition will prove very illustrative for cases in which $m = 2^n$ (see the diagram fig. 15).

7. The chain of series of Markov type. The chain with the minimum non vanishing tradition is obtained when $n_j = 1$ for all values of j. The value of x_{mt-j} then depends explicitly only upon x_{mt-j-1} since the normalized system of equations (1) can be written:

2(15) $$x_{mt-j} = a_{j,1}\, x_{mt-j-1} + \varepsilon_{mt-j} \quad (j = 0, 1, \ldots, m-1).$$

We will denote such a chain of stationary series as a *Markov type* chain. Taking the expectation of (15) after multiplication by x_{mt-j-1}, and introducing the notations:

$$r(x_{mt-j}, x_{mt-j-1}) = r_{j,1},$$

we find, since $s_x = 1$:

2(16) $$a_{j,1} = r_{j,1}.$$

Consequently, formulae (15) can be written:

2(17) $$x_{mt-j} = r_{j,1}\, m_{mt-j-1} + \varepsilon_{mt-j} \quad (j = 0, 1, \ldots, m-1).$$

The Markov chain of series of order m for a normalized variable is therefore completely known when we know m parameters, $r_{j,1}$, satisfying the conditions:

2(18) $$|r_{j,1}| \leq 1, \quad (j = 0, 1, \ldots, m-1).$$

For a non-normalized variable it is also necessary to know the m values of the standard deviations and the m mean values — see equation (1). In general, therefore, $3m$ parameters are needed to characterize completely a Markov chain of stationary time series of order m.

By successive eliminations from (17), we find the corresponding m-series:

2(19) $$x_{mt-j} = \left(\prod_{i=0}^{m-1} r_{j+i,1}\right) x_{mt-j-m} + \varepsilon_{mt-j} + \sum_{k=1}^{m-1} \left(\prod_{i=1}^{k}\right) \varepsilon_{mt-j-k}.$$

Introducing:

2(20) $$r_{j,m} = \prod_{i=0}^{m-1} r_{j+i,1},$$

and denoting by ε'_{mt-j} the total error term, we may write:

2(21) $$x_{mt-j} = r_{j,m}\, x_{m(t-1)-j} + \varepsilon'_{mt-j} \quad (j = 0, 1, \ldots, m-1),$$

which represent m simple stationary time series, because in this case the inequalities (9) and (10) reduce to equalities. Since $r_{j,m}$ is independent of j, the m-series, which are of Markov type, are identical.

(21) is a special case of the formulae, obtained by combining $n - 1$ of equations (17):

2(22) $$x_{mt-j} = \left(\prod_{i=0}^{n-1} r_{j+i,1}\right) x_{mt-j-n} + \varepsilon_{mt-j} + \sum_{k=1}^{n-1} \left(\prod_{i=1}^{k} r_{j+i,1}\right) \varepsilon_{mt-j-k}.$$

Consequently we have:

2(23) $$r\,(x_{mt-j}, x_{mt-j-n}) = r_{j,n} = \prod_{i=0}^{n-1} (r_{j+i,1})$$

enabling us to express the n-th order correlation coefficients by means of the given first order coefficients $r_{j,1}$.

(23) shows, in particular, that we always have $r_{j,n+1} \leq r_{j,n}$, so that *absolute maxima and minima* (like those observed in figs. 6, 9, 10, 11) *cannot occur in correlograms of Markov chains.* Consequently the Oslo and Bergen air temperatures at one hour of observation depend not only upon the immediately preceding temperatures, but also on temperatures in a more remote past.

Considering the tradition representation of our series, we note that in the Markov case the tradition length is 1 in the primitive series and m in the m-series, thus also in the m-series equal to one time step. Of course the tradition is much stronger in the primitive series than in the m-series. The strength of tradition may be different for the different primitive series, but all m-series have the same strength of tradition. If one or more of the coefficients $r_{j,1}$ vanishes, there will be a break of tradition in the corresponding primitive series, and *all m-series will have zero tradition*, i.e. $x_{mt-j}, x_{m(t-1)-j}, \ldots x_{m(t-n)-j}, \ldots$ will be stochastically independent for all values of j.

8. Special case when one m-series is of Markov type. When the primitive series are non-Markovian, the m-series will, as shown by (6), in general possess an infinity of terms (have an infinitely long tradition). However, if with a suitable labelling we can write the normalized series (1) in the form:

2(24) $$x_{mt-j} = \sum_{i=1}^{m-j} a_{j,i}\, x_{mt-j-i} + \varepsilon_{mt-j},$$

elimination of $x_{mt-j-1}, x_{mt-j-2}, \ldots x_{mt-j-m+1}$ leads to a stationary series of Markov type, i.e. a relation connecting x_{mt} and $x_{m(t-1)}$ with an error term uncorrelated with former errors and former x-values. The other m-series, however, contain an infinity of terms, and are of the general type (6).

Let us consider a simple example and discuss in some detail the equations:

2(25) $$x_{2t} = a_{0,1}x_{2t-1} + a_{0,2}x_{2t-2} + \varepsilon_{2t},$$

2(26) $$x_{2t-1} = a_{1,1}x_{2t-2} + \varepsilon_{2t-1}.$$

For the first m-series, to be called the *even series*, we find:

2(27) $$x_{2t} = (a_{0,1}a_{1,1} + a_{0,2})x_{2t-2} + \varepsilon_{2t} + a_{0,1}\varepsilon_{2t-1}.$$

Introducing for simplification:

2(28) $$y_t = x_{2t},\ b_1 = a_{0,1}a_{1,1} + a_{0,2},\ \varepsilon_t' = \varepsilon_{2t} + a_{0,1}\varepsilon_{2t-1},$$

we find the Markov type series:

2(29) $$y_t = b_1 y_{t-1} + \varepsilon_t',$$

which is seen to be of the ordinary type since for its random terms we have the equations:

2(30) $$\begin{cases} \overline{\varepsilon_{t_1}' \varepsilon_{t_2}'} = \overline{\varepsilon_{2t_1} \varepsilon_{2t_2}} + a_{0,1} (\overline{\varepsilon_{2t_1} \varepsilon_{2t_2-1}} + \overline{\varepsilon_{2t_1-1} \varepsilon_{2t_2}}) \\ + a_{0,1}^2 \overline{\varepsilon_{2t_1-1} \varepsilon_{2t_2-1}} = 0 \text{ for } t_1 \neq t_2, \\ \overline{\varepsilon_t' y_{t-i}} = \overline{\varepsilon_{2t} x_{2(t-i)}} + a_{0,1} \overline{\varepsilon_{2t-1} x_{2(t-i)}} = 0 \text{ for } i \neq 0, \end{cases}$$

as a consequence of conditions (4).

The second m-series, to be denoted the *odd series*, assumes the form:

2(31) $$\begin{cases} x_{2t-1} = a_{0,1} a_{1,1} \sum_{i=0}^{\infty} a_{0,2}^i x_{2t-3-2i} \\ \qquad + \varepsilon_{2t-1} + a_{1,1} \sum_{i=0}^{\infty} a_{0,2}^i \varepsilon_{2t-2-2i}. \end{cases}$$

Putting:

2(32) $$\begin{cases} z_t = x_{2t-1}, \quad c_i = a_{0,1} a_{1,1} a_{0,2}^{i-1}, \\ \varepsilon_t'' = \varepsilon_{2t-1} + a_{1,1} \sum_{i=0}^{\infty} a_{0,2}^i \varepsilon_{2t-2-2i}, \end{cases}$$

we can write the above series in the form:

2(33) $$z_t = \sum_{i=1}^{\infty} c_i z_{t-i} + \varepsilon_t''.$$

Since, as follows from (9) and (10):

2(34) $$\overline{\varepsilon''_t z_{t-i}} \neq 0,$$

and:

2(35) $$\overline{\varepsilon''_t \varepsilon''_{t-i}} \neq 0,$$

the error ε''_t at time t is neither independent of preceding errors, nor of the preceding z-values.

The first order correlation coefficients:

2(36) $$\begin{cases} r_{0,1} = r(x_{2t}, x_{2t-1}), \\ r_{0,2} = r(x_{2t}, x_{2t-2}), \\ r_{1,1} = r(x_{2t-1}, x_{2t-2}), \\ r_{1,2} = r(x_{2t-1}, x_{2t-3}), \end{cases}$$

can be expressed by means of the regression coefficients $a_{j,i}$. By direct computation, or by putting $a_{1,2} = 0$ in the general formulae (45) we find:

$$2(37)\quad \begin{cases} r_{0,1} = a_{0,1} + a_{0,2}a_{1,1}, \\ r_{1,1} = a_{1,1}, \\ r_{0,2} = a_{0,2} + a_{0,1}a_{1,1}, \\ r_{1,2} = a_{1,1}(a_{0,1} + a_{0,2}a_{1,1}) = r_{0,1}r_{1,1}. \end{cases}$$

The higher correlation coefficients $r_{0,n} = r(x_{2t}, x_{2t-n})$ and $r_{1,n-1} = r(x_{2t-1}, x_{2t-n})$ can be expressed by using equations (25) and (26). Multiplication with x_{2t-n} gives when expectations are taken:

$$2(38)\quad \begin{cases} r_{0,n} = a_{0,2}r_{0,n-2} + a_{0,1}r_{1,n-1}, & m > 2 \\ r_{1,n-1} = a_{1,1}r_{0,n-2}. & m > 3 \end{cases}$$

Using the given values of the first order coefficients and observing equations (37) we easily find:

$$2(39)\quad \begin{cases} r_{0,2n} = r^{n}_{0,2}, \\ r_{0,2n+1} = r_{0,1}r^{n}_{0,2}, \\ r_{1,2n-1} = r_{1,1}r^{n-1}_{0,2}, \\ r_{1,2n} = r_{0,1}r_{1,1}r^{n-1}_{0,2}. \end{cases}$$

Finally, the regression coefficients $a_{j,i}$ are expressed as functions of the first order correlation coefficients by the formulae:

$$2(40)\quad \begin{cases} a_{0,1} = \dfrac{r_{0,1} - r_{1,1}\, r_{0,2}}{1 - r^2_{1,1}}, \quad a_{0,2} = \dfrac{r_{0,2} - r_{0,1}\, r_{1,1}}{1 - r^2_{1,1}} \\ a_{1,1} = r_{1,1}, \qquad a_{1,2} = 0. \end{cases}$$

We may illustrate the chain just studied by using a *tradition diagram;* we assume the man to obtain his tradition via his parents, whereas the woman gets tradition information only from her husband (see fig. 15a). The left end of the diagram corresponds to the time $2t$, and time decreases towards the right; man is denoted by a cross, woman by a dot. The diagram shows that all information due to grandfather, great-grand-father, etc. passes through the father, so that the even "man" series is of Markovian type. Similarly we see that all earlier women contribute to the tradition received by a woman (thus W_2 obtains information from W_4 via M_3 and M_2 etc.). The odd "woman" series consequently must contain an infinity of terms. We note that two arrows end at M_i, and two start from M_i, but only one arrow ends, and one starts at W_i; thus the

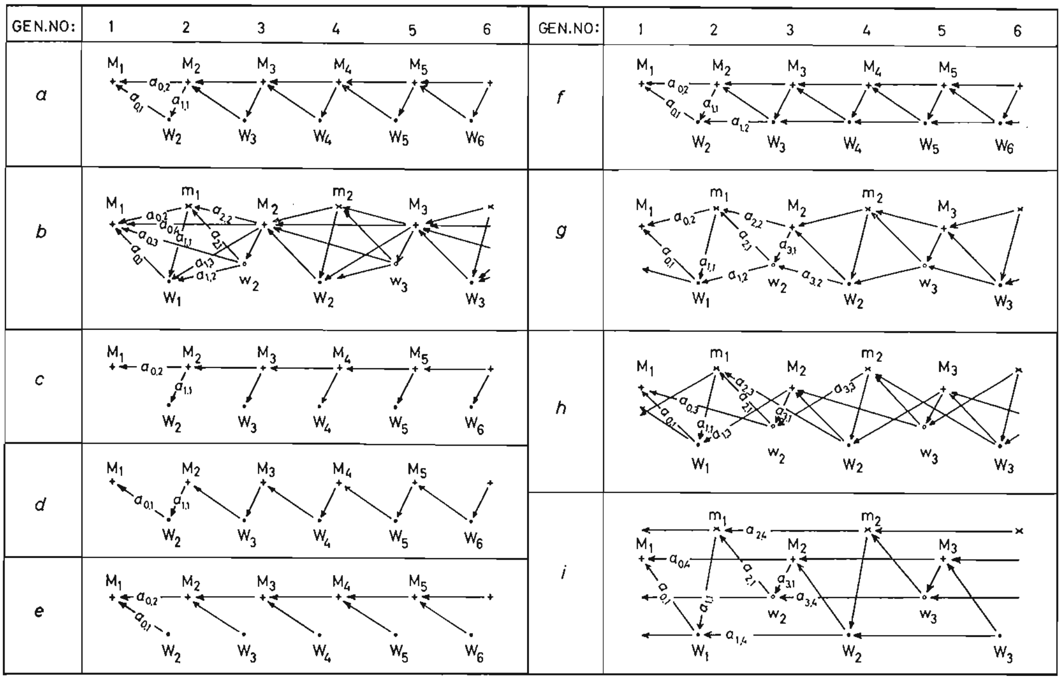

Fig. 15. Tradition diagrams for different chains of stationary series.

man receives tradition and transfers tradition in a more manifold way than the woman.

Let us now suppose all correlation coefficients to be positive, as is the case with the series for our air temperatures. Then $r_{0,2} > r_{0,1}r_{1,1}$ will correspond to $a_{0,2} > 0$; we will denote this case as "*super-Markovian*". On the other hand, $r_{0,2} < r_{0,1}r_{1,1}$ will lead to $a_{0,2} < 0$ ("*sub-Markovian*" case). In the latter case a direct "negative" influence of x_{2t-2} upon x_{2t} combines with the "positive" influence via x_{2t-1}. Moreover formulae (39) show, if no r-values are equal to 1, that:

$$r_{0,2m} > r_{0,2m+1}, \quad r_{1,2m-1} > r_{1,2m}.$$

Finally we note:

$$\frac{r_{1,2m}}{r_{1,2m+1}} = r_{0,1}r_{0,2}^{-1} = \frac{r_{0,2m-1}}{r_{0,2m}} \text{ for all } m\text{-values.}$$

According to $r_{0,1} > r_{0,2}$ or $r_{0,1} < r_{0,2}$ we thus have two essentially different cases. In the first case the correlogram slopes continuously downward from the left, in the second case *absolute maxima and minima appear*, displaced with one time unit when we pass from series (25) to series (26); examples will be given in Chapter 3. This case will necessarily also be super-Markovian, since $r_{0,2} > r_{0,1}r_{1,1}$ follows from $r_{0,2} > r_{0,1}$.

A generalization of the chain (25), (26) to $m = 4$ is given in fig. 15b. Two types of generations exist. In the first (M, w) corresponding to x_{2t} and x_{2t+1} the man has a "long" sense of tradition being influenced both by parents and grandparents, whereas his wife, with a "short" feeling of tradition, is only influenced by her husband. In the second type of generation (m, W) corresponding to x_{2t-2} and x_{2t-1} the man's sense of tradition is much "shorter", since tradition is received only via his parents; his wife, however, obtains tradition from both husband and parents-in-law. The diagram shows in a simple way that a Markov type tradition exists in the special "4-series" connecting all men with "long" sense of tradition. The women of type w receive information from one single channel, but transmit it through 3 channels, whereas conditions are opposite for women W; on the contrary the number (4 or 2) of channels carrying tradition *to* the man is equal to the number transmitting tradition *from* him to the future.

Let us consider the different types of "breaks in tradition" for the case $m = 2$. $a_{0,1} = 0$ breaks the tradition between man and mother, but a correlation between them still exists owing to the father's influence (fig. 15c). The odd series (31) degenerate, since all women represent "blind alleys" in the development of tradition. $a_{0,2} = 0$ breaks the direct connection between father and son (fig. 15d) and makes the primitive series as well as the odd and even series Markovian. If, finally, the tradition between wife and husband is broken ($a_{1,1} = 0$), as illustrated in fig. 15e, the mother just adds "noise" to the pure male tradition (her contribution is "uncorrelated with", or "orthogonal to", the male tradition). The odd female series then degenerate into a succession of independent stochastic variables. If both $a_{0,1}$ and $a_{1,1}$ vanish, the women are completely left out of the picture, and all traditions follow the Markovian pure male line. If $a_{0,1} = a_{0,2} = 0$, part of the tradition created by a man passes to his wife, but nothing to the next generation, and both y_t and z_t are independent of former values. Finally, if $a_{1,1} = a_{0,2} = 0$, only the mother can transmit tradition which, however, does not belong to the male line, and dies out with the son.

Breaks of tradition in the case $m = 4$ (fig. 15b) can be discussed in a similar way.

9. Chain of second order with length of tradition equal to 2. When $m = n_1 = n_2 = 2$, the normalized equations (1) reduce to:

2(41) $$x_{2t} = a_{0,1}x_{2t-1} + a_{0,2}x_{2t-2} + \varepsilon_{2t},$$

2(42) $$x_{2t-1} = a_{1,1}x_{2t-2} + a_{1,2}x_{2t-3} + \varepsilon_{2t-1}.$$

For the two series (6), as in section 8 called the *even series* and *the odd series*, we have:

2(43) $$\begin{cases} x_{2t} = (a_{0,2} + a_{0,1}a_{1,1})x_{2t-2} + a_{0,1}a_{1,1}\sum_{i=1}^{\infty} a_{1,2}^{i}x_{2t-2-2i} \\ \quad + \varepsilon_{2t} + a_{0,1}\sum_{i=0}^{\infty} a_{1,2}^{i}\varepsilon_{2t-1-2i}, \end{cases}$$

and:

2(44)
$$\begin{cases} x_{2t-1} = (a_{1,2} + a_{1,1}a_{0,1})x_{2t-3} + a_{1,1}a_{0,1}\sum_{i=1}^{\infty} a_{0,2}^{i}x_{2t-3-2i} \\ \quad + \varepsilon_{2t-1} + a_{1\,1}\sum_{i=0}^{\infty} a_{0,2}^{i}\varepsilon_{2t-2-2i}. \end{cases}$$

The general inequalities (9) and (10) will hold, so that the stationary m-series, even in this simple case, are not of the type generally considered.

Let us now consider two normalized series of type (6) with time step $m = 2$. If these series shall have (41) and (42) as primitive series, we find comparing (6) with (43) and (44):

$$\frac{A_{0,i+1}}{A_{0,i}} = a_{1,2}, \quad \frac{A_{1,i+1}}{A_{1,i}} = a_{0,2} \quad \text{for } i = 2, 3, \ldots$$

Moreover, $i = 2$ gives:

$$a_{0,1}a_{1,1}a_{1,2} = A_{0,2}, \quad a_{0,1}a_{1,1}a_{0,2} = A_{1,2}$$

which leads to:

$$\frac{A_{1,2}}{A_{0,2}} = \frac{a_{0,2}}{a_{1,2}}.$$

For $i = 1$ we have:

$$A_{0,1} - A_{1,1} = a_{0,2} - a_{1,2}, \quad A_{0,1} = a_{0,2} + a_{1,1}a_{0,1}.$$

If these conditions are fulfilled, $a_{0,2}$, $a_{1,2}$ and the product $a_{0,1}a_{1,1}$ are unambiguously defined through the coefficients in (43) and (44). Thus the two "coefficients of coupling", $a_{0,1}$ and $a_{1,1}$, which connect the two series, are not determined, but only their product. Passing to the general case we thus conclude that *to a system of m-series, satisfying certain conditions, corresponds an infinity of primitive series, with different types of coupling.*

The regression coefficients $a_{0,1}$, $a_{0,2}$, $a_{1,1}$, $a_{1,2}$ can be expressed by means of the corresponding first order correlation coefficients $r_{0,1}$, $r_{0,2}$, $r_{1,1}$, $r_{1,2}$. In fact, after multiplications of (41) by x_{2t-1} and x_{2t-2}, and of (42) by x_{2t-2} and x_{2t-3}, we find taking the expectations, and observing (4):

2(45)
$$\begin{cases} r_{0,1} = \dfrac{a_{0,1} + a_{0,2}a_{1,1}}{1 - a_{0,2}a_{1,2}}, \\[2ex] r_{1,1} = \dfrac{a_{1,1} + a_{1,2}a_{0,1}}{1 - a_{0,2}a_{1,2}}, \\[2ex] r_{0,2} = \dfrac{a_{0,2} + a_{0,1}a_{1,1} - a_{0,2}^{2}a_{1,2} + a_{0,1}^{2}a_{1,2}}{1 - a_{0,2}a_{1,2}}, \\[2ex] r_{1,2} = \dfrac{a_{1,2} + a_{0,1}a_{1,1} - a_{1,2}^{2}a_{0,2} + a_{1,1}^{2}a_{0,2}}{1 - a_{0,2}a_{1,2}}. \end{cases}$$

The inverse formulae, giving the regression coefficients as functions of the correlation coefficients, can be written:

$$2(46) \qquad \begin{cases} a_{0,1} = \dfrac{r_{0,1} - r_{1,1}r_{0,2}}{1 - r_{1,1}^2}, \quad a_{0,2} = \dfrac{r_{0,2} - r_{0,1}r_{1,1}}{1 - r_{1,1}^2}, \\ a_{1,1} = \dfrac{r_{1,1} - r_{0,1}r_{1,2}}{1 - r_{0,1}^2}, \quad a_{1,2} = \dfrac{r_{1,2} - r_{0,1}r_{1,1}}{1 - r_{0,1}^2}. \end{cases}$$

Since $|r_{i,j}| < 1$, certain restrictions must be put on the regression coefficients.

The higher correlation coefficients $r_{0,n} = r(x_{2t}, x_{2t-n})$ and $r_{1,n-1} = r(x_{2t-1}, x_{2t-n})$ can be expressed by the lower ones. We multiply equations (41) and (42) by x_{2t-n}, take expectations, and use (4) together with $r_{2,n} = r_{0,n-2}$, $r_{3,n} = r_{1,n-2}$:

$$2(47) \qquad r_{0,n} = a_{0,2}r_{0,n-2} + a_{0,1}r_{1,n-1}, \quad n > 2,$$

$$2(48) \qquad r_{1,n-1} = a_{1,2}r_{1,n-3} + a_{1,1}r_{0,n-2}, \quad n > 3.$$

These formulae make possible a successive computation of the correlation coefficients when $a_{0,1}$, $a_{0,2}$, $a_{1,1}$, and $a_{1,2}$ — or $r_{0,1}$, $r_{0,2}$, $r_{1,1}$, and $r_{1,2}$ — are known.

The tradition diagram corresponding to the above formulae is given in fig. 15f. The man receives direct tradition from mother and father, the woman from husband and mother-in-law. No Markov chains can be defined by considering pure male or pure female lines. The different types of breaking the flow of tradition can be discussed by means of the diagram as was done at the end of the preceding section.

10. Simple non-Markovian chains of order 4. In our discussion of the Oslo and Bergen temperature correlograms for the different months, we noted that two minima and two maxima per day could be identified. In order to reproduce this feature by a chain of stationary time series, its order must be at least 4 — and in fact ought to be much greater. Even the chain of order 4 is quite complicated to discuss; at present we will consider only some few simple cases, reserving to later a more systematic treatment.

a. Chain of order 4 and tradition length 2. The model is defined by the following equations:

$$2(49) \qquad \begin{cases} x_{4t} = a_{0,1}x_{4t-1} + a_{0,2}x_{4t-2} + \varepsilon_{4t}, \\ x_{4t-1} = a_{1,1}x_{4t-2} + a_{1,2}x_{4t-3} + \varepsilon_{4t-1}, \\ x_{4t-2} = a_{2,1}x_{4t-3} + a_{2,2}x_{4t-4} + \varepsilon_{4t-2}, \\ x_{4t-3} = a_{3,1}x_{4t-4} + a_{3,2}x_{4t-5} + \varepsilon_{4t-3}. \end{cases}$$

It is illustrated by the tradition diagram fig. 15g; two different generations exist; in both the woman receives tradition via husband and mother-in-law, the man via mother and father.

Taking expectations of the first equation multiplied by x_{4t-1}, the second by x_{4t-2}, etc., we find:

$$2(50)\quad \begin{cases} r_{0,1} - a_{0,2}r_{1,1} & = a_{0,1}, \\ r_{1,1} - a_{1,2}r_{2,1} & = a_{1,1}, \\ r_{2,1} - a_{2,2}r_{3,1} & = a_{2,1}, \\ - a_{3,2}r_{0,1} \qquad\qquad r_{3,1} & = a_{3,1}. \end{cases}$$

Similarly, by multiplying the first equation (49) by x_{4t-2} the second by x_{4t-3} etc., we find, taking expectations:

$$2(51)\quad \begin{cases} r_{0,2} = a_{0,1}r_{1,1} + a_{0,2}, \\ r_{1,2} = a_{1,1}r_{2,1} + a_{1,2}, \\ r_{2,2} = a_{2,1}r_{3,1} + a_{2,2}, \\ r_{3,2} = a_{3,1}r_{0,1} + a_{3,2}. \end{cases}$$

Equations (50) give the first order correlation coefficients as functions of the partial regression coefficients $a_{j,1}$ and $a_{j,2}$; the second order coefficients are afterwards found by (51). The higher order correlation coefficients are determined by the equations (where $r_{j+4,n} = r_{j,n}$, $a_{j+4,n} = a_{j,n}$):

$$2(52)\qquad r_{j,n} = a_{j,1}r_{j+1,n-1} + a_{j,2}r_{j+2,n-2} \qquad (j = 0, 1, 2, 3),$$

valid for $n > 2$.

Solving (50) for the correlation coefficients, we find:

$$2(53)\qquad r_{j,1} = \frac{a_{j,1} + a_{j+1,1}a_{j,2} + a_{j+2,1}a_{j,2}\,a_{j+1,2} + a_{j+3,1}a_{j,2}\,a_{j+2,2}a_{j+3,2}}{1 - \prod_{i=0}^{3} a_{i,2}}$$

where $j = 0, 1, 2, 3.$

From (50) and (51) we find:

$$2(54)\qquad a_{j,1} = \frac{r_{j,1} - r_{j+1,1}r_{j,2}}{1 - r_{j+1,1}^2}, \quad a_{j,2} = \frac{r_{j,2} - r_{j+1,1}r_{j,1}}{1 - r_{j+1,1}^2},$$

expressing the regression coefficients as functions of the correlation coefficients. Under Markov conditions $r_{j,2} = r_{j,1}r_{j+1,1}$ corresponding to $a_{j,2} = 0$. As in section 8 the primitive series (49) may be both super- and sub-Markovian; in the same chain some series may be super-, other series sub-Markovian (see also example 12).

b. Special chain of order 4 and tradition length 3. The model is defined by the equations:

$$2(55)\quad \begin{cases} x_{4t} = a_{0,1}x_{4t-1} + a_{0,3}x_{4t-3} + \varepsilon_{4t}, \\ x_{4t-1} = a_{1,1}x_{4t-2} + a_{1,3}x_{4t-4} + \varepsilon_{4t-1}, \\ x_{4t-2} = a_{2,1}x_{4t-3} + a_{2,3}x_{4t-5} + \varepsilon_{4t-2}, \\ x_{4t-3} = a_{3,1}x_{4t-4} + a_{3,3}x_{4t-6} + \varepsilon_{4t-3}. \end{cases}$$

Using the "two-generation" model mentioned on p. 192, we see (fig. 15h) that the man in this case obtains his tradition only via mother and grandmother, the woman via husband and father-in-law. Multiplying the first equation by x_{4t-1} and by x_{4t-2}, the next by x_{4t-2} and by x_{4t-3}, etc. we find, by taking expectations, 8 equations between the a-coefficients and the correlation coefficients $r_{i,1}$ and $r_{i,2}$:

$$2(56)\quad \begin{cases} r_{0,1} - a_{0,3}r_{1,2} = a_{0,1}, \\ r_{0,2} - a_{0,1}r_{1,1} - a_{0,3}r_{2,1} = 0\ , \\ r_{1,1} - a_{1,3}r_{2,2} = a_{1,1}, \\ r_{1,2} - a_{1,1}r_{2,1} - a_{1,3}r_{3,1} = 0\ , \\ r_{2,1} - a_{2,3}r_{3,2} = a_{2,1}, \\ - a_{2,3}r_{0,1} + r_{2,2} - a_{2,1}r_{3,1} = 0\ , \\ - a_{3,3}r_{0,2} + r_{3,1} = a_{3,1}, \\ - a_{3,1}r_{0,1} - a_{3,3}r_{1,1} + r_{3,2} = 0\ . \end{cases}$$

Multiplying the first equation (55) by x_{4t-3}, the second by x_{4t-4}, etc., we find analogously equations giving $r_{j,3}$ when $r_{j,1}$ and $r_{j,2}$ are known:

$$2(57)\quad r_{j,3} = a_{j,1}r_{j+1,2} + a_{j,3} \qquad (j = 0, 1, 2, 3).$$

The higher correlation coefficients are found by the analogous equation for $n > 3$:

$$2(58)\quad r_{j,n} = a_{j,1}r_{j+1,n-1} + a_{j,3}r_{j+3,n-3} \qquad (j = 0, 1, 2, 3)$$

which enable us successively to determine $r_{j,n}$.

The determinant of equations (56) is:

$$2(59)\quad \begin{cases} D = 1 - a_{0,1}a_{2,1}a_{1,3}a_{3,3} - a_{1,1}a_{3,1}a_{2,3}a_{0,3} \\ \qquad - a_{0,3}a_{1,3}a_{2,3}a_{3,3}(a_{0,1}a_{1,3} + a_{1,1}a_{2,3} + a_{2,1}a_{3,3} + a_{3,1}a_{0,3} \\ \qquad - a_{0,1}a_{1,1}a_{2,1}a_{3,1}) - a_{0,3}^2a_{1,3}^2a_{2,3}^2a_{3,3}^2, \end{cases}$$

and the general solution of the equations will be quite complicated. We will, therefore, only consider the *quasi-Markovian* case, in which $a_{j,3}$ are so small that powers and products higher than the second order are negligible. In this case (59) reduces to:

2(59') $$D \approx 1 - a_{0,1}a_{2,1}a_{1,3}a_{3,3} - a_{1,1}a_{3,1}a_{2,3}a_{0,3},$$

and the correlation coefficients $r_{j,1}$ and $r_{j,2}$ are given by the expressions:

2(60) $$\begin{cases} r_{j,1} \approx D^{-1}(a_{j,1} + a_{j,3}a_{j+1,1}a_{j+2,1} + a_{j,3}a_{j+1,3}a_{j+3,1} \\ \qquad\qquad - a_{j+1,3}a_{j+3,3}a_{j,1}^2a_{j+2,1}) \\ \quad \approx a_{j,1} + a_{j+1,1}a_{j+2,1}a_{j,3} + a_{j+3,1}(a_{j,3} + a_{j,1}a_{j+1,1}a_{j+2,3})a_{j,3} \\ r_{j,2} \approx D^{-1}(a_{j,1}a_{j+1,1} + a_{j+2,1}a_{j,3} + a_{j+1,3}a_{j,1}a_{j+2,1}a_{j+3,1} \\ \qquad + a_{j+1,3}a_{j+2,3}a_{j,1}^2 + a_{j,3}a_{j+2,3}a_{j,1}a_{j+3,1} \quad (1 - a_{j+1,1}^2)) \\ \quad \approx a_{j,1}a_{j+1,1} + a_{j+2,1}(a_{j,3} + a_{j,1}a_{j+3,1}a_{j+1,3}) \\ \qquad + a_{j,1}a_{j+3,1}a_{j,3}a_{j+2,3} + a_{j,1}^2(a_{j+2,3} + a_{j+1,1}a_{j+2,1}a_{j+3,3})a_{j+1,3}. \end{cases}$$

The expressions for $a_{j,i}$ as functions of the correlation coefficients are simpler, $a_{0,1}$ and $a_{0,3}$ only appearing in the first and second equation (56), $a_{1,1}$ and $a_{1,3}$ only in the third and fourth etc. In analogy with (54) we thus find:

2(61) $$\begin{cases} a_{j,1} = \dfrac{r_{j,1}r_{j+2,1} - r_{j,2}r_{j+1,2}}{r_{j+2,1} - r_{j+1,1}r_{j+1,2}}, \\ a_{j,3} = \dfrac{r_{j,2} - r_{j,1}r_{j+1,1}}{r_{j+2,1} - r_{j+1,1}r_{j+1,2}}. \end{cases}$$

We note that in this special model regression coefficients of order *unity* and *three* can be computed from correlation coefficients of order *unity* and *two*.

c. Special chain of order 4 and tradition length 4. Finally we consider the non-Markovian chain of series:

2(62) $$\begin{cases} x_{4t} = a_{0,1}x_{4t-1} + a_{0,4}x_{4t-4} + \varepsilon_{4t}, \\ x_{4t-1} = a_{1,1}x_{4t-2} + a_{1,4}x_{4t-5} + \varepsilon_{4t-1}, \\ x_{4t-2} = a_{2,1}x_{4t-3} + a_{2,4}x_{4t-6} + \varepsilon_{4t-2}, \\ x_{4t-3} = a_{3,1}x_{4t-4} + a_{3,4}x_{4t-7} + \varepsilon_{4t-3}. \end{cases}$$

The tradition diagram is presented in fig. 15i; the woman receives tradition via husband and his grandmother, the man via mother and grandfather. Multiplying the first equation by x_{4t-1}, x_{4t-2}, x_{4t-3}, the second by x_{4t-2}, x_{4t-3}, x_{4t-4}, etc. we find, after taking expectations, a system analogous to (56) but consisting of 12 equations serving for the determination of $r_{j,i}(j = 0, 1, 2, 3; i = 1, 2, 3)$. Multiplying the first equation (62) by x_{4t-4}, the second by x_{4t-5}, etc. we find equations analogous to (57) for determining $r_{j,4}$.

If the correlation coefficients $r_{j,1}$, $r_{j,2}$, and $r_{j,3}$ are known and satisfy the quite complex conditions which are necessary for a model of type (62) to be applicable, the corresponding a-values can be determined by one of the 3 types of formulae, analogous to (61):

$$2(63)\quad \begin{cases} a_{j,4} = \dfrac{r_{j,2} - r_{j,1}r_{j+1,1}}{r_{j+2,2} - r_{j+1,1}r_{j+1,3}}, & a_{j,1} = \dfrac{r_{j,1}r_{j+2,2} - r_{j,2}r_{j+1,3}}{r_{j+2,2} - r_{j+1,1}r_{j+1,3}}, \\ a_{j,4} = \dfrac{r_{j,3} - r_{j,1}r_{j+1,2}}{r_{j+3,1} - r_{j+1,2}r_{j+1,3}}, & a_{j,1} = \dfrac{r_{j,1}r_{j+3,1} - r_{j+1,3}r_{j,3}}{r_{j+3,1} - r_{j+1,2}r_{j+1,3}}, \\ a_{j,4} = \dfrac{r_{j,2}r_{j+1,2} - r_{j,3}r_{j+1,1}}{r_{j+1,2}r_{j+2,2} - r_{j+1,1}r_{j+3,1}}, & a_{j,1} = \dfrac{r_{j,3}r_{j+2,2} - r_{j+3,1}r_{j,2}}{r_{j+1,2}r_{j+2,2} - r_{j+1,1}r_{j+3,1}}, \end{cases}$$

where $j = 0, 1, 2, 3$, and the formulae $r_{j+4,i} = r_{j,i}$ must be observed. Only 2 of these 3 sets of formulae are independent.

The solution of the 12 linear equations for the lowest correlation coefficients as functions of the 8 regression coefficients is very laborious. Assuming a quasi-Markovian regime (see *b*) we find, by keeping only linear terms in $a_{j,4}$:

$$2(64)\quad \begin{cases} r_{j,1} = a_{j,1} + a_{j+1,1}a_{j+2,1}a_{j+3,1}a_{j,4}, \\ r_{j,2} = a_{j,1}a_{j+1,1} + a_{j+2,1}a_{j+3,1}(a_{j,4} + a_{j,1}^2a_{j+1,4}), \\ r_{j,3} = a_{j,1}a_{j+1,1}a_{j+2,1} + a_{j+3,1}(a_{j,4} + a^2_{j,1}(a_{j+1,4} + a_{j+1,1}^2a_{j+2,4})), \end{cases}$$

where $j = 0, 1, 2, 3$.

The higher order correlation coefficients are determined successively by the formulae:

$$2(65)\qquad r_{j,n} = a_{j,1}r_{j+1,n-1} + a_{j,4}r_{j,n-4}, \quad n > 5,$$

derived by the method indicated in subsection *b*.

The expressions (63) contain correlation coefficients of order 1, 2, and 3. If we estimate these 12 coefficients from an empirical correlogram, we cannot expect the model (62) to apply. However, by estimating only $r_{j,1}$ and $r_{j,2}$ we have a sufficient number of equations for the determination of $a_{j,1}$, $a_{j,4}$, and $r_{j,3}$ — but these equations are *not linear*. By laborious computations we can determine quadratic equations for one of the unknown, say $a_{0,4}$, and find that solution which vanishes in the Markovian case; the other a-values can then afterwards be determined by linear equations. We will, at present, not enter into the discussion of this fairly complicated problem.

CHAPTER 3

NUMERICAL EXAMPLES OF CHAINS OF TIMES SERIES WITH APPLICATION TO AIR TEMPERATURES

11. Examples of Markov chain of series. The following 3 examples illustrate the most important properties of the series discussed in section 8.

Example 1. Let us consider the "single-period" chain of order 8 presented in fig. 16, upper left; the value of the first order correlation coefficients $r_{j,1}$ are given at the top of the diagram. In analogy with figs. 8—11 the curves have been displaced vertically, the value j increasing from top to bottom. The curves show within the "period", equal to 8, one relative maximum and one relative minimum ("absolute" extremes cannot exist in the Markov case), which have a *time displacement of identically the same type as that observed on the temperature diagrams figs. 8—11.*

It is interesting to note that the curve for $r_{j,i}$ may be situated completely below the exponential curve (see $j = 3$ in fig. 16) or completely above it (see $j = 7$); consequently, "average" curves ought not to be drawn "by intuition".

Example 2. The "double-period" chain of order 8, represented in the lower left part of fig. 16, has within the period two relative maxima and two relative minima (one minimum and one maximum is extremely flat) and shows the characteristic time displacement. The general form of this diagram resembles the correlograms for the winter month temperatures, where no absolute maxima and minima occur.

Example 3. In fig. 16 right we have chosen for the first order correlation coefficients (top) those found for the Oslo temperatures for April, using a time lag of 3 hours. The "average" exponentials have been drawn (dashed lines) and also the empirical correlograms (dot-dashed lines). There is very small similarity between the Markov curves and the empirical ones; our former conclusion, that the temperature is not governed by time series of Markov type, is thus strongly confirmed.

12. Examples of simple chains of order 2. The simplest generalization of Markov chains of series is the chain of the type considered in section 9. Fig. 17 represents some examples of this type.

Example 4. Assuming $m = 2$, $n_0 = 2$, $n_1 = 1$, $r_{0,1} = 0.85$, $r_{0,2} = 0.70$, $r_{1,1} = 0.90$, we have, according to the remarks on p. 191 a chain of sub-Markovian type. Its 8 first correlation coefficients, computed by using equations 2(46)—2(48) are given in fig. 17, upper left diagram. The correlograms are situated systematically below the Markovian ones, corresponding to $r_{0,1} = 0.85$ and $r_{1,1} = 0.90$ respectively. Moreover, the slightly wavy shape in the Markovian correlogram, indicating relative maxima and minima, has been almost destroyed in the non-Markovian figure.

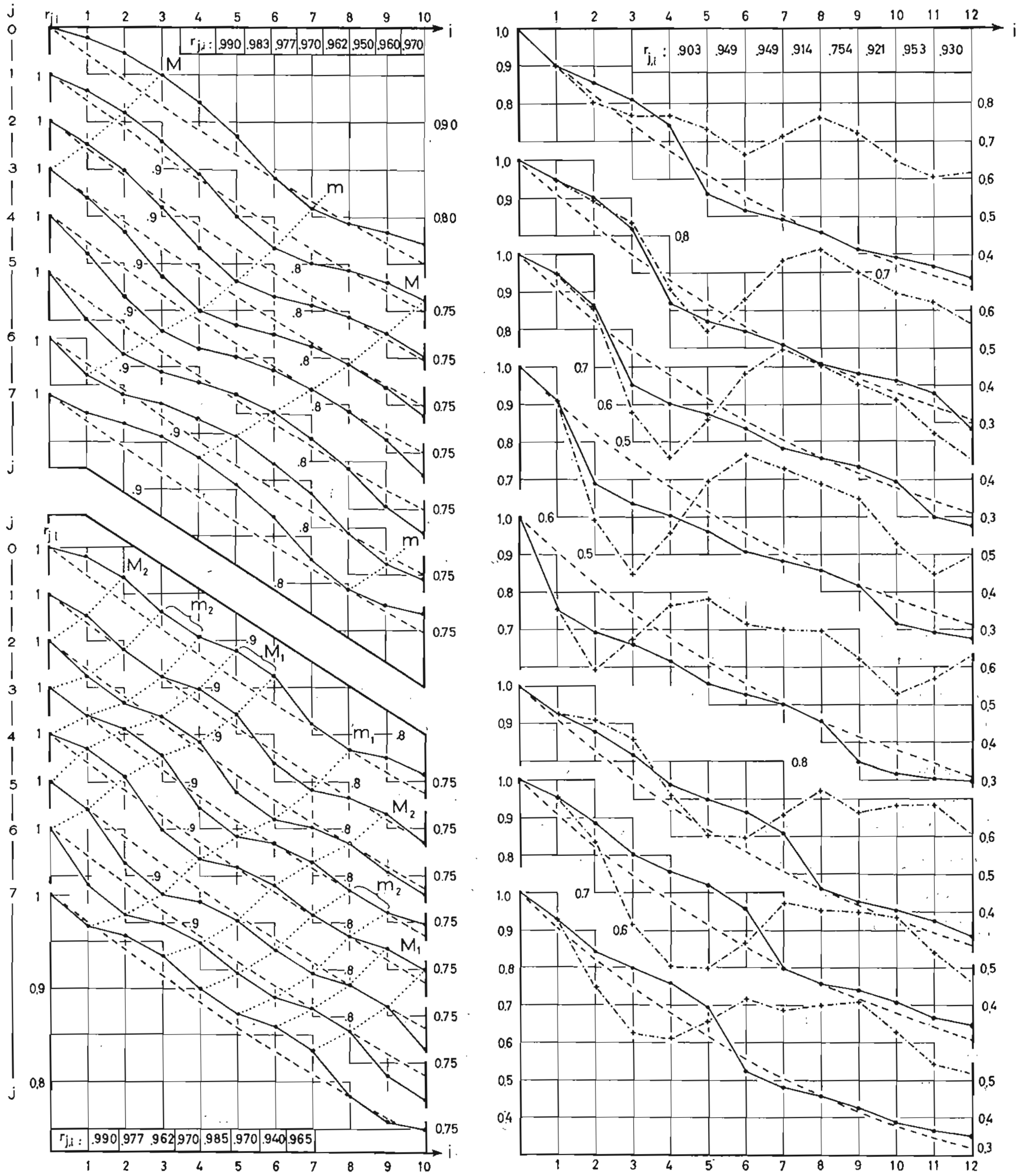

Fig. 16. Correlograms for Markov chains of series of order 8. The diagrams to the upper left show one relative maximum and one relative minimum within the period, whereas two pairs of relative extremes exist in the lower left diagram. In the diagram to the right a Markov chain (continuous curves) is fitted to the April temperature data for Oslo (dashed curves).

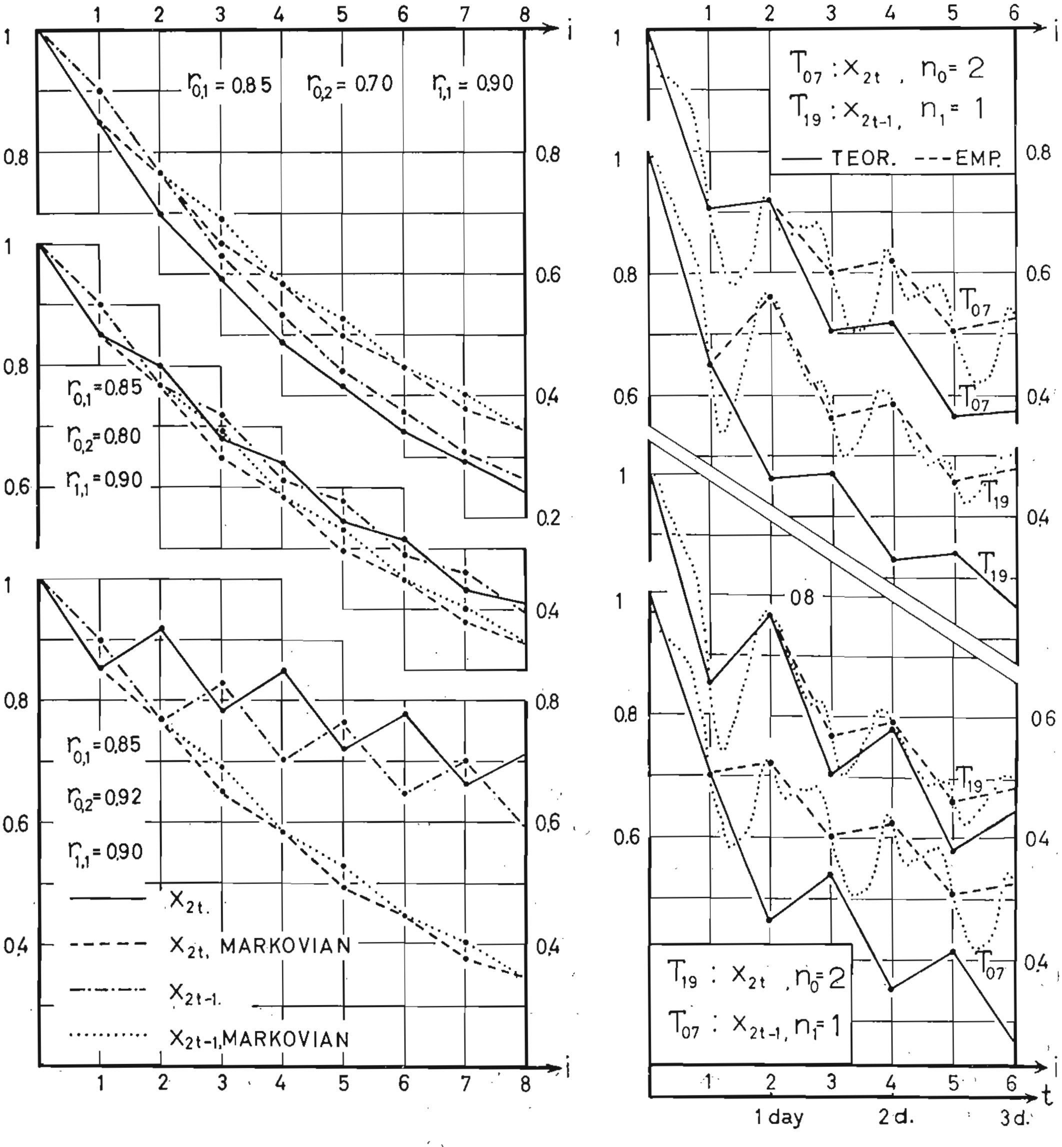

Fig. 17. Correlograms for chains of order 2, of the special type discussed in 2.4. To the left are presented one sub-Markovian chain and two super-Markovian. To the right chains of type 2.4 (continuous curves) have been fitted to Oslo April temperatures (dashed curves); in the upper diagram T_{07} has tradition length 2, and T_{19} tradition length 1, whereas the opposite is the case in the lower diagram.

Example 5. If $r_{0,2}$ in example 4 is changed to $0.80 > 0.85 \cdot 0.90$, we arrive at the super-Markovian chain whose correlogram is presented in fig. 17, left, middle. The curves lie systematically above the Markovian ones, and the wavy shape of the latter curves have been strengthened by the non-Markovian term. The "displacement" of the relative extremes is the same as in the Markovian case.

Example 6. Keeping the same value of $r_{0,1}$ and $r_{1,1}$ as in the preceding two examples, but changing $r_{0,2}$ to 0.92, so that $r_{0,2} > r_{0,1}$, we arrive at the correlograms shown in fig. 17, left, bottom. The curves are strikingly super-Markovian with absolute maxima and minima, displaced one time unit when we pass from x_{2t} to x_{2t-1}.

Example 7. Let us consider our computed correlograms for Oslo, April 07^h and 19^h; we identify x_{2t} with the normalized value of T_{07} and x_{2t-1} with normalized T_{19}. When applying the model type 9 we can find the "best model" in different ways, by estimating the correlation coefficients in the model by means of some of the empirical coefficients. The simplest procedure is to identify $r_{0,1}$, $r_{1,1}$, and $r_{0,2}$ with the observed values, i.e. by putting $r_{0,1} = 0.7070$, $r_{0,2} = 0.7212$, $r_{1,1} = 0.6553$. Using 2(39) we have computed the higher order correlation coefficients, presented by continuous curves in fig. 17, right, top. The corresponding empirical values are given by dashed lines. A comparison between "theoretical" and "empirical" curves shows that the theoretical curve for T_{07} (from which two empirical coefficients were used) is not too bad. The periodic maxima and minima are reproduced correctly in the model, but the empirical curve lies systematically above the theoretical one. The theoretical curve for T_{19} does not at all fit the empirical data, which do not show the time displacement of the extremes (the reason being that two maxima and two minima appear in the empirical curves, as indicated by the dotted lines, so that with 12 hours interval it is impossible to identify the true extremes of the correlogram).

Example 8. Let us now identify the normalized value of T_{19} with x_{t2}, the corresponding value of T_{07} with x_{2t-1}. From our data we find $r_{0,1} = 0.6553$, $r_{0,2} = 0.7634$, $r_{1,1} = 0.7070$. Computing the higher correlation coefficients, we arrive at the diagram fig. 17, right, bottom. The representation of T_{19} by our mathematical model must be characterized as very good.

Since in the model the even series (time step 24 hours) for x_{2t} is a Markov series, the good agreement between theory and experience with regard to the T_{19}-curve, seems to indicate that the *19^h-temperatures at consecutive days may to first order approximation be defined by a Markov series.* This provisional result will, however, need confirmation by study of months other than April and stations other than Oslo. We may also note that T_{07} in April, Oslo, is clearly super-Markovian; not all information needed for a prognosis 24 hours ahead is contained in T_{07} itself. One should therefore expect that the series for the average diurnal temperature also should be clearly super-Markovian.

The model for T_{07} is highly unsatisfactory; however, also the empirical curve having time unit 12 hours gives no good representation of the true conditions, as is shown by a comparison between dashed curve and dotted curve for T_{07}.

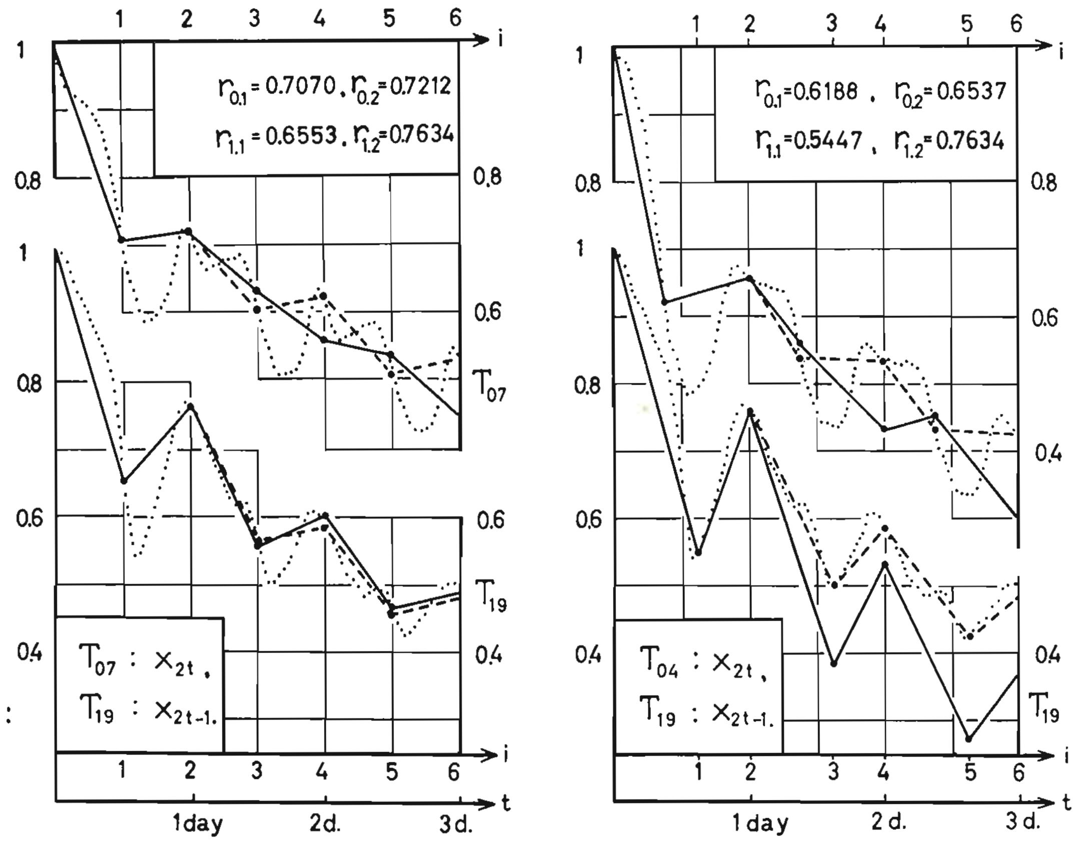

Fig. 18. General correlograms of order 2 and tradition length 2 (continuous curves), fitted to the Oslo April temperatures (dashed curves). In the diagram to the left the model has been applied to T_{07} and T_{19}, to the diagram to the right to T_{04} and T_{19}.

As a consequence of our discussion we may state that the empirical correlograms with time unit 12 hours cannot adequately be represented by the simple mathematical model defined by equations 2(25) and 2(26).

Example 9. Fig. 18, left, corresponds to a general chain of order 2 and tradition length 2, as defined by equations (41) and (42). We have identified x_{2t} with the normalized value of T_{07}, x_{2t-1} with the normalized value of T_{19} in the April temperature correlograms for Oslo; $r_{j,1}$ has been put equal to the empirical correlation coefficients (see values at the top of the diagram). The equations 2(1) corresponding to the non-normalized temperatures assume the form:

$$3(1)\quad \begin{cases} T_{07,0} = 0.330\,T_{19,-1} + 0.452\,T_{07,-2} - 0.72 + \varepsilon_{07}, \\ T_{19,-1} = 0.288\,T_{07,-1} + 0.600\,T_{19,-2} + 1.90 + \varepsilon_{19}, \end{cases}$$

where the second subscript refers to the day.

The variance reduction is 62 per cent for T_{07} and 61 per cent for T_{19}. We note that both series are highly super-Markovian, as could be expected since the Markovian chains of fig. 16, right, completely underestimate $r_{j,i}$ for higher values of i.

As shown by fig. 18 left, the model (1) gives a surprisingly good fit for the evening temperature (compare also fig. 17, right), but no good representation for the morning temperature.

Example 10. Fig. 18 right is computed for non-equidistant time steps (see p. 185); x_{2t} corresponds to the normalized value of T_{04} (near the minimum in temperature and temperature-information), whereas the normalized T_{19} has been maintained for x_{2t-1}. The model for T_{19} is not as good as in the diagram to the left; the reason being that for the not too good predictor T_{07} we have substituted an even worse predictor T_{04}, having a maximum of noise (see fig. 9).

13. Examples of chains of order 4. In the last 4 examples we have identified x_{4t}, x_{4t-1}, x_{4t-2}, x_{4t-3} with normalized values of $T_{04,0}$, $T_{22,-1}$, $T_{16,-1}$, $T_{10,-1}$ respectively for April, Oslo, the aim being to arrive at mathematical models for the empirical correlograms. Starting with the simple Markovian chain, we proceed to models of increasing complexity, making use of the formulae developed in 10. The empirical parameters used in the model are given by table 1.

Table 3.1. *Statistical parameters for the temperature correlograms for Oslo, April.*

Time	04^h	22^h	16^h	10^h
j:	0	1	2	3
$\bar{T}_j$:	1°.86	4°.18	8°.88	6°.21
s_{T_j}:	2°.92	3°.03	4°.45	3°.72
$r_{j,1}$:	0.8366	0.8046	0.8511	0.5901
$r_{j,2}$:	0.5031	0.7683	0.4579	0.7658

Example 11. The Markovian model, derived by using the first order correlation coefficients from table 1, is defined by the non-normalized equations:

$$3(2)\quad \begin{cases} T_{04,0} = 0.806\,T_{22,-1} - 1.51 + \varepsilon_{01} \\ T_{22,-1} = 0.548\,T_{16,-1} - 0.68 + \varepsilon_{22} \\ T_{16,-1} = 1.018\,T_{10,-1} - 2.56 + \varepsilon_{16} \\ T_{10,-1} = 0.752\,T_{07,-1} + 0.481 + \varepsilon_{10} \end{cases}$$

The variance reduction for T_{04}, T_{22}, T_{16}, and T_{10} respectively is 70, 65, 73, and 35 per cent. The reduction for T_{10} is especially low, the reason being that T_{04} is an extremely bad predictor, much influenced by noise. The high value of the regression coefficient in the third equation is partly due to the high correlation coefficient, but even more to the high standard deviation in the 16^h temperature.

The correlograms characteristic for model (2) are represented by continuous lines in fig. 19 left, the dashed lines giving the corresponding empirical correlograms, whereas the dotted lines represent empirical correlograms with time step one hour. The close correspondence between dashed and dotted lines indicates that a fairly good first approximation to the statistical time behaviour of the hourly air temperatures is obtained by using observations 6 hours apart, provided that the hours of observation correspond to the daily extremes in the noise level.

Practically no similarity can be found between the theoretical curves and the empirical, which are strongly super-Markovian for $i > 4$. For $i = 2$ the theoretical correlations are higher than the empirical for T_{04} and T_{16}, whose 12 hours predictability has been overestimated in the Markovian model.

Example 12. We improve our model by using a tradition length 2 (equal to 12 hours), applying the formulae 2(49), together with $r_{j,1}$ and $r_{j,2}$ from table 1. We then arrive at the equations:

$$3(3)\quad \begin{cases} T_{04,0} = 1.180\,T_{22,-1} - 0.316\,T_{16,-1} - 0.27 + \varepsilon_{04}, \\ T_{22,-1} = 0.372\,T_{16,-1} + 0.247\,T_{10,-1} - 0.66 + \varepsilon_{22}, \\ T_{16,-1} = 1.067\,T_{10,-1} - 0.104\,T_{04,-1} + 2.45 + \varepsilon_{16}, \\ T_{10,-1} = -0.215\,T_{04,-1} + 1.114\,T_{22,-2} + 1.94 + \varepsilon_{10}. \end{cases}$$

The first and third equations are sub-Markovian, since a high value of T_{16} contributes towards a low 04^h-temperature, T_{22} being kept constant, and a high value of T_{04} contributes towards a low T_{16}, when T_{10} is kept constant. Taking into consideration the influence of cloudiness, this behaviour of our model is not unreasonable from the meteorological point of view. Moreover we note that, as shown by the last equation, a low T_{04} contributes towards increasing T_{10} when T_{22} is constant — as can well be imagined by clear skies when T_{04} is low and the morning rise of temperature very rapid. The variance reduction in the above prediction equations is 77, 68, 72, and 60 per cent respectively. The prediction is thus most successful when the observations 6 hours earlier have a minimum of noise; the worst prediction is found for T_{10}, which is immediately preceded by the minimum temperature whose noise level is very great.

The correlograms for the model represented by (3) is shown by the continuous lines in fig. 19 right, the dashed lines characterizing the empirical correlograms. The general form of the correlograms for T_{10}, T_{16}, and T_{22} is reproduced by the model, which shows the correct position and displacements of the relative and absolute extremes.

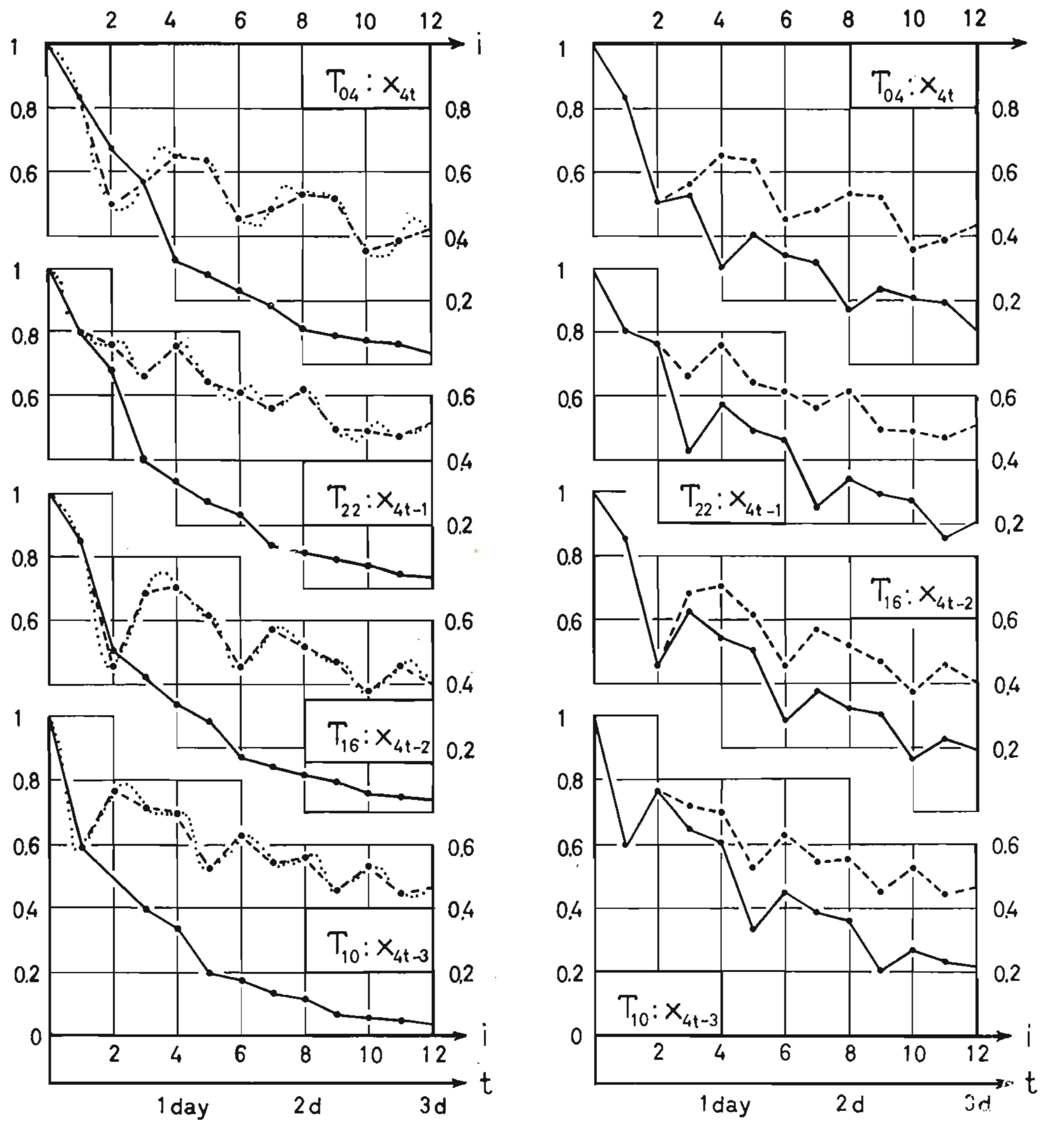

Fig. 19. Model correlograms of order 4 (continuous curves) applied to empirical correlograms (dashed curves) for the Oslo April temperatures T_{04}, T_{22}, T_{16}, T_{10}. To the left a Markov chain has been fitted to the data, to the right a chain 2.6a where all terms have a length of generation equal to 2 time units (12 hours).

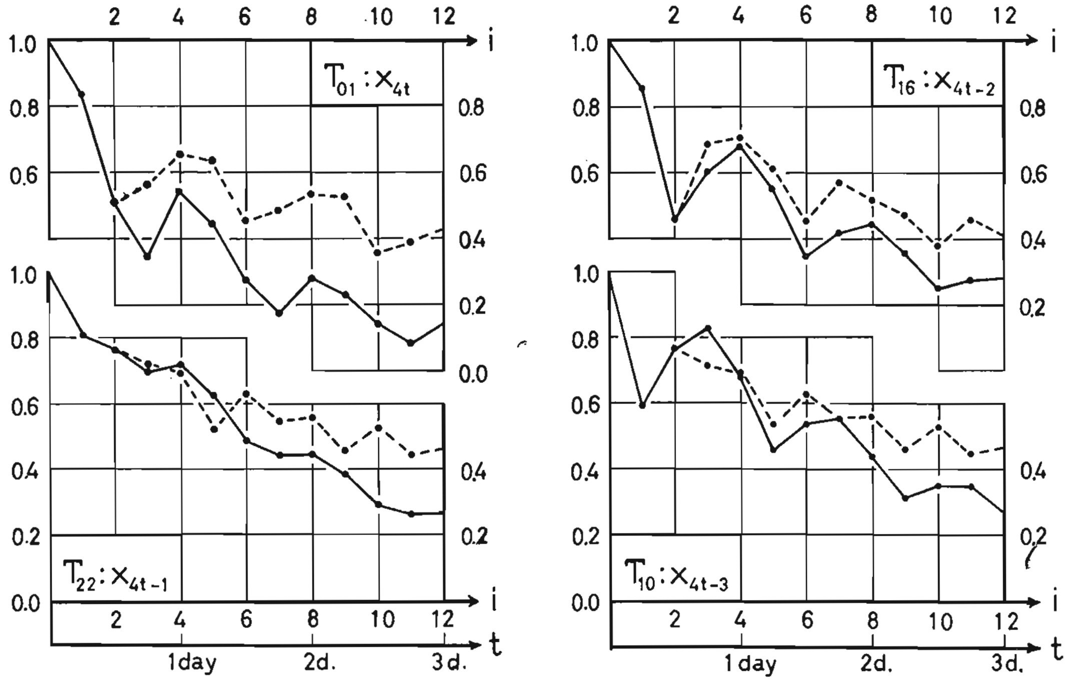

Fig. 20. The special correlogram 2.6b of order 4 (continuous curves) applied to the empirical correlograms (dashed curves) for the Oslo April temperature T_{04}, T_{22}, T_{16}, T_{10}.

However, the model gives systematically too low values of $r_{j,i}$ for $i > 2$, probably because some influences exist in Nature of time scale greater than 12 hours. For T_{04} the theoretical correlogram is highly unsatisfactory.

Example 13. The natural generalization of our model would consist in introducing additional terms of type $a_{j,3}$. We will, however, at present only apply the special chain 2(55). Its tradition length is 3, and it contains regression coefficients of order 1 and 3 only. Using 2(61) for the determination of these coefficients by means of $r_{j,1}$ and $r_{j,2}$, estimated from table 1, we find, for the non-normalized equations corresponding to 2(55):

$$3(4)\quad \begin{cases} T_{04,0} = 1.347\,T_{22,-1} - 0.573\,T_{10,-1} - 0.211 + \varepsilon_{04}, \\ T_{22,-1} = 0.418\,T_{16,-1} + 0.432\,T_{04,-1} - 0.335 + \varepsilon_{22}, \\ T_{16,-1} = 1.124\,T_{10,-1} - 0.169\,T_{22,-2} + 2.610 + \varepsilon_{16}, \\ T_{10,-1} = 0.297\,T_{04,-1} + 0.593\,T_{16,-2} + 0.394 + \varepsilon_{10}. \end{cases}$$

Two of the equations can be characterized as super-, the other two as sub-Markovian. The variance reductions are 92, 78, 76, and 72 per cent for T_{04}, T_{22}, T_{16}, and T_{10} respectively, thus greater than in the model (3). If model (4) gives a better fit than (3) to the empirical data, it consequently is preferable, containing a higher degree of predictability.

The correlograms corresponding to (4) are represented in fig. 20 (x_{4t} is to be identified with T_{04}, not with T_{01} as is erroneously stated in the diagram). A comparison with fig. 19 right, representing model (3) of tradition length 2 shows: The general model with tradition length 2 gives for T_{10}, T_{16}, and T_{22} a better picture of the *form* of the empirical correlograms, but model (4) — a somewhat artificial chain of tradition length 3 — corresponds to correlograms which *on the average* give a better representation of the empirical data. None of the models gives a good representation of T_{04}, but the special chain (4) with tradition length 3 is decidedly the better. An application of a general chain of tradition length 3, with coefficients determined from the empirical correlation coefficients $r_{j,1}$, $r_{j,2}$, and $r_{j,3}$ thus seems promising, and will be attempted later.

Example 14. An application of the special model of tradition length 4 given by 2(62) to the data of table 1 is more complicated since the determination of $a_{j,1}$, $a_{j,4}$, and $r_{j,3}$ leads to quadratic equations. Moreover, the condition 2(11) is not fulfilled by the computed values. Consequently, it is not possible exactly to represent the 2 first terms of the empirical correlograms by means of model 2(62). It may be possible to arrive at a model of this type by changing somewhat the values of table 1; some attempts made have not been successful, and the detailed discussion of the problem will be taken up later, when the computations have been programmed for the digital computer IBM 650 for different types of mathematical models.

14. General considerations about the time flow of information. A time series connecting x_t, x_{t-1}, . . . x_{t-n} describes a *flow of information* (measured by correlation coefficients) from the past to the present. n gives the number of previous "hours of observation" which transmit information directly to the time t. The hours of observations prior to $t - n$ of course also contain information of interest to the time t, but this information will always "pass through" the times $t - 1$, $t - 2$, . . . $t - n$. These informators consequently contain no "additional information" to that already obtained x_t. If $n = 1$ (Markov case), all past information passes through *the immediate past time* $t - 1$.

Let us first consider stationary series with $n = 1$; the correlation coefficients $r(x_t,x_{t-1})$ and $r(x_{t-1},x_{t-2})$ by definition are then equal. A certain amount of information passes from time $t - 2$ to time $t - 1$. Since part of this information is "destroyed" by the noise at the time $t - 1$, only a certain fraction of it is transmitted to the time t. But at the same time new information is "created" at $t - 1$, and part of it is also transferred to the future time t. Under stationary conditions the information which x_{t-1} gives about x_t is the same as the information which x_{t-2} gives about x_{t-1}. Consequently: in the time flow of information there are neither "sources" nor "sinks"; one could denote this flow as *non-divergent*, using the terminology of vector analysis. If $n > 1$, similar conditions obtain: the total information given about x_{t-1} by x_{t-2}, x_{t-3}, . . ., x_{t-1-n}

is equal to the information given by $x_{t-1}, x_{t-2}, \ldots, x_{t-n}$ about x_t. In general *a nondivergent time flow of information is characteristic of the stationary time series.*

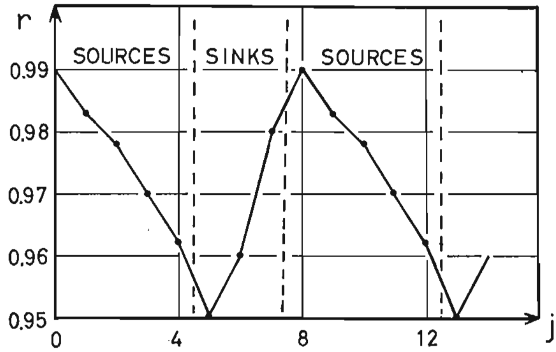

Fig. 21. Diagram illustrating sinks and sources of information.

Let us now consider, as the simplest example of a non-stationary chain of series, the Markov type 2(15). In this case the flow of information from x_{mt-j-1} to x_{mt-j} varies with j with a period equal to m. We then may encounter "relative sinks of information", in which more information is received from the immediate past than is transmitted to the immediate future, and analogous "relative sources of information". Fig. 21 gives an illustration corresponding to example 1, and to fig. 16, upper left. The times 05^h—07^h, 13^h—15^h, where r_j increases with j evidently represent relative sinks, the times 0—04^h, 08—12^h, ... where r_j decreases with j relative sources of information; the information is measured by the correlation coefficient r_j. Since, in the diagram fig. 21 the strength of sources and sinks shows a regular variation with time, their added contribution to the higher correlation coefficients is responsible for the relative maxima and minima in the corresponding correlograms of fig. 16. Owing to the periodic character of the sources and sinks and their concentration on certain time values, the information about x_{mt-j} given by x_{mt-j-m} is independent of j in the Markov chain of series.

Let us next proceed to a non-Markovian chain of series ($n > 1$); we consider n successive times of observations as an entity, their accumulated influence being measured by the variance reduction. If x_{t-1} receives more (less) information from $x_{t-2}, \ldots x_{t-n-1}$ than x_t receives from $x_{t-1}, \ldots x_{t-n}$, the latter time interval of length n can be characterized as a relative sink (source) of information. An illustration is shown in example 12 (fig. 19 right); the time interval containing $T_{10,0}$ and $T_{04,0}$ appears as a relative source, contributing a greater amount of information about $T_{16,0}$ than the information about $T_{10,0}$ which is given in the time interval containing $T_{04,0}$ and $T_{22,-1}$.

In the non-Markovian case we further meet with correlograms in which, for some values of $n > 1$, $r(x_t, x_{t-n-1})$ is greater than $r(x_t, x_{t-n})$ as shown in some of our models and empirical diagrams (see for instance fig. 6). Then for certain values of j the direct information, given by x_{t-j} about the value x_t, decreases as the time distance $t - j$ decreases. In Meteorology we have certain time intervals during which *predictability decreases with decreasing prognostic interval.* As an example we may mention (see fig. 6) that a statistic prognosis of the April temperature T_{07} in Oslo by means of *one* earlier temperature is much better when we use a time interval of 22 hours than if the time interval is only 16 hours. During the hours from 09^h to 15^h a great amount of "daytime" noise occurs which decreases the information given by the actual temperature about the future morning temperature T_{07}.

The preceeding fairly loose considerations stress the importance of the concept of the time flow of information in Meteorology. In more general problems, to be attacked later, we must also take into consideration the cross-information, which one meteorological variable gives about an other, referred to the same or to a later time, say the cross-information which is encountered when the time variations of temperature and cloudiness are discussed simultaneously (with time units 1 day or 1 hour).

REFERENCES

KENDALL, M. G.: The Advanced Theory of Statistics, Vol. I, London 1945.

NORDØ, J.: Expected Skill of long-range Forecasts when derived from daily Forecasts and past Weather Data, Oslo 1959.

WOLD, H.:A Study in the Analysis of Stationary Time Series, Uppsala 1938.

(Manuscript received November 15, 1961)