LONDON, METEOROLOGICAL OFFICE.

Met.O.3 Technical Note No.4. Monthly mean minimum temperatures 1959-77: coefficients of a set of factors and station residuals. By DONE, A.L.

London, Met.Off., Met.O.3 Tech.Note No.4, 1980, 31cm. Pp.8+[44]. 3 Refs. Abs.p.2.

An unofficial document - restriction on first page to be observed.

Met O 3 Technical Note No 4

MONTHLY MEAN MINIMUM TEMPERATURES 1959-77:
COEFFICIENTS OF A SET OF FACTORS AND STATION RESIDUALS

by

A L DONE

Climatological Services Branch (Met O 3)
Meteorological Office
London Road
Bracknell
Berkshire
RG12 2SZ

October 1980

CONTENTS

1. Summary

This paper shows how orthogonal factors, identified in one extract of data, may be used to study data of a different type, and from a different period. A multiple linear regression has been performed, using monthly minimum temperatures for 1959 to 1977, on factor scores from a principal components analysis of daily minimum temperatures for 1973 to 1977. The resulting time series of coefficients were analysed and the residuals studied with a view to their reflecting inhomogeneity of data at particular stations.

This note should be read in conjunction with Met O 3 Technical Note No 3 by Spackman (1979) which describes the method by which the factors were obtained and also presents the geographical distribution of the factors together with a short description of them.

2. Techniques

Factor analysis is a technique by which a large number of variables are summarized by a few factors; handling of the data is thus made a much simpler task. Using a particular type of factor analysis, called Principal Components Analysis with Varimax rotation, Spackman (1979) reduced a multitude of daily minimum temperatures, from 469 stations for 1973-77, to a set of fifteen factors. For all these stations there exist fifteen scores, one associated with each of these factors. These scores express the degree to which each station possesses the property that particular factor describes. The 'factor score' maps for all fifteen factors may be found in Spackman (1979). These station factor scores, together with monthly minimum temperatures, for a selected network of United Kingdom climatological stations, over the period 1959 to 1977, have been used in the following analysis.

The BMDP program P6R (Biomedical Computer Programs, Dixon and Brown (1979)), was used to regress the monthly minimum temperatures (dependent variables) on the set of 15 station factor scores (independent variables).

$$t_{ij} = c_{io} + c_{i1}f_{1j} + c_{i2}f_{2j} + \ldots\ldots + c_{i15}f_{15j} + r_{ij}$$

t_{ij} - mean monthly minimum temperature for month i, station j.
c_{io} - constant term (intercept).
$c_{i1} \ldots c_{i15}$ - set of 15 factor coefficients for month i.
$f_{1j} \ldots f_{15j}$ - set of 15 factor scores for station j.
r_{ij} - residual for month i, station j.

The initial aim of the work was to study the time series of coefficients c_i for each factor. However as the analysis continued, interest also developed in the residuals. More time was devoted to their study, with a view to discerning how they might reflect, and so help find, discontinuities in the data at individual stations.

3. Data

3.1 Network

To give adequate representation of the whole of the United Kingdom it is desirable to use a fairly dense station network. However strict limitations were set on the choice of stations by the length of data required, that is, a continuous record of minimum temperatures from 1959 to 1977.

From those stations available, some objective method of choice for the analysis was essential. For this purpose the climatological regions (produced by Spackman (private communication) from a similar factor analysis using the data of 10 climatic elements), shown in Figure 1, were used. Wherever possible, two stations from each region were included in the network. Unfortunately this was not always possible and some regions, notably in Scotland and South West England, have only one, or even no representatives. A total of 125 stations were used in the final regression (see Fig 1) and a further 36 stations included for extraction of the station residuals only.

The factor scores at each station are empirically orthogonal over the 469 stations from which they were determined. Thus the squared multiple correlation (SMC) of each factor, with all other factors, over all stations used in the analysis, should be nearly zero. The SMCs for various station networks, used during development of the technique, are shown in Fig 2. The SMCs from the 125 station network used in this analysis are all small, implying all fifteen factors may be adequately represented by this choice of stations. This is not so for the networks of 19 and 50 stations.

Although the SMCs changed dramatically as more stations were added to the network, there was little effect on the factor coefficients produced by the multiple linear regression, once a certain threshold number had been reached. Coefficients from a 19-station analysis were wildly differing, but those from a 50-station and the final 125-station analyses had few dissimilarities. These coefficients, for a selection of factors for 1959 and 1960, are shown in Fig 3(i)-3(iii). Also displayed here are coefficients produced by a regression performed using only the first five factors for 50 stations; again the coefficients show little change.

3.2 Missing values and errors

The dataset, compiled of station factor scores and the minimum temperatures, contained several missing values of temperature. The BMDP program PAM was used to estimate these by the regression method, using a ridge parameter of 0.2.

In addition to the missing values was the problem of errors in the minimum temperatures. The best method of discovering these errors proved to be to use the station residual series which will be presented in Section 5.2. All extreme residuals more than about 1.5 deg C from the mean residual for the station and out of character with the others at the station were examined. In many cases the data were found to be in error and could be replaced with the "correct" value from the Monthly Weather Report (MWR) or other sources. When a better replacement value could not be found and the residual was substantially greater than 1.5 deg C the data were set to the "missing value" code and estimated by the BMDP program PAM. For example, the residuals at Blyth Bridge (DCNN 1836) for July and August 1971 (see Fig 8(vi)) were investigated but no evidence was found to justify changing the data.

4. Computation

It was inconvenient to analyse all the data at one time with BMDP programs so the data were separated into four periods: 1959-63, 1964-68, 1969-73 and 1974-77. All estimation of data and subsequent calculations were performed within these periods using exactly the same station network (see Fig 1) each time.

5. Results

Although the factors were calculated from minimum temperatures for the years 1973 to 1977 inclusive, it was hoped that they would equally well represent the underlying spatial variation of minimum temperatures for any recent period. It was found that this is not quite the case. The squared multiple correlations shown in Fig 4 are tests of the significance of the regression for each month, on all the factors. Throughout the whole period 1959 to 1977 the SMCs are always very high, having a mimimum of 0.88 in December 1971, but for the period January 1973 to December 1977 the correlations average a slightly higher value than for the remaining months. This is explained by the fact that the factor scores were determined from data for 1973-77.

5.1 Factor coefficients

Associated with all fifteen factors used in the analysis, a coefficient was calculated by linear regression for each month of the 19 year period. These are shown in Fig 5(i) to 5(iv).

The coefficients are displayed for each factor, with the equivalent 30 year (1941-70) mean for each month superimposed (⊖-⊖-⊖). The standard error of each coefficient is shown by a cross (X), above and below the coefficient value. All the error intervals are small, even where the coefficient values have a large departure from the mean, which implies these extreme are true and not caused by large errors in the data.

The intercept offers little more information than many other analyses have done in the past. It represents the mean minimum temperature of all the stations used in the analysis and shows how the average minimum temperature varies from month to month. The other coefficients however create a new method of describing the minimum temperatures of any one month. The coefficient for factor one, for example, gives an indication of how strong the North/South variation is for any particular month, for factor two, how pronounced the coastal/inland contrast is (Spackman (1979)). January 1963 has a relatively small coefficient for factor one and large for factor two, implying an unusually small gradient of minimum temperatures North to South but a large contrast between the coastal and inland regions of England. Individuals may interpret the factors in different ways but for each month the coefficients indicate to what degree the overall pattern of minimum temperatures can be accounted for by each factor.

Annual cycles can be seen in many of the series of factor coefficients of Fig 5. Each series of coefficients was analysed using a spectral analysis package developed by Ross (1978). As anticipated, highly significant annual cycles were found in all but four (factors 3, 8, 10 and 14) of the factor coefficient series. The Bartlett Whiteness Test was used to test for significance. An example of the spectra is shown in Fig 7 for the intercept (or constant) term where the logarithmic density is plotted against frequency. The peak at about 0.08 corresponds to a 12 month cycle and is significant to at least the 1% level. Other peaks (eg. at .16 (6 months) and .25 (4 months)) are not significant at the 5% level. It was found that the annual cycle was significant at 5% for factor 14, only noticeable for factors 8 and 10 and not evident at all in factor 3.

Peaks were evident for some factors at 6, 4 and 3 months but none are even nearly significant at the 5% level. They are harmonics of the annual cycle and probably indicate merely that the 'annual cycle' is not sinusoidal.

The spectral analysis diagrams give almost constant density for all frequencies probably because the overall month to month persistence of the coefficients is small both for the intercept (constant) and for the factors.

Study of the coefficients halted with the knowledge that annual cycles only are significant in the coefficient series for factors 1, 2, 4, 5, 6, 7, 9, 11, 12, 13, 15 and the intercept (constant).

The factor coefficient series shown in Fig 5 with their associated equivalent climatological mean for 1941-70 make it difficult to appreciate whether there are other features of interest. Fig 6 shows the coefficient time series with the monthly means for 1941-70 removed (ie. as a departure from the normal for 1941-70). As in Fig 5, 'x' represents one standard error. There are many features of interest (note the persistent coolness shown by the intercept for late 1962 and early 1963, the anomalous values in many factors for early 1963, persistent positive anomalies of factor 11 during 1975 following small negative anomalies in 1974 etc.), but no clear indications of climatic trend etc., except perhaps in factor 12. If a decrease of the factor 12 coefficient of 0.1 were significant this would for instance represent a change in temperature contrast between East and West Scotland of up to 0.5 deg C with East Scotland warming with respect to West. Clearly further detailed statistical analysis of the coefficients is required.

5.2 Station residuals

The residual, r_{ij}, is defined by the equation in Section 2. It represents the difference between the "observed" value and an "estimated" value determined from the factor coefficients for the month and by the factor scores for the station.

The 19 years of residuals for each of the 125 stations used in the analysis and also for a further 36 stations (marked with an asterisk) are shown in Fig 8(i) to 8(xxvii). Many of the residual series have striking discontinuities and other interesting features which demand further investigation. Some stations have sudden changes in character of their residuals, for example Jersey (DCNN 9970) at the end of 1969, (records reveal change of site) and other residuals have what seem to be marked trends, see Exmouth (DCNN 8881).

Unfortunately over the years, details of instrument replacement and even site changes have not been recorded quite as might be wished. Records that do exist are far from perfect and are littered with omissions and errors. Study of the residual series was inevitably hampered by the quality of these station records, but nevertheless it was possible to associate certain discontinuities with recorded changes at stations.

Table 1 gives a list of some stations whose residuals show more distinct discontinuities and the approximate dates these occur. Records were not readily available for most Scottish or any Irish stations but any station change it has been possible to trace which may have been responsible for the altered minimum temperatures is noted. Other stations exhibit short period anomalies. Table 2 gives a list of some of the more notable.

It is hoped these residuals will serve as useful indications when the homogeneity of minimum temperature at any of the stations available in this analysis is in question. It should be relatively easy to repeat the above analysis for maximum temperature and so help determine whether effects are likely to be caused by instrument or site.

Acknowledgements

Thanks are given to E Spackman, whose work these investigations developed from, and for the inevitable assistance he was able to give me throughout the study. Also to G Ross for his help in applying and understanding many of the statistical techniques used.

References

DIXON, W J and BROWN, M B (editors) (1979).
University of California Press. Biomedical Computer Programs.

ROSS, G H (1978).
A Spectral Analysis Package.
(Typed manuscript available in Meteorological Office (Met O 3)).

SPACKMAN, E A (1979).
Met O 3 Technical Note No 3. A Multivariate Analysis of Temperature within the UK Climatological Network.

Figures

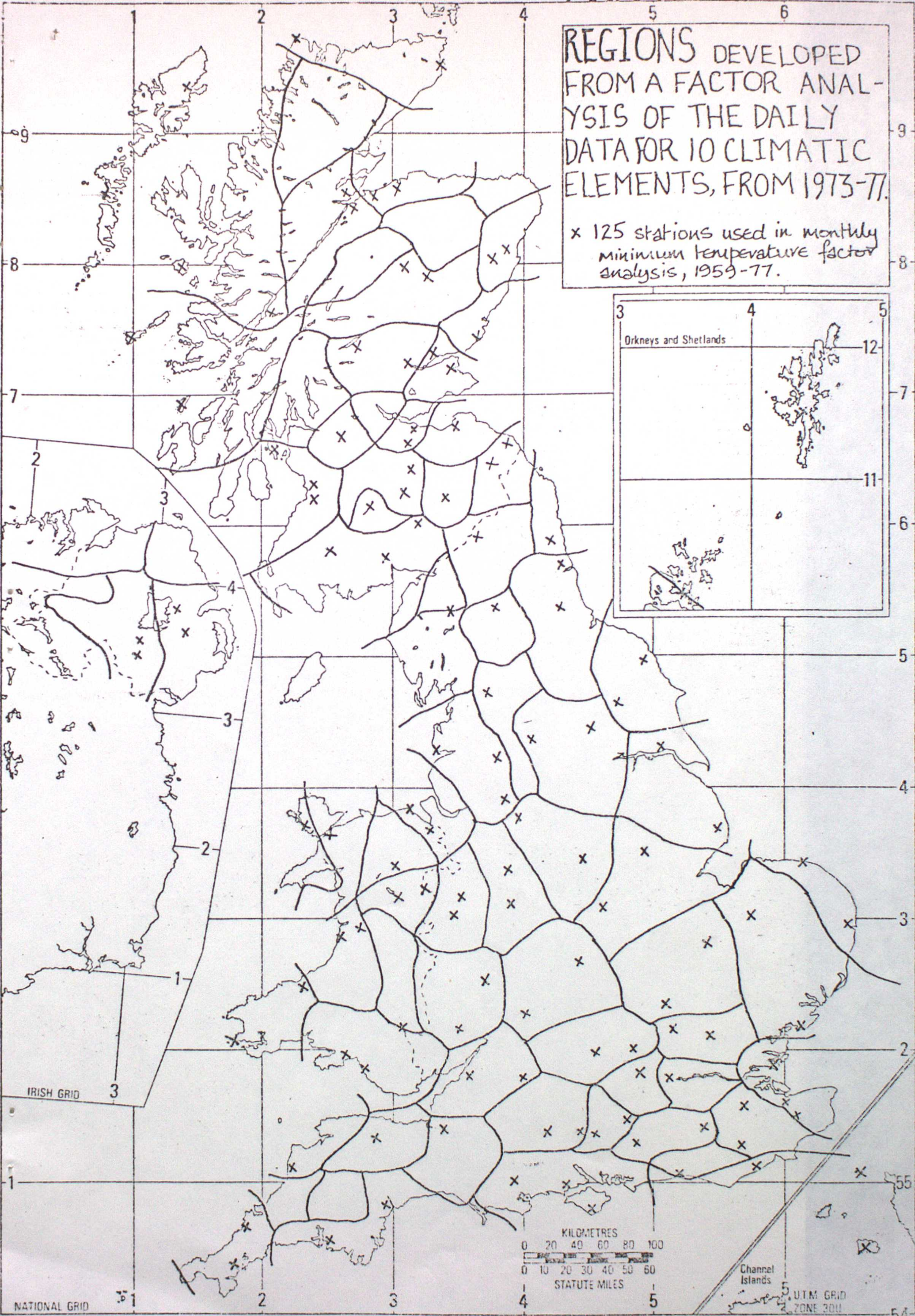
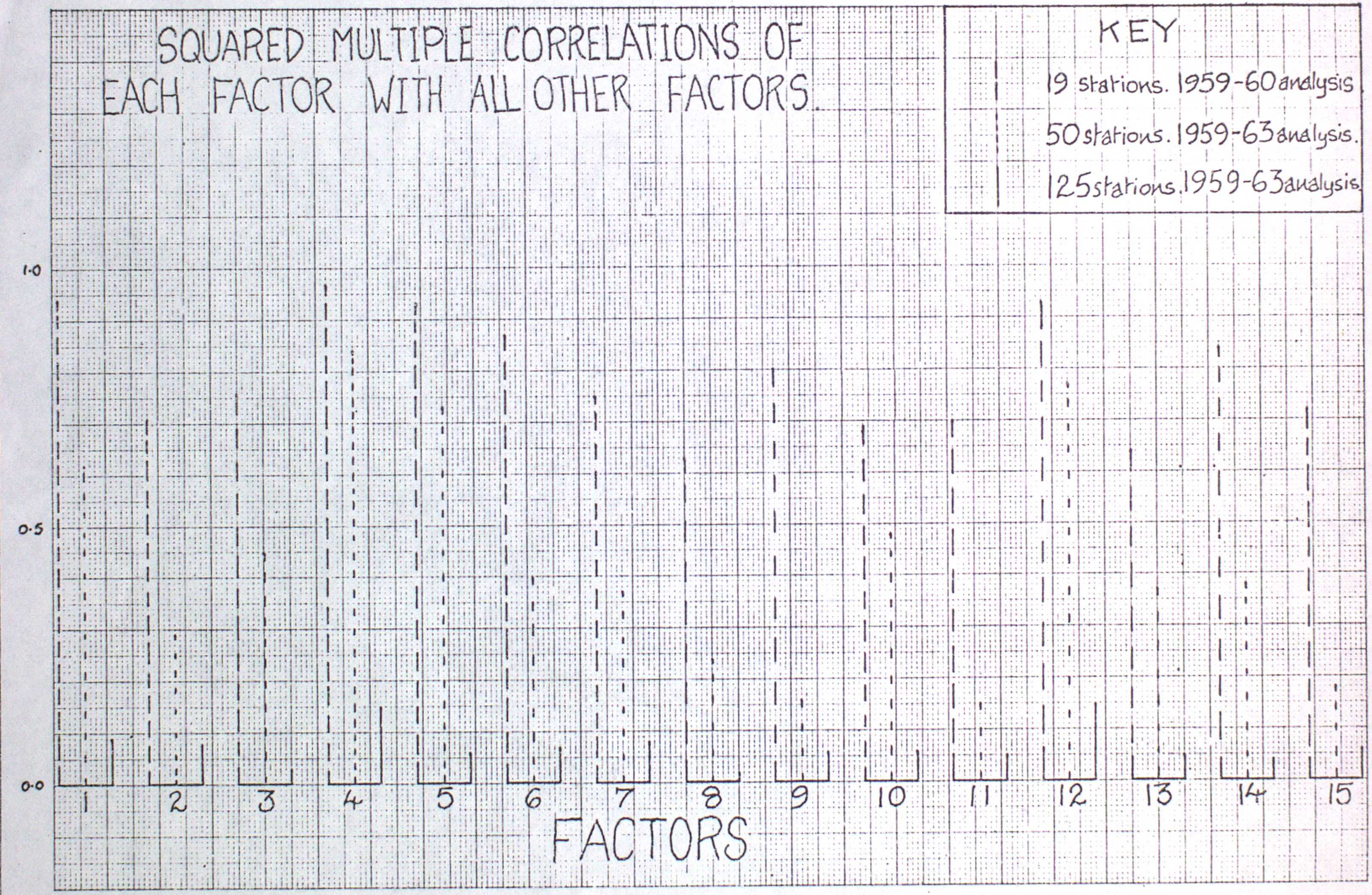

REGIONS DEVELOPED FROM A FACTOR ANALYSIS OF THE DAILY DATA FOR 10 CLIMATIC ELEMENTS, FROM 1973-77.
× 125 stations used in monthly minimum temperature factor analysis, 1959-77.
Orkneys and Shetlands
IRISH GRID
KILOMETRES
0 20 40 60 80 100
0 10 20 30 40 50 60
STATUTE MILES
Channel Islands
U.T.M. GRID ZONE 30U
NATIONAL GRID

FIGURE 2


SQUARED MULTIPLE CORRELATIONS OF EACH FACTOR WITH ALL OTHER FACTORS
KEY
19 stations. 1959-60 analysis
50 stations. 1959-63 analysis.
125 stations. 1959-63 analysis
1.0
0.5
0.0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
FACTORS

FACTOR COEFFICIENTS

FIGURE 3(i)

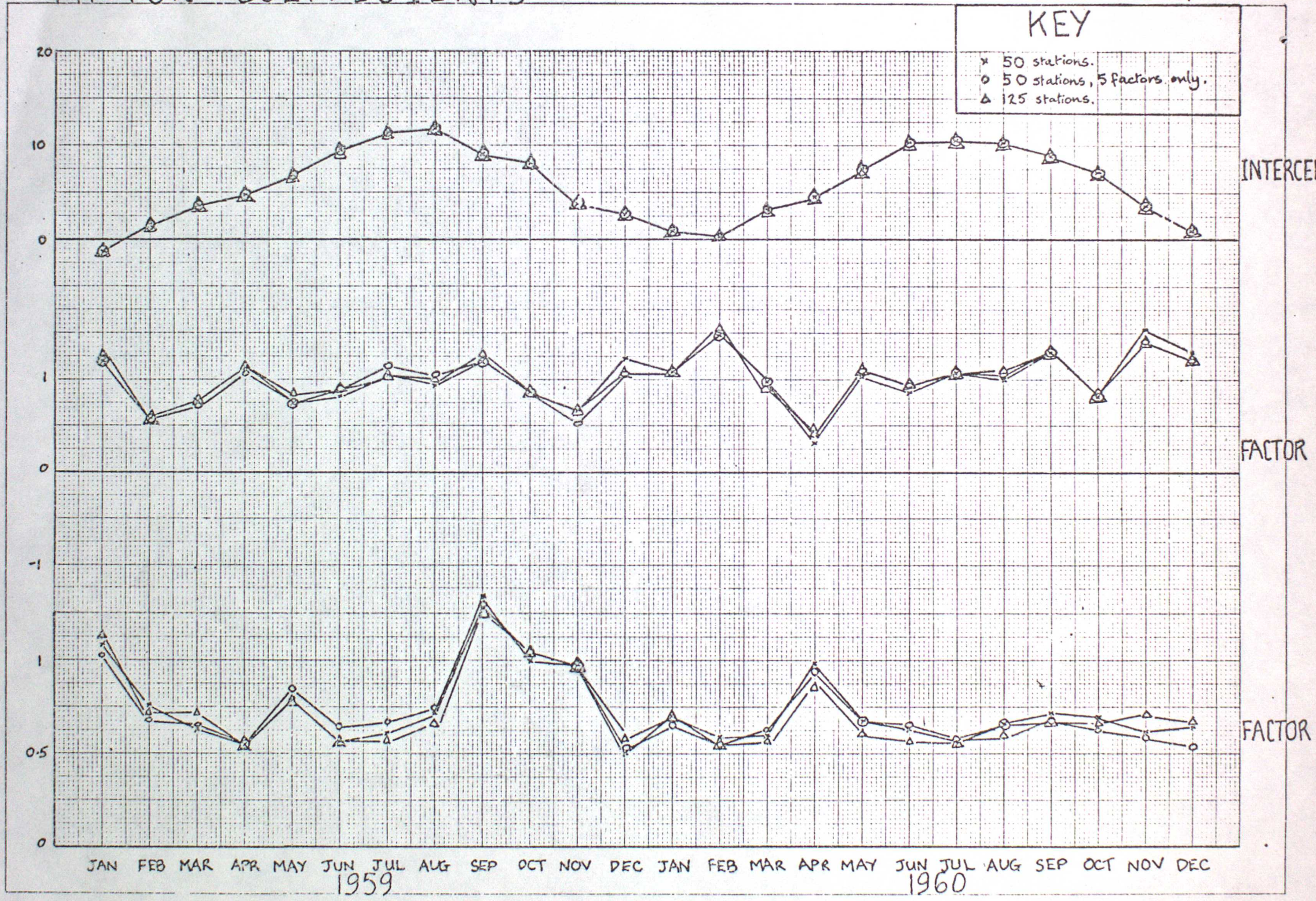
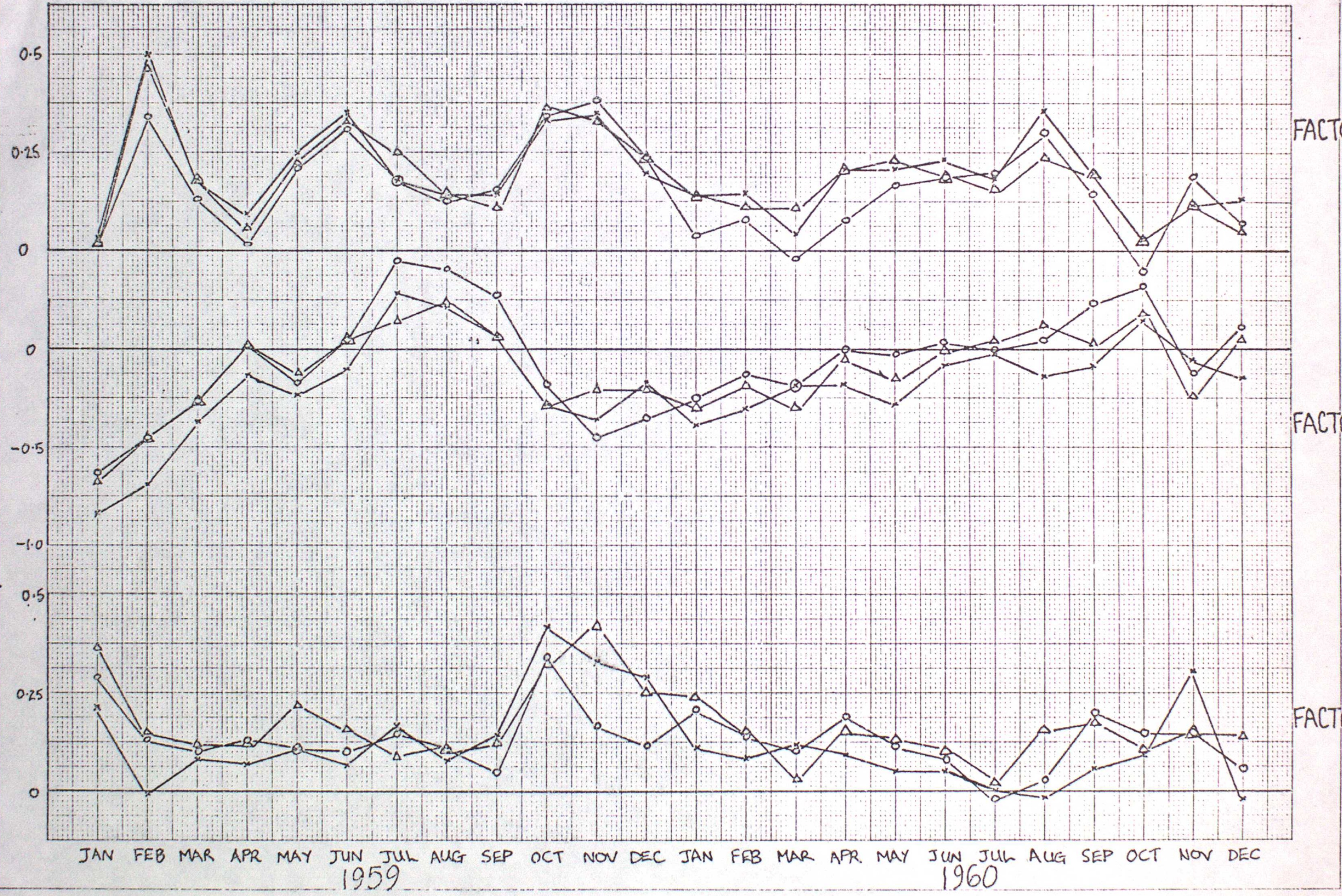
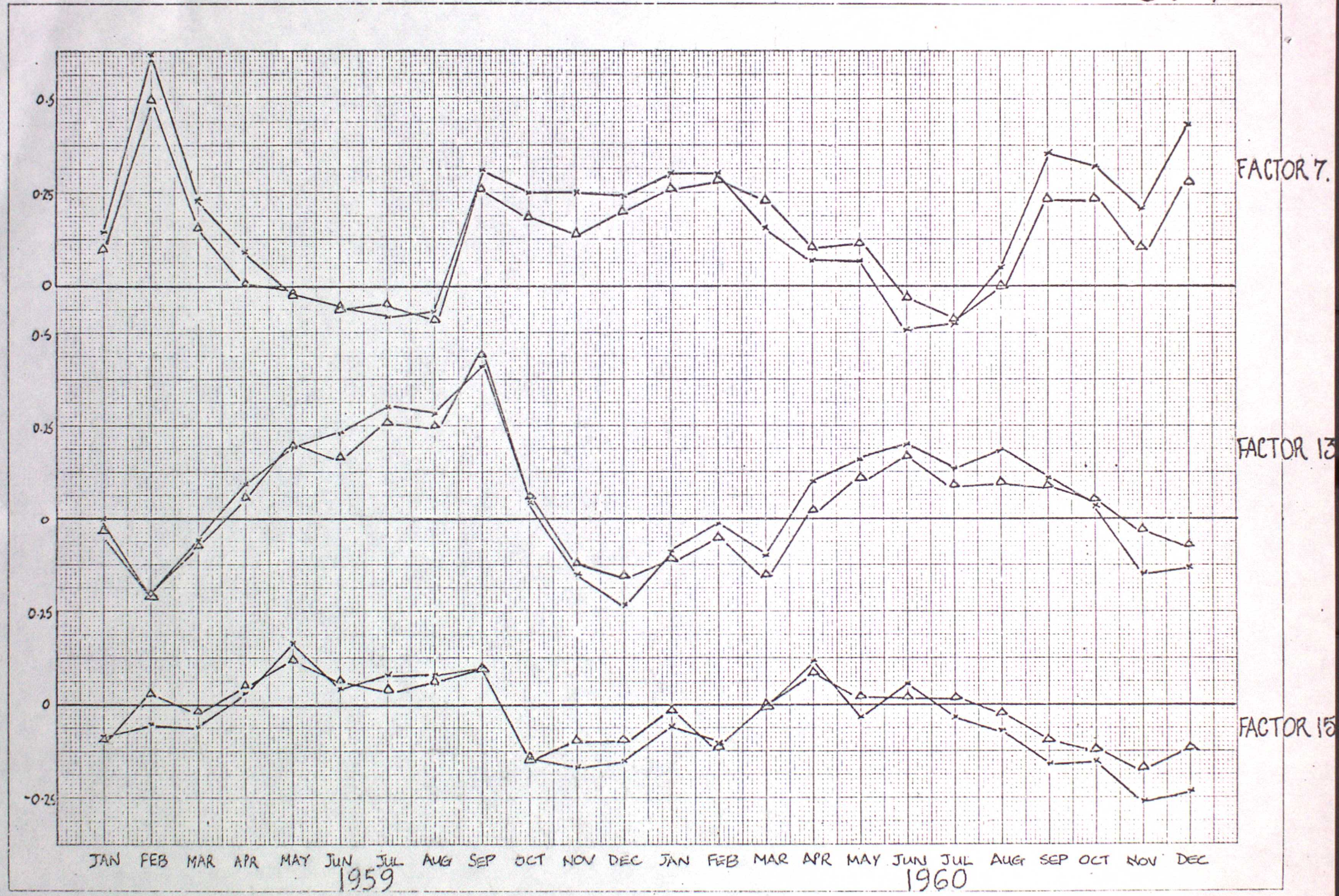


FIGURE 3(ii)



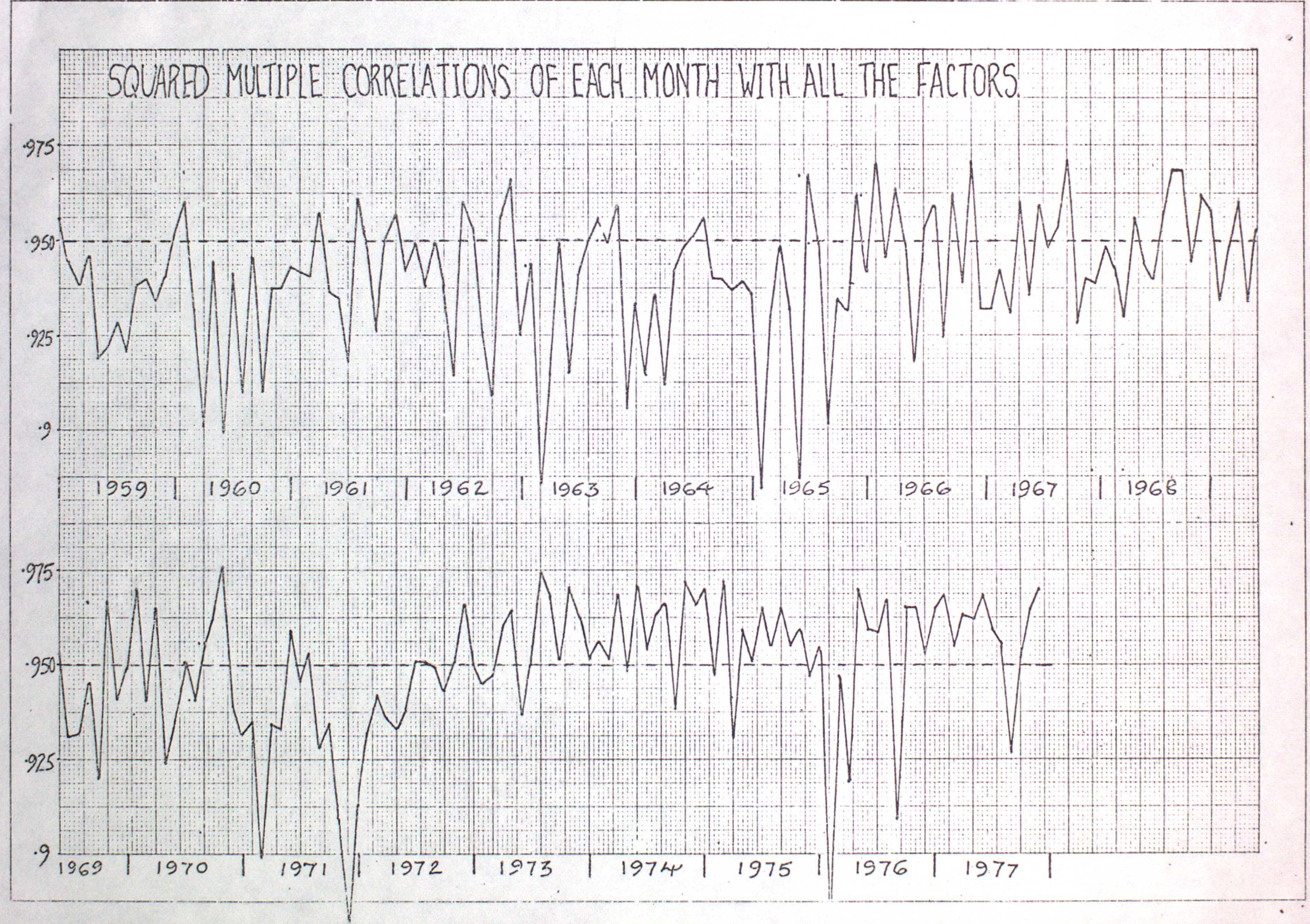
0.5
0.25
0
0
-0.5
-1.0
0.5
0.25
0
FACT
FACT
FACT
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
1959
1960

FIGURE ·3(iii)



0.5
0.25
0
0.5
0.25
0
0.25
0
-0.25
FACTOR 7.
FACTOR 13
FACTOR 15
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
1959
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
1960

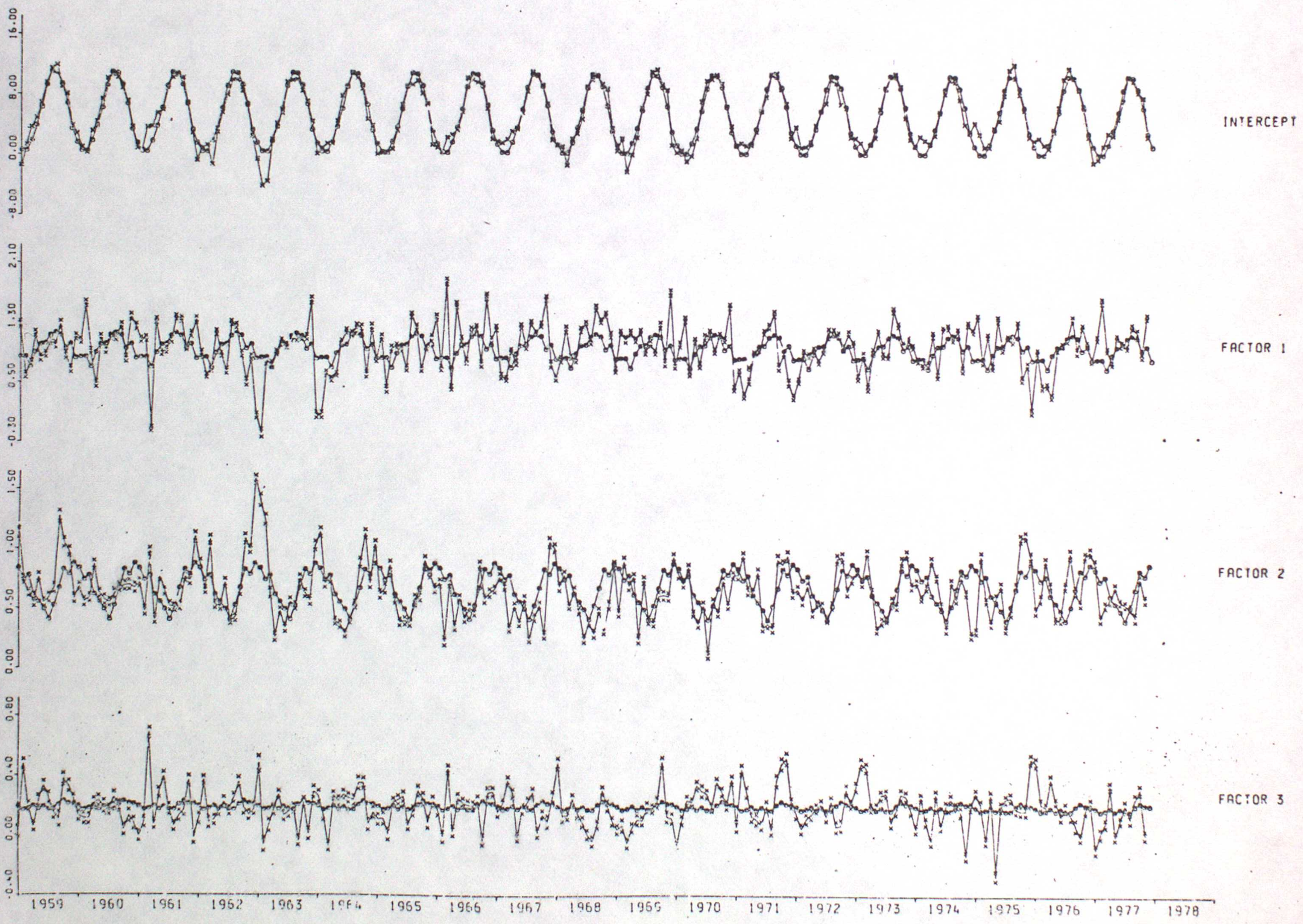
FIGURE · 4.


SQUARED MULTIPLE CORRELATIONS OF EACH MONTH WITH ALL THE FACTORS
·975
·950
·925
·9
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
·975
·950
·925
·9
1969
1970
1971
1972
1973
1974
1975
1976
1977

FACTOR COEFFICIENTS.

PRODUCED USING BMDP PROGRAM 6R.

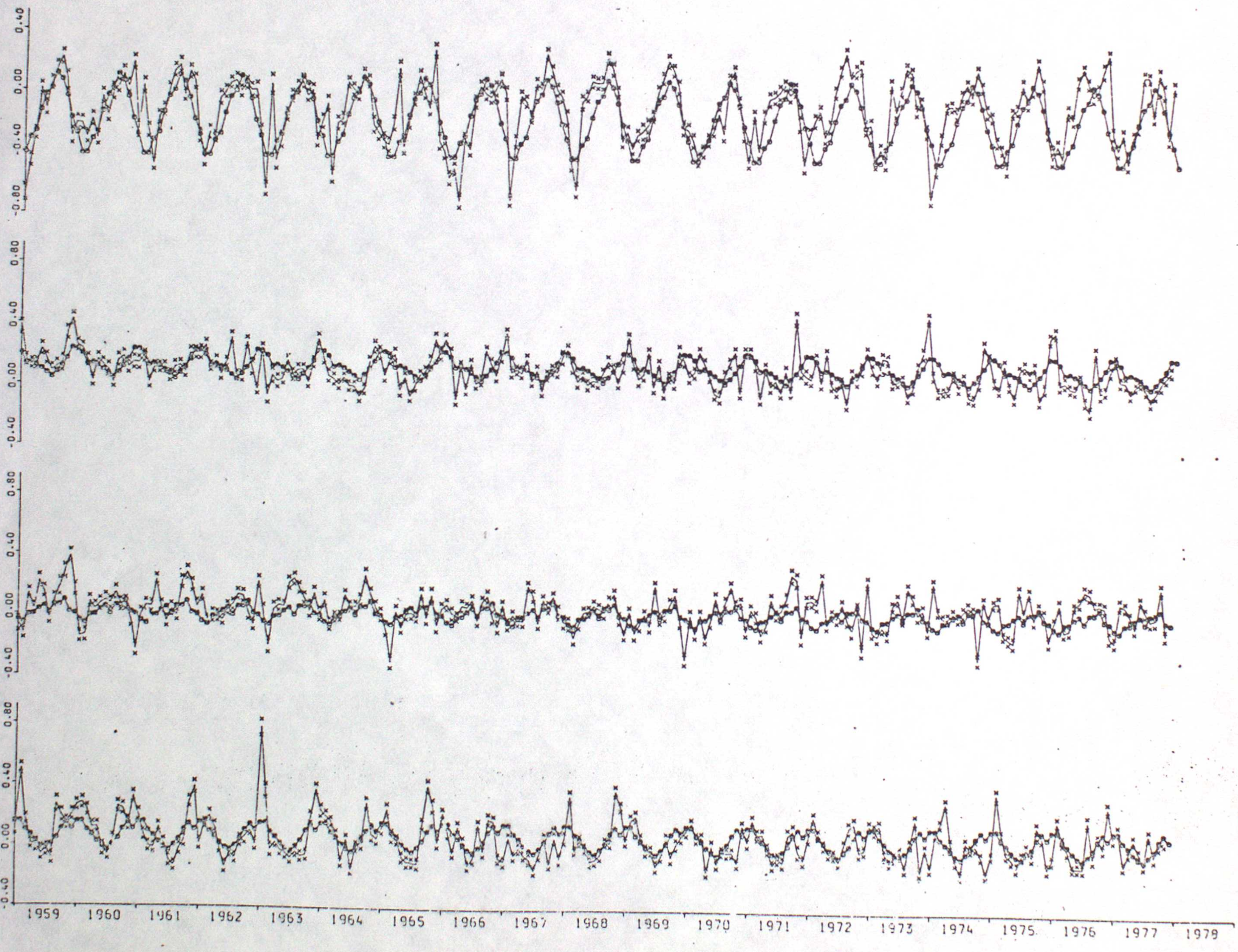
FIGURE 5(i)



FACTOR COEFFICIENTS.

PRODUCED USING BMDP PROGRAM 6R.

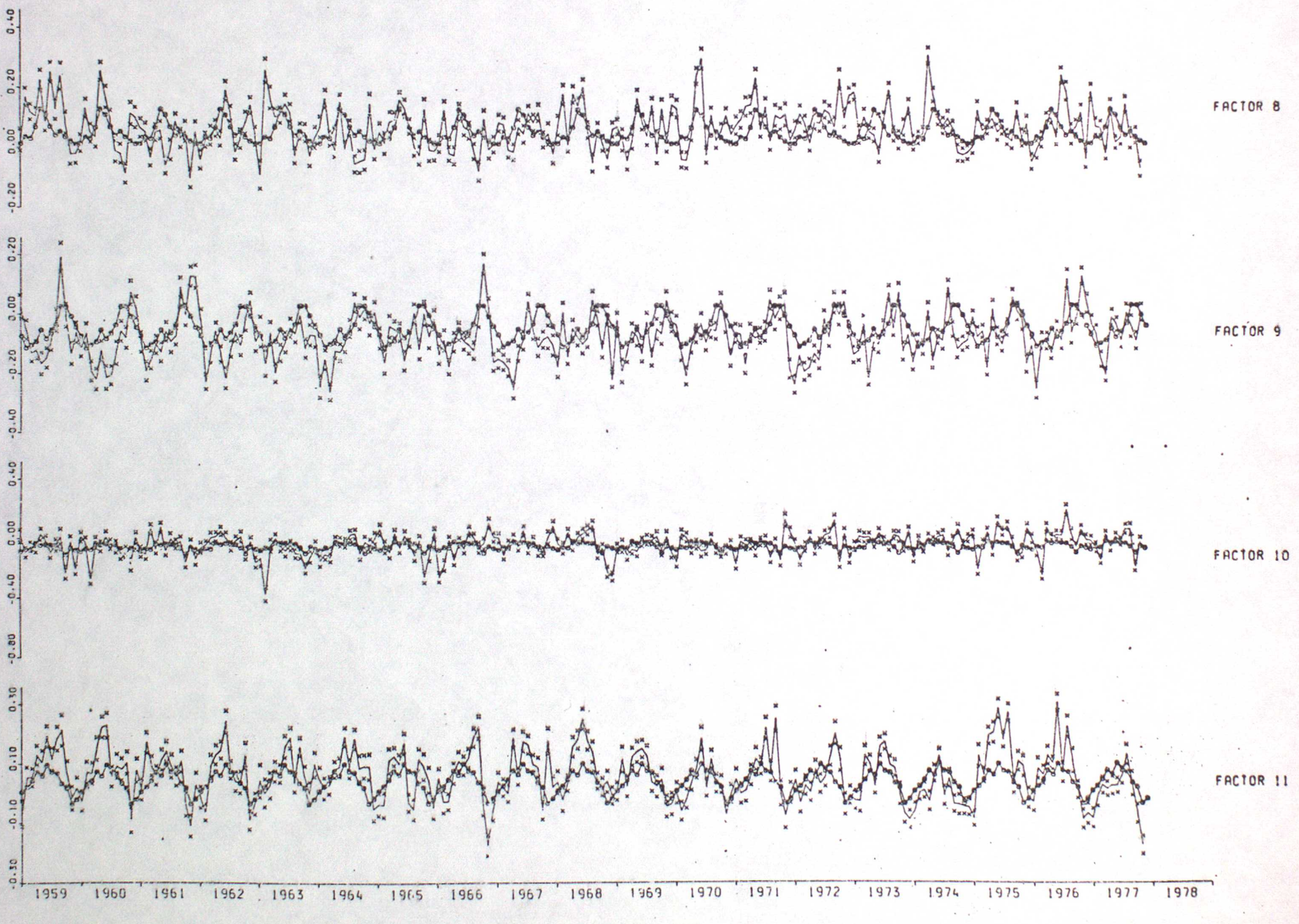
FIGURE 5(ii)



FACTOR COEFFICIENTS.

PRODUCED USING BMDP PROGRAM 6R.

FIGURE 5(iii)



FACTOR COEFFICIENTS.

PRODUCED USING BMDP PROGRAM 6R.

FIGURE 5(iv)

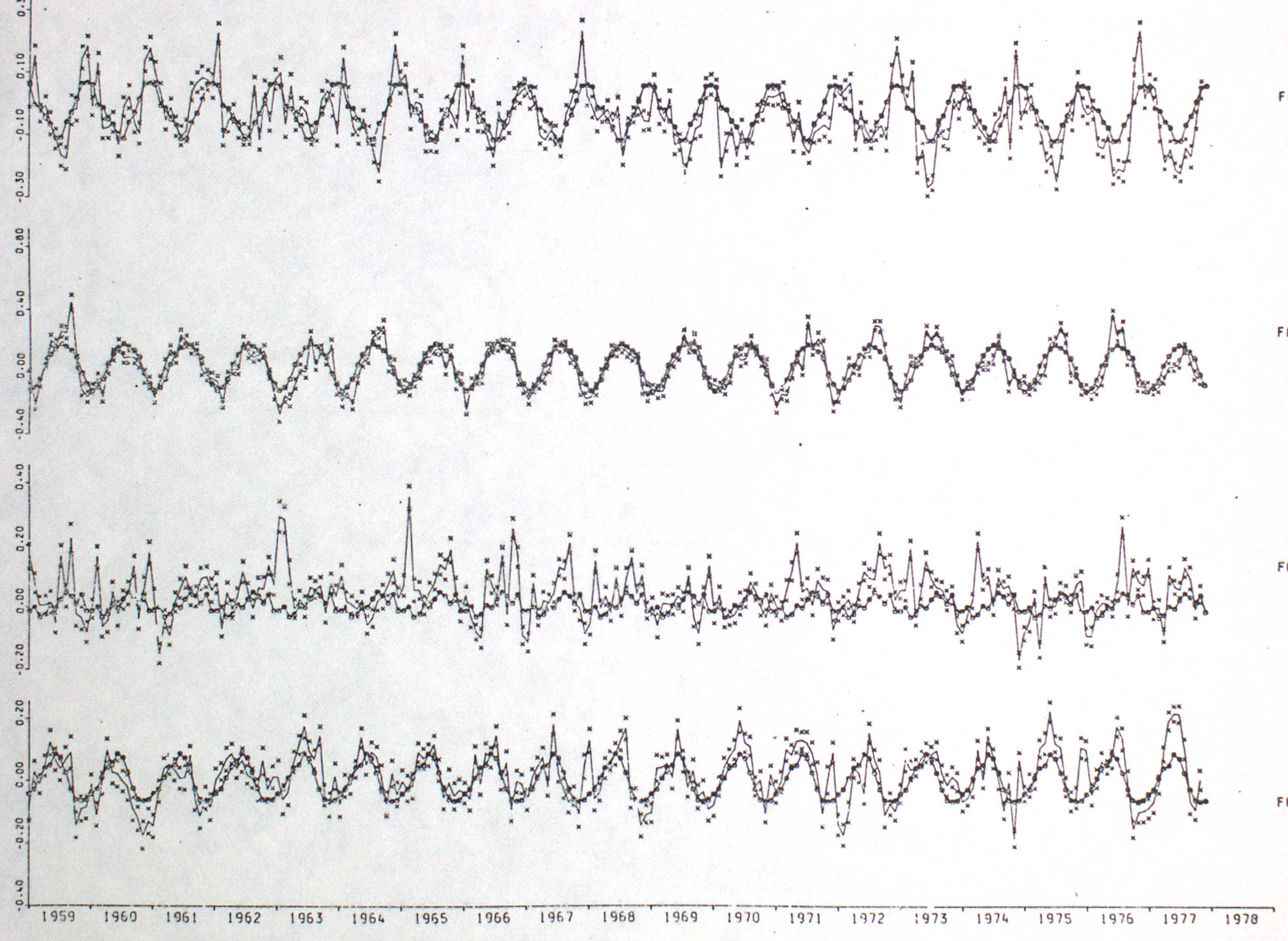
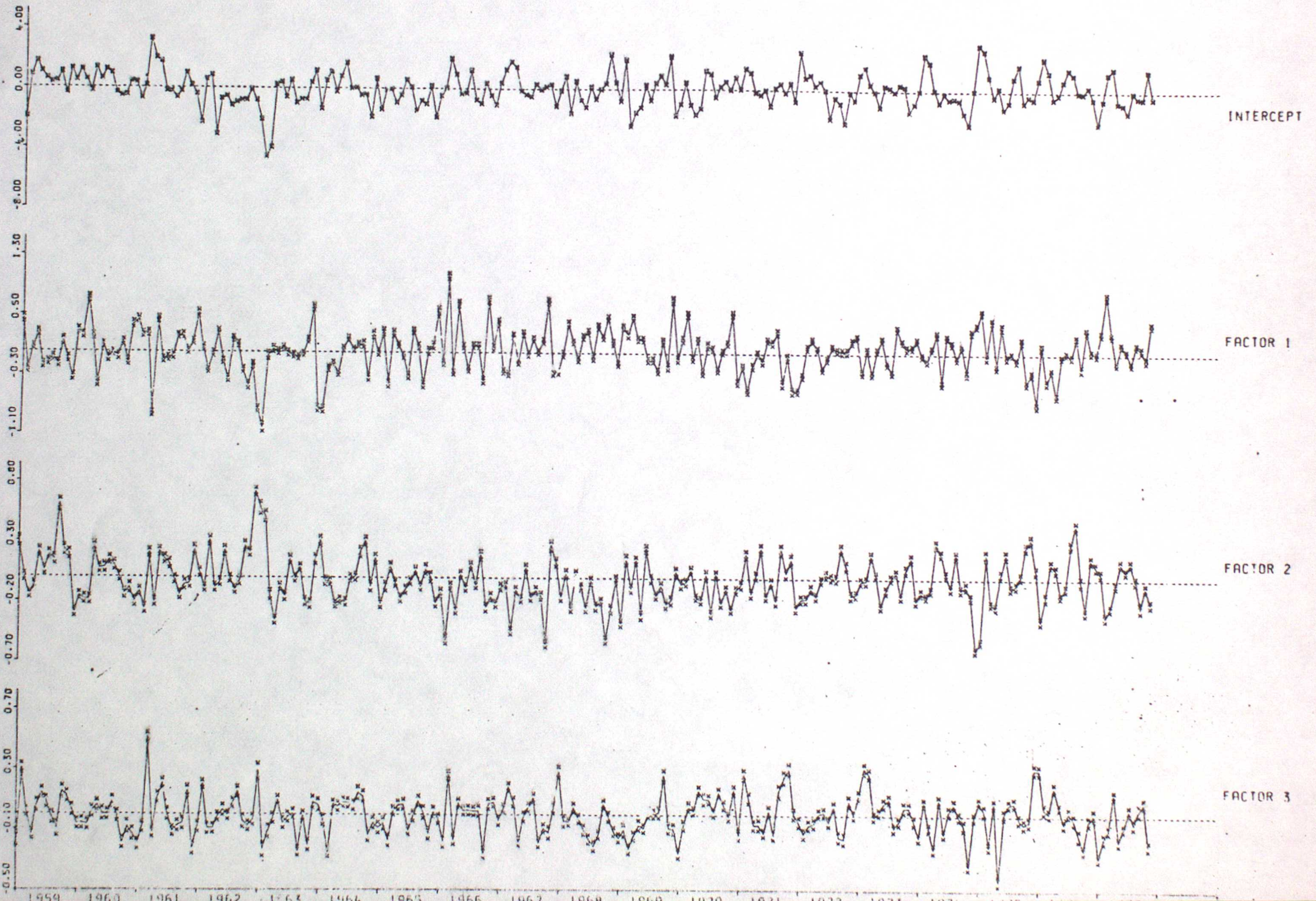


FIGURE 6(i)

FACTOR COEFFICIENTS.

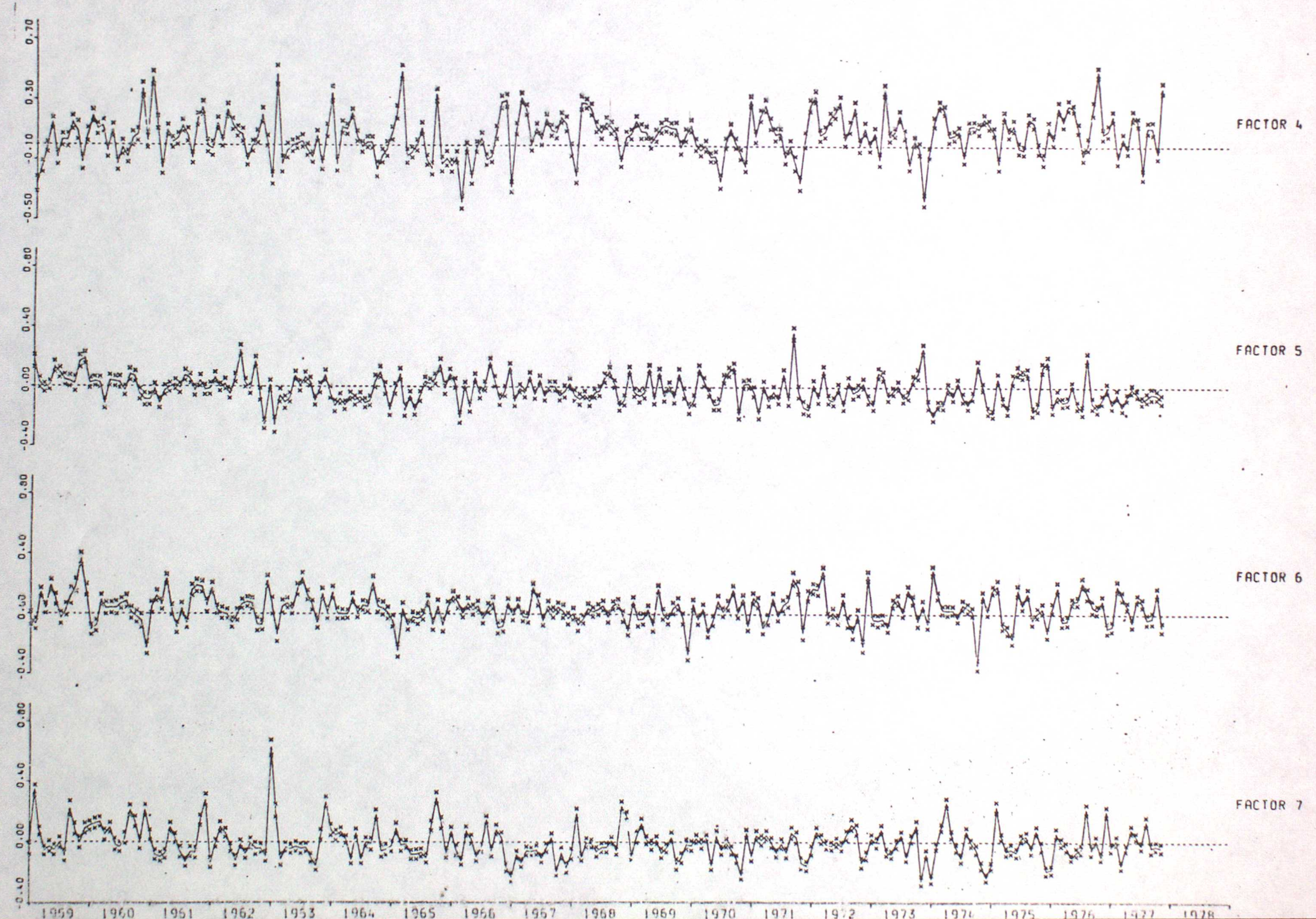
PRODUCED USING BMDP PROGRAM 6R.



FACTOR COEFFICIENTS.

PRODUCED USING BMDP PROGRAM 6R.

FIGURE 6(ii)



FACTOR COEFFICIENTS.

PRODUCED USING BMDP PROGRAM 6R.

FIGURE 6(iii)

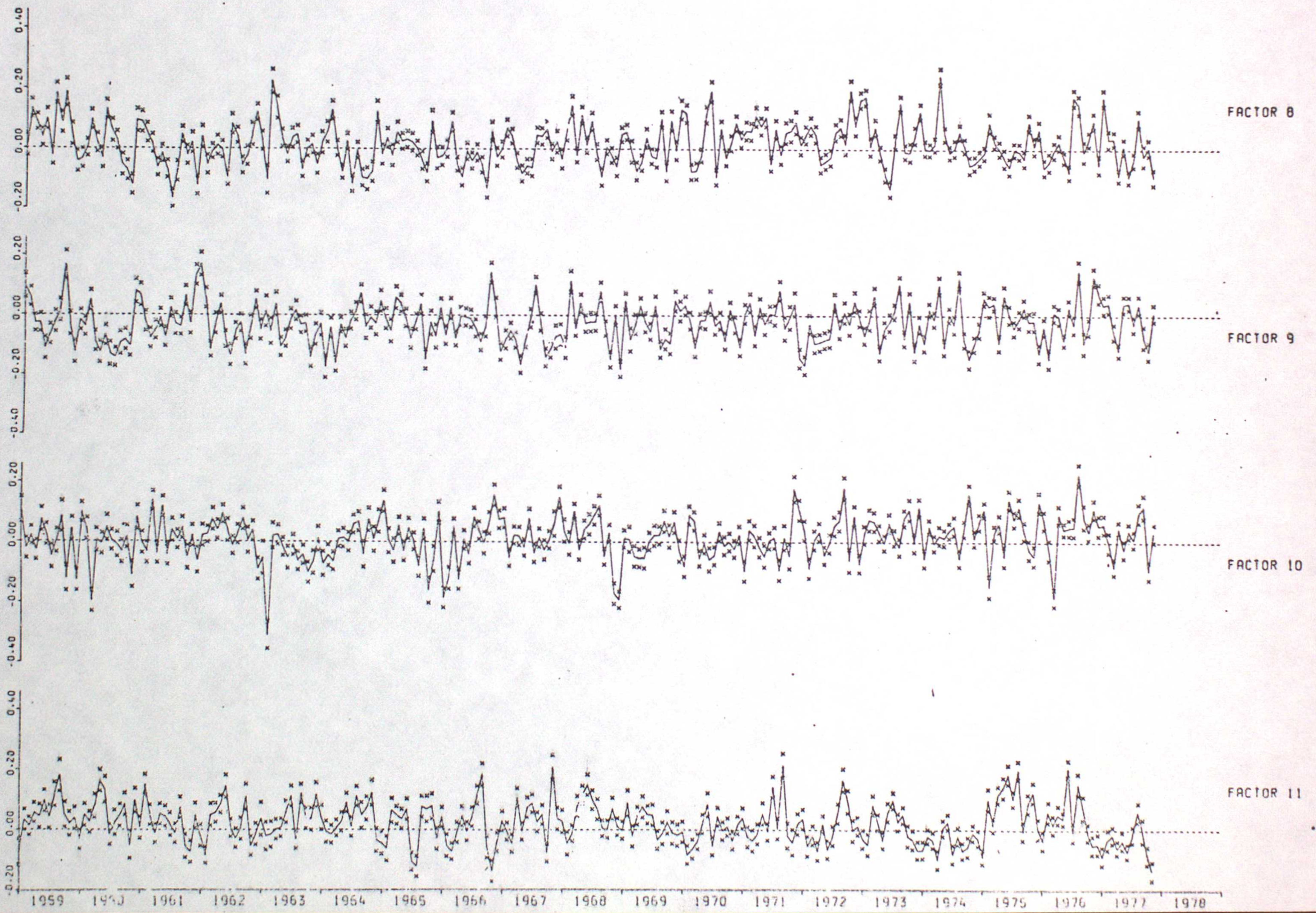


FIGURE 6(iv)

FACTOR COEFFICIENTS.

PRODUCED USING BMDP PPOGRAM 6R.

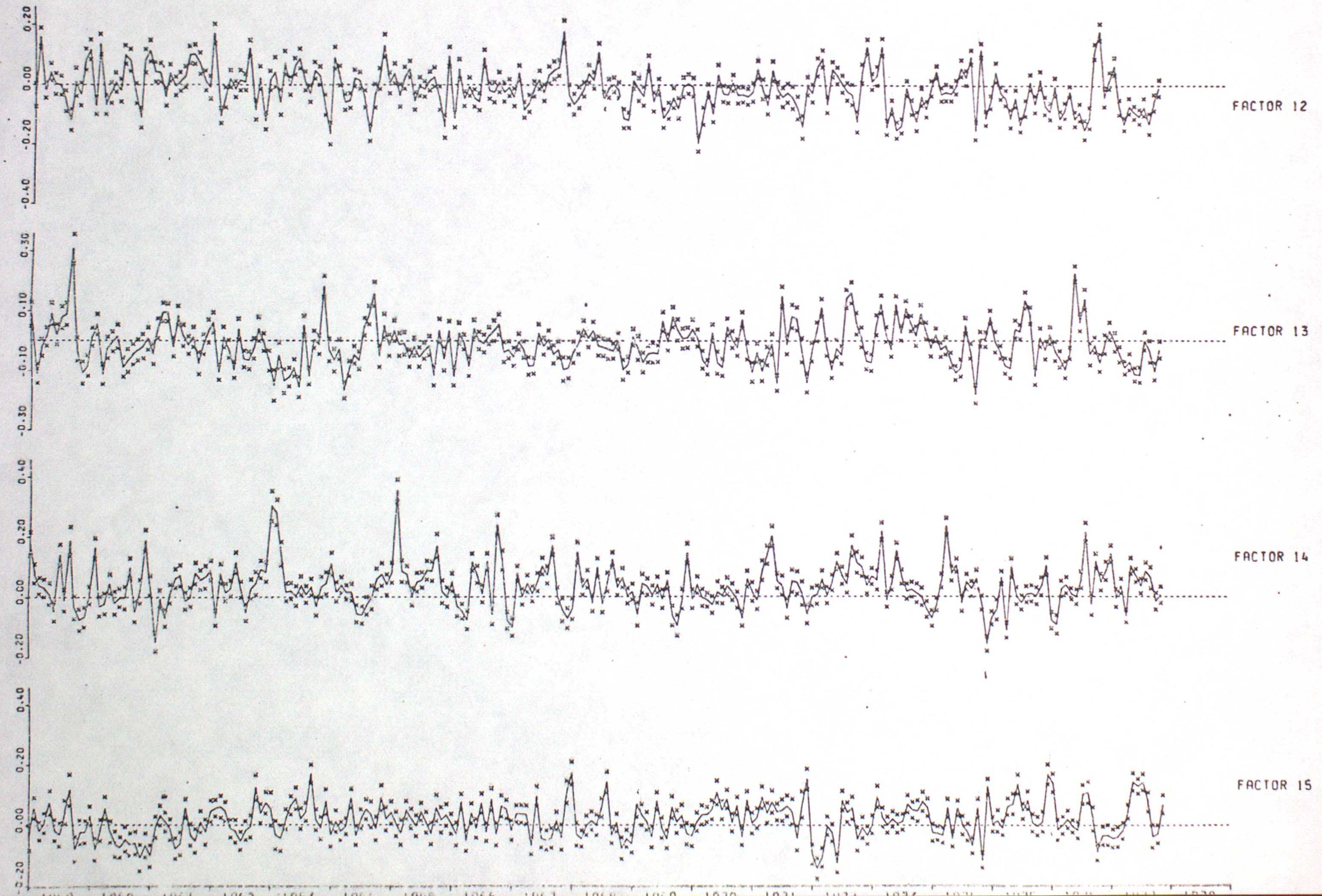
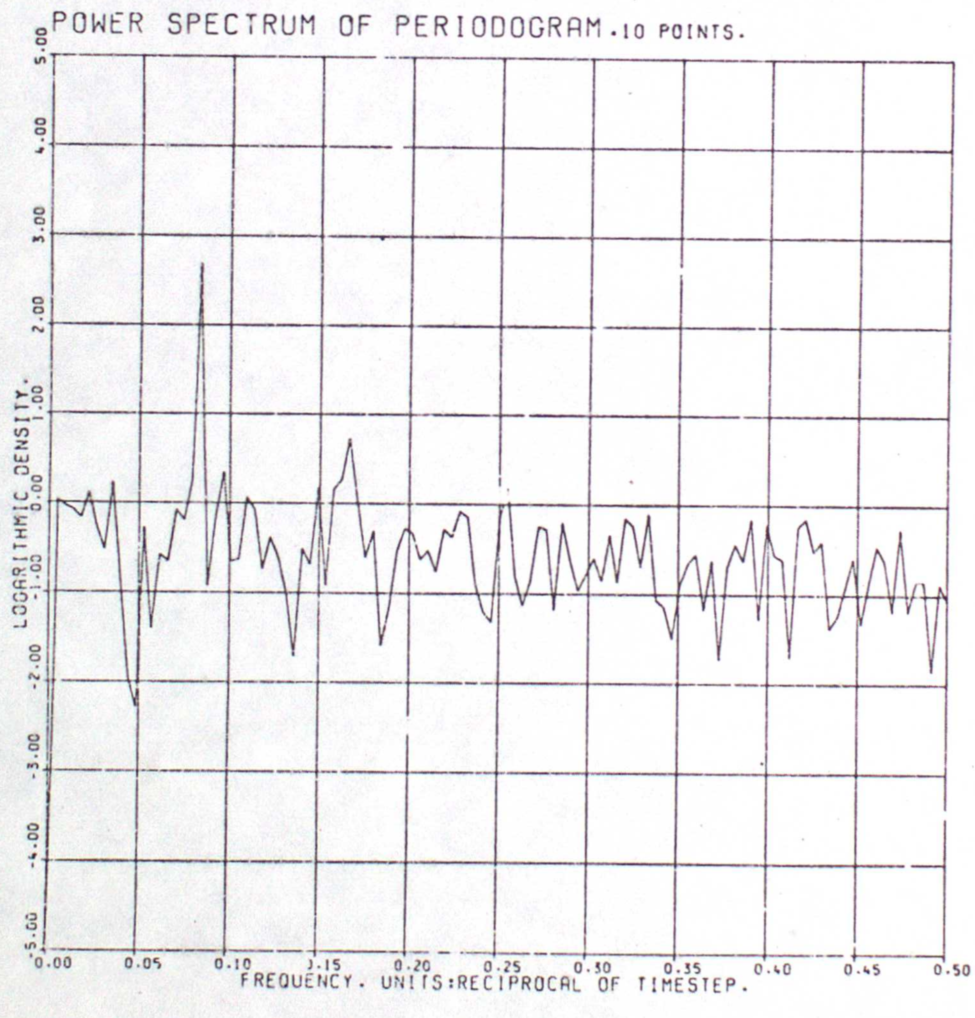


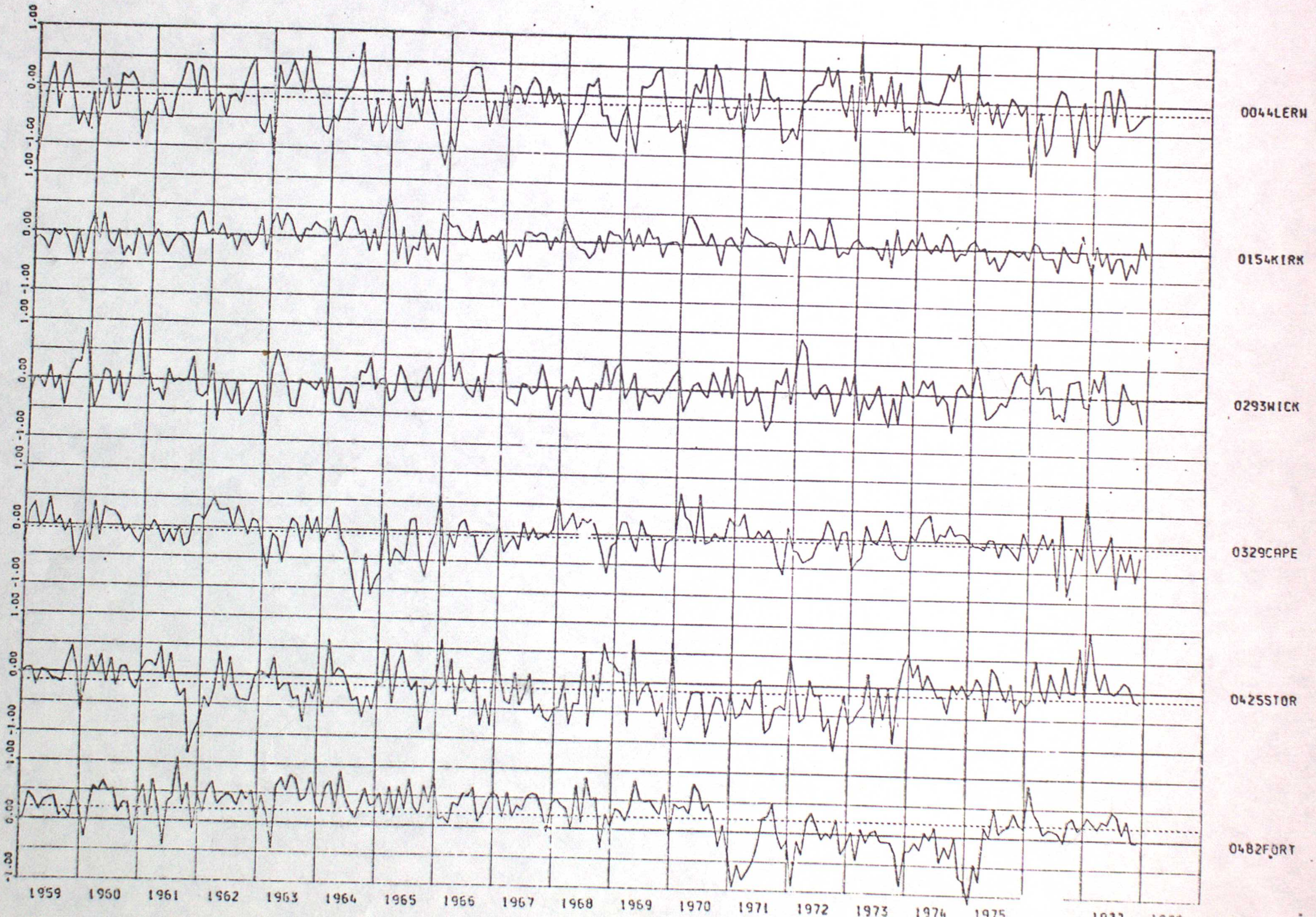
FIGURE 7

LOGARITHMIC AUTOSPECTRUM
1959-77 COEFFICIENTS FOR INTERCEPT



RESIDUALS

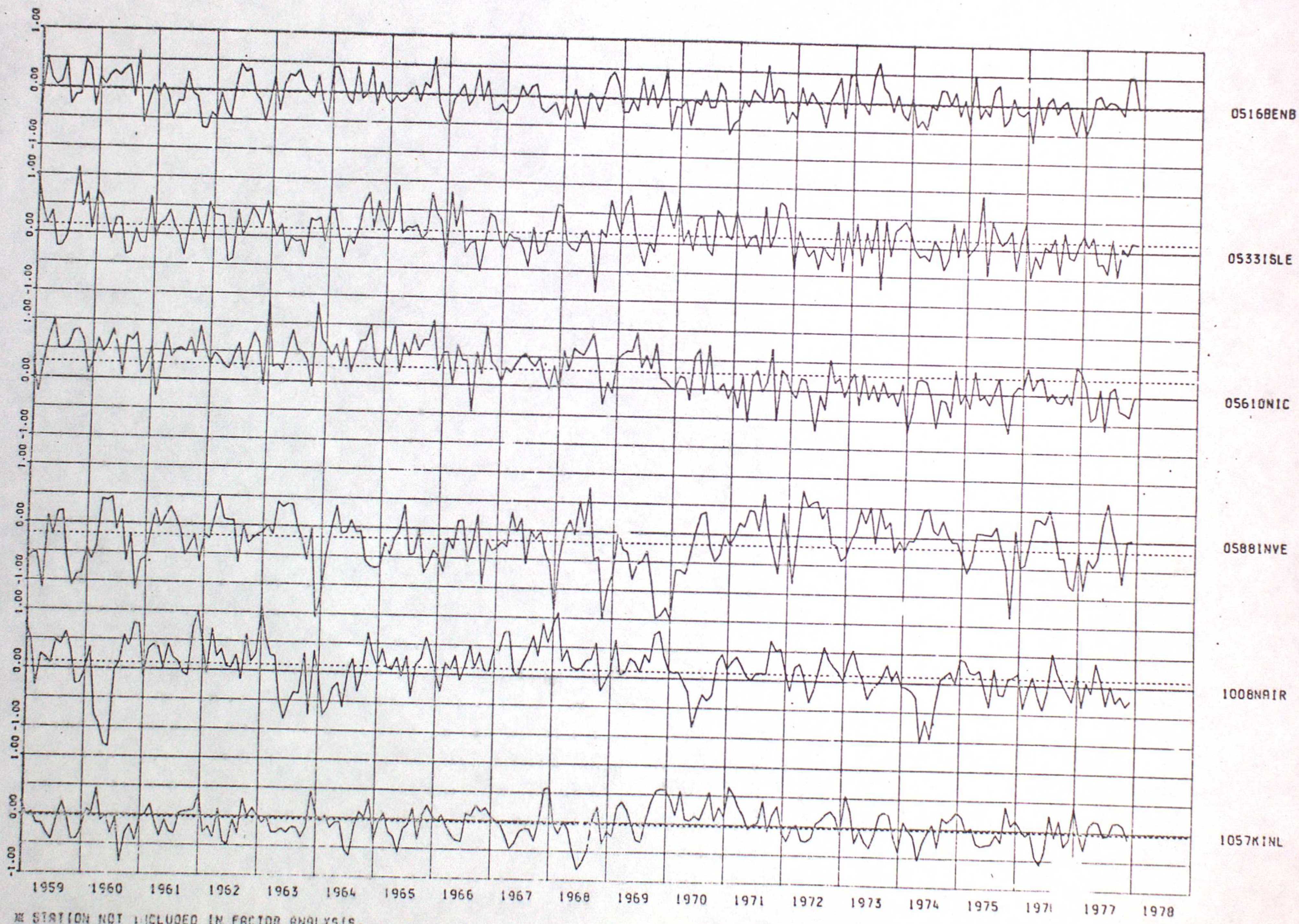
FIGURE 8(i)



* STATION NOT INCLUDED IN FACTOR ANALYSIS.

RESIDUALS

FIGURE 8(ii)



※ STATION NOT INCLUDED IN FACTOR ANALYSIS.

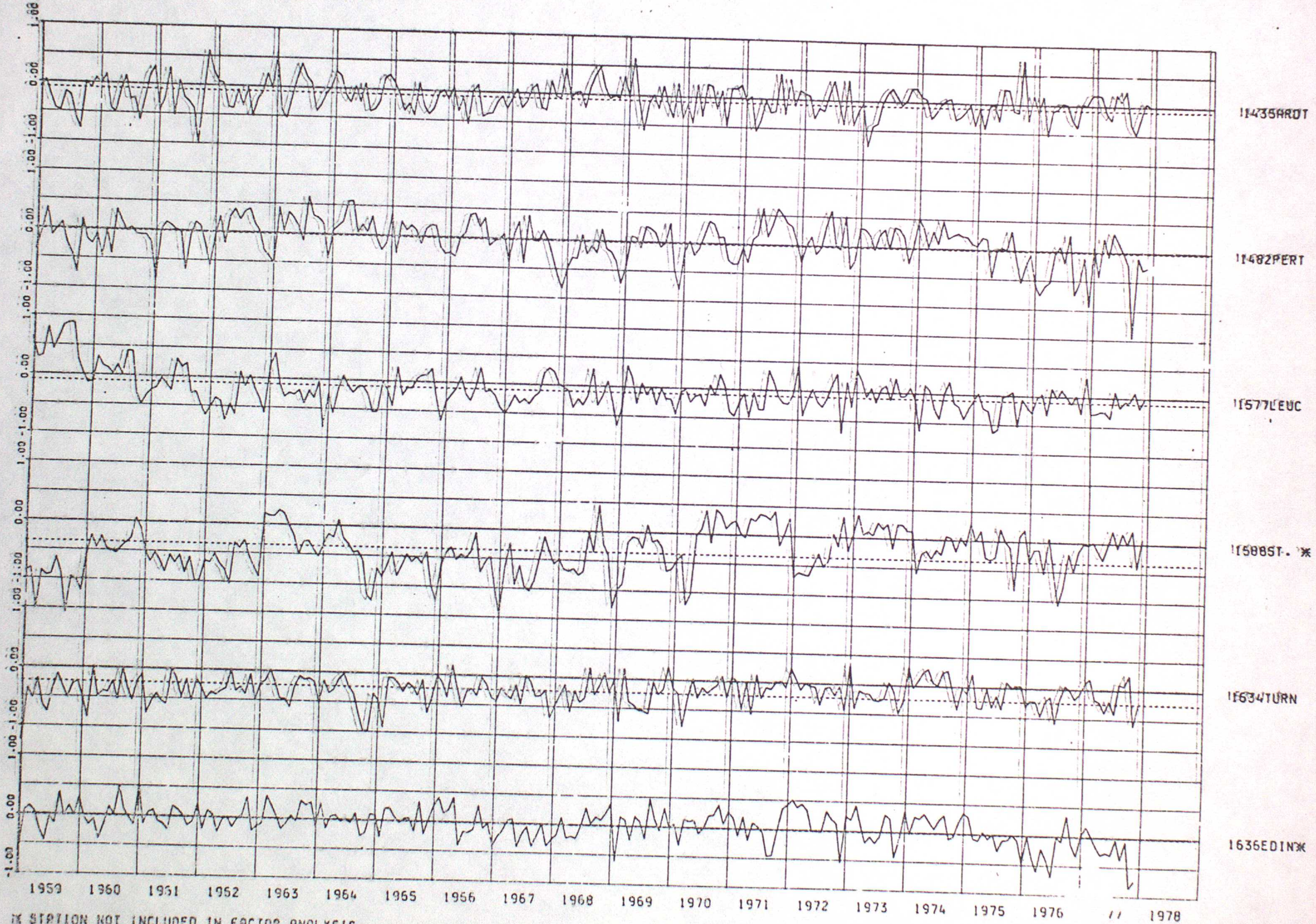
RESIDUALS

FIGURE 8(iii)

* STATION NOT INCLUDED IN FACTOR ANALYSIS.

RESIDUALS

FIGURE 8(iv)



* STATION NOT INCLUDED IN FACTOR ANALYSIS.

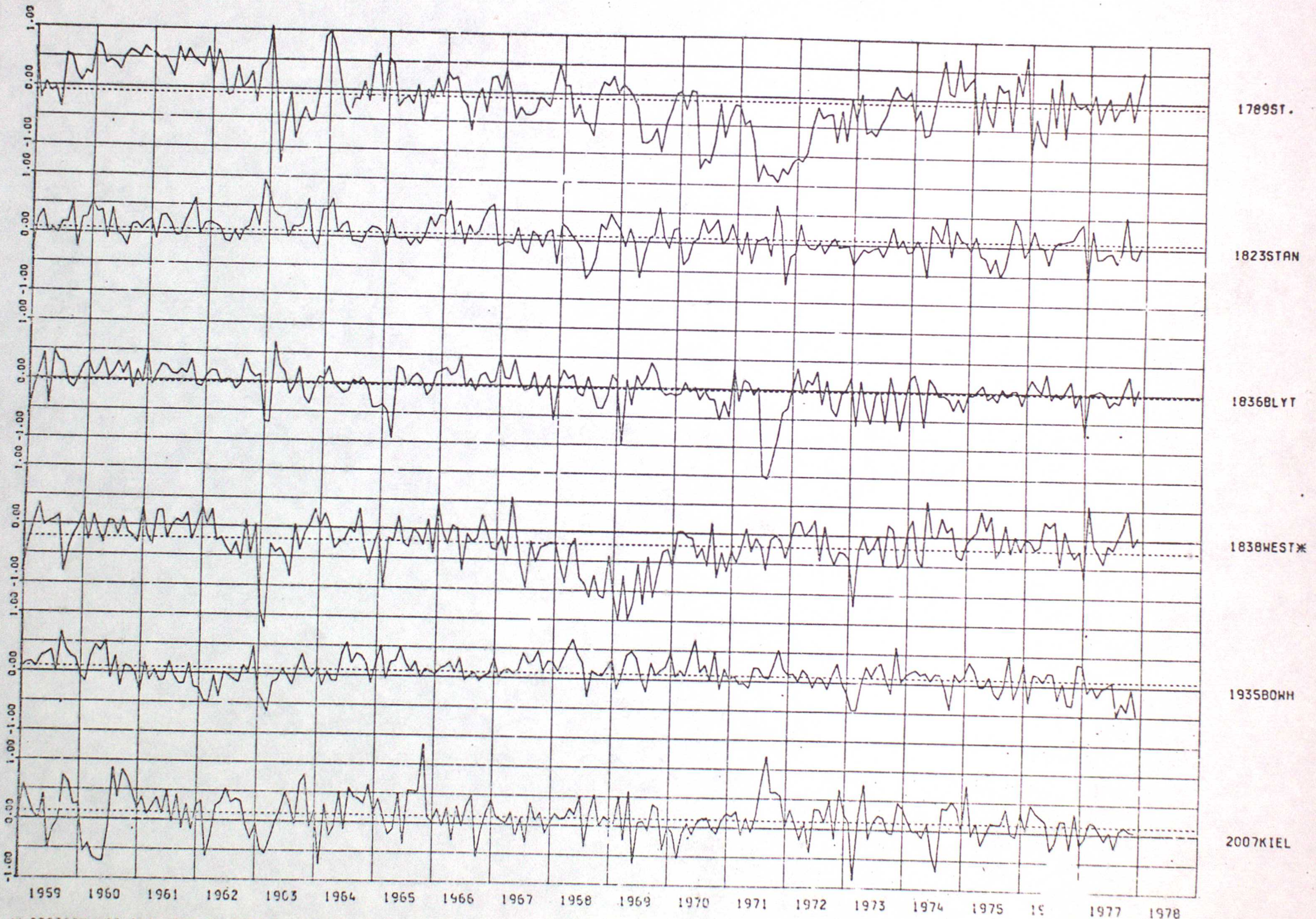
RESIDUALS

FIGURE 8(v)

* STATION NOT INCLUDED IN FACTOR ANALYSIS.

RESIDUALS

FIGURE 8(vi)



* STATION NOT INCLUDED IN FACTOR ANALYSIS.

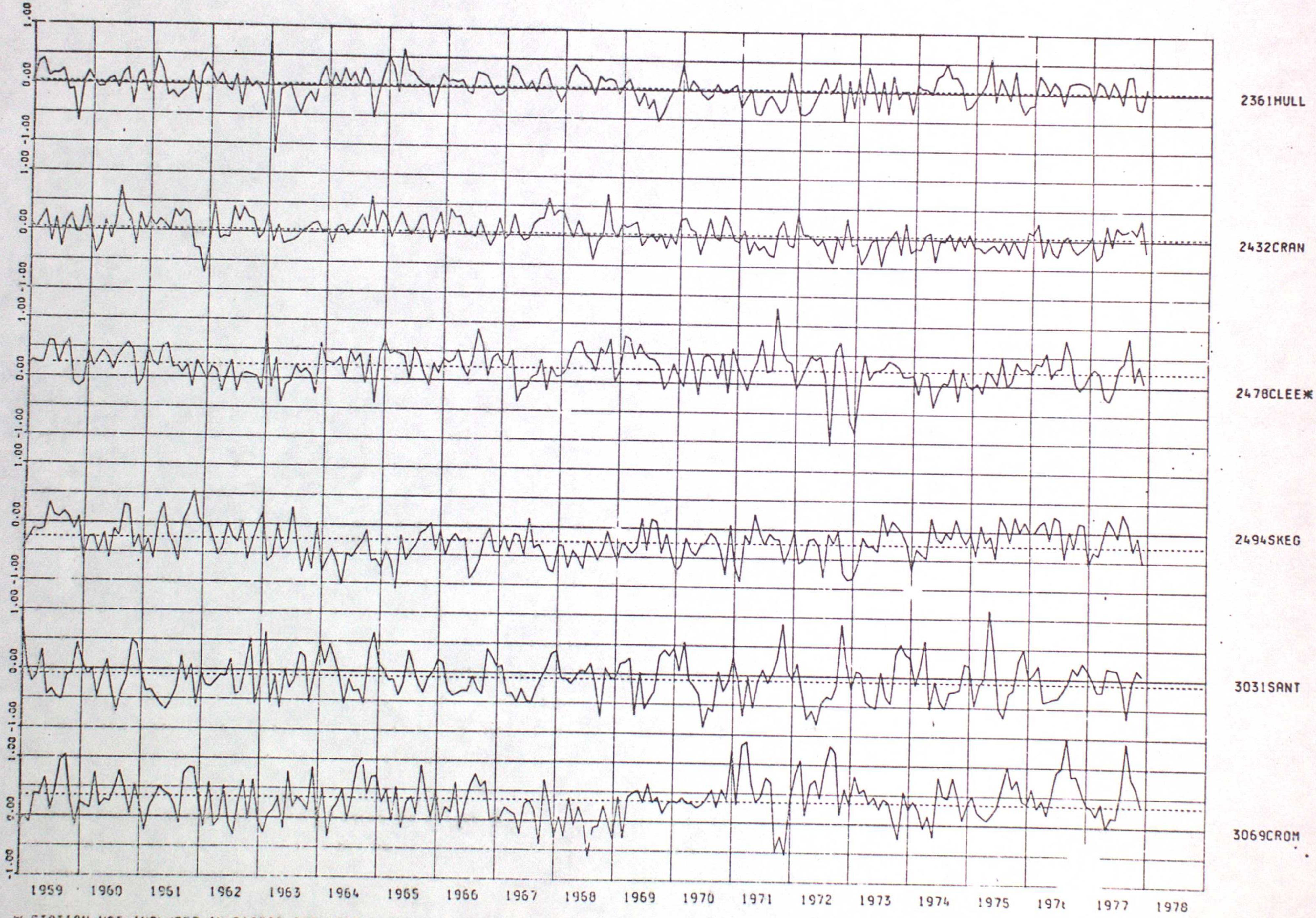
RESIDUALS

FIGURE (vii)

* STATION NOT INCLUDED IN FACTOR ANALYSIS.

RESIDUALS

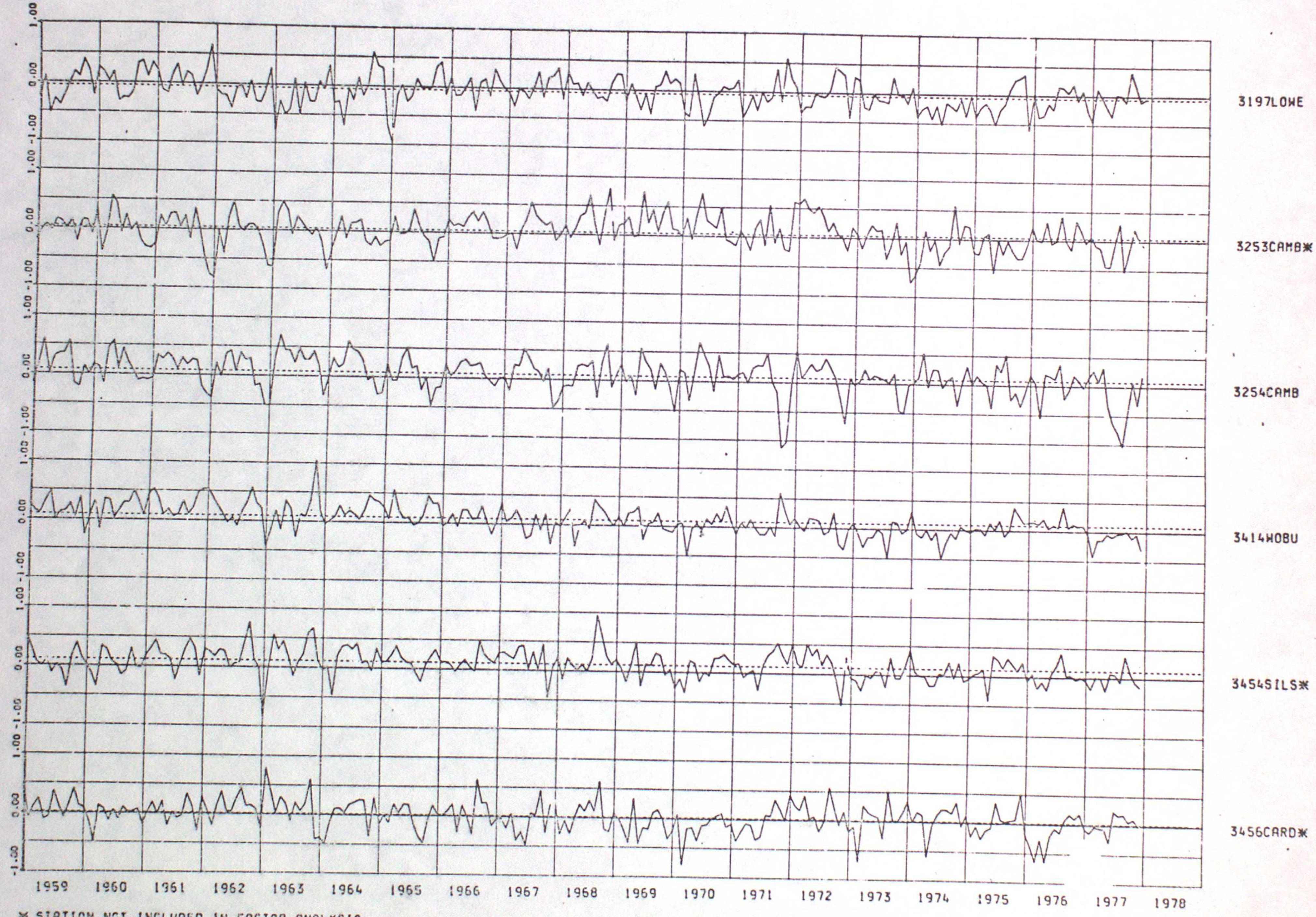
FIGURE 8(viii)



* STATION NOT INCLUDED IN FACTOR ANALYSIS.

RESIDUALS

FIGURE 8(ix)



* STATION NOT INCLUDED IN FACTOR ANALYSIS.

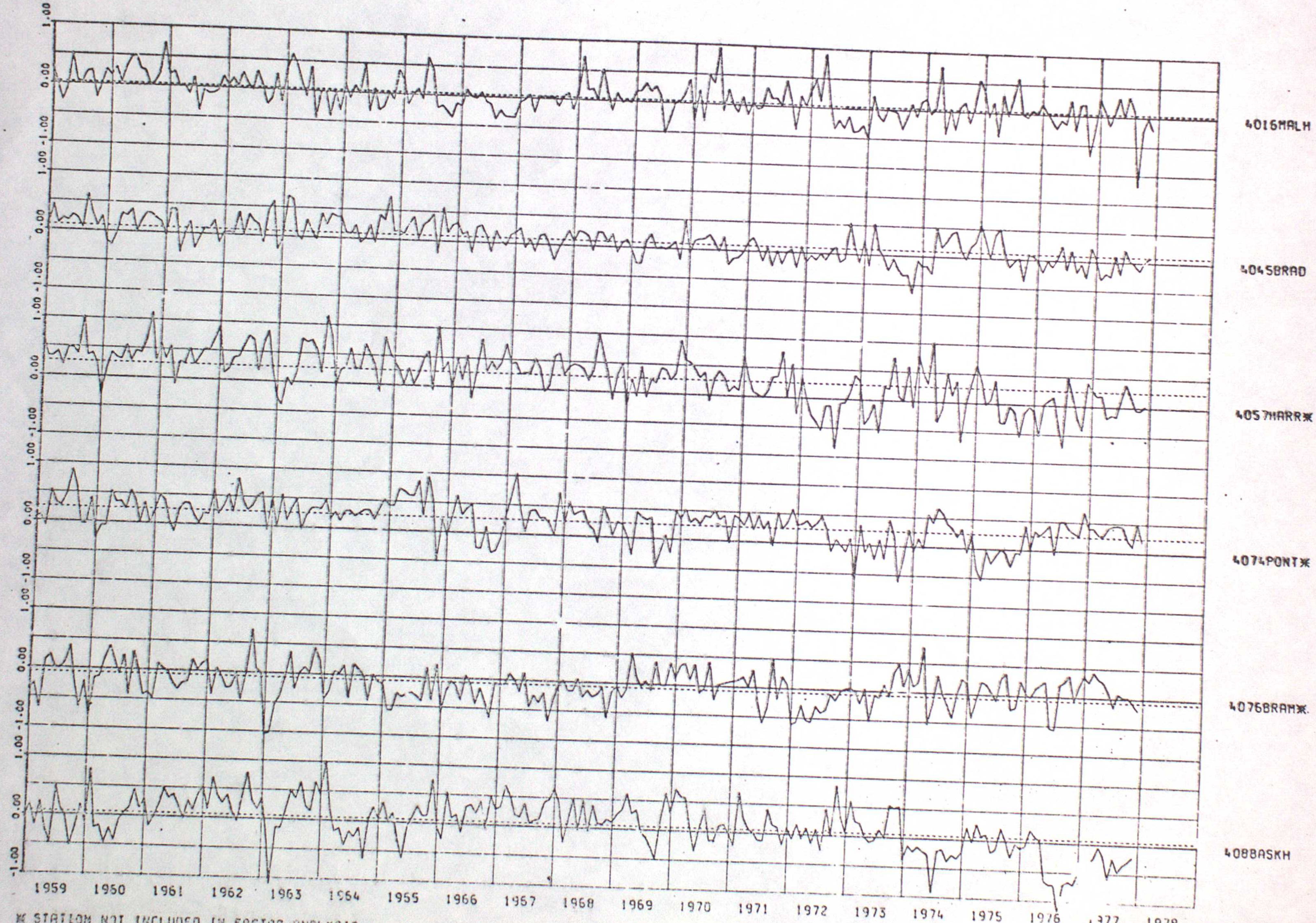
RESIDUALS

FIGURE 8(?)

* STATION NOT INCLUDED IN FACTOR ANALYSIS.

RESIDUALS

FIGURE 8(xi)



* STATION NOT INCLUDED IN FACTOR ANALYSIS.

RESIDUALS

FIGURE · 8(xii)

* STATION NOT INCLUDED IN FACTOR ANALYSIS.

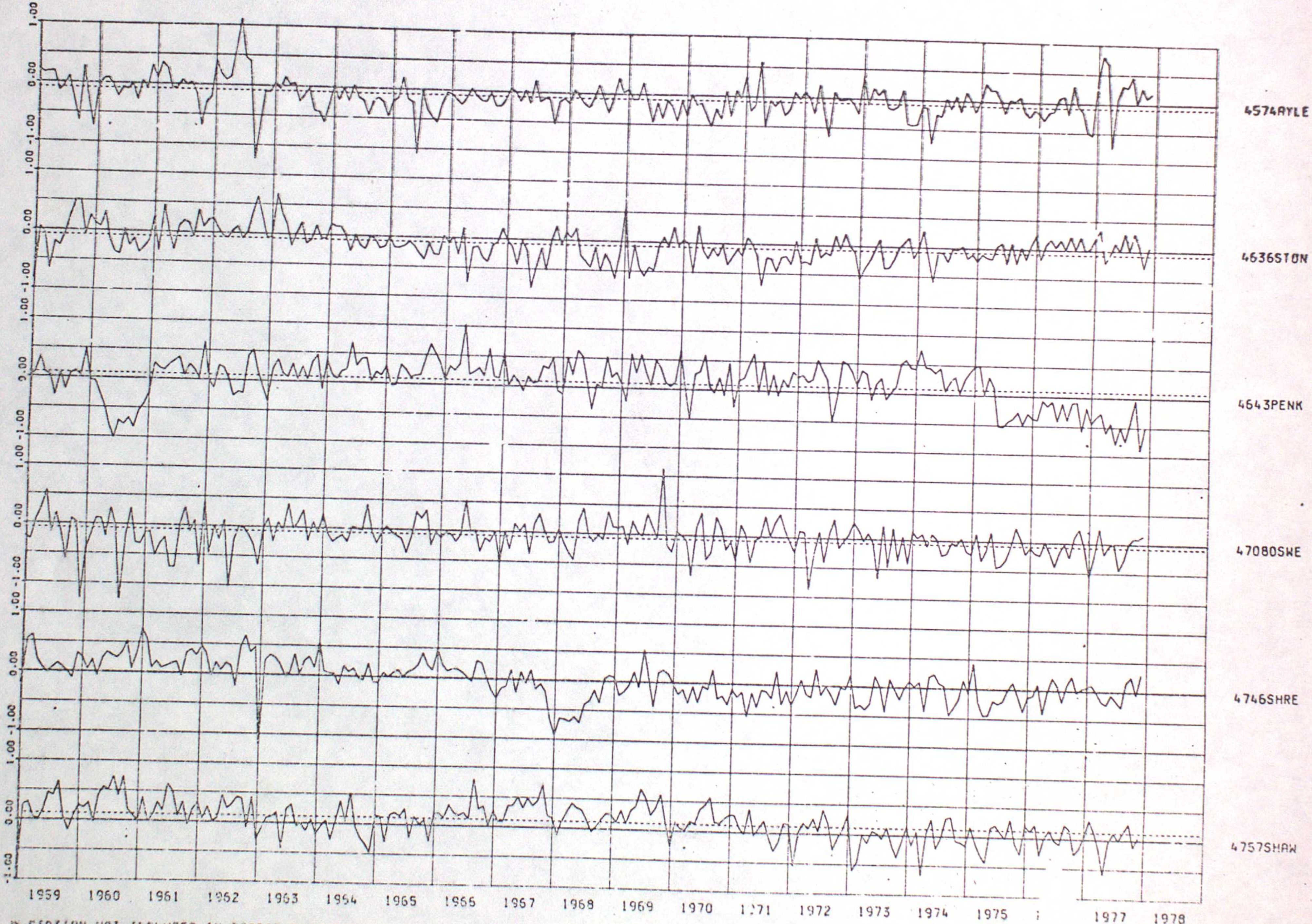
RESIDUALS

FIGURE 8(xiii)

* STATION NOT INCLUDED IN FACTOR ANALYSIS.

RESIDUALS

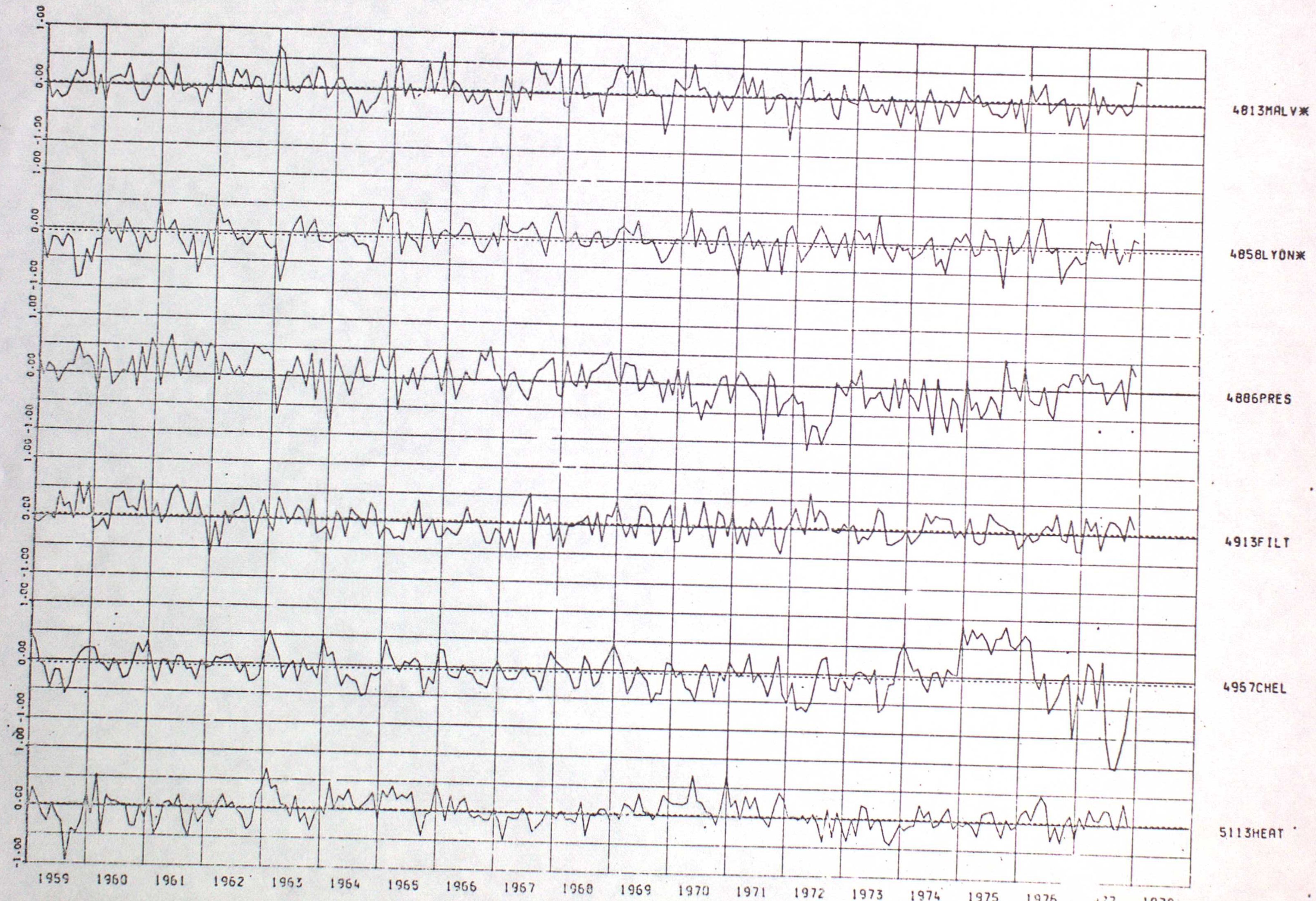
FIGURE 8(xiv)



* STATION NOT INCLUDED IN FACTOR ANALYSIS.

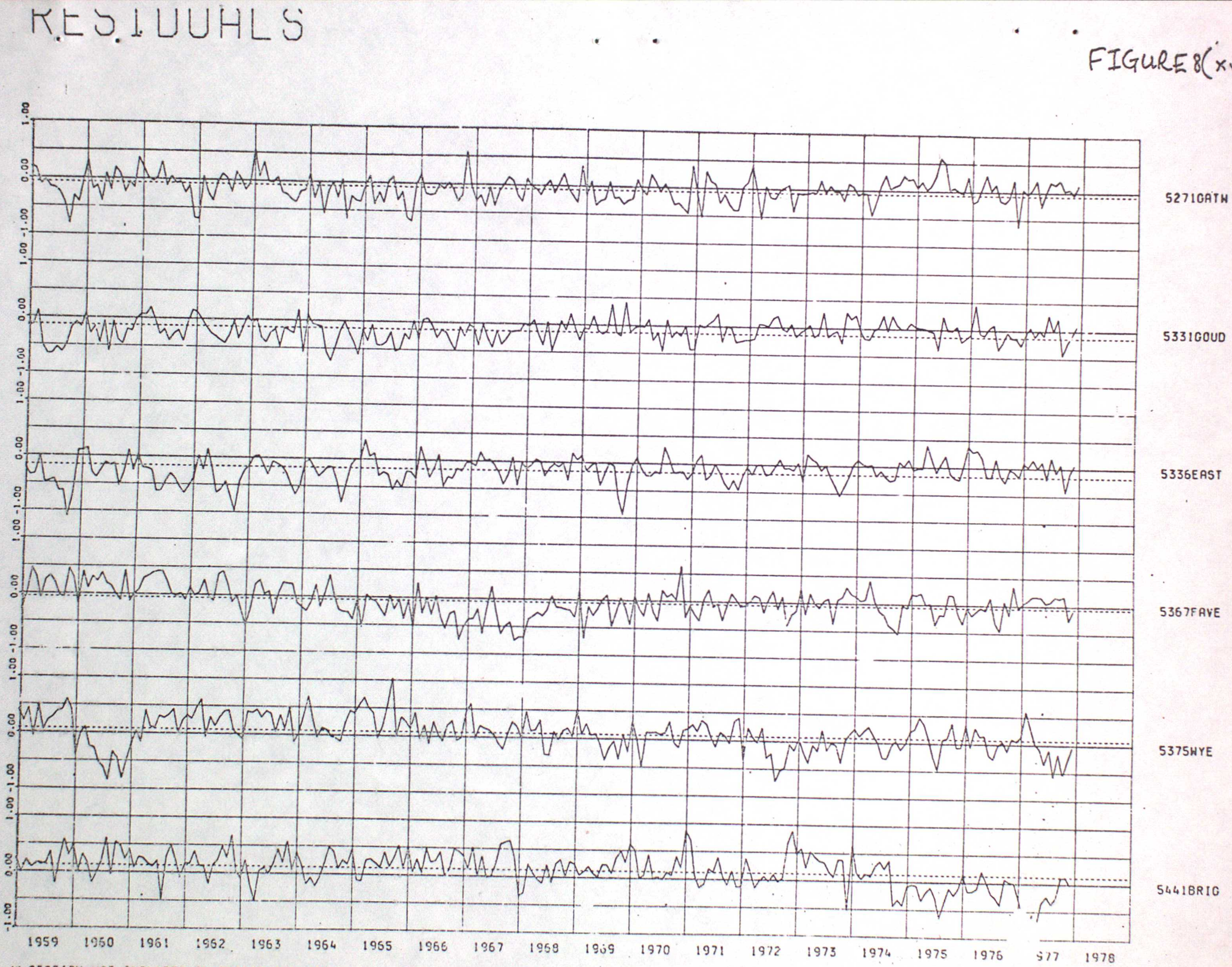
RESIDUALS

FIGURE 8(xv)



* STATION NOT INCLUDED IN FACTOR ANALYSIS.

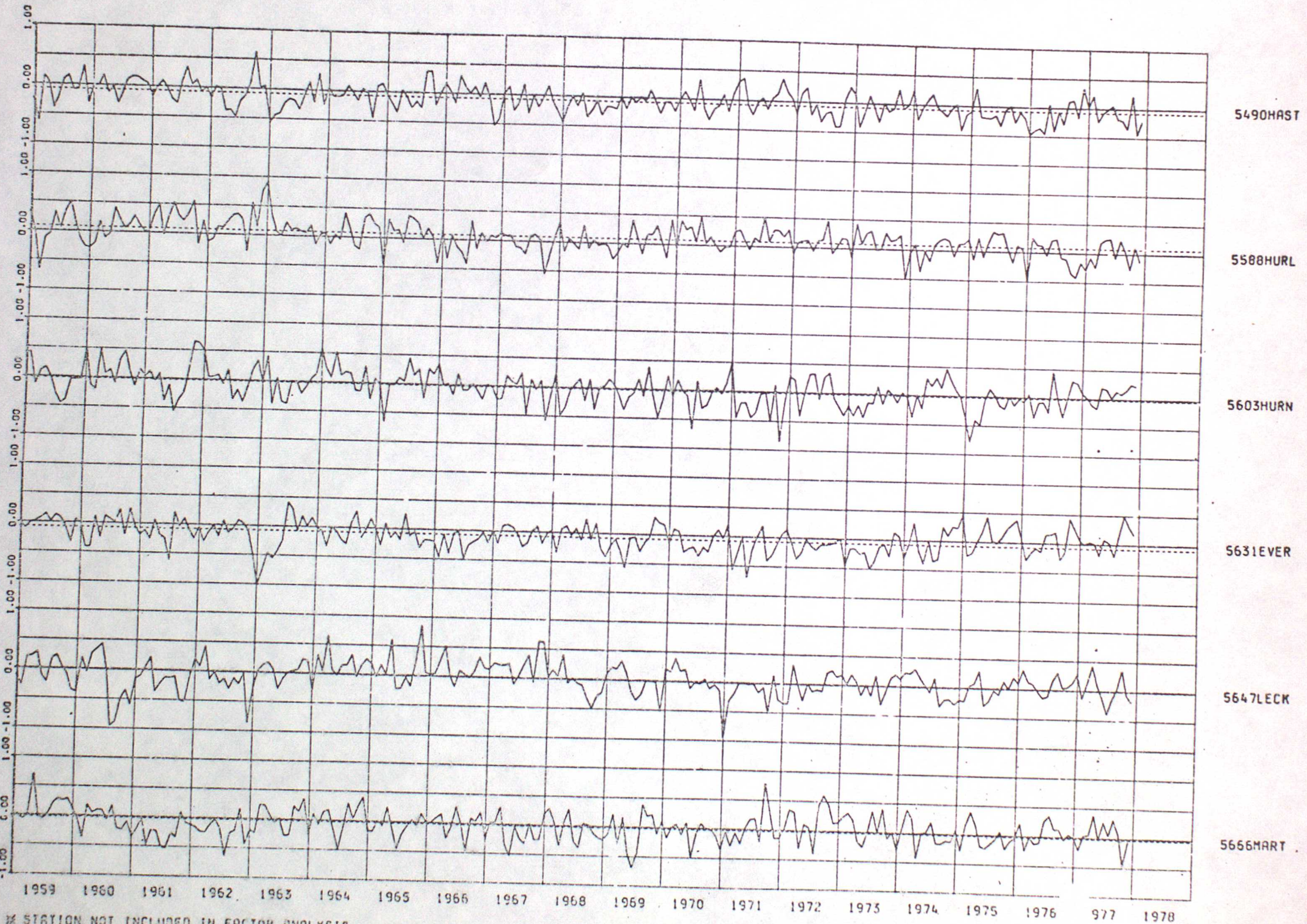
FIGURE 8(xvi)



* STATION NOT INCLUDED IN FACTOR ANALYSIS.

RESIDUALS

FIGURE 8(xvii)



* STATION NOT INCLUDED IN FACTOR ANALYSIS.

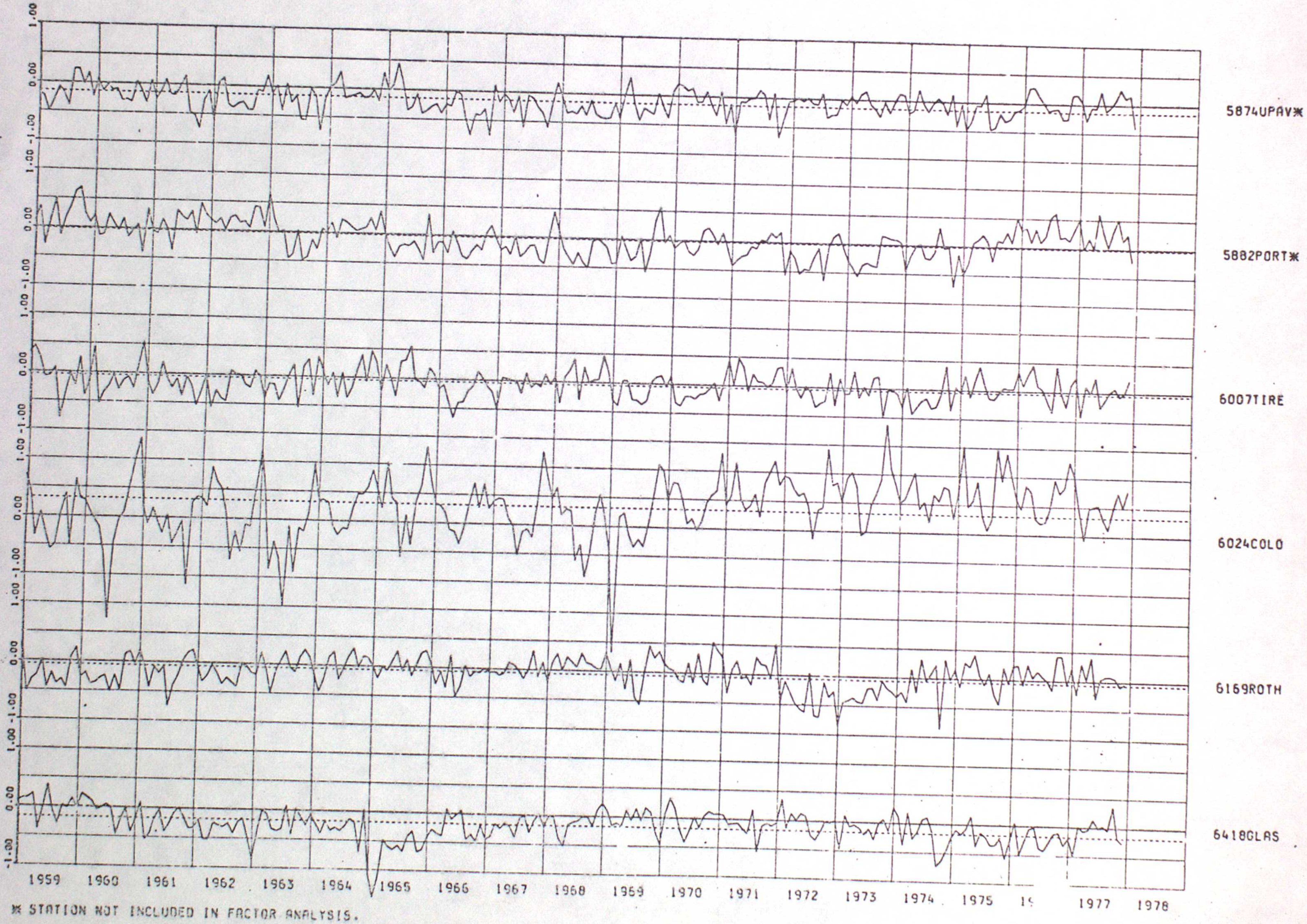
RESIDUALS

FIGURE 8(xviii)

* STATION NOT INCLUDED IN FACTOR ANALYSIS.

RESIDUALS

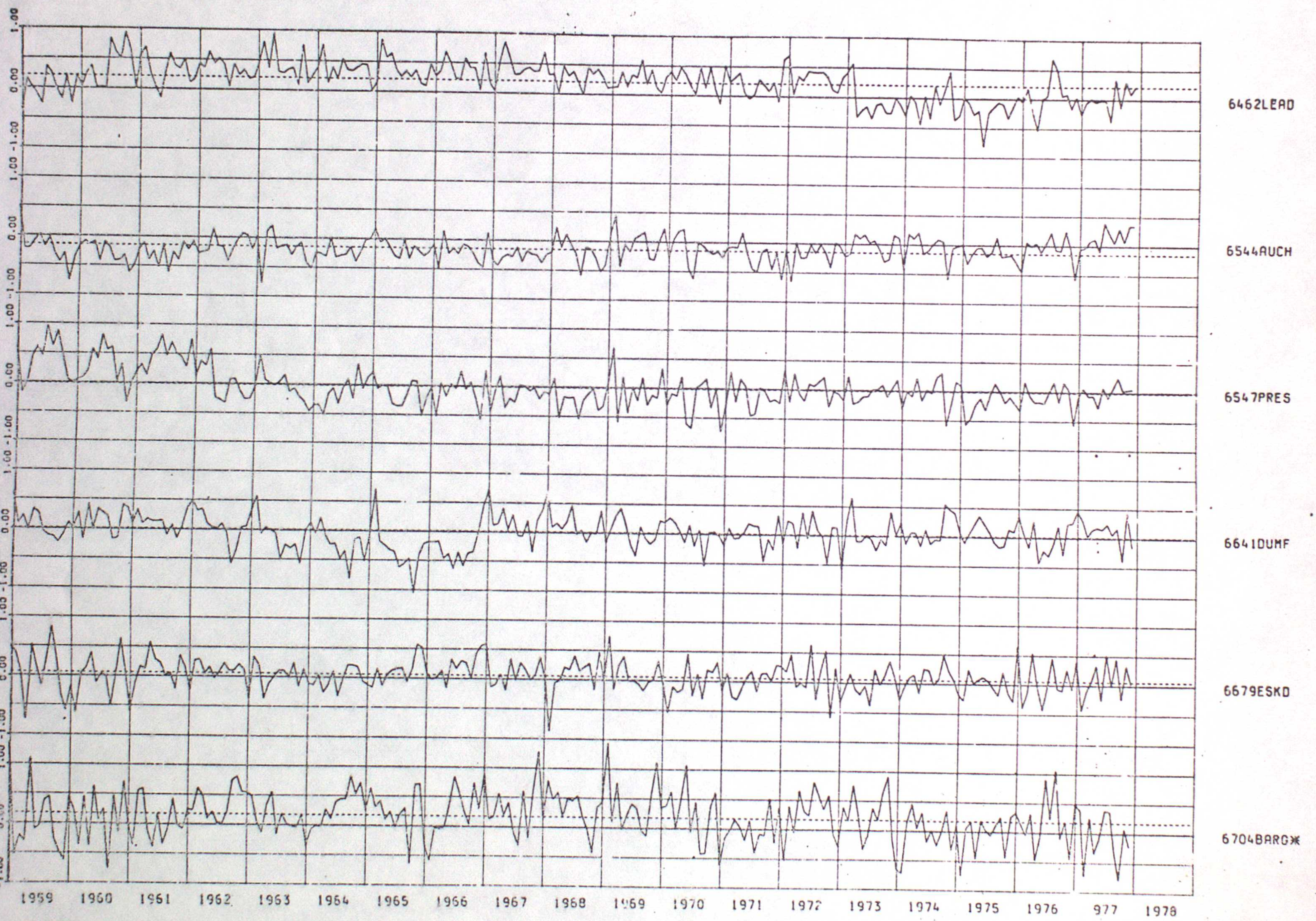
FIGURE 8(xix)



* STATION NOT INCLUDED IN FACTOR ANALYSIS.

RESIDUALS

FIGURE 8(xx)



* STATION NOT INCLUDED IN FACTOR ANALYSIS.

RESIDUALS

FIGURE 8.(xxi)

* STATION NOT INCLUDED IN FACTOR ANALYSIS.

RESIDUALS

FIGURE 8 (xxii)

* STATION NOT INCLUDED IN FACTOR ANALYSIS.

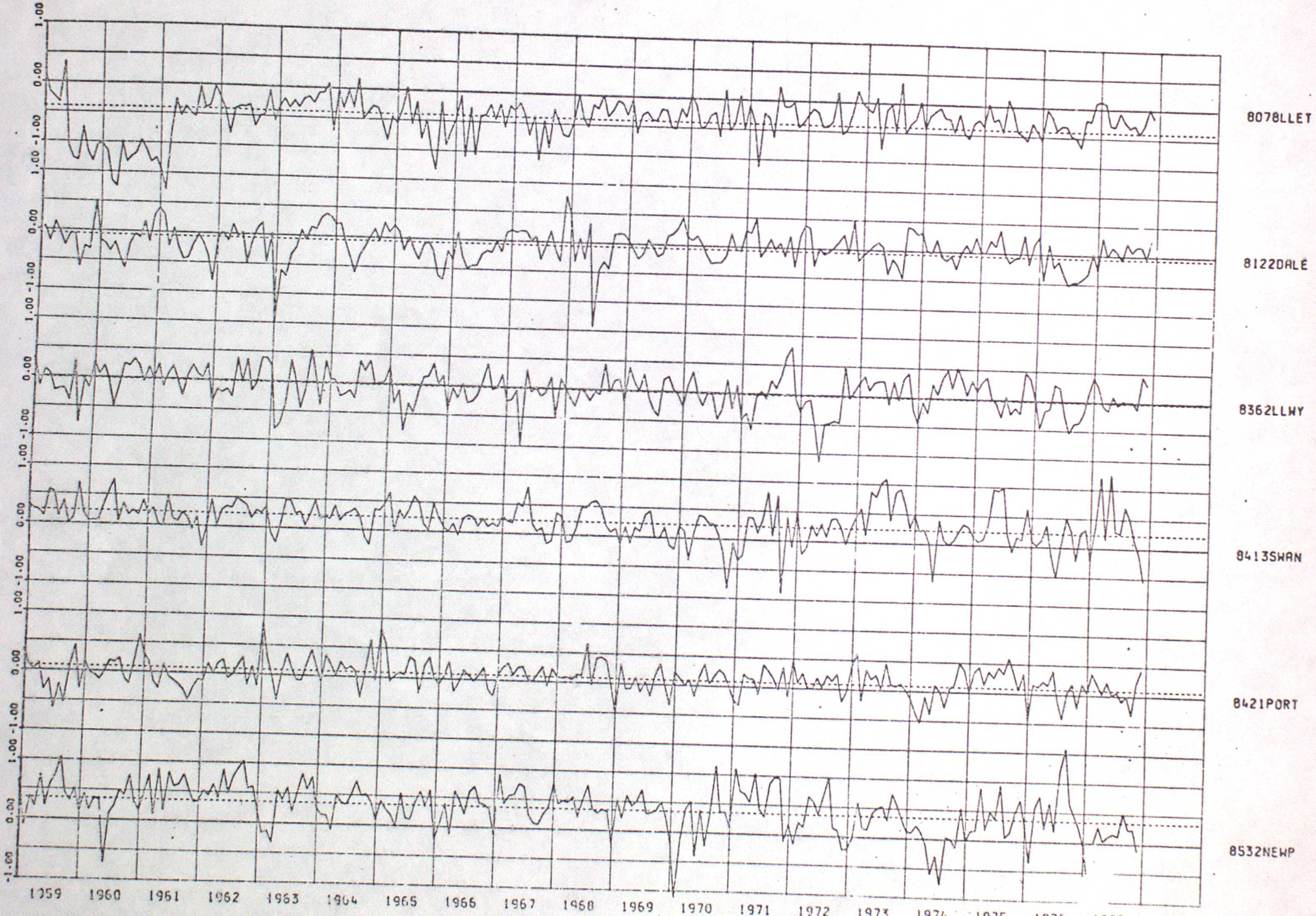
RESIDUALS

FIGURE 8(xxiii)

* STATION NOT INCLUDED IN FACTOR ANALYSIS.

RESIDUALS

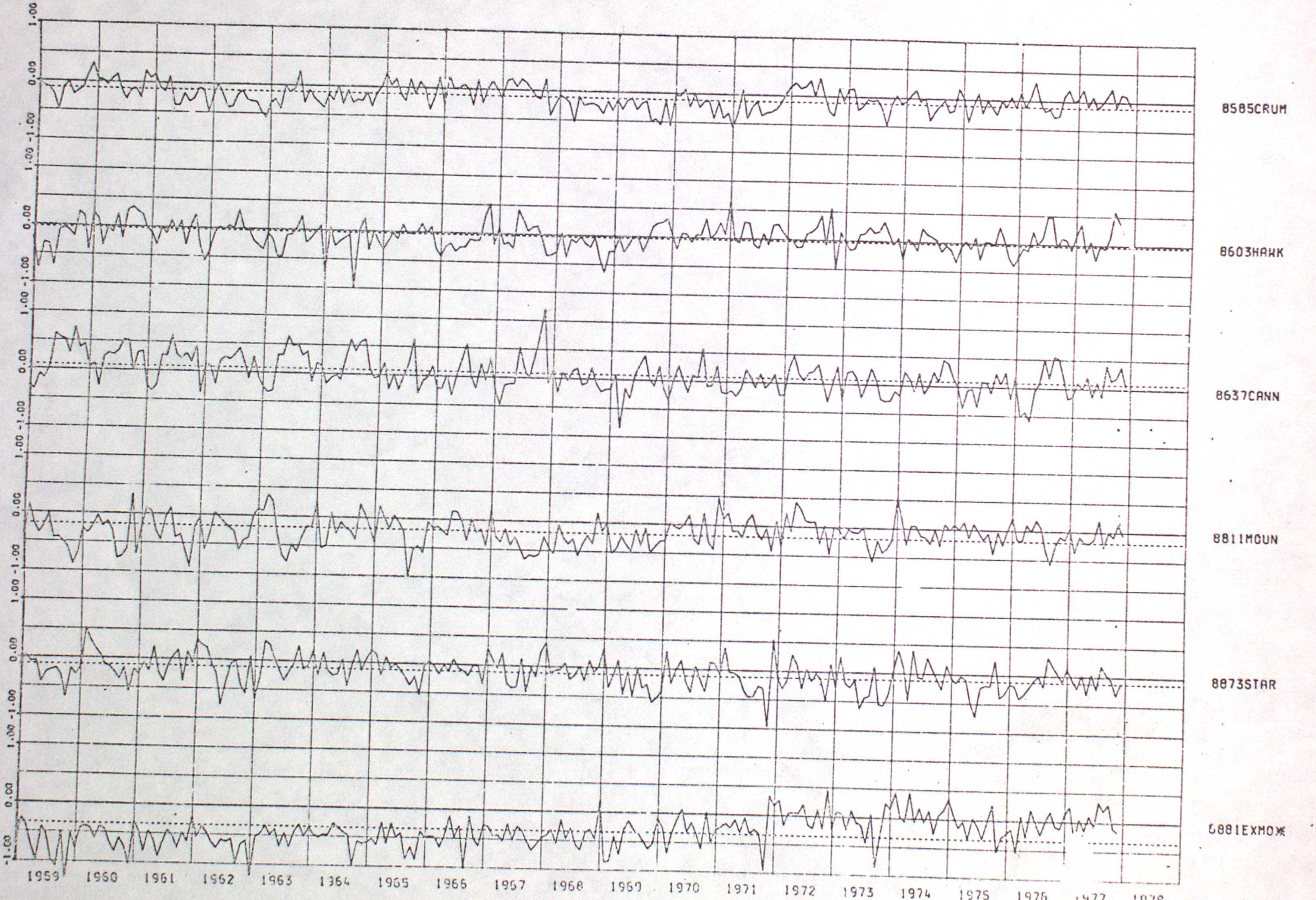
FIGURE 8(xxiv)



* STATION NOT INCLUDED IN FACTOR ANALYSIS.

RESIDUALS

FIGURE 8(xxv)



* STATION NOT INCLUDED IN FACTOR ANALYSIS.

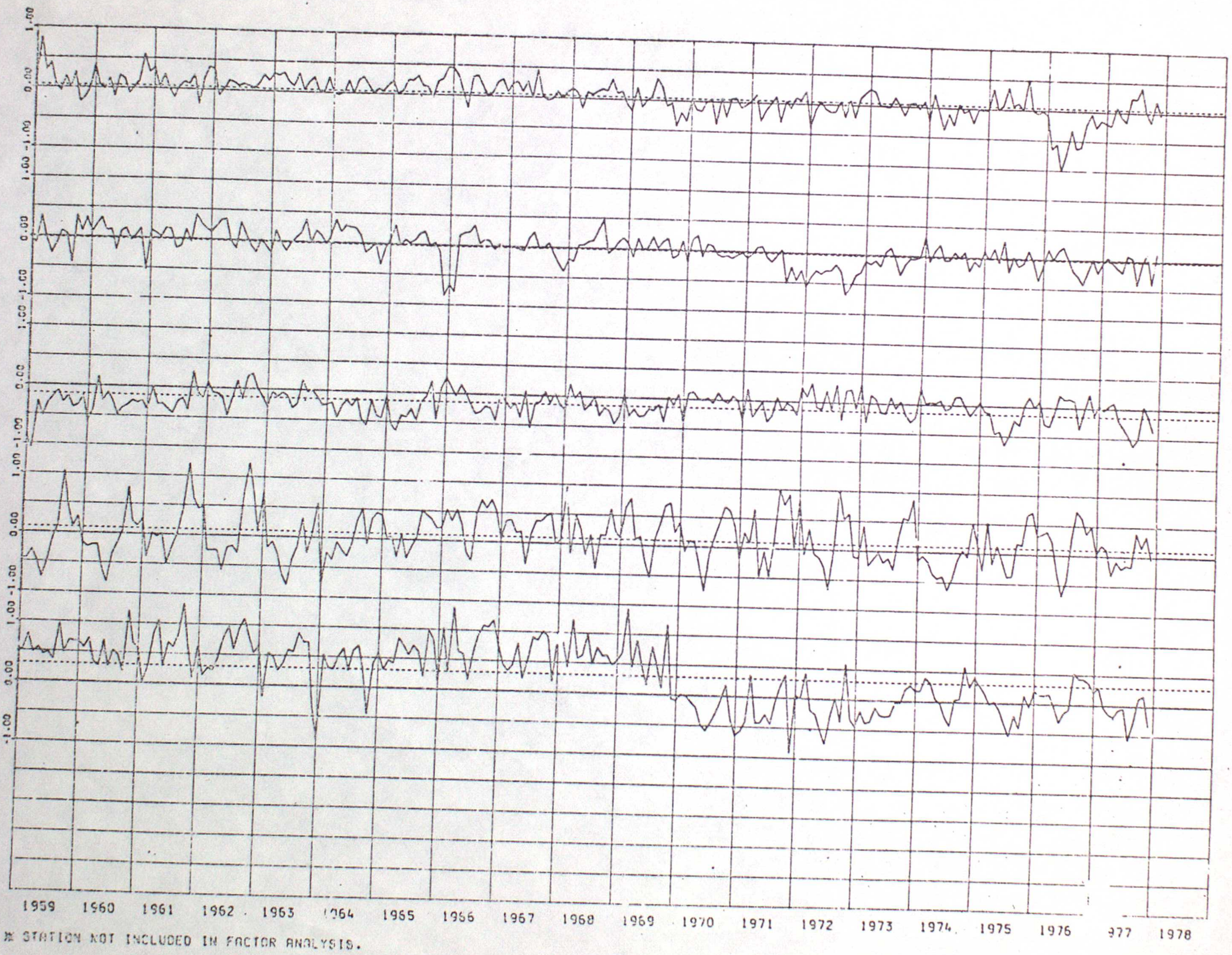
RESIDUALS

FIGURE 8(xxvi)

* STATION NOT INCLUDED IN FACTOR ANALYSIS.

RESIDUALS

FIGURE 8(xxvii)



* STATION NOT INCLUDED IN FACTOR ANALYSIS.

Table 1

Some stations with discontinuities in their residual series

DCNN	Station Name	Dates of Discontinuity		Notes
0514	KIRKWALL	↑ OCT 61		
0425	STORNOWAY	↓ SEP 61	↑ NOV 73	change of site 1/11/73
0482	FORTROSE	↓ NOV 70	↑ JUL 75	
0561	ONICH	↓ 1970		
1226	BALMORAL	↓ APR 66	↑ FEB 68	
1272	CRAIG	↓ JUN 71		
1588	ST ANDREWS	↑ JAN 60	↑ MAY 70	
1646	EDINBURGH BLACKFORD HILL	↓ JUN 71		change of screen and thermometer 1/6/79
1675	HADDINGTON	↓ JAN 72		
1789	ST ABBS HEAD	very variable throughout		
1838	WEST LINTON	↓ JUN 68	↑ NOV 69	
2165	DURHAM	↑ APR 69		
2295	SILPHO MOOR	↓ 59-64	↑ JUN 71	
3069	CROMER	↑ APR 69		
3253	CAMBRIDGE ROYAL BOTANICAL GDNS	↑ JUN 67	↓ 1973	change of site 4/5/67
4057	HARROGATE	↓ FEB 72		thermometer changed to centigrade 31/10/71
4174	MORLEY	69 becomes variable 75 steadies		thermometer changed Sep 65, May 72, Nov 73 and Sep 75
4237	WARSOP	↑ 68-69		
4341	CALDECOTT	↑ JAN 72		site change 21/12/71
4643	PENKRIDGE	↓ JUN 75		
5694	ALICE HOLT LODGE	↓ FEB 72		
5771	VENTNOR	↑ FEB 68		
5882	PORTON	↓ MAR 63	↑ MAR 75	thermometer to centigrade Mar 63, thermometer changed Feb and Mar 75
6024	COLONSAY	very variable throughout		
7748	PEN-Y-FRIDD	↓ 59-72		
8881	EXMOUTH	↑ 59-71	↑ DEC 71	
9970	JERSEY	↓ DEC 69		apparent site change 1970

KEY

↑ increase of residuals
↓ decrease of residuals

Table 2

Some stations with short period anomalies in their residual series

DCNN	Station Name	Period of Anomaly	Notes
1789	ST ABBS HEAD	MAY 71 - MAR 72	
1838	WEST LINTON	JUN 68 - OCT 69	
4341	CALDECOTT	SEP 70 - MAR 71	
4643	PENKRIDGE	APR 60 - DEC 60	
4967	CHELTENHAM	JAN 75 - MAR 76	Thermometer changed Dec 74, Mar 76, Feb 77. Inspector's report Sep 76 "..Screen in a poor state of repair, hinge broken so door falls off, inside unpainted and dirty.."
5798	RYDE	JAN 72 - DEC 74	
5863	LARKHILL	JUN 73 - SEP 73	
6764	GLENMORE LODGE	FEB 72 - AUG 74	
8078	LLET-Y-EVAN-HEN	JUN 59 - FEB 61	