Air pressure dependence of cosmic ray intensity

Methods and results of a statistical analysis on neutron monitor data

By STEN MARTINELLE, *Institute of Statistics and Cosmic Ray Group, Institute of Physics, Upsala University*

(Manuscript received November 24, 1966)

ABSTRACT

A general regression model is derived which describes the dependence of the cosmic-ray neutronic component on the air pressure. As the residuals are correlated the ordinary least-squares method is not the best one to estimate the pressure effect. Alternative methods are discussed. Five methods are used to estimate the pressure coefficient at Upsala during the years 1957–1964. It turns out that a modified two-station method gives the best result and also that there exists a solar cycle dependence of the pressure coefficient.

1. Introduction

The purpose of this paper is to point out some statistical methods, well known to the econometrician but perhaps not to the physicist, which may be of help when making regression analysis on time series. The classical regression analysis is based on the assumption that the residuals are randomly distributed. But as correlated residuals are the rule rather than the exception in time series analysis, the research worker should be cautious when applying classical least squares, for two reasons: First, the ordinary measure of precision is ruined. Second, estimation in the classical way is usually not the best way of extracting information from the data. However, more appropriate methods exist, which pay due regard to the time series aspect of the problem.

To make the exposition more concrete, we shall refer to the problem of estimating the pressure dependence of the cosmic ray intensity recorded by the neutron monitor. As almost every practical problem is specific we give some details only pertaining to this problem.

2. Statement of the problem

2.1. *Background*

Cosmic radiation consists of charged particles which enter the atmosphere of the earth from outer space. When passing through the atmosphere this radiation interacts with the air nuclei. Other types of particles are generated and an energy degradation takes place. As a result, the particles which reach the earth show other characteristics than those entering the outer atmosphere. (These latter make up what is known as primary radiation.)

The instruments used to observe the particles are as a rule fixed and located at ground level. Because the air mass above an instrument changes, particles have to pass through different masses at different times to reach the instrument. One effect is that more particles are recorded at low barometric pressure, the primary radiation being constant. On the other hand variations in primary radiation are concealed by changes in the air pressure. Consequently one wants to correct the recorded intensity for the pressure dependence.

The pressure dependence is intimately associated with the energy distribution (energy spectrum) of the primary radiation which varies with time. It is therefore of interest to study the influence of the barometric pressure on the recorded variation during relatively short periods. To this end regression analysis is used, but as the residuals are autocorrelated, the classical least squares method is not optimal and can be replaced by more efficient methods. Before we take up different methods of determining the pressure dependence, we shall derive

the regression model which applies to neutron monitor data. Confining ourselves to these data means that only the effect of the barometric pressure has to be considered.

2.2. *Derivation of the regression model*

Let N_t denote the number of particles counted and let ΔP_t be the deviation in the air pressure from the normal level at time t. Then the relation between the number of particles counted and the pressure is generally given by the formula

$$N_t = N_{ot} e^{-\alpha \Delta P_t} \tag{2.2.1}$$

where N_{0t} is the number of particles which would be counted if the recording instrument is located at the normal pressure level, while α is the pressure coefficient. We prefer, however, to use a more detailed formula, which allows for the observed fact that the pressure coefficient is not constant through time and spaec (e.g. Bachelet *et al.*, 1964, and Griffiths *et al.*, 1966).

Let $N_{ot}^{(E)}$ denote the number of particles at time t at the normal pressure level having an energy in the interval $(E, E + \Delta E)$. Then the influence of the pressure is given by

$$N_t = \sum_E c_E N_{ot}^{(E)} e^{-\alpha_E \Delta P_t} + \delta_t \tag{2.2.2}$$

where α_E is a constant, which expresses the influence of the pressure on the particles in the energy interval $(E, E + \Delta E)$, c_E are instrumental constants dependent on the size and location of the recording instrument and δ_t is the measurement error.

We introduce the symbols

$$K_t = \sum_E c_E N_{ot}^{(E)} \tag{2.2.3}$$

$$w_t^{(E)} = \frac{c_E N_{ot}^{(E)}}{\sum_E c_E N_{ot}^{(E)}} \tag{2.2.4}$$

$$\alpha_t = \frac{\sum_E c_E N_{ot}^{(E)} \alpha_E}{\sum_E c_E N_{ot}^{(E)}} = \sum_E w_t^{(E)} \alpha_E \tag{2.2.5}$$

and rewrite (2.2.2) using (2.2.4)

$$N_t - \delta_t = \sum_E c_E N_{ot}^{(E)} e^{-\alpha_t \Delta P_t} e^{(\alpha_t - \alpha_E)\Delta P_t}$$

$$= K_t e^{-\alpha_t \Delta P_t} \sum_E w_t^{(E)} e^{(\alpha_t - \alpha_E)\Delta P_t} \tag{2.2.6}$$

Take the logarithm of (2.2.6)

$$\log (N_t - \delta_t) = \log K_t - \alpha_t \Delta P_t + \log \left(\sum_E w_t^{(E)} e^{(\alpha_t - \alpha_E)\Delta P_t} \right) \tag{2.2.7}$$

By Taylor expansion of the last term it is seen that a first approximation of (2.2.7) is

$$\log (N_t - \delta_t) = \log K_t - \alpha_t \Delta P_t + \beta_t (\Delta P_t)^2 \tag{2.2.8}$$

where β_t is half the "variance" of the constants α_E, or formally:

$$\beta_t = \tfrac{1}{2} \sum_E w_t^{(E)} (\alpha_E - \alpha_t)^2 \tag{2.2.9}$$

To reduce the equation (2.2.8) to a form suitable for regression analysis, we rewrite $\log (N_t - \delta_t)$ and $\log K_t$. Since δ_t is small compared to N_t, then:

$$\log (N_t - \delta_t) = \log N_t \left(1 - \frac{\delta_t}{N_t}\right) \approx \log N_t - \frac{\delta_t}{N_t} \tag{2.2.10a}$$

Let $\bar{K}$ be the mean value of K_t during the period of analysis, and ΔK_t be the deviation of K_t from the mean at time t. Since ΔK_t is small in relation to $\bar{K}$ we get:

$$\log K_t = \log \bar{K} \left(1 + \frac{\Delta K_t}{\bar{K}}\right) \approx \log \bar{K} + \frac{\Delta K_t}{\bar{K}} \tag{2.2.10b}$$

Now we write (2.2.8) in a form adaptable to regression analysis.

$$\log N_t = \log \bar{K} - \alpha \Delta P_t + \beta (\Delta P_t)^2 + u_t \tag{2.2.11}$$

where

$$u_t = -(\alpha_t - \alpha)\Delta P_t + (\beta_t - \beta)(\Delta P_t)^2 + \frac{\Delta K_t}{\bar{K}} + \varepsilon_t \tag{2.2.12}$$

This u_t is called the residual and is composed of three parts. One part depends on α_t, β_t and ΔP_t and vanishes if α_t and β_t are constant during the period. The second part is proportional to the deviation of the cosmic ray intensity from its mean at the normal pressure level.

The third part is

$$\varepsilon_t = \frac{\delta_t}{N_t}$$

Usually, the quadratic term in (2.2.11) is overlooked. Since β_t is rather small compared to α_t this is not too serious. Moreover it is difficult to get a reliable estimate of β when using sample of moderate length. Instead of neglecting the quadratic term we prefer to use an approximate value of β in the following denoted by β_c.

So instead of (2.2.11–12) we consider:

$$\begin{cases} \log N_t - \beta_c(\Delta P_t)^2 = \log \bar{K} - \alpha \Delta P_t + u_t & (2.2.13) \\ u_t = -(\alpha_t - \alpha)\Delta P_t + (\beta_t - \beta_c)(\Delta P_t)^2 \\ \qquad + \Delta K_t/\bar{K} + \varepsilon_t & (2.2.14) \end{cases}$$

2.3. *Short time variation in the pressure coefficient*

Now, what is the meaning of the coefficient α in (2.2.13)? If α_t is constant during the period of analysis we think of α as this value. But as α_t fluctuates the meaning of α is not so obvious. We can look upon α as a mean value of the α_t-values. To better visualize this let us consider the ordinary least-squares method.

Suppose we have made observations on N_t and $\Delta P_t (t = 1, \ldots n)$, letting $\overline{\Delta P}$ denote the mean value of ΔP_t; then we may write the least squares estimator as follows, neglecting the term $(\beta_t - \beta_c)\Delta P_t^2$ in (2.2.14):

$$\hat{\alpha} = -\frac{\sum(\log \bar{K} - \alpha_t \Delta P_t + \Delta K_t/\bar{K} + \varepsilon_t)(\Delta P_t - \overline{\Delta P})}{\sum(\Delta P_t - \overline{\Delta P})^2}$$

$$= \frac{\sum \alpha_t \Delta P_t (\Delta P_t - \overline{\Delta P})}{\sum(\Delta P_t - \overline{\Delta P})^2} - \frac{\sum(\Delta K_t/\bar{K} + \varepsilon_t)(\Delta P_t - \overline{\Delta P})}{\sum(\Delta P_t - \overline{\Delta P})^2}$$

$$= \frac{\sum(\Delta P_t - \overline{\Delta P})^2 \alpha_t}{\sum(\Delta P_t - \overline{\Delta P})^2} + \overline{\Delta P} \cdot \frac{\sum(\Delta P_t - \overline{\Delta P})\alpha_t}{\sum(\Delta P_t - \overline{\Delta P})^2} - \frac{\sum(\Delta K_t/\bar{K} + \varepsilon_t)(\Delta P_t - \overline{\Delta P})}{\sum(\Delta P_t - \overline{\Delta P})^2} \quad (2.3.1)$$

If we think of the cosmic ray intensity and the measurement error as stochastic variables, we may talk about the expected value of $\hat{\alpha}$. As the intensity at the normal pressure level can be supposed to vary without any relation to the air pressure at the ground level it is natural to introduce the following restriction on this variable:

$$E(\Delta K_t \mid \Delta P_t) = 0 \quad (2.3.2)$$

This means that the conditional expected value of ΔK_t, given ΔP_t, is zero. It is assumed that the measurement error has the same property. These assumptions imply that the expected value of the last term in (2.3.1.) equals zero.

Hence

$$E(\hat{\alpha} \mid \Delta P_1, \cdots, \Delta P_n) = \frac{\sum(\Delta P_t - \overline{\Delta P})^2 \alpha_t}{\sum(\Delta P_t - \overline{\Delta P})^2} + \overline{\Delta P} \cdot b_{\alpha_t \mid \Delta P_t} \quad (2.3.3)$$

where $b_{\alpha_t \mid \Delta P_t}$ is the regression coefficient of α_t on ΔP_t. There is no physical significance to the coefficient $b_{\alpha_t \mid \Delta P_t}$, and no other reason to expect a co-variation of α_t and ΔP_t; the last term in (2.3.3) is therefore negligible.

Thus the meaning of α in (2.2.13) should be clear. We can think of it as a weighted mean of the α_t-values. It is true that the weighting depends on the estimation technique. But as α_t and ΔP_t are unrelated, we can equally well use some other estimation method having a different weighting system. There should not be any systematic difference.

The importance of the first two terms of the residual (2.2.14) should not be overemphasized. The dominating part of the residual is the cosmic ray intensity and the measurement error. The cosmic ray intensity does not vary in a simple random way. Instead it shows a high autocorrelation and necessitates a more refined estimation technique than the usual least squares method.

2.4. *Limitations of the study*

It is empirically verified that the pressure coefficient depends on the geomagnetic latitude of the place where the observation is made. This phenomenon is due to differences in the energy distributions at different latitudes. Other geographical effects have also been observed, which can be explained in the same way. Hence the pressure coefficient is assumed to vary in time and space. In this investigation we shall not consider the space variation.

There are in principle two ways of studying

the pressure dependence of the cosmic ray intensity.

One way is to measure the intensity at two or more levels of altitude. Such measurements have been made using mobile instruments (e.g. Bachelet *et al.*, 1964). These are, however, studies covering only short periods.

The other way is to estimate the pressure coefficient on time series data. We will confine ourselves to this approach.

Several investigations based on time series data are reported in the literature on this subject. Some references will be given later on. Different estimation techniques have been applied, but there seems to be some uncertainty about the relative merits of the different techniques.

The estimation problem can be split up into two parts. One problem is to decide on an "optimal" estimation method with the sample size given. The other problem is to decide on the appropriate sample size. To some extent these problems are interrelated. Primarily the former will be considered.

To conclude this section, we shall substitute for the foregoing symbols those commonly used in statistical texts as follows:

$y_t = \log N_t - \beta_c(\Delta P_t)^2$ (dependent variable)[1]

$\beta_0 = \log \overline{K}_0$

$\beta_1 = -\alpha$

$x_t = \Delta P_t$ (independent variable).

Then (2.2.13) is altered to

$$y_t = \beta_0 + \beta_1 x_t + u_t \qquad (2.4.1\,a)$$

The residual will be treated as a stochastic variable with the expected value zero for each x_t. Hence

$$E(y_t \mid x_t) = \beta_0 + \beta_1 x_t \qquad (2.4.1\,b)$$

[1] We put β_c equal to $2.5.10^{-6}$ (the air pressure being measured in mb). This value is derived from experimental results on the altitude dependence of the pressure coefficient obtained by Bachelet *et al.*, 1965. This value is in good agreement with an estimate of ours. By using a method, in this paper called the two-station method, and taking the average over several periods, we found β equal to $2.2.10^{-6}$, with a standard deviation of $0.8 \cdot 10^{-6}$.

The device of using side conditions on the parameters is in econometrics known as conditional regression analysis (Wold, 1953, p. 47).

This regression model is a special case of the general linear model

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_p x_{pt} + u_t \qquad (2.4.2)$$

which will be considered too.

The variables are assumed to be observed at equidistant time points and numbered 1, 2,..., n, where n is the sample size.

3. Methods of estimation and attempts of improvement

3.1. *Ordinary least squares*

Let us consider the residual u_t in (2.2.12). We assume that the terms containing ΔP_t are negligible.

$$u_t = \frac{\Delta K_t}{\overline{K}} + \frac{\delta_t}{N_t} \qquad (3.1.1)$$

If the intensity is constant during the period, the first term equals zero by definition. In that case the variance of the residual is:

$$\sigma^2_{u_t} = \frac{1}{N_t^2}\sigma^2_{\delta_t} \qquad (3.1.2)$$

The variance of the measurement error is proportional to the number of particles counted (e.g. Dyring, 1962). Hence:

$$\sigma^2_{u_t} = \frac{k}{N_t} \qquad (3.1.3)$$

The measurement errors are normally distributed and independent (e.g. Dyring, 1962). The minimum-variance estimates of the regression coefficients, β_0 and β_1, in (2.4.1) are obtained by minimizing the weighted sum of squares:

$$Q^2_w(\hat{\beta}_0, \hat{\beta}_1) = \sum_{t=1}^{n} N_t(y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t)^2 \qquad (3.1.4)$$

As the relative variation in N_t is small, the residual variance can be treated as a constant, and a good approximation is obtained by minimizing the unweighted sum of squares:

$$Q^2(\hat{\beta}_0, \hat{\beta}_1) = \sum_{t=1}^{n} (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t)^2 \qquad (3.1.5)$$

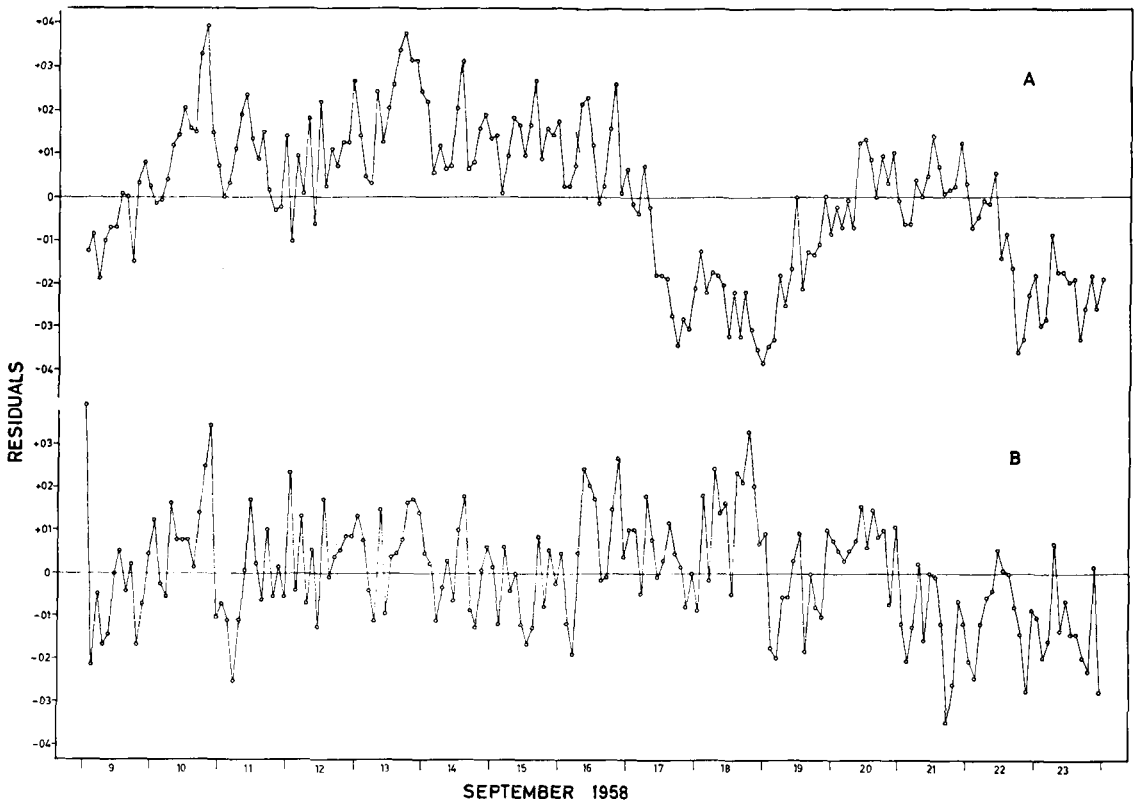

Fig. 1. Residuals from ordinary least squares regression analysis on bihourly values from: (A) Upsala, (B) Upsala and Murchison Bay.

If $\bar{N}$ denotes the mean value of N_t, and s_x^2 the variance of the air pressure, we have:

$$\sigma^2_{\hat{\beta}_1} \approx \frac{1}{n} \cdot \frac{k}{\bar{N} s_x^2} \qquad (3.1.6)$$

(3.1.6) is the smallest possible variance and is attained only when the primary cosmic ray intensity is constant. However, in general the intensity fluctuates, and the first term in (3.1.1) cannot be neglected. It is not realistic to assume that the intensity varies quite at random just as the measurement error. This means that the conditions leading to the minimumvariance property of the ordinary least squares estimates are no longer fulfilled.

Fig. 1 A is a plot of the residuals from a regression analysis on bihourly values from Upsala. The ordinary least squares method has been applied, and the residuals obviously display a non-random pattern.

In order to improve the estimation technique it is necessary to make some realistic assumptions about the variation of the cosmic ray intensity. Such assumptions may be either physical or statistical. For the time being there is no workable physical model that explains the cosmic ray intensity variation by using other observable variables. It is possible, however, to reduce the variation in the cosmic ray intensity by using observations from two stations at about the same geomagnetic latitude.

Another approach would be to describe the cosmic radiation as a realization of a stochastic process. In this way we can make use of the statistical regularities in the variation of the cosmic radiation.

3.2. *The two-station method—a specific approach*

The two-station method is a specific solution pertaining to our problem. It makes use of the physical fact that the cosmic ray intensity is almost the same on one and the same geomagnetic latitude. Due to differences in size of the

instruments the relative numbers of particles counted, divided by the mean number of particles, and not the absolute numbers are comparable.

We denote two stations with 1 and 2 respectively. From (2.4.1) and (3.1.1) it follows that:

$$\left.\begin{aligned} y_{1t} - y_{2t} &= \beta_{10} - \beta_{20} + \beta_{11} x_{1t} - \beta_{21} x_{2t} + u_t \\ u_t &= \frac{\Delta K_{1t}}{\bar{K}_1} - \frac{\Delta K_{2t}}{\bar{K}_2} + \frac{\delta_{1t}}{N_{1t}} - \frac{\delta_{2t}}{N_{2t}} \end{aligned}\right\} \quad (3.2.1)$$

If the relative number of counts are equal at the normal pressure level, we get:

$$\frac{\Delta K_{1t}}{\bar{K}_1} - \frac{\Delta K_{2t}}{\bar{K}_2} = 0 \quad (3.2.2)$$

and the residual consists of measurement errors only. The ordinary least squares method would then be appropriate for estimating $\beta_{11}$and β_{21}. It should be observed, however, that no assumption of equality between the coefficients is necessary.

A modification of this method is to use an approximate value of β_{21}, say β_{21}^*, if we want to estimate β_{11}.

We put

$$y_{2t}^* = y_{2t} - \beta_{21}^* x_{2t}, \qquad \beta_0^* = \beta_{10} - \beta_{20} \quad (3.2.3)$$

and rewrite (3.2.1) using the assumption (3.2.2):

$$\left.\begin{aligned} y_{1t} - y_{2t}^* &= \beta_0^* + \beta_{11} x_{1t} + u_t \\ u_t &= (\beta_{21}^* - \beta_{21}) x_{2t} + \frac{\delta_{1t}}{N_{1t}} - \frac{\delta_{2t}}{N_{2t}} \end{aligned}\right\} \quad (3.2.4)$$

The residual is freed of the cosmic ray intensity term, but instead it has acquired a fraction of the air pressure at the second station.

There are two reasons for this modified estimation technique: First, the official data is generally corrected for the influence of the air-pressure, and therefore, original data is not always supplied. Second, a co-variation between x_{1t} and x_{2t} enlarges the variance of the estimates obtained when fitting the model (3.2.1). This is the problem of multicollinearity frequently met with when analysing non-experimental data (Wold, 1953, p. 46). Whatever the reason, one should be aware of the consequences of using this method when β_{21}^* differs from β_{21}, while x_{1t} and x_{2t} are correlated. If we denote the regression coefficient of x_{2t} on x_{1t} by $b_{x_{2t} \mid x_{1t}}$, the expected value of the least squares estimator $\hat{\beta}_{11}$ is:

$$E(\hat{\beta}_{11}) = \beta_{11} + (\beta_{21}^* - \beta_{21}) \cdot b_{x_{2t} \mid x_{1t}} \quad (3.2.5)$$

The result is obtained in the same way as (2.3.3), the pressure variables being treated as non-stochastic variables. Hence the estimate is biased if x_{1t} and x_{2t} are linearly related and β_{21}^* does not equal β_{21}.

On the other hand, if x_{1t} and x_{2t} are uncorrelated, i.e.

$$\sum_{t=1}^{n} (x_{1t} - \bar{x})(x_{2t} - \bar{x}_2) = 0 \quad (3.2.6)$$

it can be shown that the least-squares estimates of β_{11}, obtained from (3.2.1) and (3.2.4), are identical, while the modified method will give a somewhat larger estimate of the variance of $\hat{\beta}_{11}$, because of the presence of x_{2t} in the residual.

Even if x_{1t} and x_{2t} are unrelated one cannot assume (3.2.6) to be true. In general the two estimates differ. When the pressure variables are unrelated, it is natural to look upon x_{2t} as a random variable correlated in time but independent of x_{1t}. If β_{21}^* does not equal β_{21}, the residuals in the model (3.2.4) are correlated, and they have a larger variance than the residuals in (3.2.1). Under these conditions the estimates obtained by the modified method are not biased but exhibit a larger variance. Consequently, if possible, the first method based on original data should be used.

Up to this point we have assumed the condition (3.2.2). Due to differences in instruments and location this is not generally true. Moreover for every station there is no corresponding "twin-station" at the same geomagnetic latitude Hence the variation in the residuals due to the cosmic ray intensity would be only partly removed when using the two aforementioned methods. Therefore, we have to assume correlated residuals even in this case. Especially when the analysis is based on pressure corrected data should this assumption be made, as the variable x_{2t} is probably present in the residual.

The idea of using two stations to improve the estimates goes back to Mathews (1958), McCracken & Johns (1959) and Lindgren (1961). But none of them considered the simple approach of fitting model (3.2.1). Lindgren used

this method to analyse the disturbed period between 1957–59 and found it to be a considerable improvement.

Fig. 1 B shows the residuals from a regression analysis based on the model (3.2.1). The data on which Fig. 1 A is based has been employed. The second station is Murchison Bay. We notice that the residuals are not so strongly correlated as in Fig. 1 A; yet one cannot say that they are random.

3.3. *Correlated residuals*

3.3.1. *The correlogram.* The minimum-variance property of the least-squares estimators is based on the assumption of uncorrelated residuals.

To study the correlation structure of the residuals the autocorrelation coefficients of the observed residuals (v_t) are computed. These are defined by:

$$r_k = \frac{\Sigma v_t\, v_{t+k}/(n-k)}{\Sigma v_t^2/n} \quad (k = 1, 2, \cdots, m) \qquad (3.3.1)$$

A plot of the coefficients r_k is called a correlogram. If the residuals (u_t) are uncorrelated and the regression analysis is based on a large sample, the coefficients r_k fluctuate around zero in a random way. In small samples, however, the observed residuals may be autocorrelated. This is true when the independent variable is autocorrelated and not wholly removed from the residuals. Durbin & Watson (1950) have studied the problem of testing for correlation in the residuals when the sample size is small. They have derived the distribution of a general kind of quadratic form which includes the coefficients r_k. The distributions are rather complicated for small n's but for large n's, r_k has an approximate normal distribution, with:

$$E(r_k) \approx 0, \quad \sigma_{r_k}^2 \approx \frac{1}{n-k} \quad (k = 1, 2, \cdots, m)$$
$$E(r_k r_e) \approx 0 \quad (k \neq e) \qquad (3.3.2)$$

Consequently we may use the two curves:

$$\pm 2(n-k)^{-\frac{1}{2}} \qquad (3.3.3)$$

as control limits. If the residuals are uncorrelated we expect only 5 % of the r_k's to fall outside these limits.

Fig. 2 shows two correlograms computed from the residuals in Fig. 1 A and Fig. 1 B. The control limits are also drawn (broken line). We see that the two correlograms have values outside these curves indicating correlated residuals. Moreover, both correlograms show up a systematic pattern which we would not expect if the residuals were uncorrelated.

3.3.2. *The autoregressive model.* Thus the assumption of uncorrelated residuals cannot be maintained. A more appropriate model of the residuals should be used. A general approach would be to specify the residuals, $u_t\,(t = 1, 2, \ldots, n)$ as random variables with expectation zero and a covariance matrix Σ, where the i, jth element is $\sigma_{ij} = E(u_i u_j)$. However, the practical value of such a general model is small. Usually the restriction of stationarity is introduced:

$$\sigma_{ij} = \sigma_{|i-j|} \qquad (3.3.4)$$

This condition implies that the covariance between any two residuals depends only on the distance in time ($|i-j|$) between them. Especially all the residuals have the same variance. The autocorrelation coefficients are then defined by:

$$\varrho_k = \frac{\sigma_k}{\sigma_0} \quad (k = 0, 1, 2, \cdots, n) \qquad (3.3.5)$$

If the matrix Σ is known except for a constant factor (with or without the restriction (3.3.4)), then under very general conditions the problem of finding the B.L.U. (best linear unbiased) estimators of the regression parameters has a unique solution (Aitken, 1934). By means of a linear transformation of the regression model new uncorrelated residuals are obtained. But this transformation is based on the inversion of the matrix Σ of order $n \cdot n$, and therefore in practical situations some approximate method has to be used. Moreover, it is not realistic to assume that the matrix Σ (or a matrix proportional to Σ) is known. Generally the correlation structure of the residuals has to be estimated before a transformation is made.

As a useful model of the residuals we use the linear stochastic difference equation:

$$u_t + \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + \ldots + \alpha_h u_{t-h} = \varepsilon_t \qquad (3.3.6)$$

where $\alpha_i\,(i = 1, 2, \ldots h)$ are constants and $\varepsilon_t\,(t = 1, 2, \ldots, n)$ are random variables with the following properties:

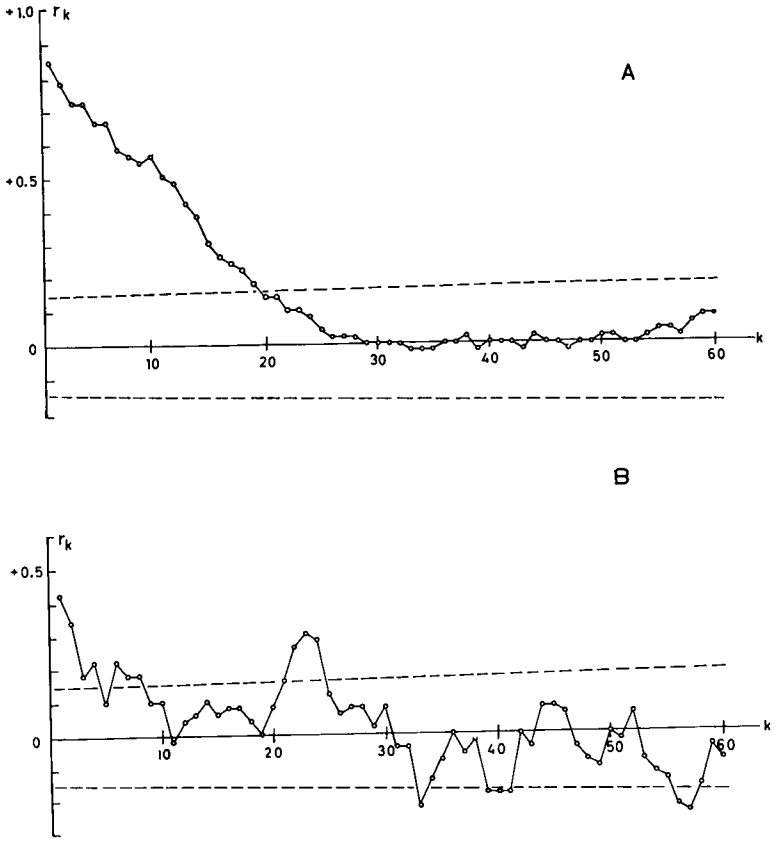

Fig. 2. Correlograms computed from the residuals in Fig. 1.

$$E(\varepsilon_t) = 0, \quad E(\varepsilon_t^2) = \sigma^2, \quad E(\varepsilon_t \varepsilon_{t+k}) = 0 \quad (k \neq 0) \tag{3.3.7}$$

The equation (3.3.6) can be solved for u_t, if all of the roots of the polynomial

$$z^h + \alpha_1 z^{h-1} + \cdots + \alpha_h = 0 \tag{3.3.8}$$

are inside the unit circle. (See e.g. Bartlett, 1962, Chapter 5.2.) This solution is stationary in the sense of (3.3.4), if it is assumed that the process started a long time ago. We make this assumption. Then the elements of the matrix Σ satisfy the equation:

$$\sigma_0 + \alpha_1 \sigma_1 + \ldots + \alpha_h \sigma_h = \sigma^2 \tag{3.3.9}$$

as well as the difference equation:

$$\sigma_i + \alpha_1 \sigma_{i-1} + \ldots + \alpha_h \sigma_{i-h} = 0 \quad (i > 0) \tag{3.3.10}$$

(Multiply (3.3.6) by $u_{t-i} (i \geq 0)$ and take expectations.)

Now, it is possible to express the inverse of the matrix Σ in the constants $\alpha_i (i = 1, \ldots, h)$ and σ^2 (Wise, 1955). Then we can write down the transformed sum of squares obtained by the linear transformation mentioned above. In the case of one independent variable and $h = 1$ we get:

$$Q^2 = \sum_{t=2}^{n} [y_t + \alpha_1 y_{t-1} - \beta_0(1 + \alpha_1) - \beta_1(x_t + \alpha_1 x_{t-1})]^2 + (1 - \alpha_1^2)(y_1 - \beta_0 - \beta_1 x_1)^2 \tag{3.3.11}$$

(Cochrane & Orcutt, 1949).

If the constants $\alpha_i (i = 1, \ldots, h)$ are known, the problem of estimating the regression coefficients is simple, because then (3.3.11) is quadratic in the unknown parameters. (Of course the same holds true for the general case, when there are

several independent variables and $h > 1$.) However, the interesting case arises when these constants are unknown and have to be estimated. Then the sum of squares (3.3.11) is no longer so easy to minimize. Q^2 is not quadratic in the unknown parameters. In other words, we have a non-linear regression model. If we neglect the last term in (3.3.11), we see that the estimating problem becomes equivalent to that of estimating the parameters β_0, β_1 and α_1 by least squares in the model:

$$y_t + \alpha_1 y_{t-1} = \beta_0(1 + \alpha_1) + \beta_1(x_t + \alpha_1 x_{t-1}) + \varepsilon_t$$

$$(t = 2, 3, \ldots, n) \qquad (3.3.12)$$

The model is non-linear because of the terms $\beta_0 \alpha_1$ and $\beta_1 \alpha_1$. The model (3.3.12), or its general form:

$$y_t + \sum_{i=1}^{h} \alpha_i y_{t-i} = \beta_0(1 + \sum_{i=1}^{h} \alpha_i) + \sum_{j=1}^{p} \beta_j(x_{jt} + \Sigma \alpha_i x_{jt-i}) + \varepsilon_t \quad (t = h+1, \cdots, n), \qquad (3.3.13)$$

can be obtained in a direct way. Let the regression model be

$$y_t = \beta_0 + \beta_1 x_{1t} + \ldots + \beta_p x_{pt} + u_t \quad (t = 1, 2, \ldots, n), \qquad (3.3.14)$$

where u_t is generated by (3.3.6). Multiply y_{t-i} in (3.3.14) by $\alpha_i (i = 1, \ldots, h)$ and add to y_t. It is convenient to use the model (3.3.13). Making an estimation of this model by least squares means using some approximations, but estimates obtained in this way should not differ too much from those obtained by minimizing the transformed sum of squares corresponding to the inverse of the matrix Σ. In large samples the difference is negligible.

A sequence of u_t-values generated by the difference equation (3.3.6) is called an autoregressive series. To gain an idea of this kind of series see Wold (1965). Briefly the series can be described as "periodic" but with continually shifting phase and amplitude. If a strict periodic component is contained in the residuals, one should include this part in the regression equation. The daily variation in the cosmic ray intensity is such a periodic component, at least for shorter periods. Usually, the first harmonic is enough to describe the daily variation. If we use bi-hourly values we can write, instead of (2.4.1*b*):

$$E(y_t \mid x_t) = \beta_0 + \beta_1 x_t + \beta_2 \sin \frac{2\pi t}{12} + \beta_3 \cos \frac{2\pi t}{12} \qquad (3.3.15)$$

3.3.3. *Iterative least squares.* In this section we will consider the problem of estimating the nonlinear model (3.3.13) by least squares.

We want to minimize the residual sum of squares:

$$Q(\alpha, \beta) = \sum_{t=h+1}^{n} \left[y_t + \sum_{i=1}^{h} \alpha_i y_{t-i} - \beta_0 \left(1 + \sum_{i=1}^{h} \alpha_i\right) - \sum_{j=1}^{p} \beta_j \left(x_{jt} + \sum_{i=1}^{h} \alpha_j x_{jt-i}\right)\right]^2 \qquad (3.3.16)$$

where $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_h\}$ and $\boldsymbol{\beta} = \{\beta_0, \ldots, \beta_p\}$.

As $Q(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is not quadratic in the unknown parameters we cannot proceed in the usual way. Some iterative procedure must be applied. We shall use a method suggested by Cochrane & Orcutt (1949). This method is based on a specific feature of the function $Q(\boldsymbol{\alpha}, \boldsymbol{\beta})$. If $\boldsymbol{\alpha}$ is fixed, $Q(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is quadratic in $\boldsymbol{\beta}$, and if $\boldsymbol{\beta}$ is fixed, $Q(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is quadratic in $\boldsymbol{\alpha}$. Hence the following iterative procedure can be used.

Step 1:a) Set $\boldsymbol{\alpha} = 0$ and determine $\boldsymbol{\beta}$ for minimum $Q(0, \boldsymbol{\beta})$. We denote this $\boldsymbol{\beta}$-vector by $\boldsymbol{\beta}^{(1)}$.

b) Determine $\boldsymbol{\alpha}$ for minimum $Q(\boldsymbol{\alpha}, \boldsymbol{\beta}^{(1)})$. This $\boldsymbol{\alpha}$-vector is denoted by $\boldsymbol{\alpha}^{(1)}$.

Step 2: The same as step 1, but $\boldsymbol{\alpha}^{(1)}$ is used to compute $\boldsymbol{\beta}^{(2)}$ which gives a new value of $\boldsymbol{\alpha}$ denoted by $\boldsymbol{\alpha}^{(2)}$.

After s steps we get $\boldsymbol{\alpha}^{(s)}$ and $\boldsymbol{\beta}^{(s)}$. Obviously:

$$0 \leqslant Q(\boldsymbol{\alpha}^{(s)}, \boldsymbol{\beta}^{(s)}) \leqslant Q(\boldsymbol{\alpha}^{(s-1)}, \boldsymbol{\beta}^{(s)}) \leqslant Q(\boldsymbol{\alpha}^{(s-1)}, \boldsymbol{\beta}^{(s-1)}) \qquad (3.3.17)$$

Hence the sequence $Q(\boldsymbol{\alpha}^{(s)}, \boldsymbol{\beta}^{(s)})$ $(s = 1, 2, \ldots)$ converges.

An attractive feature of this procedure is that the simple least squares technique can be used, twice in each step. The determination of $\boldsymbol{\beta}^{(s)}$ from $\boldsymbol{\alpha}^{(s-1)}$ is nothing more than a fitting of the equation

$$y'_t = \beta_0 x'_{0t} + \beta_1 x'_{1t} + \cdots + \beta_p x'_{pt} \quad (t = h+1, \cdots, n) \qquad (3.3.18)$$

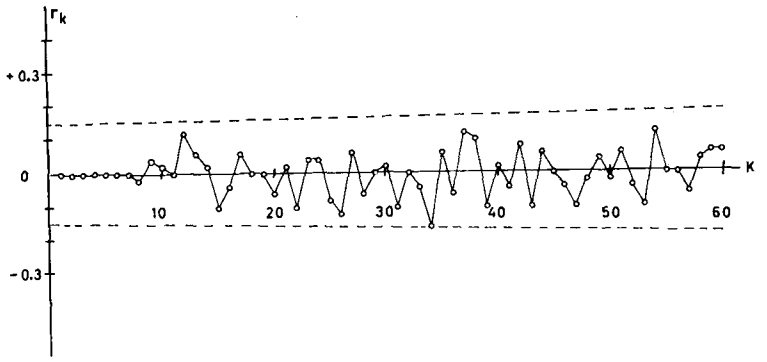

Fig. 3. Correlograms computed from residuals obtained by iterative least squares on the same data as in Fig. 1A.

where:

$$y'_t = y_t + \sum_{i=1}^{h} \alpha_i^{(s-1)} y_{t-i}, \quad x'_{0t} = 1 + \sum_{i=1}^{h} \alpha_i^{(s-1)},$$

$$x'_{jt} = x_{jt} + \sum_{i=1}^{h} \alpha_i^{(s-1)} x_{jt-i} \quad (j = 1, \cdots, p)$$

This can be done if the variables x'_{jt} $(j = 0, 1., p)$ are not linearly dependent. In the same way $\boldsymbol{\alpha}^{(s)}$ is determined by $\boldsymbol{\beta}^{(s)}$ by a fitting of the equation:

$$y'_t = \alpha_1 z'_{1t} + \cdots + \alpha_h z'_{ht} \qquad (3.3.19)$$

where

$$y'_t = y_t - \beta_0^{(s)} - \sum_{j=1}^{p} \beta_j^{(s)} x_{jt} \quad \text{and} \quad z'_{it} = -y_{t-1} + \beta_0^{(s)} + \sum_{j=1}^{p} \beta_j^{(s)} x_{jt-i} \quad (i = 1, \cdots, h)$$

This is possible if the variables z'_{it} $(i = 1, ..., h)$ are not linearly dependent. Hence each iteration consists of two regression analyses. This technique is related to the NILES-procedures studied by Wold (1966).

Different types of convergence criteria can be used to determine the number of iterations. A simple one, which we have used, is to stop when the differences $|\alpha_i^{(s)} - \alpha_i^{(s-1)}| (i = 1, ..., h)$ and $|\beta_j^{(s)} - \beta_j^{(s-1)}| (j = 0, ..., p)$ are smaller than some prescribed quantities.

On condition that the function $Q(\boldsymbol{\alpha}, \boldsymbol{\beta})$ has a single, unique minimum, we shall find this minimum and the corresponding values of the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ by the technique above. However, multiple minima may exist, and in such cases the estimates of the parameters depend on the initial values used in the first iteration cycle. It is well known that local minima may exist (Klein, 1953, Chapter III.2), but it is probably a complication pertaining to rather small sample sizes. It is possible to get an indication of the existence of multiple minima by using different starting values in the first iteration cycle. This was tried on a small part of our data, but the same estimates were obtained. The assumption that multiple minima are unlikely in practice is supported by an econometric study by Hildreth & Lu (1960). Asymptotically, for large n's there is only one minimum. This is a consequence of the consistency of the estimates obtained by the authors for the case of an autoregressive series of order one. But the proof can easily be generalized to an autogregressive series of arbitrary order.

Mere consistency of the estimates does not motivate iterative least squares, as even the ordinary least squares method gives consistent estimates (see section 3.3.5). But the estimates are also asymptotically efficient (Durbin, 1960) a property which the ordinary least squares estimates possess only under very special conditions. The distributional properties of the estimates based on small samples have not yet been investigated. Therefore the situation involved in the application of the asymptotic results is not clear and caution is recommended.

The asymptotical variances of the estimates of the parameters β_i $(i = 0, 1, ...p)$ derived by Durbin (1960) are the same as would have been given by least squares if $\boldsymbol{\alpha}$ were known. Hence

in the case of $p = 1$ and a large sample, we can estimate the variance of $\hat{\beta}_1$ as follows:

$$\sigma^2_{\hat{\beta}_1} \approx \frac{\hat{\sigma}^2}{\sum_{t=h+1}^{n} (x'_t - \bar{x}')^2} \qquad (3.3.20)$$

where $\hat{\sigma}^2 = (n-2h-2)^{-1} Q(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ and

$$x'_t = x_t + \sum_{i=1}^{h} \hat{\alpha}_i x_{t-i} \quad (t = h+1, \cdots, n)$$

As $h+2$ parameters are estimated, we have divided the residual sum of squares by $n-h-(h+2)$. However, using the number of observations instead would make no difference in large samples.

3.3.4. *Variance of the ordinary least squares estimates when the residuals are autocorrelated.* Until now we have been concerned with improving upon the estimation technique. In the next section we shall touch upon two other methods of estimation. But before we go on, we shall refer to two basic works that examine the implication of autocorrelated residuals, i.e. the papers by Wold (1950) and Grenander (1954).

Both authors consider the asymptotic variances of the ordinary least squares estimators of the parameters in the regression model (3.3.14), with stationary residuals generated by (3.3.6). In both papers the structure of the independent variables is shown to be significant. As the model (2.4.1), with only one independent variable, is of special interest to us, we shall restrict ourselves to the results pertaining to this case.

Assuming the following limits exist

$$\begin{aligned} \bar{x} &= \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} x_t, \\ c_k &= \lim_{n\to\infty} \frac{1}{n-k} \sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) \end{aligned} \qquad (3.3.21)$$

and that the moments c_k are negligible for $k > N$, then the asymptotic variance of $\hat{\beta}_1$ is, according to Wold:

$$\sigma^2_{\hat{\beta}} = \frac{\sigma^2_\varepsilon}{n c_0}(1 + 2r_1\varrho_1 + 2r_2\varrho_2 + \cdots), \qquad (3.3.22)$$

where $r_k = c_k/c_0$ and ϱ_k is defined by (3.3.5) and (3.3.9–10). Hence the variance in the case of uncorrelated residuals is multiplied by a factor containing the autocorrelations in the dependent variable as well as in the residuals. Ordinarily this factor is larger than unity, and consequently the usual variance estimate is too low. Perhaps the most interesting thing about the formula (3.3.22) is, that the autocorrelation in the independent variable is of the same significance as the autocorrelation in the residuals. Obviously this is relevant to our problem, the air pressure being a strongly autocorrelated variable.

Wold's formula applies to the case of a "stationary" independent variable taken in the sense of (3.3.21). Grenander allows, in addition, other types of independent variables, such as: $x_t = t$ and $x_t = \sin(2\pi t/T)$. He compares the ordinary least squares estimators to those obtained by the optimal method, based on a knowledge of the autocorrelations in the residuals. He shows that the ordinary least squares estimators are asymptotically efficient when the independent variables are of the type just mentioned. Regression equations containing polynomials in t or trigonometric functions are thus efficiently estimated by the ordinary least squares method.

The deviation in the air pressure from the normal level, probably well described as a stationary variable taken in the sense of (3.3.21) with $\bar{x} = 0$, has not the same structure of variation as the trend ($x_t = t$) and the sine curve ($x_t = \sin(2\pi t/T)$). The difference is best described in the frequency domain. Grenander uses the quantities

$$R_k = \lim_{n\to\infty} \frac{\sum_{t=1}^{n} x_t x_{t+k}}{\sum_{t=1}^{n} x_t^2} \quad (k = 0, 1, 2 \cdots) \qquad (3.3.23)$$

which under certain conditions imposed on the sequence of x_t-values, can be represented as follows:

$$R_k = \int_0^\pi \cos k\lambda \, d\Phi(\lambda) \qquad (3.3.24)$$

where Φ is a distribution function defined in the interval $(0, \pi)$. $\Phi(\lambda)$ is called the spectral distribution function. Grenander shows, that a necessary and sufficient condition for the least squares estimator to be assymptotically efficient is, that the spectral function has only one point of increase. As the trend has $R_k = 1$, and the sine curve has $R_k = \cos(2\pi k/T)$, both vari-

ables have discrete spectra with a single point of increase for $\lambda = 0$, and $\lambda = 2\pi/T$ respectively. The air pressure, being a variable of a random nature opposite to the two other variables mentioned, is better described as a variable with a continuously differentiable spectral function. We denote the derivative by $\varphi(\lambda)$. As the condition placed on the spectral function is necessary, the least squares estimator cannot be efficient.

The least squares estimator of β_1 is:

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{n} y_t x_t}{\sum_{t=1}^{n} x_t^2} \quad (\bar{x} = 0) \tag{3.3.25}$$

which has, according to Grenander, asymptotically, the variance:

$$\sigma^2_{\hat{\beta}_1} = \frac{\pi}{\sum_{t=1}^{n} x_t^2} \int_0^{\pi} f(\lambda)\, \varphi(\lambda)\, d\lambda \tag{3.3.26}$$

where $f(\lambda)$ is the spectral density of the residuals defined by:

$$\sigma_k = \int_0^{\pi} \cos k\lambda\, f(\lambda)\, d\lambda \quad (k = 0, 1, 2 \cdots) \tag{3.3.27}$$

$f(\lambda)$ is equivalent to $\varphi(\lambda)$. But since $\sigma_0 = \sigma_\varepsilon^2$, we get:

$$\int_0^{\pi} f(\lambda)\, d\lambda = \sigma_\varepsilon^2 \tag{3.3.28}$$

In the case of uncorrelated residuals, the covariances σ_k $(k \neq 0)$ equal zero, which implies that $f(\lambda) = \sigma_\varepsilon^2/\pi$. Inserting this spectral density in the formula (3.3.26) gives the usual variance.

The two variances (3.3.22) and (3.3.26) are equivalent. The first one, expressed in the time domain by means of autocorrelation coefficients, is convenient to use when correcting for an enlarged variance due to correlation in the residuals. The second one, expressed in the frequency domain by means of spectral densities, is appropriate when studying the effect of filtering operations, which we will do in the next section.

3.3.5. *Least squares after filtering.*[1] While the iterative method makes use of the first h autocorrelations in the residuals, an efficient method when the residuals are generated by a stochastic difference equation of order h, the method developed in this section is based on the entire correlation structure of both the residuals and the independent variable. As in the preceding section we shall confine ourselves to the case of a single independent variable with a continuous spectral density, $\varphi(\lambda)$.

The residuals are assumed to be stationary taken in the sense of (3.3.4), and to possess a continuous and positive spectral density of $f(\lambda)$, defined by (3.3.27). Stationary residuals generated by the Scheme (3.3.6) are allowed and the spectral density is given by:

$$f(\lambda) = \frac{\sigma_\varepsilon^2}{\pi} \left| \sum_{v=0}^{h} \alpha_v e^{iv\lambda} \right|^{-2} \quad (\alpha_0 = 1) \tag{3.3.29}$$

But other types of stationary residuals are also included. Hence the method in this section is more general in this respect. On the other hand power spectra have to be estimated before the method can be applied. We assume, however, that the spectral densities, $f(\lambda)$ and $\varphi(\lambda)$, are known.

The independent variable, also assumed to have a continuous spectral density, can be looked upon as a realization or a sample of a stochastic process belonging to the same broad class as the residuals. Alternatively, the independent variable can be looked upon as a set of constants obeying the conditions (3.3.21).

We shall consider the effect of a filtering operation on the variables in the regression model (2.4.1). By filtering of the sequence of x_t-values $(t = 1, \ldots n)$ is meant, a linear operation defined by:

$$x_t' = \sum_{v=-k}^{l} a_v\, x_{t-v} \quad (t = l+1, \cdots, n-k) \tag{3.3.30}$$

where a_v are constants. A new sequence of values is obtained, denoted by x_t'. The set of constants: $\mathbf{a} = (a_{-k}, \cdots, a_l)$ is called a filter. A filter is said to be symmetric if $a_{-v} = a_v$. If the sum of the constants equals zero the filter is

[1] As far as the author knows, general filtering technique has not been used in regression analysis before, even if the possibility is mentioned by Hannan (1963).

said to be trend-reducing. We have an example of a trendreducing filter when $a_0 = 1$ and $a_1 = -1$ corresponding to the difference operation: $x_t' = x_t - x_{t-1}$.

Filtering the variables in the regression model (2.4.1) does not effect the parameter β_1. In fact we get:

$$y_t' = \beta_0 \sum_{v=-k}^{l} a_v + \beta_1 x_t' + u_t' \quad (t = l+1, \cdots, n-k) \tag{3.3.31}$$

Hence we can make use of the filtered variables, y_t' and x_t' in estimating the parameter β_1. But the spectral properties of the residual and the independent variable are altered, and therefore the variance of the least squares estimator depends on the applied filter.

The effect of a filtering operation on the (power) spectrum is simply a multiplication by the (power) transfer function (see e.g. Blackman & Tukey, 1959)

$$\psi(\lambda; \mathbf{a}) = \left| \sum_{v=-k}^{l} a_v e^{-iv\lambda} \right|^2 \tag{3.3.32}$$

Denoting the spectral density of the filtered residuals by $f'(\lambda)$ we get:

$$f'(\lambda) = \psi(\lambda; \mathbf{a}) f(\lambda) \tag{3.3.33}$$

In the same way, but remembering the normalizing condition, we obtain the spectral density of the new independent variable as:

$$\varphi'(\lambda) = \frac{\psi(\lambda; \mathbf{a})\, \varphi(\lambda)}{\int_0^\pi \psi(\lambda; \mathbf{a})\, \varphi(\lambda)\, d\lambda} \tag{3.3.34}$$

The division is rendered necessary by the fact that the new independent variable has not the same variance as the unfiltered one. Hence the new variance is obtained by multiplying by the factor used in (3.3.34), giving us:

$$\lim_{n\to\infty} \frac{\sum_{t=l+1}^{n-k} x_t'^2}{\sum_{t=1}^{n} x_t^2} = \int_0^\pi \psi(\lambda; \mathbf{a})\, \varphi(\lambda)\, d\lambda \tag{3.3.35}$$

The asymptotic variance of the least squares estimator of β_1, based on the filtered variables, can now be expressed in the power spectra of the unfiltered variables. Making use of the relations (3.3.26) and (3.3.33–35), we get:

$$\sigma^2_{\hat{\beta}_1'} = \frac{\pi}{n} \frac{\int_0^\pi [\psi(\lambda; \mathbf{a})]^2 f(\lambda)\, \varphi(\lambda)\, d\lambda}{\sum_{t=1}^{n} x_t^2 \left[\int_0^\pi \psi(\lambda; \mathbf{a})\, \varphi(\lambda)\, d\lambda\right]^2} \tag{3.3.36}$$

This formula shows the effect of a filtering operation on the asymptotic variance of the regression coefficient. A lower bound of the variance (3.3.36) is obtained from the Schwartz' inequality:

$$\int_0^\pi \frac{\varphi(\lambda)}{f(\lambda)}\, d\lambda \int_0^\pi [\psi(\lambda; \mathbf{a})]^2 f(\lambda)\, \varphi(\lambda)\, d\lambda \geqslant \left[\int_0^\pi \psi(\lambda; \mathbf{a})\, \varphi(\lambda)\, d\lambda\right]^2 \tag{3.3.37}$$

Hence we get the lower bound:

$$\sigma^2_{\hat{\beta}_1'} \geqslant \frac{\pi}{n} \frac{1}{\sum_{t=1}^{n} x_t^2 \int_0^\pi \varphi(\lambda)\, f^{-1}(\lambda)\, d\lambda} \tag{3.3.38}$$

where the equality sign holds when:

$$\psi(\lambda; \mathbf{a}) f(\lambda) = \text{constant} \tag{3.3.39}$$

Thus, the lower bound is attained when the power spectrum is of the form (3.3.29), i.e. when the residuals make up a stationary autoregressive series. The lower bound (3.3.38) is the variance of the B.L.U. estimator as was shown by Grenander (1954).

The filtering method is optimal only in the case of residuals of the autoregressive type, but even in other cases the filtering method may be applied with only a small loss in efficiency, if the length of the filter and its constants are properly chosen, a rational choice being guided by formula (3.3.36).

The filtering technique is simple as soon as a filter is determined. But before we can study the effect of a particular filter, we must know or have some estimate of the power spectra involved. (When estimating power spectra see Blackman and Tukey, 1959.) The next step, to determine a filter with optimal properties is no easy task, and we do not think it is worth

while unless the filter can be used several times. The iterative least squares method in section (3.3.3) is a filtering method, but the filter is determined simultaneously with the regression coefficients.

Regression analysis on differenced daily values is commonly used to determine the pressure dependence. As the difference operation belongs to the class of filtering operations, this method is a special case of those considered in this section. The rationale of the difference method is that a dominating low frequency component is present in the residuals. When this is the case we can expect the difference method to give more reliable estimates than the ordinary least squares method, which is also found in pressure dependence calculations (Bachelet *et al.*, 1964; Bachelet, Balata & Iucci, 1965; Lapointe & Rose, 1962).

Another estimation method, also based on spectra, is proposed by Hannan (1963). We submit, however, that reliable results can be obtained with the simpler method presented in section (3.3.3), even if the assumptions are not wholly fulfilled.

4. On the choice of periods and aggregation

We have assumed the regression analysis to be based on a fixed set of data, the type of data (e.g. bi-hourly or 24-hourly values) and the number of observations being given. We have, however, at our disposal a long series of measurements covering several years (though not entirely complete due to instrumental breakdowns). In general the observations are hourly or bi-hourly values. When trying to estimate the pressure coefficient, α_t (2.2.5), as a function of time from this data, we have to decide upon the appropriate sample length as well as upon the type of data to use.

If we assume that the entire period is to be divided into non-overlapping sub-periods, each giving an estimate of the pressure coefficient; what would then be the appropriate length of such a sub-period? There is no unique answer to this question. If we want to study the short time variation up to 1 and 2 months in the pressure dependence, it is natural to make the sub-periods short, but it is questionable whether much information of this kind can be extracted from the data. If, on the other hand, we want to study the long time variation, we can use either short or long sub-periods. In the first case, we can take averages over several short periods. Let us say that we want an estimate for a period consisting of r sub-periods, with an estimate a_i of variance σ_i^2 corresponding to each sub-period. Assuming that $E(a_i) = \alpha\,(i = 1, 2, \ldots r)$, the estimator:

$$\bar{a} = \frac{\sum_{i=1}^{r} a_i/\sigma_i^2}{\sum_{i=1}^{r} 1/\sigma_i^2} \tag{4.1}$$

is unbiased and has minimum variance. We get:

$$\sigma_{\bar{a}}^2 = \frac{1}{\sum_{i=1}^{r} \frac{1}{\sigma_i^2}} \tag{4.2}$$

The estimator (4.1) is optimal only in the restricted sense that it is the best linear combination of the estimates a_i. A better estimate may be obtained by a single regression analysis based on the entire period. If we disregard the fact that the residuals are correlated, the variance σ_i^2 is given by the formula:

$$\sigma_i^2 = \frac{\sigma_\varepsilon^2}{n_i s_i^2} \quad (i = 1, 2, \cdots, r) \tag{4.3}$$

where s_i^2 is the variance in the air pressure. Let us assume that the r sub-periods are of equal length; hence: $n_i = n$ $(i = 1, 2, \ldots r)$. Then we can rewrite (4.2) as follows:

$$\sigma_{\bar{a}}^2 = \frac{\sigma_\varepsilon^2}{n \sum_{i=1}^{r} s_i^2} \tag{4.4}$$

If we denote the estimator, based on the entire period, by a, and the variance in the air pressure with s^2, we get:

$$\sigma_a^2 = \frac{\sigma_\varepsilon^2}{rns^2} \tag{4.5}$$

Because of the relation

$$rs^2 = \sum_{i=1}^{r} s_i^2 + \sum_{i=1}^{r} (\bar{x}_i - \bar{x})^2, \tag{4.6}$$

it is clear that $\sigma_a^2 \leqslant \sigma_{\bar{a}}^2$.

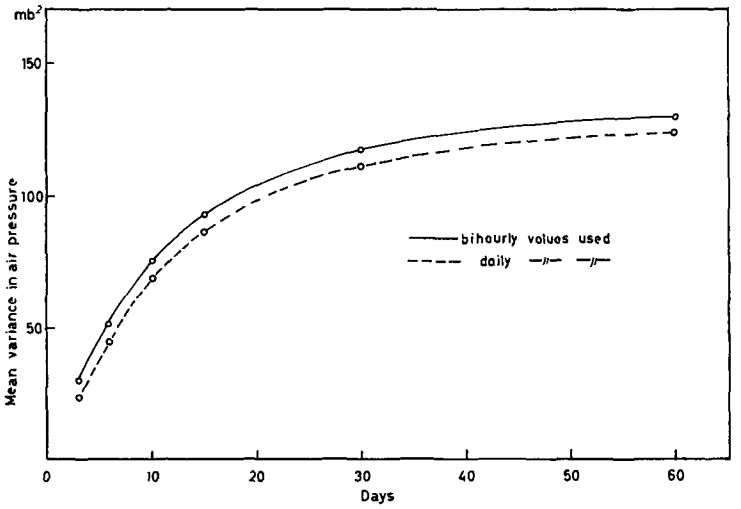

Fig. 4. The mean variance in the air pressure at Upsala 1957–62 as a function of the length of period.

The length of the sub-period is of relevance when considering the variance in the air pressure. This is demonstrated by Fig. 4.

The air pressure (in mb) at Upsala for the years 1957–62 has been used to compute mean variances for six different lengths of period (3, 6, 10, 15, 30, and 60 days). For instance the mean value corresponding to 60 days is based on the following variances. The entire period has been divided into 36 sub-periods of 60 days length, and for each period a variance has been computed using bihourly values. The mean values, plotted on the graph, have been connected by a line drawn by hand. We see that the variance is on the average an increasing function of the length of the period, rather steep to the left but flat to the right. Using formula (4.4), it follows that the variance of the estimate (4.1) is on the average approximately doubled if we use periods of 6 days instead of 20 days. Thus using very short periods may spoil much of the information in the data.

When the residuals are correlated and we apply the iterative least squares method, the pressure variable is in principle filtered and so the variance of the estimate (3.3.20) depends on the variance of the new independent variable. Hence Fig. 4 does not apply to this case. As no standard filter is used, the equivalent to Fig. 4 cannot be computed. We believe, however, that such a curve would flatten out for rather short periods, because a large part of the low frequency component is filtered out. This means that somewhat shorter periods could be used than indicated by Fig. 4.

Another problem, met with in practice, is what type of data to use. Should original values (e.g. bi-hourly) or aggregated (e.g. daily) values be used? If by neglecting the quadratic term in (2.2.11) aggregation is allowed, it is a matter of precision which type we use. By using aggregated values the sample size is reduced as well as the computation time. As seen from Fig. 4 (broken line), the variance of the air pressure is reduced very little by using daily instead of bi-hourly values. Hence, if the residuals were uncorrelated, daily values could be used without any noticeable loss of information. Because of the autocorrelation in the residuals, matters are not so simple. If we knew that, when using bi-hourly values the residuals are generated as an autoregressive series of order five, say, there would be no problem. Unfortunately, the autoregressive model is at best only a good approximation, equally probable for bi-hourly as for daily values. Obviously the order of the autoregressive series (or the length of the filter) is not only intimately connected with the type of data used but also with the length of the period of analysis. If for instance we use 30 daily values to estimate the pressure coefficient, we do not think it is meaningful to use an autogregressive series of higher order than one, or perhaps two. If instead, bi-hourly values are

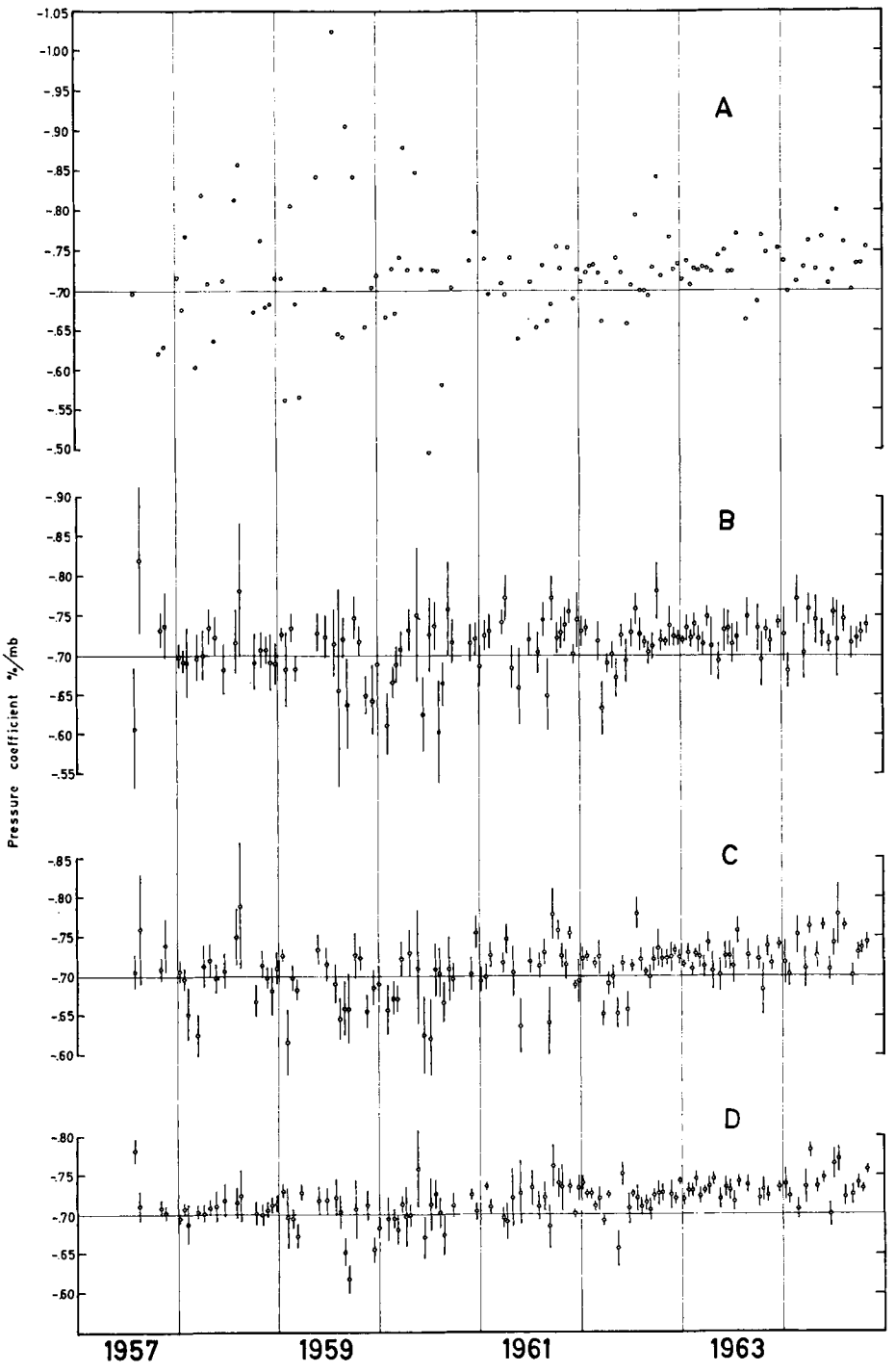

Fig. 5. Pressure coefficients obtained by four different methods: A (1*a*), B (1*b*), C (1*c*), D (2*b*).

used for the same period, a longer filter could be applied without any risk of overfitting. A longer filter does not imply better estimates, but certainly a longer computing time.

In the next section we shall present some results of the estimation methods discussed. Since the periods were rather short (15 days), we used bi-hourly values. However, after looking at these results, we are rather sceptical about the value of such short periods. Periods of say 45 days or so with a filter including 3 or 4 terms are practical and probably quite sufficient. For the time being we cannot give any theoretically motivated recommendations.

5. A numerical comparison of the methods

5.1. *About the data and the methods*

In order to get an idea of how the different estimation methods work in practice, we have applied them to a rather large amount of data covering both disturbed and undisturbed periods.

All through we have been concerned with estimating the pressure coefficient at Upsala. As an auxiliary station we have used Deep River. Pressure corrected data from this station was available from August 1957 to the end of 1964. Hence the comparison of the methods has been restricted to these years. In total, 125 periods of about 15 days length have been analysed. These periods are the same as those used in an earlier work (Dyring, Lindgren & Martinelle, 1966). Thus periods of rapid intensity changes, due to either primary variations or unsatisfactory performance of the detecting and recording instruments, have been avoided.

The following estimation methods have been used:

1. *One-station methods*
   - *a*. Ordinary least squares.
   - *b*. The difference method (the ordinary least squares method applied on differenced daily values).
   - *c*. Iterative least squares.
2. *Two-station methods*
   - *a*. Ordinary least squares.
   - *b*. Iterative least squares.

All methods but the difference method have been applied on bi-hourly values. For the difference method daily values have been used. The length of the filter used in the iterative methods is: $h = 12$. As only pressure corrected data was available from the station at Deep River, the two-station method has been applied in its modified form (3.2.4).

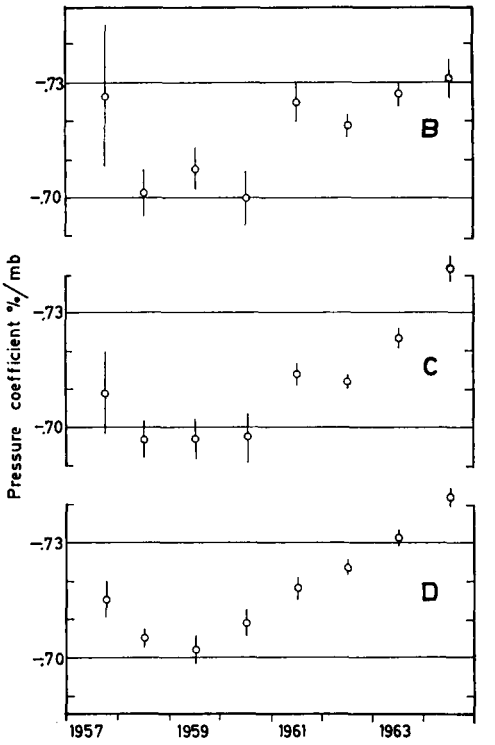

Fig. 6. Yearly mean pressure coefficients computed on the coefficients in Fig. 5.: A (2*b*), B (1*c*), C (1*b*).

5.2. *Results and conclusions*

Fig. 5 shows the estimated coefficients obtained by the methods 1*a–c* and 1*b*. The standard error is plotted on both sides of each estimate, except for the ordinary least squares estimates. As seen from the figure, the ordinary least squares estimates show a large variation, especially during the years 1957–60, which contain rather disturbed periods. The methods 1*b* and 1*c* effectively reduce the variation in the estimates. The difference between 1*b* and 1*c* is not large, but the latter results in a somewhat smaller variation. Using two stations gives, however, more closely clustered estimates (5 D). The coefficients obtained by the method 2*a* are not shown in the figure, but are somewhat more scattered than those plotted in 5 D.

Yearly mean values have been computed using formula (4.1) and are plotted in Fig. 6. The upper diagram is based on the coefficients in Fig. 5 B. We see that the coefficient is dependent on time, with a minimum in 1959. However, the location of the true minimum is uncertain and not contradictory to the results found by Griffiths *et al.* (1966).

For the reason discussed in section 4, better estimates of the yearly means could be obtained if longer periods were used. We believe, however, that the relatively small variation of the estimates in Fig. 5 D and the smooth curve in Fig. 6 D demonstrate clearly enough the effectiveness of the two station method. Using method 2*b* instead of 2*a* is an improvement, but the difference between these two methods is smaller than between 1*a* and 1*c*. For another station couple the improvement may be larger.

Acknowledgements

This work has been sponsored in part by the Air Force Cambridge Research Laboratories under Contract No. AF 61(052)-868 through the European Office of Aerospace Research (OAR), United States Air Force.

REFERENCES

Aitken, A. C. 1934. On least squares and linear combination of observations. *Proc. of Royal Society of Edinburgh 55*, 42–47.

Bachelet, F. *et al.* 1964. Attenuation coefficients of the cosmic-ray nucleonic component in the lower atmosphere. *Il Nuovo Cimento 35*, 23–35.

Bachelet, F., Balata, P. & Iucci, N. 1965. Some properties of the radiation recorded by the IGY cosmic-ray neutron monitors in the lower atmosphere. *Il Nuovo Cimento 40*, 250.

Bartlett, M. S. 1962. *An Introduction to Stochastic Processes.* Cambridge, Cambridge University Press.

Blackman, R. B. & Tukey, J. W. 1959. *Measurement of Power Spectra from the Point of View of Communications Engineering.* New York, Dover Publications.

Cochrane, D. & Orcutt, G. H. 1949. Application of least squares regression to relationships containing autocorrelated error terms. *J. Am. Stat. Ass. 44*, 32–61.

Durbin, J. & Watson, G. S. 1950. Testing for serial correlation in least squares regressin. I, *Biometrika 37*, 409–428.

Durbin, J. 1960. Estimation of parameters in time-series regression models. *J. R. Stat. Soc. B 22*, 139–153.

Dyring, E. 1962. Statistical studies of the performance of cosmic ray recording instruments. *Tellus 14*, 33–43.

Dyring, R., Lindgren, S. & Martinelle, S. 1966. Solar cycle variation in the attenuation length of the cosmic ray nucleon component. *Arkiv f. Geofysik 5*, 87–90.

Grenander, U. 1954. On the estimation of regression coefficients in the case of an autocorrelated disturbance. *Ann. Math. Stat. 25*, 252–272.

Griffiths *et al.* 1966. The variation of the barometric coefficient of the Leeds neutron monitor during the solar cycle 1954–1965. *Journal of Geophysical Research 71* (No. 7), 1895–1898.

Hannan, E. J. 1963. Regression for time series. *Proc. of the symp. on time series analysis.* Ed. M. Rosenblatt, Wiley.

Hildreth, Cliffort & Lu, John Y. 1960. Demand Relations With Autocorrelated Disturbances, Michigan State University, Agricultural Experiment Station. *Technical Bulletin* 276.

Klein, L. 1953. *A Textbook of Econometrics.* Row, Peterson and Company, New York.

Lapointe, S. M. & Rose, D. C. 1962. A statistical analysis of the barometer coefficients for cosmic-ray intensities. *Can. Journ. Phys. 40*, 687.

Lindgren, S. 1961. On the pressure dependence of the cosmic ray intensity recorded by the standard neutron monitor. *Tellus 14*, 44–48.

Mathews, P. M. 1958. Atmospheric effects on cosmic ray intensity at sea level. *Can. J. Phys. 37*, 85–101.

McCracken, K. G. & Johns, D. H. 1959. The attenuation length of the high energy nucleonic component of the cosmic radiation near sea level. *Il Nuovo Cimento 13*, 96.

Wise, 1955. The autocorrelation function and the spectral density function. *Biometrika 42*, 151–159.

Wold, H. 1950. On least squares regression with autocorrelated variables and residuals. *Bull. Inst. Int. Statist. 29* (No. 2).

Wold, H. 1953. *Demand Analysis. A Study in Econometrics*, New York, Wiley.

Wold, H. 1965, A graphic introduction to stochastic processes. In H. Wold, ed.: *Bibliography on Time Series and Stochastic Processes*, 7–76, Edinburgh, Oliver and Boyd.

Wold, H. 1966. *Nonlinear Estimation by Iterative Least Squares Procedures.* In Festschrift Jerzy Neyman, ed.: F. David, New York, Wiley and Sons.

ЗАВИСИМОСТЬ ДАВЛЕНИЯ ВОЗДУХА ОТ ИНТЕНСИВНОСТИ КОСМИЧЕСКИХ ЛУЧЕЙ. МЕТОДЫ И РЕЗУЛЬТАТЫ СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ ПОЛУЧЕННЫХ С ПОМОЩЬЮ ДАТЧИКА НЕЙТРОНОВ

Выводится общая регрессионная модель, которая описывает зависимость давления воздуха от нейтронной компоненты космических лучей. Поскольку остатки корреляционно связаны, то отмечается, что обычный метод наименьших квадратов не является наилучшим для оценки эффектов давления. Обсуждаются другие методы. Для того, чтобы оценить регрессионный коэффициент давления Упсала для 1957–1964 гг. используется пять методов. Оказывается, что модифицированный метод «двух станций» дает лучший результат; существует также зависимость регрессионного коэффициента давления от солнечного цикла.