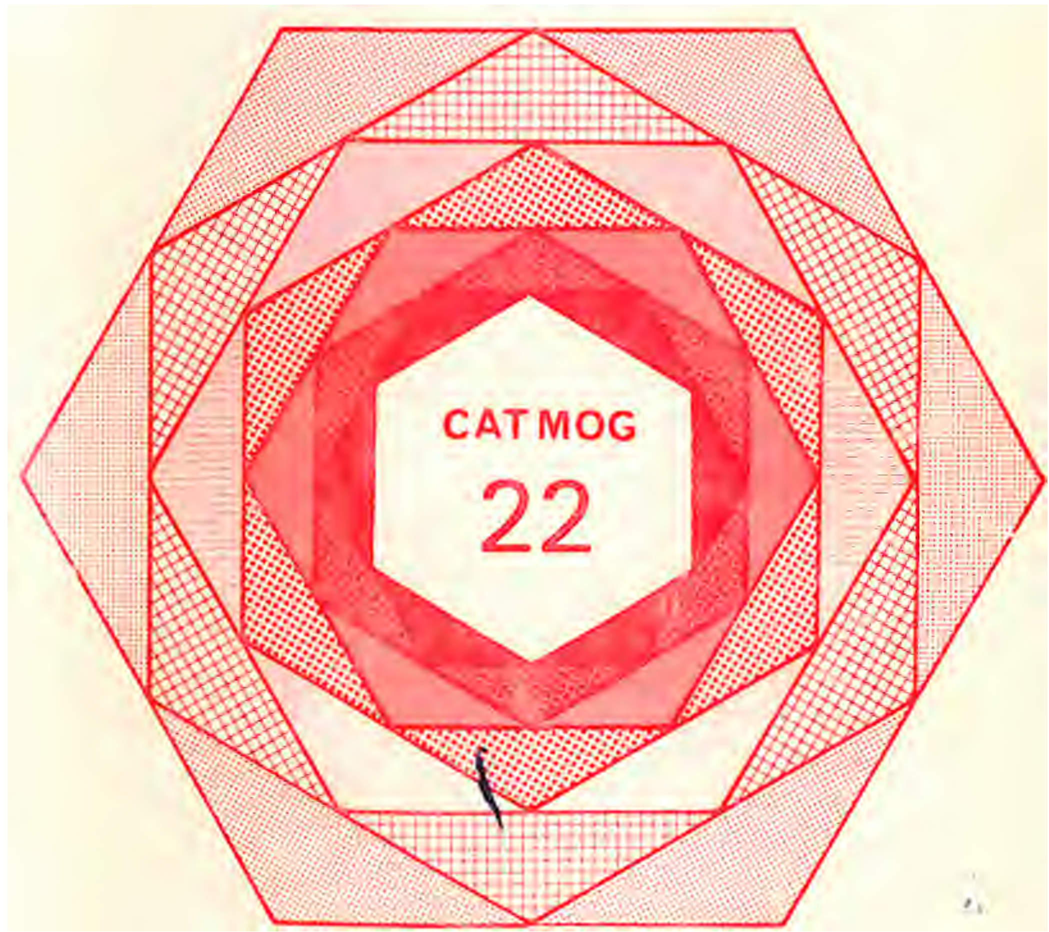
TRANSFER FUNCTION MODELLING RELATIONSHIP BETWEEN TIME SERIES VARIABLES

Pong-wai Lai



ISSN 0306 6142
ISBN 0 86094 029 2

PRINTED BY MID ANGLIA LITHO

CATMOG (Concepts and Techniques in Modern Geography)

1 An introduction to Markov chain analysis - L. Collins
2 Distance decay in spatial interactions - P.J. Taylor
3 Understanding canonical correlation analysis - D. Clark
4 Some theoretical and applied aspects of spatial interaction shopping models - S. Openshaw
5 An introduction to trend surface analysis - D. Unwin
6 Classification in geography - R.J. Johnston
7 An introduction to factor analytical techniques - J.B. Goddard & A. Kirby
8 Principal components analysis - S. Daultrey
9 Causal inferences from dichotomous variables - N. Davidson
10 Introduction to the use of logit models in geography - N. Wrigley
11 Linear programming: elementary geographical applications of the transportation problem - A. Hay
12 An introduction to quadrat analysis - R.W. Thomas
13 An introduction to time-geography - N.J. Thrift
14 An introduction to graph theoretical methods in geography - K.J. Tinkler
15 Linear regression in geography - R. Ferguson
16 Probability surface mapping. An introduction with examples and Fortran programs - N. Wrigley
17 Sampling methods for geographical research - C. Dixon & B. Leach
18 Questionnaires and interviews in geographical research - C. Dixon & B. Leach
19 Analysis of frequency distributions - V. Gardiner & G. Gardiner
20 Analysis of covariance and comparison of regression lines - J. Silk
21 An introduction to the use of simultaneous-equation regression analysis in geography - D. Todd
22 Transfer function modelling: relationship between time series variables - Pong-wai Lai
23 Stochastic processes in one-dimensional series: an introduction - K.S. Richards
24 Linear programming: the Simplex method with geographical applications - James E. Killen

The series is produced by the Study Group in Quantitative Methods, of the Institute of British Geographers. For details of membership of the Study Group, write to the Institute of British Geographers, 1 Kensington Gore, London SW7. The series is published by GEO ABSTRACTS, University of East Anglia, Norwich NR4 7TJ, to whom all other enquiries should be addressed.

CONCEPTS AND TECHNIQUES IN MODERN GEOGRAPHY No. 22

TRANSFER FUNCTION MODELLING: RELATIONSHIP BETWEEN TIME SERIES VARIABLES

by

Pong-wai Lai
(London School of Economics and Political Science)

CONTENTS

Acknowledgement

The author wishes to express his gratitude to Dr. J.B. Thornes for his discussion on various points in this monograph and also for providing the data sets for the examples in the applications of the transfer function method. The help by Miss V. Blank in typing the draft of this monograph is also acknowledged.

I INTRODUCTION

(i) Prerequisite

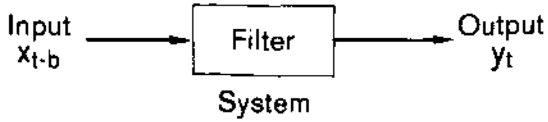
In mathematics, readers are required to have a knowledge of basic algebra and the concept of polynomial functions. A knowledge of matrix algebra is not needed to start with in single-input modelling, but is essential for understanding the calculations in the multiple-input models. Relevant sections in Wilson and Kirkby (1975) may serve as a useful reference when readers have difficulties in understanding the mathematical arguments in this paper.

In statistics, readers should know the basic descriptive statistics and the inferential statistical tests, such as the chi-square test and other significance tests. Another CATMOG monograph (K.S. Richards. Stochastic processes in one-dimensional series CATMOG 23) on univariate time series modelling should be read first to enable the reader to understand basic time series concepts such as autocorrelation, partial autocorrelation and differencing method. The chapters concerning measurement and modelling time series in Thornes and Brunsden (1977) may help readers to grasp the concept of time series in geographical modelling.

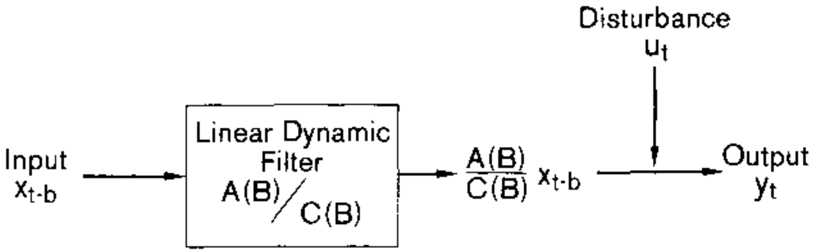
Beginners are recommended to skip the sections involving multiple-inputs modelling. The general concept of spectral analysis has to be understood in dealing with multiple-inputs models, although the calculation procedures are listed in the relevant sections.

(ii) Modelling input-output relationships

Investigation of the causal relationship between two variables is often the focus of geographical studies. The systems approach has been widely used in geographical models. Basically, a system contains a set of elements which are interrelated to fulfil a particular function. When the system receives an input stimulus, it will respond to the input signal according to the internal structure of the system, i.e. the web of relationships between the elements. The internal structure of the system, which governs the input-output relationship, filters the input signal into the output response. In terms of time, the input at time t - b acts on the system, passing through the filter, and is transformed into the output response at time t; b is called the response delay which is the time lag during which the filtering process takes place. The schematic of this concept is presented as Fig. la. Thus the major objective of the investigation of the input-output relationship is to model the filter of the system. If we know the internal mechanism of the system completely, the filter can be represented deterministically by differential equations. These equations can be solved under certain assumptions of the ideal conditions such as homogeneity. There are many geographical models of this kind, especially in physical geography. However, the ideal assumptions used in the deterministic models rarely hold in practice. The systems are affected by random or non-random disturbances which may be caused by spatial and temporal variations in the underlying environment in which the system operates. This problem leads us to model the relationship in a stochastic or probabilistic approach. The input and output of the system are considered as



a) Input-output relationship of a system



b) Input-output relationship represented by a linear dynamic filter, using the transfer function approach

Fig. 1 Schematic of the input-output relationship of a system

stochastic processes determined by probabilistic laws, and they are sequential in spatial and/or temporal sequences. This paper will examine some stochastic models of input-output relationships and introduce the transfer function modelling method which has been recently used by geographers as well as engineers and econometricians.

(iii) The regression model

Regression analysis has been widely used in geographical studies to model input-output relationships, considering the output as the dependent variable and the input as the independent variable. The general form of the multiple input-single output system is:

$$y_t = b_0 + b_1 x_{1,t} + b_2 x_{2,t} + \dots + b_k x_{k,t} + u_t \quad (t=1,\dots n)$$

where

y_t = the output variable at the t^{th} observation

$x_{k,t}$ = the k^{th} input variable at the t^{th} observation

b = the regression coefficients

u_t = the residual (noise) at the t^{th} observation

k = number of input variables

n = number of observations

The regression coefficients can be estimated by the ordinary least square (OLS) method based on the assumptions that the residuals, u_t, are independent of the input variables and the residual series (u_t) have zero mean, are uncorrelated and...have constant variance.

As we can see from equation (1), a regression model only takes into account the simultaneous response between the input and output variables. The simultaneous response of y_t to x_t is represented by the regression coefficient (b). It is often used to model the input-output relationship of a system in a state of equilibrium. In other words, the method is unable to model the transient behaviour of the system in a state of disequilibrium. In some geographical analyses we may want to explain or predict the output behaviour in terms of the short-term variations of the input variables. A regression model appears to be inadequate to cope with these conditions if there is a response delay between input and output.

On the other hand, in most cases geographers find themselves dealing with sequential data, usually in the form of spatial and temporal series. If the autocorrelation of the variables is strong the regression model will not be suitable, for the estimates of the parameter may be invalid.

(iv) The linear system model

From above, we see that the geographers dealing with time series data and modelling transient input-output relationships need a method to cope with the response of the output at time t to the input at time t, t-1, t-2, and so on. The input-output system can be described as a linear system by a linear differential equation in the continuous case, or by a linear difference equation in the discrete case. Considering only the discrete case here, the form of the model is:

$$y_t = v_0 x_t + v_1 x_{t-1} + v_2 x_{t-2} + \dots + u_t \quad \text{(single input)} \qquad (2a)$$

$$y_t = v_{1,0} x_{1,t} + v_{1,1} x_{1,t-1} + \dots + v_{2,0} x_{2,t} + v_{2,1} x_{2,t-1} + \dots + v_{k,0} x_{k,t} + v_{k,1} x_{k,t-1} + \dots + u_t \quad \text{(multiple inputs)} \qquad (2b)$$

where $\{y_t\}$ and $\{x_t\}$ are stationary series with zero means.

The parameters (v) in equations (2a) and (2b) are called the impulse response weights which are analogous to the regression coefficients (b) in equation (1). If these values are 0 for t > 1, then equation (2b) will be the same as a regression model represented by equation (1).

In general, if the values of v_1, v_2 ... are non-zero, then y_t depends not only on x_t but also on x_{t-1}, x_{t-2} ..... . Thus, a linear system model characterizes the dynamic aspects of the input-output relationship while a regression model is static. Also, by comparing equations (1) and (2), we can see that if we model a system which is not in equilibrium, the dynamic characteristics are lumped into the regression coefficients (b).

There are several difficulties encountered in this approach. First, as in equation (2a), y_t may depend on the previous values of x_t to the infinite past (i.e. x_{t-1}, x_{t-2}, ... $x_{t-\infty}$). In order to make the estimation and prediction procedures possible, one assumes that the impulse response weights will die down and become zero after some lags k. The equation (2a) becomes:

$$y_t = v_0 x_t + v_1 x_{t-1} + \dots + v_k x_{t-k} + u_t \qquad (3)$$

This assumption, which is analogous to the assumption of stationarity in the univariate time series case, is called the assumption of stability of the system. However, the determination of the value of k is not a simple matter. Second, if the impulse response weights do not die down very quickly the value of k will be quite large (e.g. 20 or 30), and a large number of parameters would have to be estimated. Third, since the time series y_t and x_t are usually autocorrelated, the estimates of the parameters are expected to be highly correlated.

To overcome the first problem it is sometimes suggested that one chooses a large value of k and estimates the impulse response weights by regression analysis. The actual value of k is then determined by deciding beyond which point the values are not significant. However, because of the second and the third difficulties, this solution is not adequate.

The third difficulty may be overcome if the estimation is carried out in the frequency domain by the cross-spectral analysis method. The estimates of the impulse responses can be obtained by taking the inverse Fourier transformation from the frequency response function. Details of this procedure will be discussed later. The estimates obtained by this procedure are too approximate for forecasting purposes. They are only accurate enough to detect the pattern of the impulse response weights.

(v) The distributed-lags model

In the field of econometric studies, a dynamic economic model can be represented as equation (3). The endogenous variable is dependent on the current and the past values of the exogenous variable. In addition to the problem of a large number of parameters to be estimated, a large number of observations would be lost due to lagging the exogenous variable. .This could be a serious problem in econometric studies in which the number of observations available is often quite small. Thus the econometricans have developed the distributed-lags models to overcome these problems.

One of the distributed-lags models is called the Koyck-lags model. In this, it is assumed that the impulse response weights decline geometrically:

$$v_i = C^i v_0$$

where:

i=1, ...k and C is a constant less than 1 and greater than 0.

Therefore equation (3) can be simplified as:

$$y_t = v_0 x_t + C y_{t-1} + (1-C) u_t$$

This largely reduces the number of observations lost and the number of parameters to be estimated. However the Koyck model does not generally satisfy the underlying statistical assumptions about the residuals and thus the estimates obtained are not necessarily consistent and unbiased. Also, the assumption of the geometric pattern of the weights makes the model very inflexible. A more flexible model called the Almon-lags model assumes that the pattern of the weights can be described by a polynomial:

$$v_i = a_0 + a_1 i + a_2 i^2 + \dots + a_m i^m$$

where m is the order of the polynomial and $a_0, \dots a_m$ are the coefficients. One may notice in the later sections that the Koyck model actually is a special case of the transfer function model. A problem of the Almon-lags model is that the pattern of the weights is unknown prior to the estimation of the model parameters.

Details of the distributed-lags models will not be covered here. The relevant section in Kelejian and Oates (1974) is a good introduction to the method.

(vi) The transfer function model

The transfer function approach involves modelling the linear system with a relatively small number of parameters. In the single- input-single output case, the model is in the form of:

$$y_t = \frac{A(B)}{C(B)} x_{t-b} + u_t \qquad (4a)$$

where u_t is the disturbance (noise)

A(B) is a polynomial of order S considering B as the variable

$$A(B) = (A_0 - A_1 B - A_2 B^2 - A_3 B^3 - \dots - A_s B^s)$$

$A_0, A_1, A_2, \dots, A_s$ are the parameters of the model

B is the backward shift operator

i.e. $x_t B = x_{t-1}$

$x_t B^2 = x_{t-2}$

$x_t B^r = x_{t-r}$

Thus, $A(B)x_{t-b} = (A_0 - A_1B - A_2B - \ldots - A_sB^s)\ x_{t-b}$

$$= A_0x_{t-b} - A_1x_{t-b}B - A_2x_{t-b}B^2 - \ldots - A_sx_{t-b}B^s$$

$$= A_0x_{t-b} - A_1x_{t-b-1} - A_2x_{t-b-2} - \ldots - A_sx_{t-b-s}$$

Similarly, C(B) is a polynomial of order r considering B as the variable.

$$C(B) = (1 - C_1B - C_2B^2 - \ldots - C_rB^r)$$

where $C_1, C_2, \ldots, C_r$ are the parameters of the model.

Multiplying both sides of equation (4a) by C(B), it becomes:

$$C(B)y_t = A(B)x_{t-b} + n_t$$

where $n_t = C(B)u_t$

Then it can be written as:

$$(1 - C_1B - C_2B^2 - \ldots - C_rB^r)y_t = (A_0 - A_1B - A_2B^2 - \ldots - A_sB^s)x_{t-b} + n_t$$

or $$y_t - C_1y_{t-1} - C_2y_{t-2} - \ldots - C_ry_{t-r}$$

$$= A_0x_{t-b} - A_1x_{t-b-1} - A_2x_{t-b-2} - \ldots - A_sx_{t-b-s} + n_t \quad (4b)$$

From equation (4b), we can see y_t is dependent upon the current and the past s values of the input variable (at time t, t-1, ...t-s) and also upon the past r values of the output variable itself.

Equation (2a) can also be written as:

$$y_t = v(B)x_t + u_t \quad (5)$$

where $v(B) = (v_0 + v_1B + v_2B + \ldots)$

is called the impulse response function.

Equation (4a) can be written as:

$$y_t = \frac{A(B)}{C(B)} B^b x_t + u_t$$

where $B^bx_t = x_{t-b}$

Comparing it with equation (5),

$$\frac{A(B)B^b}{C(B)} = v(B)$$

or $$A(B)B^b = v(B)C(B) \quad (6)$$

From equation (6), we see that the impulse response function (v(B)), which is a polynomial of very high order (K), can be expressed as a ratio of two polynomials of order s and r respectively. Box and Jenkins (1970) suggested that s and r are seldom greater than 2. Thus, a linear system originally described by a large number of parameters can now be represented by a transfer function model involving only a few parameters.

A(B)/C(B) is called the transfer function of the system. The concept of a transfer function model is like passing the input series through a stochastic-dynamic filter to generate the output series. This idea is shown as figure 1.

In case of multiple input-single output, the model is in the form of:

$$y = \frac{A_1(B)}{C_1(B)} x_{1,t-b_1} + \ldots + \frac{A_m(B)}{C_m(B)} x_{m,t-b_m} + u_t \quad (4c)$$

$\frac{A_1(B)}{C_1(B)}, \ldots, \frac{A_m(B)}{C_m(B)}$ are then the transfer functions

between the output y_t and the corresponding inputs $x_{1,t-b_1}, \ldots, x_{m,t-b_m}$

The procedure of building a transfer function model involves three steps: a) identification, b) estimation and c) model checking. In identification, we have to determine the orders of the polynomials A(B) and C(B), the b value between the output and each input. In other words, the structural form of the model has to be determined. The noise term u_t can be modelled as an autoregressive-integrative-moving average (ARIMA) model. The details of the method of this univariate stochastic modelling are covered in Richards' CATMOG. This ARIMA model has to be identified, too. In estimation, the parameters involved in the model are estimated using the non-linear least square method. Finally, after the model has been fitted, we have to check the fitted model according to several criteria.

II A DIFFUSION MODEL - AN ILLUSTRATIVE EXAMPLE OF AN INPUT-OUTPUT STOCHASTIC-DYNAMIC MODEL

A dynamic input-output system can be modelled in a deterministic way in terms of the differential equations. In the discrete case, it can be represented by the difference equations. In this section, a diffusion model is used to illustrate the close relationship between a deterministic model and a transfer function model.

Suppose that at time t, $z_{x(t)}$ and $z_{y(t)}$ are the concentrations of a chemical at two different layers in a soil mass. The gradient of concentration between the two locations is $z_{x(t)} - z_{y(t)}$. Due to the existence of this gradient, the chemical at the location of x will diffuse to the location of y. This one-dimensional diffusion process was modelled deterministically by de Wit and van Kenlan (1972). Their model can be simplified as:

$$\frac{dz_y}{dt} = g(z_{x(t)} - z_{y(t)}) \quad (7)$$

where $\frac{dz_y}{dt}$ is the change of the concentration at y over time and g is a constant.

Using the symbol $D = {}^{d}/_{dt}$, the differential operator, equation (7), becomes:

$$(1 + KD)z_{y(t)} = gz_{x(t)} \qquad (8)$$

where $K = {}^{1}/_{g}$ is also a constant.

If there is a response delay between input and output, the equation becomes:

$$(1 + KD)z_{y(t)} = gz_{x(t-b)}$$

Representing the process in a discrete way, the equation is:

$$(1 + K\nabla)z_{y(t)} = gz_{x_{t-b}} \qquad (9)$$

where ∇ is the differencing operator, i.e. $\nabla y_t = y_t - y_{t-1}$.

Recalling the use of the backward shift operator B

$$y_t B = y_{t-1}$$

and

$$\begin{aligned}(1 - \nabla)y_t &= y_t - \nabla y_t \\ &= y_t - y_t + y_{t-1} \\ &= y_{t-1} ,\end{aligned}$$

it shows the relationship between B and ∇ is that

$$B = 1 - \nabla .$$

Substituting ∇ by B in equation (9), it becomes:

$$(1 - C_1 B) z_{y_t} = A_0 z_{x_{t-b}}$$

or

$$z_{y_t} = \frac{A_0}{(1-C_1 B)} z_{x\ t-b} \qquad (10)$$

Equation (10) represents a deterministic one-dimensional diffusion model in terms of a first order difference equation. To put the model in a stochastic mode, a stochastic disturbance term u_t is included.

$$z_{y_t} = \frac{A_0}{(1-C_1 B)} z_{x_{t-b}} + u_t \qquad (11)$$

The disturbance term represents the random variations of the concentration within the soil and the effects of other inputs on the system, such as diffusion in other spatial dimensions.

Comparing the above equations, we see that a first order differential equation can be represented discretely as a first order difference equation and also in the form of a first order transfer function model. The stochastic-dynamic model of equation (11) is in a form parallel to the deterministic model. The parameters of this model can be estimated in a stochastic framework and the forecasts of the output series according to the input series can be performed in a stochastic way.

In general, dynamic systems of a higher order can be represented by higher order differential equations. Also, not only the level of the input $x_{(t)}$ but also its rate of change $({}^{dx}/_{dt})$ and its higher derivatives can influence the output of the system.

Thus, the general form of a continuous dynamic system is:

$$(1 + E_1 D + \ldots E_r D^r)y_{(t)} = g(1 + H_1 D + \ldots + H_s D^s)x_{(t-b)}$$

or it may be expressed discretely as:

$$(1 + e_1 \nabla + \ldots + e_r \nabla^r)y_t = g(1 + h_1 \nabla + \ldots + h_s \nabla^s)x_{t-b} \qquad (12)$$

In terms of the backward shift operator (B), it is expressed by:

$$(1 - C_1 B - \ldots - C_r B^r)y_t = (A_0 - A_1 B - \ldots - A_s B^s)x_{t-b} \qquad (13)$$

As a stochastic-dynamic model, it is in a form of a transfer function model of order (r,s,b):

$$y_t = \frac{A(B)}{C(B)} x_{t-b} + u_t \qquad (14)$$

III APPLICATIONS IN GEOGRAPHY

Having discussed the concept of the transfer function modelling method, it appears that the method could be very useful in geographical studies, since geographers quite often deal with spatial and/or temporal series and attempt to investigate the transient behaviour of the input-output system. In hydrological studies, Clarke (1971) applied the method in modelling relationships between effective rainfall and stream runoff. Whitehead and Young (1975) also used the method to develop a stochastic-dynamic model of rainfall-runoff system and linked it to the water quality system. R.J. Bennett has made a number of contributions to the application of the method in geography and extended it into a temporal-spatial model (Bennett, 1975a, b, Bennett and Haining, 1976). Bennett (1976) applied the concept to geomorphology in dealing with the drainage basin system. He tried to relate the channel geometry at a spatial location in a down-stream direction to other factors and to the channel geometry at previous spatial locations. Haggett (1971) and Bassett and Haggett (1971) noticed the importance of the lead-lag relationships in regional economic systems. Although they investigated the relationship in a frequency domain using cross-spectral analysis, their example suggests the application of transfer function modelling to the investigation of regional economic systems. In fact, Bennett (1975b) represented the regional economic system by a spatio-temporal transfer-function model. Martin and Oeppen (1975) attempted to extend the method to build spatio-temporal forecasting models and Lai (1977) attempted to model the relationship between the soil moisture tension at two different depths within a soil mass by the technique.

In the above, there are only some examples of the applications of the technique in geography. In the later sections, the method will be applied to a soil temperature model in the single-input case.

IV IDENTIFICATION OF A TRANSFER FUNCTION MODEL

(i) Generating stationary series with zero mean

As in equations (4a), (4b) and (4c), the transfer function model relates the input and output series which are assumed to be stationary series with zero means.

In order to obtain stationarity of a univariate time series, there are several methods which have been widely used, e.g. linear regression, harmonic regression and differencing method. These methods have already been discussed by Richards in his CATMOG monograph. No attempt will be made here to cover them in detail.

Whatever methods are used to transform the original series, one must remember that the resulting transfer function model is the model between the transformed series, e.g. the residuals after the regression detrending procedure. There is an advantage in using the differencing method in connection with transfer function model building. Box and Jenkins (1970) pointed out that if the same order of differencing applied to both the input and all the output series, the resulting transfer-functions between the differenced series are identical to the ones between the raw series. In order to illustrate this point, consider a single input model

$$y_t = V(B)x_t + u_t \tag{15}$$

$$\begin{aligned}\nabla y_t &= y_t - y_{t-1}\\ &= (V(B)x_t + u_t) - (V(B)x_{t-1} + u_{t-1})\\ &= V(B)(x_t - x_{t-1}) + u_t - u_{t-1}\\ &= V(B)\nabla x_t + \nabla u_t\end{aligned}$$

Thus, if we obtain the transfer function model of the same order differenced series as:

$$\nabla y_t = V(B)\nabla x_t + n_t$$

we can work back the transfer function model between the raw series as:

$$y_t = V(B)s_t + \frac{1}{\nabla} n_t$$

$$\text{or} \quad y_t = V(B)x_t + \frac{1}{(1-B)} n_t \tag{16}$$

Thus, although the differencing method sometimes cannot transform the series properly, it is preferable to use the method if it is applicable along with the other methods.

After transforming the series into a stationary series, if it still has a non-zero mean then the mean has to be removed by subtracting it from each observation.

(ii) Relationship between the impulse response function and the transfer function of a model

Recalling equation (6):

$$A(B)B^b = v(B)C(B)$$

$$(A_0 - A_1B - A_2B^2 - \ldots - A_sB^s)B^b = (v_0 + v_1B + v_2B^2 + \ldots)(1 - C_1B - \ldots - C_rB^r)$$

On equating coefficients of B, Box and Jenkins (1970) derived a set of simultaneous equations relating the impulse response weights and the parameters in the transfer function (Box and Jenkins, 1970, p 347).

$$\begin{aligned}
v_j &= 0 && j < b\\
v_j &= C_1v_{j-1} + C_2v_{j-2} + \ldots + C_rv_{j-r} + A_0 && j = b\\
v_j &= C_1v_{j-1} + C_2v_{j-r} + \ldots + C_rv_{j-r} - A_{j-b} && j = b+1,\ b+2, \ldots b+s\\
v_j &= C_1v_{j-1} + C_2v_{j-2} + \ldots + C_rv_{j-r} && j > b+s \qquad (17)
\end{aligned}$$

The impulse response weights $v_{b+s}, v_{b+s-1}, \ldots, v_{b+s-r+1}$ supply r starting values for the difference equation

$$C(B)v_j = 0$$

$$\text{or} \quad (1 - C_1B - \ldots - C_rB^r)v_j = 0 \qquad j > b+s$$

Thus, in general, the pattern of the impulse response weights v_j is:

a) $v_0, v_1, \ldots v_{b-1} = 0$

b) $v_b, v_{b+1}, \ldots v_{b+s-r}$ have no fixed pattern (no such values occur if $s < r$)

c) v_j $(j \geq b+s-r+1)$ follows the pattern determined by the r^{th} order difference equation with r starting values

$$v_{b+s},\ v_{b+s-1},\ v_{b+s-2},\ \ldots\ v_{b+s-r+1}.$$

Knowing the impulse response function of a model, it is then possible to identify the orders r, s and b and initially estimate the parameters A_0, $A_1, \ldots, A_s$, $C_1, \ldots, C_r$ according to the relationship between the impulse response function and the transfer function mentioned above. The impulse response function of transfer function models or orders r and s ranging from 0 to 2 are listed in Box and Jenkins (1970, p 350, Table 10.1). As pointed out before, the values of r and s rarely exceed 2 in practice. This table may serve as a useful guide to identifying the order of a transfer function model and to initially estimate the parameters involved.

In order to illustrate the above discussion more clearly, a simulation of two transfer function models with different orders and parameters was carried out. The theoretical values of the impulse response weights were then calculated and plotted in Figure 2. The orders of the models and values of the impulse response weights and the parameters are listed in Table 1.

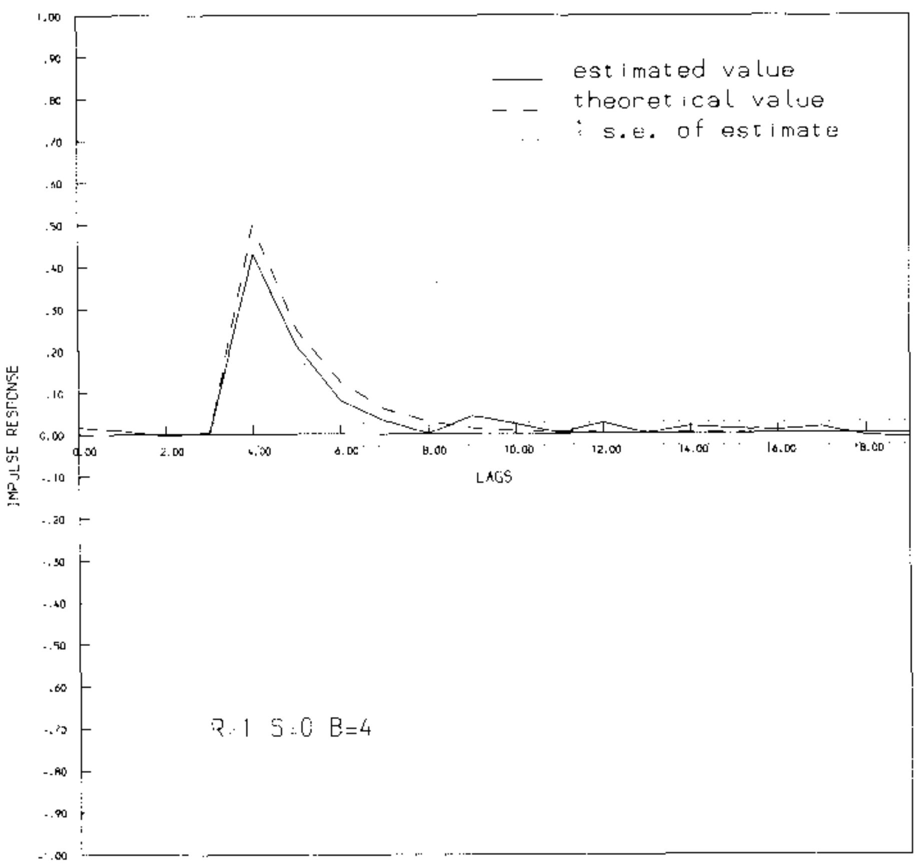


Fig. 2a Impulse response function of the simulated model of orders: r=1, s=0, b=4

(iii) Estimation of the impulse response function

A direct estimation in the time domain of the impulse response weights of models having the structure of equations (2a) and (2b) similar to regression analysis has already been discussed in Section I(iii). Due to the difficulties mentioned, this method is unsatisfactory. Thus, the estimation procedure has to be done indirectly. Box and Jenkins (1970) derived a method which is useful for estimation of the impulse response function but which is only applicable in the single input case. In the multiple input case, the estimation has to be carried out in the frequency domain. For readers who are not familiar with the transfer function modelling technique, they are recommended to start with a single input case. Below, both the single and the multiple input cases will be discussed.

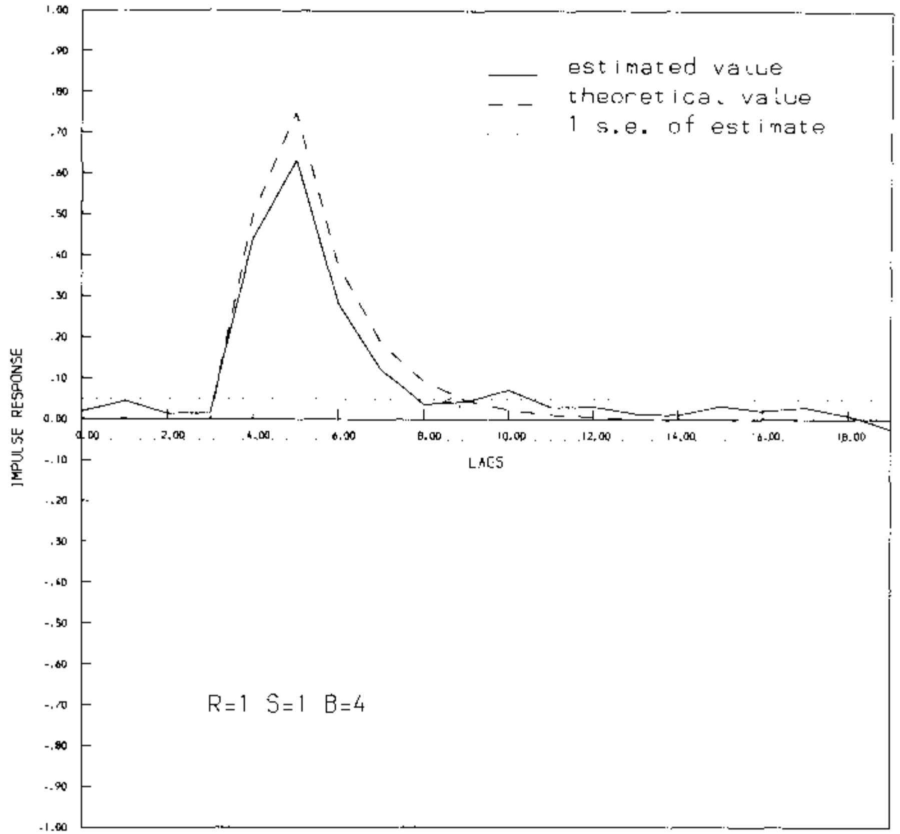


Fig. 2b Impulse response function of the simulated model of orders: r=1, s=1, b=4

a) Estimation in time domain

This method involves transforming the input series x_t into a non-autocorrelated white noise series α_t by an ARIMA filter (this is called the prewhitening procedure). Then the output series y_t is transformed by the <u>same</u> filter into a generally autocorrelated noise series β_t.

The estimation procedure is as follows:

(1) Fit an ARIMA model to the input series x_t and the residuals (noise series) α_t are obtained.

$$\nabla^d \phi_x(B) x_t = \theta_x(B) \alpha_t$$

or

$$\alpha_t = \phi_x(B) \nabla^d \theta_x(B)^{-1} x_t$$

Table 1. The transfer function and the impulse response function of two simulated models.

	Model 1 (1, 0, 4) $A_0 = .5$ $C_1 = .5$		Model = (1, 1, 4) $A_0 = .5$ $A_1 = -.5$ $C_1 = -.5$	
lag k	Theoretical k	Estimated* v_k	Theoretical v_k	Estimated* v_k
0	0	.0198	0	.0199
1	0	.0087	0	.0450
2	0	-.0032	0	.0134
3	0	.0054	0	.0145
4	.5	.4303	.5	.4397
5	.25	.2104	.75	.6328
6	.125	.0807	.375	.2828
7	.0625	.0337	.1875	.1223
8	.03125	.0018	.09375	.0372
9	.015625	.0438	.046875	.0442
10	.0078125	.0244	.0234375	.0721
11	.003906	.0026	.0117187	.0303
12	.001953	.0273	.005859	.0314
13	.000977	.0017	.002930	.0150
14	.000488	.0176	.001465	.0115
15	.000744	.0123	.000732	.0336
16	.000122	.0084	.000366	.0212
17	.000061	.0170	.000183	.0302
18	.000031	-.0069	.000092	.0101
19	.000015	-.0118	.000046	-.0238
		S.E. =.0281		S.E. =.0500

* Estimated by using cross spectral analysis.

where ϕ_x and θ_x are the parameters of the autoregressive part and the moving average part respectively.

(2) Using the same ARIMA filter of input series x_t, transform the output series y_t.

$$\nabla^d \phi_x(B)\theta_x(B)^{-1} y_t = \beta_t$$

For example, if an AR(2) model is fitted to x_t, then

$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} = \alpha_t$$

$$\text{and} \quad y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} = \beta_t$$

(3) Calculate the cross correlation between α_t and β_t. Generally, the cross covariance between input x_t and output y_t is given by:

$$C_{xy(k)} = \frac{1}{n}\sum_{t=1}^{n-k}(x_t - \bar{x})(y_{t+k} - \bar{y}) \tag{18}$$

where $C_{xy(k)}$ is the cross covariance between x_t and y_t at lag k,

$k = 0, 1, 2 \ldots$ mlag

n = no. of observations in the series

mlag = maximum lag after which $C_{xy(k)}$ are negligible.

The cross correlation is:

$$\gamma_{xy(k)} = \frac{C_{xy(k)}}{s_x s_y} \tag{19}$$

where s_x = standard deviation of x_t

s_y = standard deviation of y_t

Then, the cross covariance between α_t and β_t is:

$$C_{\alpha\beta(k)} = \frac{1}{n}\sum_{t-1}^{n-k}(\alpha_t - \bar{\alpha})(\beta_{t+k} - \bar{\beta}) \tag{20}$$

and the cross correlation is:

$$r_{\alpha\beta(k)} = \frac{C_{\alpha\beta(k)}}{s_\alpha s_\beta} \tag{21}$$

(4) The impulse response between x_t and y_t can be calculated by:

$$v_k = \frac{C_{\alpha\beta(k)}}{s^2\alpha} \tag{22}$$

$$\text{or} \quad v_k = \frac{r_{\alpha\beta(k)} s_\beta}{s_\alpha} \tag{23}$$

$k = 0, 1, 2 \ldots$ mlag

Assuming that α_t and β_t have no cross correlation, the standard deviation of the cross correlation $r_{\alpha\beta(k)}$ can be approximated by:

$$s_{r(k)} = \pm \frac{1}{\sqrt{n-k}} \qquad (24)$$

Thus, $s_{r(k)}$ can provide a statistical check whether $r_{\alpha\beta(k)}$ is significantly different from zero. A rough guide to decide if v_k is significantly different from zero can also be obtained.

This method of the estimation of v_k is only applicable in the single input case. In the multiple input case, due to the cross correlation between the prewhitened input series, v_k cannot be obtained in a similar manner. The estimation procedure must then be carried out in the frequency domain.

b) Estimation in frequency domain

In the multiple input case, we have to carry out a cross spectral analysis to estimate the frequency response function, which is actually the frequency representation of the impulse response function. Then, the impulse response function can be acquired by taking the inverse Fourier transformation from the estimated frequency response function. In this section, the estimation procedure will be outlined. However, readers are recommended to consult other references on spectral analysis if they wish to understand the concepts of analysis completely. The classical Fourier transformation will be used here, although the fast Fourier transformation is more commonly used.

Suppose we are dealing with a transfer function model with an input series $\{x_{1,t}\}$, $\{x_{2,t}\}$ ... $\{x_{m,t}\}$ and the ouput series $\{y_t\}$.

1) Calculate the cross covariances and auto-covariances of each series to all other series.

Autocovariance $$C_{xx(k)} = \frac{1}{n}\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x}) \qquad (25)$$

for $k + 0, \pm 1, \pm 2, \ldots$ lag

Cross covariance $$C_{xy(k)} = \frac{1}{n}\sum_{t=1}^{n-k}(x_t - \bar{x})(y_{t+k} - \bar{y}) \qquad (26)$$

for $k = 0, 1, 2, \ldots$ mlag

$$C_{xy}(-k) = \frac{1}{n}\sum_{t=1}^{n-k}(x_{t+k} - \bar{x})(y_t - \bar{y})$$

where mlag is the maximum lag beyond which the auto- and cross covariance are considered to be zero.

Using equations (25) and (26), the covariances matrix

$$\begin{bmatrix} C_{x_1x_1} & C_{x_1x_2} & C_{x_1x_3} & \cdots & C_{x_1x_m} & C_{x_1y} \\ C_{x_2x_1} & C_{x_2x_2} & C_{x_2x_3} & \cdots & C_{x_2x_m} & C_{x_2y} \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ C_{x_mx_1} & C_{x_mx_2} & C_{x_mx_3} & \cdots & C_{x_mx_m} & C_{x_my} \\ C_{yx_1} & C_{yx_2} & C_{yx_3} & & C_{yx_m} & C_{yy} \end{bmatrix}$$

can be obtained.

2) Calculate the spectra and cross-spectra of each series to all other series

Spectrum $$Sp_{xx}(h) = \frac{2}{\pi}\sum_{k=0}^{mlag} e_p\, C_{xx(k)} \cos\frac{hk\pi}{mlag} \qquad (27)$$

where $h = 0, 1, \ldots$ mlag

and $e_p = 0.5$ for $k=0$ or mlag

1.0 otherwise

The cross spectrum is complex in general and can be represented as:

$$Sp_{xy(h)} = CSp_{xy(h)} - iQSp_{xy}$$

where $i = \sqrt{-1}$,

$CSp_{xy(h)}$ is the co-spectrum,

and $QSp_{xy(h)}$ is the Quadrature spectrum

$$CSp_{xy(h)} = \frac{1}{\pi}\sum_{k=0}^{mlag}\left\{e_p\left[C_{xy(k)} + C_{xy(-k)}\right]\cos\frac{hk}{mlag}\right\} \qquad (28)$$

$$QSp_{xy(h)} = \frac{1}{\pi}\sum_{k=0}^{mlag}\left\{e_p\left[C_{xy(k)} - C_{xy(-k)}\right]\sin\frac{hk}{mlag}\right\} \qquad (29)$$

3) Smooth the spectra and the cross-spectra. There are many procedures of smoothing the spectral estimates. The Hamming window is used in this paper. The spectra and cross-spectra are then smoothed by using the Hamming window. Let Sp be the spectrum (or cross-spectrum) and STD be the smoothed. spectrum (or smoothed cross-spectrum).

The smoothing procedure is:

$$\overline{Sp}(h) = 0.54\ Sp(h) + 0.46\ Sp(h+1) \qquad \text{for } h=0$$
$$\overline{Sp}(h) = 0.54\ Sp(h) + 0.46\ Sp(h-1) \qquad \text{for } h=mlag \qquad (30)$$
$$\overline{Sp}(h) = 0.54\ Sp(h) + 0.23Sp(h+1) + 0.23\ Sp(h-1) \qquad \text{for } h \neq 0,\ mlag$$

4) Estimate the frequency response function of each input series. The frequency response function in general is also complex and can be defined as:

$$H_j(h) = HR_j(h) - iHI_j(h) \qquad \text{for } j = 1, \ldots m \qquad (31)$$

Following Box and Jenkins (1970) and Jenkins and Watts (1969)

$$\overline{SP}_{x_1y}(h) = H_1(f)\overline{SP}_{x_1x_1} + H_2(f)\overline{SP}_{x_1x_2}(h) + \ldots + H_m(f)\overline{SP}_{x_1x_m}(h)$$

$$\vdots$$

$$\overline{SP}_{x_my}(h) = H_1(f)\overline{SP}_{x_mx_1} + H_2(f)\overline{SP}_{x_mx_2}(h) + \ldots + H_m(f)\overline{SP}_{x_mx_m}(h) \qquad (32)$$

Combining equations (27) and (31) with (32), we obtain

$$\begin{bmatrix} HR_1 \\ HI_1 \\ \vdots \\ HR_m \\ HI_m \end{bmatrix} = \begin{bmatrix} \overline{CSP}_{x_1x_1} & -\overline{QSP}_{x_1x_1} & \ldots & \overline{CSP}_{x_1x_m} & -\overline{QSP}_{x_1x_m} \\ \overline{QSP}_{x_1x_1} & \overline{CSP}_{x_1x_1} & \ldots & \overline{QSP}_{x_1x_m} & \overline{CSP}_{x_1x_m} \\ \vdots & & & & \\ \overline{CSP}_{x_mx_1} & -\overline{QSP}_{x_mx_1} & \ldots & \overline{CSP}_{x_mx_m} & -\overline{QSP}_{x_mx_m} \\ \overline{QSP}_{x_mx_1} & \overline{CSP}_{x_mx_1} & \ldots & \overline{QSP}_{x_mx_m} & \overline{CSP}_{x_mx_m} \end{bmatrix}^{-1} \begin{bmatrix} \overline{CSP}_{x_1y} \\ \overline{QSP}_{x_1y} \\ \vdots \\ \overline{CSP}_{x_my} \\ \overline{QSP}_{x_mx_m} \end{bmatrix} \qquad (33)$$

Hence, the frequency response functions can be estimated.

5) Calculate the impulse response function

After the frequency response function is estimated, the impulse response function can be obtained by taking its inverse Fourier transform

$$V_{jk} = (1/mlag) \sum_{h=0}^{mlag} e_p \left[HR_j(h)\cos(hk\pi/mlag) + HI_j(h)\sin(hk\pi/mlag) \right] \qquad (34)$$

for k=0,1,2, ...mlag, and j=1,2,...m.

6) Estimate the noise spectrum

The noise spectrum can be estimated by:

$$SP_n(h) = \overline{SP}_{yy}(h) - H_1(h)\overline{SP}_{yx_1} - \ldots - H_m(h)\overline{SP}_{yx_m} \qquad (35)$$

7) Derive the standard error of v_k in case of a single input model

Jenkins and Watts (1969) derived the variance of the frequency response function H(h) in single-input cases as:-

$$s_H^2(h) = \frac{2}{df-2} \cdot \frac{SP_n(h)}{\overline{SP}_{xx}(h)} \qquad (36)$$

The standard error of the estimated impulse response function can then be obtained:

$$s_v = (1/mlag)\left[\sum_{h=0}^{mlag} e_p^2\ s_H^2(h) \right]^{1/2} \qquad (37)$$

where e_p =0.5 for h=0,mlag and e_p =1 otherwise.

The impulse response estimates at different lags have the same standard error. The derivation of equation (37) is listed in Appendix III.

In the multiple-inputs case, the derivation has not yet been achieved successfully. Jenkins and Watts (1969) suggested an approximate variance of the frequency response function based on the assumption that the input series are not correlated with each other. However, this assumption is usually untrue. Also, if error occurs in the calculation of $s_H^2(h)$, the errors in calculating s_v will be enormous since equation (37) will accumulate the errors in different frequencies. Thus, the standard error of v_k is calculated only in single input cases.

After the estimates of the impulse response function have been obtained, the order (r,s,b) of the transfer function can be identified and the initial estimates of the parameters involved are obtained according to the relationship between the impulse response function and the transfer function outlined above. In the simulated examples mentioned in Section IV(ii), the impulse response functions are estimated using cross-spectral analysis.

(iv) Identification of the noise model

As discussed in Section II, the noise series u_t of equation (14) can be modelled by an ARIMA model. After the transfer function is identified and initially estimated, the remaining problem in the identification procedure is to identify the noise model.

When the impulse response function is estimated in the time domain, the estimated output can be generated by using the initial estimates of the transfer function parameters in the identified model. Then, the noise series can be generated by subtracting the estimated output from the observed output.

$$u_t = y_t - \frac{A_0(B)}{C_0(B)} X_{t-b} \qquad (38)$$

where A_0, C_0 denote the initial estimates of the parameters.

Then, the autocorrelations and the partial autocorrelations of the estimated noise series can be obtained and used to identify the ARIMA noise model.

When cross-spectral analysis is used to estimate the impulse response function, the noise autocovariances can be derived by taking the inverse Fourier transform from the noise spectrum $SP_n(h)$.

$$C_{nn}(k) = (1/mlag) \sum_{h=0}^{mlag} e_p \, SP_n(h) \cos(hk\pi/mlag) \qquad (39)$$

where $e_p = 0.5$ for $h=0$, mlag and $e_p = 1$ otherwise.

The noise autocorrelations and partial autocorrelations can be calculated from the estimated noise autocovariances and the noise ARIMA model can then be identified.

Since we are only using the initial estimates of the parameters in this procedure, the noise ARIMA model has to be checked and modified, if necessary, during the actual estimation of the model parameters.

V ESTIMATION OF THE MODEL PARAMETERS

Let us consider a simple example of a transfer function model with a single-input and non-seasonal noise model of the form

$$y_t = \frac{A(B)}{C(B)} x_{t-b} + u_t \qquad (40a)$$

and $u_t = \theta(B)\phi(B)^{-1} a_t$

Equation (40a) can also be written as

$$y_t = \frac{A(B)}{C(B)} x_{t-b} + \frac{\theta(B)}{\phi(B)} a_t$$

$$C(B)\phi(B)y_t = A(B)\phi(B)x_{t-b} + \theta(B)C(B)a_t \qquad (40b)$$

For example, the polynomials $A(B)$, $C(B)$, $\theta(B)$, and $\phi(B)$ are all first order polynomials, then

$$(1-C_1B)(1-\phi_1B)y_t = (A_0-A_1B)(1-\phi_1B)x_{t-b} + (1-\theta_1B)(1-C_1B)a_t$$

$$y_t - (C_1+\phi_1)y_{t-1} + C_1\phi_1 y_{t-2}$$

$$= A_0x_{t-b}-(A_1+\phi_1)x_{t-b-1}+A_1\phi_1x_{t-b-2}+a_t-(\theta_1+C_1)a_{t-1}+\theta_1C_1a_{t-2}$$

then, $a_t = y_t-(C_1+\phi_1)y_{t-1}+C_1\phi_1y_{t-2}-A_0x_{t-b}+(A_1+\phi_1)x_{t-b-1}-A_1\phi_1x_{t-b-2}$

$$+ (\theta_1+C_1)a_{t-1} - \theta_1C_1a_{t-2} \qquad (41)$$

Alternatively, the residuals a_t can be obtained from a three-stage procedure as suggested by Box and Jenkins (1970).

a) Let $y_t' = \frac{A(B)}{C(B)} x_{t-b}$

For our particular example.

$$y_t' = A_0x_{t-b} + A_1x_{t-b-1} + C_1y_{t-1}$$

Given the initial estimates of A_0 and A_1 and the known values of x_t, y_t' can be obtained from $t = b+1$ onwards.

b) From equation (40), u_t can be obtained by:

$u_t = y_t - y_t'$ for $t = b+1, \ldots, n$

c) The residual a_t can then be calculated from:

$$a_t = \frac{\phi(B)}{\theta(B)} u_t$$

For our particular example.

$$a_t = \theta_1 a_{t-1} + u_t - \phi_1 u_{t-1}$$

Thus a_t can be obtained from $t = b+2$ onwards. Assuming that the residual series $\{a_t\}$ is a normal process with a zero mean, is not autocorrelated for non-zero lags, and is independent of the input series, the parameters can be estimated by minimizing the conditional sum of squares function

$$S = \sum_{t=1}^{n} a_t^2 \, (b, A, C, \phi, \theta \,|x_0, \; y_0, a_0)$$

where x_0, y_0, a_0 are the starting values. In practice, some of these values are unknown at $t = 1$, or so. Thus, the conditional sum of square function becomes

$$S = \sum_{t=m+p+1}^{n} a_t^2 \, (b, A, C, \phi, \theta \,|x_0, y_0, a_0) \qquad (42)$$

where m = the larger of r and $(s+b)$.

When $t=1,\ldots m+p$, x and y values are known. u_t will be available from u_{m+1} onwards. All the unknown a values can be set equal to their expected values of zero.

Since the conditional sum of squares function is non-linear, a non-linear least square algorithm has to be used to estimate the parameters involved. The method of non-linear least square estimation is itself a complicated procedure and is not discussed here. However, there are many computer library subroutines for the method available. The NAG subroutine E04FBF (NAG, 1977) was used by the author to establish a computer program for the estimation of the transfer function model parameters. Readers who wish to investigate the non-linearleast square method further should consult the relevant references.

VI CHECKING THE FITTED MODEL

In the identification procedure, we have specified the form of the transfer function model. The parameters are then estimated by employing a non-linear least square method. It is then necessary to check whether the fitted model is 'adequate' according to several criteria. An 'adequate' model has to satisfy the following requirements:-

a) It involves a small number of parameters.

b) The transfer function component represents a stable linear dynamic system. As discussed in section I(ii) above, for the system's stability, we require the impulse response weights to die down after finite lag k. In other words, the output variable depends on only the present and finite past values of the input variable.

c) The noise ARIMA model has to be stationary. To illustrate the concept of stationarity, let us consider an AR(1) model

$$u_t = \phi u_{t-1} + a_t$$

This model represents that u_t depends on its past value u_{t-1} according to the weight ϕ. It means

$$u_{t-1} = \phi u_{t-2} + a_{t-1}$$

$$u_{t-2} = \phi u_{t-3} + a_{t-2}$$

$$\vdots$$

$$u_{t-j} = \phi u_{t-j-1} + a_{t-j}$$

Substituting $u_{t-1}, u_{t-2}, \ldots u_{t-j}$ recursively in the original model,

$$\begin{aligned} u_t &= \phi(\phi u_{t-2} + a_{t-1}) + a_t \\ &= \phi^2 u_{t-2} + a_t + \phi a_{t-1} \\ &= \phi^2(\phi u_{t-3} + a_{t-2}) + a_t + \phi a_{t-1} \\ &= \phi^3 u_{t-3} + a_t + \phi a_{t-1} + \phi^2 a_{t-2} \\ &\vdots \\ &= \phi^j u_{t-j} + a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \ldots + \phi^{j-1} a_{t-j+1} \end{aligned}$$

According to this equation, if $|\phi| < 1$, the value ϕ^{j-1} will die down after a finite lag. It means that u_t depends on the finite past values of the residuals. However, if $|\phi| \geq 1$, the value ϕ^{j-1} will not die down and u_t depends on the infinite past values of the residuals. The process is then called a non-stationary process. Thus, for the requirement of stationarity, we have to constrain the limit of the autoregressive parameters.

d) The noise ARIMA model has to be invertible. Let us consider a MA(1) model.

$$u_t = a_t - \theta a_{t-1}$$

or $$a_t = u_t + \theta a_{t-1}$$

then, $$a_{t-1} = u_{t-1} + \theta a_{t-2}$$

$$\vdots$$

$$a_{t-j} = u_{t-j} + \theta a_{t-j}$$

Substituting $a_{t-1}, \ldots a_{t-j}$ recursively in the original model,

$$\begin{aligned} u_t &= a_t - \theta a_{t-1} \\ &= a_t - \theta(u_{t-1} + \theta a_{t-2}) \\ &= a_t - \theta u_{t-1} - \theta a_{t-2} \\ &= a_t - \theta u_{t-1} - \theta^2(u_{t-2} + \theta a_{t-3}) \\ &\vdots \\ &= a_t - \theta u_{t-1} - \theta^2 u_{t-2} - \ldots \theta^j u_{t-j} - \theta^{j+1} a_{t-j-1} \end{aligned}$$

We see that, if $|\theta| \geq 1$, u_t depends on the infinite past of itself ($u_{t-1}, \ldots u_{t-\infty}$), since the values θ^j do not die down for $j=1,\ldots\infty$. The process is then not invertible. Similar to the requirement of stationarity, we have to constrain the limit of the moving average parameter for the requirement of invertibility.

e) The residuals of the model should not be autocorrelated and should be independent of the input variables.

Based on the above requirements, the following are needed to check the adequacy of the fitted model.

(i) Checking the estimates of the parameters

a) Checking the estimate with its estimated standard error

From the non-linear least square estimation, along with the estimates of parameters, the estimated standard errors of the estimates are also obtained. It is necessary to test if the estimates are significantly different from zero. The value of two standard errors is a commonly used guideline. If the estimates are within the limit of twice their corresponding standard error, they are considered not significant. This means that the model can be represented by fewer parameters. For example, if a model is identified as

$$y_t = A_0 x_{t-b} - A_1 x_{t-b-1} - A_2 x_{t-b-1} - C_1 y_{t-1} + n_t$$

and after the estimation procedure, the estimates of A_2 are found to be smaller than two standard errors, this would suggest that the model should be in the form of:

$$y_t = A_0 x_{t-b} - A_1 x_{t-b-1} - C_1 y_{t-1} + n_t$$

b) Checking the stability of the fitted model

For the stability of a transfer function model, we require that for r=1,

$$-1 < C_1 < 1; \tag{43}$$

and for r=2,

$$C_2 + C_1 < 1$$

$$C_2 - C_1 < 1 \tag{44}$$

$$-1 < C_2 < 1$$

Thus, if the fitted transfer function model is of the order $r \neq 0$, we have to check whether the c parameters satisfy the above mentioned requirement. If the result shows that the assumption of stability is invalid, the model has to be re-identified.

c) Checking the stationarity and invertibility of the noise model

For the stationarity and invertibility of the noise model we require:

for r = 1,

$$-1 < \phi_1 < 1 \tag{45}$$

$$-1 < \phi_1 < 1$$

and for r = 2,

$$\phi_1 + \phi_2 < 1 \qquad \theta_1 + \theta_2 < 1$$

$$\phi_2 - \phi_1 < 1 \qquad \theta_2 - \theta_1 < 1 \tag{46}$$

$$-1 < \phi_2 < 1 \qquad -1 < \theta_2 < 1$$

(ii) Checking the autocorrelation patterns of the residuals a_t.

As shown in equation (40), a transfer function model consists of a noise model and a transfer function component. Box and Jenkins (1970) considered the autocorrelations of the residuals and the cross-correlations between the residuals and the input series in two cases:

a) when the transfer function component is correct, but the noise model is incorrect, and

b) when the transfer function component is incorrect.

In the former case, the residuals would not be cross-correlated with the input series but would be autocorrelated. In the latter case, the residuals would be cross-correlated to the input series and also would be autocorrelated. Thus, we have to check the autocorrelations of the residuals first.

If the transfer function model is correctly fitted, the estimated autocorrelations would have zero mean and variance

$$s^2 = \frac{1}{(n-m-p)} \tag{47}$$

Then the value $\pm 2\sqrt{\frac{1}{(n-m-p)}}$ can be used as an approximate guide to the significance of individual autocorrelation estimates. A helpful overall check can be done by a chi-square test. If the model fitted is adequate, the quantity given by

$$Q = (n-m-p) \sum_{k=1}^{mlag} r^2_{aa}(k) \tag{48}$$

where mlag = maximum lag beyond which the impulse response is negligible.

$r_{aa}(k)$ = the autocorrelation estimate of a_t at lag k

is approximately distributed as chi-square with (mlag - p - q) degrees of freedom. Thus, with a certain probability the model can be accepted as adequate if Q is smaller than the theoretical (table) chi-square value.

It the chi-square test shows that the model is inadequate, it would be either the transfer function component or the noise model which is incorrect. Thus, the next stage of checking is to check the cross correlations of the residuals to the input series.

(iii) Checking the cross correlation pattern between the residuals and the prewhitened input series.

Generally, the input series $\{X_t\}$ will be autocorrelated. If the residuals have no cross-correlation with $\{X_t\}$, the estimates of the cross-correlations would have the same autocorrelation function as $\{X_t\}$. This effect can be eliminated if the cross-correlation check is carried out between the residuals and the prewhitened input series $\{\alpha_t\}$. (The prewhitening procedure has already been discussed in Section IV(iiia)).

The estimates of $r_{\alpha a}$ have a variance $^1/_{(n-m-p)}$. Thus the value $\pm 2\sqrt{\frac{1}{n-m-p}}$ can be used as a rough guide to the significance of individual cross-correlations.

An overall check, similar to the chi-square test, can be applied to the cross-correlation function.

The quantity

$$W = (n-m-p) \sum_{k=0}^{mlag} r^2_{\alpha a}(k) \tag{49}$$

is approximately distributed as chi-square with (mlag -r -s) degrees of freedom.

The logical conclusions with the results on the autocorrelation and the cross-correlation checks are listed in Table 2.

Table 2. Logical conclusions of the auto- and cross correlation checks on the fitted transfer function model.

Autocorrelation Check	Cross-correlation Check	Conclusion
Adequate	—	Model adequate
Not adequate	Adequate	Noise model incorrect
Not adequate	Not adequate	(i)Transfer function incorrect or (ii)Transfer function and noise model incorrect

VII AN APPLICATION TO A SOIL TEMPERATURE MODEL - AN EXAMPLE OF A SINGLE-INPUT TRANSFER FUNCTION MODEL.

The transfer of heat through a solid medium has been investigated extensively in a deterministic way (Carslaw and Jaegar, 1962; deWit and van Kenter, 1972). The change of the soil temperature at a particular depth over time basically depends on the temperature gradient between two spatial locations. The process of the transfer of heat in soil is essentially a diffusion process analogous to the illustrative example in Section II. Consider a one-dimensional vertical soil profile. The soil temperature at a particular depth can be expressed deterministically as: (Carslaw and Jaegar, 1962)

$$H_e = -K\frac{dT}{dz}$$

where H_e is the heat flow; K is the thermal conductivity of the soil, T is the soil temperature and z is the depth below the surface.

Also, $$\rho c \frac{dT}{dz} = \nabla H_e$$
$$= \frac{d}{dz}K\left(\frac{dT}{dz}\right)$$
$$= K\frac{d^2T(t)}{d\ z^2}$$

where p and c are the density and the specific heat of the soil. Alternatively, the equation can be written as:

$$\frac{dT}{d_t} = k\frac{d^2T(t)}{d\ z^2} \qquad (50)$$

where $k = {}^K/_{\rho c}$ is called the thermometric conductivity or diffusivity.

This differential equation can be solved according to certain initial and boundary conditions.

Here this process is modelled by using the transfer function technique. This involves modelling the transient variations of the soil temperature at the 10cm depth with the ambient air temperature as the input variable. The

temperature data are obtained from a project being carried out jointly by Dr. J.B. Thornes, Dr. M.J. Waylen and the author, at the King's College Field Centre, Rogate, Sussex. The temperature series (°C) contain 144 continuous hourly records starting from 1210 hours, 8 October, 1977. The series is listed in Appendix I. The first 120 observations (Figure 3) are used to construct the transfer function model and the last 24 observations are used in the forecasting procedure after building the model.

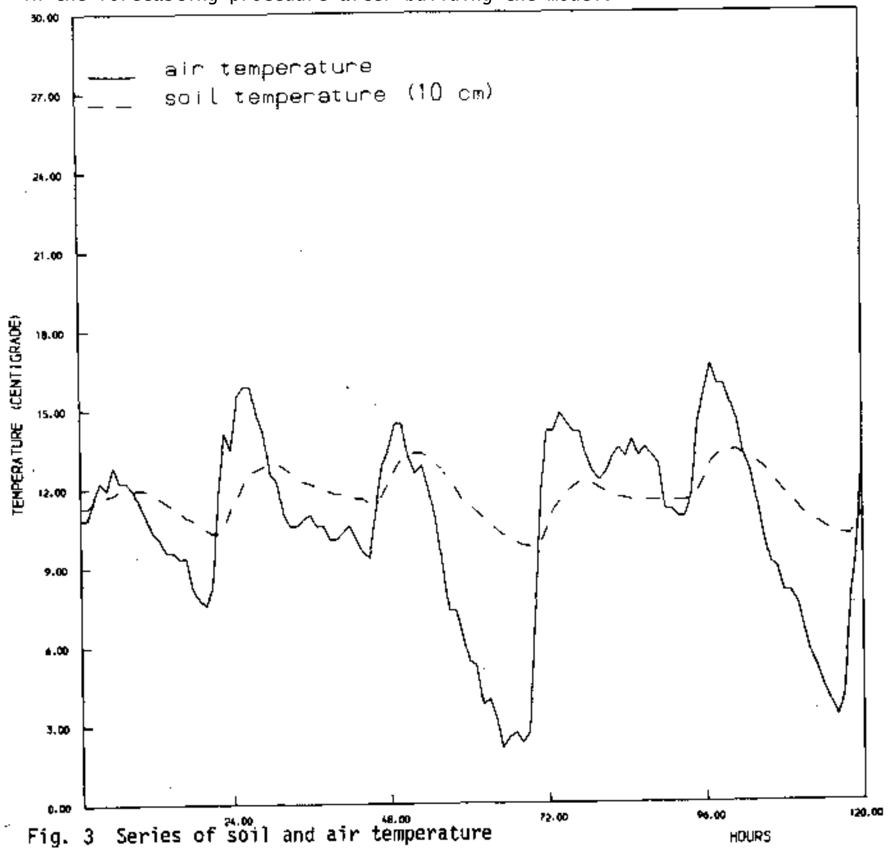


Fig. 3 Series of soil and air temperature

(i) Identification of the model

a) Generating stationary series with zero mean.

The input series (air temperature), $\{x_t\}$, and the output series (soil temperature at 10cm depth), $\{Y_t\}$, are both first-order simple differenced into stationary series with zero means.

i.e. $\nabla y_t = y_t - y_{t-1}$

$\nabla x_t = x_t - x_{t-1}$ for t=2,....n

b) Estimating the impulse response function.

Since the model is a single-input model, the estimation of the impulse response function is carried out in the time domain. The first procedure is to fit an ARIMA model to the input variable. In order to give a simple example of the application of the transfer function modelling method, a non-seasonal model is fitted here. The autocorrelations and the partial autocorrelations of the differenced input series $\{x_t\}$ are calculated for the first 20 lags (Table 3, Figure 4).

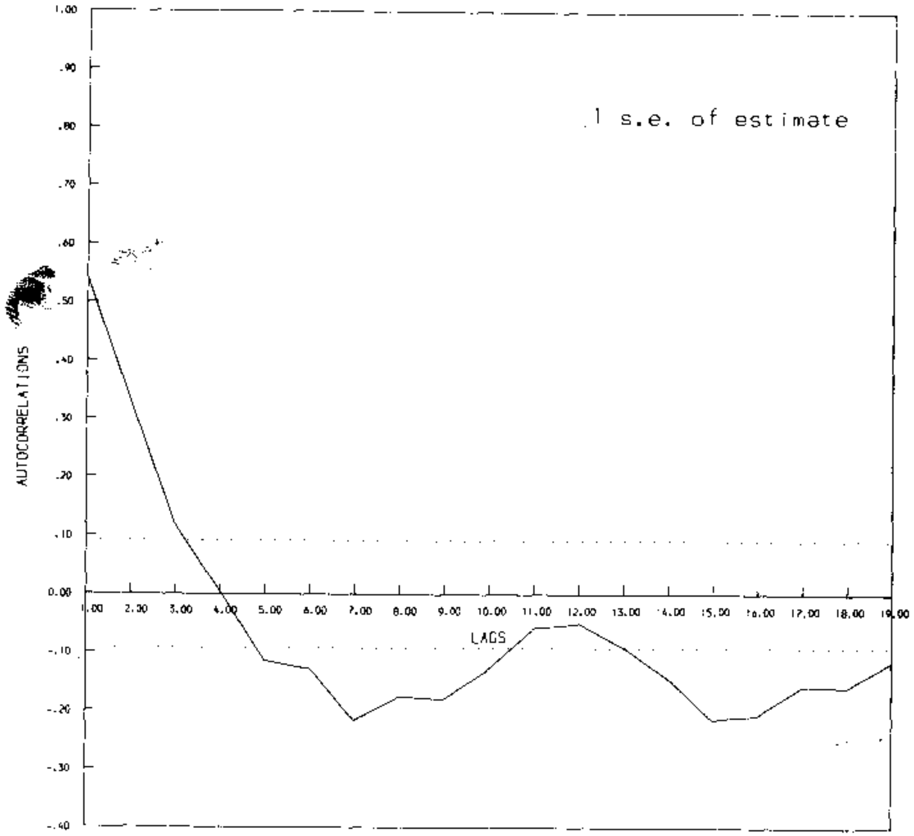


Fig. 4a Autocorrelation estimates of the differenced air temperature series

An AR(1) model is then fitted to the differenced input series.

$$\nabla x_t = .5860\ \nabla x_{t-1} + \alpha_t$$
$$(\pm .0790)$$

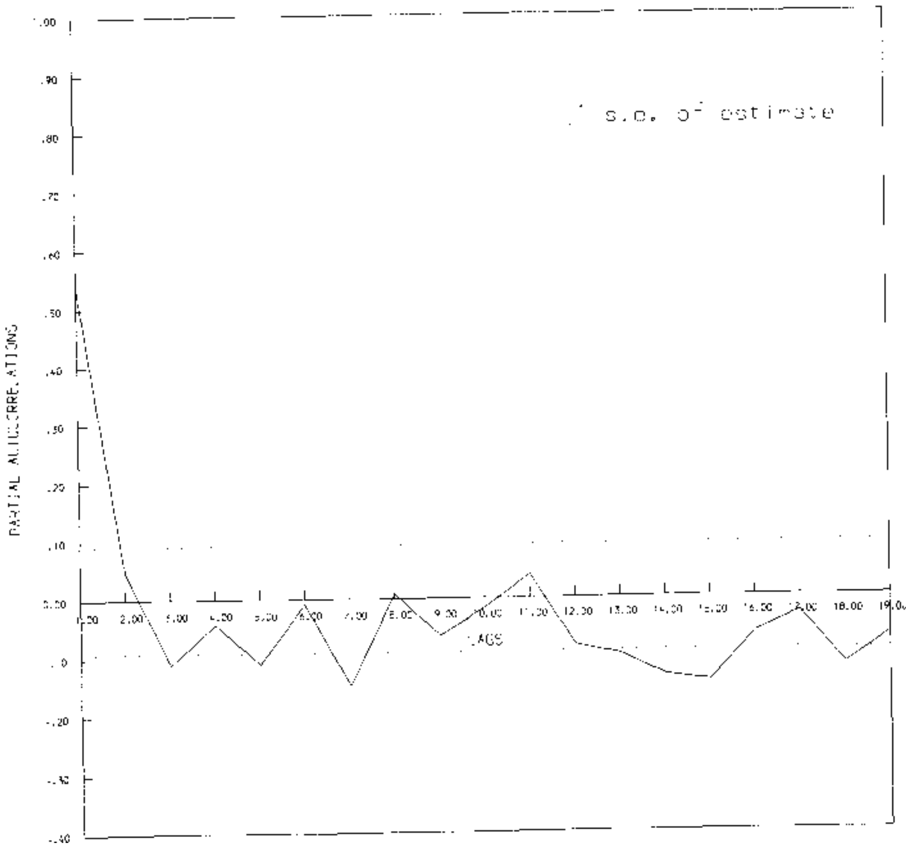


Fig. 4b Partial autocorrelation estimates of the differenced air temperature series

where the figure in brackets is the standard error of the estimate of the Parameter;

$$\text{or} \quad \alpha_t = \nabla x_t - .5860\ \nabla x_{t-1} \qquad (t=3,\ldots n) \qquad (51)$$

The differenced output series is also transformed by the same AR(1) filter:

$$\beta_t = \nabla y_t - .5860\ \nabla y_{t-1} \qquad (t=3,\ldots n\ ) \qquad (52)$$

The cross-covariance, cross-correlation and the impulse response are then estimated according to equations (20) to (23), and are listed in Table 4. The impulse response estimates are plotted in Figure 5.

Table 3. Autocorrelation and partial autocorrelation estimates of the differenced air temperature series

Lag	Autocorrelation	Partial Autocorrelation	Standard Error (±)
1	.5408	.5408	.0917
2	.3263	.0478	.0917
3	.1155	-.1111	.0917
4	.0046	-.0426	.0917
5	-.1140	-.1106	.0917
6	-.1274	-.0083	.0917
7	-.2157	-.1477	.0917
8	-.1749	.0082	.0917
9	-.1792	-.0643	.0917
10	-.1281	-.0170	.0917
11	-.0578	-.0391	.0917
12	-.0501	-.0820	.0917
13	-.0900	-.0980	.0917
14	-.1435	-.1347	.0917
15	-.2140	-.1448	.0917
16	-.2072	-.0632	.0917
17	-.1570	-.0288	.0917
18	-.1590	-.1179	.0917
19	-.1155	-.0662	.0917
20	-.1076	-.1213	.0917

c) Identification of the transfer function component

By inspecting the pattern of the impulse response estimates, we can consider:

$v_0 = 0$

$v_2, v_3, \ldots$ follow the pattern of a first order difference equation with starting value v_2.

v_1 does not follow this pattern

According to equation (17) and the discussion in section IV(ii), we can identify the transfer function model as being of the order (1,1,1), i.e., r=1, s=1, b=1, and in the form

$$\nabla y_t = \frac{(A_0 - A_1 B)}{(1 - C_1 B)} \nabla x_{t-1} + u_t$$

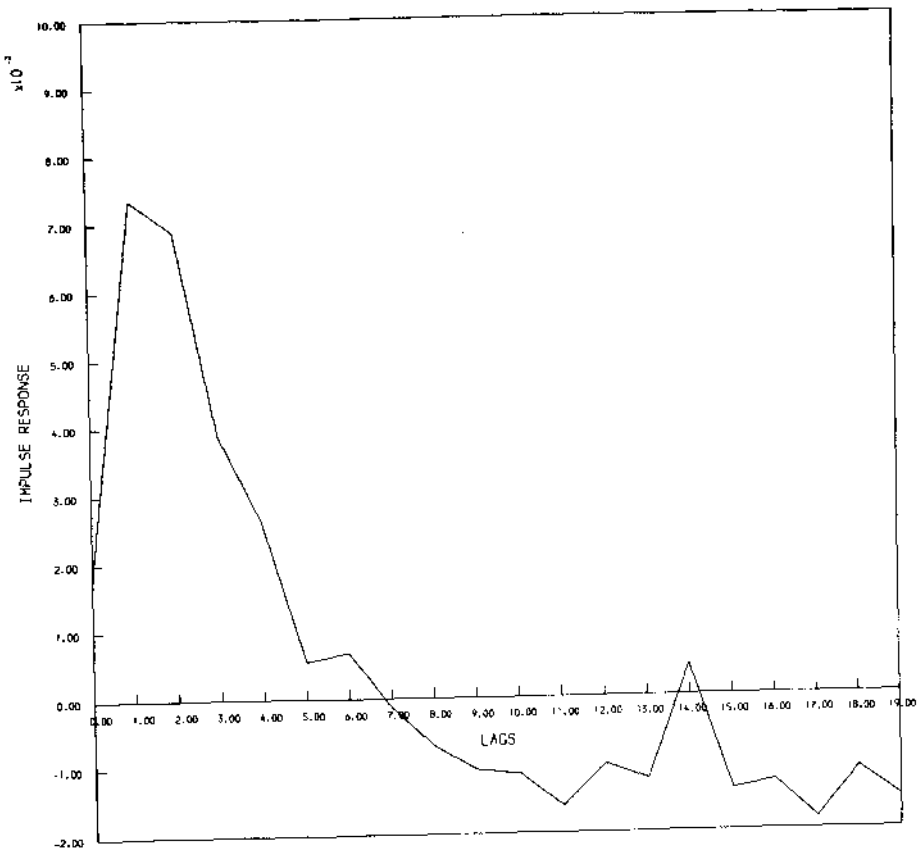


Fig. 5 Impulse response estimates between air temperature (input) and soil temperature (output)

Using equation (17) and letting b=1, r=1, s=1,

$v_0 = 0$

$v_1 = C_1 v_0 + A_0 = A_0$

$v_2 = C_1 v_1 - A_1$

$v_3 = C_1 v_2$

Then, we can initially estimate the parameters A_0, A_1, and C_1 by using the estimates of v_1, v_2 and v_3.

$A_0 = v_1 = .0735$

$C_1 = v_3/v_2 = .5617$

$A_1 = C_1 v_1 - v_2 = -.0276$

Table 4, Cross covariance, cross correlation and impulse response estimates between the prewhitened air temperature series and the transformed soil temperature series

Lag	Cross Covariance	Cross Correlation	S.E. of Cross Correlation (±)	Impulse Response
0	.0173	.1183	.0921	.0180
1	.0705	.4824	.0925	.0735
2	.0660	.4519	.0928	.0689
3	.0371	.2539	.0933	.0387
4	.0250	.1709	.0937	.0260
5	.0053	.0360	.0941	.0055
6	.0065	.0443	.0945	.0068
7	-.0012	-.0085	.0949	-.0013
8	-.0068	-.0468	.0953	-.0071
9	-.0101	-.0693	.0958	-.0106
10	-.0108	-.0742	.0962	-.0113
11	-.0155	-.1058	.0967	-.0161
12	-.0096	-.0658	.0971	-.0100
13	-.0118	-.0809	.0976	-.0123
14	.0042	.0289	.0981	.0044
15	-.0134	-.0919	.0985	-.0140
16	-.0122	-.0838	.0990	-.0128
17	-.0177	-.1209	.1000	-.0184
18	-.0106	-.0723	.1005	-.0110
19	-.0148	-.1010	.1010	-.0154

$S_\alpha = .9792$

$S_\beta = .1492$

Thus, the model is identified as:

$$\nabla y_t = \frac{(.0735 + .0276B)\nabla x_{t-1}}{(1 - .5617B)} + u_t$$

d) Identification of the noise model

According to equation (38) and the model identified above, the noise series $\{u_t\}$ can be obtained. The autocorrelations and partial autocorrelation estimates are then calculated (Table 5, Figure 6). The noise model is identified as an AR(1) model

Table 5. Autocorrelation and partial autocorrelation estimates of the noise series from the initial soil temperature model

Lag	Autocorrelation	Partial Autocorrelation	Standard Error (±)
1	-.1742	-.1742	.0933
2	.1276	.1003	.0933
3	.0353	.0759	.0933
4	.0361	.0427	.0933
5	-.0035	-.0046	.0933
6	-.0890	-.1077	.0933
7	-.0308	-.0705	.0933
8	.0615	.0722	.0933
9	-.1702	-.1309	.0933
10	.0156	-.0383	.0933
11	-.1857	-.1733	.0933
12	-.0506	-.1161	.0933
13	-.0593	-.0464	.0933
14	-.0915	-.0727	.0933
15	-.0806	-.1203	.0933
16	-.0138	-.0589	.0933
17	.0648	.0619	.0933
18	-.0649	-.0842	.0933
19	-.0182	-.0690	.0933
20	.1377	.0689	.0933

$$u_t = -.1742u_{t-1} + a_t$$

or $$u_t = a_t / (1 + .1742B)$$

Then, the identified transfer function model is:

$$\nabla y_t = \frac{(A_0 - A_1B)\nabla x_{t-1}}{(1 - C_1B)} + \frac{a_t}{(1 + \phi_1 B)}$$

It can be written as:

$$y_t = \frac{(A_0 - A_1B)}{(1 - C_1B)} x_{t-1} + \frac{a_t}{(1-B)(1-\phi_1 B)}$$

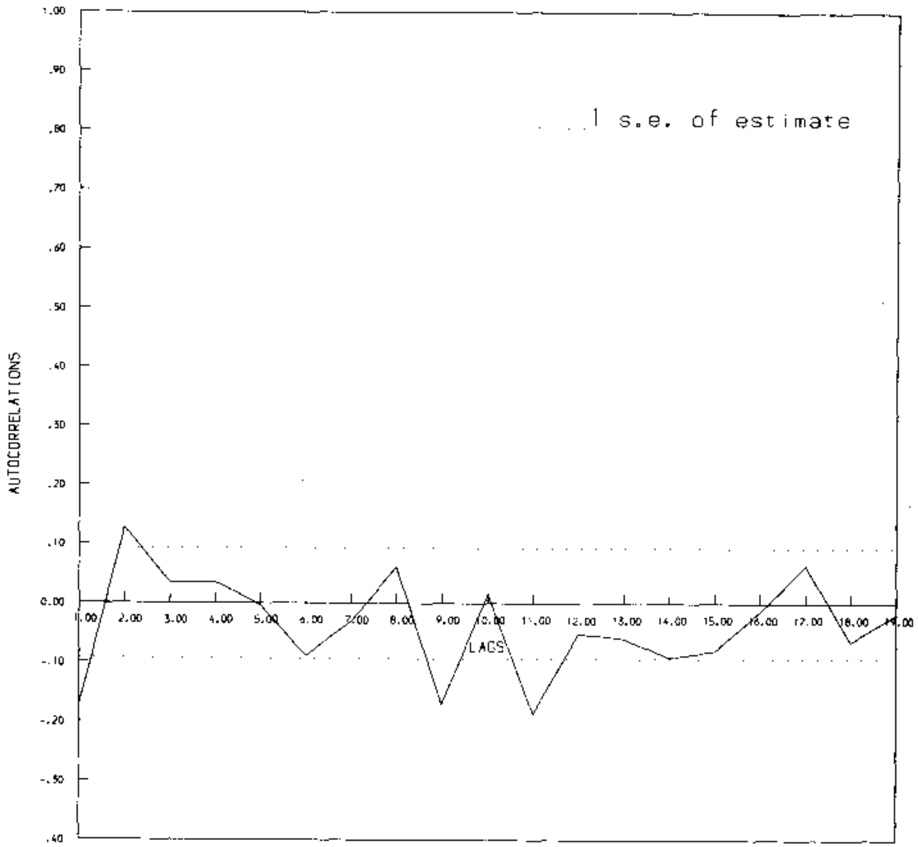


Fig 6a. Autocorrelation estimates of the noise series from the initial soil temperature model

With the initial value of the parameters:

A_0 = .0735

A_1 = -.0276

C_1 = .5617

ϕ_1 = -.1742

(ii) Estimation of the Model Parameters

Using the observations of x_t and y_t and the initial values of the parameters A_0, A_1, C_1 and ϕ_1, the residual A_t can be obtained. Using the non-linear least square method, the estimates of A_0, A_1, C_1 and ϕ_1 can be obtained iteratively until the residual sum of square has been minimized.

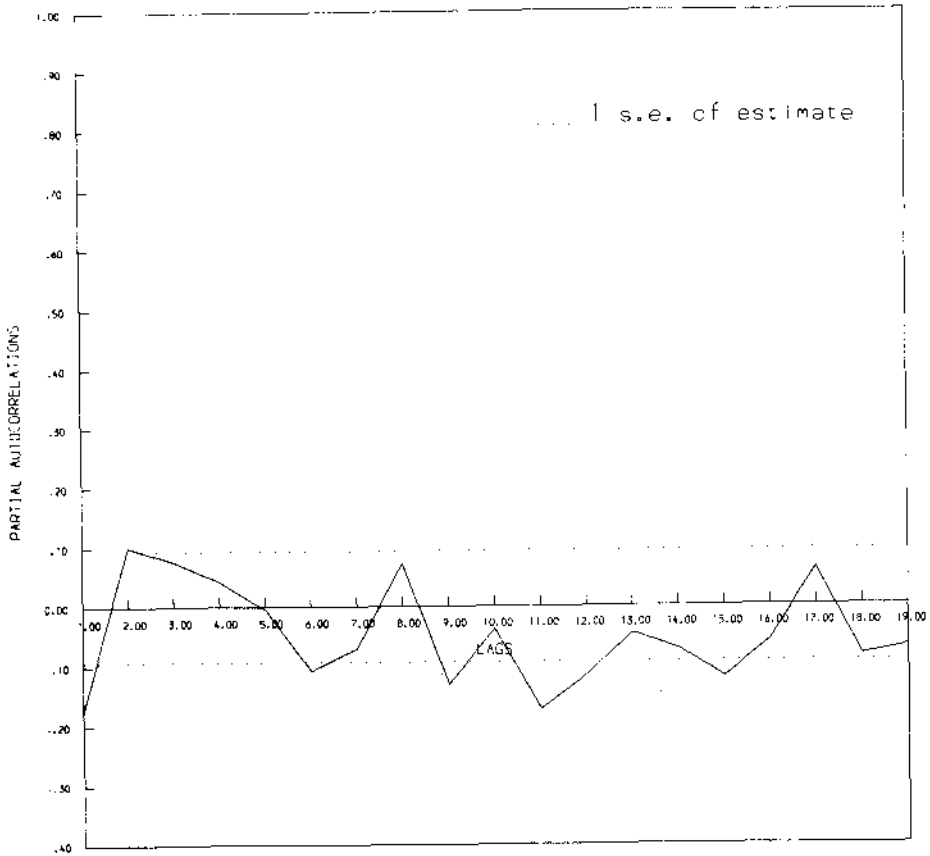


Fig. 6b Partial autocorrelation estimates of the noise series from the initial soil temperature model

The parameters are finally estimated as:

A_0 = .0846 (± .0064)

A_1 = -.0214 (± .0088)

C_1 = .6388 (± .0356)

ϕ_1 = -.2597 (± .0730)

The values in brackets are the standard error of the estimates.

(iii) Checking the Fitted Model

a) Checking the estimate of the parameters

Comparing the estimates of the parameters to their corresponding standard error, they are all significant. Since

$$-1 < C_1 = .6388 < 1$$

$$-1 < \phi_1 = -.2597 < 1,$$

the requirements of model stability and of noise stationarity are both satisfied.

b) Checking the autocorrelation of patterns of the residuals

The autocorrelation estimates of the residuals are calculated. (Table 6, Figure 7).

There is no significant pattern in the autocorrelation estimates of the residual series. An overall chi-square test shows:

$$Q = 11.4476$$

with the degree of freedom = mlag - P = 20 - 1 = 19. With a probability of .95 and 19 degrees of freedom, the theoretical chi-square is 30.144. Since the calculated chi-square is smaller than the theoretical value, there is no evidence of model inadequacy.

c) Checking the cross-correlation pattern between the residuals and the prewhitened input series.

The cross correlation estimates ($r_{\alpha a}$) are calculated (Table 7, Figure 8). The overall chi-square test shows:

$$W = 10.5830$$

with degrees of freedom = mlag - r - s = 20 - 1 - 1 = 18.

With the probability .95 and degrees of freedom 18, the theoretical chi-square is 28.969. Thus, the overall chi-square test shows no model inadequacy either.

Table 6. Autocorrelation estimates of the residuals from the fitted soil temperature model

Lag			Autocorrelations		
1 - 5	.0030	.0353	.0581	.0643	-.0177
6 - 10	-.1330	-.0478	.0737	-.1047	.0221
11 - 15	-.1347	-.0374	-.0417	-.0622	-.0685
16 - 20	.0310	.1006	-.0356	-.0067	.0984

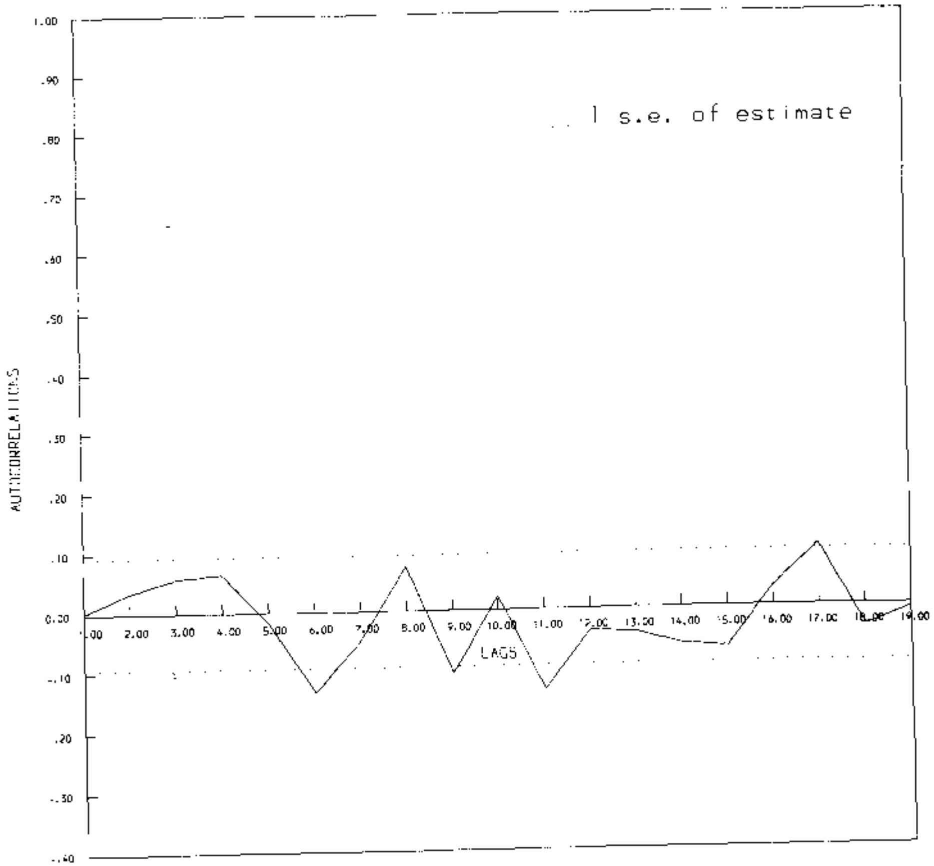


Fig. 7 Autocorrelation estimates of the residuals from the fitted soil temperature model

Table 7. Cross-correlation estimates between the residuals and the prewhitened input series

Lag			Autocorrelations		
0 - 4	.1861	-.0000	.0111	-.0010	.0441
5 - 9	-.1031	.0361	-.0195	.0464	)0249
10 - 14	.0198	-.0189	-.0161	-.1335	.0484
15 - 19	-.1046	-.0514	-.0362	.0164	-.0320
20	.0360				

Standard error = ± .0933

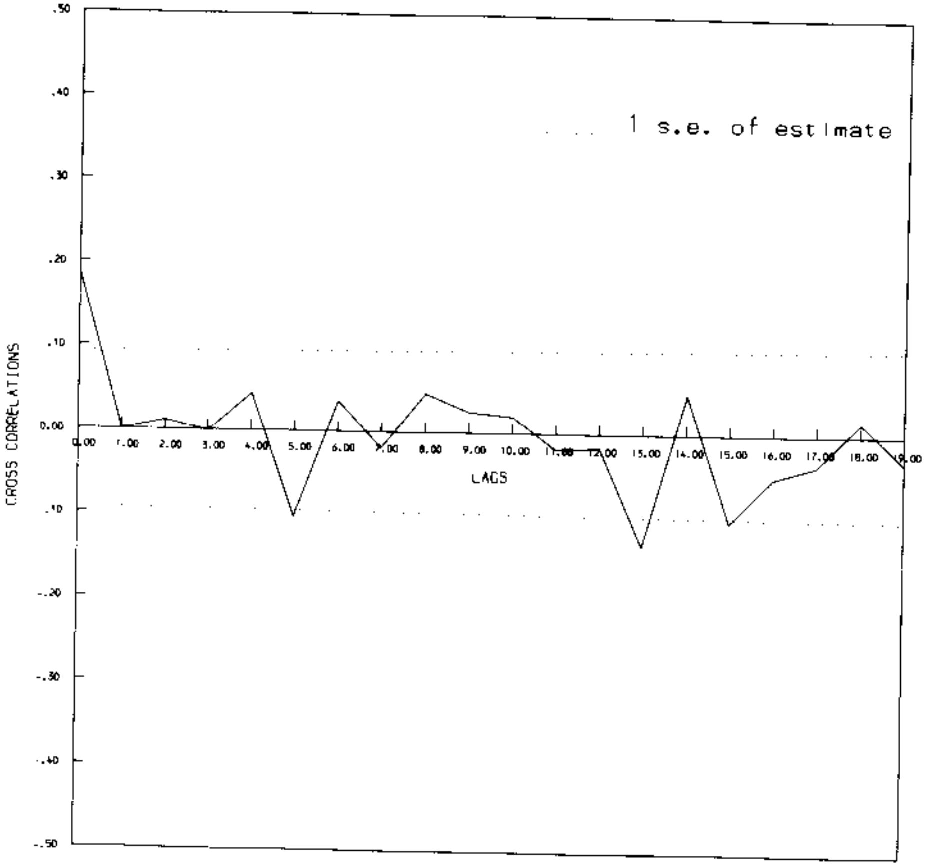


Fig. 8 Cross-correlation estimates between the residuals and the prewhitened air temperature (input) series

(iv) Discussion

The final transfer function model of soil temperature at 10cm depth using air temperature as input variable is:

$$y_t = \frac{(.0846 + .0214B)x_{t-1}}{(1 - .6388B)} + \frac{a_t}{(1-B)(1 + .2597B)} \qquad (53)$$

From this, we can extract some interesting characteristics:

a) there is a 1-hour response delay between air temperature input and soil temperature (10cm) output.

b) As r = 1, it shows the general form of the transfer function model of a simple diffusion process. It represents a first order difference equation

(differential equation in its continuous representation) with respect to time.

c) y_t depends on x_{t-2} as well as on x_{t-1}. In equation (50), the term $\frac{d^2T}{dz^2}$ is a second differential of T with respect to vertical distance.

Let T_0, T_1, T_2 be the temperature at the surface, 10cm, and 20cm respectively. The discrete representation of $\frac{d^2T}{dz^2}$ is:

$$\nabla_z^2 T = (T_0 - T_1) - (T_1 - T_2) = T_0 - 2T_1 + T_2$$

i.e., the second order differencing with respect to z (space) analogous to the second order differencing with respect to time.

This means that the soil temperature (10cm) at time t depends on the temperature gradient between air surface and 10cm ($x_t - y_t$) and also depends on the temperature gradient between 10cm and 20cm. However the model only includes air temperature as an input variable. As we found that there is a response delay of 1 hour on the soil temperature at 10cm to the air temperature, we expect the response of the soil temperature at 20cm depth to air temperature would have a delay of 2 hours. Thus, we could interpret the effect of x_{t-2} on y_t as the effect of soil temperature at 20cm depth.

VIII FORECASTING THE OUTPUT OF A TRANSFER FUNCTION MODEL

One of the purposes of modelling input-output relationships is to predict or forecast the output based on the known model structure. As discussed above, a transfer function model is able to cope with the transient behaviour of the output series in response to the variation on the input series. Thus, after an adequate transfer function model has been built, we can forecast the future output according to the known or forecasted values of the input variables and of the output variable itself. Since a transfer function model represents a stochastic-dynamic model of the transient input-output relationship, it usually gives good forecasts, especially short-term forecasts on the output variable. Box and Jenkins (1970) point out that the transfer function model forecasts also usually have smaller forecasting errors than the forecasts based on univariate models (e.g. an ARIMA model) based on the output variable itself. Thus, a usual practice after building a transfer function model is to forecast the output variable. Let us consider a single-input, non-seasonal noise model, Equation (40) can be represented as:

$$c(B)\phi(B)y_t = A(B)\phi(B)x_{t-b} + \theta(B)c(B)a_t \qquad (54)$$

This can be written as:

$$F(B)y_t = G(B)x_{t-b} + R(B)a_t$$

or

$$y_t - F_1 y_{t-1} + \ldots + F_{p+r} y_{t-p-r} = G_0 x_{t-b} - \ldots - G_{p+s} x_{t-b-p-s}$$

$$+ a_t - R_1 a_{t-1} - \ldots - R_{q+r} a_{t-q-r} \qquad (55)$$

Suppose we are at time instance t. x_1, y_1; x_2, y_2;.....; x_t, y_t are known. The forecast of y at time $t+\ell$ (denoted by $\hat{y}_t(\ell)$ and called the ℓ-steps ahead forecast) can be obtained from:

$$\hat{y}_t(\ell) = F_1(y_{t+\ell-1}) + \ldots + F_{p+r}(y_{t+\ell-p-r}) + G_0(x_{t-b+\ell}) - \ldots -$$

$$G_{p+s}(x_{t-b-p-s+\ell}) + (a_{t+\ell}) - R_1(a_{t+\ell-1}) - \ldots - R_{q+r}(a_{t-q-r+\ell})$$

where

$$(y_{t+j}) = \begin{cases} y_{t+j} & j \le 0 \quad \text{(observed)} \\ \hat{y}_{t(j)} & j > 0 \quad \text{(forecasted)} \end{cases} \qquad (56)$$

$$(x_{t+j}) = \begin{cases} x_{t+j} & j \le 0 \quad \text{(observed)} \\ \hat{x}_{t(j)} & j > 0 \quad \text{(forecasted)} \end{cases}$$

$$(a_{t+j}) = \begin{cases} a_{t+j} & j \le 0 \quad \text{(observed)} \\ 0 & j > 0 \quad \text{(expected value)} \end{cases}$$

The forecasts of the input variable $x_t(j)$ can be obtained by using other forecasting methods on $\{x_t\}$, such as an ARIMA forecasting method.

This procedure can now be applied to the fitted soil temperature model as an example of the transfer function model forecasting. The model is fitted using the first 120 observations on soil and air temperature. Recall the fitted model represented as equation (53).

Rewriting this in the same way as equation (55),

$$(1 - .6388B)(1-B)(1 + .2597B)y_t$$
$$= (.0846 + .0214B)(1-B)(1 + .2597B)x_{t-1} + (1 - .6388B)a_t$$
$$y_t = 1.3197y_{t-1} - .2132y_{t-2} - .1659y_{t-3} + .0846x_{t-1}$$
$$- .0412x_{t-2} - .0378x_{t-3} - .0056x_{t-4} + a_t - .6388a_{t-1} \qquad (57)$$

x_t and y_t were observed at time t=1,...120. Suppose we want to perform a 1-step ahead forecast of y at t=120, i.e. $y_{120}(1)$, according to equations (56) and (57). The values required are:

$x_{t+\ell-b} = x_{120}$	$y_{t+\ell-2} = y_{119}$
$x_{t+\ell-b-1} = x_{119}$	$y_{t+\ell-3} = y_{118}$
$x_{t+\ell-b-2} = x_{118}$	$a_{t+\ell} = a_{121}$
$x_{t+\ell-b-3} = x_{117}$	$a_{t+\ell-1} = a_{120}$
$y_{t+\ell-1} = y_{120}$	

Except a_{121}, all the required values are known. a_{121} can be set to equal zero which is the mean expected value of the residuals. For 2 steps ahead forecast to y_t at t=120, the unknown but required values are: x_{121}, y_{121}, a_{121}, a_{122}. a_{121} and a_{122} can be set to zero. x_{121} is the 1-step ahead forecast of x_t at t=120. (i.e., $x_{120}(1)$), and y_{121} is $y_{120}(1)$. The forecasts of x_t are obtained by using the ARIMA forecasting method based on the AR(1) model fitted (equation (51)).

At time t=121, we have now observations on x_{121} and y_{121}. These values can be used to compare with the forecasted values. We can then carry out the 1-step and 2-steps ahead forecast at t=121.

Since we have observations on x_t and y_t from t=121,...144, the 1-step and 2-steps ahead forecast of x and y were carried recursively at t=121,..., 144. The values of forecast and observations of x_t and y_t (t=121,144) are listed in Table 8 and graphically presented in Figure 9.

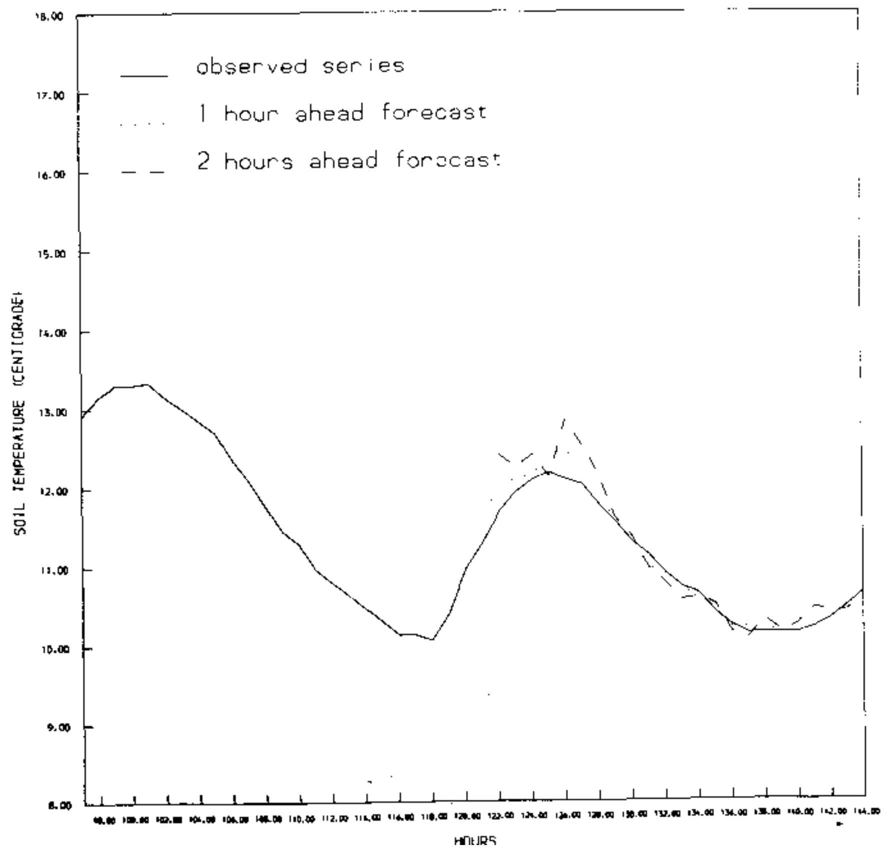


Fig. 9 Transfer function forecasting of the soil temperature model

Table 8. Forecasts and new observations of soil and air temperature

t	x_t	$x_{t(1)}$	y_t	$y_{t(1)}$	$y_{t(2)}$
121	13.14	15.23	11.29	11.69	12.40
122	13.78	13.14	11.67	11.92	12.24
123	12.53	14.15	11.93	12.11	12.40
124	15.49	11.80	12.08	12.13	12.13
125	14.78	17.23	12.18	12.41	12.83
126	13.78	14.37	12.08	12.45	12.52
127	11.66	13.19	12.02	12.20	12.13
128	10.84	10.42	11.77	11.87	11.60
129	9.33	10.36	11.55	11.56	11.35
130	8.64	8.45	11.29	11.25	10.96
131	7.56	8.24	11.13	11.00	10.78
132	7.77	6.93	10.90	10.81	10.56
133	7.56	7.89	10.73	10.70	10.60
134	5.97	7.44	10.63	10.59	10.50
135	5.79	5.04	10.40	10.41	10.15
136	7.15	5.69	10.23	10.21	10.07
137	6.95	7.95	10.13	10.21	10.31
138	7.47	6.83	10.13	10.16	10.14
139	8.64	7.78	10.13	10.18	10.24
140	8.87	9.33	10.13	10.28	10.45
141	9.10	9.01	10.20	10.30	10.39
142	9.82	9.24	10.30	10.32	10.41
143	11.38	10.24	10.47	10.44	10.57
144	14.10	12.29	10.63	10.69	10.95

IX DISCUSSION AND CONCLUSION

In the above section, the concept, methodology and applications of the transfer function modelling techniques have been described in detail. The investigations of spatial structures and spatial processes are the two major aspects in geographical studies. Geographers often find themselves dealing with non-equilibrium systems. As noted in Section I, the regression analysis method which has been used widely in various fields of geography since the quantitative revolution in the discipline is inadequate to cope with this

kind of system. The transfer function modelling technique, which recently has been extensively used in econometrics and engineering studies, is introduced here to enable geographers to model the transient input-output relationship of a non-equilibrium system. This input-output relationship is modelled in a dynamic way, instead of the static way as using the regression method, in terms of the transfer function between the input and output variables. The transfer function represents the way the input signals are filtered to become the output signals. The transfer function model is closely related to the deterministic model which represents the dynamic relationship in terms of differential equations. This point was illustrated in Section II, and in the application of the method to the soil temperature model in Section VII. The method also has the advantage that a noise term which represents the random and non-random disturbances to the system is included in the model, and the parameters are estimated under a stochastic framework.

As geographers may also be interested in the development of the system (spatially or temporally), the model can be used to forecast on the output variable. Since the model is built in a stochastic way, the forecasting procedure is also carried out under the stochastic framework. The probability on the forecasting errors can also be acquired. The derivation of these values are not covered in Section VIII. Readers interested in this aspect should consult the references on statistical forecasting and the relevant chapter in Box and Jenkins (1970).

If the forecasts of the output variable project an undesirable development, one may want to change the input of the system to bring the output into the desired limit. If the input variable is not controllable, we may put in an extra control variable in the system in order to control the output variable. The transfer function modelling method has been recently used by engineers to build stochastic-control systems. Thus, a geographical transfer function model could be very useful in application to decision-making and planning practice.

The transfer function model described in the above sections is time-invariant. The parameters in the model are assumed to remain constant over time. This assumption could be untrue if the structure of the system changes with time. This may be caused by a sudden shock on the system such as a national economic disturbance (e.g. a national strike in major economic activities) or changes of the environment surrounding the system (e.g. changes of river bed material). Therefore, one way of extending the existing model is to allow the parameters to vary with time. The structural changes or the environmental changes of the system can then be coped with in the model. The Kalman filtering technique has recently been attempted to combine with other modelling methods, e.g. regression methods, univariate (ARIMA) time series method, to allow the parameters involved to be time-variant (e.g. Bennett, 1976; Harrison and Stevens, 1976). Therefore, a possible extension of the existing transfer function modelling method is to combine it with the Kalman filtering technique to build a time-variant transfer function model.

Another limitation of the existing transfer function modelling method is that it is able to model one functional relationship between an input and an output. Suppose we have an input series $\{x_t\}$ with strong seasonality with lag 12. The existing method is able to build a transfer function between $\{x_t\}$ and the output $\{y_t\}$. However, the values of x_{t-12} may also have functional relationship with y_t analogous to the seasonal effect in an ARIMA model.

Geographers often investigate series with periodicity. The existing method can only model the input-output relationship within a period, but is unable to model the relationship between periods. The possible extension to the existing method is to enable the model to cope with inter-period relationships as well as the intra-period relationships in a similar approach as the multiplicative seasonal ARIMA model.

The existing method is for building a one-dimensional temporal or spatial model. The idea of extending the model into the spatio-temporal model has already been proposed by Bennett (1975), Bennett and Haining (1976) and Martin and Oeppen (1975). However, there are difficulties in the extension methodologically, especially in the identification procedure. Further work on this aspect is needed.

There are some difficulties which geographers may encounter in using the method. First, the method requires a continuous sequence of observations on the variables. In practice, we may have gaps in the series. Thus, the gaps have to be filled with some reasonable method before using the technique. Second, a reasonably long series (more than 50 observations) is required. There may be a problem in applying the method in the fields of physical geography due to the large number of instruments needed for measurement.

Despite the limitations and difficulties in using the method mentioned above, the method is thought to be a very useful one in geographical modelling.

BIBLIOGRAPHY

A. On applications

Bennett, R.J. (1974), Process identification for time series modelling in urban and regional planning. *Regional Studies,* 8, 157-174.

Bennett, R.J. (1975a), The representation and identification of spatio-temporal systems: an example of population diffusion in North West England. *Transactions IBG,* 66, 73-94.

Bennett, R.J. (1975b), Dynamic systems modelling of the North West Region. *Environment and Planning A,7,* 525-538, 539-566, 617-636.

Bennett, R.J. (1976), Adaptive adjustment of channel geometry. *Earth surface Processes,1,* 131-150.

Bennett, R.J. and R.P. Haining (1976), Space-time models: an introduction to concepts. *occasional Papers,* Department of Geography, University College London, 28.

Chow, V.T. and S.J. Kareliotis (1970), Analysis of stochastic hydrologic systems. *Water Resources Research,* 6(6), 1569-1582.

Clarke, R.J. (1971), The representation of a short-period of experimental catchment data by a linear stochastic difference equation. *International Symposium on Mathematical Models in Hydrology,* Warsaw.

Gudmundsson, G. and G. Sigbjarnarnson (1972), Analysis of Glacier runoff and meteorological applications. *Journal of Glaciology,ll,* 303-318.

Lai, P.W. (1977), Stochastic-dynamic models for some environmental systems: transfer function approach. *LSE Geography Discussion Paper,* 61.

Martin, R.L. and J.E. Oeppen (1975), The identification of regional forecasting models using space-time correlation functions. *Transactions IBG,* 66, 95-118.

Whitehead, P. and P. Young (1975), A dynamic stochastic model for water quality in part of the Bedford-Ouse river system. *Modelling and simulation of water resource systems,* ed G.C. Vansteenkiste, (North-Holland Publishing Co.)

B. On Methodology

Blackman, R.B. and J. Tukey (1958), *The Measurement of Power Spectra,* (Dover Publications, New York).

Box, G.E.P. and G.M. Jenkins (1970), *Time Series Analysis: Forecasting and* control.(Holden-Day, San Franisco).

Dhrymes, P.J. (1971), *Distributed Lags.* (Holden-Day, San Francisco).

Dixon, W.J. (1967), *BMD Biomedical Computer Programs.* (University of California Press).

Harrison, P.J. and C.F. Stevens (1976), Bayesian forecasting.Journal *of Royal Statistical Society, Series* B, 38(3).

Jenkins, G.M. and D.G. Watts (1969), *Spectral Analysis and its Applications,* (Holden-Day, San Francisco).

Kelejian, H.H. and W.E. Oates (1974) *, Introduction to Econometrics: principles and applications.* (Harper and Row, New York).

Marquardt, D.W. (1963), An algorithm for least-squares estimation of non-linear parameters. *Journal of the Society for Industrial and Applied Mathematics, 11,* 431-441.

Murray, W. (ed) (1972), *Numerical method for unconstrained optimization,* (Academic Press, London).

NAG (1976), *Numerical Algorithms Group Manual.*

Pindyck, R.S. and D.L. Rubinfeld (1976), *Econometric models and Economic Forecasts.* (McGraw-Hill, New York).

Powell, M.J.D. (1968), A FORTRAN subroutine for solving systems of non-linear algebraic equations, Harwell report AERE-R5947, (HMSO).

APPENDIX I. SERIES OF HOURLY OBSERVATIONS ON SOIL AND AIR TEMPERATURE

Hour	Air temp.	Soil temp.			
1	10.8400	11.2877	61	5.2300	11.1255
2	10.8400	11.2877	62	3.7700	10.8967
3	11.6000	11.4489	63	3.9500	10.6328
4	12.2300	11.6090	64	3.2200	10.4665
5	11.9400	11.7046	65	2.1000	10.2319
6	12.8300	11.7681	66	2.4800	10.1308
7	12.2300	11.9262	67	2.6700	9.9614
8	12.2300	11.9262	68	2.2900	9.7910
9	11.9400	11.9262	69	2.6700	9.7568
10	11.3800	11.9262	70	7.1500	9.7225
11	10.8400	11.9262	71	11.6600	9.8934
12	10.3200	11.7681	72	14.1000	10.4665
13	10.0600	11.6090	73	14.1000	11.0930
14	9.5700	11.4489	74	14.7800	11.4489
15	9.5700	11.2877	75	14.4400	11.7681
16	9.3300	11.1255	76	14.1000	11.9891
17	9.3300	10.8967	77	14.1000	12.1770
18	8.2000	10.7981	78	13.1400	12.1770
19	7.7700	10.6328	79	12.5300	12.0832
20	7.5600	10.4665	80	12.2300	11.9262
21	8.2000	10.2991	81	12.5300	11.7681
22	11.6600	10.2991	82	13.1400	11.6090
23	14.1000	10.6328	83	13.4500	11.5771
24	13.4500	11.2877	84	13.1400	11.5451
25	15.4900	11.6090	85	13.7800	11.4810
26	15.8600	12.2392	86	13.1400	11.4489
27	15.8600	12.6095	87	13.4500	11.4489
28	14.7800	12.7011	88	13.1400	11.4489
29	14.1000	12.8530	89	12.8300	11.4489
30	12.5300	12.9135	90	11.1000	11.4489
31	12.2300	12.9135	91	11.1000	11.4489
32	11.0000	12.7924	92	10.8400	11.4489
33	10.5700	12.6095	93	10.8400	11.4489
34	10.5700	12.4867	94	11.6600	11.4489
35	10.8400	12.2392	95	14.4400	11.7681
36	11.0000	12.1770	96	15.4900	12.3014
37	10.5700	12.0832	97	16.6100	12.9135
38	10.5700	11.9262	98	15.8600	13.1537
39	10.0600	11.8631	99	15.8600	13.3025
40	10.0600	11.7681	100	15.1300	13.3025
41	10.3200	11.7681	101	14.4400	13.3322
42	10.5700	11.6728	102	13.1400	13.1537
43	10.0600	11.6090	103	12.5300	13.0039
44	9.5700	11.6090	104	11.3800	12.8530
45	9.3300	11.4489	105	10.0600	12.7011
46	11.1000	11.4489	106	9.1000	12.3633
47	12.8300	11.6728	107	8.8700	12.0832
48	13.4500	12.1770	108	7.9800	11.7681
49	14.4400	12.7011	109	7.9800	11.4489
50	14.4400	13.1537	110	7.5600	11.2877
51	13.1400	13.3025	111	6.5500	10.9623
52	12.5300	13.3025	112	5.6000	10.7981
53	12.8300	13.3025	113	5.0400	10.6328
54	11.9400	13.1537	114	4.3100	10.4665

55	10.8400	12.9135	115	3.7700	10.2991
56	9.1000	12.7011	116	3.2200	10.1308
57	7.3500	12.3324	117	3.9500	10.1308
58	7.3500	12.0205	118	7.5600	10.0632
59	6.3600	11.6090	119	9.5700	10.3997
60	5.4100	11.3845	120	13.1400	10.9623

APPENDIX II. DERIVING THE VARIANCE OF v_k BY KNOWING THE VARIANCE OF $H_{(h)}$

Equation (34) can be represented as:

$$V_k = \frac{1}{mlag} \sum_{k=0}^{mlag} e_p H(h)^{ihk\pi/mlag}$$

where e_p = .5 for k=0 and mlag

= 1. otherwise, h=0,...mlag; k=0,...mlag

Let A = $(1,\ldots,e^{ihk\pi/mlag},\ldots, e^{i\pi h})$ a row vector of order (1xm)

where m = mlag +1,

and C = $\begin{pmatrix} .5H_{(0)} \\ \vdots \\ H_{(h)} \\ \vdots \\ .5H_{(mlag)} \end{pmatrix}$, a column vector of order (mx1)

Thus $v_k = \frac{1}{mlag} AC$

Then, the variance of v_k is:

$$S_v^2 = \frac{1}{mlag^2}\left[A(C - \mu C)\right]\left[(C - \mu C)^* A^*\right] = \frac{1}{mlag^2} A\,cov(C)A^*$$

where * represents the complex conjugate transpose of a complex matrix and μ is the mean of C.

Since $H_{(h)}$'s are mutually uncorrelated (Jenkins and Watts, 1969), the matrix cov(C) is a diagonal matrix with all the non-diagonal terms all zero. Thus, since $e^{ihk\pi/mlag}$, $e^{-ihk\pi/mlag} = 1$,

$$S_v^2 = \frac{1}{mlag^2} \sum_{h=0}^{mlag} e_p^{\,2} S_H^2(h)$$

, where $S^2H_{(h)}$ is the variance of $H_{(h)}$

$$S_v = \sqrt{S_v^2}$$