Plant and Soil XIX, no. 1 August 1963

# CAUSAL SOIL—PLANT RELATIONSHIPS AND PATH COEFFICIENTS

by TH. J. FERRARI

Institute for Soil Fertility, Groningen

## INTRODUCTION

The use of the normal regression model to interpret the relationships between soil factors and plant characteristics such as yield and mineral content is open to certain objections. One of the most important of these is connected with the imperfections and limitations of the regression model used. In such a model the assumption is made that the so called independent factors do not influence each other, in other words a change in one factor does not result in a change in another independent factor. In many cases however, this assumption is not valid; this is particularly the case with investigations into the mineral relationships of plants.

The investigation of Sluijsmans [10] into the relationship between the MgO and $K_2O$ contents of herbage on the one hand and soil fertility and other factors on the other provides one of the many examples in which the use of the regression model is not correct. Data obtained from an uncontrolled experiment in which the soil factors were not changed experimentally (Ferrari [3]) were fitted to a regression model, in which the MgO or the $K_2O$ content of the herbage were assumed to be the dependent variables, and the MgO and $K_2O$ content of the soil, the humus content, the proportion of weeds and the crude-protein content of the herbage the independent or causal factors. This model is illustrated in Figure 1 in which the assumed causal relationships are represented by arrows, directed towards the effect. The influence of the correlations between the independent factors is eliminated in the statistical analysis and the assumption is made in this regression model, that a change in *e.g.* the $K_2O$ content of the soil will result only in a change in the MgO content of the herbage and not in a change in the crude-protein content. This assumption as we know is probably incorrect with the result that the effects of the $K_2O$ content of the

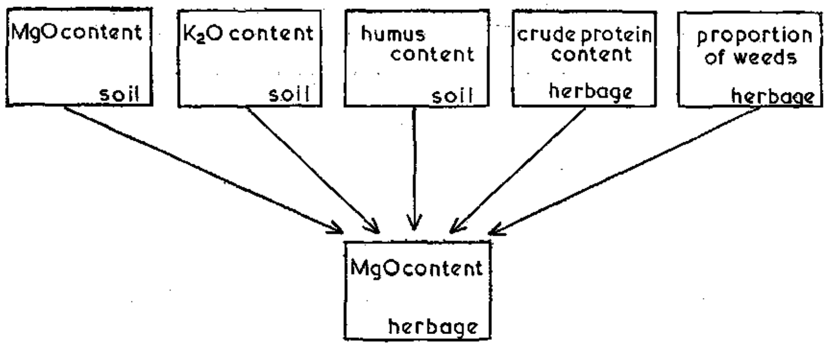


Fig. 1. Regression model with MgO content of the herbage as dependent variable, other variables as independent causal variables.

soil and of the crude-protein content on the MgO content of the herbage may be calculated incorrectly. For practical purposes it is necessary to know the total effect, not only the direct influence of the $K_2O$ content of the soil on the MgO content of the herbage, but also the indirect effects via a change in the crude-protein content and in the proportion of weeds. The use of path coefficients allows among other things the possibility of making allowance for these indirect influences.

The principle of path coefficients was first introduced by Wright [16], and in later papers he further elucidated and worked out his ideas. The method has been used up till now mainly in genetics, although Wright also discusses applications in other disciplines such as physiology. Apart from this however and the example mentioned by Turner and Stevens [14] there appear to have been no other applications outside genetics. Furthermore, the method was not popular in mathematics and genetics for a long time (Kempthorne [5]); in recent years however, it has been applied more and more in genetics. The literature on path coefficients is generally not easy to follow, especially as opinions on certain points frequently diverge. Besides the above mentioned papers, those of Li [7] and Wright [17 19 20] can also be recommended; in addition the articles of Wright [18] and Tukey [13] in "Statistics and Mathematics in Biology" are important.

Economics often requires the analysis of processes in which the variables are both cause and effect and econometrians have applied the method of *simultaneous equations* to analyze these processes. The method of path coefficients is related to this, so that a study of the simultaneous equation techniques is to be recommended. (*cf* Tintner [12], Valavanis [15], Theil and Kloek [11] and Basmann [1]). The ideas of Simon [8 9] and Koopmans [6] concerning identification problems are also very instructive.

The principles of the method will now be discussed briefly with the terminology adapted to ecological problems of soil-plant relationships using data already analyzed by means of a regression model (Sluijsmans [10]).

## PATH COEFFICIENTS

The starting-point is a model consisting of a *closed causal linear system* with $m$ primary causes ($x$) and $n$ effects ($y$). These $m$ and $n$ variables are assumed to be connected with each other by a network of causal paths. We understand by a closed linear system a network in which each variable is a linear combination of one or more other variables in this system or is one of the variables that is determined by none of the variables in this system; the latter are the primary causes, $x$. The fact that a variable is taken as a linear combination of one or more variables, means that this variable can be expressed as a linear function of one of more variables. This is not so for the primary causes, at least in the chosen model. The parameters which give the extents of the influences in these functions are called *path coefficients*.

These path coefficients, which have to be calculated, do not necessarily imply causal relationships; the investigator himself has to formulate the causal network according to already existing knowledge and/or on the basis of the hypothesis he wishes to test. By means of the path coefficients he then tries to determine quantitatively the significance of the assumed primary causes or other variables or to reject eventually, parts or even the whole of the causal model.

It is usual and practical to formulate the causal model by means of a diagram. A causal relationship between two variables is indicated by an arrow connecting cause and corresponding effect and pointing towards the effect. In this diagram the effect is represented as a linear function of one or more variables, $x$ and $y$. Otherwise the assumption of linearity is not necessary.

The calculation of the path coefficients is made on the assumption that the $x$'s and $y$'s are errorless. Another assumption is that the errors of the $y$'s (measurement and response error) have a normal distribution and a mean of zero, and are not correlated with the $x$'s or with the errors of the other $y$'s. The statistical estimation of the path coefficients can now be obtained by the method of least-squares.

The formulation of the path coefficients is so arranged that the total influence of a primary cause $x$ on an effect $y$, expressed as a simple or partial regression coefficient (*total path regression*), is given as a sum of a number of *compound path regressions.* A compound path regression represents the influence of a cause on an effect via one particular chain of variables; it is equal to the product of all the path coefficients in this chain.

We obtain the compound path regression by expressing the ultimate effect $y$, as a function of the variables of which it is the direct result (*structural equations*). When these variables are not in themselves primary causes, they are expressed in their turn as a function of other variables. This process of elimination goes on until the first mentioned ultimate effect $y$, can be expressed solely as a function of primary causes $x$ (*reduced structural equations*). A partial regression equation is then obtained in which the regression coefficients are expressed as a sum of the already mentioned compound path regressions. By writing down the compound path regressions for all effects by these substitutions (or directly by inspection of the path diagram) and equating them to the corresponding computed regression coefficients we obtain a set of simultaneous non-linear estimation equations in which the regression coefficients are known and out of which the path coefficients can sometimes be solved.

An example derived from Turner and Stevens [14] will be used for illustration; the simple causal model is presented in Figure 2. This shows that whilst $y_3$ is determined by the primary causes $x_1$ and $x_2$, the effect $y_4$ is brought about directly only by $x_2$ and $y_3$; the primary cause $x_1$ influences $y_4$ only indirectly via $y_3$. The numbering of the path coefficients p is self explanatory; our experience is that it is easier to use continuous numbering between primary causes and effects.

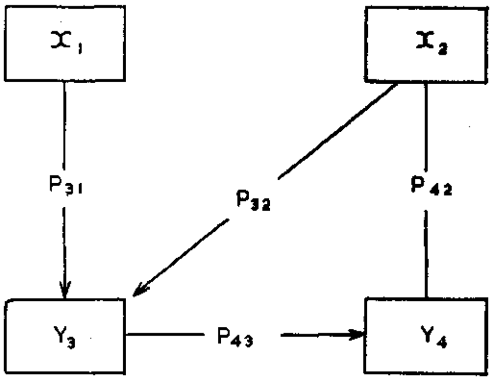


Fig. 2. Closed causal system with $x_1$ and $x_2$ as primary causal factors and an indirect influence of $x_2$ on $y_4$, via $y_3$.

The structural equations that can be written down according to the above-mentioned rules are:

$$y_3 = a_3 + p_{31}x_1 + p_{32}x_2 \qquad (1)$$

$$y_4 = a_4 + p_{43}y_3 + p_{42}x_2 \qquad (2)$$

Equation (1) is already one of the reduced structural equations. The second is obtained by substitution of (1) into (2):

$$y_4 = a_4 + a_3p_{43} + (p_{31}p_{43})x_1 + (p_{42} + p_{32}p_{43})x_2 \qquad (3)$$

These expressions (1) and (3) for $y_3$ and $y_4$ respectively, may also be taken as partial regression equations in which the regression coefficients $b_{31.2}$, $b_{32.1}$, $b_{41.2}$ and $b_{42.1}$ can be calculated. Applying these regression coefficients to the corresponding terms of equations (1) and (3) results in a set of equations, from which the four unknown path coefficients can be solved, namely:

$$b_{31.2} = p_{31} \quad \text{and } b_{32.1} = p_{32} \qquad (4)$$

$$b_{41.2} = p_{31}p_{43} \text{ and } b_{42.1} = p_{42} + p_{32}p_{43} \qquad (5)$$

Thus the coefficients $p_{31}$ and $p_{32}$ are found instantly, after which $p_{42}$ and $p_{43}$ can be calculated.

Such sets of equations are not always soluble. In the example discussed the path coefficients are identifiable because they can be computed uniquely from the regression coefficients. For this it is necessary (but not always sufficient as *e.g.* in cases with *feedback systems*), that the total number of path coefficients be equal to the total number of regression coefficients. However, *over-identification* or *under-identification* will often appear. An extreme case of under-identification is the model used in a factor analysis (Ferrari, Pijl and Venekamp [2]) in which very few regression coefficients are known. For a discussion on the problems of over- and under-identification reference may be made to the above mentioned econometric literature and to the work of Simon [8 9].

## APPLICATION TO AN ECOLOGICAL SOIL-PLANT PROBLEM

The potentialities of the method of path coefficients will be demonstrated using the same data as used by Sluijsmans [10] for determining influences on the MgO content of the herbage.

We first construct a closed causal system as shown in Figure 3, incorporating our ideas on this problem. Attention is drawn to the fact that this method of path coefficients has been used to overcome the objection that in the regression model the porportion of weeds and the crude protein content are included as independent variables; this is not the case in the model of Figure 3 which has been so constructed that both variables are cause as well as effect. In this model, the $K_2O$ and MgO contents of the soil, the humus content and the pH are taken as primary factors. Two of these factors, namely the MgO and $K_2O$ content of the soil, directly affect the MgO content of the herbage via the path coefficients $p_{73}$ and $p_{72}$. The humus content and the pH only indirectly affect the MgO content via the proportion of weeds and/or the crude-protein content. A direct influence of pH and humus content has been excluded because of the difficulty in conceiving how this could possibly be realized. The same argument also applies to the elimination of a causal connection between pH and the crude-protein content. An influence of the humus content on the proportion of weeds and on the crude-protein content is assumed on the basis of a possible influence through the water and nitrogen supplies respectively. The crude-protein content is also dependent on the botanical composition of the grass. It will be clear that the content and form of these causal models are determined primarily by the knowledge, intuition and ideas of the investigator. In any case the validity of the model discussed does not affect the demonstration of the potentialities of path coefficients.

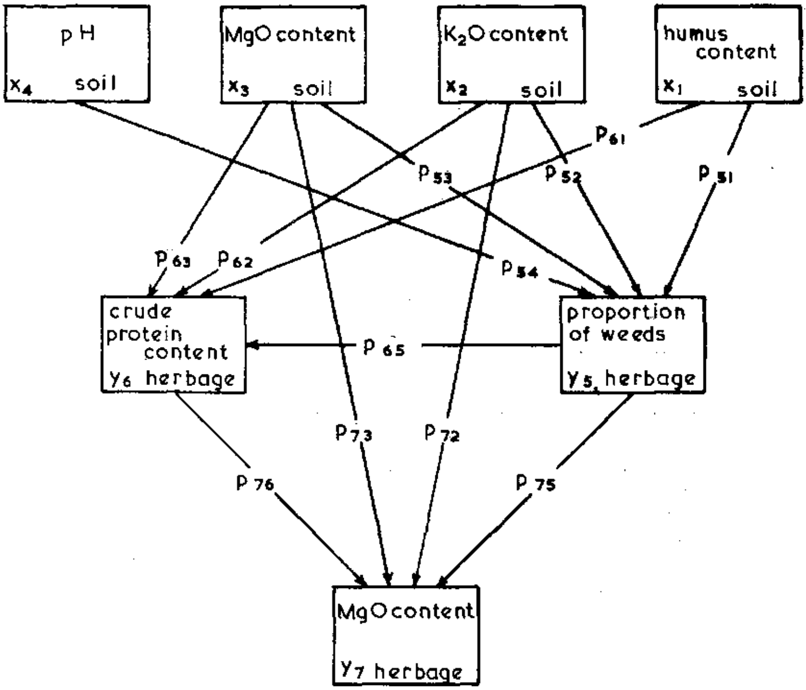


Fig. 3. Direct and indirect influences of the 4 primary causal factors on the MgO content of the herbage.

The influence of the MgO content of the soil on the MgO content of the herbage is directed trough different *paths*. First there is the direct influence via the path coefficient $p_{73}$. There is also an indirect influence represented as a change in the MgO content of the soil directly causing a change in the proportion of weeds and the crude-protein content via $p_{53}$ and $p_{63}$ respectively, which in turn both effect the MgO content of the herbage. In addition the proportion of weeds also influences the crude-protein content. Thus the MgO content of the herbage is influenced by the MgO content of the soil via one direct and three indirect paths. This influence can consequently be expressed by path coefficients as:

$$p_{73} + p_{53}p_{75} + p_{63}p_{76} + p_{53}p_{65}p_{76} \quad (6)$$

Similar formulations can also be derived for the other factors. These are obtained from the appropriate reduced structural equations or directly from the model by inspection. This model is soluble, because, among other things, the number of path coefficients is equal to the number of regression coefficients.

A summary of the data used in this analysis is given in Table 1. These data are helpful in understanding the meaning and value of path coefficients in the whole of the ecological relationships involved. The path coefficient like the regression coefficient gives the rate of change of an effect caused by unit change in the causal variable.

TABLE 1

Lowest, highest and mean values of the variables of the model, Figure 3.

| Variable | Lowest-highest value | Mean value |
|---|---|---|
| Humus content soil . . . . . . | 3.4 — 12,6 | 7.8 |
| $K_2O$ content soil . . . . . . . | 7 — 30 | 16.5 |
| MgO content soil . . . . . . . | 50 —227 | 125 |
| pH soil . . . . . . . . . . . | 4.9 — 5,9 | 5.5 |
| Proportion of weeds . . . . . | 1.1 — 37,5 | 13.5 |
| Crude-protein content herbage | 9.5 — 24,6 | 14.1 |
| MgO content of herbage . . . . | 0.209— 0,383 | 0.293 |
| $K_2O$ content of herbage . . . . | 1.95 — 4,05 | 2.82 |

The path coefficients of the model are given in Table 2.

TABLE 2

Computed values of the 12 path coefficients of the model, Figure 3.

| Effect \ Cause | Humus content ($x_1$) | $K_2O$ content soil ($x_2$) | MgO content soil ($x_3$) | pH ($x_4$) | Proportion of weeds ($y_5$) | Crude protein content ($y_6$) |
|---|---|---|---|---|---|---|
| Proportion of weeds ($y_5$) | 1.67 | −0.23 | −0.031 | 5.26 | | |
| Crude-protein content ($y_6$) | −0.74 | 0.11 | 0.011 | | 0.20 | |
| MgO content of herbage ($y_7$) | | −0.0038 | 0.0004 | | 0.0041 | 0.0083 |

For comparison, in Table 3 we also give the regression coefficients estimated according to the model in Figure 1.

TABLE 3

Regression coefficients computed according to the model, Figure 1.

| Effect \ Cause | Humus content soil | $K_2O$ content soil | MgO content soil | Proportion of weeds | Crude-protein content herbage |
|---|---|---|---|---|---|
| MgO content of herbage | 0.0001 | −0.0038 | 0.0004 | 0.0029 | 0.0059 |
| $K_2O$ content of herbage | −0.0078 | 0.040 | 0.003 | −0.002 | 0.092 |

A strict comparison between the results presented in Tables 2 and 3 is not possible, since the pH has been included as a factor only in the path coefficient model. In spite of this a comparison is still very instructive.

It appears that the regression coefficients and the path coefficients $p_{72}$ and $p_{73}$, taken as a measure of the direct influences of the $K_2O$ and the MgO contents of the soil on the MgO content of the herbage have about the same values. It is striking that the total path coefficients (not given here) are also equal to these values. The direct influences of the proportion of weeds and the crude-protein content however, are very different from those corresponding to the regression coefficients; the ratio is about 1.5. Consequently, practical advice based on the influences of the botanical composition and the crude-protein content of the herbage as indicated by the regression model, would be incorrect.

Table 2 reveals further interesting results as for example, the negative influence of the $K_2O$ and MgO content of the soil on the proportion of weeds, the positive influence of the crude-protein content, the positive total influence (direct plus indirect via the botanical composition) on the crude-protein content, the negative influence of the $K_2O$ content of the soil on the MgO content of the herbage, the unexpected negative influence of the humus content on the crude-protein content *etc.* This last effect of the humus content suggests that with humus, other factors have to be taken into account, not only the supply of nitrogen. Whilst it is beyond the scope of this paper to go further into this problem, the above should be sufficient to demonstrate the potentialities of the method of path coefficients.

## DISCUSSION

Our starting-point was the error introduced by the regression model in which variables were included which were considered independent but which were in fact interdependent. We have demonstrated the potentialities of the method of path coefficients in obtaining a solution to this problem. This method has certain other important aspects, which will be dealt with briefly. Experience in the application of the method to ecological problems is still limited and many questions of a mathematical nature, *e.g.* the testing of the statistical significance of the path coefficients, have still to be solved (Basmann [1]).

We have applied the method to data derived from an experiment in which the (soil) factors were not changed experimentally (*uncontrolled experiment*). The method can also be used for the design and analysis of experiments in which the factors are changed (*controlled experiment*); in this case it may prove very profitable to use the concept of path coefficients. In such a controlled experiment it is usually assumed that the factors left out do not change; in many cases this is not true (Ferrari [3]). The need to design a more comprehensive causal model, the problem of whether or not to take into consideration particular factors *etc.*, all lead to the question as to which observations and measurements are necessary to obtain a picture as close as possible to reality. The model also shows that the effect of a change in a factor can take place in different ways and in the absence of a direct causal relation between two variables. This emphasizes one of the limitations of a controlled experiment and the great importance of defining the factors to be considered.

The method of path coefficient compels the researcher working with data from an uncontrolled experiment to enquire into the way in which way certain phenomena come about in reality and to formulate a hypothesis.

Undoubtedly the method is of great importance in the synthesis of orientated ecological research, although of course the possibilities are not unlimited. The investigation already discussed provided one example but it is possible to proceed still further with the synthesis.

In the model of Figure 3 the $K_2O$ content of the herbage was omitted since a direct causal influence of the $K_2O$ content of the soil was assumed via the path $p_{72}$. Hence no further decision as to the nature of this influence was made. In the model of Figure 4 however, the variable $K_2O$ content of the herbage has been included assuming causal connections between this and the $K_2O$ content of the soil, the proportion of weeds and the crude-protein content and also

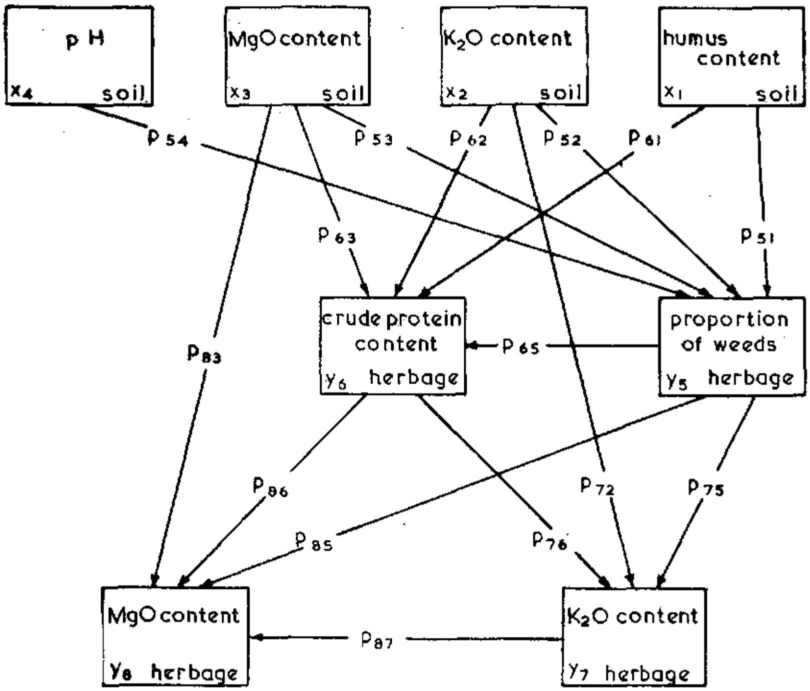


Fig. 4. Direct and indirect influences of the 4 primary causal factors on the MgO content of the herbage, showing the indirect effect of the $K_2O$ content of the soil via the $K_2O$ content of the herbage.

the MgO content of the herbage. This implies that the influence of the $K_2O$ content of the soil on the MgO content of the herbage is of a plant-physiological and not of a soil-chemical nature. This model is an example of over-identification, since the path coefficients $p_{72}$, $p_{75}$ and $p_{76}$ cannot be computed uniquely. A unique solution has however been obtained by introducing a limiting relation, namely that the proportion of weeds does not effect the $K_2O$ content of the herbage (corresponding to that coefficient which differs least from zero). The result is given in Table 4.

TABLE 4

Computed values of the path coefficients for the $K_2O$ content of the herbage according to model, figure 4.

| Effect \ Cause | $K_2O$ content soil ($x_2$) | Proportion of weeds ($y_5$) | Crude protein content ($y_6$) | $K_2O$ content herbage ($y_7$) |
|---|---|---|---|---|
| $K_2O$ content herbage ($y_7$) . | 0.029 | −0.007 | 0.184 | |
| MgO content herbage ($y_8$) . | | | | −0.1307 |

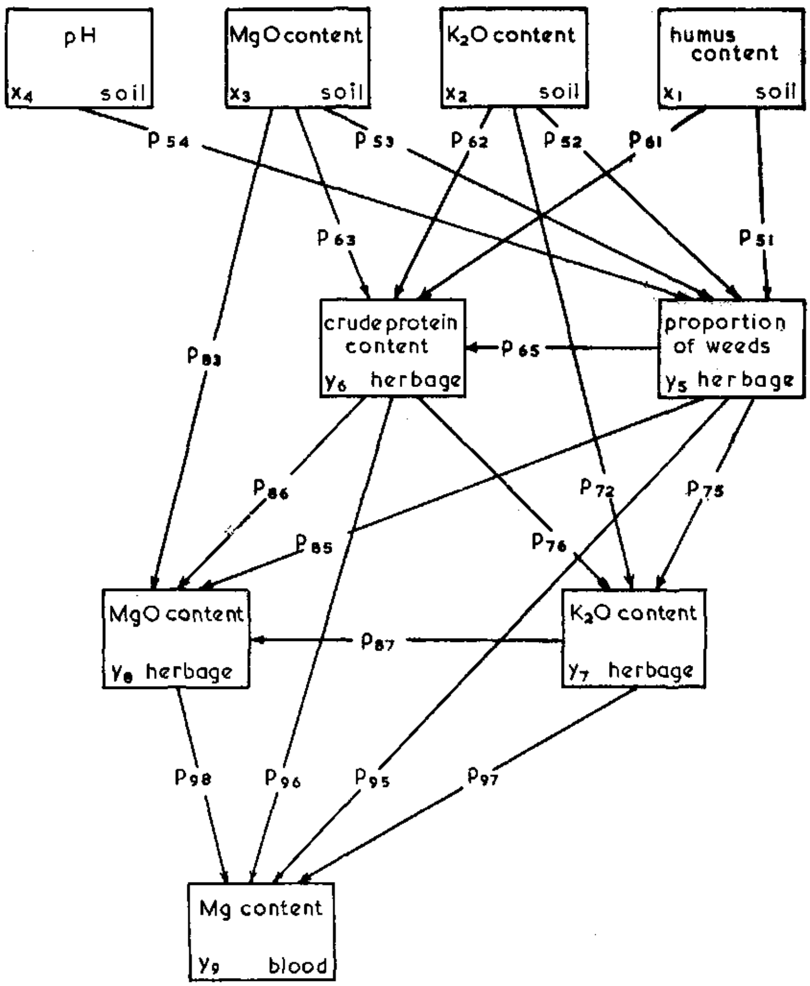


Fig. 5. Indirect influences of the 4 primary causal factors on the Mg-content of blood serum.

The antagonistic effect is demonstrated by the negative value of the path coefficient $p_{87}$.

This model however is still not ideal, since an influence of the MgO content on the $K_2O$ content of the herbage can also be assumed; an arrow from $y_8$ to $y_7$ must therefore also be drawn. Now we have a *feedback* system which is however not soluble (underidentification). A much simplified model gave as a solution, a positive effect of the MgO and a negative effect of the $K_2O$ in the herbage. Such feedback systems with opposing coefficients (so called *negative feedback*) may be considered as providing the basis for the often observed constant sum of the cations in a plant.

The method of path coefficients encourages one to proceed with the synthesis. According to Kemp [4] the Mg-content of the blood serum of cattle is of great importance for the appearance of hypomagnesaemia. At the same time, the MgO content of the herbage is assumed to be important for the Mg-content of the blood serum. We also know that the crude-protein content and the $K_2O$ content of the herbage affect unfavourably the Mg-content of the blood. All these relations are represented in the model of Figure 5, a study of which shows us the ways in which the Mg-content of the blood serum may be affected. Calculation of the path coefficients should then indicate the way in which this effect should take place most efficiently. This model is certainly soluble, but the necessary measurements of the Mg-content of the blood serum were lacking in this investigation.

## SUMMARY

Data obtained from uncontrolled experiments are often fitted to regression models, in which the dependent variable is assumed to be affected by a number of independent factors. The regression coefficient then gives the rate of a change of an effect caused by unit change in the independent variable on the assumption that this change in the causal factor does not result in a change in another independent factor (partial regression coefficient). In many cases however, this assumption is not valid; this is particularly the case with investigations into the quantitative relationships of plants.

The principle of path coefficients introduced by Wright and used up till now mainly in genetics, allows among other things for the possibility of making allowance for these indirect influences. For this purpose the investigator has to formulate a closed causal linear system with $m$ primary causes ($x$) and $n$ effects ($y$). By a closed linear system is understood a net-

work in which each variable is a linear combination of one or more other variables of this system or is one of the variables that is determined by none of the variables in this system; the latter are the primary causes, $x$. The parameters which give the extents of the influences are called path coefficients. The derivation of path coefficients is demonstrated by the equations 1–5 of the example of the simple system in Figure 2.

The potentialities of the method of path coefficients are illustrated by its application to an investigation into the effects of soil and other factors on the MgO and $K_2O$ content of herbage. The conventional regression model is given in Figure 1. Figure 3 presents a more realistic model which has been constructed that the variables, proportion of weeds and crude-protein content, are treated as cause as well as effect. The path coefficients of this model are soluble and are given in Table 2. For comparison, the regression coefficients estimated according to the model in Figure 1 are given in Table 3. In the model in Figure 4 the influence of the $K_2O$ content of the soil on the MgO content of the herbage is shown to be of a plant-physiological and not of a soil-chemical nature.

The method of path coefficients has greater potentialities than the regression merhod for the solution of certain problems. In the model of Figure 5 a synthesis between soil factors, chemical and botanical composition of the herbage, and Mg content of the blood is demonstrated; this model is soluble.

RÉSUMÉ

*Relations causales sol—plante et "coefficients path".*

Les résultats d'un essai sans intervention sont souvent analysés par une équation de régression. Dans le modèle de cette équation une des variables est expliquée par les autres, nommées variables independantes. Les coëfficients de régression formulent alors l'accroissement de l'effet pour chaque augmentation de la cause d'une unité en admettant que les autres variables explicatives ne sont pas influencées par ce changement de la cause (régression partielle). Dans beaucoup de cas cependant cette admission est inexacte.

La méthode des "coëfficients path" développée par Wright fournit la possibilité de surmonter ces difficultés. À cet effet le chercheur doit ébraucher un système causal linéair fermé avec $m$ causes primaires $x$ et $n$ effets $y$. Un système causal fermé est un réseau, dans lequel chaque variable est, soit une combination linéaire d'une ou plusieurs variables de ce système, soit une variable, qui est indépendante des variables de ce système. Ces dernières sont là-dedans les causes primaires $x$. L'intensité des influences est exprimée par les coefficients paths. Un exemple des calculs des coefficients path du système simple de la figure 2 est donné par les équations 1–5.

Les possibilités de la méthode avec les coefficients path sont démontrées à l'aide des résultats d'une recherche sur les influences des facteurs pédologiques et autres sur la teneur en MgO et $K_2O$ dans l'herbe de pâturage, vue l'importance de ces teneurs sur l'apparation de l'hypomagnesaemie. Le mo-

dèle de regression de cette recherche est donné dans la figure 1. La figure 3 présente un modèle plus réel, dans lequel les variables: teneur en mauvaises herbes et teneur en proteïne brute de l'herbe sont aussi bien cause qu'effet. Les coefficients path ce de modèle sont résolubles et mentionnés dans la tabelle 2. Pour comparaison la tabelle 3 mentionne les coefficients de régression du modèle de la figure 1. Le modèle de la figure 4 exprime par voie physiologique l'influence de la potasse du sol au moyen de la teneur en potasse de l'herbe.

La méthode des coefficients path a plus de possibilités que le modèle de régression. Elle présente de grosses possibilités pour une analyse synthétique des résultats. Le modèle de la figure 5 donne la synthèse entre les facteurs pédologiques, composition de l'herbe et teneur en Mg du sang. Les coefficients path de ce modèle sont résolubles.

## ZUSAMMENFASSUNG

*Kausale Boden—Pflanze-Zusammenhänge und Pfad-Koeffizienten.*

Die in einem Experiment ohne Eingriff erzielten Ergebnisse werden oft mit einer Regressionsgleichung ausgewertet. Im Model dieser Gleichung wird eine Variabele durch die sonstigen sog. unabhängigen Variabelen "erklärt". Die Regressionskoeffizienten geben dann die Zunahme des Effektes an wenn eine Ursache um 1 wächst, unter Annahme dass die sonstigen erklärenden Variabelen durch die Änderung dieser Ursache selbst nicht geändert werden (partielle oder Teilregression). In vielen Fällen entspricht diese Annahme nicht der Wirklichkeit. Die von Wright entwickelte Methode mit den Pfad-Koeffizienten gibt die Möglichkeit diese Schwierigkeiten bisweilen zu beseitigen. Hierzu muss der Forscher ein geschlossen kausales, lineares System mit $m$ primären Ursachen $x$ und $n$ Effekten $y$ aufsetzen. Unter einem geschlossen kausalen System wird ein Netzwerk verstanden in dem jede Variabele entweder eine lineare Kombination einer oder mehrerer Variabelen dieses Systems oder eine der Variabelen ist, welche durch keine der Variabelen des Systems bestimmt ist. Die letzten Variabelen sind darin die primären Ursachen $x$. Die Grösse eines Einflusses wird durch den Pfad-Koeffizient gegeben. Ein Beispiel der Auswertung der Pfad-Koeffizienten des einfachen Systems aus Figur 2 wird durch die Gleichungen 1–5 gegeben.

Die Möglichkeiten der Methode mit den Pfad-Koeffizienten werden vorgeführt an Hande einer Untersuchung nach den Einflüssen von Boden- und anderen Faktoren auf den MgO- und $K_2O$-Gehalt des Weidegrases in Bezug auf die Wichtigkeit dieser Zusammensetzung für das Auftreten von Hypomagnesaemie. Das Regressionsmodell dieser Untersuchung wird in Figur 1 gegeben. Figur 3 gibt ein mehr reelles Modell, worin die Variabelen Prozentsatz an Kräutern und Roheiweissgehalt des Grasses Ursache sowohl wie Effekt sind. Die Pfad-Koeffizienten dieses Modelles sind lösbar und werden in Tabelle 2 gegeben. Zur Vergleich werden in Tabelle 3 die Regressionskoeffizienten des Modelles aus Figur 1 gegeben. Im Modell von Figur 4 wird der Einfluss von Kali im Boden mit Hilfe des Kaligehaltes vom Gras physiologisch gedeutet.

Die Methode mit den Pfad-Koeffizienten hat viele Vorzüge vor dem Regressionsmodell. Die Methode gibt weiter grosse Möglichkeiten für eine synthetische Auswertung der Ergebnisse. Im Modell von Figur 5 wird eine Synthese zwischen Bodenfaktoren, Zusammensetzung des Weidegrases und Mg-Gehalt des Blutes gegeben. Die Pfad-Koeffizienten dieses Modelles sind lösbar.

Received September 28, 1962

## LITERATURE

1 Basmann, R. L., An expository nota on estimation of simultaneous structural equations. Biometrics **16**, 464–480 (1960).

2 Ferrari, Th. J., Pijl, H. and Venekamp, J. T. N., Factor analysis in agricultural research. Neth. J. Agr. Sci. **5**, 211–221 (1957).

3 Ferrari, Th. J., Vergelijking tussen proeven met en zonder ingreep. Landbouwk. Tijdschr. **72**, 792–801 (1960).

4 Kemp, A., Hypomagnesaemia in milking cows: The response of serum magnesium to alterations in herbage dcomposition resulting from potash and nitrogen dressings on pasture. Neth. J. Agr. Sci. **8**, 281–304 (1960).

5 Kempthorne, O., An introduction to genetic statistics. New York (1957).

6 Koopmans, T. C., Identification problems in economic model construction. Chapter II, Studies in Econometric Method. Cowles Commission Monograph no. 14, ed. W. C. Hood and T. C. Koopmans (1953).

7 Li, C. C., The concept of path coefficients and its impact on population genetics. Biometrics **12**, 190–210 (1956).

8 Simon, H. A., Causal ordering and identifiability. Chapter III, Studies in Econometric Method, Cowles Commission Monograph no. 14, ed. W. C. Hood en T. C. Koopmans (1953).

9 Simon, H. A., Spurious correlation: a causal interpretation. J. Am. Stat. Ass. **49**, 467–479 (1954).

10 Sluijsmans, C. M. J., Magnesium- en kaliumgehalten van gras in afhankelijkheid van bodem- en andere factoren. Tijdschr. Diergeneesk. **87**, 547–556 (1962).

11 Theil, H. and Kloek, T., The statistics of systems of simultaneous economic relationships. Statist. Neerland. **13**, 65–89 (1959).

12 Tintner, G., Econometrics. New York (1952).

13 Tukey, J. W., Causation, regression and path analysis. Chapter 3, Statistics and Mathematics in Biology, ed. Kempthorne (1954).

14 Turner, M. E. and Stevens, C. D., The regression analysis of causal paths. Biometrics **15**, 236–258 (1959).

15 Valavanis, S., Econometrics. An introduction to maximum likelihood methods. New York (1959).

16 Wright, S., Correlation and causation J. Agr. Research **20**, 557–585 (1921).

17 Wright, S., The method of path coefficients. Ann. Math. Statist. **5**, 161–215 (1934).

18 Wright, S., The interpretation of multivariate systems. Chapter 2, Statistics and Mathematics in Biology, ed. Kempthorne (1954).

19 Wright, S., Path coefficients and path regressions: alternative or complementary concepts? Biometrics **16**, 189–202 (1960).

20 Wright, S., The treatment of reciprocal interaction, with or without lag, in path analysis. Biometrics **16**, 423–445 (1960).