

INSTITUTE FOR SOIL FERTILITY - GRONINGENMODELS AND THEIR TESTING IN AGRICULTURAL RESEARCH

Considerations on the methodology of research designed
to estimate input/output relationships

by

Th.J. Ferrari

Introduction

It is the task of agronomic research to help the farmer in answering the question which of the different alternative technical possibilities he has to choose in a certain situation. The agronomist is often in a difficult position because he has to take into account a great number of factors which are difficult to measure besides. These two facts are giving many difficulties in the research too. The difficulties count particularly to the agriculturist who has to study e.g. the economics of milk production. One has to realize which alternative possibilities a farmer in his mixed farming-system of Western Europe is up to. Which are the consequences of an increase of grass yield by rising the nitrogen dressing via the chain: soil, crop and cattle for his financial results? To this livestock farming apply particularly the words of the mathematician Bross who wrote in his book "Design for Decision": "It is much more difficult to be a good farmer than a good mathematician because the farmer must deal with so many vague and complex problems". It is the task of research to give solutions for these problems.

It is clear that it is necessary for making a justified choice between the alternative possibilities to have at one's disposal a preferably qualitative description as complete as possible of the technical possibilities and their consequences. Whether one has to advice the farmer with the help of programming models or one tries to analyze a certain farming result, this knowledge is always necessary. In the first case the technical relationships are taken for granted, in the second one the accent lies on the explanation of certain phenomena. The correspondence

of both procedures is that a representation is made of the relationship or phenomenon by means of models. In the first case the model is assumed to be known, in the second one the model is constructed in the form of a hypothesis, which has to be tested on its reality value by an experiment. Consequently the research as such will have to do with the model as hypothesis mainly, although there are all sorts of nuances. In this paper we only deal with problems met in the investigation into the explanation of relationships; we know, however, that the results of this research can be used for all sorts of programming purposes.

Use of models in agricultural research

What do we understand by models and what are their functions in research? Models are simplifying abstractions of reality, in which only elements already familiar to us are taken up. Only elements of the reality are taken up, which are being studied in the science concerned. The abstraction is expressed in some language, that may be in words or in diagrams, mathematically or materially. Within the given limits we try to describe the reality as completely as possible.

For the research it is of great importance, that the models have the properties that the conclusions drawn from them must hold for the reality too. In other words, the reality value of an assumed model is closely connected with that of the conclusions. It also appears, that the hypotheses, so important for progress in science, are very suitable to be expressed in models. In this way we get the connection between model and research. This takes place as follows. As in all empirical sciences the systematic increase of our knowledge in agricultural research is acquired by the formulation of hypotheses, which are tested by the reality, viz. the observations, by means of predictions. Absence of agreement between observations and predictions rejects the hypothesis, presence makes the hypothesis more acceptable. In connection with the complex and practical character of an agricultural object it appears useful to build up models in the form of mathematical equations. In view of this

particular character of the object, e.g. plant or milk production, one will meet some difficulties in testing and quantifying the parameters of the models; such difficulties are present also in other sciences as sociology, economy and astronomy.

It is clear, that the ultimate criterion in agriculture must be the production expressed in some economic terms. In the model the production, e.g. the milk yield in kgs per ha will be brought in causal relation with a number of factors. In a simple case the function has the following form: the yield depends on the amount of roughage and concentrates. This is a very simple model, which applies perhaps for stable feeding under certain conditions. It is much more difficult, however, to relate the farm-economic results with the amounts of nitrogen dressing. No direct relation between these two factors exists and all kind of factors may interfere. It is clear, that the hypothetical model of these relationships becomes more complex and that the testing and quantifying will give difficulties. When it is examined which factors may influence the yield or the economic results of an operation, the following groups can be distinguished:

- a. factors which can be varied, such as nitrogen dressing and amount of concentrates;
- b. hardly or not changeable factors, which can be measured previously or predicted: soil profile and groundwater level, size of holding, number of cows per ha etc.;
- c. hardly or not changeable factors, which can not be predicted such as weather, diseases and pests, economic state etc.

The complex character of the production, specially under farming conditions and the peculiar attributes of the just-mentioned factor groups have certain consequences in the research for the construction and testing of models.

This applies first of all to the testing. It is a well-known fact, that testing of a hypothesis in natural sciences takes place mainly by means of an artificial variation, ceteris paribus, according to the idea that a change of a

factor assumed to be a cause must also result in a corresponding change of the effect. In this the ceteris paribus assumption is very important. The introduction of a variation is difficult or impossible, when we are dealing with factors of the second or third group, for they are not changeable. Astronomy shows, that it is possible after all to obtain important results without artificial change. For the rest it is doubtful whether the ceteris-paribus principle with an artificial change can be maintained in many cases. Changes in groundwater level or nitrogen dressing cause a chain of changes of other factors, which can affect the production in their turn. The result is, that conclusions about a factor causing a phenomenon can not be drawn. It is also clear, that it is difficult to investigate effects of certain changes under farming conditions. Restrictions by farming conditions, costs etc. prevent the introduction of experiments with artificial changes in the farm-economic research. There are perhaps possibilities in certain production branches, in which the feeding takes place in the stable and with purchased feeding stuffs only.

A second difficulty is connected with the great number of factors influencing the production, and with their interdependence. In view of the wanted usefulness in practice it has the consequence that the researcher always has to investigate many factors together. The normal experiment, by which the influence of one or two factors is investigated, is less suitable to solve practical questions. It is a well-known fact that an increase of the number of factors investigated soon becomes impossible indeed; an increase of the number of factors increases the size of the experiment, by which the rest variance, certainly of field experiments, becomes the main factor. Statisticians have tried to eliminate this drawback by introducing the principle of
x confounding; a satisfying solution has not yet been obtained.

The limitation remains, that the results of these experiments only hold for this special case, with the special conditions of soil, climate, care etc. The experience is therefore, that the results of the different investigations can diverge strongly. The investigator can try to solve

this difficulty by carrying out a large number of experiments. He tries to obtain a good sample of production circumstances. After all only an average result is obtained. A subdivision according to geographical units, usual in sociology, does not satisfy either. Without a more profound analysis of the factors causing the differences an extrapolation from the average result to the future individual cases remains risky.

Such an analysis is possible, however, as it not at all necessary to test the hypothesis by means of empirical data obtained by an artificial change only. Under the influence of the natural sciences many research-workers are of opinion that the s.c. experiment with an artificial change (controlled experiment¹⁾) is the only correct method. However, it is quite possible to test a hypothesis by means of data from an experiment without this artificial change (uncontrolled experiment¹⁾), in which the variation of nature is used. As far as the logic of experimentation is concerned, this distinction is of no account at all. The testing of the hypothesis by means of deduced predictions is deciding. The word "experiment", derived from the latin verb "experiri", i.e. to test, expresses this already. However, by the methods and results of the physical and chemical sciences the word "experiment" has got a quite other sense, viz. artificial change, and the original sense is often forgotten. Of course it must be said, that an uncontrolled experiment also involves certain difficulties, of which particularly the difficulty to obtain a sufficient separation between the possible causal factors should be mentioned. For the rest, the difficulties of a controlled experiment should not be underrated either. We mentioned already the unreal assumption of the ceteris-paribus principle. In a previous paper we compared the advantages and disadvantages of both methods. It is evident, that an experiment without artificial change gives the possibility to test and quantify models, in which factors of the second and third group are taken up, factors therefore which are not or hardly changeable and by which differences between

1) In the Dutch language "proef met ingreep" and "proef zonder ingreep", in German "Experiment mit Eingriff" and "Experiment ohne Eingriff" respectively.

experimental results can be explained.

Two-variable models with one equation

Which models and which functions are generally used in agricultural research? In the following discussion we have illustrated our statements with examples derived from soil fertility studies. It will be clear, that the statements also apply to other parts of the agricultural research. We restrict ourselves for the present to models which can be described with one equation with one or more factors.

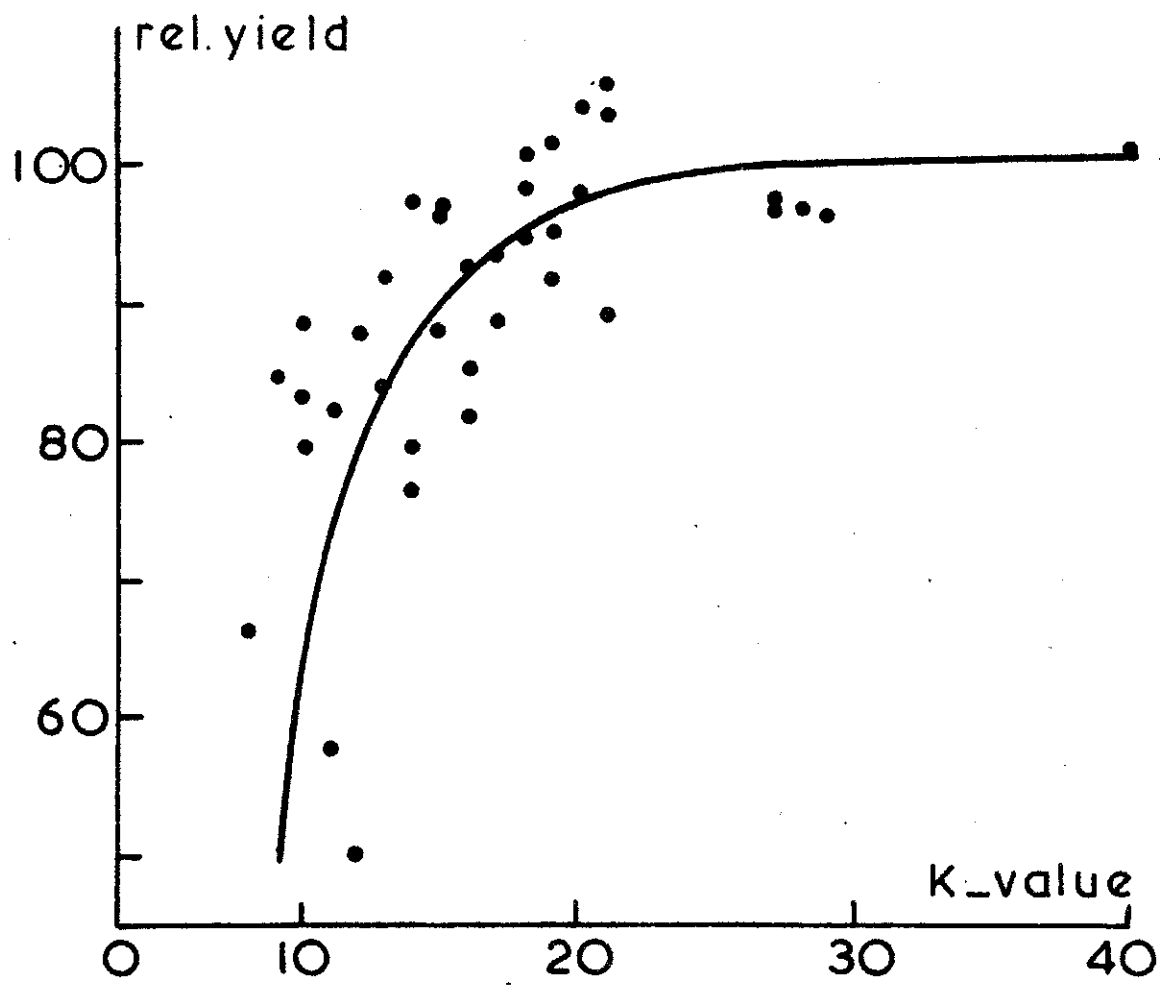
The most simple model is the hypothesis, that the yield differences can be explained by one or more factors without a further description of the functional form. This is the point of view of the analysis of variance. The drawback of course is, that a possibility to interpolate and to extrapolate is difficult because of the absence of a function. Economic interpretations are difficult in that case.

More possibilities are given by the model, of which the function is a linear equation: a unit increase of the independent variable increases the effect with a constant amount, no matter which value the first variable has. We know, that this assumption is in many cases not real, especially in milk production investigations. The linearity can be useful in a limited region of the production; according to experiences of agricultural research it would be more useful to use non-linear functions reaching a maximum. It has advantages to choose the most simple function in this case. In the literature many equations have been proposed. The most well-known is the Mitscherlich equation, afterwards with a depression. Some more functions are:

$y = A(1 - 10^{-cx})$	Mitscherlich
$y = ax^b$	Cobb-Douglas
$y = A \cdot 10^{-z \left(\log \frac{x+i}{a+1} \right)^n}$	von Boguslawski-Schneider
$y = bx - cx^2$	
$y = b\sqrt{x} - cx$	

We abandon a discussion of the general and particular properties of these functions, for which we refer to the

Fig. 1 Relationship between potash-status of the soil and loss of potato yield without potash dressing

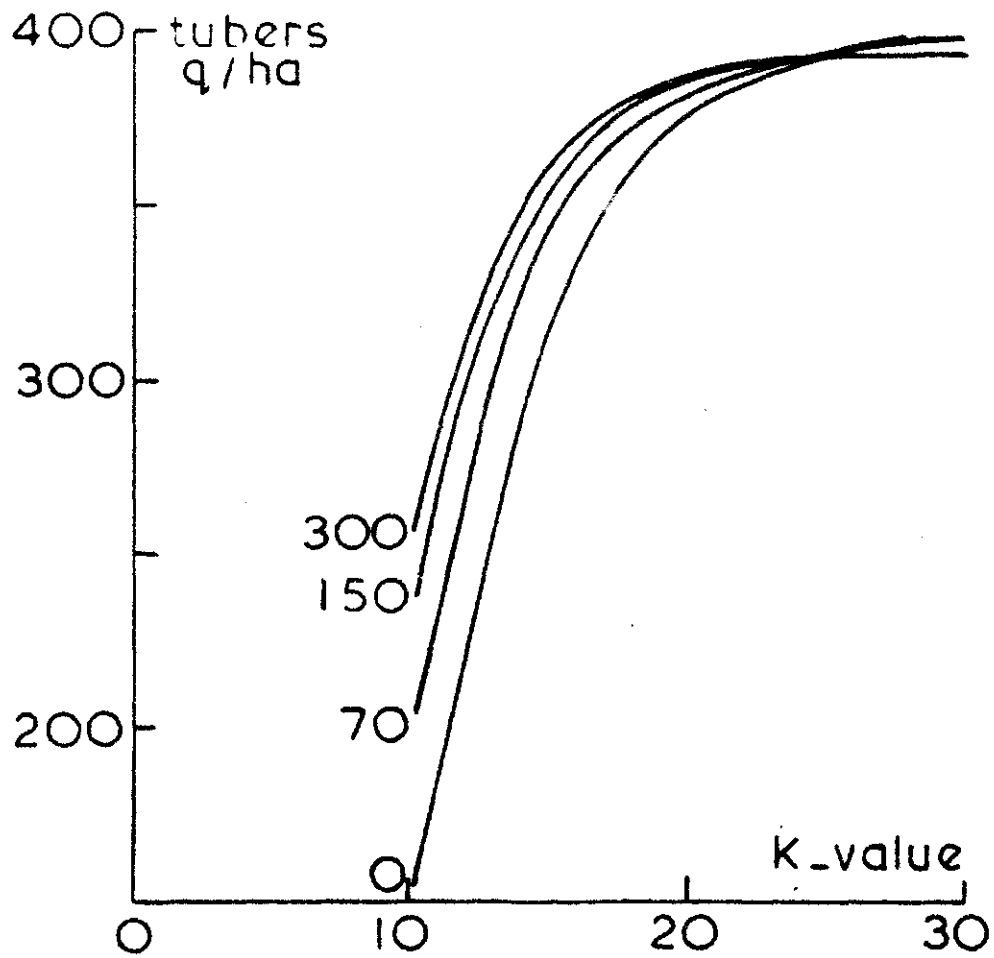


book of Heady and Dillon and to the paper of Hoffmann and Dörfel.

These equations have one thing in common. They have been developed mainly heuristic, their theoretical base is very small. By this we mean to say, that there is no preference for some equation out of a physiological or biochemical point of view. The only theoretical derivation we know is the one for the Mitscherlich equation by Linser and Kaindl in the domain of plant nutrition. It is striking indeed, that so little basic research on the production functions has been done. This does not apply to all parts of agricultural research. In soil science e.g. a number of processes have been described by functions derived from basic chemical or physical knowledge. It appears, that we are in urgent need for more biologically derived equations, specially in view of the great possibilities which the computers have for the solution of these equations.

Personal preference decides at present which equation is chosen ultimately. The choice is often made by the suggestion, which the observational data are giving. A study over the milk production by Heady, Schnittker, Jacobson and Bloom leaves the choice between three functions, viz. the logarithmic, the quadratic or the square root equations. It is clear, that the function ultimately chosen should be taken again as a hypothesis in the next investigation; experience shows, that this has often been omitted. The uncertainty about the function to be taken and the impossibility to compute - formerly we did not yet have calculating machines - have been the background to develop some 30 years ago the graphic method in soil fertility research, always using the suggestion given by the observational data. The same application has also been used in the economic research in the U.S.A. As an example we show in fig. 1 the results of an investigation into the relationship between potash-status of the soil and the loss of potato yield without potash-dressing, expressed in percentages of the maximal yield. Each point represents the result of one field experiment, the differences in potash-

Fig. 2 Relationship between potash status of the soil and potato yield with 4 amounts of potash dressing



status were acquired without artificial change by taking natural situations. We may expect, that the differences between the graphic and numerical methods will be small.

More-variable models with one equation

We know, that the two-variable models mostly do not meet the needs of a complete or satisfying description of the processes. With a view to this description functions with more factors have been developed such as:

$$y = b_1x_1 + b_2x_2 + \dots$$

$$y = A(1-10^{-c_1x_1})(1-10^{-c_2x_2})\dots \quad \text{Mitscherlich}$$

$$y = ax_1^{b_1} x_2^{b_2}\dots \quad \text{Cobb-Douglas}$$

$$y = b_1x_1 + b_2x_2 + b_{12}x_1x_2$$

The properties of these equations will not be discussed either, although important in connection with terms as isoclines, substitution rates etc. We only point to the possibility to take up in these equations terms for interaction. In the last equation the product term represents the interaction. Although the interaction in our opinion is mostly nothing else than a word to mark our lack of knowledge, the researcher is often forced to take up these terms of interactions. Fig. 2 shows a tested model with interaction, in which the effect of potash-dressing depends on the potash-status of the soil.

The extreme consequences of the possibilities of an experiment without artificial change and of more-variable models are the investigations, in which the research worker tries to find in a graphic or numerical way an explanation for the differences in yield or economic farming results by means of single plots or farms respectively. A model has been drawn up, which aims to give an explanation of the variance present in nature. The upbuilding of the model with many factors goes rather far. In contrast with the design of the analysis of variance to make the rest variance as small as possible, these multi-factorial investigations are interested especially in a great starting variance. Fig. 3 shows the possibilities of such an analysis by means of the correlation between actual and estimated yields. This analysis was based on a model with 13 factors, of which

Fig. 3 Correlation between estimated and actual yields of potatoes

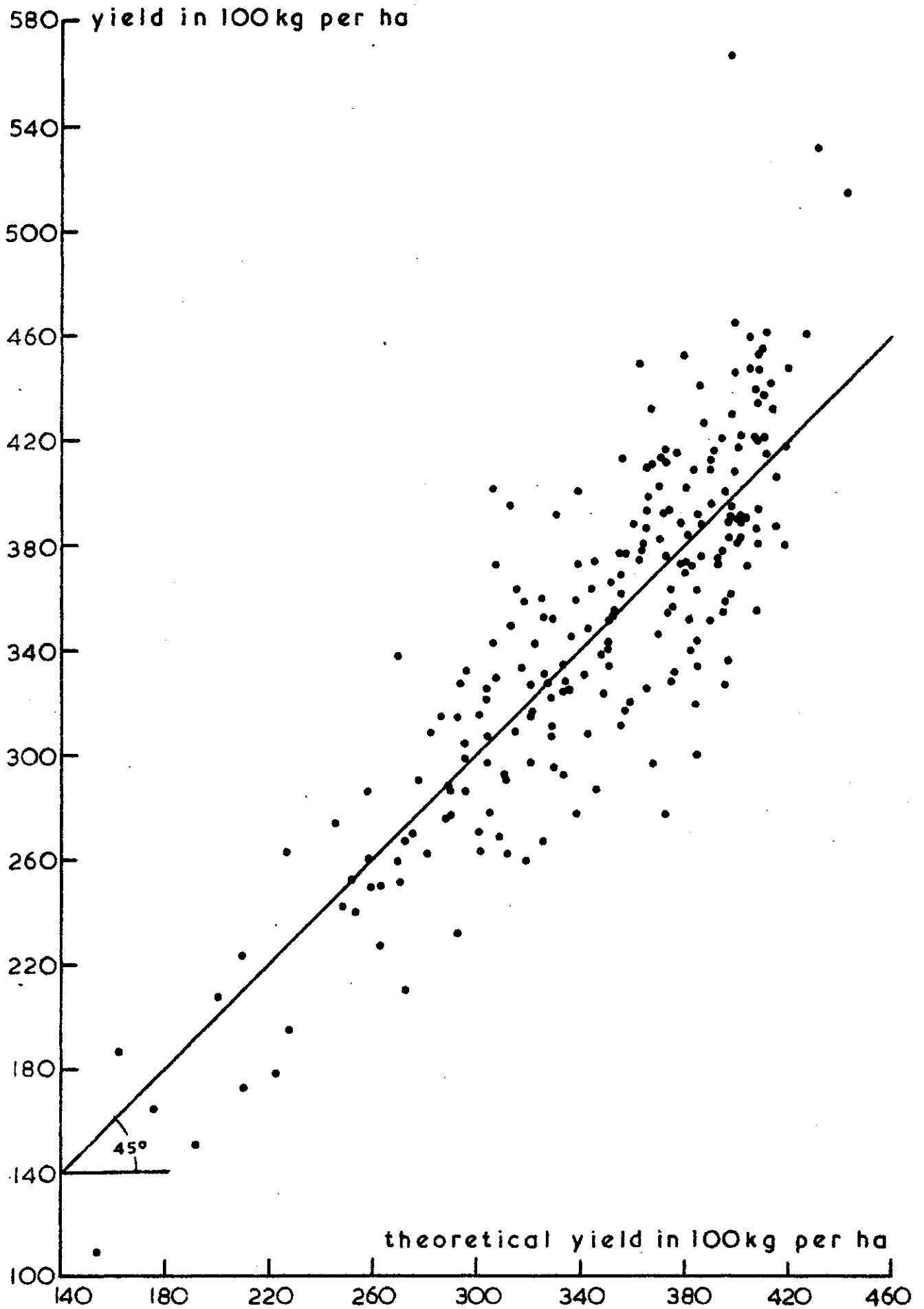
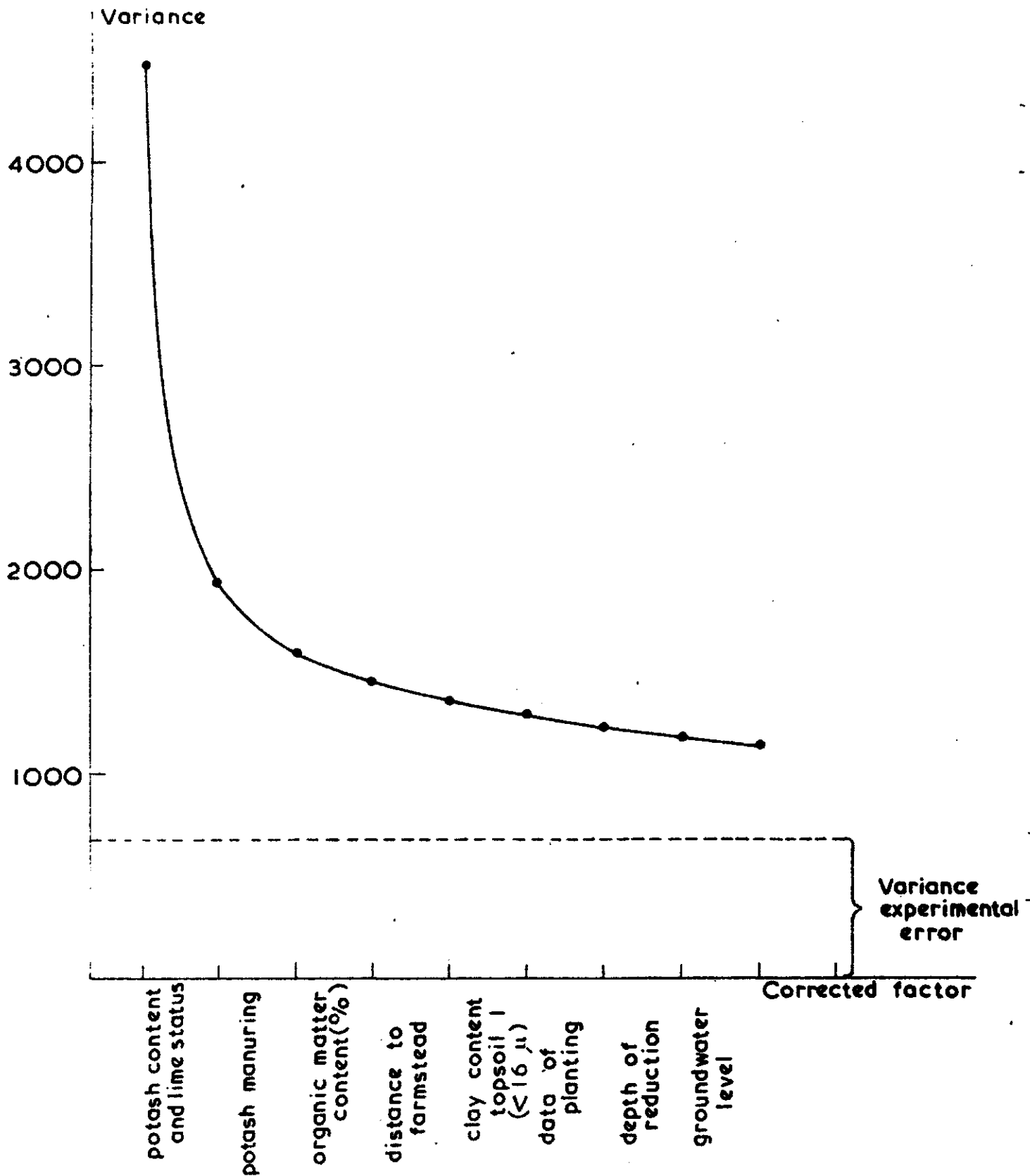


Fig. 4 Decrease in yield variance after successive elimination of the factor influences



9 had a statistically significant influence. Fig. 4 shows the decrease of the yield variance by successive elimination of the factor influences. The diagram also shows a probably more general phenomenon: many factors have a small, only few factors a relatively great influence. We shall return later on to this subject in connection with the choice of simple-structure rotation in factor analysis.

Models with more equations, chain processes

The equations of the models discussed up till now are especially normal regression equations. The regression model is characterized by the hypothesis, that a causal relationship exists between the s.c. independent or causal factors and the dependent factor or effect. It is also assumed, that a change of an independent factor affects the dependent variable only and does not affect the other factors. The same assumption must also be made in the experiment with artificial change according to the ceteris-paribus principle. We find, however, that these assumptions are not often in agreement with the facts both in the experiment with artificial change and in that without such a change. This means, that the assumed model is incorrect and can not be applied.

We can demonstrate this by means of an example out of an investigation into the factors affecting the magnesium content of herbage. At first a normal regression model was built up and tested by observations of an uncontrolled experiment. The diagram of fig. 5 shows the hypothetic model. In this the magnesium content of the herbage is the dependent variable or effect; further it is assumed, that the factors magnesium, potash and humus content of the soil, crude-protein content and proportion of weeds in the herbage will influence causally the magnesium content of the herbage. In the diagram these influences are marked with an arrow, the rate and the direction have been calculated from the observations. Thus we assume, that a change of the magnesium content of the soil only effects the magnesium content of the herbage but does not effect

Fig. 5 Regression model with magnesium content of the herbage as dependent variable, other variables as dependent causal variables

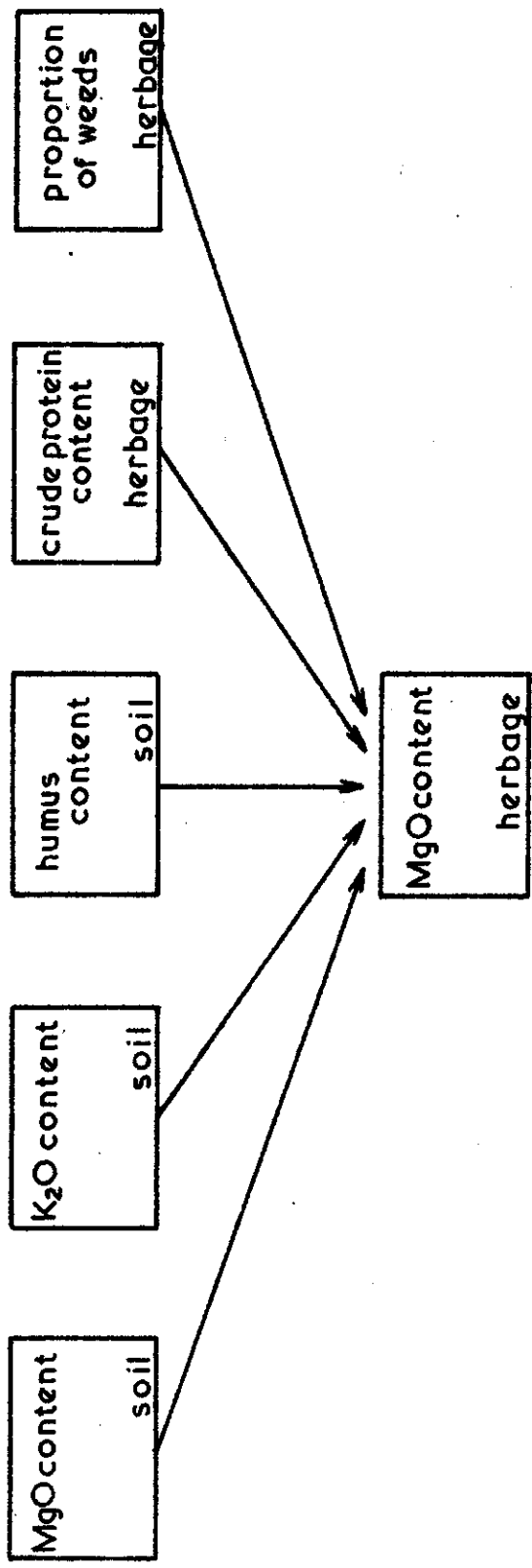
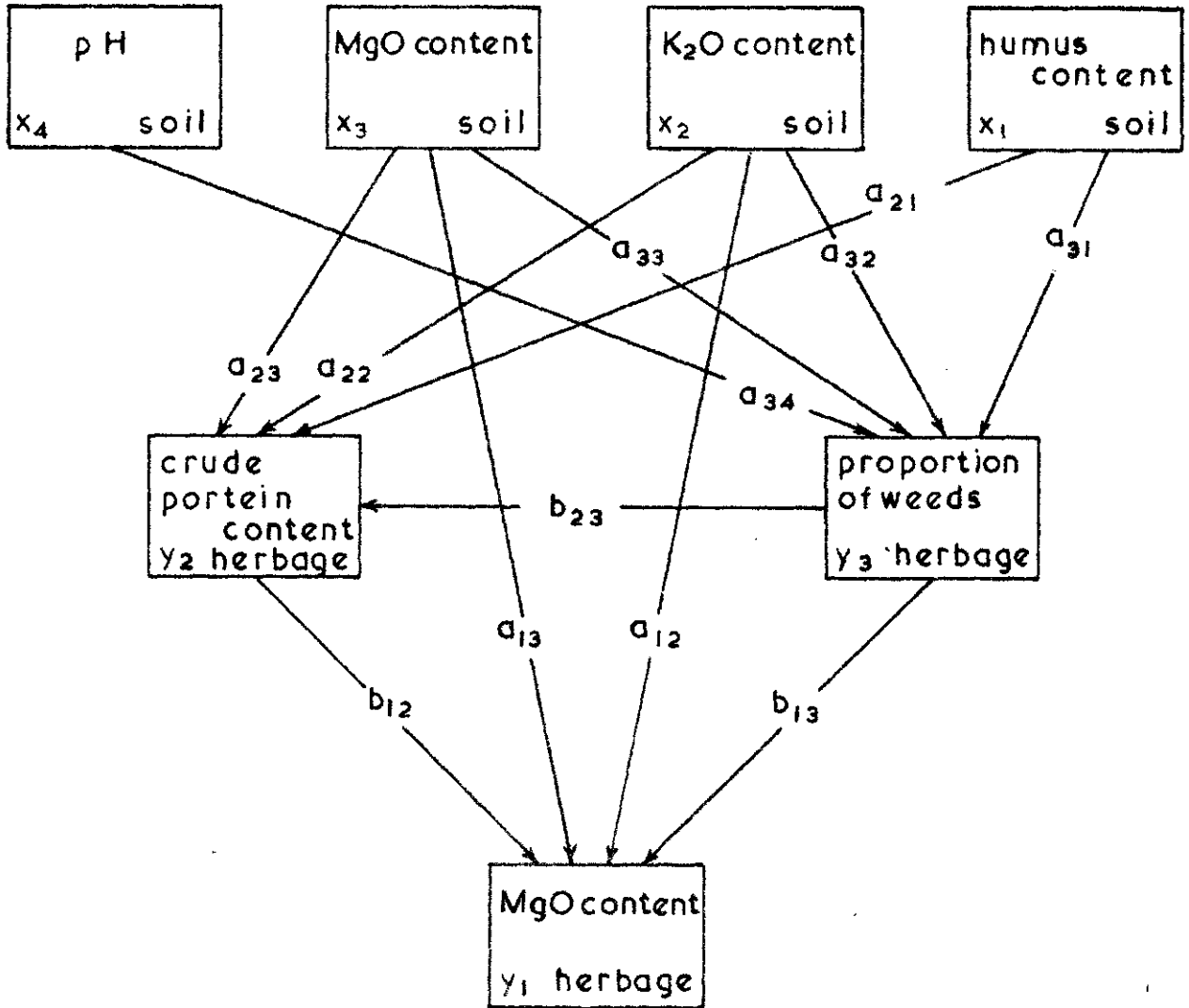


Fig. 6 Direct and indirect influences of the 4 causal factors on the magnesium content of the herbage



the crude-protein content and the proportion of weeds. However, we know by other investigations, that this is not true; the model is therefore not acceptable. Essentially we meet in this case a s.c. chain process, which is not describable by means of one equation.

The diagram of fig. 6 gives a model of these relationships probably more in agreement with reality. The variables crude-protein content and proportion of weeds are not only taken as independent variables; both variables are now cause as well as effect. A change of the magnesium content of the soil affects the magnesium content of the herbage not only directly, but also indirectly via the chain: proportion of weeds and crude-protein content. The model of fig. 5 without these chain processes can be represented by one equation:

$$y_1 = a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5,$$

the second model needs a system of the following 3 equations:

$$y_1 = b_{12}y_2 + b_{13}y_3 + a_{12}x_2 + a_{13}x_3$$

$$y_2 = b_{23}y_3 + a_{21}x_1 + a_{22}x_2 + a_{23}x_3$$

$$y_3 = a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4$$

Such systems of equations can be solved by the method of path coefficients. The term "path" has something to do with pathes via with the influence is affected. By this method the hypothesis formulated in a model, concerning these relationships is tested and quantified. The influence is represented by the path coefficient, giving the rate and direction of the effect change for every unit change of the causal variable. Table 1 gives the results of the analysis of the model shown in fig. 6.

Table 1. Computed values of the 12 path coefficients of the model of fig. 6.

Effect \ Cause	Humus content (x ₁)	K ₂ O content soil (x ₂)	MgO content soil (x ₃)	pH (x ₄)	Proportion of weeds (y ₃)	Crude protein content (y ₂)
Proportion of weeds (y ₃)	1,67	-0,23	-0,031	5,26		
Crude protein content (y ₂)	-0,74	0,11	0,011		0,20	
MgO content of herbage (y ₁)		-0,0038	0,0004		0,0041	0,0083

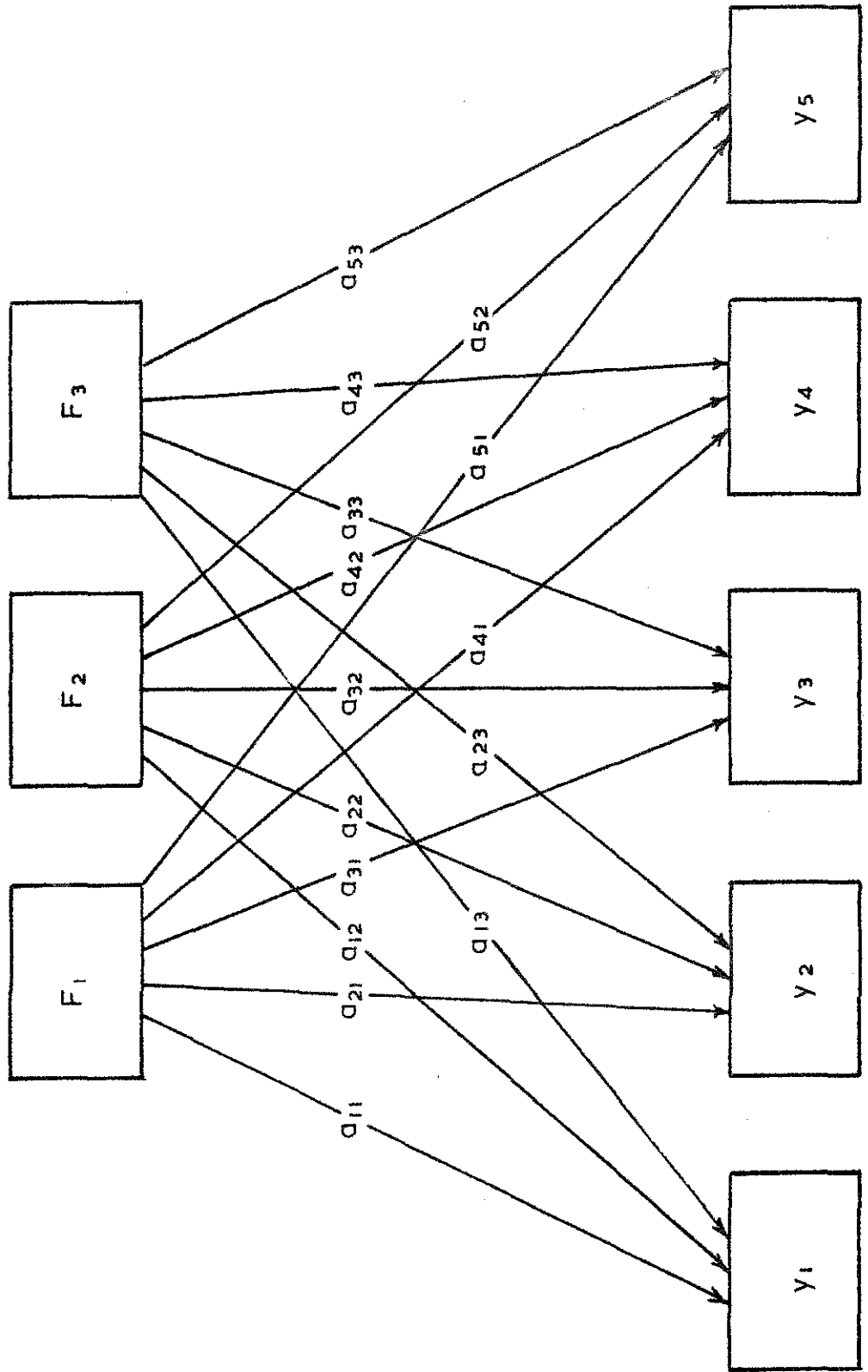
The general form of a system of equations describing a chain process is as follows:

$$\begin{aligned}
 b_{11}y_1 + \dots + b_{1MYM} + a_{11}x_1 + \dots + a_{1L}x_L &= u_1 \\
 b_{21}y_1 + \dots + b_{2MYM} + a_{21}x_1 + \dots + a_{2L}x_L &= u_2 \\
 \vdots & \\
 b_{M1}y_1 + \dots + b_{MMY} + a_{M1}x_1 + \dots + a_{ML}x_L &= u_M
 \end{aligned}$$

It is clear, that some path coefficients a and b a priori may be assumed to be zero in real models. By means of this method it is also possible to investigate models, in which feedback systems are taken up. In our opinion such models should be preferred to the normal regression models, especially by their closer correspondence to the reality. It is possible to use non-linear functions in these systems. The method is closely related to the method of simultaneous equations out of the econometry.

An extreme case of such models is the model, upon which the factor analysis is based. The number of limiting conditions in a model of the factor analysis is small, by which fact the system of equations has become s.c. unidentifiable; an exact solution can not be obtained by

Fig. 7 Diagram of a factor-analysis model



mathematical arguments only. The schematic diagram of such a model is given in fig. 7. The causal x-variables (here named F) are unknown. Next the analysis tries to calculate these F-variables as s.c. aspects. This factor analysis is not only important for the testing of such models, it also can be used to give ideas for drawing up a more limited hypothesis. The possibilities of this analysis are many. The method is very suitable e.g. to indicate and quantify the ecological properties of grasses grown under natural conditions. Starting point for the analysis is the matrix of correlation coefficients between soil factors and sociological characteristics, in this case the frequency percentages of grasses. The factor analysis with a following rotation of the aspect axes resulted in a number of aspects given in table 2, of which the first aspect represents the reaction of grass to water supply. The differences in numbers running from +1 to -1 are a measure for this reaction. The positive numbers show the hydrophile character, the negative ones the drought resistance.

Table 2. Interdependences of soil factors and frequency percentages of grasses; aspect values after rotation to simple structure.

Factor	Aspects			
	1	2	3	4
pH(KCl)	0,655	-0,246	-0,209	-0,074
humus content	0,684	-0,098	0,003	-0,240
silt content	0,811	-0,298	-0,113	0,003
sand content	-0,881	0,242	0,074	0,110
specific surface sand	0,671	-0,261	-0,258	-0,028
magnesium content soil	0,575	-0,385	0,266	-0,152
phosphate content soil (water)	-0,137	0,550	0,255	0,010
phosphate content (citric acid)	0,650	0,184	0,112	0,243
potash content	-0,049	0,691	0,396	-0,463
copper content (Asp.)	0,647	-0,340	-0,096	0,055
distance farm	0,318	0,029	-0,493	0,020
depth clay-layer	0,197	0,360	-0,040	0,380
thickness humus-layer	0,023	-0,004	-0,038	0,568
moisture content	0,611	-0,134	-0,062	0,400
groundwater level	0,626	-0,381	-0,075	-0,409
fluctuation	-0,495	-0,214	0,059	-0,023
nitrogen dressing	-0,151	0,352	0,357	0,320
phosphate dressing	0,007	0,461	-0,037	0,059
potash dressing	0,023	0,538	0,252	0,116
<i>Poa pratensis</i> L.	-0,401	0,052	0,103	0,248
<i>Festuca rubra</i> L.	0,383	-0,131	-0,412	0,298
<i>Agrostis tenuis</i> Sibth.	-0,341	-0,259	-0,213	-0,446
<i>Lolium perenne</i> L.	-0,254	-0,217	0,282	-0,166
<i>Poa annua</i> L.	-0,219	0,252	0,563	-0,069
<i>Alopecurus geniculatus</i> L.	0,321	-0,301	0,336	-0,357
<i>Agropyron repens</i> P.B.	-0,246	0,075	0,372	0,326
<i>Festuca pratensis</i> Huds.	0,787	-0,024	-0,005	0,154
<i>Poa trivialis</i> L.	0,314	-0,726	0,034	-0,181
<i>Agrostis stolonifera</i> L.	-0,105	-0,186	-0,365	-0,107
<i>Dactylis glomerata</i> L.	-0,111	0,069	0,156	0,342
<i>Achillea Millefolium</i> L.	-0,270	0,072	0,227	0,181
<i>Ranunculus repens</i> L.	0,051	-0,427	-0,086	-0,218
<i>Gardamine pratensis</i> L.	0,346	-0,521	-0,025	0,004
<i>Carex stolonifera</i> Hoppe	0,716	0,099	-0,165	0,043
<i>Glyceria maxima</i> Holmb.	0,807	-0,018	-0,138	0,214
<i>Ranunculus acer</i> L.	0,297	-0,344	-0,489	-0,181
<i>Rumex Acetosa</i> L.	0,393	-0,152	-0,414	0,272
<i>Holcus lanatus</i> L.	0,274	-0,183	-0,568	-0,134
<i>Anthoxanthum odoratum</i> L.	0,172	-0,246	-0,676	-0,020
<i>Centaurea Jacea</i> L.	0,174	-0,201	-0,251	0,202
<i>Bellis perennis</i> L.	0,099	-0,482	-0,101	0,307
<i>Cynosurus cristatus</i> L.	0,046	-0,300	-0,232	-0,373
<i>Alopecurus pratensis</i> L.	0,141	-0,047	0,022	0,252
<i>Luzula campestris</i> Lam. et D.C.	-0,137	0,033	-0,556	-0,023
<i>Trifolium repens</i> L.	-0,104	-0,203	-0,041	-0,545
<i>Bromus mollis</i> L.	-0,069	0,031	-0,216	0,337
<i>Phleum pratense</i> L.	0,024	0,011	0,309	0,124
<i>Taraxacum officinale</i> Web.	-0,116	-0,083	0,129	0,671
<i>Leontodon autumnalis</i> L.	-0,229	-0,199	0,056	0,080
<i>Phalaris arundinacea</i> L.	-0,207	0,176	0,062	-0,015
quality figure grass	0,050	-0,196	0,266	-0,103

The most remarkable result is, that it is obtained by a mathematical analysis followed by a rotation to the simple structure only. The choice of a rotation to the simple structure is based on the already mentioned phenomenon, that many factors have a small influence and only few factors have a great one. A rotation of the model to simple structure tries to reach the same situation.

Summary

We have the experience, that causation, especially in agricultural phenomena, is complex and that the method of analysis used in natural sciences is not satisfactory in all respects. Some lines to disentangle this complexity based on the following ideas are given.

The first point is connected with the thought, that also in agricultural research with its applied character the hypothesis expressed in a model and followed by testing has to supply the main contribution to new knowledge. According to our experience this is frequently forgotten.

The second point is the idea, that testing can also be carried out with observational data from experiments without artificial change (uncontrolled experiments).

The third point is the knowledge, that the research worker can choose out of many models and functions. In this it is not necessary to confine the choice to functions with few factors and to models, in which the ceteris-paribus principle must be assumed.

A definite advice which attack and which models and functions should be chosen, can not be given. Each problem asks its own method of analysis, each research worker goes his own way and chooses his own models.

Litterature quoted

- Ferrari, Th.J., Vergelijking tussen proeven met en zonder ingreep. Landbk. Tijdschr. 72 (1960) 792-801.
Ferrari, Th.J., Causal soil-plant relationships and path coefficients. Plant and Soil 19 (1963) 81-96.

- Heady, E.O., J.A. Schnittker, N.L. Jacobson and S. Bloom,
Milk production functions, hay/grain
substitution rates and economic optima in
dairy cow rations. Agr. Exp. Station. Iowa
State College. Research Bull. 444 (1956).
- Heady, E.O. and J.L. Dillon, Agricultural production
functions (Ames, 1961).
- Hoffmann, E. und H. Dörfel, Die funktionale Betrachtungs-
weise bei der Auswertung von Feldversuchen.
Landw. Versuchs- und Untersuchungsw., 9
(1963) 75-107.