Causal pathways when independent variables are co-related: new interpretational possibilities

M. Kozak[1], M.S. Kang[2], M. Stępień[3]

[1]*Department of Biometry, Warsaw Agricultural University, Warsaw, Poland*
[2]*School of Plant, Environmental and Soil Sciences, Louisiana State University Agricultural Center, Baton Rouge, L.A., USA*
[3]*Department of Soil Environment Sciences, Warsaw Agricultural University, Warsaw, Poland*

ABSTRACT

We propose a novel interpretation in classical path analysis, whereby the influence of *k* independent variables on a dependent variable can be analyzed. The approach should be useful to study a causal structure with the assumption that this structure is true for the situation investigated. We propose a new coefficient, Q_i, which provides a better interpretation of classical path analysis. We provide an example in which effects of certain soil properties on grain yield of winter rye (*Secale cereale* L.) were examined.

Keywords: causal systems; determination coefficient; indirect effects; path analysis

Path analysis, developed by Wright (1921, 1934), is a time-honored statistical technique frequently used in agricultural research (Simane et al. 1993, Samonte et al. 1998, Wang et al. 1999, Mohammadi et al. 2003, Das et al. 2004). This path analysis methodology appears to have gained popularity in the agricultural investigations following the publication of Dewey and Lu (1959). There have been numerous applications of path analysis in plant and soil investigations (Board et al. 1997, 2003, Güler et al. 2001, Seker and Serin 2004, Zhang et al. 2005, Ige et al. 2007, Zheng et al. 2007).

Besides the classical path analysis, there are some additional interpretational tools for a causal system. Structural equation modeling (SEM) – a modern tool for studying path models – is also used to interpret causal systems. Kozak and Kang (2006) argued that the extension of SEM to agricultural investigations should be useful.

The SEM methodology is not perfect for all causal situations; for example, it does not work when a model is not identified, which can be for various reasons (see, e.g. Bollen 1989, Shipley 2002, sec. 6.2), for instance because of too small number of degrees of freedom of the model. Structural equation modeling fails to handle causal systems we study in this paper because there are not enough degrees of freedom to apply a χ^2 test, which constitutes a testing part of likelihood-based path analysis and SEM. Thus, there is a need for a new approach to handle such causal systems.

In the proposed approach, there is one dependent variable and several correlated variables that influence the dependent variable; the variables are assumed to be co-related through unknown common causes. Thus, we assume that a double-headed arrow (↔) in a path diagram represents a co-relationship between the two connected variables. The co-relationship is assumed to originate from an influence of a common cause of the two variables. Therefore, this double-headed arrow is not an admission of ignorance. The proposed interpretation is based on decomposition of a coefficient of multiple linear determination of a response variable (R^2) into direct and indirect effects of the independent variables considered in the causal system. The new approach is easy and informative. It makes the interpretation clearer and the selection of traits with the largest influence on a response variable easier.

The methodology presented in this paper is important for those who intend to apply path analysis in the classical, Wright's version, that is, when SEM cannot be appropriately applied. We show what to do and what not to do to apply the analysis properly, and when such an analysis may be applied. The paper is organized as follows:

– Interpretation of Pathways – here we address some issues related to the interpretation of direct and indirect effects (which are the main tools for interpreting the path diagrams).
– Interpretation Possibilities in Path Analysis – here we present theoretical background of some new coefficients that should be helpful in interpreting path diagrams.
– Example – here we demonstrate the new methodology and interpretation. As an example we use an experiment in which effects of three soil properties, viz. exchangeable and available potassium content, and total nitrogen content, on winter rye (*Secale cereale* L.) grain yield were investigated in an acid-soil environment.
– Conclusions – here we summarize the results presented in the paper.

Interpretation of pathways

Consider Figure 1, representing a classical path diagram, wherein effects of k independent variables X_1, ..., X_k, set at the same ontogenetic level (which means they are/may be co-related, but none may be a cause or an effect of the other), on a dependent/response variable Y, are studied. We assume that none of the Xs in the system is a cause or an effect of any other X. The only possible cause-and-effect relationships in the system are those between the independent variables and the response variable. All further discussion in this paper holds true only for the causal system defined above. If these assumptions are not fulfilled, the approach may provide spurious results and interpretation.

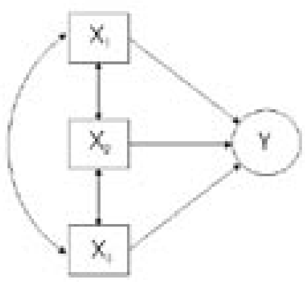

Figure 1. A classical path diagram where a dependent variable Y is affected by k independent variables X_1, ..., X_k, which are set at the same ontogenetic level

If one postulates a model under consideration without at least one of the co-relation arrows, the likelihood-based methodology, i.e. SEM, can be applied. For example, we may assume that the variables X_1 and X_2 in Figure 1 are not correlated; in that case, the correlation coefficient is set to zero. Such information may originate from the knowledge of the process being studied.

Figure 1 represents a multiple regression model expressed as a path model. Here we cannot employ the SEM methodology to test causal implications of the model (for detailed discussion, see Shipley 2002, pp. 127–129). Such an approach cannot help decide which of the independent variables are causes of the response variable (Shipley 2002, p. 130). Therefore, we cannot be sure whether or not the assumptions of the model are correct. As stated earlier, the case that we are considering assumes the model to be as follows: Y may be an effect of Xs, but an independent variable (X) cannot be a cause of another X in the system. Here, we would simply like to test whether or not the effects of the independent variables are significant. This can be done via classical testing of partial regression coefficients from the linear model Y versus Xs.

Consider the following standardized regression model:

$$y = \sum_{i=1}^{k} P_{iy} x_i + e \quad (1)$$

where: y is the standardized response variable and x_i (i = 1, ..., k) represent the standardized independent variables, P_{iy} are the partial regression coefficients (the path coefficients) for the model $E(y|x_1, ..., x_k)$, and e is a residual variable (often denoted by P_0).

The classical interpretation in path analysis (Figure 1) is based on decomposition of correlation coefficients between the response and independent variables, that is (Kang and Seneta 1980):

$$r_{yx_i} = P_{iy} + \sum_{j=1, j \neq i}^{k} P_{jy} r_{ij} \quad (2)$$

where: r_{yxi} and r_{ij} are the correlation coefficients between the i^{th} independent variable and the response variable (y), and the i^{th} and j^{th} independent variables, respectively. Thus, the correlation coefficient r_{yxi} is decomposed into terms connected with (i) direct effect of X_i on Y (or simply the path coefficient P_{iy}) and (ii) k – 1 indirect effects of the i^{th} independent variable via j^{th} independent variable on Y, that is, $P_{jy} r_{ij}$, j = 1, ..., k, $i \neq j$.

Interpreting the indirect effects from Eq. [2] is often misleading. Note that X_i is assumed not to affect X_j ($i, j = 1, ..., k, i \neq j$), so the indirect effect of X_i via X_j cannot be interpreted as the effect of X_i via X_j on Y (see Steel 2005). Such interpretation is permissible only if there is a direct path $X_i \rightarrow X_j$, which is, however, not the case in the causal situation represented here. An indirect effect in this case is simply the result of a common cause of i^{th} and j^{th} independent variables. This common cause, say C_{ij}, affects the two independent variables simultaneously, which causes them to be correlated. Because we do not/cannot measure C_{ij}, we are only able to measure the correlation between X_i and X_j and the influence of their interaction on the response variable Y. The situation described here is presented in Figure 2, where the reader acquainted with SEM will recognize a simple structural equation model. Nevertheless, this model does not fulfill the basic assumptions; viz. it lacks degrees of freedom, and one of the indicator (observed) variables (the dependent one) is not caused by any latent variable. These two assumptions must be fulfilled to make a model identified, which, in turn, helps analyse this model via SEM methodology; the issues of identification, overidentification, and underidentification in SEM are thoroughly discussed by Shipley (2002) in chapter 6.

The assessment of contributions of common causes of all pairs of independent variables to determination of a response variable is shown below. These contributions are sources of indirect effects used in the interpretation of pathways presented in Figure 1. Hence, we will (indirectly) estimate the effects that are presented in Figure 3.

When studying the importance of traits relative to their effects on response variable using path analysis (Eq. [2]), Board et al. (1997, 2003) listed the following criteria for identifying the most desirable and important traits: (a) positive correlation between the trait and dependent variable; (b) large positive direct effect of the trait on Y; and (c) small or non-negative indirect effects via other traits. After introducing the new methodology for interpreting causal systems under study, we provide new corresponding criteria.

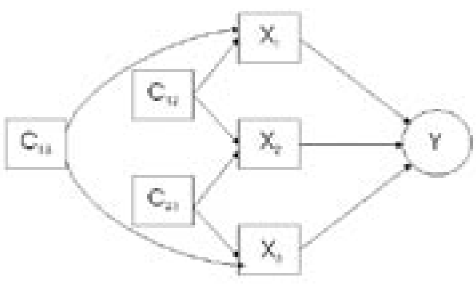

Figure 2. Explanation of common causes of independent variables (Xs) that indirectly affect the response of variable Y

Interpretation possibilities in path analysis

Below we show how we can simply interpret the causal system presented in Figures 1–3. We re-emphasize that our approach is appropriate only when a model follows the causal structure presented in Figure 1.

Let a and b be two random variables (in our case, a and b are two traits), $Cov(a, b)$ be the covariance between them, and $V(a)$ and $V(b)$ be the variances of a and b, respectively. From Eq. [1], we know that the variance of y is:

$$V(y) = V\left(\sum_{i=1}^{k} P_{iy}x_i + e\right) =$$

$$\sum_{i=1}^{k} P_{iy}^2 V(x_i) + 2\sum_{i=1}^{k} \sum_{j=1, j\neq i}^{k} P_{iy}P_{jy}Cov(x_i, x_j) + V(e) =$$

$$\sum_{i=1}^{k} P_{iy}^2 + 2\sum_{i=1}^{k} \sum_{j=1, j\neq i}^{k} P_{iy}P_{jy}r_{ij} + V(e)$$

because $V(x_i) = 1$ and $Cov(x_i, x_j) = r_{ij}$. Furthermore, $V(y) = 1$ and $R^2 = 1 - V(e)/V(y)$ (Quinn and Keough 2002, p. 92), where R^2 is the determination coefficient from Eq. [1]. Hence, we can decompose the R^2 as follows (Dofing and Knight 1992):

$$R_y^2 = R^2 = \sum_{i=1}^{k} P_{iy}^2 + 2\sum_{i=1}^{k} \sum_{j=1, j\neq i}^{k} P_{iy}P_{jy}r_{ij} \quad (3)$$

The very simple and known equality presented in Eq. [3] can be used to interpret the causal system under study. From Eq. [3] follow:

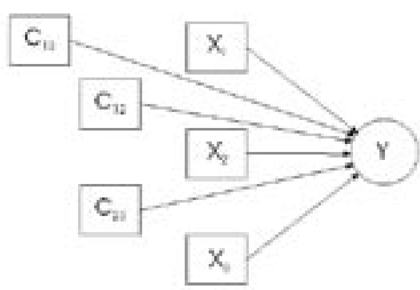

Figure 3. Indirect effects whose contribution is used in the interpretation of pathways presented in Figure 1

(a) P_{iy}^2 – the squared path coefficient connected with the influence of the i^{th} trait on Y – is part of the determination of the response variable that is related to the direct influence of the i^{th} trait on the response variable Y. Both positive and negative direct effects increase R^2.

(b) The entity $2P_{iy}P_{jy}r_{ij}$ is related to the common cause, C_{ij}, of traits X_i and X_j. Notice that the indirect effects $P_{iy}r_{ij}$ and $P_{jy}r_{ij}$ differ, but they both result from an influence of C_{ij}. Therefore, we just measure the contribution of C_{ij} to determination of the response variable; this contribution equals $2P_{iy}P_{jy}r_{ij}$. The two indirect effects under discussion should not be treated independently, because both of them are the effects of the common cause of the two independent variables (Li 1951). Therefore, we take P_{iy}^2 as the contribution to R^2 of the direct effect of the i^{th} trait, and $2P_{iy}P_{jy}r_{ij}$ as the contribution of the effect of the common cause of the i^{th} and j^{th} traits. This common cause can result in the compensation of these two traits. Note that the component $2P_{iy}P_{jy}r_{ij}$ can be negative and hence can decrease R^2. Obviously, it can happen in two ways; viz. when one of the path coefficients (P_{iy} and P_{jy}) or r_{ij} is, or all three of them are, negative.

On the basis of the explanations above, the following new coefficient is proposed:

$$Q_i = P_{iy}^2 + \sum_{j=1, j \neq i}^{k} P_{iy} P_{jy} r_{ij} \quad (i = 1, \ldots, k) \qquad (4)$$

Let us call coefficient Q_i the overall contribution of the i^{th} trait to the determination of the response variable, or simply the contribution of the i^{th} trait to R^2. An important characteristic of coefficient Q_i [Eq. 4] is that

$$R^2 = \sum_{i=1}^{k} Q_i$$

Theoretically, it looks possible that $Q_i < 0$ for some i; in such a case, the i^{th} trait would decrease the determination of the response variable. This process is difficult to understand and interpret because classical statistics does not take into account the possibility of decreasing the determination – it just assumes that a variable may positively, or may not at all, contribute to the determination of another variable. Therefore, this peculiar situation needs to be studied from the point of view of its statistical modeling and interpretation.

Such processes should be recognized via studying the decomposition in Eq. [4], and, in a breeding program, one should focus on minimizing effects that reduce the negative value of Q_i. We suggest, however, that the situation in which $Q_i << 0$ is only theoretical – in practice, it is possible to obtain $Q_i < 0$ but with a rather low absolute value.

We now concentrate on the meaning of a negative contribution of a common cause to the determination of the response variable. First, we know intuitively that such a contribution is undesirable, because we would prefer the determination coefficient of the response variable via studied traits to be as large as possible; a negative contribution decreases the R^2. If the contribution is large, it means that both traits involved in this indirect effect have a substantial effect on the response variable; it also means that they are correlated.

If two traits have a positive direct effect on the response variable and they are negatively correlated, then the contribution of the common cause is negative. Here we can point out a particular situation that is often encountered in practice. Suppose that two variables are strongly negatively correlated; the direct effect of one variable is positive and large, whereas that of the second variable is positive but negligible. The classical interpretation of path analysis would suggest that the indirect effect of the second trait via the first trait is negative and relatively large, whereas the new approach would suggest that the contribution of the common cause of the two traits is negligible. The conclusion is that the common cause makes a large (in its absolute value) contribution to R^2 only if all three coefficients involved in the contribution formula are large (irrespective of their sign). Thus, we do not have to be concerned about all negative indirect effects identified by the classical path analysis; some of them may turn out to be unimportant in the determination of the response variable, and hence we can ignore them.

Another interesting situation is when the coefficients P_{iy}, P_{jy}, and r_{ij} are all negative for some pair of i and j ($i \neq j$, $i, j = 1, \ldots, k$). Both indirect effects ($P_{iy}\, r_{ij}$ and $P_{jy}\, r_{ij}$) are then positive, but the contribution of their common cause is negative and decreases the determination of the response variable. Hence, we note that even positive indirect effects (in the classical meaning of indirect effects) can decrease R^2 and, in this sense, they are not desirable. Note that this situation is unlikely to be discovered in the classical interpretation of indirect effects. What should we conclude from such a result? Our efforts should likely be focused on reducing one or both of the direct effects then.

If this can be managed, a negative contribution of the common cause will be reduced.

Now let us determine which traits are desirable and which are undesirable in the determination of the result variable. If a trait has a positive direct effect on a response variable, then we prefer it to have a positive correlation with traits that have a positive direct effect (alternatively, the correlation may be negligible). Moreover, such a trait should be negatively correlated or uncorrelated with traits having a negative direct effect. On the other hand, if a trait has a negative direct effect on a response variable, then it should be negatively correlated or uncorrelated with traits having a positive direct effect, and positively or not correlated with those having a negative direct effect. Those are desired indirect effects; all other effects are undesirable or unimportant. In summary, we simply conclude that desired effects are those that represent a positive contribution of a common cause of any two traits. In the event of an undesirable situation, a contribution of the common cause of two traits that are involved in the undesirable indirect effect is negative.

To make the analysis clear, besides the classical path analysis table (see Williams et al. 1990), values of various quantities from Eq. [4] can be included. Let us recall that, in the classical analysis of pathways, one should avoid interpreting an indirect effect as an effect of an independent variable via another variable. An example of table construction for the new interpretation is presented in the next section.

We are now able to define new criteria for identifying relative importance of traits in determination of a response variable, alternative to the criteria given by Board et al. (1997). The new general criteria are:

(a) Overall contribution Q_i of the i^{th} trait to the determination coefficient of the response variable; the larger the overall contribution of a trait, the more important the trait in determination of the dependent variable.

(b) A correlation between the trait and the response variable (as in Board et al. 1997); [correlation coefficient in our situation is a measure of an overall effect of the i^{th} trait on Y (see Shipley 2002, p. 127)]; for a trait that would be highly desirable at a high level, this correlation should be positive.

(c) Direct effect of the trait on the response variable (as in Board et al. 1997); for a trait that would be highly desirable at high level, the direct effect should be positive.

(d) Contributions of common causes of the i^{th} trait with other traits; for a trait that would be highly desirable at a high level, the contributions of common causes with other traits should be positive (or nonexistent).

A summary table for the analysis (interpretation) via which we are able to show values of the coefficients involved in each criterion is provided. However, such a table usually does not provide all detailed information needed for interpretation; in some cases (especially when the number of variables in causal system is large and the variables are correlated), additional tables providing detailed information may be required. On the basis of the tables, we are able to draw conclusions about the importance of a trait in determining the response variable as well as about the direction of its influence. Furthermore, on such a basis, we can determine which traits we should concentrate on, and at what level we should try to set them. Moreover, we can provide a detailed explanation of a whole process of determination of the resultant variable by the traits under study.

When a particular criterion is identified, we should decide whether or not the overall contribution of a trait is important (we do not have to decide whether it is positive or negative). It is done via its magnitude (not statistical testing – we know that even a coefficient that significantly differs from zero may be unimportant). Therefore, understanding the nature of the problem/situation under consideration would be very helpful here.

The most important criterion is the first one. A near-nil Q_i represents no contribution of the trait under consideration to R^2. The next most important criteria are the second and third ones. The fourth criterion is probably the least important, especially when the first three criteria prevail. In general, we can state that the combined interpretation in path analysis, that is, interpretation that bridges the classical and proposed approaches, yields more detailed and exhaustive overall picture of the process of formation of a response variable, in comparison with classical path analysis. Following the discussion of Rencher (1998, p. 210) on interpretation of discriminant functions and determining the contribution of each variable, classical interpretation in path analysis concerns independent variables influencing response variable, whereas our approach concerns contributions of the independent variables to the determination of the response variable. Hence, these two interpretations are very different.

An example

A winter-rye experiment was carried out in 2002 at the Experimental Station of Faculty of Agriculture and Biology of the Warsaw Agricultural University, which was established in 1922 and is located in Skierniewice (51°58’ latitude, 20°10’ longitude, and 120 m above sea level). The climate of the Experimental Station is characterized by a mean annual temperature of 7.9°C (10.1°C in 2002) and an annual rainfall of 527 mm (506.3 mm in 2002), evenly distributed during the growing season, except in April. In April 2002, rainfall was about 10 mm. The trial was conducted on stagnic luvisol-loam sandy soil containing 6–8% clay in the plow layer.

A long-term fertilizer experiment was established during 1922–1924, and comprised three groups of fields: A1–A4, with treatments 0, CaNPK, NPK, PK, NP and NK and with ammonium nitrate as the source of nitrogen; A5–A8, with treatments Ca, CaNPK, NPK, CaPK, CaNP and CaNK and with the same nitrogen fertilization as for A1–A4 fields; AF1–AF2 fields with the same fertilizer treatments as for A1–A4 fields, but fertilized with ammonium sulfate. The control treatments (0 or Ca) had four replications, and the others had three replications. The doses of all nutrients were the same for all fields and amounts: N = 90 kg/ha, P = 26 kg/ha (superphosphate) and K = 91 kg/ha (potassium chloride). CaO at a rate of 1.6 t/ha was applied every four years on treatments designated to receive Ca. On all these fields, arbitrary crop rotation without legumes was practiced. Most fields were not fertilized with organic manures; on fields A4, A5 and AF2, farmyard manure had been applied every 3 to 4 years since 1992–1994.

In the present work, only the fertilizer treatments with acid and strongly acid soil are considered (we have chosen those plots where pH values were lower than 5.5). There were 47 such plots from the treatment combinations CaNPK, NPK, PK, NP and NK from fields A2, A4, AF1 and AF2, and NPK from A5 and A8 fields. These plots were treated as a sample from a population of an acid-soil environment.

The study considered grain yield of rye cv. Dańkowskie Złote. Soil properties were measured after harvesting the crop. Soil analyses comprised, among others, the following properties: pH in KCl (1 mol/dm^3), Hh – hydrolitic acidity (extraction with calcium acetate, 1 mol/dm^3, pH 8.2), exchangeable Al (1 mol/dm^3 KCl), exchangeable cations content: Ca, Mg, and K (1 mol/dm^3 ammonium acetate), available P and K (Egner-DL method), and total N content (measured using the modified Kjeldahl method).

We assumed that the soil properties studied are set at the same ontogenetic level. Some might criticize such an assumption and argue that there is no evidence that soil properties are correlated and not connected via other causal structure. In our opinion, soil properties are modeled via the system under consideration, and co-relations among them are the result of their common causes. There is no evidence or even circumstantial evidence that any of the soil properties studied is a cause or an effect of any other property included in the causal system.

The backward stepwise selection was used to select the variables that significantly affected grain yield. Three soil properties were selected: available (K_{av}) and exchangeable potassium (K_{exch}) content, and total nitrogen content (N_{tot}).

Hereafter, we use just the above-noted three selected independent variables, omitting the others (as they are assumed not to determine grain yield). Table 1 contains summary statistics for grain yield and selected soil properties. Table 2 represents a correlation matrix for the traits. Grain yield was significantly correlated with total N content

Table 1. Summary statistics (mean values – MV, standard deviations – SD, and coefficients of variation – CV) for grain yield and selected soil properties

Variable	MV	SD	CV (%)
Grain yield (t/ha)	2.96	0.88	29.9
K_{exch} (mmol/kg)	2.92	0.79	27.1
K_{av} (mg/kg)	42.85	13.72	32.0
N_{tot} (g/kg)	0.45	0.05	10.6

K_{exch} – K exchangeable, K_{av} – K available, N_{tot} – total nitrogen content

Table 2. Correlation matrix for winter rye grain yield and selected soil properties

	Grain yield	K_{exch}	K_{av}
K_{exch}	–0.31*	1	
K_{av}	0.04	0.71**	1
N_{tot}	0.76**	–0.14	–0.01

K_{exch} – K exchangeable, K_{av} – K available, N_{tot} – total nitrogen content
*, **significant at 0.05 and 0.01 probability level, respectively

Table 3. Path analysis for winter rye grain yield as affected by selected soil properties: exchangeable (K_{exch}) and available (K_{av}) K content, and total N content (N_{tot})

	K_{exch}	K_{av}	N_{tot}
K_{exch}	*–0.492***	–0.350	0.070
K_{av}	0.283	*0.398***	–0.004
N_{tot}	–0.098	–0.007	*0.685***
Corr. with grain yield (*Y*)	–0.31*	0.04	0.76**

*, **significant at 0.05 and 0.01 probability level, respectively (for direct effects and correlation coefficients); italics is used for direct effects

(a strong positive correlation) and exchangeable K (a weak negative correlation). Among the selected soil properties, only exchangeable and available K content were significantly correlated (a strong positive correlation).

Table 3 contains results of path analysis for the studied model. The largest positive direct effect on grain yield was recorded for N_{tot}. K_{exch} influenced yield negatively. The lowest influence on yield was detected for K_{av}; it was a positive direct effect. A large direct effect of N_{tot} led to a large correlation between this soil property and grain yield (no meaningful indirect effects of this variable were detected). A fairly large negative direct effect of K_{exch} was counterbalanced by the positive indirect effect of the common cause of this variable and K_{av}. The common cause of this variable and K_{exch} almost completely compensated the positive direct effect of K_{av}; it resulted in the near-zero correlation between K_{av} and grain yield.

Tables 4 and 5 contain interpretations according to the proposed approach. In Table 4, the decomposition of yield determination (R^2) into components relative to direct and indirect effects of the independent variables on yield is presented. The largest contribution to yield determination was that of the direct effect of N_{tot}. Next, a large contribution was detected for the common cause of two forms of potassium (exchangeable and available), but this component was negative and decreased the yield determination. The direct effect of K_{exch} contributed substantially to R^2, too. Contribution of the direct effect of K_{av} was positive but not large. Contributions of the common causes of N_{tot} and both potassium forms (especially the available one) were rather small.

Table 5 represents a summary for the whole analysis. Let us go over the proposed criteria for identifying important traits that determine the response variable. The first trait, exchangeable potassium content, contributed little to yield determination, was negatively correlated with grain yield, had a relatively large negative direct effect on grain yield, and the contributions of its common causes with the other traits were, in summary, negative. The second trait, available potassium content, had near-nil overall contribution to R^2, was not correlated with yield, had a relatively large positive direct effect on yield, and the contributions of its common causes with the other traits were negative or non-existent. Total nitrogen content had a large positive direct effect and a large overall contribution to R^2, was positively correlated with grain yield, and the contributions of its indirect effects were negligible.

The conclusion from the combined path analysis is that mainly total nitrogen content, especially the direct effect of this soil property, determined winter-rye grain yield. Within the range of total nitrogen content recorded in our study, we would want to maximize it. The exchangeable potassium content determined grain yield, but to a noticeably lesser extent. In the case of this soil property, we would want to achieve a lowest value (in the range of this soil property that occurred in the experiment, of course). Available potassium content was the soil property that generally weakly determined yield in spite of its positive direct effect on it; actually, a value of the available potassium content would have no influence on final grain yield.

The new interpretational approach to infer from causal systems from Figure 1 is a tool that can enrich the classical interpretation of pathways. It is based on the assumption that two independent

Table 4. Decomposition of winter rye grain yield determination (R^2)

Effect	P_1^2	P_2^2	P_3^2	$2P_1P_2r_{12}$	$2P_1P_3r_{13}$	$2P_2P_3r_{23}$	$R^{2\dagger}$
Value	0.242	0.158	0.469	–0.279	0.096	–0.005	0.682

$^\dagger P_1^2 + P_2^2 + P_3^2 + 2P_1P_2r_{12} + 2P_1P_3r_{13} + 2P_2P_3r_{23} = R^2$

Table 5. Checklist of new criteria for three studied soil properties in determining winter rye grain yield

Selection criterion	K_{exch}		K_{av}		N_{tot}	
	value	+/–†	value	+/–†	value	+/–†
Q_i (overall effect on grain yield)	0.15	ni.	0.02	ni.	0.52	+
Correlation with grain yield	–0.31	–	0.04	ni.	0.76	+
Direct effect	–0.49	–	0.40	+	0.69	+
Undesired common causes‡	–0.279	–	–0.284	–	–0.005	ni.
Desired common causes§	0.096	ni.	0	ni.	0.096	ni.
Final decision	little important trait desired at – level		not important trait		important trait desired at + level	

K_{exch} – K exchangeable, K_{av} – K available, N_{tot} – total nitrogen content

†+ and – mean that the particular criterion is at its high and low level, respectively; ni. means that the criterion is not important

‡sum of desired contributions of common causes with other traits

§sum of undesired contributions of common causes with other traits

variables in the causal system are correlated through a common cause that affects the two variables; it is, however, an inappropriate approach when this assumption is not met. By applying the proposed procedure, we can assess the contributions to the determination of the response variable of the independent variables from the model under consideration as well as the common causes of each pair of the independent variables. Researchers attempting to study a causal system should be aware of SEM methodology, which is the most powerful tool for studying causal structures. Nevertheless, if SEM cannot be applied for any reason and the model of study represents the model considered in this paper, our methodology presented above may be applied in order to enrich one's interpretation.

REFERENCES

Board J.E., Kang M.S., Bodrero M.L. (2003): Yield components as indirect selection criteria for late-planted soybean cultivars. Agron. J., *95*: 420–429.

Board J.E., Kang M.S., Harville G. (1997): Path analyses identify indirect selection criteria for yield of late-planted soybean. Crop Sci., *37*: 879–884.

Bollen K.A. (1989): Structural Equations with Latent Variables. John Wiley and Sons, New York.

Das M.K., Fuentes R.G., Taliaferro C.M. (2004): Genetic variability and trait relationships in switchgrass. Crop Sci., *44*: 443–448.

Dewey D.R., Lu K.H. (1959): A correlation and path-coefficient analysis of crested wheatgrass seed production. Agron. J., *51*: 515–518.

Dofing S.M., Knight C.W. (1992): Alternative model for path analysis of small-grain yield. Crop Sci., *32*: 487–489.

Güler M., Sait A.M., Ulukan H. (2001): Determining relationships among yield and some components using path coefficient analysis in chickpea (*Cicer arietinum* L.). Eur. J. Agron., *14*: 161–166.

Ige D.V., Akinremi O.O., Flaten D.N. (2007): Direct and indirect effects of soil properties on phosphorus retention capacity. Soil Sci. Soc. Am. J., *71*: 95–100.

Kang K.M., Seneta E. (1980): Path analysis: An exposition. In: Krishnaiah P.R. (ed.): Developments in Statistics. Vol. 3. Academic Press, New York: 217–246.

Kozak M., Kang M.S. (2006): Note on modern path analysis in application to crop science. Commun. Biometry Crop Sci., *1*: 32–34.

Li C.C. (1951): Population Genetics. The University of Chicago Press, Chicago.

Mohammadi S.A., Prasanna B.M., Singh N.N. (2003): Sequential path model for determining interrelationships among grain yield and related characters in maize. Crop Sci., *43*: 1690–1697.

Quinn G.P., Keough M.J. (2002): Experimental Design and Data Analysis for Biologists. Cambridge University Press, Cambridge.

Rencher A.C. (1998): Multivariate Statistical Inference and Applications. John Wiley and Sons, New York.

Samonte S.O.P., Wilson L.T., McClung A.M. (1998): Path analyses of yield and yield-related traits of fifteen diverse rice genotypes. Crop Sci., *38*: 1130–1136.

Seker H., Serin Y. (2004): Explanation of the relationships between seed yield and some morphological traits in smooth bromegrass (*Bromus inermi* Leyss.) by path analysis. Eur. J. Agron., *21*: 1–6.

Shipley B. (2002): Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference. Cambridge University Press, Cambridge.

Simane B., Struik P.C., Nachit M.M., Peacock J.M. (1993): Ontogenetic analysis of yield components and yield stability of durum wheat in water-limited environments. Euphytica, *71*: 211–219.

Steel D. (2005): Indeterminism and the causal Markov condition. Brit. J. Philos. Sci., *56*: 3–26.

Wang G., Kang M.S., Moreno O. (1999): Genetic analyses of grain-filling rate and duration in maize. Field Crops Res., *61*: 211–222.

Williams W.A., Jones M.B., Demment M.W. (1990): A concise table for path analysis statistics. Agron. J., *82*: 1022–1024.

Wright S. (1921): Correlation and causation. J. Agr. Res., (Wash., D.C.), *20*: 557–585.

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

Zhang H., Schroder J.L., Fuhrman J.K., Basta N.T., Storm D.E., Payton M.E. (2005): Path and multiple regression analyses of phosphorus sorption capacity. Soil Sci. Soc. Am. J., *69*: 96–10.

Zheng Z., Simard R.R., Lafond J., Parent L.E. (2007): Pathways of soil phosphorus transformations after 8 years of cultivation under contrasting cropping practices. Soil Sci. Soc. Am. J., *66*: 999–1007.

Received on March 7, 2007

Corresponding author:

Dr. Ing. Marcin Kozak, Warsaw Agricultural University, Faculty of Agriculture and Biology, Department of Biometry, Nowoursynowska 159, 02 787 Warsaw, Poland
e-mail: m.kozak@omega.sggw.waw.pl