Ontogenetic Chain and the Markov Condition in Crop Science

Marcin Kozak

Department of Biometry, Warsaw University of Life Sciences, Nowoursynowska 159, 02-787 Warsaw, Poland; E-mail: m.kozak@omega.sggw.waw.pl

Abstract: This paper discusses an ontogenetic chain, the specific form of a cause-and-effect system among traits, in relation to crop science. Two approaches to modeling associations in such a system are presented, one based on a so-called sequential approach and another, on the Markov condition. It is shown that usually the former has much stronger biological basis than the latter. [Nature and Science. 2007;5(3):5-8].

Keywords: cause, effect, interpretation, yield components.

The general definition of an ontogenetic chain in relation to plant traits is that the traits that constitute it develop in a certain order during ontogeny, a fact that may result in a particular form of relationships among the traits (Mądry et al., 2005). We may distinguish two main approaches to postulating these relationships ("postulating" because at the very beginning of the investigation we do not know which of the relationships are true and important and which are not, but we may postulate which traits *may be* causes or effects of other traits). The first approach consists in applying what we will hereafter call the all-effect sequential model (the ALL-SEQ model); and the second, in applying the model based on the Markov condition (the MC model), a model that represents the causal chain, quite common in the philosophy of causation. In fact, both of these models are "sequential" because they both take into account the sequence of events in the ontogenetic chain. The name "all-effect sequential model" comes from the fact that the model comprises all the direct and indirect effects that are possible in the ontogenetic chain (see Figure 1); methods that use this model are, for example, sequential yield component analysis (e.g., Eaton and Kyte, 1978; Eaton and MacPherson, 1978) and sequential yield analysis (e.g., Mądry et al., 2005; Kozak et al., 2006).

We are not discussing here the methods of analyzing relationships among the traits in the ontogenetic chain. Our aim is to discuss models that describe these relationships; such a model is to be postulated prior to the analysis, and it may influence its results. (Note that some methods, e.g., path analysis, may be used to study a causal process based on various models; hence the choice of a model matters.) We assume that such postulation is to be done based on the knowledge of a process one aims to study. Therefore, for example, we know that SPAD measurement in the DC 31 stage (SPAD 31; Zadoks et al., 1974) may influence SPAD 49, but the opposite situation (in which SPAD 49 would influence SPAD 31) is impossible (Samborski et al., 2006).

Figure 1 presents the ALL-SEQ model for three traits. In this system the traits that are prior in the chain to other traits may influence them, but may not be influenced by those traits that follow them. A particular, jth trait (X_j) may be, then, determined by all the traits previous to it, that is, X_i for $i = 1, \ldots, j - 1$, and may determine the traits that follow it, that is, X_k for $k = j + 1, \ldots, p$, where p is the number of traits in the ontogenetic chain. Therefore, it is assumed possible that the influence of the jth ($j = 2, \ldots, p - 2$) trait on the last, pth trait (which is often called the dependent trait) may have a direct as well as indirect character, the latter via the traits that follow X_j in the chain (of course, the last but one trait may have only the direct effect on the last one). So, the first trait in the chain may influence all the other traits, not being influenced by any of them, whereas the last trait may be influenced by all other traits, not influencing any of them. This type of causal system is often used to model the associations among traits in an ontogenetic chain in agronomy and crop science; see, for example, Grafius (1969), Rasmusson and Cannell (1970), Thomas et al. (1971), Eaton et al. (1986), McArthur and Eaton (1988), Freeman et al. (1989), Bowen and Kliewer (1990), Akwilin Tarimo (1991), Dofing and Knight (1992), Shamaila et al. (1992), Bos and Spaarnaij (1993), Spaarnaij and Bos (1993), Gołaszewski (1996), Gołaszewski et al. (1998, 2001), Spaner et al. (1996, 2000, 2001), Mądry et al. (2005), Samborski et al. (2006), and many others.

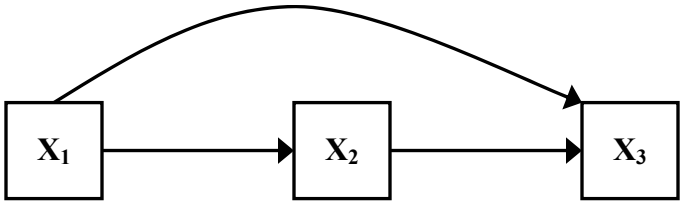

Figure 1. The all-effect sequential (ALL-SEQ) model presenting the associations among the traits in an ontogenetic chain (for three traits).

The modeling based on the MC model is different from that based on the ALL-SEQ model. A description of the Markov condition may be found, for example, in Pearl (2000); its illegible explanation in relation to ecology, in which this model has found many applications (see, e.g., Baker,1989 and Tucker and Anand 2004 and the citations therein), but also plant physiology, was presented by Shipley (2002, sec. 2.7). Under the Markov condition a *j*th ($j > 1$) trait in the ontogenetic chain (note that this paper is concerned with the ontogenetic chain so the Markov condition is discussed only in this context, although it may also be used for other models) is determined *only* by the trait that the *j*th trait follows. Analogously, every trait (except for the last one) in the chain is assumed to be a cause of only one trait: the one it precedes. In this way every trait in the model may be a cause of only one trait as well as an effect of only one trait. The model of this type, for three traits, is presented in Figure 2.

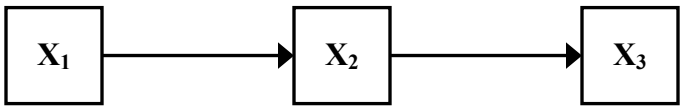

Figure 2. The model based on the Markov condition (the MC model) presenting the associations among traits in an ontogenetic chain (for three traits).

The choice of one of these two models is essential: it is likely that the interpretations based on them will be different (it is likely but not certainly because it may happen that the model in Figure 2 will be true; in such an instance this should be shown also by the analysis of the ALL-SEQ model). Below we compare the two models for one of the most common crop science problems—yield component analysis (Kozak and Mądry, 2006), although what will be said may be linked to many other causal systems in crop science.

Consider the following components of cereal grain yield (GY): number of spikes per unit area (NS), number of grains per spike (NG), and kernel weight (usually presented as 1000-kernel weight, TKW). The order of the traits in the ontogenetic chain is acknowledged to be as follows (e.g., Dofing and Knight 1992): NS, NG, TKW and GY. The MC model assumes in this case that the influence of TKW on GY is direct; the influence of NG on GY is *only* indirect, via TKW; and the influence of NS on GY is *only* indirect, via the path NS → NG → TKW → GY. Thus, the only yield component that directly affects cereal grain yield would be 1000-kernel weight. It is easy to imagine that not too many crop scientists would agree with this statement. Number of spikes per unit area may also have an influence (for example, the direct effect) on cereal grain yield of different character from that along the above-given path (e.g., Rozbicki, 1997). How, then, in the light of the above discussion should we interpret the MC model, in which number of spikes per unit area is *assumed* to affect grain yield *only* indirectly?

In contradiction to the MC model, the ALL-SEQ model assumes that each trait (yield component in our example) *may* (so does not have to) affect the final trait in the chain (grain yield) directly and, for all traits except for the last but one trait, indirectly.

Let us now come back to the example concerned with SPAD measurements. Consider the following ontogenetic chain: SPAD 31, SPAD 49 and winter triticale grain yield. Why should we assume that SPAD 31 has no direct influence on grain yield and the only influence it has on grain yield is indirect, via SPAD 49? Is there any knowledge that would support such a statement? If no, we may simply choose the ALL-SEQ model and try to find which associations (see Figure 1) are and which are not significant and important. Note once more that we do not assume that all associations from the ALL-SEQ model are significant: we assume they are possible.

These simple examples show that the MC model quite often is not correct in crop science. For many, which does not mean that for all, causal processes in crop science that are based on the ontogenetic chain the ALL-SEQ model has much stronger biological basis than the MC model. Of course, this is a researcher's task to decide which model to choose, and it never should be done by convention: the knowledge of the process should provide necessary information.

Nonetheless, looking more generally at the MC model it is easy to notice that this type of approach to analyzing cause-and-effect relationships is quite a simplification of a process represented by the ontogenetic chain. One could say that the assumption that each trait (except for the first one in the ontogenetic chain) has only one cause, that is, the preceding trait, is rather artificial and usually has nothing to do with the biology of the processes. And note that by applying the MC model we are not studying whether or not this is true: we *assume* this. Is there any crop scientist who would agree with the following conclusion: "The only influence that number of kernels per spike has on cereal grain yield is indirect, via 1000-kernel weight, and the only influence that number of spikes per unit area has on cereal grain yield is that via the following path: number of spikes per unit → number of kernels per spike → 1000-kernel weight → grain yield"? A question arises: Is such a simplification of biology appropriate?

In the light of the discussion presented in this paper we may acknowledge that the ALL-SEQ model in crop science is *usually* correct, in contradiction to the model based on the Markov condition, which usually does simplify the biology of crop science processes. This is why Dofing and Knight's (1992) choice of the model of how components influence small grain yield may be acknowledged to be correct. And this is why the choice to use the ALL-SEQ model, not the MC model, in so many analyses (see the papers cited above) may be acknowledged to be correct.

Correspondence to:
Marcin Kozak
Department of Biometry,
Warsaw University of Life Sciences,
Nowoursynowska 159, 02-787 Warsaw, Poland;
E-mail: m.kozak@omega.sggw.waw.pl

Received: July 11, 2007

References

1. Akwilin Tarimo J.P. Physiological response of groundnut to plant population density. African Crop Science Journal 1997;5:267-272.
2. Baker W.L. A review of models of landscape change. Landscace ecology 1989;2(2):111-133.
3. Bos I., Sparnaaij L.D. Component analysis of complex characters in plant breeding. II. The pursuit of heterosis. Euphytica 1993;70:237-245.
4. Bowen P.A., Kliewer W.M. Influence of clonal variation, pruning severity, and cane structure on yield component development in 'Cabernet Sauvignon' grapevines. Journal of the American Society for Horticultural Science 1990;115:530-534.
5. Dofing S. M., Knight C. W. Alternative model for path analysis of small-grain yield. Crop Science 1992;32:487-489.
6. Eaton G.W., Bowen P.A., Jolliffe P.A. Two-dimensional partitioning of yield variation. HortScience 1986;21:1052-1053.
7. Eaton, G.W. Kyte T.R. Yield component analysis in strawberry. Journal of the American Society for Horticultural Science 1978;103:578-583.
8. Eaton G.W., MacPherson E.A. Morphological components of yield in cranberry. Horticultural Research 1978;17:73-82.

9. Freeman J.A., Eaton G.W., Baumann T.E., Daubeny H.A., Dale A. Primocane removal enhances yield components of raspberries. Journal of the American Society for Horticultural Science 1989;114:6-9.
10. Gołaszewski J. A method of yield component analysis. Biometrical Letters 1996;33(2):79-88.
11. Gołaszewski J., Idźkowska M., Milewska J. The TDP method of seed yield component analysis in grain legume breeding. Journal of Applied Genetics 1998;39(4):299-308.
12. Gołaszewski J., Idźkowska M., Milewska J., Koczowska I. Yield component analysis with SYCA and TDP in fodder pea. Plant Breeding and Seed Science 2001;45(2):77-85.
13. Grafius J.E. Stress: A necessary ingredient of genotype by environment interactions. International Barley Genet. II, Wash. State Univ. Press. 1969:346-355.
14. Kozak M., Gozdowski D., Hossain S., Ahmed S.E., Laudański Z., Wyszyński Z. Canonical correlations in studying grain yield and protein content as affected by yield components: An ontogenetic approach. Plant Breeding and Seed Science 2006;54:17-27.
15. Kozak M., Mądry W. Note on yield component analysis. Cereal Research Communications 2006;34 (2-3):933-940.
16. Mądry W., Kozak M., Pluta S., Żurawicz E. A new approach to sequential yield component analysis (SYCA): Application to fruit yield in blackcurrant (Ribes nigrum L.). Journal of New Seeds 2005;7(1):85-107.
17. McArthur D.A.J, Eaton G.W. Strawberry yield response to fertilizer, paclobutrazol and chlormequat. Scientia Horticulturae 1988;34:33-45.
18. Pearl J. Causality: Models, reasoning, and inference. Cambridge: Cambridge University Press; 2000.
19. Rasmusson D.C., Cannell R.Q. Selection for grain yield and components of yield in barley. Crop Science 1970;10:51-54.
20. Rozbicki, J. Agronomical determination of winter triticale growth, development and yielding. Fundacja "Rozwój SGGW", Warsaw;1997 (in Polish, with English abstract).
21. Samborski S., Kozak M., Rozbicki J. The usefulness of chlorophyll meter SPAD-502 for winter triticale grain yield estimation. Folia Universitas Agriculturae Stetinensis, Agricultura 2006;247(100):157-162.
22. Shamaila M., Baumann T. E., Eaton G. W., Powrie W. D., Skura B. J.. Quality attributes of strawberry cultivars grown in British Columbia. Journal of Food Science 1992;57(3):696-699.
23. Shipley B. Cause and correlation in biology: A user's guide to path analysis, structural equations and causal inference. Cambridge University Press, Cambridge; 2002.
24. Spaner D., Mather D.E., Brathwaite R.A.I. Yield and quality attributes of rainfed green corn in Trinidad. HortTechnology 1996;6:131-134.
25. Spaner D., Todd A.G., McKenzie D.B. The effect of seeding date, seeding rate and N fertilization on winter wheat yield and yield components in Eastern Newfounland. Canadian Journal of Plant Science 2000;80:703-711.
26. Spaner D., Todd A.G., McKenzie D.B. The effect of seeding rate and nitrogen fertilization on barley yield and yield components in a cool maritime climate. Journal of Agronomy & Crop Science 2001;187:105-110.
27. Sparnaaij L.D., Bos I. Component analysis of complex characters in plant breeding. I. Proposed method for quantifying the relative contribution of individual components to variation of the complex character. Euphytica 1993;70:225-235.
28. Thomas R.L., Grafius J.E., Hahn S.K. Transformation of sequential quantitative characters. Heredity 1971;26:189-193.
29. Tucker B.C., Anand M. The application of Markov models in recovery and restoration. International Journal of Ecology and Environmental Sciences 2004;30:131-140,
30. Zadoks J.C., Chang T.T., Konzak G.F. A decimal code for growth stages of cereals. Weed Research 1974;14:415-421.