Multiplicative yield component analysis: what does it offer to cereal agronomists and breeders?

M. Kozak[1], **M.R. Verma**[2]

[1]*Department of Experimental Design and Bioinformatics, Warsaw University of Life Sciences, Warsaw, Poland*
[2]*Division of Agricultural Economics and Statistics, ICAR Research Complex for NEH Region Umroi Road, Umiam (Barapani), Meghalaya, India*

ABSTRACT

In this paper we have discussed the problem of decomposition of a response variable into other predictor variables whose product gives the value of the response variable. This methodology is known as multiplicative yield component analysis. We have discussed the applications of multiplicative yield component analysis in cereal science and breeding. The discussion of multiplicative yield component analysis involves the description of yield component analysis for cereal grain yield, single genotype yield component analysis and across-genotype yield component analysis.

Keywords: yield component analysis; single-genotype yield component analysis; across-genotype yield component analysis

Agronomists and plant breeders are very often faced with so-called yield components, which are such plant and crop traits the product of which gives yield. Yield component analysis is a general methodology of analyzing a causal model of how yield components affect yield. In this way the information on the importance of particular yield components is extracted, which can be interesting for agronomists and plant breeders for various reasons. We discuss these reasons in this paper.

Yield component analysis was applied in numerous applications, and its theoretical issues were studied quite deeply (see Kozak and Mądry 2006). Still there are some issues to be solved and discusses (for example, methodology for multiplicative yield components that develop in sequential order during ontogenesis, extracting direct and indirect effects from the analysis; see Kozak and Mądry 2006 and Kozak et al. 2007a). Here we will deal with one of such issues, namely how the results of yield component analysis should be used, and what kind of conclusions might be drawn based on them. Despite the fundamental importance of this issue, no account of it might be found in the literature. Applications of yield component analysis are diverse and so are the conclusions that can be drawn based on them; some conclusions are of critical importance for a given crop species, while others seem to be drawn without appropriate and desirable consideration.

First, let us formally define what yield component analysis is. A yield-component model is one in which components are traits the product of which gives yield, that is (e.g., Fraser and Eaton 1983, Spaarnaij and Bos 1993, Piepho 1995),

$$Y = \prod_{i=1}^{k} X_i \qquad (1)$$

where: Y stands for yield and X_i, $i = 1, \ldots, k$, for the i^{th} component

Owing to the multiplicative character, the model (1) should be called the multiplicative yield-component model to distinguish it from an additive yield-component model (Jolliffe and Courtney 1984). For simplicity, henceforth we will call it yield-component model, and the whole methodology yield component analysis (instead of multiplicative yield component analysis). If the product of the traits at the right-hand side of (1) does not give yield, at least in theory, such traits cannot be called yield components (Kozak and Mądry 2006). In this paper we deal only with yield components in the strict sense provided by the model (1).

Yield component analysis – hereafter YCA – is a general methodology aiming to probe into the effects of yield components on yield. This modeling should take account of the specific form of the model (1); including the full determination of yield by the components (note the lack of the error term in the model; Kozak and Mądry 2006). YCA is a general methodology because not one but many statistical approaches and methods have been constructed and applied for YCA; we will touch upon this problem in next section.

Although the methodology's name suggests it is yield what is of interest, in general YCA can be applied for any set of traits that follow the yield-component model (1) and for which the causal structure can be a priori stated through model (1). Such a model is considered, for example, for nitrogen uptake studied as the product of nitrogen-uptake efficiency and crop nitrogen supply (Samborski et al. 2008); for oil yield per plant studied as the product of number of capitula per plant, number of achenes per capitulum, achene mass and oil concentration (Abbadi et al. 2008); or in various ecological problems, for example one in which the lifetime productive success of an individual in a wild animal population is studied as the product of its reproductive lifetime in years, average birth rate per year, and average proportion of offspring that survive to be adults (Brown and Alexander 1991).

Having said that, in this paper we will focus on the classical yield component analysis in the agronomy/plant breeding sense, in which yield is analyzed as the product of at least two plant traits. We will also limit the discussion to cereals, for which grain yield is considered as the product of its three following components: number of spikes per unit area, average number of kernels per spike (commonly called number of kernels per spike), and average kernel weight (commonly presented as thousand kernel weight and called the kernel weight). Although cereal grain yield may be also studied as the product of biomass yield and harvest index (Kozak et al. 2007a), we will not consider this situation and will focus on what we could call the classical yield components. A set of such classical yield components could probably be defined for any plant species: For legumes, for example, this could be seed yield per unit area considered as the product of number of plants per unit area, number of branches per plant, number of pods per branch, number of seeds per pod and seed weight (Gołaszewski et al. 1998); for root species, root yield can be considered as the product of number of plants per unit area and root yield (Hűhn 1987); and so forth.

The main question asked by those who apply YCA is which of the components are most important in determining final yield. The questions we ask in this paper are: With the YCA results in hand, what conclusions can one draw about yield and its components? What are the implications of YCA? We will try to address these issues and find out what really underlies yield components and their influence on cereal grain yield. To meet this aim, we introduce concepts of single-genotype yield component analysis and across-genotype yield component analysis; these concepts lay the foundations for a proper interpretation of the YCA results, no matter which statistical method has been applied. We hope our considerations create a brief account of how YCA should be used to conclude about optimizing yield and its components, and how YCA should be interpreted and understood.

Yield component analysis for cereal grain yield: methodology. It was probably Engledow and Wadham (1923) who first proposed to analyze cereal grain yield as a result of its three classical components, namely number of spikes per unit area, number of kernels per spike, and thousand kernel weight. Later, this approach was taken up by many others, the effect of which was the acceptance of yield components as one of the most important sets of traits that contribute to final yield. It resulted in numerous applications of YCA for a number of plant species, among which cereals likely play the key role.

Description of statistical methods for YCA is beyond the scope of this paper; let us, however, remind what constitutes their essence. First, an efficient statistical method for YCA should take account of two specific features of the model (1), namely its multiplicative character and lack of a residual term. At first, classical statistical methods such as correlation or regression were applied (Fraser and Eaton 1983), thus these two features were disregarded. Later on, statisticians decided that the model (1) requires a specific approach. In addition, a new problem emerged when it appeared clear that some components develop in a so-called sequential order while others develop simultaneously. That cereal yield components develop in sequential order is now a generally accepted standpoint; see e.g. Dofing and Knight (1992) and Kozak and Mądry (2006). The former authors presented a discussion on the sequential development of small grain cereals, pointing out the sequence in which the yield and its components develop, namely number of spikes per unit area, mean number of kernels per spike, kernel weight, and grain yield. For discussion on sequential aspect of yield component analysis the readers are referred to Kozak and Mądry (2006).

Several statistical approaches for YCA that took account of the above issues were proposed. These were, among others, sequential yield component analysis SYCA (Eaton and Kyte 1978, Eaton and MacPherson 1978), Two-Dimensional Partitioning of Yield Variation (Eaton et al. 1986, see Kozak 2006 for the critical discussion of this method), and approaches by Hűhn (1987), Brown and Alexander (1991), Bos and Spaarnaij (1993) and its modification by Kozak and Mądry (2005), Piepho (1995), Kozak (2004), etc.

For simplicity, henceforth by yield we will understand cereal grain yield, and by yield components, its three following multiplicative components: number of spikes per unit area, number of kernels per spike, and average kernel weight; YCA will be used to denote the multiplicative yield component analysis for cereal grain yield and its three above-mentioned components.

Single-genotype yield component analysis. The most common application of YCA is to study the influence of components on yield for a particular genotype. This approach helps identify the most important components in determining final grain yield of this genotype. This in turn may suggest to which of the components most of the attention should be paid when searching for optimal agronomical practices for the genotype. If, for a given species, it is possible to control these most important components, it is generally assumed that in this way one might control grain yield. For a particular genotype this control can be done practically only through agronomic practices and for cereals such control is rather limited. For any species, number of plants per unit area is the easiest yield component to control, but the extent to which this can be done strongly depends on a species. For sugar beet, for example, this can be done very efficiently; for cereals, on the other hand, number of spikes per unit area (which is considered instead of number of plants per unit area) is much more difficult to control, yet this is the main controllable yield component.

Nevertheless, this indirect type of approach towards optimizing grain yield seems to be somewhat longer way to go than the classical optimizing grain yield, when one seeks the best agronomic practices for yield itself and discounts the components' value. In fact, whichever the values of yield components, if only grain yield is at the optimum level, the aim of searching for the optimum agronomical practices is reached. Focusing on grain yield is an overall approach and despite its generality, or rather thanks to it, it takes account of the complexity of the relationships among the components. Focusing on chosen yield components, on the other hand, leads to limited conclusions, concerned mainly with those selected components. These two kinds of approaches are not necessarily parallel, and may lead to different conclusions (i.e., different agronomic recommendations).

A different situation one faces when not only grain yield is of interest, but also one or more of the components. For example, kernels may be required to be heavier than a specified weight. YCA can then provide information whether for a given genotype it is possible to obtain high yield together with large kernels; this, however, should be studied across various environments because it is quite likely that YCA may give different results in different environments.

Note that in this example it is not necessarily YCA that should be applied. YCA's theory deals with analyzing all components from model (1) [even though in the past, univariate methods, such as simple regression, were applied; in this paper, following the contemporary point of view, we assume that YCA analyzes all the components from model (1)], not a chosen one. So in fact when one wishes to focus on the relationship between one particular component and yield it determines, it should be better to carry out a univariate analysis for influence of this component on yield, an analysis that would aim to provide interpretation and conclusions one wants to draw.

Indeed, environment is quite likely to influence the results of YCA, which is another important issue of YCA. This makes impossible to claim which of the components for a particular cultivar is the most important. For example, Samborski et al. (2005) studied two winter triticale cultivars Bogo and Fidelio. For the former cultivar, mean number of kernels per spike and kernel weight were equally most contributing components to grain yield determination, while number of spikes per m^2 had very small contribution. The results for the latter cultivar were quite similar though the contribution of mean number of kernels per spike was two times greater than that of kernel weight, number of spikes per m^2 being unimportant. Nevertheless, in other results on the same cultivars, Kozak et al. (2007b) obtained fairly different results, in which it was number of spikes per m^2 that was the main component in determining grain yield of cultivar Bogo, contributing to about 50% of yield determination. For cultivar Fidelio this difference was not so visible, but still the first component accounted for 26% of grain yield determination, while in Samborski et al. (2005) study this contribution was 0%.

From the above discussion it is clear that there is little sense in studying relationships among yield and its components for a particular species; this should rather be done at a genotype level. One example for this is the above-mentioned research by Kozak et al. (2007b), who studied 15 winter triticale genotypes in terms of yield component analysis. From that study it clearly follows that there is a strong influence of genotype on YCA, which means that the results of YCA may greatly vary among genotypes.

The above discussion suggests that usually, if not always, it makes no sense to decide that a particular yield component is the best for a particular species as well as for a particular genotype across all environments. This is because the pattern of the influence of yield components on cereal grain yield, which is reflected in the results of YCA, clearly depends on both genotype and environment.

In summary, owing to the limitations mentioned above, the benefits of applying YCA for a particular genotype are rather narrow. The dependence of the YCA results on environment and genotype may be so strong that any conclusions drawn based on a single experiment, even if repeated across two or three years, may have no real meaning for broadly understood knowledge of the species and even genotype. This is not to say that no studies should be carried out and no such results provide interesting information; this is to say, however, that one should be always careful with YCA and its interpretation for a single genotype. And most of all, single-genotype YCA is not interesting itself, and should be linked to any process one studies, for example, optimizing grain yield.

Across-genotype yield component analysis. By across-genotype YCA we understand the YCA applied for a pool of genotypes based on mean values of the components and yield; hence for one genotype there is only one yield value, which represents the average yield for this genotype (either for a particular environment or for a pool of environments considered in the study). According to Kozak and Mądry (2006), however, it is important that the values of one of the components be calculated based on values of other components and yield in order to ensure that the model (1) be not violated. The best practice is to calculate the values of that component which is the most difficult to measure or with measurement of which the largest error is associated; for cereals it is common to calculate the values of mean number of kernels per spike. It is important to note, however, that such calculated values for the across-genotype yield component analysis will be slightly biased; in the case of two components in the model, for example, this bias for a particular genotype will be equal to the covariance between the components for this genotype divided by the mean of the other component (which follows from the equations given by Hűhn 1987). It is possible that this covariance might be incorporated into the model, but this needs further research. Hence an across-genotype yield-component model needs further development.

Interpretation based on across-genotype YCA deals with relationships for genotypes. For example, if number of spikes per unit area has a positive influence on grain yield, it means that genotypes with large number of spikes per unit area are high-yielding. This type of analysis fully disregards what is going on for particular genotypes, so has nothing to do with single-genotype YCA. In fact, for some genotypes number of spikes per unit area may be the most while for the others the least important yield component in determining grain yield, which says nothing about the across-genotype YCA, which can for example prove this component to influence grain yield moderately.

While the single-genotype YCA has agronomic importance, the across-genotype YCA is rather concerned with plant breeding. Information the across-genotype YCA provides may be utilized to detect to which components high yield can be attributed. This in turn may show in which direction breeding of high-yielding genotypes should be led, and which genotypes from the pool of genotypes used in a study might be used for this purpose. As was the case with single-genotype YCA, the across-genotype YCA may be environment-dependent. This makes the YCA results vary among various genotypes, which in turn affects conclusions as to which components indeed affect grain yield most importantly. If, despite these differences, there is still one component the influence of which is major in all or at least most of the environments, this component can be considered the most important in determining grain yield. Nonetheless, if the differences among the environments are large and no such component can be pointed out, then the YCA results are unstable across the environments. In such a case, quite likely the across-genotype YCA will provide no particular conclusions as to which components should be paid most attention to in the breeding program. A more detailed analysis should be then performed in order to probe into the YCA pattern to find any information about the influence of components on grain yield that might be helpful.

From the above it follows that across-genotype YCA cannot offer any information on the optimum agronomic practices or environments in terms of optimizing yield or its components. These optimum practices and environments may wildly differ among genotypes. Such considerations are in fact remote from its focus, and as such should never be taken into account during interpretation. The across-genotype YCA should be reserved for finding the patterns of the influence of components on yield among the pool of genotypes, keeping in mind that this information may be far from (and even opposite to) the patterns for individual genotypes.

REFERENCES

Abbadi J., Gerendas J., Sattelmacher B. (2008): Effect of nitrogen supply on growth, yield and yield components of safflower and sunflower. Plant Soil, *306*: 167–180.

Bos I., Sparnaaij L.D. (1993): Component analysis of complex characters in plant breeding. II. The pursuit of heterosis. Euphytica, *70*: 237–245.

Brown D., Alexander N. (1991): The analysis of the variance and covariance of products. Biometrics, *47*: 429–444.

Eaton G.W., Bowen P.A., Jolliffe P.A. (1986): Two-dimensional partitioning of yield variation. Horticultural Science, *21*: 1052–1053.

Eaton G.W., Kyte T.R. (1978): Yield component analysis in strawberry. Journal of the American Society for Horticultural Science, *103*: 578–583.

Eaton G.W., MacPherson E.A. (1978): Morphological components of yield incranberry. Hortïcultural Research, *17*: 73–82.

Engledow F.W., Wadham S.M. (1923): Investigation on yield in the cereals. The Journal of Agricultural Science, *13*: 390–439.

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

Fraser J., Eaton G.W. (1983): Applications of yield component analysis to crop research. Field Crop Abstracts, *36*: 787–796.

Gołaszewski J., Idźkowska M., Milewska J. (1998): The TDP method of seed yield component analysis in grain legume breeding. Journal of Applied Genetics, *39*: 299–308.

Hűhn M. (1987): Stability analysis of winter-rape (*Brassica napus* L.) by using plant density and mean yield per plant. Journal of Agronomy and Crop Science, *159*: 73–81.

Jolliffe P.A., Courtney W.H. (1984): Plant growth analysis: additive and multiplicative components of growth. Annals of Botany, *54*: 243–254.

Kozak M. (2004): New concept of yield component analysis. Biometrical Letters Year, *41*: 59–69.

Kozak M. (2006): Two-dimensional partitioning of yield variation: A critical note. Plant Breeding & Seed Science, *53*: 37–42.

Kozak M., Mądry W. (2005): Primary characters in yield component analysis. Part I. Description of the method. Fragmenta Agronomica, *26*: 71–83. (In Polish)

Kozak M., Mądry W. (2006): Note on yield component analysis. Cereal Research Communications, *34*: 933–940.

Kozak M., Samborski S., Kang M.S., Rozbicki J. (2007a): Applying Statistics for nonsequential yield component analysis. Plant, Soil and Environment, *53*: 456–463.

Kozak M., Samborski S., Rozbicki J., Mądry W. (2007b): Winter triticale grain yield, a comparative study of 15 genotypes. Acta Agriculture Scandinavia Section B – Plant and Soil Science, *57*: 263–270.

Piepho H.P. (1995): A simple procedure for yield component analysis. Euphytica, *84*: 43–48.

Samborski S., Kozak M., Azevedo R.A. (2008): Does nitrogen uptake affect nitrogen uptake efficiency, or vice versa? Acta Physiologiae Plantarum, *30*: 419–420.

Samborski S., Kozak M., Mądry W., Rozbicki J. (2005): Primary characters in yield component analysis. Part II. Application for winter triticale grain yield. Fragmenta Agronomica, *4*: 84–97. (In Polish)

Sparnaaij L.D., Bos I. (1993): 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, *70*: 225–235.

Received on December 12, 2008

Corresponding author:

Dr. Marcin Kozak, Warsaw University of Life Sciences, Department of Experimental Design and Bioinformatics,
Nowoursynowska 159, 02-776 Warsaw, Poland
phone: + 482 259 327 30, fax: + 482 259 327 22, e-mail:nyggus@gmail.com