Colloquium Biometricum 41
2011, 39–47

VISUALIZING DIVERSITY OF YIELD DETERMINATION BY ITS COMPONENTS FOR WINTER WHEAT CULTIVAR WITH TERNARY PLOT

Dariusz Gozdowski[1], Daniel Sas[2], Jan Rozbicki[3], Wiesław Mądry[1], Jan Golba[3], Mariusz Piechociński[3], Lidia Kurzyńska[3], Marcin Studnicki[1], Adriana Derejko[1]

1 – Department of Experimental Design and Bioinformatics
Warsaw University of Life Sciences
Nowoursynowska 159, 02–776 Warsaw, Poland
2 – Department of Planning and Science Documentation
Research Institute ofHorticulture
Konstytucji 3 Maja 1/3, 96–100 Skierniewice, Poland
3 – Department of Agronomy
Warsaw University of Life Sciences
Nowoursynowska 159, 02–776 Warsaw, Poland
e–mails: dariusz_gozdowski@sggw.pl; daniel.sas@insad.pl; jan_rozbicki@sggw.pl; w.madry@omega.sggw.waw.pl; jan.golba@gmail.com; mariusz.piechocinski@gmail.com; kurzyska@wp.pl; studnicki.marcin@gmail.com; adrianaderejko@gmail.com

Research supported by scientific project MNiSW N N310 091136

Dedicated to the memory of Professor Wiktor Oktaba

Summary

This work presents evaluation of usefulness of ternary plots (ternary diagram) for visualization of 4–dimensional data i.e. yield and its three components. Data used for the analyses were obtained in 8 field experiments located across Poland with 28 winter wheat cultivars in 2009. Path analysis for examination of determination of the yield by its components was conducted for each cultivar and then cluster analysis was conducted based on the path coefficients. Groups of cultivars with different pattern of yield determination were presented on ternary plot created using R statistical software.

Key words and phrases: ternary plot, grain yield, winter wheat, cluster analysis, path analysis

Classification AMS 2010: 62H20, 62H30

1. Introduction

Evaluation of determination of yield of crops by yield–related traits, including yield components, is important in indication, which traits have influence on yield variability. It can be important in plant breeding process of high–yielding genotypes and optimization of crop management. One of the methods used in evaluation of determination of variability of yield by yield components is classic path analysis (Wright, 1921, 1923, 1934) based on multiple regression analysis. This method is very common in agronomic research especially in evaluation of determination of the cereals grain yield, including wheat (Acreche and Slafer 2006, Ahmed et al. 2003, Garcia del Moral et al.2005, Moragues et al.2006). When we examine differences in determination of three yield components between large number genotypes it is difficult to detect distinct patterns of yield determination. In this paper we propose the method of visualization such a data based on ternary plot. The main aim of the study was to present a method of visualization for 4–dimensional data (variables: quantity of yield and three relative values of path coefficients for yield components) for genotypes of winter wheat.

2. Material and methods

Data used for analyses were obtained from eight locations of post–registration multi–environment trials (conducted by COBORU – Research Centre For Cultivar Testing) with winter wheat in 2009. The eight trials were located in main regions of wheat production in Poland. Each field experiment was conducted in split–block design with 2 replications where the factors were: cultivar and crop management level (2 levels). 28 cultivars were examined in each location. Total number of experimental units for each cultivar was equal 32 (8 locations x 2 crop management levels x 2 replications).

The data for each cultivar were analyzed separately using path analysis i.e. multiple regression based on standardized data. Following linear model was used for analyses:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon_i$$

where

Y – standardized value of grain yield (in g per m^2);

X_1, X_2 and X_3 – standardized values of yield components (i.e.number of spikes per m^2, mean number of grains per spike and mean weight of individual grain);

β_1, β_2, β_3 – path coefficients i.e. partial regression coefficients for standardized data.

Path coefficients for each cultivar were normalized using following transformation:

$$\beta_{in} = \frac{\beta_i \times 100}{\beta_1 + \beta_2 + \beta_3} ,$$

where β_i – value of path coefficient for i–th yield component and β_{in} –normalized value of path coefficient for i–th yield component.

The normalized values of path coefficients for distinct cultivars were presented in ternary plot created using R (R Development Core Team 2009). Additionally at the ternary plot quantity of grain yield and the group of cultivars distinguished in cluster analysis based on normalized values of path coefficients (3 variables) was indicated. Squared Euclidean distance was used as a measure of dissimilarity of genotypes and Ward's method was used for agglomeration of the genotypes. Cluster analysis was performed using Statistica 7.1 software (StatSoft, 2005)

3. Results and discussion

Results of path analysis (Table 1) proved relatively strong effect of number of spikes per m^2 on yield, quite strong effect of number of grains and relatively weak effect of weight of individual grain. It means that yield variability is determined mainly by the first two yield components i.e. number of spikes and number of grains per spike. For better evaluation of effects of each yield component normalized path coefficients were calculated (Table 1). It is very difficult to distinguish different patterns of yield determination for examined cultivars because it demands multivariate evaluation of quite big number of genotypes. We propose graphical method based on ternary plot which can be useful for visualization of such kind of the data. Before visualization the examined cultivar were divided into five groups of similar pattern of yield determination using cluster analysis. Variables for cluster analysis (Ward's method) were normalized path coefficients (Table 1) for yield components i.e. number of spikes per m^2, number of grains per spike and weight of individual grain. Results of cluster analysis are presented in Fig. 1.

5 groups of cultivars which have different pattern of yield determination are presented in Fig. 1. Because some of the cultivars have very similar pattern of yield determination (eg. cv. Figura and cv. Kohelia) not all genotypes were presented in ternary plot (Fig. 2). 13 cultivars were chosen for the example of ternary plot.

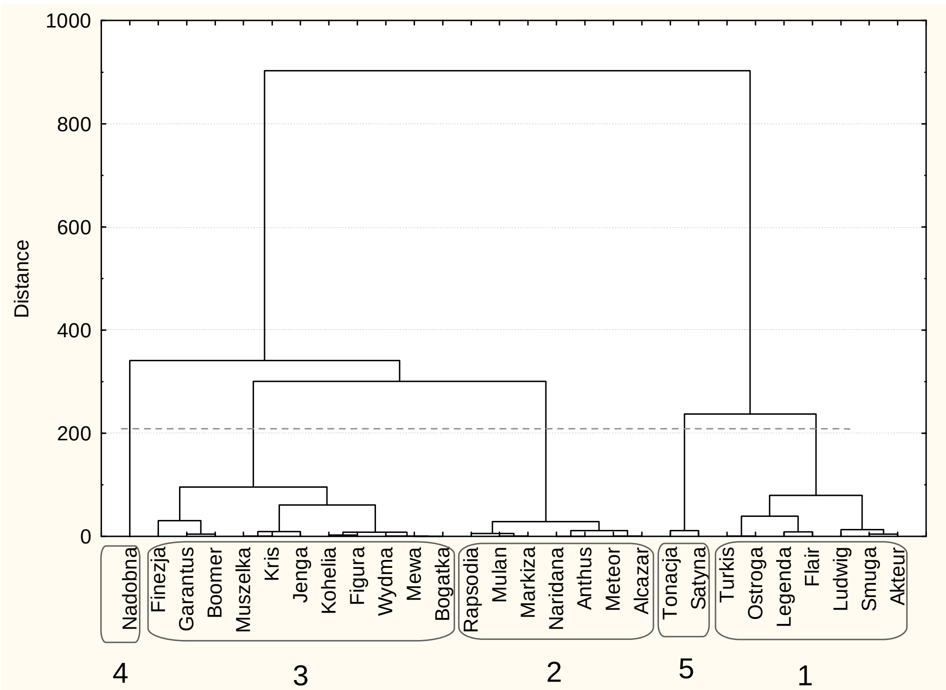


Fig. 1. Dendrogram based on results of cluster analysis for normalized values of path coefficients for yield components

Table 1. Path coefficients and normalized path coefficients for 28 cultivars of winter wheat

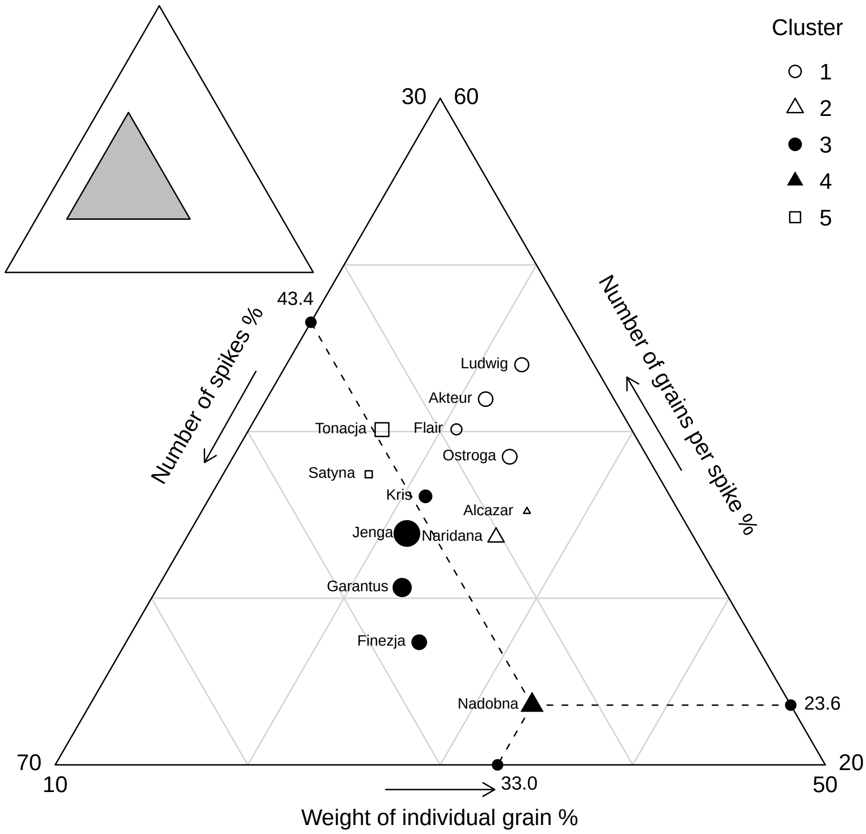
	Path coefficients			Normalized path coefficients			
Cultivar	(β_1) Number of spikes per m^2	(β_2) number of grains per spike	(β_3) weight of indivi-dual grain	(β_{1n}) Number of spikes per m^2	(β_{2n}) number of grains per spike	(β_{3n}) weight of indivi-dual grain	(Y) grain yield (g/m^2)
Akteur	0.935	1.069	0.545	36.7	42.0	21.4	798.2
Alcazar	0.944	0.877	0.670	37.9	35.2	26.9	757.5
Anthus	0.896	0.764	0.587	39.9	34.0	26.1	818.9
Bogatka	1.153	0.885	0.635	43.1	33.1	23.7	841.6
Boomer	1.015	0.720	0.517	45.1	32.0	23.0	806.8
Figura	0.921	0.683	0.543	42.9	31.8	25.3	822.7
Finezja	0.975	0.563	0.519	47.4	27.4	25.2	798.5
Flair	1.000	1.026	0.531	39.1	40.1	20.8	782.9
Garantus	1.087	0.714	0.529	46.7	30.6	22.7	814.3
Jenga	1.051	0.795	0.500	44.8	33.9	21.3	847.7
Kohelia	1.187	0.838	0.654	44.3	31.3	24.4	824.5
Kris	1.137	0.961	0.564	42.7	36.1	21.2	789.8
Legenda	0.901	0.828	0.461	41.1	37.8	21.1	819.2
Ludwig	0.666	0.868	0.438	33.8	44.0	22.2	796.3
Markiza	0.933	0.863	0.573	39.4	36.4	24.2	769.8
Meteor	0.850	0.779	0.582	38.4	35.2	26.3	818.8
Mewa	0.972	0.735	0.546	43.1	32.6	24.2	786.5
Mulan	1.004	0.962	0.636	38.6	37.0	24.4	809.6
Muszelka	0.914	0.754	0.458	43.0	35.5	21.6	820.2
Nadobna	0.860	0.467	0.653	43.4	23.6	33.0	812.0
Naridana	0.966	0.808	0.626	40.3	33.7	26.1	788.5
Ostroga	0.815	0.844	0.534	37.1	38.5	24.4	799.3
Rapsodia	0.939	0.823	0.548	40.7	35.6	23.7	820.3
Satyna	0.872	0.725	0.340	45.0	37.4	17.6	770.0
Smuga	0.657	0.773	0.428	35.4	41.6	23.0	746.7
Tonacja	1.042	0.973	0.410	43.0	40.1	16.9	803.6
Turkis	0.901	0.966	0.582	36.8	39.4	23.8	806.9
Wydma	0.938	0.739	0.533	42.4	33.5	24.1	768.1

R package *ade4* and *soiltexture* was used for creating the ternary plot and main parts of code is given below:

```
data1 <- read.table("dataset", header = TRUE, sep = "\t")
colnames(data1)[1:3] <- c("Number of spikes %", "Weight of
individual grain %", "Number of grains per spike %")
library(soiltexture)
library(ade4)
#adjusting values of variable “Yield” (point size)
data1["Yield"] <- TT.str(data1[,"Yield"], 0.5, 2.5)
data1$Cultivar <- factor(data1$Cultivar)
data1$Cluster <- factor(data1$Cluster)
levels(data1$Cl) <- c(1, 2, 19, 17, 22)
#loading modified function “triangle.plot” - package “ade4”
source("C:path_to_directory_where_function_is_saved/triangle.p
lot.R")
#plot drawing
triangle <- triangle.plot(data1[, 1:3], scale = T,
show.position = T, cpoint = 0)
#adding points
points(triangle, pch = as.numeric(as.vector(data1$Cl)), cex =
data1[, 4])
#adding labels (cultivars)
text(triangle, label = data1$Cultivar, cex = 0.7, pos = 2)
#adding legend
par(xpd = NA);legend(0.6, 1, levels(data1$Cluster), pch = c(1,
2, 19, 17, 22), pt.cex = 1.2, cex = 1, bty = "n", title =
"Cluster")
```

The structure of the data set should be following (header and three rows of the data):
Number_of_spikes, Weight_of_individual_grain, Number_of grains_per_spike, Yield, Cultivar, Cluster
36.7, 21.4, 42.0, 798.2, Akteur, 1
37.9, 26.9, 35.2, 757.5, Alcazar, 2
47.4, 25.2, 27.4, 798.5, Finezja, 3

Ternary plot seems to be very clear method of visualization of 3–dimensional data and gives possibility to present fourth variable (i.e. grain yield), which is represented by point size. Additionally we can distinguish groups of cultivars using various types of the points. In our case it is important distinguishing groups of cultivars with various pattern of yield determination. The first group of cultivars (eg. cv. Ludwig, Akteur, Ostroga, and Flair) in the Fig. 2 have relatively low grain yield and it is determined mainly by number of grains per spike. The opposite pattern of yield determination was proved for cultivars in group 3 (eg. cv. Garantus and Finezja), where yield was determined mainly by number of spikes per unit area.



Point size indicates quantity of grain yield for each cultivar

Fig. 2. Ternary plot presenting groups of cultivars with different patterns of yield determination by its components

Such a method of visualization can be useful only in particular kinds of the data, when sum of three variables is equal constant value for all objects. The ternary plots (other name: ternary diagrams or triangular diagram) were quite unusual until the mid–ninetenth century (Howarth, 1996). The particular type of graph which is presented in the paper consists of an equilateral triangle in which a given plotted point represents the relative proportions (a, b, c) where $a + b + c = 100\%$. One very common application of the ternary plot in agriculture is presentation of soil texture, where sum of three fractions soil particles is equal 100% (Marshall et al. 1996, Flemming 2000). Other adoptions of ternary plot in agricultural research are not very common, but there are some papers where ternary plot was used for visualization of diversity of crop genotypes (Wiesenberg and Schwark 2006, Kozak 2010) and chemical composition of crops (Herrera et al. 2006).

4. Conclusions

Ternary plot seems to be very clear way of visualization of 3–dimensional data, and it is possible to add fourth variable using various size for points as well distinguish groups of objects using different type of points. Number of application is limited mainly because the sums of the three variables have to be constant value. Two of the very useful packages for creation of ternary plot are *ade4* and *soiltexture* included into R.

References

Acreche M., Slafer G. (2006). Grain weight response to increases in number of grains in wheat in a Mediterranean area. *Field Crops Research* 98, 52–59.

Ahmed H., Khan B., Khan S., Kissana N., Laghari S. (2003). Path coefficient analysis in bread wheat. *Asian Journal of Plant Science* 2 (6), 491–494.

Flemming B.W. (2000). A revised textural classification of gravel–free muddy sediments on the basis of ternary diagrams. *Continental Shelf Research* 20, 1125–1137.

Garcia del Moral L., Rharrabti Y., Elhani S., Martos V., Royo C. (2005). Yield formation in Mediterranean durum wheats under two contrasting water regimes based on path–coefficient analysis. *Euphytica* 146, 203–212.

Herrera C.M, Pérez R., Alonso C. (2006). Extreme intraplant variation in nectar sugar composition in an insect–pollinated perennial herb. *American Journal of Botany* 93, 575–581.

Howarth R.J. (1996). Sources for a history of the ternary diagram. *The British Journal for the History of Science* 29, 337–356.

Kozak M. (2010). Visualizing adaptation of genotypes with a ternary plot. *Chilean Journal of Agricultural Research* 70, 596–603.

Marshall T.J., Holmes J.W., Rose C.W. (1996). *Soil physics*. Cambridge University Press, p. 453.

Moragues M., Garcia del Moral L., Moralejo M., Royo C. (2006). Yield formation strategies of durum wheat landraces with distinct pattern of dispersal within the Mediterranean basin I: Yield components. *Field Crops Research* 95, 194–205.

R Development Core Team (2009). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R–project.org.

StatSoft, Inc. (2005). *STATISTICA* (*data analysis software system*), version 7.1. www.statsoft.com.

Wiesenberg G.L., Schwark L. (2006). Carboxylic acid distribution patterns of temperate C3 and C4 crops. *Organic geochemistry* 37, 1973–1982.

Wright S. (1921). Correlation and causation. *Journal of Agricultural Research* 20, 557–585.

Wright S. (1923). The theory of path coefficients – a reply to Niles's criticism. *Genetics* 8, 239–255.

Wright S. (1934). The method of path coefficients. *Annals of Mathematical Statistics* 5, 161–215.