Use of Trellis Graphics in the Analysis of Results from Field Experiments in Agriculture

Katarina Čobanović, Emilija Nikolić-Đorić, and Beba Mutavdžić[1]

Abstract

Trellis graphics (Becker, Cleveland, and Shyu, 1996) is a very effective method for visualizing multidimensional data sets. The basic idea behind trellis graphics is to display any of a large variety of 1-D, 2-D or 3-D statistical plot types in trellis layout of panels, where each panel displays a subset of the data for different values of one or more additional discrete or continuous conditioning variables.

The data that we use for the illustration of different applications of trellis graphics are the results of a field experiment conducted at the Institute for Field and Vegetable Crops in Novi Sad in the period 1994-1998 (Čobanović et al., 2001) with three fertilizers (nitrogen, phosphorus and potassium) in three repetitions with nine variants of wheat. In the experiment, four quantities of each fertilizer were applied (0, 50, 100, 150 kg/ha) at plots of the same size in 20 from 64 possible combinations, whereby the yield of wheat (t/ha) was the measured outcome.

1   Introduction

Modern computer technology has changed the way statistical analysis and summary is done. In particular, graphical methods of analysis and summary now play more important role, and deserve increased emphasis in scientific reports (Cleveland,1993).

Statistical graphics generally have two major functions: the analysis and the presentation which is traditionally their primary function. With the development of computer technology and with the intensive use of computer software, the analysis function of statistical graphics has assumed   increasing importance (Schmid,1983). In recent times, the statistical graphics are a very attractive way of visual communication and analysis.

---

[1] Faculty of Agriculture, University of Novi Sad,   21000 Novi Sad, Trg Dositeja Obradovića 8, Serbia; katcob@polj.ns.ac.yu

Graphical methods have a central role in the Exploratory Data Analysis (EDA), the approach of data analysis introduced by John Tukey (1977). As the human brain is very powerful in processing visual information, sometimes a single glance at a graph is enough to identify a very complex structure of data and relations between the variables. Cleveland and his associates at Bell Laboratories did a lot of research in the theory of graphical perception. They focused on the psychophysical aspects of human graphical processing.

The appropriately drawn graphs may help in understanding the data and in presenting the results to others. On the other hand, the graphs not properly drawn, can lead to the wrong conclusions (Tufte, 2001)

Statistical graphics are used both in teaching statistics as well as in the research (Nikolić-Đorić et al., 2006).

The data analysis based on graphics is the first step in a great many statistical investigations.

The trellis graphics were introduced by Becker, Cleveland, and Shyu, 1996. The basic idea behind the trellis graphics is to display any of a large variety of 1-D, 2-D or 3-D statistical plot types in trellis layout of panels, where each panel displays a subset of the data for different values of one or more additional discrete or continuous conditioning variables.

The name trellis comes from the arrangement of the plots which looks like a trellis (grid, lattice). The term "trellis" comes from gardening where it is an open structure used as a support for vines.

One important application of a trellis display is uncovering the structure of multivariate data and relations of the variables in the multivariable data sets (Becker, Cleveland, and Shyu, 1996). The trellis display enables making important discoveries not found in the initial analysis. By comparing each conditioned panel on the same scale it is possible not just to explore if the relationship between the variables exists, but if it holds for all the levels of the conditioning variable as well.

Although the trellis was developed initially in the context of large data sets, it is also useful for modelling the data from the designed experiments, even small experiments, and it is a very powerful tool for revealing the structure of interactions in the studies of how a response depends on explanatory variables (Cleveland and Fuentes, 1997).

One of the first applications of the trellis diagram was to present the yields data of ten varieties of barley in an experiment arranged in randomized blocks, carried out in the State of Minnesota in the years 1930 and 1931, at six locations (Fisher, 1966). The trellis display led to the conclusion that the data are in error.

2   Software for Trellis

A trellis display was implemented in S/S-Plus system (Backer and Cleveland, 1996).

The special package "lattice" for producing trellis graphs was developed in R-language.

Also the trellis was implemented in the statistical packages GENSTAT 8, VSN, International, 2005 and STATISTICA, Statsoft Inc, where it was named a categorized graph, the term first used in 1990. S-Plus Clinical Pack developed by Insightful Corporation, enables the easy integration of powerful S-Plus graphics within SAS environment.

2.1 Software for Trellis

A trellis display consists of panels laid out into a three-way rectangular array of columns, rows and pages. For the small data sets one page is usually enough. In the case of larger data sets multi-page layouts are necessary for presenting the whole data set. The panels of a trellis display, by default are ordered left-to-right and bottom-to-top but may be changed by the user in some other way.The table ordering, for example, is left-to-right and top-to-bottom. The strip labels written at the tops of panels indicate the conditioning (slicing) variable and ranges, values or levels of it, depending on whether it is a numerical (continuous, discrete) or categorical variable.

A trellis display may be created by a single command line. It is based on repeating the same graphical specification for each element in a Cartesian product of levels of one or more factors. The programs in S-Plus and R trellising library have the structure:

$$Y \sim X / a * b$$

where Y is a continuous variable, X is a continuous variable or factor and a, b levels of factors, variables or functions of the fitted model. A scatter plot also consists of panels that are defined by Cartesian product of variables. Each of the panels is a plot of different set of variables but it is based on the entire set of observations and not on a subset as in the case of trellis display. In S-Plus and R it is possible to present a scatter-plot matrix conditioned on values on a relevant variable, i.e. in trellis form.

The very flexible object-oriented S-Plus and R languages make possible the control of display in order to present maximum information of data. It is also possible to define the aspect ratio, multi-panel layout, plotting symbols, lines, colours, character sizes.

3   Experimental data

The data that we use for the illustration of different applications of trellis graphics are the results of a field experiment conducted at the Institute for Field and Vegetable Crops in Novi Sad in the period 1994-1998 (Čobanović et al., 2001) with three fertilizers (nitrogen, phosphorus and potassium) in three repetitions with nine variants of wheat. In the experiment, four quantities of each fertilizer were applied (0, 50, 100, 150 kg/ha) at the plots of the same size in 20 from 64 possible combinations (Table 1), whereby the yield of wheat (t/ha) was the measured outcome.

Table 1: Combinations of fertilizer applied in the experiment.

Variants	N	P	K
1	0	0	0
2	100	0	0
3	0	100	0
4	0	0	100
5	100	100	0
6	100	0	100
7	0	100	100
8	50	50	50
9	50	100	50
10	50	100	100
11	100	50	50
12	100	100	50
13	100	100	100
14	100	150	100
15	100	150	150
16	150	50	50
17	150	100	50
18	150	100	100
19	150	150	100
20	150	150	150

4   Discussion

In the analysis of the results from the field experiment in agriculture, several univariate, bivariate and high dimensional trellis displays were applied. Also for the same type of trellis various partitioning of data were made in order to focus on the different features of data.

In order to explore the characteristics of data distribution one-dimensional trellis were applied: trellis dot-plot (Figure 1), box-plots (Figure 2), histograms with probability density plot (Figure 3), and quantile plots (Figure 4). With these

diagrams the subpopulation structure, the distribution shape and the presence of outliers may be quickly revealed. Figure (1) illustrates that the highest wheat yield was in 1994/95 and the lowest in 1996/7, regardless of the variety and combination of fertilizers.

Trellis notched box-plots (Figure 2) shows that Pobeda had the highest median wheat yield while the yield of variety Lasta had the highest variability. Also on the basis of 95% median confidence interval that consists of notches that are drawn about the median and are extended to $\pm 1,58 \cdot \dfrac{\mathrm{IQR}}{\sqrt{n}}$, where IQR is interquartile range, it is possible to make multiple comparisons of the median yields. The notches that do not overlap represent a significant difference between the medians.
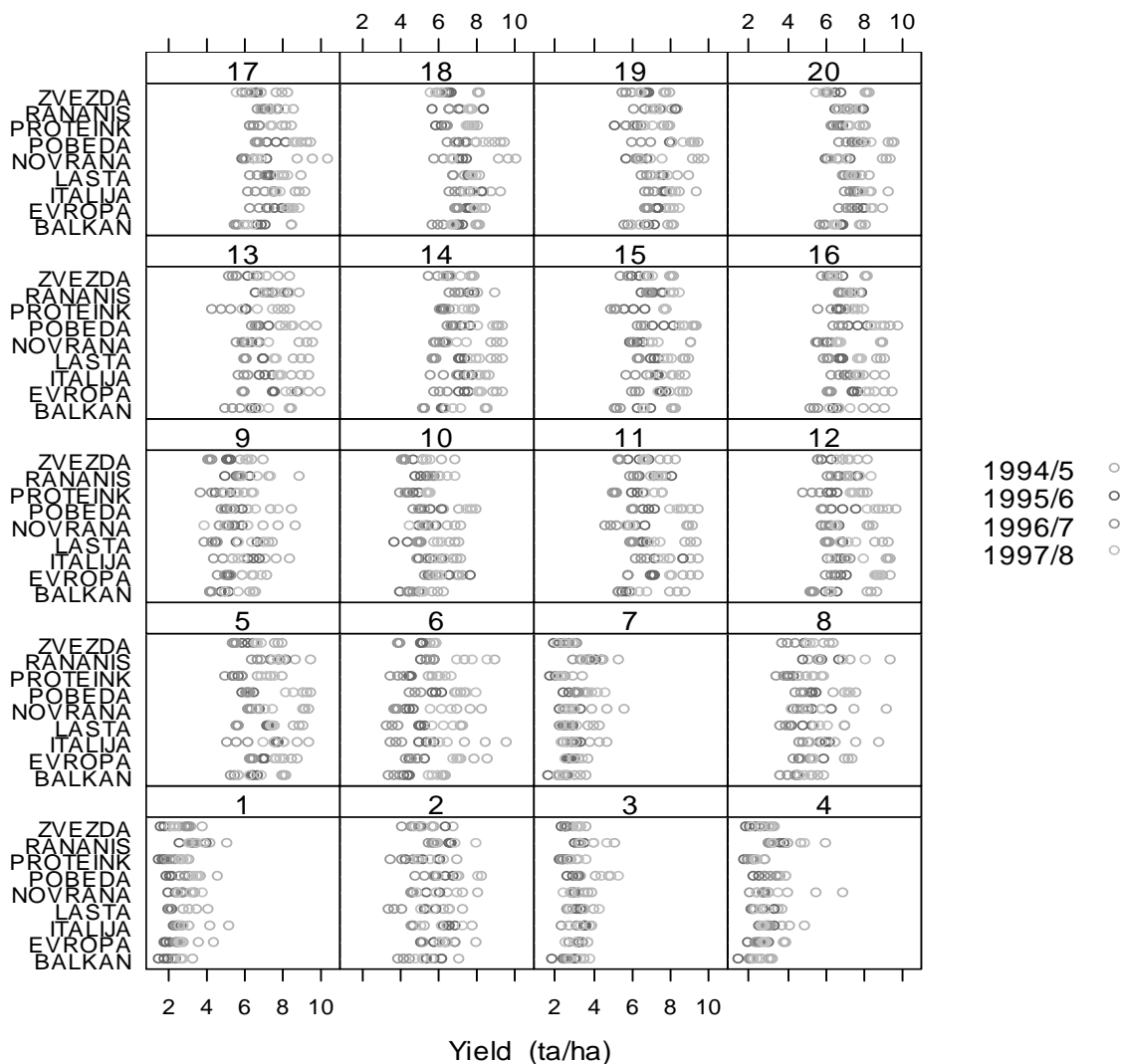
Figure 1: The yield against variety and year given the variant of fertilizer.

The trellis histograms with probability density plots (Figure 3) show that the average wheat yield strongly depends on the level of nitrogen. The greatest deviation from normal distribution is for zero level of nitrogen (Figure 4).

In order to explore the presence of interaction, series of box-plots conditioned on values of one or two factors were applied. If two predictors are not of equal importance the exposure should be taken as panel and the confounder as the condition variable.
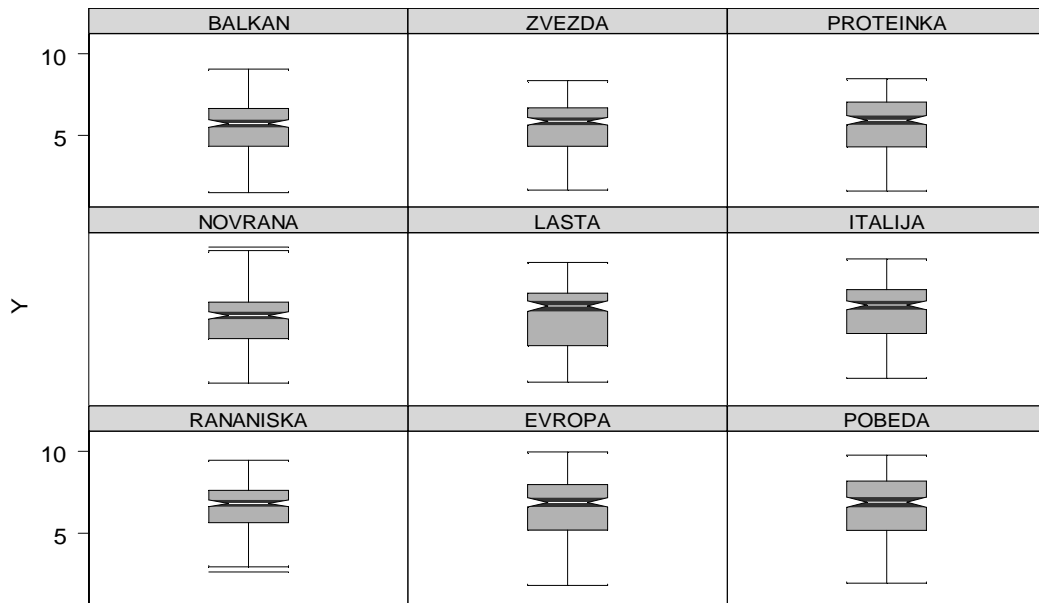
Figure 2: Trellis notched box-plots.
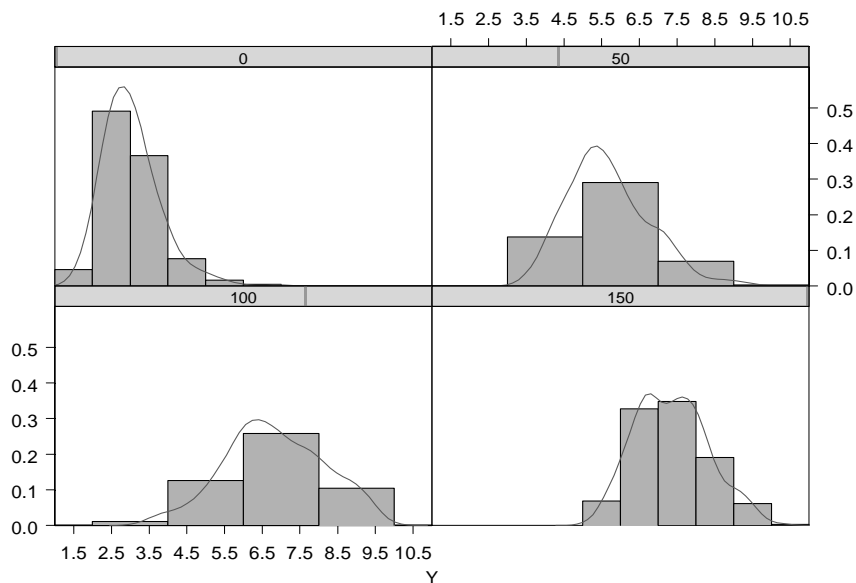
Figure 3: Trellis histograms with probability density plots.
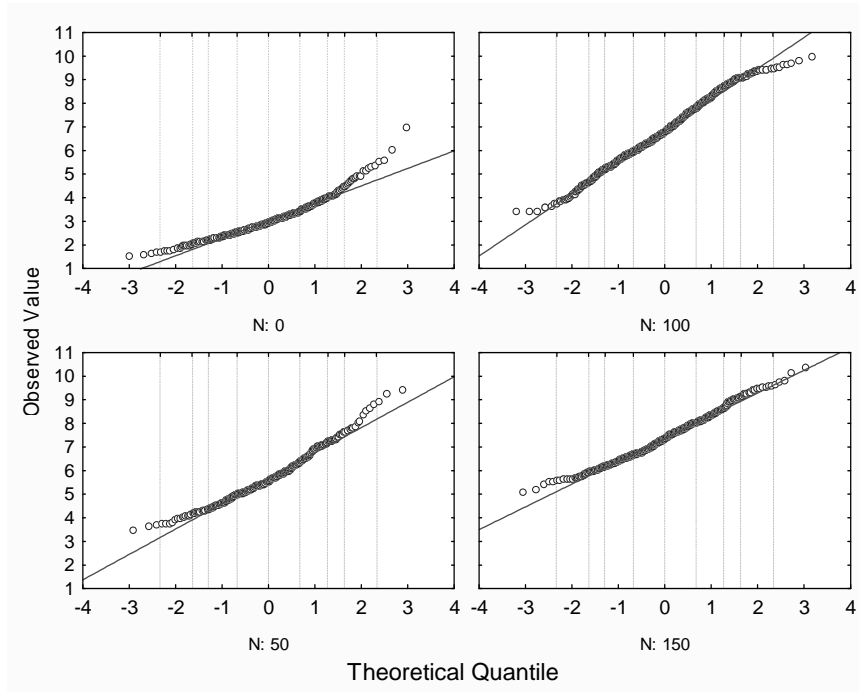
Figure 4: Trellis Q-Q plots.
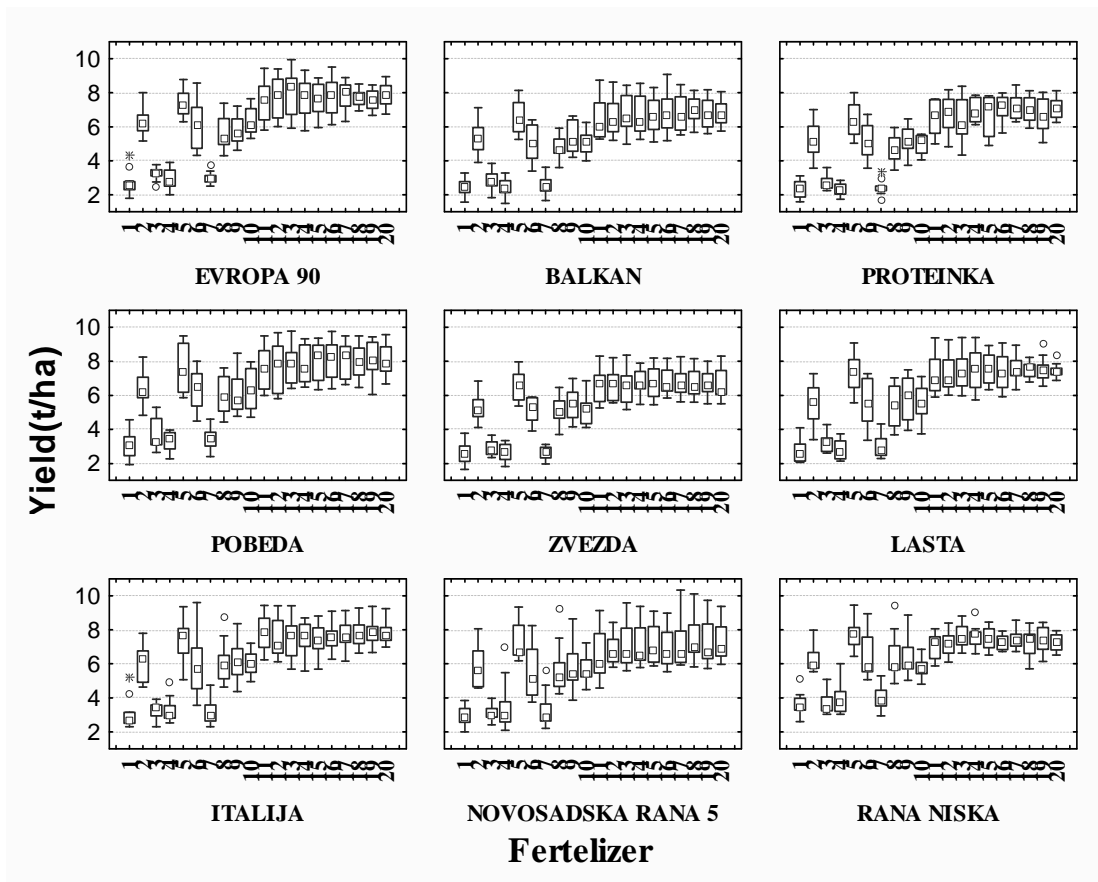
Figure 5: A box-plot of the wheat yield against variant of fertilizer given variety.
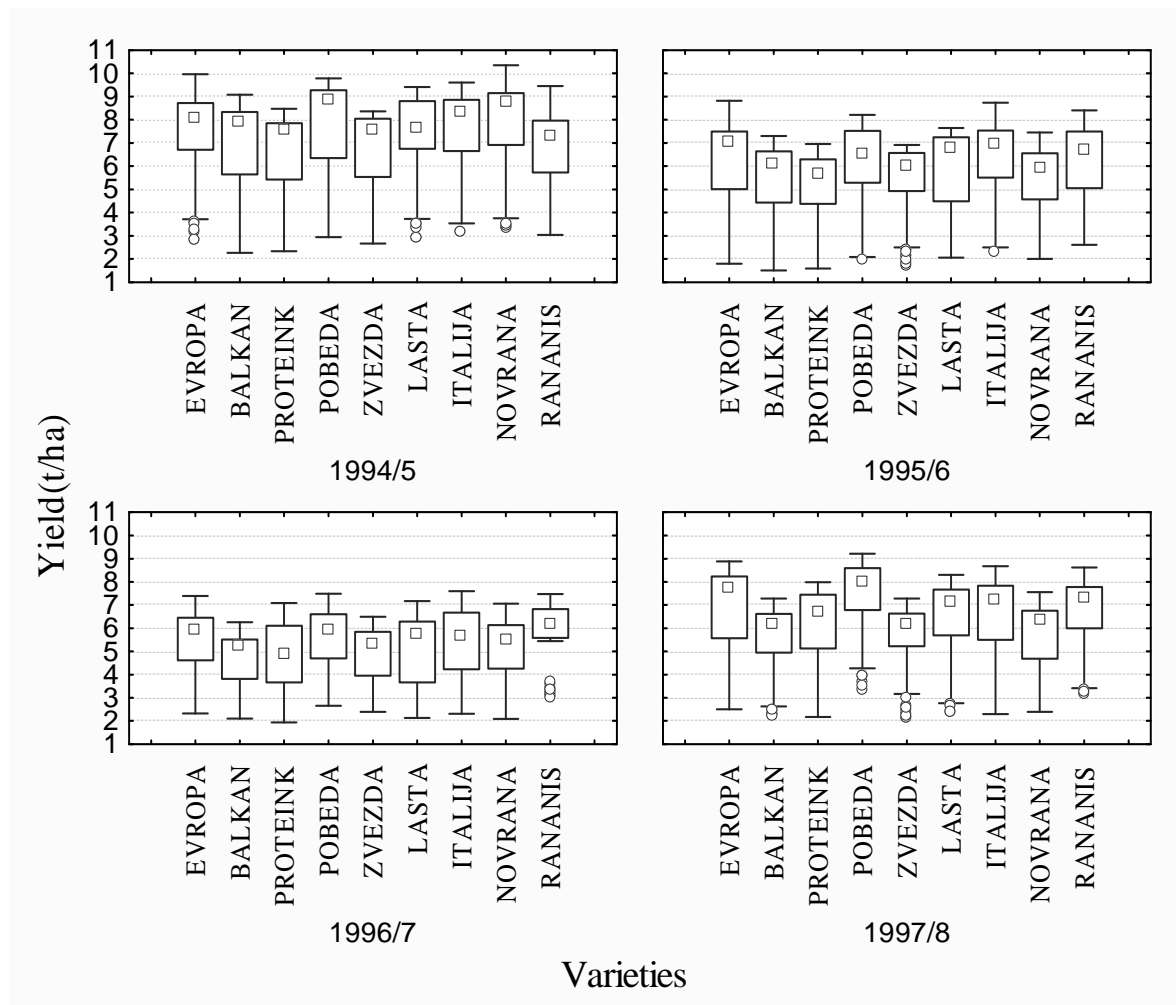
Figure 6: A box-plot of the wheat yield against varieties given year.

As we want to provide the equal representation of the effects of the two predictor variables (variety and fertilizer) the panel variables were switched.

Figures 5 and 6 are the examples of the use of Trellis graph in exploring the interactions fertilizer-variety and variety-year. The conditional dependence on two variables is presented in Figure 7, showing the wheat yield against variant of fertilizer given variety and year.

A two-dimensional trellis scatter plot is applied in order to fit the function that approximates the influence of fertilizers on the wheat yield. Figure 8 displays the influence of nitrogen that may be approximated with quadratic regression. It could be easily noticed that the parameters of regression models are not the same for all varieties.

The simultaneous influence of nitrogen and phosphorus on the wheat yield is displayed by means of 3D trellis plot (Figure 9). The 3D scatter plots for

individual wheat varieties suggest the quadratic regressions that contain linear and quadratic terms for nitrogen and phosphorus, and their interaction.

The better insight into nitrogen-phosphorus interaction is given in the trellis contours plots made on the basis of estimated quadratic response surfaces (Figure 10).
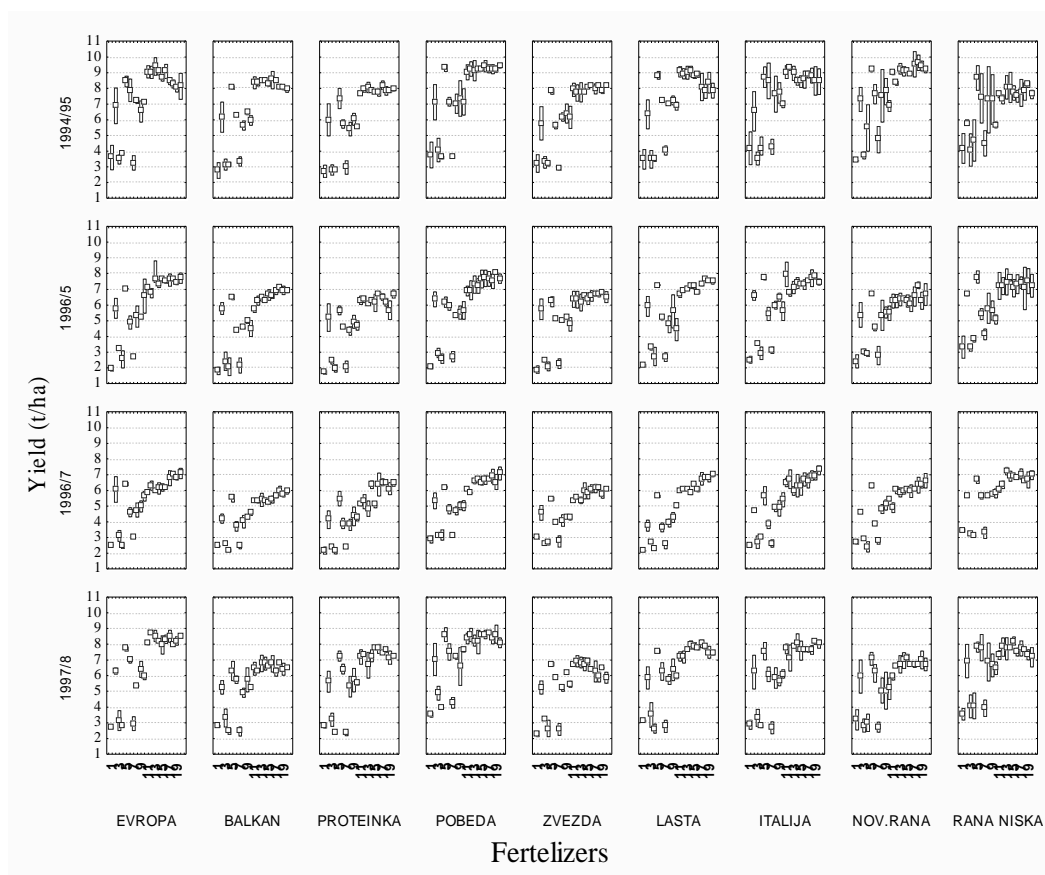

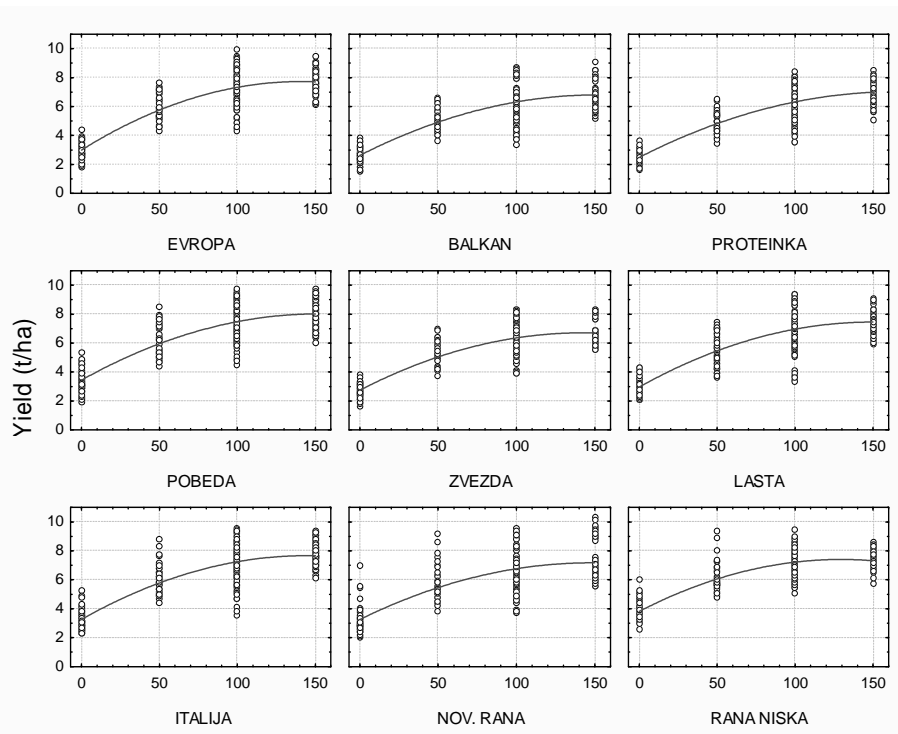**Figure 7:** The wheat yield against variant of fertilizer given variety and year.

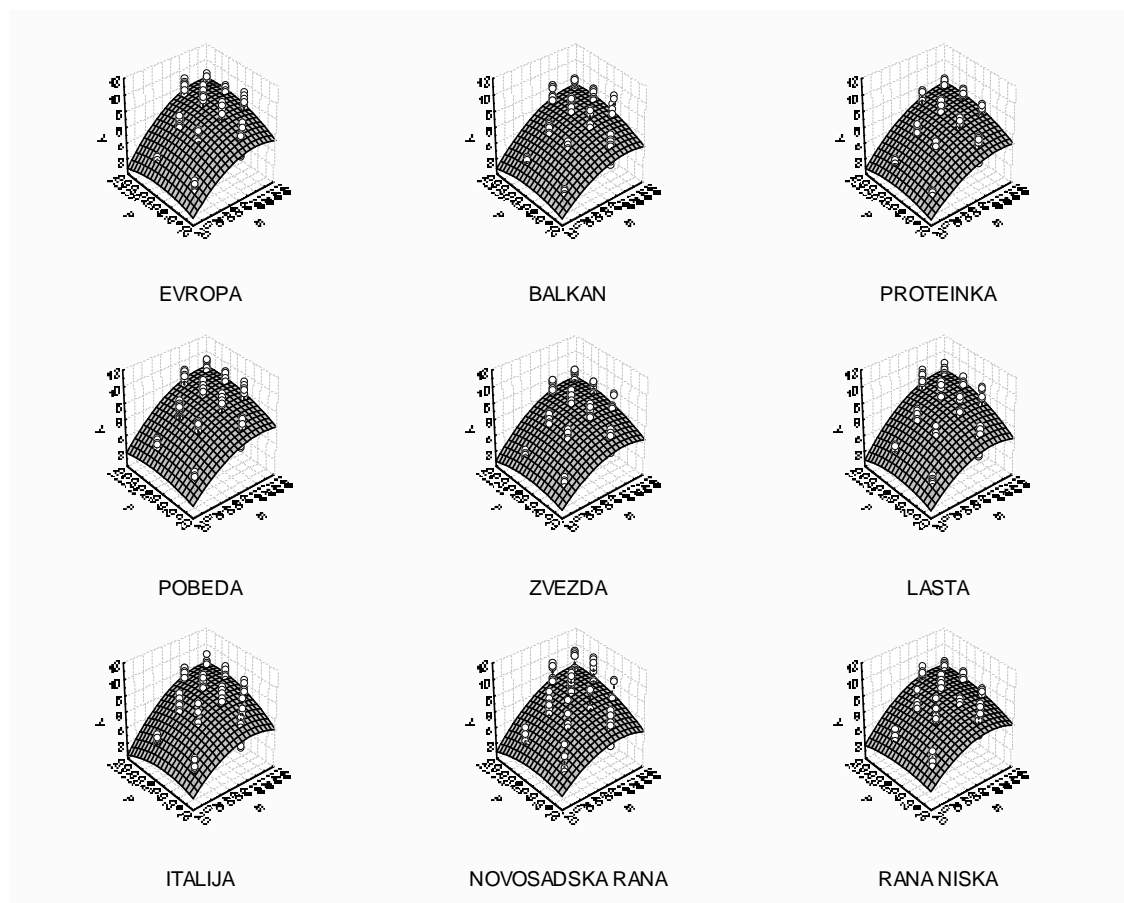**Figure 8:** The wheat yield against level of nitrogen given variety.

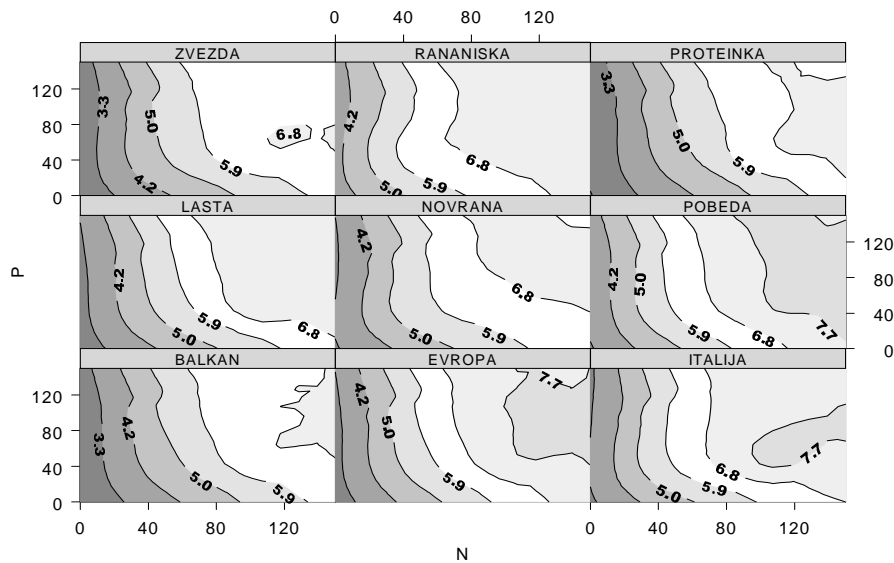**Figure 9:** 3-D trellis scatter plots and response surfaces for given variety.

Figure 10: Trellis contour plots.

5 Concluding remarks

Trellis graphics help in choosing the adequate regression model which reflects the influence of nitrogen, phosphorus and potassium on the wheat yield for individual varieties and for the whole experiment.

The separate influence of nitrogen and phosphorus may be approximated with quadratic functions that differ for the different varieties of wheat. The simultaneous effect may be modelled with response surfaces that is quadratic regression that contains linear and quadratic terms of nitrogen and phosphorus, and their interaction. The response surfaces also differ for different variants of the wheat.

The preliminary analysis based on trellis graphs suggested that the influence of the considered fertilizers on the wheat yield for the whole experiment may be modelled with quadratic function that includes dummy variables. Each dummy variable quantifies the influence of a particular variety of wheat.

Acknowledgements

References

[1] Becker, R.A., Cleveland, W.S., and Shyu, M.J. (1996): The visual design and control of trellis graphics displays. *Journal of Computational and Graphical Statistics*, **5**, 123-155.

[2] Becker, R.A. and Cleveland, W.S. (1996): *S-PLUS Trellis Graphics User's Manual*. Math, Trellis Versions 2.0 & 2.1., M, Inc. MathSoft.

[3] Cleveland, W.S. and Fuentes, M. (1997): *Trellis Display: Modeling Data from Designed Experiments* (Technical Report). Murray Hills, NJ: Bell Labs.

[4] Čobanović, K., Nikolić-Đorić, E., Jovanović, M., Malešević, M., and Mutavdžić, B. (2001): *Ispitivanje sortnih karakteristika pšenice primenom proizvodnih funkcija*. Zbornik radova sa XXVIII Jugoslovenskog simpozijuma o operacionim istraživanjima, SYMOP-IS 2001, 417-420, Beograd.

[5] Fisher, R. (1966): *Design of Experiments*. Edinburgh: Oliver and Boyd.

[6]   Nikolić-Đorić, E., Čobanović, K., and Lozanov-Crvenković, Z. (2006): *Statistical graphics and experimental data,* ICOST 7, Proceedings (Editors: Allan Rossman, Beth Chance), Brazil.

[7] STATISTICA 7.1 (2006): StatSoft,Inc. University Licence, Novi Sad.

[8] Tufte, E.R. (2001): *The Visual Display of Quantitative Information,* 2nd edn. Cheshire, CT: Graphic Press.

[9] Tukey, J.W. (1977): *Exploratory Data Analysis*, Reading, Mass: Addison-Wesley.