DOI: 10.22620/agrisci.2017.21.005

ANALYSIS OF THE DYNAMICS OF SUNFLOWER PRODUCTION IN BULGARIA BY STATISTICAL REGIONS

Diana Stoyanova*, Nikola Tsaikin

Agricultural University – Plovdiv

***E-mail: di_vest@yahoo.com**

Abstract

The paper presents the developed software for statistical analysis of time series on crop production, extracted from multidimensional database.

The software can be used by statisticians, economists and agricultural specialists to facilitate their access to statistical data and analytical methods.

The created algorithm updates automatically the database using public online sources.

The sunflower yields by regions and their sustainability have been studied.

Hierarchical clustering of Bulgarian regions is presented graphically.

Keywords: time series processing, regional analysis, sunflower production.

INTRODUCTION

Every research in the economic and social fields has to be based on sufficient and reliable information (Slaveva et al., 2016). For crop production in Bulgaria, such information can be extracted from two main sources: statistical database governed by Food and Agriculture Organization and database and paper books of National Statistical Institute of Bulgaria. These databases provide access to data, but not on-line processing. To overcome this disadvantage the Crop_Production BGEC database has been created under a Research Project 07-11, funded by the Research Centre of Agricultural University - Plovdiv. Crop_Production BGEC is a free access database, which stores time series for crop production in Bulgaria and other European countries. It has been created using Microsoft Access and contains information about 37 countries, 82 crops for the period 1961-2014. Data for Bulgaria are at the country and regional level.

The aim of this study is to expand the opportunities for updating, extraction, and analysis of data stored in Crop_Production BGEC.

MATERIALS AND METHODS

The database updating consists of three basic steps shown in Figure 1.

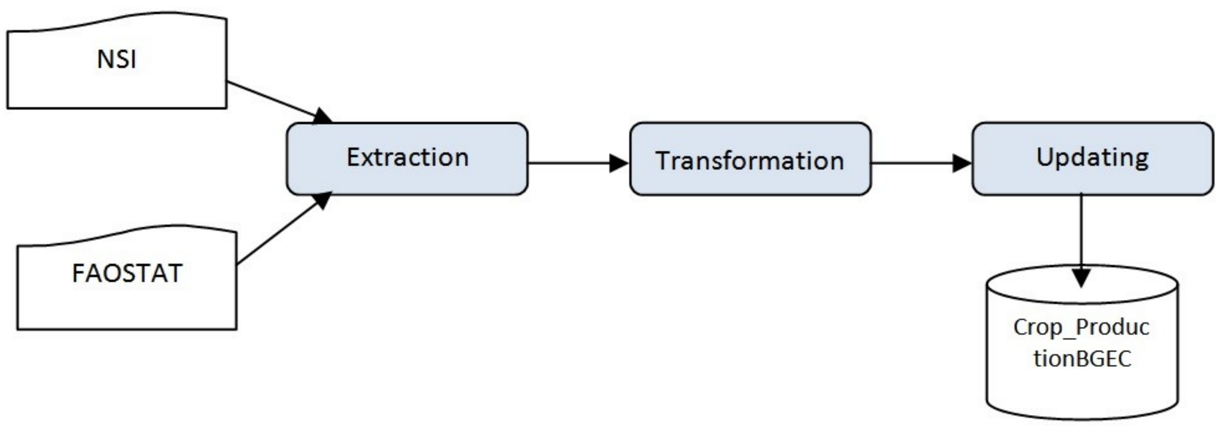

Fig. 1. *Crop_Production BGEC update – basic steps*

1. **Data extraction for a new period**. The new data used for Crop_Production BGEC update are extracted from FAOSTAT. Data by statistical regions are extracted from statistical yearbooks of NSI - Bulgaria.

   These yearbooks are available in paper or pdf format, which does not allow the direct input of information in the database.

   From the website of FAO (http://www.fao.org/faostat/en/#data/QC) the developed application automatically downloads the file *Production_Crops_E_All_Data_(Normalized).csv*, containing data about crop production and harvested area of 245 countries for the period 1961-2014.

   The file is in csv format, allowing the direct import of data to a Microsoft Access table.
2. **Transformation.** The second step is when the extracted data are converted into a suitable format for Crop_Production BGEC, and unnecessary data are deleted. For this purpose:
   - Only data on crops production and harvested area of European countries for a selected period is transformed from the downloaded information;
   - FAO number codes used to identify countries, statistical indicators and crops are replaced by their respective codes from Crop_Production BGEC.
3. **Updating**. In the final stage, the transformed data are appended into the table storing the time series.

Crops in Crop_Production BGEC are classified into four hierarchical levels, according to the Indicative Crop Classification (ICC) proposed by FAO (Food & Agriculture Org., 2005). These levels are group, class, subclass and order. It gives additional information which is not available in FAO database. Also, it can be used for statistical research and analysis (Onkov & Dimova, 2013).

The developed software facilitates the processing of time series stored in Crop_Production BGEC. This software allows:

- yield computing on the base of crop production and the harvested areas;
- computing of percentage of total area or production;
- plotting of diagrams, computing of descriptive statistics, performing one-way analysis of variance (ANOVA) and cluster analysis by direct execution of R language scripts in Microsoft Access.

  This provides an easy and affordable way for statistical processing even by users who do not have experience with this environment for statistical processing.

RESULTS AND DISCUSSION

The developed software is used for analysis of production and average yields of sunflower in Bulgaria by regions for the period 1961-2014. The six statistical regions that correspond to the NUTS2 level - North-Central (BG01), North-West (BG02), North East (BG03), South-West (BG04), South-East (BG05) and South-Central (BG06)-were officially created in 2000. That is why for the whole period data were obtained in three different ways:

- For the period 1961-1986 time series were calculated by aggregating data at a district (NUTS3) level
- From 1987 to 1997 Bulgaria was divided into 9 large districts. The North-Central region comprises the districts of Lovech and one part of Razgrad district.
- The North-East region comprises Varna and the other part of Razgrad.

  Therefore, for this period data for North-Central and North-East regions were computed by using the following equation (Poisvert et al., 2016):

$$D_{region\_year} = D_{district\_year} \frac{D_{region\_other\_years}}{D_{district\_other\_years}},$$

where:

D_{region_year} - data for a region for a given year

$D_{district\_year}$-aggregated data from the districts of Lovech, Razgrad, and Varna for a given year

$D_{region\_other\_years}$– aggregated data for a region throughout the entire study period

$D_{district\_other\_years}$– aggregated data from the three above mentioned districts throughout the entire study period

- For the period 2000-2014 time series were extracted from the statistical yearbooks of NSI.

Figure 2 shows the percentage of each region on the total production of sunflower in Bulgaria.

For the period 1961 - 2006 the North-East region (BG03) has the largest relative share in the total production - on average about 41%.

From 2007 to 2014 it shares the leading position with two other regions – North-Central (BG01) and North-West (BG02)-as their percentages are almost equal.

The South-West region (BG04) has the smallest share for the whole study period - on average about 1%.

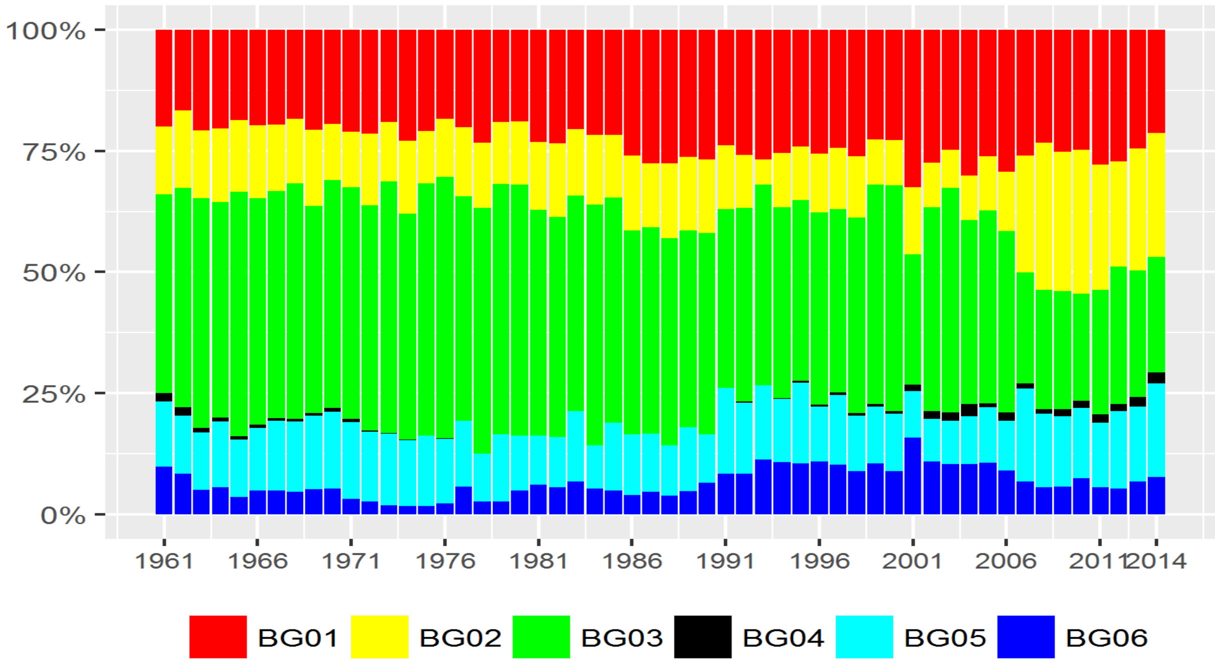

Fig. 2. *Distribution of sunflower production in Bulgaria by regions*

Figure 3 presents the grouping of statistical regions into clusters depending on their average sunflower yield for the period 1961-2014. Regions BG02 and BG05, BG01, and BG03 merge into two clusters at a small distance. This shows a high degree of similarity in their average yield values. BG06 and BG04 are considered as separate clusters, because of the large distance between them and the previous two clusters.

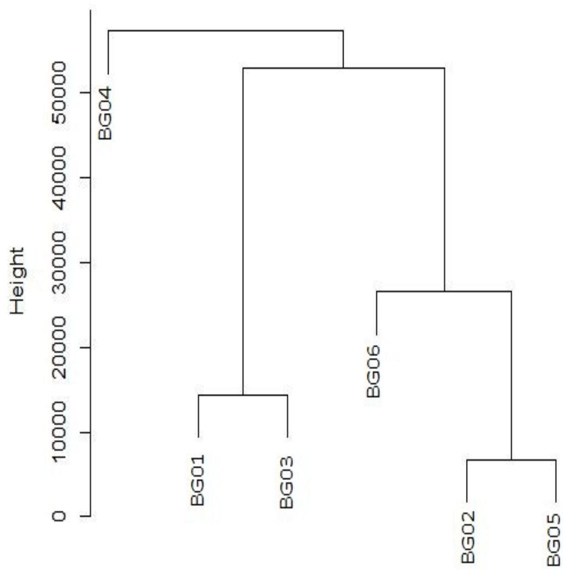

Fig. 3. *Results from the performed cluster analysis*

The results from the cluster analysis are confirmed from the performed one-way ANOVA. Table 1 shows the statistical assessment of the average sunflower yields by statistical regions and the values of standard deviation.

BG03 and BG01 regions form the group with the highest average yield, which is significantly different from the average yield of other regions. BG04 has the lowest average yield and the lowest degree of similarity with the yield of other regions. BG06 has the smallest standard deviation.

According to the results in Table 1, it can be concluded that BG01 and BG03 have the highest average sunflower yields.

Both regions have relatively high standard deviation values, which is an indicator for unstable crop production. The slight variation of yield for BG06 shows that sunflower production in this region is the most sustainable.

Figure 4 reflects the dynamics of sunflower yields by statistical regions. Although the lines show a strong variation in annual yields, the main tendency can be observed.

The entire study period can be interpreted by dividing it into three sub-periods: 1961 – 1989, 1990–2000, 2001-2014. The trend models of sunflower yields for these sub-periods have a different analytical type.

Table 1. Results from one-way ANOVA test performed on yields of sunflower by statistical regions

Statistical regions	Grouping using Tukey method		Standard deviation
BG02	1401,63	c	429,36
BG05	1319,24	bc	357,35
BG01	1634,03	d	398,58
BG03	1750,75	d	381,78
BG06	1156,19	b	300,90
BG04	916,65	a	396,61
The mean difference is significant at the 0.05 level			

From 1961 to 1989 the average yield of sunflower in Bulgaria is 1,400 kg/ha. The production for the entire period is relatively stable with slight fluctuations.

In the second period (1990-2000) there is a negative tendency in sunflower yields for all regions. The average yield during this period reaches the minimum - 1028 kg/ha.

From 2001 to 2014 the tendency becomes positive. The average yield during this period has the highest value - 1549 kg/ha.

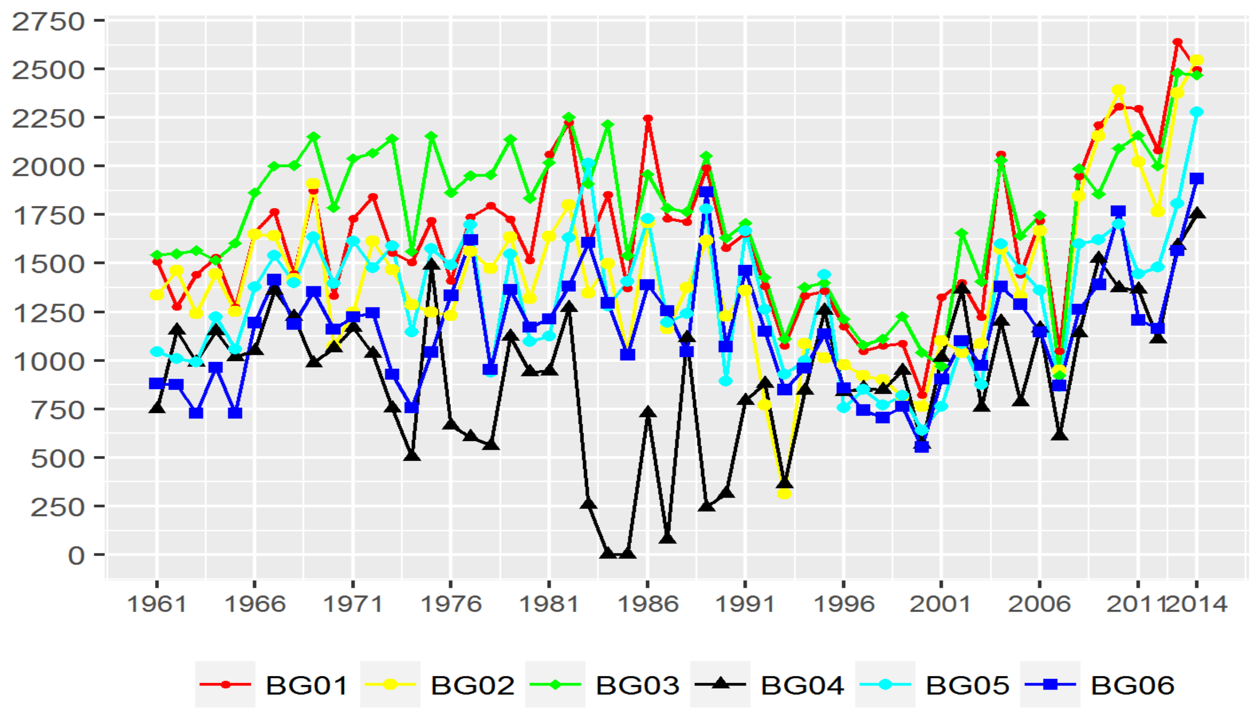

Fig. 4. *Dynamics of sunflower yields*

CONCLUSIONS

1. The developed software extends users abilities to update time series in multidimensional database Crop_Production BGEC and as well as to make statistical and cluster analysis.
2. The software is applied for studying sunflower yields in Bulgaria for a period 1961–2014.

3. The empirical results from the study may be of interest to researchers, agricultural specialists and can be used as a base for other theoretical and empirical studies in the field of agriculture.

REFERENCES

Food and Agriculture Organization of the United Nations, 2005). A System of Integrated Agricultural Censuses and Surveys: World Programme for the Census of Agriculture 2010, ISBN 978 – 92-5-105375-1.

Onkov, K., Dimova D., 2013. Data Mining Algorithm for Analyzing Spatial Time Series on Crop Production, Proceedings of the International Multidisciplinary Scientific Geoconference SGEM, Section "Informatics, Geoinformatics and Remote Sensing", vol. I, 69-75 (2013).

Poisvert, C., Curie, F., Moatar, F., 2016. Annual agricultural N surplus in France over a 70-year period. Nutrient Cycling in Agroecosystems, 1-16.

Slaveva, K., Petkov, P., Ivanov, L., Varbanov, T., Georgieva, N., 2016. Improving the training in statistics by using modern information and communication technologies, Almanach „Nauchni izsledvaniya", Academic Publishing House "D. A. Tsenov", Svishtov, 2016.