Почвознание агрохимия и екология, **53***, 3-4/2019*
Bulgarin Journal of Soil Science Agrochemisty and Ecology, **53***, 3-4/2019*

Method for determination of planting and harvesting dates by fuzzy regression

Alexander Sadovski

Bulgarian Science Center of the IEAS
E-mail: bsc.ieas@yahoo.com

Abstract

Sadovski, A. (2019). Metod for determination of planting and harvesting dates by fuzzy regression. *Bulgarin Journal of Soil Science Agrochemisty and Ecology, 53*(3-4), 10-16

In this paper we investigate planting and harvesting dates of maize, potatoes and sugarbeets grown on the territory of selected USA states, which have similar homoclimates with Bulgaria. Such data are very important in agricultural practice and particularly in scheduling irrigation of agricultural crops. Fuzzy set theory provides a strict mathematical framework in which vague conceptual phenomena can be precisely studied and described. A simple solution approach was applied to solve a general fuzzy system of linear equations with crisp explanatory variables and fuzzy unknown variable vectors.

Key words: Planting date, harvesting date, homoclimates, fuzzy regression.

Introduction

In this section we give some fundamental definition and concepts concerning to the content of this article.

Most modeling techniques and algorithms are designed for manipulating exact numerical data, or data which are uncertain in some well-defined statistical sense (Diamond, 1988). On the other hand, stochastic models may not be appropriate because necessary information is simply unavailable, or is very imprecise or even couched in terms that are not truly numerical. Fuzzy set theory has been regarded as a natural way of describing data of this type (Zadeh, 1965).

Fuzzy set theory provides a strict mathematical framework in which vague conceptual phenomena can be precisely and rigorously studied. It can also be considered as a modeling language well suited for situations in which fuzzy relations, criteria, and phenomena exist.

Following (Zimmermann, 1991), a fuzzy number may be defined as F = (b, g, h); where b denotes the center (or mode), g and h are the left spread (L) and right spread (R), respectively, L and R denote the left and right shape functions. A popular fuzzy number is the triangular fuzzy number (see Fig. 1).

The membership function of a triangular fuzzy number is defined by:

$$u(x)=\begin{cases}\dfrac{x-b}{g}+1, & b-g\le x\le b,\\ \dfrac{b-x}{h}+1, & b\le x\le b+h,\\ 0, & otherwise.\end{cases} \quad (1)$$

In this paper, the main aim is using of a method where right - hand - side is fuzzy vector and coefficients matrix is crisp. Crisp means - something clearly defined, deterministic in character. When we have crisp explanatory variables X_j, (j = 1, ..., n) and a fuzzy dependent variable Y_i≡ (b, g, h), (i = 1, ..., m), a model capable to incorporate the possible influence of the magnitude of the centers on the spreads, can be taken into account (D'Urso, 2003).

Different methods have been applied to find a solution of fuzzy linear systems. The least squares method is used (Diamond, 1988). The method for obtaining the fuzzy least squares estimators with the help of the extension principle in fuzzy sets theory is proposed in (Wu, 2003). In the work (Zareamoghaddam & Zareamoghaddam, 2014), the parameters of fuzzy linear regression based on the least squares approach is computed by ST-decomposition method. This method is not an iterative technique, however, it is a powerful method for nonsingular coefficient matrices. Different operations with fuzzy numbers are considered (Dubois & Prade, 1980), which helps a generalization of the usual tolerance analysis, so it can be applied in any scientific domain where quantities which are vaguely known have to be combined, provided that this uncertainty may be quantified. A general solution of m×n fuzzy linear systems is given in (Mikaeilvand & Noeiaghdam, 2012), where the original system is replaced by two m×n crisp linear system.

Pedomodels have become a popular topic in soil science and environmental research. They are predictive functions of certain soil properties based on other easily or cheaply measured properties. The common method for fitting pedomodels is to use classical regression analysis, based on the assumptions of data crispness and deterministic relations among variables. In modeling natural systems such as soil system, in which the above assumptions are not held true, prediction is influential and we must therefore attempt to analyze the behavior and structure of such systems more realistically. In the paper (Mohammadi & Taheri, 2004) fuzzy least squares regression as a means of fitting pedomodels are considered.

The term "homoclime" is used by Prescot (Prescot, 1938) for areas with similar climate. It refers to areas or regions, which possess similar climate. A recent study (Sadovski, 2019) reveals the similarity of Bulgarian conditions for growth and development of crops with corresponding USA states: North Dakota (ND), Washington (WA), South Dakota (SD), Wyoming (WY), Oregon (OR), Colorado (CO) and Utah (UT). The main task of the study is determination of planting and harvesting dates for some agricultural crops valid for Bulgaria (BG) by Fuzzy regression. These data are essential in solving practical tasks for scheduling irrigation of agricultural crops (Jensen, 1969).

Materials and methods

Materials for analysis are data for usual planting and harvesting dates of field crops in the USA (USDA, 2010). Data for maize, potatoes and sugarbeets from the above mentioned states are analyzed using available usual planting and harvesting dates (begin, most active period and end). For analysis purposes the given begin and end calendar dates must be converted to the corresponding serial number during the year. The following algorithm finds the consecutive number of the day in a year:

INPUT: Day = D9, Month = M9, Year = Y9;
N1 = INT(275 * M9 / 9);
N2 = INT((M9 + 9) / 12);
N3 = (1 + INT((Y9 - 4 * INT(Y9 / 4) + 2) / 3));
N = N1 - (N2 * N3) + D9 - 30;
OUTPUT: Day Number = N.

The length of the growing season (in days) of the crops in the various states is quoted from the Glenns Garden blog (Glenns-Garden, 2019).

Extensive compilation of observations was made of crop planting and harvesting dates from around the world to make a single, comprehensive crop calendar data set (Sacks et al., 2010). We shall use this data set for comparison with our results. Here it should be noted that the data file has an error in the dates for harvesting maize for Bulgaria. They do not correspond to the ones shown in the map of Bulgaria in the material Major World Crop Areas and Climatic Profiles (USDA, 2006).

System of simultaneous linear equations is important for studying and solving a large proportion of the problems in many topics in applied mathematics. Usually, in many applications, at least some of the system's parameters are represented by fuzzy rather than crisp numbers, and hence it is important to develop mathematical models and solving methods that would appropriately treat general fuzzy linear systems and solve them.

There is a simple solution approach to solve a general fuzzy system of linear equations (Mosleh et al., 2011). In case of fully fuzzy linear system with new notation where A, M and N are three crisp matrices, with the same size of , the matrices A, M and N are called the center matrix, the left and right spread matrices, respectively.

In our paper, the coefficient matrix is considered as real crisp whereas the unknown variable vectors are considered as fuzzy. In this case the matrices M and N are zero matrices. Using matrix notation, we have

$$A \otimes x = \tilde{b}, \quad (2)$$

or in expanded form (equation 3 below)

$$\begin{cases} (a_{11} \otimes x_1) \oplus (a_{12} \otimes x_2) \oplus \ldots \oplus (a_{1n} \otimes x_n) = \tilde{b}_1 \\ (a_{21} \otimes x_1) \oplus (a_{22} \otimes x_2) \oplus \ldots \oplus (a_{2n} \otimes x_n) = \tilde{b}_2 \\ . \\ (a_{m1} \otimes x_1) \oplus (a_{m2} \otimes x_2) \oplus \ldots \oplus (a_{mn} \otimes x_n) = \tilde{b}_m \end{cases}$$

where the crisp coefficient matrix is

and are nonnegative fuzzy numbers.
For calculation of (2) the following simple sequence is used:
1. Singular value decomposition is made

$$A = U \sum V^t \quad (4)$$

where U and V are orthogonal matrices; and $\sum$ is a diagonal matrix.
2. Pseudo-inverse matrices $\sum^+$ and $A^+ = V\sum^+U^t$ are found.
The following dependencies exist

$$\begin{aligned} Ax &= b, \\ Ay + Mx &= g, \quad (5) \\ Az + Nx &= h. \end{aligned}$$

3. From them consecutively the unknown values are calculated

$$\begin{aligned} x &= A^+b, \\ y &= A^+(g - MA^+b), \quad (6) \\ z &= A^+(h - NA^+b). \end{aligned}$$

From calculated values of x, y and z we can find the fuzzy solution

$$\tilde{b} = (b, g, h),$$

where $b = Ax$, $g = Ay$, $h = Az$.

All calculations by the method described are done with free software package GNU Octave, version 4.4.1.

Results and Discussion

The "usual planting dates" shown are the times when crops are usually planted in the fields. The "harvest dates" refer to the periods during which harvest of the crop actually occurs - combining, picking, cutting, pulling, and so on.

The unknown variable vectors we are looking for are:
Planting = (P-g, P-b, P-h),
Harvesting = (H-g, H-b, H-h).

Explanatory variables for the analysis are:
X_1 = average latitude (φ°N),
X_2 = elevation above sea level (m),
X_3 = season length (days),
X_4 = average daily temperatures March - June (°C),
X_5 = average amount of precipitation March -

June (mm),
X_6 = average daily temperatures August - November (°C),
X_7 = average amount of precipitation August - November (mm).

Variables X_4 and X_5 are used to find solution for planting fuzzy set. Variables X_6 and X_7 participate in calculations of harvesting fuzzy set.

We succeed with the help of the described method to obtain positive solution for fuzzy systems easily and rapidly.
Maize usual planting and harvesting dates input data are given in Table 1.

Therefore, by equations (4) - (6), the minimal solution of fully fuzzy linear system is presented on Table 2.

In the USDA manual (USDA, 2010) there are no sugarbeets records for the states of South Dakota and Utah. Unfortunately, Sachs' database (Sachs et al., 2010) does not contain data on planting and harvesting dates for potatoes and sugarbeets in Bulgaria. The comparison of the data for maize sowing indicated in this database is close to that obtained from the present analysis (start - end: respectively 84 - 147 and 86.7 - 154.2).

The estimated statistical characteristics: average, median and standard deviation allow easily the results of analyzed fuzzy sets to be interpreted. As can be expected, the dates for maize from the selected states are close to the average values obtained. Table 2 shows that the sowing and harvesting dates for maize in Bulgaria are relatively close to those of South Dakota and Utah.

The data used for potatoes and sugarbeets are shown in Table 3 and Table 5. The results of their analysis using the method described in the previous paragraph are presented in Tables 4 and Table 6.

Starting dates for sowing and harvesting potatoes vary considerably (this is evident from the corresponding standard deviation values 22.16 and 33.01). The dates for sugarbeets sowing also vary considerably, while those for harvesting are very close.

An important prospective task that can be solved by the Fuzzy set apparatus is to connect empirical data on crop growth and development with widely used, vegetation-based, empirical climate classification system developed by German botanist-climatologist Wladimir Köppen (Köppen, 1936).

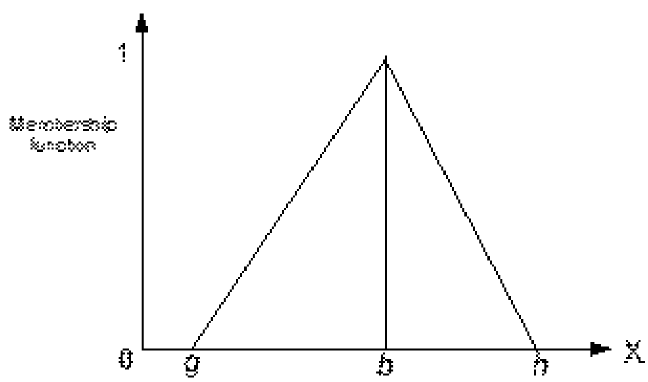

Fig. 1. Triangular fuzzy number

Table 1. Maize usual planting and harvesting dates input data

State			Explanatory variables						Planting			Harvesting	
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	P-g	P-b	P-h	H-g	H-b	H-h
ND	47.467	580	129	9.163	49.000	10.300	50.750	116	135.5	155	271	305.5	340
WA	47.275	520	153	11.938	62.500	13.563	40.750	100	126.0	152	268	298.5	329
SD	44.208	670	145	9.563	60.750	10.325	67.250	116	138.5	161	267	302.0	337
WY	43.000	2040	123	9.088	47.250	11.075	41.000	114	135.5	157	278	311.0	344
OR	44.150	1000	188	12.100	66.750	14.013	33.250	84	125.0	166	283	311.0	339
CO	39.000	2070	157	11.625	42.500	13.625	37.000	109	129.0	149	271	298.5	326
UT	39.500	1860	165	14.300	49.250	16.138	28.500	105	130.5	156	268	306.0	344
BG	42.750	472	114	12.986	57.979	18.562	53.609	84	115.5	147	220	269.5	319

Table 2. Maize planting and harvesting dates output results

State		Planting			Harvesting	
	P-g	P-b	P-h	H-g	H-b	H-h
ND	118.5	135.4	152.2	269.2	301.7	334.2
WA	102.7	132.1	161.5	269.9	303.3	336.8
SD	98.3	126.7	155.0	267.1	301.1	335.2
WY	118.9	139.0	159.2	278.5	311.4	344.3
OR	95.1	130.5	166.0	282.9	310.9	338.8
CO	108.3	129.2	150.0	272.1	303.8	335.6
UT	98.5	126.0	153.4	266.3	299.8	333.2
BG	86.7	115.9	145.2	220.1	270.0	319.9
Average	103.4	129.3	155.3	265.7	300.2	334.7
Median	100.6	129.9	154.2	269.5	302.5	335.4
St. dev.	11.28	6.94	6.67	19.33	12.98	6.92

Table 3. Potatoes usual planting and harvesting dates input data

State			Explanatory variables						Planting			Harvesting	
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	P-g	P-b	P-h	H-g	H-b	H-h
ND	47.467	580	129	9.163	49.000	10.300	50.750	118	137.5	157	145	242.5	340
WA	47.275	520	153	11.938	62.500	13.563	40.750	89	112.0	135	191	247.0	303
SD	44.208	670	145	9.563	60.750	10.325	67.250	69	104.5	140	191	247.0	303
WY	43.000	2040	123	9.088	47.250	11.075	41.000	74	102.0	130	191	255.0	319
OR	44.150	1000	188	12.100	66.750	14.013	33.250	84	125.0	166	196	257.5	319
CO	39.000	2070	157	11.625	42.500	13.625	37.000	122	136.5	151	249	269.5	290
UT	39.500	1860	165	14.300	49.250	16.138	28.500	115	143.0	171	237	270.0	303

Table 4. Potatoes planting and harvesting dates output results

State	Planting			Harvesting		
	P-g	P-b	P-h	H-g	H-b	H-h
ND	117.9	134.0	150.0	151.0	237.9	324.7
WA	90.2	119.6	149.0	179.5	249.3	319.0
SD	67.0	101.0	134.9	193.0	248.6	304.2
WY	74.6	102.3	129.9	192.9	258.2	323.4
OR	84.9	124.2	163.6	199.6	258.4	317.2
CO	122.0	141.8	161.7	235.8	266.4	297.1
UT	114.3	137.5	160.7	248.3	269.7	291.1
Average	95.8	122.9	150.0	200.0	255.5	311.0
Median	90.2	124.2	150.0	193.0	258.2	317.2
St. dev.	22.16	16.39	13.33	33.01	11.07	13.42

Table 5. Sugar beets usual planting and harvesting dates input data

State	Explanatory variables							Planting			Harvesting		
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	P-g	P-b	P-h	H-g	H-b	H-h
ND	47.467	580	129	9.163	49.000	10.300	50.750	109	127.5	146	260	279.0	298
WA	47.275	520	153	11.938	62.500	13.563	40.750	74	85.5	97	258	284.5	311
WY	43.000	2040	123	9.088	47.250	11.075	41.000	95	114.5	134	269	291.5	314
OR	44.150	1000	188	12.100	66.750	14.013	33.250	91	103.0	115	274	288.5	303
CO	39.000	2070	157	11.625	42.500	13.625	37.000	89	109.5	130	273	294.5	316

Table 6. Sugar beets planting and harvesting dates output results

State	Planting			Harvesting		
	P-g	P-b	P-h	H-g	H-b	H-h
ND	109.0	127.5	146.0	260.0	279.0	298.0
WA	74.0	85.5	97.0	258.0	284.5	311.0
WY	95.0	114.5	134.0	269.0	291.5	314.0
OR	91.0	103.0	115.0	274.0	288.5	303.0
CO	89.0	109.5	130.0	273.0	294.5	316.0
Average	91.6	108.0	124.4	266.8	287.6	308.4
Median	91.0	109.5	130.0	269.0	288.5	311.0
St. dev.	12.56	15.46	18.90	7.40	6.07	7.64

Conclusions

In this article we show the efficiency of proposed method for solving non-least-square linear fuzzy regression. This scheme for finding the positive solution of the fuzzy systems, when parameters are positive, it turns out to be quite satisfactory. Application of homoclimates approach combining with the mathematical apparatus of fuzzy regression is able to reveal qualitative and quantitative dependencies in soil science, agriculture and environmental research.

This data set could be used in different ways. The dependence of planting and harvesting dates on the climate in the season is important to bear in mind if we hope to predict how climate change might affect these dates. As a future research direction, it would be desirable to consider fuzziness of not only observations on dependent variable but also on explanatory variables.

Acknowledgements

The author would like to thank the European Union program for Research and Innovation Horizon 2020 for funding the grant agreement Grant # 826310.

References

Diamond, P. (1988). Fuzzy least squares. *Information Sciences, 46*, 141-157.

Dubois, D. & Prade, H. (1980). Fuzzy Sets and Systems: Theory and Applications. Academic Press, New York.

D'Urso, P. (2003). Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data. *Computational Statistics & Data Analysis, 42*, 47-72.

Jensen, M. (1969). Scheduling irrigations with computers. *J. Soil and Water Conservation, 24*(5),193-195.

Glenns-Garden. (2019). Average Frost Dates And Growing Season For Different US Cities. http://www.glenns-garden.com/average-frost-dates-and-growing-season-for-different-us-cities/

Köppen, W. (1936). Das geographische System der Klimate. Handbuch der Klimatologie, Band I, Teil C. Verlag Borntraeger, p. 44.

Mikaeilvand N. & Noeiaghdam, Z. (2012). The General Solutions of mxn Fuzzy Linear Systems. *Middle-East Journal of Scientific Research 11*(1), 128-133.

Mohammadi, J. & Taheri, S. M. (2004). Pedomodels fitting with fuzzy least squares regression. *Iranian Journal of Fuzzy Systems, 1*(2), 45-61.

Mosleh, M., Abbasbandy, S., & Otadi, M. (2011). A method for solving fully fuzzy linear system. *Mathematics Scientific Journal, 7*(2), 59-70.

Prescot, J. A. (1938). The climate of tropical Australia in relation to possible agricultural occupation. *Trans. Roy. Soc. S. Australia, 67*, 312-318.

Sacks, W. J., Deryng, D., Foley, J. A., & Ramankutty, N. (2010). Crop planting dates: an analysis of global patterns. *Global Ecology and Biogeography, 19*(5), 607-620.

Sadovski, A. (2019). Detection of similar Homoclimates by Numerical Analysis. *Bulgarian Journal of Soil Science, 4*(1), 69-75.

USDA (2006). Major World Crop Areas and Climatic Profiles. Agricultural Handbook Number 664, p. 293.

USDA (2010). Field Crops Usual Planting and Harvesting Dates. Agricultural Handbook Number 628, p. 51.

Wu, H. C. (2003). Linear regression analysis for fuzzy input and output data using the extension principle. *Computers & Mathematics with Applications, 45*(12), 1849-1859.

Zadeh, L. A. (1965). Fuzzy sets, *Inf. Control, 8*, 338–353.

Zareamoghaddam, H. & Zareamoghaddam, Z. (2014). A New Algorithm for Fuzzy Linear Regression with Crisp Inputs and Fuzzy Output. I*nternational Journal of Nonlinear Science, 17*(2), 128-134.

Zimmermann, H. J. (1991). Fuzzy Set Theory and its Application. Kluwer Academic Press, Dordrecht, p. 407.