Fuzzy C-means Clustering Applied to the Classification of *Glycyrrhiza uralensis* Communities in North China

Naiqi Song[1], Jintun Zhang[2, *]

[1]School of Chinese Materia Medica, Beijing University of Chinese Medicine, Beijing, China

[2]College of Life Sciences, Beijing Normal University, Beijing, China

Email address:

Zhangjt@bnu.edu.cn (Jintun Zhang)

*Corresponding author

To cite this article:

Naiqi Song, Jintun Zhang. Fuzzy C-means Clustering Applied to the Classification of *Glycyrrhiza uralensis* Communities in North China. *Automation, Control and Intelligent Systems*. Vol. 5, No. 5, 2017, pp. 73-77. doi: 10.11648/j.acis.20170505.13

Abstract: Fuzzy C-means clustering is a soft technique and has some advantages in ecological studies. *Glycyrrhiza uralensis* is an endangered medicinal plant species and mainly distributed in North China. Species data of 100 plots of 5 m x 5 m from semi-arid and arid regions in North China were analyzed by fuzzy C-means clustering. A hundred plots were classified into 12 vegetation formations by fuzzy C-means clustering. These formations were main communities dominated by *Glycyrrhiza uralensis* in North China. Each formation has its own composition, structure, distribution range and environment, and all of them should be protected effectively. The results suggest that fuzzy C-means clustering is an useful technique for classification of plant community.

Keywords: Fuzzy Set Theory, Numerical Classification, Licorice, Medicinal Plant, Ecological Conservation

1. Introduction

Quantitative methods, such as numerical classification and ordination, are significant in ecological analysis of plant communities [1]. Fuzzy C-means clustering is based on fuzzy set theory which has been developed by numerous scientists, and is now applied in various fields of sciences [2, 3]. It is applicable in ecology because the description of ecological systems is not always possible in terms of a binary approach [4]. Fuzzy C-means clustering is the only soft method in the clustering family and should have some advantages [5]. Ecological communities, such as *Glycyrrhiza uralensis* communities, have great variations in their species composition respond more or less independently to environmental variables. Because of this, methods based on fuzzy mathematics might be more appropriate in ecological analysis [6, 7].

Medicinal plant is important natural resource and significant for people health in many countries and regions, such as China, Japan, Korea, India and so forth [8, 9]. Licorice (*Glycyrrhiza uralensis*) is one of the most popular Chinese herbal medicines and a significant resource plant species.

Beside medicine, it is also widely used in food, tobacco, chemical industries [10, 11]. Like licorice, most medicinal plants are harvested in the wild and the extended use has led to some medicinal species endangered and tending to be extinct, and the protection of such medicinal species has been urgent [12, 13]. Further more, conservation of medicinal plant species is important in term of biodiversity conservation [11, 14]. The studies on medicinal plants and their communities are the basis for their conservation and restoration. Some studies on the taxonomy, biochemistry, pharmacology and genetics of *Glycyrrhiza uralensis* have been carried out [15, 16], while the research concerning its community ecology is limited [7]. This study aims to identify *Glycyrrhiza uralensis* communities and analyze their characteristics on composition, structure and environment in North China.

2. Methods

2.1. Fuzzy C-means Clustering

Classification is the process of dividing plots into clusters (communities) so that plots in the same cluster are as similar as possible, and plots in different clusters are as dissimilar as

possible. Fuzzy C-means clustering is a soft classification technique [5, 7] in which a plot can belong to more than one cluster, and associated with each plot is a set of membership levels. The Fuzzy C-means clustering attempts to divide a finite collection of n elements (plots) $X = \{x_1, x_2, ..., x_n\}$ into a collection of *C* fuzzy clusters with respect to some given criterion. Given a finite set of data matrix (species × plots), the method calculates a list of *C* cluster centres $V = \{v_1, v_2, ..., v_c\}$ and a partition matrix $U = u_{ij} \in [0,1]$, $i =$ 1, 2, …, *N; j* = 1, 2, …, *C*, where each element U_{ij} tells the degree to which plot x_i belongs to cluster C_j. Fuzzy C-means clustering aims to minimize an objective function, i.e. it minimizing the within group sum of squares, J_m (*U, V, A*), which is given by

$$J_m(U,V,A) = \sum_{i=1}^{N} \bullet \sum_{j=1}^{C} (U_{ij})^m (dA_{ij})^2 \qquad (1)$$

Where *i* = 1, 2, …, *N* = the number of plots; *j* = 1, 2, …, *C* = the number of clusters; $U = \{U_{ij}\}$ = the matrix of membership values, U_{ij} is the membership of plot *i* in cluster *j*; *V* is a matrix of cluster centers; *m* is fuzzifier which determines the level of cluster fuzziness ($1 \leq m < \infty$). A large *m* results in smaller memberships u_{ij} and hence, fuzzier clusters. In the absence of experimentation or domain knowledge, *m* is commonly set to 2. $(dA_{ij})^2$ is the distance index:

$$(dA_{ij})^2 = \| X_i - V_j \|^2 \; A = (X_i - V_j)^T A(X_i - V_j) \qquad (2)$$

X_i is the vector of attribute measurements in plots, usually a vector of ordination scores; V_j is the centre of cluster *j*, if *A* is a unit matrix, then

$$(dA_{ij})^2 = \| X_i - V_j \|^2 \qquad (3)$$

Based on the algorithm above, the procedure of fuzzy C-means clustering is as follows:

(1) Selecting a common ordination method, we use Detrended Correspondence Analysis (DCA), and do ordination analysis for species data. The first DCA axis was used as the basic data X_i in clustering.

(2) Determining the number of clusters *C* according to ecological meanings.

(3) Assigning the primary membership values, U_0. For a plot, any value can be given as its membership value in cluster *j,* but the sum of memberships for a plot must be equal to 1:

$$\sum U_j = 1$$

(4) Calculating the cluster centres V_j and distance $(dA_{ij})^2$:

$$V_j = \sum_{i=1}^{N} (U_{ij})^m X_i / \sum_{i=1}^{N} (U_{ij})^m \qquad (4)$$

$(dA_{ij})^2$ is calculated using equations (2) and (3).

(5) Calculating the new membership values based on new V_j and $(dA_{ij})^2$ values:

$$U_{ij} = \left\{ \sum_{k=1}^{C} \left[\frac{(dA_{ij})^2}{(dA_{ik})^2} \right]^{\frac{1}{m-1}} \right\}^{-1} \qquad (5)$$

$(i = 1, 2,…, N; j = k = 1, 2,…, C)$

(6) Based on the new membership values *U*, we go back to the fourth step and calculated the next turn V_j, $(dA_{ij})^2$ and U_{ij} iteratively, and until the membership values become approximately stable.

(7) Classified plots into clusters based on the final *U*. We can use *U* to identify the relationships among plots and communities directly. However for clustering purpose, a plot should belong to the cluster in which it had the maximum membership value.

2.2. *Glycyrrhiza uralensis* Community Data

Based on a general survey of *Glycyrrhiza uralensis* and its community distribution, five study regions, Chifeng (in inner Mongolia), Hengjinqi (in inner Mongolia), Minqin (in Ganshu), Aletai (in Xinjing) and Kashi (in Xinjing), were selected as sampling sites [7]. Twenty plots of 5 m × 5 m were established randomly at each site which is over 40 ha in area. The coverage, mean height, individual number for shrub and herb species was measured in each plot. The coverage of species was estimated by eyes, and the heights were measured using a tape ruler. In total, 191 plant species were recorded in 100 plots. Elevation, slope and aspect for each plot were also measured and recorded. The elevation was measured by a GPS, slope and aspect measured by a compass meter.

The Importance Value (IV) of species was calculated and used as data in clustering analysis. The importance value was calculated by the formulas [7]:

IV _Shrubs_ = (Relative abundance +Relative coverage + Relative height)/3

IV _Herbs_ = (Relative coverage + Relative height)/2

The relative abundance refers to the percentage of one species abundance over the sum of all species abundance in a plot, relative coverage to the percentage of one species coverage over the sum of all species coverage in a plot, and relative height to the percentage of one species mean height over the sum of all species mean height in a plot. The species data matrix is consisted of importance values of 191 species in 100 plots.

3. Results

Based on the investigation, we determined *C* equal to twelve. Fuzzy C-means clustering is a nonhierarchical method which provides *C* clusters and their plot compositions (Table 1). It classified 100 plots into 12 clusters, representing 12 *Glycyrrhiza uralensis* communities, e.g. 12 vegetation formations. The name and characteristics in species composition, structure and environment of each community are described below.

I. Form. *Glycyrrhiza uralensis + Stipa bungeana*. It is distributed from 380 to 605 m in hills with slope 10 – 20° in sunny and semi-sunny slope and chestnut soil. Its disturbance intensity is medium and heavy. The community has a total cover of 75%, a shrub layer cover of 20% and an herb layer cover of 65%. The common species are *Ziziphus jujuba* var. *spinosa, Lespedeza darurica, Pedicularis resupinata, Potentilla anserine, Saussurea epilobioides, Artemisia sacrorum, Artemisia mongolica, Cynanchum hancockianum,* and *Vicia amoena*. The average cover of *Glycyrrhiza uralensis* in this community is 25% with a density of 3600 ha^{-1}.

Table 1. *Plot composition of 12 Glycyrrhiza uralensis formations recognized by fuzzy C-means clustering in North China.*

Formation Number	Plot composition
I	1, 2, 3, 6, 8, 11
II	9, 12, 13, 18, 20, 35, 36
III	4, 5, 7, 10, 14, 15, 16, 17, 19
IV	21, 24, 26, 27, 28, 29, 32, 33, 34, 37, 40
V	22, 25, 30, 31, 39
VI	42, 43, 44, 45, 46, 48, 54, 56, 58
VII	41, 47, 49, 50, 51, 52, 53, 55, 57, 59, 60
VIII	63, 64, 65, 66, 67, 71, 72, 73, 74, 75, 77, 78, 79, 80
IX	61, 62, 70, 76
X	23, 38, 68, 69
XI	83, 84, 85, 87, 91, 92, 94, 95, 96, 97
XII	81, 82, 86, 88, 89, 90, 93, 98, 100

II. Form. *Glycyrrhiza uralensis + Aneurolepidium chinense*. It is distributed from 350 to 650 m in hills with slope 10 – 20° in sunny and semi-sunny slope and sandy chestnut soil. Its disturbance intensity is heavy. The community has a total cover of 80%, a shrub layer cover of 10% and an herb layer cover of 75%. The common species are *Haloxylon ammodendron, Caragana pygmaea, Lespedeza darurica, Berberis sibirica, Stipa glareosa, Artemisia mongolica, Cynanchum hancockianum, Carex duriuscul, Astragalus melilotoides,* and *Carex stenophylloides*. The average cover of *Glycyrrhiza uralensis* in this community is 28% with a density of 3950 ha^{-1}.

III. Form. *Glycyrrhiza uralensis + Potentilla anserine*. It is distributed from 350 to 580 m in hills with slope 10 – 25° in sunny and semi-sunny slope and chestnut and cinnamon soil. Its disturbance intensity is heavy and medium. The community has a total cover of 80%, a shrub layer cover of 5% and an herb layer cover of 80%. The common species are *Lespedeza darurica, Artemisia mongolica, Stipa bungeana, Vicia amoena, Carex duriuscul,* and *Saussurea amara*. The average cover of *Glycyrrhiza uralensis* in this community is 26% with a density of 4120 ha^{-1}.

IV. Form. *Glycyrrhiza uralensis + Artemisia ordosica*. It is distributed from 300 to 650 m in hills with slope 15 – 30° in sunny and semi-sunny slope and sandy chestnut soil. Its disturbance intensity is heavy and medium. The community has a total cover of 85%, a shrub layer cover of 25% and an herb layer cover of 70%. The common species are *Caryopteris mongolica, Lespedeza darurica, Artemisia mongolica, Stipa bungeana, Trigonella ruthenica, Trigonella ruthenica, Astragalus melilotoides,* and *Ephedra przewalskii*. The

average cover of *Glycyrrhiza uralensis* in this community is 27% with a density of 3680 ha^{-1}.

V. Form. *Glycyrrhiza uralensis + Carex duriuscul + Aneurolepidium chinense*. It is distributed from 300 to 600 m in hills with slope 20 – 30° in sunny，semi-sunny and semi-shady slope and chestnut soil. Its disturbance intensity is medium. The community has a total cover of 75%, a shrub layer cover of 10% and an herb layer cover of 70%. The common species are *Lespedeza darurica, Haloxylon Ammodendron, Berberis sibirica, Cynanchum hancockianum, Stipa glareosa, Caragana pygmaea, Artemisia scoparia, Carex stenophylloides, Potencilla acaulis,* and *Astragalus melilotoides*. The average cover of *Glycyrrhiza uralensis* in this community is 25% with a density of 4500 ha^{-1}.

VI. Form. *Glycyrrhiza uralensis + Polygonum bistorta*. It is distributed from 300 to 500 m in hills with slope 20 – 30° in sunny，semi-sunny and semi-shady slope and sandy chestnut soil. Its disturbance intensity is medium and heavy. The community has a total cover of 70%, a shrub layer cover of 5% and an herb layer cover of 70%. The common species are *Oxytropis myriophylla, Polygonum divaricatum, Adenophora gmeliniia, Potencilla acaulis, Suaeda prostrate, Astragalus melilotoides, Allium condensatum, Artemisia ordosica,* and *Oxytropis grandiflora*. The average cover of *Glycyrrhiza uralensis* in this community is 35% with a density of 5700 ha^{-1}.

VII. Form. *Glycyrrhiza uralensis + Ephedra przewalskii + Cancrinia discoidea*. It is distributed from 300 to 500 m in hills with slope 20 – 30° in sunny，semi-sunny and semi-shady slope and sandy chestnut soil. Its disturbance intensity is medium and heavy. The community has a total cover of 70%, a shrub layer cover of 15% and an herb layer cover of 60%. The common species are *Caragana korshinskii, Elaeagnus, mooceroftii, Suaeda prostrate, Artemisias phaerocephala, Saussurea laciniata, Saposhnikovia divariicata, Oxytropis glabra,* and *Artemisia ordosica*. The average cover of *Glycyrrhiza uralensis* in this community is 40% with a density of 59500 ha^{-1}.

VIII. Form. *Glycyrrhiza uralensis + Artemisia frigida*. It is distributed from 280 to 500 m in hills with slope 15 – 25° in sunny and semi-sunny slope and sandy soil. Its disturbance intensity is medium and heavy. The community has a total cover of 65%, a shrub layer cover of 10% and an herb layer cover of 60%. The common species are *Salicornia Bigelivii, Carex duriuscula, Stipa sareptana, Artemisias phaerocephala, Alopecurus pratensis, Saposhnikovia divariicata,* and *Carex pediformis*. The average cover of *Glycyrrhiza uralensis* in this community is 37% with a density of 4900 ha^{-1}.

IX. Form. *Glycyrrhiza uralensis + Carex pediformis + Stipa sareptana*. It is distributed from 350 to 650 m in hills with slope 15 – 35° in sunny and semi-sunny slope and sandy soil. Its disturbance intensity is heavy. The community has a total cover of 60%, a shrub layer cover of 5% and an herb layer cover of 58%. The common species are *Cleistogenes squarrosa, Caragana pygmaea, Hordeum brevisublatum, Ephedra sinica, Achnatherum sibiricum, Artemisia frigida, Viola tianschanica, Carex duriuscula,* and *Alopecurus pratensis*. The average cover of *Glycyrrhiza uralensis* in this

community is 33% with a density of 5500 ha^{-1}.

X. Form. *Glycyrrhiza uralensis + Astragalinae triloa + Stipa sareptana*. It is distributed from 400 to 700 m in hills with slope 10 – 30° in sunny and semi-sunny slope and sandy soil. Its disturbance intensity is heavy. The community has a total cover of 75%, a shrub layer cover of 10% and an herb layer cover of 70%. The common species are *Artemisia scoparia*, *Kochia prostrate, Potencilla acaulis*, *Artemisia frigida*, *Ceratoides lates* and *Atraphaxis frutescus*. The average cover of *Glycyrrhiza uralensis* in this community is 40% with a density of 6000 ha^{-1}.

XI. Form. *Glycyrrhiza uralensis + Aneurolepidium chinense +Stipa sareptana*. It is distributed from 400 to 750 m in hills with slope 15 – 35° in sunny and semi-sunny slope and sandy soil. Its disturbance intensity is heavy. The community has a total cover of 70%, a shrub layer cover of 15% and an herb layer cover of 65%. The common species are *Caragana pygmaea, Astragalinae triloa, Stipa parpurea*, *Festuca logae*, *Artemisia kaschgarica*, *Polygonum viiiparum*, *Ephedra equisetina, Glycyrrhiza inflate*, and *Alyssum desertorum*. The average cover of *Glycyrrhiza uralensis* in this community is 29% with a density of 4100 ha^{-1}.

XII. Form. *Glycyrrhiza uralensis + Festuca logae + Stipa sareptana*. It is distributed from 400 to 800 m in hills with slope 20 – 35° in sunny and semi-sunny slope and sandy soil. Its disturbance intensity is heavy. The community has a total cover of 80%, a shrub layer cover of 10% and an herb layer cover of 75%. The common species are *Artemisia parvula, Scorzonera divaricata, Roegneria kamoji, Potentilla bifurca*, *Carex duriuscula,* and *Ranunculus japonicas*. The average cover of *Glycyrrhiza uralensis* in this community is 30% with a density of 4000 ha^{-1}.

4. Discussion

Fuzzy C-means clustering successfully classified 100 plots into 12 communities dominated by *Glycyrrhiza uralensis*. Theoretically, fuzzy C-means clustering is the only one soft clustering method and can deal with much imprecise and incomplete information and has advantages in solving non-linear problems and in studying complex system [2, 17]. Vegetation systems are very complex with interactions of species, communities, environmental variables, and so on [18]. Therefore, fuzzy C-means clustering should describe ecological phenomena and rules better [5, 6]. This study proved that fuzzy C-means clustering is fully usable in classification analysis of plant communities [1, 19].

Glycyrrhiza uralensis communities recognizing by fuzzy C-means clustering varied greatly in species composition, structure and distribution area [20]. These twelve formations represent the general vegetation of *Glycyrrhiza uralensis* in northern China [21, 22]. They distribute from temperate meadow grassland, through typical temperate steppe and desert steppe to temperate desert regions from east to west in North China [15]. These communities are important not only for protection of medicinal plant species but also for conservation of ecosystems and their environments in semi-arid and arid

regions in North China [23, 24]. The classification results by fuzzy C-means clustering are reasonable according to vegetation classification system of China [25-27]. These further confirm that fuzzy C-means clustering is an effective technique in vegetation analysis [28, 29].

Acknowledgements

References

[1] Greig-Smith P. Quantitative plant ecology. 3rd edition. Blackwell Scientific Publications, London, 1-311, 1983.

[2] Kaufmann A. Introduction to the theory of fuzzy subsets: Vol. 1: foundamental theoretical elements. Academic Press, London, 1975.

[3] Sarbu C, Zwanziger HW. Fuzzy classification and comparison of some Romanian and German mineral waters. *Analytical Letters,* 2001, 34: 1541-1552.

[4] Zhang J-T, Zhang B, Qian ZY. Functional diversity of Cercidiphyllum japonicum, communities in the Shennongjia Reserve, central China. *Journal of Forestry Research*, 2015, 26(1): 171–177.

[5] Bezdek J C. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York, 1981.

[6] Podani J. Introduction to the Exploration of Multivariate Biological Data. Leiden: Backhuys Publishers, 2000.

[7] Zhang J-T. Quantitative ecology (2nd edition), Beijing: Science Press. (In Chinese), 2011.

[8] Zhou CM. The outline of the normative Cultivation Techniques for *Glycyrrhiza uralensis*. *Research and Information on Traditional Chinese Medicine*, 2003, 5(2): 25-28.

[9] Hussain S, Hore DK. Collection and conservation of major medicinal pants of Darjeeling and Sikkim Himalayas. *Indian Journal of Traditional Knowledge*, 2007, 6: 352-357.

[10] Zhou CM. Cultivation Techniques for *Glycyrrhiza uralensis*. *Xinjiang Farmland Reclamation Science and Technology*, 2006, 15 (1): 14-15.

[11] Nautiyal BP, Nautiyal MC, Khanduri VP, Rawat N. 2009. Floral Biology of Aconitum heterophyllum Wall: A Critically Endangered Alpine Medicinal Plant of Himalaya, India. *Turkish Journal of Botany*, 33: 13-20

[12] Kate KT, Laird SA. The commercial use of biodiversity. Earthscan Publications Ltd., London, 1999.

[13] Larsen HO, Olsen CS. Unsustainable collection and unfair trade? Uncovering and assessing assumptions regarding Central Himalayan medicinal plant conservation. *Biodiversity and Conservation*, 2007, 16: 1679-1697.

[14] Ji YB, Jiang W, Fan YL. Advance in studies on flavonoids of licorice. *Chinese Traditional Herb Drugs*, 2004, 35(9): 5-6.

[15] Hu JF, Shen FJ. Structure of a new alkaloid from the roots of *Glycyrrhiza yunnanensis*. *Chemistry Journal of Chinese University*, 1995, 16(8): 1245-1247.

[16] Zhang JH, Yao J, Yang YL. Analysis and menstruate contents of alkaloid in licorice. *Acta Botanica Boreal- Occidenta Sinica*, 2001, 21(5): 1259-1262.

[17] Giraudel JL, Aurelle D, Lek S. Application of the self-organizing mapping and fuzzy clustering microsatellite data: how to detect genetic structure in brown trout (*Salmo trutta*) populations. In: Lek S., Gueguan J. F. (Eds.), Artificial Neuronal Networks: Application to Ecology and Evolution, Environmental Science. Springer, Berlin. 2000.

[18] Zhang J-T, Li M，Nie EB. Pattern of functional diversity along an altitudinal gradient in the Baihua Mountain Reserve of Beijing, China. *Brazilian Journal of Botany,* 2014, 37: 37-45.

[19] ter Braak CJF, Šmilauer P. CANOCO Reference Manual and User's Guide to *Canoco for Windows*. Software for Canonical Community Ordination (version 4.5). Centre for Biometry Wageningen (Wageningen, NL) and Microcomputer Power (Ithaca NY, USA), pp. 352, 2001.

[20] Pan BR, Zhang YM. Characteristics and conservation of biodiversity in Xinjiang. *Science in China Series D – Earth Sciences*, 2002, 45 (Suppl.): 174-179.

[21] Wu ZY (ed). Vegetation of China. Science Press, Beijing, 1980.

[22] Zhang J-T, Xu B, Li M. Diversity of communities dominated by *Glycyrrhiza uralensis,* an endangered medicinal plant species, along a precipitation gradient in China. *Botanical Studies*, 2011, 52: 493-501.

[23] Zhang J-T. A study on relations of vegetation, climate and soils in Shanxi province, China. *Plant Ecology*, 2002, 162(1): 23-31.

[24] Paschke MW, Redente E F, Brown S L. Biology and establishment of mountain shrubs on mining disturbances in the Rocky Mountains, USA. *Land Degradation & Development,* 2003, 14 (5): 459-480.

[25] Shen T, Fu T, Yang G. Relationship between Hybrid Performance and Genentic Diversity Based on SSRs and ISSRs in Brassicanapus. *Agricultural Science in China*, 2003, 2 (10): 1083-1090.

[26] Glaser B, Turrion M-B, Solomon D, Ni A, Zech W. Soil organic matter quantity and quality in mountain soils of the Alay Range, Kyrgyzia, affected by land use change. *Biology and Fertility of Soils,* 2000, 31 (5): 407-413.

[27] Zhao ZH, Cao JG, Li QY. Study of variety trends of flavonoids in *Glycyrrhiza uralensis* in the west of Heilongjiang, China. *Bulletin Botanical Research*, 2004, 24(2): 235-239.

[28] Sadia S, Zhang J, Tariq A, Bai XH, Shedayi A, Cao K, Mazari P, Aslam S, Ahmad L. Species diversity, vegetation pattern and conservation of Gentiana macrophylla Pall. Communities in Dongling Mountain meadow, Beijing, China. *Pakistan Journal of Botany*, 2017, 49(5): 1725-1734.

[29] Song NQ, Zhang J, Zhao FG. The PCA index for measuring functional diversity and its application to *Juglans mandshurica* communities in the Beijing mountains, China. *International Journal of Biomathematics,* 2017, 10(7): 1750097 (13 pages).