Eurasian J Soil Sci 2016, 5 (3) 231 - 240

Eurasian Journal of Soil Science

Journal homepage : http://fesss.org/eurasian_journal_of_soil_science.asp

An application of Embedded Markov chain for soil sequences: Case study in North Western part of Algeria

Lotfi Mustapha Kazi-Tani*, Abdelaziz Gaouar

Department of Agronomy, Faculty of Nature and Life Sciences, Earth and Universe, University of Tlemcen, Algeria

Abstract

Embedded Markov chain (EMC) has long history in geological domains, particularly to define the most representative sequences from statigraphic logs. In other words, what is viewed as a meaningless and disordered stratigraphic layer stack can be reorganized in a meaningful sequence by using EMC. This method was transposed in this paper to obtain soil sequences from data retrieved from soil map made by authors, covering a part of the region of Traras (N.W. of Algeria) and containing 13 major soil types. Each major soil type occupies at least one polygon in the map and allow to establish soil adjacencies, which have been tabulated in a matrix regardless to the direction. Three EMC methods have been tested, Walker, Harper and Türk using Strati-signal software and to erect soil relationship diagrams (SRD) representing the most significant links between soils. Significant test is the main difference between the above mentioned three EMC methods. It has been shown that Harper method is quite insensitive to small number of transitions. Besides, all three methods agreed for one soil sequence made by four soils: lithics leptosols- cambisols chormics- cambisols calcarics- fluvisols representing theoretical catena the most representative to the study area. This soil sequence is relevant to the study region and even to the whole Mediterranean region, and is commanded by the topography and the Mediterranean bioclimate. Walker SRD is the most realistic but the most difficult to interpret because of the high number of soil links, Harper SRD gives interesting results. Although the results didn't bring something new to the soil interpretation and soil pedogensis but EMC applied to a finer scale may highlights other hidden relationships between soils.

Article Info

Received : 10.01.2016
Accepted : 15.02.2016

Keywords: Adjacency matrix, catena, markov chain, randomness test, soil map, soil relationship diagram

Introduction

It is well known from readings that Markov chain is based upon the assumption that knowledge of the past can sometimes be useful to explain the present and vice versa. This assumption was originally related to time series events, and has been applied to space phenomena as well. This "spatial" markovian assumption is closely akin to the first principle of Tobler (1970): "All things are related but nearby things are more related than distant things". Geologists are the most prolific for applying Markov chain models to their studies, and have developed several methods mainly in

* Corresponding author.
Department of Agronomy, Faculty of Nature and Life Sciences, Earth and Universe, University of Tlemcen, Tlemcen, 13000 Algeria
Tel.: +2130551490763 E-mail address: zarifet@yahoo.com
e-ISSN: 2147-4249 DOI: http://dx.doi.org/10.18393/ejss.2016.3.231-240

embedded Markov chain (Gingerich, 1969; Selley, 1969; Walker, 1979; Harper, 1984a; Türk, 1982; Ndiaye, 2007); unlike pedologists are less inclined to apply this stochastic model. Nevertheless, concerning Markov chain in soil studies several works have been done such as Burgess and Webster (1984a,b), Li et al. (1999), Li et al. (2004), Li and Zhang (2006), Sun et al. (2013), but none of them have worked with embedded Markov chain.

The key question is to know whether the occurrence of a soil at one point is completely dependent of the contiguous soil which means that the hypothesis of randomness must be disproved statistically.

In this article, embedded Markov chain (EMC) methods was applied to soil map covering the eastern part of Traras region located on the north western part of Algeria in Tlemcen vicinity. The purpose is to explore spatial links between soils in order to extract soil sequences the most representative. It is worth mentioning that soil sequence in this study has the same meaning as catena, and it can be defined as: "A chain, string, or connected series of soils, related by their sequence in a landscape. Synonymous in part with toposequence chronosequence. The variability of soils in a topographic sequence is a function of gradient and position on the slope" (Birkeland, 1999). The concept of catena is fundamental to explain the sequence of soils following both hill-slopes and landscapes. Water and gravity are the main forces involved to sequential variation of soils. This led us to propose the use of Markov chain as an analytical tool in the study of horizontal relationship of soil types.

Material and Methods

Brief presentation of soil map

The study area is located in the eastern part of the Mountains of Traras and covers an area of about 963km^2, under semi-arid Mediterranean bioclimate. Vegetation is structured principally by arar tree (*Tetraclinis articulate* (Vahl) link.) and from an altitude of about 450m it is associated with holm oak (*Quercus rotundifolia* Lam.). On higher altitudes cork oak (*Quecus suber* L.) is present as relict species on acidic substratum. Aleppo pine (*Pinus halepensis* Mill.) occupies naturally the shallowest soils. *Juniperus phoenicea* L. is observed mainly on littoral dunes.

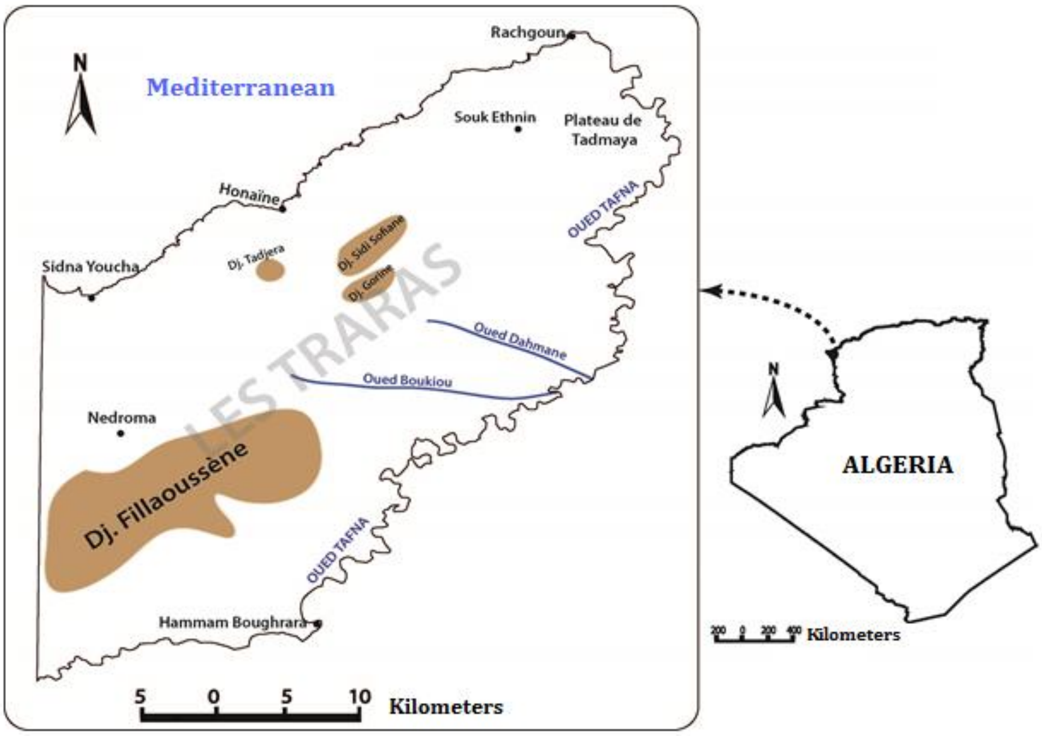

Figure 1. Geographic location of the study area

Soil map was constructed by authors following the most classical method based on JENNY's model (Jenny, 1949). Three main covariables, namely lithology, landform and vegetation cover have been used in order to build chorological laws allowing inference by GIS queries. Originally, the French soil referential (Baize and Girard, 2008) was utilized to define the 13 major soil types of the map. The corresponding soil reference group following FAO-WRB (2006) was roughly defined (Figure 2, Table 1).

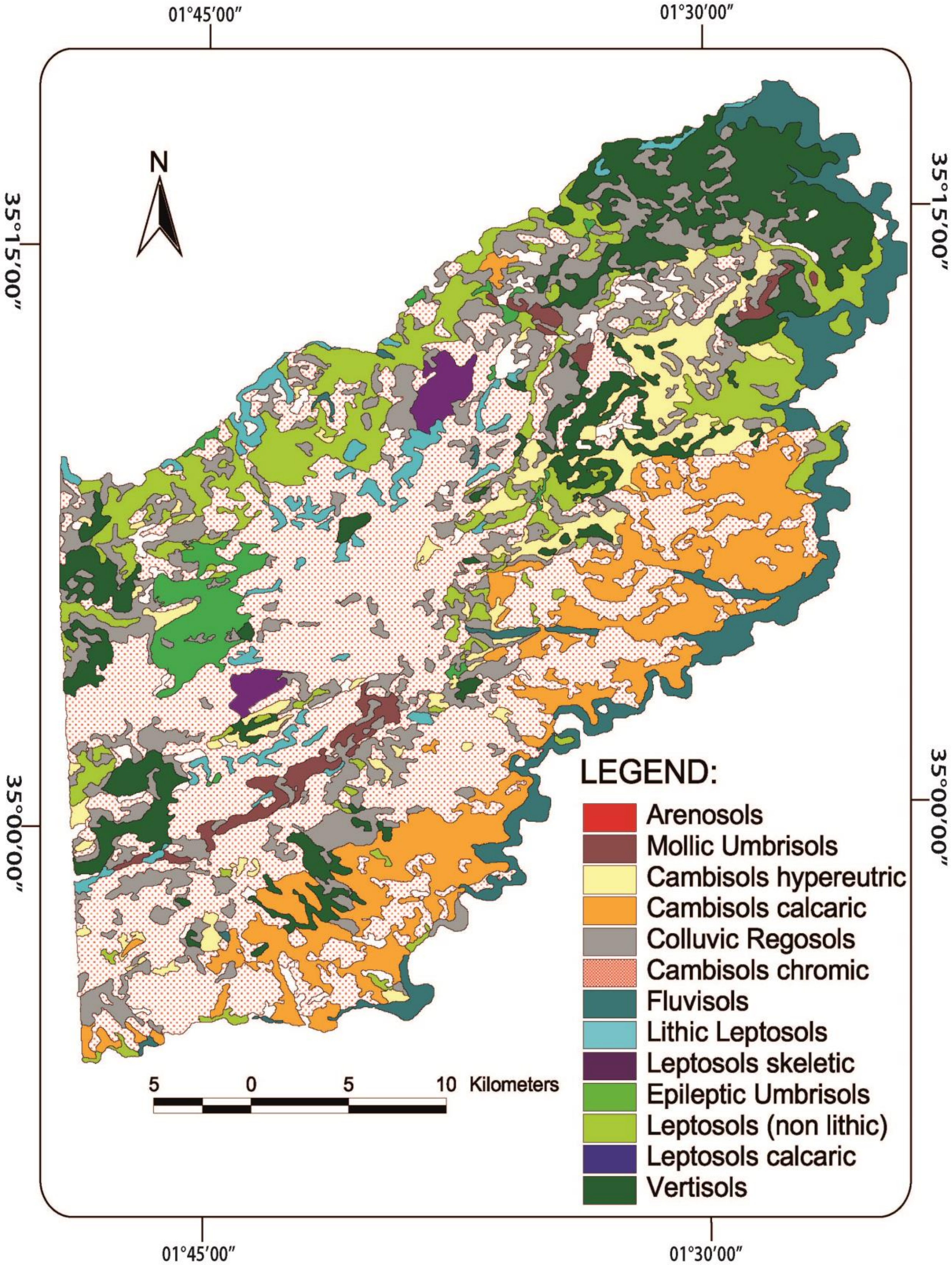

Figure 2. Soil map

Table 1. Major soil units' abbreviation and their correspondence between RP (2008) and FAO-WRB (2006)

Major soil unit following R.P. (2008)	Major soil unit following FAO-WRB (2006)	Abbreviation
Arénosols	Arenosols	ARN
Lithosols	Lithics leptosols	LLP
Régosols	Leptosols non lithics	LNL
Peyrosols	Letosols skeletics	LSK
Rendisols	Leptosols calcarics	LCA
Calcosols	Cambisols calcarics	CCA
Calcisols	Cambisols hypereutrics	CHY
Rankosols	Epileptics umbrisols	EUM
Fersialsols	Cambisols chromics	CCH
Fluviosols	Fluvisols	FLV
Colluviosols	Colluvics regosol	CLV
Brunisols saturés	Mollics umbrisols	MUM
Vertisols	Vertisols	VRS

Method

The EMC has long history and is rooted in the work of Gingerich (1969), Selley (1969), Read (1969), Türk (1979, 1982), Walker (1979), Harper (1984), Powers and Easterling (1982) and many others. Numerous techniques exist to apply EMC; all are based on the same principle, and can be divided into five steps (Xu and Maccarthy, 1998; Labat, 2004):

1. Tabulate the number of instances of the transition (transition frequency matrix) from which transition probability matrix(P_{ij}) is computed by dividing each value by the appropriate row total;
2. Compute independent trials (probability) matrix $e_{ij} = C_j/(T - C_i)$ where $e_{ij}$is the probability of passing from facies i to facies *j*, C_i and $C_j$are the number of occurrences facies *i* and facies *j*, *T* is the total number of transitions of the whole matrix (P_{ij}).
3. Deducing difference matrix $(D_{ij} = P_{ij} - e_{ij})$;
4. Test for randomness and significance;
5. Construct a relationship diagram from the positive values of the difference matrix or the resulting values of any significance test used.

Gingerich (1969) and Selley (1969) methods are considered as the earliest in the EMC field. However they did not solve correctly the problem of randomness tests as well as the determination of a given difference (for more details readers can refer to the review article of Xu and Maccarthy (1998) and the comment of this article made by Türk (2002). Some modifications have been advocated by Türk (1979, 1982), Powers and Easterling (1982), Harper (1984a,b); improvements were made in order to detect outliers in the original transition matrix and to test randomness of the sequence. All these methods were developed by geologists for lithostratigraphic column data where direction is taken into account (top to bottom or bottom to top).

The main differences that exist between geological and current data are: the first one are in one dimension (stratigraphic column) and the second in two dimensions (soil map). In addition, data have been introduced regardless to the direction, hence transition from soil A to soil B is the same from B to A and the consequence is a symmetric matrix. Furthermore, transition of one soil with itself is not allowed, so the diagonal of the transition matrix is equal to zero, which is a prerequisite of the EMC definition.

The main impediment for applying EMC methods is the tediousness of the calculation, increasing the risk of error. Thus, in this case study, Strati-signal software developed by Ndiaye (Ndiaye, 2007; Ndiaye et al. 2014) has been used to relieve calculation tasks. Strati-signal allows three main EMC methods: Walker, Harper and Türk. The simplest method is Walker which follows all steps above-mentioned without any test of significance of the D_{ij} matrix positive values. Harper method is similar to the previous one but uses binomial probability to rule out the null hypothesis of randomness of each transition link between soils. Türk method uses quite different approach, by introducing the idea of iterative proportional fitting (IPF) to calculate a valid randomized matrix to compare data, and to test it using quasi-independence concept. Hence, the transition matrix is split into noise and signal matrix (expectancy matrix) and to check the residuals with chi-square significance method (Türk, 1982).

These three methods have been used in the current case study in order to perform some comparisons, and to check out the fitness of EMC soil map spatial analysis. The results are presented on tables and soil relationship diagram (SRD) by analogy to facies relationship diagram (FRD) of geologists.

From the soil map (Figure 2) adjacency matrix was tabulated (Table 2) each value represent the number of times where soil A is neighboring (adjacent) with soil B. For practical purposes, each major soil type was designated with an abbreviation as presented in the Table 1.

Table 2. Adjacency matrix stemmed from the above soil mapR_i and C_i total rows and columns respectively.

	ARN	LLP	LNL	LSK	LCA	CCA	CHY	EUM	CCH	FLV	CLV	MUM	VRS	R_i
ARN	0	1	0	0	0	0	0	0	0	1	0	0	0	2
LLP	1	0	13	3	8	0	1	1	43	1	14	5	10	100
LNL	0	13	0	12	18	12	41	2	79	10	125	6	27	345
LSK	0	3	12	0	6	1	5	0	15	0	21	4	1	68
LCA	0	8	18	6	0	14	16	0	46	2	44	4	17	175
CCA	0	0	12	1	14	0	11	0	69	22	32	0	10	171
CHY	0	1	41	5	16	11	0	1	55	4	85	0	33	252
EUM	0	1	2	0	0	0	1	0	3	0	5	0	0	12
CCH	0	43	79	15	46	69	55	3	0	25	184	14	42	575
FLV	1	1	10	0	2	22	4	0	25	0	8	0	2	75
CLV	0	14	125	21	44	32	85	5	184	8	0	25	78	621
MUM	0	5	6	4	4	0	0	0	14	0	25	0	4	62
VRS	0	10	27	1	17	10	33	0	42	2	78	4	0	224
C_j	2	100	345	68	175	171	252	12	575	75	621	62	224	2682

The adjacency matrix is a large database (13 soil types and 2682 soil transitions counted) is also called frequency matrix, was introduced into the Strati-signal software to perform the three above mentioned EMC methods by keeping software computation defaults.

Results and Discussion

A χ^2 test has been used in order to check for the Markov property, in other words to rule-out the null hypothesis of randomness. The results are:

Computed χ^2=512.3

Number of degree of freedom=131

Critical χ^2 at 95% confidence level=124.3

The null hypothesis is rejected and data should not be random, exhibiting a strong first-order Markovian property.

From Table 3 representing Walker matrix, the positive values are ranged between 0.472 and 0.001 and represent about 58 links between soils, which is quite high to represent on SRD and difficult to unravel and interpret. That could be the weakness of this method especially if dealing with relatively big matrix. Hence, the highest values were selected, totalizing about 80% of the cumulative values, representing 16 links between soils (differences≥0.044, from Walker matrix). By assuming that the remaining 20% shared by 42 links (differences<0.044) are more likely to bring noise into the SRD impeding a clear interpretation.

Table 3. Walker matrix. Positive differences are in bold and are used to build the SRD (Soil Relationship Diagram).

	ARN	LLP	LNL	LSK	LCA	CCA	CHY	EUM	CCH	FLV	CLV	MUM	VRS
ARN	0.000	0.463	-0.129	-0.025	-0.065	-0.064	-0.094	-0.004	-0.215	0.472	-0.232	-0.023	-0.084
LLP	0.009	0.000	-0.004	0.004	0.012	-0.066	-0.088	0.005	0.207	-0.019	-0.101	0.026	0.013
LNL	-0.001	-0.005	0.000	0.006	-0.023	-0.038	0.011	0.001	-0.017	-0.003	0.097	-0.009	-0.018
LSK	-0.001	0.006	0.044	0.000	0.021	-0.051	-0.023	-0.005	0.001	-0.029	0.071	0.035	-0.071
LCA	-0.001	0.006	-0.035	0.007	0.000	0.012	-0.009	-0.005	0.033	-0.018	0.004	-0.002	0.008
CCA	-0.001	-0.040	-0.067	-0.021	0.012	0.000	-0.036	-0.005	0.175	0.099	-0.060	-0.025	-0.031
CHY	-0.001	-0.037	0.021	-0.008	-0.009	-0.027	0.000	-0.001	-0.018	-0.015	0.082	-0.026	0.039
EUM	-0.001	0.046	0.037	-0.025	-0.066	-0.064	-0.011	0.000	0.035	-0.028	0.184	-0.023	-0.084
CCH	-0.001	0.027	-0.026	-0.006	-0.003	0.039	-0.024	0.000	0.000	0.008	0.025	-0.005	-0.033
FLV	0.013	-0.025	0.001	-0.026	-0.040	0.228	-0.043	-0.005	0.113	0.000	-0.132	-0.024	-0.059

The Harper method results are presented in the Table 4 where binomial probability values are all <0.005 which means that they are highly significant with >99.5% confidence level (Xu and Maccarthy, 1998; Sarmah, 2013). Number of links is dropped from 58 links in the previous method to only 10 links.

Table 4. Binomial probability values following HARPER method.

Transitions	Binomial probability
FLV→CCA	1,21E-09
CCA→FLV	9,10E-09
CCA→CCH	2,84E-07
LLP→CCH	2,33E-06
LNL→CLV	5,61E-05
CCH→CCA	8,43E-04
VRS→CLV	9,61E-04
CCH→LLP	0,00224241
CHY→CLV	0,00245536
MUM→CLV	0,00275467

Table 5. Significance difference values according to TÜRK method.

Transitions	Significant difference values
FLV→ARN	0,952
ARN→FLV	0,952
LLP→ARN	0,935
ARN→LLP	0,935
FLV→CCA	0,802
CCA→FLV	0,802
CCA→CCH	0,573
CCH→CCA	0,573
CCA→LCA	0,470
LCA→CCA	0,470
LLP→CCH	0,436
CCH→LLP	0,436
FLV→CCH	0,402
CCH→FLV	0,402
VRS→CHY	0,367
CHY→CCH	0,367

Türk method shows more links between soil units 16 (Table 5) instead of 10 by Harper method. The resulted SRD of the three methods is represented Figure 3.

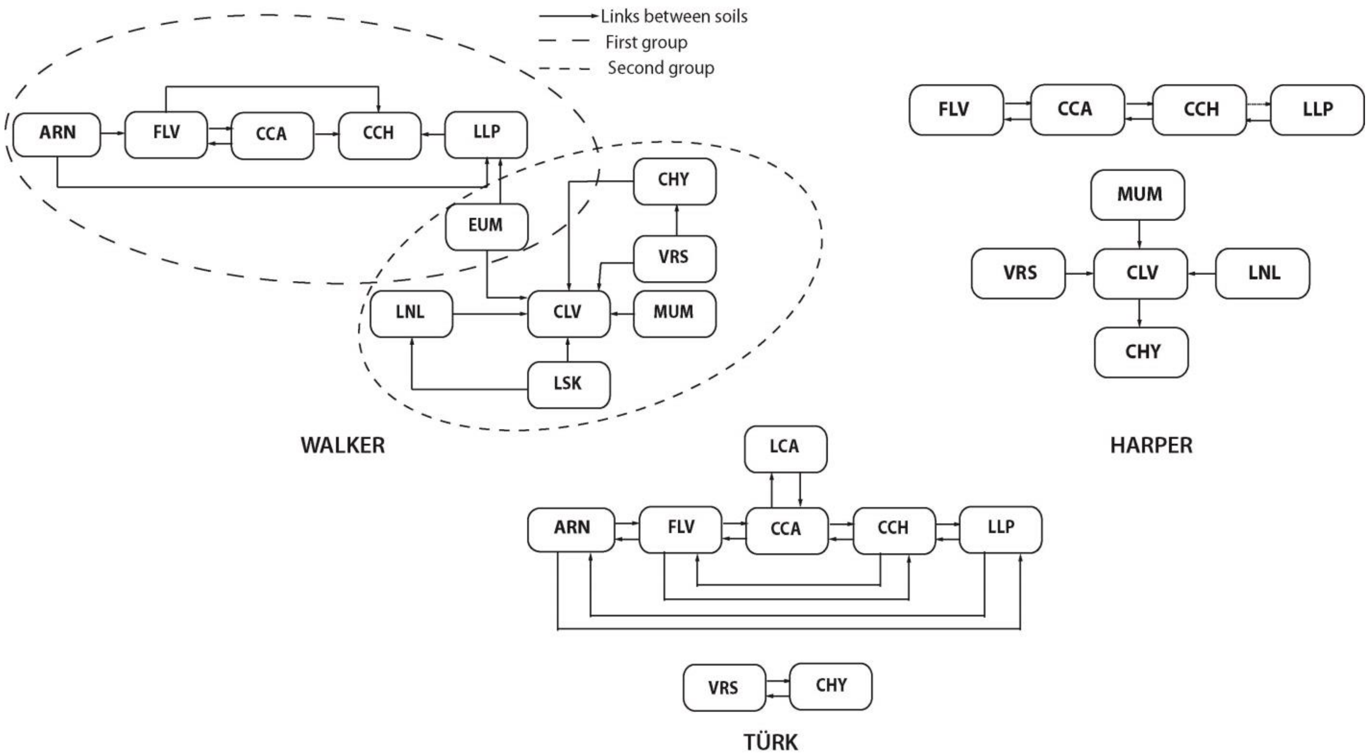

Figure 3. Soil relationship diagram (SRD) following Walker, Harper and Türk methods

The SRD drawn from Walker matrix shows two distinctive groups (Figure 3). The first one is structured around the cambisols chromics (CCH) which are considered as the climax soils in the Mediterranean region. FLV is also linked to CCH because ancient terraces are able to evolve to CCH by fersiallitization, which is a long-term pedogenic phenomenon. So FLV-CCH is more likely to be defined as a topo-chronosequence and not to complete the cyclic sequence FLV-CCA-CCH-LLP-FLV. A pedological cyclic sequence must obligatory pass through LLP or LNL which are both considered as rejuvenating step.

The second group is structured around colluvic regosols (CLV) where it plays the role of a receptacle of eroded material coming from the other soils. However, Walker SRD raised one problem concerning ARN which is represented by only one polygon on the map, but it presents the largest differences with FLV and LLP because it is linked only with these two soil types. Hence, it is considered as an outlier since it affects negatively the other links. Similar problem has been pointed out by Mastej (2002) presuming that they must have at least 5 transitions for credible results although single case with lower number of transitions is acceptable. The same observation can be made for the Türk SRD concerning ARN. Conversely, in Harper SRD the ARN is not represented at all, and the first soil sequence is reduced to only four types of soil (FLV, CCA, CCH, LLP), which can be found in the other methods. CLV is the main soil structuring the second sequence by both Walker and Harper methods but it doesn't exist in Türk method. According to this method, links are not enough significant to be represented and are considered as noise and hence, eliminated. However, it further stresses the link between VRS and CHY which is substantiated by Walker SRD. This link is confirmed in the field; cambisols hypereutrics (CHY) occur essentially on marls and can be evolved, by transition with cambisols vertics, towards vertisols (VRS) wherever topography permits in bowl-like depressions.

So if the ARN have brought confusion, it should be overlooked and rebuild a new matrix with 12 soil types instead of 13, and all the procedure is then restarted.

Computed χ^2=446.5

Number of degree of freedom=109

Critical χ^2 at 95% confidence level=124.3

The null hypothesis is rejected and data should not be random, exhibiting a strong first-order Markovian property.

In the Walker method the 80% of the cumulative values correspond to 19 out of 54 links instead of 16 out of 58. The results are shown in Figure 4.

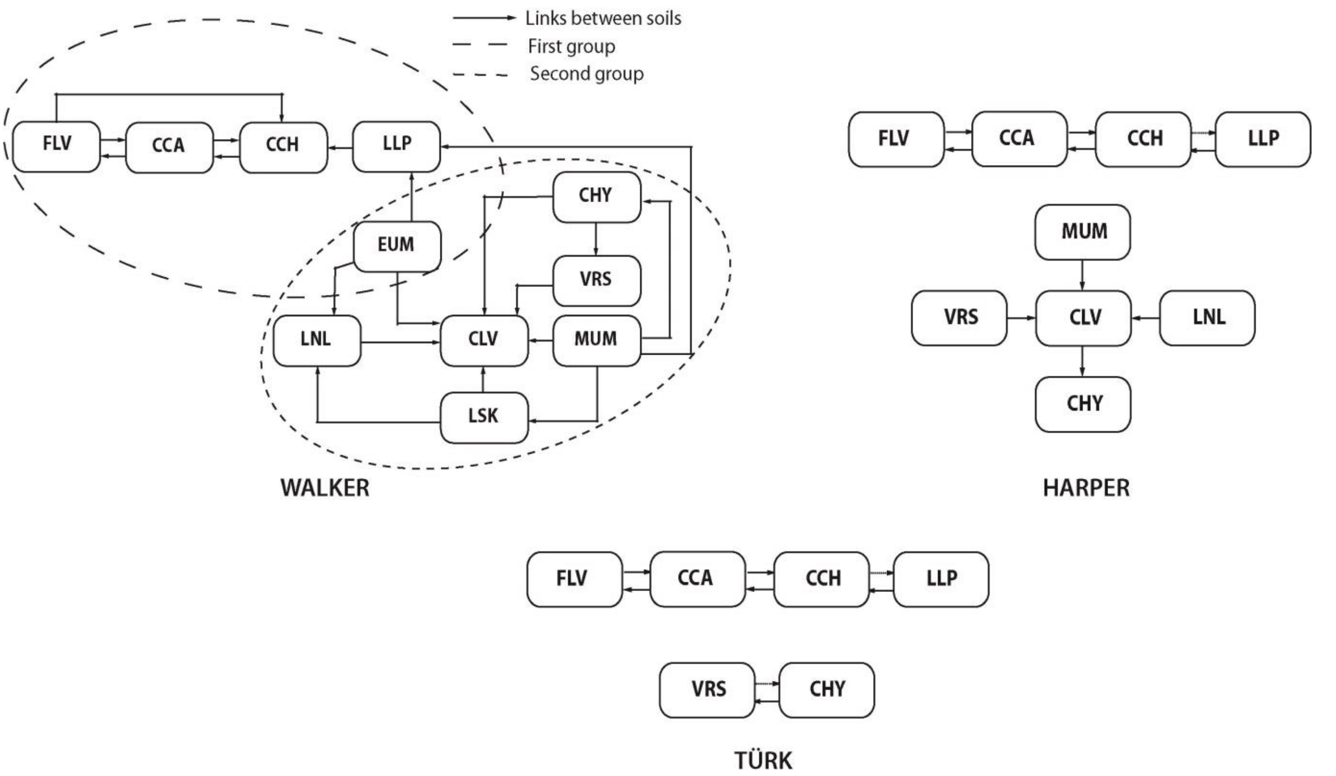

Figure 4. Soil relationship diagram (SRD) following WALKER, HARPER and TÜRK methods without ARN

With or without ARN, Harper SRD doesn't change at all. So this method is not affected by low number of transitions. However, in Türk SRD by losing 4 links and one soil type (LCA) it supports Harper results concerning the first sequence. Walker method seems to be the most realistic but the weakness of the method lies on the significance of the relationships between soil types. The ceiling 80% was defined arbitrary; conversely, by introducing significance test information might be lost that can be reliable. However, Harper SRD is the most relevant to the current data, because it shows two separate sequence of soils as it is mentioned before: the first FLV-CCA-CCH-LLP (Figure 5) begins with lithics leptosols occurring on the steepest slopes maintaining them by rejuvenation. On lower altitudes and less steep slopes, truncated cambisols chormics are widely spread, they exist on almost all kind of substratum bearing forested maquis; they are followed by cambisols calcarics on less vegetated areas because of the human impact (low altitudes and gentle slopes). In pedogenic point of view the CCA are developed from eroded material coming from the CCH. Finally, CCA are naturally connected with fluvisols bordering wadi Tafna the main stream of the region, with its tributaries wadi Boukiou and wadi Dahmane.

The second sequence is associated to the steep slopes relief with the colluvics regosols (CLV) occurring on the foothills. The relation of these two sequences can be seen in the Walker SRD, where LLP plays a pivotal role between the two sequences.

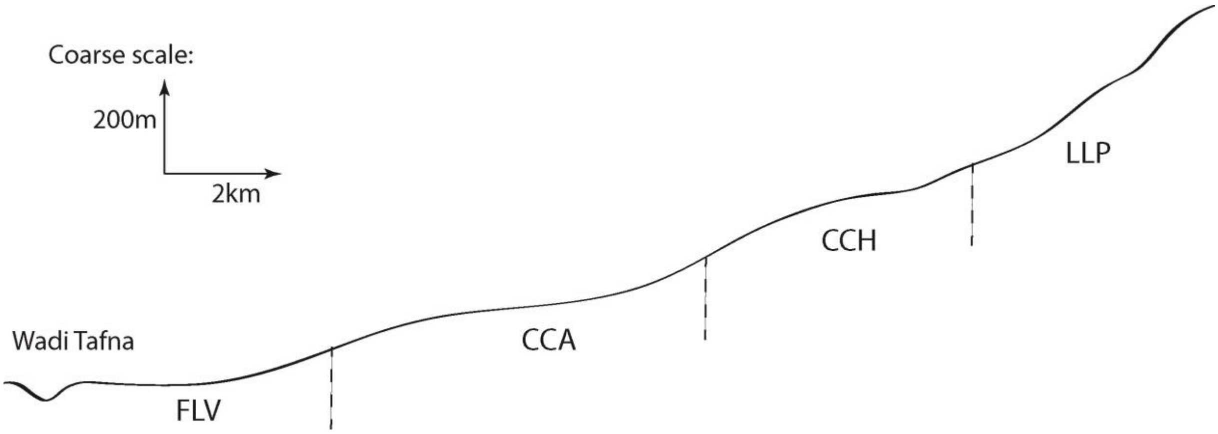

Figure 5. Theoretical catena of the Traras region with EMC method (the 1st sequence)

Conclusion

The present study reveals that it is worthwhile to apply embedded Markov chain analysis for soil map spatial analysis. However, the questioning arises from this study is, if it is mathematically fair to apply an EMC to data from a 2 dimension map with a symmetric matrix. By letting this question to statisticians, the results are conclusive and the catena is still remains a powerful means to understand soils. Theoretical catena highlighted by EMC method is the most representative of the study area and even for the whole Mediterranean region.

Acknowledgments

Authors would like to express all their thanks to Pr. Nacereddine KAZI-TANI (Société Géoressources, Pau, France) who called our attention to this method.

References

Baize, D., Girard, M. C., (coord.), 2008.Référentiel Pédologique. Editions : Quæ. 405p.

Birkeland,P. W., 1999.Soils and Geomorphology, 3rd edition. New York: Oxford University Press. New York, USA 430p.

Burgess, T. M., Webster, R., 1984a. Optimal sampling strategiesfor mapping soil types: I. Distribution of boundary spacing. *European Journal of Soil Science* 35(4): 655-665.

Burgess, T. M, Webster, R., 1984b. Optimal sampling strategiesfor mapping soil types: II. Risk functions and sampling intervals. *European journal Soil Science* 35(4): 641-654.

C.P.C.S., (Commission de Pédologie et de la Cartographie des Sols). 1967. Classification des sols, INRA, Grignon. 96p.

FAO, 2006. World reference base of soil resources. A framework for international classification, correlation and communication. Food and Agriculture Organization of The United Nations World Soil Resources Reports No. 103, Rome, Italy.

Gingerich, P. D., 1969.Markov analysis of cyclic alluvial sediments. *Journal of Sedimentary Research,* 39(1),330-332.

Harper, C. W. Jr.., 1984a. . Improved methods of facies sequence analysis. In : Facies models: Response to sea level change. Walker, R.G. (Ed.). Geological Association of Canada, Reprint Series No.1, 2nd Edition. pp.11-13.

Harper, C. W. Jr., 1984b. Facies models revisited: An examination of quantitative methods. *Geoscience Canada: Journal of The Geological Association of Canada* 11(4): 203-207.

Jenny, H., 1941. Factors of soil formation. A system of quantitative pedology. MacGraw-Hill, New York, 281 p.

Labat, L., 2004. Simulations stochastiques de faciès par la méthode des membership functions. PhD dissertation Institut National Polytechnique de Lorraine. p.105.

Li, W., Li, B., Shi, Y., 1999. Markov-chain simulation of textural profiles. *Geoderma* 92(1-2): 37-53.

Li, W., Zhang, C., Burt, J. E., Zhu, A. X., Feyen, J., 2004. Two-dimensional Markov chain simulation of soil type spatial distribution. *Soil Science Society of America Journal* 68(5): 1479-1490.

Li, W., Zhang, C., 2006. A generalized markov chain approach for conditional simulation of categorical variables from grid samples. *Transactions in GIS* 10(4): 651–669.

Mastej, W., 2002. An application of Markov chain analysis to studies on lithofacies sequences in the alluvial fans from the „Bełchatów" lignite deposit (Poland). *Annales Societatis Geologorum Poloniae* 72(3): 271-282.

Mathieu, C., 2009. Les principaux sols du monde. Voyage à travers l'épiderme vivant de la planète Terre. Tec & Doc Editions. Paris, 221p.

Ndiaye, M., Davaud, E., Jorry, S.J., 2014. A computer-assisted method for depositional model determination. *International Journal of Geosciences* 5(2): 178-183.

Ndiaye, M., 2007. A multipurpose software for stratigraphic signal analysis. Doc. ès Sciences dissertation, Faculté des sciences, Université de Genève. p. 118.

Powers, D. W., Easterling, R.G., 1982. Improved methodology for using embedded Markov chains to describe cyclical sediments. *Journal of Sedimentary Petrology* 52(3): 913-923.

Read, W. A., 1969. Analysis and simulation of namurian sediments in central Scotland using a Markov-process model. *Journal of the International Association for Mathematical Geology* 1(2): 199-219.

Sarmah, R.K., 2013. Study of cyclic pattern and lithofacies variability by application of Markov chains and entropy analysis to Tikak Parbat formation, Makum coal basin, North Eastern India. *European International Journal of Science and Technology* 2(6): 41-55.

Selley, R. C., 1969. Studies of sequence in sediments using a simple mathematical device. *Quarterly Journal of The Geological Society* 125: 557-581.

Sun, X.L., Wu, S.C., Wang, H.L., Zhao, Y.G., Zhang, G.L., Man, Y.B., Wong, M.H., 2013. Dealing with spatial outliers and mapping uncertainty for evaluating the effects of urbanization on soil: A case study of soil pH and particle fractions in Hong Kong. *Geoderma* 195-196: 220-233.

Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46: 234-240.

Türk, G., 1979. Transition analysis of structural sequences: Discussion and reply. *Geological Society of American Bulletin* 90(10): 989–991.

Türk, G., 1982. Letter to the editor: Markov chain analysis. *Journal of the International Association for Mathematical Geology* 14 (5): 539-542.

Türk, G., 2002. Comment on "Markov chain analysis of vertical facies sequences using a computer software package (SAVFS): Courtmacsherry formation", by H. Xu and I.A.J. MacCarthy. *Computer & Geosciences* 28(3): 427-429.

Walker, R. G., 1979. Facies and Facies Models: General Introduction. In : Facies models: Response to sea level change. Walker, R.G. (Ed.). Geological Association of Canada, Reprint Series No.1, 2nd Edition. pp.1-7.

Xu, H., Maccarthy, I.A.J., 1998. Markov chain analysis of vertical facies sequences using a computer software package (SAVFS): Courtmacsherry formation (Tournaisian), southern Ireland. *Computer and Geosciences* 24(2): 131-139.