Geosciences and Engineering, Vol. 6, No. 9 (2018), pp. 127–141.

EFFECTS OF EARTHQUAKE DATA CLUSTERING ON THE RESULTS OF STRESS INVERSIONS

LILI CZIROK[1*]–LUKÁCS KUSLITS[2]

[1]*University of Sopron, Roth Gyula Doctoral School, Sopron*
[2]*MTA CSFK Geodetic and Geophysical Institute, Sopron*
[*]*cziroklili@ggki.hu*

Abstract: An important criterion of stress inversions is the homogeneous stress field of the studied area. Therefore, it is necessary to take into account the variability of used focal mechanism solutions (FMS) and create some subareas for estimations. If there are a large number of earthquakes, it is worth generating the clusters automatically.

This research has two aims:

Firstly, effects of the existing clusterization methods are analysed by the authors. Clusters of earthquake data are created based on three methods: (a) manual grouping based on the FMS map, (b) k-means algorithm and (c) an automatized, nearest neighbour-based method. The input data for clustering were in each case the epicentral coordinates of the investigated earthquakes.

Secondly, the authors have begun developing a new, automatized clustering method that does not require a minimal number of cluster members or any other manually given parameter but whose reliability is similar to that of the manual method.

The applied stress inversion is a linear, iterative method that uses the Mohr-Coulomb law to analyse the fault instability of input data (using the code STRESSINVERSE). Focal mechanisms used for calculation are situated in the Vrancea Zone (SE Carpathians).

Keywords: *focal mechanism solutions, stress inversions, clustering algorithms, SE-Carpathians*

INTRODUCTION

Tectonics

The South-eastern Carpathians are an interesting part in the Eurasian plate because of their geodynamic activity characterized by an ongoing subduction processes. This has two main indications in the Carpathian Bends: the Ciomadul Volcano and the Vrancea Zone.

In the Inner Carpathians, there are post volcanic activities produced by the Ciomadul Volcano. From the Late Jurassic to the Early Cretaceous, a volcanic mountain range evolved due to the subduction of the Neo-Tethys. The Ciomadul Volcano is the youngest and the southernmost member of this chain. It erupted about 30,000 years ago, that being the last volcanic eruption in the Carpathian Basin [1].

The Vrancea Region is a complex seismic region which has been active since the closure of Ceahlau-Severin Ocean between the late Early and Late Cretaceous [2]. Nowadays, this subduction event is still in progress. Moreover, three tectonic units converge here: the East European Plate and the Intra-Alpine and Moesian subplates, evoking major seismic and geodynamic activity. As for intermediate-depth earthquakes' focal mechanism solutions, reverse faults are the dominant tectonic regime but stress orientations are variable. Normal faults and strike-slips also occur in the Vrancea Zone due to the complicated tectonic situation. *Figure 1* presents the tectonic situation in Romania [2].

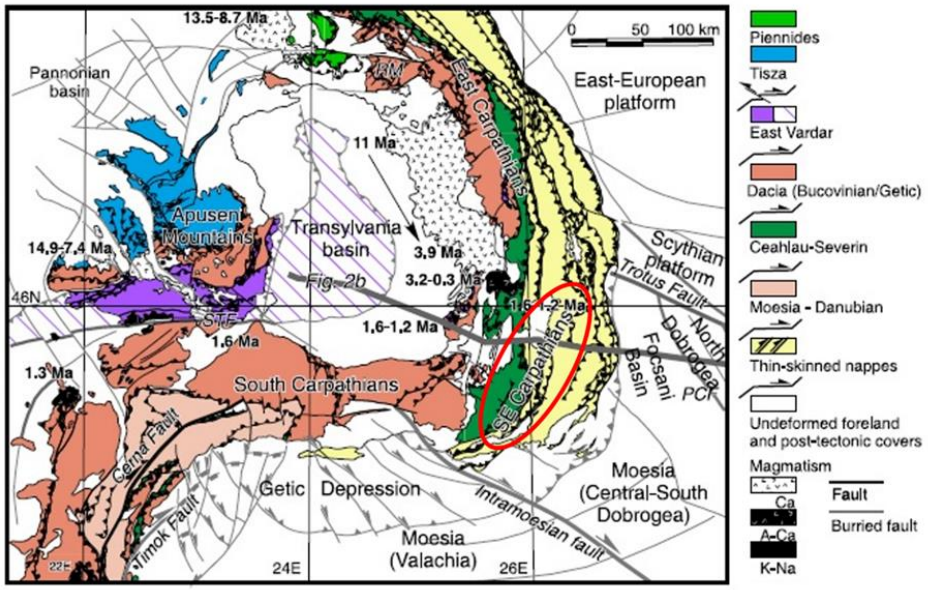

Figure 1. *The tectonic units in Romania.*
The red circle outlines the studied area (SE Carpathians) [2]

Previous studies

Martínez-Garzón et al. [3] carried out stress inversions for the region of San Jacinto Fault Zone. On the 4th April 2010, a M_w 7.2 earthquake occurred in the area followed by two larger seismic events (having local magnitudes M_L 4.3 and 5.2). In the relocated seismicity catalogue, there were roughly 5,400 events until the end of 2010, about 3,400 of which have focal mechanism solutions. In this publication [3], the authors developed a refined methodology for stress inversion. They classified seismic events using a k-means algorithm and then applied the mixture of two stress inversion methods: Michael's [4] and Vavrycuk's method [5].

Hardebeck and Michael [4] presented a damped, linear inversion method that creates groups before the carrying stress inversion. The code MSATSI (Martinez-Garcon et al. 2014) [6], has an option for setting the number of minimum events falling into a grid point. After the clustering of input data, stress inversion routines assign each

clustered set of earthquakes to one geographic point for the computations, called a grid point or node. Based on this setting, the program creates the groups for the formal stress inversion and executes the estimations. Nevertheless, the reliability of the stress tensor is lower than that computed by STRESSINVERSE, as this algorithm cannot distinguish between the correct and auxiliary fault planes. For the code STRESS-INVERSE, FMS data need to be classified before using them as input data [5].

DATA

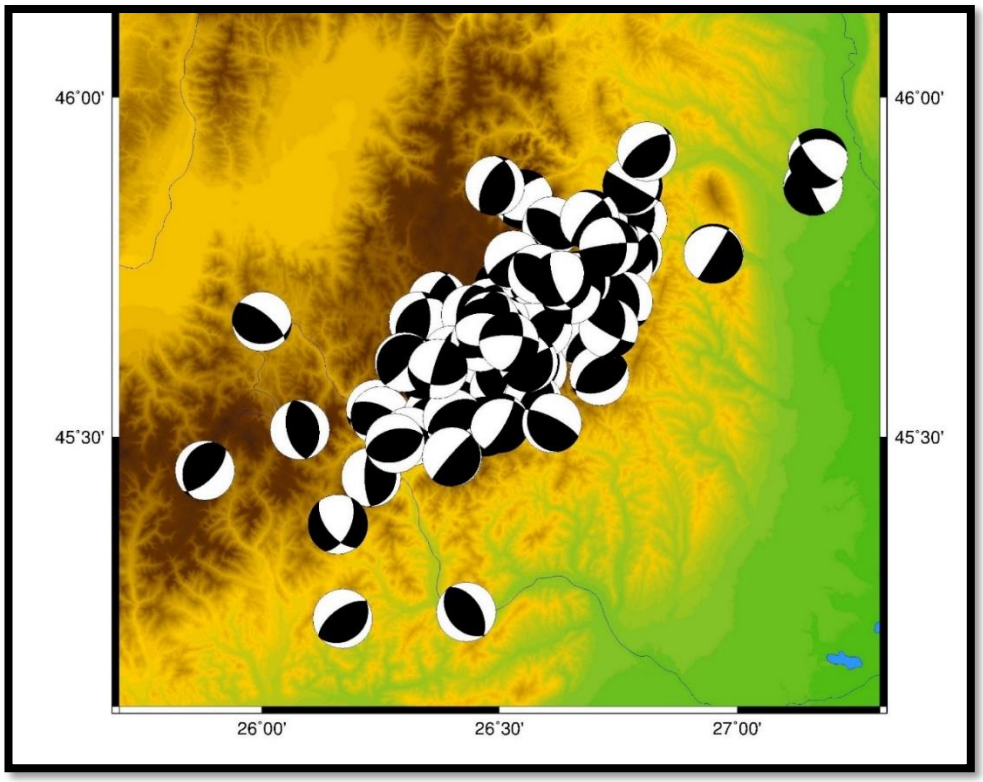

Figure 2. *Selected focal mechanism solutions in the Vrancea-zone used for the clusterization. The map is created using Generic Mapping Tools*

For these estimations, there were altogether 84 focal mechanism solutions, available from 1978 to the very recent days (the last solution was determined on March 14, 2018). The most important data source is the Romanian National Institute for Earth Physics (Institutul National de Cercetare-Dezvoltare pentru Fizica Pamantului). Seismic moment tensors were determined using two codes: FMNEAR [7] and ISOLA-GUI [8]. FMNEAR applies a fully-automated waveform inversion on seismic records of near-earthquake stations. Moment tensors are computed in near real-time. The ISOLA-GUI is also a waveform-inversion method but it is not automatic.

Earlier focal mechanism solutions (until the 1990s) originate from the websites of the National Institute of Geophysics and Volcanology [9] and Incorporated Research Institutions for Seismology [10].

Figure 2 is a map of used focal mechanism solutions in the Vrancea Region. The map was created using Generic Mapping Tools.

CLUSTERING METHODS

It is visible in *Figure 2* that most of the earthquakes are concentrated in a relatively small region (having an area of about 220 × 150 km^2). Because of that, other clustering methods (a k-means algorithm and an automatic, nearest neighbour-based algorithm) were applied to compare final results obtained using them with results using manual classification of the FMS map.

Manual clustering

Subareas for the stress inversion were generated based on the' latitude and longitude coordinates of epicentres. Moreover, the homogeneity of the stress field was also an important criterion. The FMS map *(Figure 2)* was used for classification. A total number of 8 clusters (shown in *Figure 3*) were finally created as inputs for each individual estimation. In *Figure 3*, two other thematic maps are also represented: a tectonic map of the study area [11] and a digital elevation model produced by the Shuttle Radar Topography Mission [12].

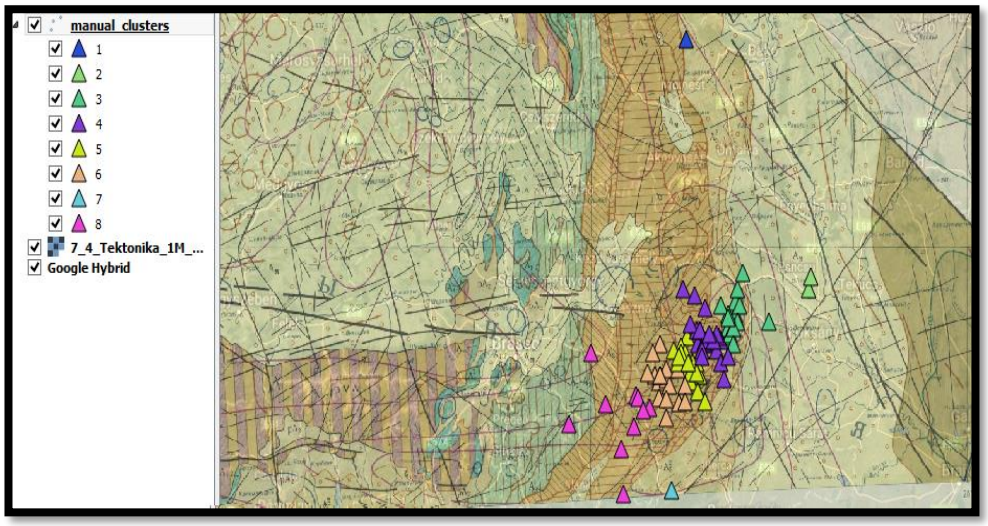

Figure 3. *Classes created manually (in a rather subjective way) based on the map of focal mechanisms (shown in* Figure 2*)*

k-means clustering

Here, k-means clusters for the inversion were created using the STATISTICA programme [13] [14] [15]. In this case, altogether 8 clusters were generated based on the seismic events' coordinates. The result of this algorithm is visible in *Figure 4*.

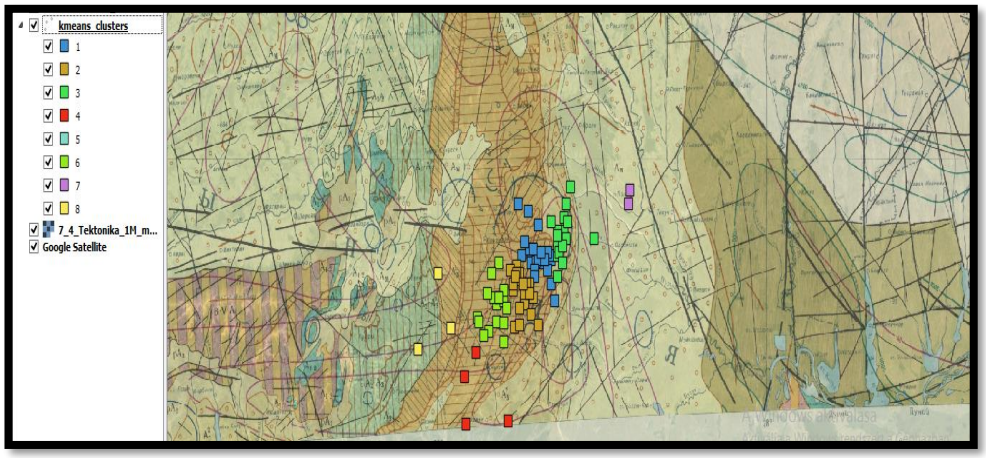

Figure 4. *Classes created using k-means clustering*

Automatized method

The automatic algorithm presented here is based on the principle of nearest neighbours, essentially resulting in a simpler version of an agglomerative hierarchical clustering method [14]. Euclidean distances were the applied metric for linkage, which only prepares the first row of the dendrogram, and only searches for the closest two points of the whole remaining dataset when creating a new cluster. It operates using the following iterative computational steps:

1) For the clustering, the distance matrix of the investigated points is determined *(Figure 5)*.
2) Non-identical points positioned closest to each other, then form the initial cluster points *(Figure 6)*.
3) Each cluster is then linked with unclassified points positioned closest to points already part of a cluster *(Figure 6)*. This step is repeated as long as there is still a new point to be identified.
4) If clusters cannot be further enlarged, the iteration returns to step 2 using the remainder of the dataset until all points are classified *(Figure 7a)*.
5) It is possible to set a threshold distance value to isolate "lonely" points that are too far away from their nearest neighbours in the dataset *(Figure 7b)*. This is an important feature, because such points can probably behave as noise for the calculations and potentially bias the stress-field estimation [16].

The steps of this clustering method are presented in *Figures 5–7*.

	1	2	3	4	5	6	7	8	9	10
1	0	5.0990	8.2462	4.4721	7.8102	1	2	10	12.0416	4.4721
2	5.0990	0	7.6158	1.4142	4.1231	6.0828	7.0711	7.0711	8.5440	4.2426
3	8.2462	7.6158	0	6.3246	5	8.5440	8.9443	4	6.0828	4
4	4.4721	1.4142	6.3246	0	3.6056	5.3852	6.3246	6.3246	8.0623	2.8284
5	7.8102	4.1231	5	3.6056	0	8.6023	9.4340	3	4.4721	4.1231
6	1	6.0828	8.5440	5.3852	8.6023	0	1	10.6301	12.7279	5
7	2	7.0711	8.9443	6.3246	9.4340	1	0	11.3137	13.4536	5.6569
8	10	7.0711	4	6.3246	3	10.6301	11.3137	0	2.2361	5.6569
9	12.0416	8.5440	6.0828	8.0623	4.4721	12.7279	13.4536	2.2361	0	7.8102
10	4.4721	4.2426	4	2.8284	4.1231	5	5.6569	5.6569	7.8102	0
11	6.7082	5.3852	2.2361	4.1231	3.1623	7.2111	7.8102	3.6056	5.8310	2.2361
12	1.4142	6	7.6158	5.0990	8.0623	1	1.4142	9.8995	12.0416	4.2426
13	8.2462	9.8995	4	8.4853	8.5440	8.0623	8	8	10.0499	5.6569
14	6	1.4142	8.9443	2.8284	5	7	8	8	9.2195	5.6569
15	11.4018	8.2462	5.0990	7.6158	4.1231	12.0416	12.7279	1.4142	1	7.0711
16	9.4340	7	3	6.0828	3.1623	10	10.6301	1	3.1623	5
17	3.1623	4.4721	5.0990	3.1623	5.3852	3.6056	4.2426	7.0711	9.2195	1.4142
18	25.4951	21.6333	18.6011	21.4009	17.8045	26.1725	26.8701	15.5563	13.4536	21.2132
19	12.7279	8.9443	7.0711	8.6023	5	13.4536	14.2127	3.1623	1	8.6023
20	12.3693	7.2801	11.1803	8.0623	6.3246	13.3417	14.3178	7.8102	7.2111	10.0499
21	24	23.5372	16.1245	22.3607	19.9249	24.0208	24.0832	17.0880	17	20.0998
22	29.0172	28.2843	21.0238	27.1662	24.5153	29.0689	29.1548	21.5870	21.1896	25.0200
23	12.5300	7.8102	9.2195	8.0623	5.0990	13.4164	14.3178	5.3852	4.2426	9.2195

Figure 5. *Distance matrix of the points of the test dataset; red cells indicate points forming the first cluster*

Figure 6. *Schematics of the algorithms operation. Application of iterative steps 2 and 3 using the principle of nearest neighbours for creating the cluster of points highlighted with orange*

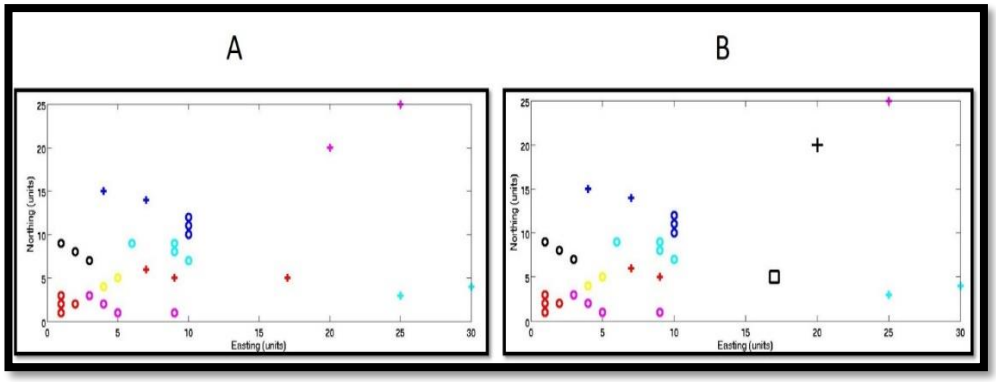

Figure 7. *Two clustering results using the same example dataset, on the left without (A) and on the right with a threshold value of 6 units for distance (B). Here this threshold value was introduced in order to separate the highlighted points (thick black cross and square, pink cross), which can be considered outliers from all the other groups*

In case of the Vrancea Region, no critical distance was applied. The input data and results can be seen in *Figure 8*. This algorithm produced a relatively large number of 28 classes for the analysis of stress relations.

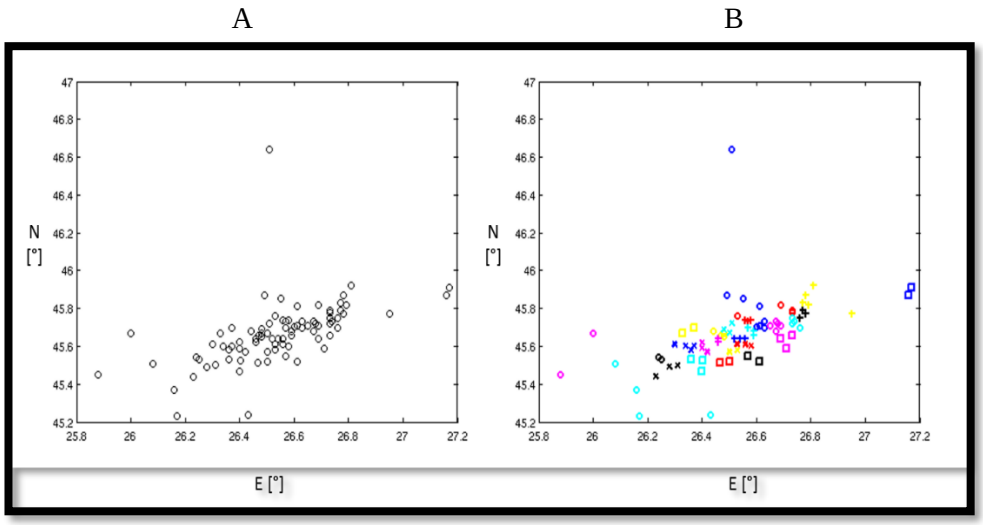

Figure 8. *The epicenters of seismic events (A) and the clusters (B) generated without critical distance in a MATLAB environment*

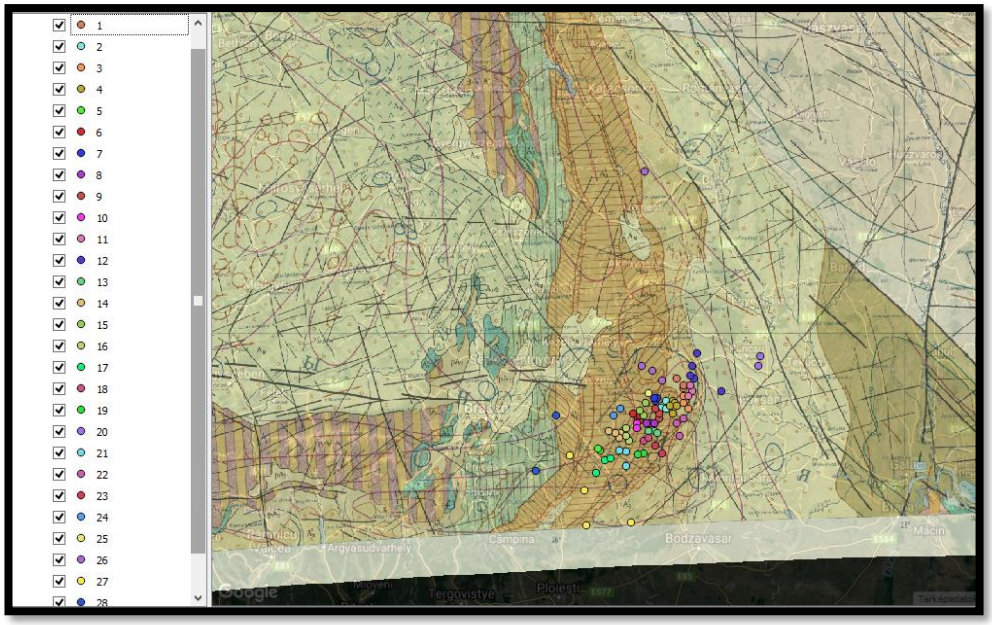

Figure 9. *Classes created using the automatized method plotted in Quantum GIS*

RESULTS OF STRESS INVERSIONS IN THE VRANCEA REGION

In the current publication, the authors applied the code STRESSINVERSE in MATLAB environment for computing the estimations.

Results using manual clustering

The azimuths of maximum horizontal stress directions (S_H) and the vertical principal (σ_V) stress axes were determined as in [17] and [18]. The best R values were calculated by the error analysis of this stress inversion method. The distributions of all R values were illustrated using histograms (shown in *Figures 10–12*). The maximum of the histograms presents the best values of shape ratios (R_{best}).

Table 1

Estimated orientations of the principal stress axes and the best R values

Number of cluster	Amount of events	Largest plunge (°) and σ_V		Azimuth (°) and S_H		R_{best}
1.	1	48.86	σ_1	150.33	$\sigma_3 + 90°$	0.5752
2.	2	55.71	σ_1	213.51	σ_2	0.6288
3.	15	76.73	σ_3	323.63	σ_1	0.5802
4.	22	78.49	σ_3	13.33	σ_1	0.1858
5.	17	83.08	σ_3	232.84	σ_1	0.2317
6.	16	83.43	σ_3	289.67	σ_1	0.5219
7.	1	80.02	σ_3	51.77	σ_1	0.7483
8.	10	81.28	σ_3	168.49	σ_1	0.4381

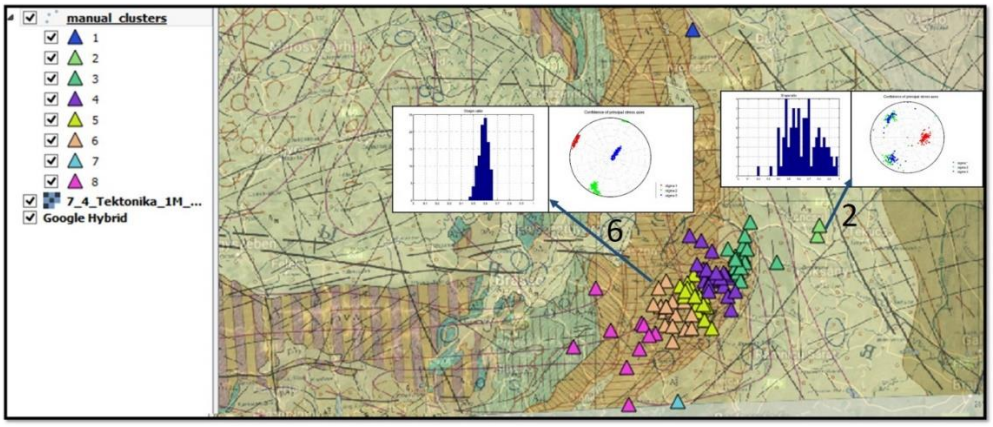

Figure 10. *R-histograms and stereograms of the 6th group and 2nd group*

In *Figure 10*, the 6th and 2nd clusters' R histograms and stereograms are shown. The 2nd cluster has the least reliable stress tensor – the spread of R values is large and the σ_2- and σ_3-axes cannot be distinguished on the stereogram. Thus, the exact stress relations cannot be evaluated by the estimated orientations. In the case of the 6th cluster, principal axes directions could be identified very clearly

Results using the k-means method

Table 2

Estimated orientations of the principal stress axes and the best R values

Number of cluster	Amount of events	Largest plunge (°) and σ_V		Azimuth (°) and S_H		R_{best}
1.	22	76.15	σ_3	358.56	σ_1	0.1744
2.	20	79.2	σ_3	256.86	σ_1	0.4983
3.	16	74.75	σ_3	313.3	σ_1	0.4572
4.	4	52.02	σ_3	18.82	σ_1	0.3253
5.	1	48.86	σ_1	150.33	$\sigma_3 + 90°$	0.5752
6.	16	86.84	σ_3	139.25	σ_1	0.5711
7.	2	48.16	σ_1	302.46	$\sigma_3 + 90°$	0.6547
8.	3	78.55	σ_3	172.29	σ_1	0.1652

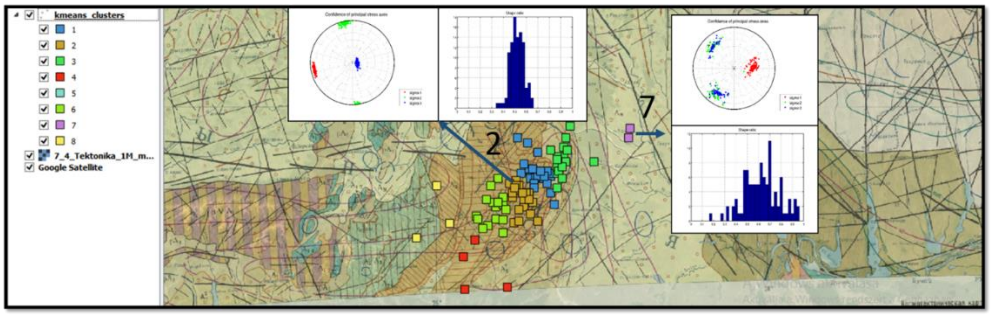

Figure 11. *R histograms and stereograms of the 7th group and 2nd group*

In the case of k-means clustering, results of the 7th group are the least reliable. This group is identical to the 2nd cluster of manual grouping (shown in *Figure 10*). As it is visible, the orientations of principal stress axes cannot be identified accurately based on their stereogram. Here, the 2nd group has the most reliable stress tensor, which also significantly overlaps the 6th group of the manual classification. In this case, the orientations of principal stress axes are similar. All the principal axes can be identified well and the best R values can be determined unambiguously.

Results using the automatized method

Table 3

Estimated orientations of the principal stress axes and the best R values

Number of cluster	Amount of events	Largest plunge (°) and σ_V		Azimuth (°) and S_H		R_{best}
1.	3	65.5	σ_3	305.79	σ_1	0.3579
2.	4	58.96	σ_3	105.53	σ_1	0.655
3.	4	66.38	σ_3	227.63	σ_1	0.1624
4.	5	72.47	σ_3	328.33	σ_1	0.195
5.	2	65.46	σ_2	341.45	$\sigma_3 + 90$	0.3977
6.	4	65.87	σ_2	88.77	σ_1	0.6479
7.	3	66.41	σ_3	83.76	σ_1	0.5996
8.	3	46.14	σ_1	167.94	σ_2	0.816
9.	3	51.4	σ_3	186.36	σ_1	0.9242
10.	2	63.14	σ_3	155.43	σ_1	0.433
11.	3	61.35	σ_2	318.16	σ_1	0.7477
12.	5	58.33	σ_3	50.19	σ_1	0.6845
13.	3	45.54	σ_2	70,36	σ_1	0.6607
14.	4	72.21	σ_2	319,54	$\sigma_3 + 90$	0.7988
15.	3	78.59	σ_3	350,12	σ_1	0.728
16.	3	73.87	σ_3	4,98	σ_1	0.5269
17.	3	83.54	σ_3	240,5	σ_1	0.3313
18.	2	64.18	σ_3	93,87	σ_1	0.6572
19.	2	43.92	σ_1	228,43	$\sigma_3 + 90$	0.3325
20.	2	48.16	σ_1	302,46	$\sigma_3 + 90$	0.6547
21.	3	73.89	σ_3	47,42	σ_1	0.1968
22.	3	50.17	σ_3	219,21	σ_1	0.335
23.	2	63.41	σ_3	70,22	σ_1	0.3363
24.	2	76.02	σ_3	281,33	σ_1	0.6185
25.	1	56.04	σ_3	325,36	σ_1	0.6008
26.	3	40.38	σ_1	206,96	$\sigma_3 + 90$	0.8026
27.	4	44.94	σ_3	16,44	σ_1	0.3358
28.	2	60.01	σ_3	169,63	σ_1	0.4579

In *Figure 12*, the results of three groups are presented: the 5th, 7th and 19th, respectively. The 7th group has similar results to the previous studies (the stereogram indicates thrust faults). The 19th group has rather special results. It seems reliable and accurate but it disagrees with the known studies. Here, the σ_1-axes has the largest plunge and the maximum horizontal direction was determined by the σ_3 axes (shown in *Table 3*). In the case of the 5th group, the calculations yielded a very poor outcome.

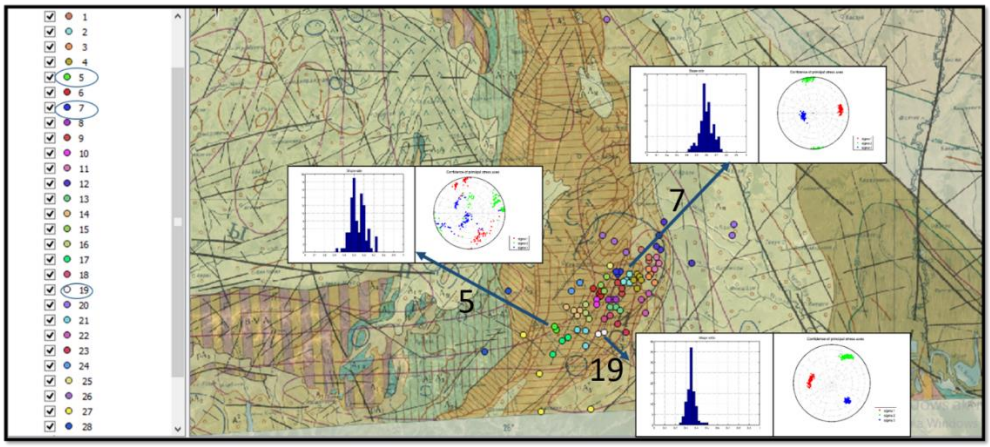

Figure 12. *R histograms and stereograms of the 5th, 7th and 19th groups*

DISCUSSION

After carrying out stress inversions, the stress relations were characterised by stress tensors. The authors used [17] and [18] for the determination of maximum horizontal stress orientations (S_H) and dominant tectonic regimes.

Figures 13–15 (the manual method, the k-means algorithm and the automatic clustering, respectively) illustrate the determined stress orientations and tectonic regimes based on the results. The maps were generated using Quantum GIS.

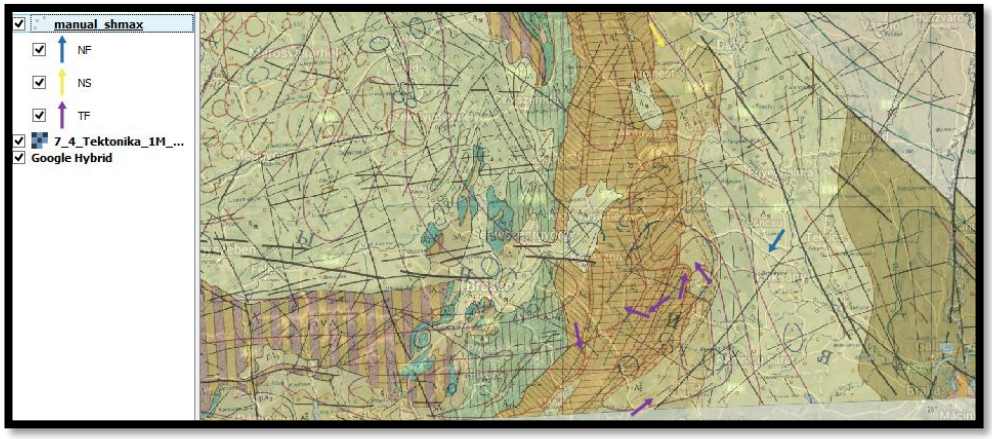

Figure 13. *The interpreted maximum horizontal stress orientations based on the azimuth of S_H and dominant tectonic regimes in the case of manual clustering. The different regimes are indicated by different colours (blue – normal faults, yellow – transtensions, purple – thrust faults)*

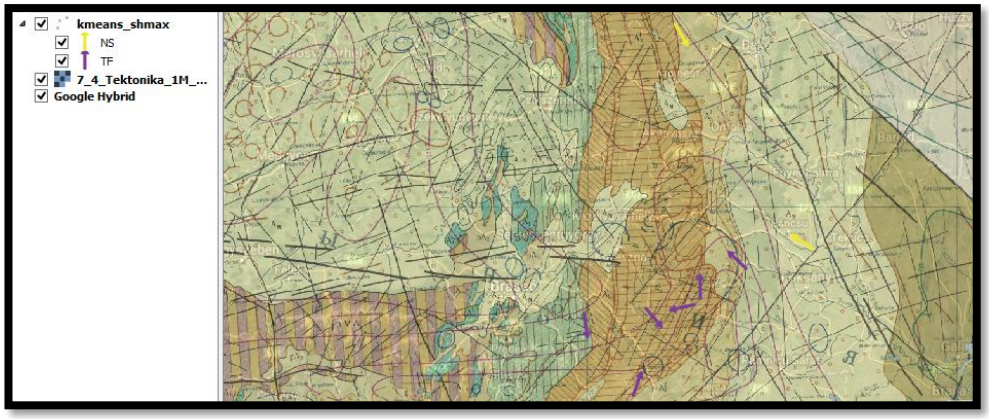

Figure 14. *The interpreted maximum horizontal stress orientations based on the azimuth of S_H and dominant tectonic regimes in the case of groups created by k-means algorithm. The different regimes are indicated by different colours (yellow – transtensions, purple – thrust faults)*

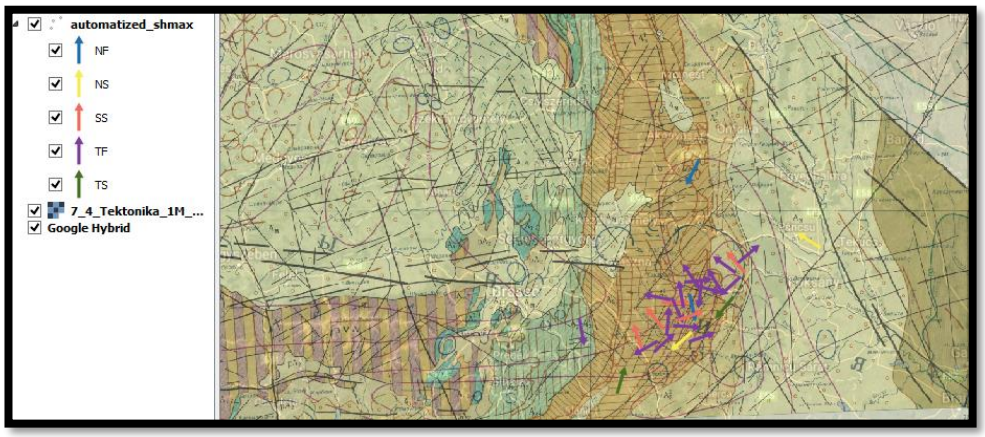

Figure 15. *The interpreted maximum horizontal stress orientations based on the azimuth of S_H and dominant tectonic regimes in the case of clusters created by the automated closest neighbour-based approach. The different regimes are indicated different colours (blue – normal faults, yellow – transtension, purple – thrust faults)*

It is visible in the comprehensive maps that the main tectonic structures are thrust faults and that stress orientations are variable. These pieces of information are similar to those obtained in previous studies and verify the complexity of the Vrancea Zone's tectonics.

However, it is necessary to take into account the difference among the clustering methods in the interpretation. While 8 subareas were created when using manual and

k-means clustering, the automated, nearest neighbour-based method generated 28 clusters. These clusters have a maximum of only 5-6 events, while the k-means and manual methods produced 15 events per cluster on average. Nevertheless, all grouping methods were able to create very small clusters (e.g. *Figure 9*, the 1st class of manual method or *Figure 10*, the 7th of k-means algorithm). Where there is a low amount of available data, results were generally less reliable than in larger subareas.

CONCLUSIONS

Based on this study, it seems to be apparent that the sets of epicentres generated by the manual and k-means method provided more reliable results for the evaluation of stress relations in the exterior SE Carpathians. The authors have begun developing an automatic method that is based on the principle of nearest neighbours. This initial version probably created too many sets for stress inversion. Thus, the next step will be to create an automatic clustering method more similar to the classical hierarchical method [19] but more suitable for taking into account the various types of different parameters (such as stress-field homogeneity) potentially relevant to stress-field estimations.

List of symbols

Symbol	Description	Unit
R	the shape ratio	–
R'	derived from shape ratio	–
σ_{11}, σ_1	the greatest principal stress axis (PSA)	°
σ_{22}, σ_2	the second greatest PSA	°
σ_{33}, σ_3	the smallest PSA	°
σ_V	the vertical principal stress	°
S_{Hmax}	maximum horizontal directions	–
M_w	momentum magnitude	–
$M_\_$	local magnitude	–

ACKNOWLEDGMENTS

This publication was created within the framework of ÚNKP-18-3 New National Excellence Program of the Ministry of Human Capacities, Hungary. Moreover, the authors would like to thank dr. Gribovszki Katalin (MTA CSFK Geodetic and Geophysical Institute) and dr. Norbert Péter Szabó for their help and support for this work.

REFERENCES

[1] Harangi, S., Novák, A., Kiss, B., Seghedi, I., Lukács, R., Szarka, L., Wesztergom, V., Metwaly, M., Gribovszki, K. (2015). Combined magnetotelluric and petrologic constrains for the nature of the magma storage system beneath the Late Pleistocene Ciomadul volcano (SE Carpathians). *Journal of Volcanology and Geothermal Research*, 290, pp. 82–96.

[2] Ismail-Zadeh, A., Matenco, L., Radulian, M., Cloetingh, S., Panza, G. (2012). Geodynamics and intermediate-depth seismicity in Vrancea (the south-eastern Carpathians): Current state-of-the art. *Tectonophysics*, 530, pp. 50–79.

[3] Martínez-Garzón, P., Ben-Zion, Y., Abolfathian, N., Kwiatek, G. & Bohnhoff, M. (2016). A refined methodology for stress inversions of earthquake focal mechanisms. *J. Geophys. Res. Solid Earth*, 121, pp. 8666–8687, doi:10.1002/2016JB013493.

[4] Hardebeck, J. L., Michael, A. J. (2006). Damped regional-scale stress inversions: Methodology and examples for southern California and the Coalinga aftershock sequence. *Journal of Geophysical Research*, Vol. 111, B11310, doi: 10.1029/2005JB004144

[5] Vavrycuk, V. (2014). Iterative joint inversion for stress and fault orientations from focal mechanisms. *Geophys. J. Int.*, 199, 69–77.

[6] Martínez-Garzón, P., Kwiatek, G., Ickrath M. & Bohnhoff, M. (2014). MSATSI: A MATLAB Package for Stress Inversion Combining Solid Classic Methodology, a New Simplified User-Handling, and a Visualization Tool. *Seismological Research Letters*, 85 (4), pp. 896–904.

[7] Moment Tensors by FMNEAR: http://fmnear.infp.ro/

[8] Moment Tensors by ISOLA-GUI: http://mt.infp.ro/

[9] National Institute of Geophysics and Volcanology (INGV): http://rcmt2.bo.ingv.it/

[10] Incorporated Research Institutions for Seismology (IRIS): https://ds.iris.edu/spud/momenttensor

[11] Meželovskij, N. V., Conseil d'assistance économique mutuelle (1987). *Space tectonic of map european countries – the CMEA members and SFRY/Council for Mutual Economic Assistance.* Echelle 1:1 000 000 (E 8°–E 30° /N 54°–N 42°), [Moskva]: Mingeo, Fédération de Russie.

[12] Shuttle Radar Topography Mission: http://srtm.csi.cgiar.org/

[13] Csanády, V., Horváth-Szováti, E. & Szalay, L. (2013). *Applied statistics.* (In Hungarian), pp. 162–172, Publishing of the University of West Hungary.

[14] Hartigan, J. A. (1975). *Clustering algorithms*.

[15] Hartigan, J. A. & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society.* Series C (Applied Statistics), 28 (1), pp. 100–108.

[16] Cesca, S., Sen, A. T. & Dahm, T. (2014). Seismicity monitoring by cluster analysis of moment tensors. *Geophysical Journal International*, Vol. 196, pp. 1813–1826.

[17] Zoback, M. L. (1992). First-and second-order patterns of stress in the lithosphere: The world stress map project. *Journal of Geophysical Research: Solid Earth*, 97 (B8), pp. 11703–11728.

[18] Barth, A., Reinecker, J. & Heidbach, O. (2008). *Stress derivation from earthquake focal mechanisms*. World Stress Map Project.

[19] Murtagh, F. (1983). A Survey of Recent Advances in Hierarchical Clustering Algorithms. *The Computer Journal*, Vol. 26 (4), pp. 354–359.