Crime "Hot-Spots" Identification and Analysis in Hungary by Computational Intelligence

László T. Kóczy[1, 2], Gábor Kovács[3], Péter Földesi[4], Szilvia Nagy[5], Boldizsár Tüű-Szabó[2], Gergő Fogarasi[2 [0000-0002-1854-2583]]

[1] Budapest University of Technology and Economics, Department of Telecommunications and Media Informatics
Műegyetem rkp. 3, 1111 Budapest, Hungary
E-mail: koczy@tmit.bme.hu

[2] Széchenyi István University, Department of Information Technology
Egyetem tér 1, 9026 Győr, Hungary
E-mail: {koczy, tuu.szabo.boldizsar, fogarasi.gergo}@sze.hu

[3] Széchenyi István University, Department of Criminal Sciences
Egyetem tér 1, 9026 Győr, Hungary
E-mail: gkovacs@sze.hu

[4] Széchenyi István University, Department of Logistics and Forwarding
Egyetem tér 1, 9026 Győr, Hungary
E-mail: foldesi@sze.hu

[5] Széchenyi István University, Department of Telecommunications
Egyetem tér 1, 9026 Győr, Hungary
E-mail: nagysz@rs1.sze.hu

Abstract: In the constantly growing and widening field of forensic science, crime maps are used in versatile ways. The representation of the data and analysis could offer some steps toward crime prevention and helps understand patterns, in terms of a timely distribution of crime types. Clustering is able to help identify criminal hot-spots and additional analysis may determine which areas require intervention. The aim of this study is to present an analysis of criminal information related to Hungary, in annual and monthly breakdown.

Keywords: clustering; Fuzzy C-Means; crime hot-spots; forensic science; crime prevention

1   Introduction

Computational intelligence (CI) methods are up-to-date IT approaches dealing with:

- Very complex problems

- Non-deterministic problems

- Uncertainty (not only Bayesian)

CI includes fuzzy systems [1] (applied in this study), evolutionary computation, and artifical neural networks.

The goal of this study is the identification and analysis of criminal hotspots (where far more crime cases happened than elsewhere, in a given area) in order to improve crime prevention, using CI methods.

To reach this goal the following steps will be carried out:

- Choose several types of crime, for detailed analysis

- Gather the related criminal data in a monthly breakdown

- Choose a (fuzzy) clustering method, for identification of geographic areas with similar patterns

- Generate (fuzzy) clusters

- Generate Dynamic maps, presenting their clusters

- Develop a Time series analysis and evolution

Considering certain types of crime in Hungary having a very low occurrence and so it is not possible to analyse them (for example car break-in). **Bodily harm** and **burglary** have been chosen for further investigation.

Pre-stat is a criminal database which was used to gather the required input data [2].

2   Preceding Research

As the initial step in the current study, in the earlier phase of this research, certain annual data for the years between 2010 and 2017 were used. The data have been graphically represented in 2D and 3D formats to show extreme areas, trends and predictable future tendencies. We attempted to interpret the charts and recorded the appropriate conclusions in a structured way. These new results were presented verbally at some conferences and published in a conference article [2].

The continuation of this research (published in the present paper), monthly breakdowns were used instead of the annual data, and a more sophisticated analysis has done, and so it would be possible to draw more in-depth conclusions and deepen the understanding of the phenomena.

3   The Public Administration Structure of Hungary

Hungary is divided into 7 regions, which include 19 Counties (in Hungarian: megye) and the capital (főváros), Budapest, as shown on Fig. 1, with the highlighted town of Győr, where this study was performed.

Counties are divided into 197 districts (járás):

- 174 within the Counties (járás)

- 23 in Budapest (kerület) – as shown on Fig. 2, with roman numbers

- Additionally, 23 individual towns are considered as districts

In this study, these 197 districts are considered single (homogenous) units.

Inside the districts, there are 322 towns, among them 23 cities with county rights. 126 "large municipalities" (város) among them and 2683 other municipalities (község) [2].
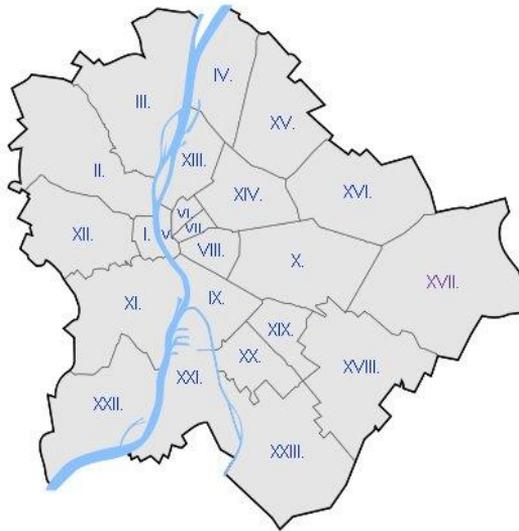

Figure 1
Public administration structure of Hungary

Figure 2
23 districts which are located in Budapest (in Hungarian: kerület)

4   Pre-Stat, the Structure of the Gathered Data

As mentioned above, Pre-Stat (available on https://prestat.lechnerkozpont.hu/bunmegelozes/ for hungarian citizens) is an interactive map of Hungarian crime data with database export features. Pre-Stat is not only for downloading databases, but it also provides generated maps as it can be seen on Fig. 3 as an example (for the given period and crime type, green means small number of cases, while red means high intensity of crime). It may be considered as a good example for an interactive up-to-date crime map system. The following data were downloaded from the Pre-Stat website for the two crime types mentioned (**bodily harm** and **burglary**, which were chosen for further investigation):

- Average criminal acts per 100,000 inhabitants, between 2010 and 2017 for 197 districts, in an annual breakdown

- Average criminal acts per 100,000 inhabitants, between 2013 and 2017 for 197 districts, in a monthly breakdown

Each district is represented with the coordinates of its seat on the related maps [2].
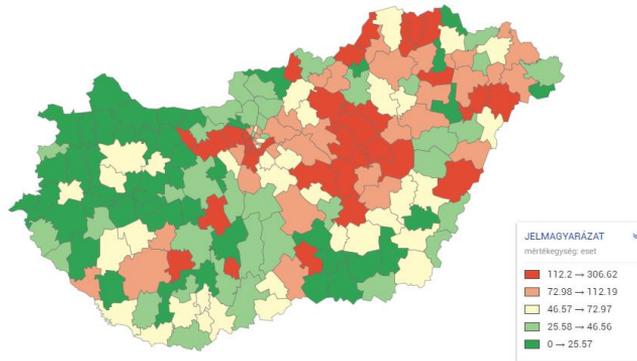
Figure 3
Sample image for Pre-Stat

5   Clustering Methods

5.1   The Aim of Using a Clustering Approach in this Study

The aim of this work is to determine the "crime hot-spots" consisting of several districts which are areas on the map with crime intensity of various levels. The crime hotspot detection and analysis can help with the decision making of the police to suggest where police actions need to be focused. Time series analysis can also help with the allocation strategy of a police force, by showing and predicting the tendencies of crime intensity [2].

5.2   Clustering in General

Clustering is an essential data mining tool that aims to discover inherent cluster structure in data [2] [3]. Its aim is to form clusters of objects (patterns), so that similar objects are grouped into the same clusters and different objects are grouped into different clusters [2] [4]. There are several methods known for clustering into vague clusters (with uncertain, maybe, overlapping borderlines), among various advantages and disadvantages [2].

5.3　Fuzzy C-Means Clustering (FCM)

The FCM algorithm is a clustering method which allows one instance of data to belong to two or more clusters with different degrees of membership. It is based on a clustering algorithm developed by Dunn [5] and later "fuzzified" and improved by Bezdek [6]. It can be used when the required number of clusters is known (is pre-determined). In each iteration of the FCM algorithm, the following objective function *J* is minimized: [2]

$$J = \sum_{i=1}^{N} \sum_{j=1}^{C} \mu_{ij} ||x_i - c_j||^2 \tag{1}$$

For a given data point x_i, the degree of its membership to cluster *j* is calculated as follows:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{C} (\frac{||x_i - c_j||}{||x_i - c_k||})^{\frac{2}{m-1}}} \tag{2}$$

where, *m* is the "fuzziness coefficient" and the center vector c_j is calculated as follows:

$$c_j = \frac{\sum_{i=1}^{N} \mu_{ij}^m . x_i}{\sum_{i=1}^{N} \mu_{ij}^m} \tag{3}$$

5.4　Generating Clusters by the FCM Method

As mentioned in this work, the FCM algorithm was used for creating clusters. The coordinates of the seats of the districts and the corresponding crime intensities served as an input of the algorithm. The number of clusters was chosen to be 30, based on expert estimation. (Experiments with less and more clusters lead to less interpretable results while choosing 30 was lead to present the best results.)

Because of presence of fuzzy clusters, technically, every district is included in every cluster, but with various (seatings low) membership degrees. For assigning a given district to a primary cluster, the cluster with the highest membership value has chosen.

If two districts were geographically close to each other, and if the number of crime acts were similar, they were merged into the same cluster.

For generating clusters by the FCM algorithm the corresponding Matlab toolbar was used [2]. Using this technique, crime hot-spots will appear, their position are formed and determined.

6   Generating Crime Maps

2D and 3D crime-maps were generated visualizing the clusters and the crime intensities [2].

In the 2D maps the centres of the clusters are marked and the average crime act values within the clusters are indicated. Based on these maps, we propose, that in the districts belonging to clusters with high criminal act numbers, further police intervention is necessary, and that very unpredictable behavior of a cluster needs deeper investigation and possibly additional intervention [2].

The 3D crime maps show the problematic regions with high crime intensity [2] (see Figs. 8-13).

To summarize, the following crime maps were prepared (272 in total):

- Average criminal acts per 100,000 inhabitants, between 2010 and 2017 for 197 districts, in annual breakdown, **burglary**, 8 pieces of 2D maps

- Average criminal acts per 100,000 inhabitants, between 2010 and 2017 for 197 districts, in annual breakdown, **bodily harm**, 8 pieces of 2D maps

- Average criminal acts per 100,000 inhabitants, between 2010 and 2017 for 197 districts, in annual breakdown, **burglary**, 8 pieces of 3D maps

- Average criminal acts per 100,000 inhabitants, between 2010 and 2017 for 197 districts, in annual breakdown, **bodily harm**, 8 pieces of 3D maps

- Average criminal acts per 100,000 inhabitants, between 2013 and 2017 for 197 districts, in monthly breakdown, **burglary**, 60 pieces of 2D maps

- Average criminal acts per 100,000 inhabitants, between 2013 and 2017 for 197 districts, in monthly breakdown, **bodily harm**, 60 pieces of 2D maps

- Average criminal acts per 100,000 inhabitants, between 2013 and 2017 for 197 districts, in monthly breakdown, **burglary**, 60 pieces of 3D maps

- Average criminal acts per 100,000 inhabitants, between 2013 and 2017 for 197 districts, in monthly breakdown, **bodily harm**, 60 pieces of 3D maps

Figs. 4-13 are presenting visualized data for each counties. The counties are represented with dots, and the colours are showing which dots are in the same cluster (Figs. 4-7), or the crime intensity (Figs. 8-13). These dots together are forming the recognizable shape of Hungary, so the axes can be matched with geographical coordinates. However, the geographical location of the country is irrelevant, therefore, they are not better marked.

6.1   2D Maps

In total, 136 pieces of 2D maps were generated for each year between 2010 and 2017, for each month between 2013 and 2017, and for the previously mentioned two crime types. Then the crime intensities of the clusters were calculated. The clusters with high membership values indicate high risk areas. The centers of the clusters in the maps are marked with an "X" (see Figs. 4-7) [2].
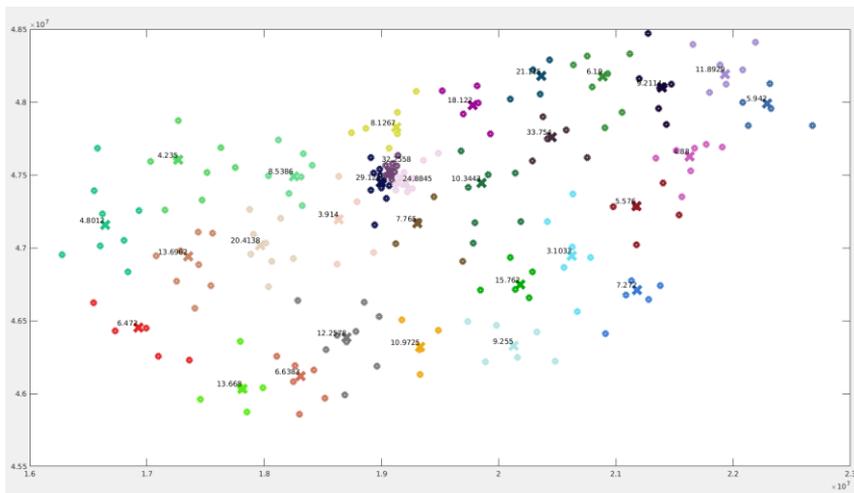

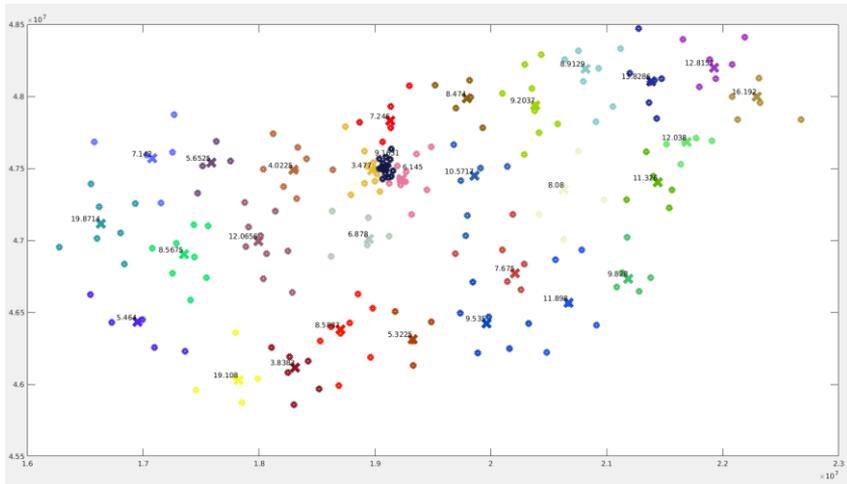Figure 4
2D map of the burglary in January of 2013

Figure 5
2D map of the bodily harm in July of 2017
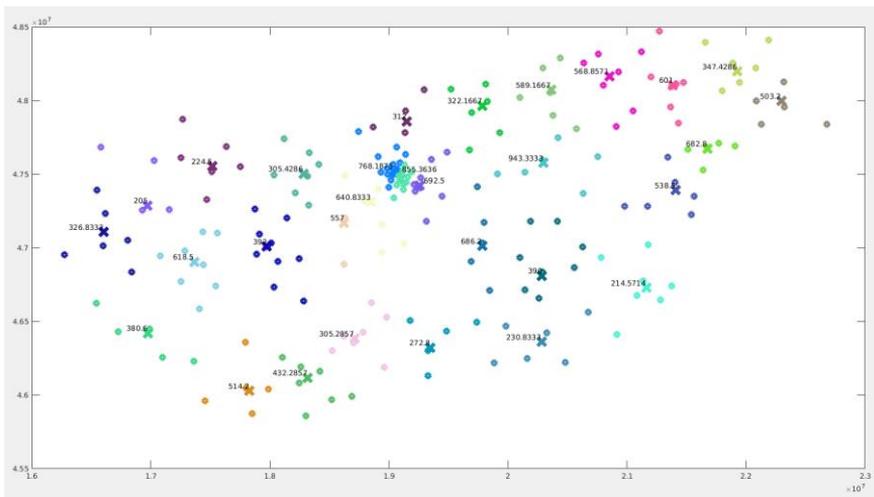
Figure 6
2D map of the burglary in 2012

Fig. 6 shows the 2D map of the **burglary** in 2012. The outlines and shape of Hungary can be easily recognized considering the dots.
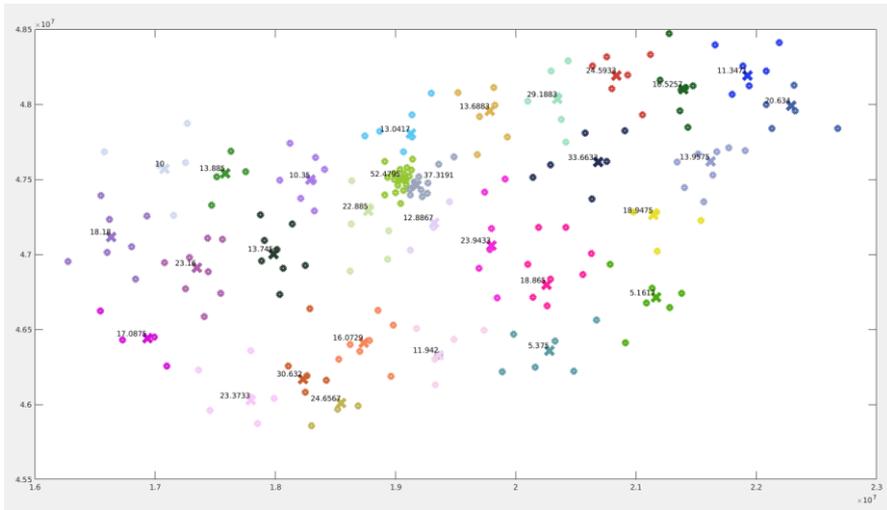
Figure 7
2D map of the burglary in March of 2013

6.2   3D Maps

Comparing the 3D maps of **bodily harm** in 2010 and 2017 (Figs. 8 and 9, show crime intensities) it can be concluded the crime prevention strategy in the North-Eastern region of Hungary is successful, however it still remained the most dangerous region in terms of **bodily harm** [2].
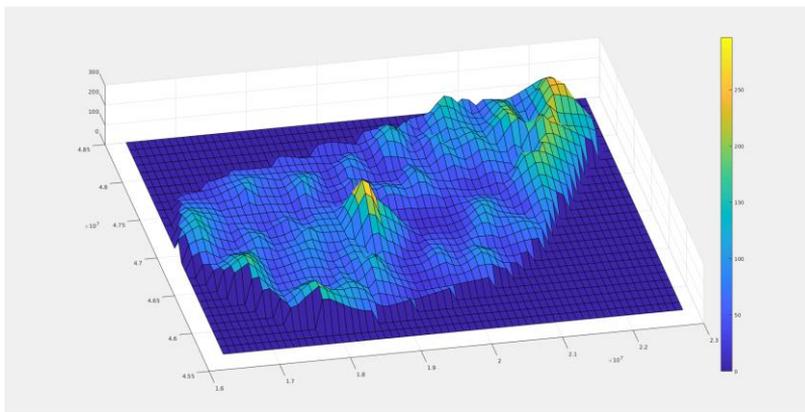
Figure 8
3D map of the bodily harm in 2010

In the early 2010 year, Budapest and its surroundings showed the highest **burglary** crime rate, but because of its decreasing tendency nowadays North-Eastern region of Hungary has the highest intensity (Figs. 8 and 9) [2].

Figure 9
3D map of the bodily harm in 2017

Figure 10
3D map of the burglary in 2011

Figure 11
3D map of the burglary in 2017

Figure 12
3D map of the burglary in November of 2014

Figure 13
3D map of the bodily harm in February of 2014

6.3   Dynamic 3D Maps for Investigating the Changes

Using the 3D maps, dynamic animated 3D maps were generated to analyse behavior in time, during the 8 years of available data [2]. They are embedded into a presentation which is uploaded to researchgate.net website [13].

Considering the data for the **burglary** in monthly breakdown between 2013 and 2017, it is possible to observe certain tendencies and features, especially if the animations are considered as 3D charts. The same phenomena in certain areas of the country may oc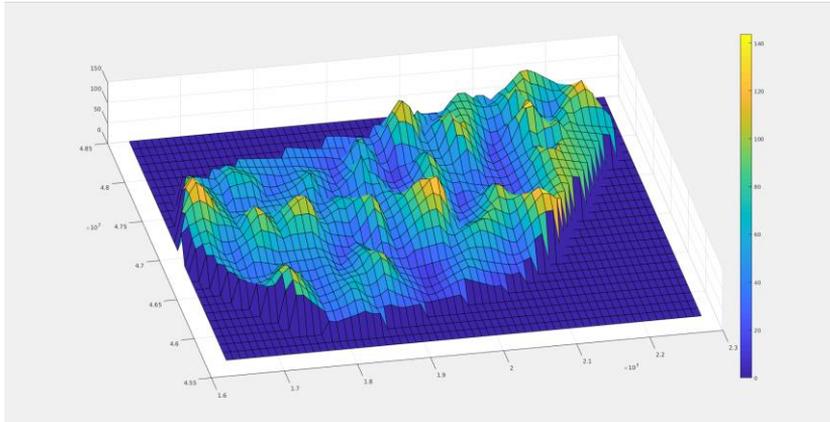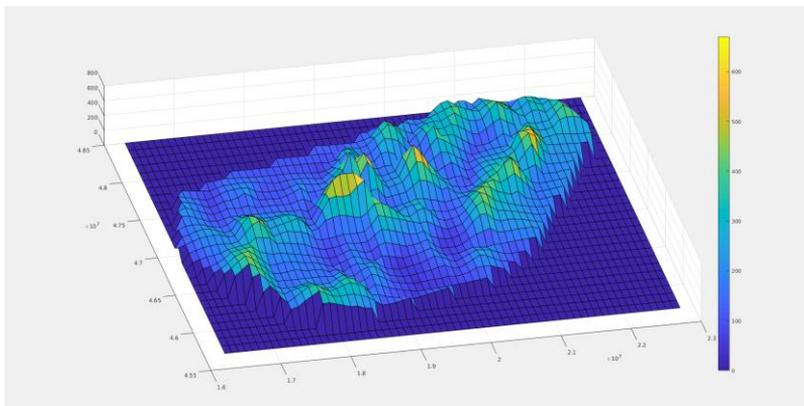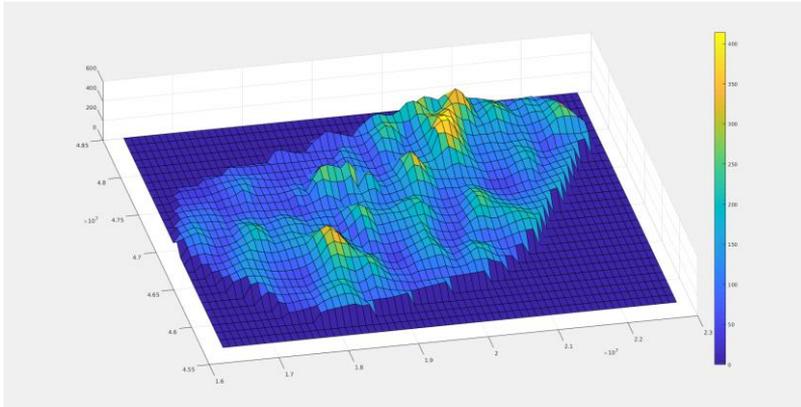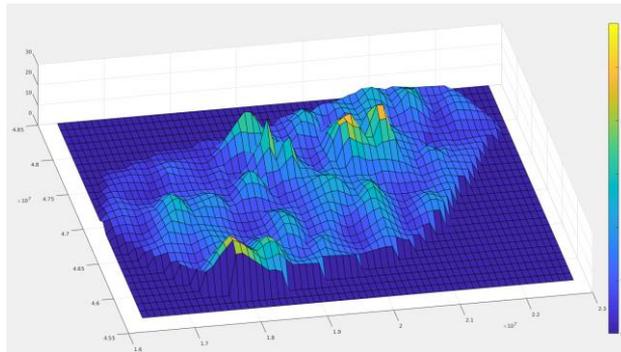cur in some given months in every single year. For instance, the values in a given area are very high in March, meanwhile they are very low everywhere in April. Conclusions drawn from the available data are as follows:

- January of 2013   The values are rather high in the Northern part of Hungary.

- March and October of 2013   The values are low for all Hungary.

- September of 2014   The values are rather high in the South-East part of Hungary.

- October of 2014   The values are high in all Hungary.

- May of 2015   The values are high in all Hungary.

- February in every year   The values are high near to the region of the lake Balaton.

- March in every year   The values are low in all Hungary.

- Jul/Aug/Oct in each year   The values are high near to Budapest.

The same for the **bodily harm**:

- Jan/Feb/Apr/Oct in each year   The values in the Northern part of Hungary are rather high.

- August in every year   The values are moderately high near to Budapest.

- November of 2013   The values are high near to the lake Balaton.

- March of 2014   The values are high near to the lake Balaton.

- June of 2014   The values are very high near to
                                  Balaton, while everywhere else the
                                  values are much more lower.

- June of 2017   The values near Kecskemét are
                                  rather high, while everywhere else the
                                  values are much lower.

7   Time Series Analysis

Time series analysis could be used because changes in time could describe the
behavior of fuzzy clusters much more precisely and indicate the success or lack of
success of the current police prevention strategy in the regions [2].

The prevailing trend could be:

- **Decreasing** – the actual crime prevention programs are working
  correctly, they should be continued

- **Increasing** – likely, the tendency will continue to grow, thus, further
  intervention is necessary

- **Random behavior** (stagnation or unpredictable) - not time-related,
  random. In the latter case, further analysis is needed, searching for
  other clusters.

Figure 14
Burglary (time independent) clusters labeled by numbers
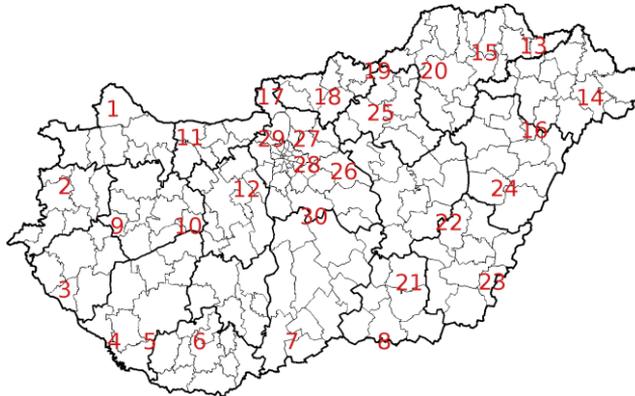
Figure 15
Bodily harm (time independent) clusters labeled by numbers

Figs. 14 and 15 show the labelled clusters for **burglary** and for **bodily harm**. The areas and the centres of these clusters are changing year to year, but not so extremely, so it is possible to identify more or less matching clusters in time. Using these numbered maps, it was possible to create a time series. It means, that the data for every available years was gathered into table form, using Figs. 14-15 helps to decide which actual annual cluster belongs to which general cluster, so where to put the data inside of the table. The result is presenting data rows for clusters on Figs. 14-15, for the entire time period which was examined. Using this, it is possible to draw as many charts as there are clusters on Figs. 14-15, for each crime types.

7.1   Burglary

Fig. 16 shows the changes of **burglary** crime intensity of the clusters in time. Most of the clusters indicate decreasing tendency or stagnation. However, cluster #19 presents randomly behavior with unexpected high values alternating with lower crime rates. Here, further detailed analysis is necessary [2].
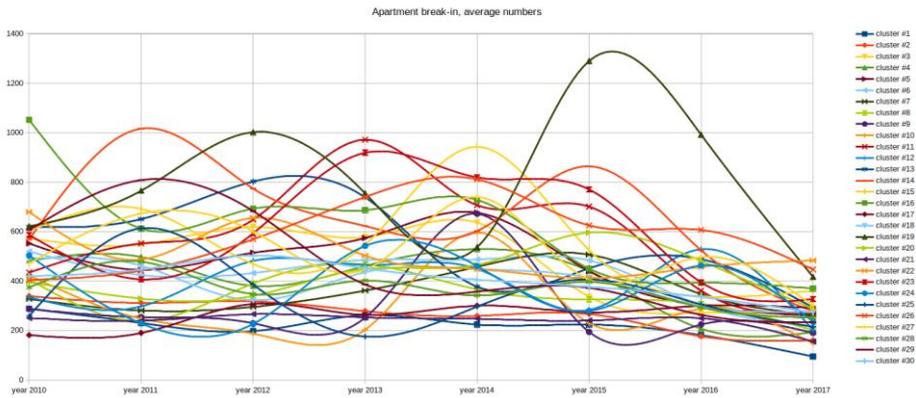
Figure 16
Time series of the burglary (2010-2017)

7.2   Bodily Harm

Similarly to **burglary** most of the clusters show decreasing tendency of **bodily harm**. There are also some clusters with random behavior (Fig. 17) [2].
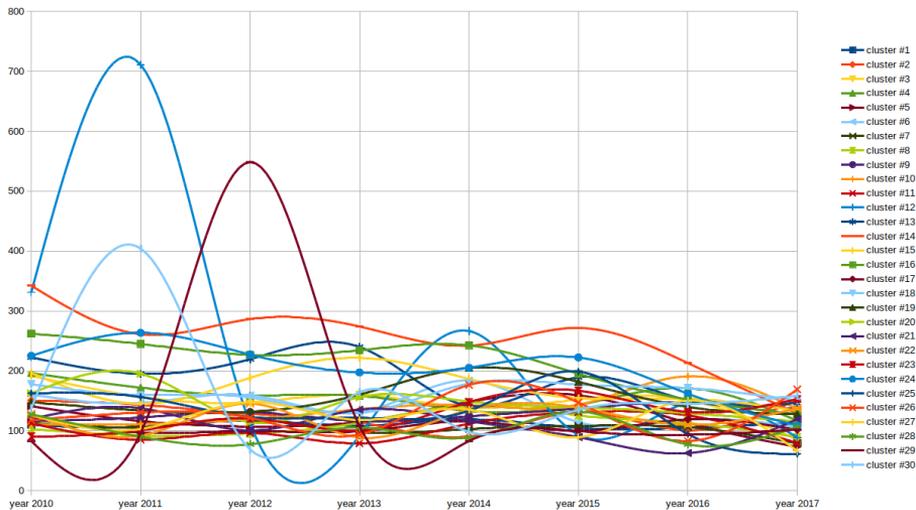

Figure 17
Time series of the bodily harm (2010-2017)

8   Further Technical Tools and Goals Suggested for the Future Research

In the preceeding work [2] of this research some initial steps were made in the field of crime map generation and its analization using the rough preprocessing of the collected data.

Then, the research was continued, using much more precise data, which made it possible to draw even further, more precise conclusions.

Fuzzy clusters help identify approximate areas of similar behavior, and thus suggest further police intervention (or lack of further investigation) areas/districts – independently, from the public administration borders. The time dependent behavior of these clusters reveal tendencies, and suggest further aspects of intervention. These tendencies and changes should be matched with historical data on past interventions and connections must be searched. Very irregular behavior may indicate unknown effects in the area and needs deeper analysis [2].

In the future research, instead of using districts in Hungary, towns, zip codes or even GPS coordinates could be used to reach more precise results.

A possible improvement could be to use the [7] DBMEA evolutionary-memetic algorithm or [8] the Gath-Geva method (instead of Fuzzy C-Means – FCM [12]) to produce more precise clusters.

Smoothing and filtering the data would be possible using [9] wavelet transform (which was first used and mentioned by A. Haar in his article about the Theory of Orthogonal Function Systems in 1910 [10] [11]).

The time series analysis part of the research could apply the traditional (statistical) Box-Jenkins method (named after G. Box and G. Jenkins) or other methodologies [2].

Considering more types of crime would make the study more comprehensive and useful, while proving cross-correlation between types of crime could help even more in crime prevention as well as crime forecast by various time series prediction approaches [2].

Conclusions

The preprocessing and the fuzzy clustering of the data clearly identified some areas of higher criminal activity (crime hot spots), and also the behavior of these areas, in real-time, was identified. The sub-normalization of the fuzzy clusters (assigning the heights of the clusters corresponding to the absolute crime intensities) point out problem areas, where the number of certain types of crimes is higher than the average. The tendencies of these crime hot spots in time (increasing, stagnating, decreasing, or random) may be the starting points for further police intervention. The first two tendencies, especially the increasing

tendency, mark the areas where immediate urgent intervention is necessary. It may mean the increase of the police force on spot, mounting more video cameras for observing public areas, or education, creating jobs, etc. A decreasing tendency is an indicator of proper police measures and regulations being carried out in any given area. Random behavior may be enigmatic, or the result of some sudden change in the general conditions, and these areas need further deeper investigation [2].

In future work, a full analysis of the time dependent behavioral patterns will be applied, and alternative clustering techniques may be compared, when they represent true images of statistical observations [2].

It is important to refine the investigations towards potentially periodical behavior within the year (connected to holiday seasons, mass events, etc.) and clusters depending on the part of the year (clusters of groups of weeks or days) should be identified. Having these results, police intervention measures may be recommended, in terms of exact places and times [2].

Acknowledgement

References

[1]   I. Batyrshin, O. Kaynak and I. Rudas: Fuzzy modeling based on generalized conjunction operators. Fuzzy Systems, IEEE Transactions on. 10. 678 - 683. 10.1109/TFUZZ.2002.803500, 2002

[2]   L. T. Kóczy, G. Kovács, P. Földesi, Sz. Nagy, B. Tüű-Szabó and G. Fogarasi: Identification and dynamic analysis of crime hotspots in Hungary by a complex Computer Intelligence approach. 23rd IEEE International Conference on Intelligence Engineering Systems, Gödöllő, Hungary, 2019

[3]   A. Adolfsson, M. Ackerman, N. Brownstein: To Cluster, or Not to Cluster: An Analysis of Clusterability Methods. Pattern Recognition. 88. 10.1016/j.patcog.2018.10.026, 2018

[4]   G. Desagulier: Clustering Methods. Corpus Linguistics and Statistics with R. DOI: 10.1007/978-3-319-64572-8_10, 2017

[5]   J. C. Dunn: A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. Journal of Cybernetics 3: 32-57, 1973

[6]   J. C. Bezdek: Pattern Recognition with Fuzzy Objective Function Algoritms. Plenum Press, New York, 1981

[7]   L. T. Kóczy, P. Földesi, B. Tüű-Szabó: A discrete bacterial memetic evolutionary algorithm for the traveling salesman problem. 3261-3267, 10.1109/CEC.2016.7744202, 2016

[8]   I. Gath and A. B. Geva: Unsupervised optimal fuzzy clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11(7), pp 773-781, 1989

[9]   A. Graps: An Introduction to Wavelets. IEEE Comput. Sci. Eng. 2, 2 (June 1995) 50-61, DOI: https://doi.org/10.1109/99.388960 1995

[10]   A. Haar: Zur Theorie der orthogonalen Funktionensysteme. Math. Ann. 69: 331, DOI: https://doi.org/10.1007/BF01456326, 1910

[11]   Sz. Nagy and J. Pipek: Wavelet Based Density Operators, Electron Density and Energy Functionals. 2005

[12]   J. Bezdek, R. Ehrlich, W. Full: FCM—the Fuzzy C-Means clustering-algorithm. Computers & Geosciences. 10, 191-203, 10.1016/0098-3004(84)90020-7, 1984

[13]   L. T. Kóczy, G. Kovács, P. Földesi, Sz. Nagy, B. Tüű-Szabó, G. Fogarasi and A. Czebe: Identification and dynamic analysis of crime hotspots in Hungary by a complex CI approach. FEE 2019 Conference, London, 2019, URL: https://www.researchgate.net/publication/331718127_ Identification_and_dynamic_analysis_of_crime_hot-spots_in_Hungary _by_a_complex_CI_approach