Intellectual analysis for Ukrainian refugees' number dynamics depending on attack frequency from russian federation

Victoria Vysotska†, Viktoriia Yakovlieva*,†, Sofiia Ivaniv† and Iryna Shakleina

Lviv Polytechnic National University, Stepan Bandera 12, 79013 Lviv, Ukraine

Abstract

The article presents the results of an intellectual analysis of the relationship between the dynamics of the number of Ukrainian refugees and the frequency of attacks by russia on the civilian and political infrastructure of Ukraine. The study aims to identify statistical relationships between these phenomena and visualize their dynamics using the R programming language. Three data samples were processed: the number of refugees, the number of attacks on civilian objects, the number of attacks on political objects, and the dynamics of the number of Ukrainian refugees abroad. The paper uses methods of preliminary statistical analysis, time series smoothing (moving average, median filtering, exponential smoothing) and also conducts correlation analysis. The results indicate a strong connection between the intensity of attacks, especially on political objects, and the growth of the number of refugees. The analysis allows for a deeper understanding of the impact of military actions on migration processes and may be helpful in predicting future trends.

Keywords

Intelligence analysis, big data analysis, refugees, R, shelling, time series, smoothing, correlation, aggression, statistical analysis, migration, infrastructure

1. Introduction

The full-scale invasion of russia in Ukraine has caused not only infrastructure destruction but also large-scale humanitarian consequences, including mass migration of the population. The number of Ukrainian refugees forced to leave the country has grown in parallel with the intensity of attacks on civilian and political targets. The scientific community is actively researching migration processes. Still, most of the work focuses on the sociological or political aspects of the problem, leaving aside the analytical relationships between the frequency of attacks and the dynamics of migration. In this context, the use of modern tools for data mining and mathematical modelling using the R programming language, which allows for deep statistical processing, visualization, and correlation analysis of large data sets, is of particular importance. The work is aimed at filling the existing scientific gap and has applied value to the development of effective strategies for responding to humanitarian crises caused by armed conflicts. For Ukraine, the results of such a study are significant from the point of view of strategic planning, social protection of the population, and international cooperation.

The purpose of the work is to identify and formalize the relationships between the dynamics of the number of Ukrainian refugees and the frequency of russian attacks on the civilian and political infrastructure of Ukraine. To achieve this goal, the following tasks must be solved:

MoMLeT-2025: 7th International Workshop on Modern Machine Learning Technologies, June, 14, 2025, Lviv-Shatsk, Ukraine

* Corresponding author.

† These authors contributed equally.

Victoria.A.Vysotska@lpnu.ua (V. Vysotska); viktoriia.yakovlieva.sa.2022@lpnu.ua (V. Yakovlieva); sofiia.ivaniv.sa.2022@lpnu.ua (S. Ivaniv); ioshakleina@gmail.com (Iryna Shakleina)

0000-0001-6417-3689 (V. Vysotska); 0009-0003-2887-173X (V. Yakovlieva); 0009-0003-1989-9300 (S. Ivaniv); 0000- 0003-0809-1480 (I. Shakleina)

- generate data samples on shelling and the number of refugees;
- perform pre-processing, normalization and visualization of data;
- apply time series smoothing methods to identify trends;
- create correlation models of dependencies between parameters;
- carry out cluster analysis of dynamics based on statistical indicators.

The object of the study is the forced migration of the population of Ukraine under the influence of military actions. The subject of the study is the dependence of the dynamics of changes in the number of Ukrainian refugees on the frequency and nature of attacks on civilian and political infrastructure. For the first time, a comprehensive approach to the analysis of the relationship between the number of refugees and the intensity of attacks is proposed, which is based on a combination of statistical methods, time series smoothing and clustering in the R environment. New results were obtained regarding the degree of influence of attacks on political infrastructure on the increase in the number of refugees. It was established that these attacks have a stronger correlation with the dynamics of migration, which was not covered in previous studies. The work improves the methodology for analysing migration processes in conditions of armed conflict, providing the opportunity for operational forecasting of the humanitarian situation in the country.

2. Related works

The issue of the impact of armed conflicts on migration processes attracted the attention of researchers long before the start of the full-scale war in Ukraine. However, it was the events after 2022 that became the impetus for the active study of changes in the structure and scale of forced migration as a result of armed aggression. The works [1,2] investigated the general trends in population displacement as a result of the war and, in particular, studied the socio-economic consequences and challenges for host countries. Specific attention was paid to Ukraine as the largest source of refugees in Europe after 2022 [3].

A number of studies [4,5] analyse the demographic characteristics of refugees, gender composition, access to services, and adaptation conditions. However, insufficient attention has been paid to statistical modelling and attempts to establish a connection between migration waves and specific military events. Study [6] is one of the few that uses regression analysis to identify a correlation between the number of attacks on infrastructure and the number of migrants. Work [7] emphasizes the importance of building forecasting models but uses only aggregated indicators without deep temporal detail.

Existing publications mainly focus on qualitative analysis of the phenomenon, which creates a need for more formalized approaches using statistical methods of time series analysis, smoothing, clustering and data normalization. Such techniques allow moving from describing the phenomenon to identifying hidden patterns and creating tools for operational forecasting, which is especially relevant for Ukraine in the context of a long-term threat from the aggressor.

The issue of forced migration as a result of russia's armed aggression against Ukraine has received considerable attention in modern scientific and intergovernmental literature. The works of Libanova, Poznyak, and Tsymbal [1] provide a fundamental analysis of demographic changes in Ukraine after the outbreak of full-scale war. The researchers emphasize the complexity of the problem, including social, economic, and security aspects. Digital methods of tracking migration flows have become an important research tool: Wycoff et al. [2] demonstrated the effectiveness of analysing digital traces, in particular data from Google and social networks, to monitor the movement of Ukrainian refugees in real-time. Minora et al. [3] worked in a similar direction, using Facebook advertising data to build a model of Ukrainian movements within the EU, confirming the feasibility of using big data in times of crisis.

The organizational and political aspects are reflected in the reports of international organizations. OECD analysis [4] emphasizes the unprecedented nature of the current flow of refugees to Europe, suggesting ways to improve integration strategies. The specific impact of massive shelling on migration dynamics is revealed in a study by the International Center for Ukrainian Victory [5], where particular data shows how waves of attacks lead to a sharp increase in the number of people leaving the country.

The focus of the study by Kovtun and Salabay [6] is the integration of Ukrainians in host countries, particularly in Germany. The work includes statistical processing of questionnaire data, which is a valuable addition to macro-level assessments. Publications by Reuters [7] and The Guardian [8] complement the scientific picture with a modern context - they emphasize the threat of “migration weapons” as an element of hybrid warfare and draw attention to the potential decrease in international support due to donor fatigue. This review allows us to conclude that although the issue of forced migration is actively researched, most of the existing work does not focus on quantitative modelling of the relationship between attacks on infrastructure and migration dynamics. Therefore, a study based on time series and statistical analysis is a logical and vital continuation of this scientific discussion.

3. Methods and materials

The study used a set of methods of mathematical statistics, data mining, and computer modelling [9-16], which allowed for a comprehensive analysis of the relationship between the number of Ukrainian refugees and the frequency of attacks on the civilian and political infrastructure of Ukraine by russia. The primary tool for implementing all stages of the analysis was the R programming language, which provides extensive capabilities for processing, visualization, and statistical modelling of data. The use of the R language is due to its openness, flexibility, and the availability of specialized packages for processing time series (zoo, forecast, TTR), plotting (ggplot2, plotly), and conducting cluster analysis (cluster, factoextra). It allowed for the effective processing of three independent samples: data on the number of refugees, attacks on civilian infrastructure, and political objects.

At the stage of primary data processing, methods of normalization, cleaning and structuring of data were used, followed by their presentation in the format of time series. Smoothing methods were chosen to study the dynamics, in particular:

- simple moving average method — to eliminate random fluctuations and identify the primary trend;
- weighted moving average — to better take into account the asymmetric effects of events;
- median filtering — as an effective means of highlighting long-term trends in the presence of outliers;
- exponential smoothing — to reflect the inertia and adaptability of processes.

Correlation analysis methods were used to identify the nature of the dependencies between variables, including the calculation of Pearson coefficients, determination and correlation relations, and the construction of correlation fields. These methods made it possible to establish the degree of connection between the number of attacks and the change in the number of refugees. In addition, hierarchical agglomerative cluster analysis was used to analyse structural changes in the data, which allowed group periods with similar characteristics of attacks and migration dynamics to identify phase transitions in the migration behaviour of the population. Thus, the selected methods and tools provided a reliable basis for a comprehensive analysis of complex nonlinear dependencies in time series, allowing the identification of hidden relationships between military events and the behaviour of the civilian population in conditions of armed conflict. For this work, the R language was chosen to be the best suited for statistical data analysis. R is an open-source programming language. It is used for statistical data processing (statistical calculations) and graphics (visualization). R can be used to work with data sets. For example, R is used to solve complex problems of mathematical statistics, perform primary data analysis, and perform mathematical modelling. Using R, you can prepare data for research and process experimental results in various areas of life, such as medicine, nature management, environmental protection, econometrics and financial analysis, marketing, engineering calculations, etc. R not only supports a wide range of statistical and numerical methods but can also be extended with software packages - libraries for specific functions or special areas of application. The first versions of R were created in the 1990s. Since then, it has been constantly evolving, adding new packages and features, improving existing ones, and fixing bugs. Thanks to an active community of users and developers, R remains relevant and is constantly being updated. The programming language has a unique syntax and framework for running programs. R is actively used in artificial intelligence and machine learning. At first glance, the programming language may seem quite complex. In fact, it is pretty logical and straightforward. R was created by developers for scientists who have experience and knowledge in the field of mathematical analysis, static methods, and probabilistic deviations. It has a number of advantages:

- Code in this programming language can be run without compilation, as it uses an interpreter that demonstrates how the program works in real-time;
- R is efficient and productive due to its vector approach.

The R programming language is used to work with data:

- collecting and analysing data from various sources;
- searching for patterns and deviations;
- testing and validating hypotheses;
- visualizing data in multiple ways;
- working with statistical data to identify anomalies.

Thus, R is and remains one of the most flexible and powerful programming languages designed specifically for data analysis. The R language is capable of processing a large number of types of various objects - vectors, matrices, lists, data tables, etc. The R programming language can also work with a large number of data types. These can be, for example, numbers with a fractional part, integers, text records, date and time values, logical operation values, etc.

Three datasets [17-18] were selected for analysis, which are directly related to the topic of the work. The first dataset is the number of Ukrainian refugees abroad. This dataset contains information about the number of Ukrainians who left the country due to the war. The second dataset is the shelling of Ukrainian civilian infrastructure. This dataset includes information on the number and scale of attacks on Ukraine's civilian infrastructure. And the third is the shelling of the political infrastructure of Ukraine. This dataset covers attacks on administrative buildings, government facilities and other political institutions. With the help of these three datasets, we want to show how the shelling of Ukraine affects the number of Ukrainians travelling abroad. The analysis will help to understand how attacks on civilian and political infrastructure are related to forced population migration. Dataset 1 (Fig. 1a) contains information on 1,821 cases of shelling of civilian infrastructure in different regions of Ukraine from January 2018 to October 2024. The data are structured by administrative units (regions and cities) with corresponding codes and the number of events.

Dataset 2 (Fig. 1b) presents a larger dataset (3,146 records) on shelling of political infrastructure. It is noteworthy that the number of events (Events) varies significantly between regions – from single cases to 40 events in the individual areas, which indicates an uneven distribution of attacks. Dataset 3 (Fig. 1c) shows the dynamics of the number of Ukrainian refugees abroad from April 25, 2022, to March 12, 2024. There is an increase in the number of refugees – from 85,000 to 5,982,920 people. The data is presented in CSV format and contains 470 rows after cleaning, presenting data values in the form of a compressed table in Table 1-3.

The graphical representation of the data is given in Fig. 2-4 for the respective datasets. For dataset 1 (Fig. 2a) in the initial period (2018-2022), the data shows a relatively low and stable level of incidents, the number of which fluctuates within 20-40 events per month. In 2022, there is a sharp jump in the number of incidents to 700-800 incidents. It is the most intense period on the graph.

After the main surge (2022-2024), the number of incidents decreased but remained significantly higher than before 2022, fluctuating between 200-400 events per month. Towards the end of the graph, there is a sharp decrease in the number of incidents.

```
 [1] "Country"      "Admin1"       "Admin2"       "ISO3"         "Admin2 Pcode"
 [6] "Admin1 Pcode" "Month"        "Year"         "Events"       "Fatalities"
# A tibble: 1,821 × 10
   Country Admin1   Admin2 ISO3  `Admin2 Pcode` `Admin1 Pcode` Month Year  Events
   <chr>   <chr>    <chr>  <chr> <chr>          <chr>          <chr> <chr>  <dbl>
 1 Ukraine Kharkiv  Khark… UKR   UA6312         UA63           Janu… 2018       1
 2 Ukraine Crimea   Simfe… UKR   UA0116         UA01           Janu… 2018       1
 3 Ukraine Dnipro…  Nikop… UKR   UA1208         UA12           Janu… 2018       1
 4 Ukraine Luhansk  Luhan… UKR   UA4406         UA44           Janu… 2018       1
 5 Ukraine Donetsk  Kalmi… UKR   UA1410         UA14           Janu… 2018       2
 6 Ukraine Donetsk  Donet… UKR   UA1408         UA14           Janu… 2018       4
 7 Ukraine Donetsk  Pokro… UKR   UA1416         UA14           Janu… 2018       1
 8 Ukraine Dnipro…  Dnipr… UKR   UA1202         UA12           Janu… 2018       1
 9 Ukraine Cherni…  Chern… UKR   UA7306         UA73           Janu… 2018       1
10 Ukraine Odesa    Odesk… UKR   UA5110         UA51           Janu… 2018       1
# ℹ 1,811 more rows
# ℹ 1 more variable: Fatalities <dbl>
```

```
 [1] "Country"      "Admin1"       "Admin2"       "ISO3"         "Admin2 Pcode"
 [6] "Admin1 Pcode" "Month"        "Year"         "Events"       "Fatalities"
# A tibble: 3,146 × 10
   Country Admin1   Admin2 ISO3  `Admin2 Pcode` `Admin1 Pcode` Month Year  Events
   <chr>   <chr>    <chr>  <chr> <chr>          <chr>          <chr> <chr>  <dbl> Розмір прочитаних даних: 470 рядків, 2 колонок
 1 Ukraine Kharkiv  Khark… UKR   UA6312         UA63           Janu… 2018       1 Оригінальні дати (перші 10 рядків):
 2 Ukraine Luhansk  Sievi… UKR   UA4412         UA44           Janu… 2018      40  [1] "2022-04-25" "2022-04-27" "2022-05-03" "20
 3 Ukraine Odesa    Bilho… UKR   UA5104         UA51           Janu… 2018       1  [6] "2022-05-19" "2022-05-21" "2022-05-22" "20
 4 Ukraine Crimea   Simfe… UKR   UA0116         UA01           Janu… 2018       1 Перші рядки очищених даних:
 5 Ukraine Kherson  Kakho… UKR   UA6506         UA65           Janu… 2018       1   RefugeesDate NoRefugees
 6 Ukraine Dnipro…  Nikop… UKR   UA1208         UA12           Janu… 2018       1 1   2022-04-25      85000
 7 Ukraine Donetsk  Krama… UKR   UA1412         UA14           Janu… 2018       7 2   2022-04-27     128300
 8 Ukraine Luhansk  Luhan… UKR   UA4406         UA44           Janu… 2018      25 3   2022-05-03     165700
 9 Ukraine Donetsk  Volno… UKR   UA1404         UA14           Janu… 2018      13 4   2022-05-09     180700
10 Ukraine Donetsk  Horli… UKR   UA1406         UA14           Janu… 2018      81 5   2022-05-18     180894
# ℹ 3,136 more rows                                                               6   2022-05-19     225894
# ℹ 1 more variable: Fatalities <dbl>                                             Кількість рядків після очищення: 470
```

Figure 1: Result presentation for the dataset – (a) Ukrainian civilian infrastructure shelling, (b) Ukrainian political infrastructure shelling and (c) Ukrainian refugees abroad number dynamics

Table 1
Results of registration of dynamics of the indicator of Ukrainian refugees abroad

Date	Value	Date	Value	Date	Value	Date	Value	Date	Value
2022-04-25	85000	2022-06-01	4773142	2022-06-17	5146852	2022-07-11	5675126	2022-07-26	6160801
2022-04-27	128300	2022-06-02	4774703	2022-06-19	5147396	2022-07-12	5827712	2022-07-27	6178196
2022-05-03	165700	2022-06-03	4774703	2022-06-20	5166105	2022-07-13	5846934	2022-07-28	6194406
2022-05-09	180700	2022-06-05	4792122	2022-06-21	5349476	2022-07-14	5863980	2022-07-29	6211291
2022-05-18	180894	2022-06-06	4824522	2022-06-22	5349487	2022-07-15	5883282	2022-07-30	6227592
2022-05-19	225894	2022-06-07	4913869	2022-06-23	5350285	2022-07-16	5902393	2022-07-31	6244872
2022-05-21	1005894	2022-06-08	4919773	2022-06-26	5350863	2022-07-17	5922269	2022-08-01	6272752
2022-05-22	1068175	2022-06-09	4942901	2022-06-27	5371814	2022-07-18	5956085	2022-08-02	6302641
2022-05-23	1143057	2022-06-10	4958105	2022-06-28	5493080	2022-07-19	5988696	2022-08-03	6305849
2022-05-24	1252598	2022-06-11	4972953	2022-06-29	5493884	2022-07-20	6011362	2022-08-07	6305867
2022-05-25	1509585	2022-06-12	4986518	2022-07-03	5498558	2022-07-21	6026076	2022-08-08	6311974
2022-05-26	1542006	2022-06-13	5076934	2022-07-04	5524054	2022-07-22	6043626	2022-08-09	6463084
2022-05-29	1565693	2022-06-14	5113791	2022-07-05	5651401	2022-07-23	6060514	2022-08-10	6490187
2022-05-30	1697393	2022-06-15	5129970	2022-07-06	5655548	2022-07-24	6076119	2022-08-13	6521187
2022-05-31	3032612	2022-06-16	5145849	2022-07-10	5656749	2022-07-25	6099745	2022-08-14	6521506

Table 2
Results of registration of dynamics of the indicator of destruction of Ukrainian political targets

N	Value	N	Value	N	Value	N	Value	N	Value	N	Value	N	Value	N	Value
1	1	14	83	27	1	40	157	53	103	66	2	79	183	92	2
2	40	15	40	28	4	41	1	54	67	67	1	80	71	93	165
3	1	16	1	29	1	42	4	55	135	68	123	81	1	94	1
4	1	17	19	30	21	43	8	56	1	69	78	82	33	95	1
5	1	18	99	31	1	44	1	57	84	70	1	83	104	96	9
6	1	19	1	32	21	45	1	58	45	71	1	84	1	97	1
7	7	20	2	33	138	46	123	59	1	72	1	85	1	98	29
8	25	21	1	34	89	47	1	60	2	73	5	86	1	99	1
9	13	22	3	35	180	48	43	61	26	74	47	87	5	100	44
10	81	23	91	36	80	49	2	62	100	75	56	88	1		
11	51	24	1	37	106	50	3	63	1	76	105	89	1		
12	1	25	96	38	1	51	24	64	2	77	60	90	1		
13	154	26	1	39	20	52	23	65	9	78	135	91	214		

Table 3
Results of Ukrainian civilian infrastructure shelling indicator dynamics registration

N	Value	N	Value	N	Value	N	Value	N	Value	N	Value	N	Value	N	Value
1	1	14	1	27	1	40	1	53	2	66	1	79	1	92	1
2	1	15	1	28	1	41	1	54	1	67	2	80	2	93	2
3	1	16	2	29	1	42	1	55	5	68	1	81	2	94	1
4	1	17	2	30	2	43	1	56	2	69	1	82	1	95	1
5	2	18	1	31	7	44	3	57	3	70	1	83	1	96	1
6	4	19	2	32	1	45	1	58	3	71	1	84	4	97	2
7	1	20	4	33	1	46	1	59	1	72	1	85	1	98	3
8	1	21	1	34	1	47	1	60	8	73	1	86	2	99	1
9	1	22	2	35	1	48	1	61	1	74	1	87	1	100	1
10	1	23	1	36	2	49	1	62	1	75	1	88	1		
11	1	24	9	37	1	50	3	63	2	76	2	89	1		
12	1	25	1	38	6	51	1	64	4	77	1	90	1		
13	1	26	1	39	3	52	1	65	4	78	1	91	2		

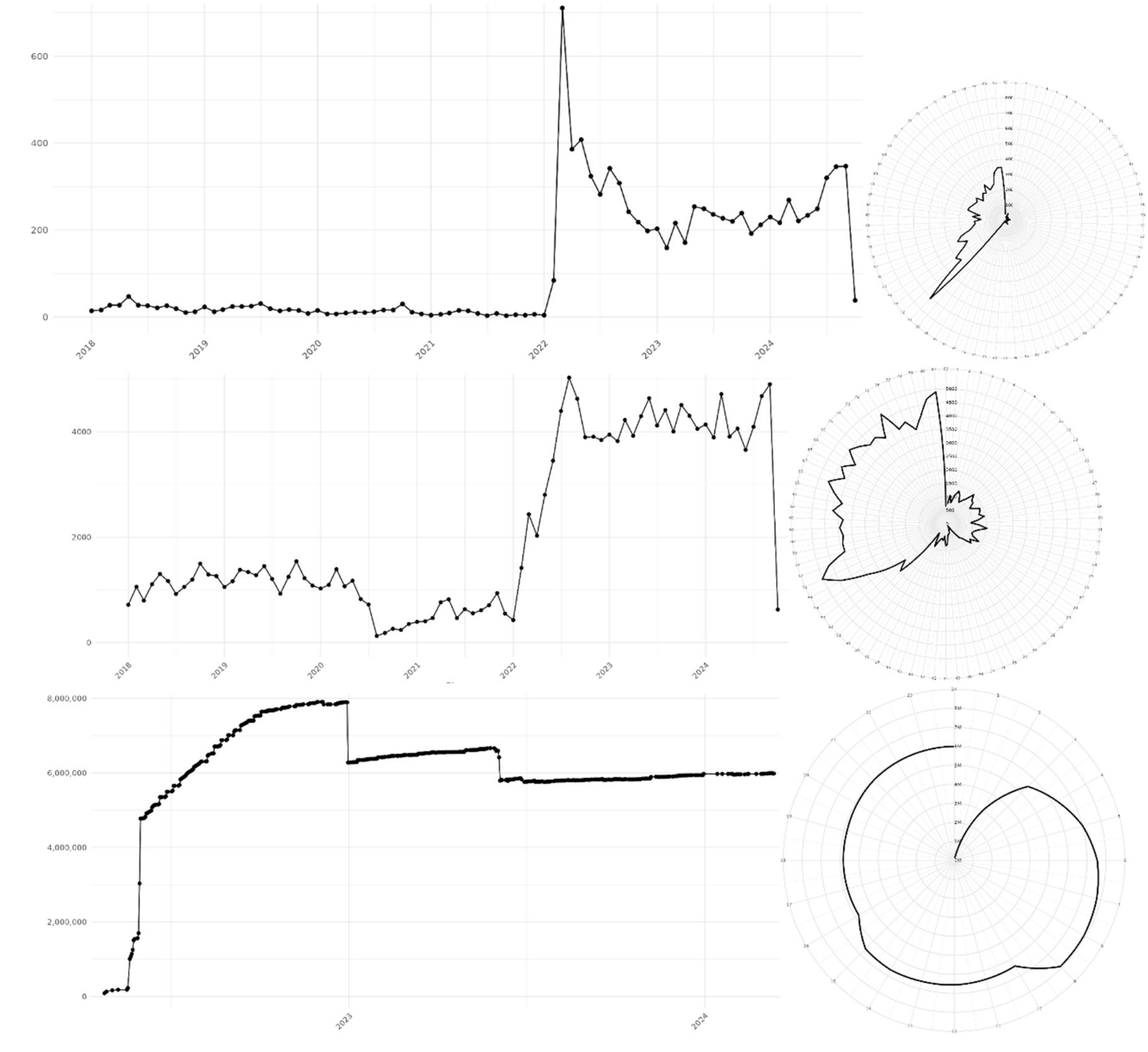


Figure 2: Graphical representation of the dynamics of the indicator for the period from January 2018 to October 2024 (a) Ukrainian civilian infrastructure shelling, (b) Ukrainian political infrastructure shelling and (c) Ukrainian refugees abroad number dynamics since the beginning of the full-scale war in the Cartesian coordinate system and in the polar coordinate system.

According to the data analysis for dataset 2 (Fig. 2b), the initial period (2018-2022) has a relatively stable level of events. The indicators fluctuate within 1000-1500 cases per month, and some seasonal fluctuations are observed. There was a noticeable decrease in the number of events in 2020-2021. The indicators decreased to approximately 500 cases per month, a relatively stable low level during this period—a dramatic increase in the number of events. Peak values reach about 5000 cases per month, the most intense period for the entire time of observations. Stabilization at a high level in 2022-2024, the indicators fluctuate within 4000-4500 cases per month, and periodic fluctuations in intensity are noticeable. Towards the end of the graph, a sharp decrease in the number of incidents is observed.

According to the data analysis for dataset 2 (Fig. 2c), the initial period (early 2022) has a sharp increase in the number of refugees from almost zero to about 4.5 million in a very short period. The most rapid growth is observed in the first weeks. The peak value of about 8 million people is reached in mid-2022, with relative stabilization at a high level. Two noticeable "step" declines in the period 2022-2023, the first decline to about 6.5 million and the second decline to about 6 million. A relatively stable level of about 6 million people in the period 2023-2024, minor fluctuations in this range, and a tendency to slow levelling off. Descriptive statistics – quantitative characteristics of the data for the datasets are presented in Fig. 3.

Dataset 1 (Fig. 3) describes events involving civilian targets. On average, there were 5.1 events with 4.8 casualties. The data have significant variability (coefficient of variation 169% for events and 524% for casualties). There is a strong right-sided skew (5.0 for events and 21.6 for casualties), indicating the presence of extreme values. The maximum number of events in a single case is 139, and the maximum number of casualties is 774. In total, 1821 observations were recorded, with 9293 events and 8696 casualties.

Name	Dataset 1		Dataset 2		Dataset 3
	Events	Fatalities	Events	Fatalities	
Mean	5.10323998	4.77539813	54.40178004	38.39224412	6166432.35744681
Standard Error	0.20221286	0.58709346	1.77672284	3.28644634	54911.00236340
Median	2.00000000	1.00000000	7.00000000	1.00000000	5986904.50000000
Mode	1.00000000	0.00000000	1.00000000	0.00000000	4774703.00000000
Standard Deviation	8.62906508	25.05314283	99.65498854	184.33419396	1190442.16359304
Sample Variance	74.46076422	627.65996584	9931.11674087	33979.09506224	1417152544860.08520508
Correlation Coefficient	169.08993347	524.62940548	183.18332317	480.13393899	19.30520104
Kurdishness	48.64826732	584.93481263	14.18376347	155.72100323	13.59620562
Skew	5.00807400	21.63346330	3.04278732	10.96561649	-2.50926634
Minimum	138.00000000	774.00000000	756.00000000	3757.00000000	7821463.00000000
Maximum	1.00000000	0.00000000	1.00000000	0.00000000	85000.00000000
Sum	139.00000000	774.00000000	757.00000000	3757.00000000	7906463.00000000
Range	9293.00000000	8696.00000000	171148.00000000	120782.00000000	2898223208.00000000
Confidence Interval	1821.00000000	1821.00000000	3146.00000000	3146.00000000	470.00000000
	0.39659367	1.15144778	3.48365346	6.44379636	107902.04091277

Figure 3: Descriptive statistics results for three datasets

Dataset 2 (Fig. 3) describes events of political violence. The average number of events is much higher - 54.4 with 38.4 victims. There is also high variability (coefficient of variation 183% for events and 480% for victims). The right-sided asymmetry is less pronounced (3.0 for events and 11.0 for victims). The maximum values are significantly higher - 757 events and 3757 victims. A total of 3146 observations are of 171148 events and 120782 victims. Dataset 3 (Fig. 3) has a different nature of the data, as it describes the number of refugees and has completely different columns. The average value is about 6.17 million—relatively low variability (coefficient of variation 19.3%). The negative asymmetry (-2.5) indicates a left-sided distribution. The range of values is from 85,000 to 7.9 million—a total of 470 observations.

4. Experiments, results and discussion

4.1. Data pre-processing and presentation of results

The histogram in Fig. 4a for dataset 1 shows a sharply asymmetric distribution with a maximum of about 1500 cases at the beginning of the scale. Most events are concentrated in the range of 0-50 attacks. There is a sharp decrease in frequency as the number of events increases. Such a distribution may indicate that a single or small series of attacks occur most often. It has a pronounced right-sided asymmetry. The shape of the distribution resembles an exponential or Poisson distribution. A sharp decrease in frequency with an increase in the number of events is characteristic of an exponential distribution law. Since the data are discrete and represent the number of events, the Poisson distribution may be the most suitable approximation. The histogram in Fig. 4b for dataset 2 also shows an asymmetric distribution but with a higher peak (over 2000 cases). The distribution is more stretched along the X-axis (up to 800 events). The frequency of events gradually decreases with an increase in their number. It indicates more intense attacks on political infrastructure than on civilian infrastructure. It also exhibits right-sided asymmetry. As in the first case, the shape corresponds to an exponential distribution. It can be approximated by a gamma distribution, which is more flexible and can better account for the "heavy tail" of the distribution. Again, given the discreteness of the data, a Poisson distribution may be appropriate.

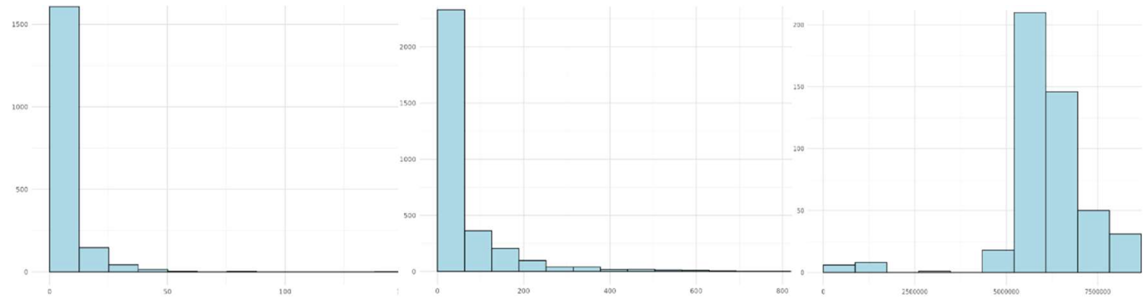


Figure 4: Histogram for three datasets (X is the number of events, and Y is the frequency)

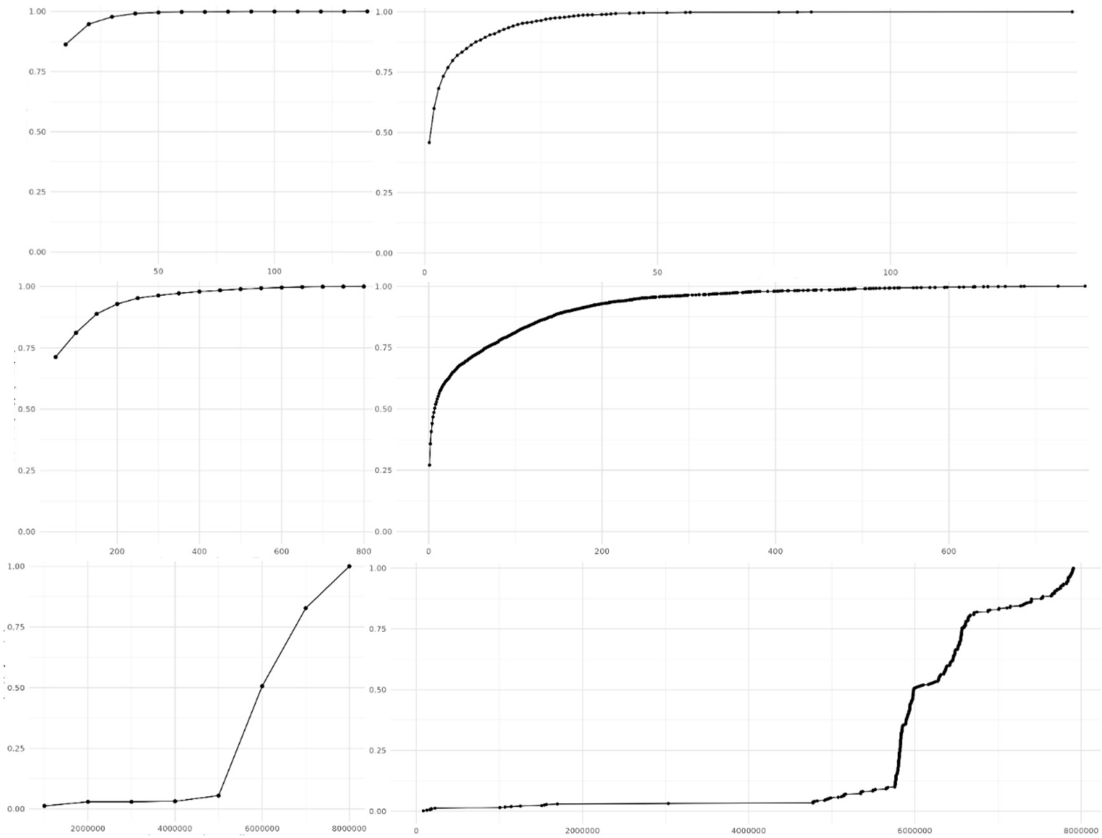


Figure 5: Cumulative for dataset 1-3 (left – histogram data and right) – interval percentage)

The histogram in Fig. 4c for dataset 3 has a fundamentally different distribution pattern - close to normal. The peak of the distribution falls in the range of about 5-6 million refugees. The distribution is more symmetrical compared to the previous histograms. There are a small number of cases with a small number of refugees (about 0-2.5 million). The bulk of the data is concentrated in the range of 5-7.5 million refugees. There is some asymmetry, but much less than in the previous cases. It can be approximated by a normal distribution or, to better account for the asymmetry, by a lognormal distribution. You can also consider the gamma distribution as an alternative since it works well with data that has a slight asymmetry.

Most events (shellings) are concentrated at the beginning of the scale (0-50). It has a very pronounced peak at the beginning, with about 1500 events. The cumulative curve increases rapidly and reaches a plateau, indicating that most events occur in the first intervals (Fig. 5a-b). After 50 events, only isolated cases are observed.

Similar distribution to the first dataset, but with a higher peak (about 2000 events). It also has an intense concentration of events at the beginning of the scale. The cumulative curve shows a similar dynamic of rapid growth with subsequent plateauing (Fig. 5c-d). The distribution is more stretched along the scale (up to 800 events). It differs significantly from the first two in the nature of the distribution (Fig. 5e-f). It has a normal distribution with a peak of about 5-6 million people. The cumulative curve has an S-shaped shape, which is typical of a normal distribution. The bulk of the data is concentrated in the range of 4-7 million. There is a small number of observations at the beginning of the scale (0-2 million).

4.2. Time series trend detection using smoothing methods

Smoothing methods are used to reduce the influence of the random component (random fluctuations) in time series. They provide an opportunity to obtain more "clean" values, consisting only of deterministic components. Some of the methods are aimed at highlighting only some elements, such as a trend. Smoothing methods can be conditionally divided into two classes, which are based on different approaches: analytical approach and algorithmic approach. The analytical approach is based on the selection of a mathematical function (for example, an exponential, polynomial or hyperbola) that best fits the data trend, determined visually. Then, the parameters of this function are estimated using mathematical or statistical methods, which form a model to describe the time series. The algorithmic approach focuses on calculating new values of the series using algorithms such as the moving average method, weighted average method, exponential smoothing method, and median smoothing method.

From 2018 to early 2022, the data (Fig. 6a) shows relatively stable low-level activity, where the moving average (red line) closely follows the actual data points (blue dots). There is a sharp spike in early 2022, after which the level remains elevated but gradually stabilizes. The moving average smoothes out the volatile spikes in the data, preserving the overall trend. The trend is clearly nonlinear, especially after 2022. The graph shows more stable activity from 2018-2020 with approximately 1000 events (Fig. 6b). Small decline in 2020-2021. As in the first graph, there will be a sharp increase in 2022. The trend is highly nonlinear, with several clear phases. Given the nonlinear nature of both data sets, a weighted moving average would be more appropriate than a simple moving average. A simple moving average tends to lag behind significant changes in the data, especially during sharp increases/decreases. The current simple moving average may not accurately reflect the actual dynamics of the processes, especially during rapid change periods.

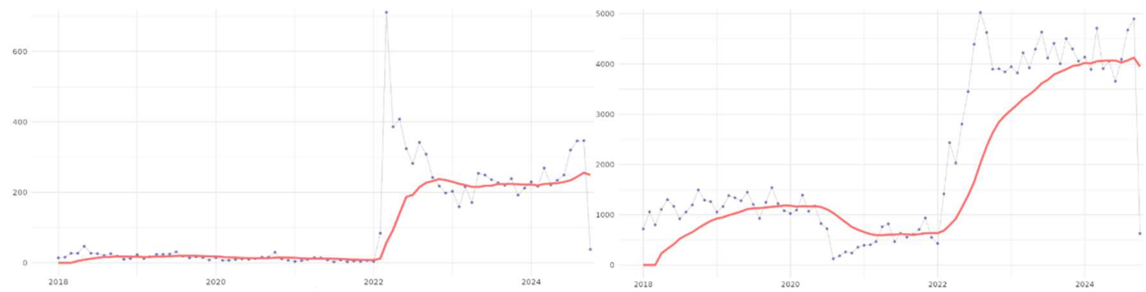


Figure 6: Graph of the results of applying the moving average method to the dataset of events (a) from shelling of civilian targets and (b) from shelling of political targets

Graph 1 in Fig. 7a (Civilian-targeted events) illustrates the weighted moving average (grey line), which better reflects the dynamics of changes compared to the simple moving average. In the period 2018-2021, the trend line reacts more sensitively to fluctuations in the data. After a sharp jump in

2022, the weighted average adapts more quickly to the new level of activity. The smoothing is less aggressive, which allows us to better track fundamental changes in the data. At the end of the period (2023-2024), the trend towards stabilization at the new level is better visible. Graph 2 in Fig. 7b (dataset 2) illustrates that by 2022, the weighted moving average more accurately reflects fluctuations in activity around the level of 1,000 events. The gradual decrease in activity in 2020-2021 is more noticeable. After the jump in 2022, the method better reflects the fundamental dynamics of growth. There is less lag from actual data during sharp changes. The formation of a new stable level in 2023-2024 is more clearly visible.

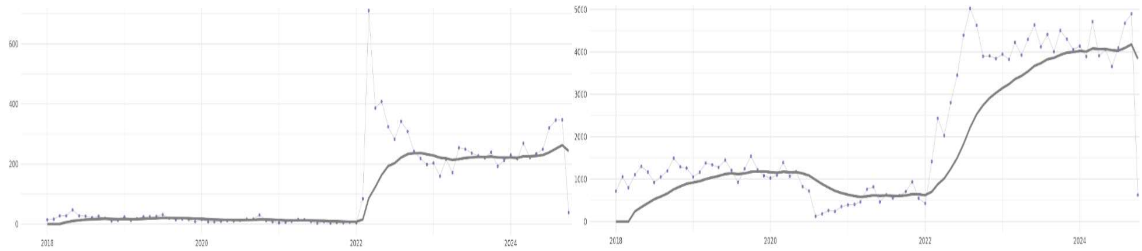


Figure 7: Graph of the results of applying the weighted moving average method

The least aggressive smoothing is shown in Fig. 8a-b for datasets 1-2, which better reflects short-term fluctuations. More responsive to data outliers. There is more detail in the process dynamics and less lag from real data.

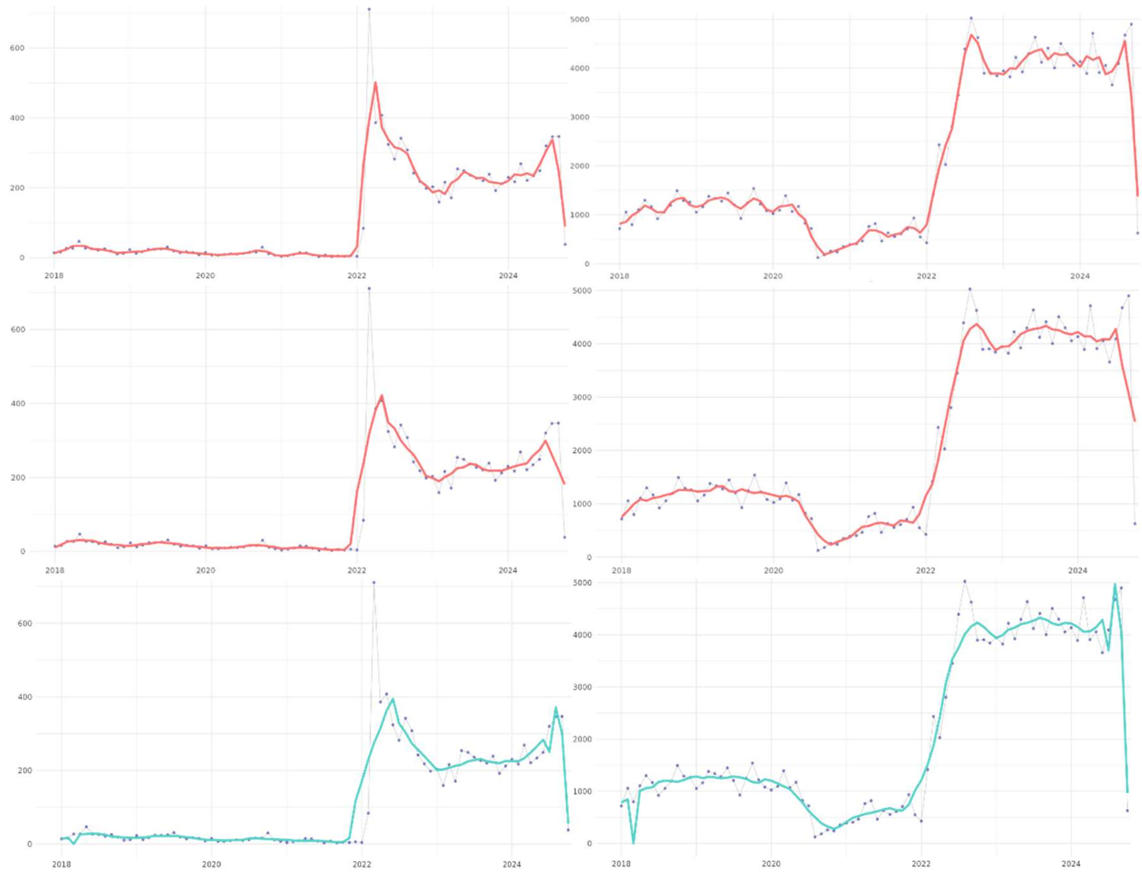


Figure 8: Simple moving average smoothing for w as 3 (1st row), 5 (2nd row), and 7 (3rd row)

Stronger smoothing compared to w=3 is shown in Figure 8c-d, which filters out random fluctuations better. There is more lag from real data. More clearly, it shows medium-term trends. Less sensitive to local extremes. The most aggressive smoothing is shown in Figure 8d-e, which best detects long-term trends. Significantly reduces the impact of outliers and the most considerable lag from real data. It is best suited for detecting a general trend. Comparative analysis of methods:

1. For events targeting civilians:
- All methods clearly show a sharp jump in 2022;
- Non-linear smoothing (w=7) best shows stabilization after the jump;
- For current monitoring, w=3 is better suited. For trend analysis - w=7;
2. For events targeting political targets:
- All methods reflect the overall dynamics well;
- Non-linear smoothing best shows the transition between different modes of activity;
- Larger values of w are better suited for analyzing long-term changes.

Therefore, for operational monitoring, linear smoothing is best suited when w = 3, for medium-term analysis – w = 5, for identifying long-term trends – nonlinear smoothing w = 7. The very low level of events (about 50) from 2018 to early 2022 is illustrated in Fig. 9a. A sharp peak in 2022 to about 700 events. Further decrease and stabilization at the level of 200-300 events during 2023-2024— a sharp drop in early 2025. A relatively stable period from 2018 to early 2022 with rates of about 1000-1500 events is shown in Fig. 9b. A sharp increase in rates in 2022 to about 4000-5000 events. Maintaining a high level (about 4000 events) throughout 2023-2024. A sharp drop in early 2025. Both graphs show a dramatic change in the situation starting in 2022, coinciding with the start of russia's full-scale invasion of Ukraine. It is noticeable that the number of events directed at political targets significantly exceeds the number of events directed at civilians. Median filtering (shown by the orange line) helps to smooth out short-term fluctuations and identify significant trends in the data.

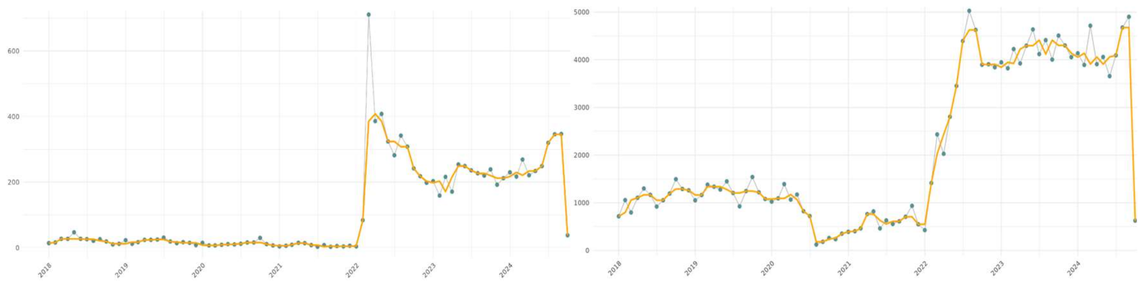


Figure 9: Median filtering result for dataset 1-2

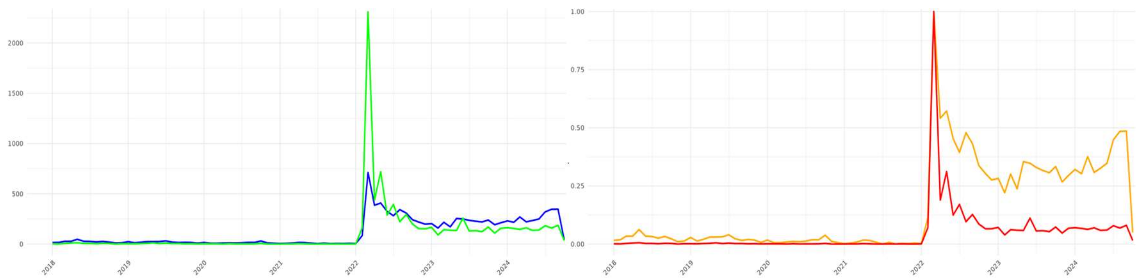


Figure 10: Result of time series normalization for the dataset on the shelling of civilian targets: (a) – original data (blue – events, green – fatalities) and (b) – normalized data (orange – minimax normalization of events, red – minimax normalization of fatalities)

The initial data in Fig. 10 shows a relatively low and stable level of both events and fatalities from 2018 to 2021, followed by a sharp spike in 2022 when over 2,000 incidents occurred. In normalized data (on a scale of 0-1), both metrics show almost zero activity until 2022. The spike in 2022 reaches the maximum normalization (1.0) for both events and fatalities. After 2022, the number of events (orange line) remains at a higher normalized level (around 0.3-0.5) compared to the number of fatalities (red line), which decreases to around 0.1. It suggests that although attacks continue, they have become relatively less lethal.

Consecutive low-level political events (blue line) in 2018-2021 are shown in Fig. 11. Sharp increase in both indicators since 2022—more erratic patterns of fatalities (green line) with extreme spikes. In normalized data: Events (orange line) show higher, more consistent normalized values (0.75-1.0) after 2022. Fatalities (red line) show more variation but generally lower normalized values. The relationship between events and fatalities is less intense than in the infrastructure dataset.

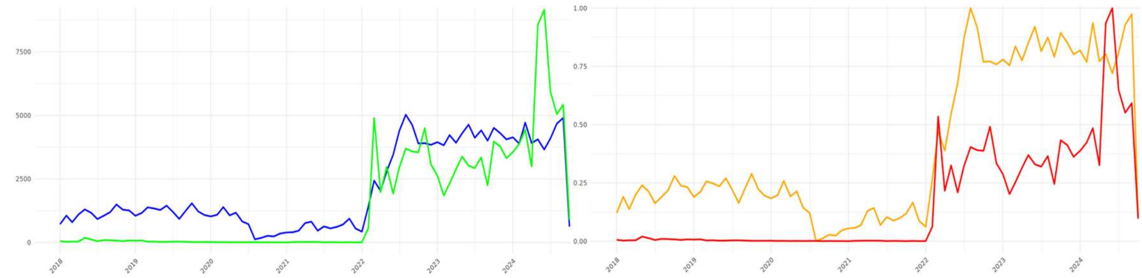


Figure 11: Result of time series normalization for the dataset on the shelling of political targets: (a) – original data (blue – events, green – fatalities) and (b) – normalized data (orange – minimax normalization of events, red – minimax normalization of deaths)

Table 4
Time smoothing efficiency criteria

Smoothing method	Dataset	The correlation coefficient between input and smoothed series	Number of turning points in the output series	Number of turning points in the smoothed series
	1	0.7503	42	17
Moving Average	2	0.861	47	15
Weighted	1	0.7819	42	19
Moving Average	2	0.8767	47	21
Linear (w = 3)	1	0.9432	42	24
	2	0.9872	47	31
Linear (w = 5)	1	0.9137	42	21
	2	0.9698	47	27
Linear (w = 7)	1	0.9035	42	20
	2	0.9813	47	23
	1	0.9633	42	11
Median filtering	2	0.9932	47	15

Dataset 1

N=3

Row	Weights	Divisor	Smoothed_Values
1	5, 2, -1	6	12.500
2	1, 1, 1	3	19.000
n	-1, 2, 5	6	89.667

N=5

Row	Weights	Divisor	Smoothed.Values
1	3, 2, 1, 0, -1	5	10.8
2	4, 3, 2, 1, 0	10	18.5
3	1, 1, 1, 1, 1	10	26.2
n-1	0, 1, 2, 3, 4	10	220.5
n	-1, 0, 1, 2, 3	5	181.0

Dataset 2

N=3

Row	Weights	Divisor	Smoothed_Values
1	5, 2, -1	6	817
2	1, 1, 1	3	857
n	-1, 2, 5	6	1376

N=5

Row	Weights	Divisor	Smoothed.Values
1	3, 2, 1, 0, -1	5	752.4
2	4, 3, 2, 1, 0	10	873.9
3	1, 1, 1, 1, 1	10	995.4
n-1	0, 1, 2, 3, 4	10	3065.3
n	-1, 0, 1, 2, 3	5	2540.0

Dataset 3

N=3

Row	Weights	Divisor	Smoothed_Values
1	5, 2, -1	6	85983.33
2	1, 1, 1	3	126333.33
n	-1, 2, 5	6	5980739.67

N=5

Row	Weights	Divisor	Smoothed.Values
1	3, 2, 1, 0, -1	5	99281.2
2	4, 3, 2, 1, 0	10	123700.0
3	1, 1, 1, 1, 1	10	148118.8
n-1	0, 1, 2, 3, 4	10	5985068.3
n	-1, 0, 1, 2, 3	5	5983876.6

Figure 12: Smoothing with Kendall formulas for datasets 1-3 at N=3 and N=5

A high correlation coefficient (≥ 0.7) indicates that the smoothed series well preserves the general trend of the original series (Table 4). At the same time, smoothing removes local fluctuations (noise) but does not destroy the structure of the data. Turning points are local maxima and minima in the series. A significant reduction in the number of turning points in the smoothed series indicates that smoothing effectively eliminates short-term fluctuations (noise).

Dataset 1, N=7

Row	Weights	Divisor	Smoothed.Values
1	13, 10, 7, 4, 1, -2, -5	28	17.929
2	5, 4, 3, 2, 1, 0, -1	14	20.714
3	7, 6, 5, 4, 3, 2, 1	28	23.500
4	1, 1, 1, 1, 1, 1, 1	7	26.286
n-2	1, 2, 3, 4, 5, 6, 7	28	242.643
n-1	-1, 0, 1, 2, 3, 4, 5	14	234.571
n	-5, -2, 1, 4, 7, 10, 13	28	226.500

Dataset 1, N=9

Row	Weights	Divisor	Smoothed.Values
1	17, 14, 11, 8, 5, 2, -1, -4, -7	45	21.600
2	56, 47, 38, 29, 20, 11, 2, -7, -16	180	22.617
3	22, 19, 16, 13, 10, 7, 4, 1, -2	90	23.633
4	32, 29, 26, 23, 20, 17, 14, 11, 8	180	24.650
5	1, 1, 1, 1, 1, 1, 1, 1, 1	9	25.667
n-3	8, 11, 14, 17, 20, 23, 26, 29, 32	180	246.567
n-2	-2, 1, 4, 7, 10, 13, 16, 19, 22	90	244.133
n-1	-16, -7, 2, 11, 20, 29, 38, 47, 56	180	241.700
n	-7, -4, -1, 2, 5, 8, 11, 14, 17	45	239.267

Dataset 2, N=7

Row	Weights	Divisor	Smoothed.Values
1	13, 10, 7, 4, 1, -2, -5	28	866.250
2	5, 4, 3, 2, 1, 0, -1	14	913.929
3	7, 6, 5, 4, 3, 2, 1	28	961.607
4	1, 1, 1, 1, 1, 1, 1	7	1009.286
n-2	1, 2, 3, 4, 5, 6, 7	28	3447.786
n-1	-1, 0, 1, 2, 3, 4, 5	14	3192.714
n	-5, -2, 1, 4, 7, 10, 13	28	2937.643

Dataset 2, N=9

Row	Weights	Divisor	Smoothed.Values
1	17, 14, 11, 8, 5, 2, -1, -4, -7	45	887.578
2	56, 47, 38, 29, 20, 11, 2, -7, -16	180	924.378
3	22, 19, 16, 13, 10, 7, 4, 1, -2	90	961.178
4	32, 29, 26, 23, 20, 17, 14, 11, 8	180	997.978
5	1, 1, 1, 1, 1, 1, 1, 1, 1	9	1034.778
n-3	8, 11, 14, 17, 20, 23, 26, 29, 32	180	3654.022
n-2	-2, 1, 4, 7, 10, 13, 16, 19, 22	90	3471.822
n-1	-16, -7, 2, 11, 20, 29, 38, 47, 56	180	3289.622
n	-7, -4, -1, 2, 5, 8, 11, 14, 17	45	3107.422

Dataset 3, N=7

Row	Weights	Divisor	Smoothed.Values
1	13, 10, 7, 4, 1, -2, -5	28	-36773.71
2	5, 4, 3, 2, 1, 0, -1	14	69407.14
3	7, 6, 5, 4, 3, 2, 1	28	175588.00
4	1, 1, 1, 1, 1, 1, 1	7	281768.86
n-2	1, 2, 3, 4, 5, 6, 7	28	5985575.82
n-1	-1, 0, 1, 2, 3, 4, 5	14	5985762.07
n	-5, -2, 1, 4, 7, 10, 13	28	5985948.32

Dataset 3, N=9

Row	Weights	Divisor	Smoothed.Values
1	17, 14, 11, 8, 5, 2, -1, -4, -7	45	-120316.33
2	56, 47, 38, 29, 20, 11, 2, -7, -16	180	25974.25
3	22, 19, 16, 13, 10, 7, 4, 1, -2	90	172264.83
4	32, 29, 26, 23, 20, 17, 14, 11, 8	180	318555.42
5	1, 1, 1, 1, 1, 1, 1, 1, 1	9	464846.00
n-3	8, 11, 14, 17, 20, 23, 26, 29, 32	180	5985010.89
n-2	-2, 1, 4, 7, 10, 13, 16, 19, 22	90	5985643.34
n-1	-16, -7, 2, 11, 20, 29, 38, 47, 56	180	5986275.79
n	-7, -4, -1, 2, 5, 8, 11, 14, 17	45	5986908.24

Figure 13: Smoothing with Kendall formulas for datasets 1-3 at N=7 and N=8

Dataset 1, N=11

Row	Weights	Divisor	Smoothed.Values
1	7, 6, 5, 4, 3, 2, 1, 0, -1, -2, -3	22	25.636
2	15, 13, 11, 9, 7, 5, 3, 1, -1, -3, -5	55	25.236
3	25, 22, 19, 16, 13, 10, 7, 4, 1, -2, -5	110	24.836
4	10, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0	55	24.436
5	15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5	110	24.036
6	1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1	11	23.636
n-4	5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15	110	244.955
n-3	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10	55	246.000
n-2	-5, -2, 1, 4, 7, 10, 13, 16, 19, 22, 25	110	247.045
n-1	-5, -3, -1, 1, 3, 5, 7, 9, 11, 13, 15	55	248.091
n	-3, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7	22	249.136

Dataset 1, N=13

Row	Weights	Divisor	Smoothed.Values
1	25, 22, 19, 16, 13, 10, 7, 4, 1, -2, -5, -8, -11	91	26.187
2	44, 39, 34, 29, 24, 19, 14, 9, 4, -1, -6, -11, -16	182	25.604
3	19, 17, 15, 13, 11, 9, 7, 5, 3, 1, -1, -3, -5	91	25.022
4	32, 29, 26, 23, 20, 17, 14, 11, 8, 5, 2, -1, -4	182	24.440
5	13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1	91	23.857
6	20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8	182	23.275
7	1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1	13	22.692
n-5	8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20	182	241.758
n-4	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	91	243.978
n-3	-4, -1, 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32	182	246.198
n-2	-5, -3, -1, 1, 3, 5, 7, 9, 11, 13, 15, 17, 19	91	248.418
n-1	-16, -11, -6, -1, 4, 9, 14, 19, 24, 29, 34, 39, 44	182	250.637
n	-11, -8, -5, -2, 1, 4, 7, 10, 13, 16, 19, 22, 25	91	252.857

Dataset 2, N=11

Row	Weights	Divisor	Smoothed.Values
1	7, 6, 5, 4, 3, 2, 1, 0, -1, -2, -3	22	857.773
2	15, 13, 11, 9, 7, 5, 3, 1, -1, -3, -5	55	906.182
3	25, 22, 19, 16, 13, 10, 7, 4, 1, -2, -5	110	954.591
4	10, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0	55	1003.000
5	15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5	110	1051.409
6	1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1	11	1099.818
n-4	5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15	110	3763.191
n-3	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10	55	3642.927
n-2	-5, -2, 1, 4, 7, 10, 13, 16, 19, 22, 25	110	3522.664
n-1	-5, -3, -1, 1, 3, 5, 7, 9, 11, 13, 15	55	3402.400
n	-3, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7	22	3282.136

Dataset 2, N=13

Row	Weights	Divisor	Smoothed.Values
1	25, 22, 19, 16, 13, 10, 7, 4, 1, -2, -5, -8, -11	91	915.791
2	44, 39, 34, 29, 24, 19, 14, 9, 4, -1, -6, -11, -16	182	947.929
3	19, 17, 15, 13, 11, 9, 7, 5, 3, 1, -1, -3, -5	91	980.066
4	32, 29, 26, 23, 20, 17, 14, 11, 8, 5, 2, -1, -4	182	1012.203
5	13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1	91	1044.341
6	20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8	182	1076.478
7	1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1	13	1108.615
n-5	8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20	182	3858.824
n-4	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	91	3754.110
n-3	-4, -1, 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32	182	3649.396
n-2	-5, -3, -1, 1, 3, 5, 7, 9, 11, 13, 15, 17, 19	91	3544.681
n-1	-16, -11, -6, -1, 4, 9, 14, 19, 24, 29, 34, 39, 44	182	3439.967
n	-11, -8, -5, -2, 1, 4, 7, 10, 13, 16, 19, 22, 25	91	3335.253

Dataset 3, N=11

Row	Weights	Divisor	Smoothed.Values
1	7, 6, 5, 4, 3, 2, 1, 0, -1, -2, -3	22	-148206.545
2	15, 13, 11, 9, 7, 5, 3, 1, -1, -3, -5	55	7721.982
3	25, 22, 19, 16, 13, 10, 7, 4, 1, -2, -5	110	163650.509
4	10, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0	55	319579.036
5	15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5	110	475507.564
6	1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1	11	631436.091
n-4	5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15	110	5984296.055
n-3	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10	55	5984992.109
n-2	-5, -2, 1, 4, 7, 10, 13, 16, 19, 22, 25	110	5985688.164
n-1	-5, -3, -1, 1, 3, 5, 7, 9, 11, 13, 15	55	5986384.218
n	-3, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7	22	5987080.273

Dataset 3, N=13

Row	Weights	Divisor	Smoothed.Values
1	25, 22, 19, 16, 13, 10, 7, 4, 1, -2, -5, -8, -11	91	-127001.26
2	44, 39, 34, 29, 24, 19, 14, 9, 4, -1, -6, -11, -16	182	23056.59
3	19, 17, 15, 13, 11, 9, 7, 5, 3, 1, -1, -3, -5	91	173114.44
4	32, 29, 26, 23, 20, 17, 14, 11, 8, 5, 2, -1, -4	182	323172.29
5	13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1	91	473230.14
6	20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8	182	623287.99
7	1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1	13	773345.85
n-5	8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20	182	5983236.09
n-4	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	91	5984172.03
n-3	-4, -1, 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32	182	5985107.97
n-2	-5, -3, -1, 1, 3, 5, 7, 9, 11, 13, 15, 17, 19	91	5986043.91
n-1	-16, -11, -6, -1, 4, 9, 14, 19, 24, 29, 34, 39, 44	182	5986979.85
n	-11, -8, -5, -2, 1, 4, 7, 10, 13, 16, 19, 22, 25	91	5987915.79

Figure 14: Smoothing with Kendall formulas for datasets 1-3 at N=11 and N=13

1

Row	Weights	Divisor	Smoothed.Values
1	29, 26, 23, 20, 17, 14, 11, 8, 5, 2, -1, -4, -7, -10, -13	120	26.850
2	91, 82, 73, 64, 55, 46, 37, 28, 19, 10, 1, -8, -17, -26, -35	420	26.100
3	161, 146, 131, 116, 101, 86, 71, 56, 41, 26, 11, -4, -19, -34, -49	840	25.350
4	35, 32, 29, 26, 23, 20, 17, 14, 11, 8, 5, 2, -1, -4, -7	210	24.600
5	119, 110, 101, 92, 83, 74, 65, 56, 47, 38, 29, 20, 11, 2, -7	840	23.850
6	49, 46, 43, 40, 37, 34, 31, 28, 25, 22, 19, 16, 13, 10, 7	420	23.100
7	77, 74, 71, 68, 65, 62, 59, 56, 53, 50, 47, 44, 41, 38, 35	840	22.350
8	1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1	15	21.600
n-6	35, 38, 41, 44, 47, 50, 53, 56, 59, 62, 65, 68, 71, 74, 77	840	239.575
n-5	7, 10, 13, 16, 19, 22, 25, 28, 31, 34, 37, 40, 43, 46, 49	420	241.750
n-4	-7, 2, 11, 20, 29, 38, 47, 56, 65, 74, 83, 92, 101, 110, 119	840	243.925
n-3	-7, -4, -1, 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35	210	246.100
n-2	-49, -34, -19, -4, 11, 26, 41, 56, 71, 86, 101, 116, 131, 146, 161	840	248.275
n-1	-35, -26, -17, -8, 1, 10, 19, 28, 37, 46, 55, 64, 73, 82, 91	420	250.450
n	-13, -10, -7, -4, -1, 2, 5, 8, 11, 14, 17, 20, 23, 26, 29	120	252.625

2

Row	Weights	Divisor	Smoothed.Values
1	29, 26, 23, 20, 17, 14, 11, 8, 5, 2, -1, -4, -7, -10, -13	120	928.267
2	91, 82, 73, 64, 55, 46, 37, 28, 19, 10, 1, -8, -17, -26, -35	420	957.152
3	161, 146, 131, 116, 101, 86, 71, 56, 41, 26, 11, -4, -19, -34, -49	840	986.038
4	35, 32, 29, 26, 23, 20, 17, 14, 11, 8, 5, 2, -1, -4, -7	210	1014.924
5	119, 110, 101, 92, 83, 74, 65, 56, 47, 38, 29, 20, 11, 2, -7	840	1043.810
6	49, 46, 43, 40, 37, 34, 31, 28, 25, 22, 19, 16, 13, 10, 7	420	1072.695
7	77, 74, 71, 68, 65, 62, 59, 56, 53, 50, 47, 44, 41, 38, 35	840	1101.581
8	1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1	15	1130.467
n-6	35, 38, 41, 44, 47, 50, 53, 56, 59, 62, 65, 68, 71, 74, 77	840	3915.990
n-5	7, 10, 13, 16, 19, 22, 25, 28, 31, 34, 37, 40, 43, 46, 49	420	3835.848
n-4	-7, 2, 11, 20, 29, 38, 47, 56, 65, 74, 83, 92, 101, 110, 119	840	3755.705
n-3	-7, -4, -1, 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35	210	3675.562
n-2	-49, -34, -19, -4, 11, 26, 41, 56, 71, 86, 101, 116, 131, 146, 161	840	3595.419
n-1	-35, -26, -17, -8, 1, 10, 19, 28, 37, 46, 55, 64, 73, 82, 91	420	3515.276
n	-13, -10, -7, -4, -1, 2, 5, 8, 11, 14, 17, 20, 23, 26, 29	120	3435.133

3

Row	Weights	Divisor	Smoothed.Values
1	29, 26, 23, 20, 17, 14, 11, 8, 5, 2, -1, -4, -7, -10, -13	120	-231175.14
2	91, 82, 73, 64, 55, 46, 37, 28, 19, 10, 1, -8, -17, -26, -35	420	-57354.87
3	161, 146, 131, 116, 101, 86, 71, 56, 41, 26, 11, -4, -19, -34, -49	840	116465.39
4	35, 32, 29, 26, 23, 20, 17, 14, 11, 8, 5, 2, -1, -4, -7	210	290285.66
5	119, 110, 101, 92, 83, 74, 65, 56, 47, 38, 29, 20, 11, 2, -7	840	464105.93
6	49, 46, 43, 40, 37, 34, 31, 28, 25, 22, 19, 16, 13, 10, 7	420	637926.20
7	77, 74, 71, 68, 65, 62, 59, 56, 53, 50, 47, 44, 41, 38, 35	840	811746.46
8	1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1	15	985566.73
n-6	35, 38, 41, 44, 47, 50, 53, 56, 59, 62, 65, 68, 71, 74, 77	840	5982174.91
n-5	7, 10, 13, 16, 19, 22, 25, 28, 31, 34, 37, 40, 43, 46, 49	420	5983175.95
n-4	-7, 2, 11, 20, 29, 38, 47, 56, 65, 74, 83, 92, 101, 110, 119	840	5984176.99
n-3	-7, -4, -1, 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35	210	5985178.02
n-2	-49, -34, -19, -4, 11, 26, 41, 56, 71, 86, 101, 116, 131, 146, 161	840	5986179.06
n-1	-35, -26, -17, -8, 1, 10, 19, 28, 37, 46, 55, 64, 73, 82, 91	420	5987180.10
n	-13, -10, -7, -4, -1, 2, 5, 8, 11, 14, 17, 20, 23, 26, 29	120	5988181.14

Figure 15: Smoothing with Kendall formulas for datasets 1-3 at N=15

Accordingly, the original series has more "noise" or "chaotic changes", which can often be insignificant for analysis. A decrease in the number of turning points is a sign that the smoothed series shows the primary trend but with less detail. Fig. 12 shows the results of smoothing using the Kendall formulas. The data show a sharp peak of activity around point 55 on the time axis, reaching approximately 700 attacks. After the peak, there is a stabilization at around 200-250 attacks. Method B provides a smoother visualization of the trend (Fig. 16). Both methods show a similar overall picture, but Method B better reflects long-term trends.

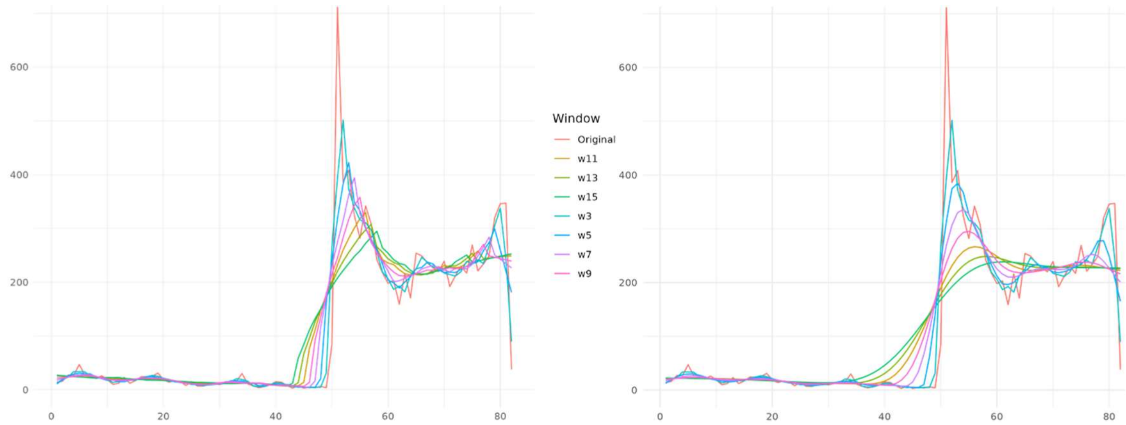


Figure 16: Based on Kendall's formulas (a) Method A: direct smoothing with different window sizes and (b) method B: sequential smoothing with increasing window size for dataset 1 - shelling of Ukrainian civilian infrastructure

There is a significant increase in the number of attacks starting from point 50 on the time axis. The peak value reaches about 5000 attacks. Method B (sequential smoothing) shows a smoother curve compared to method A (Fig. 17). Larger window sizes (w11-w15) give a smoother result but may lose important local features of the data. At the end of the period, there is a sharp decline in activity. The graph in Fig. 18 shows a rapid increase in the number of refugees at the beginning of the period (up to point 100). The maximum value reaches about 8 million people. Two noticeable declines are observed (around points 200 and 300). Both smoothing methods give very similar results, which indicates relatively “clean” initial data. At the end of the period, there is a stabilization at the level of about 6 million people.

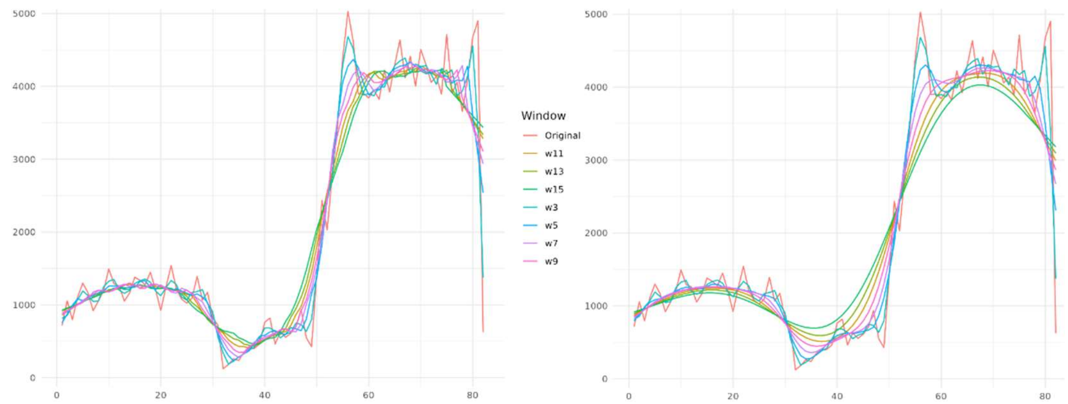


Figure 17: Based on Kendall's formulas (a) Method A: direct smoothing with different window sizes and (b) method B: sequential smoothing with increasing window size for dataset 2 - shelling of political infrastructure of Ukraine

In Fig. 19, for dataset 1, a strong positive correlation is observed between all smoothing windows (all values > 0.82). The strongest correlation is observed between neighbouring smoothing windows (for example, Window_5 and Window_7 correlate 0.9884). The correlation gradually decreases with increasing differences in the size of the smoothing windows. The original series has the strongest correlation with smaller smoothing windows (Window_3: 0.9432) and the weakest with larger ones (Window_15: 0.8205). There is a very high positive correlation between all smoothing windows (all values > 0.94) in Fig. 19 for dataset 2. Correlation values are generally higher than for civil events. There is also a trend towards a stronger correlation between neighbouring windows. The original series has a consistently high correlation with all smoothing windows (from 0.9416 to 0.9872). There is an extremely high positive correlation between all smoothing windows (all values > 0.99) in Fig. 19 for dataset 3—the highest correlation values among all three matrices. There is practically no difference between the correlations of neighbouring and distant smoothing windows. The original series has a very high correlation with all smoothing windows (all values > 0.99).

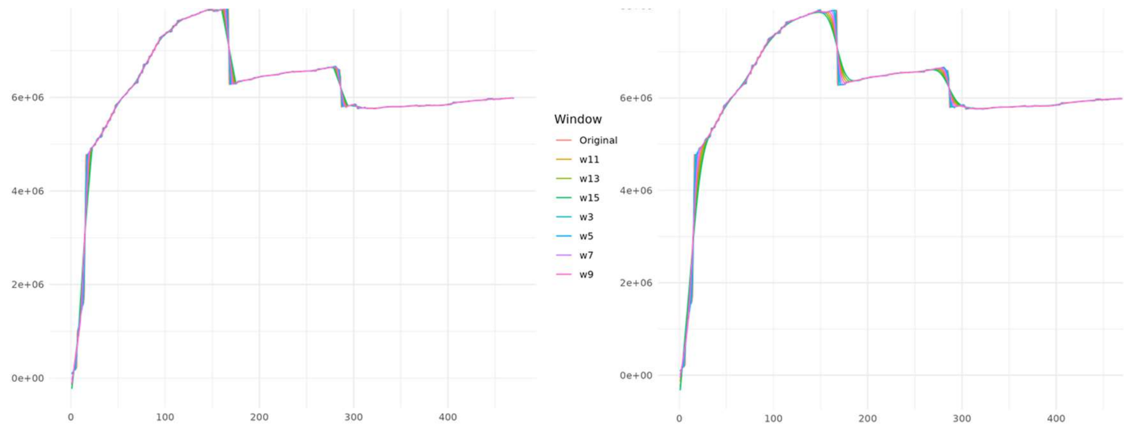


Figure 18: Based on Kendall's formulas (a) Method A: direct smoothing with different window sizes and (b) method B: sequential smoothing with increasing window size for dataset 3 - Ukrainian refugees abroad

Dataset 1

	Original	Window_3	Window_5	Window_7	Window_9	Window_11	Window_13	Window_15
Original	1.0000	0.9432	0.9137	0.8842	0.8650	0.8533	0.8386	0.8205
Window_3	0.9432	1.0000	0.9781	0.9532	0.9360	0.9206	0.9048	0.8870
Window_5	0.9137	0.9781	1.0000	0.9884	0.9755	0.9603	0.9444	0.9283
Window_7	0.8842	0.9532	0.9884	1.0000	0.9930	0.9816	0.9678	0.9534
Window_9	0.8650	0.9360	0.9755	0.9930	1.0000	0.9943	0.9846	0.9725
Window_11	0.8533	0.9206	0.9603	0.9816	0.9943	1.0000	0.9954	0.9872
Window_13	0.8386	0.9048	0.9444	0.9678	0.9846	0.9954	1.0000	0.9962
Window_15	0.8205	0.8870	0.9283	0.9534	0.9725	0.9872	0.9962	1.0000

Dataset 2

	Original	Window_3	Window_5	Window_7	Window_9	Window_11	Window_13	Window_15
Original	1.0000	0.9872	0.9698	0.9609	0.9579	0.9535	0.9484	0.9416
Window_3	0.9872	1.0000	0.9913	0.9843	0.9804	0.9756	0.9702	0.9639
Window_5	0.9698	0.9913	1.0000	0.9973	0.9943	0.9899	0.9854	0.9802
Window_7	0.9609	0.9843	0.9973	1.0000	0.9983	0.9950	0.9915	0.9875
Window_9	0.9579	0.9804	0.9943	0.9983	1.0000	0.9985	0.9962	0.9930
Window_11	0.9535	0.9756	0.9899	0.9950	0.9985	1.0000	0.9990	0.9969
Window_13	0.9484	0.9702	0.9854	0.9915	0.9962	0.9990	1.0000	0.9991
Window_15	0.9416	0.9639	0.9802	0.9875	0.9930	0.9969	0.9991	1.0000

Dataset 3

	Original	Window_3	Window_5	Window_7	Window_9	Window_11	Window_13	Window_15
Original	1.0000	0.9990	0.9975	0.9959	0.9945	0.9931	0.9918	0.9905
Window_3	0.9990	1.0000	0.9993	0.9982	0.9969	0.9956	0.9944	0.9931
Window_5	0.9975	0.9993	1.0000	0.9995	0.9986	0.9975	0.9965	0.9953
Window_7	0.9959	0.9982	0.9995	1.0000	0.9997	0.9989	0.9981	0.9971
Window_9	0.9945	0.9969	0.9986	0.9997	1.0000	0.9997	0.9992	0.9983
Window_11	0.9931	0.9956	0.9975	0.9989	0.9997	1.0000	0.9998	0.9992
Window_13	0.9918	0.9944	0.9965	0.9981	0.9992	0.9998	1.0000	0.9998
Window_15	0.9905	0.9931	0.9953	0.9971	0.9983	0.9992	0.9998	1.0000

Figure 19: Correlation matrix

In the diagram in Fig. 20a, the initial number of points is smaller for dataset 1 (about 23-24). Method A shows unstable behaviour with local peaks and troughs. Method B shows a constant decrease in the number of points. The most significant difference between the methods is observed at medium window sizes (9-11). At the maximum window size (15), both methods show the smallest number of turning points. In the diagram in Fig. 20b, the highest number of turning points is observed for dataset 2 (about 30) at the smallest window size (3). Method A shows a smoother decrease in the number of points and stabilizes at about 15-20 points. Method B shows a sharp drop at the beginning and stabilizes at about 4 points. Both methods show a tendency to decrease the number of turning points with increasing window size. At large window sizes (11-15), the difference between the methods becomes more pronounced.

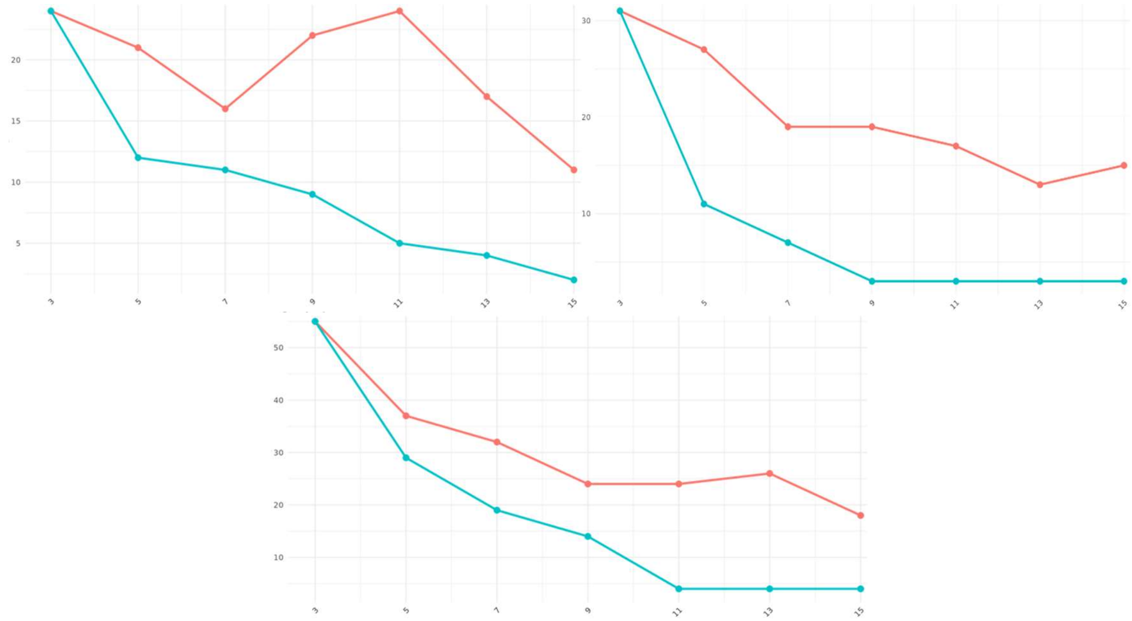


Figure 20: Diagram of the dependence of the number of turning points (Y) on the window size (X) for datasets 1-3 (red – method A, green – method B)

The highest initial number of turning points for dataset 3 (more than 50) among all three plots in Fig. 20. Both methods show a similar downward trend. Method A retains more turning points at all window sizes. After window size 11, both methods show relative stability. The difference between the methods remains almost constant at large window sizes.

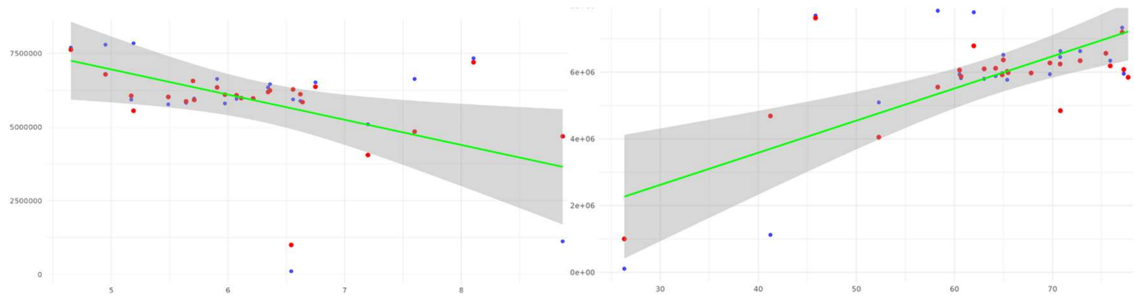


Figure 21: Correlation between shelling of civilian/political targets and the number of refugees using the Kendel method (X is the average number of shelling per month, and Y is the average number of refugees), where blue dots are unsmoothed data and red dots are smoothed data

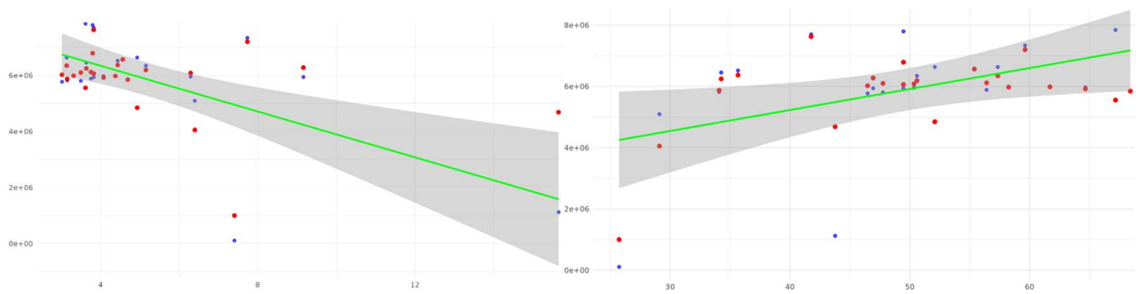


Figure 22: Correlation between fatalities caused by the shelling of civilian/political targets and the number of refugees using the Kendel method (X is the average number of deaths and Y is the average number of refugees), where blue dots – unsmoothed data and red – smoothed data

According to Fig. 21a, the correlation between civilian shelling and the number of refugees according to the Kendel method has a weak linear relationship. The correlation coefficient of the modulus < 0.5, the coefficient of determination is less than 25% (Table 5).

Table 5
Correlation coefficients and parameters according to the data from Fig. 21-22

Name	21(a)	21(b)	22(a)	22(b)
Correlation coefficient (smoothed)	-0.327	0.651	-0.339	0.439
Coefficient of determination (R-squared),%	10.66	42.43	11.5	19.31
t	-1.6204	4.0267	-1.6912	2.2947
df	22	22	22	22
p-value	0.1194	0.0005651	0.1049	0.03166
95% confidence interval	-0.64496771	0.3363726	-0.6532079	0.04382398
	0.08853236	0.8352928	0.0743871	0.71593272
Sample estimates	-0.3265258	0.6513822	-0.3391878	0.4394519

According to Fig. 21b, the correlation between the shelling of political targets and the number of refugees, according to the Kendel method, has a linear relationship of medium strength. The correlation coefficient of the modulus is less than 0.7 but more than 0.5, and the coefficient of determination is less than 50% but more than 25%. According to Fig. 22a, the correlation between the fatal cases provoked by the shelling of civilian targets and the number of refugees, according to the Kendel method, has a weak linear relationship. The correlation coefficient of the modulus is < 0.5, and the coefficient of determination is less than 25%. According to Fig. 22b, the correlation between the fatal cases provoked by the shelling of political targets and the number of refugees, according to the Kendel method, has a weak linear relationship. The correlation coefficient of the modulus is < 0.5, and the coefficient of determination is less than 25%.

The correlation ratio is 0.581, indicating a moderate relationship between the variables (Fig. 23a). A scattered nature of the points around the midline is observed. The group variance (89068252167.606) is significantly smaller than the total variance (153339421323248.07), confirming the presence of a moderate relationship. Most events are concentrated in the range of 5-7 events, which may indicate a specific pattern in the frequency of shelling (Table 6).

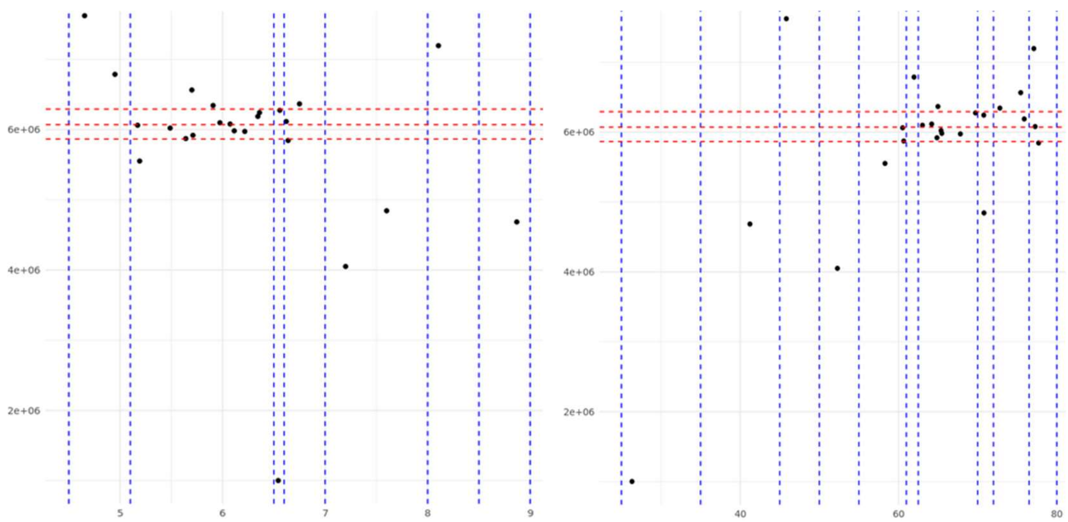


Figure 23: Division of the correlation field of the group between shelling of civilian/political targets and the number of refugees

The high correlation ratio of 0.935 indicates a powerful relationship between the variables (Fig. 23b). The points are located more densely relative to the mean line. The group variance (143336071722265.27) is close to the total (153339421323248.07), which confirms the strong relationship. There is a clear trend of an increase in the number of refugees with an increase in the number of attacks on political targets.

Table 6
Correlation coefficients and parameters according to data from Fig. 23-24

Name	23(a)	23(b)	24(a)	24(b)
Group variance	890682522167.686	1433687172265.27	1212784734739.6	1797241107307.35
Total variance	1533394213248.07	1533394213248.07	1533394213248.07	1533394213248.07
Correlation relationship	0.581	0.935	0.791	0.976

The correlation ratio of 0.791 indicates a strong relationship (Fig. 24a). The points have a noticeable spread but retain the general trend. The group variance (1212784734739.6) is significantly smaller than the total, which indicates the presence of other influencing factors. The main concentration of events is observed in the range of 4-8 fatal cases.

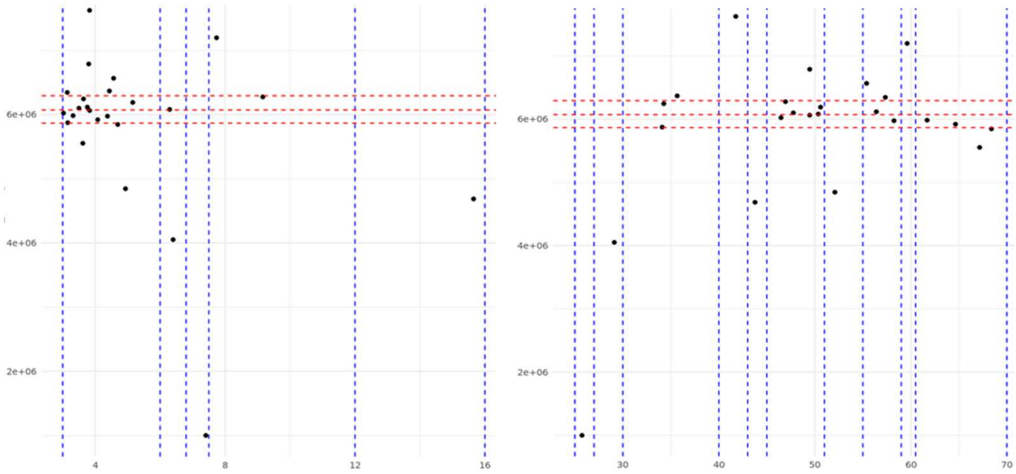


Figure 24: Dividing the correlation field between fatalities caused by shelling of civilian/political targets and the number of refugees

The very high correlation ratio of 0.976 indicates an almost functional relationship (Fig. 24b). The points are located most densely to the midline compared to other graphs. The group variance (149724110730.35) is nearly equal to the total, which confirms a powerful relationship. A direct relationship between the fatalities number and the refugees number is clearly visible.

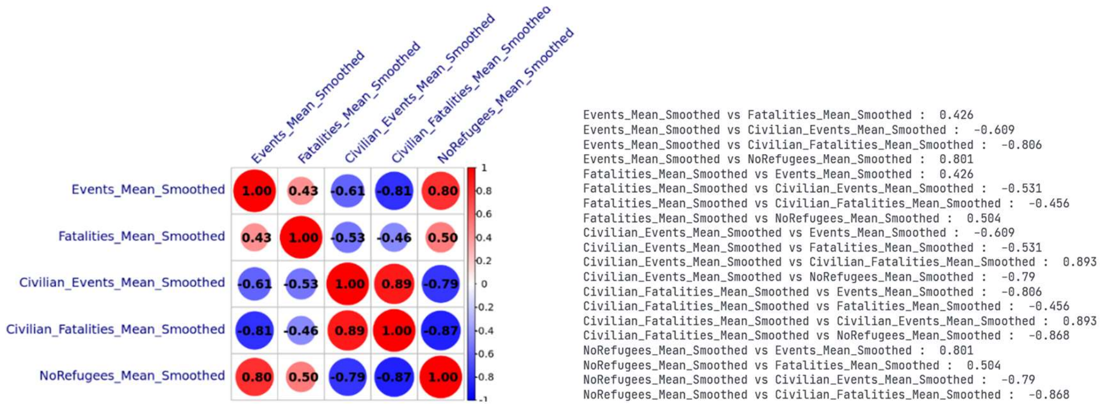
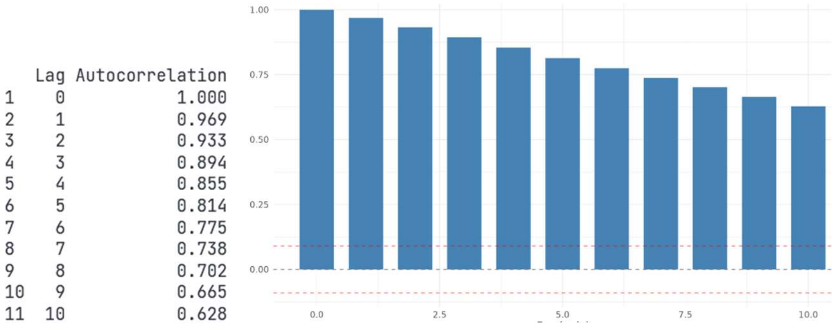


Figure 25: Correlation matrix

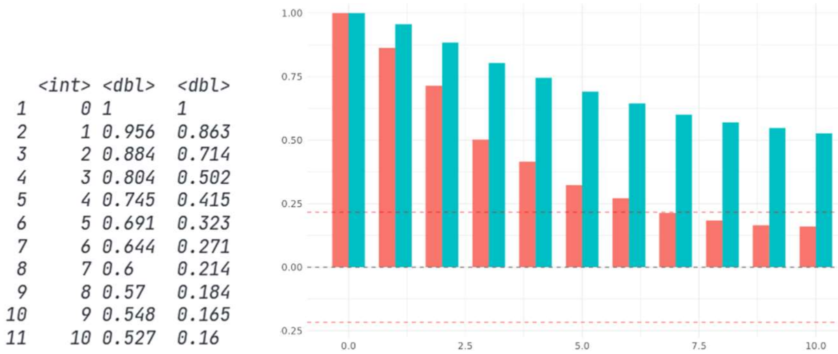
There is a very strong positive autocorrelation (0.969 at lag 1), which gradually decreases with increasing lag (Fig. 25-26). Even at lag 10, the autocorrelation remains noticeable (0.628). It indicates a stable trend and inertia of the migration process - the number of refugees in the next period strongly depends on the previous period. The smooth decrease in autocorrelation indicates a relatively stable nature of migration processes.



	Lag	Autocorrelation
1	0	1.000
2	1	0.969
3	2	0.933
4	3	0.894
5	4	0.855
6	5	0.814
7	6	0.775
8	7	0.738
9	8	0.702
10	9	0.665
11	10	0.628

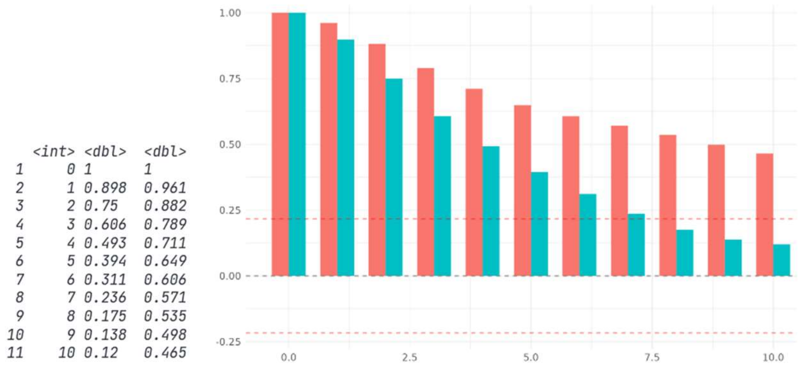
Figure 26: Autocorrelation correlogram for dataset 3 - Ukrainian refugees abroad

Fig. 27 shows the autocorrelation of events (attacks) and casualties separately. The autocorrelation of events decreases more slowly (from 0.956 to 0.527 at lag 10). The autocorrelation of the number of casualties decreases much faster (from 0.863 to 0.16). It means that the attacks themselves are more systematic, while the number of casualties is more random and less predictable.



	<int>	<dbl>	<dbl>
1	0	1	1
2	1	0.956	0.863
3	2	0.884	0.714
4	3	0.804	0.502
5	4	0.745	0.415
6	5	0.691	0.323
7	6	0.644	0.271
8	7	0.6	0.214
9	8	0.57	0.184
10	9	0.548	0.165
11	10	0.527	0.16

Figure 27: Autocorrelation correlogram for dataset 1 - shelling of Ukrainian civilian infrastructure (red - victims and green - events)



	<int>	<dbl>	<dbl>
1	0	1	1
2	1	0.898	0.961
3	2	0.75	0.882
4	3	0.606	0.789
5	4	0.493	0.711
6	5	0.394	0.649
7	6	0.311	0.606
8	7	0.236	0.571
9	8	0.175	0.535
10	9	0.138	0.498
11	10	0.12	0.465

Figure 28: Autocorrelation correlogram for dataset 1 - attacks on Ukrainian political infrastructure (red - victims and green - events)

The number of events (attacks) in Fig. 28 shows a rapid decrease in autocorrelation - from 1.0 to 0.12 over 10 months, which indicates a somewhat chaotic and less systematic nature of the attacks over time. The sharp drop is especially noticeable after the 5th lag (month). In contrast, the autocorrelation of the number of victims decreases more slowly - from 1.0 to 0.465, maintaining higher values throughout the period. It may indicate that although the attacks themselves become less predictable, their lethality retains a certain systematicity and dependence on previous periods.

Such a pattern may indicate a change in attack tactics - from regular, systematic attacks to more sporadic (irregular), but with similar effectiveness in terms of victims. Fig. 29 shows a generalized plot of the results of smoothing using the Pollard formulas. These plots demonstrate the dynamics of attacks on political infrastructure using two smoothing methods. In both cases, there is:

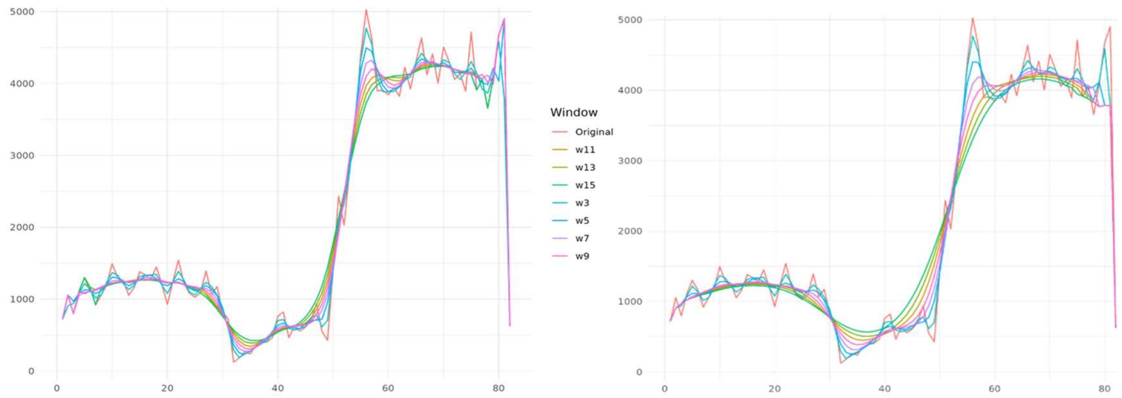


Figure 29: (a) Method A: Direct smoothing with Pollard formulas with different window sizes and (b) method B: Sequential smoothing with increasing window size for the dataset - shelling of political infrastructure of Ukraine

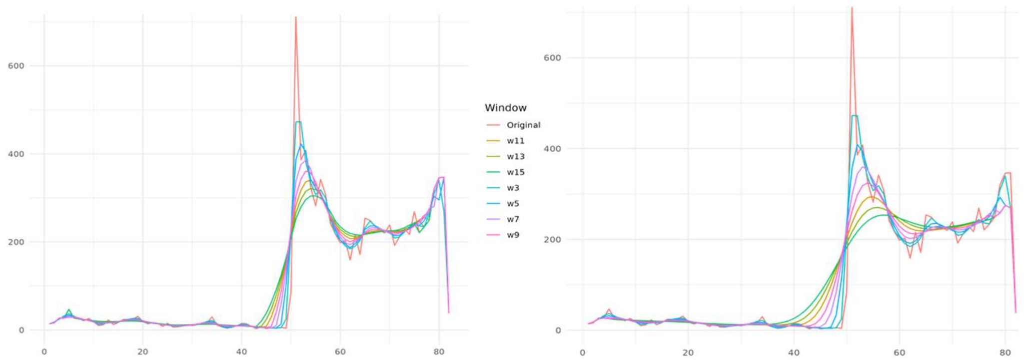


Figure 30: (a) Method A: Direct smoothing using Pollard formulas with different window sizes and (b) method B: Sequential smoothing with increasing window size for the dataset - shelling of civilian infrastructure in Ukraine

The initial period had a relatively stable event rate (around 1000-1500 events). Sharp increase after 50th period to peak around 4000-5000 events. Method B shows smoother transitions between periods, especially in the area of sharp increase. Different window sizes (w3-w15) affect the degree of smoothing, with larger windows giving a smoother curve. The graphs in Fig. 30 show a low level of events at the beginning (around 20-30 cases). A sharp peak of activity around the 50th period (up to 700 cases). Further stabilization at the level of 200-300 cases. Method B provides a smoother representation of the data, especially in the peak area. Larger window sizes (w11-w15) significantly smooth out the peak values. The graphs in Fig. 31 show a rapid increase in the number of refugees at the beginning (up to 8 million). Two sharp declines (around the 180th and 300th periods). Stabilization after the 300th period at the level of about 6 million. Both methods give almost identical results for this data set. The window size has a minimal effect on the shape of the curve, indicating greater stability of the data.

There is a robust positive correlation between all smoothing windows (coefficients from 0.9137 to 0.9999). The strongest correlation is observed between neighbouring smoothing windows (Fig. 32, dataset 1). The original series has the strongest correlation with smaller smoothing windows (Window_3, Window_5) and somewhat weaker with larger windows. As the smoothing window size increases, the correlation with the original series gradually decreases (from 0.9696 to 0.8951).

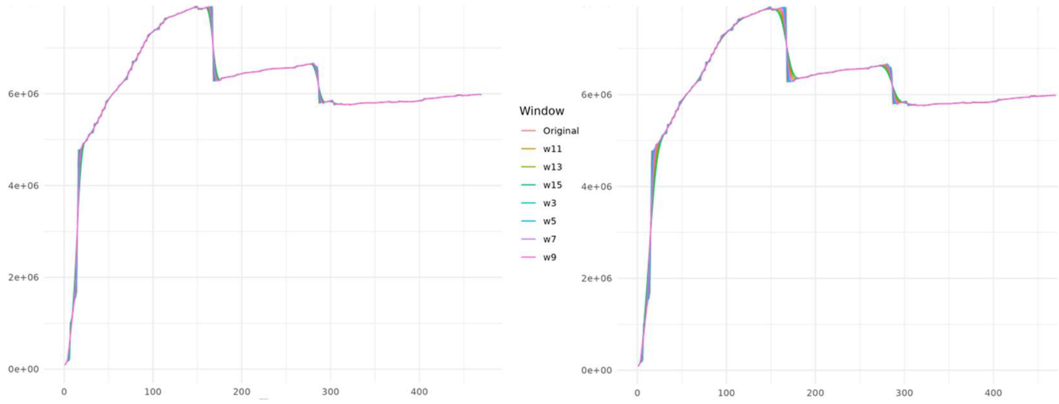


Figure 31: (a) Method A: Direct smoothing with Pollard formulas with different window sizes and (b) method B: Sequential smoothing with increasing window size for the dataset - Ukrainian refugees abroad

Dataset 1

	Original	Window_3	Window_5	Window_7	Window_9	Window_11	Window_13	Window_15
Original	1.0000	0.9696	0.9521	0.9366	0.9235	0.9137	0.9045	0.8951
Window_3	0.9696	1.0000	0.9907	0.9806	0.9703	0.9609	0.9519	0.9423
Window_5	0.9521	0.9907	1.0000	0.9954	0.9891	0.9817	0.9740	0.9656
Window_7	0.9366	0.9806	0.9954	1.0000	0.9973	0.9926	0.9867	0.9798
Window_9	0.9235	0.9703	0.9891	0.9973	1.0000	0.9984	0.9947	0.9896
Window_11	0.9137	0.9609	0.9817	0.9926	0.9984	1.0000	0.9987	0.9954
Window_13	0.9045	0.9519	0.9740	0.9867	0.9947	0.9987	1.0000	0.9987
Window_15	0.8951	0.9423	0.9656	0.9798	0.9896	0.9954	0.9987	1.0000

Dataset 2

	Original	Window_3	Window_5	Window_7	Window_9	Window_11	Window_13	Window_15
Original	1.0000	0.9937	0.9928	0.9909	0.9890	0.9876	0.9854	0.9829
Window_3	0.9937	1.0000	0.9950	0.9939	0.9922	0.9904	0.9882	0.9855
Window_5	0.9928	0.9950	1.0000	0.9982	0.9973	0.9956	0.9938	0.9915
Window_7	0.9909	0.9939	0.9982	1.0000	0.9996	0.9982	0.9969	0.9951
Window_9	0.9890	0.9922	0.9973	0.9996	1.0000	0.9993	0.9985	0.9971
Window_11	0.9876	0.9904	0.9956	0.9982	0.9993	1.0000	0.9997	0.9987
Window_13	0.9854	0.9882	0.9938	0.9969	0.9985	0.9997	1.0000	0.9996
Window_15	0.9829	0.9855	0.9915	0.9951	0.9971	0.9987	0.9996	1.0000

Dataset 3

	Original	Window_3	Window_5	Window_7	Window_9	Window_11	Window_13	Window_15
Original	1.0000	0.9994	0.9986	0.9977	0.9968	0.9959	0.9952	0.9945
Window_3	0.9994	1.0000	0.9997	0.9991	0.9984	0.9977	0.9970	0.9962
Window_5	0.9986	0.9997	1.0000	0.9998	0.9994	0.9988	0.9982	0.9975
Window_7	0.9977	0.9991	0.9998	1.0000	0.9999	0.9995	0.9990	0.9984
Window_9	0.9968	0.9984	0.9994	0.9999	1.0000	0.9999	0.9995	0.9991
Window_11	0.9959	0.9977	0.9988	0.9995	0.9999	1.0000	0.9998	0.9996
Window_13	0.9952	0.9970	0.9982	0.9990	0.9995	0.9998	1.0000	0.9999
Window_15	0.9945	0.9962	0.9975	0.9984	0.9991	0.9996	0.9999	1.0000

Figure 32: Correlation matrix

Fig. 32 (dataset 2) demonstrates extremely high correlation coefficients between all windows (from 0.9828 to 0.9999). The correlation between the original series and the smoothed data remains consistently high (from 0.9829 to 0.9937). The difference between the correlation coefficients is smaller than in the case of civil events. There is greater stability in the correlations between different smoothing windows. Accordingly, Figure 32 (dataset 3) shows the highest correlation coefficients among all three matrices (from 0.9945 to 0.9999). There is an almost perfect correlation between all smoothing windows. The original series has a robust correlation with all smoothed series (from 0.9945 to 0.9994). The slightest variation between the correlation coefficients compared to the other matrices. In Fig. 33a, for method A, the initial decline in the number of points and then stabilization at medium window sizes. Method B: Decreasing the turning points by increasing window size and reducing stabilization. For the data on civilian infrastructure attacks, method A is more stable than method B at medium window sizes. It indicates that method A handles the data better without losing essential details.

In Fig. 33b, for method A, the number of turning points gradually decreases with increasing window size. Method B: The decrease in the number of points is more abrupt, with a noticeable dip in the average values of the window size. Method B shows a greater sensitivity to smoothing compared to method A. The sharp drop in the number of turning points in method B indicates a more substantial influence of the window on the detection of local extremes. In Fig. 33c, for method A, there is a gradual, almost linear decrease in the number of points with increasing window size. Method B: The reduction in turning points is more rapid than in method A but slows down at larger window sizes. For the refugee data, method B reacts more strongly to increasing window size, which can lead to the loss of essential changes. Method A provides a more stable smoothing, preserving key features in the trend. According to Fig. 34a, the correlation between the shelling of civilian targets and the number of refugees, according to Pollard's method, has a weak linear relationship. The correlation coefficient of the modulus is < 0.5, and the coefficient of determination is less than 25% (Table 7).

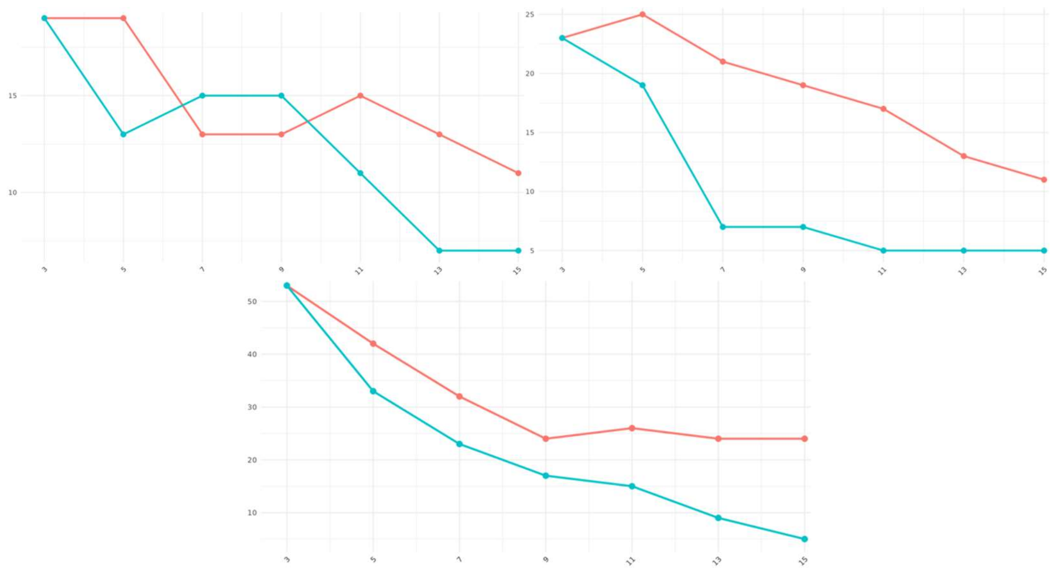


Figure 33: Diagram of the dependence of the number of turning points on the window size using Pollard's formulas for datasets 1-3 (red - method A, green - method B)

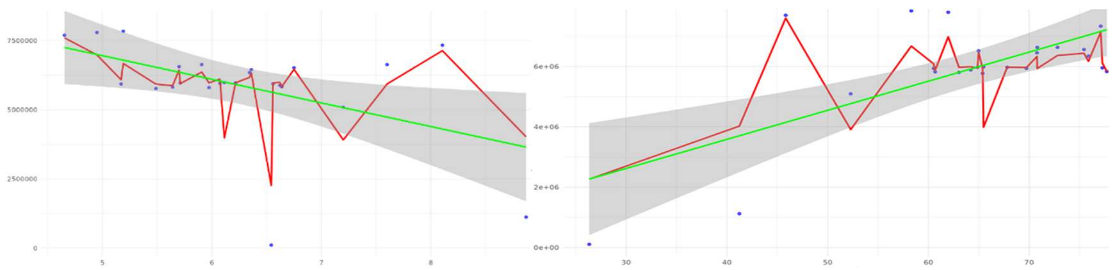


Figure 34: Correlation between shelling of civilian/political targets and the number of refugees using Pollard's method (X is the average number of shelling per month, and Y is the average number of refugees), where blue dots are unsmoothed data and red dots are smoothed data

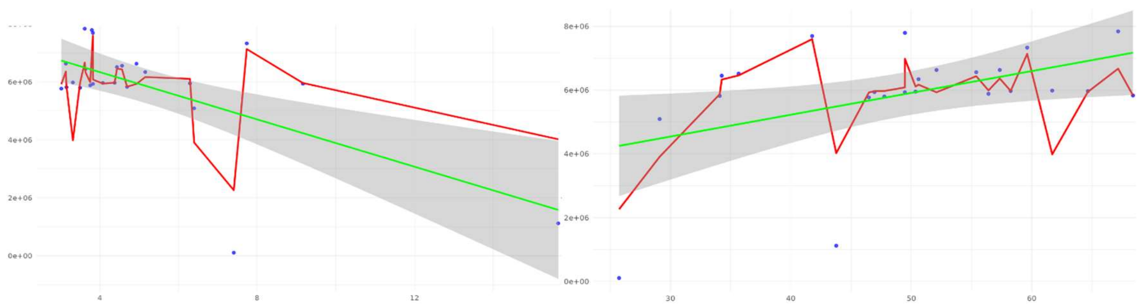


Figure 35: Correlation between fatalities caused by the shelling of civilian/political targets and the number of refugees using Pollard's method (X is the average number of fatalities and Y is the average number of refugees), where blue are unsmoothed data and red – smoothed data

According to Fig. 34b, the correlation between the shelling of political targets and the number of refugees, according to the Pollard method, has a linear relationship of medium strength. The correlation coefficient of the modulus is less than 0.7 but more than 0.5, the coefficient of determination is less than 50% but more than 25%. According to Fig. 35a, the correlation has a weak linear relationship. The correlation coefficient of the modulus < 0.5, the coefficient of determination is less than 25%. According to Fig. 35b, the correlation is a weak linear relationship. The correlation coefficient of the modulus < 0.5, the coefficient of determination is less than 25%. The correlation coefficient is 0.41, indicating a moderate positive relationship between the variables (Fig. 36a). This means that there is a specific relationship between the number of attacks on civilian targets and the number of refugees, but it is not very strong. This indicator indicates that the increase in attacks on civilian targets has some impact on the rise in the number of refugees but is not the only determining factor. The graph shows a rather scattered arrangement of points, which confirms the moderate nature of the relationship (Table 8).

Table 7

Correlation coefficients and parameters according to data from Fig. 34-35

Name	34(a)	34(b)	35(a)	35(b)
Correlation coefficient (smoothed)	-0.399	0.595	-0.417	0.385
Coefficient of determination (R-squared),%	15.92	35.38	17.39	14.82
t	-2.0408	3.4706	-2.1516	1.9568
df	22	22	22	22
p-value	0.05345	0.002172	0.04266	0.06318
95% confidence interval	-0.691137947	0.2518329	-0.70223599	-0.02174177
	0.005267275	0.8050384	-0.01629849	0.68243274
Sample estimates	-0.3989772	0.5948043	-0.4169538	0.3850319

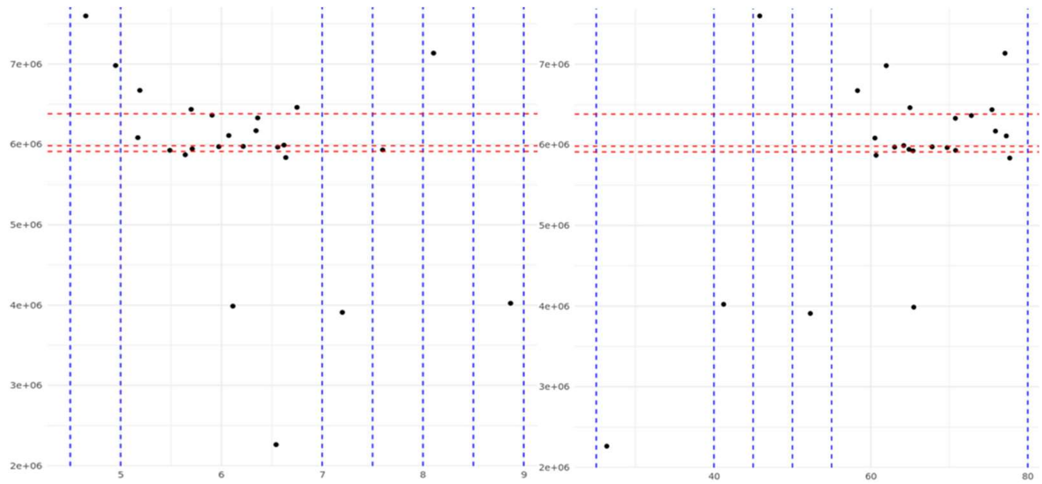


Figure 36: Division of the correlation field of the group between shelling of civilian/political targets and the number of refugees

Table 8

Correlation coefficients and parameters according to data from Fig. 36-37

Name	36(a)	36(b)	37(a)	37 (b)
Group variance	539673784831.992	1014055071914.54	882124764899.672	1211882265553.03
Total variance	13177302885549.89	13177302885549.89	13177302885549.89	13177302885549.89
Correlation relationship	0.41	0.77	0.67	0.92

The correlation coefficient is 0.77, indicating a strong positive relationship between the variables (Figure 36b). It means that there is a pronounced relationship between the number of attacks on political targets and the number of refugees. Such a high indicator indicates that the increase in attacks on political targets has a significant impact on the rise in the number of refugees. In the graph, it can be seen that the points are located more densely and form a more pronounced trend, which confirms a stronger relationship compared to the previous graph. The correlation coefficient of 0.67 indicates a moderate relationship between the number of fatalities due to attacks on civilian targets and the number of refugees (Figure 37a). It shows that although changes in fatalities have some impact on the increase or decrease in the number of refugees, the relationship is not very strong. It is likely that other factors, such as security threats, economic factors, or the availability of evacuation routes, influence migration processes. The correlation coefficient of 0.92 indicates a powerful relationship between the number of fatalities due to attacks on political targets and the number of refugees (Fig. 37b). This means that the increase in deaths in such conditions has a significant impact on the mass exodus of people. It is probably due to the high level of fear and instability that arise as a result of attacks on political structures or their objects.

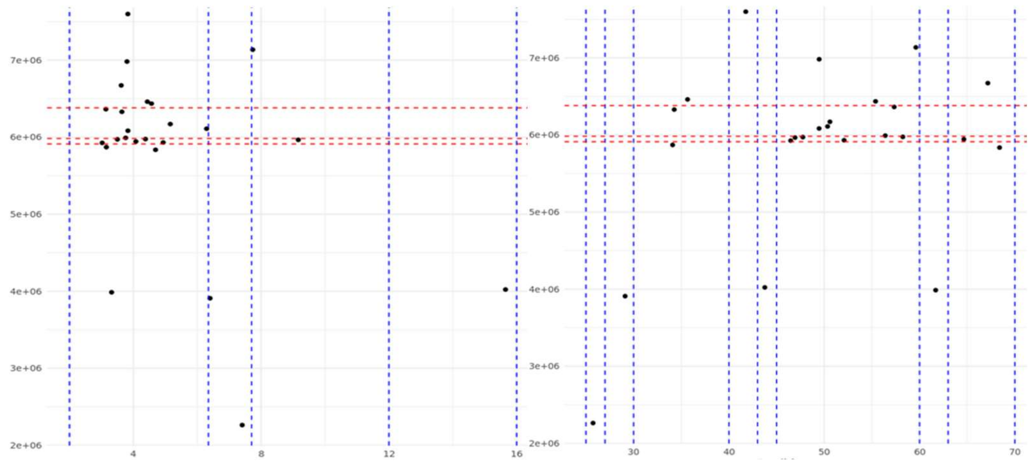


Figure 37: Dividing the correlation field between fatalities caused by shelling of civilian/political targets and the number of refugees

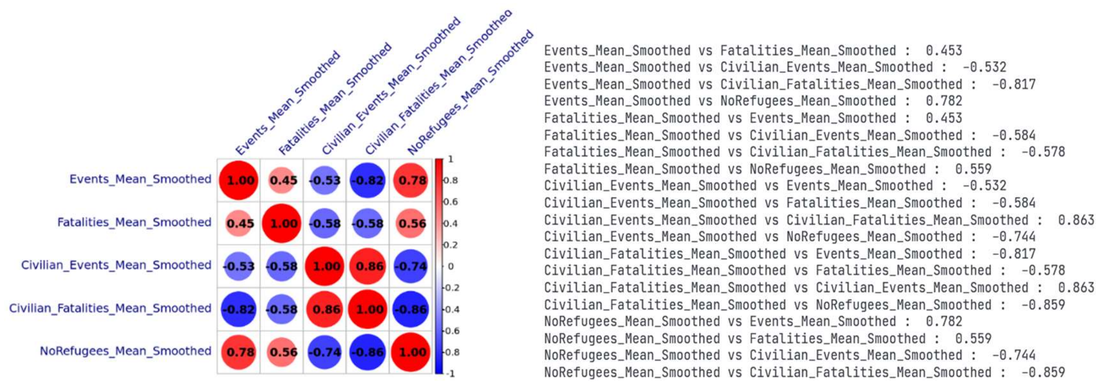


Figure 38: Correlation matrix

The correlogram in Fig. 39 shows a strong positive autocorrelation, which gradually decreases with increasing lag. Starting from lag 0 (1.000), the value gradually decreases to 0.628 at lag 10. This autocorrelation structure indicates the presence of a stable trend in the data - the number of refugees on different days is strongly correlated with each other. It is logical since people usually do not return immediately but stay abroad for an extended period.

The correlogram in Fig. 40 shows two different indicators - "Events" (green) and "Victims" (red). Both indicators demonstrate positive autocorrelation but with different intensities: "Events" have a more substantial autocorrelation, which decreases more slowly (from 1.0 to 0.527 at lag 10), "Victims" show a faster decrease in autocorrelation (from 1.0 to 0.158 at lag 10). It may indicate that the shelling itself has a more systematic nature and a certain periodicity. At the same time, the number of victims is a more random variable that depends less on previous values. This result is logical since the number of victims depends on many additional factors (time of shelling, population density at the site of the strike, the presence of shelters, etc.).

	Lag	Autocorrelation
1	0	1.000
2	1	0.969
3	2	0.932
4	3	0.893
5	4	0.854
6	5	0.813
7	6	0.774
8	7	0.737
9	8	0.701
10	9	0.664
11	10	0.628

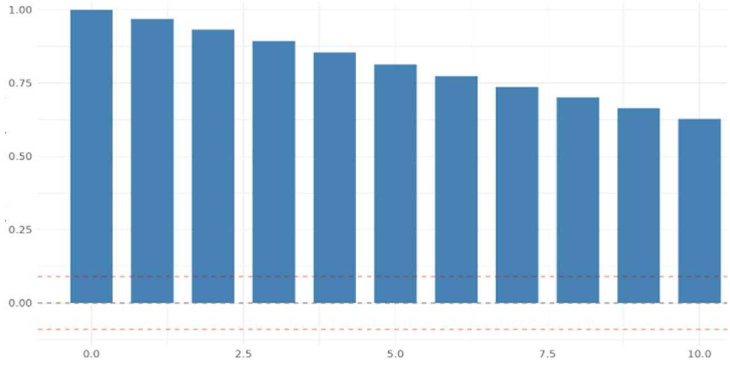


Figure 39: Autocorrelation correlogram for dataset 3 - Ukrainian refugees abroad

	<int>	<dbl>	<dbl>
1	0	1	1
2	1	0.961	0.878
3	2	0.881	0.655
4	3	0.804	0.493
5	4	0.741	0.389
6	5	0.69	0.317
7	6	0.641	0.26
8	7	0.595	0.208
9	8	0.564	0.171
10	9	0.546	0.16
11	10	0.527	0.158

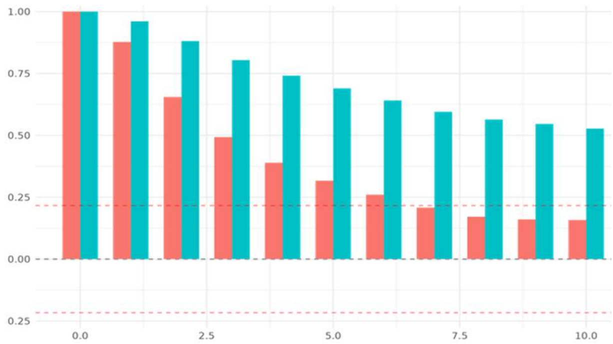


Figure 40: Autocorrelation correlogram for dataset 1 - shelling of Ukrainian civilian infrastructure (red - victims and green - events)

We see an interesting difference from the previous correlogram (Fig. 41) - here, "Victims" (red colour) have a more substantial autocorrelation than "Events" (green colour). "Victims" starts with a high autocorrelation (1.0). It slowly decreases to 0.46 at lag 10. It maintains relatively high values even at considerable lags. "Events" also starts with a high autocorrelation (1.0). It decreases much faster to 0.133 at lag 10. After lag 5, the autocorrelation becomes relatively weak (<0.4). Such a structure may indicate that in attacks on political infrastructure, the number of victims is more predictable and systematic than the events themselves. This may be due to the fact that political objects usually have a certain number of permanent personnel, so the number of potential victims is more stable. Instead, the events themselves (the shelling) may occur more chaotically and less predictably.

	Lag <int>	Події <dbl>	Жертви <dbl>
1	0	1	1
2	1	0.912	0.956
3	2	0.753	0.869
4	3	0.612	0.788
5	4	0.494	0.714
6	5	0.396	0.647
7	6	0.315	0.601
8	7	0.242	0.568
9	8	0.182	0.532
10	9	0.146	0.492
11	10	0.133	0.46

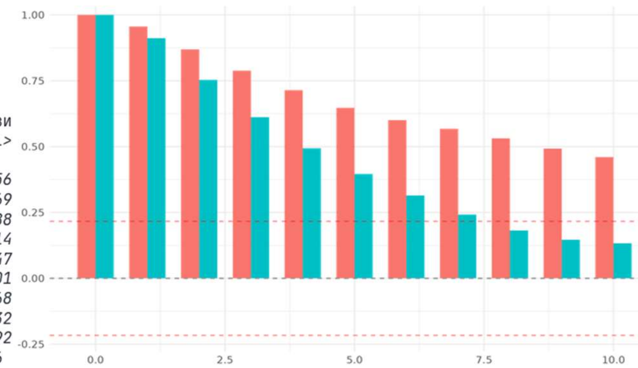


Figure 41: Correlogram for dataset 2 - attacks on political infrastructure in Ukraine (red - victims and green - events)

The overall trend in Figure 42a shows a relatively low level until period 40. There is a sharp peak of activity around period 60. Smoothing with different alpha values helps to visualize the overall trend better. As in the first graph, smaller alpha values give a smoother curve.

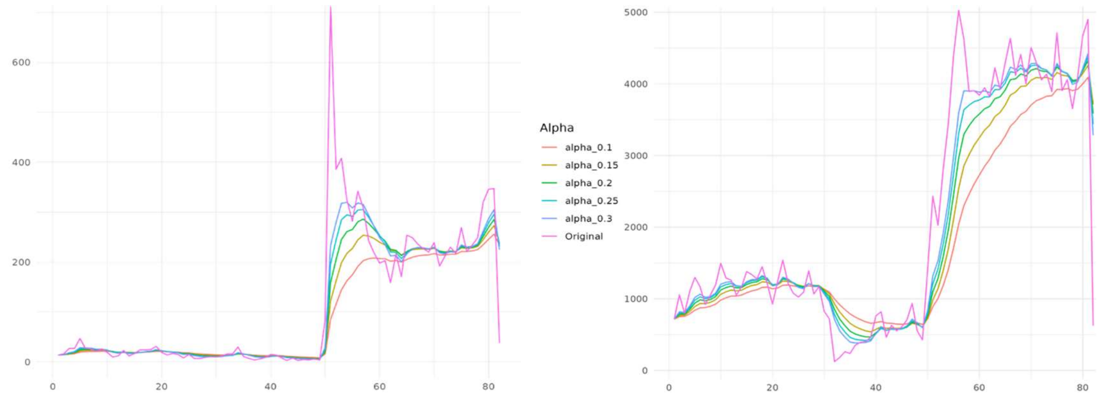


Figure 42: Exponential smoothing with different alpha values for the dataset 1-2

The data show in Fig. 42b two main periods of exponential smoothing intensity: the first - about 1000-1500 cases (up to the 40th period), and the second - a sharp increase to 4000-5000 cases (after the 60th period). Smaller alpha values (0.1, 0.15) give a smoother curve, filtering out short-term fluctuations. Larger alpha values (0.25, 0.3) better track sharp changes but retain more "noise" in the data. The original data (pink line) shows significant volatility (variability). The graph in Fig. 43 shows a rapid increase in the number of refugees at the beginning of the period (up to the 100th period). After reaching the peak, a sharp decline is observed. Then, the curve stabilizes with a slight gradual decrease. Different alpha values produce very similar smoothing results, indicating relatively "clean" data with fewer random fluctuations.

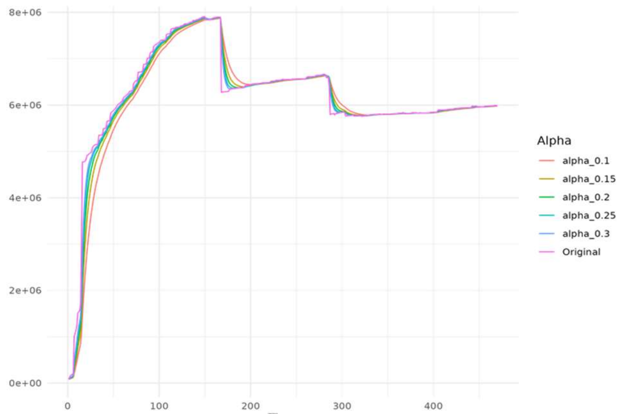


Figure 43: Exponential smoothing with different alpha values for dataset 3

Dataset 1

	Original	Alpha_0.1	Alpha_0.15	Alpha_0.2	Alpha_0.25	Alpha_0.3
Original	1.0000	0.7767	0.8190	0.8512	0.8768	0.8977
Alpha_0.1	0.7767	1.0000	0.9927	0.9777	0.9607	0.9439
Alpha_0.15	0.8190	0.9927	1.0000	0.9957	0.9862	0.9746
Alpha_0.2	0.8512	0.9777	0.9957	1.0000	0.9972	0.9907
Alpha_0.25	0.8768	0.9607	0.9862	0.9972	1.0000	0.9981
Alpha_0.3	0.8977	0.9439	0.9746	0.9907	0.9981	1.0000

Dataset 2

	Original	Alpha_0.1	Alpha_0.15	Alpha_0.2	Alpha_0.25	Alpha_0.3
Original	1.0000	0.8692	0.9062	0.9297	0.9453	0.9564
Alpha_0.1	0.8692	1.0000	0.9946	0.9840	0.9728	0.9623
Alpha_0.15	0.9062	0.9946	1.0000	0.9971	0.9912	0.9845
Alpha_0.2	0.9297	0.9840	0.9971	1.0000	0.9984	0.9949
Alpha_0.25	0.9453	0.9728	0.9912	0.9984	1.0000	0.9990
Alpha_0.3	0.9564	0.9623	0.9845	0.9949	0.9990	1.0000

Dataset 3

	Original	Alpha_0.1	Alpha_0.15	Alpha_0.2	Alpha_0.25	Alpha_0.3
Original	1.0000	0.9522	0.9728	0.9824	0.9879	0.9913
Alpha_0.1	0.9522	1.0000	0.9955	0.9888	0.9829	0.9779
Alpha_0.15	0.9728	0.9955	1.0000	0.9984	0.9956	0.9927
Alpha_0.2	0.9824	0.9888	0.9984	1.0000	0.9993	0.9978
Alpha_0.25	0.9879	0.9829	0.9956	0.9993	1.0000	0.9996
Alpha_0.3	0.9913	0.9779	0.9927	0.9978	0.9996	1.0000

Figure 44: Correlation matrix

There is a robust positive correlation between all smoothing levels (Alpha), with coefficients ranging from 0.77 to 0.99 (Fig. 44, dataset 1). The strongest relationship is between adjacent smoothing levels, and the weakest is between the original series and the smoothed data, which is the expected result. The "Political Events" matrix (Fig. 44, dataset 2) shows extremely high correlation values between all smoothing levels (0.86-0.99). The original series has a slightly stronger relationship with the smoothed data compared to civil events, which may indicate a greater regularity in political events. The "Refugee Data" matrix (Fig. 44, dataset 3) shows the highest correlation values among all three matrices (0.95-0.99), including the relationship with the original series. It indicates that the refugee data have the most stable and consistent dynamics, with fewer random fluctuations. With exponential smoothing for dataset 1, the graph shows a similar trend as the second graph but with a slightly smaller growth amplitude (from 19 to 32 points). There is a noticeable plateau at alpha values of 0.15-0.20, which may indicate some stability in the pattern of attacks on civilian infrastructure in this smoothing range.

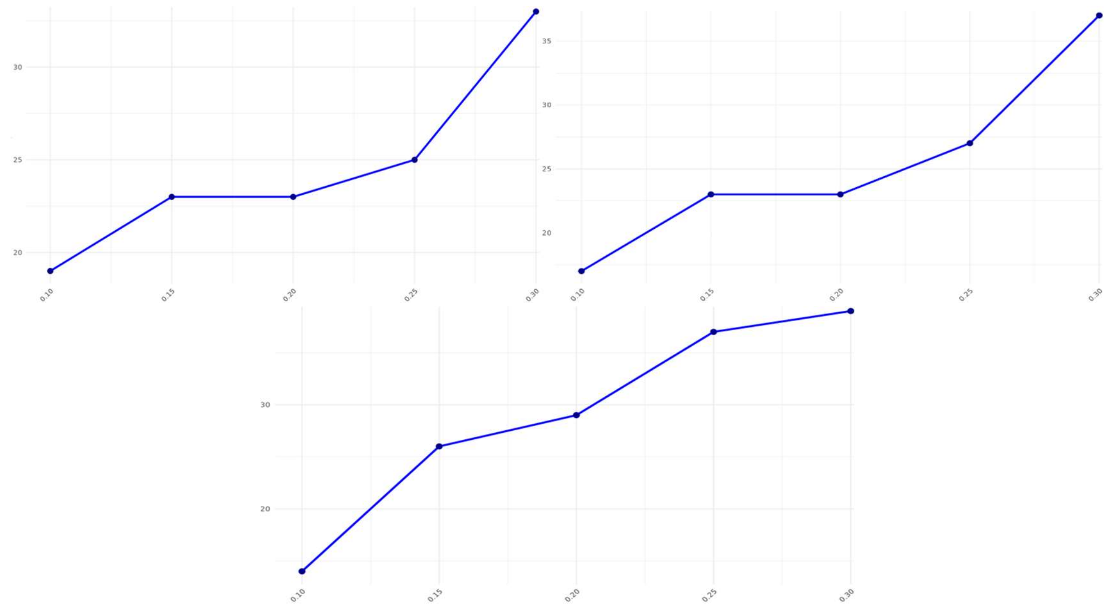


Figure 45: Diagram of the number of turning points versus different alpha values for exponential smoothing for datasets 1-3

Exponential smoothing for dataset 2 shows a gradual increase in the number of turning points from 18 to 35 as alpha increases, with the sharpest increase at alpha > 0.25. It may indicate a more complex structure and irregularity in the data on the shelling of political infrastructure at higher values of the smoothing parameter. The graph for exponential smoothing for dataset 3 shows the smoothest and most consistent increase in the number of turning points from 14 to almost 40, without obvious plateaus. It may indicate more regular dynamics of the refugee movement process and less abrupt changes compared to the shelling data. Fig. 46a shows a weak linear relationship. The modulus correlation coefficient is < 0.5, and the coefficient of determination is less than 25% (Table 9). Fig. 46b shows a dynamic relationship of medium strength. The modulus correlation coefficient is less than 0.7 but more than 0.5, and the coefficient of determination is less than 50% but more than 25%.

Fig. 47a shows a weak linear relationship. The correlation coefficient of the modulus is < 0.5, and the coefficient of determination is less than 25%. Fig. 47b shows a weak linear relationship. The correlation coefficient of the modulus is < 0.5, and the coefficient of determination is less than 25%. The correlation coefficient (Fig. 48a) is 0.815, indicating a strong positive relationship between the variables. It suggests that there is a significant relationship between the shelling and the parameter under study, where an increase in one indicator leads to a proportional increase in the other (Table 10). With a correlation ratio of 0.827 (Fig. 48b), there is a powerful positive relationship between the shelling of civilian targets and the number of refugees. It demonstrates that the intensity of shelling of civilian objects has a direct and significant impact on the increase in the number of refugees.

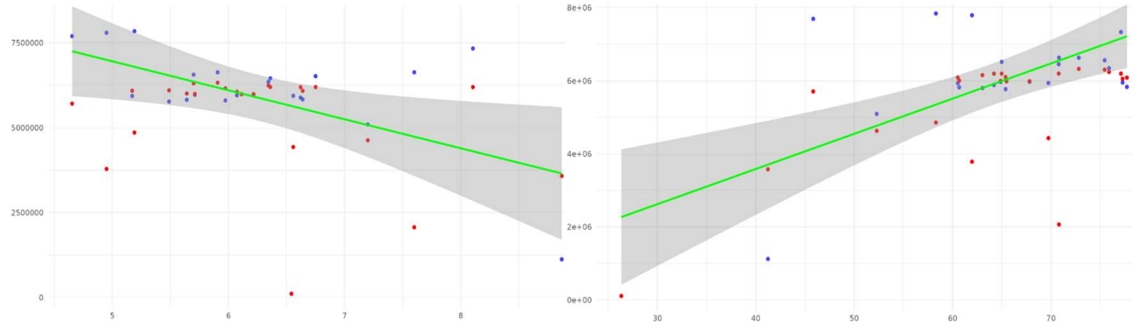


Figure 46: Correlation between shelling of civilian/political targets and the number of refugees using the exponential smoothing at α=0.3 (X is the average number of shelling per month, and Y is the average number of refugees), where blue– unsmoothed data and red – smoothed data

Table 9
Correlation coefficients and parameters according to data from Fig. 46-47

Name	46(a)	46(b)	47(a)	47(b)
Correlation coefficient (smoothed)	-0.292	0.646	-0.431	0.383
Coefficient of determination (R-squared),%	8.55	41.77	18.57	14.65
t	-1.4344	3.9728	-2.2398	1.9432
df	22	22	22	22
p-value	0.1655	0.0006444	0.03554	0.0649
95% confidence interval	-0.6224154	0.3285992	-0.71075701	-0.02442389
	0.1257902	0.8326316	-0.03330132	0.68099633
Sample estimates	-0.2924462	0.6463221	-0.4309094	0.3827438

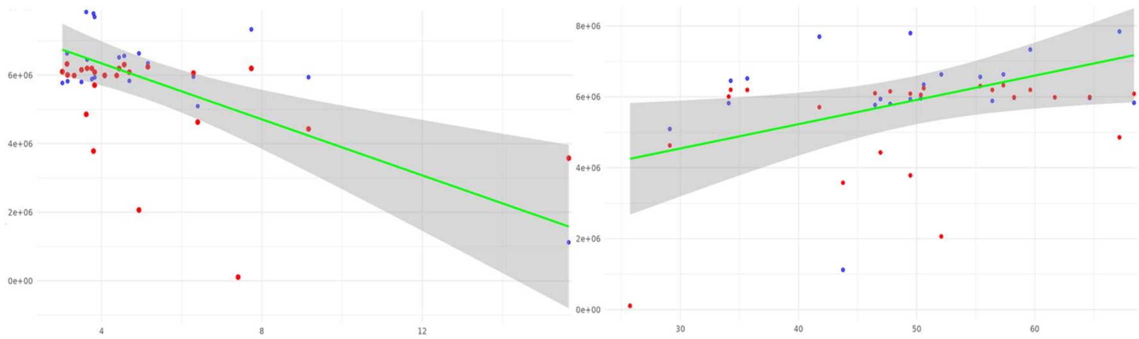


Figure 47: Correlation between fatalities caused by the shelling of civilian/political targets and the number of refugees using the exponential smoothing method at α=0.3 (X is the average number of deaths and Y is the average number of refugees), where blue dots are unsmoothed data and red dots are smoothed data

Table 10
Correlation coefficients and parameters according to data from Fig. 48-49

Name	48(a)	48(b)	49(a)	49 (b)
Group variance	1885537113281.85	1913479315163.88	1356773952809.44	2017294691518.56
Total variance	2313776260684.71	2313776260684.71	2313776260684.71	2313776260684.71
Correlation ratio	0.815	0.827	0.586	0.817

The correlation coefficient of 0.586 (Figure 49a) shows a moderate positive relationship between the number of attacks on political targets and the number of refugees. This relationship is less pronounced compared to previous indicators but still indicates some dependence between the variables. The correlation coefficient of 0.872 (Figure 49b) shows a powerful positive relationship between the number of fatalities caused by attacks and the number of refugees. It indicates that the increase in the number of victims has the most significant impact on the rise in the number of refugees, which is one of the factors studied.

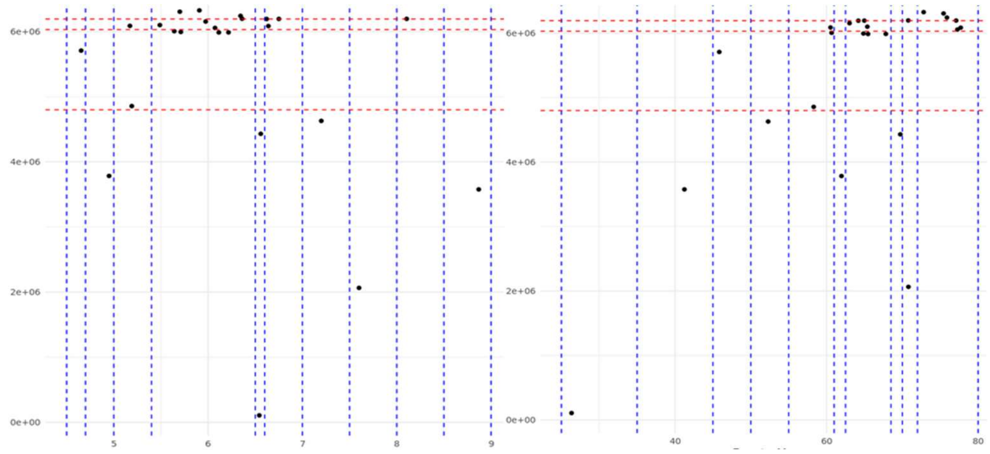


Figure 48: Division of the correlation field of the group between shelling of civilian/political targets and the number of refugees

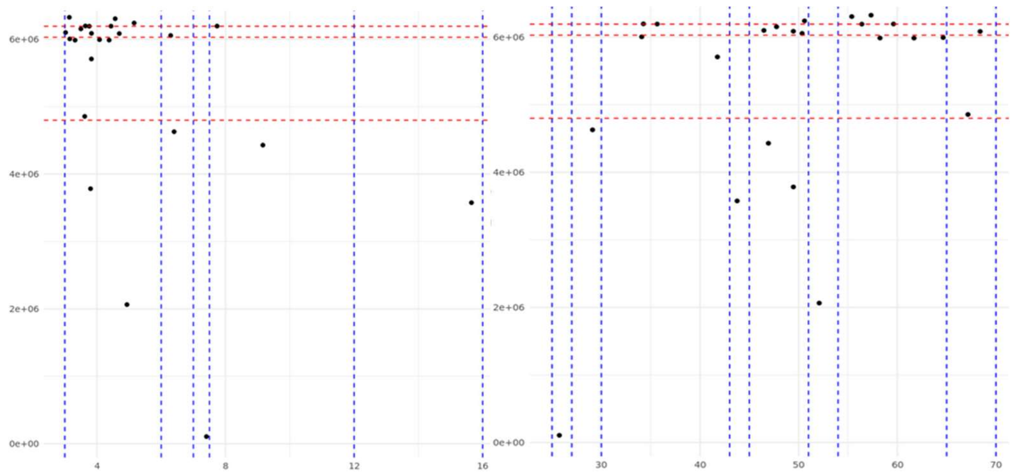


Figure 49: Dividing the correlation field between fatalities caused by shelling of civilian/political targets and the number of refugees

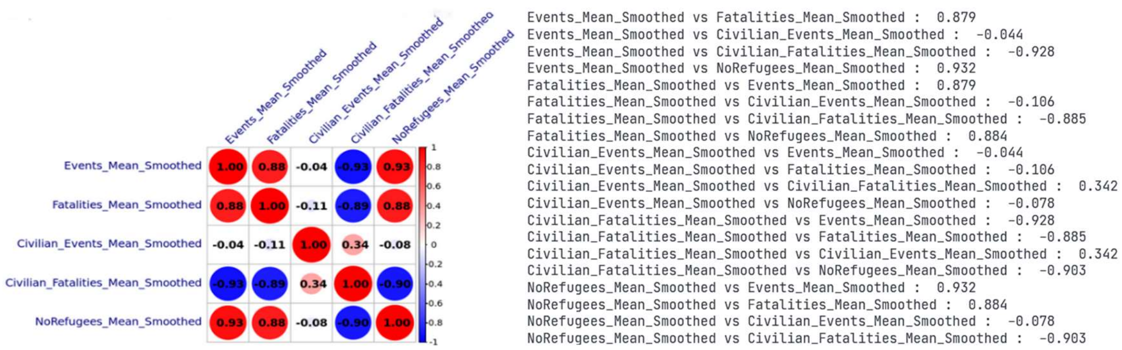


Figure 50: Correlation matrix

Figure 51 shows a strong positive autocorrelation that gradually decreases with increasing lag. It indicates a clear temporal dependence in the refugee data, where current values are strongly correlated with previous periods, suggesting persistent trends in migration processes.

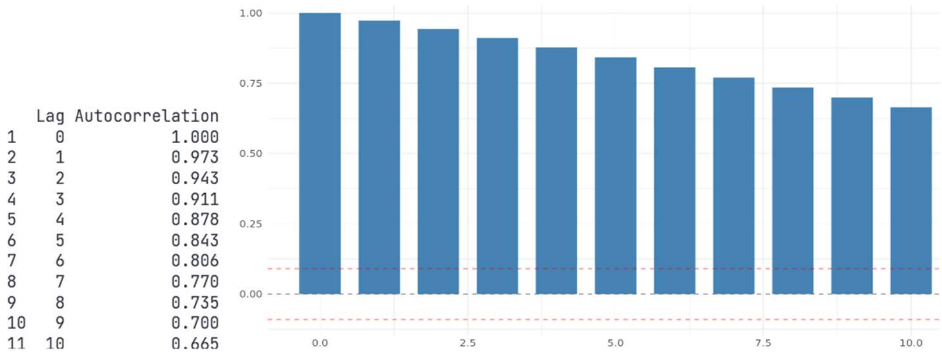


Figure 51: Autocorrelation correlogram for dataset 3 - Ukrainian refugees abroad

The graph in Fig. 52 shows a high initial autocorrelation for both metrics (casualties and events), with a sharper decline for the casualties' indicator. It suggests that while both indicators are time-dependent, the number of casualties has a less stable dynamic compared to the number of shelling events. Similar to the previous correlogram, the indicators in Fig. 53 show a high initial autocorrelation with a gradual decline but with a minor difference between the event and casualties metrics. It suggests a more consistent dynamics between the number of attacks and their consequences for the political infrastructure.

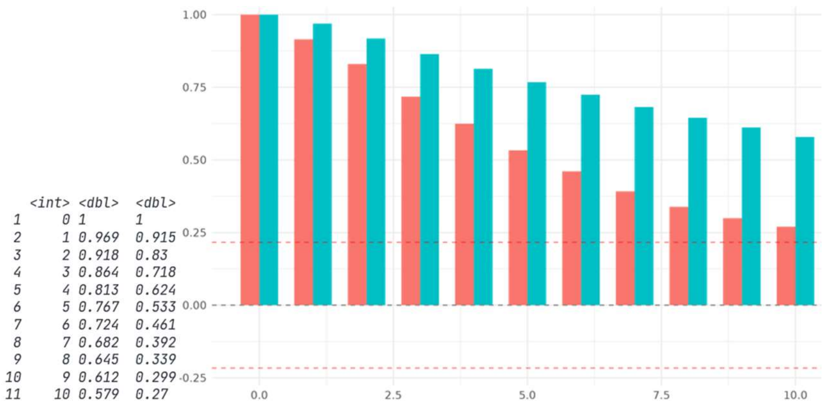


Figure 52: Autocorrelation correlogram for dataset 1 (red - victims and green - events)

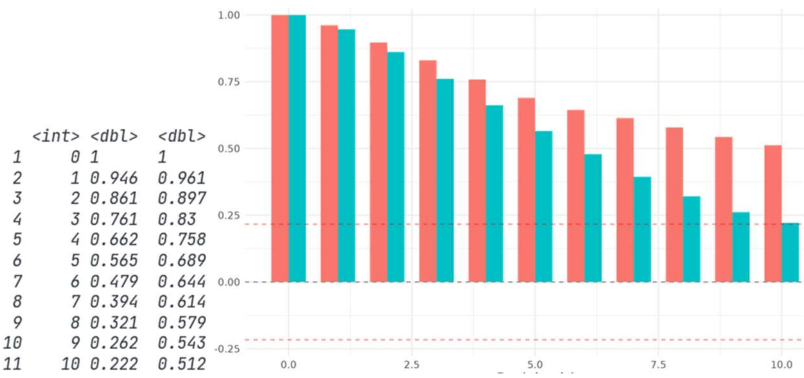


Figure 53: Autocorrelation correlogram for dataset 1 (red – victims and green – events)

The median smoothing data (Fig. 54) shows a sharp spike around time point 50, followed by fluctuations at a higher level. Both smoothing methods are effective in reducing the extreme spike while maintaining the overall pattern. The sequential smoothing in Method B creates somewhat smoother transitions between periods of change, which can be helpful for analysing long-term trends in attacks on civilian infrastructure. The original data in Fig. 55 show significant volatility with a large spike around time point 60, followed by a sharp drop around time point 80. Method A and

Method B produce similar smoothing effects, but Method B (sequential smoothing) provides somewhat more stable trends while maintaining the underlying patterns of the data. Both methods are effective in reducing noise while preserving the key features of the trend—the initial lower level of events, the sharp increase, and the final decrease.

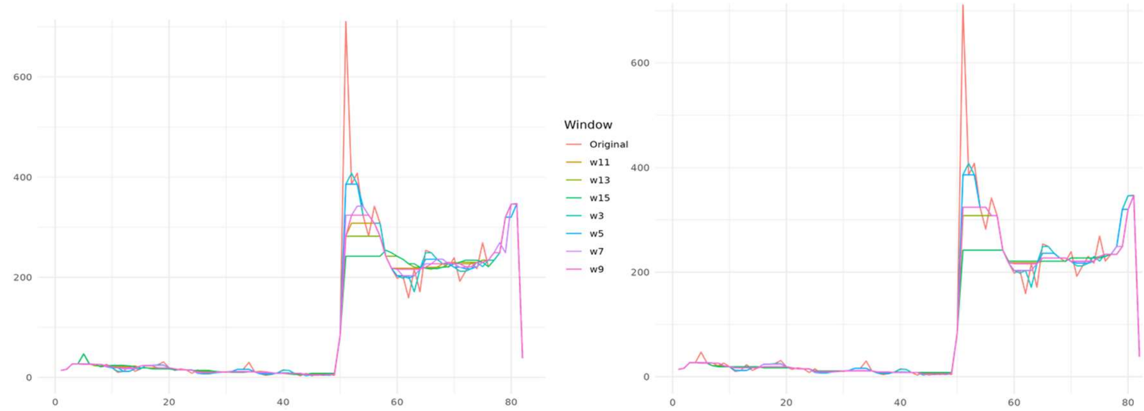


Figure 54: (a) Method A: Direct median smoothing with different window sizes and (b) method B: Sequential median smoothing with increasing window size for the dataset - shelling of civilian infrastructure in Ukraine

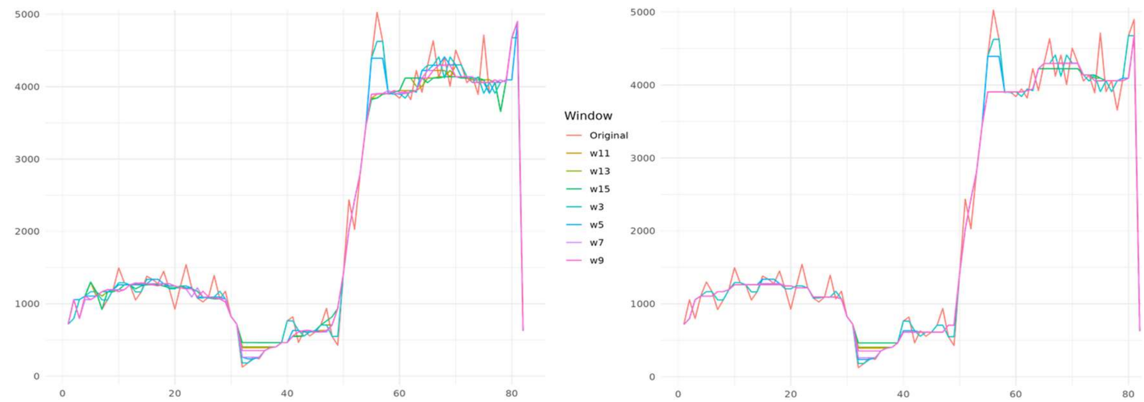


Figure 55: (a) Method A: Direct median smoothing with different window sizes and (b) method B: Sequential median smoothing with increasing window size for the dataset - shelling of political infrastructure of Ukraine

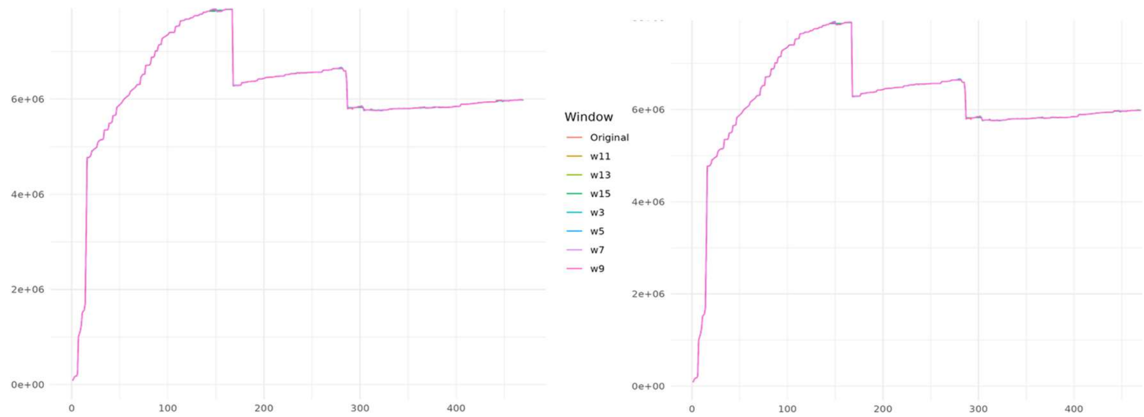


Figure 56: (a) Method A: Direct median smoothing with different window sizes and (b) method B: Sequential median smoothing with increasing window size for the dataset - Ukrainian refugees abroad

Both methods in Fig. 56 show almost identical results for this dataset, probably because the original data already has a relatively smooth trend. The data show:

- The rapid initial increase in the number of refugees;
- Plateau around the 200th time point;
- Significant drop followed by stabilization;
- There is a slight upward trend in the final period.

The correlation matrix in Fig. 57 (dataset 1) shows a high correlation between the different smoothing windows (all values above 0.92). The highest correlation is observed between neighbouring window sizes, which is logical since they similarly process the data. The original data has the lowest correlation with the most enormous smoothing windows (Window_13, Window_15), indicating more smoothing and loss of detail as the window size increases.

Dataset 1

	Original	Window_3	Window_5	Window_7	Window_9	Window_11	Window_13	Window_15
Original	1.0000	0.9633	0.9640	0.9465	0.9343	0.9338	0.9281	0.9047
Window_3	0.9633	1.0000	0.9984	0.9921	0.9901	0.9877	0.9808	0.9629
Window_5	0.9640	0.9984	1.0000	0.9942	0.9922	0.9898	0.9831	0.9660
Window_7	0.9465	0.9921	0.9942	1.0000	0.9963	0.9949	0.9901	0.9768
Window_9	0.9343	0.9901	0.9922	0.9963	1.0000	0.9992	0.9957	0.9854
Window_11	0.9338	0.9877	0.9898	0.9949	0.9992	1.0000	0.9979	0.9893
Window_13	0.9281	0.9808	0.9831	0.9901	0.9957	0.9979	1.0000	0.9957
Window_15	0.9047	0.9629	0.9660	0.9768	0.9854	0.9893	0.9957	1.0000

Dataset 2

	Original	Window_3	Window_5	Window_7	Window_9	Window_11	Window_13	Window_15
Original	1.0000	0.9932	0.9918	0.9884	0.9883	0.9895	0.9886	0.9885
Window_3	0.9932	1.0000	0.9971	0.9948	0.9946	0.9940	0.9935	0.9928
Window_5	0.9918	0.9971	1.0000	0.9969	0.9967	0.9959	0.9952	0.9947
Window_7	0.9884	0.9948	0.9969	1.0000	0.9998	0.9989	0.9985	0.9979
Window_9	0.9883	0.9946	0.9967	0.9998	1.0000	0.9992	0.9988	0.9983
Window_11	0.9895	0.9940	0.9959	0.9989	0.9992	1.0000	0.9996	0.9990
Window_13	0.9886	0.9935	0.9952	0.9985	0.9988	0.9996	1.0000	0.9996
Window_15	0.9885	0.9928	0.9947	0.9979	0.9983	0.9990	0.9996	1.0000

Dataset 3

	Original	Window_3	Window_5	Window_7	Window_9	Window_11	Window_13	Window_15
Original	1	1	1	1	1	1	1	1
Window_3	1	1	1	1	1	1	1	1
Window_5	1	1	1	1	1	1	1	1
Window_7	1	1	1	1	1	1	1	1
Window_9	1	1	1	1	1	1	1	1
Window_11	1	1	1	1	1	1	1	1
Window_13	1	1	1	1	1	1	1	1
Window_15	1	1	1	1	1	1	1	1

Figure 57: Correlation matrix

The correlation matrix in Fig. 57 (dataset 2) shows a similar pattern to the first table but with slightly higher correlation coefficients (all values above 0.93). It suggests that median smoothing produces more consistent results for political events compared to civil events.

The correlation matrix in Fig. 57 (dataset 3) shows a perfect correlation (all values = 1) between all smoothing windows. It indicates that the refugee data are very smooth, and different smoothing window sizes have little effect on the shape of the trend. With median smoothing, the high correlation with the original data is maintained, indicating that essential data characteristics are preserved during smoothing.

The diagram in Fig. 58a shows a similar pattern but with a greater difference between the methods. Method A shows unstable behaviour with oscillations, while Method B consistently reduces the number of turning points. Method B is significantly more effective in reducing the number of turning points compared to Method A for all window sizes (Fig. 58b). Method B quickly stabilizes at a low level. The plot in Fig. 58c shows the lowest number of turning points among all data sets. Method B almost eliminates turning points after a window of size 5, while Method A retains a certain number of turning points even at large window sizes.

The graph (Fig. 59a) shows a negative correlation - with an increase in the number of attacks on civilian targets, there is a tendency for the number of refugees to decrease. However, the data have significant variability (as can be seen from the scatter of blue points), and the confidence interval (grey zone) expands with an increase in the number of attacks, which indicates a lower reliability of the forecast at higher values—weak linear relationship. The correlation coefficient of the modulus is < 0.5, and the coefficient of determination is less than 25% (Table 11).

In Fig. 59b, a positive correlation is observed - with an increase in the number of attacks on political targets, the number of refugees also increases. The trend is more pronounced, although the red line shows significant fluctuations. The confidence interval expands at the edges of the graph, which indicates lower reliability of the forecast at extreme values—the linear relationship of average strength. The modulus correlation coefficient is less than 0.7 but more than 0.5, the coefficient of determination is less than 50% but more than 25%.

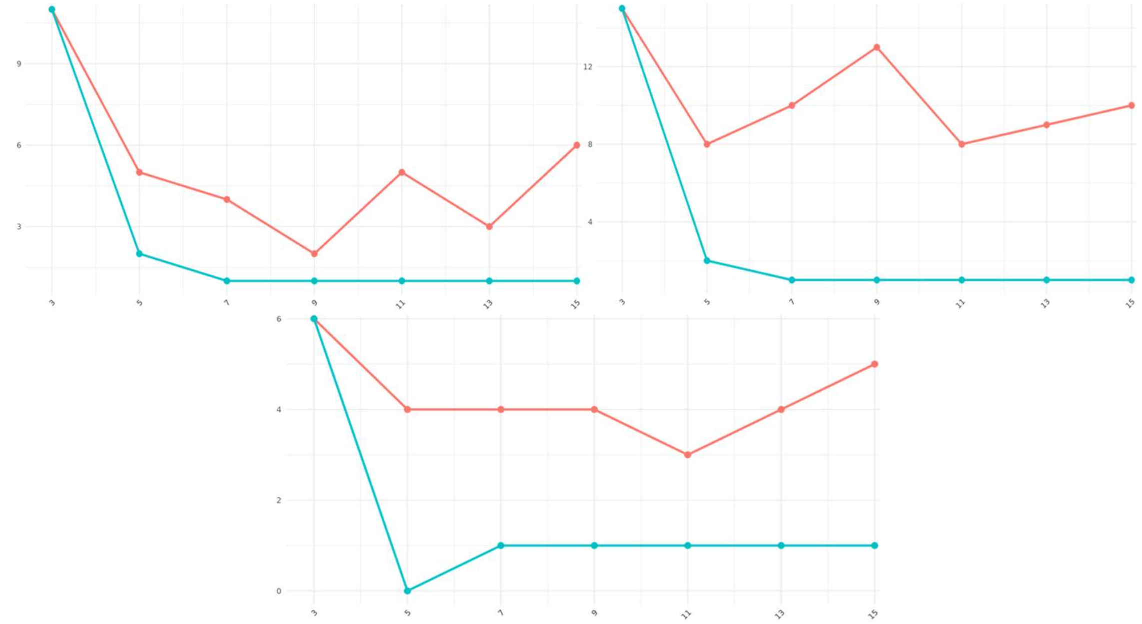


Figure 58: Diagram of the dependence of the number of turning points on the window size with median smoothing for datasets 1-3 (red – method A, green – method B)

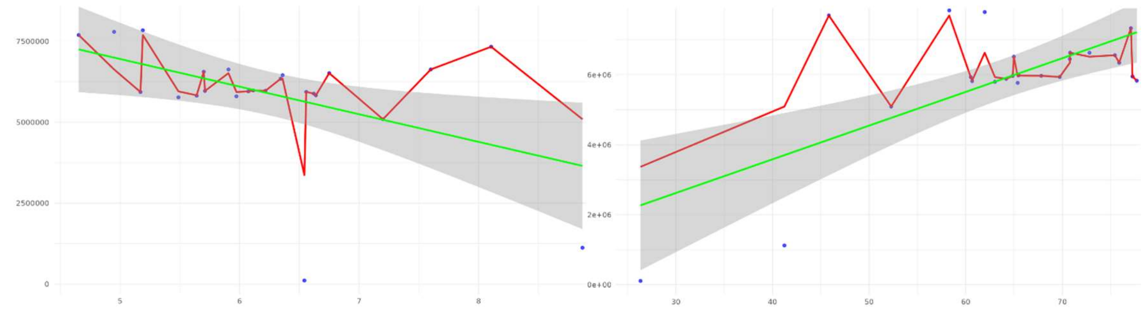


Figure 59: Correlation between shelling of civilian/political targets and the number of refugees using the median smoothing method (X is the average number of shelling per month, and Y is the average number of refugees), where blue dots are unsmoothed data and red dots are smoothed data

Table 11
Correlation coefficients and parameters according to data from Fig. 59-60

Name	59(a)	59(b)	60(a)	60(b)
Correlation coefficient (smoothed)	-0.308	0.508	-0.37	0.474
Coefficient of determination (R-squared),%	9.46	25.8	13.69	22.51
t	-1.5161	2.766	-1.8681	2.5278
df	22	22	22	22
p-value	0.1437	0.01127	0.07512	0.01916
95% confidence interval	-0.6324852	0.1315227	-0.67296683	0.08783653
	0.1094110	0.77563745	0.03923213	0.73680877
Sample estimates	-0.3075602	0.5079687	-0.3700206	0.474423

The graph in Fig. 60a shows a negative correlation - with an increase in the number of deaths, a decrease in the number of refugees is observed. Sharp fluctuations are especially noticeable at the beginning of the graph, which then smooths out. The confidence interval expands significantly with an increase in the number of cases—weak linear relationship. The modulus correlation coefficient is < 0.5, the coefficient of determination is less than 25%.

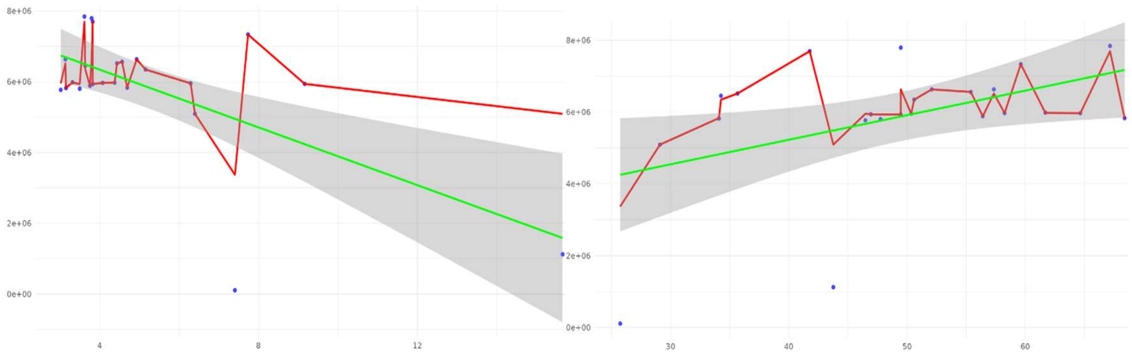


Figure 60: Correlation between fatalities caused by the shelling of civilian/political targets and the number of refugees using the median smoothing method (X is the average of deaths and Y is the average number of refugees), where blue dots are unsmoothed data and red dots are smoothed data

In Fig. 60b, a positive correlation is observed - the increase in the number of deaths correlates with the rise in the number of refugees. The trend is relatively stable, although the red line shows periodic fluctuations. The confidence interval remains relatively narrow in the middle part of the graph, which indicates a greater reliability of the forecast in this range—weak linear relationship. The correlation coefficient of the modulus < 0.5, the coefficient of determination is less than 25%.

The correlation ratio, according to Fig. 61a and Table 12, is 0.803. This value indicates a strong relationship between the variables. 80.3% of the variation in the dependent variable (number of refugees) can be explained by the shelling of civilian targets. It indicates a significant impact of the shelling of civilian targets on migration processes. The correlation coefficient, according to Fig. 61b and Table 12, is 0.753. The indicator demonstrates a strong connection between the shelling of political targets and the number of refugees. The shelling of political targets explains 75.3% of the variation in the number of refugees. This indicates a significant, although somewhat smaller compared to the first case, impact of political shelling.

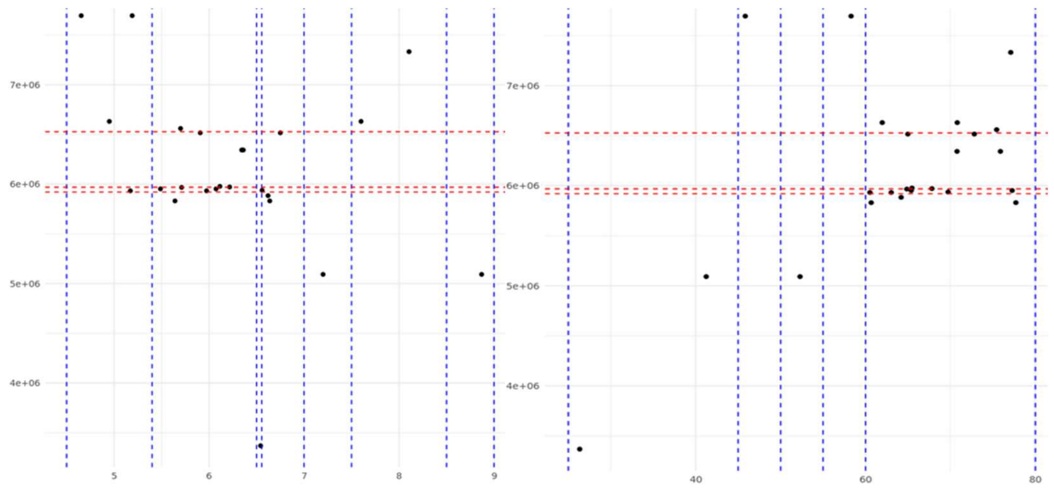


Figure 61: Division of the correlation field of the group between shelling of civilian/political targets and the number of refugees

Table 12
Correlation coefficients and parameters according to data from Fig. 61-62

Name	61(a)	61(b)	62(a)	62(b)
Group variance	591961516303.304	554998144511.364	424160636943.463	25337201030536.93
Total variance	737058240915.275	737058240915.275	737058240915.275	3005003198762.93
Correlation ratio	0.803	0.753	0.575	0.844

The correlation coefficient, according to Fig. 62a and Table 12, is 0.575. This value indicates a moderate relationship between fatalities from the shelling of civilian targets and the number of refugees. This factor can explain 57.5% of the variation. The relationship is less pronounced compared to previous indicators.

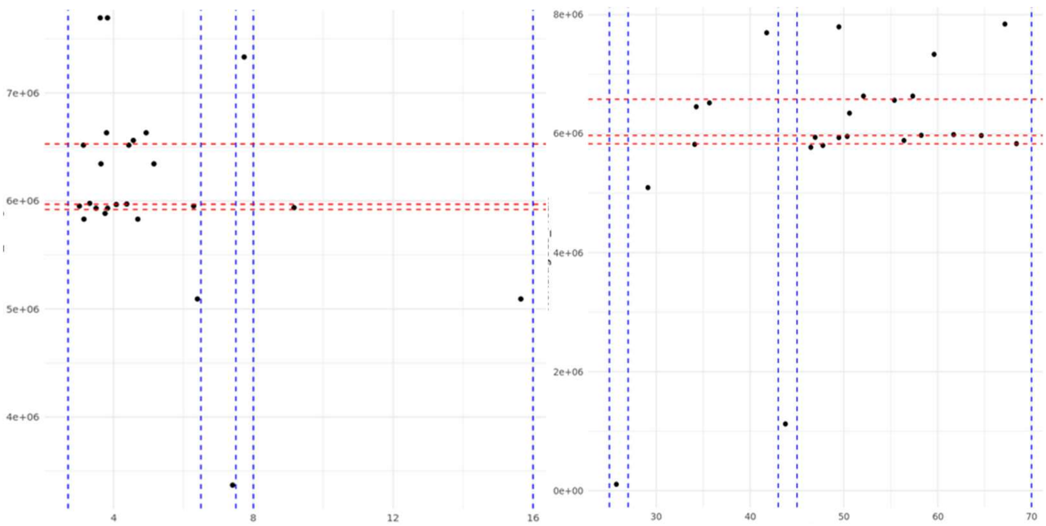
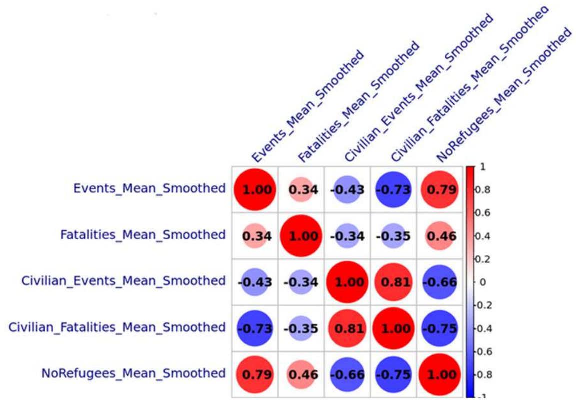


Figure 62: Dividing the correlation field between fatalities caused by shelling of civilian/political targets and the number of refugees

The correlation ratio, according to Fig. 62b and Table 12, is 0.844. The highest value of the correlation ratio among all graphs indicates a powerful relationship between fatalities from the shelling of political targets and the number of refugees. This factor can explain 84.4% of the variation in the number of refugees. It indicates the most significant impact of this indicator on migration processes. In Fig. 63, a strong positive correlation (0.793) is observed between total events (Events_Mean_Smoothed) and the number of refugees (NoRefugees_Mean_Smoothed), indicating that an increase in conflict events leads to a rise in the number of refugees. There is a robust positive correlation (0.814) between civilian events (Civilian_Events_Mean_Smoothed) and civilian fatalities (Civilian_Fatalities_Mean_Smoothed), which logically reflects a direct relationship between incidents and their consequences. There is a moderate negative correlation (-0.428) between total events and civilian events, which may indicate that not all conflict events are directly related to civilians. There is a strong negative correlation (-0.731) between total events and civilian fatalities, which may indicate that a significant proportion of events do not result in civilian casualties. It is noteworthy that the number of refugees has a negative correlation (-0.748) with civilian casualties, which may indicate that timely evacuation of the population (refugees) reduces the number of civilian casualties.




```
Events_Mean_Smoothed vs Fatalities_Mean_Smoothed :  0.342
Events_Mean_Smoothed vs Civilian_Events_Mean_Smoothed :  -0.428
Events_Mean_Smoothed vs Civilian_Fatalities_Mean_Smoothed :  -0.731
Events_Mean_Smoothed vs NoRefugees_Mean_Smoothed :  0.793
Fatalities_Mean_Smoothed vs Events_Mean_Smoothed :  0.342
Fatalities_Mean_Smoothed vs Civilian_Events_Mean_Smoothed :  -0.34
Fatalities_Mean_Smoothed vs Civilian_Fatalities_Mean_Smoothed :  -0.348
Fatalities_Mean_Smoothed vs NoRefugees_Mean_Smoothed :  0.462
Civilian_Events_Mean_Smoothed vs Events_Mean_Smoothed :  -0.428
Civilian_Events_Mean_Smoothed vs Fatalities_Mean_Smoothed :  -0.34
Civilian_Events_Mean_Smoothed vs Civilian_Fatalities_Mean_Smoothed :  0.814
Civilian_Events_Mean_Smoothed vs NoRefugees_Mean_Smoothed :  -0.662
Civilian_Fatalities_Mean_Smoothed vs Events_Mean_Smoothed :  -0.731
Civilian_Fatalities_Mean_Smoothed vs Fatalities_Mean_Smoothed :  -0.348
Civilian_Fatalities_Mean_Smoothed vs Civilian_Events_Mean_Smoothed :  0.814
Civilian_Fatalities_Mean_Smoothed vs NoRefugees_Mean_Smoothed :  -0.748
NoRefugees_Mean_Smoothed vs Events_Mean_Smoothed :  0.793
NoRefugees_Mean_Smoothed vs Fatalities_Mean_Smoothed :  0.462
NoRefugees_Mean_Smoothed vs Civilian_Events_Mean_Smoothed :  -0.662
NoRefugees_Mean_Smoothed vs Civilian_Fatalities_Mean_Smoothed :  -0.748
```

Figure 63: Correlation matrix

Overall, these correlations demonstrate a complex interdependence between different aspects of a conflict situation, where an increase in the total number of events leads to an increase in the number of refugees but not necessarily to the rise in civilian casualties, perhaps due to population evacuation. In Fig. 64, a very strong positive autocorrelation is observed for slight lags (0-3 days), where the coefficients exceed 0.9, which indicates a high inertia of the process in the short term. It means that the number of refugees on a given day is very strongly related to the number in the previous 1-3 days. With an increase in the time lag (from 4 to 10 days), a gradual decrease in the strength of the autocorrelation is observed - from 0.85 to 0.624, which indicates a weakening of the connection between observations with a larger time gap. A smooth, almost linear decrease in autocorrelation without sharp jumps indicates a stable nature of the migration process without sudden changes in trends. Even with a lag of 10 days, the autocorrelation remains moderately high (0.624), which indicates the presence of long-term trends in the migration process and the relative predictability of the dynamics of the number of refugees.

	Lag	Autocorrelation
1	0	1.000
2	1	0.966
3	2	0.927
4	3	0.889
5	4	0.850
6	5	0.809
7	6	0.768
8	7	0.733
9	8	0.698
10	9	0.662
11	10	0.624

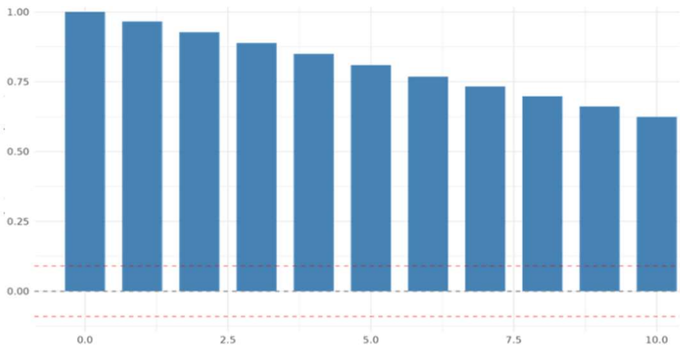


Figure 64: Autocorrelation correlogram for dataset 3 - Ukrainian refugees abroad

In general, this nature of the autocorrelation function is typical for mass migration processes. It indicates that changes in the number of refugees occur gradually, without sharp fluctuations. The current situation strongly depends on previous days, which is essential to consider when planning humanitarian assistance and developing appropriate policies.

	<int>	<dbl>	<dbl>
1	0	1	1
2	1	0.949	0.843
3	2	0.865	0.781
4	3	0.792	0.612
5	4	0.739	0.554
6	5	0.687	0.45
7	6	0.633	0.385
8	7	0.59	0.337
9	8	0.568	0.313
10	9	0.557	0.299
11	10	0.535	0.288

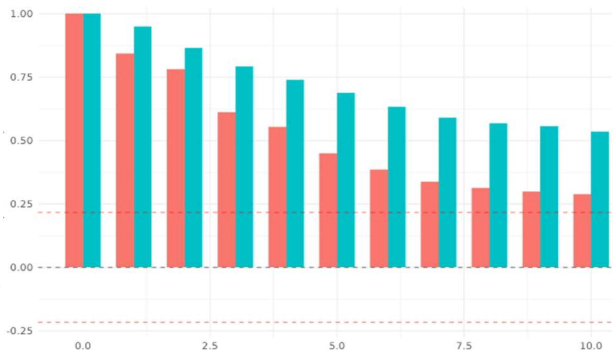


Figure 65: Autocorrelation correlogram for dataset 1 (red - victims and green - events)

The autocorrelation in Fig. 65 for both victims and events starts at a very high level (around 1.0) and gradually decreases over the 10 months. The correlation remains significant throughout the period, with events (blue) having a consistently higher autocorrelation than victims (red). It suggests a systematic and persistent nature for both events and victims, with events showing more predictable dynamics over time.

In Fig. 66, the autocorrelation pattern is noticeably different. Although both metrics start at high values, the correlation of events (blue) drops much faster and becomes very weak after 5 months. At the same time, the number of victims (red) maintains a moderately strong correlation throughout the period. It suggests that attacks on political infrastructure are more random or situational, while the number of victims resulting from them maintains a more consistent pattern over time.

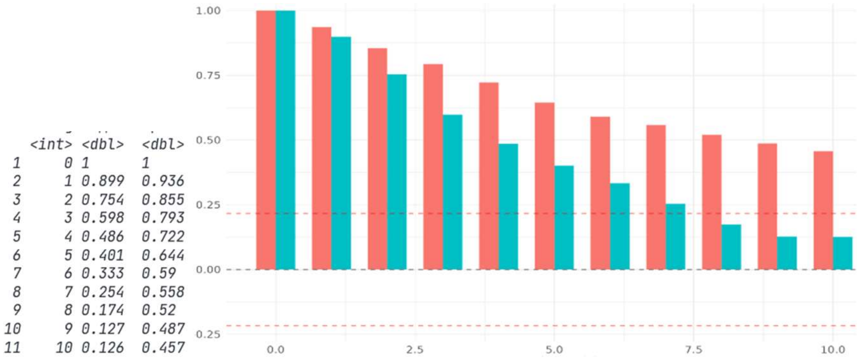


Figure 66: Autocorrelation correlogram for dataset 1 - (red - victims and green - events)

4.3. Hierarchical agglomerative cluster analysis of multidimensional data

The most significant number of shellings was recorded in Kharkiv, Kherson and Donetsk regions (Fig. 67). This can be seen from the values of "Amount" and "Maximum", which are the highest for these regions. Also, these regions have high average values ("Average"). Some regions, such as Volyn, Zakarpattia and Rivne, have the minimum number of recorded shellings. It is reflected by zero or minimum values in most columns. Significant deviations from the average ("Stand From") indicate an uneven distribution of shellings during the observation period. For example, Kyiv has a high standard deviation, which means periods of intense bombardment alternating with periods of relative calm. The "Median" and "Mode" indicators are often equal to 1, which indicates that, most often, one shelling was recorded during a specific period. However, for some regions, such as Kharkiv, Kherson, and Donetsk, these figures are higher, confirming the greater intensity of shelling in these regions.

City	Mean	Standard Deviation	Median	Mode	Standard Error	Variance	Kurdishness	Asymmetry	Range	Min	Max	Amount	Observations	Confidence
Cherkasy	1,048	0,218	1	1	0,048	0,048	19,05	4,249	1	1	2	22	21	0,099
Chernihiv	2,831	5,796	2	1	0,755	33,591	44,9	6,344	43	1	44	167	59	1,51
Chernivtsi	1,2	0,447	1	1	0,2	0,2	3,25	1,5	1	1	2	6	5	0,555
Crimea	1,071	0,261	1	1	0,04	0,068	12,077	3,328	1	1	2	45	42	0,081
Dnipropetrov:	3,359	4,048	1	1	0,327	16,39	7,018	2,119	18	1	19	514	153	0,647
Donetsk	6,766	8,359	2	1	0,455	69,871	5,275	1,685	40	1	41	2287	338	0,894
Ivano-Frankiv:	1,083	0,289	1	1	0,083	0,083	10,091	3,015	1	1	2	13	12	0,183
Kharkiv	8,875	9,616	6	1	0,68	92,462	7,272	1,956	51	1	52	1775	200	1,341
Kherson	14,126	16,451	7	1	1,416	270,648	7,289	1,92	82	1	83	1907	135	2,8
Khmelnytskyi	1,133	0,352	1	1	0,091	0,124	5,654	2,157	1	1	2	17	15	0,195
Kirovograd	1,067	0,258	1	1	0,067	0,067	13,071	3,474	1	1	2	16	15	0,143
Kyiv	4,197	17,15	1	1	2,111	294,13	59,638	7,538	138	1	139	277	66	4,216
Kyiv City	3,076	3,183	2	2	0,392	10,133	29,985	4,703	23	1	24	203	66	0,783
Luhansk	2,025	4,545	1	1	0,359	20,653	52,199	6,953	39	1	40	324	160	0,71
Lviv	1,091	0,292	1	1	0,051	0,085	9,1	2,846	1	1	2	36	33	0,104
Mykolaiv	4,766	4,662	3	1	0,583	21,738	6,06	1,814	21	1	22	305	64	1,165
Odesa	2,282	1,906	1	1	0,226	3,634	6,536	1,885	9	1	10	162	71	0,451
Poltava	1,083	0,408	1	1	0,083	0,167	22,043	4,587	2	1	3	26	24	0,172
Rivne	1	0	1	1	0	0	NA	NA	0	1	1	12	12	0
Sumy	4,161	5,815	2	1	0,549	33,812	20,183	3,81	39	1	40	466	112	1,089
Ternopil	1,111	0,333	1	1	0,111	0,111	7,125	2,475	1	1	2	10	9	0,256
Vinnytsia	1,111	0,323	1	1	0,076	0,105	7,125	2,475	1	1	2	20	18	0,161
Volyn	1	0	1	1	0	0	NA	NA	0	1	1	11	11	0
Zakarpattia	1	0	1	1	0	0	NA	NA	0	1	1	9	9	0
Zaporizhia	4,201	3,841	3	1	0,315	14,756	14,312	2,622	28	1	29	626	149	0,622
Zhytomyr	1,682	2,457	1	1	0,524	6,037	15,688	3,709	11	1	12	37	22	1,089

Figure 67: Indicators of Ukrainian cities according to descriptive statistics for dataset 1

Eastern and southern regions were most affected: Donetsk, Kharkiv, Zaporizhia, Sumy, Kherson, and Luhansk regions experienced the highest number of shelling (Fig. 68). Uneven distribution of shelling: The intensity of shelling fluctuated significantly, as evidenced by the high standard deviation for many regions. Frequency of shelling: Most often, one shelling was recorded during the observation period (Median and Mode often = 1).

Period of mass departure from Fig. 69 - April-September 2022 - the largest outflow, rapid growth in the number of refugees. High volatility of data at the beginning. Stabilization: October 2022 - January 2023 - the number of refugees remained relatively stable at a high level. Return: Since February 2023, there has been a trend towards the return of refugees to Ukraine. The number is gradually decreasing. High reliability: The data was fairly reliable throughout the period. A mass exodus of Ukrainians abroad characterized the first months of the war. Over time, the situation stabilized, and since the beginning of 2023, there has been a trend towards return. The data is generally reliable.

City	Mean	Standard Deviation	Median	Mode	Standard Error	Variance	Kurdi-shness	Asymmetry	Range	Min	Max	Amount	Observa-tions	Confidence
Cherkasy	1,404	0,712	1	1	0,104	0,507	3,517	1,438	2	1	3	66	47	0,209
Chernihiv	27,434	40,556	11	1	3,478	1644,781	7,255	2,162	200	1	201	3731	136	6,878
Chernivtsi	1,111	0,333	1	1	0,111	0,111	7,125	2,475	1	1	2	10	9	0,256
Crimea	1,524	1,055	1	1	0,095	1,113	11,079	2,667	6	1	7	189	124	0,188
Dnipropetı	13,278	24,11	3	1	1,64	581,299	8,064	2,425	118	1	119	2868	216	3,233
Donetsk	130,168	141,139	91	5	5,523	19920,11	6,926	1,949	756	1	757	85000	653	10,845
Ivano-Fran	1,357	0,488	1	1	0,092	0,238	1,356	0,596	1	1	2	38	28	0,189
Kharkiv	76,494	100,173	27	1	6,507	10034,62	4,681	1,582	443	1	444	18129	237	12,819
Kherson	69,812	97,409	13	1	7,24	9488,531	4,04	1,414	405	1	406	12636	181	14,287
Khmelnyts	2,571	1,65	2	1	0,279	2,723	4,469	1,187	7	1	8	90	35	0,567
Kirovograd	2,111	1,968	1	1	0,293	3,874	7,456	2,123	9	1	10	95	45	0,591
Kyiv	6,859	32,062	1	1	3,222	1027,959	76,43	8,344	303	1	304	679	99	6,395
Kyiv City	5,068	7,386	3	1	0,859	54,557	39,766	5,512	58	1	59	375	74	1,711
Luhansk	41,491	55,23	14	1	2,744	3050,389	7,56	1,857	362	1	363	16804	405	5,395
Lviv	1,397	0,672	1	1	0,082	0,452	3,643	1,419	2	1	3	95	68	0,163
Mykolaiv	31,372	51,588	7	1	5,321	2661,354	8,795	2,442	243	1	244	2949	94	10,566
Odesa	4,06	5,031	2	1	0,465	25,315	5,886	1,941	22	1	23	475	117	0,921
Poltava	1,803	1,276	1	1	0,163	1,627	7,803	2,122	6	1	7	110	61	0,327
Rivne	1,361	0,798	1	1	0,133	0,637	7,634	2,343	3	1	4	49	36	0,27
Sumy	85,201	115,649	38	1	9,474	13374,61	9,143	2,374	597	1	598	12695	149	18,722
Ternopil	1,211	0,535	1	1	0,123	0,287	7,901	2,443	2	1	3	23	19	0,258
Vinnytsia	1,525	0,905	1	1	0,143	0,82	4,871	1,709	3	1	4	61	40	0,29
Volyn	1,095	0,301	1	1	0,066	0,09	8,605	2,758	1	1	2	23	21	0,137
Zakarpattiɛ	1,059	0,243	1	1	0,059	0,059	15,062	3,75	1	1	2	18	17	0,125
Zaporizhia	73,824	130,944	18	1	9,576	17146,33	7,1	2,267	616	1	617	13805	187	18,891
Zhytomyr	2,812	6,205	1	1	0,896	38,496	24,273	4,704	36	1	37	135	48	1,802

Figure 68: Descriptive statistics of objects for dataset 2

City	Mean	Standard Deviation	Median	Mode	Standard Error	Variance	Kurdi-shness	Asymmetry	Range	Min	Max	Amount	Observa-tions	Confidence
2022-04	106650	30617,72	106650	85000	21650	9,37E+08	1	0	43300	85000	128300	213300	2	275089,3
2022-05	1120785	817929,6	1143057	165700	226852,8	6,69E+11	3,348	0,658	2866912	165700	3032612	14570201	13	494269,9
2022-06	5092764	230482,3	5113791	4774703	46096,47	5,31E+10	1,885	0,196	720742	4773142	5493884	1,27E+08	25	95138,43
2022-07	5937976	220750,5	5972391	5498558	43292,74	4,87E+10	2,175	-0,449	746314	5498558	6244872	1,54E+08	26	89163,07
2022-08	6631375	241868,5	6708922	6880285	51566,55	5,85E+10	1,777	-0,049	739498	6272752	7012250	1,46E+08	22	107238,5
2022-09	7332829	143919,6	7352374	7401475	30009,31	2,07E+10	2,616	-0,499	529892	7008505	7538397	1,69E+08	23	62235,49
2022-10	7695192	54958,1	7689105	7540367	11717,11	3,02E+09	3,898	-0,76	237056	7540367	7777423	1,69E+08	22	24367,05
2022-11	7841196	37859,42	7843714	7785514	9775,26	1,43E+09	2,067	0,032	116975	7785514	7902489	1,18E+08	15	20965,85
2022-12	7794427	358027,4	7881501	6276837	80057,36	1,28E+11	17,888	-4,098	1629626	6276837	7906463	1,56E+08	20	167562
2023-01	6343790	41435,48	6354771	6280218	8126,166	1,72E+09	1,882	-0,355	141132	6280218	6421350	1,65E+08	26	16736,15
2023-02	6452001	19216,98	6457707	6421484	4297,048	3,69E+08	1,812	-0,363	58173	6421484	6479657	1,29E+08	20	8993,824
2023-03	6516837	26123,06	6517826	6477863	5447,033	6,82E+08	1,704	-0,254	77898	6477863	6555761	1,5E+08	23	11296,46
2023-04	6560636	8857,043	6561828	6549287	1807,936	78447214	7,679	1,647	42804	6549287	6592091	1,57E+08	24	3740,001
2023-05	6631123	19824,11	6639172	6607874	4325,975	3,93E+08	1,668	0,212	57108	6607874	6664982	1,39E+08	21	9023,826
2023-06	5952921	290331,1	5827968	5759723	61898,8	8,43E+10	3,902	1,672	833970	5759723	6593693	1,31E+08	22	128725,6
2023-07	5769923	7874,935	5770252	5758340	1544,402	62014598	3,886	0,87	33697	5758340	5792037	1,5E+08	26	3180,755
2023-08	5801119	4952,439	5801389	5802110	971,253	24526653	3,839	0,05	22435	5791490	5813925	1,51E+08	26	2000,333
2023-09	5819495	7823,584	5816362	5812085	1749,407	61208469	2,686	0,906	25339	5812085	5837424	1,16E+08	20	3661,55
2023-10	5831668	6212,59	5831988	5822169	1295,415	38596280	2,931	0,711	23451	5822169	5845620	1,34E+08	23	2686,526
2023-11	5884629	22956,43	5894045	5896513	5266,566	5,27E+08	2,114	-0,926	64003	5846613	5910616	1,12E+08	19	11064,65
2023-12	5933238	13568,86	5936357	5910678	2769,732	1,84E+08	5,171	0,686	64381	5910678	5975059	1,42E+08	24	5729,627
2024-01	5971741	10045,25	5974106	5951697	4100,957	1,01E+08	3,897	-1,608	27057	5951697	5978754	35830445	6	10541,84
2024-02	5967267	8368,122	5966893	5953024	2415,669	70025464	1,853	-0,116	26987	5953024	5980011	71607198	12	5316,851
2024-03	5983959	4457,688	5982321	5979958	1409,645	19870984	2,667	1,088	12244	5979958	5992202	59839589	10	3188,838

Figure 69: Descriptive statistics of objects for dataset 3

The most significant deviation from the average (more shelling), according to Table 70, is Kherson (3.679), Kharkiv (1.953), and Donetsk (1.259) regions. It confirms that these regions experienced significantly more shelling than the average for Ukraine. Close to the average: Dnipropetrovsk (0.139), Kyiv (0.415), Zaporizhia (0.416), Sumy (0.403). The most significant deviation from the average (fewer shelling): Rivne (-0.637), Volyn (-0.637), and Zakarpattia (-0.637) regions. These regions experienced significantly less shelling than the average.

City	Mean	Standard Deviation	Median	Mode	Standard Error	Variance	Kurdishness	Asymmetry	Range	Min	Max	Amount	Observations	Confidence
Cherkasy	-0,621	-0,687	-0,475	-0,196	-0,665	-0,446	0,148	0,524	-0,636	0	-0,636	-0,532	-0,621	-0,675
Chernihiv	-0,035	0,481	0,175	-0,196	0,809	-0,008	1,803	1,752	0,682	0	0,682	-0,302	-0,14	0,808
Chernivtsi	-0,571	-0,639	-0,475	-0,196	-0,348	-0,444	-0,863	-1,087	-0,636	0	-0,636	-0,557	-0,824	-0,196
Crimea	-0,613	-0,678	-0,475	-0,196	-0,682	-0,446	-0,298	-0,016	-0,636	0	-0,636	-0,495	-0,355	-0,694
Dnipropetr	0,139	0,115	-0,475	-0,196	-0,083	-0,232	-0,622	-0,724	-0,103	0	-0,103	0,248	1,051	-0,099
Donetsk	1,259	1,017	0,175	-0,196	0,184	0,466	-0,734	-0,979	0,588	0	0,588	3,059	3,395	0,161
Ivano-Fran	-0,609	-0,672	-0,475	-0,196	-0,592	-0,445	-0,425	-0,199	-0,636	0	-0,636	-0,546	-0,735	-0,587
Kharkiv	1,953	1,28	2,776	-0,196	0,653	0,761	-0,606	-0,82	0,933	0	0,933	2,247	1,647	0,631
Kherson	3,679	2,711	3,427	-0,196	2,188	3,088	-0,605	-0,841	1,906	0	1,906	2,457	0,823	2,165
Khmelnyts	-0,593	-0,659	-0,475	-0,196	-0,576	-0,445	-0,709	-0,702	-0,636	0	-0,636	-0,54	-0,697	-0,574
Kirovograd	-0,615	-0,679	-0,475	-0,196	-0,626	-0,446	-0,235	0,07	-0,636	0	-0,636	-0,541	-0,697	-0,629
Kyiv	0,415	2,858	-0,475	-0,196	3,638	3,394	2,746	2,451	3,663	0	3,663	-0,127	-0,051	3,653
Kyiv City	0,046	-0,066	0,175	4,903	0,052	-0,314	0,848	0,79	0,054	0	0,054	-0,245	-0,051	0,044
Luhansk	-0,3	0,219	-0,475	-0,196	-0,017	-0,177	2,27	2,108	0,556	0	0,556	-0,053	1,14	-0,033
Lviv	-0,607	-0,672	-0,475	-0,196	-0,659	-0,445	-0,489	-0,298	-0,636	0	-0,636	-0,51	-0,469	-0,67
Mykolaiv	0,602	0,243	0,825	-0,196	0,451	-0,163	-0,683	-0,903	-0,008	0	-0,008	-0,083	-0,077	0,446
Odesa	-0,215	-0,334	-0,475	-0,196	-0,294	-0,399	-0,653	-0,861	-0,385	0	-0,385	-0,31	0,012	-0,305
Poltava	-0,609	-0,647	-0,475	-0,196	-0,592	-0,444	0,34	0,722	-0,605	0	-0,605	-0,525	-0,583	-0,598
Rivne	-0,637	-0,733	-0,475	-0,196	-0,766	-0,446	NA	NA	-0,667	0	-0,667	-0,548	-0,735	-0,779
Sumy	0,403	0,485	0,175	-0,196	0,38	-0,005	0,221	0,267	0,556	0	0,556	0,172	0,532	0,366
Ternopil	-0,6	-0,663	-0,475	-0,196	-0,534	-0,445	-0,615	-0,516	-0,636	0	-0,636	-0,551	-0,773	-0,51
Vinnytsia	-0,6	-0,665	-0,475	-0,196	-0,607	-0,445	-0,615	-0,516	-0,636	0	-0,636	-0,535	-0,659	-0,61
Volyn	-0,637	-0,733	-0,475	-0,196	-0,766	-0,446	NA	NA	-0,667	0	-0,667	-0,549	-0,748	-0,779
Zakarpattia	-0,637	-0,733	-0,475	-0,196	-0,766	-0,446	NA	NA	-0,667	0	-0,667	-0,552	-0,773	-0,779
Zaporizhia	0,416	0,071	0,825	-0,196	-0,108	-0,254	-0,155	-0,43	0,211	0	0,211	0,426	1	-0,125
Zhytomyr	-0,412	-0,218	-0,475	-0,196	0,327	-0,368	-0,067	0,207	-0,322	0	-0,322	-0,508	-0,609	0,366

Figure 70: Data normalization for Dataset 1

City	Mean	Standard Deviation	Median	Mode	Standard Error	Variance	Kurdishness	Asymmetry	Range	Min	Max	Amount	Observations	Confidence
Cherkasy	-0,603	-0,666	-0,438	-0,196	-0,699	-0,524	-0,526	-0,702	-0,692	0	-0,692	-0,383	-0,526	-0,7
Chernihiv	0,138	0,197	0,087	-0,196	0,396	-0,241	-0,282	-0,24	0,166	0	0,166	-0,167	0,107	0,397
Chernivtsi	-0,611	-0,675	-0,438	-0,196	-0,697	-0,524	-0,29	-0,041	-0,697	0	-0,697	-0,386	-0,796	-0,692
Crimea	-0,6	-0,659	-0,438	-0,196	-0,702	-0,524	-0,031	0,082	-0,675	0	-0,675	-0,375	0,021	-0,703
Dnipropetr	-0,265	-0,16	-0,333	-0,196	-0,2	-0,424	-0,229	-0,072	-0,189	0	-0,189	-0,218	0,675	-0,202
Donetsk	3,061	2,375	4,284	4,903	1,06	2,911	-0,303	-0,376	2,577	0	2,577	4,604	3,78	1,05
Ivano-Fran	-0,604	-0,671	-0,438	-0,196	-0,703	-0,524	-0,668	-1,238	-0,697	0	-0,697	-0,384	-0,661	-0,703
Kharkiv	1,534	1,488	0,926	-0,196	1,379	1,206	-0,45	-0,61	1,22	0	1,22	0,678	0,824	1,375
Kherson	1,344	1,428	0,192	-0,196	1,617	1,112	-0,492	-0,717	1,055	0	1,055	0,355	0,426	1,617
Khmelnyts	-0,57	-0,646	-0,385	-0,196	-0,642	-0,524	-0,464	-0,862	-0,671	0	-0,671	-0,381	-0,611	-0,641
Kirovograd	-0,583	-0,639	-0,438	-0,196	-0,638	-0,524	-0,268	-0,265	-0,662	0	-0,662	-0,381	-0,54	-0,637
Kyiv	-0,448	0,013	-0,438	-0,196	0,313	-0,347	4,248	3,7	0,613	0	0,613	-0,347	-0,156	0,318
Kyiv City	-0,499	-0,522	-0,333	-0,196	-0,454	-0,515	1,847	1,895	-0,449	0	-0,449	-0,364	-0,334	-0,453
Luhansk	0,538	0,514	0,244	-0,196	0,158	0,002	-0,262	-0,434	0,869	0	0,869	0,6	2,018	0,153
Lviv	-0,603	-0,667	-0,438	-0,196	-0,706	-0,524	-0,518	-0,714	-0,692	0	-0,692	-0,381	-0,377	-0,708
Mykolaiv	0,25	0,436	-0,123	-0,196	0,994	-0,065	-0,181	-0,062	0,353	0	0,353	-0,213	-0,192	1,004
Odesa	-0,527	-0,573	-0,385	-0,196	-0,582	-0,52	-0,371	-0,381	-0,606	0	-0,606	-0,359	-0,028	-0,583
Poltava	-0,592	-0,654	-0,438	-0,196	-0,68	-0,524	-0,246	-0,266	-0,675	0	-0,675	-0,38	-0,426	-0,681
Rivne	-0,604	-0,664	-0,438	-0,196	-0,689	-0,524	-0,257	-0,125	-0,688	0	-0,688	-0,384	-0,604	-0,69
Sumy	1,782	1,823	1,503	-0,196	2,342	1,782	-0,158	-0,105	1,888	0	1,888	0,359	0,199	2,346
Ternopil	-0,608	-0,67	-0,438	-0,196	-0,693	-0,524	-0,239	-0,061	-0,692	0	-0,692	-0,385	-0,725	-0,692
Vinnytsia	-0,6	-0,662	-0,438	-0,196	-0,686	-0,524	-0,438	-0,529	-0,688	0	-0,688	-0,383	-0,576	-0,687
Volyn	-0,612	-0,675	-0,438	-0,196	-0,711	-0,524	-0,193	0,14	-0,697	0	-0,697	-0,385	-0,711	-0,712
Zakarpattia	-0,613	-0,676	-0,438	-0,196	-0,713	-0,524	0,23	0,772	-0,697	0	-0,697	-0,385	-0,739	-0,714
Zaporizhia	1,458	2,154	0,454	-0,196	2,375	2,433	-0,292	-0,173	1,97	0	1,97	0,424	0,469	2,374
Zhytomyr	-0,563	-0,547	-0,438	-0,196	-0,442	-0,518	0,833	1,38	-0,545	0	-0,545	-0,379	-0,519	-0,438

Figure 71: Data normalization for Dataset 2

The most significant deviation from the average (more attacks on political infrastructure) according to Table 71 are Donetsk (3,061), Sumy (1,782), Kharkiv (1,534), Zaporizhia (1,458), Kherson (1,344) regions. These regions experienced significantly more attacks on political infrastructure than the average for Ukraine. Close to the average: Many areas have values close to zero, indicating that the number of attacks on political infrastructure is close to the average for the country. The most significant deviation from the average (fewer attacks on political infrastructure): Most western regions, as well as some central ones, such as Chernihiv, have negative values, indicating that the number of attacks on political infrastructure is lower than the average. The most significant outflows (significantly above average), according to Table 72, are mainly in the first months after the start of the full-scale invasion: April (-3.257), May (-2.684), September (0.824), October (1.029), November (1.111), December (1.085) 2022. April and May stand out in particular with tremendous negative values, indicating a sharp jump in the number of refugees immediately after the start of the war. The positive values from September to December show that the number of refugees remained significantly above average throughout the fall and early winter of 2022. Gradual stabilization and decline (close to average or below): Since the beginning of 2023, the "Average" values have been closer to 0, and since June 2023, they have been primarily negative, indicating a gradual return of refugees and a decrease in their number abroad relative to the average for the entire period.

City	Mean	Standard Deviation	Median	Mode	Standard Error	Variance	Kurdishness	Asymmetry	Range	Min	Max	Amount	Observations	Confidence
2022-04	-3,257	-0,427	-3,254	-3,048	-0,094	-0,319	-0,744	0,027	-0,509	-3,068	-3,695	-2,487	-2,672	1,847
2022-05	-2,684	3,836	-2,67	-3,004	4,152	4,554	-0,049	0,59	3,749	-3,024	-1,886	-2,191	-1	3,775
2022-06	-0,441	0,655	-0,432	-0,513	0,412	0,062	-0,482	0,195	0,513	-0,498	-0,353	0,135	0,823	0,264
2022-07	0,036	0,602	0,051	-0,122	0,354	0,03	-0,396	-0,358	0,551	-0,1	0,114	0,694	0,975	0,212
2022-08	0,428	0,717	0,467	0,625	0,525	0,101	-0,514	-0,015	0,541	0,325	0,592	0,519	0,367	0,371
2022-09	0,824	0,186	0,829	0,906	0,079	-0,175	-0,266	-0,401	0,225	0,728	0,92	0,988	0,519	-0,025
2022-10	1,029	-0,295	1,019	0,981	-0,299	-0,304	0,113	-0,624	-0,217	1,02	1,069	1,002	0,367	-0,358
2022-11	1,111	-0,388	1,106	1,114	-0,34	-0,315	-0,428	0,054	-0,398	1,154	1,147	-0,065	-0,697	-0,388
2022-12	1,085	1,346	1,127	0,298	1,115	0,609	4,252	-3,484	1,883	0,327	1,149	0,725	0,063	0,901
2023-01	0,265	-0,369	0,267	0,3	-0,374	-0,313	-0,483	-0,277	-0,362	0,329	0,224	0,912	0,975	-0,425
2023-02	0,326	-0,489	0,325	0,377	-0,453	-0,323	-0,504	-0,284	-0,487	0,406	0,261	0,171	0,063	-0,493
2023-03	0,363	-0,452	0,359	0,407	-0,429	-0,321	-0,536	-0,191	-0,457	0,437	0,308	0,601	0,519	-0,473
2023-04	0,388	-0,545	0,384	0,446	-0,504	-0,325	1,232	1,438	-0,51	0,477	0,331	0,757	0,671	-0,54
2023-05	0,428	-0,486	0,427	0,477	-0,452	-0,323	-0,546	0,208	-0,488	0,509	0,376	0,382	0,215	-0,493
2023-06	0,045	0,979	-0,03	0,019	0,739	0,289	0,115	1,459	0,683	0,044	0,332	0,211	0,367	0,56
2023-07	-0,059	-0,55	-0,062	0,018	-0,51	-0,325	0,11	0,772	-0,524	0,043	-0,168	0,604	0,975	-0,544
2023-08	-0,041	-0,566	-0,045	0,042	-0,522	-0,326	0,096	0,07	-0,541	0,061	-0,154	0,62	0,975	-0,555
2023-09	-0,031	-0,551	-0,036	0,047	-0,506	-0,325	-0,245	0,803	-0,536	0,072	-0,139	-0,09	0,063	-0,54
2023-10	-0,024	-0,559	-0,028	0,053	-0,515	-0,326	-0,173	0,636	-0,539	0,078	-0,134	0,276	0,519	-0,549
2023-11	0,006	-0,469	0,007	0,093	-0,433	-0,322	-0,414	-0,766	-0,478	0,091	-0,094	-0,185	-0,089	-0,475
2023-12	0,034	-0,52	0,031	0,101	-0,485	-0,325	0,49	0,614	-0,477	0,126	-0,054	0,447	0,671	-0,522
2024-01	0,055	-0,539	0,052	0,123	-0,457	-0,325	0,113	-1,351	-0,534	0,149	-0,051	-1,753	-2,064	-0,48
2024-02	0,053	-0,548	0,048	0,123	-0,492	-0,325	-0,491	-0,073	-0,534	0,15	-0,051	-1,014	-1,152	-0,526
2024-03	0,062	-0,569	0,057	0,138	-0,513	-0,326	-0,251	0,959	-0,556	0,164	-0,043	-1,257	-1,456	-0,544

Figure 72: Data normalization for Dataset 3

According to Fig. 73, the Kharkiv region is most similar to Donetsk (3.44), Sumy (4.26), and Kherson (4.26). It confirms that these regions, which are located in the east and south of Ukraine, experienced similar intensity and nature of shelling. Kherson region: Most identical to Kharkiv (4.26), Donetsk (6.68) and Mykolaiv (7.31). Again, these are regions that are relatively close to each other and experience intense shelling. Donetsk region: Most similar to Kharkiv (3.44) and Luhansk (6.51). These are neighbouring regions that were the epicentre of hostilities. Sumy region: Most identical to Kharkiv (4.26) and Chernihiv (3.83). Western regions (Zakarpattia, Volyn, Rivne, Ternopil, Ivano-Frankivsk, Chernivtsi) show the highest values, with most of the eastern and southern regions. This means that the nature of shelling in the western regions was significantly different from that of shelling in the east and south, which is reasonably expected, given the geographical location and intensity of hostilities in other regions.

	Cherkasy	Chernihiv	Chernivtsi	Crimea	Dnipropetr	Donetsk	Ivano-Fran	Kharkiv	Kherson	Khmelnyts	Kirovograd	Kyiv	Kyiv City
Cherkasy	0	3,8339858	1,9985029	0,7506237	2,8389551	6,6459596	0,9368527	6,7917755	10,016456	1,5043457	0,6019958	10,655889	5,4974112
Chernihiv	3,8339858	0	4,8657541	4,2271197	4,1496811	6,3797014	4,3404348	5,8513254	8,1687466	4,7144441	4,1526767	7,2893839	5,4948484
Chernivtsi	1,9985029	4,8657541	0	1,4328015	2,4736354	6,4538657	1,0969872	6,5413462	9,6870356	0,6197677	1,4204686	11,04183	5,9551274
Crimea	0,7506237	4,2271197	1,4328015	0	2,3782931	6,3182171	0,4649957	6,5706309	9,8910321	0,8859481	0,371229	10,881505	5,6037504
Dnipropetr	2,8389551	4,1496811	2,4736354	2,3782931	0	4,1886973	2,5346225	4,8769609	8,2957677	2,4293254	2,6166591	9,9907976	5,6981537
Donetsk	6,6459596	6,3797014	6,4538657	6,3182171	4,1886973	0	6,5199236	3,4407056	6,685505	6,4327364	6,555376	10,101165	7,6339571
Ivano-Fran	0,9368527	4,3404348	1,0969872	0,4649957	2,5346225	6,5199236	0	6,6519043	9,882039	0,5796508	0,3360595	10,899626	5,670796
Kharkiv	6,7917755	5,8513254	6,5413462	6,5706309	4,8769609	3,4407056	6,6519043	0	4,2651992	6,5907132	6,69344	9,2606777	7,4444293
Kherson	10,016456	8,1687466	9,6870356	9,8910321	8,2957677	6,685505	9,882039	4,2651992	0	9,8337413	9,9276926	8,1235881	9,8902134
Khmelnyts	1,5043457	4,7144441	0,6197677	0,8859481	2,4293254	6,4327364	0,5796508	6,5907132	9,8337413	0	0,9094344	11,098546	5,83803
Kirovograd	0,6019958	4,1526767	1,4204686	0,371229	2,6166591	6,555376	0,3360595	6,69344	9,9276926	0,9094344	0	10,813932	5,5958421
Kyiv	10,655889	7,2893839	11,04183	10,881505	9,9907976	10,101165	10,899626	9,2606777	8,1235881	11,098546	10,813932	0	10,352139
Kyiv City	5,4974112	5,4948484	5,9551274	5,6037504	5,6981537	7,6339571	5,670796	7,4444293	9,8902134	5,83803	5,5958421	10,352139	0
Luhansk	3,8746014	2,0140668	5,2909147	4,2699183	4,1912106	6,1079025	4,5589914	6,4199874	9,1994227	4,9757136	4,311264	8,2612522	5,6895749
Lviv	1,051439	4,437464	1,1020045	0,3611149	2,3518882	6,3385095	0,3119792	6,5642582	9,8675731	0,530066	0,5057361	10,986459	5,6935829
Mykolaiv	3,2628224	3,9244741	2,5712036	2,940856	1,9858127	4,9645883	2,8681857	4,221449	7,3122369	2,7439708	2,9911839	9,5718186	5,6934473
Odesa	1,9240572	4,3459447	1,103479	1,3098779	1,4182475	5,5066355	1,2922302	5,8036487	9,1046875	1,0547559	1,489954	10,556117	5,6730438
Poltava	0,30412	3,5936163	2,2368203	1,0123196	2,9122203	6,6520743	1,2081722	6,7837392	9,9770587	1,7732888	0,8798204	10,472779	5,4337244
Rivne	0,2121089	3,6634948	0,7920576	0,4414165	2,8074004	7,075844	0,2929744	7,2773175	10,817157	0,3221739	0,238888	11,116791	5,9452827
Sumy	3,1548341	2,4468827	3,4486777	3,1106408	1,9868689	4,5073461	3,2293806	4,2696419	7,2935256	3,3730345	3,1808089	8,3610751	5,3312329
Ternopil	1,3162971	4,5507899	0,7244253	0,7645286	2,4791055	6,4904293	0,3840742	6,612758	9,8280645	0,2350617	0,7188734	10,984763	5,7685894
Vinnytsia	1,2937299	4,5880536	0,8094745	0,6764059	2,4257508	6,4291665	0,3786423	6,5973579	9,8541971	0,2173545	0,7002999	11,044058	5,7731746
Volyn	0,2206441	3,6660626	0,7904637	0,4546396	2,8173974	7,0852989	0,2933206	7,2827416	10,819675	0,3242926	0,241733	11,117777	5,9471035
Zakarpattia	0,2386437	3,6712328	0,7880883	0,4803539	2,8370739	7,1041629	0,2958986	7,2937195	10,824915	0,3300657	0,2494103	11,119772	5,9507618
Zaporizhia	3,1577989	3,6652488	3,1508738	2,8828583	1,5187008	4,0377146	3,0484329	3,8657537	7,5063716	3,0539761	3,0628061	9,6452089	5,5438506
Zhytomyr	1,6377109	3,1261593	1,8797971	1,6627486	2,3270486	6,1450131	1,5846564	6,0854115	9,0189882	1,8375486	1,5527608	9,5514878	5,3551651

	Luhansk	Lviv	Mykolaiv	Odesa	Poltava	Rivne	Sumy	Ternopil	Vinnytsia	Volyn	Zakarpattia	Zaporizhia	Zhytomyr
Cherkasy	3,8746014	1,051439	3,2628224	1,9240572	0,30412	0,2121089	3,1548341	1,3162971	1,2937299	0,2206441	0,2386437	3,1577989	1,6377109
Chernihiv	2,0140668	4,437464	3,9244741	4,3459447	3,5936163	3,6634948	2,4468827	4,5507899	4,5880536	3,6660626	3,6712328	3,6652488	3,1261593
Chernivtsi	5,2909147	1,1020045	2,5712036	1,103479	2,2368203	0,7920576	3,4486777	0,7244253	0,8094745	0,7904637	0,7880883	3,1508738	1,8797971
Crimea	4,2699183	0,3611149	2,940856	1,3098779	1,0123196	0,4414165	3,1106408	0,7645286	0,6764059	0,4546396	0,4803539	2,8828583	1,6627486
Dnipropetr	4,1912106	2,3518882	1,9858127	1,4182475	2,9122203	2,8074004	1,9868689	2,4791055	2,4257508	2,8173974	2,8370739	1,5187008	2,3270486
Donetsk	6,1079025	6,3385095	4,9645883	5,5066355	6,6520743	7,075844	4,5073461	6,4904293	6,4291665	7,0852989	7,1041629	4,0377146	6,1450131
Ivano-Fran	4,5589914	0,3119792	2,8681857	1,2922302	1,2081722	0,2929744	3,2293806	0,3840742	0,3786423	0,2933206	0,2958986	3,0484329	1,5846564
Kharkiv	6,4199874	6,5642582	4,221449	5,8036487	6,7837392	7,2773175	4,2696419	6,612758	6,5973579	7,2827416	7,2937195	3,8657537	6,0854115
Kherson	9,1994227	9,8675731	7,3122369	9,1046875	9,9770587	10,817157	7,2935256	9,8280645	9,8541971	10,819675	10,824915	7,5063716	9,0189882
Khmelnyts	4,9757136	0,530066	2,7439708	1,0547559	1,7732888	0,3221739	3,3730345	0,2350617	0,2173545	0,3242926	0,3300657	3,0539761	1,8375486
Kirovograd	4,311264	0,5057361	2,9911839	1,489954	0,8798204	0,238888	3,1808089	0,7188734	0,7002999	0,241733	0,2494103	3,0628061	1,5527608
Kyiv	8,2612522	10,986459	9,5718186	10,556117	10,472779	11,116791	8,3610751	10,984763	11,044058	11,117777	11,119772	9,6452089	9,5514878
Kyiv City	5,6895749	5,6935829	5,6934473	5,6730438	5,4337244	5,9452827	5,3312329	5,7685894	5,7731746	5,9471035	5,9507618	5,5438506	5,3551651
Luhansk	0	4,5627034	4,7809363	4,5805366	3,6193683	3,2334969	3,0564407	4,8365761	4,8107549	3,2424883	3,2600729	3,880257	3,7948248
Lviv	4,5627034	0	2,8415503	1,1303548	1,3240419	0,3450889	3,2030991	0,4459282	0,3263847	0,3569902	0,3804335	2,9195102	1,7162363
Mykolaiv	4,7809363	2,8415503	0	2,0521969	3,3251675	3,1963718	1,9532752	2,7506228	2,7855845	3,199693	3,2064277	1,6517954	2,294976
Odesa	4,5805366	1,1303548	2,0521969	0	2,1021111	1,3493569	2,6162729	1,1309217	1,0946383	1,3580047	1,3751083	2,2365697	1,6750854
Poltava	3,6193683	1,3240419	3,3251675	2,1021111	0	0,3458791	3,0726098	1,5794977	1,5662829	0,3528385	0,3674611	3,1724383	1,6107452
Rivne	3,2334969	0,3450889	3,1963718	1,3493569	0,3458791	0	3,569574	0,3980808	0,281046	0,0140831	0,0412715	3,3888067	1,8965731
Sumy	3,0564407	3,2030991	1,9532752	2,6162729	3,0726098	3,569574	0	3,3048831	3,3080806	3,575216	3,5864498	1,5113127	2,2601117
Ternopil	4,8365761	0,4459282	2,7506228	1,1309217	1,5794977	0,3980808	3,3048831	0	0,169071	0,3968715	0,3959474	3,0625261	1,6724129
Vinnytsia	4,8107549	0,3263847	2,7855845	1,0946383	1,5662829	0,281046	3,3080806	0,169071	0	0,2855185	0,2958888	3,0189858	1,7576883
Volyn	3,2424883	0,3569902	3,199693	1,3580047	0,3528385	0,0140831	3,575216	0,3968715	0,2855185	0	0,0271968	3,3969266	1,8976573
Zakarpattia	3,2600729	0,3804335	3,2064277	1,3751083	0,3674611	0,0412715	3,5864498	0,3959474	0,2958888	0,0271968	0	3,4130106	1,9000627
Zaporizhia	3,880257	2,9195102	1,6517954	2,2365697	3,1724383	3,3888067	1,5113127	3,0625261	3,0189858	3,3969266	3,4130106	0	2,7102279
Zhytomyr	3,7948248	1,7162363	2,294976	1,6750854	1,6107452	1,8965731	2,2601117	1,6724129	1,7576883	1,8976573	1,9000627	2,7102279	0

Figure 73: Proximity matrix for dataset 1

According to Fig. 74, the Kharkiv region has the most significant similarity with Donetsk (3.46), Sumy (7.29) and Luhansk (3.82). The similarity with Donetsk remains very high, which is expected since these regions were on the front line. However, unlike the shelling of civilian infrastructure, the similarity with the Kherson region is lower here. Kherson region: The most significant similarity with Mykolaiv (3.00), Zaporizhia (3.37) and Dnipropetrovsk (2.44). Shifting emphasis to southern regions may indicate a different nature of attacks on political infrastructure in this region. Donetsk region: Most similar to Kharkiv (3.46) and Luhansk (4.09). As in the previous analysis, similarities with neighbouring regions remain high. Sumy region: Most identical to Chernihiv (2.02) and Kharkiv (7.29). Western regions (Zakarpattia, Volyn, Rivne, Ternopil, Ivano-Frankivsk, Chernivtsi) again show the highest values, with most of the eastern and southern regions. It confirms that the nature of attacks on political infrastructure in the western regions was significantly different from the nature of attacks in the east and south.

	Cherkasy	Chernihiv	Chernivtsi	Crimea	Dnipropeti	Donetsk	Ivano-Fran	Kharkiv	Kherson	Khmelnyts	Kirovograd	Kyiv	Kyiv City	Luhansk	Lviv	Mykolaiv	Odesa	Poltava	Rivne	Sumy	Ternopil	Vinnytsia	Volyn	Zakarpattia	Zaporizhia	Zhytomyr
Cherkasy	0	2,49893	0,7521895	1,076851	1,8300445	12,404273	0,5707775	5,7408657	5,4348627	0,2204155	0,518027	6,9490855	3,5629416	4,1654923	0,15009	3,2977136	0,6595953	0,5292164	0,6415699	7,3739462	0,7300658	0,201415	0,924436	1,6703973	7,5357113	2,5249511
Chernihiv	2,49893	0	2,5397327	2,383402	1,3498415	10,751873	2,7108954	3,457032	3,1649942	2,4601098	2,3556649	6,1010597	3,5636796	2,3931216	2,4737567	1,0364044	2,1277906	2,3829954	2,4559082	5,0278249	2,5063374	2,4624486	2,5480257	2,7722035	5,2607918	2,8482007
Chernivtsi	0,7521895	2,5397327	0	0,866782	1,8989339	12,505776	1,2625942	5,8455872	5,5392472	0,866531	0,356731	6,403381	2,9690116	4,344135	0,8252357	3,2609066	0,8801517	0,4378093	0,2131033	7,3788587	0,0902386	0,5559272	0,2235911	0,9671422	7,560462	1,8845745
Crimea	1,076851	2,383402	0,866782	0	1,3685514	12,216273	1,6169425	5,6794101	5,431566	1,2204257	0,7076821	6,1097012	2,6844424	3,8615064	1,0148483	3,2015798	0,6216896	0,6068113	0,6966692	7,2966553	0,788222	0,9467323	0,7530352	1,0599948	7,4557364	1,7054296
Dnipropeti	1,8300445	1,3498415	1,8989339	1,3685514	0	11,286028	2,1867634	4,5064753	4,2809335	1,8852979	1,6497518	6,0728135	3,1104758	2,5906202	1,7433778	2,2394508	1,2327737	1,6064224	1,7413268	6,2054324	1,8374033	1,7886959	1,85449	2,0976482	6,3374187	2,3051835
Donetsk	12,404273	10,751873	12,505776	12,216273	11,286028	0	12,486187	8,4510088	9,1948675	12,37569	12,357056	12,764877	12,4246	9,3384676	12,355094	10,720636	12,078836	12,34247	12,423197	8,4946796	12,473919	12,409483	12,486254	12,5503	8,7299018	12,384515
Ivano-Fran	0,5707775	2,7108954	1,2625942	1,6169425	2,1867634	12,486187	0	5,8192168	5,4961638	0,4461031	1,0648357	7,4052489	4,0669502	4,3342251	0,6146885	3,4831991	1,1348714	1,0865188	1,188072	7,4676946	1,2544939	0,7507403	1,4585044	2,2029335	7,6252959	3,054302
Kharkiv	5,7408657	3,457032	5,8455872	5,6794101	4,5064753	8,4510088	5,8192168	0	1,0163149	5,6777864	5,6739423	7,4849907	6,2426491	2,9298462	5,7132402	3,075589	5,4222718	5,6966586	5,7720727	2,1056512	5,8181592	5,7375321	5,8587943	6,0150915	2,3872863	5,8933737
Kherson	5,4348627	3,1649942	5,5392472	5,431566	4,2809335	9,1948675	5,4961638	1,0163149	0	5,364044	5,3712443	7,3706602	6,0504464	3,1286551	5,417231	2,577839	5,1514937	5,4063921	5,4747837	2,344777	5,5157637	5,4310094	5,5649792	5,7423654	2,3491203	5,6354587
Khmelnyts	0,2204155	2,4601098	0,866531	1,2204257	1,8852979	12,37569	0,4461031	5,6777864	5,364044	0	0,6348913	6,9882251	3,6347278	4,1661981	0,3055847	3,2443791	0,776525	0,6658821	0,7717208	7,3009463	0,8462571	0,3484092	1,0507278	1,784733	7,4740591	2,6158932
Kirovograd	0,518027	2,3556649	0,356731	0,7076821	1,6497518	12,357056	1,0648357	5,6739423	5,3712443	0,6348913	0	6,465789	3,0613479	4,0997329	0,550744	3,1263847	0,556178	0,133514	0,1779157	7,2563693	0,2934859	0,3269098	0,4630853	1,1744203	7,4283462	2,0082246
Kyiv	6,9490855	6,1010597	6,403381	6,1097012	6,0728135	12,764877	7,4052489	7,4849907	7,3706602	6,9882251	6,465789	0	3,5806832	6,7116008	6,947157	5,9751465	6,5554079	6,4651429	6,4067923	7,9085599	6,3690736	6,7767263	6,2272622	5,5805032	8,0931733	4,6213605
Kyiv City	3,5629416	3,5636796	2,9690116	2,6844424	3,1104758	12,4246	4,0669502	6,2426491	6,0504464	3,6347278	3,0613479	3,5806832	0	4,8268243	3,5625578	3,9008876	3,2065647	3,0498657	2,9746339	7,3993979	2,9339197	3,379298	2,7726552	2,084781	7,6234433	1,1672305
Luhansk	4,1654923	2,3931216	4,344135	3,8615064	2,5906202	9,3384676	4,3342251	2,9298462	3,1286551	4,1661981	4,0997329	6,7116008	4,8268243	0	4,0794636	2,8060344	3,6803467	4,0633252	4,2010534	4,8120914	4,2898219	4,1738927	4,3171776	4,4906032	4,9256056	4,4346308
Lviv	0,15009	2,4737567	0,8252357	1,0148483	1,7433778	12,355094	0,6146885	5,7132402	5,417231	0,3055847	0,550744	6,947157	3,5625578	4,0794636	0	3,2954878	0,5639264	0,528548	0,6835459	7,3677214	0,7910935	0,2848456	0,9730103	1,7026673	7,5243967	2,5361496
Mykolaiv	3,2977136	1,0364044	3,2609066	3,2015798	2,2394508	10,720636	3,4831991	3,075589	2,577839	3,2443791	3,1263847	5,9751465	3,9008876	2,8060344	3,2954878	0	2,9875172	3,1765587	3,2136333	4,3748187	3,2392559	3,2477295	3,2686152	3,3992443	4,5677666	3,3057533
Odesa	0,6595953	2,1277906	0,8801517	0,6216896	1,2327737	12,078836	1,1348714	5,4222718	5,1514937	0,776525	0,556178	6,5554079	3,2065647	3,6803467	0,5639264	2,9875172	0	0,4796207	0,6814059	7,0847394	0,8149221	0,615165	0,9168359	1,506161	7,2435997	2,201152
Poltava	0,5292164	2,3829954	0,4378093	0,6068113	1,6064224	12,34247	1,0865188	5,6966586	5,4063921	0,6658821	0,133514	6,4651429	3,0498657	4,0633252	0,528548	3,1765587	0,4796207	0	0,2290109	7,2979507	0,3645271	0,3592771	0,5026191	1,1857846	7,4661357	2,0109764
Rivne	0,6415699	2,4559082	0,2131033	0,6966692	1,7413268	12,423197	1,188072	5,7720727	5,4747837	0,7717208	0,1779157	6,4067923	2,9746339	4,2010534	0,6835459	3,2136333	0,6814059	0,2290109	0	7,3379936	0,1384413	0,4436215	0,2951034	1,0303121	7,5130768	1,9081908
Sumy	7,3739462	5,0278249	7,3788587	7,2966553	6,2054324	8,4946796	7,4676946	2,1056512	2,344777	7,3009463	7,2563693	7,9085599	7,3993979	4,8120914	7,3677214	4,3748187	7,0847394	7,2979507	7,3379936	0	7,3608114	7,3473162	7,3848403	7,448508	1,3615304	7,1945236
Ternopil	0,7300658	2,5063374	0,0902386	0,788222	1,8374033	12,473919	1,2544939	5,8181592	5,5157637	0,8462571	0,2934859	6,3690736	2,9339197	4,2898219	0,7910935	3,2392559	0,8149221	0,3645271	0,1384413	7,3608114	0	0,5301547	0,2086337	0,9565778	7,5408078	1,8590487
Vinnytsia	0,201415	2,4624486	0,5559272	0,9467323	1,7886959	12,409483	0,7507403	5,7375321	5,4310094	0,3484092	0,3269098	6,7767263	3,379298	4,1738927	0,2848456	3,2477295	0,615165	0,3592771	0,4436215	7,3473162	0,5301547	0	0,7263195	1,4722035	7,5143482	2,3323919
Volyn	0,924436	2,5480257	0,2235911	0,7530352	1,85449	12,486254	1,4585044	5,8587943	5,5649792	1,0507278	0,4630853	6,2272622	2,7726552	4,3171776	0,9730103	3,2686152	0,9168359	0,5026191	0,2951034	7,3848403	0,2086337	0,7263195	0	0,7610171	7,5666951	1,6851237
Zakarpattia	1,6703973	2,7722035	0,9671422	1,0599948	2,0976482	12,5503	2,2029335	6,0150915	5,7423654	1,784733	1,1744203	5,5805032	2,084781	4,4906032	1,7026673	3,3992443	1,506161	1,1857846	1,0303121	7,448508	0,9565778	1,4722035	0,7610171	0	7,6429035	0,9983541
Zaporizhia	7,5357113	5,2607918	7,560462	7,4557364	6,3374187	8,7299018	7,6252959	2,3872863	2,3491203	7,4740591	7,4283462	8,0931733	7,6234433	4,9256056	7,5243967	4,5677666	7,2435997	7,4661357	7,5130768	1,3615304	7,5408078	7,5143482	7,5666951	7,6429035	0	7,3942935
Zhytomyr	2,5249511	2,8482007	1,8845745	1,7054296	2,3051835	12,384515	3,054302	5,8933737	5,6354587	2,6158932	2,0082246	4,6213605	1,1672305	4,4346308	2,5361496	3,3057533	2,201152	2,0109764	1,9081908	7,1945236	1,8590487	2,3323919	1,6851237	0,9983541	7,3942935	0

Figure 74: Proximity matrix for dataset 2

According to Fig. 75, The first months after the start of the full-scale invasion (April-June 2022): April and May 2022 show a relatively high closeness value (9.45), which indicates a similarity of dynamics during this period (rapid growth in the number of refugees). June 2022 shows less similarity with these months, which may indicate the beginning of changes in dynamics. Summer-autumn 2022 (July-November): July, August, September, October and November 2022 show relatively low closeness values to each other, which indicates a similar dynamic during this period (relative stabilization and further growth). The similarity between July and August (1.28), as well as between September, October and November (values around 1-2), is especially noticeable. Winter 2022 - Spring 2023 (December 2022 - May 2023): This period is characterized by greater variability. December 2022 shows relatively low similarity with previous months, which may indicate the beginning of a new phase. Starting from January 2023, there is a tendency for the proximity values between neighbouring months to increase, although with some fluctuations. Summer 2023 - Spring 2024 (June 2023 - March 2024): Starting from June 2023, the proximity values decrease again, which indicates the formation of a new trend, different from the previous one. July and August 2023 have the lowest value (0), which indicates the identity of the dynamics.

	2022-04	2022-05	2022-06	2022-07	2022-08	2022-09	2022-10	2022-11	2022-12	2023-01	2023-02	2023-03	2023-04	2023-05	2023-06	2023-07	2023-08	2023-09	2023-10	2023-11	2023-12	2024-01	2024-02	2024-03
2022-04	0	9,4564525	8,0251293	9,0836515	9,7099128	10,573871	10,947258	10,564961	12,30445	9,6534418	9,2414664	9,5765964	10,048847	9,5482029	9,0568203	9,0542273	9,0637988	8,5327033	8,8036132	8,5036365	9,0786951	8,0426179	8,1183605	8,1197352
2022-05	9,4564525	0	10,056635	10,824658	11,155145	12,42436	13,236798	13,174638	12,141813	12,441772	12,384345	12,518385	12,771693	12,542374	10,200289	12,191883	12,245664	11,969878	12,103681	11,958047	12,175775	11,994983	11,898538	11,909404
2022-06	8,0251293	10,056635	0	1,2846505	2,2044598	3,1785048	3,9078021	4,221858	6,9841658	2,5616059	2,7811339	2,7303141	3,5645283	2,8679648	1,9978939	2,4491345	2,4234958	2,5014028	2,3855396	2,5863325	2,5435131	4,4741866	3,2889625	3,676793
2022-07	9,0836515	10,824658	1,2846505	0	1,3778926	2,0939207	2,8909808	3,5804673	6,2939224	1,8343358	2,3127566	2,0986205	3,2831148	2,3760684	2,1808964	2,3683726	2,1469977	2,6255363	2,3560811	2,3709608	2,3868406	4,5121823	3,401147	3,9599222
2022-08	9,7099128	11,155145	2,2044598	1,3778926	0	1,3798927	2,339582	2,6822778	6,2651152	2,1235277	2,2109163	2,0930843	3,1949997	2,1409094	1,943282	2,7852571	2,6758593	2,728193	2,625212	2,5947985	2,6727514	4,3428925	3,2326198	3,6630485
2022-09	10,573871	12,42436	3,1785048	2,0939207	1,3798927	0	1,0512012	2,0737975	6,129775	1,7181973	1,9370351	1,6615258	2,9007297	1,7994813	3,0014028	2,7621113	2,523959	2,8881248	2,6635454	2,5980668	2,5916018	4,4649635	3,4303424	3,9528476
2022-10	10,947258	13,236798	3,9078021	2,8909808	2,339582	1,0512012	0	1,7672249	6,1618935	1,9221217	1,9500321	1,7415223	2,7964637	1,8420035	3,7628863	2,9120606	2,6151633	3,0265546	2,8066621	2,6683311	2,6563708	4,3429942	3,4515688	3,9844862
2022-11	10,564961	13,174638	4,221858	3,5804673	2,6822778	2,0737975	1,7672249	0	7,057938	2,7352645	1,9771573	2,2027328	3,1495839	1,8585734	3,8032038	3,3228685	3,2157809	2,8000759	2,9331065	2,6883461	3,0320861	3,5639376	2,6226126	2,917017
2022-12	12,30445	12,141813	6,9841658	6,2939224	6,2651152	6,129775	6,1618935	7,057938	0	6,9746793	7,0314559	7,0483751	7,0951311	7,2541008	6,8618543	7,4646874	7,1011993	7,6601074	7,5116787	6,9351771	7,052253	7,1211145	7,5685398	8,0713063
2023-01	9,6534418	12,441772	2,5616059	1,8343358	2,1235277	1,7181973	1,9221217	2,7352645	6,9746793	0	1,2010899	0,6200048	2,4830479	1,1300748	3,1656517	1,4720465	1,0597118	1,8958473	1,4379416	1,715868	1,540877	4,2596548	2,9313954	3,537965
2023-02	9,2414664	12,384345	2,7811339	2,3127566	2,2109163	1,9370351	1,9500321	1,9771573	7,0314559	1,2010899	0	0,6421472	2,5927873	0,6050669	3,2268564	1,8069294	1,487005	1,4022354	1,3435807	0,9491575	1,633619	3,1821156	1,819917	2,5144771
2023-03	9,5765964	12,518385	2,7303141	2,0986205	2,0930843	1,6615258	1,7415223	2,2027328	7,0483751	0,6200048	0,6421472	0	2,4193786	0,5712136	3,1670638	1,5706047	1,2352202	1,5995327	1,3045275	1,4017807	1,5150327	3,8047269	2,4337381	3,0403983
2023-04	10,048847	12,771693	3,5645283	3,2831148	3,1949997	2,9007297	2,7964637	3,1495839	7,0951311	2,4830479	2,5927873	2,4193786	0	2,2447909	3,0290916	1,6845783	2,0529579	2,1374354	1,9345017	3,1289382	1,4030021	4,8369803	3,506401	3,3996704
2023-05	9,5482029	12,542374	2,8679648	2,3760684	2,1409094	1,7994813	1,8420035	1,8585734	7,2541008	1,1300748	0,6050669	0,5712136	2,2447909	0	3,0425368	1,6083345	1,482224	1,3297804	1,21572	1,5095178	1,4961688	3,6522527	2,1514962	2,613005
2023-06	9,0568203	10,200289	1,9978939	2,1808964	1,943282	3,0014028	3,7628863	3,8032038	6,8618543	3,1656517	3,2268564	3,1670638	3,0290916	3,0425368	0	2,8650396	3,1317781	2,8171006	2,837461	3,4713144	2,7941985	4,9510544	3,6853366	3,6242686
2023-07	9,0542273	12,191883	2,4491345	2,3683726	2,7852571	2,7621113	2,9120606	3,3228685	7,4646874	1,4720465	1,8069294	1,5706047	1,6845783	1,6083345	2,8650396	0	0,7041136	1,2018844	0,6484204	2,1062773	0,578549	4,4008154	2,877297	3,1005217
2023-08	9,0637988	12,245664	2,4234958	2,1469977	2,6758593	2,523959	2,6151633	3,2157809	7,1011993	1,0597118	1,487005	1,2352202	2,0529579	1,482224	3,1317781	0,7041136	0	1,410921	0,8488315	1,6669871	0,7818536	4,1158654	2,7574978	3,2244713
2023-09	8,5327033	11,969878	2,5014028	2,6255363	2,728193	2,8881248	3,0265546	2,8000759	7,6601074	1,8958473	1,4022354	1,5995327	2,1374354	1,3297804	2,8171006	1,2018844	1,410921	0	0,6127079	1,5968162	1,1229234	3,4782634	1,7866208	1,9333057
2023-10	8,8036132	12,103681	2,3855396	2,3560811	2,625212	2,6635454	2,8066621	2,9331065	7,5116787	1,4379416	1,3435807	1,3045275	1,9345017	1,21572	2,837461	0,6484204	0,8488315	0,6127079	0	1,6233918	0,7191982	3,8544412	2,2561093	2,5296379
2023-11	8,5036365	11,958047	2,5863325	2,3709608	2,5947985	2,5980668	2,6683311	2,6883461	6,9351771	1,715868	0,9491575	1,4017807	3,1289382	1,5095178	3,4713144	2,1062773	1,6669871	1,5968162	1,6233918	0	1,9262432	2,6454467	1,5260983	2,4622995
2023-12	9,0786951	12,175775	2,5435131	2,3868406	2,6727514	2,5916018	2,6563708	3,0320861	7,052253	1,540877	1,633619	1,5150327	1,4030021	1,4961688	2,7941985	0,578549	0,7818536	1,1229234	0,7191982	1,9262432	0	4,0412425	2,6264038	2,8478336
2024-01	8,0426179	11,994983	4,4741866	4,5121823	4,3428925	4,4649635	4,3429942	3,5639376	7,1211145	4,2596548	3,1821156	3,8047269	4,8369803	3,6522527	4,9510544	4,4008154	4,1158654	3,4782634	3,8544412	2,6454467	4,0412425	0	1,8383166	2,4684977
2024-02	8,1183605	11,898538	3,2889625	3,401147	3,2326198	3,4303424	3,4515688	2,6226126	7,5685398	2,9313954	1,819917	2,4337381	3,506401	2,1514962	3,6853366	2,877297	2,7574978	1,7866208	2,2561093	1,5260983	2,6264038	1,8383166	0	1,1297907
2024-03	8,1197352	11,909404	3,676793	3,9599222	3,6630485	3,9528476	3,9844862	2,917017	8,0713063	3,537965	2,5144771	3,0403983	3,3996704	2,613005	3,6242686	3,1005217	3,2244713	1,9333057	2,5296379	2,4622995	2,8478336	2,4684977	1,1297907	0

Figure 75: Proximity matrix for the dataset - Ukrainian refugees abroad

We use the nearest neighbour strategy to perform agglomerative hierarchical cluster analysis (Fig. 76). The distance between two groups is defined as the distance between the two nearest elements of these groups. This strategy is monotonic and firmly compresses the feature space, and its parameters are $\alpha_i = \alpha_j = 0.5,\ \beta = 0,\ \gamma = -0.5$.

1	19+23	d27	0.014
2	24+d27	d28	0.027
3	21+22	d29	0.169
4	1+d28	d30	0.212
5	10+d29	d31	0.217
6	11+d30	d32	0.239
7	d31+d32	d33	0.281
8	7+d33	d34	0.293
9	18+d34	d35	0.304
10	15+d35	d36	0.312
11	4+d36	d37	0.361
12	3+d37	d38	0.62
13	17+d38	d39	1.055
14	5+d39	d40	1.418
15	20+25	d41	1.511
16	d40+d41	d42	1.519
17	26+d42	d43	1.553
18	16+d43	d44	1.652
19	2+14	d45	2.014
20	d44+d45	d46	2.447
21	6+8	d47	3.441
22	d46+d47	d48	3.866
23	9+d48	d49	4.265
24	13+d49	d50	5.331
25	12+d50	d51	7.289

1	3+21	d27	0.09
2	11+18	d28	0.134
3	19+d27	d29	0.138
4	1+15	d30	0.15
5	d28+d29	d31	0.178
6	22+d30	d32	0.201
7	23+d31	d33	0.209
8	10+d32	d34	0.22
9	d33+d34	d35	0.327
10	7+d35	d36	0.446
11	17+d36	d37	0.48
12	4+d37	d38	0.607
13	24+d38	d39	0.761
14	26+d39	d40	0.998
15	8+9	d41	1.016
16	2+16	d42	1.036
17	13+d40	d43	1.167
18	5+d43	d44	1.233
19	d42+d44	d45	1.35
20	20+25	d46	1.362
21	d41+d46	d47	2.106
22	14+d45	d48	2.393
23	d47+d48	d49	2.578
24	12+d49	d50	3.581
25	6+d50	d51	8.451

1	12+14	d25	0.591
2	16+21	d26	0.599
3	11+d25	d27	0.626
4	18+19	d28	0.634
5	10+d27	d29	0.642
6	d26+d28	d30	0.671
7	17+d30	d31	0.729
8	20+d29	d32	0.982
9	6+7	d33	1.088
10	d31+d32	d34	1.097
11	23+24	d35	1.169
12	3+4	d36	1.33
13	5+d36	d37	1.426
14	d33+d37	d38	1.428
15	13+d34	d39	1.452
16	d35+d39	d40	1.58
17	d38+d40	d41	1.72
18	8+d41	d42	1.829
19	22+d42	d43	1.903
20	15+d43	d44	2.011
21	9+d44	d45	6.345
22	1+d45	d46	8.307
23	2+d46	d47	9.788

Figure 76: Cluster analysis results for datasets 1-3

First steps (1-10) for dataset 1 (Fig. 76a): Mergers occur at relatively small metric values (from 0.014 to 0.312). It means that at the beginning of the algorithm, regions with very similar shelling patterns are merged. Middle steps (11-17): Metric values begin to increase (from 0.361 to 1.553). It means that less similar clusters are merged. Last steps (18-25): A significant increase in metric values is observed (from 1.652 to 7.289). It indicates the merging of large and relatively heterogeneous clusters. The last step (25) stands out in particular, where the metric value reaches 7.289. It means that at this step, two large clusters with very different natures of shelling were combined, which actually completes the process of hierarchical clustering (Fig. 77).

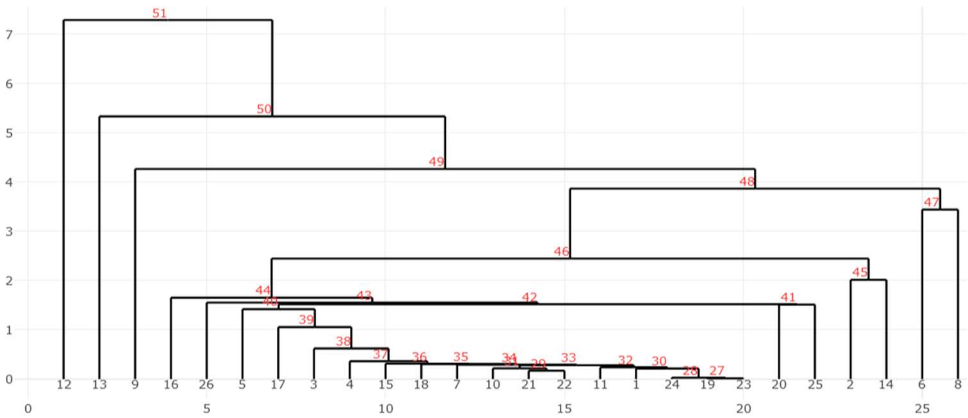


Figure 77: Dendrogram for dataset 1

First steps (1-10) for dataset 2 (Fig. 76b): Mergers occur at relatively small metric values (from 0.09 to 0.446). It indicates that at the beginning of the algorithm, regions with very similar patterns of shelling of political infrastructure are merged. Middle steps (11-17): Metric values gradually increase (from 0.48 to 1.167). It means that less similar clusters or individual regions with clusters are merged. The growth rate of the metric here is smoother than in the analysis of the shelling of civilian infrastructure. Last steps (18-25): A faster growth of metric values is observed (from 1.233 to 8.451). It indicates the merging of large and relatively heterogeneous clusters. As in the previous case, the last step (25) is characterized by a tremendous metric value (8.451), which indicates the combination of two large clusters with the most different nature of attacks on the political infrastructure (Fig. 78).

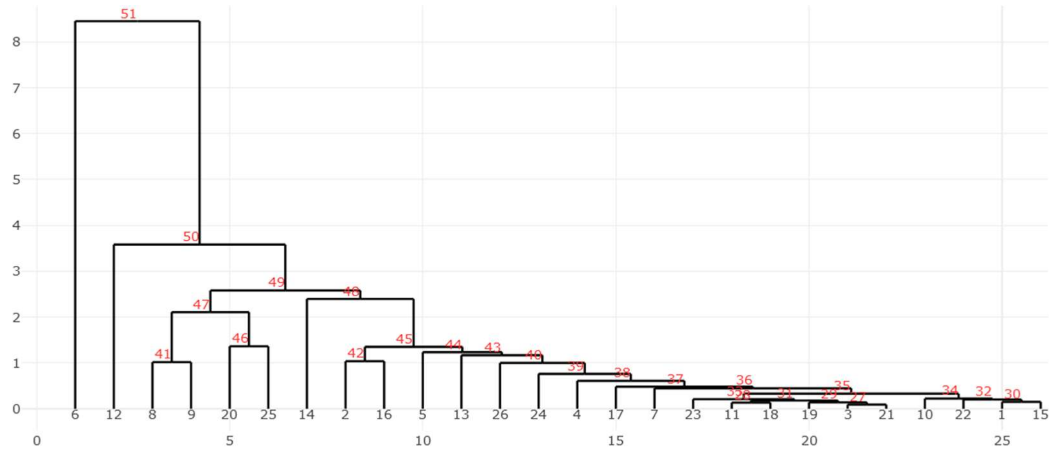


Figure 78: Dendrogram for dataset 2

First steps (1-10) for dataset 3 (Fig. 76c): Mergers occur at relatively small metric values (from 0.591 to 1.097). It means that at the beginning of the algorithm, months with very similar trends in the number of refugees are merged. Middle steps (11-20): Metric values gradually increase (from 1.169 to 2.011). It means that clusters with clusters that are less similar or individual months with clusters are merged. Last steps (21-23): A sharp increase in metric values is observed (from 6.345 to 9.788). It indicates the merging of large and very heterogeneous clusters. The last two steps stand out in particular, where metric values become very large. It means that clusters that differ significantly in the dynamics of the number of refugees were merged at these steps (Fig. 79).

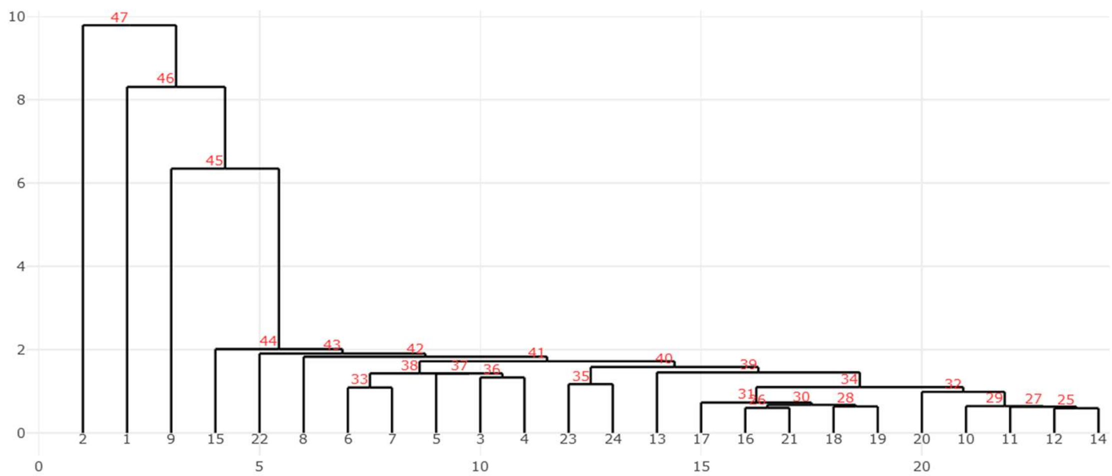


Figure 79: Dendrogram for dataset 3

5. Conclusions

Based on the multi-stage comprehensive study of the relationship between russian aggression and migration processes in Ukraine, which included statistical analysis, time series and cluster analysis, the following solid conclusions can be drawn:

1. Methodological aspects of the study were based on the integrated application of various analysis methods:

- The use of the R programming language for statistical data processing demonstrated high efficiency due to its powerful data processing and visualization capabilities;
- The use of various time series smoothing methods (Kendall, Pollard, and exponential smoothing) allowed us to identify fundamental trends and patterns;

- The sequential smoothing method (method B) showed better results compared to direct smoothing (method A), providing smoother curves and better preservation of long-term trends;
- Hierarchical agglomerative cluster analysis effectively revealed hidden patterns in the data and allowed us to group areas by the similarity of situation;

2. The study of attacks on civil infrastructure showed a clear evolution of the intensity of attacks:

- Until 2022, a relatively low and stable level of incidents was observed (20-40 events per month);
- A sharp surge in activity in 2022 to 700-800 incidents;
- Further stabilization at the level of 200-400 events per month;
- A clear geographical pattern was identified: the eastern and southern regions of Ukraine (Kharkiv, Kherson, Donetsk) suffered the most significant number of attacks;
- Cluster analysis showed the formation of stable groups of regions with similar patterns of attacks;

3. The characteristics of attacks on political targets demonstrate distinct dynamics:

- A stable level of 1,000-1,500 cases per month in 2018-2022;
- A dramatic increase to 5,000 cases per month in 2022;
- Stabilization at a high level of 4000-4500 cases;
- More intense attacks compared to civilian infrastructure;
- Donetsk, Sumy, Kharkiv, Zaporizhia and Kherson regions formed the core of the most affected areas;

4. Research on refugee dynamics revealed a clear structure of migration processes:

- Rapid growth from almost zero to 4.5 million in a short period in early 2022;
- Peaking at around 8 million;
- Two notable “stepped” declines to 6.5 and 6 million;
- Relative stabilization at around 6 million;
- Three key periods are identified:
    a. Initial period (April-May 2022) with a massive outflow of population;
    b. Stabilization period (summer-autumn 2022);
    c. Period of gradual return (from early 2023);

5. A complex system of correlations between different aspects of the conflict has been identified:

- There is a strong relationship between the intensity of attacks and the growth of the number of refugees (correlation 0.793);
- The powerful impact of shelling of civilian infrastructure (correlation ratio 0.803);
- Attacks on political infrastructure show a significant impact (correlation ratio 0.753);
- The highest correlation ratio (0.844) between fatalities from attacks on political targets and the number of refugees;
- Negative correlation (-0.748) between the number of refugees and civilian casualties;

6. The analysis of time characteristics revealed:

- High inertia of migration processes, especially in short-term periods (0-3 days);
- The gradual decrease in the strength of autocorrelation over time, but maintaining significance even with a lag of 10 days;
- More predictable dynamics of shelling of civilian infrastructure compared to attacks on political objects;
- A clear change in the nature of all studied indicators with the beginning of a full-scale invasion;

7. The practical significance of the research results has wide practical application:

- Forecasting migration flows and planning humanitarian aid;

- Risk assessment for different regions of Ukraine;
- Planning civil protection measures;
- Development of strategies for the restoration of affected territories;
- Optimization of the distribution of humanitarian aid;
- Documentation of russian war crimes;
- Improvement of early warning systems for threats;

8. The following limitations of the study were identified:

- Potential delay in data registration;
- Difficulty in taking into account all factors influencing migration;
- Limitations of statistical methods in the analysis of extreme events;
- Potential delay in data registration;
- The analysis is limited by available periods;

9. Directions and prospects for further future research were identified:

- Expansion of the period of analysis;
- Inclusion of additional influencing factors;
- Development of predictive models;
- Detailing regional features;
- Improvement of data analysis methods;
- Application of other cluster analysis methods;
- In-depth analysis of cause-and-effect relationships.

The conducted research convincingly demonstrates the systemic nature of russian aggression aimed at destroying both the civilian and political infrastructure of Ukraine, which led to large-scale forced migrations of the population. Clear patterns and relationships between the intensity of military actions and the scale of migration were identified, which is of critical importance for understanding the nature of the conflict and its impact on the population. The use of a set of statistical methods made it possible to identify hidden patterns and trends that can be used to predict and plan a humanitarian response. Of particular importance is the identification of different dynamics of attacks on civilian and political targets, which is of fundamental importance for understanding the aggressor's strategy and developing effective protective measures. The results of the research create a methodological basis for further analysis and forecasting the development of the situation. It can also be used both for scientific purposes and for practical planning of humanitarian assistance and management of migration processes.

Acknowledgements

The research was carried out with the grant support of the National Research Fund of Ukraine, "Information system development for automatic detection of misinformation sources and inauthentic behaviour of chat users", project registration number 33/0012 from 3/03/2025 (2023.04/0012). Also, we would like to thank the reviewers for their precise and concise recommendations that improved the presentation of the results obtained.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] E. Libanova, O. Pozniak, O. Tsymbal, "Scale and consequences of forced migration of the population of Ukraine as a result of armed aggression of the russian federation," *Demohrafiia ta sotsialna ekonomika*, vol. 48, no. 2, pp. 37–57, 2022. doi:10.15407/dse2022.02.037.

[2] N. Wycoff, L. O. Singh, A. Arab, K. M. Donato, H. Marahrens, "The digital trail of Ukraine's 2022 refugee exodus," *Journal of Computational Social Science*, vol. 7, no. 2, pp. 2147–2193, 2024. doi:10.1007/s42001-024-00304-4.

[3] U. Minora, C. Bosco, S. Iacus, S. Grubanov-Boskovic, F. Sermi, S. Spyratos, *The potential of Facebook advertising data for understanding flows of people from Ukraine to the European Union*, arXiv preprint, 2022. URL: https://arxiv.org/abs/2206.12352.

[4] OECD, *russia's war of aggression against Ukraine generates historic migration flows: More support needed for integration now and possible future return*, OECD Press Release, 2022. URL: https://www.oecd.org/en/about/news/press-releases/2022/10/russia-s-war-of-aggression-against-ukraine-generates-historic-migration-flows.html.

[5] International Center for Ukrainian Victory, *How Migration was Affected by russia's Targeted Shelling of Ukraine in October 2022 – January 2023*, 2023. URL: https://ukrainianvictory.org/publications/research-how-migration-was-affected-by-russias-targeted-shelling-of-ukraine-in-october-2022-january-2023/.

[6] N. V. Kovtun, M. V. Salabai, "Statistical Characteristics of the Integration Process of Ukrainian Refugees in Germany (Based on a Pilot Survey)," *Statistics of Ukraine*, vol. 99, no. 1, 2024. URL: https://su-journal.com.ua/index.php/journal/article/view/451.

[7] A. Pope, *In Kyiv, UN migration chief warns donor fatigue could compound problems*, Reuters, 2024. URL: https://www.reuters.com/world/kyiv-un-migration-chief-warns-donor-fatigue-could-compound-problems-2024-04-11/.

[8] K. Kallas, *Putin seeking to weaponise threat of mass migration, warns Estonian PM*, The Guardian, 2024. URL: https://www.theguardian.com/world/article/2024/may/17/putin-is-seeking-to-weaponise-threat-of-ukraine-mass-migration-says-estonian-pm.

[9] V. Lytvyn, A. Kowalska-Stychen, V. Vysotska, Preface: Machine Learning Workshop, CEUR Workshop Proceedings, Vol-3373, 2023.

[10] M. Emmerich, V. Vysotska, V. Lytvynenko, Preface: Modern Machine Learning Technologies and Data Science Workshop, CEUR Workshop Proceedings, Vol-3426, 2023.

[11] V. Lytvyn, A. Kowalska-Stychen, V. Vysotska, Preface: Machine Learning Workshop, CEUR Workshop Proceedings, Vol-3664, 2024.

[12] V. Kovtun, V. Vysotska, Preface: Modeling, Optimization, and Controlling in Information and Technology Systems Workshop, CEUR Workshop Proceedings, Vol-3668, 2024.

[13] M. Emmerich, V. Vysotska, V. Lytvynenko, Preface: Modern Machine Learning Technologies and Data Science Workshop, CEUR Workshop Proceedings, Vol-3711, 2024.

[14] A. Chiche, Hybrid decision support system framework for crop yield prediction and recommendation, Int. J. Comput. 18 (2) (2019) 181–190. doi:10.47839/ijc.18.2.1416.

[15] H. Lipyanina, V. Maksymovych, A. Sachenko, T. Lendyuk, A. Fomenko, I. Kit, Assessing the investment risk of virtual IT company based on machine learning, in: S. Babichev, D. Peleshko, O. Vynokurova (Eds.), Data Stream Mining & Processing. DSMP 2020, Commun. Comput. Inf. Sci., vol. 1158, Springer, Cham, 2020, pp. 122–134. doi:10.1007/978-3-030-61656-4_11.

[16] S. Bhatia, M. Sharma, K.K. Bhatia, P. Das, Opinion target extraction with sentiment analysis, Int. J. Comput. 17 (3) (2018) 136–142. doi:10.47839/ijc.17.3.1033.

[17] Ukraine - Refugee Situation, 2024. URL: https://data.humdata.org/dataset/ukraine-refugee-situation

[18] Ukraine - Conflict Events, 2024. URL: https://data.humdata.org/dataset/ukraine-acled-conflict-data