Relationships between crime and everyday factors

Zbigniew Wawrzyniak, Grzegorz Borowik, Eliza Szczechla,
Paweł Michalak, Radosław Pytlak, Paweł Cichosz, Dobiesław Ircha,
Wojciech Olszewski and Emilian Perkowski

September 2, 2018

Relationships between crime and everyday factors

Zbigniew M. Wawrzyniak[a], Grzegorz Borowik[b], Eliza Szczechla[c], Paweł Michalak[c], Radosław Pytlak[a], Paweł Cichosz[a], Dobiesław Ircha[c], Wojciech Olszewski[c], and Emilian Perkowski[c]
[a]Warsaw University of Technology, Warsaw, Poland
[b]Police Academy in Szczytno, Szczytno, Poland
[c]Scott Tiger S.A., Warsaw, Poland
z.wawrzyniak@ise.pw.edu.pl, g.borowik@wspol.edu.pl, eliza.szczechla@tiger.com.pl, pawel.michalak@tiger.com.pl, r.pytlak@mini.pw.edu.pl, p.cichosz@elka.pw.edu.pl, dobieslaw.ircha@tiger.com.pl, wojciech.olszewski@tiger.com.pl, emilian.perkowski@tiger.com.pl

***Abstract*—Technological development in every aspect of human life has formed wider analytical approach to the crime. The genesis and structure of crime, its intensity, and dynamics are subjects of intense scientific research carried out by researchers in various fields of science. At the same time, the crowd-sourced open data-sets as social media and Internet data-sets can be a valuable source of knowledge about various behavior patterns, and social phenomena, including those of criminal nature. This paper aims to present some results of observations, simulations, and prediction of crimes and their correlations to the physical environmental factors like weather or distance to the point of interest. The results are based on qualitative and quantitative data obtained from existing open resources. The research may be useful to model forces and police means to prevent and combat crime effectively.**

I. Introduction

Currently, the development of individual areas of forensic science, criminology, and investigative work depends mainly on other sciences, such as mathematics, physics, chemistry, or IT [1]. Data analysis and development of models in the application of forecasting crime events requires authorized persons with access rights to datasets, making specific decisions as well as repeatability in conducting these activities in the form of continuous and systematic monitoring and management of processes [2]. These processes in the sense of continuity and complexity essentially include data mining and analysis for discovery of hidden relationships between crucial data and predictor of occurring events. They also include key indicators of quality and effectiveness of forecasting; study of the natural language, so-called text analysis in the categorization of complex concepts from qualitative and unstructured data, especially from open sources. The open data includes telecommunications connections and transaction supervision records, various forms of criminals' notes, telecommunications data in the form of SMS, blogs, communicators and similar sources without specific information structure. Data finally is presented by visualization and panels for reporting (dashboards). The other processes are managing data and data operations in real time [3], as well as risk analyzes supporting the optimization of the use of resources of law enforcement authorities (police or judicial structures) [4].

This paper aims to present some results of observations, simulations, and forecast of crimes and their correlations to the physical environmental factors like weather or distance to the point of interest. The results are based on qualitative and quantitative data obtained from existing open resources. The research may be useful to model forces and police means to prevent and combat crime effectively.

II. Background

Despite the development of science and technology, the inquiry process of data gathering is not simple, quick, and cheap. The use of appropriate analytical and presentation methods may, however, reduce the time required to develop a final opinion or decision [5]. By automating specific subtasks in the investigation processes, commonly, the automatic acquisition of knowledge may help in achieving this goal [6], [7].

It should be noted that the high-tech crime forecasting resulting from the need of ensuring safety, as well as the simulation and modeling of crime risk development and the probability of events, requires careful monitoring and obtaining information about new technological solutions according to the developed and implemented models [8].

The study of factors determining the development of crime boils down to the research of methods of action of the perpetrator and behavior of the victim. Furthermore, so-called PESTLE attributes (Political, Economic, Socio-Cultural, and Technological) [9], as well as others, as notable event and atmospheric conditions, are known as influencing ones [10], [11]. The every-day factors, including weather characteristics, relate to the current history of human behavior in the environment and society, and also modify criminal or common illegal behaviors [12].

For spatiotemporal simulation of the occurrence of a future criminal event, many data analysts also use hotspot models [13], [14] and those based on causal econometric. Knowledge discovery uses auto-regressive models, neural network, and other machine learning methods [15], [16].

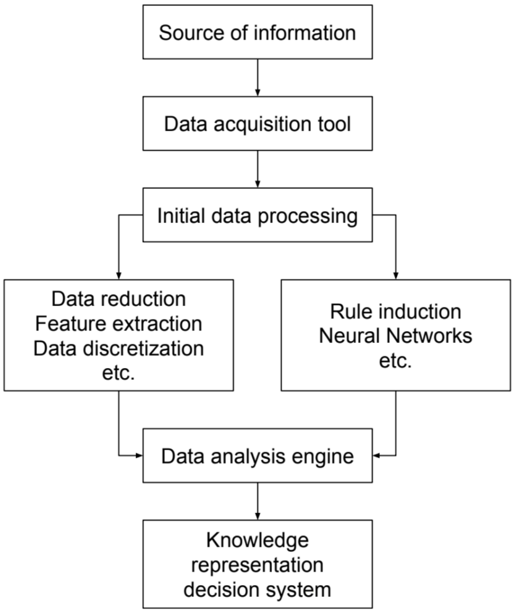

Fig. 1. Data processing scheme

III. Methodology

It is proposed to use the method of data processing by the methodology shown in Fig. 1. There are many IT fields in which this scheme is applicable. These are expert systems, decision support systems, image recognition, and others. Applications implementing the following proposed scheme can be used in industry, medicine, social sciences, and process control. The analysis is most often performed on incomplete and uncertain data sets. This analysis leads to the solution of such problems as:

- finding data dependencies,
- generating decision algorithms,
- data reduction,
- data classification,
- pattern recognition,
- learning from examples,
- generating and minimizing control circuits.

Analysed data covers daily recorded crimes and offenses from 1st January 2013 to 30th June 2016 from across the territory of the Republic of Poland within three police districts of big cities, i.e., Białystok (BA), Elbląg (NE), Olsztyn (NO). The data were mainly obtained from the internal sources (unstructured data), but some aggregated data on records is available here (https://danepubliczne.gov.pl/dataset/dane-o-przestepczosci). Meteorological data was acquired from the Institute of Meteorology and Water Management - National Research Institute (https://danepubliczne.imgw.pl/) and Weather Online (https://www.weatheronline.pl/). The weather information for the location of a given unlawful act was assigned from the nearest meteorological station (Euclidean distance) with smoothing spline model to interpolate meteorological measurements on a regular grid based on station locations. The records used, consist the number of reported crimes by district, date and time, and category of crime; to be aggregated according to the need of a prediction model. Our empirical analysis focuses on the different type of crimes: especially police interventions (INT), communication offenses, and domestic violence: battery (BOJ), hooliganism (CHU), theft (KRA), robbery (ROZ). The dataset consists of 1 192 591 total instances including about 112 predictive and non-predictive features and parameters. Data analysis was completed using TIBCO Statistica and several Python libraries including Pandas, SciPy, and scikit-learn. In the study, the generalized linear mixed models were estimated, and a Poisson distribution and log link were used for almost all of the reported models, avoiding problems with overdispersion effect.

IV. Experiments and Results

In this examination, we show elements of cyclical description for time series as a way of parameterization of models containing the main information characterizing the processes of creation and propagation of criminal activities.

In this section, we present the results from three police districts of big cities and their suburbs, i.e., BI, NE, NO. To hide the number of criminal cases all the data presented in charts were normalized to the interval of $(0, 1)$. For some instances to flatten the peaks, we also re-scaled the data into the log scale.

A crucial experiment was to perform frequency analysis of the events. The domain of frequency is an alternative field of signal description and analysis, closely related to the domain of time. Frequency analysis methods in the frequency domain are called frequency methods or spectral methods. They are more effective tools for studying signals than time domain methods. The mathematical formalism of Fourier's integral transformations is a basement for frequency analysis methods of continuous signals. They determine the mutual relations between the time domain and the frequency domain. Fourier's integral transformation known for over two hundred years remains the fundamental and most widespread signal analysis tool.

The frequencies presented in Figs. 2, 3 show initial evidence of seasonality. The most police interventions have been carried out on Saturday at 11 pm. Communication offenses happen mainly during working days between 7 am to 6 pm. We may also observe an intensiveness of communication offenses around 3 pm due to the traffic jams as well the migration of people from work to shops or home. For some cities (not included in this paper) we also observed the seasonality of communication offenses between 3 pm and 6 pm on Fridays correlated to the beginning of the weekend and the migration of people between cities.

For the theft (Fig. 4), one may notice the events mostly at the beginning of the week (Monday). Most of the events happened around 10 am and 3 pm since the perpetrators exploit the knowledge that the owners are absent from home or apartment.

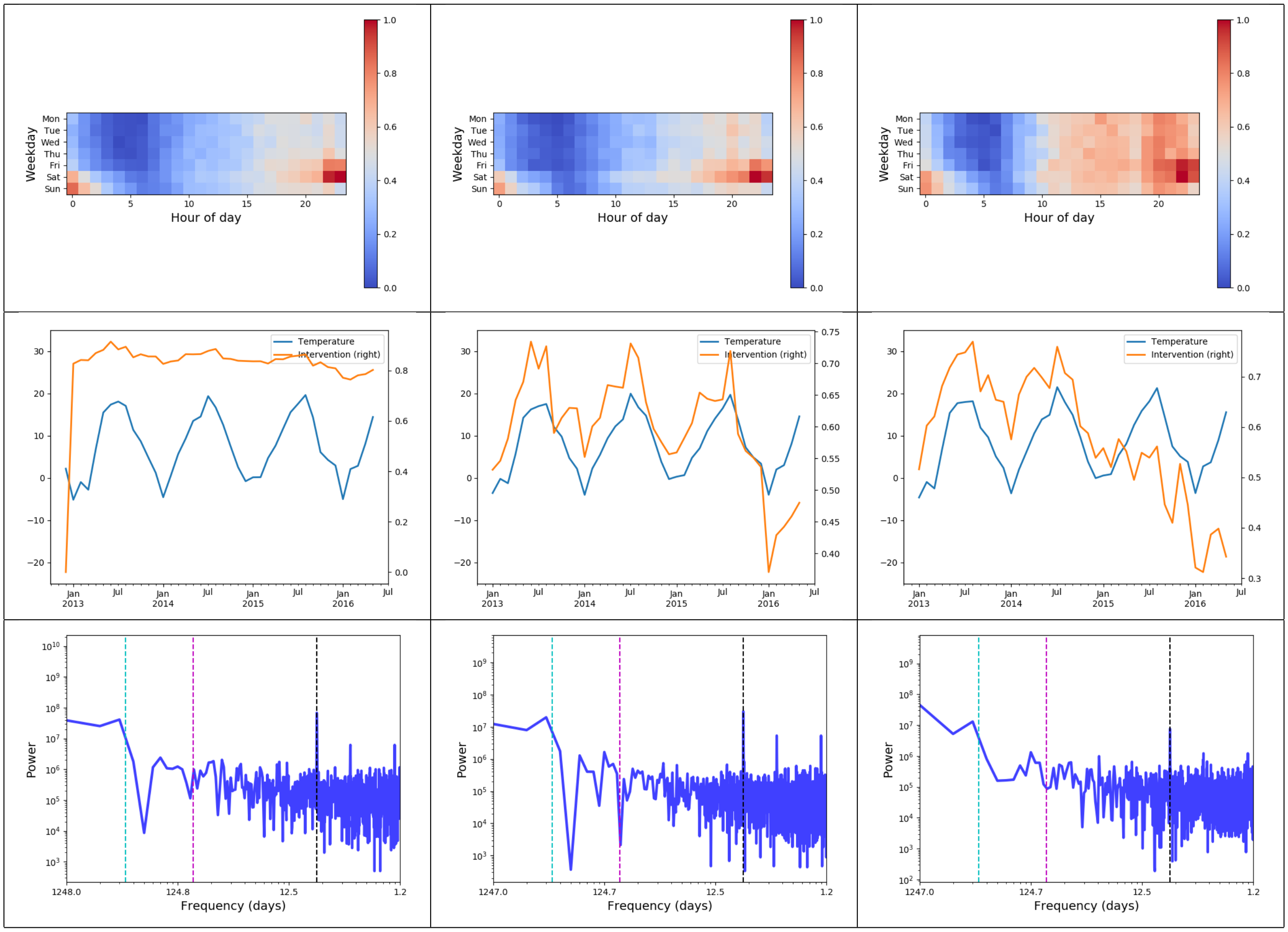

Fig. 2. Distribution of the relative difference in INT for weekday and hours profiles (upper row), time series of INT (left scale) with relative temperature change (middle row), and frequency characteristics of INT (lower row) for the areas of BA, NE, NO (left to right); INT – police intervention number

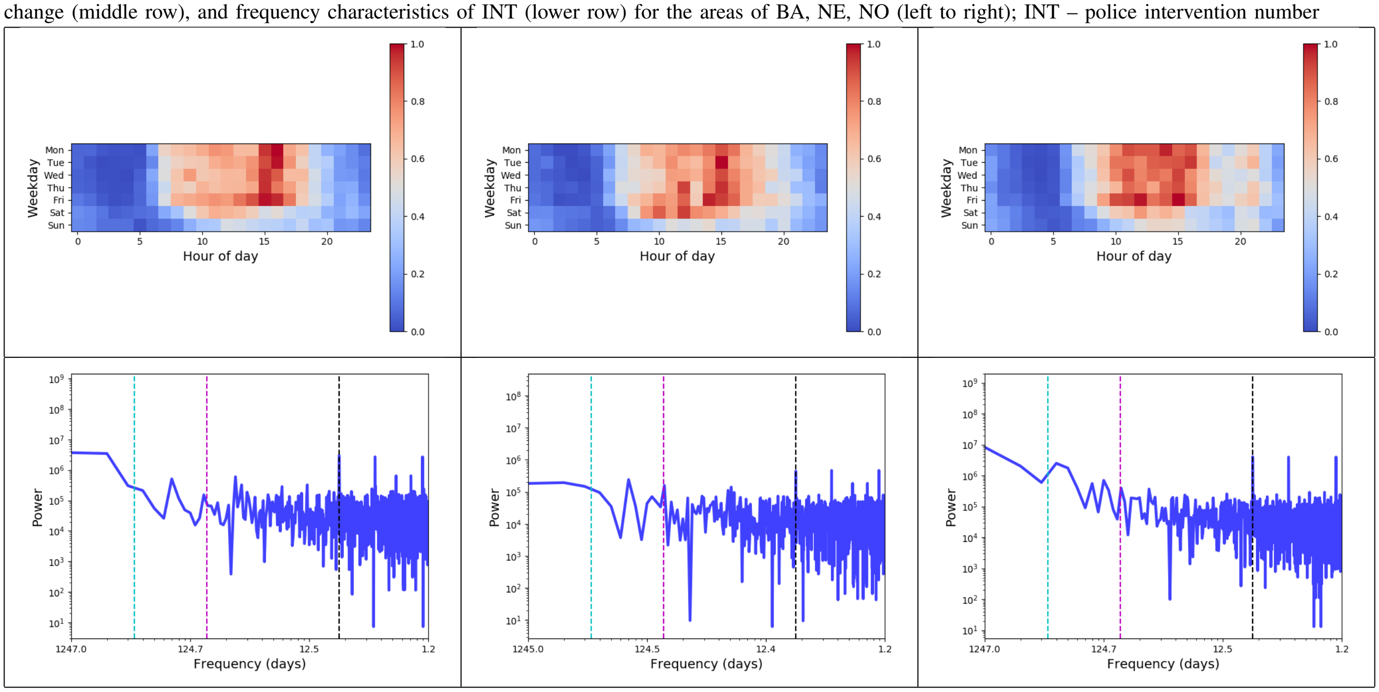

Fig. 3. Distribution of relative difference in KOM for weekday and hour's profiles (upper row), frequency characteristics of KOM (lower row) for the areas of BA, NE, NO (left to right); KOM – communication offense number

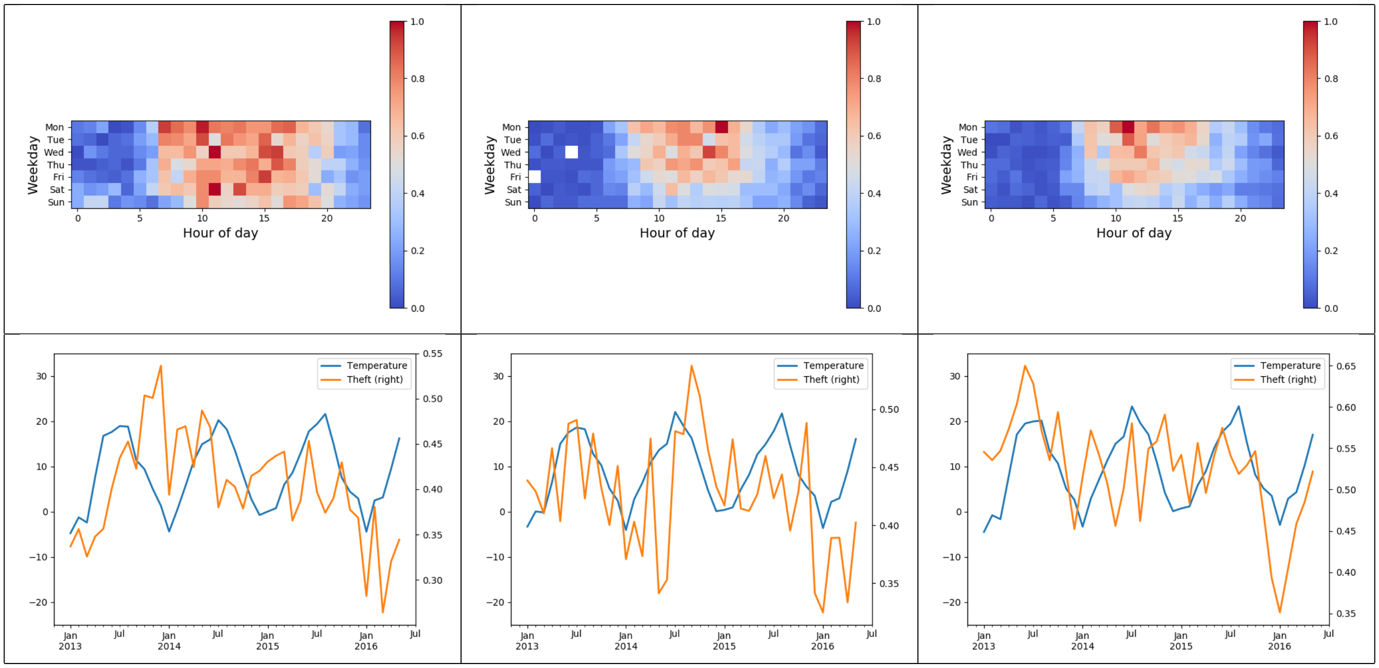

Fig. 4. Distribution of relative difference in KRA for weekday and hourâĂŹs profiles (upper row), frequency characteristics of KRA (lower row) for the areas of BA, NE, NO (left to right); KRA – theft number

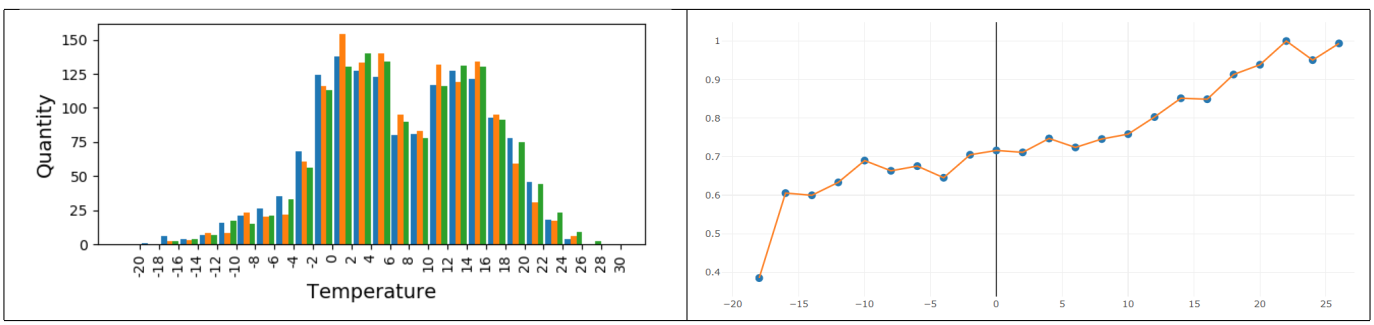

Fig. 5. Distribution of temperature profile the areas of BA (blue), NE (orange), NO (green) – left to right (left side) and relative fluctuation for the area of BI (right side); INT - police intervention number

Next type of experiment was an examination of the aggregation influence on understanding the phenomena using visualization basing cyclical estimators and relations with crime and weather attributes.

In Fig. 5 we present experimental results where on the left-hand side we have the temperature distribution (provided in orange) – the number of days with a given average daily temperature in the analyzed period, finally normalized to the sale [0, 1]. In blue we have the number of police interventions at a given daytime temperature. For example, if the average daily temperature X occurred twice and in the first day there were two offenses, in the second day there were three offenses, then crime rate = 2 + 3 = 5 (before normalization to [0, 1]). On the right-hand side, we have the number of crimes at a given average daily temperature, divided by the number of days with a given average daily temperature, in the range [0, 1]. We may observe the growing trend in the number of police interventions according to the growing temperature.

As part of the study of external parameters that could affect the number of crimes, we examine the distance of the event localization from the 24/7 open stores. We may observe the natural intensification of events near those stores (Fig. 6).

Finally, the weekly, daily and annually crimes of selected categories were examined. Figure 7 shows the daily cycle for three categories of offenses. Similarities are visible (smaller number around 5 am); however, there is variation: for the hooliganism category, the period of the reduced number of occurrences is longer.

By observing an influence of seasonality of weather on crime and corresponding data from the plots, we found the strong correlation in the examined period, i.e., 2013 – 2015 (Pearson and Spearman correlation coefficient at the level higher than 0.700, $p < 0.05$) (see Tab. I). For the whole 3-years range data, the relation of interventions and temperature profile correlate with slightly lower values than for the one-year ranges, and for shortened aggregation (monthly vs. weekly).

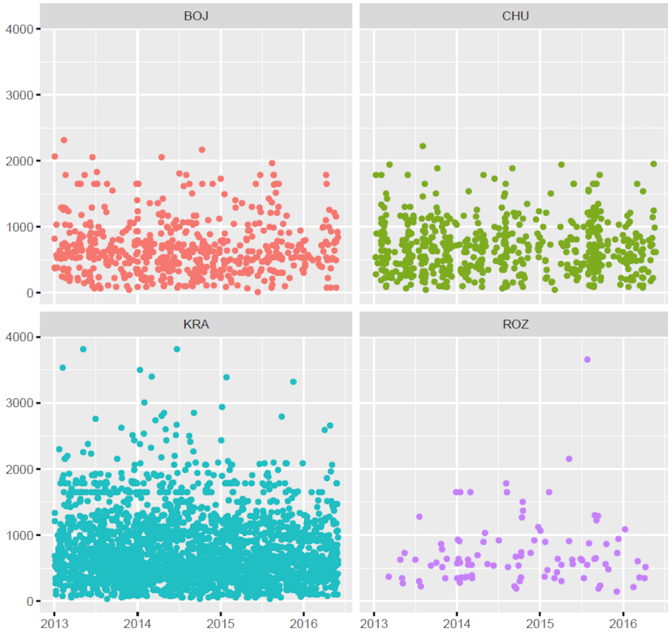

Fig. 6. Distribution of distance to the point of interest for different crimes; BOJ – battery, CHU – similar-hooliganism, KRA – theft, ROZ – robbery

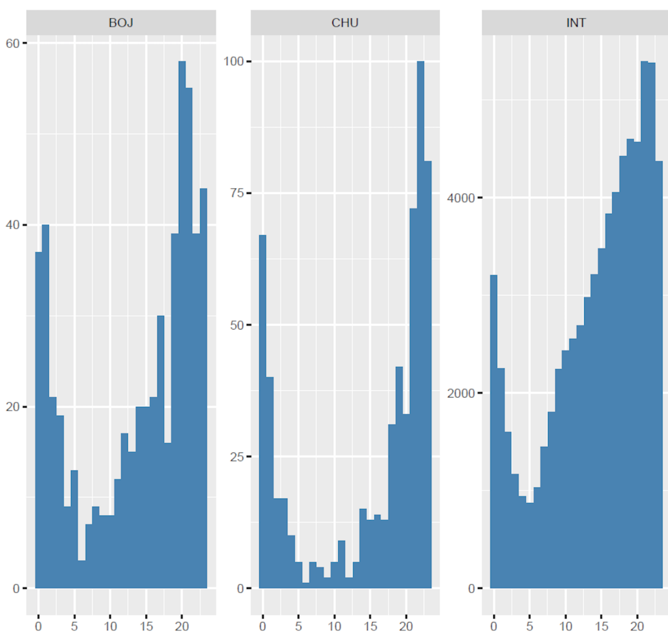

Fig. 7. Distribution of events (location has been hidden); BOJ – battery, CHU – hooliganism, INT – police intervention

The correlation for BI area is 0.725 vs. 0.674, for NE: 0.825 vs. 0.763, except NO area with some lowering to medium range as 0.401 vs. 0.437, respectively for monthly vs. weekly aggregation method. For other types of crimes, we obtained similar results.

The figures show strong seasonal patterns in all areas for all variables (INT, KRA). Seasonal weather variation is most substantial during most cooling temperature ranges (below 0^0C), where the mean temperature differences between January and July for 3-years range equal to 21.9, 20.5 and 21.8 for BI, NE, NO, respectively.

TABLE I
CORRELATION COEFFICIENTS

Area of the city	Aggregation method	2013	2014	2015	2013 – 2015
BI	weekly	0.835†	0.838	0.552	0.674†
	monthly	0.898	0.917	0.642	0.725†
NE	weekly	0.783†	0.853	0.700	0.763†
	monthly	0.855	0.921	0.765	0.825
NO	weekly	0.867†	0.748	0.053†	0.437†
	monthly	0.936	0.799	0.097	0.401

† denotes Spearman correlation coefficient, others – Pearson correlation coefficient

V. SUMMARY

The temperature is an essential weather factor associated with violence and crime in our study. Our findings suggest potential implications for police and intervention services as well as public administration (e.g., transportation) to be explored in future studies.

We observe the linear regression algorithm to be very useful and accurate in predicting the crime data based on the training set input.

Considering the number of crimes, their location, recurrence, seasonality, correlation with weather factors, and possible predictions, all of those help the police and public administration to adapt planning and intervention strategies to ensure public safety. Obtaining the spatial information about crimes and offenses, their weekly distribution, and especially predicting the occurrence from the data gathered in the past can be very valuable for management and planning of police means to protect people against violence and preserve human lives from criminal acts. Summarizing the general conclusions, exploratory analysis, and data mining have become a vital part of crime detection and prediction. Even though the scope of this part of the project was to prove how useful and accurate algorithms can be at predicting violent crimes, there are other applications of such analyses in the realm of law enforcement such as determining criminal hot spots, creating criminal profiles, and learning crime trends from large volumes of data. However, the precision in which one could infer and create new knowledge on how to slow down crime is well worth the safety and security of people.

ACKNOWLEDGMENT

The methods of criminal event prediction and weather modeling were tested in a frame of the project “Creating a system for forecasting the development of crime, as an element of building a security strategy and public policy” co-financed by the funds of the National Centre for Research and Development (NCBiR) in the framework of programme “Defence and Security BiO-7” No. DOB-BIO7/05/20/2015.

References

[1] M. M. Houck and J. A. Siegel, *Fundamentals of forensic science*. Academic Press, 2009.

[2] R. B. Santos, *Crime analysis with crime mapping*. Sage publications, 2016.

[3] T. Nakaya and K. Yano, "Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation and scan statistics," *Transactions in GIS*, vol. 14, no. 3, pp. 223–239, 2010.

[4] L. W. Kennedy, J. M. Caplan, and E. Piza, "Risk clusters, hotspots, and spatial intelligence: risk terrain modeling as an algorithm for police resource allocation strategies," *Journal of Quantitative Criminology*, vol. 27, no. 3, pp. 339–362, 2011.

[5] R. Bowles, M. G. Reyes, and N. Garoupa, "Crime reporting decisions and the costs of crime," *European journal on criminal policy and research*, vol. 15, no. 4, p. 365, 2009.

[6] J. Schroeder, J. Xu, H. Chen, and M. Chau, "Automated criminal link analysis based on domain knowledge," *Journal of the Association for Information Science and Technology*, vol. 58, no. 6, pp. 842–855, 2007.

[7] A. Iriberri and C. J. Navarrete, "Internet crime reporting: Evaluation of a crime reporting and investigative interview system by comparison with a non-interactive reporting alternative," in *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*. IEEE, 2010, pp. 1–9.

[8] A. B. Sakpere, A. V. Kayem, and T. Ndlovu, "A usable and secure crime reporting system for technology resource constrained context," in *Advanced Information Networking and Applications Workshops (WAINA), 2015 IEEE 29th International Conference on*. IEEE, 2015, pp. 424–429.

[9] H. L. Larsen, J. M. Blanco, R. P. Pastor, and R. R. Yager, *Using Open Data to Detect Organized Crime Threats: Factors Driving Future Crime*. Springer, 2017.

[10] K. M. Beaver, "Nonshared environmental influences on adolescent delinquent involvement and adult criminal behavior," *Criminology*, vol. 46, no. 2, pp. 341–369, 2008.

[11] M. Warr, *Companions in crime: The social aspects of criminal conduct*. Cambridge University Press, 2002.

[12] J. M. Caplan, L. W. Kennedy, and J. Miller, "Risk terrain modeling: brokering criminological theory and gis methods for crime forecasting," *Justice Quarterly*, vol. 28, no. 2, pp. 360–381, 2011.

[13] H. Berestycki and J.-P. Nadal, "Self-organised critical hot spots of criminal activity," *European Journal of Applied Mathematics*, vol. 21, no. 4-5, pp. 371–399, 2010.

[14] F. Ahmad, S. Syal, and M. S. Tinna, "Criminal policing using rossmo's equation by applying local crime sentiment," in *Data Engineering and Intelligent Computing*, S. C. Satapathy, V. Bhateja, K. S. Raju, and B. Janakiramaiah, Eds. Singapore: Springer, 2018, pp. 627–637.

[15] E. Cesario, C. Catlett, and D. Talia, "Forecasting crimes using autoregressive models," in *Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), 2016 IEEE 14th Intl C*. IEEE, 2016, pp. 795–802.

[16] H.-W. Kang and H.-B. Kang, "Prediction of crime occurrence from multi-modal data using deep learning," *PloS one*, vol. 12, no. 4, p. e0176244, 2017.

[17] M. Ranson, "Crime, weather, and climate change," *Journal of environmental economics and management*, vol. 63, no. 3, pp. 274–302, 2014.

[18] D. J. Lemon and R. Partridge, "Is weather related to the number of assaults seen at emergency departments?" *Injury*, vol. 48, no. 11, pp. 2438–2442, 2017.

[19] X. Hu, J. Wu, P. Chen, T. Sun, and D. Li, "Impact of climate variability and change on crime rates in tangshan, china," *Science of the Total Environment*, vol. 609, pp. 1041–1048, 2017.

[20] J. Tiihonen, P. Halonen, L. Tiihonen, H. Kautiainen, M. Storvik, and J. Callaway, "The association of ambient temperature and violent crime," *Scientific Reports*, vol. 7, 2017.

[21] S. J. Michel, H. Wang, S. Selvarajah, J. K. Canner, M. Murrill, A. Chi, E. D., and E. B. Schneider, "Investigating the relationship between weather and violence in Baltimore, Maryland, USA," *Injury*, vol. 47, no. 1, pp. 272–276, 2016.

[22] B. Hołyst, *Prognozowanie kryminologiczne w wymiarze społecznym – Metodologia, Analiza, Tendencje rozwojowe*. Tom 1 (in Polish). Wydawnictwo Naukowe PWN S.A., 2017.

[23] J. M. Blanco and J. Cohen, "Macro-environmental factors driving organised crime," in *Using Open Data to Detect Organized Crime Threats*. Springer, 2017, pp. 137–166.

[24] M. Kruse and A. D. Svendsen, "Foresight and the future of crime: Advancing environmental scanning approaches," in *Using Open Data to Detect Organized Crime Threats*. Springer, 2017, pp. 73–101.

[25] D. Bjerk, "How much can we trust causal interpretations of fixed-effects estimators in the context of criminality?" *Journal of Quantitative Criminology*, vol. 25, no. 4, pp. 391–417, 2009.