



<http://www.diva-portal.org>

Preprint

This is the submitted version of a paper presented at *European Intelligence and Security Informatics Conference (EISIC)*.

Citation for the original published paper:

Boldt, M., Borg, A. (2017)

A statistical method for detecting significant temporal hotspots using LISA statistics

In: IEEE Computer Society

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:bth-15166>

A Statistical Method for Detecting Significant Temporal Hotspots using LISA Statistics

Martin Boldt and Anton Borg
Department of Computer Science and Engineering
Blekinge Institute of Technology, Sweden
Email: {martin.boldt, anton.borg}@bth.se

Abstract—This work presents a method for detecting statistically significant temporal hotspots, i.e. the date and time of events, which is useful for improved planning of response activities. Temporal hotspots are calculated using Local Indicators of Spatial Association (LISA) statistics. The temporal data is in a 7x24 matrix that represents a temporal resolution of weekdays and hours-in-the-day. Swedish residential burglary events are used in this work for testing the temporal hotspot detection approach. Although, the presented method is also useful for other events as long as they contain temporal information, e.g. attack attempts recorded by intrusion detection systems. By using the method for detecting significant temporal hotspots it is possible for domain-experts to gain knowledge about the temporal distribution of the events, and also to learn at which times mitigating actions could be implemented.

Index Terms—Temporal analysis, temporal hotspot, computational criminology, LISA statistics, local indicators of spatial association.

I. INTRODUCTION

Law enforcement are often required to handle large amounts of crimes, often with limited resources. Further, proactive policing are preferred, rather than reactive policing. As such, predictive policing is of interest to law enforcement [1].

Crime prediction makes use of quantitative analytical techniques for developing statistical predictions of for instance future targets [2]. Predicting future crime trends require accurate spatial and temporal crime scene data. Unfortunately, many crime categories generally do not include accurate spatial and/or temporal data, e.g. various property thefts where no witness exist. As a consequence, these crimes are typically reported to have occurred within a time interval of various length rather than at a specific spot in time [3], [4].

The lack of accurate temporal data regarding certain crime categories is unfortunate since it renders the analysis of crime frequency variation over time harder [5]. However, to address this program there is a need for crime prediction methods with the ability to, based on imprecise temporal data, estimate a more accurate time of occurrence for crime events. One such approach that exists today is aoristic analysis [3].

Based on the need for such temporal analysis methods, the motivation behind the present work is to investigate a novel method for detection of temporal hotspots, which can provide increased knowledge about temporal patterns of criminal offenses. Such increased understanding of how different crime categories distribute over time could prove valuable for law enforcement during both tactical and strategic work, as well

as in operations analysis when it comes to the allocation of resources when criminal activity is more likely to occur [4]. This paper presents initial work in this area.

II. RELATED WORK

The various existing analytical methods that could be used for predicting crimes make use of historical crime data. Some examples include for instance hotspot identification, crime mapping and risk terrain analysis [5], [6], [1], [4], [7]. In addition to this, the analysis of both repeat and near-repeat victimization is common [5].

Previous research on temporal analysis methods have been less explored, since spatial analysis methods have attained more focus [8]. Albeit, temporal data is just as important during crime prediction [9], [10].

Studies on temporal characteristics of different crime types have been conducted to increase the understanding of criminal behavior, e.g. when offenders commit crimes [11], [10]. Offenses have been found to be more likely to be committed in the afternoon or early evening [8], [9], while nighttime burglaries were found to be more likely in weekends [12]. Consequently, knowledge of temporal trends is essential to understanding criminal behavior, e.g. late mornings and early afternoon is more likely to result in empty houses due to work hours.

Producing accurate temporal maps of when crimes occur is a difficult task [4]. Nevertheless, several methods exist that can represent crimes either as moments or to account for the estimated time spans. Circular statistics has been used to analyze the number of incidents by time of the day and day of the week [13]. To allow comparison of crime temporal distribution between geographical areas, mapping the crimes into quartile minutes has been investigated [9].

The accuracy of temporal analysis methods have been evaluated on their ability to estimate the exact offense time for bicycle thefts [8]. The study evaluates how accurately five methods can estimate the true offense time in terms of hour of the day. It was concluded that the aoristic method best manages to estimate the offense time in that context. Which was supported when investigating residential burglaries [14].

The aoristic method can handle crimes without a known specific occurrence time, i.e. represented as a time span instead [15], [11], [3]. It works by choosing a structured time or temporal unit, e.g. hours or days. It works by giving

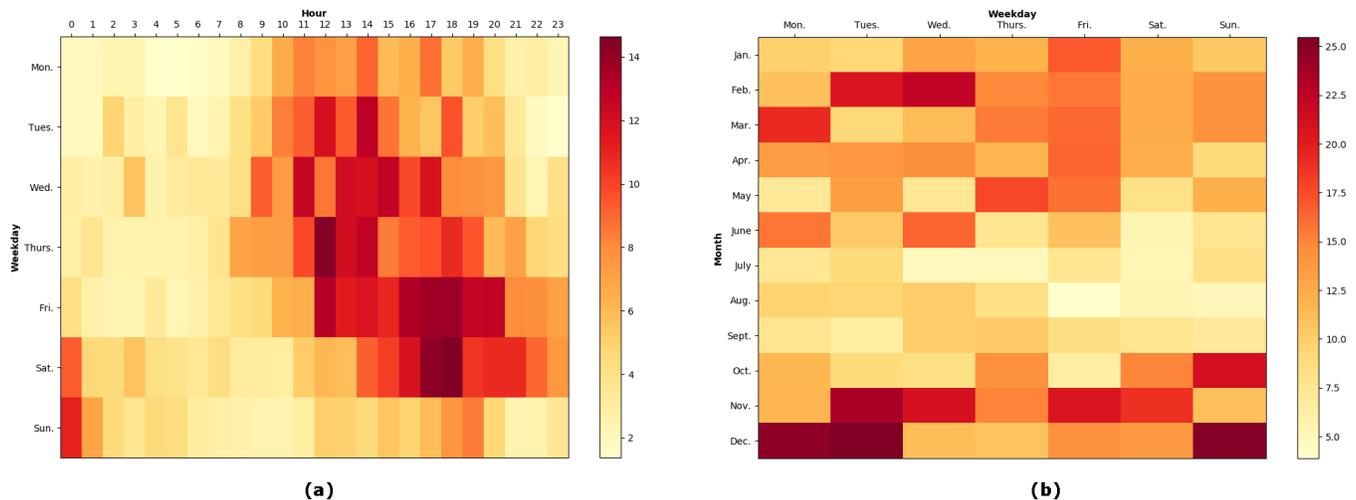


Fig. 1: Two examples of temporal matrix representations (with different resolutions) of residential burglaries in the city center of Malmö in Sweden for 2015 and 2016 ($N=1,059$). (a) shows the weekday by hour-of-the-day matrix, while (b) shows the weekday by month matrix.

each offense a point of 1.0, which then is evenly divided among the units within the time span. This is repeated for every crime investigated. The points of every temporal unit are then summarized and used to indicate high profile time periods, which evenly distributes the likelihood of when a crime occurred across the possible temporal units.

The research gap identified in the present work, involves the lack of methods for detecting statistically significant hotspots in temporal matrices representations that combine two temporal variables, e.g. year and week number. A novel method for detecting significant temporal hotspots is therefore described in this work and its operation is shown with regards to residential burglaries in Sweden.

III. METHOD FOR TEMPORAL HOTSPOT DETECTION

This section shows how temporal data can be represented in various formats, how LISA statistics work, and how such statistics was used in the presented novel method for temporal hotspot detection.

A. Temporal Matrix Representations

Traditional representations of crime frequencies most often rely on one temporal variable, e.g. crimes per year, per month, per weekday, or per hour-of-the-day. Although, such analyses provide general knowledge about the temporal distribution of crimes, they do not provide more detailed knowledge about the occurrence of criminal events at given points in time, e.g. Thursdays around 13 o'clock. However, by combining two temporal variables (each with different resolution) it is possible gain knowledge about such temporal patterns. Some examples of such combined temporal resolutions are:

- weekday by time-of-the-day (7×24)
- weekday by month (7×12)
- weekday by week-in-the-year (7×52)
- month by time-of-the-day (12×24)

Fig. 1 shows two examples of the matrices above from an aoristic analysis of residential burglary data from the city of Malmö in Sweden between 2015 and 2016. The temporal heatmap plots are generated in a decision-support system that the authors have developed together with Swedish law enforcement [16]. In Fig. 1 (a) the 7×24 matrix shows that there are increased crime frequencies on working days from around 10 until about 19 in the evening. For Saturdays there is a similar interval, although delayed with a few hours. Fig. 1 (b) shows the 7×12 matrix with a general decrease of residential burglaries during the summer months, and a later increase in the autumn and winter months.

B. LISA and G_i^* Statistics

This work focuses on temporal representations in the form of matrices where the aim is to determine which cells in a given matrix that deviates more than can be accounted for by chance alone. As such, Local Indicators of Spatial Association (LISA) statistics tests are appropriate [17]. In this work the Getis and Ord G_i^* statistic is used [18], since it is commonly used for crime mapping and is also implemented in the major GIS applications that are used for crime analysis [5]. However, there exists other LISA statistics as well, e.g. Local Moran's I and Local Geary's C [5].

Given a matrix of e.g. crime frequencies, the Getis Ord G_i^* method identifies clusters of cells in the matrix that have higher values than is expected by random chance, by comparing local to global averages in the matrix. The method calculates a Z score (also known as standard score) per cell by taking both the cell's own crime frequency, as well as surrounding cells' frequencies into account. The Z score indicates the place of each value in the data relative to the standardized mean of all data. Therefore, a cell with $Z = 0$ equals the mean, while $Z > 0$ indicates a value greater than the mean, and vice versa. Exactly how many cells that are

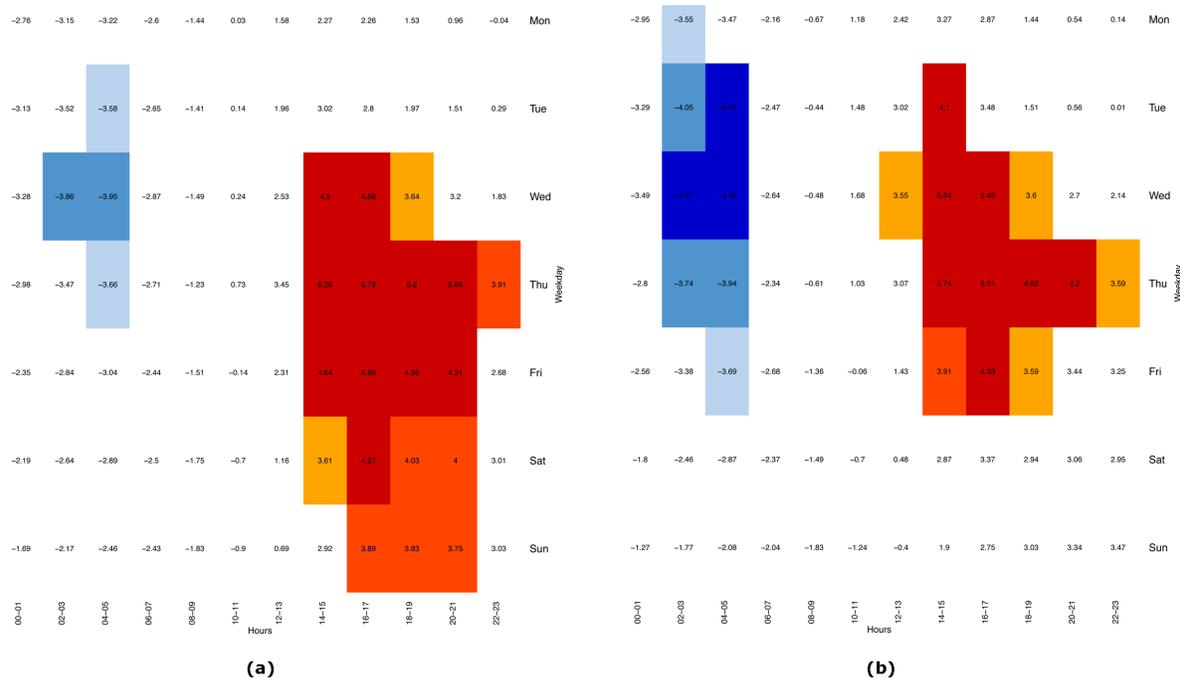


Fig. 2: Two examples that show statistically significant temporal hotspots for residential burglaries in the county of Skåne in Sweden during two arbitrarily chosen months in 2014. Each cell in the matrices holds the Z-statistic that represent the crime frequency. (a) shows the weekday by hour-of-the-day matrix for burglaries in January 2014 (N=422), while (b) shows the temporal hotspots for October 2014 (N=449).

included in the calculation of local averages is determined by the distance parameter d , which is specified a priori. A value $d = 1$ results in that a total of 9 cells will be included, i.e. the cell itself as well as 8 neighboring cells. The method calculates how much each cluster of cells deviates from the mean frequency of the whole matrix. All details for calculating Getis and Ord G_i^* are explained in [18].

The resulting Z score for a particular cell relates the frequencies of its neighbors. Thus, a high value in a particular cell indicates that neighboring cells have high frequencies as well, and vice versa. The further away from 0 that a Z score deviates, the stronger the association with neighboring cells are. Z scores close to 0 indicate that there are no clear association. Since the Z score represents the number of standard deviations that a data point deviates from the mean, it is easy to translate them to confidence or significance levels. By comparing the Z score in each cell with the critical values of the Z distribution, it is possible to determine at which confidence level (if any) that the cell frequency deviates.

C. Proposed Method

In our prototype, the horizontal aspect of the distance variable d in the Getis and Ord G_i^* statistical method ties together the last hour of one day with the first hour in the proceeding day, to handle bordering effects. The vertical aspect of d is motivated since this reflects similar temporal behavior of the offenders. For instance, increased aoristic crime frequencies on Wednesdays at 21 o'clock is related to the crime frequencies of both Tuesdays and Thursdays around that time. The correctness of the prototype implementation was

validated using the Getis and Ord G_i^* example 16×16 matrix in [5] on page 165, by feeding the matrix into the prototype and checking that the result is identical.

The proof-of-concept component was implemented in R, which is a freely available open source programming language and software environment for statistical computing and graphics. The package RMySQL was used for database integration and the package Lubridate was used for managing date and time aspects in the prototype. The visualizations and heatmaps were produced using the gplots and ggplot2 packages, which used various shades of red for indication of increased burglary frequencies. Each shade is determined by the critical Z values related to the following four confidence levels: .90, .95, .99 and .999. Similarly, various shades of blue were used to signify decreased frequencies.

D. Data and Data Representation

The burglary data used is the official Swedish crime reports for 2014 and 2015 that was provided by the Swedish law enforcement agencies. The motivation for using data from these years was due to convenience because we already had access to the full data for these years. The data was kept in a MySQL relational database and this approach primarily used the start and end date and time of burglaries in the format YYYY-MM-DD and hh:mm:ss respectively.

IV. EXAMPLES OF TEMPORALLY SIGNIFICANT HOTSPOTS

As examples of the operation of the proposed method, we have used the method for calculating statistically significant temporal hotspots for the county of Skåne in Sweden for two arbitrarily chosen months in 2014. In Fig. 2 (a) the hotspots

for January 2014 is shown and it is clear that there are significant hotspots of both over and under-represented burglary frequencies. Fig. 2 (b) shows a similar visualization, but for October in 2014 instead. In both figures there are statistically significant hotspots during the afternoon and early night, while there are statistically significant decreased frequencies during the night.

V. DISCUSSION

This work presents examples of the proposed method that focus on burglary events. However, the method works equally well on other types of events. It could for instance be interesting to generate significant hotspots for different crime categories, e.g. robberies, various types of burglaries and Diesel thefts. Further, an aggregated visualization of all criminal events could also be interesting as it gives an overview of time slots with high criminal frequencies. Regardless of which approach is used, it is important to carefully select the underlying geographical area that is studied. The smaller the area the more useful for tactical and operational actions, but the less criminal events are (generally) included.

This method allows comparison of temporal hotspots between geographical areas, e.g. are the temporal hotspots the same for geographically neighboring hotspots? As such, this would allow law enforcement to investigate how different types of events spread across a larger geographical area.

The proposed method is not limited to only criminal events, but is applicable to events where a temporal pattern is present, e.g. intrusion attempts recorded by IDS or determining peaks of use/request for various resources. For example service providers could analyze temporal attack patterns against their clients to establish whether a connection exists between the attacks. Although, it is important to keep in mind the problem with offenders in different time zones that blur temporal analyses.

The proposed method could also be used as basis for automating detection or prediction of significant temporal hotspots. As such, it would be interesting to investigate the predictive-value of the presented method by evaluating to what extent past significant temporal hotspots could predict future ones. This could be investigated using e.g. a rolling-horizon setup where the past couple of months of temporal data is used for calculating the statistically significant hotspots and then test to what extent these still hold a few weeks into the future. Further, based on known spatio-temporal hotspots it would be interesting to investigate whether it is possible to detect changes over time (trends) and in those trends detect statistically significant hotspots. Thus, there are several avenues for interesting continuations.

VI. CONCLUSIONS AND FUTURE WORK

This paper presents a novel method for detecting statistically significant temporal hotspots of events, e.g. criminal events. Determining which hotspots that are statistically significant is useful for various reasons, for instance when planning when to implement actions to respond to the occurring events. For

future work it would be interesting to evaluate the predictive value of the proposed method. It would further be interesting to use the method on criminal events in known spatial hotspots. Finally, it would also be interesting to analyze how these temporal hotspots relates to the ones in geographically neighboring spatial hotspots.

REFERENCES

- [1] W. L. Perry, B. McInnis, C. C. Price, S. Smith, and J. S. Hollywood, "Predictive Policing," RAND Corporation, Tech. Rep., 2013.
- [2] W. J. Bratton and S. W. Malinowski, "Police performance management in practice: taking COMPSTAT to the next level," *Policing*, vol. 2, no. 3, 2008.
- [3] J. H. Ratcliffe, "Aoristic Signatures and the Spatio-Temporal Analysis of High Volume Crime Patterns," *Journal of Quantitative Criminology*, vol. 18, no. 1, pp. 23–43, Feb. 2002.
- [4] R. B. Santos, *Crime Analysis With Crime Mapping*, 3rd ed. SAGE Publications, Inc, 2013.
- [5] S. Chainey and J. Ratcliffe, *GIS and crime mapping*. John Wiley & Sons, Inc., 2005.
- [6] S. J. Rey, E. A. Mack, and J. Koschinsky, "Exploratory Space–Time Analysis of Burglary Patterns," *Journal of Quantitative Criminology*, vol. 28, no. 3, pp. 509–531, Nov. 2011.
- [7] E. Johansson, C. Gahlin, and A. Borg, "Crime hotspots : An evaluation of the kde spatial mapping technique," in *Proceedings - 2015 European Intelligence and Security Informatics Conference, EISIC 2015* :, 2015, pp. 69–74, conference of European Intelligence and Security Informatics Conference, EISIC 2015 ; Conference Date: 7 September 2015 Through 8 September 2015; Conference Code:119026.
- [8] M. P. Ashby and K. J. Bowers, "A comparison of methods for temporal analysis of aoristic crime," *Crime Science*, vol. 2, no. 1, pp. 1–16, 2013.
- [9] M. Felson and E. Poulson, "Simple indicators of crime by time of day," *International Journal of Forecasting*, vol. 19, no. 4, pp. 595–601, Oct. 2003.
- [10] L. Tompson and M. Townsley, "(Looking) Back to the Future: using space–time patterns to better predict the location of street crime," *International Journal of Police Science & Management*, vol. 12, no. 1, pp. 23–40, Feb. 2012.
- [11] J. H. Ratcliffe, "Aoristic analysis: the spatial interpretation of unspecific temporal events," *International Journal of Geographical Information Science*, vol. 14, no. 7, pp. 669–679, 2000.
- [12] L. E. Cohen and M. Felson, "Social Change and Crime Rate Trends: A Routine Activity Approach," *American Sociological Review*, vol. 44, no. 4, pp. 588–608, Feb. 1979.
- [13] C. Brunson and J. Corcoran, "Using circular statistics to analyse time patterns in crime incidence," *Computers, Environment and Urban Systems*, vol. 30, no. 3, pp. 300–319, May 2006.
- [14] M. Boldt and A. Borg, "Evaluating temporal analysis methods using residential burglary data," *ISPRS International Journal of Geo-Information*, vol. 5, no. 9, 2016. [Online]. Available: <http://www.mdpi.com/2220-9964/5/9/148>
- [15] J. H. Ratcliffe and M. J. McCullagh, "Aoristic crime analysis," *International Journal of Geographical Information Science*, vol. 12, no. 7, pp. 751–764, Nov. 1998.
- [16] A. Borg, M. Boldt, N. Lavesson, U. Melander, and V. Boeva, "Detecting serial residential burglaries using clustering," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5252–5266, 2014.
- [17] L. Anselin, "Local indicators of spatial association—lisa," *Geographical Analysis*, vol. 27, no. 2, pp. 93–115, 1995. [Online]. Available: <http://dx.doi.org/10.1111/j.1538-4632.1995.tb00338.x>
- [18] A. Getis and K. J. Ord, *Local Spatial Statistics: An Overview*. John Wiley and Sons, 1996.