

Forecasting of Crime Events Applying Risk Terrain Modeling

Milena Kocher¹ and Michael Leitner²

¹Department of Geoinformatics – Z_GIS, University of Salzburg, Austria ·
milena.kocher@stud.sbg.ac.at

²Department of Geography and Anthropology, Louisiana State University, USA

Full paper double blind review

Abstract

This research project deals with the implementation and evaluation of the Risk Terrain Modeling (RTM) technique, which allows localizing places, where the probability is high that a crime event will take place. RTM does not focus on previous events that happened, but on risk factors, which have an influence on the environment and can increase the probability of the risk that a crime will be committed. RTM is a recently developed approach, that has not yet been tested in Austria. Using the example of the city of Salzburg, predictions are made for the crime events assault, auto theft, burglary, and robbery for 2013 and 2014. In addition, the results of 2013 are evaluated and compared. Using the RTMDx Utility software, risk factors that correlate with the crime event as well as their influence can be identified. Based on these results, the risk factors can be operationalized to risk map layers, which is done using two models developed within ArcGIS. After that, the risk map layers are combined to a final risk terrain map, which is classified and finalized according to cartographic aspects. The evaluation for the predictions of 2013 is done using the Predictive Accuracy Index (PAI), based on a model developed in ArcGIS. The results of the evaluation and the percentage of correctly predicted crime events in respect to the size of the predicted areas are shown. In sum, 27 models were calculated and predictions made, because crime events have been partly separated into seasons or sub-types. The best predictive accuracy is reached for assaults, which includes results for the seasons spring and summer and for robberies, with PAI values of 31, 23, and 18, respectively. In contrast, the predictions for burglaries and auto thefts performed rather poorly, with PAI values between two and four. Overall, the RTM technique can be applied to Austrian cities, although the accuracy of the predictions varies. Additionally, the availability and quality of the risk factor data are crucial for the accuracy of the predictions.

1 Introduction

In 2012, the city of Salzburg reported 15,201 crimes, which constitutes more than 40 reported crimes every day, whereby property crimes have the largest proportion with about two-thirds of all reported criminal offenses (SALZBURGER NACHRICHTEN 2013). The tasks of the police not only include recording criminal offenses and their investigations, but also focus on crime prevention. Appropriate measures have to be taken into account in order to protect potential victims and prevent future crimes. One part of the preventive work is the

aversion of crimes (JURA FORUM 2014). For this purpose, several techniques exist in geoinformation, “(...) such as Hotspot Mapping, Near Repeat Analysis, and Risk Terrain Modeling (RTM)” (KENNEDY et al. 2011, 1), among others.

Motivated by the number of reported crimes and the fact that Risk Terrain Modeling can support the police, the primary goal of this research project is to implement and evaluate the RTM technique for the first time for an Austrian city, using the example of the city of Salzburg. The project is carried out within the context of the research project Criminal Predictive Analytics (CriPA), which focuses on predictions of future criminal activity (JOANNEUM RESEARCH 2014). Risk terrain models are implemented for 2013 and 2014 for the four crime types assault, auto theft, burglary, and robbery, and varying spatial influences of risk factors. For each crime type for 2013 four risk terrain models are implemented, including different spatial influences. After the model and the most appropriate spatial influences are selected, the prediction for 2014 is made. The predictions are based on a one-year period. This timeframe is reasonable for a general predictive model¹. If no results can be calculated, the risk terrain models are implemented for seasons. The results can then be compared and evaluated, regarding the impact of available risk factor data and different spatial influences of those on the final risk terrain map. Furthermore, the amount of correctly predicted crime events for 2013 is evaluated. The research results show if and for which crime types the method is appropriate, or otherwise outline possible reasons, why it is not appropriate.

2 Literature Review and State of the Art

The following sections provide an overview of crime analysis and prediction, together with the assumptions behind criminogenic factors. The approach of RTM is described in more detail.

2.1 Crime Analysis and Prediction

Crime analysis methods, in particular predictive crime analytics, gained in importance over the last 5-10 years. Instead of reacting to committed crimes the developed approaches show a trend towards predictive policing, which allows predictions as to where and when a crime event is likely to happen in the future (PERRY et al. 2013). This includes the usage of Geographic Information Systems (GIS) and spatial statistical methods that allow the forecasting of future crime locations based on crime event data from the past or other data, such as environmental data (FERGUSON 2012).

2.2 Criminogenic Factors

Years of research and experience of police agencies revealed that some areas have a high concentration of crime, also called hotspots (CAPLAN & KENNEDY 2010). A city is structured through basic elements, such as the location of the city in a larger region, the road network, land use, and socio-economic factors. These factors result in locations that have a

¹ CAPLAN, J. M., Personal communication, 19 February 2014.

higher probability of crime concentrating there (BRANTINGHAM & BRANTINGHAM 2008). RTM is based on so-called criminogenic factors, or risk factors, that correlate with crime events, and could be identified through research. The RTM Compendium lists these factors for 17 different crime events. The factors do not produce the crime, but influence the environment and can point out locations which are at a higher risk that a crime event will occur (CAPLAN & KENNEDY 2011), due to the fact that offenders are influenced by “(...) situational and environmental features” (GROFF & LA VIGNE 2002, 32).

2.3 Risk Terrain Modeling

Geoinformation technologies play an essential role in fighting against crime in the USA since the 1990s. While hotspot mapping is a widely available method, which is also used in the Federal Criminal Police Office in Austria (KAMPITSCH et al. 2008), RTM is a more recent approach and has not yet been tested in Austria. Criminal predictive analytics is still in its infancy in Austria. However, the Austrian police is very keen on the RTM and other predictive crime modeling approaches, and would like to implement those approaches into their proactive decision-making process.

RTM is a proactive approach to risk analysis and was developed at the School of Criminal Justice at Rutgers University, USA, by Joel M. Caplan and Leslie W. Kennedy (CAPLAN & KENNEDY 2010). Instead of focusing on previous events that happened, as retrospective methods do, RTM focuses on the “(...) dynamic interaction between social, physical, and behavioral factors (...)” (PIZA et al. 2010, 1). The entire process of implementing RTM includes ten steps, which are briefly described in the following. Recently, the Risk Terrain Modeling Diagnostics (RTMDx) Utility software was released, which automates most steps of RTM. In the first step, the study area, the crime event, and the time period have to be determined. Then risk factors are identified and each risk factor is operationalized. This means that for each risk factor, a risk map layer is created, which shows the influence of the risk factor on the environment. There are two operationalization types: proximity, which means that there is a higher risk to be near a risk factor, and density, which defines a higher risk with a higher concentration of risk factors at a particular location. In the end all layers are combined to a risk terrain map, partially with different weights, because some risk factors might have a bigger influence on the environment than others. As a last step, everything is finalized to a final risk terrain map. The risk terrain map shows locations where the risk is high for a future criminal offense. This information can be used by the police and decision-makers for strategic planning and preventive operations (CAPLAN & KENNEDY 2010).

3 Methodology

The following sections present the required data for the analysis and the implementation of the risk terrain models. Further, the necessary steps for the finalization and visualization, as well as the method to evaluate the results are described. Figure 1 gives an overview of the workflow for the implementation.

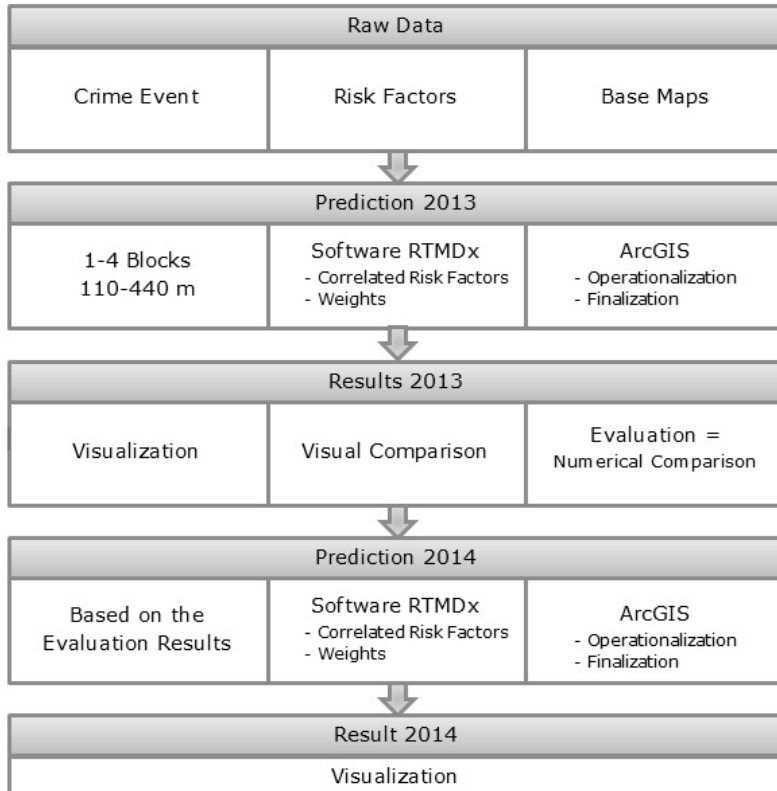


Fig. 1: Workflow for the implementation

3.1 Required Data

For the analysis and visualization, crime data, risk factor data, and base map data are required. Crime data are necessary for testing the correlation with an outcome event and for the evaluation. The risk factors are used for the calculation of the risk terrain models. That can include banks, pawn shops, bus stops, or bars and pubs, just to mention a few. The data representing the base layer are used for the visualization, in order to present the information in a meaningful way, and can include rivers, forests, or the street network. The required data were provided by the Austrian Federal Criminal Police Office, the State Police Headquarters of the city of Salzburg, the Geographic Information System of the city of Salzburg, or self-captured.

3.2 Implementation of Risk Terrain Models

The basic workflow to create risk terrain models includes two steps: First the RTMDx Utility software is used to identify correlated risk factors and their influence. Subsequently, the risk factors are operationalized to risk map layers in ArcGIS and then combined to create the risk terrain map.

For the calculation with the RTMDx software, the required parameters have to be chosen. This includes a shapefile of the study area, the block length, the cell size, as well as the model type, the outcome event data, a name for the model, and its output location. The average block length is calculated based on a digitized street network of the city of Salzburg and results in 110 meters. The value for the raster cell size is determined to half of this size, as recommended by CAPLAN et al. (2013). The outcome event data (crime locations) are only used for the calculation of the correlation with the risk factors. Crime locations are not included in the risk terrain models, as it is done by retrospective methods².

Each risk factor that is included in the model is tested for its spatial influence on the outcome event (CAPLAN et al. 2013). The maximum spatial influence determines within which distance the risk factor has an influence on the environment. This can be up to four blocks, because empirical research has shown that this is the maximum spatial influence. In this research, the models are calculated for four different spatial influences, from one to four block lengths. The operationalization type includes how the risk factor should be operationalized and can be set to proximity, density, or both. We chose to follow the description of the risk factor in the RTM Compendium, or decided to apply both proximity and density, if the option was not obvious. The last parameter defines if half- and whole-block increments, or only whole-block increments are applied. We decided to set the increment parameter to half, in order to get more accurate results (CAPLAN et al. 2013).

After all parameters are chosen and the calculation is completed, the result is summarized in form of a report. This includes the best model and lists the correlated risk factors, their weight, the operationalization type, and the spatial influence on the outcome event.

The statistics implemented in the RTMDx Software identify which risk factors are significantly correlated with the outcome event and calculate their spatial influence. While the input has to be in vector format, the calculations and the output are in raster format. The operationalization type “proximity” measures whether a raster cell is within a particular distance to the risk factor, which is represented by the spatial influence parameter. For “density”, it is tested if a raster cell is within a highly concentrated risk factor area. If that is the case, then cells with values greater than two standard deviations above the mean are extracted. For the operationalization process, a series of variables (risk factors) is tested to identify the optimal model that only includes the most correlated risk factors and their spatial influences (HEFFNER 2013).

Two versions of the RTMDx Utility exist, namely a professional version and an educational one. For this research project the free educational version is used, which does not automatically produce a final map, compared to the professional version, where the final risk terrain map is provided in the form of a GeoTIFF image. This task has to be implemented manually in the educational version by operationalizing the risk factors to risk map layers (CAPLAN et al. 2013).

After the risk terrain models are calculated, the operationalization and finalization can be done. For the operationalization of the types “proximity” and “density”, models are developed using the Model Builder. Both models operationalize a risk factor showing its influence on the environment, as explained above, and result in binary values that show cells

² CAPLAN, J. M., Personal communication, 5 April 2014.

which have an influence or not. After all risk factors are operationalized, the individual risk map layers have to be combined with different weights. This can be done using the “Raster Calculator”, based on the provided formula included in the report. The risk values of the risk terrain map are then shown by a raster through stretched values by default.

3.3 Finalization and Visualization

In order to compare the risk terrain maps visually, a standardized classification scheme has to be defined. For all produced risk terrain models, four classes are created, which are based on standard deviations, whereby the break points are set manually. The first class, “Low Risk” includes all cells with values below the average value. The second class “Medium Risk” includes all cells with values above the mean and below one standard deviation (above the mean). The third class “High Risk” represents values above one standard deviation (above the mean) but below two standard deviations (above the mean). All values greater than two standard deviations (above the mean) make up the fourth class “Highest Risk”. The classes are represented by different red hues to show the risk. The class with the lowest risk is not assigned a color, in order to not overload the map. After the classification, the map can be designed using different base maps, which enables an easier interpretation for the user.

3.4 Evaluation and Comparison

Testing the predictive validity of the final model is an important task, built upon a simple principle: A predictive model is implemented and then it is tested how many crimes indeed happened in the predicted areas. Through the evaluation it can be shown with which certainty the results can be used and how good a method is to predict future crime events (CHAINEY et al. 2008). Also differences between the created models can be determined, thus the most appropriate one can be identified.

First, a visual comparison of the predictions for 2013 is made, adding the outcome event data to the map to get an impression of how many crime events fall into predicted areas. For the numerical evaluation of the results the Predictive Accuracy Index (PAI) is used. In its calculation, the PAI does not only include the size of the predicted areas, but also the size of the whole study area. To calculate the PAI, the hit rate is divided by the percentage of the predicted areas in relationship to the whole study area. The hit rate is defined as the number of crime events, which reside in the predicted areas divided by the total number of crime events (CHAINEY et al. 2008).

For the calculation of the PAI another model was created to calculate the PAI value automatically. Each model is evaluated with the risk classes “Highest Risk”, “Highest and High Risk” and “Highest, High, and Medium Risk”.

4 Results

In sum, 42 different models were calculated with the RTMDx Utility software, whereby 27 of these models identified at least one correlated risk factor. For the other 15 models, no risk factors correlating with the crime event could be found. The 27 successfully calculated

models were further operationalized and finalized to a risk terrain map, showing the prediction for a particular outcome event for 2013 or 2014.

For the crime type assaults, the calculation for one year was not successful, thus the models were created for each season. For spring and summer, predictions could be made, for fall and winter no risk factors correlated with assaults. This can be an indicator that the spatial distribution of assaults varies during the year. For auto thefts, only the models based on two, three, and four blocks were successful for 2013. For 2014, no model could be calculated. The crime event burglary was separated into all burglaries and burglaries into buildings. For both categories all models could be calculated. For robberies, all models for 2013 and 2014 could be calculated.

The PAI values for the above described crime events are given in Figure 1. The highest PAI value (31.37) was calculated for assaults in spring. Also, the model for assault in summer showed a high value of 23.40. The models for auto theft with a value of 1.72 and for burglaries with a value of 4.46 show the lowest PAI values. The PAI for robberies resulted in a value of 18. It can be seen that the models for assaults as well as for robberies have high PAI values compared with the values for auto thefts and burglaries.

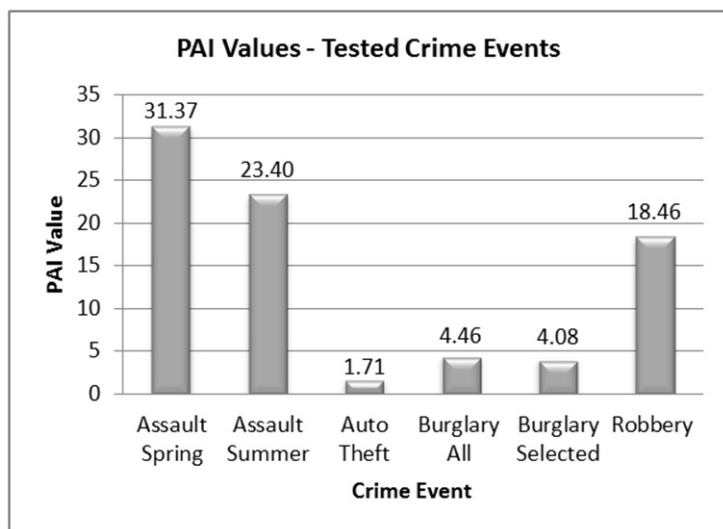


Fig. 1: Comparison of PAI values for different crime events

Table 1 summarizes the correctly predicted crime events and the area for the six predictions for 2013. This information is particularly interesting for the police and other decision-makers. In general, between 25% and 44% of crimes of a particular type could be predicted correctly. The modeled areas range from 0.78km² to 9.9km².

Table 1: Comparison of correctly predicted crime events

Crime Event	Correctly Predicted Crimes	Area in km ²
Assault Spring	153 of 412	37.14%
Assault Summer	198 of 446	44.39%
Auto Theft	9 of 35	25.71%
Burglary all	629 of 2473	25.43%
Burglary selected	252 of 765	32.94%
Robbery	37 of 85	43.53%

The same information of correctly predicted crime events is shown in Figure 2. It is visible that the result for auto thefts, depicted by the triangle symbol, is not very good. Only 25% were predicted correctly and the area of nearly 10km² is rather large. The results for burglary (all burglaries and selected burglaries) are relatively good, compared to the results of the other crime types. Between 25% and 32% were predicted correctly for an area of 3.7 and 5.3km², respectively. The best results were achieved for assaults (spring and summer) as well as for robbery. The result for assault in spring correctly predicted 37% of the crimes, located inside the smallest area of only 0.78km². The results for assaults for summer and for robbery for the entire year correctly predicted 44% and 43%, for an area of 1.25 and 1.32km².

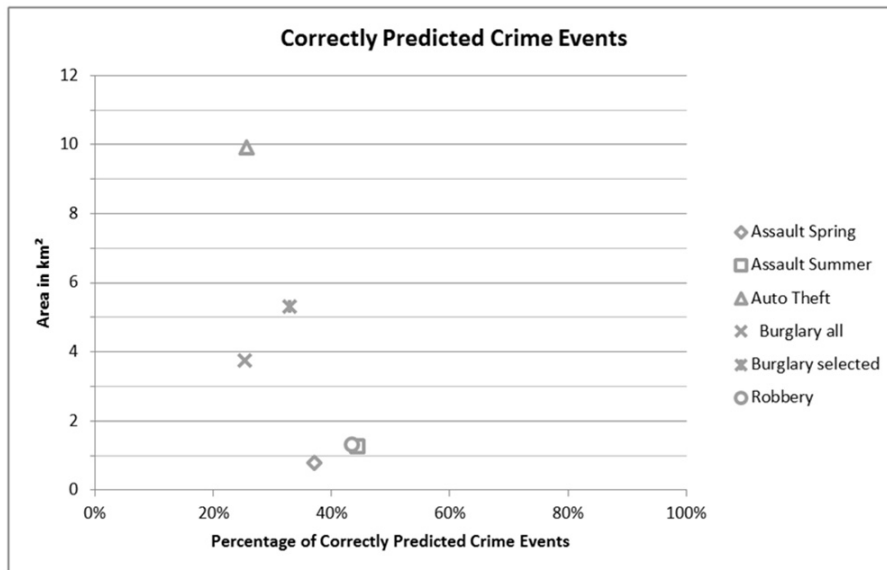


Fig. 2: Comparison of correctly predicted crime events

The reason for the poor prediction of auto thefts might be mainly due to the few available risk factors. Although there was no result for assaults based on a one year time period, the results for spring and summer are good. For robberies, all risk factors were available and also the high data quality of robberies may have led to such a good result. The high data

quality can be explained by the fact that the time and location of a robbery can be reported exactly.

As an example, the finalized risk terrain map for the prediction of robberies for 2014 is presented in Figure 3. Areas with a different degree of risk that a robbery will take place in 2014 are shown. The map represents the city of Salzburg and the street network. Further base layers are not included, due to visualization issues for the black and white image. Additionally, the risk value classes can be seen. The “Highest Risk” class predicts areas with a size of 1.67 km², additionally including the “High Risk” class, an area of 2.33 km², and in combination with the “Medium Risk” class almost 10 km² are predicted.

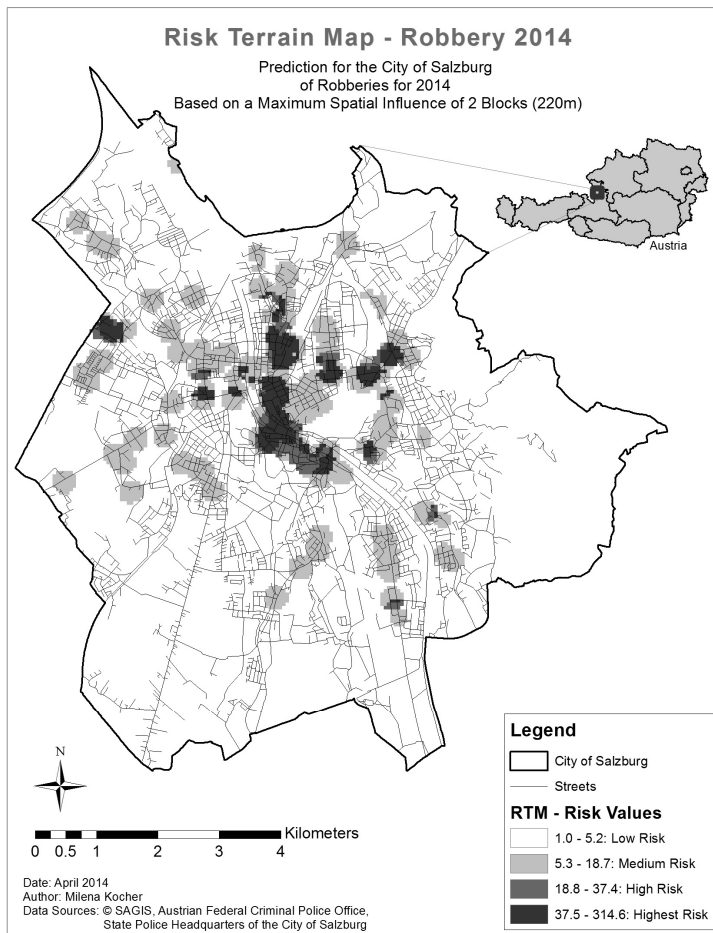


Fig. 3: Risk terrain map of robberies for 2014

5 Conclusion and Outlook

How good the RTM technique works for a specific project area is affected by different factors. Above all, the availability and quality of the risk factor data is crucial. Since this technique was developed and implemented in the USA, the selected risk factors listed in the RTM Compendium are influenced by research done in the USA. It might be possible that for the project area of this research, the city of Salzburg, further risk factors exist that were not included in this research. In addition, not all risk factor data listed in the RTM Compendium were available or could be obtained because of a lack of expertise, as is true for the factor social disorganization. The other risk factor data was difficult, if not impossible, to obtain within the context of this work, e.g. for household income or location of older vehicles parked. A further aspect that has to be considered is that both the obtained as well as the self-captured risk factor data were partly too old (e.g. from 2009), or the timeliness of data was not declared.

The expected results and goals of this research project were reached. Risk terrain models for the crime events assaults, auto thefts, burglaries, and robberies were implemented and their results evaluated. Furthermore, based on the evaluation results, predictions for 2014 were done, which can be used by the Salzburg Police. For the operationalization and the evaluation processes, models could be implemented to enable an automated process.

In summary, this research project showed that the RTM technique can be applied to an Austrian city; however, there are differences regarding the accuracy of the predictions. For future projects, it has to be considered that, above all, the availability and quality of the risk factor data are crucial for the accuracy of the predictions. Another factor is the spatial distribution of crime events, which can vary during the year, and might lead to the conclusion that for some crime events predictions should be made on a seasonal basis rather than on a one-year period. Risk factors, which are specific for the study area, but were not included in this research, as well as socio-economic factors, which were also not implemented, could have improved the risk terrain models, too. Different evaluation methods, which could include the number of crime incidents per year, could be evaluated to see if the results differ regarding the methods tested.

Acknowledgement

This study has been funded by the Austrian security research programme KIRAS of the Federal Ministry for Transport, Innovation and Technology (bmvit).

References

- BRANTINGHAM, P. J. & BRANTINGHAM, P. L. (2008), Crime pattern theory. In: WORTLEY, R. & MAZEROLLE, L. (Eds.), *Environmental Criminology and Crime Analysis*, 2008, 78-93.
- CAPLAN, J. M. & KENNEDY, L. W. (2010), *Risk Terrain Modeling Manual*.
- CAPLAN, J. M. & KENNEDY, L. W. (Eds.) (2011), *Risk Terrain Modeling Compendium*.

- CAPLAN, J. M., KENNEDY, L. W., PIZA, E. L. (2013), Risk Terrain Modeling Diagnostics Utility User Manual (Version 1.0). http://www.rutgerscps.org/docs/RTMDxUserManual_FinalDraft_CaplanKennedyPiza.pdf (17 March 2014).
- CHANEY, S., TOMPSON, L., UHLIG, S. (2008), The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime. *Security Journal*, 2008, 21, 4-28. <http://www.palgrave-journals.com/sj/journal/v21/n1/full/8350066a.html> (8 April 2014).
- FERGUSON, A. G. (2012), Predictive Policing and Reasonable Suspicion. *Emory Law Journal*, 62 (2), 259-325. Download via LSU Libraries Homepage.
- GROFF, E. R., LA VIGNE, N. G. (2002), Forecasting the Future of Predictive Crime Mapping. *Crime Prevention Studies*, 3, 29-57. http://www.popcenter.org/-library/crimeprevention/volume_13/03-groff.pdf (20 May 2014).
- HEFFNER, J. (2013), Statistics of the RTMDx Utility. In: CAPLAN, J., KENNEDY, L. & PIZA, E. (Eds.), Risk Terrain Modeling Diagnostics Utility User Manual, 35-39. http://www.rutgerscps.org/docs/RTMDxUserManual_FinalDraft_CaplanKennedyPiza.pdf (9 April 2014).
- JOANNEUM RESEARCH (2014), CRIPA Crime Predictive Analytics. <http://www.joanneum.at/policies/referenzprojekte/cripa.html> (13 May 2014).
- JURA FORUM (2014), Gefahrenabwehr. <http://www.juraforum.de-/lexikon/gefahrenabwehr> (11 May 2014).
- KAMPITSCH, K., LEITNER, M. & SCHABAUER, H. (2008), Der Einsatz von Geografischen Informationssystemen im österreichischen Bundeskriminalamt. http://www.agit.at/php_files/myagit/papers/2008/6834.pdf (11 May 2014).
- KENNEDY, L. W., CAPLAN, J. M., PIZA, E. L. (2011), Hotspot mapping, near repeat analysis, and risk terrain modeling. Presentation to UCL Department of Security and Crime Science. <http://www.ucl.ac.uk/secret/secret-student-seminars-11/hotspot-mapping> (11 May 2014).
- PERRY, W. L., MCINNIS, B., PRICE, C. C., SMITH, S. C., HOLLYWOOD, J. S. (2013), Predictive Policing – The Role of Crime Forecasting in Law Enforcement Operations. http://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.sum.pdf (21 February 2014).
- PIZA, E. L., KENNEDY, L. W., CAPLAN, J. M. (2010), Risk Clusters, Hotspots, and Spatial Intelligence: Risk Terrain Modeling as an Algorithm for Police Resource Allocation Strategies. *Journal of Quantitative Criminology*, 27 (3), 339-362. http://www.rutgerscps.org/docs/NewarkRTM_CaseStudy_Brief.pdf (25 February 2014).
- SALZBURGER NACHRICHTEN (2013), Kriminalstatistik Salzburg: Leichte Anzeigen-Zunahme. <http://www.salzburg.com/nachrichten/salzburg/chronik/sn/artikel/-kriminalstatistik-salzburg-leichte-anzeigen-zunahme-46249> (11 May 2014).