Analyzing Spatio-Temporal Effects of Social-Economic Factors on Crime

Sebastian Baumbach
German Research Center for Artificial Intelligence (DFKI)
Technische Universität Kaiserslautern
Kaiserslautern, German
Email: sebastian.baumbach@dfki.de

Nikita Sharm
Technische Universität Kaiserslautern
Kaiserslautern, German
Email: nikita.sharma1108@gmail.com

Sheraz Ahmed
German Research Center for Artificial Intelligence (DFKI)
Kaiserslautern, German
Email: sheraz.ahmed@dfki.de

Andreas Dengel
German Research Center for Artificial Intelligence (DFKI)
Technische Universität Kaiserslautern
Kaiserslautern, German
Email:andreas.dengel@dfki.de

***Abstract*—Rampant increase in crime incidents has led to the need of crime analysis in greater detail. Existing crime analysis approaches focused on higher spatial granularity (i.e., country or state levels) and consider each data observation independent of each other. However, data can exhibit spatial and temporal relationships among them. Such interrelationships must be taken into consideration if precise crime analysis is intended. Therefore, a two-stage approach is proposed for predicting crime by analyzing its relationship with socio-economic factors: the first stage applies a spatio-temporal analysis on the data and these results are utilized for the spatio-temporal prediction, which forms the second stage. For evaluation, more than 450 different socio-economic factors and crime data for county level in Germany were analyzed. The evaluation results exhibit a mean absolute percentage error of 6.79% for spatio-temporal crime predictions, outperforming traditional regression techniques with an error rate of 37.1% - 37.8%.**

Keywords–Staptio-temporal Data Mining; Crime Analysis; Prediction Models; Location Factors;

I. INTRODUCTION

Crime has been a recurrent activity since the beginning of society evolution. Crime incidents can be traced to as early as imperial era in history. As McCollister et al. stated in their work [1], such incidents have been a deterrent to social harmony and have affected the development of communities. The authors specified the effects of crime in society in terms of economic development, as well as society integration and suggest effective measures for government policies to reduce criminal activities.

A detailed insight on crime is required as concluded in [1], as it can benefit the inland security services for effective police force deployment. Cunningham et al. [2] described that crime analysis is necessary to provide better law enforcement in a region and maintain integrity as well as peaceful environment of the society. Crime analysis assists surveillance forces to make preemptive decisions and hence, ensures better vigilance and control of a crisis situation [1] [2].

Consequently, crime analysis is an active research field. There are numerous studies like [3] [4] that explored social media data (e.g., from Twitter) to predict crime. These works conducted sentiment analysis on Twitter posts collected to predict crime at a specific location. There is a wide range of works studying effect of social-economic factors on crime [5] [6]. For instance, Chainey et al. [7] assessed the relationship between crime and social-economic factors at the state level in the USA. Similarly, Entorf and Spengler [8] analyzed the effects of social-economic data on crime at state level in Germany. Caruso and Schneider conducted a similar analysis in their work [9] by focusing on crime trends at a higher geographical level, i.e., comparing crime trends between different European countries.

All the above-mentioned work focused either on higher spatial granularity (e.g., [9]) or on data content, such as analysis of social media data for sentiment detection (e.g., [4]). None of these studies focused on the detailed relationship of crime and its influencing factors at lower spatial and temporal granularity. Neither of these researchers assessed the interrelationship that exists in crime data.

Referring to the problem stated for the existing approaches, there arises the need for analyzing crime data on a deeper spatial and temporal granularity. Additionally, there is the *First Law Of Geography* by Walder Tobler' which says "*everything is related to everything else, but near things are more related than distant things*" [10]. Thus, the proposed two-stage approach utilizes spatial and temporal data correlations to predict crime intensity and applies it at lower spatio-temporal granularity. The first stage, spatio-temporal analysis, focuses on validating the existence of spatio-temporal relationships in data and allows for the selection of the best feature subset for crime prediction. The second stage, spatio-temporal prediction, exploits the spatio-temporal proximity in data and predicts crime rate by utilizing spatio-temporal prediction models. The evaluation was carried out on a county level collection of social-economic factors and crime data.

Summarizing, the main contributions of this paper are defined as follows: i) a spatio-temporal analysis approach that gives a detailed insight into crime and social-economic data trends in space and time domain, ii) a significant improvement of 32% in crime prediction over existing regression approaches by utilizing spatio-temporal prediction, and iii) a spatio-temporal prediction approach, which can predict crime at county level in Germany. Consequently, spatio-temporal relations in a dataset decrease the error rate in crime prediction and enhance the performance of existing prediction models.

The rest of this paper is organized as follows. Section II summarizes and assesses the state-of-the-art in crime data

 ISBN: 978-1-61208-617-0

analysis by focusing on the explored data sources, as well as utilized methodologies. Section III presents the proposed approach for analyzing crime with a high spatio-temporal granularity. Section IV depicts the applied crime dataset together with the socio-economic location factors, which are used in this paper for evaluating the proposed approach. Section V presents the findings of this paper, stating that spatio-temporal data interactions increase the prediction capability for crime analysis. Finally, Section VI discusses and summarizes the results.

II. Related Work

In the scientific community, there is a vast range of work on crime analysis. The existing approaches focused on enhancing traditional regression techniques or deep learning for crime analysis and prediction. Their authors focused on tuning these models based on crime relationship with influencing factors, such as Twitter data, social-economic factors, or background data of criminals. However, each observation in these datasets is considered as independent, i.e., these approaches consider no relation to be existing between individual records in the dataset.

A. Exploring Data Sources for Crime Analysis

Data sources, such as social media, criminal records, or ideological beliefs of listed terrorist organizations, were explored in various studies to gain detailed insights into crime trends. Acquiring social media data with a high spatial granularity is difficult as geo-referenced social media content is hardly available [11]. Furthermore, analyses of crime with social media data, criminal data, and their ideological beliefs are based on subjective analysis. The sentiments respective the intentions of people are evaluated based on a list of words termed as 'hate' words. These approaches do not necessarily amount to crime intentions.

1) Social Media Content: Wang et al. [3] applied semantic analysis and natural language processing on Twitter data to find topics of discussion on social media. The authors proved that these topics can be indicators of future crime incidents by analyzing previous crime incidents and the topics of discussion on social media at their time of occurrence. Gerber [4] used a Twitter-specific linguistic analysis and a statistical topic modeling to automatically identify discussion topics across a major city in the United States. Other studies [12] [13] focused on determining the general population's sentiment in a certain regions by conducting sentiment analysis on microblogging sites like Twitter.

2) Social-Economic Data: Caruso and Schneider [9] performed an empirical evaluation on social-economic determinants of crime. Their work was based around the hypothesis stating that social-economic factors (such as population, migration, and poverty) determine factors of crime. Edmark [5] explored the relation between unemployment and crime using regression methods. Freytag et al. [6] applied regression techniques to conclude whether social-economic factors have an impact on crime. Entorf and Spengler [8] utilized panel regression on social-economic and crime data to predict crime incidents at state level in Germany.

3) Crime Data and Criminal Beliefs: There exists a number of researches that explored past records of criminals and their ideological beliefs to analyze crime. Martinez et al. [14] assessed the relationships between past actions of criminals and their associated behavior. This relationship was utilized to predict actions based on the current observed behavior of criminals. Sampson and Groves [15] measured the society integration, i.e., how well people are connected and integrated in a community. The authors explored the direct relationship of social integrity with the number of crime incidents as listed by the Federal Bureau of Investigation, USA.

B. Methodologies of Crime Analysis

Traditional regression techniques and neural networks are based on a frequentist approach and rely on a large data sample to train, learn and estimate crime incidents. Thus, Bayesian approaches have their advantages over neural network and other probability based regression techniques (frequentist approach) when it comes to the analysis of (spare) crime datasets. Bayesian approaches add a degree of uncertainty to the prediction methods and thus, emulate a real world situations closely compared to the frequentist approach.

1) Regression Techniques: A range of work on crime prediction is based on regression analysis. Edmark [5] performed a panel regression on data from Swedish counties over the time period 1988 - 1999. The author focused on analyzing the impact of unemployment on crime. Entorf and Spengler [8] utilized logarithmic panel regression on demographic and economic data of German states to predict crime. Caruso and Schneider [9] applied a negative binomial regression on social-economic panel data of western European countries. Freytag et al. [6] applied the same approach on data of 110 countries to test the hypothesis that poor socio-economic development leads to rise in terrorism.

2) Neural Network Techniques: A large section of studies in the research community analyzed crime with neural network techniques. Olligschlaeger [16] incorporated the predictions with neural networks by using a geographical information system to forecast the emergence of drug hot-spot areas. The input data are the number of distress calls made to security department in a certain region which were fed to a pre-trained neural network to predict crime prone areas. Caulkins [17] compared the performance of neural network based approaches over statistical methods for crime analysis. The dataset used is an information set about offenders and criminals that includes their imprisonment terms, level of punishments, number of crimes committed. Palocsay et al. [18] researched on neural network approaches to locate recidivists from a dataset of criminals and listed offenders.

C. Spatial Temporal Analysis

A vast range of work on crime analysis applied visual exploration approaches to understand crime patterns and used the derived information to predict crime occurrences. Cheong and Lee [13] performed visual analysis of Twitter data to generate insights on how Twitter data could be a facilitator of crime. Nakaya and Yano [19] conducted an exploratory analysis of crime to facilitate the visualization of the geographical extent and duration of crime clusters.

Wang and Brown [20] proposed the Spatio-Temporal Generalized Additive Model (S-T GAM) to discover the underlying factors related to crime and predict future crime incidents. Wang et al. [20] extended the S-T GAM approach by adding Twitter analysis and concluded that the additive Twitter analysis enhance the predictive performance of S-T GAM.

Other research works emphasized that it is important to find external factors that facilitated the varying spatial or temporal crime patterns. Ivaha et al. [21] devised a crime prediction model that incorporated the effects of weather conditions on changing crime patterns in space and time. Townsley et al. [22] focused on discovering space and time dependencies with crime. The authors investigated the relation between crime incidents that have spatial and temporal proximity. They concluded that the existing proximity relationship between crime data can be used to forecast future crime locations and time of occurrence.

III. Proposed Approach for Crime Analysis

To perform crime analysis with a high spatio-temporal granularity, the proposed approach consists of two subprocesses: Spatio-Temporal Analysis and Spatio-Temporal Prediction. Figure 1 presents the workflow of the proposed approach.

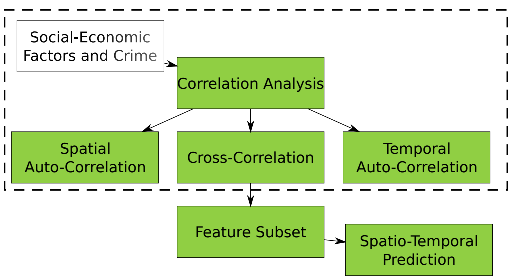
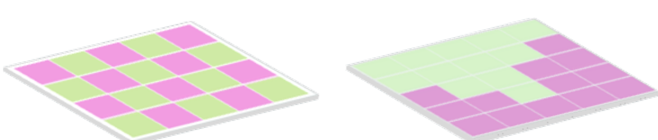


Figure 1. Workflow for proposed approach.

A. Correlation Analysis

The first stage is the spatial and temporal data analysis, for which the spatial, temporal, and cross-correlations are discussed.

1) Spatial Auto-correlaion Analysis: Spatial auto-correlation is an analysis process that measures the association of a variable with itself along the spatial dimension. There exists a number of statistical measures that can be computed for spatial analysis. Moran's I was chosen for this work [23]. The statistical measures for spatial analysis are based on a spatial weight matrix that defines the intensity of the distance relationship among observations (crime data) in a geographical space. The Moran's I ranges from -1 to $+1$ depending on whether the observations are spatially dispersed or clustered.



(a) Negative Auto-correlation. (b) Positive Auto-correlation.

Figure 2. Example of observations with Spatial Autocorrelation.

Figure 2a displays an example of observations with negative spatial auto-correlation. In this case, Moran's I is close to -1 for such values because geographically nearby locations exhibit negative relationship, i.e., they are dispersed and do not form a cluster. Similarly positive auto-correlation is depicted in Figure 2b, where data from geographically close locations form a cluster. In general, an observation dataset with Moran's I close to $+1$ indicates a positive auto-correlation. Moran's I with a value 0 indicates no spatial auto-correlation.

2) Temporal Auto-correlation Analysis: Temporal auto-correlation is a measure of how data at one timepoint is related to data at other timepoints. Figure 3 explains the temporal auto-correlation plot for the social-economic factor "Migration data". The plot depicts how migration data are related to itself at time lags of $0, 1, 2, 3$, and 4. There is a positive correlation at lag 1, i.e., the relation between migration data at consecutive timepoints is a positive slope. The blue dashed line denotes the significant level of correlation. Correlation at any lag that is intersecting this line is defined to be a significant auto-correlation at this lag. The lag with a positive temporal auto-correlation is used in spatio-temporal prediction models as an input parameter.

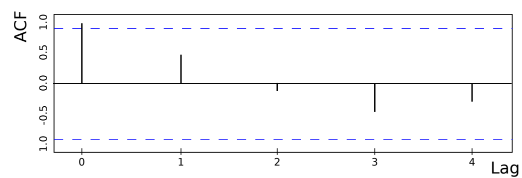


Figure 3. Temporal Auto-correlation for Migration Data Time Series.

3) Cross-Correlation Analysis: Cross-Correlation between two time series is a measure of the lateral effect of one time series over the other. Correlations are calculated between every social-economic factor at timepoint t_{+h} and crime at timepoint t for $h \in \mathbb{N}, h \leq 0$. A negative value for h is a correlation between a social-economic factor at a time before t and the crime variable at time t. When a time series x with h negative are predictors of a time series y at t, it is referred to as x leads y.

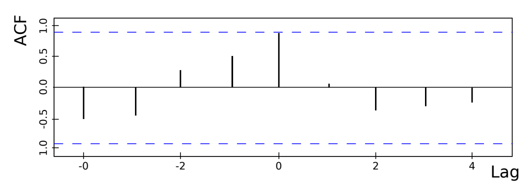


Figure 4. Cross-correlation between Migration and Crime Data Time Series.

Figure 4 depicts an example of cross-correlation between migration data and crime series from the current dataset. At lag -1 and -2, the plot shows a positive cross-correlation. This concludes that at time $t-1$ and $t-2$, migration data can be a positive influence in predicting crime data at time t.

B. Spatio-Temporal Prediction

The proposed approach utilizes spatio-temporal models, which can be categorized into general and dynamic models.

Copyright (c) IARIA, 2018. ISBN: 978-1-61208-617-0

1) General Spatio-Temporal Models: General spatio-temporal models are of 3 types, which differs in the choice of distribution for the process stage model.

Gaussian Process Models are defined as follows:

$$Z_t = O_t + \epsilon_t, \tag{1}$$

$$O_t = X_t\beta + \eta_t, \tag{2}$$

where ϵ_t is the independent normally distributed nugget effect or the pure error term at time unit t and η_t denotes the spacial-temporal random effects following an independent normal distribution. Z_t depicts the observed spatio-temporal data while O_t represents the overall random effects. For the p covariates and n number of observations, X_t denotes the $n \times p$ covariate matrix. $\beta = (\beta_1, ...\beta_p)$ denotes the $p \times 1$ vector of regression coefficients.

To perform predictions at location s at time t, the posterior predictive distribution for $Z(s,t)$ is defined as an integration over the parameters with respect to the joint posterior distribution as:

$$\begin{aligned}\pi(Z(s,t)|z) = \int \pi(Z(s,t)|O_l(s,t), \sigma^2_\epsilon, z)\\ \pi(O(s,t)|\theta)\pi(\theta|z)\partial O(s,t)\partial\theta\end{aligned} \tag{3}$$

where $\theta = (\beta, \sigma^2_\epsilon, \phi, \nu)$ denotes all the model parameters.

Auto-Regressive Models are defined as follows:

$$Z_t = O_t + \epsilon_t, \tag{4}$$

$$O_t = \rho O_{t-1} + X_t\beta + \eta_t, \tag{5}$$

where ρ denotes the unknown temporal correlation parameter assumed to be in the interval $(1, 1)$. The initial value for O_0 is assigned a prior distribution with independent spatial model with mean μ and the covariance matrix $\sigma^2 S_0$.

To perform predictions at location s at time t, the posterior predictive distribution for $Z(s,t)$ is defined as an integration over the parameters with respect to the joint posterior distribution as:

$$\begin{aligned}\pi(Z(s,t)|z) = \int, \pi(Z(s,t)|O_l(s,t), \sigma^2_\epsilon, z)\\ \pi(O(s,t)|\theta, z^*)\pi(\theta, z^*|z)\partial O(s,t)\partial z^*\partial\theta\end{aligned} \tag{6}$$

where $\theta = (\beta, \sigma^2_\epsilon, \phi, \nu)$ denotes all the model parameters. z^* refers to the missing data while z refers to the non-missing data [24].

Gaussian Predictive Model introduces random effects $\eta(s,t)$ at a smaller number, m, of locations, called the knots, and then use kriging to predict those random effects at the data and prediction locations. Hence, the basic Gaussian predictive process model can be represented as:

$$Z(s) = \mu(s) + \eta(s) + \epsilon(s), \tag{7}$$

where, Z(s) denotes vector of observed data for a location s at all timepoints, $\mu(s)$ is the mean function at location s. The residuals are represented in two parts: $\eta(s)$ is the spatially correlated error with a distribution of zero mean and stationary Gaussian process. The second part is the $\epsilon(s)$ which is a non-spatial uncorrelated pure error also distributed normally with mean zero and variance $\sigma^2_\epsilon I$, where I is the identity matrix [24]. The posterior predictive distribution of an unknown location s^* as described in [24] is defined as:

$$\pi(Z(s^*)|z(s)) = \int \pi(Z(s^*)|\theta, z(s))\pi(\theta|z(s))\partial\theta \tag{8}$$

2) Dynamic Spatio-Temporal Models: Bayesian frameworks have the advantage of working with short time series data and can also deal with uncertainties in data by introducing the concept of priors. A detailed explanation of Bayesian modeling can be found in [25]. Bayesian modeling is a statistical inference approach where the Bayes theorem is used to update the probability of unknown variables as more data become known. The Bayesian models involve drawing inference from the posterior distribution of unknown parameters which is proportional to the likelihood of data times a prior knowledge of various model parameters [26], which as can be seen in (9).

$$posterior \propto likelihood \times prior \tag{9}$$

With respect to the crime dataset, Bayesian modeling can be explained as follows: let $x = x_1....x_n$ be the observed social-economic data and $Q = Q_1....Q_p$ be the model parameters with an assumed prior distribution of $\pi(Q)$, the posterior distribution of parameters can be defined as follow:

$$\pi(Q|x) \propto f(x|Q) \times \pi(Q) \tag{10}$$

where $f(x|Q)$ is a likelihood function which determines the probability of observing the data for different values of Q.

Spatial-Temporal processes contain observations in space and time with varying spatial and temporal support and complicated underlying dynamics [27]. Because of the complexity of these processes, hierarchical models are deemed suitable because of their ability to represent joint covariance relationships among process and model parameters into disjoint covariance structures at lower level of the hierarchy model. There are two main variants of Bayesian spatio-temporal process.

General Spatio-Temporal Models are beneficial when data are available across time and space domain. A general spatio-temporal model focuses on spatio-temporal interactions by modeling a joint space-time covariance structure [28]. However, due to high dimensional and complexity of non-linear spatio-temporal behavior, formulating joint covariance structures is highly complicated.

Dynamic Spatio-Temporal Models represent spatio-temporal interactions in a hierarchical framework. The current state of the process is evaluated as a function of previous states [29]. The joint spatio-temporal process Y can be factored into conditional models based on a Markovian assumption. That is,

$$[Y|\theta_t, t = 1, ..., T] = [y_0]\prod_{t=1}^{T}[y_t|y_t - 1, \theta_t] \tag{11}$$

 ISBN: 978-1-61208-617-0

where $y_t = (y(s_1, t), ...y(s_n, t))$ with $y(s_n, t)$ is the process at spatial location s and time t. The conditional distribution $[y_t|y_t - 1, \theta_t]$ depends on a vector of parameters θ_t which govern the dynamics of the spatio-temporal process of interest. Arab et al. gives a detailed explanation of these models in [30].

IV. Dataset

Crime data were obtained from the Federal Criminal Police Office of Germany. The dataset contains the total number of offences for different crime categories per year and location. These data are only publicly available for counties belonging to cities with more than $100,000$ inhabitants. In Germany, there are only a total of 81 sites beyond this population count, which were taken into consideration for this paper. Ultimately, crime data were modeled as a time series for a constant time period from 2009 to 2013.

Furthermore, more than 450 socio-economic location factors were assessed, which are offered by the Federal Statistical Office of Germany. For the evaluation in this paper, 18 social-economic factors were selected based on expert knowledge [6] [8] [9]. These factors include, among others, Gross Domestic Product (GDP), population division, migration population, index of health services, social secured and insured population, literacy level, employment rate, birth and death rate, number of enterprises and businesses, and index of child day care facilities.

V. Evaluation

The conducted evaluation aimed to prove that spatio-temporal data interactions enhance the prediction capability of existing approaches. Thus, a similar comparison approach is followed as described in [31] [32]. The existing models are tested against the given dataset and a comparison is drawn between the error in crime prediction of the proposed approach and existing state-of-the-art approaches.

A. Prediction Evaluation and Ground Truth

The ground truth for this evaluation was generated by two traditional regression models [33]. Ordinary Least Square (OLS) Regression was performed by minimizing the sum of the squares of the differences between observed and predicted values [34]. Panel Regression was calculated over panel data (cross-sectional data across space and time). Table I shows the prediction efficiency of OLS regression and Panel regression on the socio-economic and crime dataset, measured in Mean Absolute Percentage Error (MAPE).

TABLE I. PREDICTION RESULTS FOR GROUND TRUTH MODELS.

	OLS Regression	Panel Regression
MAPE	38%	37.1%

B. Spatio-Temporal Correlation Evaluation

The spatio-temporal analysis' results allowed to reject the null hypothesis, which states that there is no spatial or temporal auto-correlation between observation data and the data spread is random. Hence, spatio-temporal auto-correlation indicates the presence of spatial and temporal interactions and thus, validates the choice for using spatial-temporal models for prediction.

1) Local Spatial Auto-correlation of Crime: Spatial auto-correlation was depicted by visualizing the Local Interactions of Spatial Association (LISA) cluster maps [35]. LISA maps were generated based on the neighborhood weight matrix that represents the relation between locations based on their distance proximity. The spatial association of a region is plotted based on the significance of its Local Moran's I.

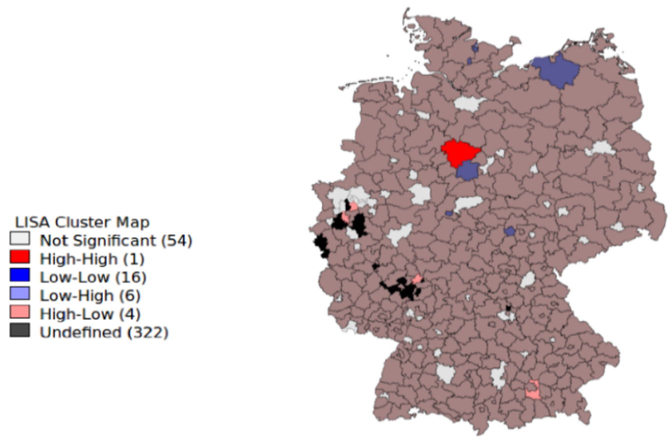


Figure 5. LISA Map for Crime Data of 81 County Sites of Germany.

Figure 5 shows the LISA cluster map for 81 county sites of Germany. The weight matrix is based on a fixed distance band (average distance between two farthest location within the same state). There were 27 such locations with significant spatial clustering. 17 locations depicted a positive spatial correlation and consists of the categories high-high and low-low. Ten regions fell under the category of high-low and low-high and hence, depict a significant negative spatial auto-correlation. For the remaining sites in Germany labeled as "not significant", there were not enough data available to draw conclusions.

2) Temporal Auto-correlation of Crime: Figure 6 shows the temporal auto-correlation in crime data, which is positive at lag 1. Thus, crime occurrence at time $t - 1$ has a positive effect on crime at time t. However, it is below the significant level. The reason for that is most likely that the given time series only have 5 time points. For this research, the lag of 1 for prediction was taken into consideration. However, more data is necessary to eventually clarify this fact.

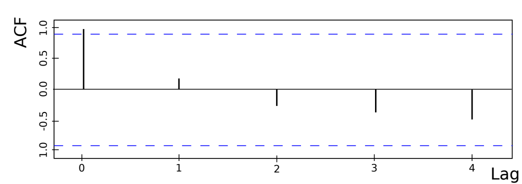


Figure 6. Temporal Auto-correlation Plot for Crime Data of Germany.

3) Cross-Correlation of Crime and Social-Economic Factors: Figure 7 visualizes cross-correlation between the social-economic factor "disposable income" and crime rate in Germany. It depicts the positive correlation between the disposable income of households in Germany and crime rate at the temporal $lag - 1$ and $lag - 2$.

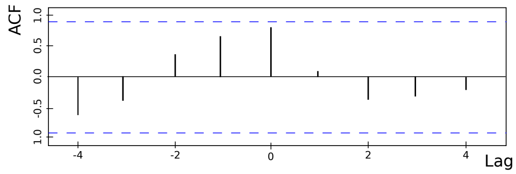


Figure 7. Cross-correlation between Disposable Income and Crime Rate in Germany.

Table II presents the resulting factors selected from a subset of 450 social-economic attributes. These factors have a significant cross-correlation with crime data at various temporal lags. A factor was selected if there was a significant correlation (close to 1 or -1) at lag 0. When there was a weak correlation at lag 0, subsequent cross-correlations at lag -1 and -2 were taken into consideration.

TABLE II. CROSS-CORRELATION BETWEEN SOCIAL-ECONOMIC FACTORS AND CRIME.

Factors	**Lag 0**	**Lag -1**	**Lag -2**	**Lag -3**	**Lag -4**
Migration Data	0.819	0.483	0.252	-0.366	-0.482
Population Data	0.214	-0.425	-0.749	0.420	0.248
Disposable Income	0.741	0.563	0.285	-0.271	-0.580
No. of Employees	0.801	0.596	0.161	-0.388	-0.453
No. of Employer	0.830	0.570	0.131	-0.380	-0.442
No. of Enterprises	-0.058	0.696	0.584	-0.116	-0.586
GDP	0.670	0.554	0.366	0.225	-0.631
No. of Hospital Beds	-0.783	-0.282	-0.358	0.472	0.352
Real Estate Price	0.319	0.469	0.423	0.170	-0.775
Graduate/Dropout Ratio	-0.581	0.134	0.449	0.472	-0.453
No. Social Insured Persons	0.932	0.389	0.048	-0.259	-0.445

C. *Evaluation of General Spatio-Temporal Models*

This section evaluates the crime prediction efficiency of three Gaussian processes and compares it with the ground truth.

Table III shows the result for the Gaussian process model, which produced a MAPE of 36% with the given dataset. Comparing the results with the ground truth, there is not much improvement in the prediction results with Gaussian model.

TABLE III. PREDICTION RESULTS OF GAUSSIAN PROCESS MODEL.

Models	**MAPE**
OLS Regression	38%
Panel Regression	37.1%
Gaussian Process Model	**36%**

Table IV presents the comparison between the ground truth and the Gaussian predictive process model, which gave a MAPE of 37.1%. This model showed a mere improvement of 0.5% over OLS regression. The prediction accuracy was, however, less than for the panel regression and the Gaussian process model.

Table V shows the performance of the auto-regressive model and the comparison with the ground truth. The auto-regressive models produced a MAPE of 28.23%. Comparing the results with the ground truth, auto-regressive models perform better over traditional regression models. Among the spatio-temporal prediction models, auto-regressive models produce the best prediction result.

TABLE IV. PREDICTION RESULTS OF GAUSSIAN PREDICTIVE PROCESS MODEL.

Models	**MAPE**
OLS Regression	38%
Panel Regression	37.1%
Gaussian Predictive Process Model	**37.5%**

TABLE V. PREDICTION RESULTS OF AUTO-REGRESSIVE MODEL.

Models	**MAPE**
OLS Regression	38%
Panel Regression	37.1%
Auto-regressive Model	**28.23%**

D. *Evaluation of Dynamic Spatio-Temporal Models*

Table VI displays the comparison between these models and the ground truth. As a result, the dynamic spatio-temporal model outperforms all other spatio-temporal prediction models and the traditional regression models referred in the ground truth.

TABLE VI. PREDICTION RESULTS OF DYNAMIC ST MODEL.

Models	**MAPE**
OLS Regression	38%
Panel Regression	37.1%
Dynamic Spatio-Temporal (ST) Model	**6.79%**

VI. Conclusion and Future Work

The proposed approach validated general and dynamic spatio-temporal models for crime prediction. The results showed that the relationships among this spatio-temporal data i) have a positive impact on the prediction accuracy and ii) can be utilized to analyze crime data with a high spatial and temporal granularity. The conducted experiments, however, are only a proof of concept for spatio-temporal predictions at lower spatial granularity.

Upstream spatio-temporal analysis improved the spatio-temporal predictions. The spatial analysis yielded that some locations have a spatial-relation with their neighbors. The temporal analysis confirmed positive correlations between consecutive year's crime incidents. Cross-correlation further identified social-economic factors having relations with crime.

Taking these spatio-temporal patterns into account, the proposed prediction approach outperformed traditional OLS and panel regression that ignores any spatio-temporal relationships in the data. In detail, the dynamic spatio-temporal Bayesian model lowered the error rate in prediction by 31.6% when compared with the ground truth. In contrast, Gaussian based prediction models and auto-regressive models only decreased the error rate by 1% respective 7% (compared with ground truth).

In the future, evaluations have to be extended with more complete data from all geographical hierarchy levels. In practice, however, these data are hard to obtain. Furthermore, more complex analysis models can be designed that accommodate a larger number of independent variables (social-economic factors) for predictions. Having more complete datasets, network models like Bayesian neural networks and LSTM can be evaluated. Especially LSTM archived high time series prediction accuracies when applied on large datasets.

Additionally, the proposed approach can be combined with social media analysis to create a hybrid prediction model that consider the prediction results from two different data sources (social-economic dataset and social media). This way, statistical data (i.e., social-economic factors) anonymously describing (large) groups of people are combined with concrete and precise information about individuals and thus, likely enhance prediction performance.

REFERENCES

[1] K. E. McCollister, M. T. French, and H. Fang, “The cost of crime to society: New crime-specific estimates for policy and program evaluation,” Drug and alcohol dependence, vol. 108, no. 1, 2010, pp. 98–109.

[2] W. C. Cunningham, J. J. Strauchs, and C. W. VanMeter, Private security trends, 1970 to 2000: The Hallcrest report II. Butterworth-Heinemann Boston, MA, 1990, vol. 2.

[3] X. Wang, M. S. Gerber, and D. E. Brown, “Automatic crime prediction using events extracted from twitter posts,” in International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction. Springer, 2012, pp. 231–238.

[4] M. S. Gerber, “Predicting crime using twitter and kernel density estimation,” Decision Support Systems, vol. 61, 2014, pp. 115–125.

[5] K. Edmark, “Unemployment and crime: Is there a connection?” The Scandinavian journal of economics, vol. 107, no. 2, 2005, pp. 353–373.

[6] A. Freytag, J. J. Krüger, D. Meierrieks, and F. Schneider, “The origins of terrorism: Cross-country estimates of socio-economic determinants of terrorism,” European Journal of Political Economy, vol. 27, 2011, pp. S5–S16.

[7] S. Chainey, L. Tompson, and S. Uhlig, “The utility of hotspot mapping for predicting spatial patterns of crime,” Security Journal, vol. 21, no. 1-2, 2008, pp. 4–28.

[8] H. Entorf and H. Spengler, “Socioeconomic and demographic factors of crime in germany: Evidence from panel data of the german states,” International review of law and economics, vol. 20, no. 1, 2000, pp. 75–106.

[9] R. Caruso and F. Schneider, “The socio-economic determinants of terrorism and political violence in western europe (1994–2007),” European Journal of Political Economy, vol. 27, 2011, pp. S37–S49.

[10] W. R. Tobler, “A computer movie simulating urban growth in the detroit region,” Economic geography, vol. 46, no. sup1, 1970, pp. 234–240.

[11] A. Beltran, C. Abargues, C. Granell, M. Núñez, L. Díaz, and J. Huerta, “A virtual globe tool for searching and visualizing geo-referenced media resources in social networks,” Multimedia tools and applications, vol. 64, no. 1, 2013, pp. 171–195.

[12] X. Chen, Y. Cho, and S. Y. Jang, “Crime prediction using twitter sentiment and weather,” in Systems and Information Engineering Design Symposium (SIEDS), 2015. IEEE, 2015, pp. 63–68.

[13] M. Cheong and V. C. Lee, “A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via twitter,” Information Systems Frontiers, vol. 13, no. 1, 2011, pp. 45–59.

[14] V. Martinez, G. I. Simari, A. Sliva, and V. S. Subrahmanian, “Convex: Similarity-based algorithms for forecasting group behavior,” IEEE Intelligent Systems, vol. 23, no. 4, July 2008, pp. 51–57.

[15] R. J. Sampson and W. B. Groves, “Community structure and crime: Testing social-disorganization theory,” American journal of sociology, vol. 94, no. 4, 1989, pp. 774–802.

[16] A. M. Olligschlaeger, “Artificial neural networks and crime mapping,” Crime mapping and crime prevention, 1997, pp. 313–348.

[17] J. Caulkins, J. Cohen, W. Gorr, and J. Wei, “Predicting criminal recidivism: A comparison of neural network models with statistical methods,” Journal of Criminal Justice, vol. 24, no. 3, 1996, pp. 227–240.

[18] S. W. Palocsay, P. Wang, and R. G. Brookshire, “Predicting criminal recidivism using neural networks,” Socio-Economic Planning Sciences, vol. 34, no. 4, 2000, pp. 271–284.

[19] 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, 2010, pp. 223–239.

[20] X. Wang and D. E. Brown, “The spatio-temporal generalized additive model for criminal incidents,” in Intelligence and Security Informatics (ISI), 2011 IEEE International Conference on. IEEE, 2011, pp. 42–47.

[21] C. Ivaha, H. Al-Madfai, G. Higgs, and J. A. Ware, “The dynamic spatial disaggregation approach: A spatio-temporal modelling of crime.” in World congress on engineering, 2007, pp. 961–966.

[22] M. Townsley, S. D. Johnson, and J. H. Ratcliffe, “Space time dynamics of insurgent activity in iraq,” Security Journal, vol. 21, no. 3, 2008, p. 139.

[23] M. Tiefelsdorf and B. Boots, “The exact distribution of moran’s i,” Environment and Planning A, vol. 27, no. 6, 1995, pp. 985–999.

[24] K. S. Bakar, S. K. Sahu et al., “sptimer: Spatio-temporal bayesian modelling using r,” Journal of Statistical Software, vol. 63, no. 15, 2015, pp. 1–32.

[25] J.-M. Marin, K. Mengersen, and C. P. Robert, “Bayesian modelling and inference on mixtures of distributions,” Handbook of statistics, vol. 25, 2005, pp. 459–507.

[26] C. Fernández and M. F. Steel, “On bayesian modeling of fat tails and skewness,” Journal of the American Statistical Association, vol. 93, no. 441, 1998, pp. 359–371.

[27] A. O. Finley, S. Banerjee, and A. E. Gelfand, “spBayes for large univariate and multivariate point-referenced spatio-temporal data models,” ArXiv e-prints, Oct. 2013.

[28] N. Cressie and H.-C. Huang, “Classes of nonseparable, spatio-temporal stationary covariance functions,” Journal of the American Statistical Association, vol. 94, no. 448, 1999, pp. 1330–1339.

[29] J. R. Stroud, P. Müller, and B. Sansó, “Dynamic models for spatiotemporal data,” Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 63, no. 4, 2001, pp. 673–689.

[30] A. Arab, M. B. Hooten, and C. K. Wikle, “Hierarchical spatial models,” in Encyclopedia of GIS. Springer, 2008, pp. 425–431.

[31] B. Huang, B. Wu, and M. Barry, “Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices,” International Journal of Geographical Information Science, vol. 24, no. 3, 2010, pp. 383–401.

[32] K. Bakar, “Bayesian analysis of daily maximum ozone levels,” Ph.D. dissertation, University of Southampton, June 2012. [Online]. Available: https://eprints.soton.ac.uk/340039/

[33] H. R. Moon and M. Weidner, “Dynamic linear panel regression models with interactive fixed effects,” Econometric Theory, vol. 33, no. 1, 2017, pp. 158–195.

[34] B. Craven and S. M. Islam, Ordinary least squares regression. Sage Publications, 2011.

[35] L. Anselin, “Local indicators of spatial associationlisa,” Geographical analysis, vol. 27, no. 2, 1995, pp. 93–115.