Assessing the Spatiotemporal Relation between Twitter Data and Violent Crime

Marco Stam[1], Charlotte Gerritsen[2], Ward van Breda[3], and Elias Krainski[4]

[1] Department of Economics, Leiden University, The Netherlands
[2] Netherlands Institute for the Study of Crime and Law Enforcement, The Netherlands
[3] Sentimentics bv, The Netherlands
[4] Departamento de Estatistica, Univresidade Federal do Paraná

Abstract. Social media have grown over the past decade to become an important factor in the way people share information. This paper explores the feasibility of using Twitter data for predictive policing models within the city of Amsterdam, the Netherlands. A novel, powerful approach to Bayesian inference is used to assess the correlations between the spatiotemporal distributions of tweets containing predefined keywords and that of violent crime incidents. Spatiotemporal log-Gaussian Cox processes are considered using the stochastic partial differential equations (SPDE) approach for space correlated over time with autoregressive dynamics and fitted using the integrated nested Laplace approximations (INLA) method. The findings show that the occurrence of such tweets raises the probability of incidents occurring nearby in space-time. This novel insight has unveiled the promising potential of Twitter-based predictive policing models within the city of Amsterdam.

1 Introduction

Social media usage has grown at an astonishing pace over the past decade, with users sharing information ranging from news stories to their personal mood states. Having been launched in 2006, Twitter is one of the most popular social media platforms. Its approximately 320 million monthly active users generate more than 500 million messages daily, providing a powerful insight into the public's sentiment and activities (https://about.twitter.com/company). As such, its usefulness as a data source for various stakeholders in fields that rely on human factors is drawing great attention. An advantage of Twitter is its (partially) open nature, with freely available application programming interfaces (APIs) and high amounts of publicly available messages, which makes it a very accessible source for data mining.

The main focus of this paper is the potential of Twitter for issues related to the field of criminology. The human factor in crime leads to the assumption that insights into people's activities, preoccupations and mood states can be useful for monitoring, predicting and preventing various types of crimes. The potential of social media such as Twitter has therefore not escaped the attention of researchers within the field of criminology, especially of those focused on 'predictive policing', which encompasses the usage of statistical analysis of data to anticipate and prevent future crime or to respond more adequately to it [16]. In general this enables a more effective and efficient allocation of the often limited resources of law enforcement agencies compared to traditional policing approaches. This is mainly done through analysis of the spatial or spatiotemporal distributions of crime. By assessing the risks of various types of crimes in certain geographical areas, law enforcement can allocate its resources accordingly. A limitation however of traditional predictive policing models is their dependency on historical crime data [29]. These models are therefore locally descriptive and lack the possibility of implementation in other geographical areas. Models based on data derived from social media sources can overcome this lack of portability, as they are not based on specific geographical attributes.

This paper aims to explore the potential of Twitter-based predictive policing models within the city of Amsterdam, the Netherlands. To this end, a novel approach to Bayesian inference is applied to approximate parameter values that describe the distributions of certain tweets and violent crime incidents in geographical space and time. This is done by fitting spatiotemporal log-Gaussian Cox processes onto the individual point processes, which is considered a good method to model the collected data and its spatial and temporal dependence [15]. The applied approach enables us to subsequently correlate the spatiotemporal distributions of the point processes with each other.

2 Related Work

The potential of Twitter as a data source has not gone unnoticed by the academic community. A sizeable body of research surrounding Twitter has developed in areas such as computer science, economics, sociology and criminology. Many of these studies focus on its predictive power, covering a broad range of future events. By monitoring the public's general mood state via Twitter it has been shown possible to predict the up- or downward trend in the closing values of the Dow Jones Industrial Average of the following day [5, 31].

Other research has shown the ability of Twitter-based models to accurately predict phenomena such as the development of an influenza epidemic, civil unrest and box office results [1, 2, 3, 27]. The accuracy of these models trumped that of the traditionally employed methods of prediction and monitoring, respectively those of the Centers for Disease Control and Prevention and the Hollywood Stock Exchange [1, 3]. A prime example of the important role that social media fulfills can be found in the "Arabic Spring" uprisings in the Middle East and North Africa, which were largely facilitated by the communicational means offered by social media platforms [9]. Analyzing the activity on Twitter showed that changes therein preceded certain incidents such as mass protests. The general implication of these researches is that Twitter provides a useful insight into phenomena that contain human factors.

The body of research surrounding Twitter within the Netherlands is considerably smaller. However, there is some work confirming the predictive potential of Dutch Twitter corpora as well. By analyzing the content of Dutch tweets it has been proven possible to predict the activities its senders were likely to engage in later on that day [30]. Analysis of the prevalence of party names and accompanying sentiment in Twitter messages enabled the prediction of the Dutch senate and parliament election results [18, 25]. This relatively simple approach resulted in a less accurate prediction than that of traditional polling companies.

These researches show that Twitter offers insight into phenomena related to varying academic fields. The most relevant related work for the current project however is that of Gerber [8]. This comprehensive research has assessed the feasibility of predicting 25 types of crime using Twitter. Coherent topics within the collected tweets were identified via automatic semantic analysis and subsequently used to train a predictive model. This approach was applied over the complete corpus of geotagged tweets sent from within the city of Chicago, Illinois (USA), which resulted in improved accuracy of the prediction of 19 types of crime compared to the traditional kernel density estimation approach. This includes crimes of the legal definitions of aggravated assault and battery within the state of Illinois, which correspond in part to the definition of violent crime incidents used in this research.

Although Gerber's research shows promising results for this study, they differ from each other in several ways. An important difference is that the proportion of tweets written in the Dutch language which are geotagged is lower than those written in English, with 2.9 versus 3.4 percent respectively [24]. Although this can be considered a fairly small difference, it is amplified by Chicago having a much larger population size than Amsterdam. There is also the potential that differences between the Dutch and English Twitter corpora may affect the results. Apart from the clear language difference there is the possibility that Dutch Twitter users deviate from their American counterparts in additional unforeseen ways as to the way they share information via Twitter. These differences emphasize the relevance of this research and the need to assess the potential of Twitter for predictive policing in the Netherlands.

3 Data Collection

Data have been collected over two separate time periods of 68 and 108 days, concerning the spatiotemporal distribution of two main variables within the city of Amsterdam. The first variable are tweets sent within those time periods, originating from within the city boundaries. The second main variable is violent crime incidents that have occurred over those periods within the city. Information concerning these incidents has been gathered by aggregating RSS-feeds, i.e., Rich Site Summary feeds, a collection of formats primarily designed for web syndication. Two publicly accessible web feeds were aggregated of the p2000 communications network of the emergency services within the region of Amsterdam-Amstelland. 'Incidents' are defined as "violent crime incidents in the public domain to which the police has responded". These violent crime incidents include physical assaults and assaults with the use of a deadly weapon, i.e., firearms or sharp objects. Street names are used over the p2000 communications network to indicate the locations of these incidents. The point location coordinates for the incidents were subsequently approximated as the center points of those streets by using the Google Maps API. As the region Amsterdam-Amstelland not only covers the city of Amsterdam but also adjacent towns, a subsection was made of incidents originating within Amsterdam itself. Of the 714 captured incidents, 682 were located within the city boundaries.

The tweets were gathered using the Sentimentics software, which is primarily designed for text-based sentiment analysis (http://sentimentics.com/). Its application in this current project was however limited to the usage of predefined keywords to filter tweets from the entire Dutch twitter corpus. To facilitate the analyses the keywords have been grouped into three 'tweet keyword clusters' (TKCs), which contain keywords respectively related to: "physical violence", "agitation/aggression" and “substance abuse”. As the tweets are extracted from the Dutch Twitter corpus, Dutch keywords were used. The three TKCs are composed to correspond to their identically numbered sub-hypotheses. As such, TKC_1 mainly contains keywords directly related to physical violence (I), such as “fight” (vechtpartij) and “fighting” (vechten). Keywords more indirectly related to physical violence are however also incorporated in this cluster. These include keywords not directly indicating physical assault, but rather various acts indicating public disorder that are often accompanied by acts of violence, e.g., “rioting” (rellen). Keywords referring to agitated or aggressive mood states (II) were incorporated in TKC_2, such as, “angry” (boos) and “enraged” (woedend). The third cluster (TKC_3) includes keywords related to substance abuse (III), e.g., “drunken” (dronken) and “intoxicated” (beschonken).

Only 2.9 percent of the collected tweets include geotags, resulting in a number of 5.960 geotagged tweets in dataset 1 and 6.744 geotagged tweets in dataset 2. Less than half of these tweets originated from within the municipality of Amsterdam, specifically 2.209 and 3.012 tweets for dataset 1 and 2 respectively. A large proportion of these tweets however did not originate from private individuals, but rather entities engaged in reporting about crime or police activity. Such tweets were deemed unsuitable and therefore excluded from the analysis. Data cleaning lead to a final count of 193 usable tweets over the first time period (dataset 1). The second time period over which data was collected resulted in a total count of 221 usable tweets (dataset 2).

Both the tweets as well as the incidents are spatiotemporal, random point processes that are considered to be aggregated due to a stochastic environmental heterogeneity. The most frequently applied models fur such data are Cox processes, which are ‘doubly stochastic’ as they are inhomogeneous Poisson processes with a random intensity surface ($\lambda(s)$) [15]. More specifically, the collected data is to be modeled as log-Gaussian Cox processes (LGCPs), i.e. Cox processes of which the logarithm of the intensity surface ($\log(\lambda(s))$) is a Gaussian process. One of the useful properties of LGCPs is that the intensity surface and the underlying Gaussian process can be predicted using Bayesian methods.

4 Analysis

The data preparation and analyses were performed in the R software environment for statistical computing, version 3.2.2 (R Development Core Team, 2015). To adequately analyze the spatiotemporal distributions of our data a model is needed which incorporates Tobler’s ‘first law of geography’ [26]: “Everything is related to everything else, but near things are more related than distant things". This quote is mostly aimed at spatial autocorrelation, but can be extended to correlation between multiple variables. For spatiotemporal analyses, the concept can be extended into a third, temporal dimension, i.e., that events near to each other in time are more related than events more distant in time. Accounting for these terms is possible using the R-INLA package for Bayesian computing with R (http://r-inla.org). This package contains a fairly new alternative for Bayesian inference to the computationally intensive Markov chain Monte Carlo (MCMC) method. This so-called integrated nested Laplace approximation (INLA) approach enables fast approximate inference for latent Gaussian models [17]. This is especially useful as MCMC methods become increasingly hard to run with the advances in data collection and modeling leading to increases in model size and complexity. Shortly summarized, the INLA approach is fast by taking advantage of the precision matrix sparsity when considering Gaussian approximations in the nested

Laplace approximations. The first Laplace approximation is used to find the parameter modal configuration. The second is considered for improving the latent field marginal distributions. In addition, efficient exploration around the mode is done when doing the numerical integration over the parameter space in order to account for uncertainty. Whereas MCMC has to explore the posterior distributions in a stochastic way, INLA does it in an analytical way [12, 14].

The R-INLA package was created to make the approach more accessible to applied researchers from various fields and contains a myriad of functions for various methods of Bayesian modeling. Particularly useful for spatial analyses are the 'stochastic partial difference equation' (SPDE) models. These SPDE-models provide an additional gain in computational efficiency when modeling spatial data with Gaussian fields in the Matérn class [13, 21, 22]. This is achieved by using SPDEs to link the continuously indexed Gaussian fields to the discretely indexed Gaussian Markov random fields. In doing so the continuous representation of space remains, while the computations are being performed on a discrete scale. Gaussian Markov random fields (GMRFs) have sparse precision matrices, which enable a drastic reduction in computational burden compared to the Gaussian fields dense matrices. As such, it also largely eliminates the often encountered 'big n problem' of Gaussian fields, which can quickly become computationally unfeasible. Removing this limitation has great potential for spatial statistics, wherein Gaussian fields play an essential role. The supported class of latent Gaussian models, which is a very broad range of models in itself, ranges from linear mixed models to spatiotemporal models [14]. The additional integration of the SPDE approach in R-INLA expands upon this further, enabling us to use R-INLA for point process models, including the spatiotemporal LGCPs applied in this research. Inference for log-Gaussian Cox processes can be performed via INLA due to the model assumption that there is a latent Gaussian Random Field [10, 20].

Bayesian inference is widely used in spatial statistics. Main factors to its popularity in this field are its ability to incorporate uncertainty and to deal with missing data. Despite its popularity, using this approach for statistical analysis with unknown parameter values is quite controversial [4]. This is mainly due to the significant effect the often arbitrarily specified prior distributions can have on the resulting posterior distributions. Accordingly, the INLA approach also approximates posterior marginal distributions for all of the model parameters. R-INLA however offers the ability to use Penalized Complexity priors, which largely negate the effect of weak prior information and subjectivity [23].

5 Model Properties

To answer our research questions, the incidents were modeled as a response variable and the tweets as covariates. At the time of this research there was no single comprehensive method for this in R as the data is misaligned, i.e., both point processes occur at different spatiotemporal locations. However, the R-INLA package offers a multistage solution due to its many general functions. One SPDE model can be considered for each of the tweets and one for the incidents, which is particularly useful for misaligned data [4, 11]. The joint model was built as a simplified coregionalization model where the tweets fields are treated as covariates for the incidents field. This approach for multivariate spatial point processes is based on Wackernagel's linear model of coregionalization [7, 19, 28]. We have considered the spatiotemporal model in [6] implemented in a way to allow complete misalignment of the tweets and incidents, following Section 3.2 in [12], and [11]. It only requires the data to be bounded by identically space-time observation windows [11]. The final model specification will be discussed in the next section.

6 Specification and Results

The first steps in using the SPDE approach to fit a spatiotemporal LGCP, are to specify a temporal and spatial mesh upon which the model is fitted. To accommodate the comparison of multiple point processes they must be fitted on the same meshes, bounded by one space-time observation window [11]. The applied model was fitted over both datasets separately using the same specifications. These datasets are however bounded by different temporal observation windows and therefore require individual temporal meshes. The aforementioned time knots were manually specified on the temporal meshes. Although generally limited by computational capacity, this computationally feasible maximum could not be reached due to the limited amount of tweets. To maintain a high enough number of tweets per time knot, the highest achievable amount of time knots was 12. These were applied equally spaced over the time window, resulting in relative counts of 1 time knot per 5.7 days for dataset 1 and 1 knot per 9 days for dataset 2.

The creation of the spatial mesh is done using constrained refined Delaunay triangulation by default. This method partitions the spatial area into triangles around mesh nodes in such a way that any location within it is closer to its own mesh node than to any other mesh node [11]. For geostatistical analysis it is recommended to define the mesh nodes on the event locations to increase efficiency and precision. This is however not necessary for LGCPs and would also defeat its purpose if it is defined on only one of multiple point processes that are to be fitted upon it. As all of the LGCPs need to be fitted on one spatial mesh, it was defined separately from the event locations. This is quite straightforward with R-INLA, although it requires some tuning to get the ideal specification. A finer mesh can be specified where it is preferable and conversely empty areas can be covered in a cruder mesh to save unnecessary computational costs. A shapefile of the municipality of Amsterdam was used to subset the tweets and incidents originating from within the city of Amsterdam. The spatial mesh was initially placed over the entire city, using the city borders as boundaries. The limited amount of tweets however required a reduction in spatial mesh nodes to accommodate raising the number of time knots to a level that adequately models the temporal variance. To accommodate this, a subsection of the area was made, which encompasses the vast majority of the tweets and incidents while greatly reducing the number of spatial mesh nodes. Combined with a cruder spatial mesh specification and the usage of convex hulls, the number of nodes was lowered from 948 to 62 nodes for dataset 1 and 65 nodes for dataset 2. This led to lower absolute counts of 182 tweets and 207 incidents in dataset 1 and 208 tweets and 379 incidents in dataset 2. The relative count however is higher with this specification, with 2.9 tweets and 3.3 incidents per mesh node for dataset 1 and 3.2 tweets and 5.8 incidents per node for dataset 2.

The initial model specification was to run four LGCPs separately for each of the four point processes and subsequently assess the correlations using a spatiotemporal coregionalization approach. However, to reduce the risk of oversmoothing, the temporal resolution needs to be high enough to catch enough variation. Raising the number of time knots increases the model complexity, which needs to be supported by a proportionate amount of data.

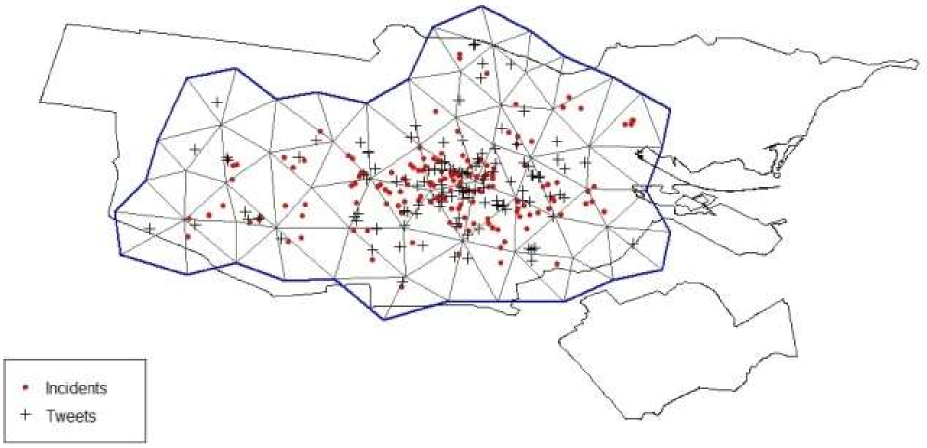

Figure 1. The final specification of the spatial mesh with 62 mesh nodes projected over the city of Amsterdam overlaid by the spatial distributions of the incidents and tweets of dataset 1.

The limited number of tweets resulted in the necessity to compensate this added complexity by reducing the number of model parameters. The coregionalization approach consists of assessing the correlation between each variate and the spatial and temporal correlation of each variate. The complexity of the algorithm for parameter estimation increases proportionately to the number of possible combinations of the model parameters, reducing the number of parameters was therefore determined to be the most effective method of model simplification. These simplifications were implemented in multiple steps. To answer our research questions the correlations between the TKCs were not of interest and therefore dropped from our model. A subsequent test run of this model specification showed that the three TKCs have identical spatiotemporal process parameters. Combined with the aforementioned assumption of conditional independence and absence of a likelihood hyper-parameter (due to the Poisson likelihood), this enabled modeling the TKCs jointly on one set of parameters. Although they now share one set of hyper-parameters, the latent field is replicated for each TKC independently leading to independent realizations for each TKC. This final simplification reduced the number of model parameters further to a final count of nine parameters. Combined with a reduced number of spatial mesh nodes, these simplifications facilitated the aforementioned final specification of temporal mesh ($k = 12$).

By jointly fitting the three TKCs, our final model specification consists of three calls to INLA in which the posteriors are approximated. All of these calls have been specified using the aforementioned Penalized Complexity priors, ensuring that the posterior probabilities are derived from the data alone [23]. The first two calls consist of fitting the LGCPs on the TKCs and the incidents separately. The computational time of these calls for dataset 1 were 46 and 16 seconds respectively on a Windows 8.1 (64-bit) platform with a quad-core 4.4 GHz Intel i7 processor with 16GB of RAM. The larger spatial mesh and sample size of dataset 2 is reflected in slightly longer computational times, with 52 and 25 seconds respectively. The resulting approximated posterior values of the individual temporal correlation parameters (ρ) are shown in Table 1.

Table 1. *Posterior estimates (mean, standard deviation and 95% credible interval) of the temporal correlation parameter (ρ) of the fitted log-Gaussian Cox processes per dataset*

	Dataset 1		Dataset 2	
	TKCs	Incidents	TKCs	Incidents
Mean	.985	.998	.987	.999
σ	.009	.003	.010	.002
$Q_{0.025}$	.964	.990	.963	.995
$Q_{0.975}$	.997	1	.998	1

These near perfect correlations show that the spatial intensity surfaces ($\lambda(s)$) of both LGCPs remain constant over time. Plots of the latent fields show that the spatial intensity for all four of the point processes is higher near the city center, as displayed in figure 3. The relative risk of events taking place therefore seems to be related to population density.

The approximated posterior values of all of the relevant model parameters of the latent field (θ) were used in the final call to INLA as initial values to speed up the approximations. On the aforementioned system, this resulted in a computational time of two hours and 49 minutes for dataset 1 and four hours and 55 minutes for dataset 2.

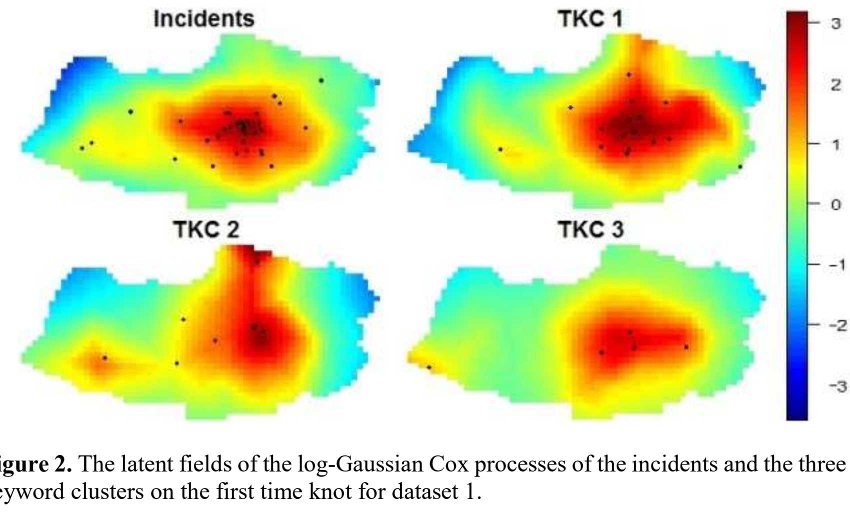

Figure 2. The latent fields of the log-Gaussian Cox processes of the incidents and the three tweet keyword clusters on the first time knot for dataset 1.

To answer our research questions the TKCs were specified as covariates to the incidents. The final model was specified as follows;

$$\eta TKC_m = \beta_0 TKC_m + \xi TKC_m$$

$$\eta Incidents = \beta_0 Incidents + \xi Incidents + \sum_{m=1}^{3} \beta TKC_m \xi TKC_m$$

Here η is the linear predictor, ξ the spatiotemporal field, β_0 the intercept and β the correlation coefficient of the respective TKC with the incidents, which indicate the proportion of the spatiotemporal field shared with that of the incidents. The resulting estimated posterior values of the βs are shown in table 2.

Table 2. *Posterior estimates (mean, standard deviation and 95% credible interval) of the correlation coefficients of the three tweet keyword clusters to the incidents and sample sizes*

	Dataset 1					Dataset 2				
	Mean	σ	$Q_{0.025}$	$Q_{0.975}$	n	Mean	σ	$Q_{0.025}$	$Q_{0.975}$	n
β TKC_1	.366	.106	.165	.581	110	.355	.110	.155	.584	130
β TKC_2	.276	.106	.058	.471	52	.022	.081	-.141	.176	63
β TKC_3	.299	.107	.088	.508	20	.438	.127	.217	.716	15

These results show a great consistency in the approximated values of β TKC_1 between the two datasets, with means of .366 and .355 for dataset 1 and 2 respectively. This extends to their respective 95% credible intervals of [.165, .581] and [.155, .584]. The consistent findings show that the spatiotemporal field of the first tweet keyword cluster correlates significantly with that of the incidents. These results support our first sub-hypothesis, proving that the spatiotemporal distribution of tweets containing keywords related to physical violence is correlated positively with the spatiotemporal distribution of violent crime incidents. From this it can be concluded, with respect to our central research question, that the spatiotemporal distribution of certain tweets correlates positively with the spatiotemporal distribution of violent crime incidents. The approximated values of β TKC_2 and β TKC_3 however show a great inconsistency. For β TKC_2 the mean value estimated from dataset 2 (.022) even lies outside of the 95% credible interval derived from dataset 1 [.058, .471]. The latter of which indicates a significant correlation, which is however contradicted by the 95% credible interval of dataset 2 [-.141, .176]. Although the approximated values for β TKC_3 are significant in both datasets, with 95% credible intervals of [.088, .508] and [.217, .716], they differ greatly from each other ($\Delta\bar{x}$=.139). The obtained results for β TKC_2 and β TKC_3 are too inconsistent to be considered reliable. With respect to the sub-questions whether the spatiotemporal distributions of tweets containing keywords related to agitation or aggression (II) and intoxication (III) are correlated with the spatiotemporal distribution of violent crime incidents, the findings are therefore considered to be inconclusive.

7 Discussion

The goal of this study was to explore the potential for predictive policing models based on Twitter data within the city of Amsterdam, the Netherlands. To this end a novel approach to Bayesian inference was used to assess the correlations between the spatiotemporal distributions of violent crime incidents and that of tweets containing predefined, hypothetically relevant keywords. This paper discussed the implementation of spatiotemporal log-Gaussian Cox processes with first-order autoregressive dynamics which have been jointly modeled and correlated using the SPDE approach offered by the R-INLA package. This approach enables using more complex models than traditional (MCMC) methods of Bayesian inference due to its computational efficiency. The R-INLA package offers an unprecedented level of flexibility and versatility, which is reflected in the multiple simplifications made to our model and the option to expand upon it further by adding terms specifying any type of random effect conceivable. To model the spatial and temporal dependency of the analyzed point processes, the applied statistical methods are considered an optimal approach. The excellent capability of dealing with uncertainty inherent to Bayesian inference makes it most suitable to account for the stochastic nature of the log-Gaussian Cox processes. Despite the small dataset this paper proves its feasibility for modeling the spatiotemporal relation between twitter messages and violent crime incidents, a topic previously unexplored within the Netherlands.

References

[1] Achrekar, H., Gandhe, A., Lazarus, R., Yu, S., and Liu, B. (2011). Predicting flu trends using twitter data. *Proceedings of the IEEE Conference on Computer Communications Workshops*, 702–707.

[2] Aramaki, E., Maskawa, S. and Morita, M. (2011). Twitter catches the flu: detecting influenza epidemics using Twitter. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 1568-1576.

[3] Asur, S. and Huberman, B.A., (2010). Predicting the future with social media. *Proceedings of the Web Intelligence and Intelligent Agent Technology Conference*, 492-499.

[4] Blangiardo, M. and Cameletti, M., (2015). Spatial and Spatio-temporal Bayesian Models with R-INLA. West Sussex: Wiley.

[5] Bollen, J., Mao, H. and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1–8.

[6] Cameletti, M., Lindgren, F., Simpson, D., and Rue, H., (2013). Advances in Stat.An. , vol. 97, issue 2, pp. 109-131.

[7] Gelfand, A.E., Schmidt, A.M. and Sirmans, C.F., (2002). Multivariate spatial process models: conditional and unconditional Bayesian approaches using coregionalization. Storrs, CT: Center for Real Estate and Urban Economic Studies, University of Connecticut.

[8] Gerber, M.S., (2014). Predicting crime using Twitter and kernel density estimation. *Decision Support Systems, 61*(1), 115-125.

[9] Howard, P.N. and Duffy, A., (2011). Opening closed regimes, what was the role of social media during the arab spring? *Project on Information Technology and Political Islam*, 1–30.

[10] Illian, J.B., Sørbye, S.H., and Rue, H., (2012). A toolbox for fitting complex spatial point process models using integrated nested laplace approximation (inla). *Annals of Applied Statistics, 6*(4),1499-1530.

[11] Krainski, E.T., Lindgren, F., Simpson, D. and Rue, H. (2015). The R-INLA tutorial: SPDE models (draft). Available from http://r-inla.org

[12] Lindgren, F. and Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software. 63*(19), 1–25.

[13] Lindgren, F., Rue, H. and Lindstrom, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The SPDE approach (with discussion). *Journal of the Royal Statistical Society (Series B), 73*(4),423–498.

[14] Martins, T.G., Simpson, D., Lindgren, F. and Rue, H. (2013). Bayesian computing with INLA: new features. *Computational Statistics & Data Analysis, 67*, 68-83.

[15] Møller, J., Syversveen, A.R. and Waagepetersen, R.P. (1998).Log Gaussian Cox processes. Scandinavian journal of statistics, 25(3), 451-482.

[16] Pearsal, B. (2010). Predictive Policing: The Future of Law Enforcement? *National Institute of Justice Journal, 266*, 15-19.

[17] Rue, H., Martino, S. and Chopin, N., (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society (Series B), 71*(2), 319-392.

[18] Sanders, E. and Bosch, A. van den (2013). Relating political party mentions on Twitter with polls and election results. *Proceedings of the 13th Dutch-Belgian Workshop on Information Retrieval.*

[19] Schmidt, A.M., and Gelfand, A.E., (2003). A Bayesian coregionalization approach for multivariate pollutant data. *Journal of Geophysical Research: Atmospheres, 108*(24).

[20] Simpson, D.P., Illian, J., Lindgren, F., Sørbye, S.H., and Rue, H. (2011). Going off grid: Computationally efficient inference for log-Gaussian Cox processes. *ArXiv e-prints.*

[21] Simpson, D.P., Lindgren, F. and Rue, H. (2012a). In order to make spatial statistics computationally feasible, we need to forget about the covariance function. *Environmetrics, 23*(1), 65-74.

[22] Simpson, D. P., Lindgren, F., and Rue, H., (2012b). Think continuous: Markovian Gaussian models in spatial statistics. *Spatial Statistics, 1*, 16-29.

[23] Simpson, D.P., Rue, H., Martins, T.G., Riebler, A., and Sørbye, S.H., (2014). Penalising model component complexity: A principled, practical approach to constructing priors. *ArXiv e-prints*.

[24] Sloan, L. and Morgan, J., (2015). Who Tweets with Their Location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on twitter. *PLoS ONE, 10*(11).

[25] Tjong Kim Sang, E. and Bos, J. (2012). Predicting the 2011 Dutch Senate Election Results with Twitter. *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 53–60.

[26] Tobler, W.R., (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 2*(46), 234-240.

[27] Van Noort, R. Kunneman, F, and Bosch, A. van den. (2016). Predicting civil unrest by categorizing Dutch Twitter events. In: *Proceedings of the 28th Benelux Conference on Artificial Intelligence, BNAIC 2016* (eds. T. Bosse and B. Bredeweg), Springer CCIS.

[28] Wackernagel, H. (1998). Multivariate geostatistics: An introduction with applications (2nd ed.). New York, NY: Springer-Verlag.

[29] Wang, X., Gerber, M.S. and Brown, D.E. (2012). Automatic crime prediction using events extracted from twitter posts. In N. Agarwal, K. Xu & N. Osgood (Eds.) *Social Computing, Behavioral-Cultural Modeling and Prediction* (pp. 231-238). New York, NY: Springer.

[30] Weerkamp, W. and Rijke, M. de (2012). Activity prediction: A twitter-based exploration. *SIGIR Workshop on Time-aware Information Access*.

[31] Zhang, X., Fuehres, H., and Gloor, P.A. (2011). Predicting Stock Market Indicators Through Twitter: I hope it is not as bad as I fear. *Social and Behavioral Sciences, 26*, 55–62.