GEOGRAPHIC PROFILING: BAYESIAN MODELS FOR RESIDENTS AND NON-RESIDENTS

Keywords: Bayesian data analysis; geographic profiling; distance decay function; anchor point; resident; non-resident.

ABSTRACT. We consider the problem of geographic profiling and offer an approach to choosing a suitable model for each offender. Based on the analysis of the examined dataset, we divide offenders into several types with similar behavior. According to the spatial distribution of the offender's crime sites, each new criminal is assigned to the corresponding group. Then we choose an appropriate model for the offender and using bayesian methods we determine the posterior distribution for the criminal's anchor point. Our models include directionality, similarly to models of Mohler and Short (2012). Our approach also provides a way to incorporate two possible situations into the model – when the criminal is resident or non-resident. We test this methodology on a data set of offenders from Baltimore County and compare the results with Rossmo's approach. Our approach leads to substantial improvement over Rossmo's method, especially in the presence of non-residents.

1. INTRODUCTION

The problem of geographic profiling aims to find the common location of a criminal (place of residence, workplace, favourite pub, etc.). Given the knowledge of places, where the criminal commited a series of crimes, we want to estimate the so called anchor point $\mathbf{z} = \left(z^{(1)}, z^{(2)}\right) \in \mathbb{R}^2$.

There are several approaches to locate the anchor point. One approach is based on *spatial distribution strategies*, which estimate directly the anchor point by various methods. Such techniques include *the centroid method*, *center of minimum distance* or *the circle method* (see Canter, 1996).

Another group of techniques, usually called probability distribution strategies, uses *hit score functions*. In order to construct such a function one has to choose a distance metric and a *distance decay function*, which distinguish individual methods in this group. The most popular ones include *Rossmo's model CGT*, *Canter's method* and *Levine's method* (see Canter, 1996; Canter et al., 2000; Levine, 2008; O'Leary, 2009a; Rossmo, 1995). The hit score function indicates a prioritised search area.

We can write the hit score function for all $\mathbf{y} \in \mathbb{R}^2$ in the form

$$S(\mathbf{y}) = \sum_{i=1}^{n} f\left(d\left(\mathbf{x}_i, \mathbf{y}\right)\right), \tag{1}$$

where $d(\mathbf{x}_i, \mathbf{y})$ denotes a distance metric between points $\mathbf{x}_i$ a $\mathbf{y}$, the function f is a distance decay function and $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ denote known crime sites corresponding to the given offender.

Formula (1) is subject to criticism, since it does not provide a probability density and does not include geographic features of the given region and other variables related to the criminal's behaviour (Mohler and Short, 2012). Several studies point out an important connection between a series of crimes and geography of the region (Brantingham and Brantingham, 1993; Canter et al., 2000; Rossmo, 2000). This is one reason why an appropriate tool to treat the problem is provided by bayesian methods (see Levine, 2008; O'Leary, 2009a, 2010a; Mohler and Short, 2012). They allow to implement also information available before analyzing data. Moreover, as a result we obtain a posterior, which corresponds to a true probability distribution.

The fundamental problem is to choose a suitable model, which would best characterize the searched criminal. *O'Leary* proposes several options in O'Leary (2009a) or O'Leary (2009b). So far there exists no universal model which would well describe behaviour of an arbitrary criminal. On the other hand, *Mohler* and *Short* offer a more general model, which for a suitable choice of parameters could cover a broad range of criminals (Mohler and Short, 2012). The question remains which one to choose and how to choose the parameters.

It is obvious that every criminal has his own particularities. However, when we are at the stage of investigation, it is difficult to specify and determine what his style of behaviour is, hence his considerations when choosing the crime location. This paper offers an approach to deal with this problem.

In Section 2, we introduce bayesian approach to geographic profiling and show how we can incorporate required parameters into the model. Section 3 describes the data set that we use, explains necessity to distinguish between coordinate systems and outlines conversion of geographic coordinates recorded in the dataset into to the plane coordinates, for the possibility of using the euclidan metric in our calculation. In Section 4, we deal with the different types of offenders in our data set, explain differences between resident and non-resident type of criminals and offer various models that are suitable for each type or subtype of offenders in our data set. Moreover, we suggest a method how to divide our offenders into mentioned categories of types and subtypes based on the spatial distribution of their crime sites. Section 5 describes how we can estimate prior distribution for used parameters. To obtain this, we use kernel smoothing or logspline density estimation. In Section 6, we consider several cases of modelling with our data set. Then we illustrate the effectiveness of our methodologies in comparison with Rossmo's approach.

2. Bayesian methods in geografic profiling

The choice of a model and suitable parameters for the offenders behaviour is one of the key parts of analysis. The probability function (or density) p captures our knowledge about the given offender. In other words, this function suggests how the offender chooses the crime site. It represents our uncertainty and lack of information. From this point of view it is possible to use bayesian approach (unlike the frequentist aproach, which could be only used if the offender acts randomly).

Geographic profiling assumes, that the crime site selection is influenced by the offender's anchor point[1] $\mathbf{z}$, hence p will depend on $\mathbf{z}$. Denote $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_k\}$ the vector of k additional parameters, which also influence the offender's behaviour, therefore p. It may include the average distance, which the offender is willing to travel to the crime site, or a direction preferred by the offender (e.g. related to the transport infrastructure of the area), size of the buffer zone [2], etc. The choice of such parameters depends on the knowledge and experience of the analyst, and available information.

If we denote by $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ the known sites of a series of crimes, we can describe a model of the offender's behaviour by a function $p\left(\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\} | \mathbf{z}, \boldsymbol{\theta}\right)$. It expresses the probability that the offender with a unique anchor point $\mathbf{z}$ and with given values of parameters $\boldsymbol{\theta}$ commits crimes at the locations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$.

We would like to find the best way to estimate the anchor point $\mathbf{z}$. One possible approach is to use the maximum likelihood method. For criminalistic purposes, it is not very convenient, since it provides a single point estimate. In geographic

[1] For this approach, it is important that the offender has a single anchor point that is stable during the crime series.

[2] Buffer zone is an area around the offender's anchor point where he does not commit crimes because of his conspicuousness.

profiling, we would prefer to obtain an area, which contains the anchor point with high probability. This can be done using bayesian analysis (Bolstad, 2007; Carlin, 2000; Damien et al., 2013; Robert, 2007; Sivia, 2006).

Using Bayes rule we obtain

$$p(\mathbf{z},\boldsymbol{\theta}|\{\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_n\}) = \frac{p(\{\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_n\}|\mathbf{z},\boldsymbol{\theta})\cdot p(\mathbf{z},\boldsymbol{\theta})}{p(\{\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_n\})}. \tag{2}$$

Let us recall that $p(\mathbf{z},\boldsymbol{\theta}|\{\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_n\})$ denotes a posterior distribution, $p(\mathbf{z},\boldsymbol{\theta})$ is a prior distribution, $p(\{\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_n\}|\mathbf{z},\boldsymbol{\theta})$ denotes a likelihood function and the denominator $p(\{\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_n\})$ is called evidence. For our purpose, it is just a normalization constant. We can thus omit it from (2) while replacing equality ($=$) by proportionality ($\propto$).

Since we are only interested in the probability distribution for $\mathbf{z}$, we can get rid of the nuisance parameters by integrating over all possible values of $\boldsymbol{\theta}$. If we assume in addition independence of the anchor point $\mathbf{z}$ and further parameters $\boldsymbol{\theta}$ and mutual independence of the crime sites of the offender, we obtain

$$p(\mathbf{z}|\{\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_n\}) \propto \int_{M_{\boldsymbol{\theta}}}\cdots\int p(\mathbf{x}_1|\mathbf{z},\boldsymbol{\theta})\cdot\ldots\cdot p(\mathbf{x}_n|\mathbf{z},\boldsymbol{\theta})\cdot h(\mathbf{z})\cdot g(\boldsymbol{\theta})\,\mathrm{d}\theta_1\ldots\mathrm{d}\theta_k, \tag{3}$$

where $M_{\boldsymbol{\theta}} \subseteq \mathbb{R}^k$ denotes the domain of integration, $h(\mathbf{z})$ is a prior corresponding to the anchor point $\mathbf{z}$ and $g(\boldsymbol{\theta})$ is a prior distribution corresponding to the other parameters $\boldsymbol{\theta}$.

There are numerous possibilities for choosing $p(\mathbf{x}_i|\mathbf{z},\boldsymbol{\theta})$, or $p(\{\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_n\}|\mathbf{z},\boldsymbol{\theta})$ in modelling the offender's behavior. Also, we need to consider the most reasonable choice of the prior and which parameters enter the prior. All this depends on available information, data, geographic area under consideration and other factors. Hence we will next consider in more detail the data set which we use in this study. After careful analysis of this data set, we will choose appropriate models and tools for analysis.

3. THE DATA SET

In this study we have used freely available data[3] about 88 serial criminals, who committed crimes in Baltimore County in the time period 1993-1997. Most offenders in the dataset committed more crime types. These included forty-five criminals who predominantly committed larceny, twelve who predominantly committed assaults, ten who committed mostly burglaries and the same number of offenders who committed mainly vehicle theft. There were nine criminals who were mostly robbers and remaining two offenders cannot be categorized.

The number of crimes per criminal varied from three to thirty-three. In total, the data set contains 962 crimes. Each crime includes information about the identification number of the crime, identifier of the offender, UCR code [4], latitude and longitude of the crime site and latitude and longitude of the anchor point.

[3]Data can be found at `http://www.icpsr.umich.edu/CrimeStat/download.html`.

[4]UCR stands for *Uniform Crime Reporting* – the uniform coding system for indicating the type of the crime (see City of Tucson, nd; FBI, 2012).

For the calculation of the distance and choice of the distance metric, we have to distinguish between geographic coordinate system and projected coordinate system. The use of the euclidean metric can lead to easier calculation – some functions can be expressed analytically. Additionally, all crime sites and anchor points of all offenders from the data set are located only in an area of 72,5 km $\times$ 48,7 km. Distances between a resident's crime site and his anchor point are mostly a few kilometers. Hence, the choice of a coordinate system, or the choice of a distance metric should not lead to significantly different results.

Coordinates in the data set are recorded in the geographic coordinate system, thus, for using the euclidean metric we have to convert them into the plane coordinates. We use the most common projected coordinate system UTM (*Universal Transverse Mercator coordinate system*) that divides the Earth between 84° S and 80° N latitude into 60 zones, each 6° of longitude in width (see Kennedy, 2000). To use Euclidean metric all investigated points have to lie in the same zone. Our data set is compliant for using this system because all points are located in the zone 18. To transfer coordinates, we use a simplified version of relations which *Johann Heinrich Louis Krüger* derived in 1912 (see Kawase, 2011, 2012).

4. Procedures and models

When analyzing the data set, we will generally assume as known the fact that the crime series was commited by the same offender. This assumption is realistic in practice, since investigators may, according to the way the crime was commited, DNA analysis or other signs and evidence tell that the crimes are related to one person, just do not know which one and where to find him.

Different types of offenders require different models, hence we need to analyze character of offenders in the given data set, and the way they choose the crime site.

4.1. **Models for residents.** First we will consider resident offenders[5]. They commit crimes near their anchor point. Hence their criminal and anchor regions overlap, at least to a large extent. (see Fig. 1).

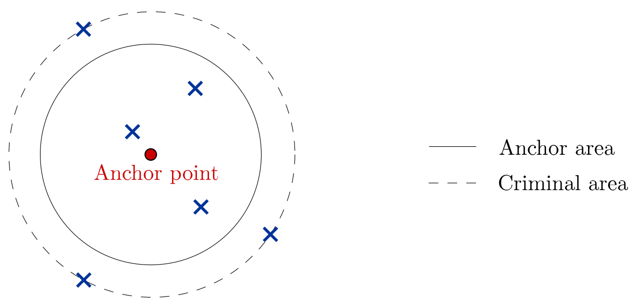


FIGURE 1. *Anchor and criminal area of the resident offenders - the dashed line circle denotes the criminal area, the solid line circle is the anchor area, blue crosses indicate crime sites, the red circle is the anchor point of the offender.*

[5]Our division of offenders into two groups (residents and non-residents) is similar to the marauder and the commuter hypothesis of Canter (1996). But in our case, the radius of the anchor area is especially based on the distances of the crime sites from the anchor points of the observed offenders. The criminal area can be imagined as the smallest circle that contains all crime sites of the given offender.

After analyzing the data set, we can classify offenders further into two subtypes, each of them requiring a different model.

(A) *Without a buffer zone.* (B) *With a buffer zone.*

FIGURE 2. *Two basic subtypes of residents (red triangle indicates the crime site, the black circle denotes the anchor point).*

The two basic subtypes of residents (Fig. 2) are characterized by the existence of a buffer zone. In both cases the distribution of crime sites is depicted on an approximately same area 25 km $\times$ 15 km. However, in the first case the offender commits crime in an immediate vicinity of the anchor point, having no buffer zone (see Fig. 2a). In the second case, the offender commits crimes several kilometres from the anchor point, hence there is some buffer zone (see Fig. 2b).

It was generally observed that if a resident has a small distance between crime sites (up to 2 km), the offender does not consider any buffer zone, and the behaviour is similar to the offender from Fig. 2a. Conversely, if the crime sites have larger distances and the sites are irragularly spaced, the offender's behaviour is similar to that of Fig. 2b and we have to take some buffer zone into account

We can also observe residents whose crime sites create clusters (see Fig. 3), where the distances within some clusters are small, but larger within others.

(A) *Without a significant buffer zone.* (B) *With a buffer zone.*

FIGURE 3. *Residents whose crime sites create clusters (The red triangle indicates the crime site, the black circle denotes the anchor point)).*

Some residents with crime sites in clusters have similar bahavior as in Fig. 3a. They commit some crimes close to their anchor points, but other crimes in larger distances. Distant crime sites can, but do not have to, create clusters. On the other hand, some offenders (see Fig. 3b) create clusters although their anchor point does not occur in any of them. Such criminals only prefer some locations. Therefore, if the crime sites of an offender create clusters or a cluster with other isolated and distant crime sites, the offender can, but does not have to, create a buffer zone around himself. In modelling we have to take this fact into account.

Undoubtedly, each criminal has an individuality with a unique behavior. Thus, we could find various specifics for all of them. However, the detailed examination of our data set shows that each offender significantly tends to one of these subtypes. According to the space distribution of the criminal's crime sites we can decide which subtype is the most suitable for the investigated offender. But in the case of clusters we are not able to decide on the existence of the criminal's buffer zone only on the basis of the space distribution of his crime sites.

For residents modelling we will use various modifications of the normal distribution. Compared to the use of the exponential distribution, the results are not significantly different. We use euclidean metrics for all models because it leads to an analytical expression for the models.

4.1.1. *Residents without buffer zone and without clusters (subtype M1).* For offenders, whose behaviour is similar to Fig. 2a, we will use a model suggested in O'Leary (2009a). Using Euclidean metric we obtain

$$p\left(\mathbf{x}_i|\mathbf{z},\alpha\right)=\frac{1}{4\alpha^2}\cdot\exp\left(-\frac{\pi}{4\alpha^2}\left[\left(x_i^{(1)}-z^{(1)}\right)^2+\left(x_i^{(2)}-z^{(2)}\right)^2\right]\right), \tag{4}$$

which corresponds to two dimensional normal distribution with mean at the anchor point and standard deviation $\sigma=\sqrt{\frac{2}{\pi}}\,\alpha$. The reason for this choice of σ can be found in O'Leary (2009b). The parameter α denotes average distance that the offender is willing to travel to commit a crime.

The most probable is committing crime directly at the anchor point or its immediate vicinity, which is exactly the behavior we expect for this type of offenders. Let us just remark that committing crime at the anchor point may seems strange, but is not at all impossible. In fact in our data set this situation occurs quite often, since the anchor point is not necessarily the offender's place of residence, but his favourite bar, workplace, etc.

4.1.2. *Residents with buffer zone and without clusters (subtype M2).* For an offender as in Fig. 2b, the highest probability will not be at the anchor point, but some distance α away from the anchor point, where α denote the average distance to the crime site. This is captured by a probability distribution of the form

$$p\left(\mathbf{x}_i|\mathbf{z},\alpha,\sigma\right)=\frac{1}{N\left(\alpha,\sigma\right)}\cdot\exp\left(-\frac{1}{2\sigma^2}\left[\sqrt{\left(x_i^{(1)}-z^{(1)}\right)^2+\left(x_i^{(2)}-z^{(2)}\right)^2}-\alpha\right]^2\right), \tag{5}$$

where the choice of σ determines the decay of the probability as the distance varies from α.

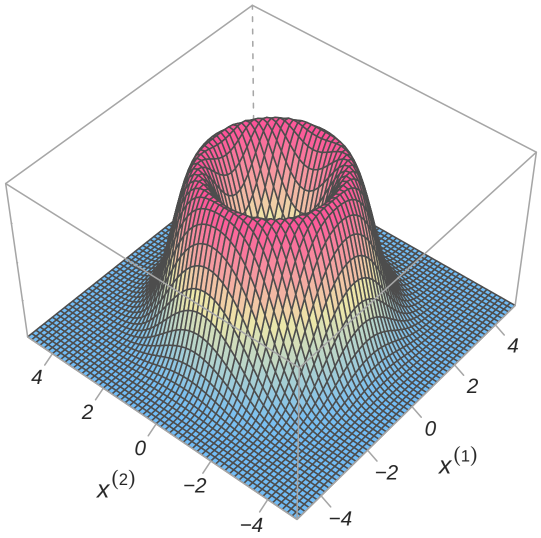
In order to obtain a probability distribution, we must have

$$\iint\limits_{\mathbb{R}^2}\frac{1}{N\left(\alpha,\sigma\right)}\cdot\exp\left(-\frac{1}{2\sigma^2}\left[\sqrt{\left(x_i^{(1)}-z^{(1)}\right)^2+\left(x_i^{(2)}-z^{(2)}\right)^2}-\alpha\right]^2\right)\mathrm{d}x_i^{(1)}\mathrm{d}x_i^{(2)}=1, \tag{6}$$

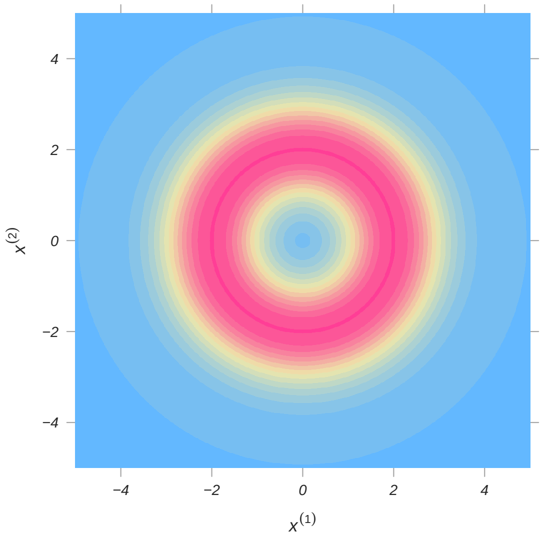
A simple calculation gives the normalizing factor

$$N\left(\alpha,\sigma\right)=2\pi\sigma^2\cdot\exp\left(-\frac{\alpha^2}{2\sigma^2}\right)+2\pi\sqrt{2\pi}\alpha\sigma\left(1-\Phi\left(-\frac{\alpha}{\sigma}\right)\right), \tag{7}$$

where Φ is the distribution function of the standard normal distribution.



(A) *Three-dimensional plot.*



(B) *Level plot.*

FIGURE 4. *Model for residents given by* (5) *with the anchor point* $\mathbf{z} = [0,0]$, $\alpha = 2$ *and* $\sigma = 0{,}8$.

Fig. 4 illustrates a model for this type of residents. The anchor point is again the origin, $\alpha = 2$ a $\sigma = 0{,}8$. For different values of α a σ, the probability of committing crime at the anchor point will be different, but nonzero, although for certain values very close to zero. The buffer zone does not have to be interpreted as an area where the criminal commits no crimes, but as an area around the anchor point, with lower probability of commiting crimes. This interpretation corresponds better to reality, where we can hardly claim with certainty that there will be no crime in the buffer zone.

Remark. *As defined, these offenders are called residents because their anchor and criminal areas overlap. However, they commit crimes at larger distances like non-residents (see Subsection 4.2). If we choose an appropriate priors for angle and distance, we can use the model of non-residents (see* (10)*) for this subtype of residents.*

4.1.3. *Residents with clusters (subtype M3).* For offenders, whose crime sites create clusters, as in Fig. 3, we are not able to decide whether their anchor point lies away from the clusters (they have a buffer zone), or whether it is in one of the clusters, and which one it is.

In this situation we will use multimodel inference (see Burnham et al., 2002; O'Leary, 2010b), where we obtain the resulting estimate as a weighted average of the estimates of anchor points in individual models, for the likelihood, or prior distribution. Let us assume that for estimating the anchor point we work with R models. We obtain as a result R posteriors, namely $p_i(\mathbf{z}|\{\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_n\})$ for $i = 1,2,\dots,R$. The resulting estimate for multimodel inference is then

$$p(\mathbf{z}|\{\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_n\}) = \sum_{i=1}^{R} w_i \cdot p_i(\mathbf{z}|\{\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_n\}), \tag{8}$$

where $w_i \geqslant 0$ denotes the weight corresponding to the i-th estimate, and the weights satisfy $\sum_{i=1}^{R} w_i = 1$. It guarantees that the resulting estimate also provides a probability distribution. For residents with $R-1$ clusters we will construct one model for each cluster and another model for the situation that the offender has a buffer zone.

The choice of weights depends on the analyst. If we have no reason to prefer one of the models, we simply set $w_i = \frac{1}{R}$ for all i. We can also use previous results and with offenders of a similar type find out which of the models better described the behavior of the offender. According to the relativities we can then assign weights to individual models. It is also possible that some evidence during the investigation prefers one of the models. In this case we set the weights based on preferences of an experienced investigator.

4.2. **Models for non-residents.** Non-resident is an offender who commits crimes relatively far from his anchor point (see Fig. 5).

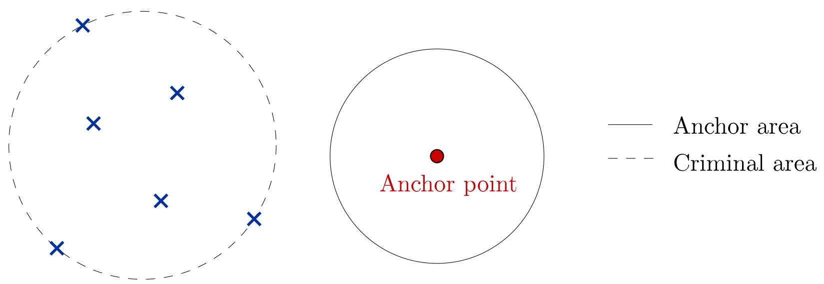


FIGURE 5. *Anchor and criminal area of the non-resident offenders - the dashed line circle denotes the criminal area, the solid line circle is the anchor area, blue crosses indicate crime sites, the red circle is the anchor point of the offender.*

The criminal and anchor areas essentially do not overlap in this case. However, it does not mean that this area is unknown for the criminal. It is only significantly away from the territory in which he normally lives, works and acts as a "non-offender".

In our dataset, non-residents typically commit crimes at a distance greater than 10 km, but it is often at least twice that distance. Another specific feature for this type of offenders is that they prefer a certain angle (measured from the horizontal axis with the origin at the anchor point) for the choice of their crime location.

In our dataset, there is just a small number of non-residents and we did not observe any markedly different behavior among them. Therefore we do not divide this type of offenders into other subtypes as in the case of residents in the previous subsection.

In Mohler and Short (2012), there is used a kinetic model that is derived on the basis of the following stochastic differential equation

$$\mathrm{d}\mathbf{x}_t = \boldsymbol{\mu}(\mathbf{x}_t) + \sqrt{2D}\,\mathrm{d}\mathbf{W}_t\,, \tag{9}$$

where $\mathbf{x}(t)$ denotes the position of the offender at time t, $\mathbf{W}_t$ is two-dimensional standard Wiener process, D denotes the diffusion parameter and $\boldsymbol{\mu}$ is the drift. The drift term could be used to describe more complex criminal behavior.

For different choices of the parameter values, the model is suitable for different type of offenders. If we want to obtain the most realistic model which is applicable to non-residents, we have to use numerical methods. However, the solution can be approximated by the product of a function of the distance and a function of the angle (again measured from the horizontal axis with the origin at the anchor point). This fact supports the idea that the criminal behavior could be generally modeled as the product of a suitable function which influences the most likely distance from the offender's anchor point and another function that affects probability of the angle preferred by the criminal.

Again, the model is based on the normal distribution, and is given by

$$p\left(\mathbf{x}_i|\mathbf{z},\alpha,\vartheta,\sigma_1,\sigma_2\right)=\frac{1}{N\left(\alpha,\vartheta,\sigma_1,\sigma_2\right)}\cdot q_1\left(\mathbf{x}_i|\mathbf{z},\alpha,\sigma_1\right)\cdot q_2\left(\mathbf{x}_i|\mathbf{z},\vartheta,\sigma_2\right), \tag{10}$$

where

$$q_1\left(\mathbf{x}_i|\mathbf{z},\alpha,\sigma_1\right)=\exp\left(-\frac{1}{2\sigma_1^2}\left[\sqrt{\left(x_i^{(1)}-z^{(1)}\right)^2+\left(x_i^{(2)}-z^{(2)}\right)^2}-\alpha\right]^2\right) \tag{11}$$

and

$$q_2\left(\mathbf{x}_i|\mathbf{z},\vartheta,\sigma_2\right)=\exp\left(-\frac{1}{2\sigma_2^2}\left[\arg\left(\left(x_i^{(1)}-z^{(1)}\right)+\mathrm{i}\left(x_i^{(2)}-z^{(2)}\right)\right)-\vartheta\right]^2\right), \tag{12}$$

where α is the average distance of the offenses, σ_1 denotes the standard deviation corresponding to the function q_1, ϑ is the average angle from the anchor point to the crime locations measured from the horizontal axis with the origin at the anchor point, and σ_2 denotes the standard deviation corresponding to the function q_2 (see Fig. 6).

The use of the function q_1 is analogous to the model (5) for residents with buffer zone, and therefore it can be interpreted in the same way. The function q_2 achieves the highest values at the angle of ϑ and its functional values around this angle decrease at a rate that depends on the choice of the value of σ_2. This function itself cannot be normalized since the double integral over all its possible values is infinite. But the product of q_1 and q_2 in (10) can already be normalized.

Let us note that the normalization factor has the form

$$N\left(\alpha,\vartheta,\sigma_1,\sigma_2\right)=N_1\left(\alpha,\sigma_1\right)\cdot N_2\left(\vartheta,\sigma_2\right), \tag{13}$$

where

$$N_1\left(\alpha,\sigma_1\right)=\sigma_1^2\cdot\exp\left(-\frac{\alpha^2}{2\sigma_1^2}\right)+\sqrt{2\pi}\alpha\sigma_1\left(1-\Phi\left(-\frac{\alpha}{\sigma_1}\right)\right) \tag{14}$$

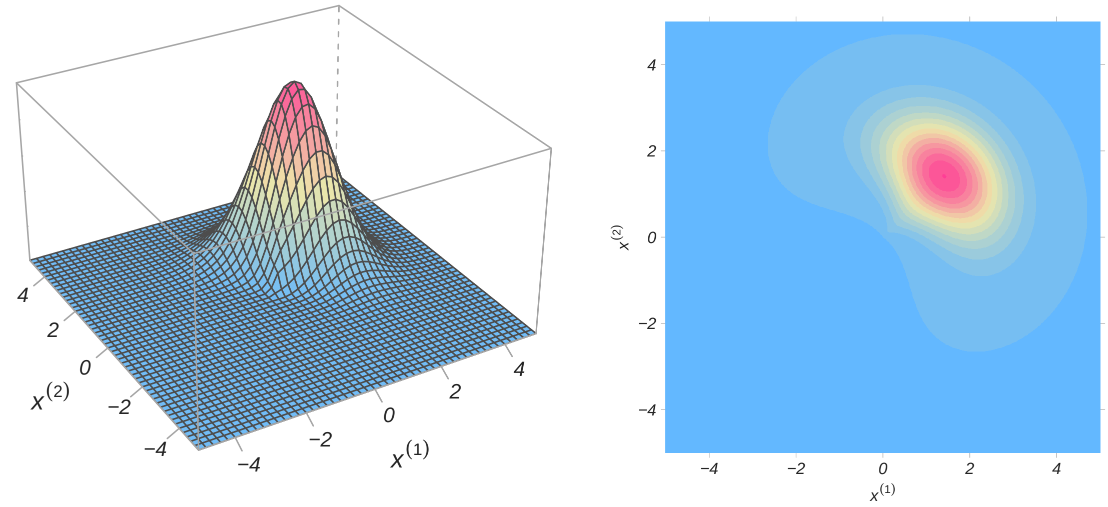
and

$$N_2\left(\vartheta,\sigma_2\right)=\sigma_2\sqrt{2\pi}\left(\Phi\left(\frac{2\pi-\vartheta}{\sigma_2}\right)-\Phi\left(-\frac{\vartheta}{\sigma_2}\right)\right), \tag{15}$$

where Φ represents the distribution function of the standard normal distribution. We obtain this value of the normalization factor from the requirement that the function p in (10) has to be a probability density, hence

$$\iint\limits_{\mathbb{R}^2}\frac{1}{N\left(\alpha,\vartheta,\sigma_1,\sigma_2\right)}\cdot q_1\left(\mathbf{x}_i|\mathbf{z},\alpha,\sigma_1\right)\cdot q_2\left(\mathbf{x}_i|\mathbf{z},\vartheta,\sigma_2\right)\mathrm{d}x_i^{(1)}\mathrm{d}x_i^{(2)}=1\,. \tag{16}$$

If we know in advance whether the offender is a resident or a non-resident, we choose an appropriate function as proposed above, to model his behavior. For



(A) *Three-dimensional plot.* (B) *Level plot.*

FIGURE 6. *The function given by* (10) *with the anchor point* $\mathbf{z} = [0, 0]$, $\alpha = 2$, $\sigma_1 = \frac{4}{5}$,$\vartheta = \frac{\pi}{4}$ *and* $\sigma_2 = \frac{\pi}{6}$.

residents, we have to decide on the offender's subtype based on the distribution of the criminal's crime locations.

If we are not able to determine in advance, whether the offender is a resident or a non-resident, we can use multimodel inference again, as discussed in Subsection 4.1.

5. THE CHOICE OF A PRIOR DISTRIBUTION

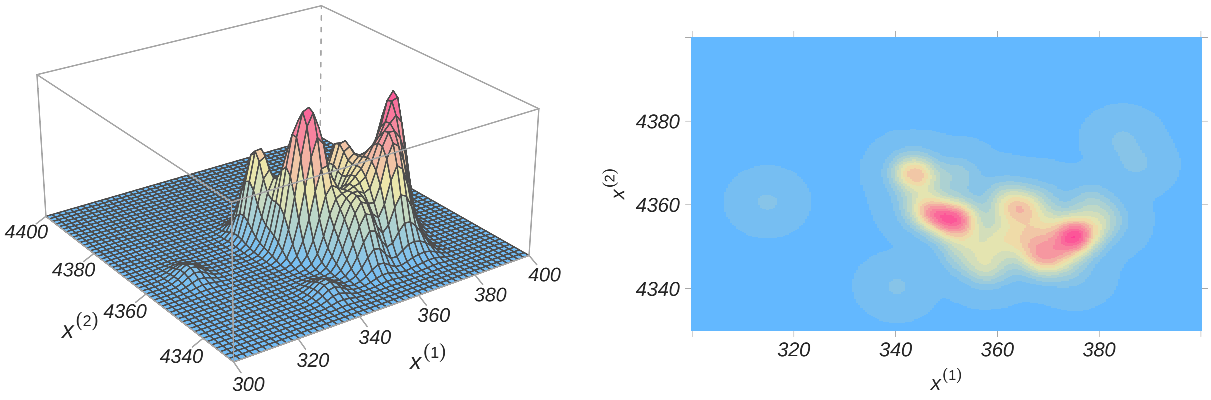
When processing the dataset, we work with offenders of different types and subtypes. Depending on a specific offender, we need to select appropriate priors for the parameters. In this study, we used kernel smoothing, and logspline density estimation (see for example Kooperberg and Stone, 1991), which allows to limit the range of values that the parameter can take. We always assume that all information about each offender contained in the data set is known to us. We only exclude knowledge about the examined offender [6].

Picture 7 graphically represents our prior for the anchor point. To obtain it, we used kernel smoothing, based on the anchor points of known offenders.

Next, we need to know the prior for average distance α to the offence. Since the distance cannot take negative values we use logspline density estimation with lower limit equal to zero. In this case, however, we have to use solely the data corresponding to the particular types of offenders. This is because the distance is just one of the main factors that the types and subtypes of offenders differ from one another. The function $g_1(\alpha)$ corresponds to the subtype M1, the $g_2(\alpha)$ to the subtype M2 and the function $g_3(\alpha)$ is obtained from the known average distance for non-residents and it will be used for this type of offenders. These functions are illustrated in Fig. 8.

If we use the model given by (10), we have to know the prior distribution for the angle ϑ. In Figure 9 we plot the estimated distribution of ϑ for offenders of subtype M2 and for non-residents. We can see that offenders of subtype M2 prefer

[6]Although all information in the dataset is known to us, in the following paragraphs we will use the word "known" for the data of all offenders except the investigated criminal.



(A) *Three-dimensional plot.* (B) *Level plot.*

FIGURE 7. *Graphical representation of the prior $h(\mathbf{z})$ for the anchor point.*

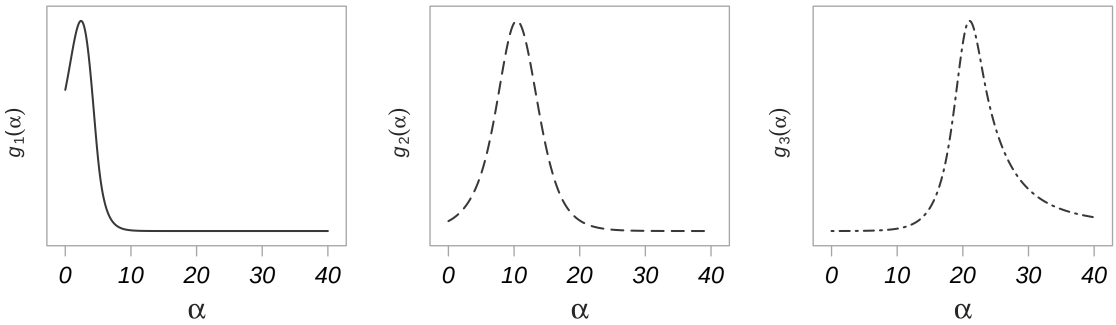


FIGURE 8. *Graphical representation of the prior $g_1(\alpha)$ (solid line), $g_2(\alpha)$ (dashed line) and $g_3(\alpha)$ (dash-dotted line) for the average distance α, that the offender is willing to travel to commit a crime.*

angle between $\frac{\pi}{2}$ and π, non-residents favour more directions –– the south-west direction (angle between π and $\frac{3}{2}\pi$) and east direction (angle around 0, or 2π). Significant preference of some directions can be caused by the existence of major transport networks for these angles.

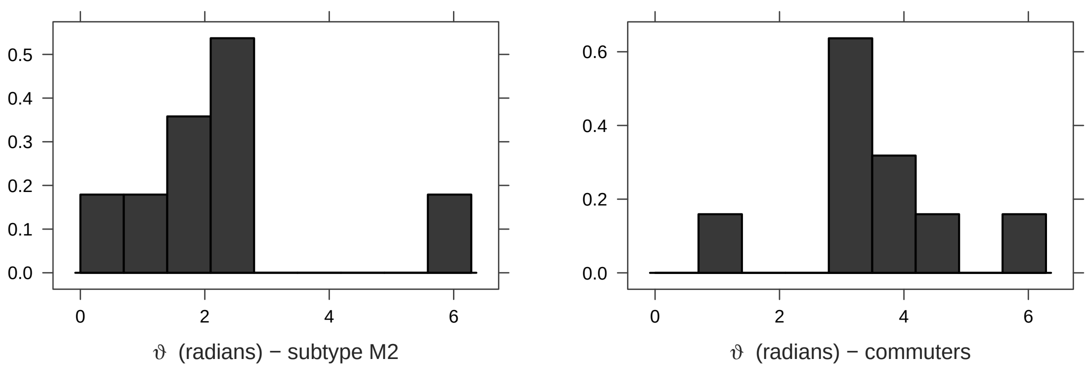


FIGURE 9. *Histograms of the average angle ϑ (measured from the horizontal axis with the origin at the anchor point), in which the offender commits crimes - for residents M2 and for non-residents.*

Similarly we obtain prior distribution for other required parameters.

6. MODELLING AND EVALUATION

In Section 4 we considered four types of offenders (residents M1, M2 and M3 and non-residents). For modelling of each of these types we can choose some of three models given by (4), (5) and (10), or use multimodel inference, a combination of some of them.

Based on this, we will separate the modelling with our dataset into the following cases:

(1) We choose only residents from the dataset.
  (a) For modelling of residents M1 we use the model (4), for residents M2 the model given by (5), for residents M3 we use multimodel inference, a combination of (4) and (5).
  (b) For modelling of residents M1 we use the model (4), for residents M2 the model given by (10), for residents M3 we use multimodel inference, a combination of (4) and (10).

(2) We deal with all offenders without knowing in advance the type of the investigated offender. For each offender we admit the possibility that the offender is a resident of a certain type or a non-resident. Then we appropriately combine these two possibilities.
  (a) For modelling of residents M1 we use the model (4), for residents M2 the model given by (5), for residents M3 we use multimodel inference, a combination of (4) and (5), for non-residents the model (10).
    (i) Multimodelling weights are the same both for residents and for non-residents.
    (ii) Multimodelling weights for residents and non-residents are derived by frequencies of these types in our dataset.
  (b) For modelling of residents M1 we use the model (4), for residents M2 the model given by (10), for residents M3 we use multimodel inference, a combination of (4) and (10), for non-residents the model (10).
    (i) Multimodelling weights are the same both for residents and for non-residents.
    (ii) Multimodelling weights for residents and non-residents are derived by frequencies of these types in our dataset.

We compare these methods with Rossmo's approach. Rossmo works with hit score function given by (1) and distance decay function of the form

$$f\left(d\left(\mathbf{x}_i,\mathbf{y}\right)\right)=\begin{cases}\dfrac{k}{\left(d\left(\mathbf{x}_i,\mathbf{y}\right)\right)^h} & \text{pro } d\left(\mathbf{x}_i,\mathbf{z}\right)>b\\[2ex] \dfrac{kb^{g-h}}{\left(2b-d\left(\mathbf{x}_i,\mathbf{y}\right)\right)^g} & \text{pro } d\left(\mathbf{x}_i,\mathbf{z}\right)\leqslant b\,,\end{cases}\tag{17}$$

where b denotes the radius of the buffer zone, distance $d\left(\mathbf{x}_i,\mathbf{z}\right)$ is calculated by the Manhattan metric and exponents f and g are equal to 1,2. These values of the parameters are recommended by Rossmo based on his research. We set the value of the parameter b equal one half of the average distance of the nearest neighbour between crimes in the given crime series [7]. The choice of parameter k is not important because the hit score function is the sum of the individual distance decay functions, the values of the hit score function are compared among themselves.

[7]This choice for the parameter b is presented eg. in Raine et al. (2009).

The constant multiplies these values of the hit score function but does not change ratios between them. Rossmo's formula assumes that the offender's anchor point is located close to his crime sites (how close – it depends on the optional parameter b). This is the reason why this relationship is suitable especially for residents.

Let the investigated jurisdiction lie in the area of the rectangle with sides 100 km and 70 km, defined by the UTM coordinates 300 km west, 400 km east, 4330 km south and 4400 km north [8]. We divide this rectangle into a grid 70×100, thus the dimensions of each cell are approximately 1,4 km $\times$ 1,4 km.

For each investigated offender we plot his crime locations. Based on the space distribution, distances between crimes and occurrence or absence of clusters, we can determine the most appropriate criminal type, or subtype for the given offender. According to this type we choose a suitable model.

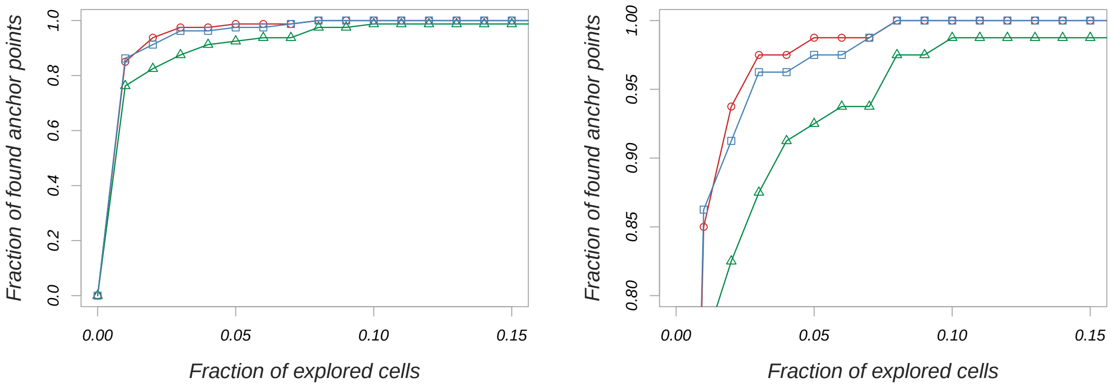
If we only deal with residents, in each cell of jurisdiction we evaluate the posterior for the considered methods described at the beginning of this section and for Rossmo's approach. In the case when we examine all criminals without knowing the type (resident or non-resident) of the offender, we evaluate also the posterior in each cell for the situation that the criminal is a non-resident. Then we multiply the posterior for residents and the posterior for non-residents by the appropriate weights and after their sum we obtain an estimate of the anchor point $\mathbf{z}$, if the offender committed crimes in locations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. Again, we apply this process to all methods described at the beginning of this section and compare them with Rossmo's approach.

For each method we order all cells based upon the value of the posterior, from highest to lowest, thus from the cell that contains the anchor point with the highest probability to the cell that includes the anchor point with the lowest probability. The efficiency of the method depends on how many cells we have to examine until we find the anchor point of the investigated offender. If we divide this number of cells by the total number of cells in the given jurisdiction, we obtain the percentage of the area that we have to explore to find the offender's anchor point. Thus, the method with the lowest percentage is the most effective for the particular series.

Figure 10 shows a comparison of methods [1a], [1b] and Rossmo's approach in terms of this evaluation. We only selected residents from the dataset; we did not consider the possibility that any of the criminals could be a non-resident. Although Rossmo's approach is appropriate just for residents, both methods [1a] and [1b] indicate better, and very similar, efficiency.

Fig. 11 gives a comparison of methods [2ai] and [2bi] and Rossmo's approach (Fig. 11a and Fig. 11c) in terms of the described means of evaluation. It also compares methods [2aii] and [2bii] and Rossmo's approach (Fig. 11b and Fig. 11d). Now we consider all offenders without knowing the type of each criminal. We can see that Rossmo's approach almost always exhibits the lowest efficiency in both cases. However, the difference between efficiencies of the methods is not as remarkable as in the situation considered above when we dealt only with residents. Due to the inclusion of all criminals to the analysis, the estimation of the anchor point for methods [2ai], [2bi] and [2aii], [2bii] worsened for residents since we had to admit the possibility that the investigated offender is a non-resident. These methods achieve better results for non-residents than Rossmo's approach.

[8] The size of the jurisdiction was chosen to include all crimes and anchor points in the dataset $\pm$ approximately 5 km.



(A) *Normal scale.* (B) *Detail.*

FIGURE 10. *The relationship between the proportion of the found anchor points and the proportion of the explored area, when we only deal with residents; method [1a] (red circles), method [1b] (blue squares), Rossmo's approach (green triangles).*

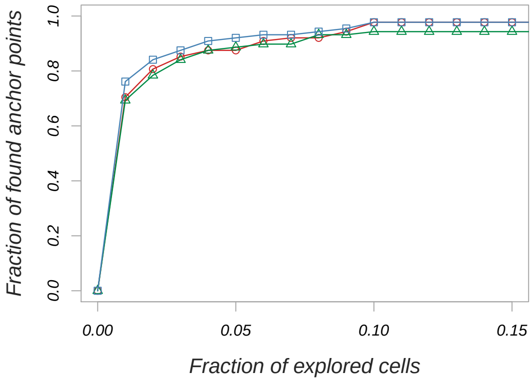
However, in our datasets, non-residents constitute only $\frac{1}{11}$ of all offenders (it corresponds to the choice of multimodel weights for methods [2aii] and [2bii]).

Nevertheless, we can say, that methods [2ai], [2bi] and [2aii], [2bii] exhibit better results than Rossmo's approach (although due to the previous case when we dealt only with residents the difference between the results of the considered approaches is not so significant). We can assume even higher efficiency for datasets with greater proportion of non-residents. We can see that for our dataset, the best choice for residents with buffer zone in both cases is the model of the form (10). The option of multimodel weights distinguishes considered method as expected. If we assign the same weights to both types of criminals (Fig. 11a, respectively Fig. 11c), non-residents are caught earlier. On the contrary, the choice of weights based on frequencies of the types in the dataset (Fig. 11b, respectively Fig. 11d), when the weight for the model of residents corresponds to the value of $w_1 = \frac{10}{11}$ and for the model of non-residents takes the value of $w_2 = \frac{1}{11}$, , leads to earlier capturing of residents at the expense of later finding of non-residents. Overall, Fig. 11 shows that the use of the prior knowledge about the structure of the dataset to determining the weights results in faster capture of a larger part of the offenders.

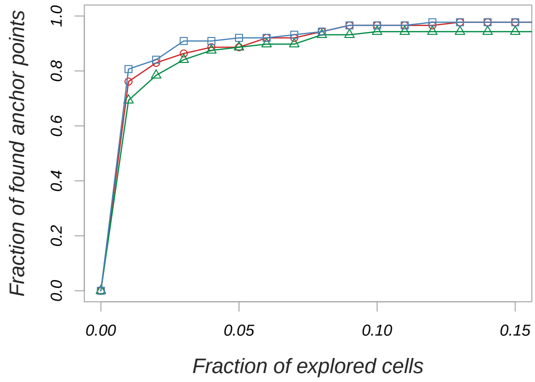
In Fig. 12 we can see estimates of the offender's anchor point – for each case we use the method [2ai] and Rossmo's approach. For residents without buffer zone, both methods are similarly effective (Fig. 12a and 12b). However, in the case of residents with buffer zone (Fig. 12c and 12d) and also in the case of non-residents (Fig. 12e and 12f), the hit score function using Rossmo's distance decay function still assumes that the anchor point lies close to any of the offender's crime sites. Conversely, the model [2ai] admits the possibility that the criminal's anchor point is located at a greater distance from his crime sites.

7. CONCLUSION

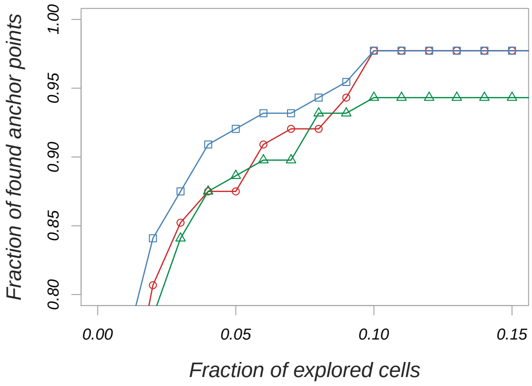
We offered a more complex approach, how to treat the problem of geographic profiling. After analyzing the data set on serial criminals from certain area, we can observe similar tendencies for choosing the crime site. Based on this, we classify the criminals into different types. It often turns out that similar distribution of



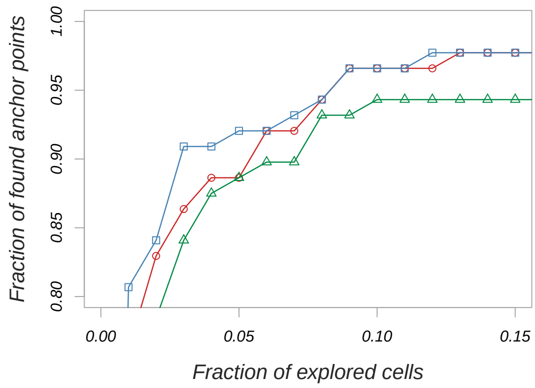
(A) *The same weights for model of residents and model of non-residents.*



(B) *The weights according to the frequencies of residents and non-residents in our data set.*



(C) *The same weights for model of residents and model of non-residents (detail).*

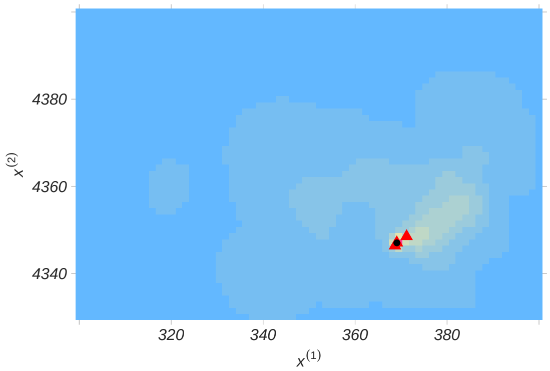


(D) *The weights according to the frequencies of residents and non-residents in our data set (detail).*

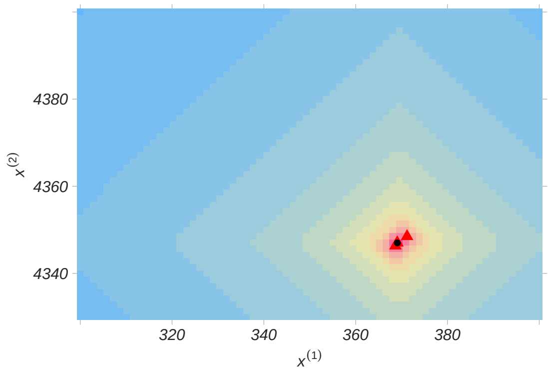
FIGURE 11. *The relationship between the proportion of the found anchor points and the proportion of the explored area, when we deal with all offenders without knowing their types in advance. The parts (a) and (c): method [2ai] (red circles), method [2bi] (blue squares), Rossmo's approach (green triangles); The parts (b) and (d): method [2aii] (red circles), method [2bii] (blue squares), Rossmo's approach (green triangles).*

crime sites corresponds to very similar behaviour when choosing the crime site. Thanks to this we can, for a given criminal, construct a model based on behaviour of criminals of the same type.

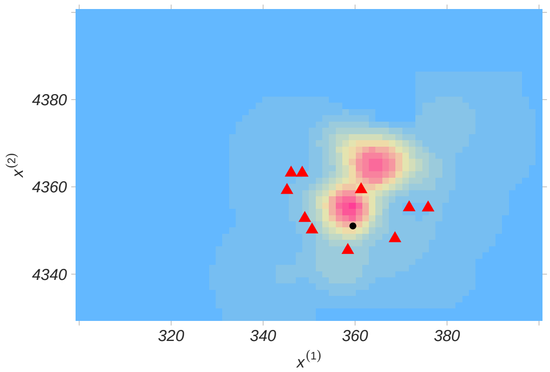
We divided offenders from our data set into two basic groups – residents and non-residents. This distinction is very similar to the marauder and the commuter hypothesis. Main idea of the overlap of criminal and anchor area is the same. But models, which are able to describe the behaviour of commuters (in our case non-residents) are very rare in the literature. One of the few is suggested in the paper of *Mohler* and *Short* (see Mohler and Short, 2012). Their approach, based on solving a stochastic differential equation, leads to a model which for a suitable choice of parameters characterizes well the nature of commuters (non-residents). Here we offered a simpler, purely bayesian model, which preserves the main features and flexibility of models suggested in Mohler and Short (2012).



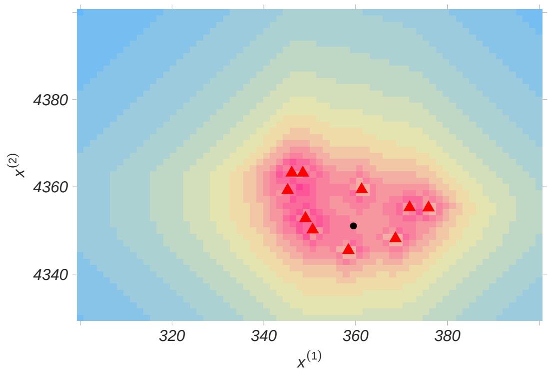
(A) *Model [2ai] for a resident without a buffer zone.*



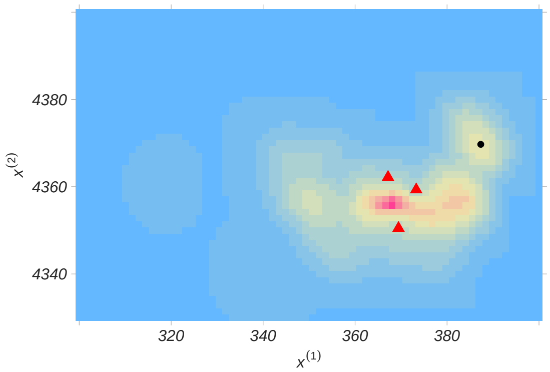
(B) *Rossmo's model for a resident without a buffer zone.*



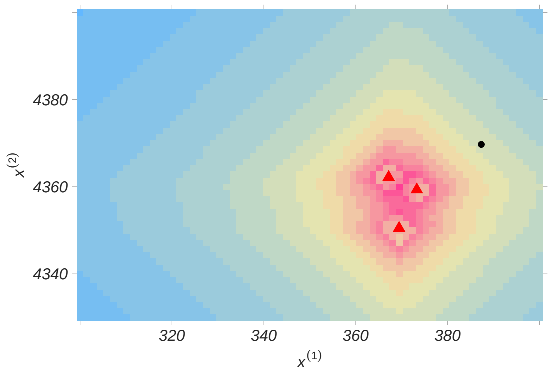
(C) *Model [2ai] for a resident with a buffer zone.*



(D) *Rossmo's model for a resident with a buffer zone.*



(E) *Model [2ai] for a non-resident.*



(F) *Rossmo's model for a non-resident.*

FIGURE 12. *Level plots indicating how likely is that the area contains the anchor point of the offender (regions with the highest probability are pink, areas with the lowest probability are blue). The red triangles denote crime sites of the offender, the black circle indicates his anchor point.*

Considered models better capture the criminal's behaviour than the models of O'Leary (2009a), since our setup is based on data from similarly behaving criminals. At the same time, they are easier to apply and interpret than the models of *Mohler* and *Short*, while keeping the ability to cover a broad spectrum of offenders. Second, the paper showed how to distinguish the types of offenders based on the spatial distribution of their crime sites.

In case we were not able to asign the given criminal to a certain group of already investigated criminals, it was natural to use multimodel inference. It allows to incorporate into the model all possibilities which we consider. In particular, this approach was used to deal with the case when the criminal can be a resident or a non-resident. In our case it turned out that it is suitable to derive weights for multimodel inference by frequencies of residents and non-residents in the dataset.

We used the standard Rossmo's model as a benchmark. Our model substantially overperformed the benchmark, especially in the presence of non-residents.

References

Bolstad, W. M. (c2007). *Introduction to Bayesian statistics*. John Wiley, Hoboken, N.J., 2nd ed. edition.

Brantingham, P. L. and Brantingham, P. J. (1993). Nodes, paths and edges. *Journal of Environmental Psychology*, 13(1):3–28.

Burnham, K. P., Anderson, D. R., and Burnham, K. P. (c2002). *Model selection and multimodel inference*. Springer, New York, 2nd ed. edition.

Canter, D. (1996). *Psychology in action*. Dartmouth, Aldershot, England Brookfield, Vt., USA.

Canter, D., Coffey, T., Huntley, M., and Missen, C. (2000). Predicting serial killers' home base using a decision support system. *Journal of Quantitative Criminology*, 16(4):457–478.

Carlin, B. (2000). *Bayes and Empirical Bayes methods for data analysis*. Chapman & Hall/CRC, Boca Raton.

City of Tucson (n.d.). Uniform crime reporting codes. Retrieved March 8, 2014, from `http://www.tucsonaz.gov/police/ucr-codes`.

Damien, P., Dellaportas, P., Polson, N. G., Stephens, D. A., and Smith, A. F. (2013). *Bayesian theory and applications*. Oxford University Press, Oxford, first edition. edition.

FBI (2012). Uniform crime reports. Retrieved March 8, 2014, from `http://www.fbi.gov/about-us/cjis/ucr`.

Kawase, K. (2011). A general formula for calculating meridian arc length and its application to coordinate conversion in the gauss-krüger projection. *Bulletin of the Geospatial Information Authority of Japan*, 59:1–13.

Kawase, K. (2012). The environmental range of serial rapists. *Bulletin of the Geospatial Information Authority of Japan*, 60:1–6.

Kennedy, M. (2000). *Understanding map projections : GIS by ESRI*. ESRI, Redlands, CA.

Kooperberg, C. and Stone, C. J. (1991). A study of logspline density estimation. *Computational Statistics & Data Analysis*, 12(3):327–347.

Levine, N. (2008). Crimestat: A spatial statistical program for the analysis of crime incidents. In *Encyclopedia of GIS*, pages 187–193. Springer US.

Mohler, G. O. and Short, M. B. (2012). Geographic profiling from kinetic models of criminal behavior. *SIAM Journal on Applied Mathematics*, 72(1):163–180.

O'Leary, M. (2009a). The mathematics of geographic profiling. *Journal of Investigative Psychology and Offender Profiling*, 6(3):253–265.

O'Leary, M. (2009b). A new mathematical approach to geographic profiling. *Towson University*.

O'Leary, M. (2010a). Implementing a bayesian approach to criminal geographic profiling. In *Proceedings of the 1st International Conference and Exhibition on Computing for Geospatial Research & Application*, pages 59:1–59:8, New York, NY, USA. ACM.

O'Leary, M. (2010b). Multimodel inference and geographic profiling. *Crime Mapping*, 2(1).

Raine, N. E., Rossmo, D. K., and Le Comber, S. C. (2009). Geographic profiling applied to testing models of bumble-bee foraging. *Journal of The Royal Society Interface*, 6(32):307–319.

Robert, C. (2007). *The Bayesian choice from decision-theoretic foundations to computational implementation*. Springer, New York.

Rossmo, D. (2000). *Geographic profiling*. CRC Press, Boca Raton, Fla.

Rossmo, D. K. (1995). *Geographic profiling: Target patterns of serial murderers*. PhD thesis, Theses (School of Criminology)/Simon Fraser University.

Sivia, D. S. (2006). *Data analysis a Bayesian tutorial*. Oxford University Press, Oxford New York.