Natural Bayesian Killers

Team #8215

February 22, 2010

Abstract

In criminology, the real life CSI currently faces an interesting phase, where mathematicians are entering the cold-blooded field of forensics. Geographic profiling uses existing geocoded data to predict, solve and prevent crime. In this study, we examine probability distance strategies and a Bayesian approach used to predict the residential location of a serial offender. We compare different distance decay functions used for estimating the probability density of the residential location. Additionally, we derive a Bayesian approach for predicting the next target location of the serial offender following the formulation of O'Leary [2009].

For the measure of performance, we choose the *search cost* ranking. The model is tested using the *leave-one-out cross-validation* with a small example dataset that consists of serial murder cases. Different strategies of combining the six different methods are discussed. The tests suggest that the best predictions are achieved using the Bayesian approach with exponential decay function. However, this result is not statistically significant due to the lack of test data.

Contents

1 Introduction

For the untrained eye, a set of crime sites of a serial killer might seem totally random when visualized on a map. However, the behavior of the serial killer usually has some underlying statistical regularities which can be discovered by computerized data analysis. The same sort of behavior can be found for all kinds of serial offenders: killers, rapists, burglars, etc. This type of analysis is called *geographic profiling*.

Understanding the statistical regularities may help predicting the offender's anchor point (residence, workplace, etc.). This can be crucial for catching the offender before further crimes are committed. However, the geographic profiling methods usually neglect some essential pieces of information regarding the situation, such as that it is very unlikely to find the offender's residence from the middle of a lake. Therefore, geographic profiling should be considered a decision support system rather than an expert system [Canter et al., 2000].

In Section 2, we first present the *spatial distribution strategies* which are a simple approach to the problem [Snook et al., 2005]. They predict the offender's anchor point by calculating, e.g., the centroid of the crime sites. Another category of strategies is the *probability distribution strategies* which are the most common way of addressing the problem of finding the anchor point. They rely on *distance decay functions*, which estimate the distribution of the distances the offender is willing to travel to the crime site. Several studies that compare different decay functions exist [Levine, 2004, Canter and Hammond, 2006, Snook et al., 2005].

In recent years, there has been a rise in the popularity of the Bayesian approaches tackling the problem of locating the anchor point. This can be seen in the number of papers published recently [O'Leary, 2009, Mohler and Short, 2009, Levine, 2009]. Following O'Leary [2009], we derive the probability density function (PDF) for the anchor point from the Bayes' rule. We then discuss the assumptions behind the Bayesian formalism and also address the question, how to measure the performance of the prediction.

At the end of Section 2, we present methods for combining the predictions of different methods. The most advanced combination methods can take the different strengths of different models into account.

A crime investigator might also be interested in knowing the next target of the offender. Section 3 deals with this problem presenting a Bayesian approach for estimating the PDF of the next crime site. O'Leary [2009] shows how this approach comes almost free of charge after we have derived the PDF for the anchor point.

Section 4 discusses the data used for testing the geographic profiling methods. It also describes the *leave-one-out cross-validation* method, which is a useful method for small data samples.

In Section 5 we present the results and some sensitivity analysis. After drawing the conclusions in Section 6, we give a technical summary that describes how a crime investigator can utilize the Bayesian approach for locating the offender's anchor point and the next target location.

2 Predicting the Residential Location of the Offender

Geographic profiling is used as a decision support system for locating the anchor point (residence, workplace, etc.) of a serial offender [Canter et al., 2000]. A prediction for the anchor point is given based on the offender's crime site locations. The prediction can be either a single spot or a probability density. This criteria divides the prediction methods into *spatial distribution strategies* and *probability distance strategies* [Snook et al., 2005].

2.1 Spatial Distribution Strategies

Spatial distribution strategies calculate a central point of the crime site locations. This point is an estimate of the anchor point. Various methods for calculating the central point have been proposed. Snook et al. [2005] compare 6 different spatial distribution strategies: *centroid*, *centre of the circle*, *median*, *geometric mean*, *harmonic mean* and *center of minimum distance*.

The centroid method is the simplest of these strategies and it calculates the mean of the x and y coordinates. Even this simple method can provide useful predictions, as in the case of the "Yorkshire Ripper" where the method was able to predict correctly the killer's home town. Furthermore, Snook et al. [2005] suggest that complexity of the strategy does not guarantee a better accuracy.

A restraint of the spatial distribution strategies is that they only provide a single point as the prediction. If the offender is not found there, the system does not give any advice for the crime investigator where to continue the search.

2.2 Probability Distance Strategies

Probability distance strategies (PDS), also referred as *journey-to-crime* estimation [Levine, 2004], create a geographical profile which is usually displayed on a map. The height of the profile indicates the likelihood of the anchor point to be found on the corresponding place. Thus, PDS provide a prioritized search strategy which can give them a significant advantage over the spatial distribution strategies [Snook et al., 2005].

PDS are based on distance decay functions that characterize the distribution of distances between the anchor point and the crime scenes. Comparison of different decay functions has been conducted by Levine [2004], Canter and Hammond [2006] and Snook et al. [2005]. Popular decay functions are: *negative exponential, lognormal, normal, truncated negative exponential,* and *linear functions*.

To create the geographical profile, we discretize the search area and go through each point $\mathbf{y}$. We calculate the overall effect of all the crime sites $\mathbf{x}_i$ on $\mathbf{y}$ and get the hit score $S(\mathbf{y})$ as referred by O'Leary [2009]

$$S(\mathbf{y}) = \sum_{i=1}^{n} f(d(\mathbf{x}_i, \mathbf{y})) = f(d(\mathbf{x}_1, \mathbf{y})) + \cdots + f(d(\mathbf{x}_n, \mathbf{y})). \tag{1}$$

The distance function d can be selected, e.g., as the Euclidean or the Manhattan distance. In our analysis, we use the Euclidean distance. The hit score gives a priority order to the search. Figure 1 shows an example of a geographic profile regarding the case of serial killer Peter Sutcliffe, the Yorkshire Ripper. Crime sites are marked with red circles and Sutcliffe's residence with a black circle. The centroid point of the crime sites is marked with a black cross.

A red value in the profile denotes a good chance of finding the anchor point. We can see that the actual location of the residence — the anchor point — is clearly in the red zone of the map, but the peak of the profile is located somewhat away from the anchor point.

2.2.1 Distance Decay Functions

Distribution of the distances between the anchor point and the crime scenes is characterized by a distance decay function [Canter and Hammond, 2006]. Common choices for the decay function include: *negative exponential, lognormal, normal, truncated negative exponential,* and *linear functions*. Comparison of different decay functions has been conducted by Levine [2004], Canter and Hammond [2006] and Snook et al. [2005].

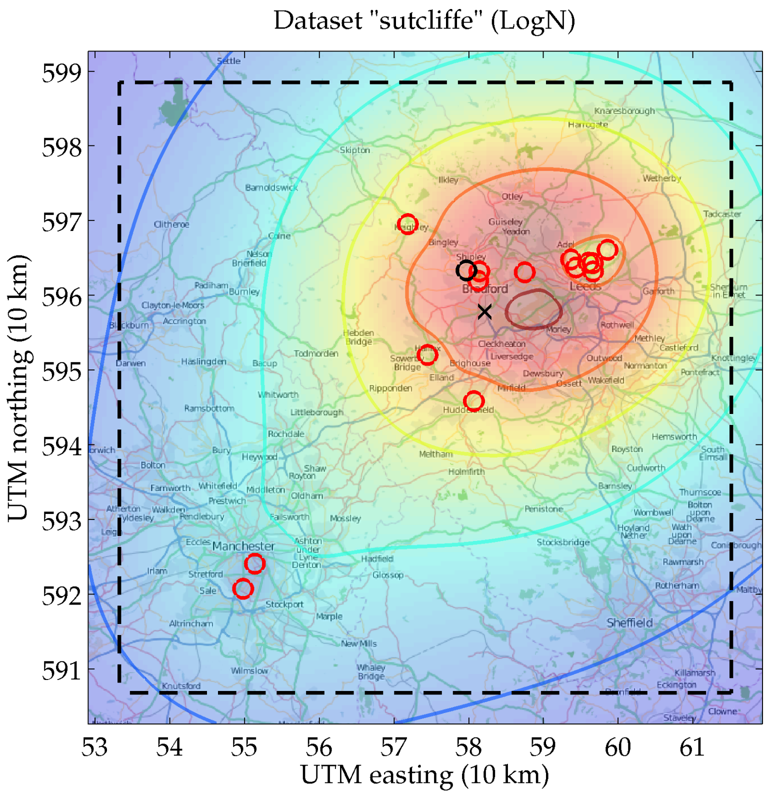


Figure 1: A geographic profile regarding the case of serial killer Peter Sutcliffe, the Yorkshire Ripper. Crime sites are marked with red circles and Sutcliffe's residential location with a black circle. The centroid point of the crime sites is marked with a black cross.

Decay functions usually give high values for low distances, i.e. the offender is not willing to travel far to commit the crime. Lognormal and truncated negative exponential functions also implement a *buffer zone* which corresponds to the behavior that the offender might not feel comfortable committing a crime right next to their residence.

In this paper, we study three different decay functions: lognormal, negative exponential and gamma functions. All three are continuous and differentiable probability density functions

$$f_{LogN}(d;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}, \tag{2}$$

$$f_{Exp}(d;\lambda) = \lambda e^{-\lambda x}, \tag{3}$$

$$f_{Gamma}(d;\mu,\sigma) = x^{k-1}\frac{e^{-x/\theta}}{\theta^k\Gamma(k)}. \tag{4}$$

The gamma distribution is chosen since the exponential distribution is a special case of it where $k = 1$. However, by increasing the k parameter a buffer zone is created.

In Figure 2 we see the three decay functions fitted to real data.

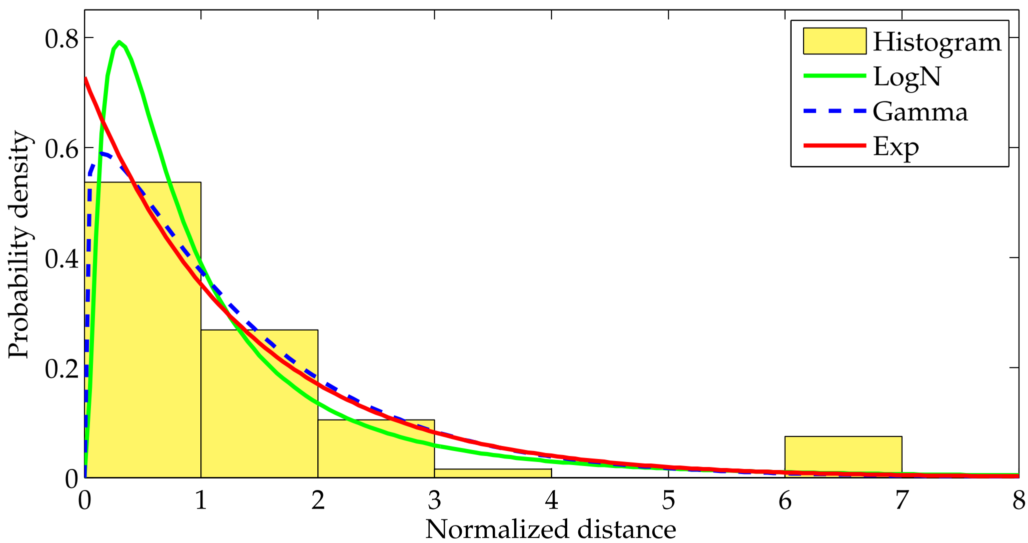


Figure 2: A histogram of normalized distances and the three different decay functions that have been fitted to the data.

The parameters of the decay function are estimated using the *maximum likelihood estimation* [Alpaydin, 2004, p. 62]. The advantage of this method is that it gives confidence intervals for the parameters, which allow us to assess the uncertainty in the predictions.

2.3 A Bayesian Approach

O'Leary [2009] applies the Bayes' rule to derive the probability distribution $P(\mathbf{z}|\mathbf{x}_1, \ldots, \mathbf{x}_n)$ for the anchor point $\mathbf{z}$ based on the crime sites $\mathbf{x}_1, \ldots, \mathbf{x}_n$ [1].

The Bayes' rule gives us

$$P(\mathbf{z}|\mathbf{x}_1, \ldots, \mathbf{x}_n) = \frac{P(\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{z})P(\mathbf{z})}{P(\mathbf{x}_1, \ldots, \mathbf{x}_n)}. \tag{5}$$

The prior probability density function $P(\mathbf{z})$ is set to the constant value 1 in our analysis. However, it could be straightforwardly included in order to take the geographic properties of the search area into account. For example, we could define

$$P(\mathbf{z}) = \begin{cases} 0 & \text{, when } \mathbf{z} \in \text{uninhabitable areas} \\ 1 & \text{, when } \mathbf{z} \notin \text{uninhabitable areas} \end{cases} \tag{6}$$

to exclude the uninhabitable from the search. Mohler and Short [2009] have used this approach and merged the housing density into $P(\mathbf{z})$.

$P(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ in Formula (5) can be ignored, since it is merely a scaling factor that is independent of the anchor point $\mathbf{z}$. Assuming that the crime sites are statistically independent, we get

$$P(\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{z}) = P(\mathbf{x}_1|\mathbf{z}) \cdots P(\mathbf{x}_n|\mathbf{z}). \tag{7}$$

Thus we may write

$$P(\mathbf{z}|\mathbf{x}_1, \ldots, \mathbf{x}_n) \propto P(\mathbf{x}_1|\mathbf{z}) \cdots P(\mathbf{x}_n|\mathbf{z}). \tag{8}$$

The term $P(\mathbf{x}_1|\mathbf{z})$ depends on the distance decay function $f(d)$ and the prior probability density function of the crime sites $P(\mathbf{x})$. The offender might be prone to commit crimes, e.g., in the less populated areas. This

[1] O'Leary also includes an α parameter which defines the shape of the offender's distance decay function. However, we assume that the same shape parameters hold for all offenders. We normalize the distances in order to take into account the individual scales of the crime site distributions.

behavior could be encoded in $P(\mathbf{x})$. In our analysis, we restrict our focus to the case $P(\mathbf{x}) = 1$, and thus we get

$$P(\mathbf{x}_i|\mathbf{z}) = f(d(\mathbf{x}_i, \mathbf{z})) \tag{9}$$

and finally

$$P(\mathbf{z}|\mathbf{x}_1, \ldots, \mathbf{x}_n) \propto f(d(\mathbf{x}_1, \mathbf{z})) \cdots f(d(\mathbf{x}_n, \mathbf{z})). \tag{10}$$

We notice that without the priors $P(\mathbf{z})$ and $P(\mathbf{x})$, we get almost the same formula that is traditionally used (1). Only the summation is replaced by multiplication.

The advantage of the Bayesian approach is that it is mathematically well-founded given the assumptions it makes (see Section 2.3.1). It also provides a natural way to include information about geographical properties of the area (e.g., by excluding uninhabitable areas from the search area).

2.3.1 Assumptions

The Bayesian approach makes the following assumptions

1. **Independent crime sites:** $P(\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{z}) = P(\mathbf{x}_1|\mathbf{z}) \cdots P(\mathbf{x}_n|\mathbf{z})$

   The offender chooses the next crime site regardless of the previous crime sites. This assumption is also made by O'Leary [2009] and Mohler and Short [2009].

2. **Distance decay function parameters are the same for all offenders**

   Warren et al. [1998] suggest that the distance travelled by a serial rapist depends on the age and race of the rapist. To take this into account, our model normalizes all distances by the mean inter-point distances, as done by Canter et al. [2000]. However, Warren et al. [1998] also suggest that the more rapes there are on the rapist's account, the farther the rapist tends to travel. Our model ignores this possibility.

3. **Uniform anchor point distribution:** $P(\mathbf{z}) = 1$

Our analyses have been conducted giving the same prior probability to all locations. This ignores the geographical properties, such as uninhabitable areas and housing densities, which naturally affect the distribution. However, these properties could be easily included in our model, if there was access to relevant data of the search area.

4. **Uniform crime site distribution:** $P(\mathbf{x}) = 1$

   We do not make any prior assumptions of the crime site distribution. We reckon that for example for sexual offences very public places are not as probable crime sites as more private ones. Again, this information can be included in our model if the data is available.

2.3.2 Normalization

The average distance an offender is willing to travel to commit a crime varies between individuals. Warren et al. [1998] suggest that these individual differences can be explained by variables such as the offender's age and race.

Canter et al. [2000] describe two normalization methods for taking the individual differences into account. The first method calculates the mean inter-point distance (MID) between all offenses. When estimating the decay function parameters or applying the decay function in Formula (10), the distances are always normalized by dividing them by the MID.

The second method, called the *QRange*, calculates a linear regression of the crime sites. Instead of the MID, the perpendicular distance from every crime site to the regression line is calculated. This method takes into account any linear structure in the crime site distribution which might derive, e.g., from an arterial pathway.

In our analysis, the normalization is conducted with the MID metric.

2.3.3 Performance of the Prediction

To compare a variety of methods, we need a way to measure the effectiveness of different techniques. Canter et al. [2000] propose that reducing police resources would be a suitable objective. The potential search area is discretized to a grid. The *cost* of carrying out a search on a particular spot (or cell in the grid) depends on the number of searched cells before. The assumption is that the police start the search at the most probable spot and move from spot to spot by decreasing probability.

Different decay functions can this way be compared by their cost-effectivity. This measure reflects the ability of a method to prioritize a cell and identify the location of the anchor point. A search area is estimated by calculating the mean inter-point distance, in a Manhattan sense, between crime sites with respect to the x coordinates

$$x_m = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|. \tag{11}$$

and y coordinates

$$y_m = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j|. \tag{12}$$

separately. Rossmo [1995a] introduces the concept of a rectangular search area — also called the *hunting area*. The rectangle is fitted to the farthest extent of the known linked crime sites, and it is extended by half of the mean Manhattan distance x_m on the horizontal axis and half the y_m on the vertical axis in all directions. We use this approach, but we extend the search area by the mean value in every direction, i.e. by $2x_m$ horizontally and $2y_m$ vertically.

The effectiveness of the search of the rectangular area is calculated by assigning each point on the grid a weighting indicating the *likelihood* of residence. Canter et al. [2000] use this method so that the weightings are used to calculate a *search cost rank index* for each grid cell. These search cost rank indexes are derived from the decay function that the method uses.

Each cell point inside the array is then searched for the anchor point, starting from the highest rank index. When the anchor point cell is reached, the search is terminated and a *search cost* value is calculated. This value reflects the proportion of the rectangle that has been searched before termination. A search cost value of 0 would indicate that the criminal was found in the first cell (i.e. the cell with the highest priority). If the search is terminated without reaching the known location of the criminal, the anchor point was not within the search area. This means that the method was a failure.

2.4 Combining Models

Different decay functions give different geographical profiles. Calculating an ensemble of these profiles can make the predictions more accurate and more stable. Finding the underlying profile can be seen as a regression problem and thus, we have a regression ensemble problem, as referred in the machine learning community.

Alpaydin [2004] shows several simple strategies for calculating the ensemble, two of which are the average and median strategies. The average strategy iterates over the search area and at each point it calculates the average height of the profiles given by different models. The median strategy takes the median of the heights. According to Alpaydin [2004], the sum strategy is the most widely used in practice. The advantage of the median strategy is that it is more robust to outliers.

Several studies about more advanced regression ensemble methods exist (see e.g. Brown et al. [2005]; Rätsch et al. [2002]). The more advanced methods take into account the strengths and weaknesses of different models. For example, one method can give more accurate predictions in the cases where the mean inter-point distance (MID) of the offences is high, whereas another one is more suitable for cases with a low MID.

A straightforward strategy for exploiting the different strengths of the models is to divide the range of the MID values into n intervals. For each interval, we calculate which model gives the best results for the cases within that interval. When we get a new case, we calculate the new MID, and then check which interval the new MID goes into and let the best model within that interval calculate the geographic profile.

In our case, the decay functions tested did not give statistically different results due to limited amount of test data. Therefore, no ensemble methods were tested in our study. To adapt the models to different types of offenders, bagging or boosting methods could be adopted, as described in [Alpaydin, 2004, pp. 430–434].

3 Predicting the Next Target Location

Usually, geographic profiling is used merely for estimating the offenders residential location — the anchor point. However, it might as well be to the authorities interest, to predict the target location of the next offense in order to prevent it and catch the offender red-handed.

3.1 A Bayesian Approach

Formally the task is to calculate the conditional probability density distribution $P(\mathbf{x}_{next}|\mathbf{x}_1, \dots \mathbf{x}_n)$. The peak of this distribution gives the most probable location of the next offense. O'Leary [2009] points out that the Bayesian approach for the anchor point prediction also gives an estimate for the next target location. This is achieved by calculating the *posterior predictive distribution*

$$P(\mathbf{x}_{next}|\mathbf{x}_1, \dots \mathbf{x}_n) = \iint P(\mathbf{x}_{next}|\mathbf{z}) P(\mathbf{z}|\mathbf{x}_1, \dots, \mathbf{x}_n) dz_x dz_y. \tag{13}$$

Using the notation of Section 2.3 and formula (10) we get

$$P(\mathbf{x}_{next}|\mathbf{x}_1, \dots \mathbf{x}_n) = \iint f(d(\mathbf{x}_{next}, \mathbf{z})) f(d(\mathbf{x}_1, \mathbf{z})) \cdots f(d(\mathbf{x}_n, \mathbf{z})) dz_x dz_y. \tag{14}$$

In practice, we discretize the search space, which transforms the integrals into summations

$$P(\mathbf{x}_{next}|\mathbf{x}_1, \dots \mathbf{x}_n) = \sum_{z_x} \sum_{z_y} f(d(\mathbf{x}_{next}, \mathbf{z})) f(d(\mathbf{x}_1, \mathbf{z})) \cdots f(d(\mathbf{x}_n, \mathbf{z})) dz_x dz_y, \tag{15}$$

where z_x and z_y, denoting the coordinates of the anchor point, range from $-\infty$ to ∞. To calculate this in a computer program, we must choose some threshold distance (from the crime sites to $\mathbf{z}$) beyond which there is no need to go since the product of functions f is practically zero.

Figure 3 shows an example of the next target's estimated probability density function based on the previous offenses. The figure is read similarly to the geographical profiles, i.e. the next target is likely to be found in the red area.

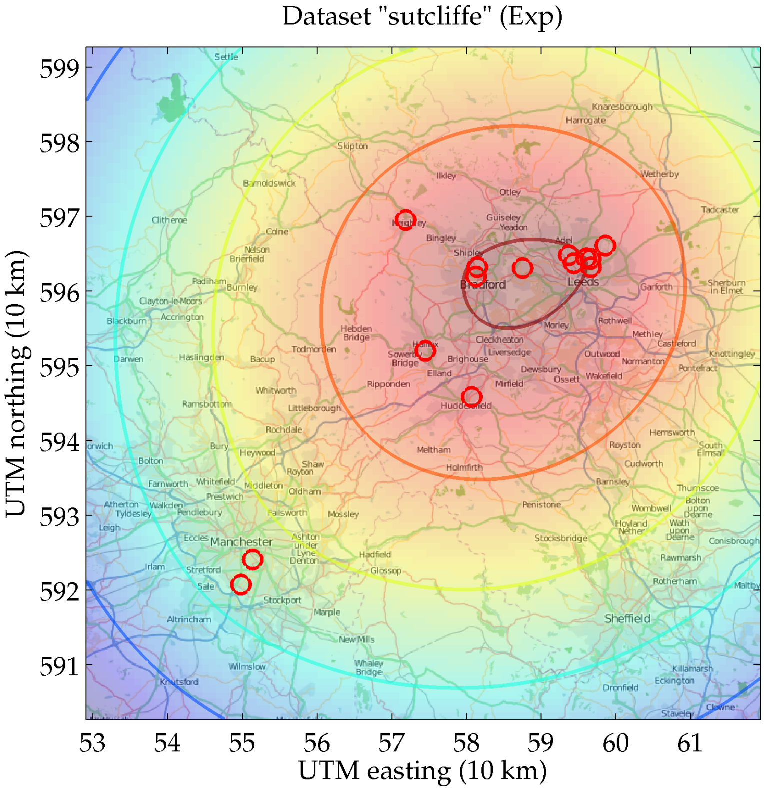


Figure 3: A profile of the next target of Peter Sutcliffe. Crime sites are marked with red circles and Sutcliffe's residential location with a black circle. A red value indicates high probability density.

3.2 Assumptions

We make the same assumptions that were presented and discussed in Section 2.3.1. Again, we assume that the offenses are statistically independent meaning that the predictions are invariant to the order — and time — of the offenses.

3.3 Performance of the Prediction

The search cost approach, described in Section 2.3.3, can be utilized in this case with some minor modifications. Instead of calculating the search cost of the anchor point, we calculate it for some of the previous crime sites. This crime site is left out from the calibration phase.

Calculating the profile for the next target is very similar to calculating the geographic profile for the anchor point. However, the time complexity grows from $O(n)$ to $O(n^2)$, where n is the number of cells in the grid. In consequence, calculating the next target profile takes approximately one hour with a 100x100 grid on an average desktop computer.

4 Experiments

4.1 Data

Both the spatial distribution strategies and the probability distance strategies described earlier are applicable to any data series of activity that includes geographical locations. Snook et al. [2005], Gorr [2004], Rossmo [1995b] and many more have used sample datasets of solved serial crime to validate and test their models. Most of these datasets are either classified or kept out of public for some other reason.

Our small test dataset consists of six cases of serial crime, some of which have gained notable public attention. These datasets are used for parameter estimation and cross validation of the models. The cases included are named after convicted criminals, which are Peter "The Yorkshire Ripper" Sutcliffe [Wikipedia], Chester Turner [Iniguez, L.], Gary "Green River Killer" Ridgway [Nowlin M. and Chaumont K., 2003], John Allen "the Beltway Sniper" Muhammad [Wikipedia, 2010], Steve "the Suffolk Strangler" Wright [Harris, 2008] and Terry Blair [News, 2004]. These cases are shortly introduced in table 1. In addition to the location data regarding victims, also a residential home location (an anchor point) of the convict was part of every case.

Table 1: The six cases of serial murder used for validating and testing.

Dataset	Name of Criminal	Datapoints	Years active	Country
Sutcliffe[2]	Peter Sutcliffe	18	1975-80	UK
Turner	Chester Turner	12	1987–98	US
Ridgway	Gary Ridgway	30	1982–90	US
Muhammad[3]	John Allen Muhammad	15	2002	US
Wright	Steve Wright	5	2006	UK
Blair	Terry Blair	5	2004	US

[2] The dataset *Sutcliffe* includes four datapoints of attempted murder.

[3] In the *Muhammad* set, the home location is specified by the place of arrest, due to the fact that Muhammad lived in his van.

4.2 Calibration and Validation

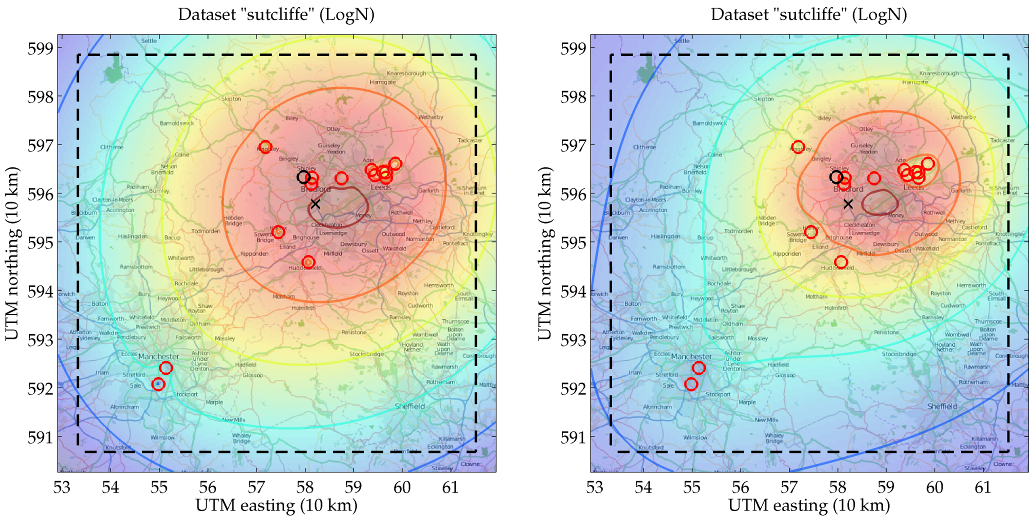
We use datasets from table 1 to get a number of training and validation set pairs. The purpose is to train the method using a dataset $\mathcal{X}$ (after having left out some part as the test set). The small dataset limits the effective use of this method. Repeated use of the same data split differently corrects some drawback. This method is called *cross-validation*. In K-fold cross-validation, the dataset $\mathcal{X}$ is divided randomly into K equal-sized parts, $\mathcal{X}_i$, $i = 1, ..., K$. To generate the pairs, we keep one of the K parts out as the validation set and use the remaining $K - 1$ parts as the training set. Doing this K times — each time leaving out a different part — we get K parts. [Alpaydin, 2004, pp-486-7]

With a small dataset the only practical option is the extreme case of K-fold cross-validation called *leave-one-out*, where only one part is left out as the validation set and training uses the $N - 1$ remaining parts. This way we get N separate pairs by leaving out a different instance at each iteration. [Alpaydin, 2004, p. 487]

5 Results

Our approach uses three different decay functions: lognormal, negative exponential and the gamma function. To compare the results each of decay function, we use the cross-validation method of leave-one-out to estimate the decay functions' parameters from five datasets and then predict the anchor point in the sixth dataset. The results of the different cases are compared with the help of a $[0, 1]$ scaled search cost estimate, where a zero cost is optimal.

In addition to the three different decay functions, we compare the Bayesian approach (equation 5) and the traditional summation method (equation 1).



(a) Bayesian Approach. (b) Summation method.

Figure 4: Predictions for the *sutcliffe* dataset which has been calibrated with the other five sets. The search area is marked with a dotted rectangle. Crime sites are marked with red circles, the location of residence with a black circle, and the center of mass is marked with a black cross.

Figure 4(a) shows the search cost on a map with the help of a color map. The more red a particular spot, the better the search cost. The dotted rectangle depicts the search area. Crime sites are marked with red circles, the location of residence with a black circle, and the centroid spot, "the center of mass", is marked with a black cross. To clarify the effect of rising search costs some contour lines are shown on the map.

Figure 4(a) uses the lognormal decay function and the Bayesian approach. Figure 4(b) uses the summation method. Differences between these two methods are marginal, although the coloring varies. By calculating the search cost rank index, we are only interested in the order of the cell list (See Section 2.3.3). The effect of the lognormal *buffer zone* can be seen as holes around the crime sites.

In Table 2, the search costs are calculated for each case and method with the help of cross validation. The value reflects the proportion of the search area that has been searched before termination at the murderer's doorstep.

Table 2: Search cost values calculated with the help of cross-validation. The value reflects the proportion of the search area that has been searched before termination at the murderer's doorstep.

	Exp		Gamma		LogN	
Dataset	Bayes	Sum	Bayes	Sum	Bayes	Sum
Sutcliffe	0.0235	0.0297	0.0275	0.0355	0.0835	0.0895
Blair	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Wright	0.3998	0.4983	0.3738	0.4629	0.3777	0.5112
Ridgway	0.0304	0.0137	0.0913	0.0566	0.1335	0.1187
Muhammad	0.5222	0.4984	0.5220	0.4977	0.5062	0.5039
Turner	0.0887	0.0982	0.0855	0.0960	0.0737	0.0863
Average	0.3441	0.3564	0.3500	0.3581	0.3624	0.3849

The search area in the *blair* dataset fails to include the real spot, so the search is terminated without success. This can be seen in Table 2 as a row of ones. This can also be confirmed by looking at the Figure 5(d). By comparing the search cost values, we see that in most cases the Bayesian approach scores better than the summation method. In addition, we note that the negative exponential decay function seems to have better results than the other two.

To check the hypothesis that the Bayesian approach with exponential decay function scores better than the others, we use the paired difference test (Student's t-test). The test gives each pair the same result; there are no significant differences between the methods in our datasets. The null hypothesis of equal means is confirmed.

In Figures 5(a) and 5(c) we see the results of the *wright* and *turner* datasets. In these two cases the home location of the criminal is consistent with the predicted anchor point. In addition, we notice that even the center of mass succeeds well in predicting the location. On the other hand, in Figures 5(b) and 5(d), where we see the results of the *muhammad* and *blair* datasets, the model is not quite successful. In fact — as noted before — the model terminates without any success in the *blair* dataset.

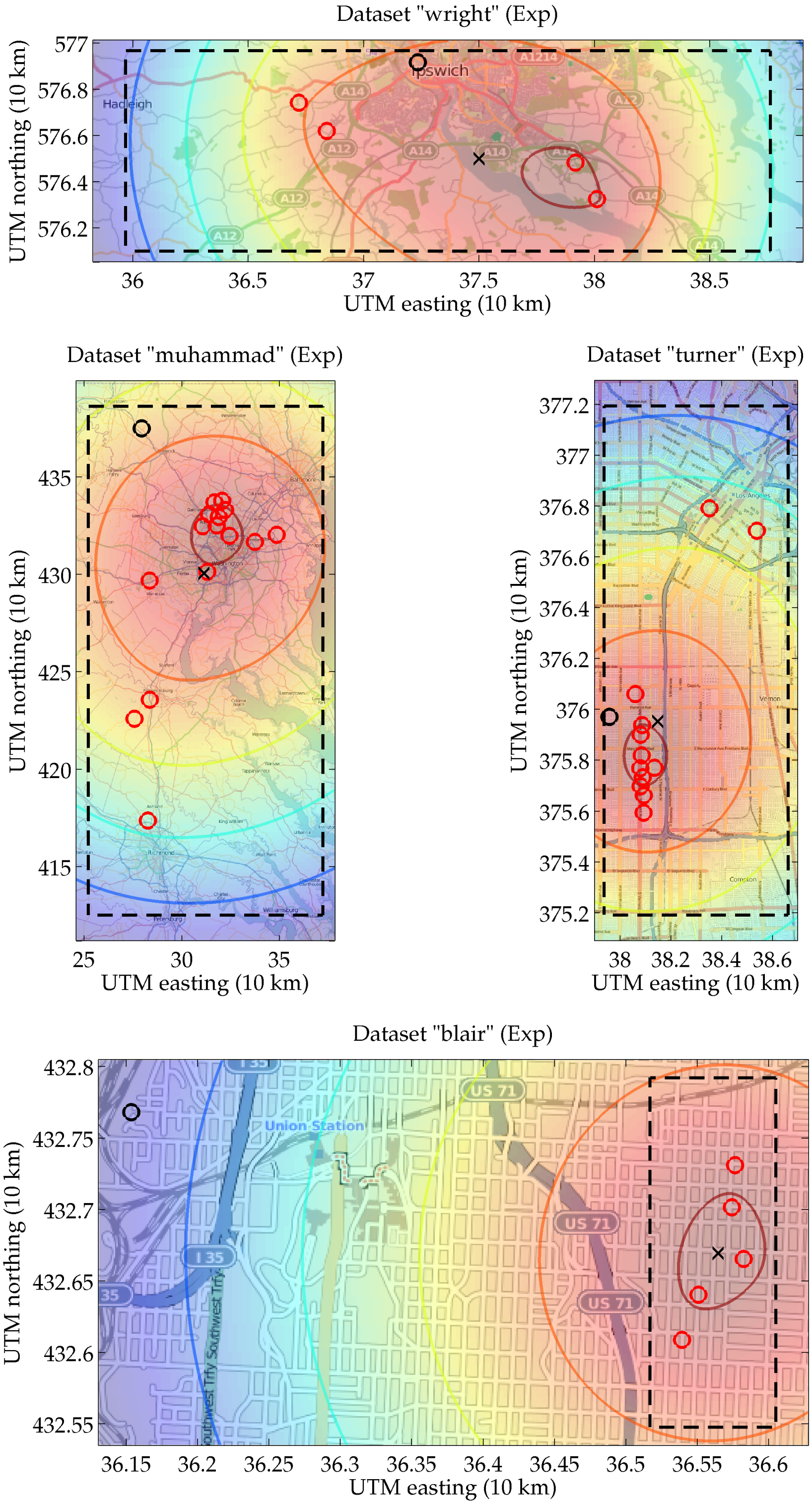


Figure 5: Predicting the residence location in different solved crimes by cross-validating the model with the other models. All these results use the inverse exponential decay function and the Bayesian approach.

5.1 Sensitivity Analysis

The decay function parameter selection may affect our measure of performance, the search cost. The parameters are estimated from the data sample and since the sample size is very small ($n = 6$), there is a lot of uncertainty in the parameters.

As an example, we take the Sutcliffe case that use the exponential decay function, and we examine how much the uncertainty in the parameter λ affects the resulting geographic profile and the search cost. Our maximum likelihood estimate (MLE) with 95 % confidence intervals is $\hat{\lambda} = 1.3746$, $\lambda_L = 1.0969$ and $\lambda_U = 1.7738$. These intervals are presented in Figure 6.

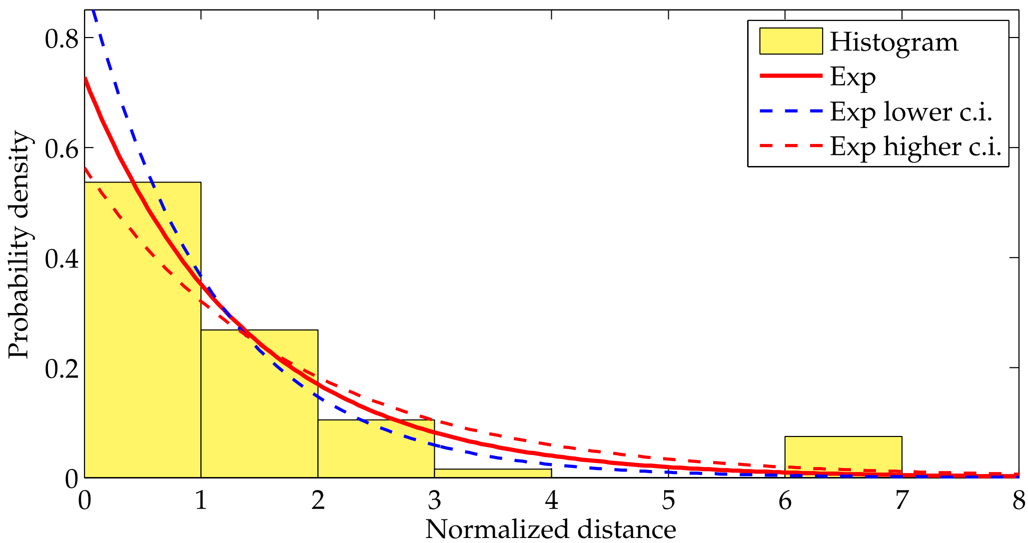


Figure 6: Our maximum likelihood estimate for the parameter λ and the 95 % confidence intervals.

Figure 7 shows the geographical profile using λ_L (Fig. 7(a)) and λ_U (Fig. 7(b)). We notice that λ_L gives a wider peak which is intuitively correct since the plot of the $Exp(\lambda_L)$ is also wider. The peaks are small compared to the other figures since linear instead of logarithmic normalization is used.

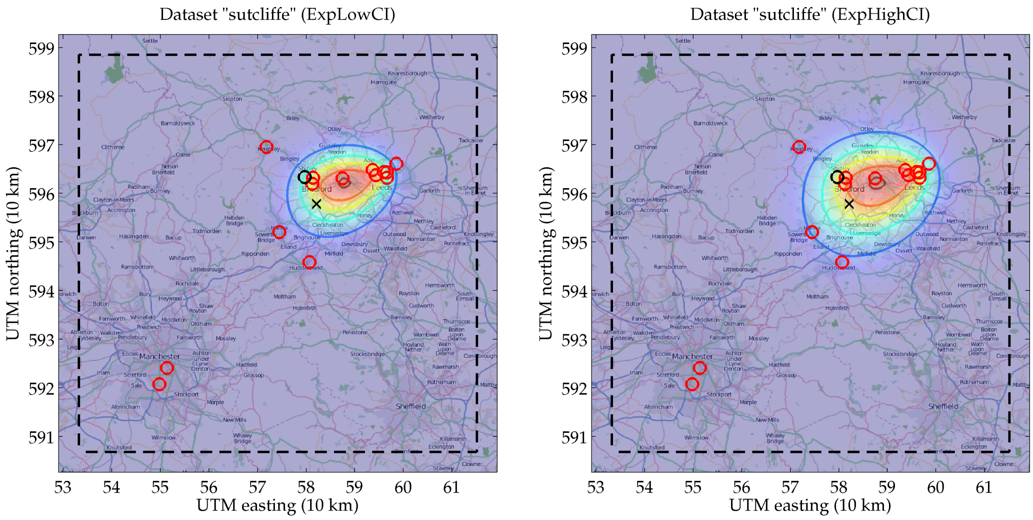


Figure 7: Geographical profile of the Sutcliffe case using the MLE confidence intervals λ_L (left) and λ_U (right). Linear normalization for the color map is used.

An interesting result is that using the exponential decay function, the search cost does not seem to change even though the geographical profile probability density does. That is, the absolute values of the geographical profile height in different grid cells change but their order of magnitude remains the same.

6 Conclusions

Over several decades, the probability distance strategies (PDS) have been the *de facto* method for predicting an offender's anchor point. In the recent years, however, the Bayesian approach has gained more popularity. Given certain assumptions, it turns out that the Bayesian approach reduces into a PDS with only a summation replaced by multiplication.

We have compared the PDS and the Bayesian approach varying the decay function used in both. The Bayesian approach gives consistently better results measured by the search cost and the leave-one-out cross-validation. Of the three tested decay functions, the negative exponential function gives the best results. Yet, the Student's t-test reveals that the differences are not statistically significant. This is an expected result since our dataset consists of only 6 serial killers.

Even though we are not able to determine which approach is better, we find the Bayesian approach more suitable for this problem. The Bayesian approach provides a natural way of handling the prior distributions of anchor points and offenses. Thus, it is able to, e.g., exclude all the uninhabitable areas from the potential anchor points if provided with the appropriate geographical data.

The Bayesian approach also provides a way of predicting the offender's next target based on the previous crime sites. However, the calculation becomes computationally expensive, which is why we are not able to systematically assess the prediction performance. One of the limitations of the model is that it assumes the previous crime sites statistically independent. In practice, this means that the model neglects the time dimension of the previous offenses.

References

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

B. Snook, M. Zito, C. Bennell, and P.J. Taylor. On the complexity and accuracy of geographic profiling strategies. *Journal of Quantitative Criminology*, 21(1):1–26, 2005.

N. Levine. Chapter 10: Journey to Crime Estimation. *CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations. Houston: National Institute of Justice*, 2004.

D. Canter and L. Hammond. A comparison of the efficacy of different decay functions in geographical profiling for a sample of US serial killers. *Journal of Investigative Psychology and Offender Profiling*, 3(2):91–103, 2006.

M. O'Leary. The mathematics of geographic profiling. *Institute for Pure and Applied Mathematics*, 2009.

G.O. Mohler and M.B. Short. Geographic profiling from kinetic models of criminal behavior. *Preprint. Retrieved June*, 15:2009, 2009.

N. Levine. Introduction to the special issue on Bayesian journey-to-crime modelling. *Journal of Investigative Psychology and Offender Profiling*, 6(3), 2009.

E. Alpaydin. *Introduction to machine learning*. The MIT Press, 2004.

J. Warren, R. Reboussin, R.R. Hazelwood, A. Cummings, N. Gibbs, and S. Trumbetta. Crime scene and distance correlates of serial rape. *Journal of Quantitative Criminology*, 14(1):35–59, 1998.

D.K. Rossmo. *Geographic profiling: Target patterns of serial murderers*. PhD thesis, Simon Fraser University, 1995a.

G. Brown, J.L. Wyatt, and P. Tiňo. Managing diversity in regression ensembles. *The Journal of Machine Learning Research*, 6:1650, 2005.

G. Rätsch, A. Demiriz, and K.P. Bennett. Sparse regression ensembles in infinite and finite hypothesis spaces. *Machine Learning*, 48(1):189–218, 2002.

W.L. Gorr. Framework for validating geographic profiling using samples of solves serial crimes. *Unpublished manuscript, Carnegie Mellon University, HJ Heinz III School of Public Policy and Management, Pittsburgh, PA*, 2004.

D.K. Rossmo. Place, space, and police investigations: Hunting serial violent criminals. *Crime and place*, 4, 1995b.

Wikipedia. Peter Sutcliffe — Wikipedia, The Free Encyclopedia. `http://en.wikipedia.org/w/index.php?title=Peter_Sutcliffei&oldid=345391111`. [Online; accessed February 21, 2010].

Iniguez, L. The Crime Scenes. `http://www.latimes.com/news/local/la-me-serial-crimescenes-gr,1,1209938.graphic`. [Online; accessed February 21, 2010].

Nowlin M. and Chaumont K. The 48 victims. `http://seattletimes.nwsource.com/news/local/greenriver/graphics/bodymap06.html`, 2003. [Online; accessed February 21, 2010].

Wikipedia. Beltway sniper attacks — wikipedia, the free encyclopedia. `http://en.wikipedia.org/w/index.php?title=Beltway_sniper_attacks&oldid=345273229`, 2010. [Online; accessed February 21, 2010].

P. Harris. The five troubled victims of the suffolk prostitute slayer. `http://www.dailymail.co.uk/news/article-508727/The-troubled-victims-Suffolk-prostitute-slayer.html`, 2008. [Online; accessed February 21, 2010].

Kansas City News. Police: 1 person likely responsible for all 6 killings. `http://www.kmbc.com/news/3709298/detail.html`, 2004. [Online; accessed February 21, 2010].

Appendix A:
Technical Summary for a Crime Investigator

Predicting the Residential and Next Target Location

Our method can be used to predict the residential location and the next target location of a suspected serial offender. The idea is to narrow down the area where the investigation is carried out. The prediction is based on the location information about earlier victims associated with the offender, e.g., body dump site locations. This location information can be obtained by using a handheld GPS device on the location or web mapping services, such as *Google Maps*.

From the given victim information, an investigation priority map is calculated and drawn over a regular street map. In Figure 8(a), the red area is where the investigation should begin. The more red a particular spot, the higher it should be priorized.

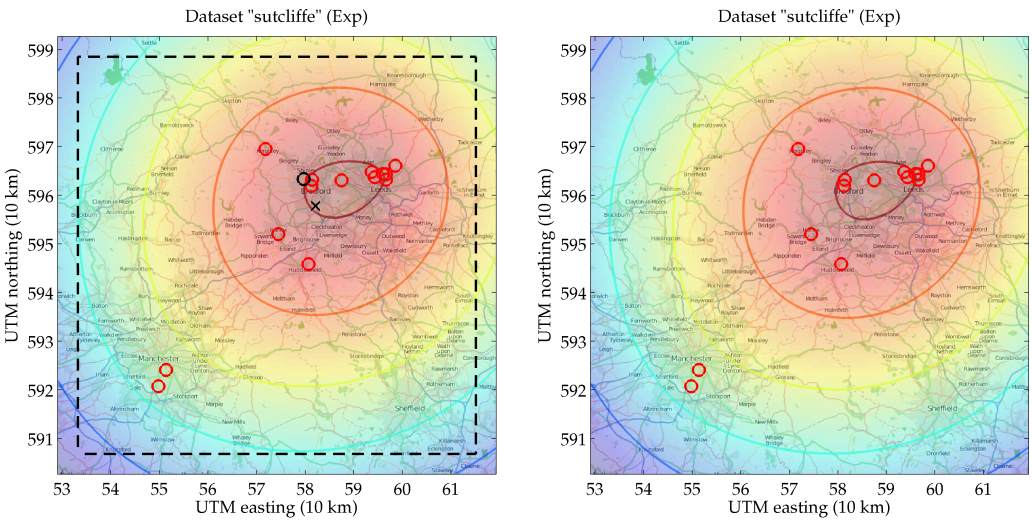


Figure 8: Examples of residential location (left) and next target (right) predictions. Victim locations are visualized with bright red circles. A red value suggests a high priority and blue a low priority.

Similarly, known victim locations can be also used to predict the next target location of the suspected serial offender. The most probable next target area is painted in red and the least probable with blue. See Figure 8(b) for an example.

Important Remarks

Geographical information in the maps is not taken into account by the method. This means that the red area might be above a sea, a lake or otherwise uninhabitable area, and this needs to be taken into account when prioritizing the search. In these cases, the investigation should concentrate on the inhabitable areas near the red or yellow areas. Also, if the previous offenses have clearly occurred on same geographical line, such as in Figure 5(d), the prediction can be rather unreliable. The predications produced by the method are meant to support the investigation process, but not control it.