Modeling and Prediction of Criminal Activity Based on Spatio-Temporal Probabilistic Risk Functions

Paulina Flores[1], Mario Vergara[1], Pablo Fuentes[1], Francisco Jaramillo[1], David Acuña[1], Aramis Perez[1], Marcos Orchard [1]

[1] *Universidad de Chile, Department of Electrical Engineering,Av. Tupper 2007, Santiago, Chile*
pflores@ing.uchile.cl
mario.vergara.l@ing.uchile.cl
pablo.fuentes@ing.uchile.cl
francisco.jaramillo@ing.uchile.cl
davacuna@ing.uchile.cl
aramis.perez@ing.uchile.cl
morchard@ing.uchile.cl

ABSTRACT

Security forces need to model risk patterns associated with criminal activity to study cause-effect relationships and predict new crimes. In this regard, criminal risk models are important to obtain relevant information for better resource allocation and prevention of future crime activity. This paper proposes a method to model and predict future criminal activity based on spatial probabilistic risk functions and a characterization of their temporal evolution as new data become available. This method uses geo-referenced information of public services (e.g., shopping centers, banks) and criminal incidents to approximate the prior risk function as a Gaussian Mixture Model (GMM). Temporal evolution of crime activity is characterized using an algorithm that is based on Sequential Monte Carlo Methods and Importance Sampling. This algorithm incorporates information from new measurements, in a recursive manner, to approximate the posterior spatial probabilistic risk function by updating particle positions in the map. Finally, we propose a novel prediction scheme for criminal activity that uses Gaussian fields centered on hypothetical future criminal events, which are sampled from a GMM that characterizes the spatial distribution associated with recent crime activity. The optimum number of centroids for each Gaussian kernel is evaluated using Silhouette algorithm. The time index related to each hypothetical future crime event is probabilistically characterized using an exponential distribution. Results using real data show that the majority of future events occur within risk modeled zones, information which can be used for resource allocation and improvement of intervention plans.

1. INTRODUCTION

Around the world, order and security forces focus on monitoring criminal incidents that occur in a determined area and time period. Location technologies and services have made it possible to include geo-referenced information, used by analysts to find spatio-temporal patterns in reported events, so as to predict when and where a new crime will occur.

Many techniques and models have been developed to achieve this objective. Among the most used we can find Hot-Spots theories (Eck, Chainey, Cameron, & Wilson, 2005); where criminal incidents are located on a plane, forming clusters that are assumed to be invariant for any prediction horizon. Unfortunately, this technique fails to reflect changes in crime patterns as the environment changes. So as to avoid this problem, more sophisticated statistical models have been developed. For instance, in Liu & Brown (2003) a point-pattern-based transition density model is implemented, which depends on geographic and demographic information. In Xue & Brown (2006) and in Smith & Brown (2007), a model based on spatial decisions has been developed: criminals are assumed to choose places that can be modeled in terms of profit maximization, which depends simultaneously on the gain in committing the crime and the likelihood of being arrested. Other model can be reviewed at Brown, Dalton, & Hoyle (2004) and Rodrigues, Diggle, & Assuncao (2010). The disadvantage of such models is that they do not directly incorporate the temporal component, and when it is modeled using time series (for example), space-time interactions are not considered (Ivaha, Al-Madfai, Higgs, & Ware, 2007). Recent studies have developed some approaches that apply generalized additive models (Generalized Additive Models, GAM) to combine spatial and temporal data, as well as diverse characteristics, for prediction (Wang & Brown, 2012a).

This paper proposes a mixed approach, which considers Gaussian fields and spatial probabilistic risk models based on information of public services associated with an area of interest and criminal event data. The concept of posterior probability distribution is used to yield a spatial notion of crime risk. The time component is incorporated via exponential models and the sequential inclusion of information from new crimes committed within the same area, updating the prior risk distribution. This scheme provides an estimate for the probability of occurrence for a crime in a certain area and time, conditional to nearby public services and events that have occurred in the past.

The article is structured as follows. Section 2 presents a theoretical background focused on concepts such as Gaussian Mixture Models, Important Sampling, Resampling, and Model Evaluation Methods. Section 3 presents the proposed methodology for modeling, temporal characterization, and prediction of criminal events. Section 4 focuses on the analysis of generated results. Finally, conclusions and future work are presented in Section 5.

2. THEORETICAL BACKGROUND

This section presents an overview of the main concepts that will help to understand our proposal for spatio-temporal models, and the associated solution for crime prediction. These concepts include Gaussian Mixture Models, Bayesian Inference and Monte Carlo integration methods such as Particle Filtering, Importance Sampling and Resampling algorithms. Finally, a brief discussion on ad-hoc performance measures is included.

2.1. Gaussian Mixture Models (GMM)

Gaussian Mixture Models (GMM) are parametric probability density functions widely used in the literature due to their capability for approximating multimodal probability density distributions. They are defined as a weighted sum of single Gaussian (Yu, Sapiro, & Mallat, 2011) distributions as stated in Eq.(1):

$$G(x) = \sum_{i=1}^{M} \alpha_i \cdot f_i(x), \tag{1}$$

where x is a D-dimensional random vector, α_i are the mixture weights satisfying $\sum_{i=1}^{M} \alpha_i = 1$, and $f_i(x)$ are multivariate Gaussian distributions of dimension D, given by:

$$f_i(x) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)\right) \tag{2}$$

where μ_i and Σ_i are the mean vector and the covariance matrix of the i-th Gaussian of the mixture.

As mentioned before, GMMs are parameterized by their mixture weights, mean vectors, and covariance matrices. The usual calculation of these parameters, conditional to a data set, is done by applying the Expectation Maximization (EM) algorithm, which is an iterative method that provides maximum likelihood or Maximum a Posteriori (MAP) estimates of these parameters.

There are three types of GMMs, depending on the choice of covariance matrices. In the first case, there is one covariance matrix per Gaussian component (nodal covariance). In the second case, the covariance matrix is the same for all Gaussian components (grand covariance). The third form is a single covariance matrix shared by the overall Gaussian model (global covariance) (Reynolds, 1995). In this paper we use two of these forms, the nodal model covariance and global covariance.

2.2. Bayesian Inference and Monte Carlo Integration

A common problem to be solved when dealing with uncertain dynamical systems is the manner in which statistical inference can be performed. In this regard, Bayesian approaches are well suited, as they provide a general framework for estimating hidden dynamical variables of a system through sequential update. These approaches involve two stages that are executed sequentially after a new measurement is available. The first stage consist of computing a prior probability density function (PDF); task that requires a stochastic representation for state transitions in the dynamic system. The second stage incorporates new measurements into the analysis by correcting the prior PDF through their likelihood, thus yielding a posterior PDF.

Mathematically, the evolution in time of the state sequence is considered as the set of N_x-dimensional vectors $x_{0:k} = \{x_i,\ i = 0, ..., k\}$, and the observable events (or measurements) as the set of N_y-dimensional vectors $y_{1:k} = \{y_i,\ i = 1, ..., k\}$. In Bayesian theory, the posterior distribution defined by $p(x_{0:k}|y_{1:k})$ is decomposed in terms of the prior distribution $p(x_{0:k})$, its likelihood $p(y_{1:k}|x_{0:k})$, and the evidence $p(y_{1:k})$ as:

$$p(x_{0:k}|y_{1:k}) = \frac{p(y_{1:k}|x_{0:k}) \cdot p(x_{0:k})}{p(y_{1:k})}. \tag{3}$$

Under Markovian assumptions, the posterior distribution in Eq. (3) can be expressed recursively (S. a. Doucet, 2001) as:

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k) \cdot p(x_k|y_{1:k-1})}{p(y_k|y_{1:k-1})}. \tag{4}$$

The prior distribution is then decomposed as stated in the Chapman-Kolmogorov equation:

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1}) \cdot p(x_{k-1}|y_{1:k-1}) dx_{k-1} \quad (5)$$

The recurrence relation established by Eqs. (4)-(5) defines the optimal Bayesian solution for the filtering problem. However, this relation cannot be determined analytically in general, and a close-form solution can only be found in a restrictive set of cases (e.g., the well-known Kalman filter for linear, and Gaussian systems). In these cases, assuming a Gaussian distribution with unbiased, and consistent, estimates for the mean and covariance matrix of the prior PDF, the filter can then optimally derive the mean and covariance matrix of the posterior PDF. In nonlinear systems, though, the posterior PDF is not necessarily Gaussian (Arulampalam, Maskell, Gordon, & Clapp, 2002). In this regard, we consider in this article a class of sub-optimal nonlinear Bayesian algorithms that allow better characterization of the posterior PDF in dynamic, non-Gaussian systems: Particle Filters (PF). In PF, the key idea is to represent the posterior density function by a finite set of weighted random samples $\{x_i, w_i\}_{i=1}^{N_s}$ in order to perform statistical inference.

2.3. Importance Sampling

The estimation of the posterior distribution requires to determine the prior distribution, the evidence, and its likelihood. As the likelihood is usually known, and the evidence corresponds to a normalizing constant, the most difficult task correspondsto the computation of the prior PDF $p(x_k|y_{1:k-1})$ described in Eq. (5). However, that expression includes an integral for probability densities that do not have a closed-form in general.

As the Eq. (5) involves the computation of an intractable integral, the idea of a sample-based approximation seems to be suitable. Nevertheless, it is usually hard to sample from the distribution $p(x_k|y_{1:k})$ at any time k. Importance Sampling (IS) (Bergman, 1999) solves this problem by sampling for an alternative PDF, which is known in the literature as *importance distribution* and is denoted by $q(x_{0:k}|y_{1:k})$. The support of this distribution must, at least, include the support of $p(x_{0:k}|y_{1:k})$. Moreover, as the samples are drawn from an alternative distribution a weight is associated to them, and thus

$$p(x_{0:k}|y_{1:k}) \approx \sum_{i=1}^{N_s} w_k^i \delta(x_{0:k} - x_{0:k}^i). \quad (6)$$

The challenge is to compute weights adequately. It is assumed that $p(x_{0:k}|y_{1:k}) \propto \pi(x_{0:k}|y_{1:k})$ is difficult to sample, but $\pi(x_{0:k}|y_{1:k})$ can be evaluated. Let $x^i \sim q(x_{0:k}|y_{1:k})$, $i = 1, ..., N_s$, be samples that are easily generated from the *importance density* $q(\cdot)$ (Arulampalam et al., 2002), then:

$$w_k^i = \frac{p(x_{0:k}^i|y_{1:k})}{q(x_{0:k}^i|y_{1:k})} \propto \frac{\pi(x_{0:k}^i|y_{1:k})}{q(x_{0:k}^i|y_{1:k})}. \quad (7)$$

The art of IS is about choosing the importance distribution $q(\cdot)$ which approximates $p(\cdot)$ as closely as possible. This is the principal factor that affects the performance of this approach (Candy, 2007). Furthermore, if this condition does not hold, a degeneracy phenomenon appears yielding sample impoverishment and thus, undesirable inefficiencies in the method.

2.4. Resampling

A common problem with IS is the degeneracy phenomenon, where after a few iterations all but one particle have negligible weight. In (A. Doucet, Godsill, & Andrieu, 2000) it has been shown that the degeneracy phenomenon is impossible to avoid.

Resampling is a method for addressing the effects of the degeneracy phenomenon in order to reduce them. The basic idea is to eliminate particles with low weights and concentrate on particles with higher weights. The algorithm (Arulampalam et al., 2002) consists of drawing a new set of N_s particles $\{x_k^{i_*}\}_{i=1}^{N_s}$ by resampling (with replacement) from:

$$p(x_k|y_{1:k}) = \sum_{i=1}^{N_s} \omega_k^i \delta(x_k - x_k^i), \quad (8)$$

so that $\mathcal{P}r(x_k^{i_*} = x_k^j) = w_k^j$. The new samples are i.i.d, so they are equally weighted and $\omega_k^{i_*} = \frac{1}{N_s}$. The method is described in Algorithm (1).

2.5. Exponential distribution

Recently, an analysis over real-crime data has demonstrated that the frequency of crime (robbery, thefts, or burglaries) events over a geographic area, and considering a fixed time interval, follows an exponential distribution (Furtado, Melo, Coelho, Menezes, & Belchior, 2008).

Thus, assuming that a criminal event is equally likely to occur in a small time interval (no matter how far or near the last criminal event occurred), then it is possible to model the criminal events rate by an exponential distribution (Blumstein, Cohen, Roth, & Visher, 1986), as it is the only continuous distribution that possesses this memory-less property.

The exponential distribution can be parametrized by its rate $1/\beta$ and is given by:

$$h_\beta(x) = \begin{cases} \frac{1}{\beta} \cdot \exp(-\frac{1}{\beta}x) & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (9)$$

Algorithm 1: Resampling Algorithm

Input: A set of particles with degeneracy phenomenon $\left[(x_k^{j*}, w_k^j, i^j)_{j=1}^{N_s}\right]$

Output: A new set of particles with w_k weight $\left[(x_k^i, w_k^i)_{i=1}^{N_s}\right]$

Initialize the CDF: $c_1 = 0$;
for i=2:N_s **do**
  Constructed CDF: $c_i = c_{i-1} + w_k^i$;
end
Start at the bottom of CDF: i = 1 ;
Draw a starting point: $u_i \sim \mathbb{U}[0, N_s^{-1}]$;
for j=1:N_s **do**
  Move along the CDF: $u_j = u_1 + N_s^{-1}(j-1)$;
  while $u_j > c_i$ **do**
    $j* = i + 1$;
  end
  Assign sample: $x_k^{j*} = x_k^i$;
  Assign weight: $w_k^j = N_s^{-1}$;
  Assign parent: $i^j = i$;
end

where the maximum likelihood estimator of β is:

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^{N} x_i \tag{10}$$

and x_i are i.i.d. samples from $h_\beta(x)$. In this paper, the probabilistic characterization of criminal incidents rates related to a determined area is used for prediction purposes.

2.6. Model evaluation

To evaluate the performance of the proposed risk model, this research effort compares high-probability areas predicted by the model with the number of crimes that actually occur in those areas. A model is said to be good as long as the number of incidents that occur within a fixed time period are proportional to the predicted for that area. The characterization of risk model performance at a time t_j is given by the curve that relates the High-Risk Percentage (HRP_θ) vs. True Incident Percentage (TIP_θ), a method proposed by (Wang & Brown, 2012b) in which:

$$HRP_\theta = \frac{||\{a_i | \mathbb{P}(inci_{a_i,t_j}=1) > \theta\}||}{||\{a_i\}||} \tag{11}$$

$$TIP_\theta = \frac{||\{inci_{a_i,t_j}=1 | a_i \subset \{a_i | \mathbb{P}(inci_{a_i,t_j}=1) > \theta\}\}||}{||\{inci_{a_i,t_j}=1\}||} \tag{12}$$

where $||\cdot||$ is the cardinality of a set, $\theta \in [0,1]$ is a threshold and $\mathbb{P}(inci_{a_i,t_j} = 1)$ is the probability that criminal incidents happen in a area subdivision a_i and a time window t_j. In this case, HRP represents the percentage of high-risk areas predicted by the model, whereas TIP represents the incidents from a test set that took place within the high-risk areas. Both measures are computed for different θ and plotted against each other obtaining a graphic similar to the operating characteristic curve (ROC) (Fawcett, 2006). If many crimes incident take place in high-risk areas, a curve closer to the upper left corner is expected. In the opposite case, a curve similar to a linear relationship is expected.

To measure the model quality, we use the concept of Area Under the Curve (AUC). This area takes values between 0.5 and 1, corresponding to the worst and the best possible cases, respectively.

3. PROPOSED METHODOLOGY

The proposed methodology provides a probabilistic characterization of criminal activity, using for this purpose a set of samples that are distributed in the space (geographic area) accordingly to a risk function. Furthermore, it also includes a mechanism to model the temporal evolution of the sample spatial distribution, where samples are reallocated sequentially as soon as the notification of new criminal incidents is available. Also, a prediction strategy is presented to evaluate the risk level within a specific area and future time period.

3.1. Required Information

For the generation of probabilistic risk models associated with criminal incidents, it is necessary to analyze three types of information sources: (i) definition of a geographic area of interest, (ii) geo-referenced data of public services (e.g., banks, supermarkets), and (iii) geo-referenced data of criminal events. These items are described below.

3.1.1. Definition of Area of Interest (A)

The definition of the area $A \subset \mathbb{R}^2$, for which the probabilistic model is needed, depends on various factors such as: main motivation of the study, processing capacity of the machine, available information, among others.

3.1.2. Geo-referenced Data of services

Within A, it is possible find many services that define places where people naturally gather (e.g., hospitals, schools, parks, supermarkets, pharmacies, among others). As soon as the area A is defined, it is necessary to note the amount of services that are included on it. When the area under study is large and it contains few services (located in a sparse manner), then it would be more difficult to implement the concept of Hot-Spot for criminal risk. However, if the area is very small and it contains several services, then the risk function could be characterized almost as a constant.

3.1.3. Geo-referenced Data of Criminal Events (d_j)

Criminal events are understood as any type of crime that occurred within a time interval $T \subset \mathbb{R}^+$. A probabilistic model requires a considerable amount of data to have statistical validity. For the case in which this paper is framed; this "considerable amount" will depend on the area of interest. The type of crimes for which the model will be generated may vary: they can incorporate many types of crime events (when requiring a general risk model for a particular area), or they can be focused on a special set of crime events (high social impact crimes, such as homicide, burglary, robbery, among others).

3.2. Generation of Spatial Risk Models

For the generation of a probabilistic model that achieves the characterization of the risk associated with criminal activity, it is of paramount importance to relate the current services to the occurrence of criminal events.

This part of the methodology is divided into a sequence of steps, which are listed below and summarized as a flowchart in Figure 1.

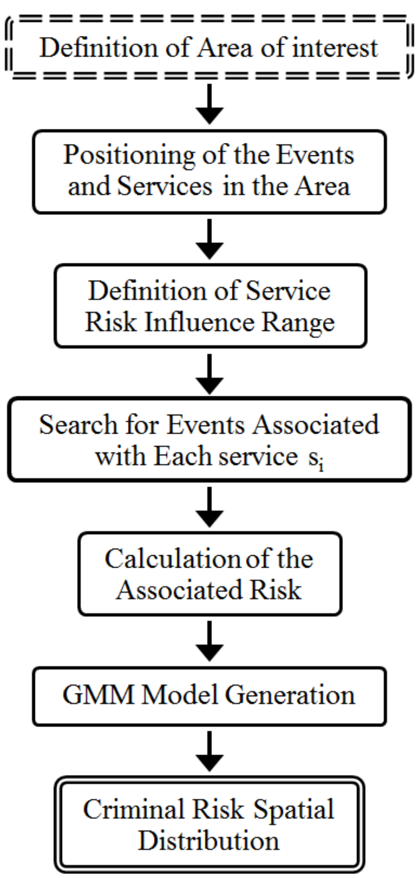


Figure 1. Detailed flowchart of methodology for the automatic generation of risk models.

3.2.1. Definition of Area of Interest

Area $A \subset \mathbb{R}^2$ where services and criminal events that are representative for the case study are defined (Figure 2.A).

3.2.2. Positioning of the Events and Services in the Area

The information of the set of criminal events denoted by $\{d_j\}_{j=1}^{D}$, and the set of services $\{s_j\}_{i=1}^{M}$ that exist in A during the time interval T (Figure 2.B), is provided in terms of spatial positioning via Global Positioning System (GPS) coordinates, preferably.

3.2.3. Definition of Service Risk Influence Range

Let $r_i \subset \mathbb{R}^+$ be the risk influence range associated with a service s_i such that r_i defines the radius of a circle centered on the corresponding service coordinates. They were considered identical for all services (Figure 2.C) in this particular study.

3.2.4. Search for Events Associated with Each Service s_i

Let the location of a crime d_j be given by the coordinates $\vec{x}_d = \left(d_{j_x}, d_{j_y}\right) \in \mathbb{R}^2$, and let the location of a service s_i be given by the coordinates $\vec{x}_s = \left(s_{j_x}, s_{j_y}\right) \in \mathbb{R}^2$. The relationship between d_j and s_i will depend on

$$dist_{ij} = \sqrt{\left(d_{j_x} - s_{i_x}\right)^2 + \left(d_{j_y} - s_{i_y}\right)^2}, \tag{13}$$

where the crime d_j is said to be "associated" with a service s_i if it is fulfilled that $dist_{ij} \leq r_i$. Repeating this procedure for all crimes $\{d_j\}_{j=1}^{D}$, a new set of events that are associated with a particular service will be obtained, defined as $\mathfrak{D}_i = \{\vec{x}_d | dist_{ij} \leq r_i\}$.

3.2.5. Calculation of the Associated Risk

Once the sets $\{\mathfrak{D}_i\}_{i=1}^{M}$ are defined, then the elements of $\mathfrak{D}_i$ are used to adjust the parameters of the i-th component of a GMM that is described in Eq. (2). In fact, if $||\mathfrak{D}_i||$ denotes the number of elements in the set $\mathfrak{D}_i$, then

$$\mu_i = x_{s_i}, \tag{14}$$

$$\Sigma_i = \frac{1}{||\mathfrak{D}_i||} \sum_{j=1}^{||\mathfrak{D}_i||} (x_j - \mu_i)(x_j - \mu_i)^T, \tag{15}$$

where $x_j \in \mathfrak{D}_i,\ j = 1, ..., ||\mathfrak{D}_i||$. Therefore, the risk associated to the i-th service is assumed to distribute following a Gaussian probability density determined by the aforementioned parameters.

3.2.6. GMM Model Generation

As each service has its own risk probability density, the risk function that covers the whole area of services will be de-

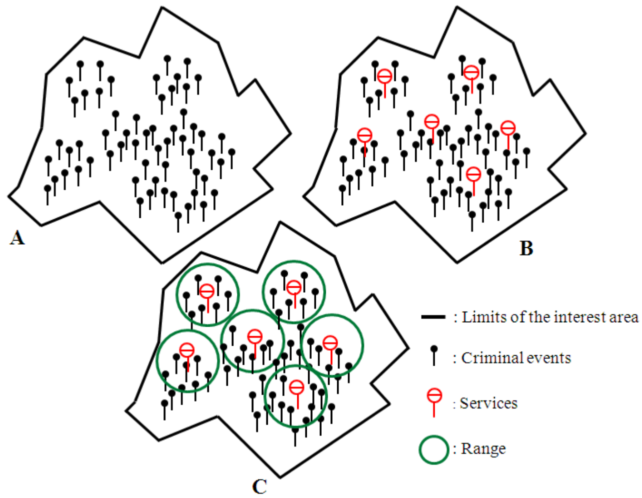


Figure 2. A) Event localization scheme within the area of interest; B) Events and services; C) Events, services and ranges defined for each service.

scribed by a GMM, as stated in Eq. (1). Hence,

$$PDF_{prior}\left(\vec{x}\right) = \sum_{i=1}^{M} \alpha_i \cdot f_i\left(\vec{x}\right), \tag{16}$$

where

$$\alpha_i = \frac{1}{M} \Rightarrow \sum_{i=1}^{M} \alpha_i = 1. \tag{17}$$

Finally, a GMM of equally weighted components is obtained for describing a prior PDF of the criminal risk of a particular area of interest.

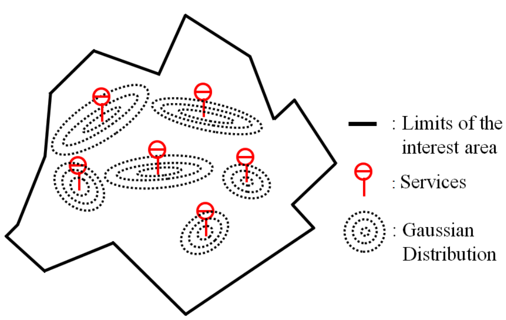


Figure 3. Representation through level curves of a GMM that characterizes the criminal risk.

3.3. Characterization of Temporal Evolution of Spatial Risk Distribution

To allow adaptation (update) of the risk spatial distribution, we propose to represent this distribution by a set of samples. These samples can then be reallocated in space accordingly to new notifications of criminal activity, procedure which can be understood as the computation of a posterior distribution. Additionally, in the absence of new events, the samples may also change their position in order to incorporate underlying uncertainty sources, property that requires a risk prediction model capable of propagate uncertainty along a period of time.

This part of the methodology is divided in two main stages: off-line and on-line, as it is depicted in Figure 4.

During the off-line stage, a set of samples (particles) are arbitrarily located at certain positions and weighted, following the Importance Sampling methodology. Then, a resampling procedure is carried out to obtain a new set of equally weighted samples.

During the on-line stage, a temporary evolution strategy defines a dynamic equation for the movement of the particles according to the inclusion of new observations (sequential incorporation of new criminal events). Therefore, every time that a new criminal notification arrives, some particles will be attracted to the area where the event was reported. Once the positions of the particles have been updated, a posterior distribution is obtained by fitting a GMM (each particle is associated with a Gaussian bi-variate probability distribution).

Finally, a different strategy is required to generate predictions for the evolution of the posterior distribution. A strategy for prediction is presented here which employs the same methodology for reallocating particles used when criminal events were registered. The additional feature of this strategy is that it simulates future criminal activity via sampling procedures, using a GMM denoted GMM_{pred}. This GMM is fitted using only recent criminal activity. The maximum number of step-ahead predictions is defined equal to the number of registered crimes events used for fitting GMM_{pred}. Besides this, the time step between two simulated crimes is obtained by sampling from an exponential distribution, as described in Eq. (9), whose parameter β is set to the average time of data registries that were used to fit GMM_{pred}.

3.3.1. Spatial Risk Distribution

The spatial risk distribution (prior distribution) defined and calculated in Section 3.2 is used (Figure 5.A).

3.3.2. Important Sampling

In order to have a better way to manipulate the spatial risk distribution with lesser computational cost, Important Sampling is used. Then, the prior distribution is empirically approximated using a set of sample (particles).

3.3.3. Resampling

After applying important sampling, it is necessary to ensure the existence of more particles in high risk probability areas, as well as less particles in low-risk probability areas,

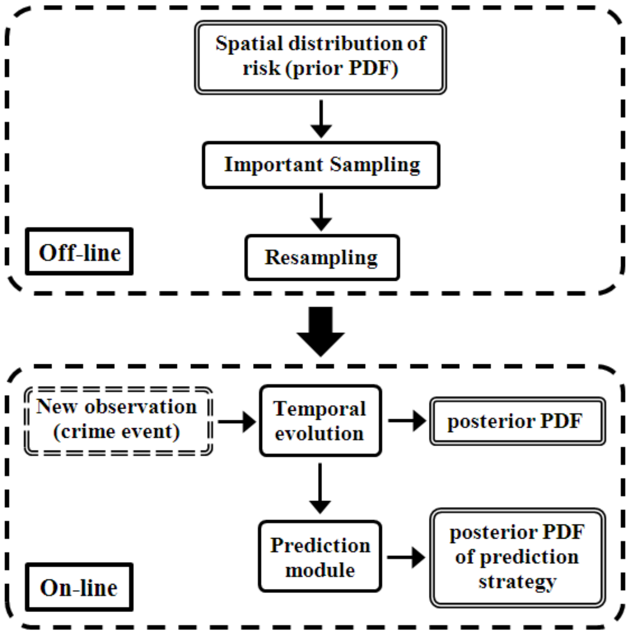


Figure 4. Flowchart of methodology for calculating the posterior PDF and predicted PDF.

Moreover, it is highly desired every particle to have the same weight. Due to this, Resampling (Section 2.4) must be applied (Figure 5.B).

3.3.4. New Observation (Crime Event)

Geo-referenced data of new crime events is chronologically stored and used by the Temporal Evolution module.

3.3.5. Temporal Evolution (Posterior PDF)

Temporal evolution of the posterior risk function is based on the movement of the particles as new observations keep arriving. Thus, it becomes necessary to define a time-varying model to represent the manner in which the particles will move.

The model must meet three requirements: i) Particles located far away from the criminal event should not be significantly affected, ii) Particles located nearby the criminal event should maintain their proximity to it, iii) Particles located at reasonable distance from the criminal event should move towards the observation, since the number of particles located in a determined area is an indicator of the associated risk (Figure 5.C). Following these guidelines, the transition model is defined as:

$$x(k) = x(k-1) + f(d)\left\{y(k) - x(k-1)\right\} + w(k), \quad (18)$$

where:

- $x(k)$: Particle position at k^{th} time instant.
- $x(k-1)$: Particle position at the previous time instant.
- $y(k)$: Observation (crime event) at k^{th} time instant.
- $w(k)$: Process noise.
- $f(d)$: Non-linear function which depends of the distance d between the observation and the particle.

The function $f(d)$ is defined as:

$$g\left(d\right) = \frac{1}{2\pi\left|\Sigma\right|^{1/2}}\exp\left(-\frac{1}{2}\left(\vec{x}-\vec{y}\right)\Sigma^{-1}\left(\vec{x}-\vec{y}\right)\right) \quad (19)$$

$$f\left(d\right) = \frac{g\left(d\right)}{max\left(g\left(d\right)\right)} \quad (20)$$

The covariance matrix Σ is a design parameter and depends of the area of interest.

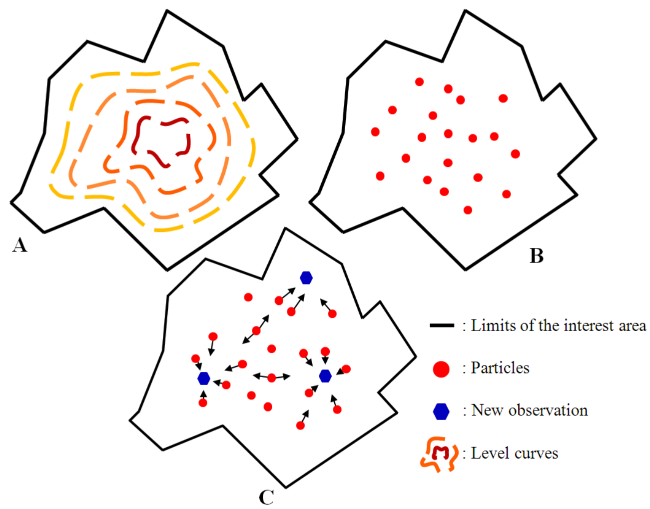


Figure 5. A) Prior distribution seen as level curves; B) Particles after Important Sampling and Resampling; C) Particles move towards the new observations.

3.3.6. Posterior PDF

Once the temporary evolution is implemented using a set of recent criminal activity records, a new GMM can be approximated to obtain a criminal risk spatial distribution. Therefore, each particle becomes to the centroid of a Gaussian bi-variate distribution, and the variance corresponds a design parameter.

3.3.7. Prediction Module

Th prediction strategy assumes that there should not be major changes on the spatial position of the criminal activity in the short term. This is, we consider that the crimes are distributed according to a stationary probability density. Hence, the historical data of recent criminal events is used to update the posterior risk PDF, as shown in Figure 6.A in color blue. The procedure then uses these events to generate a GMM representing the risk PDF associated with recent criminal activity, denoted by GMM_{pred}, as shown in Figure 6.B.

Now, to propagate uncertainty throughout time, future crime events are simulated by sequentially drawing samples from GMM_{pred}, which are coloured in black in Figure 6.C. Future temporal evolution is then characterized by the movement of particles when driven by these simulated events.

To consider the temporary component associated to the prediction (prediction time step, prediction horizon) the time between observations is modeled as an exponential random variable, whose rate estimated as the inverse of the average time registered between criminal felonies observations that were used to fit GMM_{pred}.

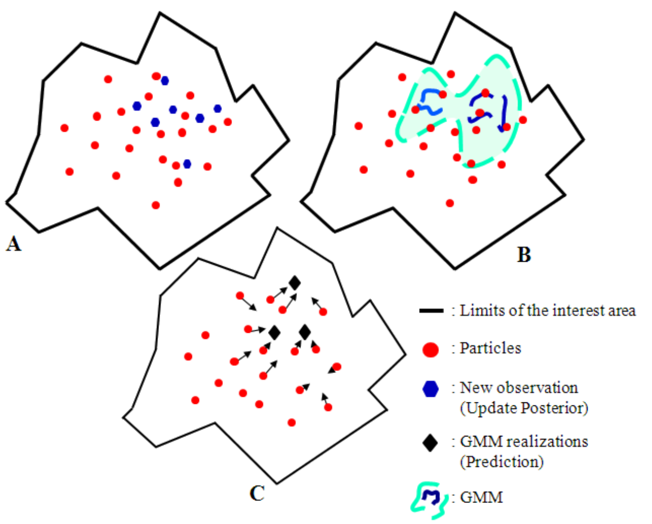


Figure 6. A) Recent criminal activity; B) GMM using recent criminal activity; C) Simulations for future criminal events.

3.3.8. Predicted Risk PDF

Once the prediction stage is finished, a Gaussian kernel is centered at each particle, in the same manner as when characterizing the posterior risk PDF, and a GMM is approximated to obtain a criminal risk spatial distribution.

4. RESULTS

The proposed methodology is applied to an interesting case of study, which considers actual records of criminal activity over a populated urban area. The database includes:

- Geo-referenced information on 4262 public services. These services mainly relate to: stores, banks, bars, fire stations, liquor stores, automatic teller machines (ATMs), police stations, shopping centers, schools, parks, hospitals, clinics, among others. Each record is labeled and has its respective GPS coordinates (latitude and longitude).
- Geo-referenced information of criminal incidents that occurred within the area (23109 records).

For analysis purposes, only robbery offenses with force are considered in this study (crimes that are classified as events with "high social impact"). Among those, 1870 of 2240 records are considered part of the training set, and will be used to characterize the prior risk distribution. From the remaining events, 185 are used to test the filtering stage, and 185 are used to validate the proposed risk prediction approach.

4.1. Prior Spatial Risk Probabilistic Function

Geo-referenced information on 4262 public services is used to generate different prior distributions for the spatial risk probabilistic function, considering for this purpose three values for the service influence range: three, five, and seven blocks respectively; see Figure 7, Figure 8, and Figure 9. Using the AUC measure described in Section 2.6, results show better risk characterization when the influence range is set to three blocks.

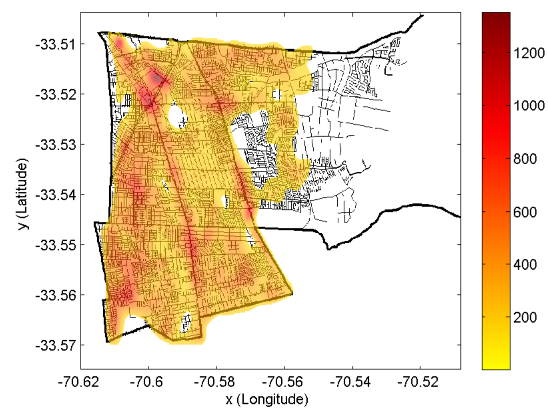


Figure 7. Prior Spatial Risk Probabilistic Function considering service influence range of three blocks.

4.2. Posterior Spatial Risk Probability Function

Once the prior spatial risk distribution is determined, 185 crime events are sequentially used to compute the posterior spatial risk probabilistic function. This procedure, described in Section 3.3.5, requires first to obtain samples from the prior distribution. Figure 10 shows the obtained results when applying an importance sampling and resampling to particles that are first allocated using an arbitrary grid over the area of interest.The grid is generated considering 23 and 16 subdivisions for the horizontal and vertical dimensions, respectively. As a result, after the resampling procedure, some coordinates in the grid contain more than one particle.

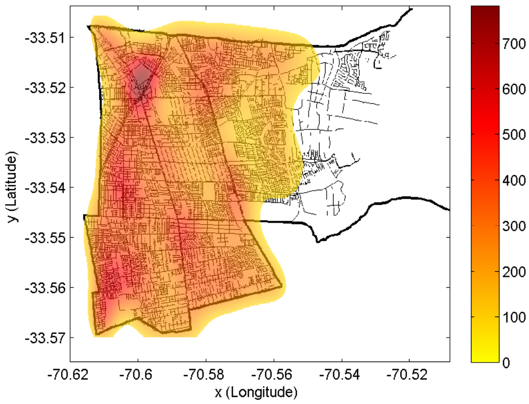


Figure 8. Prior Spatial Risk Probabilistic Function considering service influence range of five blocks.

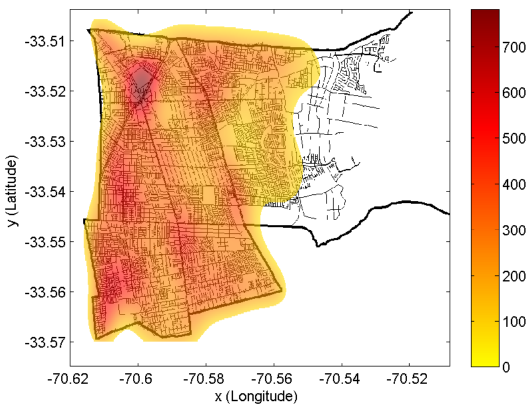


Figure 9. Prior Spatial Risk Probabilistic Function considering service influence range of seven blocks.

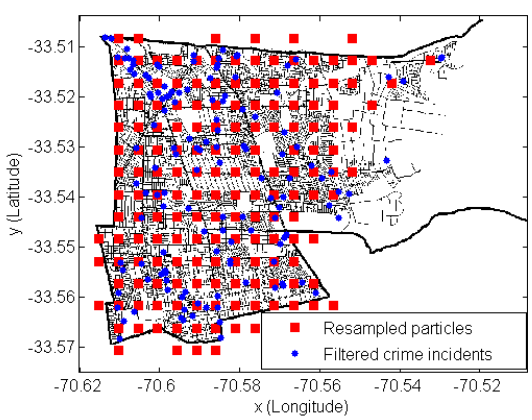


Figure 10. Grid and location of 368 particles associated with the prior spatial risk function. Blue dots indicate the location of 185 criminal events that are used to compute the posterior risk function.

The characterization of the posterior spatial risk probabilistic function is based on the position of particles after incorporating the impact of crime events that are sequentially registered during the filtering stage. The particle movement is governed by the Gaussian attraction field described by Eq. (19) and Eq. (20), where the covariance matrix is given by:

$$\Sigma = \begin{bmatrix} 1.5 \cdot 10^{-5} & 0 \\ 0 & 1.5 \cdot 10^{-5} \end{bmatrix} \tag{21}$$

During the update (or filtering stage), we identify those particles that are significantly affected by the appearance of criminal events. This is done by setting a threshold for the attraction force $f(d)$ equal to 0.6. In this case, 276 particles are in the aforementioned condition. Those are the particles that will modify their position during the prediction stage. This is done to minimize the computation effort associated with the prediction stage.

After sequentially incorporating 185 criminal events into the analysis, the posterior distribution is built by using a GMM, as explained in Section 3.3.6, every particle is used as the centroid of a Gaussian kernel (Figure 11).

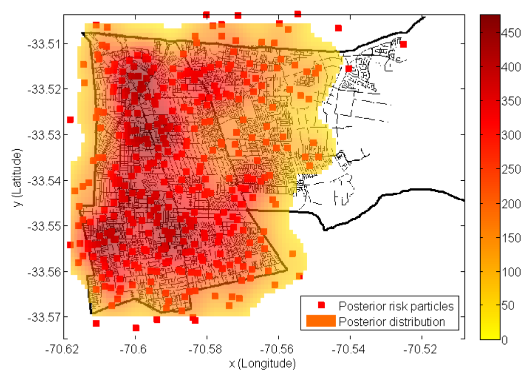


Figure 11. Posterior spatial risk function built from the 368 particles that are significantly affected by the appearance of 185 criminal events. Gaussian kernels used for this purpose consider a diagonal covariance matrix that characterizes a range of influence of three blocks.

The predictive capability of the posterior spatial risk distribution is evaluated using HRP and TIP measures accordingly to Eq. (11) and Eq. (12), for different values of influence ranges and grid sizes; see Table 1.

	1	2	3	4	5	6	7	8	9	10
3	0.934	0.915	0.907	0.902	0.907	0.881	0.885	0.890	0.872	0.861
4	0.925	0.899	0.883	0.883	0.890	0.868	0.860	0.871	0.863	0.859
5	0.922	0.892	0.878	0.879	0.871	0.863	0.848	0.854	0.852	0.854
6	0.911	0.875	0.859	0.861	0.858	0.852	0.843	0.837	0.844	0.841
7	0.915	0.871	0.858	0.851	0.852	0.842	0.835	0.838	0.841	0.846

Table 1. AUC considering [3,7] for particle influence range and [1,10] for grid resolution in the posterior spatial risk function (in blocks).

4.3. Predicted Spatial Risk Probability Function

Once the posterior spatial risk distribution is obtained by using 185 criminal events to update the position of 368 particles (Figure 11), we proceed to generate the GMM Hot-Spot distribution related to recent criminal activity. After applying clustering analysis to recent criminal activity, and testing the number of clusters using the Silhouette algorithm, three clusters as found as the optimal choice for the centroids of the Hot-Spot GMM (Figure 12). This Hot-Spot GMM, associated with recent criminal activity, is used to simulate future criminal activity. This is done by sampling 185 events from the GMM (using a combination of Importance Sampling and Resampling algorithms); see Figure 12.

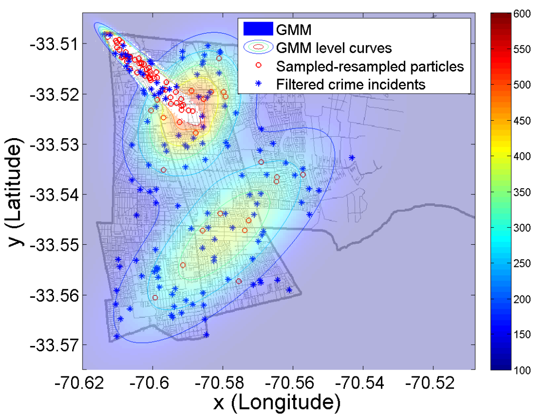
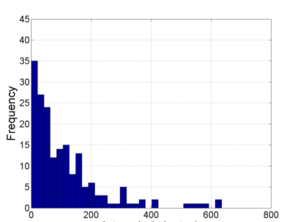
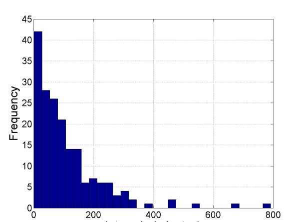


Figure 12. Hot-Spot GMM for recent criminal activity with 3 centroids, and 185 simulated events that are used for prediction purposes.

The inclusion of temporal information related to each simulated crime event is characterized by an exponential distribution with $\beta = 113.09$ [minutes] (Figure 13). This exponential distribution is sampled 185 times to assign time intervals between each simulated crime event (Figure 15).



(a) Histogram for time between recent criminal events



(b) Histogram for time between simulated criminal events.

Figure 13. (a) Histogram is constructed considering recent criminal activity (185 events). This information is used to fit an exponential distribution. (b) 185 samples are obtained from this exponential distribution to obtain time intervals related to simulated crime events.

As a result, the simulated 185 future crime events define a prediction window of 2 weeks approximately. These events are used to modify the position of particles, according to Gaussian attraction fields (Eq. (19)). As a result, the predicted spatial risk function is obtained (Figure 14).

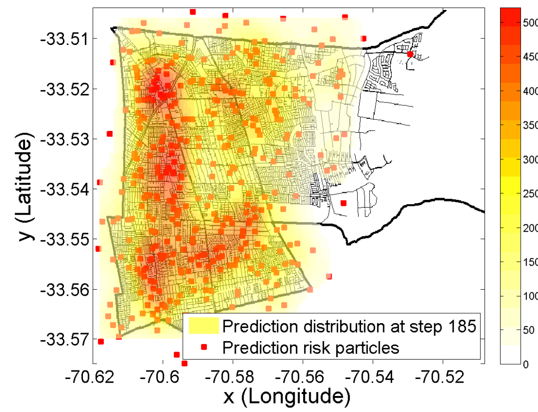


Figure 14. Predicted Spatial Risk Function (GMM after 185 prediction steps).

To evaluate the performance of the predicted spatial risk function, HRP and TIP measures are calculated according to Eq. (11) and (12). In this regard, the AUC is computed assuming different influence ranges for particles (when building the GMM), diagonal covariance matrices, and different grid sizes (measured in blocks); see Table 2. Figure 15 shows the HRP vs. TIP curve with the best AUC=0.93.

	1	2	3	4	5	6	7	8	9	10
3	0.932	0.913	0.906	0.906	0.902	0.89	0.882	0.893	0.884	0.874
4	0.932	0.911	0.901	0.903	0.892	0.881	0.873	0.877	0.876	0.866
5	0.922	0.892	0.881	0.874	0.874	0.865	0.855	0.863	0.86	0.856
6	0.916	0.886	0.868	0.87	0.861	0.855	0.849	0.849	0.849	0.854
7	0.919	0.88	0.86	0.858	0.857	0.846	0.839	0.84	0.843	0.851

Table 2. AUC considering [3,7] for particle influence range and [1,10] for grid resolution in the posterior spatial risk function (in blocks).

5. DISCUSSION

One of the main objectives behind the implementation of sequential approaches for the characterization of the posterior risk distributions is to be able to understand changes in patterns of criminal activity. In this regard, the performance of the "filtering" stage strongly depends on two aspects: (i) sampling strategies, and (ii) the definition of the attraction field that will modify the position of particles as a function of the appearance of new criminal events.

Regarding the former aspect, it is first necessary to determine the impact of sampling strategies when trying to characterize the prior spatial risk function. Although it is possible to use importance sampling and define weights for particles whose location is determined by samples from an uniform distribution (over the area of interest), we propose instead to assign

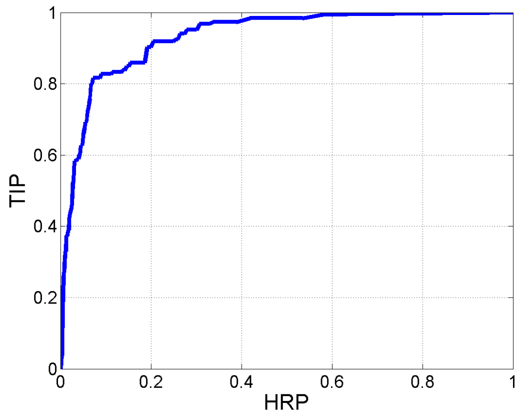


Figure 15. HRP vs. TIP curve, using influence range of 4 blocks and grid resolution = 1 block (AUC=0.93).

weights proportional to the prior risk function to particles that are allocated on an arbitrarily defined grid.

The first problem found when sampling from a bi-variate uniform distribution is to define the amount of particles. A small number of particles results in a scarce representativeness of criminal focuses related to the prior risk function. In our case study, this strategy misses the greatest criminal focus, the one located near the point [-33.52, -70.6] (see Figure 16). Opposite case happens when considering a large number of particles: risk is overestimated because some particles are located in areas where originally there is no criminal risk. Moreover, the processing time for the prediction stage is proportional to amount of sampled particles and thus, the algorithm losses computational efficiency (Figure 17).

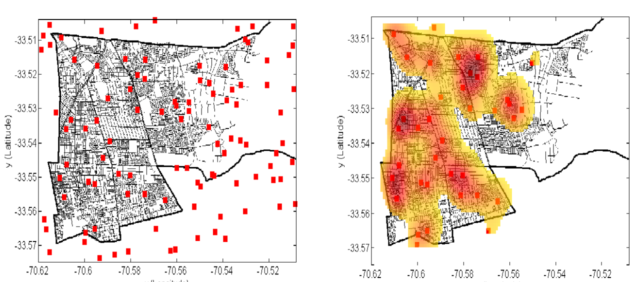


Figure 16. Using 100 particles for importance sampling with a uniform distribution bi-variate and resampling. The prior distribution is generated using particle location as centroids in a GMM and diagonal covariance matrices with influence range of three blocks.

The second problem is the confidence degree associated with samples whose location is generated via uniform sampling. Although the amount of particles may be high enough (avoiding the lack of particles in specific areas, but ensuring reasonable processing times), it is not guaranteed that particles will satisfactorily approximate the original prior distribution. It

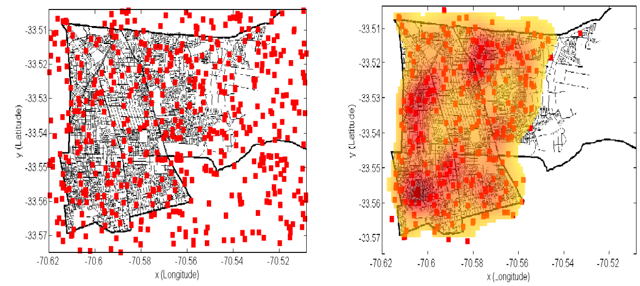


Figure 17. Using 500 particles for importance sampling with a uniform distribution bi-variate and resampling. The prior distribution is generated using particle location as centroids in a GMM and diagonal covariance matrices with influence range of three blocks.

is observed that the implementation of importance sampling strategies (with 370 particles), where the position of particles is obtained from a bi-variate uniform distribution, leads to different prior risk functions performing particle resampling; see Figure 18.

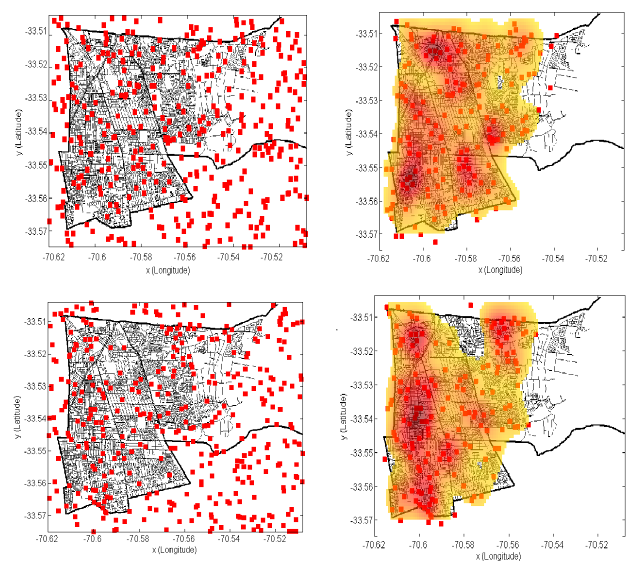


Figure 18. Difference between two estimates for the prior spatial risk function generated using 370 particles and locations that are sampled from a bi-variate uniform distribution, with resampling.

In this regard, the proposed method, where the position of particles is defined by a arbitrary grid and where the weights are assigned proportional to the prior risk function, avoids the two problems described above if resampling is used as a method for ensuring adequate risk characterization. The use of a grid allows to explore the area in a more intuitive manner and, furthermore, resampling allows to efficiently represent criminal risk within the whole area of interest.

The second aspect to be considered in the analysis is related to the definition of the function $f(d)$ given by Eq. (20), where the covariance matrix of $g(d)$, given by Eq. (19), is a design parameter that defines the strength of the particle movement as a function of the distance between the particle and the new criminal event. Our proposal considers that the covariance matrix is diagonal. If the value of diagonal elements is four times greater than the magnitude of the process noise variance, then particles that are located far away from the observation approach abruptly. As a result, after just a few iterations, all particles would converge to Hot-Spot centroids thus losing the capability of uncertainty characterization and representativeness.

On the other hand, if values of diagonal elements in the covariance matrix are equal to the process noise variance, then particles tend stay unaffected by the appearance of new crime events. This two situations provide boundaries that need to be considered in the algorithm design. The matrix given by Eq. (21) provided good results in our case study. This matrix basically represented an influence range of eight blocks around the criminal event.

The implementation of these suggestions allowed to obtain a posterior spatial risk distribution that significantly improves the characterization of criminal activity over time. Using performance measures such as HRP y TIP, Eq. (11) and Eq. (12), it is possible to obtain AUC over 0.9 (Table 1) indicating that the model provides consistent information on the recent crimes that occurred within the area of interest.

In terms of the analysis of future criminal activity, the prediction stage plays a fundamental role. The prediction window (equivalent to the number of prediction steps) should be set accordingly to the number of crimes used for estimating the Hot-Spots distribution. In other words, if 185 crime events are used to compute the posterior risk function, then it is only safe to make predictions between 1 and 185 steps-ahead in time. Although it is possible to make long-term predictions using a risk function solely based on recent criminal activity (in this case, 185 events), these predictions would be biased. The latter, because information associated with Hot-Spot distributions would discard prior knowledge and would be (in that case) based on a much reduced spatio-temporal window.

Regarding the inclusion of the temporal variable in the prediction stage, it is difficult to establish regular time periods between crimes, since criminal activity occur at irregular time intervals. Thus, the temporal analysis shown in Figure 13 is justified. Additionally, to calculate the prediction time, it is only necessary to compute the sum of realizations from an exponential distribution: a method that is simple and computationally efficient. In this case considering 185 criminal incidents, the time between events was satisfactorily characterized using an exponential distribution with parameter $\beta =$ 113.09[minutes], resulting in a prediction window of approximately 2 weeks.

Analyzing the data provided in Table 2 it must be noted that, independently of the influence range associated with each particle, as the resolution becomes smaller the AUC of the prediction model improves. This result is intuitive since it implies that crime could be better predicted if every block is monitored independently. However, police resources are limited, and there is an optimum AUC subject to that constraint. However, there is an optimum influence range in terms of the model performance.

6. CONCLUSION

This article provides a method to characterize the evolution in time of criminal risk in a specific area. A case study with real data that includes location of public services and criminal incidents is also presented. The novel methodology for quantifying risk of criminal events uses a particle-based empirical representation. Two different stages are distinguished in this method: off-line and on-line. The former considers the location of services and 1870 crimes that occurred during a time period of 6 months, yielding a probabilistic characterization of the risk using prior knowledge of historical crimes in the area. The on-line stage approximates the posterior spatial risk distribution using a sequence of 185 new crimes. This task is done by sequentially updating the location of samples (particles), using concepts of importance sampling and resampling. In addition, a strategy for criminal risk prediction is presented. For this prediction strategy, a GMM is fitted using historical registers of recent criminal activity. This GMM is used to simulate 185 future crime events that help to explore the evolution in time of particle positions. Each of these simulations has an associated time of occurrence, modeled by an exponential distribution.

The sequential estimation of a posterior distribution expressed as the movement of samples in the space requires the adjustment of some parameters. One of them is the noise variance (see Eq. (18)), which accounts for the uncertainty in the movement of the samples (it determines how far can a particle move from its initial position on each iteration). For a more realistic reallocation, this hyperparameter should take into account several factors like rate of samples per area and time dependence, to name a few. In this study, each criminal event is considered equally important and thus, each one is given the same importance in terms of the effect on particle movement. Future work will consider different types of crimes in order to cover a wider and more realistic scenario.

For the implemented prediction stage, it is important to consider the temporal analysis of processed criminal incidents, because that allows to generate predictions at irregular time periods; constituting a major advantage over methods where prediction assumes equally spaced time intervals. It is important to note that in this case we were able to generate predictions for two weeks in advance, but it is also possible to provide those results in terms of days, and even police shifts. Also, the prediction can be improved even more considering updates of the Hot-Spot distribution every fixed set of incoming criminal incidents. Finally, we should emphasize the importance of methodologies implemented since they become a good complement to the police, helping in managing its resources to cover the areas of high criminal risk.

ACKNOWLEDGMENT

This work has been partially supported by FONDECYT Chile Grant Nr. 1140774. The work of Pablo Fuentes has been supported by research grant CONICYT-PCHA / Magister Nacional / 2015-22150857. The work of Francisco Jaramillo has been supported by a Ph.D scholarship from CONICYT, Chile. CONICYT-PCHA/ Doctorado Nacional/2014-21140201. The work of Aramis Pérez was supported by the University of Costa Rica (Grant for Doctoral Studies) and CONICYT-PCHA/ Doctorado Nacional/2015-21150121.

REFERENCES

Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *Signal Processing, IEEE Transactions on*, *50*(2), 174–188.

Bergman, N. (1999). Recursive bayesian estimation. *Department of Electrical Engineering, Linköping University, Linköping Studies in Science and Technology. Doctoral dissertation*, *579*.

Blumstein, A., Cohen, J., Roth, J., & Visher, C. (1986). Criminal careers and "career criminals,". *Washington, DC: The National Academies Press*, *2*, 361-364.

Brown, D., Dalton, J., & Hoyle, H. (2004). Spatial Forecast Methods for Terrorist Events in Urban Environments. In H. Chen, R. Moore, D. Zeng, & J. Leavitt (Eds.), *Intelligence and security informatics se - 33* (Vol. 3073, pp. 426–435). Springer Berlin Heidelberg.

Candy, J. V. (2007). Bootstrap particle filtering. *Signal Processing Magazine, IEEE*, *24*(4), 73–85.

Doucet, A., Godsill, S., & Andrieu, C. (2000). On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing,Department of Engineering, University of Cambridge*, *10*, 197-208.

Doucet, S. a. (2001). Sequential monte carlo methods in practice. *New York: Springer-Verlag*.

Eck, J., Chainey, S., Cameron, J., & Wilson, R. (2005). Mapping crime: Understanding hotspots.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern recognition letters*, *27*(8), 861–874.

Furtado, V., Melo, A., Coelho, A., Menezes, R., & Belchior, M. (2008). Simulating crime against properties using swarm intelligence and social networks. *Artificial Crime Analysis Systems: Using Computer Simulations and Geographic Information Systems*, *1*, 300-317.

Ivaha, C., Al-Madfai, H., Higgs, G., & Ware, J. (2007). The Dynamic Spatial Disaggregation Approach: A Spatio-Temporal Modelling of Crime. In *World congress on engineering* (pp. 961–966).

Liu, H., & Brown, D. (2003, October). Criminal incident prediction using a point-pattern-based density model. *International Journal of Forecasting*, *19*(4), 603–622.

Reynolds, D. (1995). Robust text-independent speaker identification using gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing*, *3*(1), 72-83.

Rodrigues, A., Diggle, P., & Assuncao, R. (2010). Semiparametric approach to point source modelling in epidemiology and criminology. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *59*(3), 533–542.

Smith, M. A., & Brown, D. E. (2007). Discrete choice analysis of spatial attack sites. *Information Systems and e-Business Management*, *5*(3), 255–274.

Wang, X., & Brown, D. (2012a). The spatio-temporal modeling for criminal incidents. *Security Informatics*, *1*(1), 1–17.

Wang, X., & Brown, D. E. (2012b). The spatio-temporal modeling for criminal incidents. *Security Informatics*, *1*(1), 1–17.

Xue, Y., & Brown, D. (2006, March). Spatial analysis with preference specification of latent decision makers for criminal event prediction. *Decision Support Systems*, *41*(3), 560–573.

Yu, G., Sapiro, G., & Mallat, S. (2011). Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity,. *IEEE Trans. Image Process*.

BIOGRAPHIES

B.Eng. Paulina V. Flores was born in Santiago, Chile. She received her degree in Electrical Engineering from Universidad de Chile in 2014 . She participated in 1st International Summit on Scientific Criminal Analysis. She currently works in electrical power distribution market, in the design and implementation of fraud and meter malfunction detection models (Chilectra, ENEL Group). Her current interest include NTL Detection, Big Data Analysis and Business Intelligence.

B.Sc. Mario A. Vergara was born in Santiago, Chile in 1992. He received the B.Sc degree in Electrical Engineering at Universidad de Chile in 2014. Currently he is a M. Sc. student at the same institution under Dr. Marcos E. Orchard supervision. His reseach interests include Bayesian signal processing, non-linear filtering and elements of information theory.

B.Sc.Pablo F. Fuentes was born in Santiago, Chile in 1992. He is a student of Electric Engineering from the Faculty of Physical and Mathematical Sciences, Universidad de Chile. He is currently preparing his M. Sc. in fault diagnosis and prognosis, receiving support from Conicyt under Dr. Marcos E. Orchard supervision. He is interested in investigating criminal risk characterization and its temporal evolution within the area of Santiago, Chile.

B.Sc. Francisco Jaramillo received the B.Sc. degree in Electronics Engineering from Universidad de La Frontera, Temuco, Chile, in 2009. Currently he is a doctorate student at the Department of Electrical Engineering at the University of Chile under Dr. Marcos E. Orchard supervision. His research interests include machine learning, control systems, and estimation and prognosis based on Bayesian algorithms with applications to nitrogen removal in pilot-scale Sequencing Batch Reactors for Wastewater Treatment Plants.

B.Sc. David E. Acuña was born in Santiago, Chile. He received his B.Sc.degree in Electrical Engineering (2014) from Universidad de Chile. He currently pursues a M.Sc. degree in Electrical Engineering at the same university and is a Research Assistant at the Lithium Innovation Center (Santiago, Chile). His research interests include signal processing, nonlinear filtering, and prognostics and health management for energy storage devices.

M. Sc. Aramis Pérez is a Research Assistant at the Lithium Innovation Center (Santiago, Chile) and professor at the School of Electrical Engineering at the University of Costa Rica. He received his B.Sc. degree (2002) and Licentiate degree (2005) in Electrical Engineering from the University of Costa Rica. He received his M.Sc. degree in Business Administration with a General Management Major (2008) and also he is a M.Sc. candidate in Industrial Engineering from the same university. Currently he is a doctorate student at the Department of Electrical Engineering at the University of Chile under Dr. Marcos E. Orchard supervision. His research interests include parametric/non-parametric modeling, system identification, data analysis, machine learning and manufacturing processes.

Dr. Marcos E. Orchard is Associate Professor with the Department of Electrical Engineering at Universidad de Chile and was part of the Intelligent Control Systems Laboratory at The Georgia Institute of Technology. His current research interest is the design, implementation and testing of real-time frameworks for fault diagnosis and failure prognosis, with applications to battery management systems, mining industry, and finance. His fields of expertise include statistical process monitoring, parametric/non-parametric modeling, and system identification. His research work at the Georgia Institute of Technology was the foundation of novel real-time fault diagnosis and failure prognosis approaches based on particle filtering algorithms. He received his Ph.D. and M.S. degrees from The Georgia Institute of Technology, Atlanta, GA, in 2005 and 2007, respectively. He received his B.S. degree (1999) and a Civil Industrial Engineering degree with Electrical Major (2001) from Catholic University of Chile. Dr. Orchard has published more than 50 papers in his areas of expertise.