

Received November 8, 2021, accepted December 30, 2021, date of publication January 18, 2022, date of current version February 8, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3143858

# Spatio Temporal Sparsity in Homicide Prediction Models

ÁLVARO J. RIASCOS VILLEGAS<sup>1,2</sup>, JUAN S. MORENO PABÓN<sup>1</sup>, MATEO DULCE RUBIO<sup>1</sup>, SEBASTIÁN QUINTERO<sup>1</sup>, JOHAN GARCÍA VARGAS<sup>1</sup>, AND HERNÁN GARCÍA<sup>1</sup>

<sup>1</sup>Quantil, Bogotá 110221, Colombia

<sup>2</sup>Economics Department, Universidad de los Andes, Bogotá 111711, Colombia

Corresponding author: Álvaro J. Riascos Villegas (alvaro.riascos@quantil.com.co)

This work was supported by Colciencias with resources from the Sistema General de Regalías through the Project “Diseño y Validación de Modelos de Analítica Predictiva de Fenómenos de Seguridad y Convivencia Para la Toma de Decisiones en Bogotá” under Grant BPIN 2016000100036.

**ABSTRACT** Homicide prediction is a challenging task due to the spatio-temporal sparsity of these crime events. In this paper we report the results of using several approaches to mitigate this sparsity condition in machine learning models specially tailored towards modeling homicides events. Since spatial resolution is a direct determinant of sparsity, we focus on the performance of these models across different resolutions of interest to police authorities. We use a simple count model as benchmark and propose some enhancements of it directed towards improving prediction performance. We then compare the results to more complex models motivated by manifold learning and graph signal processing methods. We found that the simple benchmark models are as good as state of the art models for low resolution, but, as resolution increases, the performance of machine learning models outperform the benchmark. These results provide a rationality for the use of state of the art machine learning models for homicide prediction at the high resolution of interest for the deployment of police resources.

**INDEX TERMS** Crime prediction, homicides, signal processing, kernel warping, spatio-temporal sparsity.

## I. INTRODUCTION

Understanding homicides dynamics is challenging due to the particular spatio-temporal distribution of these phenomena. Homicides, compared to other forms of crime, are an infrequent phenomena in time and space. This sparse distribution makes it difficult for statistical models to capture spatio-temporal patterns useful for making predictions.

The goal of this study is to understand the performance of state of the art machine learning models, [1], [2], when the sparsity of the training and test data vary according to the spatial resolution used. For this purpose, we use homicide incidents in Bogotá, from 2018 to 2019, to train simple models that seek to overcome the difficulties encountered in predicting homicides. We then study how these benchmark models compare with the models in [1], [2] as the spatial resolution is varied. This allows us to understand how these more complex models manage to overcome the sparsity issue.

The associate editor coordinating the review of this manuscript and approving it for publication was Anubha Gupta<sup>1</sup>.

All models in this paper are trained and tested on a spatio-temporal discretization of the data. The geographical area of Bogotá is represented by a grid of cells of equal size and fixed side length, while time is divided for all the exercises into weekly periods. The grid size is a parameter that defines the sparsity of the data and plays a crucial role in the capabilities and limitations of the tested models.

We start with a naive and intuitive predictive model that predicts new homicides in places where homicides have occurred frequently in the past. The frequency of homicides in an area of the city is determined by the count of homicides that occurred there during a fixed set of weeks in the past. We refer to this model as the *static count model*. While, by construction, the static count model guarantees the detection of new homicides occurring in locations with a high frequency of homicides in the past, this model cannot identify homicides in locations with a low or null number of previous events.

The lack of predictive capacity of the static count model can be explained by two main limitations. The first one is the timeline used to train the model: since just a fixed set of weeks

is considered in the count model, as time goes by, information from the near past is not taken into account. In order to overcome this limitation, we use the *count model with sliding window* which considers the  $N$  weeks immediately preceding the week to be forecasted. Furthermore, we use the *count model with increasing window*, which always considers the entire past homicide data set. These two extensions of the count model always include the most recent homicide events in the training set.

The proposed modifications have in common that they consider time as a variable in the problem. However, giving equal relevance to all past weeks is counterintuitive: weeks closer to the week to be predicted should be more important than those in the more distant past. Thus we incorporate a *time decay factor* in the count model with increasing window that reduces the weight of the week in the count according to its temporal distance from the week to be predicted.

The second limitation of the count model is the impossibility to predict homicides in locations where they have not occurred before. In this vein, we assume that the hotspots of the count model have an influence on their neighbouring cells. Since homicides and street fights are highly correlated in Bogotá (80% of homicides occur in circumstances of street fights), we add the influence of street fights to the count model, obtaining an *augmented count model*, in the same spirit as [3] and [4]. Adding weekly street fights reports to the model also helps to deal with the sparsity.

While predictive models like the one in [3], using street fights both as a predictive element and as a tool to reduce sparsity, give satisfactory results, the model in [1] presents a more stylized way of using street fights as complementary information. [1] model incorporates the street fights information by optimally warping a kernel density of homicides to a kernel density of homicides plus street fights, following the ideas in [5] and [6]. The ideas proposed in [1] are closely related to manifold learning theory [7].

Another method to predict homicides is presented in [2]. The authors present an approach, based on Graph Signal Processing theory [8] and the proposal in [9] to use the Graph Laplacian of Gaussian (GLoG) as a processing technique to manipulate and understand spatio-temporal data. The way to deal with sparsity in this model is by adding new features from the aforementioned technique and the manipulation of a weekly homicide-induced graph.

The models in [1] and [2] expose different approaches to using complementary information to homicides, providing different views of homicides dynamics and improvements in predictive performance.

Finally, we use different cell-size values to evaluate the impact of sparsity on the predictive performance of the models. When a finer resolution is considered compared to a coarser one, the number of cells without homicides increases. Since the number of homicides for a given week does not change with the resolution, the increase in cells marked with homicides at the finer resolution remains in low proportion compared to the entire set of cells. The size of the cells has

another important consequence in practice: police patrols are only effective when the region to be covered is of a reasonable size (less than or equal to 300 meters).

The results of comparing the above models when the resolution varies reveal interesting conclusions. To compare the performance of the different models we use the Hit Rate (HR) vs. Percentage of Area Covered (PAC) curve. While in low resolutions the count model (and its extensions) is at least equivalent and sometimes even better than the sophisticated models, in high resolutions the performance of the advanced models significantly exceeds the performance of the count model.

## II. RELATED WORK

A large literature has explored spatio-temporal data for urban applications. However, there are few articles focused on homicides and/or the problem of spatio-temporal sparsity. For example, in a recent paper [10], the authors use graph structured recurrent neural networks (GSRNN) to study homicide prediction in the city of Los Angeles. To deal with the spatio-temporal sparsity they build a spatio-temporal weighted graph (STWG) where each node is a zip code region of the city but edges are sparse. Edges are constructed between nodes by solving an optimization problem. In this problem they maximize the log likelihood of the intensity of homicides that resembles a Hawkes process (with exponential kernel) on a directed graph but penalizing for the number of edges in the graph. By doing this, every time they estimate their model to incorporate new data and update predictions, they achieve a sparse spatio-temporal graph representation of homicides. This is the input they use to train their GSRNN reporting good performance.

In [11], the authors address sparsity issues in a crime model in the city of Chicago using hierarchical and multi-task models. In these models they explicitly allow for information sharing across different regions (zip codes) of the city. They show that these models outperform standard models such as Kernel Density Estimation (KDE), when compared according to the hit rate - area flagged as hotspot curve. Other authors have relied on methods that have to deal explicitly with sparse matrices such as recommendation systems. For example, [12] uses an analogy between user (space location) and item (time) to infer the time probability of a crime at a specific location. This is the fundamental idea on which they capitalize and start improving by considering, for example, georeferenced social media data that reflects the context in which crimes occur and potentially the biases inherit in crime reporting. They report results comparable to KDE type of models.

Another potential strategy,<sup>1</sup> is interpreting the crime data as images where each image is a map of the region of interest with homicides events geocoded as pixels. By interpreting the data as images, image and video processing techniques can be used to deal with the sparsity problem (e.g., using super-resolution or standard data augmentation techniques).

<sup>1</sup>We thank this observation to two generous anonymous referees.

For example, in [13] the authors propose a deep generative adversarial network (D-GAN) with external factors such as point of interest, weather data and weekend/weekday to predict taxi and bike demand in New York City. However, although we find this a potentially interesting strategy, our preliminary work in this direction has not been completely successful. In a companion paper we successfully used a cGANs architecture to predict robberies, but the same strategy was not successful when applied to only homicides (where sparsity is indeed a major problem).

### III. METHODOLOGY

#### A. DATA DESCRIPTION

Our dataset contains daily information on the spatial location of each criminal event reported in Bogotá from January 2018 to December 2019, from the Criminal, Contraventional and Operating Information System (SIEDCO) dataset assembled by the Colombian National Police. The dataset was provided by the Security Office of Bogotá. The information on homicides occurring in Bogotá, which is our main event of interest, is reviewed and consolidated weekly through a comparison between different sources including the Metropolitan Police of Bogotá, the National Institute of Forensic Medicine, the Security Office of Bogotá, among others. This process ensures the quality of the data, especially its georeferencing, and seeks to avoid the underreporting of these events.

We grouped our data in weekly periods to deal with a reasonable number of homicides in each period. The training set consists of 78 weeks from January 1st, 2018 to June 30, 2019. The test set is the set of 26 weeks from July 1st to December 29, 2019 (Figure 1).

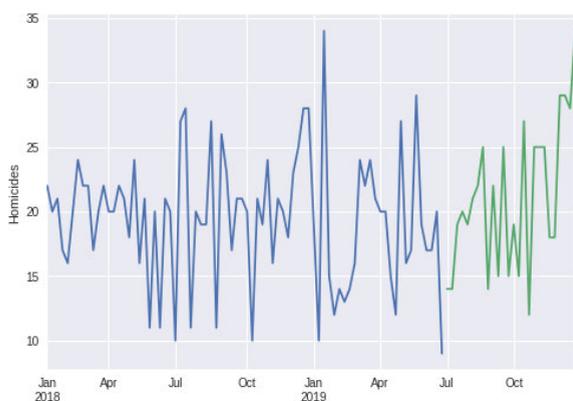


FIGURE 1. Weekly count of Homicides during the train (blue) and test (green) periods.

The area of the city is also discretized by considering a regular grid of homogeneous square cells over the city. For a given week, each cell is then labeled with the number of homicides occurring within it. The size of the cells is controlled by the cell-size parameter, also referred to as resolution. We study how this parameter impacts the performance of the algorithms considered. Figure 2 illustrates how the spatial sparsity of homicides increases at higher resolutions.

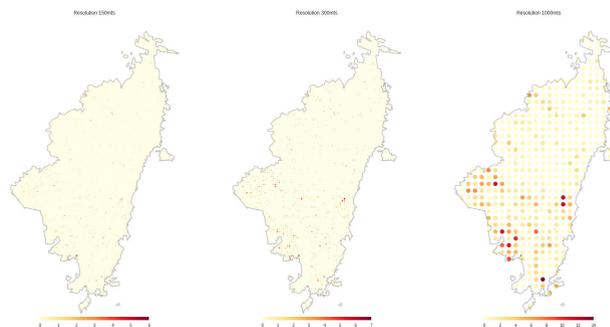


FIGURE 2. Number of homicides per cell during the test period.

#### B. VALIDATION

All models considered assign to each cell a crime intensity score for each time window of the test period (week). These scores are then compared to the actual location of the events that occurred in the test set. In detail, in each time window the cells are sorted according to this score and the first  $x\%$  are marked as hotspots. Then, the Hit Rate is computed as the number of crime events that occurred in any of these predicted hotspots divided by the total number of events:

$$\text{Hit Rate} = \frac{\text{Homicides that occurred on flagged hotspots}}{\text{Total homicides}} \tag{1}$$

Since the Hit Rate depends on the percentage of cells marked as hotspots, we vary the percentage of the area of Bogota flagged by the model as hotspots and produce a Hit Rate (HR) vs. Percentage of Area Covered (PAC) curve. The area under this curve (AUC) gives a metric of the global predictive performance of each model.

In addition, we compute the hit rate and the HR-PAC AUC for specific (low) percentages of area flagged as hotspots. The motivation behind measuring the area under the curve at low PAC values is the practical use of the model: in a city as large as Bogotá with limited resources, police cannot cover large areas of the city permanently. This is why a model with a more concave curve in the initial part of the graph might be more valuable relative to one with a larger total AUC.

#### C. COUNT MODELS

Given a fixed cell-size, we can rank the cells in a one-dimensional array  $C$ . Let  $Tr$  be the training period of time considered as a fixed set of past weeks. For a week  $t \in Tr$  and a cell  $x \in C$  we can consider the *intensity function*

$$\lambda(x) := \sum_{t \in Tr} I(x, t), \tag{2}$$

where  $I(\cdot, \cdot)$  is the *homicides count function*, i. e.,  $I(x, t)$  counts the number of homicides occurring within the cell  $x$  during the week  $t$ . Then, this model labels as hotspots the cells with the highest number of homicides during the training period.

In the following, we describe some variations of this static count model that seek to improve its predictive capacity.

The first one is the *count model with sliding window*. This model considers a time dependent training set  $Tr_{\hat{t},N}$ , with fixed size  $N$ , to predict week  $\hat{t}$ . The set  $Tr_{\hat{t},N}$  consists of the  $N$  weeks immediately preceding  $\hat{t}$ . Then, to predict homicides for week  $\hat{t}$  we consider the *intensity function up to  $\hat{t}$*  using the window of size  $N$  as the function

$$\lambda_{\hat{t},N}(x) := \sum_{t \in Tr_{\hat{t},N}} I(x, t). \tag{3}$$

We further consider an *increasing window* version, where the training set  $Tr_{\hat{t}}$  consists of the entire set of weeks prior to  $\hat{t}$  (instead of only the previous  $N$  weeks). In this case, the *intensity function* results

$$\lambda_{\hat{t}}(x) := \sum_{t \in Tr, t < \hat{t}} I(x, t). \tag{4}$$

The intensity function of the increasing window model gives the same weight to all weeks, regardless of how recent they are. A natural alternative is to consider a model that assigns more weight to more recent data. This motivates the inclusion of a time decay parameter that penalizes weeks far in the past. Consider now the *intensity function up to  $\hat{t}$  with time decay* with parameter  $\omega \geq 0$  as

$$\lambda_{\hat{t}}^{\omega}(x) := \sum_{t \in Tr, t < \hat{t}} \exp(-t\omega) I(x, t). \tag{5}$$

The factor  $\exp(-t\omega)$  ensures that weeks closer to  $\hat{t}$  have more relevance for the prediction.

Another important fact to keep in mind to understand the spatial dynamics of homicides in Bogotá is that about 80% of the homicides occur in the context of a street fight, as mentioned above. Therefore, we include this information in the count model considering the function  $\mu(x, t)$  that counts the street fights in the cell  $x$  during week  $t$ . Then, for a parameter  $\beta \in [0, 1]$  and  $\omega \geq 0$ , we have the *smoothed intensity function with time decay*

$$\lambda_{\hat{t}}^{\omega,\beta} := \sum_{t \in Tr, t < \hat{t}} \exp(-t\omega) (I(x, t) + \beta\mu(x, t)). \tag{6}$$

Note that in all the models above, the prediction step works exactly the same as in the static count model. While these extensions of the count model improve its performance, sparsity remains an issue.

#### D. KERNEL WARPING

The Kernel Warping methodology assumes that the homicide data lives in a subarea (manifold) embedded in the spatial region of Bogotá. Using that about 80% of the homicide events occur during street fights, we assume that such manifold is defined by the location of the historical occurrences of street fights across the city. Thus, to define this manifold, we consider a point cloud  $\mathcal{Z}$  of historic homicides and street fights events present in the training set. These events give an approximation of the manifold on which we wish to warp a Kernel Density Estimation trained on homicides events.

We refer the interested reader to [1] for full details about this methodology for homicides prediction.

We use the adjacency graph defined on the point cloud data as an empirical discrete approximation of the manifold of interest. Specifically, we used a binary adjacency matrix  $A$  if two event are among their  $n$ -nearest neighbors. In a semi-supervised fashion, we use the set of homicide events  $\mathcal{S}$  as the labeled data we want to predict, while the point cloud  $\mathcal{Z}$  includes both the labeled and unlabeled data from which we aim to learn the region where the labeled data lives.

We then construct the graph Laplacian matrix  $L = D - A$  from the adjacency graph matrix  $A$  and the diagonal degree matrix  $D$ , with its diagonal entries equal to the row sum of  $A$ . The graph Laplacian matrix  $L$  gives an empirical discrete approximation of the Laplace-Beltrami operator on the manifold and penalizes differences between adjacent nodes [5].

Finally, using a Gaussian kernel  $k_{\sigma}(\cdot, \cdot)$ , we construct a warped kernel  $\tilde{k}$  following [5], [6], [14], towards the point cloud data:

$$\tilde{k}_{\sigma}(x, s) = k_{\sigma}(x, s) - k_{xz}^T (I + \lambda K_{zz})^{-1} \lambda L k_{sz}. \tag{7}$$

The warped kernel in equation (7) is computed for every  $x$  in Bogotá and  $s \in \mathcal{S}$ , where  $k_{xz} = [k_{\sigma}(x, z_1), \dots, k_{\sigma}(x, z_Z)]$  and  $k_{sz} = [k_{\sigma}(s, z_1), \dots, k_{\sigma}(s, z_Z)]$  are vectors of kernels evaluated at  $x$  and  $s$ , respectively, with respect to the point cloud data  $\mathcal{Z}$ . The matrix  $K_{zz} = [k_{\sigma}(z_i, z_j)]_{i,j}$  is a symmetric matrix of kernels evaluated at all pairs of points of the cloud data, and  $I$  is a  $Z \times Z$  identity matrix. Finally,  $\lambda$  accounts for the degree of deformation: if  $\lambda = 0$  then  $\tilde{k} = k$ , while  $\lambda \rightarrow \infty$  implies  $\tilde{k}$  approaches a positive constant over the point cloud.

Furthermore, we use a temporal exponential decay component to place a larger weight to more recent events. Therefore, the crime intensity for a given point  $x$  in the city is estimated to be:

$$\hat{f}(x) = \frac{1}{|\mathcal{S}|} \sum_{s_i \in \mathcal{S}} \exp(-\omega(t_x - t_{s_i})) \tilde{k}_{\sigma}(x, s_i). \tag{8}$$

To predict homicides in Bogotá we use the parameters that maximize the predictive capacity of the model found in [1]. In detail, we set the number of neighbors to  $n = 7$ , the kernel bandwidth  $\sigma = 0.001$ , the deformation parameter  $\lambda = 10$ , and the temporal decay parameter  $\omega = 0.1$ .

#### E. GLoG

A recent work [9] introduced a promising methodology for the identification of dynamic spatio-temporal patterns. The main idea is to detect abrupt changes in a signal by using methods borrowed from graph signal processing theory. The authors proposed a precise definition of a Laplacian of Gaussian boundary detection filter operating in graph domains and introduced the concept of time-slice entropy, which allows visualizing expected and unexpected spatial phenomena over time. Specifically, [9] presents a tool for visualizing time slices where a signal exhibits unexpected (high entropy) or predictable (low entropy) edge node configurations. We use

these techniques to create features for homicides prediction, with the aim of capturing uncommon and unexpected changes in the number of homicides per unit of time in a node. For a detailed description of the methodology we refer the reader to [9].

1) GRAPH FOURIER TRANSFORM

Consider the graph  $G = (V, E)$ , where the set of nodes  $V$  corresponds to a discretization of the city and the edge set  $E$  captures their spatial contiguity. Then we constructed the graph Laplacian from  $G$  which is a symmetric matrix ensuring a complete set of orthonormal real eigenvectors  $u_l$ ,  $l = 1, 2, \dots, n$ . Then, for a signal  $f : V \rightarrow \mathbb{R}$  (e.g., the number of homicides per node) its Graph Fourier Transform (GFT), denoted by  $\hat{f}$ , is obtained via matrix multiplication

$$GFT = U^T f, \tag{9}$$

where  $U$  is the orthogonal matrix with the eigenvectors  $u_l$  as columns. Similarly, the Inverse Graph Fourier Transform (iGFT) is given by

$$iGFT = U \hat{f}. \tag{10}$$

2) BOUNDARY AND EDGE NODES DETECTION

The boundaries of a signal correspond to the points where abrupt changes occur. The Laplacian of Gaussian (LoG) [15] has been specially developed for identifying boundaries or edges in unstructured data, such as graph signals. This LoG filter relies on two main principles: i) Abrupt changes are presented where the Laplacian of the signal is zero - therefore, these locations are called zero-crossings of the Laplacian; and ii) Using a smoothing filter reduces the noise and provides an optimal location in space - the smoothing filter chosen is the Gaussian filter, since it provides optimal localization in the space and frequency domains [15].

The classical LoG filter can be defined as

$$LoG(f) = \nabla^2 G * f, \tag{11}$$

where  $\nabla^2$  is the Laplacian operator,  $G$  is the Gaussian function,  $f$  is the signal, and  $*$  is the convolution operator which, for two functions  $f$  and  $g$ , is defined by [8]:

$$f * g = iGFT(\hat{f} \cdot \hat{g}) \tag{12}$$

where  $\cdot$  is the element-wise multiplication,  $\hat{f}$  and  $\hat{g}$  are the GFT of  $f$  and  $g$  (defined in Eq. (9)), and iGFT is the inverse GFT. Therefore, the Graph Laplacian of Gaussian (GLoG) filter can be defined as:

$$GLoG(f) = iGFT(\widehat{\nabla^2 G} \cdot \hat{f}) \tag{13}$$

Once the GLoG filter is applied to a signal, the result is another signal defined on the nodes of the graph, which is denoted by  $GLoG_f$ . A pair of adjacent nodes  $\tau_i, \tau_j \in V$  is a zero-crossing pair if  $GLoG_f(\tau_i)GLoG_f(\tau_j) < 0$ . Therefore, the edge nodes are those nodes belonging to a zero-crossing pair. The score  $|GLoG_f(\tau_i) - GLoG_f(\tau_j)|$  can be associated to each zero-crossing pair, the larger the score, the stronger the

signal variation for that pair. Based on this score, weak pairs can be filtered out by considering only zero-crossing pairs whose score is amongst the largest ones. After computing the stronger edge nodes, it is possible to generate a binary signal  $f_e$  where  $f_e(\tau_i) = 1$  if  $\tau_i$  is an edge node and  $f_e(\tau_i) = 0$  otherwise.

3) GLoG FOR BOGOTÁ ON GRAPHS

The GLoG is used to construct features that capture the spatio-temporal dynamics of homicides occurring in Bogotá, which then are fed into a classification model. Specifically, the GLoG methodology is applied to the signal of the weekly number of homicides within each cell in the city.

These sequential features are comprised of: (i) an indicator of whether a node was classified by the GLoG filter as an edge node, (ii)-(iii) the number and the fraction of neighbors that were classified as edge nodes, (iv) the probability of a node of being an edge node, (v)-(vii) the minimum, maximum and mean probabilities among the node’s neighbors, (viii) and the total number of homicides that occurred in the node within a given time window. At each window, the features were calculated using the data of the current time slice and the previous 23 slices (weeks). This corresponds approximately to a six-months time window.

In addition, for each node, we constructed non-sequential features including the number of neighbors, the betweenness centrality, the eigenvector centrality, and the local clustering coefficient. Betweenness centrality of a node refers to the proportion of shortest paths among any two nodes that cross that node. A node  $k$  with a high betweenness centrality means that, it is often the case that to go from one node  $i$  to any other node  $j$  ( $i, j \neq k$ ), it is necessary to cross node  $k$ . Eigenvector centrality measure gives information about how “well” connected a node is. The score of a node is calculated based on the score of its neighbors, meaning that a high score node is itself connected to high scoring nodes. Lastly, the local clustering coefficient is related to the degree to which the nodes in the graph tend to form groups among themselves.

Once the features were constructed, sequences of 24 time slices were created as the input data for a classification model. In summary, the feature set for the predictive model was composed of eight sequential features for each of the 24 time slices plus the four non-sequential features.

IV. RESULTS

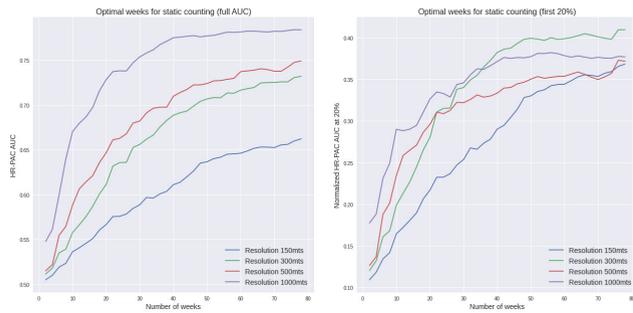
In this section, we present and discuss the performance of the different models considered at several resolutions of interest. The results are presented in two stages. First, we illustrate and discuss the forecasting behavior of the count model and its variants. Afterwards, we use the count model as benchmark to analyze the performance of the Kernel Warping an GLoG models.

We use four metrics of performance: the hit rates at 5% and 10% of area covered, the total area under the HR-PAC curve and the normalized first 20% of that area. The smaller resolution we consider has 421 cells of size 1, 000 meters,

and the largest resolution consist of 18, 712 cells of size 150 meters. We refer to each resolution level by the size of its cells.

**A. ANALYSIS OF THE COUNT MODEL**

In this subsection we study the performance of the count model and its variants at various resolutions. First, the static count model depends on a unique parameter, which is the number of weeks used. Figure 3 shows how the HR-PAC AUC changes when the number of weeks increases; on the left we see the total AUC and on the right the AUC for the first 20% of area covered.



**FIGURE 3.** Area under the HR-PAC curve for the Static Count model against number of weeks at different resolutions.

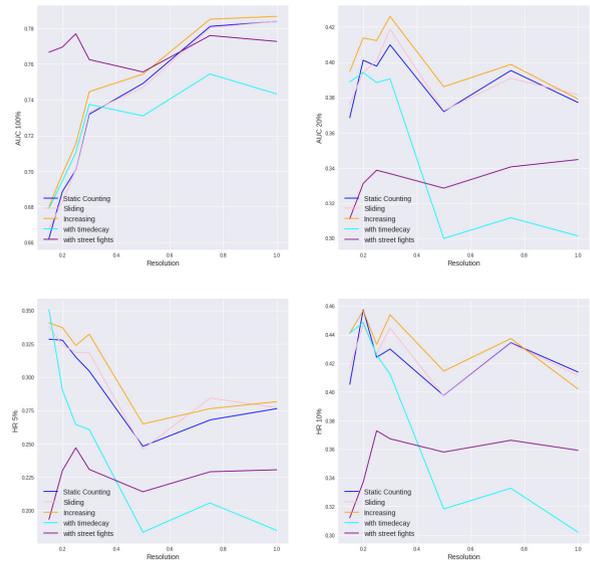
Note that in both graphs all the curves show an elbow when approximately 50 weeks are considered for training. While the grow rate after this point declines, the curves remain increasing.

Comparing across the different resolutions, we observe that for a larger cell size we have a higher total AUC, as expected since the homicide data is in consequence less sparse. However, when considering the AUC for the first 20% of the area covered we see a better performance at intermediate resolutions. This can be explained by the trade off between a larger area covered and a steeper slope at the beginning of the HR-PAC curve. See Figure 5.

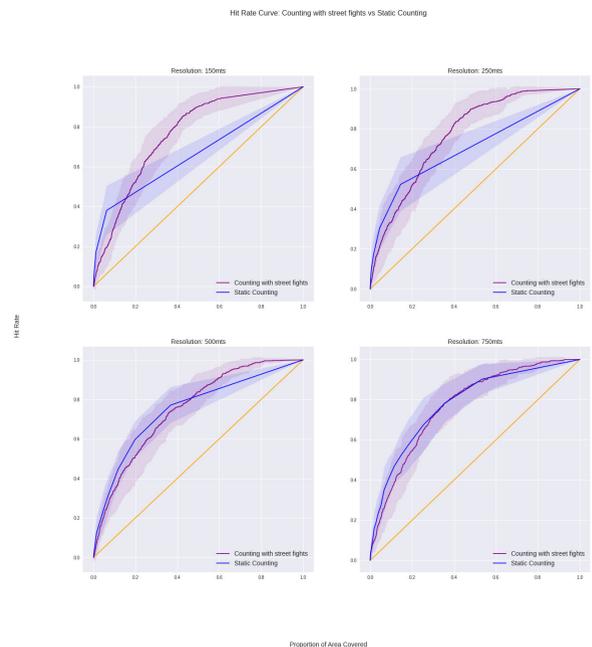
Now, let us consider the effects on performance of the proposed enhancements of the count model. The Figure 4 shows the four performance metrics, in the vertical axis, for each of the models at various resolutions.

Note that incorporating sliding and increasing time windows provides small but consistent performance improvements across all metrics and resolutions, with the greatest improvement in the case of the increasing window. On the other hand, multiplying by a time decay factor gives an improvement only at the finest resolution. At lower resolution, the decrease in performance is noticeable, which again relates to the sparsity of the data.

Moreover, adding the street fights information also hurts the performance of the model, except for the full AUC at high resolution. Figure 5 illustrates this last point. It shows the mean hit rate curves for the static count model with and without adding street fights at four different



**FIGURE 4.** Comparison of performance metrics at different resolutions for the count model and its extensions.



**FIGURE 5.** HR-PAC curves for count model with and without adding street fights in different resolutions.

resolutions. The shaded contour regions correspond to adding and subtracting one standard deviation to the mean curve.

One feature that appears on the HR-PAC curve for the count model is a straight line joining a pronounced elbow of the curve to the end point of the reference line. This elbow marks a natural limit to the performance of the count model, since by counting we cannot predict homicides occurring in cells where no previous homicide has happened. The higher the resolution, the more dispersed the data and the higher

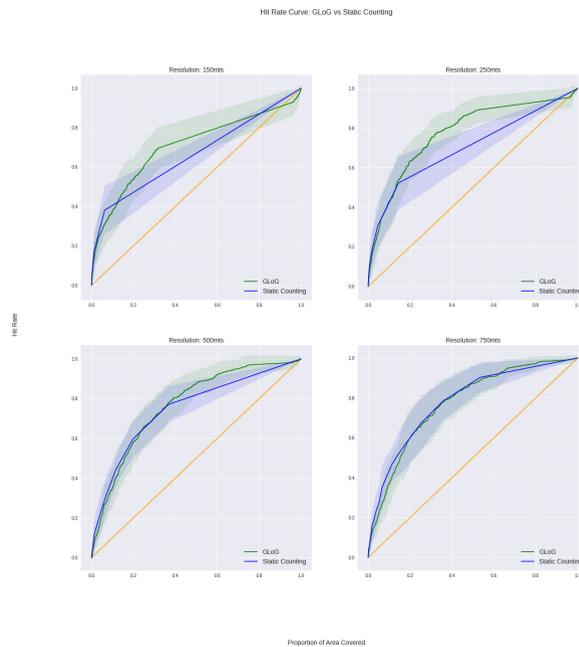
the proportion of cells with 0 homicides. This critical point moves to the left as the resolution increases.

Figure 5 also suggests that enriching our dataset with records of street fights is a promising path to overcome sparsity. While the hit rate for a 5% of area flagged as hotspots remains above 0.4 at all resolutions, it decreases considerably for 20% as resolution increases. The count model has an AUC of 0.78 when the cell-size is 300 meters and drops to 0.69 when the cell-size is 150 meters. On the other hand, the model that takes into account the street fights has a more stable performance at the different resolutions and coverage cutoff points.

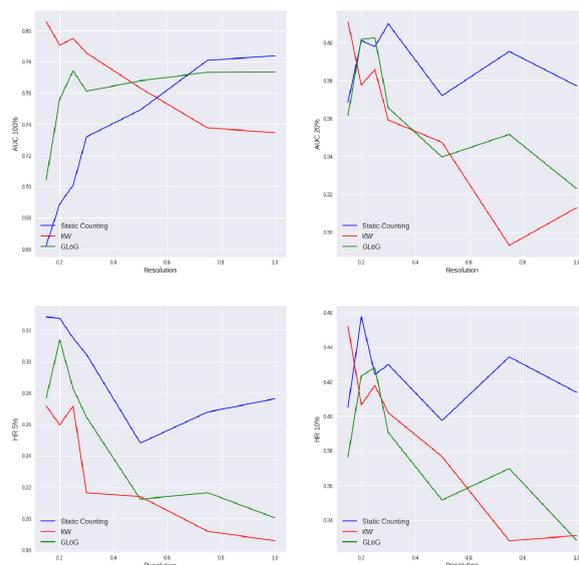
**B. COUNT MODEL AS BENCHMARK FOR STATE OF THE ART MODELS**

Now we proceed to compare the count model to the ones based in the kernel warping (KW) and the Graph Laplace of Gaussian (GLoG) methodologies. The models are trained over the same weeks and their optimal parameters are set as in [1] and [2], respectively.

The comparison of the average hit rate curves with the KW model, Figure 6, shows us the consistency of the KW methodology across resolutions. The HR-PAC curves of both models coincide for the proportions of area covered for which the count model has the best performance. After this critical point, the KW model remains performing well, while the count model fails to predict many of the new homicides. This demonstrates that KW extends the count model beyond its natural limit at high resolution. Intuitively, KW makes an effective use of the street fights data, allowing us to predict



**FIGURE 7.** Hit rate curves for the GLoG model at different resolutions.



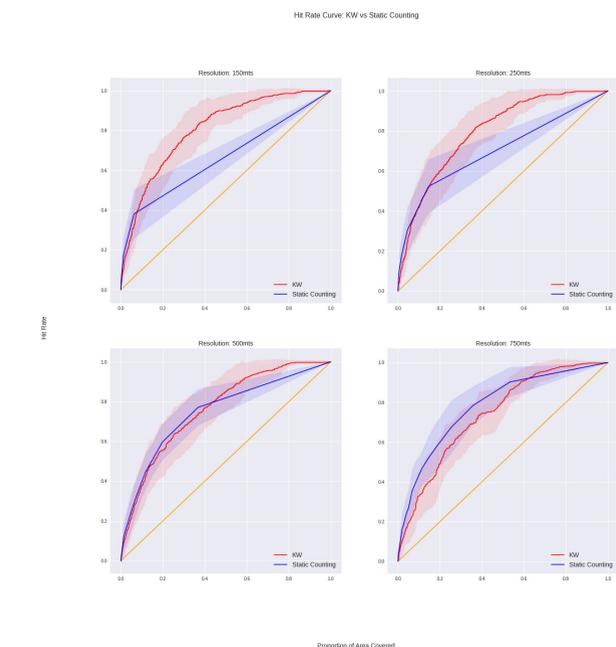
**FIGURE 8.** Comparison of performance metrics at different resolutions.

homicides in locations where this specific type of crime has not occurred before.

Furthermore, the GLoG model obtains its best results when the cell size is between 200 and 300 meters. The Figure 7 shows the mean HR-PAC at various resolutions.

Finally, Figure 7 presents the performance metrics for the three models. It is worth noticing that at low resolution the count model performs better across all metrics, at intermediate resolutions GLoG performs better and at high resolution KW performs better.

Moreover, an interesting feature of the KW model is that its performance for all metrics improves when resolution increases.



**FIGURE 6.** Hit rate curves for the Kernel Warping model at different resolutions.

## V. CONCLUSION

This study evaluates the merits of using state of the art machine learning models for homicide prediction, tailored to mitigate the problem of spatio-temporal sparsity of homicide events. To do so, we studied the performance of the models presented in [1], [2] and compared them to a reference naive model based on the historical count of events. We controlled the sparsity of our data by considering different spatial resolutions, starting with cells of size 1000 meters to cells of size 150 meters. Our results show that, although the naive count models performs well in terms of prediction when the resolution is low, machine learning models outperform the naive models when the resolution is high ( $\sim 150$  mts). These results rationalize the use of complex models for homicides predictions as well as their strategies to overcome sparsity. Our results are important because, from a policy perspective, optimal deployment of scarce police resources should be guided by high resolution spatial strategies.

## ACKNOWLEDGMENT

The opinions expressed are solely those of the authors.

## REFERENCES

- [1] J. S. M. Pabon, M. D. Rubio, Y. Castano, A. J. Riascos, and P. R. Diaz, "A manifold learning data enrichment methodology for homicide prediction," in *Proc. 7th Int. Conf. Behavioural Social Comput. (BESC)*, Nov. 2020, pp. 1–4.
- [2] S. Q. Soto, J. S. M. Pabon, M. D. Rubio, A. J. Riascos, and L. G. Nonato, "Graph restrictions for signal processing of homicides data," in *Proc. Int. Conf. Appl. Artif. Intell. (ICAPAI)*, May 2021, pp. 1–6.
- [3] G. Mohler, "Marked point process hotspot maps for homicide and gun crime prediction in Chicago," *Int. J. Forecasting*, vol. 30, no. 3, pp. 491–497, 2014.
- [4] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita, "Self-exciting point process modeling of crime," *J. Amer. Statist. Assoc.*, vol. 106, no. 493, pp. 100–108, 2011.
- [5] Z. Zhou and D. S. Matteson, "Predicting Melbourne ambulance demand using kernel warping," 2015, *arXiv:1507.00363*.
- [6] S. H. Garrido-Mejía, "Predicting crime in Bogota using kernel warping," M.S. thesis, Dept. Econ., Univ. de los Andes, Bogotá, Colombia, 2018.
- [7] M. Belkin and P. Niyogi, "Semi-supervised learning on Riemannian manifolds," *Mach. Learn.*, vol. 56, nos. 1–3, pp. 209–239, 2004.
- [8] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, Apr. 2013.
- [9] L. G. Nonato, F. P. do Carmo, and C. T. Silva, "GLOg: Laplacian of Gaussian for spatial pattern detection in spatio-temporal data," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 8, pp. 3481–3492, Aug. 2021.
- [10] B. Wang, X. Luo, F. Zhang, B. Yuan, A. L. Bertozzi, and P. J. Brantingham, "Graph-based deep modeling and real time forecasting of sparse spatio-temporal data," 2018.
- [11] M. A. Boni and M. S. Gerber, "Area-specific crime prediction models," in *Proc. 15th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2016, pp. 671–676.
- [12] Y. Zhang, P. Siriwaraya, Y. Kawai, and A. Jatowt, "Predicting time and location of future crimes with recommendation methods," *Knowl.-Based Syst.*, vol. 210, Dec. 2020, Art. no. 106503. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0950705120306328>
- [13] D. Saxena and J. Cao, "D-GAN: Deep generative adversarial nets for spatio-temporal prediction," 2019, *arXiv:1907.08556*.
- [14] V. Sindhwani, P. Niyogi, and M. Belkin, "Beyond the point cloud: From transductive to semi-supervised learning," in *Proc. 22nd Int. Conf. Mach. Learn. (ICML)*, 2005, pp. 824–831.
- [15] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Roy. Soc. London. B, Biol. Sci.*, vol. 207, pp. 187–217, Feb. 1980.



**ÁLVARO J. RIASCOS VILLEGAS** received the master's and Ph.D. degrees in applied mathematics from the Institute of Pure and Applied Mathematics in Rio de Janeiro (IMPA). He has been a Visiting Professor with the University of California in Los Angeles, IMPA in Rio de Janeiro, a Kellogg at Northwestern University, a Visiting Researcher at the International Monetary Fund in Washington, DC, USA, the Cowles Foundation for Economic Research at Yale University, a JP

Morgan at New York; Consultant for the World Bank and the Inter-American Development Bank and, for nine years, and a Researcher at the Economic Studies Sub-Management of the Banco de la República (1996–2005). He is currently a Mathematician with the Universidad de los Andes, Bogotá. He has published in the *European Competition Law Review*, *Economics of Peace and Security*, *Latin American Journal of Economics*, *Journal of Mathematical Economics*, *Journal of Monetary Economics*, *Advances in Theoretical Economics*, *Topics in Theoretical Economics*, *Energy journal*, *Applied Behavioral Statistics*, and *International Journal of Knowledge Discovery in Bioinformatics*, among others. He is the author of the text *Mathematical and Computational Methods in Macroeconomics* (Ediciones UNIANDES) and since the second semester of 2005, he has been linked as a Professor and a Researcher at the Faculty of Economics, Universidad de los Andes. He is the Founder and the Director of Quantil ([www.quantil.com.co](http://www.quantil.com.co)), a mathematics company applied to the industry, the Main Member of the Board of Directors of XM, of the Investment Committee of Asset Management of Grupo Bancolombia, and the Chairperson of the Commission of Applied and Industrial Mathematics of the Colombian Society of Mathematics. His research interests include general equilibrium theory, auction theory, game theory, and mathematics applied to the energy sector, public health, and competition policy.



**JUAN S. MORENO PABÓN** currently works at Pinpoint Predictive, a Stanford StartX that enables insurers to more effectively forecast and influence human behavior. As a Senior Research Data Scientist, he is in charge of a range of cutting-edge AI research and development projects to further advance in digital psychometrics. Previous to joining Pinpoint, he spent four years as a Researcher, and later as the Associate Director of the Data Mining Department, Quantil, a top consultancy

in applied math and machine learning. He led teams of researchers and developed a wide variety of AI-powered solutions across different sectors, including healthcare, public security, banking, transport, education, legal, and manufacturing, among others. He also worked as an Adjunct Professor with the Faculty of Economics at his Alma Mater, Universidad de los Andes, where he's taught courses on discrete math, macroeconomics, machine learning, Python, and R for data analysis. As a Computational Social Scientist, he has researched topics related to crime prediction, fairness in machine learning applications, and segregation in social networks. He also serves on the Board of Directors of the Center of Analytics for Public Policy (CAPP).



**MATEO DULCE RUBIO** is currently pursuing the Ph.D. degree in statistics and public policy with Carnegie Mellon University. He is also an Associate Researcher at Quantil, an applied mathematics company dedicated to the design, development, and implementation of statistical and machine learning models. He worked on the development of crime prediction models in Bogotá to identify hotspots, design optimal patrol routes, prioritize video surveillance systems, and optimally allocate

new police stations. His research interests include the development of crime prediction models that balance accuracy and equity to mitigate the disparate impacts of deployed models on heterogeneous populations. He has experience in text mining and natural language processing, causal inference, survival analysis, spatial analysis, among others, with applications to healthcare, criminal justice, and public safety. He works on the Board of Directors of the Center of Analytics for Public Policy (CAPP), Colombia.



**JOHAN GARCÍA VARGAS** received the master's degree in mathematics from the National University. He is currently pursuing the Ph.D. degree with the Universidad de los Andes. He is also a Mathematician. He worked as a Teacher for more than seven years in universities: National, Antioquia, and the Andes, among others. In pure mathematics, he has an interest in logic, algebra, and category theory; his Ph.D. thesis studies a generalization of Galois theory to quantum groups.

In applied mathematics, he has an interest in statistical analysis, functional programming, image processing, and data mining.



**SEBASTIÁN QUINTERO** was born in London, U.K. He received the B.S. degree in physics from the Universidad Nacional de Colombia, Bogotá, in 2017, where he is currently pursuing the M.S. degree in physics. He was a Research Assistant at the Universidad de los Andes, Bogotá, in 2019, where he worked on a project related to pollution prediction using deep learning models. He is also a Senior Researcher at Quantil, Bogotá. His main research interests include crime prediction and anomaly detection of industrial processes.



**HERNÁN GARCÍA** is currently pursuing the Ph.D. degree in mathematics with the Universidad de los Andes, Bogotá, Colombia. His research interests include optimization, control theory, machine learning, and graph theory. An important part of his research is devoted to the application of polynomial optimization techniques to the solution of optimal control problems. In the field of computational social sciences his works are developed as part of the Data Mining Department at the consultancy firm Quantil. There he is focused on applications of graph theory and machine learning tools to predict and analyze criminal dynamics.

...