

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Time Series from Clustering: An Approach to Forecast Crime Patterns

*Miguel Melgarejo, Cristian Rodriguez, Diego Mayorga  
and Nelson Obregón*

## Abstract

This chapter presents an approach to forecast criminal patterns that combines the time series from clustering method with a computational intelligence-based prediction. In this approach, clusters of criminal events are parametrized according to simple geometric prototypes. Cluster dynamics are captured as a set of time series. The size of this set corresponds to the number of clusters multiplied by the number of parameters per cluster. One of the main drawbacks of clustering is the difficulty of defining the optimal number of clusters. The paper also deals with this problem by introducing a validation index of dynamic partitions of crime events that relates the optimal number of clusters with the foreseeability of time series by means of non-linear analysis. The method as well as the validation index was tested over two cases of reported urban crime. Our results showed that crime clusters can be predicted by forecasting their representative time series using an evolutionary adaptive neural fuzzy inference system. Thus, we argue that the foreseeability of these series can be anticipated satisfactorily by means of the proposed index.

**Keywords:** crime, crime pattern theory, Fuzzy clustering, neuro-fuzzy systems, evolutionary computation

## 1. Introduction

A crime is an event that emerges from opportunities configured by the interaction of offenders, victims, and the surrounding environment [1]. The process behind a crime event has a decisional nature which evaluates the benefits and risks for the offender [2]. Because of the social value in understanding and preventing crime, it has been studied from different perspectives. It has been noted that crime is a complex phenomenon [3] since criminal activity is connected to the complex dimension of social systems and their actors [4].

Environmental criminology recognizes that crime is not uniformly distributed over space, time, or society [5]. Finding rules that explain the non-randomness of crime dynamics has become an intense research area. Typically, stochastic process analysis has been used in the study of crime dynamics [6–8]. However, other epistemological approaches, like fuzzy set theory [9], non-classical topology [10], and complex network theory [11] have been also applied to describe this phenomenon.

Crime pattern theory points out that crime forms patterns in space, time, and society [5, 10]. A pattern is the interconnection between objects, rules, or processes. This interconnection can be either physical or conceptual. This theory deals with the problem of forecasting when and where a criminal event will occur. The observation of patterns can come from evidence or theoretical considerations. Therefore, the analysis of criminal patterns can be described in terms of agents, rules, or clusters of events taking as a reference the structure of the urban form [12].

Crime is ubiquitous in modern cities [13]. It has been observed that there are urban areas with high crime concentration [8]. Therefore, the term space-time dynamics of crime indicates how crime patterns evolve in time and space. A recent work [11] approaches the analysis of crime dynamics by suggesting that robberies are geographically correlated with urban form. A previous work addressed a similar perspective by means of fuzzy topology [10]. In this case, it was observed that crime patterns correlate to the fuzzy edges of neighborhoods, disperse into vulnerable neighborhoods, and concentrate on some main roads.

Some techniques for non-hierarchical clustering have been employed to detect spatial patterns of criminal activity [14], including basic fuzzy clustering (i.e., Fuzzy C-means algorithm) [9]. However, the problems of criminal directionality and crime dynamics using this perspective have been recently addressed in a work by Mayorga et al. [15]. Fuzzy clustering algorithms for spatio-temporal data have been introduced recently. However, these algorithms are mainly focused on clustering of time series (CTS). An enhanced version of the Fuzzy C-means algorithm was proposed to consider both spatial and temporal components of data [16]. This algorithm deals with the clustering of time series produced by spatial sources (i.e., sensors, monitoring stations, etc.). The method is well adjusted to cluster time series that come from a structured sampling of spatial variables. However, when analyzing criminal events, time series provided by sensors are not available, only discrete points of criminal activity in space and time. Thus, the algorithm is not suitable for analyzing this kind of data. Another spatio-temporal fuzzy clustering algorithm was reported in a study by Ji et al. [17]. Clusters are assigned over time series by introducing a switching function that establishes the correspondence between a section of a time series and a cluster in the partition. This algorithm also assumes that time series are available, which is not how spatio-temporal criminal data are collected and studied.

This work deals with clustering the dynamics of criminal events and to forecast it. It contributes by introducing a method that: (1) uses a clustering reorganization algorithm that tracks the dynamics of crime clusters by producing time series of their geometric parameters, (2) sets the optimal number of clusters by minimizing a fuzzy partition index that quantifies the predictability of time series, and (3) forecasts the time series by means of evolutionary-fuzzy predictors. Conceptually, the main contribution of this work is focused on strengthening the time series from clustering (TSC) method in forecasting crime patterns and connecting the quality of a dynamic partition of crime events with concepts of non-linear analysis.

The method is applied over two independent study cases: (1) house burglary in San Francisco, USA and (2) cellular phone robbery in Bogota DC, Colombia. A comparison between the best and worst situations predicted by the introduced index is also provided in each case, giving a preliminary validation. Results reveal that our approach is promising in terms of prediction capabilities, which motivates its application to forecast spatial crime patterns over fine temporal scales. Hence, this method may be considered as a working tool in the practice of predictive policing [18].

The paper is organized as follows: Section 2 describes our method in detail, reviews our clustering organization algorithm, introduces the dynamic fuzzy

partition quality index, and describes the forecasting approach. Section 3 presents our results for the two study cases. Section 4 discusses our main findings and presents some conclusions.

## 2. Method

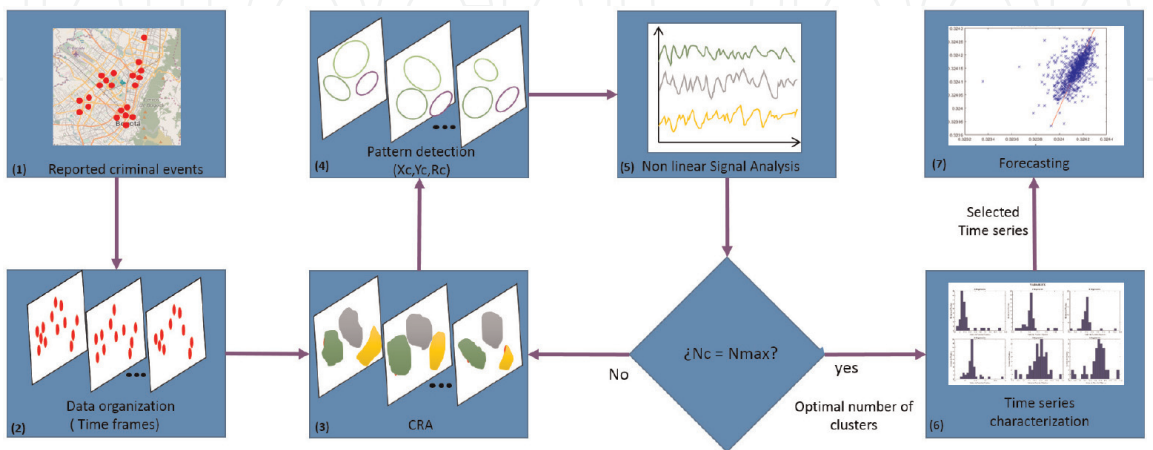
The overall TSC method is depicted in **Figure 1**. Reported spatial events are organized and filtered in frames according to a time scale (i.e., daily, weekly, etc.). A clustering algorithm is applied over these data frames in order to detect and track spatial patterns. These clusters are synthesized in terms of relevant spatial variables computed from each data frame, resulting in several time series. This process is repeated until the maximum number of clusters is reached. Non-linear signal analysis is used to compute a foreseeability index for all time series. The optimal number of clusters is selected as the one that minimizes this index. Selected time series are characterized in order to perform their forecasting. These stages are described in detail as follows.

### 2.1 Data organization

Reported criminal events are grouped in time frames. These events must have a specified date, and longitude and latitude of occurrence to be organized. Time can be defined daily, weekly, monthly, etc., depending on the number of events in the database. Time frames can be generated independently (i.e., without sharing any events) or with some amount of overlapping (i.e., sharing some events).

### 2.2 Clustering reorganization algorithm

Following data organization, clustering of criminal events in a given number of clusters is performed. The result is a set of time series that represents the spatio-temporal dynamics of crime. The number of time series is equivalent to the predefined number of clusters  $C$  multiplied by the number of parameters per cluster. It is important to ensure that these resulting time series have spatial and temporal structure. The clustering reorganization algorithm (CRA) was proposed in [15]. This algorithm ensures that the order of the identified clusters remains the



**Figure 1.** The time series from clustering method uses a clustering reorganization algorithm over reported data to produce clusters of crime events that evolve in time and space. The evolution is represented by means of time series of the clusters' parameters. A designed index relates the number of clusters with the foreseeability of time series by using non-linear analysis.



same throughout the time. In CRA, the clusters are identified by the Fuzzy C-Means algorithm (FCM) [19]. The outline of clusters can also be determined by several other clustering algorithms, such as the Gustafson-Kessel [20], among others. The identification of the clustering algorithm is related to the possible prototypes that can be found in the data frames. These prototypes are related to the urban form where the events take place [11].

The FCM and other algorithms alike initialize their parameters randomly. In this case, the centers of the clusters, their order and the membership values vary from one clustering experiment to another. Because of this random initialization, the clusters identified will not remain in the same zone, nor will the order of the fuzzy partitions created be preserved throughout time.

Therefore, if a study of  $C$  clusters is carried out, there must be certainty that the order of the partitions is maintained. It was considered for this case that the spatio-temporal trends in crime tend to be regular due to the normal population behavior. Thus, if a cluster is identified in a time frame, call it  $C_l$ , then in all future time frames clusters must be identified near that previously identified cluster. Euclidean distance is taken as a basis for the evaluation between the centers of the clusters of the different time frames (Time frame 0 and Time frame  $k$ ). The CRA obtains a time series from clustering in five steps. These steps are: Initialization, Iteration, Distance Evaluation, Discarding non-minima, and Assignment.

First, in the initialization stage, the CRA requires the number of clusters to be set, the  $k$ -th time frame in which the dataset is in, as well as, the center guide matrix. The center guide determines the order of the clusters in the future time frames. As a second step, the CRA ensures that throughout all time frames, the number of clusters is maintained. In this step, the algorithm verifies that for each reorganization process carried out, each center identified in the time frame  $k$  is assigned to a different cluster in the center guide (time frame 0). If the number of clusters is not maintained, this stage is responsible for executing the FCM algorithm once more to re-evaluate the organization of the clusters.

Third, in the distance evaluation stage, the process to measure the Euclidean distance  $d_{ij}^{(k)}$  between the center of clusters identified in time frames  $t_0$  and  $t_k$ , where  $i$  and  $j$  represent the cluster order of the  $k$ -th frame and the first frame, respectively, takes place as follows:

$$d_{ij}^{(k)} = \sum_{j=1}^C \sqrt{\left(x_j^{(0)} - x_i^{(k)}\right)^2 + \left(y_j^{(0)} - y_i^{(k)}\right)^2}, \quad i = 1, 2, \dots, C. \quad (1)$$

where  $d_{ij}^{(k)}$  is the Euclidian distance between the center of the cluster  $j$  in time 0 and the cluster  $i$  in time  $k$ . Based on this distance evaluation, a criterion for the reorganization is established, where the clusters order is assigned according to the minimum distance of each cluster to the centers in the center guide matrix.

Even though clusters will tend to occupy the same zones, in certain cases there may be some clusters that are identified further away from their usual locations. To prevent the CRA confusing the cluster organization, the “discarding non-minima” stage is introduced. In this stage, it is assumed that the  $i$ -th center  $cr_i$  in time frame  $k$  to the closest  $j$ -th center  $cr_j$  in the time frame 0 is selected for the order in which the  $i$ -th cluster is organized.

Finally, the assignment stage takes place once the iteration condition is met. The centers and the membership values have already been assigned in the correct order, according to the identification of the clusters in the first time frame. This stage assigns the correct order to the variables of the centers of the clusters as well as to the membership value of each event. By doing so for each time frame, once the

whole time window is processed there will be a set of time series that represents the spatio-temporal dynamics of the groups that have been identified by the clustering algorithm.

An example of three time series that surrogate the dynamics of a cluster is depicted in **Figure 2**. The spatial dynamics of a cluster is presented as a circle that moves in different frames and whose radius also changes. The dynamics of this cluster is summarized by three time series:  $X(t)$  displacement of the cluster centroid in dimension  $x$ ,  $Y(t)$  displacement of the cluster centroid in dimension  $y$ , and  $R(t)$  radius of the cluster measured as the Euclidean distance from its centroid to the farthest criminal event that belongs to the cluster.

### 2.3 Non-linear signal processing and optimal number of clusters

#### 2.3.1 The MIG index

The memory, information, and geometry (MIG) index is proposed in this work as a scalar that quantifies the predictability of a time series obtained from the CRA. The index is constructed as follows:

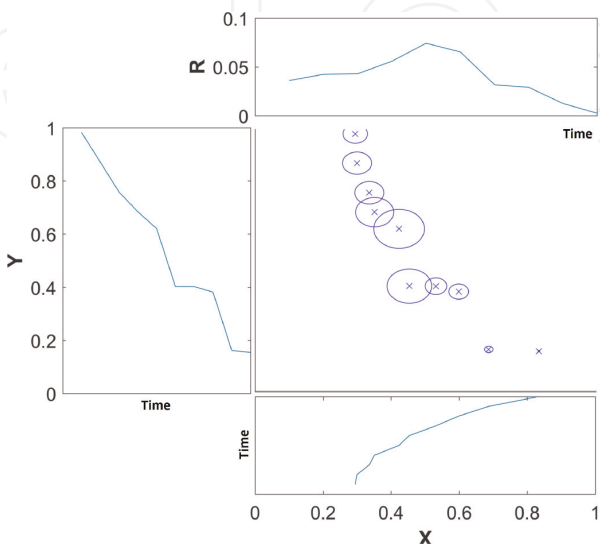
$$q = h^I \frac{G}{M} \tag{2}$$

The predictability of a time series becomes more plausible as the MIG index is minimized.  $M$  evaluates the amount of non-linear correlation inside the time series (i.e., Memory).  $I$  is an indicator of how much information is produced by the signal.  $G$  accounts for the size of the phase space in which the dynamics can be embedded (i.e., Geometry).

In practical terms, the MIG index for a time series  $S(t)$  can be evaluated from statistics grounded in non-linear signal processing and chaos theory as:

$$q_S = q(S(t)) = e^{\lambda} \frac{D}{\tau} \tag{3}$$

where  $\lambda$  corresponds to the estimated Largest Lyapunov Exponent (LLE) of  $S(t)$ ,  $D$  is the size of the embedding space of the series obtained from the estimation of



**Figure 2.**  
Example of time series that describe a cluster dynamics. Parameters of a circular cluster tracked by the CRA evolve in time producing three time series that correspond to: Horizontal displacement  $X(t)$ , vertical displacement  $Y(t)$ , and radius of the cluster  $R(t)$ .

the False Nearest Neighbors in the attractor dynamics of the signal and  $\tau$  is the correlation lag located in the first minimum of the mutual average information of  $S(t)$ . A detailed explanation about the computation of these quantities can be found in [21], while a practical estimation of LLEs is proposed in [22].

### 2.3.2 Optimal number of clusters

Dynamics of the  $i$ -th cluster is represented by three signals (i.e.,  $X_i(t)$ ,  $Y_i(t)$ , and  $R_i(t)$ ), then the MIG index  $q_C$  for the dynamics of a partition with  $C$  clusters is computed as follows:

$$q_C = \frac{1}{C} \sum_{i=1}^C q_{X_i} + \frac{1}{C} \sum_{i=1}^C q_{Y_i} + \frac{1}{C} \sum_{i=1}^C q_{R_i} \quad (4)$$

The optimal number of clusters  $C_{opt}$  is obtained by minimizing  $q_C$ :

$$C_{opt} = \arg\left(\min(q_{C_j})\right) \quad C_j = 2, 3, \dots, C_{max} \quad (5)$$

where  $C_{max}$  is the largest number of clusters considered in the optimization process.

The minimum MIG index  $q_C^{opt}$  suggests that in average the time series produced when  $C_j = C_{opt}$  is more predictable than in any other case. Thus the dynamics should be studied over  $3 * C_{opt}$  signals.

## 2.4 Time series characterization

MIG index is proposed as an a-priori estimation of the predictability of a time series produced by the CRA. The signals that belong to the most and least predictable scenarios should be forecasted so the MIG validity can be corroborated, however it is computationally expensive. Instead, these signals are characterized through several statistics (independent from the MIG index) in order to select some as the most representative signals for each scenario.

### 2.4.1 Characterization

Characterization is performed through a selected group of statistics: the first and second estimated statistical moments of the distribution  $p$  of the series  $S(t)$ , the Shannon entropy and the number of lags beyond which the autocorrelation of  $S(t)$  is effectively zero. The estimators [23] of the two first statistical moments are computed as follows:

$$\mu = \frac{1}{L} \sum_{k=1}^L S_k \quad (6)$$

$$\sigma = \frac{1}{L-1} \sum_{k=1}^L |S_k - \mu|^2 \quad (7)$$

where  $S_k$  is the discrete representation of  $S(t)$  and  $L$  is the number of samples.

The third measure, the entropy described in (8), is related to the information uncertainty inside the series. A high entropy value is interpreted as a low-predictability variable.

$$E = - \sum_{i=1}^L p(S) * \log(p(S)) \quad (8)$$

The autocorrelation function is shown in Eq. (9). The proposed measure represents a confidence bound of the number of regressors with a linear dependence on the signal. It establishes the grade of possible foreseeability in the basis of its periodical behavior as long as forecasting crime is given in terms of stochastic processes [24].

$$r_h = \frac{c_h}{c_0} \quad (9)$$

$$c_h = \frac{1}{L-1} \sum_{k=1}^{L-h} (S_k - \bar{S})(S_{k+h} - \bar{S})$$

$\bar{S}$  the average of the signal. The number of lags  $Lg$  is obtained when  $c_h$  drops below  $0.2c_0$  for  $h > 0$ .

#### 2.4.2 Series analysis

The chosen statistics are applied over time series whereby a four-dimensional vector (Eq. (10)) is obtained per signal:

$$v_S = [\mu, \sigma, E, Lg] \quad (10)$$

where  $S$  is the class identification of the corresponding series:  $X$ —displacement of the cluster in dimension  $x$ ,  $Y$ —displacement of the cluster in dimension  $y$ , and  $R$ —radius of the cluster. These vectors are arranged in matrices as in Eq. (11). In this way, the analysis is done for each matrix separately:

$$M_S = \begin{pmatrix} v_{S_1} \\ \vdots \\ v_{S_C} \end{pmatrix} \quad (11)$$

where  $C$  represents the number of clusters.

- Normalization: the aim is to find the most representative signals (i.e.,  $X, Y, R$ ). The statistics are considered as a way to describe each series, but most of them have a dependence of the scale avoiding a comparison in value. Consequently, a difference of one order of magnitude could be significant in the lags of autocorrelation, but insignificant in the mean. Therefore, normalization is a way to establish a common reference. In this case, the normalization interval is  $(0, 1]$  avoiding the value 0. Each  $M_S$  matrix is independently normalized.
- Finding the average vectors: an average vector is built from the normalized matrices. As per the last step, this process is independently applied. It is important to emphasize that in a determined clustering, there is an average vector  $\bar{v}_S$  for each  $M_S$ :

$$\bar{v}_S = \left( \frac{1}{C} \sum_{i=1}^C v_{s_i} \right) \quad (12)$$



- Determining distances: the distance between each row of the matrices  $M_S$  and the corresponding average vector is computed. Thereby, the Euclidean distance is calculated for all vectors  $v_{S_i}$  as follows:

$$d_{\bar{v}_S v_{S_i}} = \left( (\bar{v}_S - v_{S_i})(\bar{v}_S - v_{S_i})^T \right)^{1/2} \quad (13)$$

The matrices  $D_S$  are organized by collecting the distances calculated from the matrices  $M_S$ .

$$D_S = \begin{pmatrix} d_{\bar{v}_S v_{S_1}} \\ \vdots \\ d_{\bar{v}_S v_{S_C}} \end{pmatrix} \quad (14)$$

- The fittest vectors: finally, the series corresponding to the position of the minimum value in each matrix  $D_S$  is chosen as the most representative signal. Thus, there are three representative signals obtained from  $D_X$ ,  $D_Y$ , and  $D_R$  which are referenced as  $X_f(t)$ ,  $Y_f(t)$ , and  $R_f(t)$ , respectively. These three signals are taken in the forecasting stage as representation of the clustering dynamics for a given number of clusters. Particularly, the representative signals for the best and worst conditions of the MIG index are considered in this work.

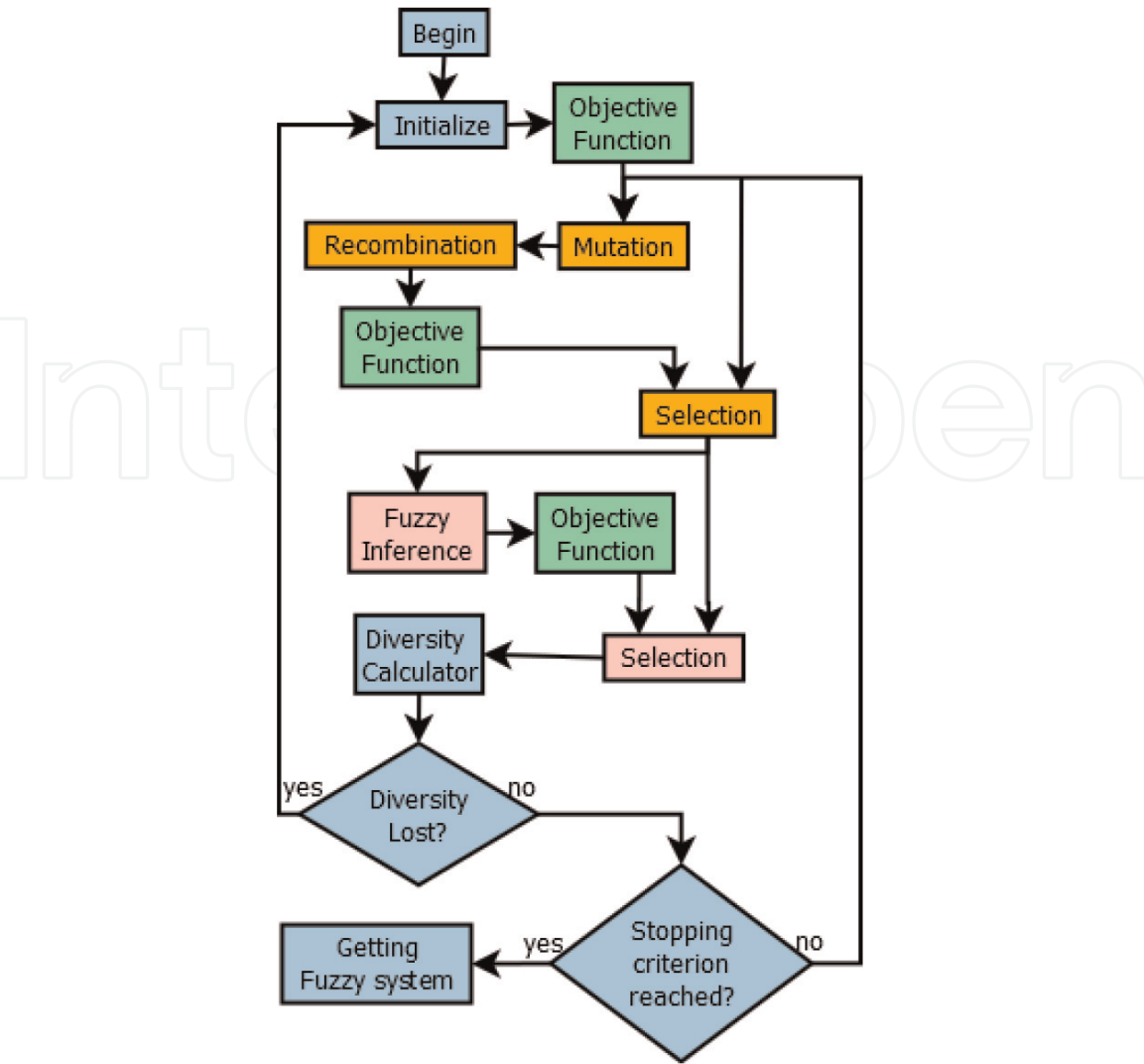
## 2.5 Evolutionary-fuzzy forecasting of representative time series

Forecasting of representative time series is carried out by means of a custom memetic algorithm [25, 26]. The heuristic was proposed in [27] as the combination of the differential evolution algorithm [28] and the adaptive neuro-fuzzy inference system (ANFIS) [29]. The general flow graph of the algorithm is shown in **Figure 3**.

### 2.5.1 Differential memetic neuro-fuzzy algorithm

A memetic algorithm allows to take advantage of both global and local search. In this manner, the optimization space can be explored widely and deeply. The differential evolution algorithm (highlighted in orange) is chosen as the global optimizer, on account of its use for optimizing multidimensional real-valued functions. In addition, it has been successfully applied for model optimization of complex systems [30]. Several variants of this algorithm have been proposed in literature, taking into account the multiplicity of elitism strategies, chromosome representations, and mutation operators. ANFIS (highlighted in pink) is implemented as the supervised learning strategy of the memetic algorithm. It uses the gradient of the objective function on the basis of its differentiability to search for solutions around a locality of the optimization landscape. ANFIS is supported on Sugeno-type fuzzy systems with gaussian membership functions in the inputs [29, 31]. In addition, the proposed fuzzy system has  $m$  rules, which also corresponds to the number of membership functions in each input and the amount of singleton fuzzy sets in the output. Each rule relates the  $r - th$  group of membership functions in the inputs with the  $i$ th singleton value in the output [30].

As shown in the flowgraph in **Figure 3**, when the population is adapted by ANFIS, it is necessary to re-evaluate the solutions and later, according to the fitness



**Figure 3.**  
The evolutionary algorithm used to tune fuzzy forecasters. This algorithm uses the script of the differential evolution algorithm with a step of local search provided by an adaptive neural fuzzy inference system.

function, select them. This re-evaluation must be carried out since ANFIS relies on the RMSE to evaluate the solutions but the memetic algorithm relies on the fitness function. Thus, an interesting individual for ANFIS may not necessarily be good according to its fitness score.

In order to guarantee population diversity, and avoiding a fast-convergence caused by ANFIS, the diversity calculator considered is the FANO factor [26], which calculates the average variation of the mean in the membership functions according to Eq. (15). The population is re-initialized if the FANO factor is greater than a given threshold:

$$\hat{F} = \left( \frac{1}{n * m} \sum_{i=1}^{n * m} \frac{\sum_{j=1}^j (D_{fi} - \bar{D}_i)^2}{\bar{D}_i} \right) * \frac{1}{gen} \quad (15)$$

where

$$\bar{D}_i = \frac{1}{j} \sum_{i=1}^j D_{li}$$

And  $n * m$  is the number of membership functions in the inputs,  $D_{li}$  is the average of the  $i$ -th membership function of the  $l$ -th individual, and  $gen$  is the current generation.

### 2.5.2 Fitness function

The root mean squared error (RMSE) index used by ANFIS may not be appropriate for guiding the global search of an evolutionary algorithm [30] while dealing with complex behaviors. Thus, a problem-oriented fitness function to evaluate candidate solutions must be designed. This function is significant because it provides the criteria to judge a solution and its performance. In addition, it is imperative to include several performance indicators to check solutions in a more precise and proper manner. This process is carried out by minimizing the following expression:

$$f = \frac{(1 - NSE) * MAE}{POCID} \quad (16)$$

The fitness function is composed of three indexes. Each one has been chosen to guide the evolution regarding features in the best solution. Thus, many possible solutions that do not have a near-zero mean error but can forecast the series properly, will be explored.

The Nash-Sutcliffe efficiency (NSE) is conceived as an overall performance measure. It is a normalized statistic that determines the relative magnitude of the residual variance compared to the measured data variance [32]. A value of 1 corresponds to a perfect match of the modeled  $\hat{d}$  to the observed data  $d$ .

$$NSE = 1 - \frac{\sum_{k=1}^N (d_k - \hat{d}_k)^2}{\sum_{k=1}^N (d_k - \bar{d})^2} \quad (17)$$

The mean absolute error (MAE) represents a clear interpretation of the absolute difference between the series and its forecasting, relative to the series [33]. It is chosen instead of RMSE because the series has several peaks and it is necessary to rest importance on them in the forecasting. If it was not done, the fitness function would penalize errors in the lowest values (i.e., near the mean) more than in the highest ones.

$$MAE = \frac{1}{N} \sum_{i=1}^N \left| \frac{d_k - \hat{d}_k}{d_k} \right| \quad (18)$$

Prediction of change in direction (POCID) is a non-linear index included to ensure that a forecast that does not fit the series well, but follows its direction changes reliably, gets a good score [34]. As a multiplicative inverse, it causes an increase in the fitness function as large as how far the fitness score is from the optimal value.

$$POCID = \frac{\sum_{k=1}^N B_k}{N} \quad (19)$$

where

$$B_k = \begin{cases} 1, & \text{if } (d_k - d_{k-1})(\hat{d}_k - \hat{d}_{k-1}) \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (20)$$

$NSE = 1$  and  $MAE = 0$  represent the two roots of  $f$ , hence these are enough reasons to consider a solution as the best, but the POCID operates as a non-linear penalizer.

### 3. Results

The method outlined in the previous section was applied to forecast criminal patterns in two cities: San Francisco, USA and Bogota, Colombia. Results obtained from evolved fuzzy predictors are described in this section. In addition, evidence is presented from these cases about the validity of the MIG index for setting the optimal number of clusters of a dynamic partition of crime events.

#### 3.1 Results for San Francisco, USA

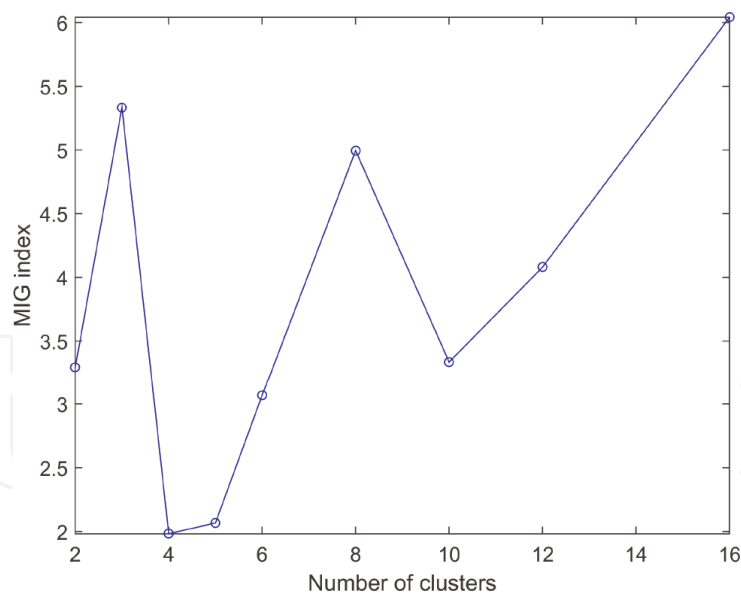
##### 3.1.1 Data organization

The criminal dataset for the city of San Francisco was obtained online through the open data services provided by the local government of the city. The dataset used in this work registers about 70,000 criminal events between years 2003 and 2015. Each criminal register contains attributes such as latitude, longitude, time, date, type of crime, among others. For the purposes of this study, only house burglary registers were considered, and from each register only latitude, longitude, and date were taken into account. Each spatial register was projected by means of the universal mercator system taken the location of the city of San Francisco as reference. Relative distances were expressed in meters. Time frames of criminal activity were created by aggregating the crime events of the last 7 days. Frames iterate from 1 day to the next. A total of 3195 time frames were produced.

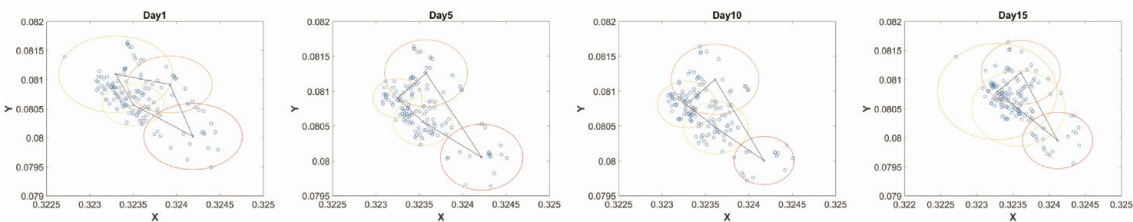
##### 3.1.2 Optimal number of clusters

Optimization of the MIG index was carried out over time series from clustering with  $C = 2 \dots 16$ , as shown in **Figure 4**. There was not be a clear criteria to set the maximum number of clusters to be considered in the optimization of any clustering validation index. In this case, the maximum was determined by taking into account that San Francisco city is divided in 10 police districts. Thus, the optimization was computed considering just three additional clustering settings (i.e.  $C = 12, 16, 20$ ). It was observed for these values that the MIG index diverged too fast, since in  $C = 20$ , it reached a value greater than the first maximum by about two orders of magnitude. The optimal MIG index is obtained for  $C = 4$ , whose value is around 30% of the maximum. However, the index found in  $C = 5$  is quite similar to the optimum. It is expected to find a similar predictability potential for these two cases.

The difference between the extreme cases may be explained by examining **Figure 5**. Note that for  $C_{opt} = 4$ , big areas are covered so that clusters concentrate a greater number of crime events. Therefore, the spatial variability of clusters in the optimal case is less subject to fluctuations than that of the worst case. In other



**Figure 4.** MIG index computed for house burglaries in the city of San Francisco, USA. The minimum of the MIG index appears at four clusters, where it is expected that time series from clustering are the most predictable.



**Figure 5.** An example of cluster dynamics for house burglaries in the city of San Francisco, USA. In this example, four clusters were tracked with the clustering reorganization algorithm (CRA).

words, the dynamics of clusters in the optimum case would exhibit the lowest level of disorder.

An example of cluster dynamics for house burglaries in the city of San Francisco, USA. In this example, four clusters were tracked with the clustering reorganization algorithm (CRA).

### 3.1.3 Characterization of time series

According to the MIG index, only clustering results for  $C = 4$  and  $C = 16$  groups were considered. These cases represent the most and least predictable dynamic partitions. It is necessary to choose a representative time series (Rep) in the two scenarios, denoting them as the optimal scenario (i.e., minimum MIG index) and the control scenario (i.e., maximum MIG index). **Table 1** summarizes some results from this process. In column  $S$ , the field “[i]” refers to the cluster number. As shown in this table, selected series are from different clusters since the first aim of the forecasting in this study is to have a preliminary validation of the MIG index instead of developing an exhaustive forecast. Representative series (Rep) are selected by finding the minimum distance between their characterization vector and the mean vector of a class  $S$  (i.e.,  $X$ ,  $Y$ , and  $R$ ) according to Eqs. (10)–(14).

### 3.1.4 Forecasting results

Once the series were selected, the memetic algorithm was configured to compute the forecasting with 70% of the data (i.e., 2236 days) and compute the



validation with the other 30% of the data (i.e., 959 days). In order to initialize the algorithm in such a way that the search space could be explored widely with the least possible computational cost, guaranteeing the simplicity of the solutions, a set of experiments was launched for the combination of parameters presented in **Table 2**. The number of generations and population size were selected as the maximum values in which at least a difference of 0.1% between consecutive generations was found in the fitness function of the fittest solutions. Number of Epochs was chosen to avoid frequent loss of diversity so that the FANO factor was close to the selected threshold. Regarding the fuzzy predictor, the number of regressors in the inputs and the number of rules were selected by finding the maximum values in which the Pearson coefficient for the forecast and real data did not change in more than 2%.

From selected configuration values, due to the random initialization of the memetic algorithm, over 10,000 experiments were run per series. **Table 3** summarizes the performance of the best fuzzy predictors found by the memetic algorithm. Pearson's correlation coefficient between the model and data for the optimal scenario was the greatest in both training and validation. The same observation can be stated for the Nash coefficient. The relative difference between Pearson coefficients of the optimal and control scenarios was about 17% in validation whereas for the Nash index, the relative difference was 26%. Regarding the estimated LLE, smaller values were obtained for the optimal scenario, and even a negative LLE was obtained. Visual results are depicted in **Figure 6**. Selected time series exhibited an interesting texture, which was more accentuated in the control scenario since more peaks appeared randomly and recurrence was not easily observed.

S	$\mu$		$\sigma \times 10^{-8}$		$E \times 10^9$		Lg		$D_s$
	Rep	Mean	Rep	Mean	Rep	Mean	Rep	Mean	Norm
x [1]	0.323	0.323	1.139	1.534	4.692	4.821	9	12.75	0.052
y [2]	0.0810	0.0806	2.3229	4.230	1.765	2.4516	7	11	0.337
R [3]	5.23e-04	5.14e-04	1.5880	1.574	4.292	4.4145	7	19.75	0.044

*These series are a representative sample of the entire, set which contains 12 time series (four clusters each one with three parameters).*

**Table 1.**  
*Features of representative time series obtained from the TSC method for house burglaries in the city of San Francisco, USA, considering four clusters.*

Parameter	Range	Step	Value
Number of generations	50–500	50	250
Population size	10–50	10	20
Epochs number	1–20	1	5
Number of regressors	3–12	1	8
Number of rules	4–24	4	8

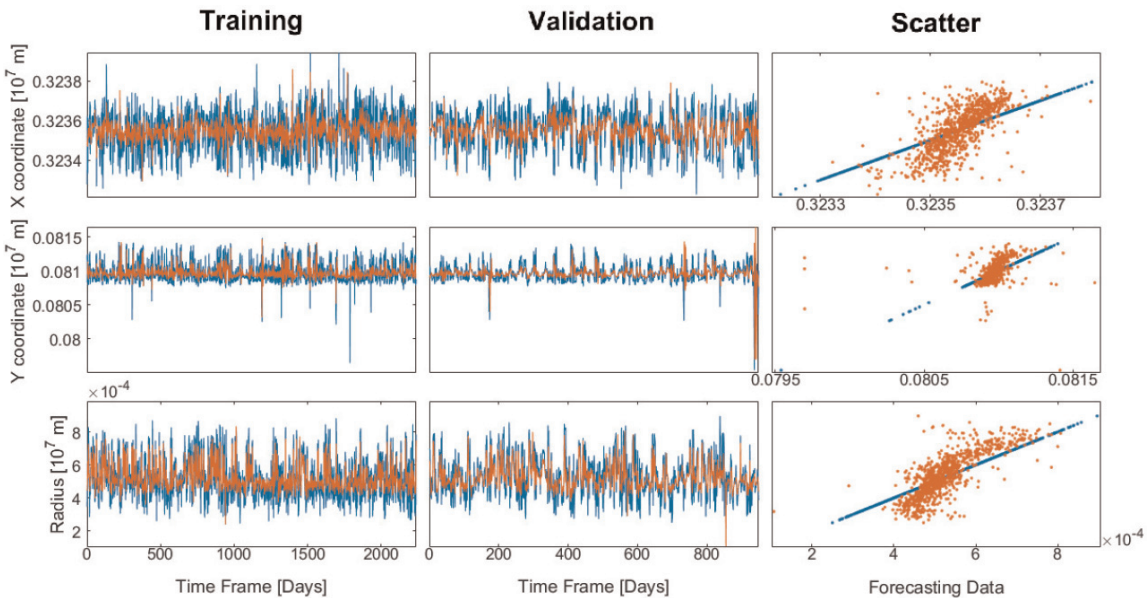
*The first three parameters were used to configure the differential evolution algorithm whereas the last two were used to adjust the Adaptive Neural Fuzzy Inference System.*

**Table 2.**  
*Parameters of the evolutionary fuzzy predictor used for the city of San Francisco, USA.*

S	Pearson coefficient				Nash		LLE	
	Training		Validation		4	16	4	16
	4	16	4	16				
x	0.6099	0.3986	0.6942	0.4581	0.3430	0.1534	0.0339	−0.0359
y	0.5447	0.7114	0.4982	0.5546	0.2908	0.5021	−0.0118	0.0796
R	0.8353	0.6645	0.7803	0.6260	0.6296	0.2779	0.0388	0.0677
Average	0.6633	0.5915	0.6576	0.5463	0.4212	0.3111	x	x

Dynamic partitions with minimum (4 clusters) and maximum (16 clusters) MIG index were considered.

**Table 3.**  
Forecasting indices for representative time series of two dynamic partitions (house burglaries in the city of San Francisco, USA).



**Figure 6.**  
Forecasting results for the three representative time series of the cluster dynamics (House burglaries in the city of San Francisco, USA). Series were obtained from a dynamic partition with four clusters.

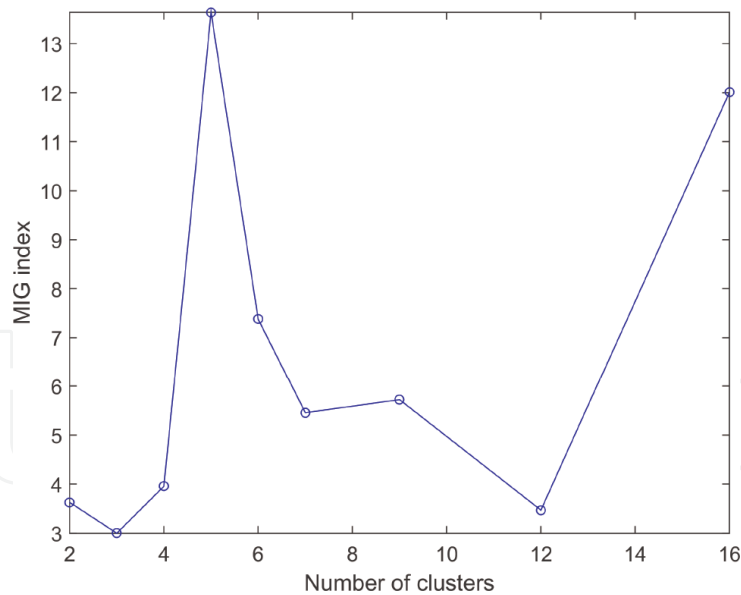
### 3.2 Results for Bogota, Colombia

#### 3.2.1 Data organization

The dataset for the city of Bogota, Colombia was provided by the Non-Governmental Organization (NGO) “Fundacion ideas para la paz,” which contains about 25,000 events of cellular phone robbery registered between the years of 2012 and 2015. Criminal registers contain X-coordinates, Y-coordinates, and dates of events. No transformation was required since the coordinates were already expressed in a geographical system adapted to the city. All differential spatial measurements were processed in meters. As in the previous case, time frames were created by the aggregation of the criminal events recorded in the last 7 days. A total amount of 1417 time frames was generated.

#### 3.2.2 Optimal number of clusters

According to **Figure 7**, the MIG index for Bogota series exhibited a global minimum at  $C_{opt} = 3$ , showing non-convex behavior with respect to the number of



**Figure 7.**  
MIG index computed for cellular phone robbery in the city of Bogota, Colombia. The minimum of the MIG index appears at four clusters, where it is expected that time series from clustering are the most predictable.

clusters. Note the similarity of the optimal MIG index and the one obtained for  $C = 12$ . The optimization process was carried out for  $C = 2 \dots 20$ ; however, the last result with  $C = 20$  was omitted since the MIG index grew exaggeratedly.

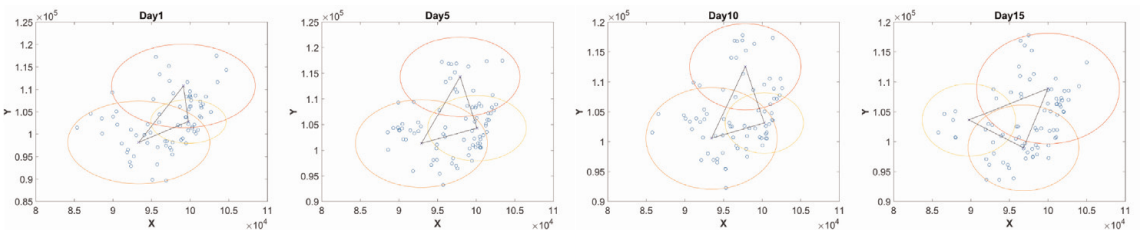
A sample of the cluster dynamics is presented in **Figure 8** for the optimal case. However, it is not a straightforward task to infer the result of the optimization process from this figure. Centroids of clusters are moving, as can be noticed from the change in shape of the connecting polygon. Note that the radiuses of the clusters changed in size so different areas were covered from one time frame to another.

### 3.2.3 Characterization of time series

Characterization of time series of dynamic partitions are summarized in **Table 4**. The field “[i]” refers to the cluster number. For these series it is supposed a-priori that the three-group clustering is more predictable than the five-group, considering the respective entropies. This result is in accordance with the MIG index optimization. Regarding the radius series, the number of lags is higher with respect to the series collection. Representative series (Rep) were selected by finding the minimum distance between their characterization vector and the mean vector of a class  $S$  (i.e.,  $X$ ,  $Y$ , and  $R$ ) according to Eqs. (10)–(14).

### 3.2.4 Forecasting results

**Table 5** presents the combination of parameters that were used for running the optimization of the fuzzy forecasters. In this case, 70% of generated samples (i.e.,



**Figure 8.**  
An example of cluster dynamics for cellular phone robbery in the city of Bogota, Colombia. In this example three clusters were tracked with the clustering reorganization algorithm (CRA).

S	$\mu \times 10^4$		$\sigma \times 10^6$		$E \times 10^{-5}$		Lg		$D_s$
	Rep	Mean	Rep	Mean	Rep	Mean	Rep	Mean	Norm
x [3]	9.8	9.7	8.115	7.251	1.036	1.4336	7	9	0.251
y [3]	10.9	10.4	11.951	11.265	0.894	0.755	6	8	0.099
R [3]	9.8	0.6	3.854	4.4214	1.723	1.484	163	158	0.15

These series are a representative sample of the entire set, which contains nine time series (three clusters each one with three parameters).

**Table 4.**  
Features of representative time series obtained from the TSC method for cellular phone robbery in the city of Bogota, Colombia, considering three clusters.

Parameter	Range	Step	Value
Number of generations	100–400	50	200
Population size	10–50	10	30
Epochs number	1–20	1	4
Number of regressors	3–12	1	8
Number of rules	4–24	4	8

The first three parameters configured the differential evolution algorithm, whereas the last two were adjusted for the Adaptive Neural Fuzzy Inference System.

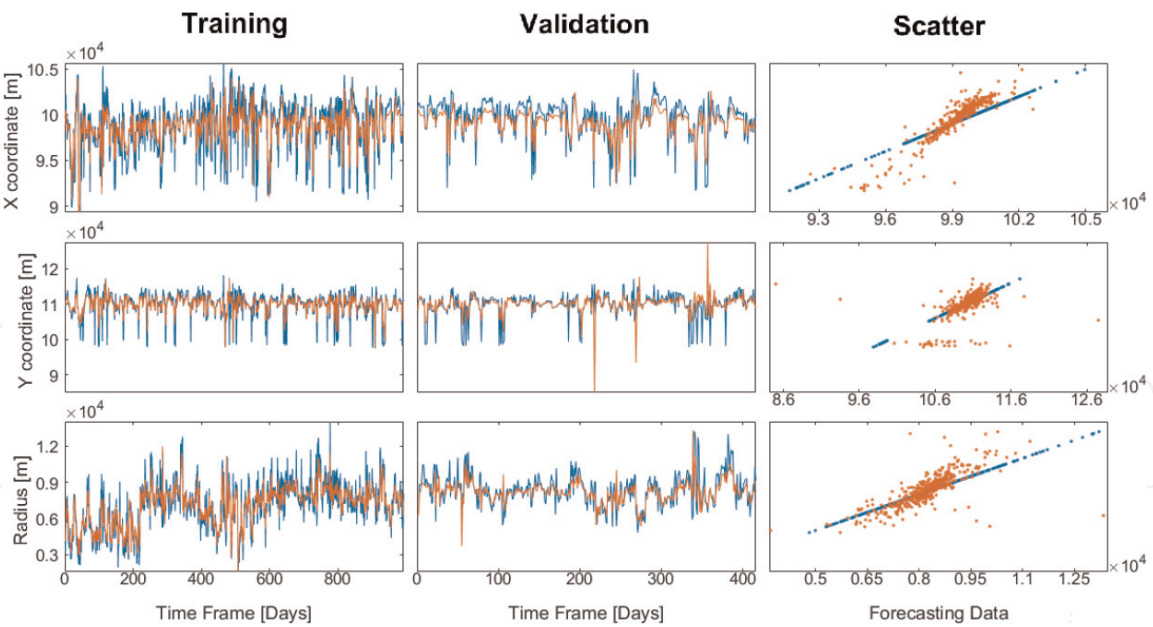
**Table 5.**  
Parameters of the evolutionary-fuzzy predictor used for the city of Bogota, Colombia.

992 days) were used for training, and the remaining 30% for validation (i.e., 425 days). The best forecasting results for Bogota series are presented in **Table 6**. The Pearson correlation coefficient between the model and real data was greater for the optimal scenario (i.e., three clusters) with respect to the control scenario (i.e. five clusters) in both training and validation. The relative difference was about 18 and 12%, respectively. In terms of the Nash index, the optimal scenario exhibited the highest performance with a relative difference of 31%. Estimated LLEs were smaller in the optimal scenario with a negative LLE in the case of the  $X(t)$  signal. **Figure 9** depicts visual results. It can be seen that predicted signals in the optimal scenario are correlated to the real ones. Signals in the control scenario exhibited more random peaks, although the recurrence is similar to the optimal case.

Signal	Pearson Coefficient				2*Nash		2*LLE	
	Training		Validation					
	3	5	3	5	3	5	3	5
x	0.8443	0.7388	0.9154	0.6833	0.6546	0.5025	−1.5102	−0.3230
y	0.7238	0.6833	0.4009	0.4755	0.5069	0.2795	1.5157	1.5100
R	0.8879	0.5025	0.7538	0.6441	0.7740	0.5499	1.2457	2.1515
Average	0.8186	0.6708	0.6900	0.6010	0.6452	0.4440	x	x

Dynamic partitions with minimum (three clusters) and maximum (five clusters) MIG index were considered.

**Table 6.**  
Forecasting indices for representative time series of two dynamic partitions of cellular phone robbery in the city of Bogota, Colombia.



**Figure 9.**  
*Forecasting results for the three representative time series of the cluster dynamics (cellular phone robbery in the city of Bogotá, Colombia). Series were obtained from a dynamic partition with three clusters.*

#### 4. Conclusion

Qualitatively speaking, similar results were obtained for the two study cases. In both cases, the forecasting of time series from the optimal scenario outperformed that of the control scenario in terms of two statistical indices (i.e., Pearson correlation and Nash coefficients) and one from chaos theory (i.e., LLE). Therefore, we argue that the MIG index might be useful to evaluate the goodness of a dynamic partition of crime events. Moreover, it may give a confidential a-priori insight about the predictability of series synthesized from TSC. Hence, this method produces coherent series that preserve a temporal structure, which a forecasting method can take advantage of.

TSC series for both cities exhibited an interesting behavior in terms of their texture. No evident periodicity was observed and abundant peaks appeared over the observed time window. Positive LLEs computed in these signals revealed the presence of a possible chaotic nature of the phenomena. Chaotic texture of these series speaks about non-stationary spatial crime patterns that evolve continuously producing information. Thus, this observation reflects a footprint of complexity for urban crime as noted in previous studies [35].

As a future work, the proposed method will be tested over other dynamic phenomena characterized by non-uniform sampling of relevant variables in both space and time. The method will be also refined by considering other techniques such as auto-encoder deep neural networks among others.

#### Conflict of interest

The authors declare no conflict of interest.



IntechOpen

### Author details

Miguel Melgarejo<sup>1\*†</sup>, Cristian Rodriguez<sup>1†</sup>, Diego Mayorga<sup>1†</sup> and Nelson Obregón<sup>2†</sup>

1 Laboratory for Automation and Computational Intelligence, Universidad Distrital Francisco José de Caldas, Bogotá DC, Colombia

2 Water Institute, Pontifical Xavierian University, Bogotá DC, Colombia

\*Address all correspondence to: [mmelgarejo@udistrital.edu.co](mailto:mmelgarejo@udistrital.edu.co)

†All authors are contributed equally.

### IntechOpen

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

## References

- [1] Felson M. Routine activity approach. In: *Environmental Criminology and Crime Analysis*. Abingdon, UK: Routledge; 2008. p. 7077
- [2] Cornish D, Clarke R. The rational choice perspective. In: *Environmental Criminology and Crime Analysis*. Abingdon, UK: Routledge; 2008. p. 2145
- [3] D'Orsogna M, Perc M. Statistical physics of crime: A review. *Physics of Life Reviews*. 2014;**12**:1-21
- [4] Perc M, Donnay K, Helbing D. Understanding recurrent crime as system-immanent collective behavior. *PLoS One*. 2013;**8**:e76063
- [5] Brantingham PJ, Brantingham P. Crime pattern theory. In: *Environmental Criminology and Crime Analysis*. New York: Willian Publishing; 2008. pp. 78-93
- [6] Short M et al. A statistical model of criminal behavior. *M3AS*. 2008;**18**: 1249-1267
- [7] Rey S, Mack E, Koschinsky J. Exploratory space-time analysis of burglary patterns. *Journal of Quantitative Criminology*. 2012;**28**(3): 509-531
- [8] Mohler G. Marked point process hotspot maps for homicide and gun crime prediction in Chicago. *International Journal of Forecasting*. 2014;**30**(3):491-497
- [9] Grubestic T. On the application of fuzzy clustering for crime hot spot detection. *Journal of Quantitative Criminology*. 2006;**22**(1):77-105
- [10] Brantingham P et al. Crime analysis at multiple scales of aggregation: A topological approach. In: *Putting Crime in Its Place*. New York: Springer; 2009. pp. 87-107
- [11] Davies T, Johnson S. Examining the relationship between road structure and burglary risk via quantitative network analysis. *Journal of Quantitative Criminology*. 2015;**31**(3):481-507
- [12] Malleson N, Andresen M. Spatio-temporal crime hotspots and the ambient population. *Crime Science*. 2015;**4**(10):1-8
- [13] Bettencourt L et al. Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities. *PLoS One*. 2010; **5**(11):20-22
- [14] Murray A, Grubestic T, Leitner M. Exploring spatial patterns of crime using non-hierarchical cluster analysis. In: *Crime Modeling and Mapping Using Geospatial Technologies*. Vol. 8. Netherlands: Springer; 2013. pp. 105-124
- [15] Mayorga D, Melgarejo M, Obregon N. A fuzzy clustering based method for the spatiotemporal analysis of criminal patterns. In: *2016 IEEE International Conference on Fuzzy Systems*; 2016. pp. 738-744
- [16] Izakian H, Pedrycz W, Jamal I. Clustering spatio-temporal data: An augmented fuzzy C-means. *IEEE Transactions on Fuzzy Systems*. 2013; **21**(5):855-868
- [17] Ji M, Xie F, Ping Y. A dynamic fuzzy cluster algorithm for time series. *Abstract and Applied Analysis*. 2013; **2013**:1-7
- [18] Hardyns W, Rummens A. Predictive policing as a new tool for law enforcement? Recent developments and challenges. *European Journal on Criminal Policy and Research*. 2018;**24**:201
- [19] Bezdek JC. *Pattern Recognition with Fuzzy Objective Function Algorithms*.

New York: Springer Science, Business Media; 2013

[20] Babuska R. Fuzzy modeling for control. In: International Series in Intelligent Technologies. Netherlands: Springer; 1998

[21] Abarbanel H. Analysis of Observed Chaotic Data. New York: Springer; 1996

[22] Rosenstein MT et al. A practical method for calculating largest Lyapunov exponents from small data sets. *Physica D*. 1993;65:117-134

[23] Ayyub B, McCuen R. Probability, Statistics, and Reliability for Engineers and Scientists. Boca Raton: Chapman and Hill/CRC Press; 2002. pp. 65-72

[24] Ting S, Gang C. Selection of fuzzy time series model based on autocorrelation theory. In: 29th Chinese Control and Decision Conference; Chongqing, China; 2017. pp. 4365-4369

[25] Neri F, Cotta C. Memetic algorithms and memetic computing optimization: A literature review. *Swarm and Evolutionary Computation*. 2012;2:1-14

[26] Moscato P, Cotta C. A Gentle Introduction to Memetic Algorithms. Boston MA: Glover, Fred and Kochenberger; 2003. pp. 105-144

[27] Rodriguez C, Mayorga D, Melgarejo M. Forecasting time series from clustering by a memetic differential fuzzy approach: An application to crime prediction. In: 2017 IEEE Symposium Series on Computational Intelligence (SSCI); Honolulu, HI; 2017

[28] Storn R, Price K. Differential evolution: A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*. 1997;11:341-359

[29] Jang J. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*. 1993;23(3):665-685

[30] Chivata J et al. Complex system modeling using TSK fuzzy cellular automata and differential evolution. In: 2013 IEEE International Conference on Fuzzy Systems; Hyderabad; 2013. pp. 1-5

[31] Mendel J. Fuzzy logic systems for engineering: A tutorial. *Proceedings of the IEEE*. 1995;83(3):345-377

[32] Nash J, Sutcliffe J. River flow forecasting through. Part I: A conceptual models discussion of principles. *Journal of Hydrology*. 1970;10(3):282-290

[33] Chai T, Draxler R. Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*. 2014;7:1247-1250

[34] Silva D, Alves G, Mattos Neto P, Ferreira T. Measurement of fitness function efficiency using data envelopment analysis. *Expert Systems with Applications*. 2014;41(16): 7147-7160

[35] Oliveira M, Bastos-Filho C, Menezes R. The scaling of crime concentration in cities. *PLoS One*. 2017; 12:113