

# The CriLiM Methodology: Crime Linkage with a Fuzzy MCDM Approach

Fabrizio Albertetti\*, Paul Cotofrei\*, Lionel Grossrieder†, Olivier Ribaux†, Kilian Stoffel\*

\*Information Management Institute  
University of Neuchatel, Switzerland

†Institut de Police Scientifique  
University of Lausanne, Switzerland

**Abstract**—Grouping events having similarities has always been interesting for analysts. Actually, when a label is put on top of a set of events to denote they share common properties, the automation and the capability to conduct reasoning with this set drastically increase. This is particularly true when considering criminal events for crime analysts; conjunction, interpretation and explanation can be key success factors to apprehend criminals. In this paper, we present the CriLiM methodology for investigating both serious and high-volume crime. Our artifact consists in implementing a tailored computerized crime linkage system, based on a fuzzy MCDM approach in order to combine spatio-temporal, behavioral, and forensic information. As a proof of concept, series in burglaries are examined from real data and compared to expert results.

**Keywords**—Crime analysis; crime linkage; fuzzy MCDM

## I. INTRODUCTION

Crime linkage is quite a particular challenge. As stated by Grubin et al., crime linkage consists in determining if a set of events has been committed by the same offender [2]. Basically, two tracks provide an answer to this problem: behavioral and situational similarities. The first aspect relies on the behavior of the offender, described by particular methods of crime (aka *modus operandi*). The second is based on spatio-temporal similarities: analyses are conducted in accordance with crime mapping theories. However, a third aspect can be considered: forensic cased data (such as DNA, shoemarks or earmarks). Depending on the organizations, the availability of forensic data within the criminal investigation process is limited. It is also based on the assumption that the offender physically leaves a trace of a satisfactory quality in more than one crime scene. Nevertheless, forensic information is much more reliable and always more integrated into intelligence databases [3].

As several studies (e.g. [2], [4], [5], [6]) mainly focus on behavioral linkage analysis (BLA) and on situational analysis, this current study attempts to combine behavioral, situational, and forensic information. Covering these three aspects is a way to combine techniques for both serious and high-volume crimes analysis. To do so, several mathematical methods that

effectively compute these similarities can be put altogether. Our approach is based on a fuzzy multi-criteria decision making (MCDM) method in order to deal with that issue. Indeed, in a crime analysis context, fuzziness permits to model experts' experience and handle vagueness whereas MCDM is useful to evaluate and combine similarities stemming from multiple criteria.

As an extension of a preliminary study [1] describing a crime linkage method, this current paper provides further work with a step-by-step methodology, thorough theoretical foundations, detailed experiments and interpreted results.

The remainder of this paper is structured as follows: Section II introduces the key concepts in both fuzzy MCDM and crime linkage. A review of similar research is also presented. In Section III, we detail the proposed methodology with the objective of implementing a computerized crime linkage system. A proof of concept is described in Section IV with real data about serial burglaries. To sum up, Section V concludes on the perspectives of the artefact.

## II. LITERATURE REVIEW

In this section, multi-criteria decision making concepts are introduced and a comparison of the underlying techniques is undertaken. Then, fuzzy sets are presented with the focus on translating experts' informal reasoning knowledge into membership functions. Third, we review a set of related work focusing on combining these methods. To finish, a forensic framework is set up with the purpose of revealing the big picture of the problematic we are confronting.

### A. Multi-criteria Decision Making and Fuzziness

Multi-criteria decision making is a vast research area. However, the main concept is simple: helping decision-makers in choosing, ranking, or sorting alternatives according to a set of criteria. In the literature, two basic approaches are frequently used: multiple objective decision making and multiple attribute decision making (MADM). Our interest lies in this latter: MADM.

In an MADM problem, the *decision maker* wants to select a particular subset from a set of *alternatives*. These alternatives are described by *criteria* (at least two), and each criterion can have a different importance (the *weight*) in regard to the other criteria. These weights usually reflect the subjectivity of

---

This paper is an extension of a preliminary study [1]. It is supported by the Swiss National Science Foundation (SNSF) [project 135236].

the decision maker. From that point of view, three challenges can be thought of: (a) finding the best alternative (*choice* problem); (b) ranking all the alternatives in decreasing order (*ranking* problem); or (c), assigning the alternatives to pre-defined ordered categories (*sorting* problem).

Plenty of MADM methods can be used to tackle these problems, such as maximin, maximax, AHP, TOPSIS, multi-attribute utility theory (MAUT), etc. MAUT is an interesting approach and is used in many different fields [7]. For our problematic, it is a consistent choice in the sense it can capture decision maker's perceptions of preferences. Basically, MAUT defines a model with functions for mapping a utility value to each attribute, and this for each criterion. Then, a weight is assigned to these criteria and a composition function is chosen to aggregate utilities. Once an aggregated utility is computed for each alternative, we can easily establish preferences among them. A performance matrix is then used to answer the respective questions.

Fuzzy sets are interesting when we need to translate these preferences into functions. Introduced by Zadeh [8], fuzzy sets form the basis of multi-valued membership functions. In crisp traditional logic, the truthfulness of a statement is a dichotomy, i.e. a statement can only be *true* or *false*. On the other hand, with fuzzy sets, the real unit interval  $[0, 1]$  can be used to assign the degree of truth of a statement. The smallest value of the interval denotes a total lack of membership of a class and the highest value denotes a full membership.

To illustrate this concept, let  $x$  ( $x \in X$ ) be the financial penalty a suspect is expected to pay ( $X \subset \mathbb{N} \mid x \leq 100\,000, \forall x \in X$ ). If the suspect is guilty, the judge has to determine the degree of implication of the suspect (for the sake of brevity, we assume that the penalty depends only on the degree of implication and simplify the procedure). For that, he might only consider a restricted set of linguistic variables: partially implicated, mainly implicated and fully implicated (assuming these alternatives are ordinal). These alternatives can be respectively represented by membership functions:  $\mu_{part}(x)$ ,  $\mu_{main}(x)$  and  $\mu_{full}(x)$ . Membership functions will map these qualitative statements to numbers for each alternative (namely  $x$ , the damage to pay), and vice versa. These values are then denoted by fuzzy sets ( $N$  being the number of linguistic variables):  $\langle x, \mu_1(x), \mu_2(x), \dots, \mu_N(x) \rangle$  (when a full partition is considered,  $\sum_{i=1}^N \mu_i(x) = 1$ ). The main characteristics of fuzzy sets lie in the choice of the membership functions and the aggregation technique.

Several ideas combine fuzzy logic with MCDM: an interesting summary of this approach is presented by Kahraman [9]. Straccia [10] presents how to embed fuzzy description logic in MCDM with the perspective to use ontologies. But there are many others: for example, personnel selection in a human resources problem is tackled with a TOPSIS method and an ordered weighted average (in a study undertaken by Dursun and Karsak [11]); the best strategy for oil spill problems is determined by taking into account several experts' opinions with fuzzy evaluation methods (presented by Krohling and Rigo [12]).

## B. Computational Forensics and Crime Analysis

*Computational forensics* (CF) is about applying computational methods from several disciplines in the forensic domain. Franke and Srihari [13] define three axes in which these methods support forensic sciences: (a) they provide a set of tools to overcome limitations of human cognitive abilities; (b) very large sets of data are potentially usable for analyses and are not anymore constrained by the human capacities; and (c) human expert knowledge can be modeled and made explicit to be used in inference mechanisms.

With the support of CF, crime analysis becomes an interesting research field with the goal to overcome traditional limitations. Crime analysis is defined as "the systematic study of crime and disorder problems as well as other police-related issues (including socio-demographic, spatial, and temporal factors) to assist the police in criminal apprehension, crime and disorder reduction, crime prevention, and evaluation" [14]. Even considering this broad definition, detecting similarities with computational methods between events in a crime analysis perspective has not particularly been an active research area over the past years. Nevertheless, a few researchers have provided some answers. A particular emphasis has been given to analyzing spatio-temporal similarities (such as for predicting serial killers' home [15], or for finding a pattern of offenses of serial rapists [16]). Another focus has been directed towards behavioral crime linkage (e.g. [17], [4], [18], [6], [19], [5]) relying mainly on the crime method (modus operandi) inferred from crime scene investigation.

Even though there exists techniques and methods related to serial crime analysis, some critical questions remain: can we really detect series in our data sets? Do series really exist? A track attempting to provide answers to these questions is based on the Ratcliffe's 4P model [20]: *prevention* requires *proactivity*, which requires *predictability* which in turn requires *patterns*. So can we infer that some patterns are present in the data? In the context of serial crimes, our assumption is yes, when we assume that most of the crimes are committed by very few offenders [21], and moreover when these latter leave a genuine "signature" [5]. In some other cases, situational crime approaches [22] also suggest a positive answer to our question given that crime occurrences are not randomly distributed and are highly correlated to the physical, spatial, and temporal environment. Repeat victimization phenomena [23] and hot spot locations [24] are also factors increasing the probability of patterns.

A recent study [25] confirms a certain amount of predictability in certain crimes, but a gap still exists. The current study intends to practically support these theoretical assumptions.

Furthermore, a critical overview of the existing crime linkage systems is presented in [26]. A famous but controversial<sup>1</sup> system (ViCLAS) is assessed, and four assumptions for an effective implementation are formulated based on this analysis. As stated by Bennell et al., "most of the issues discussed in this article apply to all linkage systems". Therefore, these

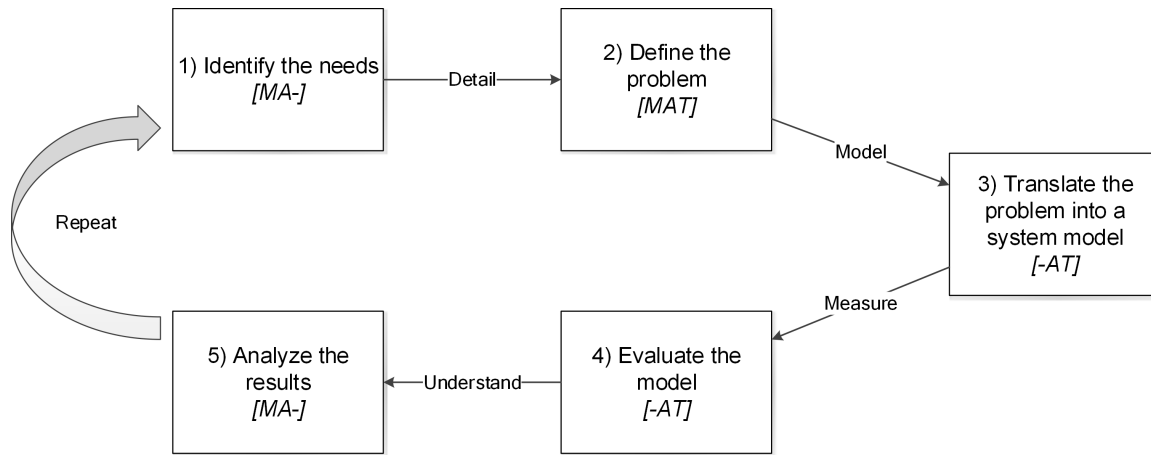


Fig. 1. The 5 iterative steps of the CriLiM methodology. The *[MAT]* notation indicates for each step the required key competences. ‘M’ stands for *managerial* competences, ‘A’ for *analytic* competences (i.e. crime analysts), and ‘T’ for *technical* competences (e.g. statistical, computational or legal).

hypotheses are considered in the proposed methodology.

### III. THE PROPOSED CriLiM METHODOLOGY

The proposed crime linkage multi-criteria methodology (denoted hereafter by CriLiM for the sake of brevity) is a set of sequential iterative steps defining a coherent framework for crime linkage. Handling behavioral, situational, and forensic link analysis, the methodology is intended for crime analysts. The idea is to let them express their own perceptions of crime linkage in their own application domain, in order to implement a system tailored to their specific needs.

The problem of similarities is addressed with MAUT methods and fuzzy sets. Using a MAUT method permits to use different criteria as input to define the notion of similarity, whereas fuzzy sets allow to express the experts’ preferences with membership functions.

The conceptual steps of CriLiM are based on the work of Checkland [28]. In a thirty year retrospective of soft systems methodology, he undertakes the challenge to “see if systems ideas could help us tackle the messy problems of ‘management’ [...]”. Actually, a managerial incentive is a requirement in such cases. Many aspects can impact on the level of data quality and on the availability of the data set. Therefore, the overall context is of high importance in order to successfully carry out such projects.

Considering the general MCDM approach, Henig and Buchanan suggest in [29] some tracks to follow in order to determine the key success factors of a “good” decision making process. They argue the fact that good processes should not conceal the preferences of the decision makers themselves but give them a certain degree of liberty, which we will try to achieve with CriLiM.

Based on this analysis, the five following steps constitute the foundation of our methodology (see Figure 1):

- 1) identify the needs and the boundaries of the system;
- 2) clearly define the problem from a business point of view;
- 3) translate the business problem into a system model;
- 4) evaluate the model based on expert results;
- 5) analyze the results.

The remainder of this section describes these five steps and considers case linkage for non-specific crimes.

#### A. Step 1: Identify the Needs and the Boundaries of the System

The first step of the CriLiM methodology is to identify the needs and the boundaries of the system. Basically, the implementation of the methodology should be considered as a project. A business leader is designated to ensure that all the steps are undertaken in an appropriate manner. The stakeholders are defined (i.e. business experts, technicians, and managerial representatives) and an approximate scope (including time, economic, and purpose aspects) is given. The required data and legal authorizations are analyzed in accordance to the given departments and the local laws. A clear definition of who will use the system, for what purpose, and to which extent is formulated.

#### B. Step 2: Clearly Define the Problem from a Business Point of View

The goal of crime linkage is supposedly to answer a business problem: “Are some offenders behaving in the same way? Is one single criminal responsible for many burglaries, considering the given data set? Are these events related? How are victims related to offenders?” If an answer to these questions is detected, the crime analyst would then try to explain the possible causes.

Experts and managers together should possess enough knowledge to properly define the purpose and what exactly is expected from the system. They also should state a way to measure the degree of conformance to these statements, i.e. which criteria have to be fulfilled and to which extent.

<sup>1</sup>As stated in [27], despite the fact that ViCLAS exists for nearly 20 years, “there is no published account of its effectiveness being evaluated systematically”.

TABLE I. PERFORMANCE MATRIX OF  $m$  CRIMES WITH  $n$  PROPERTIES AGAINST  $c_{ref}$  WITH MAUT FUNCTIONS. UTILITIES ARE DEFINED AS FUNCTIONS OF PROPERTIES AND SCORES ARE FUNCTIONS OF UTILITIES.

	$w_1$	$w_2$	...	$w_n$	score of $c$ against $c_{ref}$
	$p_1$	$p_2$	...	$p_n$	
$c_1$	$u_{1,1}^{ref} = \phi_1^{ref}(p_1(c_1))$	$u_{1,2}^{ref} = \phi_2^{ref}(p_2(c_1))$	...	$u_{1,n}^{ref} = \phi_n^{ref}(p_n(c_1))$	$s_{1,ref} = \eta(u_{1,1}^{ref}, u_{1,2}^{ref}, \dots, u_{1,n}^{ref}, w_1, w_2, \dots, w_n)$
$c_2$	$u_{2,1}^{ref} = \phi_1^{ref}(p_1(c_2))$	$u_{2,2}^{ref} = \phi_2^{ref}(p_2(c_2))$	...	$u_{2,n}^{ref} = \phi_n^{ref}(p_n(c_2))$	$s_{2,ref} = \eta(u_{2,1}^{ref}, u_{2,2}^{ref}, \dots, u_{2,n}^{ref}, w_1, w_2, \dots, w_n)$
...	...	...	...	...	...
$c_m$	$u_{m,1}^{ref} = \phi_1^{ref}(p_1(c_m))$	$u_{m,2}^{ref} = \phi_2^{ref}(p_2(c_m))$	...	$u_{m,n}^{ref} = \phi_n^{ref}(p_n(c_m))$	$s_{m,ref} = \eta(u_{m,1}^{ref}, u_{m,2}^{ref}, \dots, u_{m,n}^{ref}, w_1, w_2, \dots, w_n)$

The identified information that will be used in the linkage system should also be assessed: data quality and data coding are key success factors to successful implementations.

### C. Step 3: Translate the Business Problem into a System Model

Once the problem clearly framed, the definition of a technical system solving the above-mentioned questions can be undertaken. The purpose of this step being to explicit a system detecting similarities within a data set, we will directly strive to define the basis of such system. Actually, the smallest element of this system is an event, which stems from a police report. This list of reported events denotes a set of crimes of similar type for a certain period and for a certain region (e.g. a list of burglaries committed in 2012 in Switzerland).

In order to understand this system, let us define its foundations. The finite list of crimes  $C = \{c_1, \dots, c_m\}$  is the set of  $m$  elements for which we are interested in finding similarities. Each crime  $c_i, \forall i \in [1, m]$  is described by  $n$  properties, denoted by the set  $P = \{p_1, \dots, p_n\}$  (which can also be regarded as a function of  $C$ : the property  $p_j$  of the crime  $c_i$  is the function  $p_j(c_i)$ ). Moreover, a degree of importance, denoted as the *weight*, is attached to these properties. Subjective to the decision maker and his experience, weights can be represented by the vector  $W$  ( $W = \{w_1, \dots, w_n\}$ ).

To create series from the set  $C$ , the following steps are applied:

- set a crime as a reference  $c_{ref}$  to which all other crimes  $c_i$  will be compared;
- compute the utility  $u$  for each property  $p$  of each crime  $c_i$  against the reference crime  $c_{ref}$ ;
- aggregate the utilities of each crime  $c_i$  to get the similarity coefficients  $s$  against  $c_{ref}$ ;
- repeat the first three steps  $m$  times;
- create the series based on thresholds according to the similarity coefficients.

The first step consists in choosing a *reference crime* ( $c_{ref}$ ) and comparing it to all other crimes  $c_i$  ( $c_{ref}, c_i \in C$ ) in regard to their properties. The reason of this step is purely iterative. Among all iterations,  $c_{ref}$  will varies from  $c_1$  to  $c_m$ .

As each crime may have up to  $n$  properties, the second step is to compute the utility for each of these properties. It might be computed in many ways. In our situation, we chose to handle the problem by using the multi-attribute utility theory technique. MAUT relies on the functions we choose and devise. The idea behind this choice is to give a chance to the experts to express their preferences with the help of fuzzy sets and membership functions. A value function is

therefore chosen by the expert according to his experience, which may be subjective. However, in the case of crime analysis, the experts conduct such kind of reasoning on a daily basis and have a wide knowledge of the exogenous factors (many decision criteria often lie beyond the system and beyond the simple concept of weights, because of their intrinsic difficulty to be explicitly defined). This method is therefore an appropriate way to integrate uncertain and subjective decisions into the system.

The utility of the property  $p_j$  of the crime  $c_i$  against the reference crime  $c_{ref}$  is therefore defined by the utility function  $\phi$  ( $\phi : C \rightarrow [0, 1]$ ) :

$$u_{i,j}^{ref} := \phi_j^{ref}(p_j(c_i)), \quad (1)$$

which takes its maximum value in  $c_i = c_{ref}$ . In the context of using a membership function from fuzzy sets, we suggest a Gaussian curve, a generalized bell-shape, a triangular-shape, or a singleton for  $\phi$ .

The third step is to aggregate these utilities in order to have a single value denoting the overall similarity of a crime. A weighted sum can be used (the SAW method) or any other function can be defined in accordance with constrains. From a generic point of view, the aggregation function  $\eta$  ( $\eta : [0, 1]^{2n} \rightarrow [0, 1]$ ) defines the similarity of the crime  $c_i$  with the crime  $c_{ref}$ :

$$s_{i,ref} := \eta(c_i) = \eta(u_{i,1}^{ref}, u_{i,2}^{ref}, \dots, u_{i,n}^{ref}, w_1, w_2, \dots, w_n). \quad (2)$$

This aggregation function has to be valid for a numeric or an ordinal property (see Table I). In the case we are facing nominal non-ordinal properties (such as a multi-valued category or a crisp category), we suggest to compute a coefficient regrouping these categories. For this purpose, the Jaccard's index [31] seems appropriate (which is a similarity coefficient that has been used in many BLA [6] for modi operandi). For a pair of crimes ( $c_a, c_b$ ) and crisp memberships of categories, the Jaccard coefficient of similarity is defined as such:

$$J(c_a, c_b) := \frac{\delta_{a,b}}{\delta_{a,b} + \delta_a + \delta_b}, \quad (3)$$

where  $\delta_{a,b}$  is the number of categories common to  $c_a$  and  $c_b$ ,  $\delta_a$  the number of categories that  $c_a$  belongs to but not  $c_b$ , and  $\delta_b$  the number of categories that  $c_b$  belongs to but not  $c_a$ .

The fourth step is to *repeat the same computations, in regard to each crime*. Which means we will compute the similarities between all the crimes. From these results an adjacency matrix (see Table II) can be defined, representing the similarity of each pair of crimes. This matrix is symmetric (the similarity

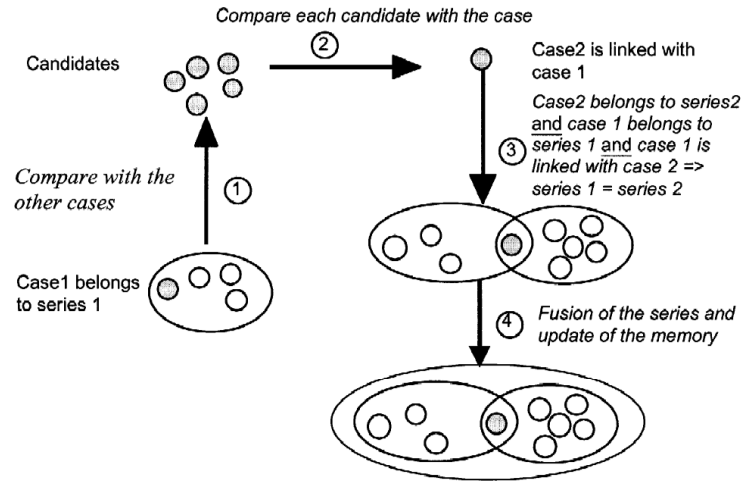


Fig. 2. Inference structure in [30] derived from experts: comparing a case.

TABLE II. ADJACENCY MATRIX, REPRESENTING THE SIMILARITY  $s_{i,j}$  BETWEEN EACH PAIR OF CRIMES ( $c_i, c_j$ ). THE MATRIX IS SYMMETRIC ( $s_{i,j} = s_{j,i}$ ), AS SIMILARITY FUNCTION ARE SYMMETRIC.

	$c_1$	$c_2$	...	$c_m$
$c_1$	$s_{1,1} = 1$	$s_{1,2}$	...	$s_{1,n}$
$c_2$	...	$s_{2,2} = 1$	...	$s_{2,n}$
...	...	...	...	...
$c_m$	...	...	...	$s_{m,m} = 1$

is assumed to be reflexive) and the diagonals are equal to 1 (a crime is fully similar to itself).

The last step is to create series in accordance with the adjacency matrix. The method used is based on an inference structure proposed by Ribaux and Margot [30], which basically describes crime analysts' implicit reasoning for cases comparison (see Figure 2). All pairs of crimes having a similarity above a specific threshold will be merged into a series, depending on the minimum size set for a series. It means that for merging a crime with an existing series, it will be compared to each crimes belonging to the series (all coefficients will have to be above the threshold).

#### D. Step 4: Evaluate the Model with Expert Results

The identified series and the adjacency matrix are the output of the model that experts can evaluate. Some specific series should be thoroughly assessed. These series should be clearly presented with all the properties to let the experts visualize the results. Both statistical and visual comparisons – a series can only be evaluated if the offender is actually known – are necessary to evaluate the adequacy of the linkage process in regard to the stated objectives. Evaluation metrics should also be measured. Furthermore, as suggested in [26], four hypotheses should be evaluated: the data provided in the system should be accurate, the data should be coded reliably, patterns should be present in the data, and experts should possess enough knowledge to link crimes accurately

#### E. Step 5: Analyze the Results

The purpose of this step is to analyze the results within the global context, and to identify the subsequent required changes for the next iteration of the methodology.

Basically, the relevance of the results have to be interpreted in regard to the whole project. The use of decision support systems are suggested to analyze properly these results. Pivot tables, dashboards, crime mapping and statistical systems should be considered. According to the results, a list of ad hoc changes has to be created in order to perform a new iteration.

### IV. CRIME LINKAGE OF RESIDENTIAL BURGLARIES

The goal of this section is twofold: (a) to provide a proof of concept of the proposed methodology, evaluated with the comparison of expert results, and (b) to illustrate the application of this general methodology with a real case study.

#### A. Description of the Data

Thanks to the Police de sûreté du Canton de Vaud (a canton police in Switzerland), we were able to access to a data set about serial and itinerant crime (high-volume crime). From about 55,000 events related to residential burglary involved in 6 cantons for a 4 year period, we focused on a particular subset for this experiment. This subset contains 2320 crimes conducted by 1141 distinct offenders (during all the study, we will assume that each crime has exactly one offender, even if sometimes crimes are committed by several offenders). The distribution of crimes is of 2 per offender on average (with a minimum of 1, a maximum of 52, and a standard deviation of 3.14). A detailed distribution of the series and the relations between the events of the entire data set is presented in [3]. The reason why we chose this subset is that we only considered the crimes that met some basic requirements to compute a similarity, i.e., an offender, a location, a date, and a modus operandi being not null. Two main reasons might explain a missing field: the intrinsic problem of gathering forensic data

TABLE III. ILLUSTRATION OF THE BURGLARIES DATA SET (THE ATTRIBUTE LIST IS NOT EXHAUSTIVE AND THEIR VALUES ARE FICTITIOUS). A ‘-’ DENOTES A MISSING VALUE.

	Modus Operandi 1	Entrance	Phenomenon	Forensic Trace	Date	Type of Place	Coordinates
evt248	Grabbed cylinder	Door	Giorno Cilindro	-	2012-10-03	Flat	(3459,4309)
evt994	Tool	Door	Giorno Piatto	-	2012-10-04	Villa	(3407, 5021)
evt093	Climbing	Roof	Notte Sera	ID-3095	2012-10-04	Villa	(9730, 2098)
evt605	Tool	-	-	-	2012-10-04	Flat	(3439,3372)
evt606	Grabbed cylinder	Door	Giorno Cilindro	-	2013-05-02	Flat	(3993, 4840)
evt038	Broken Window	Door	-	ID-3095	2012-10-05	Cellar	(9081, 2100)

from a crime scene, or, heterogeneous methods used among distinct police entities to gather data. In order to illustrate the data used for the proof of concept, Table III represents a fictitious excerpt. Basically, each crime is described with a method of crime, the main entrance used to commit the crime, the type of the place, the GPS coordinates, forensic case data (DNA, shoemarks, earmarks or toolmarks) and a phenomenon type describing the specificity of the crime (a coded value defined by the crime analysts, depending mainly on the modus operandi and the interval of time during the day when the crime occurred). For more information about the nature of the data and how the structure was adapted from police reports, see [32]. The remaining of this section describes step by step how we applied CriLiM to this subset.

### B. Applying the CriLiM Methodology

First, in response to the need of analyzing an increasing amount of data for crime investigation purposes, we focused on a particular problematic: the lack of cognitive resources to conduct exhaustive analyses (of course, crime linkage systems are only a partial answer to this problem). When considering the implementation of some existing solutions, their lack of transparency, their complexity, and their controversy may be dissuasive (e.g. the ViCLAS system, as discussed in [27]). The choice of conducting our own analysis seemed therefore appropriate. Then, the idea was to devise a generic methodology to evaluate the potential of such a system according to our own needs.

Second, we decided to focus on the particular crime of residential burglaries to implement a crime linkage analysis system. An important amount of data was already available and a way to evaluate the results was present in most of the crimes, i.e. the offender was known.

The next point was to develop the computerized crime system tailored to the data set (N=2320) and our needs. The result of a first analysis was the need to distinguish utility function types according to the nature of the variables, namely numeric (date and coordinates), categorical (modus operandi, entrance, phenomenon and the type of the place), and unique identifiers (forensic information).

The utility for the first variable type was evaluated with fuzzy sets. For example, the utility of the date property was represented by the fuzzy set  $\langle d_i, \mu^{ref}(d_i) \rangle$  (see Figure 3). The date was first normalized to the unit interval, denoted by  $d$ , and its membership value was evaluated using a bell-shaped function centered on the reference crime date ( $\mu^{ref}(d_{ref}) = 1$ ).

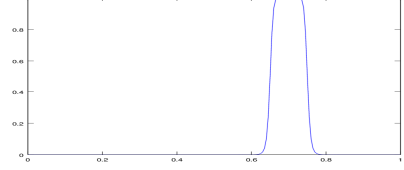


Fig. 3. Membership function for the date property, represented by a generalized bell-shaped centered on the reference date  $d_{ref} = 0.7$  (with parameters of  $[0.05, 6, 0.7]$  respectively the width, the slope, and the center of the curve). A full membership denotes a total similarity with the reference date ( $\mu^{ref}(d_{ref}) = 1$ ). An interval of the abscissa of approximately 0.2 denotes a period of 10 months.

Then, to compute the score of each pair of events, a simple weighted sum (the SAW technique) was used as an aggregation function for the sake of simplicity. Concerning the weights, a slightly higher importance was given to the date and the location of the crime compared to the other attributes. However, because physical evidence is present in very few cases (11% of N), we decided to add the weighted utility only when the field was not null (meaning that a similarity of 1 can be found even when forensic case data is missing).

The implementation of this model was done using Octave, a language for numeric computations (a free alternative of Matlab). The selection of the series is based on the degree of similarity (see the adjacency matrix in Table II) and on the inference structure of comparing cases extracted by experts as described in [30]. Results were obtained with different thresholds and series size.

The results of our implementation can be evaluated in many ways. Our choice was to analyze how series were detected by the system (how crimes were linked together as series when a particular threshold of similarity was reached). We compared these results with 4 different configurations (see Table IV), i.e. by changing the minimum coefficient of similarity for linking 2 crimes (the threshold  $t$ ), and the maximum number of crimes per series ( $s_{max}$ ). We defined two metrics for each of these configurations: (a) the number of distinct offenders per series, describing the diversity of criminals in a series; and (b), the number of distinct series per offender, counting in how many series crimes of the same offenders are linked.

The most interesting metric is the number of series in which a unique offender has been linked. On average, we can see that all crimes of the same offender are split into only one or two series (the best findings occur for the first and the last configuration, i.e. with a number of 10 crimes per series, for a total of respectively 65 and 96 series).

TABLE IV. RESULTS FOR THE IMPLEMENTED SOLUTION ( $N = 2320$  CRIMES). THE STATISTICS DESCRIBE THE NUMBER OF OFFENDERS PER SERIES (#o per series), AND THE NUMBER OF SERIES PER OFFENDER (#s per offender). THE PARAMETERS ARE THE FOLLOWING:  $t$  IS THE SIMILARITY THRESHOLD,  $s_{max}$  IS THE MAXIMUM SIZE OF A SERIES. ON AVERAGE, 890 CRIMES WERE LINKED IN SERIES.

Statistics	$t=0.8, s_{max} = 10$		$t=0.9, s_{max} = 7$		$t=0.95, s_{max} = 5$		$t=0.95, s_{max} = 10$	
	#total linked crimes: 960		#total linked crimes: 894		#total linked crimes: 1060		#total linked crimes: 645	
	#o per series ( $s_{tot} = 96$ )	#s per offender ( $o_{tot} = 533$ )	#o per series ( $s_{tot} = 128$ )	#s per offender ( $o_{tot} = 452$ )	#o per series ( $s_{tot} = 212$ )	#s per offender ( $o_{tot} = 534$ )	#o per series ( $s_{tot} = 65$ )	#s per offender ( $o_{tot} = 326$ )
Min	2	1	1	1	1	1	1	1
Max	10	11	7	11	5	15	10	7
Mean	7.47	1.35	5.04	1.43	3.77	1.5	6.63	1.32
Median	8.00	1.00	5.00	1.00	4.00	1.00	7.00	1.00
Std Dev.	2.27	0.94	1.68	1.10	1.18	1.34	2.40	0.83

These results are encouraging, in the sense that a main objective of a computerized crime linkage system is to give to crime analysts a way to reduce linkage blindness. On the other hand, even if crimes of different offenders are linked, it is not a significant issue when considering the concept of false negatives to be more important than the concept of false positives (when the purpose of case linkage focuses on decreasing linkage blindness).

To sum up, the results of such systems are to be used in a context of pure investigation. The main purpose is to enhance cognitive capacities of analysts, but the conclusions still have to be drawn with extreme care, always considering the assumptions of the model.

## V. CONCLUSIONS

Computerized crime linkage is an emerging research area and many challenges still need to be confronted. In this paper, we presented a methodology for implementing a tailor-made crime linkage system. The steps of CriLiM were devised to consider the system as a complete solution for experts, instead of providing a single technical tool without context. To cover both high-volume and serious crime, we integrated contextual, behavioral, and forensic information. The approach we adopted derives from fuzzy MCDM, in order to deal with similarities articulated with these three dimensions and to express experts' knowledge.

To illustrate and evaluate the methodology, an implementation of CriLiM has been applied to residential burglaries. The results illustrated the importance of a good understanding of case linkage systems and its environment, helping in determining the following key success factors: the data quality, the parameters of the system (the thresholds, the size of series, etc.), and the objectives of the system. More generally, results have to be drawn with extreme care because they are highly linked to the underlying assumptions of the model.

The impacts of CriLiM for crime analysts may be numerous: provide a basis for implementing their own crime linkage system, provide a method combining several distinct information types for crime linkage, adapt or compare an existing crime linkage system in accordance with the methodology, etc.

With the potential to benefit to all police organizations, CriLiM lays the ground for conducting new experiments and sharing results in the area of crime linkage.

## ACKNOWLEDGEMENTS

The authors would like to thank the Police de sûreté du Canton de Vaud, especially Sylvain Ioset and Damien Dessimoz for their active cooperation and the data provided.

## REFERENCES

- [1] F. Albertetti, P. Cotofrei, L. Grossrieder, O. Ribaux, and K. Stoffel, "Crime linkage: a fuzzy mcdm approach," in *Intelligence and Security Informatics (ISI), 2013 IEEE International Conference on*, 2013, pp. 1–3.
- [2] D. Grubin, P. Kelly, C. Brunson, and G. Britain, *Linking serious sexual assaults through behaviour*. Home Office London, 2001.
- [3] Q. Rossy, S. Ioset, D. Dessimoz, and O. Ribaux, "Integrating forensic information in a crime intelligence database," *Forensic Science International*, no. 0, InPress. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0379073812004641>
- [4] P. Santtila, T. Pakkanen, A. Zappalà, D. Bosco, M. Valkama, and A. Mokros, "Behavioural crime linking in serial homicide," *Psychology, Crime & Law*, vol. 14, no. 3, pp. 245–265, 2008. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/10683160701739679>
- [5] R. R. Hazelwood and J. I. Warren, "Linkage analysis: modus operandi, ritual, and signature in serial sexual crime," *Aggression and Violent Behavior*, vol. 9, no. 3, pp. 307 – 318, 2004. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1359178904000138>
- [6] T. Melnyk, C. Bennell, D. J. Gauthier, and D. Gauthier, "Another look at across-crime similarity coefficients for use in behavioural linkage analysis: an attempt to replicate woodhams, grant, and price (2007)," *Psychology, Crime & Law*, vol. 17, no. 4, pp. 359–380, 2011. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/10683160903273188>
- [7] J. Wallenius, J. S. Dyer, P. C. Fishburn, R. E. Steuer, S. Zionts, and K. Deb, "Multiple criteria decision making, multiattribute utility theory: Recent accomplishments and what lies ahead," *Management Science*, vol. 54, no. 7, pp. 1336–1349, 2008.
- [8] L. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338 – 353, 1965. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S001999586590241X>
- [9] C. Kahraman, *Fuzzy multi-criteria decision making: theory and applications with recent developments*. Springer, 2008, vol. 16.
- [10] U. Straccia, "Multi criteria decision making in fuzzy description logics: A first step," in *KES (1)*, 2009, pp. 78–86.
- [11] M. Dursun and E. E. Karsak, "A fuzzy mcdm approach for personnel selection," *Expert Systems with Applications*, vol. 37, no. 6, pp. 4324–4330, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0957417409010094>
- [12] R. Krohling and D. Rigo, "Fuzzy group decision making for management of oil spill responses," *Applications of Soft Computing*, vol. 58, pp. 3–12, 2009. [Online]. Available: <http://www.springerlink.com/index/E526143465739426.pdf>

- [13] K. Franke and S. Srihari, "Computational forensics: An overview," *IWCF 2008*, vol. 5158, pp. 1–10, 2008, 10.1007/978-3-540-85303-9-1.
- [14] R. Boba, *Crime Analysis with Crime Mapping*. Thousand Oaks, CA: Sage, 2012.
- [15] D. Canter, T. Coffey, M. Huntley, and C. Missen, "Predicting serial killers' home base using a decision support system," *Journal of Quantitative Criminology*, vol. 16, no. 4, pp. 457–478, 2000.
- [16] J. Warren, R. Reboussin, R. Hazelwood, A. Cummings, N. Gibbs, and S. Trumbetta, "Crime scene and distance correlates of serial rape," *Journal of Quantitative Criminology*, vol. 14, pp. 35–59, 1998. [Online]. Available: <http://dx.doi.org/10.1023/A-3A1023044408529>
- [17] G. Oatley, J. Zeleznikow, and B. Ewart, "Matching and predicting crimes," in *Applications and Innovations in Intelligent Systems XII*, A. Macintosh, R. Ellis, and T. Allen, Eds. Springer London, 2005, pp. 19–32. [Online]. Available: <http://dx.doi.org/10.1007/1-84628-103-2-2>
- [18] C. Bennell and D. Canter, "Linking commercial burglaries by modus operandi: tests using regression and roc analysis," *Science & Justice*, vol. 42, no. 3, pp. 153 – 164, 2002. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1355030602718200>
- [19] M. Tonkin, T. Grant, and J. W. Bond, "To link or not to link: a test of the case linkage principles using serial car theft data," *Journal of Investigative Psychology and Offender Profiling*, vol. 5, no. 1-2, pp. 59–77, 2008. [Online]. Available: <http://dx.doi.org/10.1002/jip.74>
- [20] J. Ratcliffe, *Strategic Thinking in Criminal Intelligence*. Federation Press, 2009. [Online]. Available: <http://books.google.ch/books?id=YJNdPgAACAAJ>
- [21] M. E. Wolfgang, R. Figlio, and T. Sellin, *Delinquency in a Birth Cohort*, ser. Studies in Crime and Justice. University of Chicago Press, 1987. [Online]. Available: <http://books.google.ch/books?id=aWNcXWv0C>
- [22] L. E. Cohen and M. Felson, "Social change and crime rate trends: A routine activity approach," *American Sociological Review*, vol. 44, no. 4, pp. pp. 588–608, 1979. [Online]. Available: <http://www.jstor.org/stable/2094589>
- [23] D. L. Weisel, *Analyzing repeat victimization*. US Department of Justice, Office of Community Oriented Policing Services, 2005.
- [24] L. W. Sherman, P. R. Gartin, and M. E. Buerger, "Hot spots of predatory crime: Routine activities and the criminology of place\*," *Criminology*, vol. 27, no. 1, pp. 27–56, 1989.
- [25] L. Grossrieder, F. Albertetti, K. Stoffel, and O. Ribaux, "Des données aux connaissances, un chemin difficile: réflexion sur la place du data mining en analyse criminelle," *Revue Internationale de Criminologie et de Police Technique et Scientifique*, vol. 66, pp. 99–116, 2013.
- [26] C. Bennell, B. Snook, S. Macdonald, J. C. House, and P. J. Taylor, "Computerized crime linkage systems: A critical review and research agenda," *Criminal Justice and Behavior*, vol. 39, no. 5, pp. 620–634, 2012. [Online]. Available: <http://cjb.sagepub.com/content/39/5/620.abstract>
- [27] B. Snook, K. Luther, J. C. House, C. Bennell, and P. J. Taylor, "The violent crime linkage analysis system a test of interrater reliability," *Criminal Justice and Behavior*, vol. 39, no. 5, pp. 607–619, 2012. [Online]. Available: <http://cjb.sagepub.com/content/39/5/607.short>
- [28] P. Checkland, "Soft systems methodology: a thirty year retrospective," *Systems Research and Behavioral Science*, vol. 17, pp. 11–58, 2000.
- [29] M. I. Henig and J. T. Buchanan, "Solving mcdm problems: Process concepts," *Journal of Multi-Criteria Decision Analysis*, vol. 5, no. 1, pp. 3–21, 1996. [Online]. Available: [http://dx.doi.org/10.1002/\(SICI\)1099-1360\(199603\)5:1;3::AID-MCDA85;3.0.CO;2-6](http://dx.doi.org/10.1002/(SICI)1099-1360(199603)5:1;3::AID-MCDA85;3.0.CO;2-6)
- [30] O. Ribaux and P. Margot, "Inference structures for crime analysis and intelligence: the example of burglary using forensic science data," *Forensic Science International*, vol. 100, no. 3, pp. 193 – 210, 1999. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0379073898002138>
- [31] P. Jaccard, *Nouvelles recherches sur la distribution florale*, 1908.
- [32] F. Albertetti and K. Stoffel, "From police reports to datamarts: Towards a crime analysis framework," in *Computational Forensics*, ser. Lecture Notes in Computer Science, S. Srihari and K. Franke, Eds. Springer Berlin / Heidelberg, 2012.