NONPARAMETRIC APPROACHES FOR DISCOVERING TRIGGERING EVENTS
FROM SPATIO-TEMPORAL PATTERNS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

BERNA BAKIR BATU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF
DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

OCTOBER 2014

NONPARAMETRIC APPROACHES FOR DISCOVERING TRIGGERING EVENTS FROM SPATIO-TEMPORAL PATTERNS

submitted by **BERNA BAKIR BATU** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in **Information Systems** Department, **Middle East Technical University** by,

Prof. Dr. Nazife Baykal
Director, Graduate School of **Informatics Institute**

———————————

Prof. Dr. Yasemin Yardımcı Çetin
Head of Department, **Information Systems**

———————————

Assist. Prof. Dr. Tuğba Taşkaya Temizel
Supervisor, **Information Systems**

———————————

Prof. Dr. Hafize Şebnem Düzgün
Co-supervisor, **Mining Engineering**

———————————

Examining Committee Members:

Prof. Dr. Yasemin Yardımcı Çetin
Information Systems, METU

———————————

Assoc. Prof. Dr. Pınar Karagöz
Computer Engineering, METU

———————————

Assoc. Prof. Dr. Selim Aksoy
Computer Engineering, Bilkent University

———————————

Assoc. Prof. Dr. Banu Günel
Information Systems, METU

———————————

Assist. Prof. Dr. Tuğba Taşkaya Temizel
Information Systems, METU

———————————

Date: 08.10.2014

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: BERNA BAKIR BATU

Signature   :

ABSTRACT

NONPARAMETRIC APPROACHES FOR DISCOVERING TRIGGERING
EVENTS FROM SPATIO-TEMPORAL PATTERNS

Bakır Batu, Berna

Ph.D., Department of Information Systems

Supervisor   : Assist. Prof. Dr. Tuğba Taşkaya Temizel

Co-Supervisor   : Prof. Dr. Hafize Şebnem Düzgün

October 2014, 90 pages

Temporal or spatio-temporal sequential pattern discovery is a well-recognized important problem in many domains such as seismology, criminology and finance. The majority of the current approaches are based on candidate generation which necessitates parameter tuning such as definition of a neighborhood, an interest measure and a threshold value to evaluate candidates. However, their performance is limited as the success of these methods relies heavily on parameter settings. In this thesis, two sequential pattern mining algorithms are developed for the multi-type spatio-temporal point patterns based on the nonparametric stochastic declustering methodology. The algorithms use multivariate conditional intensity model to define triggering relations within and among the event types and employs the estimated model to extract significant triggering patterns. They initially estimate pairwise triggering probabilities of all instances according to the multivariate Hawkes model, and then generate candidate patterns by using a rank selection method. Since a pair of instances is associated with a triggering probability, the proposed approaches also allow user to evaluate the significance of the pairwise pattern of any event type.The proposed methods are tested with synthetic data sets exhibiting different characteristics. The method gives good results that are comparable with the methods based on candidate generation in the literature. It is observed that the discretization of the density function based on the

significant interaction ranges obtained by Diggle D-function maximizes the triggering probabilities of the patterns that exist at similar scales. The method is tested with real data to estimate the effects of the speed bumps on the number of accidents reported in METU Campus.

Keywords: Spatio-temporal sequences, Hawkes processes, stochastic declustering, space-time clustering, Diggle D function

ÖZ

MEKANSAL-ZAMANSAL ÖRÜNTÜLERDEN BİRBİRİNİ TETİKLEYEN OLAYLARI BULMAK İÇİN PARAMETRİK OLMAYAN YAKLAŞIMLAR

Bakır Batu, Berna

Doktora, Bilişim Sistemleri Bölümü

Tez Yöneticisi : Yrd. Doç. Dr. Tuğba Taşkaya Temizel

Ortak Tez Yöneticisi : Prof. Dr. Hafize Şebnem Düzgün

Ekim 2014 , 90 sayfa

Zamansal veya mekansal-zamansal sıralı desen keşfi, deprem bilimi, suç bilimi ve finans gibi birçok alanda tanınan önemli bir problemdir. Mevcut yaklaşımların çoğunluğu aday üretme yöntemine dayalı olup, bu adayların değerlendirilmesinde kullanılmak üzere, komşuluk, anlamlılık ve eşik değeri gibi parametrelerin belirlenmesine ihtiyaç duyar. Ancak, bu yöntemlerin performansı seçilen parametre değerlerinden oldukça etkilenmektedir. Bu tezde, farklı olay tiplerini içeren mekansal-zamansal nokta desenleri için parametrik olmayan stokastik ayrıştırma metodolojisine dayalı iki adet sıralı örüntü çıkarımı algoritması geliştirilmiştir. Yöntemler, çok değişkenli koşullu yoğunluk modeli kullanarak aynı ve farklı olay tipleri arasındaki ilişkileri modelleyip bu modelden anlamlı tetikleme ilişkilerini çıkarır. Öncelikle, tüm örnek ikilileri arasındaki tetikleme olasılıkları çok değişkenli Hawkes modeli ile tahmin edilir, daha sonra anlamlı ikililer rank yöntemi ile seçilerek örüntüler belirlenir. Her bir ikili bir olasılık değeri ile tanımlandığı için, önerilen yöntemler ile, tüm olay tipleri için olası ikili tetikleme örüntülerinin anlamlılığı değerlendirilebilir. Yöntem farklı özelliklere sahip sentetik veri setlerinde denenmiş ve literatürdeki aday üretme yaklaşımına dayalı yöntemler ile kıyasla iyi sonuçlar vermiştir. Yoğunluk fonksiyonunda kullanılan kesikleştirme parametreleri Diggle D fonksiyonu ile elde edilen anlamlı etkileşim mesafeleri kullanılarak tanımlandığında, benzer ölçekte etkileşim gösteren örüntülerin olasılık değerlerinin maksimuma

ulaştığı gözlenmiştir. Önerilen metodoloji ODTÜ kampüsünde bulunan hız kasisleri-nin kampüs içinde kaydedilen trafik kazalarının sayısı üzerindeki etkisinin incelenmesi amacı ile gerçek bir veride test edilmiştir.

Anahtar Kelimeler: Mekansal-zamansal diziler, Hawkes süreçleri, olasılıksal yeniden kümelendirme, mekansal-zamansal kümeleme, Diggle D fonksiyonu

dedicated to my endless love Umut Batu

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1   Motivation

A spatio-temporal triggering pattern of event types is a sequence of spatio-temporal events where a former event generates the one that is following it. A spatio-temporal event is defined by the time and location at which the event took place and the type of the event. Despite the growing interest in the analysis of spatio-temporal events, the triggering relationships between the different spatio-temporal event types have not been sufficiently investigated. However, in many real life processes, a sequence of different event types might be observed and this sequence may suggest a triggering or a cause-and-effect type relationship within some spatio-temporal interaction scales. For example, it is probable to observe subsequent natural disasters such as flooding after hurricane, tsunami after earthquake, etc. In addition, an environmental damage by human action, i.e., deforestation or land destruction, can cause natural disasters. An example can be given from Black Sea coast of Turkey where a set of natural hazards occurred after construction of the coastal road. The region has a long coastline with rocky steep slopes rising directly to the mountains which makes transportation difficult in the region. Although the coastal road was considered as a solution to the transportation problem of the region, it induced several hazards due to the inappropriate risk management. For instance, in Rize, a deadly flooding occurred after heavy rain in 2010 [2]. It was stated that the road, as a dam, prevented water flowing from the valleys to the sea (See Figure 1.1). 12 people were dead and about 400 houses located between the road and the valleys suffered from flooding and landslides. Another hazard was in Espinye, Giresun where landslides caused huge rocks to drop onto the road in 2008 [18]. It was claimed that quarrying too much rock from the hillside during the construction of the road weakened rock and soil beds (See Figure 1.2).

Examples can be extended to various application areas. Crime events, for instance, can be observed after some other activities such as assaults or gun use after football matches or burglary events observed in a district after a shopping center is opened; a disease type may cause an increase to observe another disease type within a neighborhood as well as possibility of triggering new instances of itself; traffic accidents observed in an area may decrease after some preventive activities applied such as placing a camera or speed bump, and so on. Although some of these examples might be trivial in terms of expected causality, some others may suggest unknown but useful relationships. It can still be important to find the degree of relationship for those which is already expected in order to assess the risk. These information can be crucial for the government
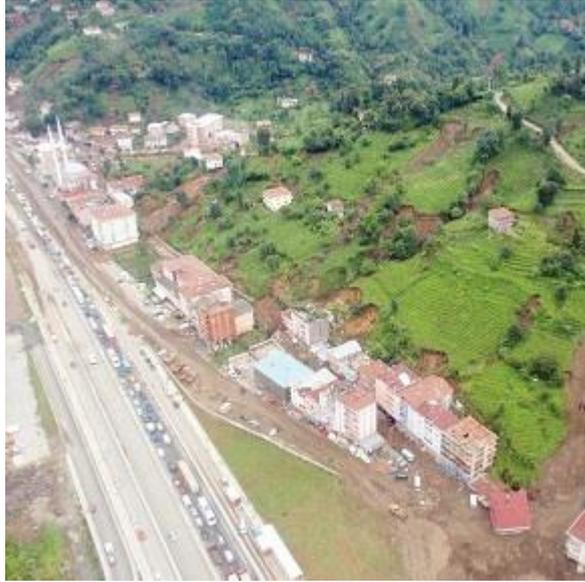
Figure 1.1: Flooding occurred in Rize, Turkey, in 2010. Photo: http://www.hurriyet.com.tr

Figure 1.2: Landslides occurred in Giresun, Turkey, in 2008. Photo: http://www.hurriyet.com.tr

agencies in order to take the necessary precautions to avoid disease and crime spread, mitigate the occurrence of disasters and improve road traffic safety.

A spatio-temporal triggering pattern of event types has four dimensions as illustrated in Figure 1.3: spatial, temporal, casual, type. They play an important role to understand the behavior of the underlying processes. Besides complexity of the problem due to the multidimensionality, each dimension may also introduce additional complexity by the characteristics of the individual processes.

Spatial Dimension

In geography, the spatial dimension describes the location of a spatial feature. Based on the underlying phenomena, a spatial feature can be represented by points (i.e., crime incidents), lines (i.e., roads) or polygons (i.e., city) in a coordinate space. One or more random variables explaining properties of the spatial features can also be associated with their location. If each location of a point process is labeled with one or more value, it is called marked point process. For example, the location and the magnitude of an earthquake correspond to the point process' spatial dimension and mark respectively. The spatial dimension may introduce additional complexity to the problem due to the following characteristics:

1. A spatial process may be stationary or not. In a stationary process, the distribution is the same across the region.

2. A spatial process may be isotropic or not. In an isotropic process, the distribution is the same across the region regardless of the direction.

3. Spatial interaction may exist at different scales. Based on the scale used, a global view of the processes may be explored or local interactions may be identified.
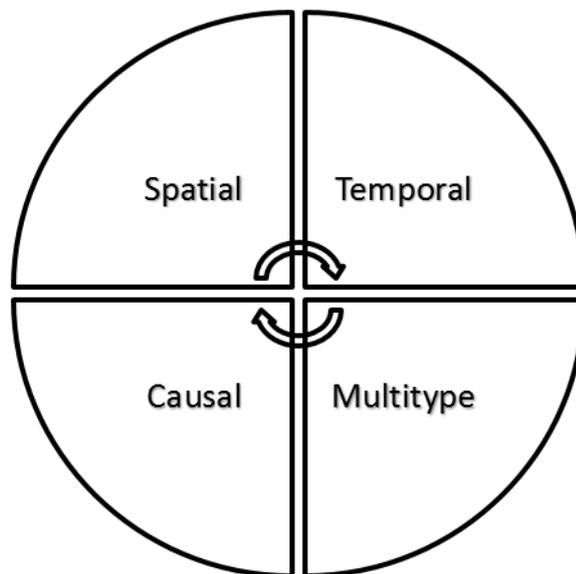
Figure 1.3: Dimensions of the problem

Temporal Dimension

Temporal dimension of a feature describes its existence or occurrence within time and exhibits the following characteristics:

1. There may be a seasonal effect in the data due to the non-stationarity.

2. Temporal behavior of the process may be independent of the spatial behaviour or there may be space-time interaction in the process.

3. Temporal interaction may exist at different scales.

Casual Dimension

Causality is the relationship between the cause and the effect. Early philosophers usually discussed causality as a concept. David Hume, on the other hand, gave a concrete definition of causality which forms a basis for causal modeling [15]. According to Hume, A causes B if

- A precedes B in time.

- A and B are contiguous in space-time.

- A and B always co-occur (A is sufficient for the B)

Another philosopher, John Stuart Mill argued that causality could be confirmed with experimentation [25]. In other words, any other explanations of the relationship have to be removed by controlled experiments. His perspective of causality provided more practical implementations with four general methods: the method of agreement, the method of difference, the method of concomitant variation, and the method of residues. The third rule of causality defined by Hume (and also as in deterministic causality) has some practical difficulties in real life examples because of its regularity. According to this definition, almost none of the phenomena could be explained by causal relationship. A clear example discussed in many studies is the relationship between smoking and lung cancer. Several empirical studies suggest that "smoking causes lung cancer". The third rule of Hume claims that anyone who smokes would have lung cancer if smoking causes it. It is known that there are some people who smoke and do not have lung cancer, therefore, this relationship does not hold given the rule. However, in reality, there may be some other causes of lung cancer that cannot be controlled or observed directly so that smoking may not be sufficient itself. Causality can exist in different forms:

1. Causal relation may exist between the same type of events or different types of events.

2. Causal structure can be in the form of chain of events, common cause, common consequence, or mixture of three.

Type Dimension

Each event type introduces a new count process in the problem. Each process usually has different behavior. It is hard to define a joint distribution for all of them. Some difficulties can be:

1. There may be interaction between the same type of events, or among the different types of events (co-occurrence).

2. Strength of an interaction may vary for different event types.

3. Different event types may have different probabilistic behavior.

1.2   Problem Definition and Research Questions

Let D denotes a spatio-temporal event database. Each record in D consists of fields representing spatial index (s) where the event occurs, occurrence time (t) of the event and the event type (e). D can be defined as $D = \{(s, t, e)|s \in R^d, t \in T, e \in E\}$ where R^d is d-dimensional continuous space, T is continuous time domain and E is set of event types. A spatio-temporal triggering pattern of interest is an ordered list of event types and can be in one of the following form according to underlying casual relationship.

1. Chain of events: $[e_1, e_2, \ldots, e_n]$ is a chain of events where e_i and e_j are different event types for $i \neq j$; $t_1 < t_2 < \cdots < t_n$; (t_{i+1}, s_{i+1}) is significantly close to (t_i, s_i) based on a given spatio-temporal neighborhood function; and e_{i+1} is conditionally independent from $e_k, k = 1, \ldots, i - 1$, given e_i.

2. One or more consequent events with a common cause event: $[e_1, (e_2, \ldots, e_n)]$ is an event list with a common cause event e_1 where e_i and e_j are different event types for $i \neq j$; $t_1 < (t_2, \ldots, t_n)$; (t_i, s_i) is significantly close to (t_1, s_1) for $i = 2, \ldots, n$ based on a given spatio-temporal neighborhood function; and there is no order among the consequent events. So, consequent events form an association instead of sequence.

3. One or more cause events with one consequence: $[(e_1, e_2, \ldots, e_{n-1}), e_n]$ is event list with a common consequence event where e_i and e_j are different event types for $i \neq j$; $(t_1, t_2, \ldots, t_{n-1}) < t_n$; (t_n, s_n) is significantly close to (t_i, s_i) for $i = 1, \ldots, n - 1$ based on a given spatio-temporal neighborhood function; and there is no order among the cause events. So, cause events form an association instead of sequence.

In this thesis, particularly, pairwise triggering patterns which can be the subset of any type of causal structure defined above are studied. A pairwise triggering pattern can be, for example, $e_k \rightarrow e_{k+1}$ where $1 \leq k < n$, $e_1 \rightarrow e_k$ where $1 < k \leq n$, and $e_k \rightarrow e_n$ where $1 \leq k < n$ for the chain, common cause and common consequence structure, respectively. The following research questions are answered:

1. What kind of algorithm can be used to extract pairwise triggering patterns of event types by considering causal relationship?

2. Can a spatio-temporal neighborhood which is not bounded by threshold be defined ?

3. Can a probabilistic neighborhood function which allows pair of sample instances to participate in support or significance value of the pattern with some degree be defined?

4. Can the probability of being a cause and effect pair between the pair of sample instances be calculated? Can degree of casual relationship between the types be measured based on these values? (i.e., summary statistics gives the estimated probability of extracted pattern.)

5. How significant patterns are identified?

1.3   Contributions of the Thesis

In this study, two algorithms are proposed in order to discover pairwise triggering patterns of spatio-temporal event types. In the algorithms, the relationships between events are modeled by using multivariate Hawkes model which is a class of mutually exciting point process. In order to estimate the mean rate and the total intensity of a cluster process, stochastic declustering methodology is used with multivariate Hawkes model. The model assumes that the process consists of events that occur randomly with a mean rate and events triggered by these random events. The conditional intensity of the entire process is the intensity at a particular time and location including both random and triggered events. Specifically, the conditional intensity model is known as self-exciting or Hawkes model which has a branching structure [11]. A typical application of the stochastic declustering methodology is to form earthquake catalog by discriminating main shocks and aftershocks probabilistically [44] which utilize parametric and nonparametric approaches based on the space-time branching models. The proposed methodologies are based on the model independent stochastic declustering (MISD) algorithm [24].

The main contributions of this thesis are as follows:

1. In this thesis, stochastic declustering methodology is utilized in sequential pattern mining problem which is a new approach in this domain. In addition, multivariate spatio-temporal Hawkes model is utilized in the algorithm while previous studies have used univariate models.

2. The proposed approach does not require candidate generation, instead it calculate triggering probabilities between the pairs of event instances and the probabilities for being a random or triggered event. Based on these values, it extracts significant triggering patterns from the data. The significant pairs are filtered with rank selection and pairwise triggering probabilities are used to calculate interest measure.

3. The algorithms proposed use similar approach for pattern extraction. However, they differ in terms of the estimation procedure. This difference is the result of the different conditioning used to define random variables for the branching

6

structure. More specifically, in one method the random variable is conditioned on the type of the previously occurred events whereas in the other method it is conditioned on the type of the current and the previously occurred events.

4. Unlike existing algorithms, this approach does not require a neighborhood threshold since the interaction is explained by the spatio-temporal density function of distances. In this thesis, histogram density estimation method is used.

5. The level of discretization used to estimate the density function has an effect on the pairwise triggering probabilities, thus, significance of patterns may differ based on the discretization level chosen. This influence is empirically handled with Diggle D function [8] the ranges of which are used for supervising the learning process. D function explores possible space time clustering or interaction with respect to separation distances. If the discretization level corresponds to spatio-temporal interaction range suggested by D function for a given pairwise pattern, the proposed algorithms produces the highest probability for the pattern. As the smaller or larger ranges are used, the pattern probabilities decrease. On the contrary, one can use the algorithm for different discretization levels and find the interaction scale of the extracted patterns as the discretization values that maximizes the pattern probability.

6. The proposed algorithm is able to extract most of the patterns regardless of the discretization level used since the rank of the pattern probabilities are still high despite their low probability values.

1.4 Organization of the Thesis

A brief introduction is already given in this chapter. Chapter 2 discusses related studies in the literature and gives some preliminary information about the Hawkes processes and stochastic declustering algorithm. Chapter 3 explains the methodologies proposed in this thesis. Chapter 4 gives the descriptions and statistics of the synthetic datasets for the simulated scenarios and discusses the results of the methods on these datasets. Chapter 5 explains the case study for a real world problem and evaluates models on real data sets. Chapter 6 gives summary and conclusion for the study and discusses future directions.

Some part of works in Chapter 3 and Chapter 4 were presented at Spatial Statistics Conference held in Columbus, Ohio in 2013. The presented work can be found in Appendix D.

CHAPTER 2

LITERATURE REVIEW AND BACKGROUND

In this chapter, relevant studies in the literature are discussed and the methods used as a basis to investigate the research problem and develop proposed method are given. In Section 2.1 and Section 2.2, the studies and the methods related to causal modeling and sequential pattern mining are addressed. In the rest of the chapter, a summary of the methods used as a background in this thesis is provided.

2.1   Causal Modeling

Causality is defined as a relationship between a cause and its effect. Identifying causal relationships between different phenomena is a challenging issue and has attracted many researchers for years since better understanding of such relationships enables to control consequences or effects.

Probabilistic causation has emerged in 70s and aims to identify the relationship between cause and effect using probability theory [34]. The central idea is the following: causes increase the probabilities of their effects and they can be expressed by conditional probabilities. This notion made casual analysis of phenomena in natural and social science feasible and verifiable based on the observations. There are several methodologies used for causal modeling. Two common approaches are functional models and graphical models. Structural equation model is a functional model which is used to hypothesize causal relationships among variables and test it with a linear equation system. Its formal definition using counterfactuals is made by Pearl [34]. Model may include hidden variables as well as observed ones. Structural equation consists of a number of multiple and simple regression models whose structure defines the relationships among the variables. Structural equation model is widely used in social science and economics. In [3], for example, long term effect of parental divorce on children is examined with functional models. Another modeling approach is causal Bayesian networks which is a type of directed acyclic graph. A Bayesian network represents probabilistic conditional dependency structure among the variables graphically [19]. Each node of the network stands for a variable or state and is associated with a conditional probability distribution. Directed arcs correspond to direct influence or dependency between variables. By this way, probability of a certain outcome in a node can be inferred given the parent nodes. A particular Bayesian network represents a specific decomposition of the joint probability distribution of the variables into product of conditional probabilities. The advantage of this representation is to decrease

complexity of the problem by decomposing it into smaller problems [4].

The methodologies used to model causal relationships may result in misleading conclusions because of the uncontrolled observational data used in the modeling. It is not certain whether variation in effect is only caused by the observed variables. Causal structure is usually complex and unknown. The correlation, time order and spatial closeness are all important clues about the causal structure. However, these dependencies may also be the result of the non-causal relations. Pearl [34] discussed identifiable causal relations under some sampling schema. It is important to perform further investigations on the discovered relationships in order to obtain more valid results.

2.2 Sequential Pattern Discovery

2.2.1 Temporal Sequence Mining

Sequential pattern mining studies can be found in both temporal and spatio-temporal domains. In temporal domain which is more dominant in the literature, two main problems, pattern matching and trend analysis were examined. The goal of pattern matching is to find specific order of occurrences whereas trend analysis aims to model deterministic or stochastic trends. Pattern matching problem has been extensively studied after it was first introduced by Agrawal and Srikant to solve sequential pattern mining problem in customer transaction databases [1]. Apriori based candidate generation approach they proposed has provided a basis for many researches. The algorithm finds frequent maximal k-sequential patterns that exceed a user specified minimum support threshold in a data set. In the basic algorithm, the time dimension is implicitly involved in the learning phase to define the order of events as before/after relationship; thus, all realization of a particular order contributes to its support count equally regardless of the time interval between them. However, relating events distant in time may be insignificant since their realization may be independent. As a solution to this problem, several window-based approaches were proposed to define closeness of events in time. For example, Srikant and Agrawal added time constraints called minimum gap and maximum gap to their first algorithm [38]. By this way, only the events which satisfy the time constraints were allowed to be in a sequence. Mannila et al. presented a general framework for discovering frequent episodes in an event sequence [23]. An event is an element of a set of different event types. Unlike the transaction data consisting of a set of customer sequences, here the data to be mined is a long event sequence as a whole which is ordered by occurrence time of the events. Patterns to be mined a.k.a. episodes are set of events which are serially ordered or parallel and are given in advance. Since the data consists of a single sequence, support count used in transaction data cannot be used directly to determine frequent episodes. Therefore, a time window with a user-specified width is used and by sliding it along the sequence, the number of windows that contain the episode is counted to determine support value. The fraction of windows in which an episode occurs gives the support value for the episode. Users also specify the value for the minimum support as well as for the confidence of the extracted rules. One drawback of these window-based solutions is the sensitivity of the algorithms to these thresholds. The results may vary significantly depending on the threshold values. Although in some domains these values can be

known and adequately provided by the domain experts, in some other domains they may be unknown. For example, time gaps between natural phenomena may not be expressed sufficiently with a constant window width. Garofalakis et al. [10] added a new constraint to the problem which is called regular expression that has to be specified by the user. The main idea is the fact that users are not interested in all frequent sequences. Instead, they are seeking for some specific patterns. Therefore the algorithm prunes the candidates according to given constraints. It is computationally efficient compared to Apriori although the general structure of the algorithm is similar. This approach provides no benefit if users want to discover previously unknown patterns. Alternatively, it can provide some benefits in terms of efficiency since user can limit the search with a subsequence that is desired to be included in the resulting sequences even though the exact pattern is not known. Similar to the problem in [23], Cao et al studied pattern matching in a long sequence [5]. The data, however, is not limited to time ordered sequence. It can be any sequence such as characters which have a meaningful order. This type of problem can be found in some domains, for instance, DNA or amino acid sequence in biology. In the study, they searched for periodic patterns in the data which are not known previously. Corresponding to time window, a period must be defined in order to find a periodically observed pattern. The difference between the time window and the period is that the time windows are overlapping windows since they are slid through the sequence but periods are non-overlapping windows which divide the sequence into equal length subsequences.

2.2.2   Spatio-temporal Sequence Mining

Spatio-temporal analysis of point pattern is a more recent topic compared to the extensive studies on spatial point patterns analysis. In these studies, taking snapshots of time is a popular approach however it has some drawbacks [16], [17], [7]. The main drawback is the possibility to miss the patterns at different time granularities; therefore, domain knowledge is essential in order to choose an appropriate discretization.

In the context of spatio-temporal sequential pattern mining, most of the studies focus on object trajectory problem which deals with the movement of the same object. The main objective of these studies is to find frequent routes followed by an object [6] or to detect motion in video streams [39], [20]. These studies are relevant but different from the problem studied in this thesis in a variety of aspects. The main difference is the fact that moving object studies deal with how the location of a same object changes over time. They assume that sequences are the consecutive locations of the same object whereas this study does not concern with the movement. In this study, sequences consist of different objects or events types which are realized in one location and do not move such as a traffic accident at a particular location and time. Observation of an event at another location and time is just another realization of the process. In some papers, moving object problem is handled by event-based approaches [41], [42], [12]. For example, Hornsby and Cole [12] model dynamic objects such as automobiles, planes, boats, etc. in geospatial domain to track the movement of the objects by using an event-based approach. They partition the spatial domain into non-overlapping contiguous regions and define the object movements as events such as "change zone event", "unexpected zone event", etc. By this way, movement of an object forms a sequence of event locations. Their approach to find patterns aggregates events in

a pairwise manner according to a time window specifying minimum and maximum allowable gaps between events. The window defines a neighborhood that relates the events in a sequence according to time. Depending on the definition of event; a pattern can be a movement of the same object, collocations of objects or movement of different objects of the same or different types together. Collocation studies usually operate on spatial domain and aims to identify different types of spatial features, instances of which are frequently located together. This relationship can be measured by cross-K function or mean nearest neighbor distance [13]. An extension of collocation patterns in spatio-temporal domain is conducted by Çelik et al. [7] which investigates patterns at different time slots. Their study deals with the collocated objects which are not located at one location, instead they are moving in space together. An example application discussed in the paper is the movements of players in different roles such as kicker or holder in a football match. Time is involved as snapshots and their method employs candidate generation approach with a proposed interest measure.

Studies similar to or directly related to the focus of the problem in this thesis are very limited. An analysis of spatio-temporal sequence of different event types can be found in [40], [14] and [26]. In all three studies, mining algorithms are suitable for spatio-temporal event data and they use candidate generation approach. The main contributions of the last two are new interestingness measures proposed. Wang et al. [40] develop an algorithm to extract flow patterns of event sets. An event set is the set of events that occur at the same time or time interval. A flow is a relation between the event sets within and among the regions. In other words, a flow pattern shows how events change or evolve over regions and time. They study on a discrete spatio-temporal domain and use window based neighborhoods. The approach in the study is based on candidate generation, however, the candidates are generated by using only 2-length sequences. The authors use a summary tree to keep the frequent patterns information and use dept first approach for mining. The main contributions of this study are to incorporate spatio-temporal neighborhood relation in sequential pattern mining problem and improve efficiency of the candidate generation approach by eliminating irrelevant patterns identified by the 2-length sequences and spatial constraints. In [14], a framework is proposed for mining event-based spatio-temporal sequences. They define a follow predicate to identify sequences based on a spatio-temporal neighborhood which is bounded by user specified thresholds and a significance measure called sequence index to determine significant sequences. They also give the dynamic neighborhood definitions such as a spatial neighborhood which is a function of time as well as different neighborhood definitions for different event types. The significance measure they proposed is calculated based on the average density ratio of subsequent events in the given neighborhood relative to its overall density within the whole space. One advantage of calculating significance measure relative to individual distributions is to prevent sequences to be labeled as significant just because of the high density of an individual event. A sequence index can be interpreted as cross-K function; however, it is a more generic measure since it allows incorporation of a temporal predicate and inclusion of more than two types. Mohan et al. [26], aim to identify cascading events in an event dataset over a common spatio-temporal framework and extend the sequence discovery problem for the partially ordered patterns. They propose an algorithm for mining cascading spatio-temporal patterns (CSTP) and filtering strategies to prevent generating uninteresting candidates. They define a new significance measure called cascade participation index (CPI) which is the minimum of participation ratios of

each event type in the sequence. They define participation ratio of an event type in a sequence as conditional probability of the sequence given a participating event type. It is estimated as the number of instances of the event type participating in the sequence over the number of instances of the event type in the database. It is shown that CPI is an upper bound for the space-time K-function. The authors state that the focus of the study conducted in [26] is the computational performance of the CSTP algorithm, hence they do not address the effect of the parameter choice such as neighborhood and significance measure thresholds and grid cell size defined for filtering.

2.2.3 Summary

As a summary, a great majority of temporal and spatio-temporal sequence studies use a window-based approach to define closeness of events or objects in time and space. The window is constant and bounded by a user specified threshold. All the events falling in the same window are considered as neighbors. Small threshold values may result in missing the true patterns at larger scales whereas higher values may produce irrelevant patterns. In addition, the time distances between events in a window are treated as the same and contribute to the significance measures such as support count equally while the ones larger than the threshold have no contribution. The window based neighborhood definition may be suitable for some domains, such as DNA or character sequences; however, it is not a realistic assumption in some other domains where the events of a sequence interact with each other according to a decay law. For instance, a main shock earthquake produces its aftershocks with a decreasing influence as time passes. A time window cannot capture the true relations in such sequences. Candidate generation is a common approach in sequential pattern mining, yet it poses a challenge of exponential increase in the number of candidate patterns with respect to the number of types. In some studies such as [26], a number of filtering strategies are used to tackle this challenge. Generally the number of windows in which a specific pattern is observed gives the support value for that pattern. A minimum support value is also required to be specified which is domain dependent and should be provided by domain experts. An important problem of this approach is the fact that in some domains these values may be provided easily but sometimes specifying such values may be hard especially if the underlying behavior is unknown. Another limitation is the necessity of examining individual distribution of event types or elements in a sequence before performing a mining task. A frequent sequence is not necessarily a meaningful pattern. For example, two regular events may be observed in a time window long enough to include both with a specific order although their occurrences are independent. Furthermore, a frequent event sequence extracted by measures like support does not necessarily imply a triggering or causal relationship between event types although time order is a clue about causal relationship. It does not show whether the preceding event triggers the next one in the sequence, rather it shows how likely this sequence is observed. In fact, these shortcomings have significant importance if the sequential pattern mining problem aims to discover sequential triggering relationship.

2.3 Hawkes Processes and Conditional Intensity Models

A Hawkes point process defines random events which are either an immigrant or a descendant. It was first introduced in [11] for temporal domain with possible application areas such as epidemiology and modeling neuron firing. The process has an underlying branching structure. If the event locations are discarded, it is reduced to a simple branching process consisting of immigrants and descendants. Besides the univariate case, Hawkes also proposes models for mutually-exciting point processes and also examines a special case where the decay is exponential. Self-exciting Hawkes models are very popular and well-recognized in seismological studies after Ogata introduced Epidemic Type Aftershock Sequences (ETAS) model defined by the Hawkes process [31]. It defines earthquake process with a marked point process model where marks are the magnitude of the earthquake events (See for example, [30, 32, 44]). Several extensions of ETAS model have been studied for both temporal and spatio-temporal domain. Univariate Hawkes model can be defined by the intensity function conditioned on the history H_t of the process at time t. It is assumed that some events in the process occur independently with a mean rate μ, and the value of μ may be independent of the time and the location or may depend on the spatio-temporal dimensions. In this thesis, independence of μ is assumed. Intensity at a particular location and time (s,t) depends on both the mean rate and the previously occurred events with some degree based on the spatio-temporal distance of point (s,t) to these historical events. The definition of the intensity for an unmarked point process is given as;

$$\lambda(s,t|H_t) = \mu + \sum_{j:t_j<t} g(s-s_j, t-t_j; \Theta) \tag{2.1}$$

In the Equation 2.1, $g(x,y,t;\Theta)$ is the triggering function of the process where Θ represents the distribution parameters. Multivariate Hawkes model generalizes the univariate case by defining a conditional intensity for each component or event types k as follows:

$$\lambda_k(s,t|H_t) = \mu_k + \sum_{r} \sum_{v:t_v<t} \eta_{rk} f_{rk}(s-s_v, t-t_v; \Theta) \tag{2.2}$$

where η_{rk} is the average number of event type k triggered by event type r, f_{rk} is the triggering function for the pairs of event types r and k which defines the probability of an event type of r at (s_v, t_v) triggering an event type k at (s,t). The univariate Hawkes model with or without marks is employed in many studies, for example, to analyse earthquakes and crime events [44], and to model temporal triggering behaviour of activities in social media such as Twitter and Wikipedia [22, 37]. Multivariate temporal models have received significant attention in the recent years. Embrechts et. al. conduct a study on financial data to investigate triggering effect on extremes [9]. Zhou and Zha propose an algorithm for learning kernel functions of multivariate Hawkes processes where they focus on the temporal characteristics of the domain for different event types [43]. To our knowledge, there is no study yet dealing with spatio-temporal multivariate Hawkes model.

2.4 Stochastic Declustering with Parametric and Non-Parametric Models

Stochastic declustering algorithm was first introduced by Zhuang, Ogata and Vere-Jones to discriminate main shock earthquakes from their aftershocks with a probabilistic thinning procedure by using the estimated conditional intensity model describing the process [44]. The process is assumed to be consisting of the main events and the aftershocks clustered around these main events. The earthquake occurrences are considered as a marked spatio-temporal point process where marks are the magnitudes. Based on the common features of the earthquake models, the intensity function is defined by the Equation 2.3. In the equation, $\mu(s, t, M)$ is the mean rate of the background process, $\kappa(M_j)$ is the expected number of aftershocks that is excited by a main event with magnitude M_j, and g, f and z are the response functions defining relation between the main and aftershock events given the magnitude of the main event.

$$\lambda(s, t, M | H_t) = \mu(s, t, M) + \sum_{j:t_j < t} \kappa(M_j) g(t - t_j) f(s - s_j | M_j) z(M | M_j) \qquad (2.3)$$

In their paper, the authors make some additional independence assumptions and simplify the intensity model used in the algorithm as given in Equation 2.4.

$$\lambda(s, t, M | H_t) = z(M)[\mu(s, t) + \sum_{j:t_j < t} \kappa(M_j) g(t - t_j) f(s - s_j | M_j)] \qquad (2.4)$$

Well-developed theoretical aftershock models describing the branching behavior of the process, i.e., ETAS models, allow researchers in seismological science to define intensity function with parametric models the parameters of which can be estimated by maximum likelihood procedure. The probability of the events being drawn from the aftershock or background processes is calculated from the estimated intensity model of the entire process. The probability of an event being an aftershock is the sum of the pairwise excitation probabilities between the event and the events that causally precede it. Their definitions are given in Equation 2.5 and Equation 2.6.

$$\rho_{i,j} = Pr\{\text{the } i^{th} \text{ event is generated by } j^{th} \text{ event} | H_t\} = \frac{\kappa(M_j) g(t_i - t_j) f(s_i - s_j | M_j)}{\lambda(s_i, t_i | H_{t_i})}$$
$$(2.5)$$

$$\rho_i = Pr\{\text{the } i^{th} \text{ event to be an aftershock} | H_{t_i}\} = \sum_{j=1}^{i-1} \rho_{i,j} \qquad (2.6)$$

Using the kernel density estimation method, the total rate of the process can be estimated. Then, the mean rate of the background process is calculated approximately as given in Equation 2.7. The derivation steps can be found in [44].

$$\mu(s) = \frac{1}{T} \sum_i (1 - \rho_i) k_d(s - s_i) \qquad (2.7)$$

The estimation of the model is made through an iterative procedure. The iteration algorithm estimates simultaneously the mean rate and the branching behavior. The algorithm steps are given below:

Intensity Model Estimation Algorithm

Step 1: Set $l = 0$ and $\mu^{(l)} = 1$.

Step 2: Fit the conditional intensity function using MLE.

Step 3: Calculate ρ_i from Equation 2.5 and Equation 2.6 for each $i = 1, 2, \ldots N$.

Step 4: Calculate $\mu^{(l+1)}$ from Equation 2.7

Step 5: If $max|\mu^{(l+1)} - \mu^{(l)}| > \epsilon$ then set $l = l + 1$ and go to Step 2. Otherwise, take $\mu^{(l+1)}$ as the mean rate.

Once the model is estimated, the events are declustered probabilistically. Based on the final obtained intensity model, the following steps are used for probabilistic declustering:

Stochastic Declustering Algorithm

Step 1: For each event calculate probability ρ_i from the final solution of the Intensity Model Estimation Algorithm.

Step 2: Generate N uniform random numbers U_i in [0,1].

Step 3: If $U_i < 1 - \rho_i$ then keep the ith event; otherwise, remove it from the catalog as it is an aftershock. The remaining events can be considered the background events.

Although decay behavior in space and time can be explained by the specific models in the case of seismology, in many other application domains, there is no prior information about the form of decay function which describes the triggering behavior. Even in seismology, aftershock effect can be more complicated due to second generation aftershocks produced by the previous aftershocks. Mohler et al. [28] state that there is a need to refine parametric models in seismology because of the researches last over decades. Marsan and Lengline [24] propose a nonparametric approach called model independent stochastic declustering algorithm (MISD) to solve the same problem which is later used in many other domains. For example, it is used in the crime analysis to estimate the mean rate of the crime events and the dense areas instead of the hot spot analysis which is common in this domain [28].

The MISD algorithm use univariate Hawkes model and estimate the conditional intensity and the mean rate with a nonparametric kernel estimation procedure. Since the branching structure is unobservable, in other words, it is not known which event actually triggers the others, the algorithm utilizes Expectation-Maximization method. The random variable X_{ij} is defined for the branching structure and it takes value of

1 if i^{th} event is caused by j^{th} event, otherwise it takes the value of 0. Similarly, the random variable X_{ii} takes the value of 1 if it is a triggering casual event, otherwise it takes the value of 0. Since these variables cannot be observed, their expected values are estimated. Given the intensity function of unmarked point process in Equation 2.1, the algorithm works as follows:

MISD Algorithm

Step 1: Start with the initial values for the intensity and the mean rate.

Step 2: Calculate the expected values for the triggering (p_{ij}) and the background (p_{ii}) probabilities as follows:

$$p_{ij} = \begin{cases} \frac{g(s_i - s_j, t_i - t_j)}{\mu + \sum_{r=1}^{i-1} g(s_i - s_r, t_i - t_r)} & t_i > t_j \\ 0 & \text{otherwise} \end{cases} \tag{2.8}$$

$$p_{ii} = \frac{\mu}{\mu + \sum_{r=1}^{i-1} g(s_i - s_r, t_i - t_r)} \tag{2.9}$$

where, p_{ij} is the probability of the ith event being triggered by the jth event and p_{ii} is the probability of ith event being a background event.

Step 3: Update the intensity and the mean rate based on the probabilities calculated in the expectation step as follows:

$$g(\Delta s, \Delta t) = \frac{1}{\delta s \delta t} \sum_{i,j \in \Lambda} p_{ij} \tag{2.10}$$

where $\delta s, \delta t$ are the discretization parameters and Λ is the set of pairs such that $t_i - t_j \in [t - \delta t, t + \delta t)$, $s_i - s_j \in [s - \delta s, s + \delta s)$.

$$\mu = \frac{1}{TR} \sum_{i=1}^{n} p_{ii} \tag{2.11}$$

where T is the duration of time, R is the area of the studied region and n is the number of events. Step 2 and Step 3 are repeated until g and μ converge.

2.5 Spatial Statistics for Point Patterns

Ripley's K-function or second order moment measure is a widely used method for the analysis of point patterns. It can be employed in spatial or spatio-temporal domain to understand the extent of the interaction among the events or the spatial dependence at different scales [35]. Formally, it is the average number of other events within the separation distance from an arbitrary event. The definitions for spatial and spatio-temporal domains are given in Equation (2.12) and Equation (2.13), respectively, where λ is the intensity of the process.

$$K(s) = \lambda^{-1} \text{E}\{\# \text{ of events within the distance } s \text{ of an arbitrary event}\} \tag{2.12}$$

$$K(s,t) = \lambda^{-1} \text{E}\{\# \text{ of events within the distance } s \text{ and time } t \text{ of an arbitrary event}\} \tag{2.13}$$

If a point pattern is clustered due to the spatial dependence, then the plot of the K function for the observed data falls above the reference envelope representing Poisson distribution. For example, in the Figure 2.1 (a), up to distance of 4.5, the values of K function fall above the envelope. This indicates spatial dependence between the events. However, for those separated by a distance more than 4.5 there is no more dependency. To find the scale of possible clustering L function, normalized form of K, can be used. In the plot of L function peaks in the positive values indicate clustering at that scale. In the Figure 2.1 (b), for example, there are two peaks at scale of 1 and 4 which indicates clustering at these scales.
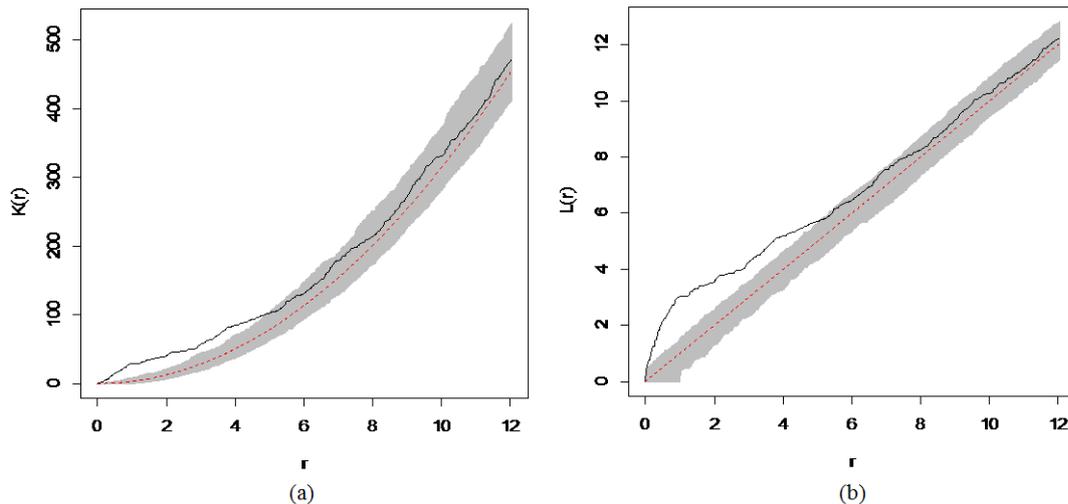
Figure 2.1: (a) K-function for sample data (b) L-function for sample data

If the spatial and temporal components of the spatio-temporal process are independent, then $K(s,t)$ can be factorized as the composition of individual components, such as,

$$K(s,t) = K(s)K(t) \tag{2.14}$$

where $K(s)$ and $K(t)$ are separate space and time functions.

In the analysis of spatio-temporal dependence Diggle [8] used Ripley's K-function as a diagnostic tool and explore significant space-time clusters with D-function which is defined as,

$$D(s,t) = K(s,t) - K(s)K(t) \tag{2.15}$$

If there is no space-time interaction or clustering Equation (2.14) holds, and thus, D-function is equal to 0. If there is an interaction, the surface of $D(s,t)$ shows peaks at the corresponding ranges. In Figure 2.2(a), an illustrative example is given. In the figure there is a peak at $s = 1.5$ and $t = 0.8$ which indicates space-time clusters at the corresponding range. These clusters are also significant based on the test statistics obtained by Markov Chain Monte Carlo (MCMC) simulation the distribution of which is given in Figure 2.2(b). There may be significant clusters at different scales. In other words, there may be local or global clusters in the data which can be explored by using different resolutions.

18
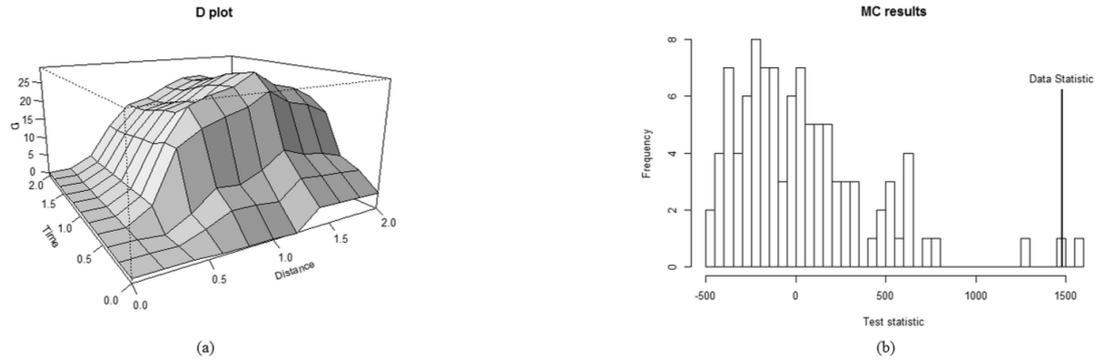
Figure 2.2: (a)D plot and (b) MCMC simulation for significance test.

2.6 Evaluations

The methods proposed in Chapter 3 were evaluated by their ability to predict actual pairs, cause events and mean rate. For the assessment, *sensitivity* (recall) and *precision* measures were used. Sensitivity shows true positives rates of the predictions, whereas precision shows how precise the positive predictions are. They are defined as

$Sensitivity = \frac{TP}{TP+FN}$

$Precision = \frac{TP}{TP+FP}$

where TP is the number of correctly classified positive samples, FP is the number of negative samples incorrectly classified as positive and FN is the number of positive samples incorrectly classified as negative. In order to compare the results with a single value, F_1 measure were also used. F_1 measure can be considered as the weighted average of the sensitivity and precision values, and defined as

$F_1 = 2 \times \frac{Sensitivity \times Precision}{Sensitivity + Precision}$

CHAPTER 3

METHODOLOGY

In this chapter, the proposed methods for the triggering pattern extraction from the spatio-temporal event data sets are explained. The proposed methods either use result of the existing intensity estimation and stochastic declustering algorithms or modify them to improve the algorithms for the multi-type pattern extraction problem. The main contributions of this thesis are provided in Section 3.3 and Section 3.4.

3.1 Method 1: Triggering Pattern Extraction by Post Processing the Results of Intensity Model Estimation Algorithm

Recall that stochastic declustering methodology proposed in [44] uses conditional intensity model with parametric density functions and it is generally used in seismology domain. In Method 1, the results of the original algorithm is used for finding triggering patterns of spatio-temporal event types.

The original algorithm does not use event types, instead, it models self-exciting behavior of the same type of events. In addition, the primary concern of the algorithm is to estimate the mean density of the background process and the branching structure, and finally to determine the main events probabilistically. Therefore, pairwise triggering probabilities which are used to calculate the probability of an event being an offspring are not examined individually. The focus in this study, on the other hand, is these individual pair relations. To evaluate whether this approach is suitable for the triggering pattern extraction problems, in other words, whether the fitted model describes the triggering behavior, a post processing approach is proposed. In this approach, the fitted model parameters are used as neighborhood thresholds and the pairs satisfying these thresholds are determined. Then, event types of selected pairs are used to extract patterns. The post processing algorithm is presented in Figure 3.1.

In seismology, there are well studied parametric models which describe behavior of seismic activities in temporal and spatial domains. However, for many application domains in real world, particularly those in social and behavioral science, domain specific models are usually unavailable and as a consequence certain known density functions such as Gaussian or exponential functions are utilized for modeling [27].

In line with the literature, certain density functions were utilized in the definition of conditional intensity model. Specifically, Gaussian density for the space function,

Input:

1) $Data = \{(s_i, t_i, y_i) | \ i \in (1,2,...,n), s_i \in R^2, t_i \in R, y_i \in (1,2,...,K)\}$.

2) Estimated parameter values, θ_1 and θ_2, for the decay in space and time.

Output:

Pairwise triggering patterns and their significance: $TP = \{(r \rightarrow k, support_{r \rightarrow k}, \ confidence_{r \rightarrow k})\}$

Algorithm:

1: $Pairs = \emptyset$

2: **for each** $i, j \in (1,2,...,n)$ and $t_j < t_i$

3:     Calculate $d(s_i, s_j)$ and $d(t_i, t_j)$

4:     **if** $d(s_i, s_j) \leq \theta_1 \wedge d(t_i, t_j) \leq \theta_2$ **then** $Pairs = Pairs \cup \{(j,i)\}$

5: Find all unique pattern $r \rightarrow k \in \{(y_j, y_i) | \ (j,i) \in Pairs\}$

6: Calculate support and confidence of each pattern over m, the cardinality of $Pairs$

7: **return** $TP = \{(r \rightarrow k, support_{r \rightarrow k}, \ confidence_{r \rightarrow k})\}$

Figure 3.1: Triggering Pattern Extraction (TPEX) by Post Processing

$f(x, y)$, and exponential and gamma densities for the time function, $g(t)$ were used. Individual behaviors of these functions are examined to understand their tendency during parameter fitting in the maximum likelihood estimation. These results can be found in Appendix B.

The results of the Method 1 on the synthetic datasets are given in Section 4.2.1 in Chapter 4. According to the results, small sample size affects the stability of the predictions. For each replication of the same data settings, the model parameters' values that maximize the likelihood function varies. In addition, these values describe the behavior of the process, as a result it is expected to have similar values for the similar data sets. The reason might be using density functions which are not appropriate to explain entire process. As discussed in Chapter 4, the data sets used in the analysis have space-time clusters. However, independence of the space and the time dimensions was assumed and factorization of the individual functions was used in the model since a complete model is not available. The proposed method relies on the use of an appropriate model and the correct estimation of the model parameters describing the interaction ranges between the events.

3.2 Method 2: Triggering Pattern Extraction with MISD Algorithm

As demonstrated in Section 4.2 in Chapter 4, parametric models are limited in terms of sensitivity in the choice of the kernel function and its parameters. In this section, MISD algorithm was studied to evaluate whether (1) nonparametric approach has superiority over parametric approach in terms of explaining the data sets which follow either a known or an unknown distribution (2) spatial and temporal dimensions can individually explain the relationship with an acceptable error rate. Two triggering kernels which are the function of time and the function of two-dimensional space were examined as well as spatio-temporal triggering kernel. Intensity model for the first two are shown in Equations 3.1 and 3.2 below. The third model is already defined in Equation 2.1. In this study, histogram estimator for the nonparametric estimation

of density functions was used. Any other approach can also be employed for learning triggering densities.

$$\lambda(t|H_t) = \mu + \sum_{j:t_j<t} g(t - t_j) \tag{3.1}$$

$$\lambda(s|H_{t_s}) = \mu + \sum_{j:t_j<t} g(s - s_j) \tag{3.2}$$

The original MISD algorithm is implemented for three intensity models. Based on the results obtained from the original algorithm, the two-dimensional spatial model was improved by using a binary constant in the definition of triggering kernel. The function was modified as follows:

$$g(\Delta s) = 1_{(\Delta s < r)} \frac{1}{\delta s} \sum_{i,j \in \Lambda} p_{ij} \tag{3.3}$$

In the equation, r is the separation distance which shows the extent of the significant spatial interaction the value of which is obtained from Ripley's K-function. This modification ensures that the value of g is zero outside the range of significant spatial clusters. It is made due to the fact that if an event in space triggers another one, it is expect that there is an interaction between the events at some spatial scales. If the pattern is random at some scales triggering kernel should be zero for the separation distances that correspond to the scales. Observed peaks at the separation distances higher than the value of r are more likely to happen by chance. Using such binary constants smoothen the function for higher distances and eliminates any effects of random noise.

The results of the original algorithm and the improved version on the synthetic datasets are discussed in Chapter 4. According to the results, the temporal dimension, alone, is not able to detect cause-effect pairs as the temporal rate increases. There is no such influence for the spatial dimension and significant improvement is observed over temporal model. Spatial and spatio-temporal triggering functions are ideal for detection of cause effect pairs. Due to the computational costs, spatial model can be preferred over spatio-temporal model since their predictive performance is comparable for the simple data sets where space-time interaction exists. However, none of the models are able to recover mean rate of the process due to the incorrect estimation of p_{ii} values although their corresponding pairwise triggering probabilities are high. To use of weights not only increase sensitivity and precision values based on the ground-truth pairs created but also result in successful predictions for μ and p_{ii} values.

3.3 Method 3: Multivariate Triggering Pattern Extraction Based on MISD Algorithm Conditioned on the Type of Triggering Events

All aforementioned studies except [9, 43], and the Method 1 and Method 2 proposed in this study employ univariate Hawkes process. However, in many real world problems, there are many event types which interact with each other as well as there are possible interactions among the same types. For example, in ecology, loss of predators due to the uncontrolled hunting causes an increase in the population of the prey species. It then

causes shortage of the food resources of the prey species which results in fighting for food and finally extinction. There are m^2 number of different pair of types including self relation where m is the number of event types. The relationship between each pair of event types is usually different. For example, the elevated risk of food shortage caused by drought can be different from the risk caused by increasing population. Decay of the effect can vary for different types as well. Therefore, modeling relationships among the large number of event types by using a univariate model which excludes the type information cannot explain such processes well and might produce insufficient results. In the results of the Method 1 and Method 2, even for the simplest case where there are only two types of events and causes are necessary and sufficient for the effects, it is observed that the results are sensitive to the parameters' values and the estimation of the mean rate is affected by noise in the triggering function.

For the limitations above it is proposed to define (1) a random variable for the branching structure which is conditioned on the type of triggering event occurred in the past and (2) a random variable for being a casual event which is conditioned on its own type.

Let Y is a discrete random variable representing the type of spatio-temporal event occurred at a particular time t and location s and there are n number of observations, $y_1, y_2, \ldots, y_n$ of $Y \in \{1, 2, \ldots, K\}$. Following conditional random variables are defined.

$$X_{ij}|Y_j = y_j = \begin{cases} 1 & \text{if event } i \text{ is caused by event } j \text{ given the type of } j^{th} \text{ event is } y_j \\ 0 & \text{otherwise} \end{cases}$$

(3.4)

$$X_{ii}|Y_i = y_i = \begin{cases} 1 & \text{if event } i \text{ occurred randomly given the type of } i^{th} \text{ event is } y_i \\ 0 & \text{otherwise} \end{cases}$$

(3.5)

Due to the unobservable branching structure, the expected values are considered as follows:

$$E(X_{ij}|Y_j = y_j) = P(X_{ij}|Y_j = y_j)$$

$$E(X_{ii}|Y_i = y) = P(X_{ii}|Y_i = y_i)$$

The estimated values for the probabilities above can be defined as:

$$p_{ij} = P(X_{ij}|Y_j = y_j) = \sum_i P(X_{ij}|Y_i = y_i, Y_j = y_j)P(Y_i = y_i) \qquad (3.6)$$

$$p_{ii} = P(X_{ii}|Y_i = y_i) = P(X_{ii}|Y_i = y_i)P(Y_i = y_i) \qquad (3.7)$$

For example, if there are two types of events, and the type of the events observed in the past is 2, then Equation 3.6 will be

$$P(X_{ij}|Y_j = 2) = P(X_{ij}|Y_i = 1, Y_j = 2)P(Y_i = 1) + P(X_{ij}|Y_i = 2, Y_j = 2)P(Y_i = 2)$$

Considering the multivariate Hawkes model in Equation 2.2, each element of the sum in Equation 3.6 can be estimated as follows:

24

$$P(X_{ij}|Y_i = y_i, Y_j = y_j)P(Y_i = y_i) = \frac{\eta_{y_j y_i} f_{y_j y_i}(s_i - s_j, t_i - t_j)}{\lambda_{y_i}(s_i, t_i)} * \frac{n_{y_i}}{n(t_i)} \qquad (3.8)$$

where $f_{y_j y_i}$ is non-parametric density estimated by the subset of the data such that all pairs are (y_j, y_i) type, n_{y_i} is the number of type y_i events and $n(t_i)$ is the total number of events up to time t_i. Similarly, the probability in Equation 3.7 can be calculated as:

$$P(X_{ii}|Y_i = y_i) = \frac{\mu_{y_i}}{\lambda_{y_i}(s_i, t_i)} * \frac{n_{y_i}}{n(t_i)} \qquad (3.9)$$

$g_{rk}(s, t)$ is defined as $g_{rk}(s, t) = \eta_{rk} f_{rk}(s, t)$ and then the mean rate μ_k and the $g_{rk}(s, t)$ are defined as:

$$\mu_k = \frac{1}{TR} \sum_{i=1}^{n} 1_{Y_i = k} p_{ii} \qquad (3.10)$$

$$g_{rk}(\Delta s, \Delta t) = \frac{1}{N_{rk} \delta s \delta t} \sum_{i,j \in \Lambda} 1_{Y_j = r} p_{ij} \qquad (3.11)$$

where T is the duration of time, R is the area of the study region, 1 is the indicator function, n is the number of observed events, N_{rk} is the number of pairs of type $(Y_i = k, Y_j = r)$ where $t_i > t_j$, δs and δt are the discretization parameters and Λ is the set of pairs such that $t_i - t_j \in [t - \delta t, t + \delta t)$, $s_i - s_j \in [s - \delta s, s + \delta s)$.

This method assumes that the type of the current event is unknown but the type for those occurred in the past are known. Two selection procedures were applied to filter out related sample instances. The first one is to use a threshold value such as 0.5 for the pairwise probabilities and the second one is to select the highest triggering probability between a given sample instance and its preceding events.

The results given in Chapter 4 show that the method is able to produce high true positive rate for the simple data sets. One drawback is low precision values which means observing high false positives. The false positive pairs are usually have smaller probability values than the true positive pairs. The distribution of the p_{ij} values has either heavy left tail or is bi-modal. This drawback is eliminated with an adjustment after selection procedure. Otsu thresholding is used to discriminate incorrect estimations from the true ones. The Otsu thresholding is a technique used in image processing to find the optimum threshold value which separates the two class of pixels in an image [33]. It assumes the values comes from a bimodal distribution. The distribution of the probability values in the initial results satisfies this assumption. For example, in Figure 3.2 and Figure 3.3,the distributions of the values obtained by the threshold and rank selection from the data set $DS1$ are illustrated. Otsu's method was applied to the cases, after selecting the pairs by using a threshold or rank. In both cases, after filtering the results with Otsu threshold, false positive rates significantly decrease while true positive rates remain high. Initial selection by using a threshold value which is 0.5 produces considerably high precision value than the initial selection by using rank. However, the result of the first one is sensitive to the threshold value used.
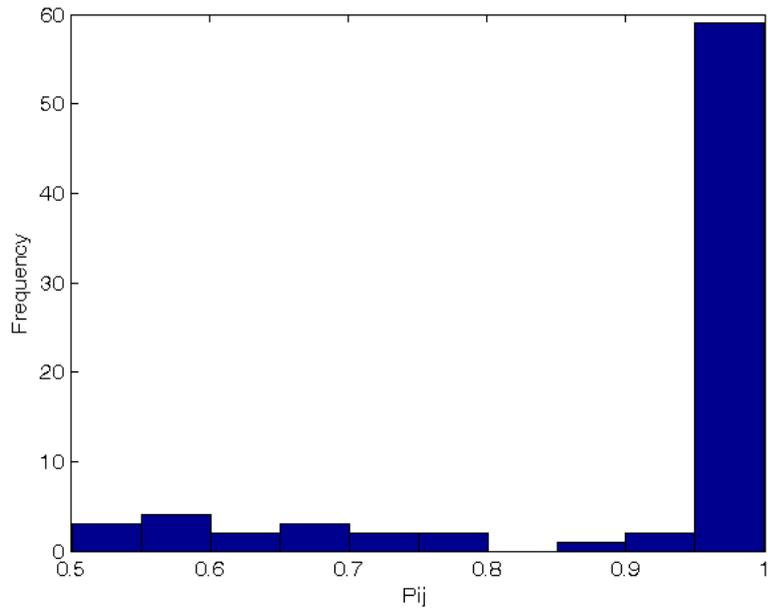
Figure 3.2: For a sample simulated data the distribution of p_{ij} which are selected by using threshold value, 0.5.
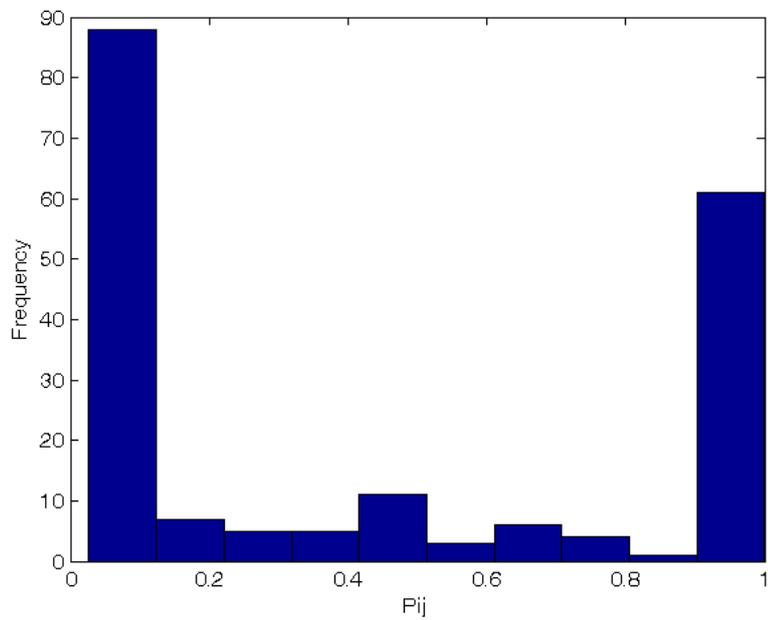
Figure 3.3: For a sample simulated data the distribution of p_{ij} which are selected by using ranking

3.4 Method 4: Multivariate Triggering Pattern Extraction Based on MISD Algorithm Conditioned on the Types of Triggering and Triggered Events

As discuss in Section 3.3, relationships among the different types can be modeled better with a multivariate model since these interactions can be too complex to represent with a simple model excluding types. Therefore, MISD algorithm was improved by defining new conditional random variables and their estimator by using multivariate Hawkes model. The proposed algorithm (Method 3) gives good results on simple datasets, however, its success is limited in the terms of complexity of the problem. This might be due to the incomplete information about the current event types used during the learning phase of the algorithm. In order to handle the problems faced in Method 3, it is proposed to define (1) a random variable for the branching structure which is conditioned on the type of triggering event occurred in the past and the type of the triggered event occurred currently, and (2) a random variable for being a casual event which is conditioned on its own type. The conditional expectations for the random variables X_{ij} and X_{ii} with respect to the type of the events were defined and the MISD algorithm was reformulated for the multivariate model.

Let Y be a discrete random variable representing the type of a spatio-temporal event occurred at a particular time t and location s. Assume that the sample space of Y consists of K number of event types and there are n number of observations, $y_1, y_2, \ldots, y_n$ of Y. The following random variables conditioned on the value of Y are defined.

$$X_{ij}|Y_i = y_i, Y_j = y_j = \begin{cases} 1 & \text{if } i^{th} \text{ event is caused by} \\ & j^{th} \text{ event given } Y_i, Y_j \\ 0 & \text{otherwise} \end{cases}$$

$$X_{ii}|Y_i = y_i = \begin{cases} 1 & \text{if } i^{th} \text{ event occurs randomly} \\ & \text{given } Y_i \\ 0 & \text{otherwise} \end{cases}$$

Then, the conditional expectations are calculated using conditional densities.

$$E(X_{ij}|Y_i, Y_j) = P(X_{ij}|Y_i, Y_j)$$

$$E(X_{ii}|Y_i) = P(X_{ii}|Y_i)$$

The conditional probability distribution of the random variables is estimated from the model as follows:

$$p'_{ij} = P(X_{ij}|Y_i, Y_j) = \frac{\eta_{Y_j Y_i} f_{Y_j Y_i}(s_i - s_j, t_i - t_j)}{\lambda_{Y_i}(s_i, t_i|H_{t_i})} \tag{3.12}$$

$$p'_{ii} = P(X_{ii}|Y_i) = \frac{\mu_{Y_i}}{\lambda_{Y_i}(s_i, t_i|H_{t_i})} \tag{3.13}$$

For the nonparametric estimation of triggering structure, take $g_{rk}(s,t) = \eta_{rk} f_{rk}(s,t)$ in Equation 2.2 and estimate the entire model. Then the mean rate μ_k and the $g_{rk}(s,t)$ can be defined as:

$$\mu_k = \frac{1}{TR} \sum_{i=1}^{n} 1_{Y_i = k} p'_{ii} \tag{3.14}$$

$$g_{rk}(\Delta s, \Delta t) = \frac{1}{N_{rk}\delta s \delta t} \sum_{i,j \in \Lambda} 1_{Y_i=k,Y_j=r} p'_{ij} \qquad (3.15)$$

where T is the duration of time, R is the area of the study region, 1 is the indicator function, n is the number of observed events, N_{rk} is the number of pairs of type $(Y_i = k, Y_j = r)$ where $t_i > t_j$, δs and δt are the discretization parameters and Λ is the set of pairs such that $t_i - t_j \in [t - \delta t, t + \delta t)$, $s_i - s_j \in [s - \delta s, s + \delta s)$.

The primary assumption for the pairwise relationship is that if there is an interaction at a particular scale, there should be many pairwise space-time distances less than the interaction range while other distance values are evenly distributed. The estimated density function will reflect this relationship with peaks if appropriate discretization values are utilized. There may be different scales where space-time interaction exists. Large discretization intervals may suppress the relation at smaller scales, therefore, the use of different values during the estimation of g_{rk} is worthy. Significant space-time clusters at different scales can be explored and tested with D function. The estimation can be supervised with these values.

The method finds pairwise triggering patterns of spatio-temporal event types. It first fits conditional intensity model by using multivariate Hawkes process and calculates pairwise probabilities for all pair of instances. It then selects significant probabilities with rank selection method and generate distinct pairwise patterns from this reduced set. It returns each significant pattern with its interest measure. The algorithm is given in Figure 3.4. A spatio-temporal dataset consisting of locations and the types of the events and the discretization values for the density functions are the inputs of the algorithm. Stopping criterion is satisfied by the convergence of the mean rate and density functions which also ensures convergence of probabilities. If the interaction ranges in space and time suggested by D-function are given to the algorithm as discretization values, it produces optimal results in terms of significance and rank values for a certain group of patterns interacting at similar scales. High discretization values

---

Input:
    1) $Data = \{(s_i, t_i, y_i)| i \in (1,2,...,n), s_i \in R^2, t_i \in R, y_i \in (1,2,...,K)\}$.
    2) Discretization values: $\delta s, \delta t$.
Output:
    Pairwise triggering patterns and their significance: $TP = \{(r \rightarrow k, significance_{r \rightarrow k})\}$
Algorithm:
1: Initialize μ_k and g_{rk}, $r, k \in (1,2,...,K)$
2: **while** μ and g not converged **do**
3:    Calculate p'_{ij} and p'_{ii} from (3.12) and (3.13)
4:    Update μ_k and g_{rk} from (3.14) and (3.15)
5: $Pair\ Probabilities = \emptyset$
6: **for each** $i \in (1,2,...,n)$
7:    $Pair\ Probabilities = Pair\ Probabilities \cup \quad \{(i,j,p_{ij}{}^*)\,|\,p_{ij}{}^* = \max(p'_{i1}, p'_{i2}, ..., p'_{ii}) \vee$
    $\max(p'_{i1}, p'_{i2}, ..., p'_{ii}) - p_{ij}{}^* < \varepsilon\}$
8: Determine set of distinct patterns $r \rightarrow k$ from $Pair\ Probabilities$
9: Calculate $significance_{r \rightarrow k} = \quad midmean(p_{ij}{}^*\,|\,y_j = r, y_i = k)$ from (16)
10: **return** $TP = \{(r \rightarrow k, significance_{r \rightarrow k})\}$

Figure 3.4: Triggering Pattern Extraction (TPEX) by Multivariate Hawkes Model

may result in missing patterns at small scales, on the other hand, small discretization values cause a decrease in significance values of patterns at larger scales.

The significance of the resulting patterns is evaluated probabilistically. The algorithm extracts patterns of event types from the pair of instances which are obtained by performing rank selection. Each pair of instances contributes to a pattern with a probability value. In the experimental study, it is realized that the distribution of the estimated probabilities for a particular pattern is not normal, instead it usually shows heavy-tailed and left-skewed distribution. Mosteller and Tukey [29], discuss the robustness of the location measures in terms of two concepts when the data is not normal. Robustness of validity ensures that regardless of the underlying distribution the confidence interval for the population location will cover the true value with a 95% of chance. Robustness of efficiency means confidence interval for the population location is as narrow as the one we can have when the true shape of the distribution is known. The mean and the median are the best in terms of these two robustness measures when the data is normal. However, in the case of non-normality, the robustness of validity does not hold for the mean whereas the median lacks robustness of efficiency. There are a number of location measure alternatives to the mean and the median, which try to balance two robustness concepts such as mid-mean, trimmed mean and winsorized mean. In that sense, mid-mean measure was selected as a location measure for the pair probabilities. Mid-mean measure is the mean value of the observations between the 25th and 75th percentiles. The probability of a triggering pattern of two types is the mid-mean of the probability values of the pair of instances which constitute the pattern. It can be written as:

$$P(r \rightarrow k) = mid - mean(p_{ij}^* : Y_j \rightarrow Y_i = r \rightarrow k) \qquad (3.16)$$

where p_{ij}^* is the probability of the pair of instance which contributes to the pattern.

CHAPTER 4

EXPERIMENTS WITH SYNTHETIC DATA SETS

In this chapter, the state-of-the art methods and the proposed algorithms are evaluated using synthetic data sets. The data sets are explained in Section 4.1. The rest of the chapter gives the experimental results and discussions.

4.1   Data Sets

There are two types of synthetic data sets used in this study. The difference between the two sets is the complexity in terms of the causal structure, the number of event types, uncertainty about producing the effects and existence of different interaction ranges among the events.

In both cases, stationarity holds and several values of the parameters for the distribution the data comes from are examined. These data sets are generated following a standard data set generation technique which is in line with the literature [36]. The details of the data set generation can be found in Appendix A.

4.1.1   Scenario I: Simple Pair of Two Event Types

In this setting, the datasets comprise two types of events one of which generates the other within a constant spatio-temporal neighborhood. Each cause event has one effect event. In other words, cause and effect relation satisfies sufficiency and necessary condition, definitions of which are given below.

Sufficiency Condition: If an event C is sufficient for an event E it means that, the existence of C guarantees the existence of E ($C \rightarrow E$ holds). However, the existence of event E does not imply the existence of event C since there may be some other events which are sufficient to produce E ($E \rightarrow C$ does not hold).

Necessary Condition: If an event C is necessary condition for an event E, it means that the event E cannot be happened without the event C ($E \rightarrow C$ holds). However, the existence of the event C may not be sufficient by itself for the event E, therefore, the opposite is not true. ($C \rightarrow E$ does not hold).

There are three varying factors: the mean rate of the spatial process (λs), the mean rate of the temporal process (λt) and the sample size (N). Their values are given
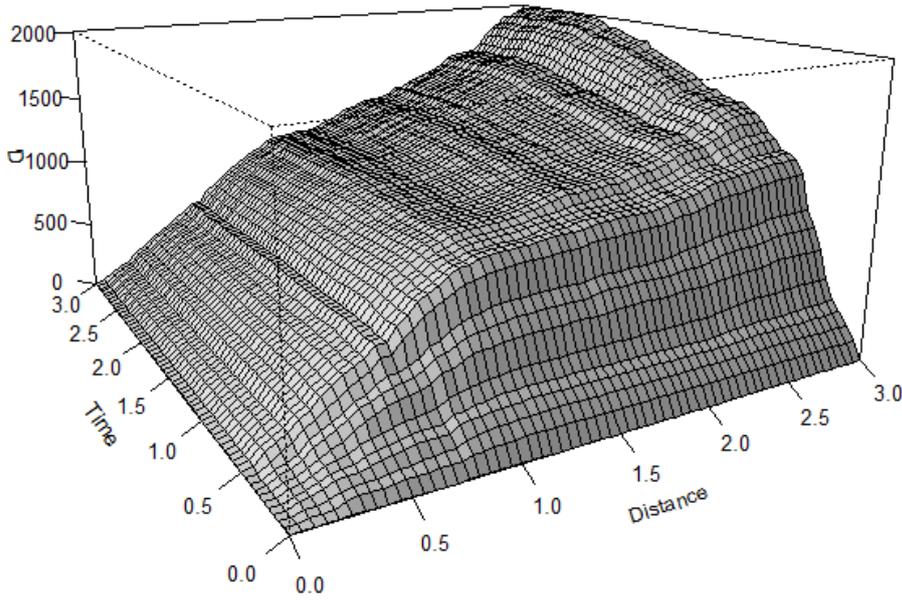
Figure 4.1: D plot of a sample data

in Table 4.1. The sample size and λs are positively correlated. In other words, if the mean rate of a Poisson process increases, then the sample size also increases for a constant region. The sample size for the replications of the same Poisson process varies due to the randomness. However, in the early findings of the study it was observed that variation in sample size for the same experimental settings increases the standard error of the average performance. Therefore, the effect of within group variation was eliminated by selecting the samples, the size of which is equal to the mean of sample size distribution. As a result, two factors with three levels producing nine different data settings were used.

The location of the effect events were generated randomly within one unit neighborhood of the cause events. In Figure 4.1, D plot of a sample data shows the scale at which space-time clustering is observed. The scale represents the neighborhood threshold used to generate effect events. These clusters are significant based on the MC simulation results. The small scale local clusters in space can be seen from the L function and density maps as well. According to the L function in Figure 4.2, the

Table 4.1: Characteristics of the simulated datasets. DS indicates the data set.

Factor	DS1	DS2	DS3	DS4	DS5	DS6	DS7	DS8	DS9
λs	0.5	0.5	0.5	2	2	2	4	4	4
λt	0.1	0.5	1	0.1	0.5	1	0.1	0.5	1
N	30	30	30	100	100	100	200	200	200

32
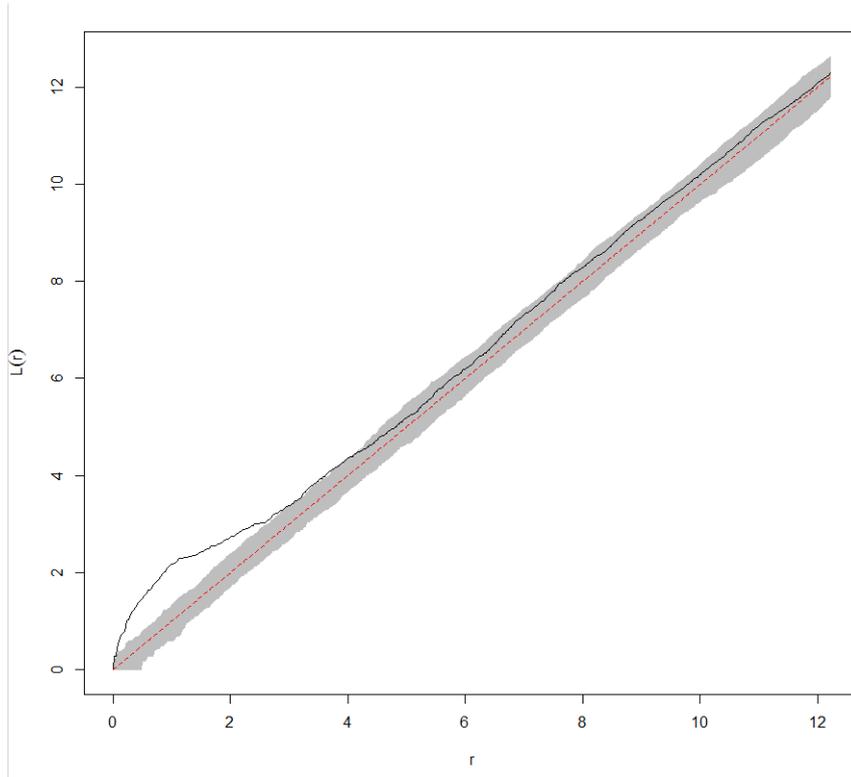
Figure 4.2: Significance of spatial clustering and clustering scales

pattern is random beyond the distance larger than 2.5 unit, in other words, there is
no interaction between the observations at those ranges. In Figure 4.3 (a), dense areas
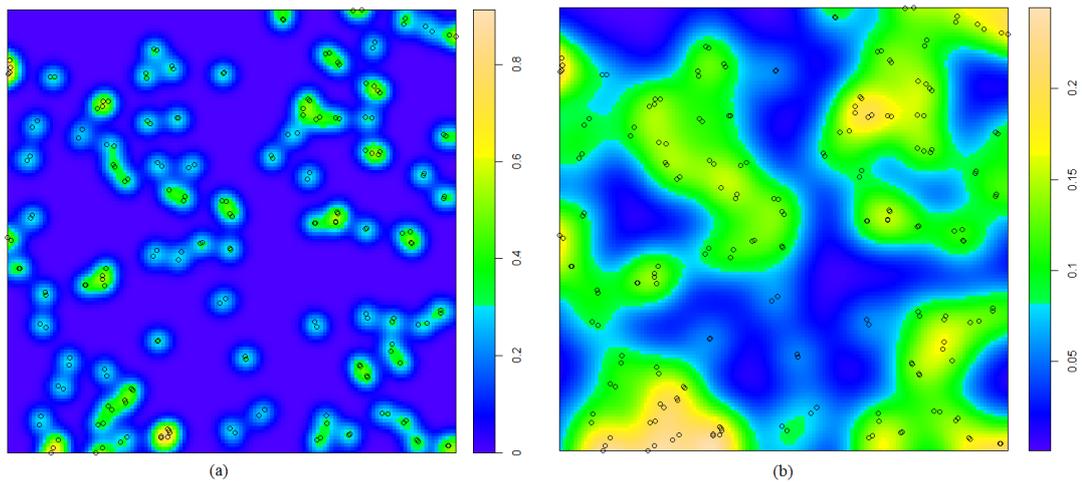
Figure 4.3: Kernel density estimations of Simple Simulation Data (a) bandwidth=1
(b) bandwidth=2.5

are the local clusters the pairs comprise whereas in (b) they are observed by chance.

The spatial locations of the data were simulated within a region of $[0, 50]$ x $[0, 50]$. The time points are generated according to the corresponding mean rate and then mapped to $[0, 50]$ interval. Five replications were applied for each setting.

4.1.2 Scenario II: Pairs among the Six Event Types

An example scenario of a triggering pattern including six event types is represented in Figure 4.4. In the figure, discount day (DD) and concert (C) events at a shopping mall are the triggering (random) events which trigger violating parking rules (VPR), pickpocketing (P) and clogging (CL) events. CL events further trigger fainting (F) events. Here, VPR is the common consequence of DD and C events, however, both ancestors can trigger the consequence individually. An event of the type triggered may be a casual event. If an event type triggers another type of event, it increases the likelihood of the triggered one. In other words, a triggered event type arises with a higher rate than usual after its triggering event type was observed. Lets assume that a VPR event occurs randomly with a specific mean rate. When a C event takes place at a specific time and location, the number of VPR events observed may be more than usual within the neighborhood of the C event.

There are 36 possible patterns including both self (e.g., $VPR \rightarrow VPR$) and mutual (e.g., $DD \rightarrow F$) excitation of events. The ground truth patterns are $DD \rightarrow VPR$, $C \rightarrow VPR$, $C \rightarrow P$, $C \rightarrow CL$ and $CL \rightarrow F$. The patterns $CL \rightarrow F$ and $C \rightarrow CL$ exist in very small and small scales, respectively, whereas the remaining ones interact at higher ranges. The scale of the interaction for each true pattern can be explored by D-function. Small scale interactions are visible when very small discretization was used while constructing D-function.
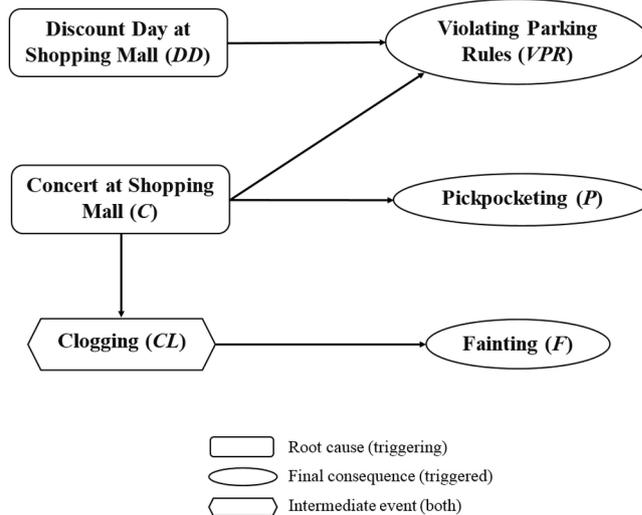
Figure 4.4: An example scenario for the triggering pattern of event types.

Several data sets of the scenario were generated based on the different parameter settings. The parameters describe the mean rates of the triggering events, cluster size of the triggered events and the noise included in the data. Six shopping malls were selected in Ankara, Turkey, where discount and concert events are organized. The sample sizes of the triggering events were generated from Poisson distributions with mean rates $\lambda_{discount} \in \{4, 8\}$ and $\lambda_{concert} \in \{3, 6\}$. The two levels for each are coded as small and large. The spatial locations of the shopping malls and concert stands are represented by 2-dimensional space. The time of the generated events in terms of the day of the year are assigned randomly.

The number of events triggered by each discount day or concert are generated from Poisson distribution with mean rates $\lambda_{parking} \in \{3, 6, 10, 20\}$, $\lambda_{pickpocketing} \in \{2, 4, 8, 16\}$, and $\lambda_{clogging} = 3$. The number of fainting events triggered by a clogging event is generated with mean rate $\lambda_{fainting} = 1$. These parameters controls the cluster size of the triggering events. In other words, they define average number of triggered events around a triggering event. Four levels of mean rates to generate VPR and P events are coded as small, medium, large and very large. For each particular parameter setting, the datasets are generated without noise and with noise such that the noise constitutes either 5% or 10% of the total sample size. The noise added to the data are the random events of the type triggered events such as P and VPR which occur independently from the triggering events DD and C. To ensure that the noise generated is repre-

Figure 4.5: Neighborhood regions for different event types.

sentative of the real case, the region for the random P events was defined based on the regions reported by the government agencies as usual areas for P events. Random VPR events are generated within the study region. Ten runs were performed for each particular setting and the averages over these runs were reported. The summary of data settings, coding for dataset name and average sample size for the generated datasets are given in Table 4.2

Table 4.2: Experimental Data Settings

$\lambda_{triggering}$	$\lambda_{triggered}$	Noise($N\%$)	DataSet Name	Average of N
small	small	0	ss	331
small	small	5	ss5	310
small	small	10	ss10	347
small	medium	0	sm	439
small	medium	5	sm5	520
small	medium	10	sm10	457
small	large	0	sl	691
small	large	5	sl5	767
small	large	10	sl10	783
small	very large	0	svl	1123
small	very large	5	svl5	1502
small	very large	10	svl10	1385
large	small	0	ls	562
large	small	5	ls5	675
large	small	10	ls10	692
large	medium	0	lm	901
large	medium	5	lm5	1023
large	medium	10	lm10	1065
large	large	0	ll	1286
large	large	5	ll5	1488
large	large	10	ll10	1585
large	very large	0	lvl	2516
large	very large	5	lvl5	2727
large	very large	10	lvl10	2759

The neighborhood regions around the shopping malls where triggered events can occur were defined such as indoor and outdoor parking areas, concert areas, clogging areas and pickpocketing areas. A sample view of neighborhood regions around a shopping mall is illustrated in Figure 4.5. After simulating the number of triggered events generated by each DD and C events according to the defined Poisson distributions, their locations are determined randomly within the neighborhood region of the corresponding triggering events. The temporal neighborhood is defined either within a day or within hours according to the event type. For example, a clogging event must occur during a concert which lasts 2 or 3 hours. Spatio-temporal locations of the all generated events are normalized to the interval $[0, 100]$ using min-max normalization.

4.2 Results and Discussions

4.2.1 Method 1

In this section, the objective is to evaluate the performance of the Intensity Model Estimation and Stochastic Declustering algorithms to discriminate cause/triggering (C) and effect/triggered (E) events in synthetic datasets and to evaluate Method 1 in terms of triggering pattern extraction.

In the experiments, the density of distances in space is explained by Gaussian function. The density of distances in time is explained either exponential and gamma functions. As a result, two distinct intensity models were used during learning. The parameters of the exponential function are K and W. K is the probability of elevated risk to observe an effect after observing a cause, W^{-1} stands for the extent of the effect range in time. d, the parameter of the Gaussian function, is the bandwidth in space. During the MLE the values that maximize the likelihood function were searched from the following sets: $W = \{0.1, 0.3, 0.5, 1, 2, 3.3, 10\}$, $K = \{0.1, 0.3, 0.5, 0.75, 1\}$ and $d = \{0.005, 0.02, 0.04, 0.5, 1, 2, 4, 8, 15\}$. All the models were converged in maximum of 13 iterations. The algorithm produces a value, ϕ, for each event which represents the probability of the event being a cause. As described in the Stochastic Declustering algorithm, if the generated random number U_i is smaller than ϕ, the event was assigned as a cause, otherwise, it was assigned as an effect. Since the objective of the stochastic declustering algorithm is to estimate the main events and the mean rate of the background process consisting of the main events, being a cause event was considered as positive and performance measures such as true positive (TP) rates and true negative (TN) rates were calculated accordingly.

Table 4.3: Results of Model Fitting for Gaussian-Exponential Intensity Model

DataSet Name	W	K	d
DS1	3.3, 10	1	0.005, 0.5
DS2	0.5, 1, 2	1	0.005, 0.5
DS3	0.5, 1	1	0.01, 0.02, 0.04, 0.5
DS4	10	1	0.5
DS5	3.3, 10	1	0.5
DS6	1, 2, 3.3	1	0.01, 0.04, 0.5
DS7	10	1	0.5
DS8	10	1	0.5
DS9	3.3	1	0.5

Over the five replications of the sample designs, the estimation of the parameters and the mean rate are given in Table 4.3. There is a variation in the fitted values not only for the datasets with different settings but also that of those for the replications of the same setting. For all of the cases the probability of the elevated risk is estimated as 1 which means each main event produces an effect. This estimation is true since the generated datasets satisfy sufficiency and necessary condition for the cause and effect

Table 4.4: Accuracy of Gaussian-Exponential Intensity Model

DataSet Name	Sensitivity	Precision	F-measure
DS1	1	0.9	0.95
DS2	1	0.87	0.93
DS3	1	0.63	0.77
DS4	0.98	1	0.99
DS5	0.95	1	0.97
DS6	0.97	0.8	0.88
DS7	0.93	1	0.96
DS8	0.96	1	0.98
DS9	0.9	0.98	0.94
Overall	0.97	0.91	0.93

relationship. The other two parameters have different values among the cases. The fitted values for W are close to the inverse of the maximum time distance between the cause and the effect events. Thus, they are able to represent the data behavior. However, the estimated values for the parameter d is not representative for the data. It is expected to have a value close to the maximum spatial distance between the cause and the effect events, which is 1 in the simulated cases. About in half of the cases it is estimated as 0.5 (mean spatial distance between the cause and the effect events), and for the remaining data it is estimated as very small number such as 0.005. The mean rate of the processes are also incorrectly estimated for most of the cases. The accuracy of the results in terms of sensitivity (TP rate), specificity (TN rate) and precision of positive estimates are presented in Table 4.4. The effect of incorrect parameter estimations on the performance is significantly high and sensitivity and precision values for those cases dramatically decrease. For the cases where the bandwidth d is estimated as the mean spatial distance between the cause and the effect events, most of the events are correctly classified although the mean rate of the process is incorrectly estimated.

Table 4.5: Results of Model Fitting for Gaussian-Gamma Intensity Model

DataSet Name	α	β	d
DS1	1.5	0.1	0.1, 0.5
DS2	2	0.5	0.1, 0.5
DS3	1.5	1	0.1
DS4	1	0.1	0.5
DS5	2	0.1	0.5
DS6	2	0.3	0.1
DS7	1	0.1	0.5
DS8	1.5	0.1	0.5
DS9	2	0.1	0.5

Table 4.6: Accuracy of Gaussian-Gamma Intensity Model

DataSet Name	Sensitivity	Precision	F-measure
DS1	1	0.92	0.96
DS2	1	0.87	0.93
DS3	1	0.63	0.77
DS4	0.98	1	0.99
DS5	0.94	1	0.97
DS6	1	0.6	0.75
DS7	0.87	0.99	0.93
DS8	0.94	1	0.97
DS9	0.95	0.99	0.97
Overall	0.96	0.89	0.92

Similarly, the fitted Gaussian-Gamma Intensity models do not represent the data well in terms of the mean rate and the spatial extent. The worst performance values are observed for those cases the parameters of which are incorrectly estimated. The other two parameters α and β are the shape and the rate parameters of the Gamma function, respectively. The parameter estimations and the performance values are given in Table 4.5 and Table 4.6. There are significant differences between the accuracy of the two models for the small datasets such as $DS1$, $DS2$ and $DS3$. Both models have tendency to assign events as a cause or random event when the fitted model is not adequate, however, Gaussian-Gamma Intensity model is able to classify correctly more effect events even the model parameters are incorrectly estimated.

The variation among the results of the data sets having the same settings might be due to the small sample size which may affect the result of the MLE. Although the sample datasets are very simple such that the commonly used density functions utilized in the study are expected to be able to model the data, in real cases the processes are more complex and there is a need to use more specific functions describing each of them. In addition to the necessity of knowing the form of the domain specific functions, if spatial and temporal dimensions of the process are dependent, then, to use the product of the separate functions is not appropriate. In the Figure 4.1 in Section 4.1, it is shown that the data sets in the simple design have significant space-time clusters. Therefore, the product of the space and time functions might be inappropriate to model the data.

4.2.2 Method 2

Based on the findings in Section 4.2.1, in this section the objective is to evaluate non-parametric approaches in intensity modeling. The ability of the models with the temporal and spatial intensity functions to describe the relationships in the spatio-temporal datasets was evaluated and the results of the Method 2 were discussed. In the experiments, different discretization values were used for the definition of the nonparametric function. It is observed that these values affect the results significantly. The reported findings are based on the best results obtained from each individual case.

Table 4.7: Accuracy Comparison of Nonparametric Intensity Models and Nonparametric Spatial Models with Binary Weighting

DataSet Name	Temporal		Spatial		Weighted Spatial	
	Sensitivity	Precision	Sensitivity	Precision	Sensitivity	Precision
DS1	0.68	0.83	0.97	0.88	0.99	0.99
DS2	0.29	0.36	0.99	0.8	1	1
DS3	0.22	0.3	0.97	0.86	0.97	0.99
DS4	0.8	0.81	0.97	0.79	0.97	0.94
DS5	0.17	0.66	0.95	0.8	0.95	0.93
DS6	0.09	0.23	0.94	0.79	0.94	0.81
DS7	0.64	0.9	0.91	0.88	0.91	0.9
DS8	0.12	0.55	0.9	0.9	0.9	0.9
DS9	0.04	0.4	0.89	0.89	0.92	0.96
Overall	0.34	0.56	0.94	0.84	0.95	0.93

4.2.2.1 Modeling with Temporal Intensity Function

The success of the algorithm to predict actual cause effect pairs depends on the density in time. As the temporal density increases, sensitivity and precision values decrease dramatically. In other words, if pairs are well separated in time it works well. For the dataset exhibiting low temporal density, sensitivity is 0.7 and precision is 0.85 on average. However, these values drop to 0.12 and 0.31 on average for the dataset exhibiting high temporal density. The high number of false positives (FP), i.e., events paired incorrectly, arises from predicting cause events as pairs although they occur randomly in time. This is because the mean distance in time between the cause events are close to the interaction scale of the cause and effect events which is hard to discriminate.

Although the model can be considered as successful to discover pairs for very specific cases such as in the case of well separated clusters, it has limitation in prediction of the mean rate. For none of the experimented data sets the mean rate is correctly estimated. This is because for all of the events except for the first in time, p_{ii} values converge to zero which means none of them were predicted as cause or background event. A temporal intensity for self-exciting point process was also studied by Lewis and Mohler [21]. They used MISD algorithm and extend it with maximum penalty likelihood estimation. In their study with simulated data, they assumed that triggering kernel needs to be zero outside the interval of $[0, 4]$ which leads to successful estimates for the mean rate, however, they neither provide reasoning for this restriction nor for the cut-off point.

4.2.2.2 Modeling with Spatial Intensity Function

Unlike the temporal intensity function, the spatial intensity model is capable of predicting the actual cause effect pairs without using time information. Very high prediction accuracy for the actual cause effect pairs were observed for all cases. As the spatial

density increases, sensitivity decreases from 0.98 to 0.90. Precision of the predictions varies between 0.79 and 0.90. A relationship between the precision values and the spatial density of the data were not observed. On average, sensitivity and precision values are 0.94 and 0.84, respectively. The problem faced during the prediction of μ and p_{ii} still remains unsolved here.

4.2.2.3 Improvement: Modeling with Spatial Intensity Function Weighted based on Ripley's K Function

Simulated data sets were analyzed with Ripley's K function and the range where clustering is significant was determined. This information was incorporated into the algorithm as defined in Section 3.2 in Chapter 3. The results were compared with those obtained by using spatial intensity model without weighting. The use of weights does not only increase sensitivity and precision values but also results in successful predictions for μ and p_{ii} values. This modification significantly increases precision of the estimates compared to the original algorithm. Both has similar sensitivity values. In a few cases binary weighting has small improvements on sensitivity values. Finally, on average, 90% of cause events are labeled correctly as cause in this method whereas the original algorithm is not able to predict cause events.

For all the methods, there is a trade-off between TP and FP as the discretization value changes. In general, the use of small discretization values increase both TP and FP.

4.2.3 Method 3

In this section, the objective is to evaluate performance of Method 3 which uses type information during learning. The algorithm was applied to both simple and complex scenarios described in Section 4.1. The maximum number of iterations for convergence is 15. The pairs were selected based on two methods. In the first method, a threshold

Table 4.8: Accuracy Comparison of Nonparametric Intensity Models and Nonparametric Spatial Models with Binary Weighting (F-measures)

DataSet	Temporal	Spatial	Weighted Spatial
DS1	0.75	0.92	0.99
DS2	0.32	0.88	1
DS3	0.25	0.91	0.98
DS4	0.8	0.87	0.95
DS5	0.27	0.87	0.94
DS6	0.13	0.86	0.87
DS7	0.75	0.89	0.9
DS8	0.2	0.9	0.9
DS9	0.07	0.89	0.94
Overall	0.42	0.89	0.94

value was used. If $p_{ij} > 0.5$, the event pairs are selected. In the second method, rank selection method was used to extract pairs.

Table 4.9: Average predictive accuracy of Method 3 with threshold selection and threshold+Otsu selection for Scenario I.

DataSet	Threshold			Threshold+Otsu		
	Sensitivity	Precision	F-measure	Sensitivity	Precision	F-measure
DS1	1	0.94	0.96	1	0.94	0.96
DS2	0.95	0.87	0.90	0.91	0.88	0.90
DS3	0.81	0.88	0.84	0.77	0.89	0.82
DS4	0.96	0.6	0.74	0.96	0.73	0.84
DS5	0.99	0.65	0.78	0.99	0.75	0.86
DS6	0.99	0.64	0.78	0.98	0.74	0.84
DS7	0.99	0.62	0.76	0.99	0.87	0.92
DS8	0.84	0.55	0.66	0.84	0.77	0.8
DS9	0.98	0.65	0.78	0.98	0.92	0.94
Overall	0.95	0.71	0.82	0.94	0.83	0.88

Table 4.10: Average predictive accuracy of Method 3 with rank selection and rank+Otsu selection for Scenario I.

DataSet	Rank			Rank+Otsu		
	Sensitivity	Precision	F-measure	Sensitivity	Precision	F-measure
DS1	1	0.13	0.22	1	0.94	0.96
DS2	0.96	0.13	0.22	0.95	0.88	0.92
DS3	0.83	0.09	0.16	0.81	0.88	0.84
DS4	0.97	0.19	0.32	0.97	0.59	0.74
DS5	1	0.16	0.28	1	0.65	0.78
DS6	1	0.18	0.30	1	0.62	0.76
DS7	1	0.49	0.66	0.99	0.78	0.88
DS8	0.86	0.42	0.56	0.84	0.68	0.74
DS9	1	0.5	0.66	0.98	0.83	0.90
Overall	0.96	0.25	0.4	0.95	0.76	0.84

The summary of the performance on Scenario I based on the selection methods are given in Table 4.9 and Table 4.10. A trade-off between sensitivity and precision values was observed according to the selection method used. The rank selection method produces slightly higher sensitivity values than the method based on thresholding, however, it produces significantly lower precision values. Generally, the method based on thresholding has lower precision values as the spatial density increases, whereas the rank selection method behaves oppositely such that it produces higher precision values in the cases where spatial density is high. As mentioned in Section 3.3, the incorrect pairs usually have small probability values. For that reason, Otsu thresholding was

used after both methods to adjust predictions. The most significant improvements is achieved for the rank selection method considering the initial performance. For example, for the data set $DS3$, precision of the predicted pairs is 0.09 with rank selection method. It increases to 0.88 after Otsu thresholding. The highest decrease in sensitivity after Otsu method applied is 0.04 and the most of the sensitivity values remains the same. The average sensitivity and precision values of the methods over all data sets are 0.95 and 0.71 for the method based on thresholding and 0.96 and 0.25 for the rank selection method. These values are adjusted to the 0.94 and 0.83 for the former and 0.95 and 0.76 for the latter after applying Otsu method. The limitation of the method is the estimation of mean rate of the process, thus, p_{ii} values.

Table 4.11: Average predictive accuracy of Method 3 with threshold selection and threshold+Otsu selection for Scenario II.

DataSet	Threshold			Threshold+Otsu		
	Sensitivity	Precision	F-measure	Sensitivity	Precision	F-measure
ss	0.07	0.27	0.06	0.01	0.1	0
ss5	0.02	0.14	0.02	0.01	0.25	0.01
ss10	0.01	0.11	0	0	0	0
Average	0.04	0.19	0.03	0.01	0.17	0.01

Table 4.12: Average predictive accuracy of Method 3 with rank selection and rank+Otsu selection for Scenario II.

DataSet	Threshold			Threshold+Otsu		
	Sensitivity	Precision	F-measure	Sensitivity	Precision	F-measure
ss	0.05	0.1	0.03	0.02	0.07	0.01
ss5	0.09	0.05	0.03	0.02	0.09	0.02
ss10	0.05	0.05	0.02	0.02	0.15	0.02
Average	0.07	0.07	0.03	0.02	0.09	0.02

In the Table 4.11 and Table 4.12, the results for the Scenario II are given. In these tables, only the results for the first simplest data sets are provided. This is because the method is unsuccessful for the remaining data sets. In other words, for the remaining data sets, all of the events in the data were predicted as random or cause events. For the simplest cases, the method still shows low performance since the TP rate is less than 0.1 for all. Low precision values are observed for both selection methods as well, however, the method based on thresholding performs better in terms of precision values. Its performance ranges between 0.11 and 0.27 whereas performance of rank selection method is at most 0.1.

4.2.4 Method 4

The performance of Method 4 (TPEX algorithm) is evaluated on the complex dataset scenario. The evaluation is made in terms of predictive accuracy, computational performance and user defined parameters required. The accuracy of the results was evalu-

ated based on the ability of the algorithm to extract ground truth patterns within the top ten ranked patterns ($recall@10$) and the distribution of their significance values. $recall@10$ illustrated in Table 4.13 shows the proportion of actual pairwise patterns correctly identified within the top ten ranked patterns. In the table, the results are reported for different data sets and discretization values. ($\delta s = 1, \delta t = 0.05$) defines the scale of the pattern $CL \to F$. ($\delta s = 4, \delta t = 0.5$) and ($\delta s = 8, \delta t = 1$) are representative for the remaining patterns. ($\delta s = 16, \delta t = 2$) doubles the extent of the interaction in space and time for the maximal range. The performance was also examined by using a very large discretization level ($\delta s = 25, \delta t = 25$) which produces highly rough density function. The ranking is carried out among the significance values calculated for all possible patterns. For the best case, it is desirable to find a true pattern with high rank value within the top ten, and with a significance value equal or close to 1. The rank distributions of the ground truth patterns are visualized in Figure 4.7 with a rank map where darker areas represent higher ranks, and the distributions of their significance values among the datasets in Figure 4.8. The results were compared with a recent study which uses CPI measure to identify significant patterns [26].

4.2.4.1 Comparison based on Predictive Performance

For 17 out of 24 datasets, TPEX achieves $recall@10 = 1$ when the density function is estimated with the smallest discretization intervals. It decreases to 0.8 for the remaining seven datasets in most of which triggering events are not dense. CSTP, on the other hand, extracts only the pattern $CL \to F$ since given threshold defines significantly small neighborhood which allows to capture interactions at very small scales.

When the discretization values are set to ($\delta s = 4, \delta t = 0.5$), ($\delta s = 8, \delta t = 1$) or ($\delta s = 16, \delta t = 2$) TPEX finds all patterns except $CL \to F$ ($recall@10 = 0.8$). The reason is that larger discretization intervals smoothen the distribution function $g_{CL,F}$ such that local peaks representing small range interactions are not visible anymore. However, an indirect relation between C and F due to the cascade of 3 events, $C \to CL \to F$, is extracted as a direct relation by TPEX. The larger ranges are more representative of the range between the root cause and final consequence. Once the relation between the intermediate cause and final consequence is discovered by using appropriate discretization, the significance of the relation between C and F decreases to 0. This behavior is meaningful in the sense of conditional independence. All the patterns TPEX found has high ranks as can be seen from the relatively darker areas in the rank map in Figure 4.7. For the same discretization values, CSTP finds three of the five patterns in all datasets. The missing patterns are $DD \to VPR$ and $C \to VPR$ which have the longest interaction ranges. The given neighborhood thresholds are greater than the distance between the events of the pairs, therefore, these patterns can be captured based on the CPI threshold defined by the user. However, they are not in the top ten ranked patterns since their ranks change between 11 and 17. In other words, some irrelevant patterns are found more significant than these two patterns. As a result, they are shown with light colors on the rank map of CSTP.

For the largest discretization which produces the roughest density function, $recall@10$ takes one of the values, 0.6, 0.8, and 1, with TPEX algorithm. All 0.6 values are

Table 4.13: recall@10: The percentage of ground truth patterns found within the top ten ranked patterns.

Discretization DataSet	$\delta s = 1, \delta t = 0.05$		$\delta s = 4, \delta t = 0.5$		$\delta s = 8, \delta t = 1$		$\delta s = 16, \delta t = 2$		$\delta s = 25, \delta t = 25$	
	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP
ss	0.8	0.2	0.8	0.6	0.8	0.6	0.8	0.6	0.8	0.8
ss5	0.8	0.2	0.8	0.6	0.8	0.6	0.8	0.6	0.8	0.8
ss10	1	0.2	0.8	0.6	0.8	0.6	0.8	0.6	0.8	0.8
sm	1	0.2	0.8	0.6	0.8	0.6	0.8	0.6	0.8	0.6
sm5	0.8	0.2	0.8	0.6	0.8	0.6	0.8	0.6	0.8	0.8
sm10	0.8	0.2	0.8	0.6	0.8	0.6	0.8	0.8	0.8	0.8
sl	1	0.2	0.8	0.6	0.8	0.6	0.8	0.6	0.8	0.6
sl5	0.8	0.2	0.8	0.6	0.8	0.6	0.8	0.6	1	0.8
sl10	1	0.2	0.8	0.6	0.8	0.6	0.8	0.6	0.8	0.8
svl	1	0.2	0.8	0.6	0.8	0.6	0.8	0.6	0.8	0.6
svl5	1	0.2	0.8	0.6	0.8	0.6	0.8	0.6	0.8	0.6
svl10	1	0.2	0.8	0.6	0.8	0.6	0.8	0.6	1	0.8
ls	1	0.2	0.8	0.6	0.8	0.6	0.8	0.6	0.8	0.8
ls5	1	0.2	0.8	0.6	0.8	0.6	0.8	0.6	0.6	0.6
ls10	1	0.2	0.8	0.6	0.8	0.6	0.8	0.6	0.6	0.6
lm	0.8	0.2	0.8	0.6	0.8	0.6	0.8	0.6	1	0.8
lm5	1	0.2	0.8	0.6	0.8	0.6	0.8	0.6	0.8	0.8
lm10	1	0.2	0.8	0.6	0.8	0.6	0.8	0.6	0.6	0.6
ll	0.8	0.2	0.8	0.6	0.8	0.6	0.8	0.6	0.8	0.8
ll5	1	0.2	0.8	0.6	0.8	0.6	0.8	0.6	0.8	0.8
ll10	1	0.2	0.8	0.6	0.8	0.6	0.8	0.6	0.8	0.6
lvl	1	0.2	0.8	0.6	0.8	0.6	0.8	0.6	1	0.8
lvl5	1	0.2	0.8	0.6	0.8	0.6	0.8	0.6	0.6	0.8
lvl10	1	0.2	0.8	0.6	0.8	0.6	0.8	0.6	0.8	0.6

observed for the datasets where the density of the triggering events are large. CSTP achieves both 0.6 and 0.8 for *recall*@10 and there is no distinct pattern according to datasets.

TPEX and CSTP measure the significance of a pattern using midmean and CPI, respectively. Figure 4.8 shows their distributions for the ground truth patterns over 24 datasets by each discretization level. The patterns $DD \rightarrow VPR$ and $C \rightarrow VPR$ have considerably high significance values for TPEX than CSTP for all discretization levels except the largest one. Although use of a rough density function significantly decreases the pattern probabilities calculated by TPEX, it still produces high ranks for them. For example, it can be seen from Figure 4.7, $DD \rightarrow VPR$ is always extracted within the top ten rank regardless of discretization level. It is even the top five for most of the cases. The highest significance and rank values are obtained when the discretization is $\delta s = 8, \delta t = 1$. This pattern is found either in 11th or 12th rank by CSTP algorithm. TPEX shows similar performance for the pattern $C \rightarrow VPR$. There is a small decrease in probability and rank values for this pattern, but, it is still discovered within the top five for most of the cases. The ranks that CSTP produce, however, is usually greater than 13. Although CSTP has remarkably higher significance values for both patterns on the largest discretization TPEX still has higher ranks for them.

Similarly, TPEX is more succesful in discriminating the pattern $C \rightarrow P$ than CSTP on the first three discretization levels. For the discretization values ($\delta s = 16, \delta t = 2$) it is observed that TPEX produces higher significance values for the datasets where density of triggering events are small whereas CSTP has higher values when the density is large. Another observation is that for the ranks of the $C \rightarrow P$, CSTP is affected by the noise. The more noise the data set has, the lower the rank of the pattern is. No such effect is observed for TPEX.

CSTP is more successful in overall for the pattern $C \rightarrow CL$. TPEX still produces comparable results based on discretization levels. For example, although TPEX ranked within the top five in most of the cases, CSTP performed better than TPEX. Similarly, TPEX is able to find the pattern $CL \rightarrow F$ having the smallest interaction range if the smallest discretization is used. CSTP finds it with high significance and ranks for all cases.

The behavior of the algorithm was further examined for the pattern $CL \rightarrow F$. The first nine datasets were selected for further evaluation since the significance values calculated for these datasets are the worst which is 0.27 on average. In addition, the significance value of the pattern $C \rightarrow F$ was calculated as 0.56, on the average, for the same set. Two aspects that affect the performance in terms of accuracy were found. Firstly, when smaller discretization values are used for the density function an increase was observed in the significance of $CL \rightarrow F$ from 0.27 to 0.36 and a decrease in the significance of $C \rightarrow F$ from 0.56 to 0.03. However, it also causes the significance of remaining true patterns to decrease while keeping the ranks to be high. This is expected, because it is also claimed in this study that the significance of a pattern has its maximum if the discretization is representative for the interaction range of the events the value of which can be obtained by D-function. Secondly, the effect of null additions on the significance value of a pattern was observed. Null addition can be defined as the existence of irrelevant data in the set with respect to a pattern. In the scenario, small density values were used for the CL and F events to be representative for the real word situations. However, in such settings, these events represented in the datasets with small frequencies relative to the other events such as VPR and P. Once the data is simulated with similar densities for all four event types, regardless of the discretization values used the method is able to find the pattern $CL \rightarrow F$ within the top ten ranked patterns and with a significance value around 0.2 while the remaining patterns have high significance as well.

The detailed results of the algorithm for each replication and the datasets can be found in Appendix C.

4.2.4.2   Comparison based on Computational Performance

In this thesis, the focus is the predictive performance of the pairwise triggering pattern extraction, however, its computational performance was also evaluated with respect to discretization level and the sample size. The same scenario with different density values was used to generate the data sets with sample sizes from 200 to 5000. To evaluate the effect of discretization level, *ss* data set was used with the discretization levels from 1 to 20. Figure 4.6 (a) shows execution time of the algorithm in seconds with respect to discretization level. Since the range of each dimension of the density
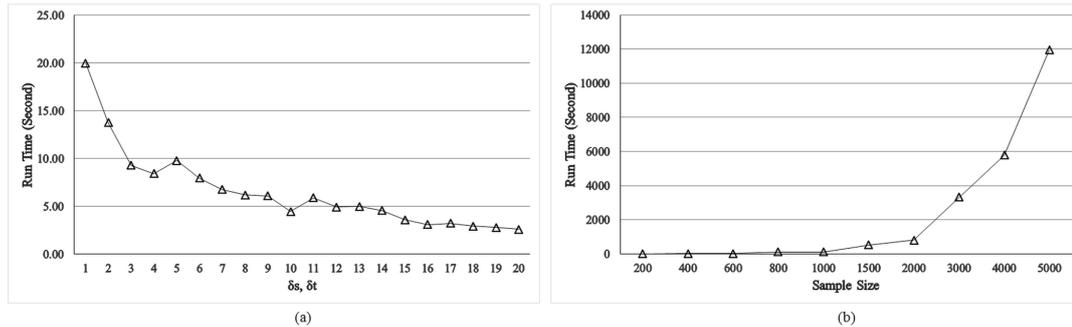
Figure 4.6: (a) Effect of data size on performance. (b) Effect of discretization value on performance.

function is 100, the discretization values can be interpreted as percentage. For example, if $\delta t = 2$ then it means the time dimension of the density function consists of 50 bins each of which constitutes 2% of the range. It can be seen from the Figure 4.6 (a), there is a fast decrease in execution time until $\delta t, \delta s = 3$ and then it decreases slowly as discretization level increases. The fluctuations around 5 and 10 is attributed to the number of iterations required to converge. For example, at $\delta t, \delta s = 4$ and $\delta t, \delta s = 5$ TPEX converges after 6 and 8 iterations, respectively. There is a decrease in execution time per iteration for $\delta t, \delta s = 5$.

The most influential factor on computational performance is the size of the dataset. Figure 4.6 (b) shows execution time of the algorithm in seconds with respect to sample size. Sharp increase was observed in the run time after $N = 2000$. CSTP algorithm performs better than TPEX with respect to the computational time reported in [26].

The number of event types increases complexity little since there are m^2 number of pairwise density function in the model where m is the number of event types. For the experiments in this study, the effect of the number of event types was insignificant up to six event types.

4.2.4.3   Comparison based on User Defined Parameters

TPEX requires the user to define discretization values for the density function. There is no other threshold needed to be defined by the user, for example, a threshold for the significance or neighborhood since it employs rank selection for the pattern extraction and the neighborhood is defined by the density function of distances. Note that a parametric density estimation procedure can also be employed in TPEX methodology. In such cases, discretization values are not needed to be defined, rather, a parameter estimation method such as MLE can be used in the algorithm. In such case where the form of the function is known, once the density is estimated, TPEX can be considered as a input free algorithm for triggering pattern extraction. In the non-parametric case, the suitable discretization values can be obtained by using D-function if the users are not sure about the appropriate values. On the other hand, if such values are not available, users may try different discretization levels and evaluate them based on the calculated pattern probabilities which is maximum.

In the candidate generation approach, users are required to define a neighborhood

appropriate for the domain as well as a threshold for the significance measure. A small neighborhood may cause missing patterns at global scales whereas larger ones may produce many irrelevant patterns.

4.2.5 Summary

Based on the average performance of the methods over all data settings in Scenario I, all methods except Method 2 using nonparametric temporal intensity function show comparable performance for predicting background events. Their success differ in terms of precision of the predictions. In other words, some of them tend to estimate the effects as random events. For example, in Method 2, if spatial density is used, only 84% of the positive predictions are true causes whereas it increases to 93% when proposed method that uses weighted spatial density is used. Their F measures are 0.89 and 0.94, respectively. There is significantly high differences between the precision of the positive predictions of the initial results of Method 3 with threshold selection where threshold value is 0.5 and with rank selection. The first one produce 71% precise predictions, however, it is 25% for the second one. This high differences reduced with the use of Otsu method such that Otsu thresholding adjusts precision values as 83% and 76%, respectively. Their F measures are 0.82 and 0.4 before Otsu thresholding and 0.88 and 0.84 after Otsu thresholding. Although Method 3 with threshold selection produce better results for both, its success might change based on the threshold value used. Therefore, Method 3 with rank selection adjusted by Otsu thresholding can be preferred since it is threshold free and more robust method.

None of these methods are successful on the data in Scenario II. Method 4 is the only approach which successfully model and identify triggering patterns of event types simulated in the scenario.

Table 4.14: Overall Accuracy of Methods for Scenario I

Method	Sensitivity	Precision	F-measure
Method 1 - exp	0.97	0.91	0.93
Method 1 - gamma	0.96	0.89	0.92
Method 2 - temporal	0.34	0.56	0.42
Method 2 - spatial	0.94	0.84	0.89
Method 2 - weighted spatial	0.95	0.93	0.94
Method 3 - threshold	0.95	0.71	0.82
Method 3 - threshold+Otsu	0.94	0.83	0.88
Method 3 - rank	0.96	0.25	0.4
Method 3 - rank+Otsu	0.95	0.76	0.84
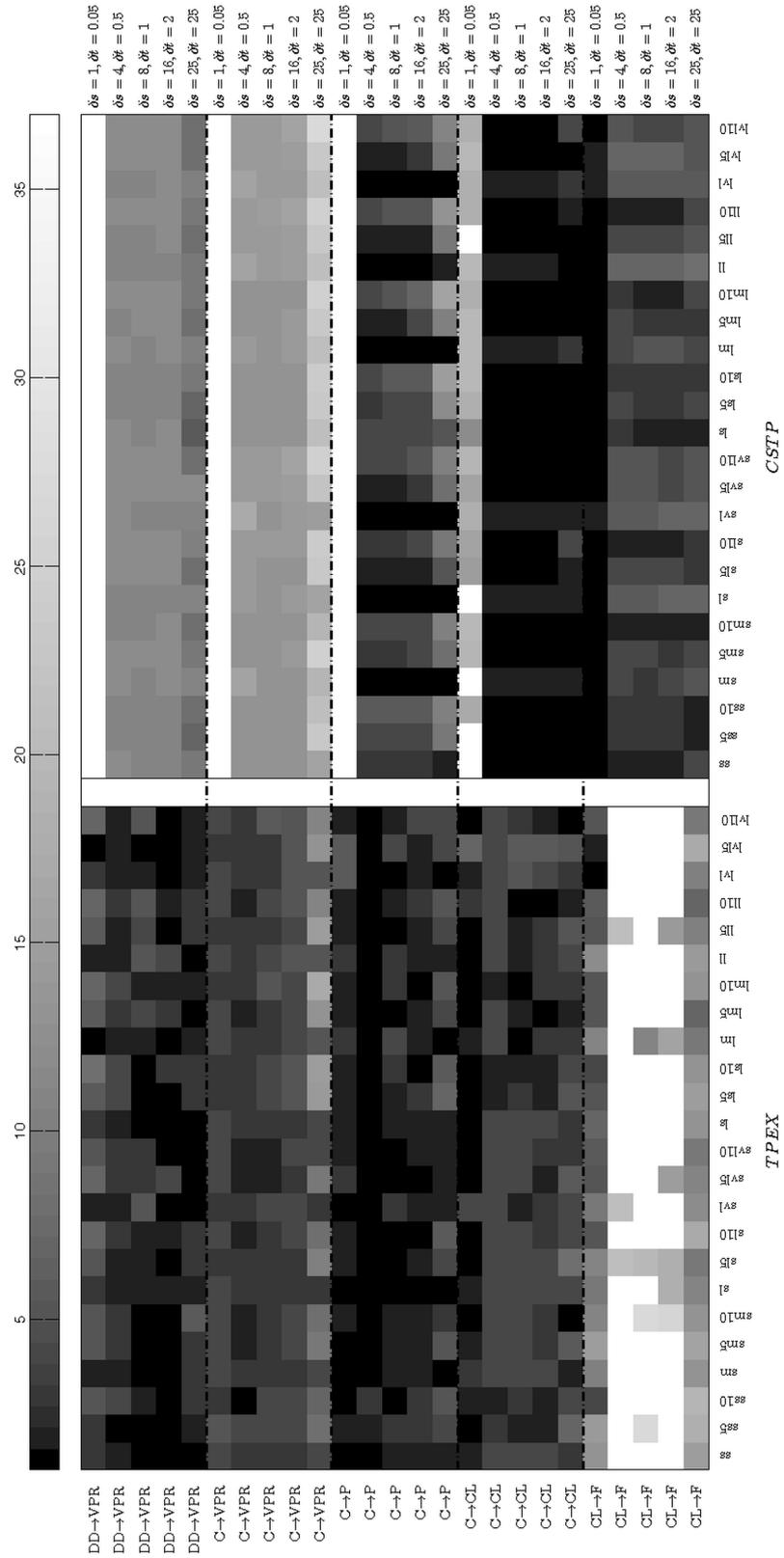
Figure 4.7: Patterns' rank maps for different datasets and discretizations for TPEX and CSTP algorithms.
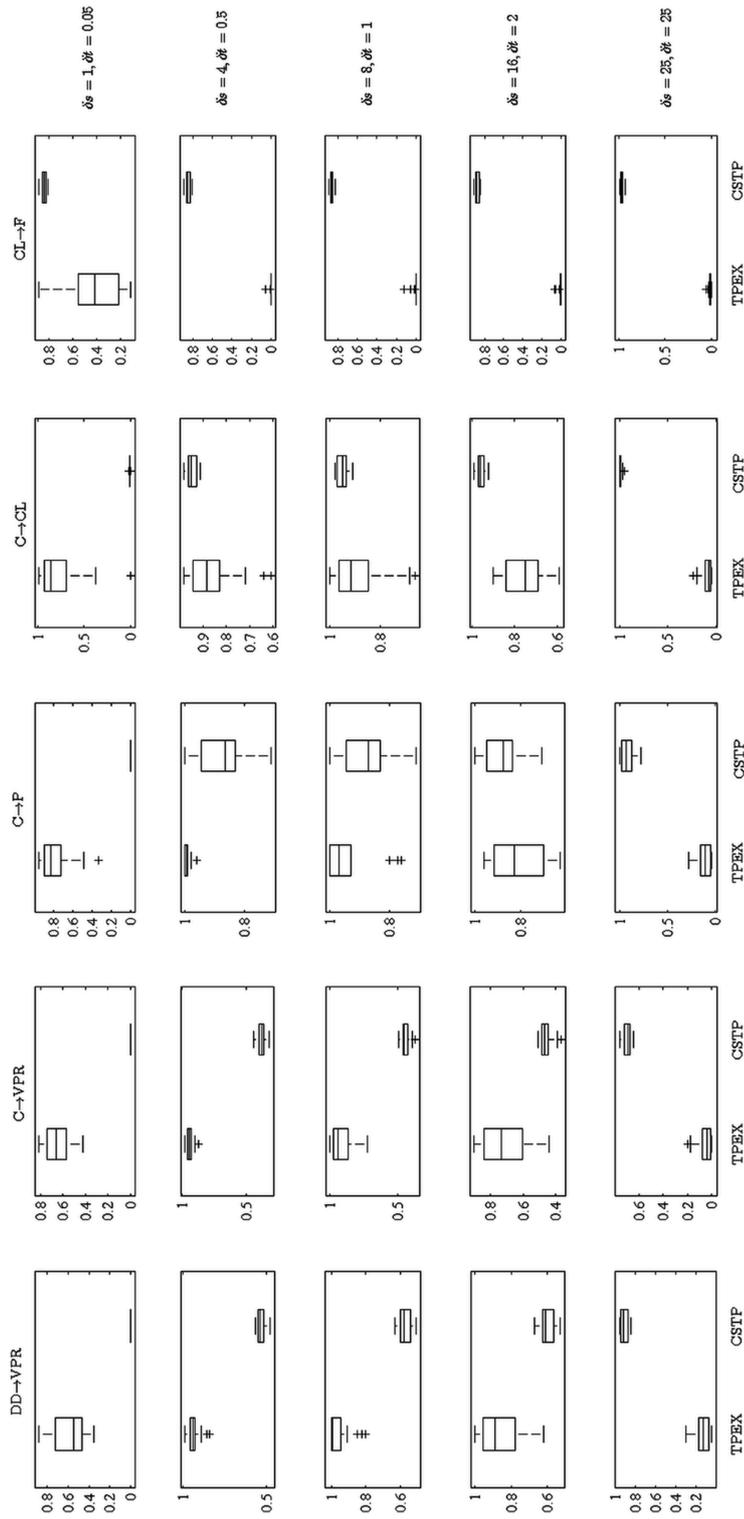
Figure 4.8: Distribution of patterns' significance for different discretizations for TPEX and CSTP algorithms.

CHAPTER 5

CASE STUDY: ARE SPEED BUMPS AND TRAFFIC ACCIDENTS RELATED?

Middle East Technical University (METU) is a state university having over 26000 students by the end of 2013. It has relatively large campus area. There is significant traffic flow in the campus, especially at certain times of the day and on particular routes. To increase safety of pedestrians, to control the speed of the vehicles within the campus and to reduce the number of traffic accidents, several speed bumps were built on the roads where pedestrians are dense and the drivers tend to speed up, in addition to the speed limit already applied in the campus.

Traffic officers may want to know if the existence of the speed bumps causes a decrease in the number of accidents within the campus. They also may wander if the effects of speed bumps have regional differences. For example, some of them may result in higher decrease in the number of accidents within those regions compared to the others, or they may have opposite effects at some regions. Based on the knowledge acquired, traffic officers may suggest to the university authorities to construct new speed bumps or to change the location of the existing ones.

In this chapter, traffic accident data collected by METU Traffic Office was analyzed and a triggering relationship between the speed bumps and the accidents occurred in the campus was examined. The research questions can be listed as:

1. Is there a triggering relationship between the existence of speed bumps and the traffic accident occurrences throughout the campus? If it exists, is it positive or negative?

2. Do all of the speed bumps in the campus affect the traffic accident occurrences similarly? Is there any regional differences? If so, which one has the highest effect?

5.1 Data Set Construction

Selecting Relevant Data

Each traffic accident is recorded by an official report by the Traffic Office of METU if there is any damaged property of METU or injured people. A sample official report

is illustrated in Fig 5.1. All the accident reports recorded from 2002 to the middle of 2014 were examined. About half of the reported accidents occur at the gates of METU, specifically at Gate A1, Gate A4 or Gate A7, which cause damage to gate barriers. According to the reports, these accidents are caused by the drivers who rush to pass the already raised barrier without waiting for the reader to register their vehicles, thus, having the barrier come down before the offending vehicle can clear the gate. These accidents are discarded from the analysis for simplification since their causes are known and are not related to any other events such as existence of a speed bump on the road. A few of the accidents are also removed due to the missing information in the reports such as driving direction of the vehicles. This information is important as the distance between a speed bump and an accident is determined based on the driving direction. Finally, a number of reported accidents which occurred outside of the campus and cause damages at the borders of the campus area are removed since they are both irrelevant and outside the study region. After eliminating irrelevant and incomplete data, there are 64 in-campus accidents spanning over years from 2002 to 2014.

GENEL SEKRETERLİK MAKAMINA,

14 Mayıs 2004 tarihinde saat 02.00 sularında Havacılık Mühendisliği Bölümü öğrencisi ██████████ ODTÜ KENT yönünden maden kavşağına aşırı hızla girerek direksiyon hakimiyetini kaybetmiş ve tek taraflı, maddi hasarlı kazaya neden olmuştur.

Kaza sonucu 20-25 yaşlarında iki adet çam ağacı kökten kırılmış, bir ağaç kısmen hasar görmüş ve bir adet trafik levhası da kullanılamaz hale gelmiştir.

Gereğini bilgilerinize arz ederim.

Saygılarımla.

Figure 5.1: : A sample official report recorded by Traffic Office of METU.

In the METU campus, there are 41 speed bumps built in different years range from 1999 to 2011. Some of them are on the one way roads whereas some others are on the two way roads. It is assumed that a speed bump constructed on a two way road as two separate speed bumps according to the driving direction of the lane. Thus, 62 speed bumps was presumed in total in the campus. The locations of the speed bumps built in the campus are shown in Fig 5.2 where the labels represent approximate construction year of each.

Determining Spatio-temporal Locations

The approximate geographic locations of the accidents are identified from the explanations in the reports and their coordinates are determined by using Google Earth. The temporal locations, on the other hand, are definite since the date and the time of the accidents are recorded. The temporal locations were represented as days where the range is 4575 days. The decimal place shows the time of the day that the accident happened.

Figure 5.2: : Locations and approximate construction years of the speed bumps in METU.

Representing a speed bump with a spatio-temporal point require further work. Actually, a speed bump covers an area instead of a single point in space. In addition, its existence in time is a duration once it is constructed. Therefore, some transformation is needed to represent each speed bump as a spatio-temporal point data. The spatial location can be expressed by a representative point such as center. Because a possible relation between the existence of speed bumps and accidents is investigated, the time of the speed bumps can be specified based on the occurrence time of the accidents. By this way, a variable that represent existence or nonexistence of all speed bumps which exist currently in the campus, at the time snapshot of each accident can be defined. For each snapshot, the time value of a speed bump was calculated as the time of the accident minus a value proportional to the spatial distance. For example, if the distance between the accident and the speed bump is small then the time distance should be small too since the speed bump should have been passed by the vehicle just a little while ago.

Data Sampling

Generating speed bump map of the campus at the time snapshot of the accidents produces 62 speed bump records for each accident which results in 3968 speed bumps in total. This class imbalance usually cause difficulties during the modeling. For example, pairs between the types related with speed bumps such as (exist, exist), (exist, non-exist) and (non-exist, non-exist) have significantly higher frequency than the pairs such as (exist-accident) and (non-exist-accident) in the data. In addition, pairs between the types related with speed bumps may show spatio-temporal clusters since there are usually more than one speed bump within a small region. As a result, those pairs are expected to be extracted, and due to the null addition effect, they may suppress the significance of those contain accidents.

According to the domain knowledge, speed prevention effect of a speed bump subsides approximately within 100 meters. In order to have a continuous prevention, there would be a speed bump about every 100 to 150 meters. However, in reality, there are usually at most two or three speed bumps close to each other in the campus. A group of speed bump that are close to each other can force a vehicle to be within speed limits, however, their effect will not be seen anymore after about more than 100 meters from the last one. The nearby speed bumps might produce a combined effect on an accident as well. The triggering pattern extraction method proposed estimates the direct effect of an event on another one. Different causes can be explored by examining individual probabilities. By using the domain knowledge on the effect range of speed bumps, those already known as irrelevant can be eliminated to reduce null addition effect. As a result, only the speed bumps within 250 meters neighborhood of each accident were sampled which yield 204 data samples including both accidents and the speed bumps.

5.1.1 Limitation of Data Preparation

The first limitation of the dataset is the missing information about the construction dates of the speed bumps. The exact dates are not recorded, thus, the construction years are asked during the interview with the head of Office of Domestic Services. The most of them are provided as year, some are determined approximately, such as within 3 years. However, the exact dates of the accidents are known. Even the construction year is correct it might be uncertain if a speed bump exists during the occurrence of an accident. Therefore, three points in time are considered; beginning, middle and the end of the year or 3 years.

Second limitation is the missing information about the exact routes of the vehicles before accidents occur since there are more than one alternative due to the junctions. For example, driving route of a driver having an accident in front of Doyurucu Patisserie can be from gate A4 or from Registrar Office, as a result, the speed bump actually passed by the driver is not certain. Unlike the scenario described in Chapter 4, Eucledian distance is not suitable for this case. Instead, distances are calculated according to the road network in the campus. The shortest path is considered while calculating the distances.

Third limitation is the necessity of data type conversion for the speed bumps to obtain spatio-temporal point data. A speed bump, in fact, covers an area instead of a single point, however, it can be expressed by a representative point such as center. The time, on the other hand, is not a point, instead, it is a duration once the speed bump is constructed. For that reason, a new variable which shows the existence of each speed bump at the time of the accident is needed to be defined

5.1.2 Encoding of Event Types

Based on the research questions stated at the beginning of this chapter, the event types related to speed bump were defined in two different ways. First, the sampled speed bumps are labeled as "exist" and "not exist" based on at the time of the accident. This encoding is made to examine the overall effect of the speed bumps on accidents.

Another approach is to examine the effects of the speed bumps at a specific region. For this purpose subgroups for the types "exist" and "not exist" were specified according to the region they are constructed. For example, the type of the speed bumps around the Department of Basic English was coded as "exist at Basic English Department" and "not exist at Basic English Department". Each speed bump can be labeled separately as well to analyze individual relations. However, such labeling were not used in this study due to the insufficient number of samples per event type.

5.2   Results and Discussion

An interview was made with the head of Office of Domestic Services to understand the domain and to discuss and evaluate the results. According to the domain knowledge, speed bumps decrease the number of accidents, especially for the region known by the drivers such as campus. This is because of the fact that, the drivers learn the location of the speed bumps and drive slowly when they get closer to them. The results for the overall effect are given in Table 5.1. P_1, P_2 and P_3 are the probability values obtained by using three different construction date (min, mean and max, respectively) based on the provided construction year. Based on this value, there is significant variation in the results. Unlike the general expectations of the domain experts, the probability of an accident to be happen when a speed bump exists around is higher than when it does not exist. The difference is minimum when the construction year is considered as the first day of the construction year. In addition, the other possible patterns also have high probabilities. Some of them can be explained by the domain knowledge. For example, speed bumps at certain regions are usually made at the same time period. Therefore, the patterns $Exist \rightarrow Exist$ or $NotExist \rightarrow NotExist$ might have high probabilities since they are usually sampled together. There might be positive or negative correlations between the speed bumps and the accidents at different regions. These local correlations might be invisible when global behavior is examined. It is hard to evaluate the correctness of the results due to the limitations posed by the data such as incomplete information about the dates and the routes.

Table 5.1: Prediction for the Overall Effect of Speed Bumps

Pattern	P_1	P_2	P_3
Accident → Accident	0.5	0.16	0
Not Exist → Accident	0.27	0.21	0.16
Exist → Accident	0.38	0.41	0.5
Accident → Not Exist	0.49	0.17	0
Not Exist → Not Exist	0.29	0.37	0.18
Exist → Not Exist	0.26	0.46	0.26
Accident → Exist	0.45	0.18	0.6
Not Exist → Exist	0.15	0.49	0.25
Exist → Exist	0.6	0.44	0.27

In Table 5.2, the results for the regional effects are given. The results obtained by

using minimum date for the construction year are more likely to be expected by the domain experts. Therefore, it is assumed that the values in column P_1 are correct and the discussions were made based on these values. For example, the probability of an accident occurred around the Mining Engineering department is 0.27 when there is no speed bump whereas it is 0 after construction. Similarly, around the Environmental Engineering, Courts and gate A1, these probabilities are 0.54, 0.85 and 0.74, respectively, when there is no speed bump. After constructions, probability of accident drops to zero. In some region, construction of speed bump reduces the risk of accident little, for example, at Residents area, it is decreases by 0.06, however, elevated risk is still 0.78. There might be some other factors in that region that trigger accidents.

Table 5.2: Prediction for the Regional Effect of Speed Bumps

Pattern	P_1	P_2	P_3
Mining- $\rightarrow$ Accident	0.27	0	0
Mining+ $\rightarrow$ Accident	0	0	0
BasicEng- $\rightarrow$ Accident	0	0	0
BasicEng+ $\rightarrow$ Accident	0.5	0.4	0.4
Environmental- $\rightarrow$ Accident	0.54	0	0
Environmental+ $\rightarrow$ Accident	0	0	0
Residents- $\rightarrow$ Accident	0.84	0	0.56
Residents+ $\rightarrow$ Accident	0.78	1	-
Court- $\rightarrow$ Accident	0.85	0	0.5
Court+ $\rightarrow$ Accident	0	0.34	0.39
Civil- $\rightarrow$ Accident	0	0	0
Civil+ $\rightarrow$ Accident	1	1	0.7
A1- $\rightarrow$ Accident	0.74	1	1
A1+ $\rightarrow$ Accident	0	0	0.5
Education- $\rightarrow$ Accident	1	1	1
ISBank+ $\rightarrow$ Accident	0.5	0	0
PresidentsOffice+ $\rightarrow$ Accident	1	0	0
Shopping- $\rightarrow$ Accident		1	1
Shopping+ $\rightarrow$ Accident	0	0.77	0.56
Cafeteria+ $\rightarrow$ Accident	0	0.41	0.86
Museum- $\rightarrow$ Accident	0.53	0.61	0.66

However, the effect is opposite in the vicinity of the Department of Basic English, Civil Engineering, Is Bank and President's Office. In other words, it is found that in these areas, speed bumps increase the risk for accidents. When I discussed these results with the domain expert, he stated that the causes of the accidents vary. For example, around the Department of Basic English, there are some other known reasons for the accidents. There is a bus bulb for the service buses and taxies to discharge and pick up passengers. However, drivers of the taxies do not usually use this area, instead, pick up or drop passenger on the road which results in impatient drivers overtaking.

It is claimed that this is the one of the main causes of accidents in this region. There is also a junction nearby which is on the main road. As a result, explaining accidents only with the speed bumps might be insufficient for this case.

It is clear that there may be many other causes of accidents and even tough a complete data set of the speed bumps and accidents is obtained, only a limited portion of the variation in accidents can be explained by speed bumps. In addition to speed bumps, the weather, psychology of the driver, pedestrians, or other factors might be the reasons for the accidents.

As discussed in data preprocessing steps, the data set is limited to evaluate the capability of the proposed algorithm. The most important reason is incomplete information. Besides, the domain is not a realization of a point process. Therefore,sa data type conversion was made to have a spatio-temporal point data. It might be better to evaluate every speed bump individually by considering each as different type. However, several speed bumps in the same region were grouped since the number of samples per each is insufficient otherwise.

CHAPTER 6

CONCLUSION

In this thesis, sequential triggering pattern of spatio-temporal event types was studied. A spatio-temporal event can be an observed disease, the location where a forest fire has started, a crime committed at a location, a traffic accident, and so on. We believe that exploring such relationships can provide valuable information for several domains such as public safety, crime prevention, epidemiology and environmental studies. In the study, new methodologies were proposed to solve sequential pattern mining problem which is usually handled by frequency based candidate generation approaches in the literature. The methodologies operate on continuous spatio-temporal domain. Specifically, the focus is spatio-temporal sequences the elements of which are related according to an unknown triggering or branching structure. In other words, the objective is to discover causal relationships among the events in a data set. Traditional frequency based approaches and the constant neighborhood thresholds commonly used in the sequential mining literature are not capable of explaining such complex and unknown relationships. In these approaches, the relationships between the observations are determined based on only the closeness of the data samples which are needed to be defined by the user. However, spatio-temporal correlation as a result of the closeness does not necessarily imply a causal relationship although it may provide a clue for understanding causation. Such sequences can be the result of any relationships or other factors. If the process is assumed to be the result of any causal structure, it is appropriate to extract the sequences based on a model that is able to describe its behavior.

Due to the limitations of the existing sequential pattern mining approaches to discover causal patterns, novel methodologies were proposed which use conditional intensity model to define the data as a realization of self-exciting or mutually exciting point process. Conditional intensity model assumes that the process consists of events which are either an immigrant or a descendant. In the literature, studies which use conditional intensity models can be found mostly as the application of seismological science. Recently, these models are utilized to describe the processes in some other domains such as criminology and social media. None of these studies concerned with the triggering pattern extraction problem. The proposed methodologies in this thesis use conditional intensity models and stochastic declustering for sequential pattern mining, particularly to solve triggering pattern extraction problem. However, the estimations of the mean rate of the immigrants and the total intensity of the process are challenging issues. The difficulty arises due to the fact that decay law which controls interaction scale in space and time, and the productivity which controls number of

triggered events per independent causes may differ based on the domain, event type, measurements describing the events and spatio-temporal proximity of the events. The triggering behavior of the process can be a consequence of either short range or long range interaction. Moreover, interaction at different scales can also be observed in the same process which results in highly mixture models. As a result, using simple parametric functions in the model is not usually capable of describing the complex process behavior. It is recommended to use domain specific functions in such models, however, they are usually unknown. Nonparametric approaches might have potential to get insight about the underlying behavior of a process which is not known well. In this perspective, the model independent stochastic declustering method was used for pairwise triggering pattern extraction. The original algorithm was improved for the processes where there are more than one type by using multivariate Hawkes model. The triggering probabilities were estimated between the events and then causal relationships between the types were extracted based on these probability values.

6.1 Discussions Based on Research Questions

In the first research question, the objective is to develop an algorithm to extract pairwise triggering patterns of event types by considering causal relationship. The proposed methodologies use conditional intensity models which describe excitation behavior in a process. Therefore, these methods can be considered as triggering pattern extraction algorithms.

In order to relate the observations, the existing methods use neighborhoods bounded by thresholds. In the second research question, the objective is to define a spatio-temporal neighborhood which is not bounded by threshold. This is because of the fact that the threshold values are domain specific, usually unknown, and the results are sensitive to these values. In Method 1, second objective is not achieved. Here, the original stochastic declustering algorithm was used with some basic parametric functions such as exponential and Gaussian. Then, the values of model parameters from the final solution were used as neighborhood thresholds during the post processing where triggering events are extracted. In this method, the success of the results is highly affected by the parameters' value. In other words, its success heavily relies on correct estimation of the model parameters which might be hard to achieve when the sample size is small. In Method 2, Method 3 and Method 4, such thresholds are not required. In the conditional intensity models, the density functions of distances are considered as neighborhood functions. Based on these functions, the observations are considered as neighbors with some degree according to the decay law defined by the functions. Thus, a neighborhood threshold is not needed. However, in Method 2, a threshold was used to prune the tail of the spatial function since otherwise it is not able to estimate the mean rate of the process and the ground-truth pairs. The threshold value used in the distance function is the significant interaction range obtained by Ripley's K-function. In Method 3, however, the ground-truth pairs are successfully estimated without using a cut off point although the mean rate of the process is still estimated incorrectly. This issue is not observed in Method 4 where event types of pairs are used during the learning. Therefore, the best solution for the second research question is Method 4 which does not require any sort of thresholds to determine neighbors.

Regarding the third research question, use of density function of distances as a neighborhood function in all methods allows evaluation of each pair of instances individually based on the distribution of distances. The decay law defined by the density functions represents subsiding strength of the relationship as the distance increases. In Method 1, on the other hand, constant thresholds obtained from the fitted models were used to assign pairs as neighbor or not. Therefore, the third objective is achieved by the remaining methods.

As defined in MISD algorithm, Method 2 and the multivariate approaches proposed in Method 3 and Method 4, the triggering probability between the pair of sample instances can be estimated by using the density function of distances and the entire intensity model. Thus, the probability of being a cause and effect can be calculated for each pair. If there is a triggering relationship between the events at some scale, it is more likely to observe many pairs the distances of which are within the interaction scale compared to the random distribution. Thus, for a generated event and its generator, the value of density function accounts for higher portion of the total intensity at the time of generated event. As a result, high probability value is expected for that pair. In all nonparametric methods proposed, triggering probabilities can be calculated by similar way. The differences in the calculations arise due to the definition of the random variables representing triggering relationship. For example, in Method 2, there is no type information during the modeling and the estimations are made based on the univariate Hawkes model. In Method 4, on the other hand, the triggering relationship is defined by conditional random variables which use event types for conditioning and the estimations are made by using multivariate Hawkes model. According to the results on the simulated data sets, Method 2, which uses binary constant in density function based on the significant spatial interaction ranges, successfully finds most of the true pairs in the simple cases described in Scenario I with higher probabilities than the irrelevant pairs. It is observed that TP rates decreases as the spatial density increases in this method. The precision of the predictions are high. Method 3 is also successful for the same cases. It produces higher probabilities for the true pairs, however, the precision of the predictions is low if small threshold value is used or rank selection is utilized. This disadvantage is eliminated by filtering initial results with Otsu thresholding method. It significantly improved the initial results. Such relation between TP rates and spatial density is not observed in this method. Unlike the simple cases, Method 3 is unsuccessful to discover triggering relations simulated in Scenario II. Method 4 which employs type information of both triggering and triggered events in the model is the most successful approach among the others to produce higher probabilities for the ground-truth pairs. It is able to extract pairs in complex cases described in Scenario II. Moreover, the estimated probability values for the true pairs are far from the irrelevant ones. In Method 4, the degree of causal relationship for a pairwise triggering pattern is calculated from the estimated probabilities of the pairs constituting the pattern. Due to the skewed distribution of the selected pairs' probabilities for a particular pattern, mid-mean, the summary statistic of the probabilities, is used to estimate significance of the triggering pattern. Although mid-mean statistic is used to calculate significance of extracted patterns only in Method 4, it can be used in Method 2 and Method 3 as well.

Finally, regarding the fifth research question, in Method 4, significant patterns are identified based on their ranks in terms of pattern significance. It is observed that

ground-truth patterns have high rank and are usually placed within the first five or ten. Moreover, most of the true pairs constituting the patterns are selected during the pattern extraction phase, and thus, contribute the significance values. However, there are some limitations which can be handled by paying particular attention. For example, if frequency of a pattern is very small compared to the others in the data, then its contribution to the total intensity of the process might have been suppressed by those having higher frequencies. As a result, the probability estimations for the corresponding pairs might be very small or zero. This issue can be eliminated by balancing the data distributions. If the interaction scale of the infrequent pattern is different than the others, then, discretization of the nonparametric density function might be done appropriately such that it represents peaks for the infrequent pattern as well.

The case study discussed in this thesis is limited to evaluate capability of the presented work. The results are affected by incomplete information. However, the approach can be considered as successful in some aspects. The domain experts claim that they observed a decrease in the number of accidents after construction of speed bumps. Based on this knowledge, the most representative results were determined and evaluated. Therefore, in this case study, the aim is not to extract unknown patterns, instead, it is to determine the strength of the relation. Triggering pattern probabilities for each region before and after construction of the speed bumps can be used to determine the strength of the relation. If the difference is high it can be considered as the existence or non-existence of speed bump in the region has high effect on accidents.

By considering the general characteristics, the capability of the proposed methods can be summarized as in Table 6.1. The first two methods do not use type information during the learning phase whereas Method 3 and Method 4 use information about all four dimensions of the problem. The neighborhood threshold required in Method 1 is provided by the system after model fitting. Candidate generation based approaches require user to define the neighborhood threshold. The remaining three do not need this information. In both Method 1 and Method 2 a significance threshold is needed to be defined to extract the patterns as similar to the studies in the literature. In Method 3, one may use a significance threshold during the pattern extraction phase, or may use rank method without providing a threshold. The former surpasses the latter in terms of precision of the initial results, however, after applying Otsu thresholding to eliminate the irrelevant patterns, their difference is comparable. Method 4, on the other hand, is able to eliminate irrelevant patterns with rank selection since the probability of the true patterns and the irrelevant ones are well separated. For all of the proposed methods, the resulting patterns represent causal relationship at some degree since they are extracted based on the triggering probabilities obtained by using a conditional intensity model. The existing sequential pattern mining algorithms does not necessarily extract causal patterns from the data, instead, they generate patterns of correlated events since interaction is explained only by closeness. Spatio-temporal data handling is provided by all.

Table 6.1: General Characteristics of Proposed Methods

Feature	Candidate Generation Based Approaches	Method 1	Method 2	Method 3	Method 4
Spatio-temporal data handling	✓	✓	✓	✓	✓
Casual relationship	not necessarily	✓	✓	✓	✓
Multitype handling	during learning	during post processing	during post processing	during learning	during learning
Neighborhood threshold	required	required	not required	not required	not required
Significance threshold	required	required	required	not required	not required

6.2 Future Work

A number of tasks to improve proposed methods is considered as a future work.

In method 1, the performance is limited due to the incorrect parameter estimates in spatial density function for some cases. Gaussian family was used for spatial densities, in line with the literature, however, the functions used in those studies are special case of the Gaussian family developed for corresponding domain. Therefore, simple models might be incapable to reflect the process behavior. Other parametric models or kernels with dynamic bandwidth might be used for further evaluation.

More importantly, computational performance of the methods are needed to be improved. The proposed methods were used for small to medium size data sets. However, scalability issue is faced when large data sets are studied. Performance is affected by discretization level in addition to sample size. Besides the improvement that can be made in the implementation, sampling methods can be developed for large data sets similar to the method Mohler et. al. proposed in [28] for the same performance issues.

Another important future work is to extend pairwise triggering patterns extraction algorithms to construct a causal network. In this future study, one task is to generate entire causal network efficiently, and the second task is to define significance of any causal patterns within the network.

Finally, the proposed methods can be tested with new real world cases which are more suitable for the problem definition in this thesis.

REFERENCES

[1] R. Agrawal and R. Srikant. Mining sequential patterns. In *Data Engineering, 1995. Proceedings of the Eleventh International Conference on*, pages 3–14. IEEE, 1995.

[2] O. Avcı, M. Kaçar, U. Kozan, and O. Şişko. Karadeniz Sahil Yolu suya set çekti, http://www.hurriyet.com.tr/gundem/15661993.asp, 2010.

[3] M. N. Bhrolcháin. Divorce effect and causality in the social sciences. *European Sociological Review*, 17(1):33–57, 2001.

[4] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[5] H. Cao, D. W. Cheung, and N. Mamoulis. Discovering partial periodic patterns in discrete data sequences. In *Advances in Knowledge Discovery and Data Mining*, pages 653–658. Springer, 2004.

[6] H. Cao, N. Mamoulis, and D. W. Cheung. Mining frequent spatio-temporal sequential patterns. In *Data Mining, Fifth IEEE International Conference on*, pages 8–pp. IEEE, 2005.

[7] M. Celik, S. Shekhar, J. P. Rogers, and J. A. Shine. Mixed-drove spatiotemporal co-occurrence pattern mining. *Knowledge and Data Engineering, IEEE Transactions on*, 20(10):1322–1335, 2008.

[8] P. J. Diggle, A. G. Chetwynd, R. Häggkvist, and S. E. Morris. Second-order analysis of space-time clustering. *Statistical methods in medical research*, 4(2):124–136, 1995.

[9] P. Embrechts, T. Liniger, and L. Lin. Multivariate hawkes processes: an application to financial data. *Journal of Applied Probability*, 48:367–378, 2011.

[10] M. N. Garofalakis, R. Rastogi, and K. Shim. Spirit: Sequential pattern mining with regular expression constraints. In *VLDB*, volume 99, pages 7–10, 1999.

[11] A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[12] K. S. Hornsby and S. Cole. Modeling moving geospatial objects from an event-based perspective. *Transactions in GIS*, 11(4):555–573, 2007.

[13] Y. Huang, S. Shekhar, and H. Xiong. Discovering colocation patterns from spatial data sets: a general approach. *Knowledge and Data Engineering, IEEE Transactions on*, 16(12):1472–1485, 2004.

[14] Y. Huang, L. Zhang, and P. Zhang. A framework for mining sequential patterns from spatio-temporal event data sets. *Knowledge and Data Engineering, IEEE Transactions on*, 20(4):433–448, 2008.

[15] D. Hume. *A treatise of human nature*. Courier Dover Publications, 2012.

[16] S. D. Johnson and K. J. Bowers. The burglary as clue to the future the beginnings of prospective hot-spotting. *European Journal of Criminology*, 1(2):237–255, 2004.

[17] S. D. Johnson and K. J. Bowers. The stability of space-time clusters of burglary. *British Journal of Criminology*, 44(1):55–65, 2004.

[18] H. Kabahasanoğlu. Giresun-Trabzon yolunda heyelan, http://www.hurriyet.com.tr/gundem/8405921.asp, 2008.

[19] D. Koller and N. Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[20] D.-J. Lan, Y.-F. Ma, W.-Y. Ma, and H.-J. Zhang. Spatio-temporal pattern mining in sports video. In *Advances in Multimedia Information Processing-PCM 2004*, pages 306–313. Springer, 2005.

[21] E. Lewis and G. Mohler. A nonparametric em algorithm for multiscale hawkes processes. *Preprint*, 2011.

[22] E. Lewis, G. Mohler, P. J. Brantingham, and A. L. Bertozzi. Self-exciting point process models of civilian deaths in iraq. *Security Journal*, 25(3):244–264, 2011.

[23] H. Mannila, H. Toivonen, and A. I. Verkamo. Discovery of frequent episodes in event sequences. *Data Mining and Knowledge Discovery*, 1(3):259–289, 1997.

[24] D. Marsan and O. Lengline. Extending earthquakes' reach through cascading. *Science*, 319(5866):1076–1079, 2008.

[25] J. S. Mill. *A System of Logic Ratiocinative and Inductive: 1*. Longmans, 1872.

[26] P. Mohan, S. Shekhar, J. A. Shine, and J. P. Rogers. Cascading spatio-temporal pattern discovery. *Knowledge and Data Engineering, IEEE Transactions on*, 24(11):1977–1992, 2012.

[27] G. Mohler, M. Short, P. Brantingham, F. Schoenberg, and G. Tita. Self-exciting point processes explain spatial-temporal patterns in crime. *Submitted for Review, accessed*, 17, 2010.

[28] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493), 2011.

[29] F. Mosteller and J. W. Tukey. Data analysis and regression. a second course in statistics. *Addison-Wesley Series in Behavioral Science: Quantitative Methods, Reading, Mass.: Addison-Wesley, 1977*, 1, 1977.

[30] F. Musmeci and D. Vere-Jones. A space-time clustering model for historical earthquakes. *Annals of the Institute of Statistical Mathematics*, 44(1):1–11, 1992.

[31] Y. Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27, 1988.

[32] Y. Ogata. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402, 1998.

[33] N. Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.

[34] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY, USA, 2000.

[35] B. D. Ripley. *Spatial statistics*, volume 575. Wiley. com, 2005.

[36] S. Ross. *Simulation, Third Edition*. Academic Press, 2002.

[37] A. Simma and M. I. Jordan. Modeling events with cascades of poisson processes. In *Proceedings of the 26th Conference on UAI*, pages 546–555, 2010.

[38] R. Srikant and R. Agrawal. Mining sequential patterns: Generalizations and performance improvements. In *Proceedings of the 5th International Conference on Extending Database Technology: Advances in Database Technology*, EDBT '96, pages 3–17, London, UK, UK, 1996. Springer-Verlag.

[39] A. Stefanidis, K. Eickhorst, P. Agouris, and P. Partsinevelos. Modeling and comparing change using spatiotemporal helixes. In *Proceedings of the 11th ACM international symposium on Advances in geographic information systems*, pages 86–93. ACM, 2003.

[40] J. Wang, W. Hsu, and M.-L. Lee. Flowminer: Finding flow patterns in spatiotemporal databases. In *Tools with Artificial Intelligence, 2004. ICTAI 2004. 16th IEEE International Conference on*, pages 14–21. IEEE, 2004.

[41] M. Worboys. Event-oriented approaches to geographic phenomena. *International Journal of Geographical Information Science*, 19(1):1–28, 2005.

67

[42] M. Worboys and K. Hornsby. From objects to events: Gem, the geospatial event model. In *Geographic Information Science*, pages 327–343. Springer, 2004.

[43] K. Zhou and H. Zha. Learning triggering kernels for multi-dimensional hawkes processes. In *Proceedings of the 30 th International Conference on Machine Learning*, 2013.

[44] J. Zhuang, Y. Ogata, and D. Vere-Jones. Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, 97(458):369–380, 2002.

APPENDIX A

DATA GENERATION

To simulate the defined cases which have arbitrary distributions, the values of corresponding random variables are simulated by using inverse transform method described in [36].

Poisson Random Variable

Probability mass function: $p_n = P(X = n) = e^- \lambda \frac{\lambda^n}{n!}$

Cumulative distribution function: $F(n) = P(X \leq n) = \sum_{k=0}^{n} P(X = k)$

To generate random number X, the relation $p_n = \frac{\lambda}{n} p_{n-1}$ is utilized.

Pseudo Code:

Step1: Generate a random uniform number U over (0,1).

Step2: $n = 0, p_n = e^- \lambda, F = p_n$.

Step3: If $U < F$, set $X = n$ and stop.

Step4: $n = n + 1, p_n = \frac{\lambda}{n} p_{n-1}, F = F + p_n$.

Step5: Go to Step3.

Exponential Random Variable

Probability mass function: $p_n = f(x) = \lambda e^{-\lambda x}$

Cumulative distribution function: $F(x) = P(X \leq x) = 1 - e^{\lambda x}$

To generate the value of random variable X from exponential distribution, the relation $x = F^{-1}(U)$ is utilized, where U is a random uniform number over (0,1). From this equation $X = \frac{1}{\lambda} log(U)$ can be derived which produces value of the exponential random variable.

APPENDIX B

KERNEL FUNCTIONS FOR INTENSITY ESTIMATION

B.1 Gaussian

Gaussian space function defined in equation B.1 has parameter D, representing bandwidth or spatial extent.

$$f(x,y) = \frac{1}{2\pi D} e^{-(x^2+y^2)/(2D^2)} \tag{B.1}$$

The value of D that maximizes Gaussian density over x and y is examined. Figure B.1 shows changes in value of $f(x = x_i, y = y_i$ as D changes. It can be seen from the figure, the maximum value of $f(x, y; D)$ decreases as x and y increase. There is also
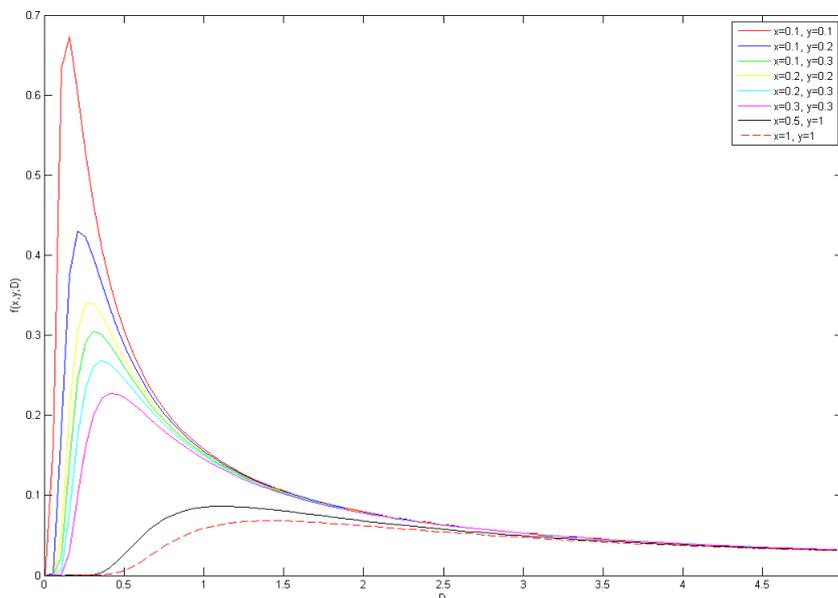

Figure B.1: Changes in Gaussian function over D given x and y.

a relation between the value of (x, y) and $\hat{D} = \arg\max_D f(x_i, y_i; D)$. This relation is correlated with quadratic mean or root mean square (RMS) value which is a measure

of the spread of a quantity. Definition of RMS and the relation between $\hat{D}$ and RMS are given in equation B.2 and B.3.

$$RMS(x_1, x_2, \ldots, x_n) = \sqrt{\frac{1}{n}(x_1^2 + x_2^2 + \cdots + x_n^2)} \tag{B.2}$$

$$\hat{D} \approx \sqrt{(x^2 + y^2)} = \sqrt{2}RMS(x, y) \tag{B.3}$$

In Table B.1, $\hat{D}$ and $\sqrt{2}RMS(x, y)$ values are given for the evaluated value of x and y shown in Figure B.1.

Table B.1: Experimental outputs for the relation between D and RMS

x	y	$\sqrt{2}RMS(x,y)$	$\hat{D}$
0.1	0.1	0.14	0.16
0.1	0.2	0.22	0.21
0.2	0.2	0.28	0.26
0.1	0.3	0.32	0.31
0.2	0.3	0.36	0.36
0.3	0.3	0.42	0.41
0.5	1	1.12	1.11
1	1	1.41	1.41

B.2 Exponential

The exponential function given in equation B.4 is considered to explain temporal behaviour. This function is used to model spatio-temporal patterns of crime in the study of Mohler et al. [27]. The function has two parameters K and W. In their study, they define the parameters as the elevated risk of effects following the crime events and the decay of this elevated risk, respectively. K takes value between 0 and 1, whereas $W^{(-1)} \in (0, n]$ ($W^{(-1)}$: unit of time). The definition of the function is

$$g(t; W, K) = KWe^{-Wt} \tag{B.4}$$

$g(t; W, K)$ has its maximum at $K = 1$ and $W = 1/t$. If $K = 1$, $g(t; W, K)$ is equal to the exponential probability distribution function, and therefore, its maximum is at the mean value of the distribution which is $W^{(-1)}$.

In Figure B.2, surface plot for $g(t = 0.1; W, K)$ is given. Its maximum value is at the point where $W = 1/t = 10$. For all t, the function has the same shape around the $1/t$. As t increases, the maximum value of the function decreases.
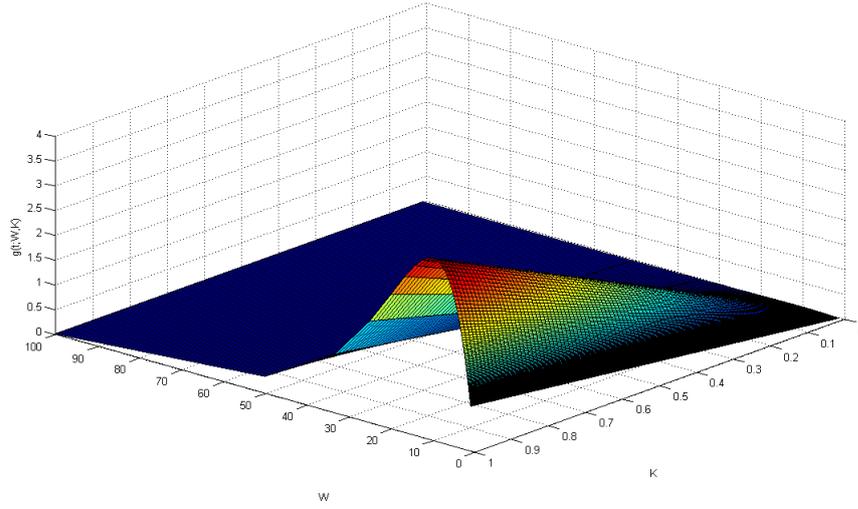
Figure B.2: Changes in Exponential function over W and K given $t = 0.1$
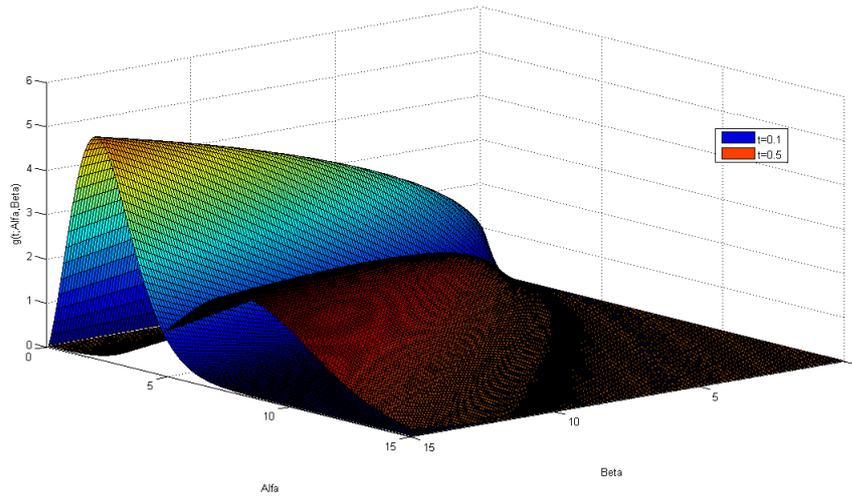
Figure B.3: Changes in Gamma function over α and β given $t = 0.1$ and $t = 0.5$

B.3   Gamma

Another function used to model temporal behaviour is the gamma distribution which is from two parameters distribution family. It is defined as in equation B.5.

$$g(t; \Gamma, \beta) = \beta^{\alpha} \frac{1}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t} \tag{B.5}$$

where $t, \alpha, \beta > 0$. α is the shape parameter and β is the rate or inverse scale parameter of the distribution. Given t, $g(t; \alpha, \beta)$ has its maximum where $t = \alpha/\beta$ such that α is the possible maximum value which satisfies this equality. This relation represents the

expected value of the distribution. In other words, the maximum value of the function is at the point where t is equal to the expected value satisfied by the possible maximal value of the parameters. In Figure B.3, the surface plots for the gamma function over the parameters given $t = 0.1$ and $t = 0.5$ are illustrated. The maximum value of the function decreases as t increases and the value of α that maximize the function shifted towards upper bound of the α as t increases.

In the intensity definition, x and y of the space function correspond to difference between the latitude values of an event and the previously occurred events, and the difference between the longitude values of an event and the previously occurred events, respectively. Similarly, t of time function corresponds to difference between the time of an event and the time of the previously occurred events. Therefore, if events are distant from each other in space-time these functions will produce very small values.

APPENDIX C

RESULTS OF TPEX ALGORITHM ON SYNTHETIC DATA

Table C.1: Pattern Rank

DataSet	Discretization	$DD \rightarrow VPR$		$C \rightarrow VPR$		$C \rightarrow P$		$C \rightarrow CL$		$CL \rightarrow F$	
		TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP
ss	$\delta s = 1, \delta t = 0.05$	3	-	4	-	1	-	2	-	13	1
	$\delta s = 4, \delta t = 0.5$	2	12	3	13	1	3	4	1	-	2
	$\delta s = 8, \delta t = 1$	1	11	3	13	2	3	4	1	-	2
	$\delta s = 16, \delta t = 2$	1	11	3	13	2	3	4	1	-	2
	$\delta s = 25, \delta t = 25$	1	9	4	16	2	2	3	1	15	4
ss5	$\delta s = 1, \delta t = 0.05$	3	-	5	-	2	-	1	-	14	1
	$\delta s = 4, \delta t = 0.5$	1	11	4	13	2	4	3	1	-	3
	$\delta s = 8, \delta t = 1$	1	11	4	13	3	4	2	1	27	3
	$\delta s = 16, \delta t = 2$	1	11	4	13	3	4	2	1	-	3
	$\delta s = 25, \delta t = 25$	2	7	8	23	4	9	7	1	18	2
ss10	$\delta s = 1, \delta t = 0.05$	5	-	3	-	1	-	2	17	4	1
	$\delta s = 4, \delta t = 0.5$	4	11	1	13	3	6	2	1	-	3
	$\delta s = 8, \delta t = 1$	2	11	4	13	1	6	3	1	-	3
	$\delta s = 16, \delta t = 2$	1	11	4	13	3	6	2	1	-	3
	$\delta s = 25, \delta t = 25$	3	8	7	21	5	10	4	1	19	2
sm	$\delta s = 1, \delta t = 0.05$	2	-	4	-	1	-	3	-	10	1
	$\delta s = 4, \delta t = 0.5$	2	12	3	16	1	1	4	2	-	4
	$\delta s = 8, \delta t = 1$	1	11	3	13	2	1	4	2	-	3
	$\delta s = 16, \delta t = 2$	1	12	3	13	2	1	4	2	-	4
	$\delta s = 25, \delta t = 25$	3	11	4	19	1	1	2	2	13	5
sm5	$\delta s = 1, \delta t = 0.05$	5	-	4	-	1	-	2	19	15	1
	$\delta s = 4, \delta t = 0.5$	3	12	2	13	1	3	4	1	-	4
	$\delta s = 8, \delta t = 1$	1	12	3	13	2	3	4	1	-	4
	$\delta s = 16, \delta t = 2$	1	12	4	14	2	4	3	1	-	3

Continued on next page

Table C.1 Pattern Rank – continued from previous page

DataSet	Discretization	DD→VPR		C→VPR		C→P		C→CL		CL→F	
		TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP
	$\delta s = 25, \delta t = 25$	3	9	9	25	5	8	6	1	16	4
sm10	$\delta s = 1, \delta t = 0.05$	5	-	4	-	2	-	1	20	11	1
	$\delta s = 4, \delta t = 0.5$	3	11	2	13	1	4	4	1	-	2
	$\delta s = 8, \delta t = 1$	1	11	3	13	2	4	4	1	27	2
	$\delta s = 16, \delta t = 2$	1	12	4	13	2	4	3	1	26	2
	$\delta s = 25, \delta t = 25$	6	8	8	19	3	10	1	1	13	2
sl	$\delta s = 1, \delta t = 0.05$	3	-	4	-	1	-	2	-	9	1
	$\delta s = 4, \delta t = 0.5$	2	11	3	14	1	1	4	2	-	6
	$\delta s = 8, \delta t = 1$	2	11	3	13	1	1	4	2	-	6
	$\delta s = 16, \delta t = 2$	2	11	3	14	1	1	4	2	18	7
	$\delta s = 25, \delta t = 25$	2	11	3	16	1	1	4	2	11	7
sl5	$\delta s = 1, \delta t = 0.05$	5	-	3	-	2	-	1	15	11	1
	$\delta s = 4, \delta t = 0.5$	2	12	3	14	1	2	4	1	21	4
	$\delta s = 8, \delta t = 1$	2	12	3	13	1	2	4	1	20	4
	$\delta s = 16, \delta t = 2$	1	12	3	13	2	2	4	1	18	4
	$\delta s = 25, \delta t = 25$	3	8	10	23	4	5	8	2	9	3
sl10	$\delta s = 1, \delta t = 0.05$	7	-	4	-	2	-	1	16	5	1
	$\delta s = 4, \delta t = 0.5$	3	12	2	14	1	3	4	1	-	2
	$\delta s = 8, \delta t = 1$	2	12	3	14	1	3	4	1	-	2
	$\delta s = 16, \delta t = 2$	2	12	4	14	1	4	3	1	-	2
	$\delta s = 25, \delta t = 25$	3	10	8	24	6	9	4	4	17	3
svl	$\delta s = 1, \delta t = 0.05$	2	-	3	-	1	-	4	18	9	2
	$\delta s = 4, \delta t = 0.5$	2	12	3	17	1	1	4	2	21	6
	$\delta s = 8, \delta t = 1$	5	11	4	13	3	1	2	2	-	6
	$\delta s = 16, \delta t = 2$	1	11	4	14	2	1	3	2	-	7

Continued on next page

Table C.1 Pattern Rank – continued from previous page

DataSet	Discretization	DD→VPR		C→VPR		C→P		C→CL		CL→F	
		TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP
	$\delta s = 25, \delta t = 25$	1	11	3	15	2	1	4	2	12	7
svl5	$\delta s = 1, \delta t = 0.05$	7	-	4	-	3	-	1	16	5	1
	$\delta s = 4, \delta t = 0.5$	3	12	2	14	1	2	4	1	-	5
	$\delta s = 8, \delta t = 1$	3	12	2	14	1	2	4	1	-	5
	$\delta s = 16, \delta t = 2$	4	12	3	15	1	3	2	1	15	4
	$\delta s = 25, \delta t = 25$	1	12	9	22	2	8	6	1	11	5
svl10	$\delta s = 1, \delta t = 0.05$	5	-	4	-	2	-	1	19	6	1
	$\delta s = 4, \delta t = 0.5$	3	12	2	14	1	4	4	1	-	5
	$\delta s = 8, \delta t = 1$	3	12	2	14	1	4	4	1	-	5
	$\delta s = 16, \delta t = 2$	1	12	4	16	2	5	3	1	-	4
	$\delta s = 25, \delta t = 25$	1	8	4	25	2	10	3	1	9	5
ls	$\delta s = 1, \delta t = 0.05$	3	-	4	-	2	-	1	12	7	1
	$\delta s = 4, \delta t = 0.5$	2	12	3	13	1	4	4	1	-	3
	$\delta s = 8, \delta t = 1$	1	11	3	13	2	4	4	1	-	2
	$\delta s = 16, \delta t = 2$	1	12	3	13	2	4	4	1	-	2
	$\delta s = 25, \delta t = 25$	1	6	4	21	2	5	3	1	13	2
ls5	$\delta s = 1, \delta t = 0.05$	6	-	3	-	2	-	1	18	5	1
	$\delta s = 4, \delta t = 0.5$	4	11	3	13	1	3	2	1	-	4
	$\delta s = 8, \delta t = 1$	1	11	4	13	2	4	3	1	-	3
	$\delta s = 16, \delta t = 2$	1	11	5	13	3	4	2	1	-	3
	$\delta s = 25, \delta t = 25$	3	7	14	23	7	12	5	1	15	4
ls10	$\delta s = 1, \delta t = 0.05$	8	-	3	-	2	-	1	20	4	1
	$\delta s = 4, \delta t = 0.5$	4	11	3	13	1	4	2	1	-	3
	$\delta s = 8, \delta t = 1$	1	11	4	13	3	6	2	1	-	3
	$\delta s = 16, \delta t = 2$	3	11	5	13	1	6	2	1	-	3

Table C.1 Pattern Rank – continued from previous page

DataSet	Discretization	DD→VPR		C→VPR		C→P		C→CL		CL→F	
		TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP
	$\delta s = 25, \delta t = 25$	3	9	15	24	6	15	4	1	13	3
lm	$\delta s = 1, \delta t = 0.05$	1	-	4	-	3	-	2	20	11	1
	$\delta s = 4, \delta t = 0.5$	2	12	3	14	1	1	4	2	-	4
	$\delta s = 8, \delta t = 1$	2	11	3	13	4	1	1	2	11	5
	$\delta s = 16, \delta t = 2$	1	12	4	14	2	1	3	2	16	5
	$\delta s = 25, \delta t = 25$	2	10	5	21	1	1	3	3	9	4
lm5	$\delta s = 1, \delta t = 0.05$	6	-	4	-	2	-	1	20	5	1
	$\delta s = 4, \delta t = 0.5$	3	11	2	13	1	2	4	1	-	4
	$\delta s = 8, \delta t = 1$	4	12	3	13	1	2	2	1	-	3
	$\delta s = 16, \delta t = 2$	3	12	4	14	2	4	1	1	-	3
	$\delta s = 25, \delta t = 25$	1	8	13	23	4	10	2	1	7	3
lm10	$\delta s = 1, \delta t = 0.05$	7	-	3	-	2	-	1	18	5	1
	$\delta s = 4, \delta t = 0.5$	4	12	3	13	1	4	2	1	-	3
	$\delta s = 8, \delta t = 1$	2	12	5	13	3	5	1	1	-	2
	$\delta s = 16, \delta t = 2$	2	12	4	13	1	7	3	1	-	2
	$\delta s = 25, \delta t = 25$	2	9	17	25	5	16	3	1	13	4
ll	$\delta s = 1, \delta t = 0.05$	2	-	4	-	3	-	1	20	12	1
	$\delta s = 4, \delta t = 0.5$	2	11	3	16	1	1	4	2	-	7
	$\delta s = 8, \delta t = 1$	5	11	4	14	3	1	2	2	-	7
	$\delta s = 16, \delta t = 2$	4	11	5	15	2	1	3	2	-	7
	$\delta s = 25, \delta t = 25$	1	9	5	21	2	2	4	1	14	8
ll5	$\delta s = 1, \delta t = 0.05$	6	-	3	-	2	-	1	-	5	1
	$\delta s = 4, \delta t = 0.5$	2	11	3	14	1	2	4	1	21	4
	$\delta s = 8, \delta t = 1$	4	11	3	14	1	2	2	1	-	4
	$\delta s = 16, \delta t = 2$	1	12	4	15	2	2	3	1	14	4

Continued on next page

Table C.1 Pattern Rank – continued from previous page

DataSet	Discretization	$DD \rightarrow VPR$		$C \rightarrow VPR$		$C \rightarrow P$		$C \rightarrow CL$		$CL \rightarrow F$	
		TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP
	$\delta s = 25, \delta t = 25$	3	8	15	23	4	9	5	1	10	5
ll10	$\delta s = 1, \delta t = 0.05$	7	-	4	-	2	-	3	18	6	1
	$\delta s = 4, \delta t = 0.5$	3	12	2	14	1	4	4	1	-	2
	$\delta s = 8, \delta t = 1$	5	12	4	15	2	5	1	1	-	2
	$\delta s = 16, \delta t = 2$	2	12	5	16	3	5	1	1	-	2
	$\delta s = 25, \delta t = 25$	3	9	11	25	5	13	2	2	7	4
lvl	$\delta s = 1, \delta t = 0.05$	3	-	4	-	6	-	2	18	1	2
	$\delta s = 4, \delta t = 0.5$	2	11	3	16	1	1	4	2	-	6
	$\delta s = 8, \delta t = 1$	2	11	3	14	1	1	5	2	-	6
	$\delta s = 16, \delta t = 2$	1	12	5	14	2	1	4	2	-	6
	$\delta s = 25, \delta t = 25$	2	10	5	21	1	1	3	3	10	6
lvl5	$\delta s = 1, \delta t = 0.05$	1	-	3	-	6	-	7	20	2	2
	$\delta s = 4, \delta t = 0.5$	2	12	3	14	1	2	4	1	-	7
	$\delta s = 8, \delta t = 1$	1	12	3	14	4	2	6	1	-	7
	$\delta s = 16, \delta t = 2$	1	12	5	15	2	3	6	1	-	7
	$\delta s = 25, \delta t = 25$	2	8	13	23	4	9	5	1	17	5
lvl10	$\delta s = 1, \delta t = 0.05$	7	-	4	-	2	-	1	18	5	1
	$\delta s = 4, \delta t = 0.5$	2	12	3	14	1	4	4	1	-	5
	$\delta s = 8, \delta t = 1$	5	12	6	14	2	5	3	1	-	4
	$\delta s = 16, \delta t = 2$	1	12	5	16	4	6	2	1	-	4
	$\delta s = 25, \delta t = 25$	2	8	11	27	4	11	1	4	9	5

Table C.2: Pattern Probabilities

DataSet	Discretization	$DD \to VPR$		$C \to VPR$		$C \to P$		$C \to CL$		$CL \to F$	
		TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP
ss	$\delta s = 1, \delta t = 0.05$	0.75	0	0.74	0	0.85	0	0.82	0	0.18	0.83
	$\delta s = 4, \delta t = 0.5$	0.98	0.54	0.96	0.35	1	0.83	0.89	0.92	0	0.83
	$\delta s = 8, \delta t = 1$	1	0.6	0.99	0.45	0.99	0.83	0.86	0.92	0	0.84
	$\delta s = 16, \delta t = 2$	0.98	0.61	0.9	0.46	0.93	0.85	0.78	0.92	0	0.85
	$\delta s = 25, \delta t = 25$	0.18	0.86	0.09	0.68	0.16	0.94	0.11	0.97	0.02	0.94
ss5	$\delta s = 1, \delta t = 0.05$	0.78	0	0.66	0	0.92	0	0.93	0	0.18	0.86
	$\delta s = 4, \delta t = 0.5$	0.99	0.53	0.97	0.37	0.99	0.84	0.97	0.97	0	0.86
	$\delta s = 8, \delta t = 1$	1	0.54	0.93	0.41	0.94	0.84	0.94	0.97	0.02	0.87
	$\delta s = 16, \delta t = 2$	0.97	0.56	0.85	0.42	0.86	0.84	0.89	0.97	0	0.88
	$\delta s = 25, \delta t = 25$	0.16	0.88	0.07	0.66	0.14	0.87	0.09	0.99	0.03	0.95
ss10	$\delta s = 1, \delta t = 0.05$	0.55	0	0.69	0	0.93	0	0.92	0.01	0.58	0.88
	$\delta s = 4, \delta t = 0.5$	0.86	0.5	0.97	0.35	0.96	0.71	0.97	0.95	0	0.88
	$\delta s = 8, \delta t = 1$	0.93	0.52	0.89	0.39	0.93	0.71	0.92	0.96	0	0.88
	$\delta s = 16, \delta t = 2$	0.89	0.54	0.8	0.39	0.82	0.71	0.88	0.96	0	0.89
	$\delta s = 25, \delta t = 25$	0.14	0.85	0.08	0.64	0.11	0.77	0.12	0.98	0.01	0.97
sm	$\delta s = 1, \delta t = 0.05$	0.8	0	0.64	0	0.81	0	0.64	0	0.21	0.86
	$\delta s = 4, \delta t = 0.5$	0.97	0.52	0.94	0.38	0.99	0.99	0.88	0.96	0	0.86
	$\delta s = 8, \delta t = 1$	1	0.56	0.96	0.46	0.96	0.99	0.83	0.97	0	0.87
	$\delta s = 16, \delta t = 2$	0.98	0.58	0.82	0.48	0.87	0.99	0.73	0.98	0	0.88
	$\delta s = 25, \delta t = 25$	0.22	0.88	0.16	0.65	0.27	1	0.23	0.99	0.01	0.93
sm5	$\delta s = 1, \delta t = 0.05$	0.57	0	0.7	0	0.93	0	0.92	0.01	0.13	0.84
	$\delta s = 4, \delta t = 0.5$	0.96	0.51	0.97	0.41	1	0.87	0.94	0.96	0	0.85
	$\delta s = 8, \delta t = 1$	0.99	0.53	0.95	0.45	0.98	0.87	0.92	0.96	0	0.86
	$\delta s = 16, \delta t = 2$	0.95	0.55	0.81	0.45	0.89	0.87	0.82	0.96	0	0.88

Table C.2 Pattern Probabilities – continued from previous page

DataSet	Discretization	DD→VPR		C→VPR		C→P		C→CL		CL→F	
		TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP
	$\delta s = 25, \delta t = 25$	0.2	0.89	0.11	0.68	0.17	0.89	0.12	1	0.02	0.97
sm10	$\delta s = 1, \delta t = 0.05$	0.55	0	0.65	0	0.89	0	0.92	0.01	0.24	0.82
	$\delta s = 4, \delta t = 0.5$	0.95	0.55	0.96	0.32	0.98	0.78	0.94	0.93	0	0.83
	$\delta s = 8, \delta t = 1$	0.97	0.58	0.93	0.37	0.95	0.78	0.9	0.93	0.01	0.83
	$\delta s = 16, \delta t = 2$	0.92	0.59	0.77	0.37	0.84	0.79	0.82	0.93	0.02	0.84
	$\delta s = 25, \delta t = 25$	0.13	0.87	0.04	0.64	0.14	0.82	0.2	0.98	0.02	0.96
sl	$\delta s = 1, \delta t = 0.05$	0.69	0	0.67	0	0.72	0	0.71	0	0.23	0.84
	$\delta s = 4, \delta t = 0.5$	0.93	0.54	0.93	0.37	1	1	0.82	0.91	0	0.84
	$\delta s = 8, \delta t = 1$	0.98	0.6	0.95	0.47	0.99	1	0.77	0.92	0	0.85
	$\delta s = 16, \delta t = 2$	0.94	0.62	0.88	0.48	0.96	1	0.7	0.94	0.01	0.85
	$\delta s = 25, \delta t = 25$	0.23	0.88	0.2	0.67	0.27	1	0.16	0.99	0.02	0.94
sl5	$\delta s = 1, \delta t = 0.05$	0.54	0	0.61	0	0.83	0	0.9	0.01	0.18	0.84
	$\delta s = 4, \delta t = 0.5$	0.97	0.52	0.96	0.42	1	0.91	0.85	0.92	0.05	0.84
	$\delta s = 8, \delta t = 1$	0.99	0.54	0.98	0.45	1	0.91	0.81	0.92	0.05	0.85
	$\delta s = 16, \delta t = 2$	0.97	0.55	0.88	0.48	0.96	0.91	0.75	0.92	0.05	0.86
	$\delta s = 25, \delta t = 25$	0.13	0.92	0.04	0.72	0.12	0.94	0.06	0.98	0.06	0.96
sl10	$\delta s = 1, \delta t = 0.05$	0.47	0	0.64	0	0.86	0	0.93	0.01	0.51	0.85
	$\delta s = 4, \delta t = 0.5$	0.96	0.52	0.98	0.38	1	0.83	0.94	0.92	0.05	0.85
	$\delta s = 8, \delta t = 1$	0.97	0.54	0.95	0.42	0.97	0.83	0.91	0.92	0.05	0.85
	$\delta s = 16, \delta t = 2$	0.89	0.57	0.83	0.44	0.93	0.83	0.86	0.92	0.05	0.86
	$\delta s = 25, \delta t = 25$	0.12	0.86	0.08	0.67	0.1	0.87	0.11	0.95	0.06	0.96
svl	$\delta s = 1, \delta t = 0.05$	0.5	0	0.44	0	0.56	0	0.38	0.01	0.12	0.81
	$\delta s = 4, \delta t = 0.5$	0.93	0.54	0.91	0.35	0.99	1	0.64	0.97	0.01	0.82
	$\delta s = 8, \delta t = 1$	1	0.59	1	0.46	1	1	1	0.97	0	0.83
	$\delta s = 16, \delta t = 2$	1	0.61	0.88	0.49	0.9	1	0.89	0.99	0	0.84

Table C.2 Pattern Probabilities – continued from previous page

DataSet	Discretization	DD→VPR		C→VPR		C→P		C→CL		CL→F	
		TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP
	$\delta s = 25, \delta t = 25$	0.3	0.85	0.18	0.68	0.25	1	0.12	0.99	0.01	0.93
svl5	$\delta s = 1, \delta t = 0.05$	0.41	0	0.56	0	0.71	0	0.87	0.02	0.55	0.88
	$\delta s = 4, \delta t = 0.5$	0.93	0.48	0.96	0.44	0.99	0.92	0.79	0.96	0	0.89
	$\delta s = 8, \delta t = 1$	1	0.51	1	0.49	1	0.92	0.93	0.96	0	0.89
	$\delta s = 16, \delta t = 2$	0.75	0.54	0.78	0.51	0.95	0.92	0.85	0.96	0.05	0.91
	$\delta s = 25, \delta t = 25$	0.11	0.87	0.02	0.72	0.07	0.94	0.04	1	0.01	0.97
svl10	$\delta s = 1, \delta t = 0.05$	0.47	0	0.54	0	0.83	0	0.86	0.01	0.39	0.82
	$\delta s = 4, \delta t = 0.5$	0.96	0.49	0.98	0.39	1	0.85	0.87	0.97	0	0.82
	$\delta s = 8, \delta t = 1$	1	0.51	1	0.43	1	0.85	0.86	0.97	0	0.84
	$\delta s = 16, \delta t = 2$	0.96	0.53	0.89	0.44	0.94	0.85	0.9	0.98	0	0.85
	$\delta s = 25, \delta t = 25$	0.19	0.9	0.08	0.67	0.09	0.89	0.08	1	0.04	0.97
ls	$\delta s = 1, \delta t = 0.05$	0.88	0	0.81	0	0.89	0	0.96	0.02	0.44	0.86
	$\delta s = 4, \delta t = 0.5$	0.97	0.54	0.93	0.35	1	0.86	0.89	0.94	0	0.87
	$\delta s = 8, \delta t = 1$	0.96	0.62	0.8	0.46	0.93	0.87	0.66	0.94	0	0.88
	$\delta s = 16, \delta t = 2$	0.94	0.64	0.67	0.48	0.78	0.89	0.59	0.95	0.02	0.9
	$\delta s = 25, \delta t = 25$	0.13	0.96	0.05	0.73	0.1	0.97	0.06	0.99	0.04	0.98
ls5	$\delta s = 1, \delta t = 0.05$	0.54	0	0.81	0	0.95	0	0.95	0.01	0.65	0.82
	$\delta s = 4, \delta t = 0.5$	0.89	0.56	0.96	0.38	1	0.84	0.96	0.95	0	0.83
	$\delta s = 8, \delta t = 1$	0.91	0.59	0.74	0.43	0.8	0.84	0.75	0.95	0	0.85
	$\delta s = 16, \delta t = 2$	0.81	0.63	0.51	0.47	0.65	0.84	0.66	0.96	0.01	0.87
	$\delta s = 25, \delta t = 25$	0.07	0.95	0.02	0.7	0.04	0.88	0.05	0.99	0.01	0.97
ls10	$\delta s = 1, \delta t = 0.05$	0.41	0	0.76	0	0.93	0	0.99	0.01	0.72	0.84
	$\delta s = 4, \delta t = 0.5$	0.84	0.51	0.9	0.38	0.99	0.73	0.98	0.95	0	0.85
	$\delta s = 8, \delta t = 1$	1	0.55	0.72	0.43	0.77	0.73	0.89	0.95	0	0.86
	$\delta s = 16, \delta t = 2$	0.62	0.59	0.44	0.45	0.68	0.74	0.64	0.96	0	0.88

Table C.2 Pattern Probabilities – continued from previous page

DataSet	Discretization	DD→VPR		C→VPR		C→P		C→CL		CL→F	
		TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP
lm	$\delta s = 25, \delta t = 25$	0.07	0.93	0.01	0.69	0.03	0.82	0.05	0.99	0.01	0.98
	$\delta s = 1, \delta t = 0.05$	0.82	0	0.79	0	0.8	0	0.8	0.01	0.22	0.82
	$\delta s = 4, \delta t = 0.5$	0.94	0.54	0.9	0.38	0.99	0.97	0.84	0.91	0	0.83
	$\delta s = 8, \delta t = 1$	1	0.61	0.95	0.49	0.93	0.97	1	0.91	0.12	0.85
	$\delta s = 16, \delta t = 2$	0.89	0.64	0.63	0.51	0.7	0.98	0.65	0.92	0.02	0.86
	$\delta s = 25, \delta t = 25$	0.12	0.94	0.03	0.72	0.14	0.99	0.07	0.98	0.02	0.97
lm5	$\delta s = 1, \delta t = 0.05$	0.47	0	0.74	0	0.91	0	0.94	0.01	0.55	0.84
	$\delta s = 4, \delta t = 0.5$	0.93	0.56	0.94	0.4	0.99	0.88	0.91	0.94	0	0.85
	$\delta s = 8, \delta t = 1$	0.8	0.59	0.82	0.44	1	0.88	0.97	0.95	0	0.86
	$\delta s = 16, \delta t = 2$	0.7	0.62	0.52	0.47	0.7	0.88	0.76	0.96	0	0.88
	$\delta s = 25, \delta t = 25$	0.05	0.95	0.01	0.74	0.03	0.91	0.04	0.99	0.02	0.99
lm10	$\delta s = 1, \delta t = 0.05$	0.52	0	0.77	0	0.89	0	0.96	0.01	0.7	0.85
	$\delta s = 4, \delta t = 0.5$	0.9	0.53	0.96	0.37	1	0.78	0.97	0.97	0	0.86
	$\delta s = 8, \delta t = 1$	1	0.55	0.85	0.42	0.95	0.78	1	0.97	0	0.87
	$\delta s = 16, \delta t = 2$	0.75	0.59	0.59	0.45	0.76	0.78	0.7	0.98	0	0.89
	$\delta s = 25, \delta t = 25$	0.06	0.92	0.01	0.69	0.03	0.85	0.05	1	0.01	0.98
ll	$\delta s = 1, \delta t = 0.05$	0.75	0	0.69	0	0.73	0	0.8	0.01	0.25	0.85
	$\delta s = 4, \delta t = 0.5$	0.95	0.57	0.91	0.35	1	1	0.78	0.98	0	0.85
	$\delta s = 8, \delta t = 1$	1	0.63	1	0.46	1	1	1	0.98	0	0.86
	$\delta s = 16, \delta t = 2$	0.83	0.67	0.7	0.49	0.84	1	0.83	0.98	0	0.88
	$\delta s = 25, \delta t = 25$	0.13	0.95	0.04	0.72	0.13	1	0.06	1	0.01	0.96
ll5	$\delta s = 1, \delta t = 0.05$	0.53	0	0.66	0	0.77	0	0.84	0	0.55	0.83
	$\delta s = 4, \delta t = 0.5$	0.94	0.57	0.94	0.4	1	0.91	0.88	0.92	0.05	0.84
	$\delta s = 8, \delta t = 1$	0.85	0.6	0.9	0.45	0.97	0.91	0.96	0.94	0	0.86
	$\delta s = 16, \delta t = 2$	0.81	0.63	0.62	0.47	0.79	0.91	0.75	0.95	0.06	0.87

Table C.2 Pattern Probabilities – continued from previous page

DataSet	Discretization	DD→VPR		C→VPR		C→P		C→CL		CL→F	
		TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP	TPEX	CSTP
ll10	$\delta s = 25, \delta t = 25$	0.07	0.96	0.01	0.72	0.06	0.94	0.05	0.99	0.02	0.96
	$\delta s = 1, \delta t = 0.05$	0.45	0	0.53	0	0.82	0	0.76	0.02	0.45	0.86
	$\delta s = 4, \delta t = 0.5$	0.94	0.55	0.95	0.37	1	0.82	0.94	0.94	0	0.86
	$\delta s = 8, \delta t = 1$	0.82	0.58	0.91	0.4	1	0.82	1	0.94	0	0.87
	$\delta s = 16, \delta t = 2$	0.71	0.62	0.57	0.43	0.69	0.83	0.74	0.95	0	0.9
	$\delta s = 25, \delta t = 25$	0.05	0.93	0.01	0.69	0.04	0.88	0.06	0.99	0.02	0.98
lvl	$\delta s = 1, \delta t = 0.05$	0.67	0	0.58	0	0.33	0	0.67	0.01	0.88	0.85
	$\delta s = 4, \delta t = 0.5$	0.93	0.57	0.87	0.37	0.99	1	0.61	0.95	0	0.85
	$\delta s = 8, \delta t = 1$	1	0.63	0.97	0.46	1	1	0.93	0.96	0	0.87
	$\delta s = 16, \delta t = 2$	0.85	0.65	0.64	0.49	0.75	1	0.7	0.97	0	0.89
	$\delta s = 25, \delta t = 25$	0.12	0.95	0.05	0.73	0.16	1	0.08	0.99	0.01	0.97
lvl5	$\delta s = 1, \delta t = 0.05$	0.61	0	0.51	0	0.48	0	0.44	0.01	0.51	0.81
	$\delta s = 4, \delta t = 0.5$	0.94	0.55	0.93	0.43	0.98	0.92	0.72	0.96	0	0.81
	$\delta s = 8, \delta t = 1$	1	0.59	0.88	0.47	0.76	0.92	0.68	0.97	0	0.83
	$\delta s = 16, \delta t = 2$	0.83	0.62	0.62	0.5	0.69	0.92	0.6	0.97	0	0.84
	$\delta s = 25, \delta t = 25$	0.09	0.95	0	0.76	0.05	0.95	0.05	1	0	0.98
lvl10	$\delta s = 1, \delta t = 0.05$	0.35	0	0.42	0	0.58	0	0.65	0.01	0.36	0.83
	$\delta s = 4, \delta t = 0.5$	0.94	0.54	0.94	0.4	0.99	0.85	0.87	0.95	0	0.84
	$\delta s = 8, \delta t = 1$	0.8	0.57	0.76	0.44	0.93	0.85	0.9	0.95	0	0.86
	$\delta s = 16, \delta t = 2$	0.71	0.61	0.48	0.47	0.63	0.85	0.68	0.95	0	0.89
	$\delta s = 25, \delta t = 25$	0.06	0.96	0.01	0.72	0.04	0.9	0.07	0.99	0.01	0.98

Table C.3: Range of Pattern Ranks

Discretization	$DD \rightarrow VPR$ TPEX		$DD \rightarrow VPR$ CSTP		$C \rightarrow VPR$ TPEX		$C \rightarrow VPR$ CSTP		$C \rightarrow P$ TPEX		$C \rightarrow P$ CSTP		$C \rightarrow CL$ TPEX		$C \rightarrow CL$ CSTP		$CL \rightarrow F$ TPEX		$CL \rightarrow F$ CSTP	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
$\delta s = 1, \delta t = 0.05$	1	8	-	-	3	5	-	-	1	6	-	-	1	7	12	20	1	15	1	2
$\delta s = 4, \delta t = 0.5$	1	4	11	12	1	4	13	17	1	3	1	6	2	4	1	2	21	-	2	7
$\delta s = 8, \delta t = 1$	1	5	11	12	2	6	13	15	1	4	1	6	1	6	1	2	11	-	2	7
$\delta s = 16, \delta t = 2$	1	4	11	12	3	5	13	16	1	4	1	7	1	6	1	2	14	-	2	7
$\delta s = 25, \delta t = 25$	1	6	6	12	3	17	15	27	1	7	1	16	1	8	1	4	7	19	2	8

Table C.4: Range of Pattern Probabilities

Discretization	$DD \rightarrow VPR$ TPEX		$DD \rightarrow VPR$ CSTP		$C \rightarrow VPR$ TPEX		$C \rightarrow VPR$ CSTP		$C \rightarrow P$ TPEX		$C \rightarrow P$ CSTP		$C \rightarrow CL$ TPEX		$C \rightarrow CL$ CSTP		$CL \rightarrow F$ TPEX		$CL \rightarrow F$ CSTP	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
$\delta s = 1, \delta t = 0.05$	0.35	0.88	0	0	0.42	0.81	0	0	0.33	0.95	0	0	0.38	0.99	0	0.02	0.12	0.88	0.81	0.88
$\delta s = 4, \delta t = 0.5$	0.84	0.99	0.48	0.57	0.87	0.98	0.32	0.44	0.96	1	0.71	1	0.61	0.98	0.91	0.98	0	0.05	0.81	0.89
$\delta s = 8, \delta t = 1$	0.8	1	0.51	0.63	0.72	1	0.37	0.49	0.76	1	0.71	1	0.66	1	0.91	0.98	0	0.12	0.83	0.89
$\delta s = 16, \delta t = 2$	0.62	1	0.53	0.67	0.44	0.9	0.37	0.51	0.63	0.96	0.71	1	0.59	0.9	0.92	0.99	0	0.06	0.84	0.91
$\delta s = 25, \delta t = 25$	0.05	0.3	0.85	0.96	0	0.2	0.64	0.76	0.03	0.27	0.77	1	0.04	0.23	0.95	1	0	0.06	0.93	0.99

APPENDIX D

PRESENTED WORK

Figure D.1: Poster presented at Spatial Statistics Conference, 2013.

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: BAKIR BATU, BERNA
Nationality: Turkish (TC)
Date and Place of Birth: 30.08.1978, Ankara, Turkey
Marital Status: Married
Phone: +90(505)5428444
email: bernabakir@gmail.com

EDUCATION

Degree	Institution	Year of Graduation
M.S.	METU Informatics Institute	2007
B.S.	Ankara University, Statistics	2002
High School	Hasanoğlan Atatürk Anatolian Teacher High School	1995

PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
2003 - present	METU Informatics Institute	Research Assistant

PUBLICATIONS

International Conference Publications

B. Bakır, S. Özkan, R. Köseler, T. Taşkaya Temizel, D. İncebacak and M. Kaya. The Attitudes of students with diverse backgrounds on computer and information literacy subjects: Evidence from a first year course. *In Proceedings of the 39th Annual Frontiers in Education (FIE) Conference*, 2009.

R. Köseler, T. Taşkaya Temizel, B. Bakır, D. İncebacak, M. Kaya and S. Özkan. Work in progress: Iterative curriculum development for an interdisciplinary online-taught IT course. In *Proceedings of the 39th Annual Frontiers in Education (FIE) Conference*, 2009.

B. Bakır, İ. Batmaz, F.A. Güntürkün, A.İ. İpekçi, G. Köksal, N.E. Özdemirel. De-

fect Cause Modeling with Decision Tree and Regression Analysis. In *Proceedings of XVII. International Conference on Computer and Information Science and Engineering*, Cairo, Egypt, 266-269, 2006.

International Conference Presentations

B. Bakır Batu, T. Taşkaya Temizel, H.Ş. Düzgün. Spatio-temporal self-exciting process modelling with k-function weighted MISD algorithm. *2nd Conference on Spatial Statistics*, 2013.

B. Bakır, D. Ayhan, B. Yenidünya, G. Köksal. A Re-sampling Approach and Its Applications for MTS Classification Based on Imbalanced Data. *23rd European Conference on Operational Research*, 39-40, Bonn, 2009.

G. Köksal, Z. Anaklı, İ. Batmaz, F. Yerlikaya Özkurt, M.C. Testik, E. Kartal, S. Kayalıgil, B. Bakır, E. Karasakal. Comparison of Data Mining Algorithms for Classification and Prediction in Quality Improvement. *23rd European Conference on Operational Research*, 39, Bonn, 2009.

A.İ. Ipekci, ,B. Bakir, ,İ. Batmaz, and M.C. Testik. Optimization of Casting Process Using Artificial Neural Networks. *International Conference on Multivariate Statistical Modeling and High Dimensional Data Mining*, HDM'08, 128, Kayseri, Türkiye, 2008.

National Conference Presentations

B. Bakır and G. Köksal. Ürün Tasarımı ile İlgili Müşteri Sesinin Modellenmesi: Bayes İnanç Ağları. *28. Ulusal Yöneylem Araştırması ve Endüstri Mühendisliği Konferansı*, İstanbul, 2008.

Z. Anaklı, B. Bakır, A.İ. İpekçi, G. Köksal and S. Kayalıgil. Kalite Geliştirmede Veri Madenciliği: Sınıflandırmada Performans Ölçüleri Etkileşim Değerlendirmesi. *28. Ulusal Yöneylem Araştırması ve Endüstri Mühendisliği Konferansı*, İstanbul, 2008.