UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

DEALING WITH INCOMPLETE AND UNCERTAIN CONTEXT DATA IN GEOGRAPHIC INFORMATION

TESIS PARA OPTAR AL GRADO DE DOCTOR EN CIENCIAS, MENCIÓN COMPUTACIÓN

JONATHAN FREZ ZACHARY

PROFESOR GUIA:
NELSON BALOIAN TATARYAN

MIEMBROS DE LA COMISIÓN:
SERGIO OCHOA DELORENZI
ANDREA RODRIGUEZ TASTEST
FRADA BURSTEIN

SANTIAGO DE CHILE
2021

Resumen

Un problema recurrente en la toma de decisiones espaciales (TDE) puede ser descrito de la siguiente manera: encontrar una zona apta para hacer “algo”. Según la literatura, los pasos para encontrar una zona apta son básicamente los mismos que en temáticas de toma de decisiones en general. Sin embargo, en los sistemas de apoyo a la toma de decisiones espaciales (SADE) hay tres pasos altamente complejos que seguir: i) Combinar datos desde múltiples fuentes, ii) Generar escenarios de idoneidad utilizando diferentes hipótesis de una manera flexible y sistemática; iii) Comparar diferentes escenarios.

Frecuentemente en los SADE es necesario tener que lidiar con problemas pocamente estructurados, en donde los objetivos pueden no estar completamente claros y/o la información requerida para resolver el problema de una manera certera puede ser insuficiente. Incluso, los problemas poco estructurados pueden tener múltiples alternativas de solución y múltiples criterios de evaluación. Entonces los SADE deben tener un grado de flexibilidad y herramientas que permitan realizar análisis bajo estas condiciones.

Para afrontar este problema, esta tesis propone un framework para lidiar con problemas espaciales no estructurados, esto soportando los pasos complejos de TDE de una manera flexible pero sistemática. Este trabajo desarrolla y valida herramientas tanto computacionales como teóricas para soportar un SDSS que opera en escenarios espaciales complejos. Específicamente este framework contiene:

1. Un modelo para combinar fuentes de información y generar escenarios espaciales, integrando con la Teoría de Dempster-Shafer.
2. Un lenguaje de generación de escenarios (SGL) para poder describir y modelar los requerimientos de toma de decisiones a ser aplicados sobre un área de análisis determinada. El lenguaje permite realizar un proceso de toma de decisión continua, permitiendo a los usuarios intentar y evaluar diferentes escenarios.
3. Una metodología para soportar la colaboración entre diferentes actores (cada uno con su propia experiencia y visión) en un proceso de toma de decisiones. Esta metodología se implementa mediante el framework permitiendo el análisis, combinación y comparación de múltiples escenarios espaciales (generados con SGL).

El framewok también incluye operadores espaciales y de comparación de escenarios. Estos son útiles para especificar relaciones espaciales o filtrar datos utilizando SGL. Este framework fue probado en tres escenarios: un sistema preciso de predicción de crimen basado en conocimiento experto, un estimador de demanda de transporte público, y una herramienta de planificación colaborativa. El sistema de predicción de crimen está enfocado en probar el uso de la teoría de Dempster-Shafer para la generación de escenarios de toma de decisiones de na manera flexible y sistemática, los resultados en términos cuantitativos fueron similares o mejores a otras técnicas. El estimador de demanda de transporte permite realizar evaluaciones preliminares de rutas sin invertir recursos en las tareas necesarias para una evaluación comprensiva. Finalmente, el sistema de planificación colaborativa probó si el framework es apropiado para analizar colaborativamente. Permitiendo a múltiples actores proponer hipótesis y utilizar fuentes de información, permitiendo combinar sus propuestas individuales con el fin de obtener conclusiones en conjunto.

Abstract

A recurrent problem in Spatial Decision-Making (SDM) can be stated in the following way: to find a suitable area to do "something". According to the literature, the steps to find a suitable area are basically the same as in a general decision-making. However, in Spatial Decision Support Systems (SDSS), there are three highly complex steps to follow: i) Combining data from different sources; ii) Generating suitability scenarios, using different hypotheses, in a flexible and systematic manner; iii) Comparing different scenarios.

Frequently the SDSS have to deal with ill-structured problems, where their goals might not be totally clear and/or there is insufficient information to solve them in a certain optimal way. Furthermore, ill-structured problems may have multiple solutions and multiple criteria for evaluating their outcome. Therefore, SDSS must have a degree of flexibility and tools that allow to make an analysis under these conditions.

As a way to address that problem, this thesis proposes a framework that allows to deal with spatial ill-structured problems by supporting the complex steps in SDM in a flexible but systematic manner. This work develops and validates computational and theoretical tools to support SDSS that operate in complex spatial scenarios. Specifically, this framework contains:

4. A model to combine different information sources and generate spatial scenarios, integrating conceptual decision-making frameworks as the Dempster-Shafer Theory. The Dempster-Shafer Theory is a general framework for reasoning when facing uncertainty. The benefits of using this theory are, its applicability to different contexts, the flexibility to combine models based on probabilities and the fact that it allows to specify a degree of ignorance about a certain subject.
5. A Scenario Generation Language (SGL) to describe and model the decision-making requirements to be applied to the area under analysis. The language builds a continuous decision-making process, allowing users to try and evaluate a scenario.
6. A methodology to support the collaboration among different stakeholders (each of them with their own background and vision) in decision-making processes. Its implementation within the framework allows the analysis, combination and comparison of spatial scenarios (generated using SGL).

The framework also includes scenario-comparison operators and spatial operators. They are useful to specify spatial relationships or filter data by using SGL. This framework was tested in three different scenarios: an accurate crime prediction system based on expert knowledge, a public transportation demand estimator, and a collaborative planning tool. The crime prediction system is focused on test the integration of the Dempster-Shafer Theory to generate spatial decision-making scenarios in a flexible and systematic manner, the results in terms of quantitative prediction metrics were similar or better in some cases than existing ones. The public transportation demand estimator allowed to make preliminary evaluations of a route without investing time and resources to perform all the necessary tasks to gather the necessary information for a comprehensive evaluation. Finally, the collaborative planning tool tested if this framework is appropriate for analyzing collaborative scenarios. The tool allowed multiple stakeholders to apply hypotheses and use uncertain data over geo-referenced information and combine their individual work in order to draw conclusions.

Acknowledgments

Quisiera comenzar agradeciendo a mi mejor amiga, compañera de vida desde la juventud y esposa. Juntos nos hemos propuesto muchas metas en la vida, apoyándonos mutuamente para lograr cumplirlas, incluyendo esta.

A mis padres quienes me incentivaron el seguir el camino de la ingeniería y en particular la computación desde niño, consiguiendo y ofreciéndome a oportunidad de estudiar formalmente a temprana edad dándome una ventaja frente a lo que vendría más adelante.

Por supuesto quisiera agradecer a mi profesor guía por todo el apoyo recibido y las oportunidades ofrecidas para poder llevar acabo este proceso.

También quisiera incluir a Sergio Mujica, quien me mostró una visión de la ingeniería mucho más amplia y relevante respecto a la aplicación de las ciencias y la importancia del saber en el hacer. En la misma línea a Jorge Elliott, de quien tuve la oportunidad de aprender de su experiencia y sus certeros consejos, con una mirada equilibra entre lo personal y profesional permitiéndome no solo mejorar en lo profesional, si no que a su vez pasar por este proceso de una mejor manera.

A Luciano Ahumada, gran amigo, con quien hemos trabajado juntos por varios años, cuya visión y experiencia como investigador me ayudó mejorar mi trabajo y orientar adecuadamente mis esfuerzos.

Contents

CHAPTER 1 INTRODUCTION 1

1.1 SPATIAL DECISION SUPPORT SYSTEMS 2

1.2 PROBLEM TO ADDRESS 4

1.3 RESEARCH OBJECTIVES 5

1.4 STRUCTURE OF THE THESIS DOCUMENT 6

CHAPTER 2 RELATED DOMAINS REVIEW 8

2.1 SPATIAL OBJECT REPRESENTATION 8

2.2 SPATIAL DECISION SUPPORT SYSTEMS 10

2.2.1 Decision-making models 10

2.2.2 Spatial decision support systems 12

2.3 THE DESIGN PROBLEM OF SDSS 14

CHAPTER 3 GEOGRAPHICAL DATA REPRESENTATION 16

3.1 GIS DATABASES 16

3.1.1 Vector data model 17

3.1.2 Raster data model 18

3.1.3 Comparison between vector and raster data models 19

3.1.4 Coordinate systems and Open Geospatial standards 19

3.2 GEOGRAPHICAL DATABASES OPERATIONS 20

3.3 SIMPLE FEATURE ACCESS 22

3.4 SPATIAL INDEXING 24

3.5 DISCUSSION 26

CHAPTER 4 DEMPSTER-SHAFER THEORY AND BELIEF FUNCTIONS 27

4.1 FORMAL DESCRIPTION 29

4.2 BELIEF FUNCTIONS AND COMBINATION RULES 30

4.3 DISCOUNT AND COMBINE METHOD 31

4.4 YAGER MODIFIED DEMPSTER RULE 32

CHAPTER 5 ADAPTING DEMPSTER-SHAFER THEORY TO SPATIAL CONTEXTS 33

5.1 MASS IN SPACE 35

5.2 MASS FUNCTION BY INTERVAL 37

5.3 MASS FUNCTION ALONG TIME ... 38
5.4 BELIEF AND PLAUSIBILITY DEFINITION BY SPACE, INTERVAL, AND TIME ... 41
5.5 DISCUSSION ... 42

CHAPTER 6 SDSS SCENARIO GENERATION MODEL ... 44

6.1 SUITABLE OBJECTS ... 44
6.2 DATA OBJECTS ... 44
6.3 GEOOBJECT ... 45
6.4 GEOOBJECTS MAP AND SUITABILITY MAP GENERATION ... 46
6.5 THE ALGORITHM ... 47
6.6 EVALUATION EXAMPLE ... 49
6.7 DISCUSSION ... 52

CHAPTER 7 SCENARIO GENERATION LANGUAGE ... 53

7.1 GRAMMAR ... 55
7.2 ANALYSIS DEFINITION ... 55
7.2.1 Hypothesis definitions ... 55
7.2.2 Model constraints ... 56
7.2.3 Saving ... 57
7.2.4 Operator syntax ... 57
7.3 SGL EXAMPLES ... 58
7.3.1 Example 1: People possible concentration points ... 58
7.3.2 Example 2: Burglary risk ... 59
7.3.3 Example 3: Burglary risk in a street ... 59
7.3.4 Example 4: Path evaluation and saving ... 59
7.4 DISCUSSION ... 60

CHAPTER 8 URBAN CRIME PREDICTION AND POLICE PATROLLING OPTIMIZATION ... 61

8.1 DEMPSTER-SHAFER RISK TERRAIN MODELING ... 63
8.2 FORECASTING EXPERIMENTS ... 68
8.2.1 Medium scale scenario ... 68
8.2.2 Medium scale scenario ... 69
8.2.3 Large scale scenario ... 71
8.3 DISCUSSION ... 74

CHAPTER 9 TRANSPORTATION NETWORK PLANNING EXAMPLE......... 75

9.1 TRANSPORTATION MODELING 76

9.2 DETERMINING AN OD ROUTE 77

9.3 BELIEF ROUTES IN REAL WORLD 80

9.4 AN APPLICATION FOR DEVELOPING BELIEF ROUTES 86

9.4.1 Discussion 87

CHAPTER 10 MULTIPLE STAKEHOLDERS, SCENARIO COMPARISON AND COMBINATION........ 89

10.1 MERGE AND COMPARE OPERATORS........ 91

10.1.1 Sum operator........ 92

10.1.2 Average operator 92

10.1.3 Subtraction operator........ 93

10.2 ORDERED WEIGHT AVERAGE (OWA) OPERATOR 93

10.2.1 OWA-DESC........ 93

10.2.2 OWA-ASC 94

10.2.3 Weighted OWA operator........ 94

10.2.4 Induced OWA operator 95

10.3 OPERATORS APPLIED TO THE TRANSPORTATION NETWORK PROBLEM........ 95

10.4 TOOL DESIGN 95

10.5 DISCUSSION........ 101

CHAPTER 11 CONCLUSIONS AND FUTURE WORK........ 103

11.1 SUMMARY AND CONCLUSIONS OF THE CONDUCTED RESEARCH 103

11.2 FUTURE WORK........ 107

11.2.1 Automatic SGL statements generation from recurrent characteristics or patterns........ 107

11.2.2 Non-spatial belief relationships........ 107

11.2.3 Integration with RDF and GEO-RDF........ 107

BIBLIOGRAPHY 109

APPENDIX A 120

Figures

Figure 1. Decision-Making Process. 1

Figure 2. Structure of the Thesis Document.. 6

Figure 3 Representation of a Fuzzy Region and its Fuzzy Set Membership Function, Source: (Zhan & Lin 2003).. 9

Figure 4. Conceptual views of decision-making according to three different authors (Crossland, Wynne, & Perkins, 1995). 11

Figure 5. Decision-making models according to various perspectives (Crossland, Wynne, & Perkins, 1995). 12

Figure 6. Suitability Map Example for land spreading of pistachio solid waste, source: http://www.agrostrat.gr/en/suitabilitymaps 14

Figure 7. Stages of reality digital representation process. 17

Figure 8. Representation examples. 17

Figure 9. Vector model geometry types. 18

Figure 10. Geoid and ellipsoid. 20

Figure 11. Representation examples. 25

Figure 12 Tree of Buckets. 25

Figure 13 R-tree example, M=3. 25

Figure 14 . Quadtree example. 26

Figure 15 Basic evaluation model, H1 to H6 are hypotheses. 33

Figure 16 Resulting belief of each cell is calculated combining the mass of all hypotheses that apply to that cell. 34

Figure 17 Space Mass functions, F(d) examples using u = 50. 35

Figure 18. Normal CDF example. 36

Figure 19 ms(d) example using $\mu = 50$ $\sigma = 10$. 37

Figure 20 ms(d) example using $\mu = 200$ $\sigma = 50$. 37

Figure 21. The hierarchal organization of time scales (Klyne, 2002). 40

Figure 22 Time mass assignment metaphor for a single scale. 40

Figure 23 Mass of each time segment for a single scale. 41

Figure 24 Mass distribution in space. 42

Figure 25. Structure of a GeoObject 46

Figure 26. Data objects location example. 46

Figure 27. Propagated belief vector example representing the influence of Object of class A and class B on the blank cell. 47

Figure 28. Polygons of different types of areas. 49

Figure 29. Certainty of each source representation in % scale. Red: 0% - Green: 100%. 50

Figure 30. Plausibility of each source representation in % scale. Red: 0% - Green: 100%. 50

Figure 31. Plausibility of each source representation in % scale. Brown: 0%, Red: near 0%, Green 100%. 51

Figure 32. Plausibility of each source representation. Max height equals to 100% of belief. 51

Figure 33. Comparison between S-T GAM and spatial GLM (Wang, 2012) 63

Figure 34. Street Robbery Analysis Triangle (ONK, 2010). 64

Figure 35. Top left figure shows crime related to assaults, robberies and drugs in Texas (USA) between October and December of 2017. Top right, bottom left and bottom right figures show the result of the crime prediction using our method for a weekday of January. 69

Figure 36. Spatial Dempster- Shafer method performance for crime prediction. 71

Figure 37. Prediction performance in percentage in different areas and threshold areas. 72

Figure 38. Prediction performance in different areas and thresholds. 73

Figure 39. Route 1. 78

Figure 40. Route 2. 78

Figure 41. Belief route 1 80

Figure 42. Belief route 2. 80

Figure 43. Congestion route 1. 80

Figure 44. Congestion route 2. 80

Figure 45. Real bus route, 12km, and 36min trip. 81

Figure 46. Proposed route, 17km, and 48min trip. 81

Figure 47. Zone types of the analyzed area. 81

Figure 48. Transportation demand in Residential area #1 82

Figure 49. Transportation demand in Industrial area. 82

Figure 50. Transportation demand in Residential area #2. 82

Figure 51. Real demand vs predicted demand for residential area #1 83

Figure 52. Real demand vs predicted demand for industrial area. 83

Figure 53. Real demand vs predicted demand for residential area #2. 84

Figure 54. Original route with real demand vs new proposed route with its predicted demand for residential area #1. 84

Figure 55. Proposed route on the right, with its predicted demand for industrial area and residential areas vs. real demand of the original route on the left. 85

Figure 56. Real demand. 86

Figure 57. Predicted demand. 86

Figure 58. Evaluation of an OD using the developed application. 86

Figure 59. Left Iquique from the sea, Right: Evacuation area. 89

Figure 60. Playa Brava. Left: Population belief. Right: Buildings. 90

Figure 61. Hierarchical combination process. 91

Figure 62. Hypothesis Dimension and Time Dimension: here we see that for the first time 91

Figure 63. Initial Scenarios. 92

Figure 64. Sum of the three scenarios. 92

Figure 65. Average of scenarios. 93

Figure 66. Subtraction of scenarios. 93

Figure 67. OWA-DESC applied to 3 scenarios. 94

Figure 68. OWA-ASC applied to 3 scenarios. 94

Figure 69. Interpretation of the four combinations of OWA values. 95

Figure 70. Screenshots of the application showing the available projects for user Nelson. 96

Figure 71. Screenshot showing the user Nelson exploring the maps that have been generated inside the “Evacuation planning Iquique” project. 97

Figure 72. Screenshot of the application when generating a new map by defining hypotheses and their weights. 98

Figure 73. Screenshot of the application when Nelson combines the scenarios generated for people in amenities and people at shops with the sum operator. 99

Figure 74. Screenshot showing Jonathan two combined two maps using the OWA-DESC operator. 99

Figure 75. Screenshot showing Alvaro combining the two maps generated by him using the OWA-ASC operator. 100

Figure 76. Screenshot showing Alvaro reviewing a scenario. 101

Chapter 1

Introduction

Decision Support Systems are defined as interactive computer-based systems that help decision makers in the use of data and models to solve unstructured problems (Keen, 1987). A simplified model for the decision-making process (DM) includes the following stages (Power, 2015) (Bahl, 1984): 1) Identifying the problem or opportunity, 2) identifying and modeling the objective(s) of the decision, 3) collecting, generating and combining data to generate alternative scenarios 4) evaluating alternatives according to established objectives, 5) choosing an alternative and conducting a sensitive analysis. If a decision maker estimates that there is enough information, the process ends with a final decision, otherwise, the flow goes back to the identifying objectives or it generates alternatives stages (see Figure 1). Like in other areas of computer science, the boundaries for defining what falls under the category of Decision Support Systems (DSS) are rather diffuse. However, most of the authors who have tried to define what a DSS is, agree that one of the most important characteristics is that the human judgment remains in the decision-making process cycle as a key actor. Human judgment is necessary for generating alternatives, re-defining, and remodeling objectives. Since this is a task involving creativity, it cannot be mechanized (Bahl, 1984). Computers can help people gathering data, generating various decision alternatives, evaluating their outcome according to the decisions goals, visualizing these results and communicating them to others.

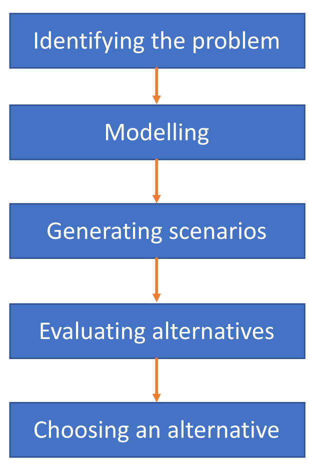


Figure 1. Decision-Making Process.

DSS systems also use various models and analysis techniques that are intended for use by non-computing experts (Shim, 2002). Hence, a DSS must be very interactive, flexible and adaptable to support different solution approaches.   Spatial decision-making problems have some particular characteristics.

In order to explain this, I am going to use a classical problem in agriculture: “how to evaluate the land capability for growing plants”. Land suitability analysis is necessarily involved in agriculture for mass food production, and selecting an area involves analyzing the soil, topographic, climatic, economic and land use characteristics. Specifically, land comprises the physical environment, including climate, relief, soil, hydrology and vegetation. Furthermore, all these attributes are spatially related, and the relationships are not always clear enough. It can also include the results of past and present human activity. A typical study in natural resource surveys is land-mapping units. In some cases, a single land mapping unit may include two or more distinct types of land, with different suitability and capabilities (Natural Resources Management and Environment Department, 1976).

1.1 Spatial decision support systems

Spatial Decision Support Systems (SDSS) aims to help users solve complex geographic problems by creating scenarios. Heugens and van Oosdterhout (Heugens, Oosdterhout ,2001) define scenarios as "Stories about the future". More rather than trying to predict the future, scenarios are possible descriptions of what the future will look like. Scenario development (or "analysis" or "planning" of scenarios) is a systematic method for thinking creatively about futures, dealing with complexity and uncertainty, identify strategies to prepare for a range of possible scenarios results (Madlenner et al., 2007). Scenarios could identify desirable futures in which people want to work.

The scenarios may also include undesirable futures that people may want to avoid. Rather, trying to reduce uncertainty through increasingly accurate predictions, scenarios can be flexible by incorporating potential feedback and surprises, to investigate and prepare uncertainties that are fundamental to complex systems (Walz et al., 2007).

Scenario analysis is the process of evaluating possible future events through the consideration of possible alternatives, however, not equally probable, states of the world (scenarios) according to (Shine, 2009) a scenario is:

"A scenario is a coherent, internally consistent and possible description of a possible future state of the world. This is not a prediction: preferably each scenario is an alternative picture of how the future can open up”.

According to this definition, scenarios are not forecasting or predictions. Instead, they provide a dynamic view of the future by exploring various trajectories of changes that lead to a broadening of the range of possible future alternatives. Scenarios are typically used in the context of planning for large horizons or short-term decisions that have long-term consequences, for greater perspectives and the illumination of key problems that might otherwise be lost. Rather than relying on predictions, scenarios allow a creative and flexible approach to preparing for an uncertain future (Schwartz, 1991).

The information required to build a scenario usually belongs to several fields, and it can be huge (Densham, 1991). Furthermore, specialized algorithms are needed to calculate the alternatives and evaluate them. SDSS are currently used in a large variety of tasks, for example, (public transport) route planning, locating safe areas during natural disasters, finding a good location for opening a particular store, analyzing transportation traffic, logistics or comparing the occurrence of a disease in various places.

SDSS must be able to model the environment and to evaluate the impact of changes. In addition, spatial information is inherently fuzzy and uncertain (Morris & Jankowski, 2001), meaning that fuzzy analysis techniques are needed (Robinson, 2010) (Laurent, 2010).

From the available literature on Geographic Information Systems (GISs) used to support DM, I realize that there are many modeling tools available that can generate a certain scenario for a geographical area, applying certain evaluation functions and showing the output, like for vegetation winter survival, wind farm's locations (Baban & Parry, 2001) or forest production estimation (Ekanayake & Dayawansa, 2003). However, the majority of the existing GISs are not explicitly designed to support a DM cycle. Therefore, the process of creating various alternative scenarios according to different criteria and comparing them using the available systems is, in most cases, a time-consuming and challenging task. In order to implement a DM process, we need to abstract the modeling part from the DM cycle and allow the decision maker to generate multiple scenarios according to different problem conditions and compare the outcomes in a simple and systematic way.

Many types of decision problems involving spatial issues can be solved with the help of a GIS. Most recurrent problems in this area can be stated in the following generic way: find a suitable area to "do" something. As seen in (Ghayoumian, Ghermezcheshme, Feiznia, & Noroozi, 2005), authors explain how to find specific locations for constructing artificial water recharge aquifers using floods. In this example, decision-makers must be experts in recharge aquifers, but they will need historical information and spatial data to design a formula that reflects the correct criterions for selecting the suitable area(s) (Chountas, Rogova, & Atanassov, 2010) and this formula is used to build a suitability map using a GIS. This visualization typically shows the suitability level on each point of the map that satisfies the requirements. However, in ill-structured problems, the suitability criterions are not completely clear. Furthermore, an ill-structured problem has no clear path to be solved and may have many different solutions. Multiple decision makers will also tend to define different goals, according to their knowledge (Densham,1991).

A decision-making process supported by GIS typically starts with two inputs: data and expert knowledge. Models are built using an expert's knowledge, and alternative scenarios are constructed using different data input. These scenarios can be compared because they are based on the same model, although each model is based on different expert knowledge. Nevertheless, the knowledge can change during the evaluation process, for instance by including or removing a person from the experts' team. A change in the knowledge of the experts' team is challenging to represent in the model because it leads to changes in the databases used (Shim, 2002).

In the DSS area, mathematical tools have been designed to include expert knowledge and manage incomplete and uncertain information. Over the last 20 years, there have been studies using Belief Functions in different spatial decision-making problems (Ghayoumian, Ghermezcheshme, Feiznia, & Noroozi, 2005) (Binaghi, Luzi, Madella, Pergalani, & Rampini, 1998) (Park, 2011) (Tangestani & Moore, 2002) with good results. However, in these experiments, the calculation method and results are based on a specific spatial problem. As a result, functions used to evaluate the suitability of a place differ from each other and thus are hard to replicate or adapt to other spatial problems.

Spatial Decision Support systems are especially relevant to a vast number of scientific, economic and humanistic areas. Among others, the most referenced in the literature are (Burrough, McDonnell, Burrough, & McDonnell, 1998) (Harley & Britain, 1975) (FAO & FOODS, 2004):

- Socioeconomics: urban planning, industrial planning, agricultural land use, housing, educations, natural resources, and many new smart city applications.

- Environmental: forestry, fire and epidemic control, floods and earthquake predictions, pollution, and also smart city applications.
- Management: organization logistics, electricity and telecommunication network planning, real-time vehicle tracking, and other public services planning like health services, security, fire protection, and also smart city applications.

According to some authors (Malafant et al., 1998), a GIS is always a DSS because they are used to support some stage of a decision-making process. According to (Morris & Jankowski, 2001), GIS offers appropriate techniques for data management, information extraction, routine manipulation, and visualization. However, they do not have the necessary analytical capabilities to manage a decision-making process. Furthermore, in (Morris, Jankowski, Bourgeois, & Petry, 2010) authors claim that, at the moment of publication (2010), the existing SDSS tools do not provide the needed characteristics, and recent literature does not show progress in this field.

1.2 Problem to address

Frequently, SDSS deal with ill-structured problems (Cats-Baril, 1987). This means that the goals to reach might not be totally clear and/or there is insufficient information to solve the problem in a certain optimal way. Because of this, in ill-structured problems there might be many stakeholders and decision-makers involved in.

However, the problem is that no one knows exactly what actions are required for resolution. Furthermore, after they tackle the problem, the definition of the problem may change. Even after the stakeholders propose a solution, they will never be sure they have made the right decision. The problem involved in SDSS is the need to provide the necessary support in order to make the best possible decision based on the information at hand.

According to the state-of-the-art (Jankowski, 2018) (Bordogna, Pagani, & Pasi, 2010), the necessary processes for supporting spatial ill-structured problems are the following:

1. Combining data from different sources, providing degrees of certainty of each and a method to evaluate and combine data of the same type and area, but with different levels of certainty.
2. Generating suitability scenarios using different assumptions in a flexible and systematic process. This is needed because it replaces the criteria/formula specification, providing a method to combine the hypothesis of multiple experts and decision makers.
3. Offering suitable tools to compare different scenarios. This is needed to complete a decision-making process. The comparison of two or more scenarios can provide a better solution and more information on the problem itself.

The required knowledge to tackle these problems has not been developed yet, thus the objective of this research is to create a general method (a framework) that helps address spatial ill-structured problems. This method should have the flexibility to be replicated or adapted to different spatial problems. The relevance of this research is based on the problems that can be supported using a general method. However, in order to achieve this objective, I defined a general research question that takes into account the characteristic needed to address spatial ill-structured problems:

How do we combine spatial data from different sources when each source can have different certainty levels?

In order to answer this question, we need to establish a separation between a spatial decision-making problem and the data combination problem. This led us to research two specific research questions. The first one focuses on the spatial decision-making problem and the second research question comes from the data combination problem:

***RQ1**: How do we generate suitability scenarios combining knowledge from various sources in a flexible and systematic process including multiple experts involved in the decision-making process?*

***RQ2:** How do we compare or combine two or more scenarios providing a better solution and more information on the problem itself?*

1.3 Research objectives

The general goal of this thesis work is to develop a conceptual framework to deal with spatial ill-structured problems by supporting the complex steps in SDM in a flexible but systematic manner. This framework involves two main components: 1) a scenario evaluation and combination engine using our adaptation of the Dempster-Shafer Theory to spatial context, 2) a scenario specification language that allows representing expert's knowledge and including available spatial data.

Considering the stated general goal, this work proposes the following research objectives:

1. *Adapting and applying belief functions and mathematical decision-making theory in spatial contexts, to combine data with different certainty levels and scenarios*.
2. *Developing a method to generate spatial decision-making scenarios in a flexible and systematic process to allow the inclusion collaboration between multiple stakeholders in the process*.

The specific goal 1 (SG1) is the first step to reach the SG2. This latter goal can be divided in two parts according to the decision-making process: "generating scenarios *in a flexible and systematic process*" and "the *inclusion of multiple stakeholders allowing them to* compare, evaluate, and combine scenarios", therefore, it is decomposed as follows:

- SG2a: Developing a method to generate spatial decision-making scenarios in a flexible and systematic process based on mathematical decision-making theory.
- SG2b: Developing a method to evaluate, compare and work collaboratively with previously generated scenarios.

The framework committed in this thesis work also includes scenario comparison operators and spatial operators allowing specification of spatial relationships or data filtering in the scenarios. This framework was tested for its ability to support decision-making in three different scenarios in order to evaluate the appropriateness and generality of the framework to the task: an accurate crime prediction system based on expert knowledge, a public transportation demand estimator and a collaborative planning tool. Each example scenario was designed to test different characteristics of the architecture in order to fulfill the research goals.

1.4 Structure of the thesis document

This document is structured in three main areas: background, the proposed solution (scenario specification method) and finally a thesis validation based on three different applications and problems (see Figure 2):

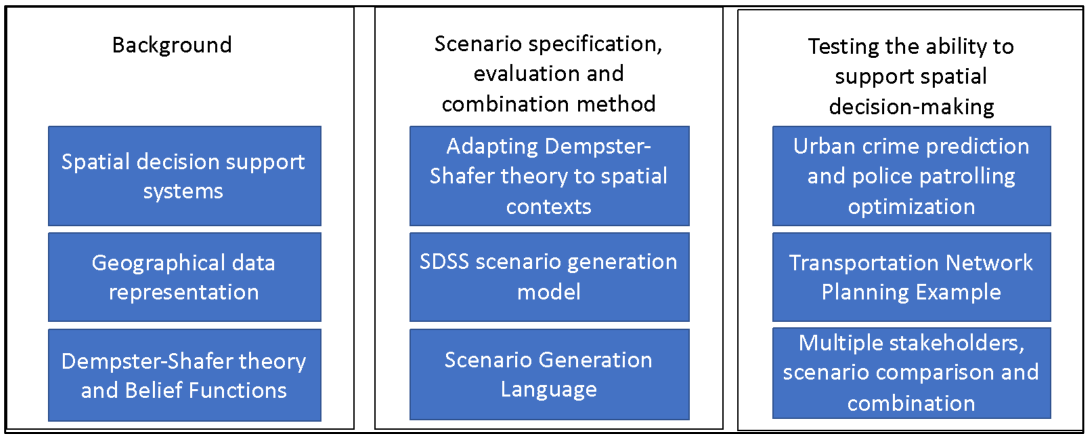


Figure 2. Structure of the Thesis Document.

- **Background:** Discusses DSS, SDSS characteristics and models in **Chapter 2**, **Chapter 3** includes the **fundamentals** of geographic data indexing and describes the technical problems and approaches to deal with spatial data. **Chapter 4** describes the Dempster-Shafer Theory providing simple examples and applications.
- **Scenario specification, evaluation and combination method:** The core of this research is a mathematical model used to represent geographical information based on the Dempster-Shafer Theory in a spatial context (described in **Chapter 4**). **Chapter 5** presents the scenario's generation method based on the extension of the mathematical model. **Chapter 6** describes the software design approach to compute the proposed method**.** Finally, **Chapter 7** describes a scenario generation language.
- **Testing the ability to support spatial decision-making:** The validation of the model is distributed in three different applications:
    - **Chapter 8:** Presents a real case scenario in which the theory and algorithms featured in the previous chapters are used to implement a tool for predicting crime in a big city. This part of the work was possible with the collaboration of the national police planning and prevention area. Based on an expert's knowledge and crime prevention literature, this chapter validates the scenario modeling method by building a crime prediction method. The results were tested using the standard metrics using large crime datasets available from different places of the world.
    - **Chapter 9:** Focuses on validating the resulting scenarios against the data of a whole city using the combination of multiple sources to build new information. It describes a public transport network-planning tool used to estimate the demand of the people. This part of the work presents a method to use existing crowdsourced data (like Waze and OpenStreetMap) and cloud

services (like Google Maps) to support a transportation network decision-making process. This chapter shows an example of the current availability of public transportation stops in order to discover its weak points. The results were validated using the transportation data of Santiago.

- **Chapter 10:** Provides a method to compare and work with previously generated scenarios. The method was presented using an example where experts must evaluate and compare many scenarios, which arise from different hypotheses. The example requires us to know where people may be at the time of the emergency and how they will react. This chapter explains this process by helping a hypothetical group of experts, generating, visualizing and comparing the outcomes of the different hypotheses. It should be noted that in this work, I am going to use the word hypothesis as the expert knowledge that can be used to build a spatial model or scenario, but it needs to be tested, such as "Plants grow better near water".
- **Chapter 11:** Presents the conclusions and future work.

Chapter 2

Related Domains Review

In the last two decades an increasing amount of multi-criteria analysis in spatial planning has be seen, the same tendency applies to multi-criteria spatial decision support systems. The evolution of this area offers assessment techniques to different problems, for instance, urban and regional developments plans, according to multiple, conflicting, and often incommensurate dimensions, which are measured via well-specified criteria (Ferretti & Montibeller, 2016). However, the main problems in this area are the integration in many domains and the spatial relation between objects.

In order to understand the aggregated complexity of SDSS, this review starts analyzing the relation between objects problem continues with the integration of different domains and ends with the decision-making models that can be applied in this area.

2.1 Spatial object representation

Kettani and Moulin (Kettani & Moulin, 1999) developed conceptual maps describing routes in a qualitative way. Their spatial models include the natural notion of objects influence on areas. Each area determines how people reason about objects, evaluating and qualifying distances between objects, etc. This approach continues in the last decade with the particular interest of the AI community on the Region Connection Calculus (RCC) theory (Randell et al., 1992), with is basically a formalism to reason about regions by defining a relation between two regions based on their topological properties and therefore independently of any coordinate system, for example, two neighboring areas have a externally connected (EC) type relationship. However spatial problems often have to deal with some form of uncertainty or imprecision. For example, the urban area of a city can be different that the real location of houses around the city. In order to deal with this kind of uncertainties fuzzy logic algorithms were introduced in this topic.

Hans Guesgen (Guesgen ,2006), adds a quantitative measure of relation for RCC between [0,1], allowing representations like the following:

DC(X,Y) -> 0.7 , X and Y areas have a 0.7 probability of being disconnected.

O(X,Y) -> 0.1, X and Y areas have a 0.1 probability of being overlapped.

EC(X,Y) -> 0.1, X and Y areas have a 0.1 probability of being externally connected.

This kind of representation offers multiples possibilities for the same relationship between X and Y, also there is a 10% of uncertainty that is omitted at by definition. The composition of fuzzy relations plays a central role in a number of algorithms for reasoning about fuzzy relations, however this composition could lead to inconsistent or incoherent states, for example a possible output can be: X contains Y, Y contains Z, Z is externally connected to X.

Another interesting work were made by Sungsoon Hwang and Jean-Claude Thill (Hwang & Thill, 2007), they incorporate the a “locality perspective” into the modeling process. With means the relations are made from the perspective of an object, and two objects can have different perspectives and relationships types of each other. A particular and relevant problem is checking if a point belongs to an area, for example if a city is in a warm region. A solution approach is to define a fuzzy area with certain characteristics, Zhan and Lin (Zhan & Lin, 2003) proposes that each element must be composed of three elements: core, boundary and exterior.

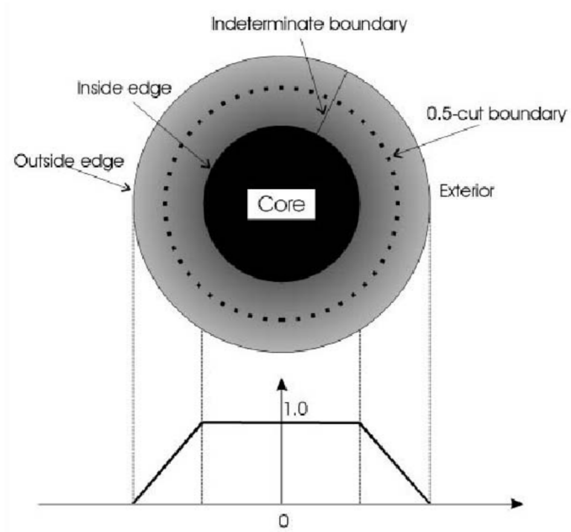


Figure 3 Representation of a Fuzzy Region and its Fuzzy Set Membership Function, (Zhan & Lin 2003)

Figure 3 shows the limits of the area and where the objects belong. Objects inside the core definitely belong to the area, inside the cut boundary perhaps be-long, and the outside definitely do not belong to a specified locality. Now, if we imagine two areas X and Y and one object, we can have different states for the same object, for example that the objects belong to area X and perhaps belongs to Y.

Maybe the most interesting part of this kind of membership method is that the areas can have different boundaries functions and cut limits, providing a modelling option for local perception of nearness without falling into inconsistent states as the fuzzy version of RCC.

The Zhan and Lin object representation method has can be used in modern geographical database system by developing custom functions to determinate different levels of boundaries around the objects. However, in this work I will use the boundaries as an input of a decision support theory, this theory and method will be explained in the following Chapters.

The next knowledge domain that we need to study is are the Spatial Decision Support Systems (SDSS), which consist of the integration of Geographical Information Systems (GIS) and Decision Support Systems (DSS).

2.2 Spatial Decision Support Systems

This section describes decision-making models and how they have been applied to the spatial domain applications such as territorial management, policy making, and governance.

According to literature, the spatial decision-making models must be integrated with spatial data (Ghayoumian, Ghermezcheshme, Feiznia, & Noroozi, 2005) (Coggins, Coops, Hilker, & Wulder, 2013) (Tran, O'Neill, & Smith, 2012) (Baloian, Frez, Janser, & Zurita, 2011) (Baloian, Frez, Zurita, & Milrad, 2012). However, these data are traditionally managed with GISs, and modeled to be indexed and queried according to spatial or logical restrictions (Meagher, 1982) (Samet, 1995) (Sozer, Ouguztuzun, Petry, & others, 2010). These restrictions are relevant to the integration and implementation of a DSS for geographical data.

In our case, I am working in a DSS area related to spatial information; this particular topic is called Spatial Decision Support Systems (SDSS) (Morris, 2003). In the following lines, I highlight some characteristics of SDSS, decision-making models. I also describe the basic models in a spatial context and analyze the difference or gap between the existing SDSS and the requirements mentioned in the literature. Finally, I mention the computational limitations of indexing methods.

A specific area in SDSS is an emerging category designated as Collaborative Spatial Decision-Making. According to state of the art, CSDM deals with the provision of the following functionality (Crossland, Wynne, & Perkins, 1995): collecting spatial-related data, identifying locations according to a set of criteria, exploring relationships and displaying and analyzing the data, some examples of SDSS can be seen in (Frez, Baloian, Zurita, & Pino, 2014) (Frez, Baloian, & Zurita, 2012) (Frez, Baloian, & Zurita, 2012).

A SDSS includes general decision-making models and a decision-making process. These models are described in the following sub section and a full description of SDSS is provided in Section 2.2.

2.2.1 Decision-making models

According to (Antunes, Sapateiro, Zurita, & Baloian, 2010), there are three conceptual views of the decision-making process. The first one views the decision process as a production system with three main components: inputs, process, and outputs (Jankowski, Nyerges, Smith, Moore, & Horvath, 1997). The second regards the process as a composition of data management, model management, and dialog management. This view focuses on controlling the strategic, tactical and operational decisions of the decision makers (Pearson & Shim, 1995) (Rosenheck, 2001).

The third model conceptualizes the different aspects of a DSS: the way of thinking, the way of controlling a problem conceptualization, the way of working and modeling as a cycle including adjusting control methods (Seligmann, Wijers, & Sol, 1989) (De Vreede & Briggs, 2005). See Figure 4.

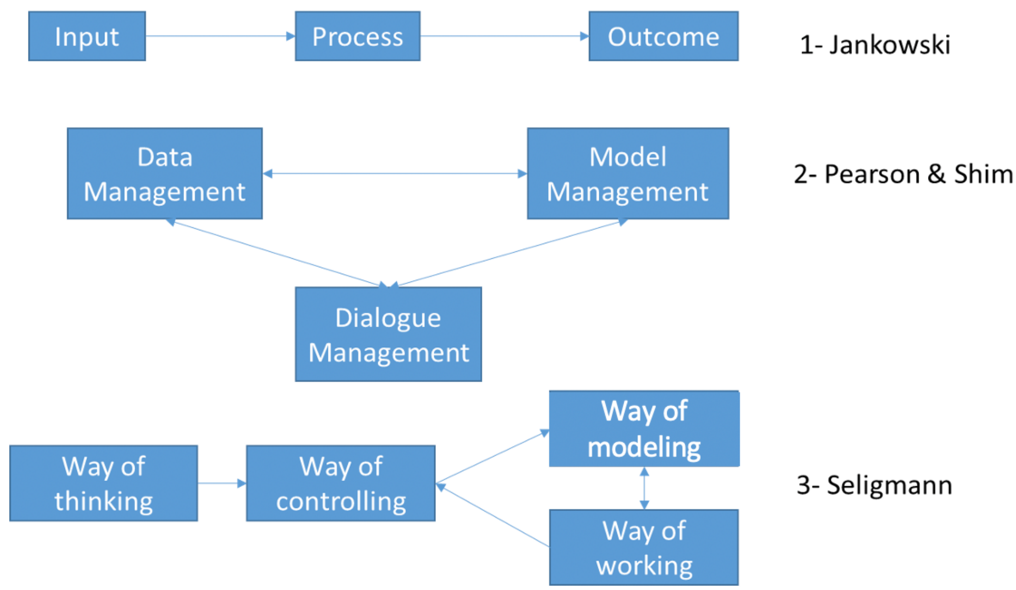


Figure 4. Conceptual views of decision-making according to three different authors (Crossland, Wynne, & Perkins, 1995).

The conceptual views can be seen as meta-models because they should be used to design more specifics models. Along these same lines, many models can explain the decision process. For instance, there is the Subjective Expected Theory (Fishburn, 1981) (Ramsey, 1931). This theory considers that experts are being faced with choosing from a set of alternatives and outcomes. This decision is made by defining a utility function to determine which choice is the most suitable. Another theory called Analytic Hierarchy Process (Saaty, 1990) defines four steps in making decisions: break the problem into a hierarchical set of decision elements; collect data associated to the decision elements; estimate relative weights of decision elements; aggregate the relative weights to obtain a set of rating of the alternatives. The work in (Simon et al., 1987), emphasizes that real-world decision makers do not find the perfect condition necessary to solve the problem. Continuing with this statement, this work uses a generic problem solver model (Crossland, Wynne, & Perkins, 1995) with three main elements: representing the problem, finding alternatives and selecting alternatives. Furthermore, unlike other models, this one includes heuristics to explain the problem and simplify the process. The output of this model is not a maximized solution, but a solution that meets reasonable conditions.

The recognition-primed decision-making theory (Klein, 1997) introduced a naturalistic perspective in the field (Lipshitz, Klein, Orasanu, & Salas, 2001). This perspective includes accuracy uncertainty by the limited time to make the decision, ill-defined goals, and other factors that affect the decision makers. The objective of this theory is to bring forward the concept of situation awareness as mechanisms to apprehend expectancies, cues, and goals. On a similar note, the Cooperative Decision-Making model (Wong, 1994) emphasizes the importance of negotiating differences in the decision-making process. The Participatory

Decision-making model (AbuSabha, Peacock, & Achterberg, 1999) makes a distinction between the Convergence and Divergence phases in collaboration. The differences between the mentioned models are shown in Figure 5.

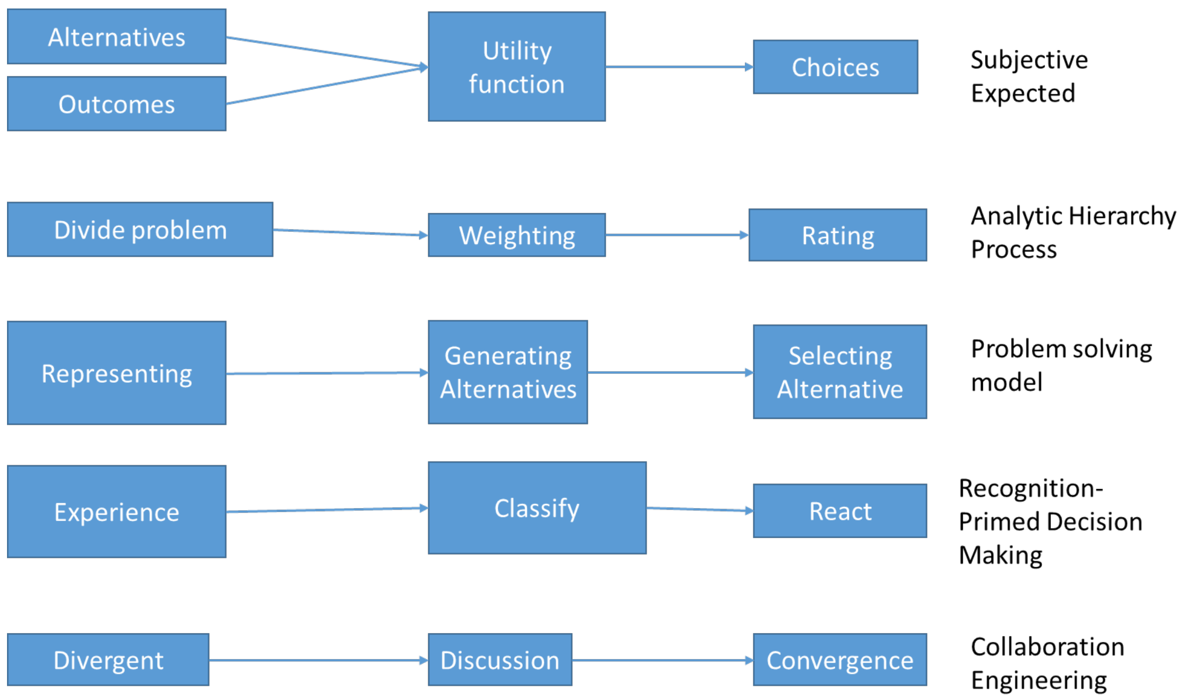


Figure 5. Decision-making models according to various perspectives (Crossland, Wynne, & Perkins, 1995).

From the decision-making models, we can identify three different mains activities in the process; the first one is related to the problem identification and evaluation of alternatives. The second is a processing methodology including weighting, classifying, defining utility functions and generating more alternatives. The third one involves alternatives evaluation and final selection of a scenario.

2.2.2 Spatial decision support systems

Geographical information systems have advantageous techniques to collect, integrate and manage large amounts of geographical data (Gao, Paynter, & Sundaram, 2004). Furthermore, they also offer techniques for data management, information extraction, and visualization. However, they lack analytical capabilities to support a decision-making process efficiently (Densham, 1991). In contrast, spatial decision support systems are defined as follows:

> *"An interactive, computer-based system designed to support a user or group of users in achieving the higher effectiveness of decision making while solving a semi-structured spatial problem" (Densham P. J., 1991)."*

To achieve this higher effectiveness, the multiple criteria-based research area has gained more attention in SDSS. It is also known as Spatial Multiple Criteria Decision-Making (SMCDM) or Spatial Multi-criteria Evaluation. The primary objective of this area is to evaluate multiple geographic choice alternatives using a multiple decision criteria technique.

According Morris, Jankowski, Bourgeois, & Petry (2010), traditional SMCDM approaches suffer from an inappropriate logic and faulty intelligence. Weighting techniques assign weight to some criteria. However, the evaluations are mostly Booleans (Morris, Jankowski, Bourgeois, & Petry, 2010). Fuzzy logic and fuzzy theory have been proposed as a solution (Morris & Jankowski, 2001) (Robinson, 2010) (Rocchini, 2014). Ashley Morris also proposed analysis techniques that allow fuzzy functions to be used (Morris, 2003). He also recognized the relevance of dynamic generation of maps as a useful decision support characteristic. The previous statement is validated in (Morris, Jankowski, Bourgeois, & Petry, 2010) (Wood & Dragicevic, 2007) (Binaghi, Luzi, Madella, Pergalani, & Rampini, 1998).

Decision problem characteristics can be used to classify a process. A specific problem may need certainty or uncertainty analysis methods depending on the amount and accuracy of available information. If the available information is complete (perfect), the problem will have certain conditions. Otherwise, the problem will have uncertain conditions. This basic difference is important to this research for two reasons (Zadrożny & Kacprzyk, 2010) (Bordogna, Pagani, & Pasi, 2010) (Malczewski, 2006):

- Real-world decisions involve uncertain data, lack of knowledge and predictions limitations.
- Approximately 77% of the research in SDSS considerers certain conditions.

According to (Malczewski, 2006) the remaining 23% of research can be divided into two types of analysis methods: Probabilistic and Fuzzy.

Most works mentioned in the literature refer to specific analysis techniques for specific problems. However, to represent suitability analysis results, they tend to use a similar visualization (Jankowski, Nyerges, Smith, Moore, & Horvath, 1997) (Andrienko, 2007). This visualization has been named in (Chen, 2010) as “Suitability map” or S-map.

An S-map is defined as a spatial distribution of the overall degree of suitability for a specific use. Each overall degree of suitability is represented by a real number in the [0, 1] interval. Lower values denote unsuitable locations, and higher values are most suitable locations (see Figure 6).

A suitability value for a particular area depends on a finite number of attributes, which can be produced by GIS systems. The resulting maps represent the distribution of geometric objects in a two-dimensional space or another kind of information that can be used for planning or decision-making. The geometric objects can represent borderlines, cities, roads, airports, sewage systems, telecommunications and transport infrastructure, among others. However, other necessary characteristics could be the altitude, materials, distance to main roads, distances from electrical systems.

Each relevant attribute for the decision-making problem must be associated with a representation, and handling based on what the decision makers know about it. In order to build a suitability map, each criterion must be evaluated. As a result, a set of suitability degrees can be obtained for each alternative. These values can be aggregated to guarantee an efficient representation of the decision makers domain into an S-Map.

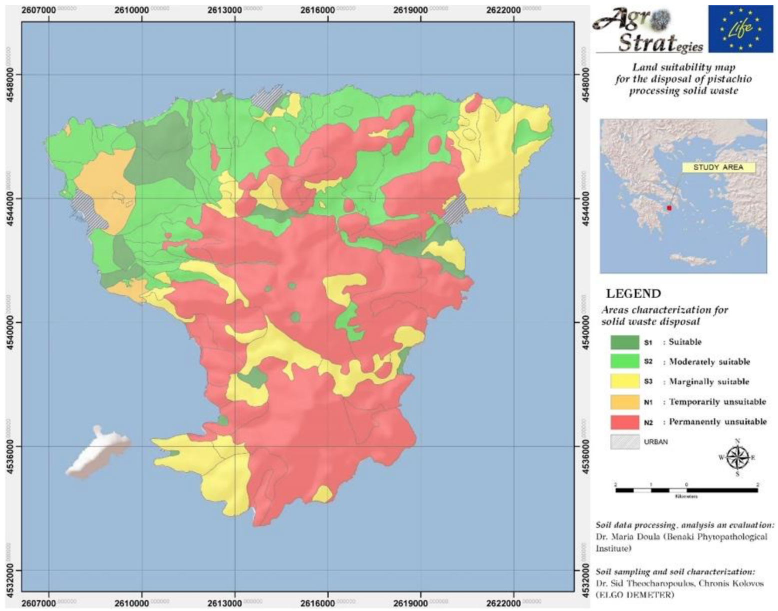


Figure 6. Suitability Map Example for land spreading of pistachio solid waste, source: http://www.agrostrat.gr/en/suitabilitymaps

In this work I maintain the S-Map visualization. However, the proposed method must use GIS standard representations and indexation. To correctly understand the computation restrictions (Meagher, 1982) (Samet, 1995) (Sozer, Ouguztuzun, Petry & others, 2010), the following section describes some basic spatial data computing techniques used to index and query data.

2.3 The design problem of SDSS

Facing spatial decision problems usually involves different dimensions and several stakeholders. SDSS can support decision makers in handling and solving the problem by providing group decision features, and steps to enhance the process.

The decision-making methodologies are mostly focused on identifying the stakeholders, a consistent set of criteria's, weights assignment and aggregation procedures. For example, facing a facilities location requires the participation of different categories of stakeholders, depending of the facility type, the location could be a logistic problem or a political problem. Currently there are mechanism and methodologies to define stakeholders and criteria's, however weights and aggregation procedures are strongly related to the modelling method to be used.

Since there is not a general accepted method, I can say that there is a need of a method that allows to use different weighting techniques and aggregation procedures. However, designing this kind of method requires divide the design into layers, starting from: A spatial object representation and indexing methods; finding and adapting a decision-making theory that allows multiple stakeholders with different weighting techniques and criteria's; a customizable aggregation procedure; unifying the object representation, decision making theory and aggregation procedures into a scenario generation model; and finally a user interfaces that allows to parametrize all the mentioned options and variables.

This work proposes a method that allows the mentioned flexibility. However, the design and its fundamentation are divided in five layers starting from Chapter 3 to Chapter 7:

- Geographical Data Representation
- Dempster-Shafer Theory and Belief Functions
- Adapting Dempster-Shafer Theory to Spatial Contexts
- SDSS scenario generation model
- Scenario Generation Language

Finally, this works ends testing the proposed SDSS method by testing the needed characteristics presented at the introduction.

Chapter 3

Geographical Data Representation

Geographic information systems are designed to perform spatial queries on geographic data (such as the distance or the location where two objects intersect). A spatial query requires knowledge about locations as well as attributes to such locations. This work is based on different types of spatial queries. In order to understand the contribution of this research we need to see the big picture of geographical data representation.

This chapter discusses some essential aspects of this topic: I first explain how the real world is represented using simpler and digitally stored elements, then explain what a spatial data model is. The vector data model and review the raster data model is specified. Finally, I describe how GISs data is symbolized.

3.1 GIS Databases

GISs databases are a digital representation of the real world. Once we identify the aspect of the real world that we want to represent, we break it down into objects or fields. We next decide on what the best way is to encode the objects and fields we have identified; this Data model stage includes encoding approaches such vectors and raster (Smith, 2007). Then we save the encoding into databases with specific structures (Smith, 2007), see Figure 7. Objects are the usual name used to identify a group of elements with a well-defined spatial boundary such as points, lines or areas (Smith, 2007). These objects have a unique ID, and attribute tables that provide additional information about the object.

Fields (Smith, 2007) are elements of reality for which their boundaries are continually changing over time with a gradual transition from one representation to another. For example, natural features change over time.

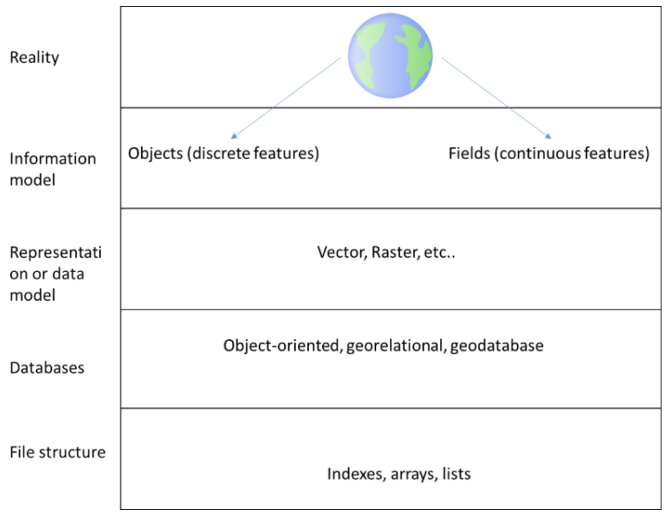


Figure 7. Stages of reality digital representation process.

In order to explain the raster and vector models, I am going to represent a small lake as an example.

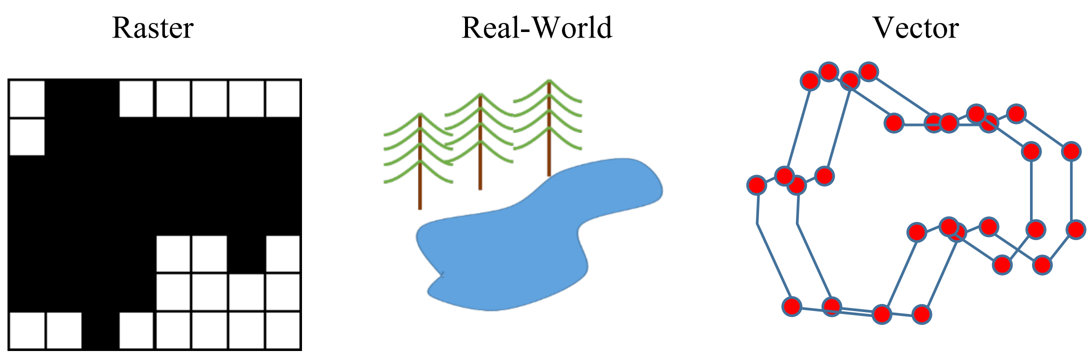


Figure 8. Representation examples.

In Figure 8, the real world corresponds to the lake in the middle. The vector data model, (right) takes that element and represents it as points, lines, or an area, which can also be called polygon. The raster data model (left) superimposes a mesh on top of the element and then encodes each cell of the mesh to determinate the absence or presence of the element. We can take the real world and decompose it into either a vector or raster representation. Most GISs databases can contain both, raster and vector layers to allow us to have different descriptions of reality.

3.1.1 **Vector data model**

Specifically, the vector data model contains points, lines or polygons. A point has a unique identifier such as house locations that can be represented using a point layer.

A series of connected points is called a vector line feature; each line segment has unique identifiers. The beginning and the end parts of a line are named nodes and the intermediate points are called vertices. This can be seen when roads are represented as a vector lines layer in GISs databases.

If the beginning and the end of a line are the same, then we have a vector area. Each polygon has a unique identifier.

In the practice, points, lines and polygons are the most often represented as follows (Smith, 2007):

- Points: as a pair of coordinates, in latitude/longitude or some standard system.
- Lines: as an ordered sequence of points connected with straight lines.
- Areas: as ordered rings of points, also connected with straight lines to form polygons. In some cases, they may contain holes or include separated areas. In those cases, each area is represented as a separated polygon.

All these features are called Geometry of the Vector Spatial Data Representation. These geometries can have links to an attribute table (see Figure 9).

Geometric Type	Representation
Point	[lat,lng]
Line segment	[[lat1,lng1], [lat2,lng2]]
Polyline	[[lat1,lng1], [lat2,lng2], [lat3,lng3], [lat4,lng4]]
Polygon (Area)	[[lat1,lng1], [lat2,lng2], [lat3,lng3], [lat4,lng4], **[lat1,lng1]**]

Figure 9. Vector model geometry types.

Using the geometry type in the vector model it is possible con calculate attributes such as lengths, angles, areas, etc.

The geometries can be extended using the concept of topologies to obtain additional spatial properties, such as connections between features, features close to a certain point or event features next to another feature. Topology can be considered as a set of rules that specify links between features. An example of topology can be roads, establishing relations between points where a vehicle can exit a road and enter another one.

3.1.2 Raster data model

As was previously discussed, the raster data model can be imaged as a mesh on top of the real-world elements, deciding which elements will be encoded into the cells. The mesh is an array of rows and columns with a unique identifier per cell. Each cell has a value, usually called the digital number (DN). There will be a resolution defined by the width and height of the cells; smaller cell sizes mean higher resolution.

3.1.3 Comparison between vector and raster data models

A raster data model is used to represent elements in the natural world, such as forest changes and landslides. The vector data model is more useful for human-created elements, such as roads or population census data.

This work focuses on data represented using the vector data model. I also include data that can be associated between a spatial object and a known location and shape (Samet, 1990). Furthermore, the spatial object can be defined as a surface, volume or variation of a single object during time. These objects can be associated with a specific context, as seen in borders, streets, buildings, and others. This data can also be associated with events, for example: a remote sensor network can provide measurement data based on specific events (Aasman, 2008) (Cugola & Margara, 2012). In this case, I am talking about event-driven systems (Weigand, 2011) (Tatbul & Zdonik, 2004) (Tatbul & Zdonik, 2006) (Michelson, 2006) (Gross, Digate, & Lee, 1994) (Ghalsasi, 2009).

Finally, following the stages of the reality digital representation process presented in Figure 6, the geographical information data indexation can vary given the objective. For instance, (Zadrożny & Kacprzyk, 2010) establish that the same query can be answered with a discrete or a fuzzy algorithm.

3.1.4 Coordinate systems and Open Geospatial standards

Currently, the Open Geospatial Consortium (OGC) provides open standards for the global geospatial community. These standards are defined through a consensus process between different actors (currently 521 companies, government agencies, and universities) involved with the reality representation process. The following section presents an overview of these standards.

The OGC defines an architecture supporting its vision of geospatial technology and data interoperability called the OGC Abstract Specification. This specification provides a conceptual foundation for most OGC standards enabling interoperability between different brands and different kinds of spatial processing systems.

The specification starts with the coordinate system definition. The spatial references specify the location of the elements in the real world, and this can be done by using a coordinate system or establishing a geographical identifier. This work focuses on the vector data model using the coordinate system method.

Since the beginning of geography science, representations of space were based on different references systems. A coordinate is a scalar value that defines the position of a single point in 1D. A coordinate tuple is an ordered list of n coordinates that define the position of a single point in 2D.

Coordinates are ambiguous until a Coordinate Reference System is used. Given this, the definition of a coordinate set is a collection of tuples referenced by the same Coordinate System.

The selection of a coordinate system represents the spatial property with precision. All coordinate systems can induce errors if they are not correctly selected. The reason is that the earth has an irregular shape, and most coordinate systems are designed to fit with a geometric shape. The Earth shape is called geoid, and the coordinate system approximation is usually called ellipsoid, see Figure 10.

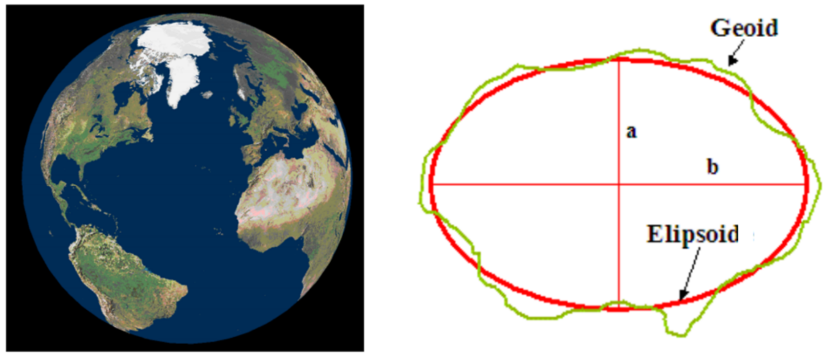


Figure 10. Geoid and ellipsoid.

A coordinate system is also called datum, and there are various types of datums. These types are grouped by the method used to deliver an approximation of the shape of The Earth. The most used types are the following:

- Geodetic: a coordinate reference system that is associated with a geodetic datum.
- Vertical: a coordinate reference system that is associated with a vertical datum.
- Engineering: a coordinate reference system that is associated with an engineering datum.

In literature, we can find multiple techniques and data structures to geographic index data. The objective of this chapter is to understand the spatial indexation methods and provide definite conclusions about what kind of queries can be used and what type of data structure must be used in some instances. In the following sections, I am going to describe the most common data structures, storage format, and query methods briefly.

3.2 Geographical databases operations

Focusing on the vector data model, the literature defines three base types of geographical data: points, line segments, and areas (polygons). These three types can be used to build more complex geographical types as polylines and geometry collections.

The interactions between operations and their results based on these three types of geometries are defined by ROSE algebra (Guting & Schneider, 1995). The original formal definition of ROSE algebra work defines two sets of data and operation as follows:

- EXT : {lines, areas}
- GEO: {points, lines, areas}

Using these sets, four groups of operations can be computed:

- Spatial operators and data sets

- $\forall obj \in OBJ. \forall geo, geo1, geo2 \in GEO$
    - $set(obj) * (obj \to geo) \to geo$ Sum
    - $set(obj) * (obj \to geo1) * geo2 \to set(obj)$ Neighborhoods
- Sum
    - Returns the geometric union of all the geo-objects in the set.
- Neighborhoods
    - Return the set of objects within a distance from geo2.

- Spatial objects generation
    - $\forall geo \in GEO$:
        - $lines * lines \to point$, intersection point
        - $areas * areas \to areas$, intersection area
        - $geo * geo \to geo$ union, difference
        - $areas \to lines$, shape
- Numerical evaluations
    - $\forall geo1, geo2 \in GEO$:
        - $geo * geo \to real$ distance
- Spatial relations:
    - $\forall geo \in GEO. \forall ext1, ext2 \in EXT. \forall area \in areas$
        - $geo * areas \to boolean$ inside
        - $ext1 * ext2 \to boolean$ intersects
        - $area * area \to boolean$ adjacent
    - Inside:
        - Compare points, lines, and areas with areas. Return true if they are inside the area.
    - Intersection
        - Returns true if two elements are intersected.
    - Adjacent
        - Returns true if two areas are adjacent.
        - $areas \to real$

Using the ROSE algebra definitions, it is possible to build representations of real elements. For instance, we can define buildings and districts as follows:

- Building: {name: STRING; center: POINT; perimeter: AREA}
- District: {name: STRING; area: AREA; habitants: INTEGER}

In the same line as was described in the vector data model section, using the ROSE algebra, it is possible to extend three types of relations between the representations of real objects. A sample would be if an object is within another, or they are close. There are three types of relations:

- Topological
    - Relations are given by a location and shape of the elements such as spatial intersections between elements.
- Directional
    - Relations are given by a relative position of the elements. For example, Up, Down, at the north of, etc.
- Metrical
    - The relation is given by numerical evaluations. For instance, if two elements are at 200 meters or less to each other.

Using these relations, it is possible to build queries like "Buildings in Santiago". This query can be structured using spatial predicates (Güting R., 1994):

building select [center inside Santiago]

The building will be a set of objects; the center is the attribute to be evaluated from the set of buildings, inside is the operation (relation), and Santiago is the District to be evaluated using the area attribute.

This kind of query is typically implemented as SQL extensions or similar languages (Brickley & Miller, 2010) (Brickley, 2006). Some of these extensions are still under discussion. However, I can provide a base example on the most accepted implementations:

select building.name from building, district where building.center inside district.area;

Data must be standardized before performing a query, and there are several ways to store data, calculate distances, and evaluate relations. In order to ensure calculation correctness, standards must be used.

Currently, the Open Geospatial Consortium (OGC), provides open standards for the global geospatial community. These standards are stated through a consensus process between different actors involved with the reality representation process. The following section presents an overview of these standards.

3.3 Simple feature access

The Open Geospatial Consortium has defined a Simple Feature Access (SFA) standard based on the SQL schema, supporting storage, retrieval, query and update of spatial collections via SQL language. Spatial and non-spatial attributes define a feature. The spatial attributes are geometries, and they are based on two or more dimensions, for example, points, curves or surfaces. The 2D spatial data types defined in the standard are the following:

- GEOMETRY
- POINT
- LINESTRING
- POLYGON
- CURVE
- SURFACE
- POLYHEDRALSURFACE
- MULTIPOINT
- MULTILINESTRING
- MULTIPOLYGON
- GEOMCOLLECTION

The standard also defines several methods and operation between features (geometries); some of them are the following:

- ST_Dimension: Returns the inherent dimension of a geometry.
- ST_GeometryType: Returns the type of geometry as a string.
- ST_AsText: Returns the Well-Known Text representation of geometry.
- ST_AsBinary: Returns the Well-Known Binary representation of the geometry.
- ST_SRID: Returns the spatial reference identifier (Coordinate system).
- ST_IsEmpty: Returns true if geometry is an empty geometry.
- ST_Boundary: Returns the closure of the combinatorial boundary of a geometry.
- ST_Envelope: Returns the minimum bounding box for the supplied geometry.
- ST_Equals: Returns true if two geometries are spatially equal.
- ST_Disjoint: Returns true if two geometries do not spatially intersect.
- ST_Intersects: Returns true if two geometries spatially intersect in 2D.
- ST_Touches: Returns true if the geometries have at least one point in common.
- ST_Crosses: Returns true if the supplied geometries have some, but not all, interior points in common.
- ST_Within: Returns true if the geometry is entirely inside another geometry.
- ST_Contains: Returns true if and only if no points of a geometry lie in the exterior of another one.
- ST_Overlaps: Returns true if the two geometries share a space, but they are not entirely contained by each other.
- ST_Relate: Return the relation code of two geometries based on the Dimensionally Extended Nine-Intersection Model (DE-9IM).

- ST_Distance: Returns the 2-dimensional Cartesian minimum distance between two geometries in projected units.
- ST_Intersection: Returns a geometry that represents the shared portion of two geometries.
- ST_Difference: Returns a geometry that represents that part of a geometry that does not intersect with another one.
- ST_Union: Returns a geometry that represents the union point set of two or more Geometries.
- ST_SymDifference: Returns a geometry that represents the portions of two geometries that do not intersect.
- ST_Buffer: Returns a geometry that represents all points whose distance from a Geometry is less than or equal to specified distance.
- ST_ConvexHull: The convex hull of a geometry represents the minimum convex geometry that encloses all geometries within the set

In order to present the necessary operations defined by the Rose algebra in the OGC standard, I made an equivalency with the ROSE algebra definitions:

ROSE Algebra	**OGC SFA Runtimes**
Sum	ST_Union
Neighborhoods	In order to return a set of features, use ST_Touches in the Where statement.
Intersection	ST_Intersection
Difference	ST_Difference
Shape	ST_Boundary
Distance	ST_Distance
Inside	ST_Within
Intersects	ST_Intersects
Adjacent	ST_Touches

3.4 Spatial indexing

Geographical elements can be described by points and a coordinate system. Usually, Latitude and Longitude are used to describe a location. However, a different coordinate system can be chosen. In this section, I am going to describe the spatial indexing techniques for 2-dimensional coordinates systems. Also, more complex geometries as lines and polygons will be described using points (See Figure 11).

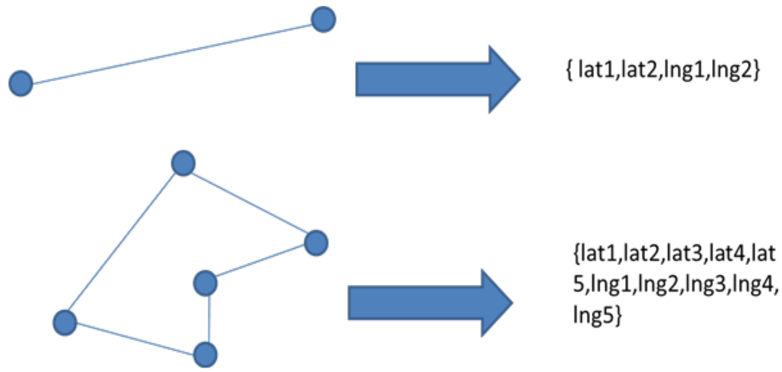


Figure 11. Representation examples.

The elements can be indexed by associating elements to square areas or buckets. The buckets can be organized into a tree structure. The top buckets contain one or more buckets (see Figure 12).

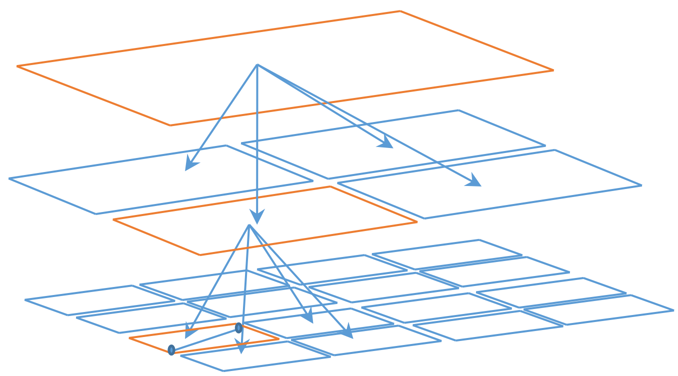


Figure 12 Tree of Buckets.

This type of structure can be implemented by extending a B-Tree (Bayer, 1997). Maybe the most used spatial data structure is the R-Tree (Guttman, 1984). The R-tree indexes the points that represent the spatial objects. Each node or bucket must be the smallest rectangle that contains all the points. In consequence, a leaf is a rectangle that contains a single geometry. Also, the number of buckets per level must be provided (M). E.g., using M=3 an R-tree could look like Figure 13:

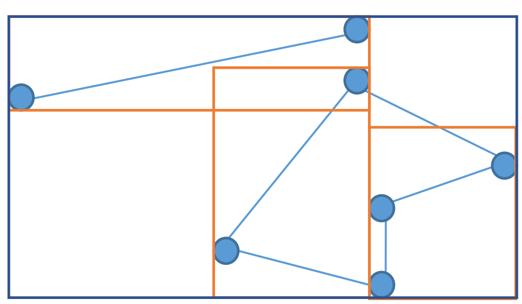


Figure 13 R-tree example, M=3.

The R-Tree is optimized to store information about its geometry type. However, some applications require retrieving large areas with the information associated with it. For example, a satellite image can contain a significant amount of data provided by different sensors. To store and index data by regions, the full area must be divided into four equal rectangles (bucket). Moreover, each bucket will also be divided, in the same way, see Figure 13.

Each rectangle is called quadrant or quad, and the full tree representation is called Quadtree (Samet, 1984) (Figure 14). Quadtree is a Tree index with four children per node.

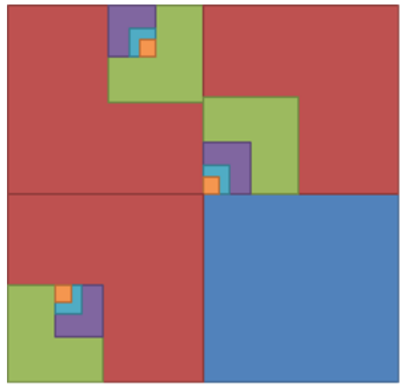


Figure 14 . Quadtree example.

The differences between the R-tree and Quadtree can be summarized as follows:

- A Quadtree can have empty leaves (buckets without elements). However, most implementations do not store empty buckets.
- An R-tree can answer spatial queries like the nearest neighbor more efficiently than Quadtree. However, the Quadtree is meant to work with large areas and frequent update operations.
- A Quadtree cannot correctly represent geodesic data (earth as a sphere) because of the partition method. However, R-Tree does not present problems indexing curved areas.

3.5 Discussion

Most geographical data representation is designed to answer specific query types. However, I can identify two different types of structures depending on the organization methods. R-Tree and Quadtree-based spatial indexes can be used to accomplish two different purposes. This research aims to enable a method to specify a “big picture” query and translate it into specific queries and analysis. Most of our analysis methods will require an association between the elements and a large area. Furthermore, organizing the input and output of decision support method into “analysis areas” it is also essential to improve affordability. In order to support these characteristics, I am going to use an indexing method based on Quadtree. Along these same lines, a R-Tree based indexation will be used to support some specific processing requirements.

The next chapter explains the Dempster-Shafer theory in practical and formal ways, presents the proposed spatial application of the theory, and concludes analyzing the interpretation of the definitions in space-time.

Chapter 4

Dempster-Shafer Theory and Belief Functions

According to Dempster (Dempster, 1967), belief functions provide a non-Bayesian method of applying probabilities to quantify subjective judgments. In these areas Bayesian assesses probabilities directly to the answer of a question of interest, a belief-function user associates a probability of related questions and then considers the implications of these probabilities to the question of interest. Along the same lines, the Dempster-Shafer theory (Shafer, 1976) was developed in 1967 by Dempster and extended by Shafer.

Belief functions are based on two main ideas:

1. Obtaining degrees of a belief of one question from subjective probabilities for a related question.
2. Using the Dempster theory to combine such degrees of belief when they are based on independent items of evidence.

The methodology to apply a belief function to reasoning problems consists of three different components:

1. Representation of incomplete knowledge. In some cases, the knowledge cannot be fully specified by probabilities, and belief functions can interpret a portion of it as *Bel(A),* which is a degree of strength of arguments in favor of A.
2. Belief updating. It should be a method to assimilate the impact of new evidence into the partial knowledge/belief.
3. Pooling of evidence. When multiple parts of evidence are available, a combination method must be provided. This method should be able to encode each piece of evidence as belief functions and combine them. In the Dempster-Shafer Theory, this method is implemented as *combination rules* of the orthogonal sum. The combination rules assume independence among the pieces of evidence and a renormalization of weights to redistribute the proportions of belief.

The composed theory proposes to use sets of hypotheses regarding a variable (e.g., the temperature values in X are always between t1 and t2) associated with a probability of being correct: *Bel(A)*. The theory defines three parallel answers to queries:

- Plausible: is the probability that the variable takes values within the range of the query.
- Certain: is the probability that the whole range of the distribution of the variable (D) is within the range of the query.
- Uncertain: no valuable information can be derived from this data.

To explain this theory, I will use an example: Table 1 shows the mean number of person values associated to a particular location. Also, we have a query Q = [13-23] looking for locations with more than 13 and less than 23 people. In this case, 3/5 of the locations meet this condition (locations 1 and 3 do not).

Location	Mean of #people
1	12
2	20
3	7
4	19
5	17

Table 1. Location versus mean number of people.

Location	Range of #people
1	[9-21]
2	[12-23]
3	[5-10]
4	[17-20]
5	[14-22]

Table 2. Location versus range of number of people.

Table 2 contains a "range" of people registered for each location. In this case, only 2/5 of the locations fully satisfy the condition (locations 4 and 5), and 2/5 may have a possibility to satisfy it (locations 1 and 2). One location does not fall within any interval of the query range (location 3).

Using the Dempster-Shafer evaluation, we can calculate the hypotheses (Plausibility, Certainty, and Uncertainty) for each location. In the example, as shown in Tables 1 and 2, we see that the Certainty level is 40%, and Plausibility level is 80% (Table 3). These values are considered as lower, and upper bounds of plausibility, i.e., between 40% and 80% of the locations have some possibility to have a similar number of people to the queried range. Location 3 adds an uncertainty level and does not provide valuable data to support the query statement.

Location	People	Hypothesis
1	[9-21]	Plausible
2	[12-23]	Plausible
3	[5-10]	Uncertain
4	[17-20]	Certain
5	[14-22]	Certain

Table 3. Location/People D-S for Q=[13-23].

In addition to this information the theory states that a particular weight should be given to each hypothesis, assigned by a human expert or a heuristic. Table 4 features an example where an expert assigns this weight.

Location	People	Weight
1	[9-21]	20%
2	[12-23]	15%
3	[5-10]	35%
4	[17-20]	20%
5	[14-22]	10%

Table 4. Example of weights assigned by an expert.

In this case, since Q = [13-23], the certainty is 30% (20% from location 4 plus 10% from location 5) and the plausibility is 65% (certainty plus 20% from location 1 plus 15% from location 2 which are plausible, plus 30% from the certainty location 4 and 5).

The previously described example can give us a background of how the theory can be applied to specific kinds of problems. However, to generate suitability scenarios using different hypotheses in a flexible and systematic process we must examine the original definition of the Dempster-Shafer theory.

The following sections describe the theory as it is presented by the author and highlights some characteristics that can be used to extend it to a spatial context.

4.1 Formal description

Following the description presented in (Dempster, 1967), let us consider a pair of spaces X and S with a multiple-valued mapping Γ which assigns a subset $\Gamma x \subset S$ to every $x \in X$. Continuing with this idea, suppose that μ is a probability measure of the members of a class τ of being subsets of X. If μ is acceptable for probability judgments about $x \in X$, and if x corresponds to an uncertain outcome of $s \in \Gamma x$, what judgments can be made about $s \subset S$? To answer this question, we must consider a multivalued Γ, and as result, upper and lower probabilities must be considered.

For any $T \subset S$ define:

$$T^* = \{x \in X, \Gamma x \cap T \neq \emptyset\} \tag{1}$$

and

$$T_* = \{x \in X, \Gamma x \neq \emptyset, \Gamma x \subset T\} \tag{2}$$

On a similar note, defining ε as the class of subsets T of S, The upper probability of $T \in \varepsilon$:

$$P^*(T) = \mu \frac{(T^*)}{\mu} (S^*) \tag{3}$$

and

$$P_*(T) = \mu \frac{(T_*)}{\mu} (S^*) \tag{4}$$

The definition of $T\ and\ \ T^*$ tells us that it consists of those $x \in X$ which can possibly correspond from Γ to an $s \in S$ (which is a multivalued single map to X), this is the largest possible amount of probability for the measure μ, and it can be transferred to the outcomes $s \in T$. Similarly, $T\ \ T_*$ is the minimal amount of probability that can be transferred to the outcomes in the previous section $s \in T$. T^* and T_* have been named as plausibility and certainty. However, according to the author, this definition can be extended when the *variate* term is included. The *variate* represents a real-valued function defined over *S*. Further, any *variate V* as an *upper distribution function* $F^*(v)$ and a *lower distribution function* $F_*(v)$*:*

$$F^*(v) = P^*(V \leq v) \tag{5}$$

$$F_*(v) = P_*(V \leq v) \tag{6}$$

Also, the expected values of $E^*(v)\ E\ \ (v)E\ \ (v)$ and $E_*(v)$ are defined by:

$$E^*(v) = \int \quad vdF_*(v) \tag{7}$$

$$E_*(v) = \int \quad vdF^*(v) \tag{8}$$

From (5)(6)(7)(8), I want to highlight the order relations between them:

$$P_*(T) \leq P(T) \leq P^*(T) \tag{9}$$

$$E_*(V) \leq E(V) \leq E^*(V) \tag{10}$$

Finally, P_* and P^* are known as *bel* and *pl* functions in the DSS theory.

4.2 Belief functions and combination rules

The Dempster-Shafer combination rules aim to combine independent sources of information. Further, it considers that a probability measure may be regarded as a belief that quantifies a state of partial knowledge. The combination results are based on pooled evidence [80]. In the Shafer extension to the theory, the probability assignment is called *mass*, and it is denoted as $m \to [0,1]$, the basic probability of a subset *A* of space *X* is:

$$bel(A) = \sum_{\Gamma e_i = A} \quad p(e_i) \tag{11}$$

Also, the mass assignment has some properties:

$$\sum_{A \subset X} \quad bel(A) = 1 - bel(\emptyset) = 0 \tag{12}$$

The Shafer extension provides a hierarchical hypothesis of space. One way to apply the theory is considering a space *X* as an "evidence space" and S as a "hypotheses space." An evidence space is mutually exclusive possible values of an evidential source.

A joint mass definition is given for same class evidence from different sources (*B* and *C)*:

$$bel(A)_{1,2} = \frac{\sum_{B_i \cap C_j = A} \quad bel_1(B_i) bel_2(C_j)}{K} \tag{13}$$

$$K = \sum_{B_i \cap C_j \neq \emptyset} \quad bel_1(B_i) bel(C_j) \tag{14}$$

In (14) K represents the mass with conflict (independence), in (13) K is used as a normalization factor.

A well-known problem with the combination rule is that it will ignore conflicts and assign the resulting mass to a null set (Yager, 1987). The consequence leads to non-intuitive results. In (Zadeh, 1984) provides an example of erroneous results. "Suppose that two physicians see a patient regarding the neurological symptoms. The first doctor believes that the patient has either meningitis with a probability of 0.99 or a brain tumor, with a probability of 0.01. The second physician believes the patient suffers a concussion with a probability of 0.99, however, admits the possibility of a brain tumor with a probability of 0.01. Since the only common element in both groups is a brain tumor, the result of combination leads to 1.

This rule of combination yields a result that implies complete support for a diagnosis that both physicians considered to be very unlikely".

There are other proposed combination rules in literature (Shafer, 1976) (Yager, 1987) (Inagaki, 1991) (Zhang, 1994), In the following section two popular combination rules are reviewed. These two were selected because they are starting points for other combination rules available in the literature.

4.3 Discount and combine method

In (Shafer, 1976), Shafer discusses dealing with conflict using a discount function. Before applying a combination rule, the decision maker can weigh the absolute reliability of the source making distinctions between the type of source: for instance, expert experience, sensor type or databases reliability.

Shafer calls $1 - \propto_i$ a degree of trust, where $0 \leq \propto_i \leq 1$ and i is the index of the discunt function to apply to a particular belief value. As it is described in [82], $bel^{\propto_i}(A)$ represents the discount function:

$$bel^{\propto_i}(A) = (1 - \propto_i) bel(A) \tag{15}$$

The resulting combination method uses the following average function for all subsets A of X.

$$bel(A) = \frac{1}{n}\sum_{i=1}^{n} bel^{\propto_i}(A) \tag{16}$$

According to Shafer, this method can be used when all belief functions to be combined are highly conflicting. This can also be used to eliminate the influence of intense conflicts, remaining belief functions only.

4.4 Yager modified Dempster rule

Maybe the most popular alternative to the classical combination rule are the Yager rules (Yager,1987). Yager states that an essential characteristic of a combination method is the ability to update an already combined structure when new information becomes available.

As it is described in (Zhang, 1994), Yager proposes that in many cases a non-associative operator is necessary to perform combinations. For example, the arithmetic average is not by itself associative; it is not possible to add a new measure to an existing average value. However, the average can be updated by adding the new value at the sum of the pre-existing data. This concept is called Quasi-associative operator, and it means that an operator can be decomposed into associative operators. In order to address this characteristic, Yager defines a ground probability assignment denoted by q. The combination of ground probabilities is described as follows:

$$q(A) = \sum_{B \cap C = A} m_1(B)m_2(C) \tag{17}$$

In the Yager rule, $q(A)$ is a combined structure, and it can contain many pieces of evidence represented by a probability assignment m_i.

$$q(A) = \sum m_i(A_i) \tag{18}$$

As Yager describes, the quasi-associative operator allows updating the combined structure $q(A)$. However, the basic probability assignments are not the same as with Dempter's rules. Yager provides the relation between the ground assignments and Dempster's rules.

$$q(\emptyset) \geq 0 \wedge m(\emptyset) = 0 \tag{19}$$

$$m(X) = \frac{q(x)}{1 - q(\emptyset)} \tag{20}$$

$$m(A) = \frac{q(A)}{1 - q(\emptyset)} for A \neq \emptyset, X \tag{21}$$

Chapter 5

Adapting Dempster-Shafer Theory to Spatial Contexts

The Dempster-Shafer Theory provides a general method that allows us to deal with uncertain information. It provides lower and upper bounds of probabilities. Furthermore, it has been already applied to GIS data successfully (Ghayoumian, Ghermezcheshme, Feiznia, & Noroozi, 2005). However, the spatial attributes are usually ignored when this theory is applied in SDSS. In this section, I develop spatiotemporal adaptations and interpretations for this theory. This contribution is compatible with the formal definition of the Dempster-Shafer theory; although, some basic definitions now include spatial characteristics.

The Dempster-Shafer theory was described in the previous section. The central concept of this theory is a mass assignment to each hypothesis. A spatial hypothesis may be that in and around schools we might have some concentration of people. Another might be that in malls we would also have a higher density of people, while in parks, there might be less. If we think that schools will attract more people than malls, then we give a higher mass to the corresponding hypothesis. Having composed hypotheses, and given the mass to them, the certainty and plausibility for a geographical area are obtained from single hypotheses (4) (3). Spatial hypotheses will generally have different values in different places if conditions vary. As a result, we can have multiples versions of the belief. Restricting the analysis to homogeneous areas can solve this issue. I defined a limited number of evaluations places (cells or squares). Defining a grid over the area to be analyzed does this. Each element of the grid will be called evaluation cell, and the elements inside will support the hypotheses evaluation for each cell independently at first (see Figure 15).

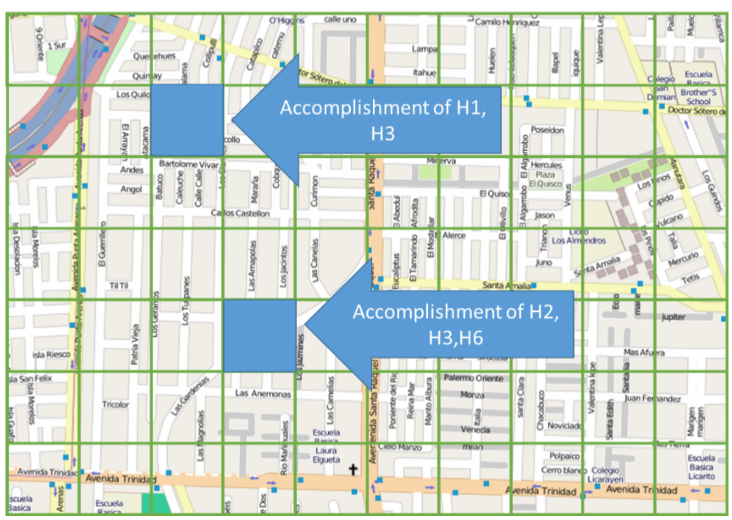


Figure 15 Basic evaluation model, H1 to H6 are hypotheses.

At this point, it is important to mention that for this work, the belief interpretation is different from traditional Dempster-Shafer applications. Furthermore, the belief does not recommend a decision: it recommends locations. In the classical Dempster-Shafer model, the hypotheses with a higher value imply more support for those hypotheses. In this spatial model, the absence of data to support the hypotheses in a certain place implies a zero-mass value for that cell. As a result, the places with more data to support hypotheses will have "composed" mass, and as I defined in the last chapter, the belief and plausibility values of composed hypotheses are higher than the single ones (see Figure 16).

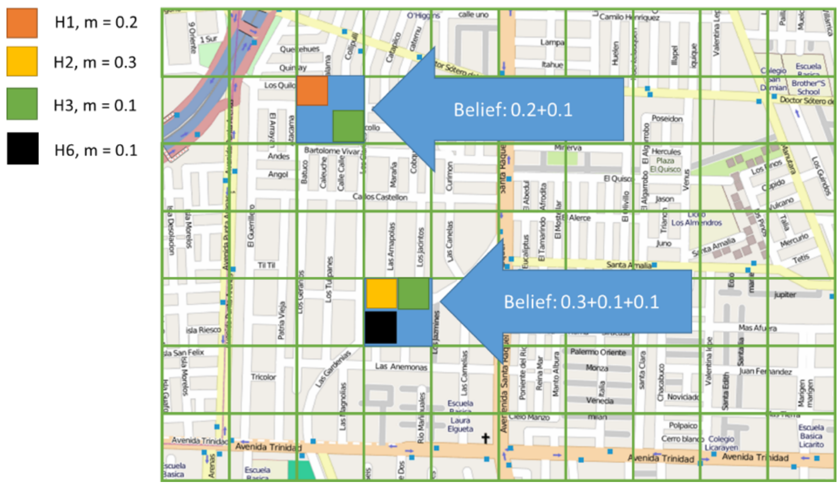


Figure 16 Resulting belief of each cell is calculated combining the mass of all hypotheses that apply to that cell.

This model can be implemented to support decisions that involve static data, for example, the belief of the existence of a particular type of a mineral given some hypotheses (Binaghi, Luzi, Madella, Pergalani, & Rampini, 1998) (Tangestani & Moore, 2002). In this case, static data support each hypothesis. However, there are scenarios based on dynamic data, like the belief of having a certain level of sound, many people, traffic congestion, or the crowd density around buildings depend on time and day. In these scenarios, data is dynamic because it can vary in space and time.

In order to provide a better modeling tool, spatial-temporal modeling must be considered, because I am modeling the mass, which can vary over time. This particularity leads to substantial differences with most spatial-temporal modeling techniques. For instance, a mall will produce different spatial behavior according to the query, and not to the mall object itself. The same effect can be expected on temporal behavior. A mall will produce different mass behavior if we are querying places to put a new shop than places with high burglary risk.

In the following sections, the model spatial-temporal modeling process of mass is presented. Each modeling method is presented as discount functions; this implies that the combination method to be used will be the "Discount and combine method" discussed in 0.

5.1 **Mass in space**

As previously described, objects can be represented in space using several approaches. In our case, mass is not the object; mass is the support that a particular object can provide a hypothesis. However, a spatial mass definition needs a spatial representation.

The spatial representation of mass must be strongly related to the spatial representation of the object(s). However, some of its characteristics must remain associated with the "spatial decision." To simplify the explanation, I am going to use two different examples:

1. "Where can people be?"
2. "Where is there a higher risk of suffering an assault?"

The scenarios for both problems will have different hypotheses sets. However, everyday objects can support these sets. For example:

1. "People can be in bars."
2. "There is more assault risk at bars."

Apparently, the assigned mass for bars must remain intact at the bars' locations. However, the mass value of the surrounding area space can vary. The first statement can be hard to model in space because the mass representation outside a bar is not explicit. However, if we are talking about a "bar area" we can say that people can be within 500 meters of a bar. The second statement is a well-known modeling issue in a criminal analysis, and a commonly used distance to represent the risk propagation in space, which is 200 meters from a "hotspot" (Perry, 2013). Both hypotheses refer to people. However, the decision-making context defines the propagation of the mass in space.

To model the mass propagation in space, I define the space mass function as follows:

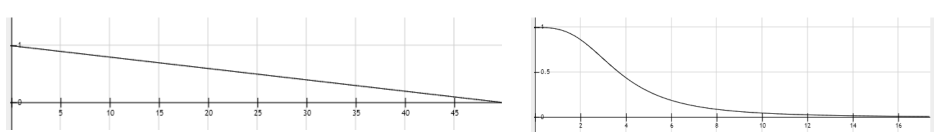
$$ms(d) = \{ \tag{22}$$

$$1 \; if \; d \; = 0 \; (inside)$$

$$F(d) if d > 0 \; (outside)$$

$$\}$$

In (22), d is the distance between the "mass generation element" and the evaluated cell. If the element is contained in the cell, the distance will be 0. Otherwise, the mass will be defined by an F(d) function with [1-0] range. Some alternatives to F(d) are illustrated in Figure 17.



$$F(d) = 1 - \frac{d}{u} \qquad\qquad F(d) = \frac{d^3}{(d^3+u)}$$

Figure 17 Space Mass functions, F(d) examples using u = 50.

In Figure 17, the u parameter refers to the distance associated to the decision-making context, for example, 200 meters for a certain type of criminality.

From the F(d) definition, it is easy to understand that any function with a [0-1] range in a$[0-u]$ domain can be used to model the mass propagation. However, in order to provide some kind of compatibility with existing models, I prefer to use probability distribution functions to define F.

Probability distribution functions meet a set of properties that are different when working with discrete functions than with continuous functions. Like $P\{x = x\} = 0$ in the continuous case. However, for the discrete case:

$$P\{x = x_i\} = F(x_i^-) - F(x_i) = p_i$$

Since they are in ranges, there is always a probability associated with a value. I prefer to use the continuous functions to achieve more accurate results.

A distribution function can be characterized by its cumulative distribution functions (CDF). The CDF return the probability of X <= x. In simple words and using our context, it is the probability that an object can be closer than x. To simplify the modeling, I am going to use a normal distribution function as a default P(x). The CDF associated with normal distribution looks like Figure 18.

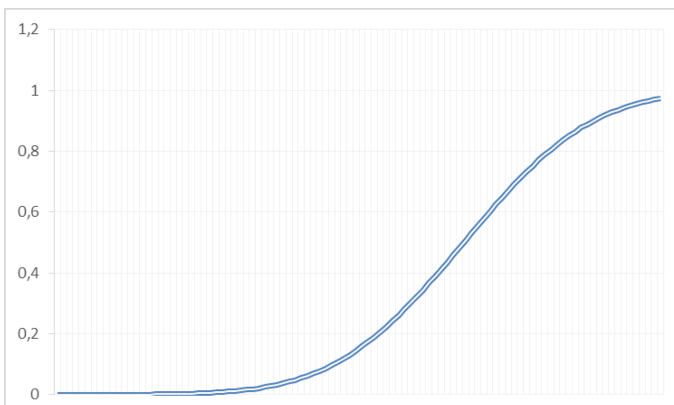


Figure 18. Normal CDF example.

As previously described, the interpretation of CDF is X <= x. However, I need the corresponding probability: "X > x ", formerly named "Complementary CDF (CCDF)", this implies that F(d) based on the normal distribution will be defined as follows:

$$F(d) = 1 - P(d) \tag{23}$$

$$P(d) = \frac{1}{2}\left[1 + erf\left(\frac{d-\mu}{\sigma\sqrt{2}}\right)\right]$$

$$erf(d) = \frac{2}{\sqrt{\pi}}\left(d - \frac{d^3}{3} + \frac{d^5}{10} - \frac{d^7}{42} \ldots\right)$$

In (23) μ and σ are the mean and standard deviation of the distribution function; these values must be selected from the decision-making context. The erf(x) corresponds to the Gaussian error function expanded using Taylor series. In order to be computable, the series length must be fixed. In my case, I use 100 elements.

Finally, the visual behavior of the ms(d) function is described in Figure 19:

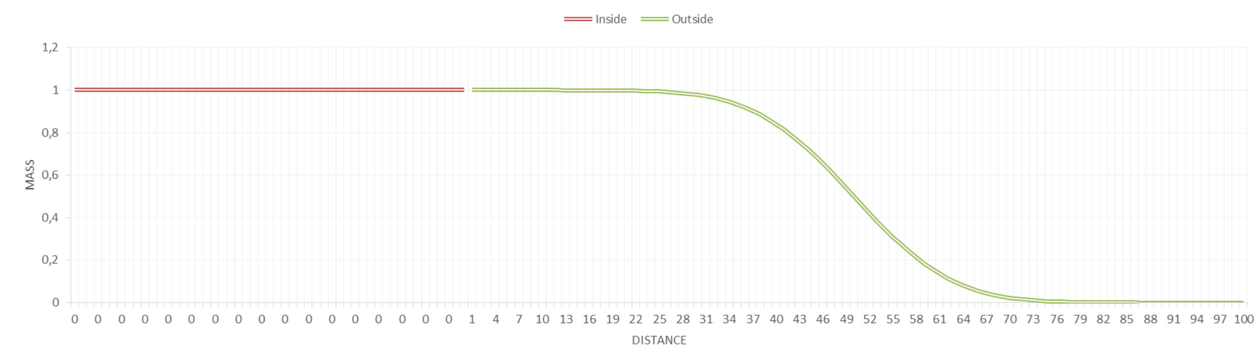


Figure 19 ms(d) example using $\mu = 50$ $\sigma = 10$.

Using the CCDF, it is possible to characterize the decision-making context easily using existing knowledge, such as the criminality context provides $\mu = 200$ and $\sigma = 50$, see Figure 20.

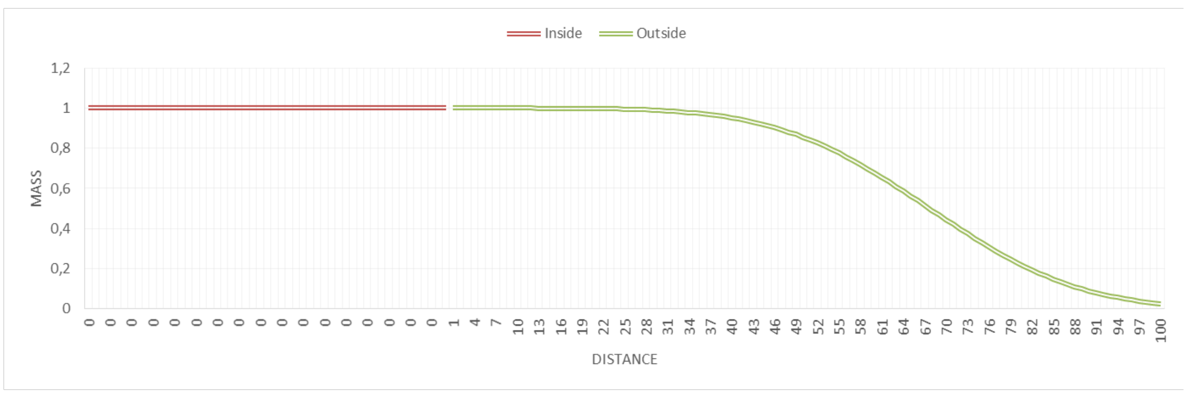


Figure 20 ms(d) example using $\mu = 200$ $\sigma = 50$.

A primary conclusion at this point is that the configuration parameters used by the ms(d) function are directly related to the decision support problem modeling. For example, if the problem is related to transportation, the width of the street can be used as an input to define the parameters.

5.2 Mass function by interval

Geographic information is usually represented discretely, with discrete geometries and deterministic values. However, a variable like temperature over a period of time becomes a random variable with frequency distributions and average values. For instance, we can have historical data about climate conditions in a particular area. This data can be represented building a frequency distribution function. This function is usually approximated by an existing probability distribution function. For example, if we want to know the probability of having 23ºC at a specific location at a precise moment, then we have two alternatives:

querying the data, modeling the probability and returning a result; or pre-modeling an existing distribution function and returning its probability value.

If we have data that can be modeled using probability functions, then we can have an interval probability. This means that it is possible to answer queries like "what is the probability of having between 20ºC and 25ºC".

Assuming F as a probability distribution function, the interval probability will be defined by a query interval as the probability of x >v1 and x<v2 is defined by (Papoulis, 2002):

$$P[v1, v2] = P\{x > v1, x < v2\} = \int_{v1}^{v2} f(x)dx = F(v2) - F(v1) \tag{24}$$

In (24), f(x) is the density probability function and F() is the CDF. Using equation (24), it is possible to estimate the probability of having a variable value inside an interval. This equation is the mathematical base for the definition of the interval mass function (im):

$$im(v1, v2) = F(v2) - F(v1) \tag{25}$$

In (25), v1 and v2 are the lower and upper values of a query range. This type of query can be seen as a useful feature when using raw data as input. However, most data sources already provide mean values.

5.3 Mass function along time

Time-dependent behaviors are included in this work as "temporal relations" between hypotheses objects and the decision-making problem. For example, the probability of finding people in a certain place is different during the day and at night.

The temporal relation is related to intrinsic properties of the object and the decision problem. For instance, if we are modeling a hypothesis using libraries as an attraction point for mass, the schedule and thus the time when it will attract more or less mass will be different than for a subway station or a cinema. This schedule is related to the characteristics of the object. However, if we are looking for a different kind of analysis, the temporal behavior could be very different.

Another example can be the crime seasonality. Some types of crimes are strongly influenced by cyclical patterns, such as the day of the week or even the season (Harries, 1980). To illustrate, during summer, when children are not at school, there may be a peak in petty crimes and burglary.

The temporal behavior of an object can be related to different time scales (Allen, 1991), for example, years, months, weeks, days, hours, etc. Each scale may have different behaviors. Moreover, a library can have a high mass related to the presence of people between 8:00 to 18:00, a low mass on Sunday and higher mass in winter and during university exam periods.

To model multiple temporal-scale mass behavior, I will first define the temporal mass function (tm) as the weight of each scale S.

$$tm(timestamp) = \prod_{s \epsilon S} SF_s(timestamp) \qquad (26)$$

In (26), *S* represents the time scale to be used; the scale functions (*SF*) return values between 0-1; and *timestamp* is the moment to be evaluated. Each scale function is composed by mass values assigned to segments of time. The timescale determines the number of segments and its mass values. For example, if we are modeling people around a subway station, we must consider a weekly schedule. This schedule (or scale) can be represented by assigning highest mass from Monday to Friday and lowest on Saturday and Sunday. We must also consider the various hours of a day in the modeling process assigning mass in the high demand hours and zero mass when it is closed. The resulting evaluation of a timestamp will be the weighting of both scales at same timestamp.

This work will focus on modeling time as discrete elements. Furthermore, I assume that stable features are exposed to sudden events in stepwise constants values. I named those steps as time segments. The time segment length will be defined according to the scale to be used.

In order define the different time scales it is necessary to analyze time-related data that can be extracted from databases. The ISO 8601 (Klyne, 2002) defines the representation of dates and time. Furthermore, those representations are frequently implemented in databases as a combination of data types and procedures. Usually, a single time point is represented using a *timestamp* or a *datetime*, depending on the database management software and design requirements. A period is usually stored as a start and end point in time, or as a start and duration. The information that can be directly extracted through procedures from a single timestamp or date time is shown in Table 5.

Scale	**Values example or range**	**Duration (seconds)**
Millennium	For year 2016, returns 3.	31,605,120,000
Century	For year 2016, returns 21.	3,160,512,000
Decade	For the year 2016, returns 201.	316,051,200
Year	For date 2016/09/20, returns 2016.	31,536,000 or 31,622,400
Quarter of year	For date 2016/09/20, returns 3.	7,776,000 – 7,948,800
Week of year	For date 2016/02/16, returns 7.	604,800
Day of year	Range: 1-365/366	86,400
Month	Range: 1-12	2,419,200 – 2,678,400
Day of month	Range: 1-31	86,400
Day of week	Range: 0-6, (Sunday is 0)	86,400
Hour of day	Range: 0-23	3,600

Minute of hour	Range: 0-59	60
Second of minute	Range: 0-59	1

Table 5. Extractable time scales from databases (Klyne, 2002).

Time representation is divided into two main categories: Date and Time. Each with hierarchal organizations, see Figure 21.

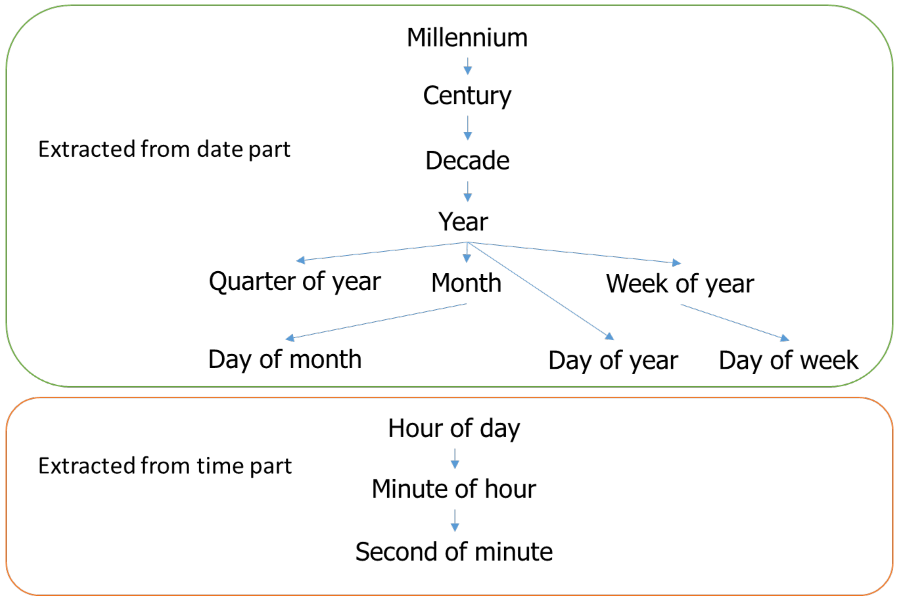


Figure 21. The hierarchal organization of time scales (Klyne, 2002).

In order to explain the modeling method, a *music score metaphor* was designed to represent time segment (duration) and mass magnitude (pitch). The metaphor can also represent a smooth transition between time segments by defining a transition mass function in time. The probability distribution functions are used as scale functions (*SF*), representing each value among the range, see Figure 22.

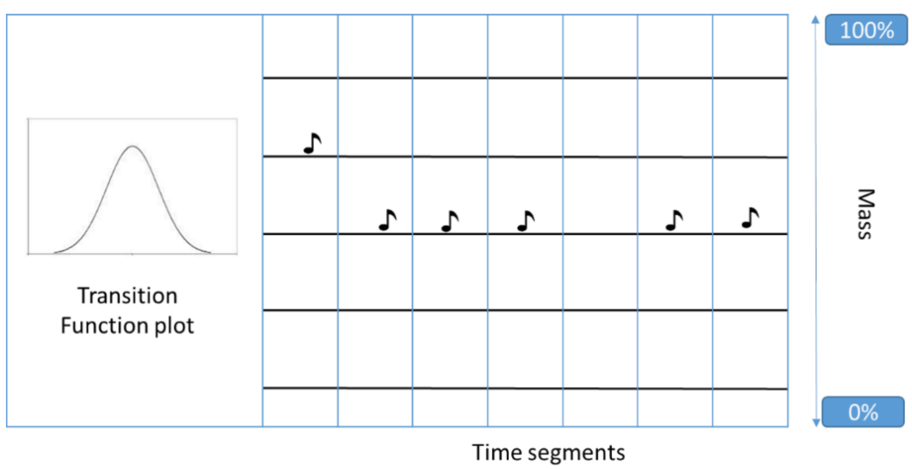


Figure 22 Time mass assignment metaphor for a single scale.

The values assigned to each time segment are aggregated by calculating the highest value (highest sound) present just in the middle of each segment (See Figure 23).

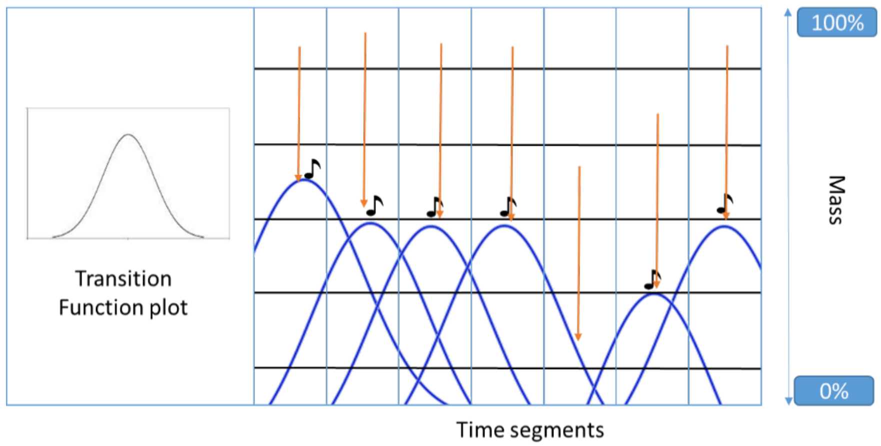


Figure 23 Mass of each time segment for a single scale.

As a starting point for modeling the mass distribution among time, I included only two kinds of probability functions as transition functions. The first one is a uniform probability function and the second one is a normal probability distribution function with the mean value placed on the note segment.

Using the tm function (26), we can query the mass at a specific date and time. As an illustration, 30/06/2015 at 13:00 requires combining at least four different scales: Month of year (06), day of Month (30), day of week (03), hour of day (13).

5.4 Belief and plausibility definition by space, interval, and time

As introduced in the previous section, there is a relation between Belief & Plausibility values and the ms() function (22). The relation can be described by the Belief and Plausibility definitions. We must remember that Belief is the probability that existing evidence supports a hypothesis, and plausibility is the probability that the same hypotheses are compatible with the evidence. The mathematical-spatial representation of these definitions concludes that belief is the sum of all masses entirely contained by the hypotheses, and plausibility is the sum of all intersecting masses. The following function evaluates the mass for a single hypothesis, given a timestamp t an interval evaluation $v1$ and $v2$.

$$Mass(d,t,v1,v2) \quad = \sum_{e\ in\ Elements} \quad ms_e(d) * tm_e(t) * im_e(v1,v2) * Interact_e(hypothesis \tag{27}$$

In (27), the "Elements" is a list of geographical elements that supports a single hypothesis, t is the evaluation timestamp, $v1$ and $v2$ are the interval evaluation of a hypothesis, and the interact functions return the interaction between the element e and the hypothesis.

From section 5.1, we see that objects associated with hypotheses contribute to the mass in areas other than the cell where it is placed, see Figure 24.

Figure 24 Mass distribution in space.

The resulting mass of all elements influencing a cell must be combined into one. However, there are different alternatives to aggregate masses. The first one is using Dempster-Shafer combination rules; this implies considering different masses as different sources. However, the combination rules are meant to be applied to expert criteria. In this case, it can lead to different combinations (see 4.2). The second alternative can use an aggregate operator as average. However, the resulting values will be directly related to the geographical distribution of elements. The third one sums merely the masses of the various elements and then normalizes all the resulting masses in the evaluated area. This alternative is similar to the discount and combination method discussed by Shafer.

By combining the masses using the discount method, I can provide a magnitude perception since it lacks a normalizing factor. This result can be beneficial to model scenarios with a statement as like "people are in shops." I named this combination alternative "Suitability mass."

5.5 Discussion

Based on the previous definition of *ms*, im and *tm* functions it is possible to generate a mass function with three parameters: distance from the source, timestamp, and values interval. However, in most cases, all these parameters are not needed simultaneously, and according to the belief definition, the resulting belief must be low if there is a lack of support for the hypothesis.

Another vital definition to discuss is the belief and plausibility meaning and values in spatial problems. Based on ms functions (22) functions that display a decreasing mass with distance, plausibility and certainty can be defined as follows:

- According to the plausibility definition based on (13), the function can be described as the combination of the masses of all elements available and discounted by (13).
- The certainty definition based on (13) function can be described as the suitability mass using only masses of elements inside the evaluation cell and not discounted (or weighted by 1) in space or time.

Based on (9), I can say that $P_*(T)$ evaluates only the masses of elements inside the evaluation and $P^*(T)$ includes the masses of all elements available. Finally, replacing $P(T)$ by (27) a definition of Certainty, Plausibility can be provided.

$$\begin{gathered} Certainty = Mass(0) \\ Plausibility = Mass(\infty) \\ Certainty \leq Mass(d) \leq Plausibility \end{gathered} \tag{28}$$

Chapter 6

SDSS scenario generation model

In order to provide tools to design scenarios and analyze the outputs of an evaluation, I designed an object-oriented representation of the elements previously described, allowing for specification of the spatiotemporal functions needed for the belief functions method in spatial contexts.

The design differentiates between objects that support the hypotheses (data objects), and objects used to describe decision-making problem characteristics (suitable objects). The problem characteristics object is called suitable objects. These two kinds of objects are used to generate dynamically specific instances of objects that support and characterize a scenario. In order to represent these instances, I created a base object called GeoObject, which is dynamically extended and overloaded by the scenario generation engine. In the following sections, I will describe the Suitable Objects, Data Objects and the dynamic generation of the scenario support data based on a primary structure.

6.1 Suitable objects

A Suitable Object is an analogy to a suitability map in GIS. A suitability map represents the capacity of a region to comply with some requirements or achieve an objective. A Suitable Object is the characterization of the scenario using mass distributions in space and time. These objects must be defined according to expert's knowledge. For example, if we need to evacuate a large area, the solution requires knowing the people concentrations points in all the area. In this case, the Suitable Object must characterize the distribution of mass in space and time of people concentrations. Using this characterization, the processing engine can build a scenario using the supporting data associated with each hypothesis that supports the belief of each Suitable Object.

6.2 Data objects

A Data Object is the instance of a particular type of geographic element on the map. The Data object describes the physical representation of data, and it contains its attributes. For example, the temperature in a region can be represented using a georeferenced polygon and the temperature values or probability function associated with it.

A Data Object cannot have mass or belief values; it is used as support hypotheses. However, a mass is associated with hypotheses that belong to a specific scenario. This relation of mass and Data Objects is designed to implement possible future extensions of the data object characterization like associating a mass to a GeoRDF (or similar standards) (Vatant & Wick, 2006; Lassila, Swick, & others, 1998) (Kolas, 2008) (Klein & Visser, 2004) (Reed, Singh, Lake, Lieberman, & Maron, 2006) (Reed & Lenat, 2002) (McGuinness, Van Harmelen, & others, 2004) (Lopez-Pellicer, Silva, Chaves, Javier Zarazaga-Soria, & Muro-Medrano, 2010) (Rackham, 2008) (Rackham, 2011) (Auer, Lehmann, & Hellmann, 2009).

The concepts of *Suitable* and *Data* Objects can be implemented using two different structures. Suitable objects are present on each cell; however, data objects can be absent, leading to suitable objects with lack of support, thus, not generating mass. In order to allow the scenario generation to work with complete and incomplete data, I will define a single object that can be extended to implement the Suitability Objects. This single object is called GeoObject, and it will be used to create geographical instances based on belief functions.

6.3 GeoObject

The GeoObject is designed to be instanced as a Suitable Object and interact with Data Objects in the same cell. As a result, the GeoObject will be the spatial-temporal characterization of the spatial decision problem, not the answer. The instance of a single GeoObject is composed of:

- A single hypothesis mass.
- The space mass function, ms (22).
- The interval mass function, im (25).
- The temporal mass function, tm (26).
- Model restrictions parameters through interactions with other Data Objects.

These components are defined in the previous chapter. However, at this point, a new concept called Model Restrictions must be introduced.

Model Restrictions complement and correct some mass behaviors resulting from the ms, im and tm functions (see Chapter 5). For instance, if we are generating a scenario about where people can be, and the spatial location is near a famous beach, the ms will propagate the mass from the hypotheses located on the beach into the sea. This spatial propagation is evidently an error since the plausibility of finding people in the water is much lower. Another function, reflecting a much faster decreasing of mass could be used to represent this fact, but it will complicate the modeling process. A more straightforward method is introducing Mass Interactions between objects. A Mass Interaction can be represented as a function that returns the weight resulting from interactions between the Suitable Object and a Data Object: e.g., the weight of the interaction between the Suitable Object “people” and the Data Object “sea” must be almost zero.

The GeoObject is used as a standard structure to dynamically build a class hierarchy representing the hypotheses, which are used to generate the suitability maps. Furthermore, the resulting set of classes is used to evaluate the mass values of each hypothesis on each cell. The GeoObject description using UML can be seen in Figure 25. The hypotheses class framework generation process is described in the following section.

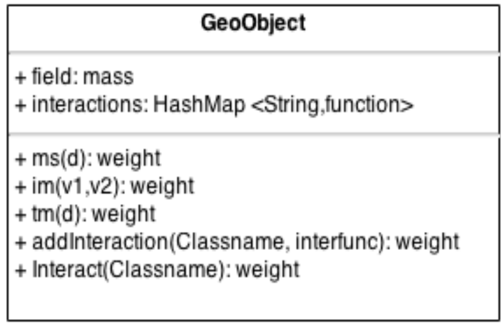


Figure 25. Structure of a GeoObject

6.4 GeoObjects map and suitability map generation

A GeoObject map is a representation of an area. Figure 26 shows two different Data Objects, with a different type, shape, and location. Each Object will have a mass, a ms()(22), im()(25) and tm()(26) methods, and interactions between Suitable Objects and other Data Objects.

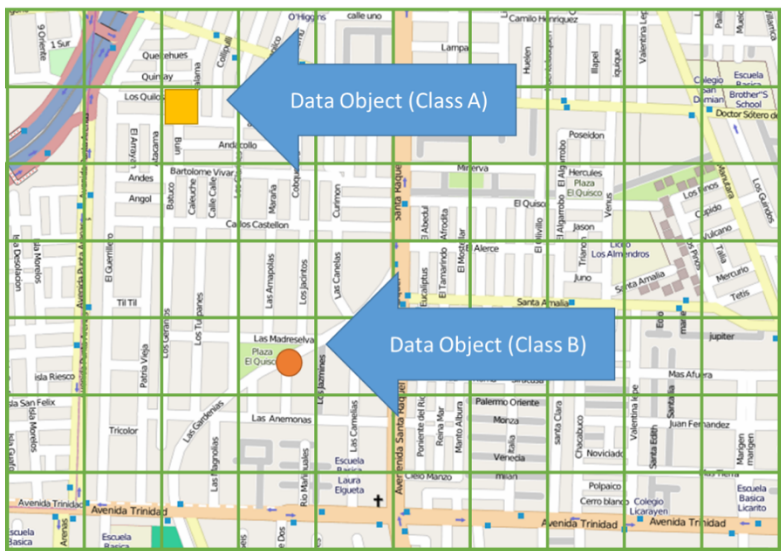


Figure 26. Data objects location example.

Using the GeoObject map, it is possible to evaluate the belief propagation in space and time of each Data Object supporting the decision-making hypotheses using the Dempster-Shafer Theory.

As was previously discussed, a suitability map evaluates the "level" of accomplishment of a requirement achieved by a particular area. This area is typically divided into cells, each one with a particular level of accomplishment. In our case, the suitability map will represent the propagated belief level on each cell (see Figure 27).

Figure 27. Propagated belief vector example representing the influence of Object of class A and class B on the blank cell.

6.5 The algorithm

In this section describes the algorithm to evaluate the belief values on each cell in the evaluated area. The first step is to divide the whole area into a grid. Based on a Quadtree structure (see section 3.3, fig. 13), the grid is divided into quads. However, I will define a minimal quad size, and thus, the whole area will be divided into these min-quads.

The algorithm (not optimized) uses as input the Quadtree leaves in the selected area, the Suitability Objects list, all the Data Objects per quad, the evaluation timestamp, the list of interval evaluations and the combination rules as described in 4.2.

The whole algorithm is similar to a map-and-reduce operation (Dean, 2008). The first part (map) evaluates all the Suitability Objects on each quad generating a list of masses. In turn, each mass is represented by a list of multiple masses coming from different sources. The second part (reduce) combines the multiples sources evaluation using Dempster-Shafer combination rules, and finally evaluates the belief value for each quad.

```
quads = genetareGrid();  //Quadtree leafs of selected area
suit = getSuitabilityObjects();  //suitability objects
temp = getEvaluationTimestamp();  // evaluation timestamp
intervals = getIntervals();  //intervals tuples associated with each suitability obj
combinationMethod = getCombinationMethod();
   dataobjs = getDataObjects(area);
quadrants = Array();   //output
forall the quads as quad do
  quadrants[quad]=Array();
    forall the dataobjs as data do
      mass=1;
      mass = mass*suit.ms(distance(data,quad));  //spatial evaluation
      mass = mass*suit.tm(temp, data);  //temporal evaluation at timestamp temp
      mass = mass*suit.interact(data);  //applying interaction model
      quadrants[quad][type(data)].append(mass);
  end
end
/* Replaces de list of masses in quadrants[quad] into a single belief value using a combination method */
forall the quads as quad do
  forall the quadrants[quad] as data do
   quadrants[quad][data]=combineHypoteses(quadrants[quad][data],combinationMethod);
  end
  quadrants[quad] = DempsterShafferEvaluation(quadrants[quad]);
end
/* values are normalized between 0 -1 */
quadrants=normalize(quadrants);
return quadrants;
```

6.6 Evaluation example

To better explain the whole process and the results, I will use an example case based in the area of the Shibuya Station in Tokyo, Japan, since this is one of the busiest areas of the world, and consists of underground lines, surface, broad avenues, shopping centers, etc. The analysis focuses on evaluating the concentration of people in a certain area.

In the proposed method, we need to define hypotheses, which will assign a specific mass of belief to objects about their potential of attracting people. These objects could be subway stations, shops, and restaurants, among others.

Let us suppose we have some knowledge about the area, and we can use polygons to define areas of different types on the map, like commercial areas, a train station area, etc. A particular location will retrieve a mass of being crowded according to hypotheses assigning mass to this kind of objects. However, there are commercial areas in the train station. Furthermore, most of the polygons overlap, so there are various hypotheses, which apply at the same time for one location. This uncertainty is frequent in GIS applications.

Figure 28. Polygons of different types of areas.

As it is described in the algorithm, the area must be divided into cells. Then Data Objects representing instances of all the polygons should be instantiate, and for each cell, a Suitability Object should be created.

After this, the algorithm (in 6.5) is applied to evaluate the combinations and finally applying the Depster-Shafer analysis method described in this work. As described in Figure 28, we can obtain an interval of values, given by the certainty as the lower bound and plausibility as the upper bound. In Figure 29 and Figure 30 we can see the difference between both extremes.

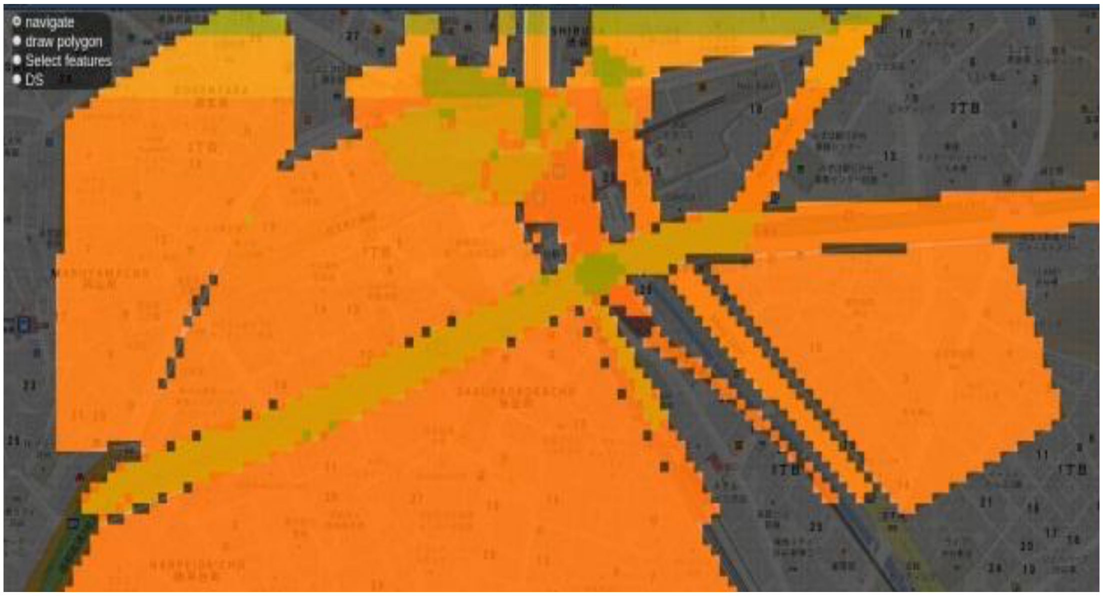


Figure 29. Certainty of each source representation in % scale. Red: 0% - Green: 100%.

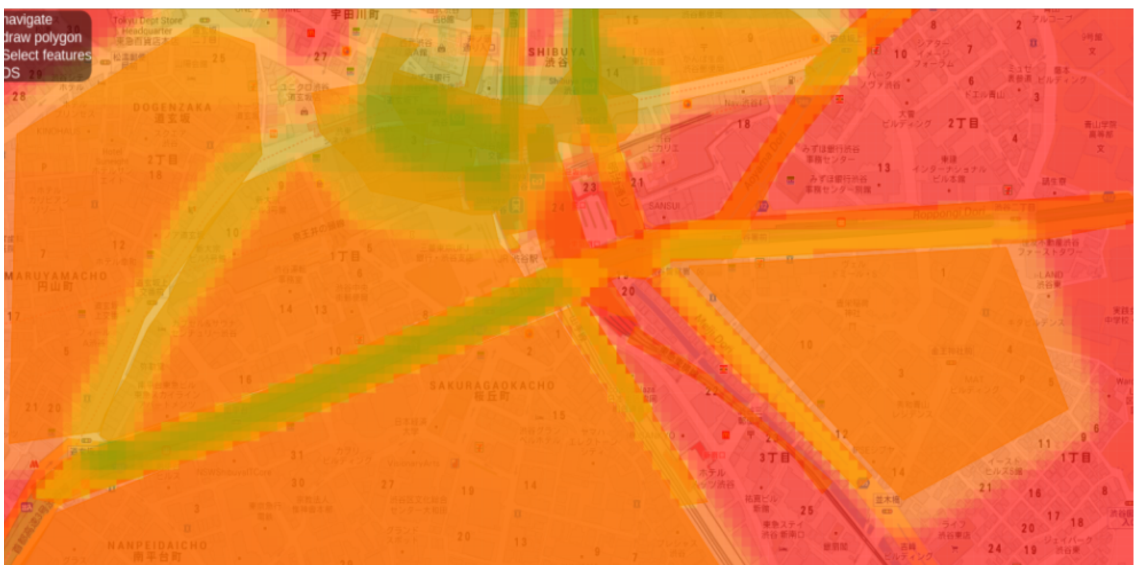


Figure 30. Plausibility of each source representation in % scale. Red: 0% - Green: 100%.

In Figure 29 there are no cells without belief because the probability functions used never goes to 0. Very low certainty values replace the dark areas from Figure 29. From Figure 30 the reader realizes that there is a certainty of finding people in the train station and on the rails). However, this is a wrong conclusion, since the certainty propagates over the map. This issue can be fixed by applying Model Restrictions. In this case the rule must reflect that people cannot be on the railways. Figure 31 shows plausibility after applying the restrictions.

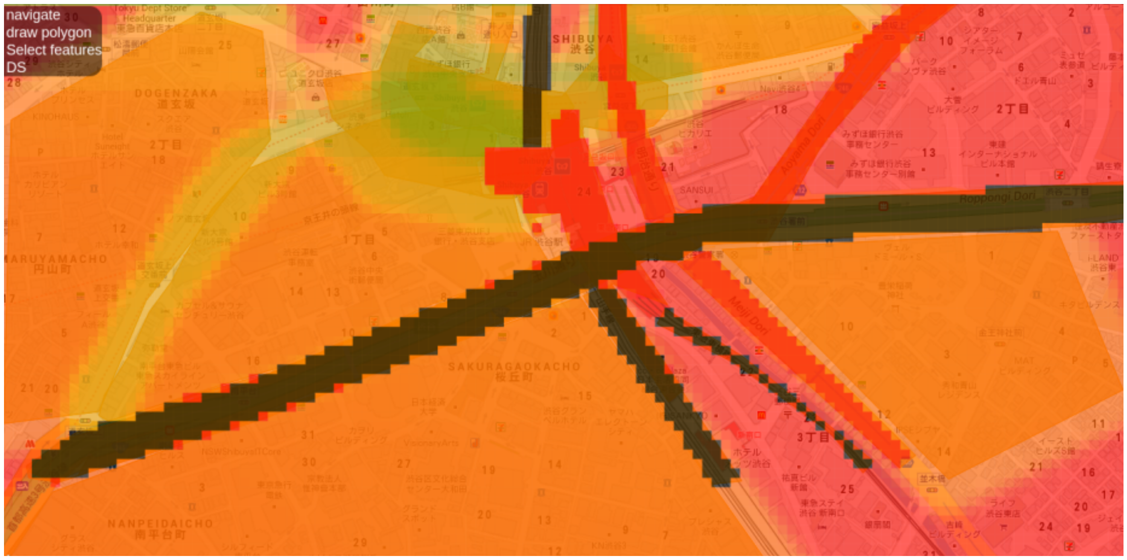


Figure 31. Plausibility of each source representation in % scale. Brown: 0%, Red: near 0%, Green 100%.

The 3D view of the evaluations is shown in Figure 32

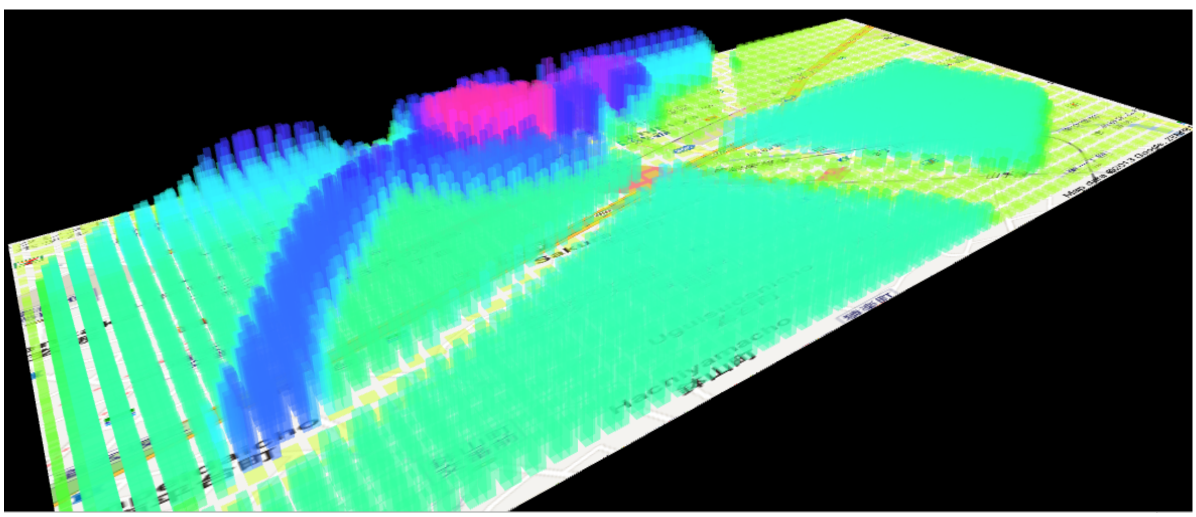


Figure 32. Plausibility of each source representation. Max height equals to 100% of belief.

In this example, I have presented a process that allows the use of geo-referenced data from various sources, which have different degrees of belief to be combined in order to answer challenging questions, which include probabilities and uncertainty. Here I presented the example regarding crowded areas. However, we can see that this method can also be applied in many scenarios, e.g., finding a certain vegetation or animal species in a geographical region given its climate and geomorphology or calculating the outbreak of a particular disease given the data of occurrences in nearby regions and general conditions of the weather.

In order to achieve the objectives of this work, the following chapter focuses on a flexible and versatile way to model scenarios using the method.

6.7 Discussion

The suitability function always returns a value between belief and plausibility. Furthermore, this property makes a suitability value compared with other scenarios only if the ms property is the same. A simple way to ensure comparable scenarios is using the same Suitability Object. However, the suitability object is directly related to the DSS problem, so the suitability object will be constant, and the scenarios are always comparable.

Finally, the definition of suitability is based in the definition of $P(T)$ described in (9), and it also asserts the condition $P_*(T) \leq P(T) \leq P^*(T)$, the reason is described in section 4.4. As a result, the $P(T)$ depends on distances, time, probability intervals and model restrictions. The difference between $P_*(T)$ and $P^*(T)$ is given by the suitable object characteristics in space. For example, $P^*(T)$ can be defined using $ms(d) = 1$ and $P_*(T)$ with $ms(d) = 0$. However, in order to avoid zero results for $P_*(T)$ we should consider the distance d as the distance between the quad and the data object.

Chapter 7

Scenario Generation Language

In the previous chapters, I described a process for generating suitability maps with uncertain data generated by various sources. To take advantage of that process, a decision maker must have the possibility of generating them in a flexible and versatile way to compare the outcome when different hypotheses are used.

Currently, the most used language to analyze GIS data is SQL. However, it is difficult to incorporate the necessary spatial concepts, and it is not designed to generate alternative scenarios —all crucial when dealing with spatial data in SDSS.

In order to simplify the alternative generation, modification and impact evaluation of various hypotheses on the resulting scenarios, I designed a simple Scenario Generation Language (SGL). SQL inspires the SGL design. However, it is not designed to query data; it is designed to generate a scenario based on expert knowledge, empirical data, and existing environmental models. The scenario generation is a complicated process; a single SGL statement will produce processing tasks. These tasks will require spatial predicates (Güting, 1994). Finally, these predicates are computed using a spatial database.

First, I am going to present an informal description of the language, showing some basic examples and some grammar. The next section describes the language in a formal method.

Using the SGL, I target to make scenarios based on queries like the followings:

1. Which are the areas with most concentration of people if we assume that there will be people at cinemas, workplaces and around schools
2. What is the point of high demand for public transportation if we assume that people will require it near shopping areas, schools, and hospitals?
3. What is the burglary risk in a specific area if we assume that it is more likely to occur where there have been some cases in the near past and in the neighborhood?

The examples and the kind of queries I want to support have an **object of analysis**: 1- People, 2- Transportation demand, and 3- Burglary risk. These objects must be modeled based on the Suitability Object definition. The combination rules determine the generated scenario values applied. In this case I use the discount and combine method (described in 4.3) to generate scenarios. Furthermore, as most applications require a suitability map to represent an analysis result, I am going to use the keyword “suitable” in SGL to define the use of the discount and combine method. However, other combination rules can be used.

In order specify the **hypotheses** needed to find an answer to the queries, Data Objects must be used as support data. For 1, people use to concentrate in shops, cinemas and around schools. For 2, we can use shops, cinemas, schools and bus stops. For 3, places where a repetitive burglary has been taken place are more dangerous. In the Dempster-Shafer Theory, each of the statement must be weighted by a mass. Specific properties of the data objects can be included as a filter, like including only shops with a certain capacity for attracting mass.

Finally, the queries must include some logical restrictions and consistent behaviors (**Model Interactions**): for example, restricting burglary risk areas to public areas, or restricting the subway station crowds at peak hours for people concentration scenarios.

The main components of an SGL query are the following:

1. Object of Analysis
2. Hypotheses
3. Model Interactions

Moreover, an SGL sentence will look like:

<Object of Analysis> **hypotheses** <Hypotheses> **model** <Model Interactions>

When the Object of Analysis component is to generate the suitability map for people, the user should use the following sentence:

“suitable **@people”**

In the **Hypothesis** component, the expert can express his/her knowledge using multiple hypotheses, which are combined using Dempster-Shafer combination rules. Furthermore, these complex scenarios are designed by the expert without requiring any GIS expertise. In the following example, the expert is looking for persons; then one hypothesis may be “people use to be in cinemas with a mass of 20%” or “people are in schools or workplaces with a mass of 30%” and applying a filter for cinemas with more than 200 people capacity:

“hypothesis {@cinema}20% {@school, @workplace}30% **where @cinema.capacity > 200**”

Finally, the model characteristics statement is designed to represent real-world interactions between the elements in the Data Object set. For example, a high density of people is not expected in a lake or sea. On the contrary, we expect a high density of people inside a sports stadium at an event. These kind of interactions complements the behavior statement by adding environmental rules. This rule can be expressed using values in an interval. For example, if we are generating a scenario for persons, we add in the model statement a 0% value for lake areas and 100% value for stadiums at 1pm and 5pm. This value can decrease or increase the certainty level in the indicated area:

“Model @stadium{day{13:100},day{17:100}} @lake{0}”

A full Scenario definition will then look like the following:

“suitable @persons hypothesis {@cinema}20% {@school,@workplace}30% {@shops}20% where @shops.capacity > 20 Model @stadium{100} @lake{0} @school{day{8:100}, day{14:100}}”

The SGL language explained above has been designed to be extended in order to allow users to specify other types of maps, which could be generated (not only suitability maps based on certainty but belief or plausibility maps) and to incorporate other filters like defining a polygon where the map should be shown, and depending on the scenario and data available, specify temporal attributes for the data, which should be used to generate the map.

In the following sections, I describe the grammar of SGL and give further details about each component of SGL.

7.1 Grammar

The full grammar starts with productions.

<S> ::= <QMAP> <HYPOTHESES> |<QMAP> <HYPOTHESES> <S2>

<S2>::= <MODEL> | <MODEL><S3>|<S3>

<S3>::= <OPER>|<OPER><S4>|<S4>

<S4>::= <INTER>|<INTER><S5>|<S5>

<S5>::= <SAVE>

According to the grammar, an SGL sentence is composed by the analysis part (<QMAP>) followed by the hypothesis (<HYPOTHESES>) and finally the different options of model restrictions (<S2>).

The <QMAP> non-terminal represents the analysis definition component; the <HYPOTHESES> represents the expert's knowledge using the Dempster-Shafer framework. <MODEL> represents real-world interactions between the elements in the data source. <SAVE> allows us to save the result of an SGL evaluation (scenarios), <OPER> permits combining previously saved scenarios. Finally, <INTER> allows intersecting the resulting scenario with a geometry returning the scenario evaluated with a specific shape. <S2>, <S3>,<S4> and <S5> non-terminals are used to make <MODEL>, <OPER>, <INTER> and <SAVE> optional.

7.2 Analysis definition

Currently, the SGL allows generating scenarios based on pure belief values or based on a discount and combine method. However, SGL can be extended to represent more complex scenarios, for example, based on a custom combination method. These characteristics can be specified, allowing other words to represent different combination methods by instantiation of the <METHOD> non-terminal; however, this will be explored in future works. The <string> non-terminal is normally instantiated by objects for which experts have some knowledge.

<QMAP> ::= <MAPS>

<MAPS> ::= <METHOD> <CLASS>

< METHOD> ::= <STRING>

<CLASS> ::= "@"<STRING>

<METHOD> ::= suitable | . . .

7.2.1 Hypothesis definitions

The hypotheses definitions allow the adding of lists of hypotheses. Each hypothesis can also be a list of Data Objects. This allows SGL to specify classical "where" filters like x > y , x< y , x =y , x!=y. It also provides spatial conditions like "inside" and "outside", both are evaluated using the ST_Touches runtime defined by the OCG.

<HYPOTHESES>::= "hypotheses" <HYPOTHESESLIST> | "hypotheses" <HYPOTHESESLIST> <WHERE>

<HYPOTHESESLIST> ::= <HIPO> | <HYPOTHESESLIST>

<HIPO> ::= "{" <CLLIST> "}"<NUM>

<CLLIST> ::= <CLASS> | <CLASS> "," <CLLIST>

<CLASS> ::= "@"<STRING>

<WHERE>::=<STRING> <DSCONDITION> | <ATTRCONDITION> | <GEOCONDITION> | <CONJUNCTION> <WHERE>

<CONJUNCTION> ::= <AND> | <OR>

<DSCONDITION> ::= <COMP> <VAL>

<ATTRCONDITION>::= <CLASS> "." <ATTRIBUTE> <COMP> <VAL>

<GEOCONDITION> ::= <GEOOPERATOR> <GEOMETRY>

<GEOOPERATOR> ::= "inside" | "outside"

<ATTRIBUTE>::= <STRING>

<GEOMETRY> ::= // Well Known Text (WKT) geometry format

<COMP> ::= <GT> | <LT> | <EQ> | <DIFFERENT>

<GT> ::= ">"

<LT> ::= "<"

<EQ> ::= "="

<DIFFERENT> ::= "!"

<VAL> ::= <NUMBER>|<STRING>

7.2.2 Model constraints

The non-terminal <MODEL> must be able to represent two types of interactions:

a) Spatial interactions between a data object and the suitable class.

b) Temporal behavior of the data class given a suitable class.

These two specifications are used to split the model constraints grammar into a spatial and a temporal production. The spatial production can produce simple class-weight interactions. However, the temporal production is designed to produce four types of time-interval specifications on different time scales. The scales considered are a 24-hour range, a Weekly range (Monday to Sunday), a Monthly range (1-31 days) and finally a year scale (January – December:1-12). If a temporal production is used, an evaluation timestamp must be provided, for example, showing two spatial models (multiplying mass by 150% in the stadium area, multiplying mass by 0 in lakes) and a temporal model (at cinemas, all the mass at 19 hours) and of course, we should provide a day and time for evaluating the scenario at a specific moment.

Model @stadiumarea{150} @lake{0} @cinema{day{19:100}} at '2014/08/09 19:00'"

The next example specifies that all the stadiums in the area will increase the resulting mass by 200% on Sundays at 2pm. Also, the primary streets will always duplicate the mass and the evaluation moment is 2014/08/09 14:00.

Model @stadium{week{0:100},day{14:200}} @primarystreet{200} at '2014/08/09 14:00'

The grammar for producing the Model part of a SGL statement is the following

<MODEL> ::= "model" <MODEL2> | "model" <MODEL2> at <TIMESAMP>

<MODEL2> ::= <SPATIAL> | <TEMPORAL> | <MODEL2> <SPATIAL> | <MODEL2> <TEMPORAL>

< SPATIAL > ::= <CLASS> ”{“ <NUMBER> ”}”

<TEMPORAL> ::= <CLASS> <TEMPORAL2>

<TEMPORAL2> ::= <DAY> | <WEEK> | <MONTH> | <YEAR>

<DAY> ::= DAY < CLLIST > ”{“ <NUMINT> “:” < NUMBER > ”}”

< WEEK > ::= WEEK < CLLIST > ”{“ <NUMINT> “:” < NUMBER > ”}”

< MONTH > ::= MONTH < CLLIST > ”{“ <NUMINT> “:” < NUMBER > ”}”

< YEAR > ::= YEAR < CLLIST > ”{“ <NUMINT> “:” < NUMBER > ”}”

<NUMINT> ::= < NUMBER > | < NUMBER > “-“ <NUMINT>

<TIMESAMP> ::= <STRING>

7.2.3 **Saving**

In order to provide the functionalities needed to combine multiples scenarios, two types of saving operations are defined: expression and evaluation.

Saving an expression in SGL can be compared to creating a *VIEW* in SQL: each time the scenario is used, a re-evaluation must be done. The evaluation depends on the Data-Objects existing in the sources at the evaluation. The results can change over time.

Saving an evaluation can be compared to making an “INSERT INTO table SELECT …” statement in SQL. The resulting evaluation of an area is statically stored. It can be used directly without a new evaluation process. The results are independent of the existing data, and they will be no change over time. It cannot be used in other areas.

Both saving alternatives could be used for different purposes. For example, an expression could be used as an ‘evaluation template’ on similar decisions problems. On the other hand, previous evaluations could be used to compare the scenario changes given different data sources. To provide both, I added two SGL expressions called soft-save and hard-save.

In a soft-save, a change in data will affect the resulting evaluation. A hard-save is the implementation of saving an SGL expression.

Following the same lines, a hard-save will save a copy of the evaluation result, implementing the second alternative.

The extension to the grammar is simple:

<SAVE> ::= <SAVETYPE> <SAVENAME>

<SAVETYPE> ::= “hardsave” | “softsave”

<SAVENAME> ::= <STRING>

7.2.4 **Operator syntax**

An operation is an independent statement, is not meant to generate a scenario, its designed to combine previously saved scenarios.

The syntax is simple; it requires the operator name (with or without arguments) and two previously saved scenarios.

<OPER> ::= <OPERPART> | <OPERPART> <SAVE>

<OPERPART> ::= <OPERATOR> <OPERNAME>"{"<OPARGS> "}" <SCNNAME> < SCNNAME >

<OPERATOR> ::= "operator"

<OPERNAME> ::= "sum" | "sub" | "avg" | "owa" | "owaasc" | "owadesc"

<OPARGS> ::= <NUM> | <NUM> "," <OPARGS> | ""

< SCNNAME > ::= <STRING>

The <OPERNAME> non-terminal can be dynamically implemented; this implies that the list of available operators will be not defined by the grammar.

7.2.5 Intersection syntax

To propose a general method to focus the evaluation into geometries, I use the "intersection" definition of the ROSE algebra (described in Chapter 3) between a set of areas (scn) and a geometry. The "intersection" operation in SGL is the following:

- $\forall scn \in scenarios. \forall wkt \in GEOMETRIES.$
  - $scn * wkt \rightarrow scn$ intersection

In simple words, a WKT is a description of the geometry; in this case, a polyline (route). By applying the intersection operation to an SGL expression result will trigger a re-evaluation of the SGL including only the intersected cells. This operator can be used to restrict the evaluation inside a WKT geometry.

<INTER> ::= <INTERSECTION> < WKT >

<INTERSECTION> ::= "intersection"

< WKT > ::= "Well Known Text Format."

For example, the following INTER expression will constrain the analysis to a specific square area in Santiago, Chile.

intersection "POLYGON((-70.74783325195312 -33.421358879825746,-70.67230224609375 -33.421358879825746,-70.67230224609375 -33.47635938775272,-70.74783325195312 -33.47635938775272,-70.74783325195312 -33.421358879825746))"

7.3 SGL examples

In this section, some SGL examples are explored, increasing the number of features used in each.

7.3.1 Example 1: People possible concentration points

This example aims to evaluate where people can be concentrated at "2014/08/09 19:00" based on data available on public databases as OpenStreetMap.

"suitable @people hypothesis {@cinema}20% {@school,@workplace}30% Model @stadiumarea{150} @lake{0} @cinema{day{19:100}} at '2014/08/09 19:00'"

In this SGL example, @people is the suitability object. The data objects are @cinema, @school, @workplace, @stadiumarea, and @lake. There are two different hypotheses: people are in @cinemas with 20% of mass; people are in @school and @workplace with 30% of the mass. If there is a stadium in the evaluation locations, the belief will be increased by 50% in that area. If there is a lake in the evaluation location, the belief will be decreased to 0%. Also, the cinemas have a temporal behavior: the maximum value of belief will be accomplished each day at 19:00 hours.

7.3.2 **Example 2: Burglary risk**

This example analyzes a burglary risk at "2014/08/09 14:00" based on data available on databases as OpenStreetMap and burglary history data, which may be provided by police or other security institutions.

"suitable @burglary hypothesis {@busstops}20% {@school,@workplace}30% {@burglaryevents where inside "An area in WKT" }30% Model @stadium{week{0:200},day{14:200}} @primarystreet{200} at '2014/08/09 14:00' " intersection "POLYGON((-33.221 -78.3322, -33.231 -78.3312, -33.231 -78.3311 , -33.221 -78.3322))"

In this case, the burglary object must be carefully configured; the propagation distance parameters must have been setup previously. The data objects are bus stops with 20% of mass, schools, and workplaces together with 30%, and of course, the burglary history data with 30%. Furthermore, the burglary data is restricted to a certain area. The model part increases by 100% the risk at stadiums as 14:00 on Sundays. It also increases by 100% on primary streets. Finally, the evaluation is restricted to the same area as the burglary events.

7.3.3 **Example 3: Burglary risk in a street**

Continuing with the previous example, we evaluate the burglary risk at "2014/08/09 14:00" in a single street:

"suitable @burglary hypothesis {@busstops}20% {@school,@workplace}30% {@burglaryevents where inside "POLYGON((-77.2 -33.1 , -77.2 -33.2, -77.3 -33.2, -77.2 -33.1))" }30% Model @stadium{week{0:100},day{14:100}} at '2014/08/09 14:00' intersection "POLYLINE(-33.221 -78.3322, -33.231 -78.3312, -33.231 -78.3311)"

In this case, the intersection operation allows for restricting the evaluation to the cells (or squares) that intersect with the specified geometry; in this case a polyline representing a specific street.

7.3.4 **Example 4: Path evaluation and saving**

The following example is similar to the people concentration example. However, the hypotheses are different, adding the saving feature to the expression. The example also shows the results for a path geometry instead of an area, which could correspond to a particular street.

"suitable @people hypothesis {@shops}20% {@busstop}20% {@amenity}30% Model @shops{day{9:50}, day{11:200}, day{17:200}} at '2014/08/09 14:00' intersection "POLYLINE(-33.221 -78.3322, -33.231 -78.3312, -33.231 -78.3311)" hardsave mypath"

In this SGL example, @people is the suitability object. The data objects are bus stops with 20% of mass and amenity places with 30%. The model part decreases the possibility by 50% in shops at 9 am, and it increases by 100% at 14:00 and 17:00. It evaluates the expression at '2014/08/09 14:00' inside a specific geometry (polyline) and saves the result as "mypath".

7.4 Discussion

The SGL is a language designed to specify scenarios, which will associate the Suitability Object of a location with the spatial-temporal distributions of belief. It allows specifying multiple hypotheses with mass support (or expert knowledge). Finally, it provides a model restriction, temporal characteristics, the interactions between the suitability object and data objects, saving methods, and basic GIS operations at intersections. It can be implemented using existing geodatabases. However, the SGL interpreter must translate some parts of the expression into spatial predicates to provide compatibility. In the following Chapters, I use a first version of the SGL interpreter with different geodatabases to test the DSS capabilities in different scenarios.

The following chapters describes some analysis methods developed to support three easy to understand spatial decision-making problems using the proposed method. The first one touches upon the scenario of crime forecasting. It consists of trying to predict the places with high risk of burglaries, assaults, or other types of crimes that usually occur in public areas. The applications of crime forecasting are not only restricted to police patrol planning, but also needed in logistics and transportation.

The second application deals with the combination of multiples scenarios, following an example of evaluating the evacuation possibilities of a densely populated area in case of the occurrence of a tsunami.

Finally, the third example focuses on the public transportation network planning problem, supporting one of the hard questions of the area: how to estimate the transportation demand with incomplete transportation data.

Chapter 8

Urban crime prediction and police patrolling optimization

This chapter presents a real case scenario in which the theory and algorithms presented in the previous chapters were used to implement a tool for predicting Crime in a big city. Here, the Dempster-Schafer theory is used to calculate the suitability (based on belief) that a crime may occur in a particular place, during a specific window of time.

Urban crime prediction has been an object of study by police and experts in the area. Several police institutions around the world have criminal analysis units in charge of planning police patrolling to prevent crimes.

Criminologists and police focus their efforts on reducing the number of persistent offenders by increasing the risk awareness. The most common preventive tactic is patrolling. However, patrolling requires a significant amount of resources including time. Identifying the most effective times and places to patrol can make a substantial difference on the crime reduction effects.

To find these times and places, crime analysis methods and prediction algorithms have been developed and increased over twenty years, mostly as a result of partnerships between researchers and police institutions. Moreover, there has been an increasing development of IT in police institutions. In particular, police departments now typically have systems to record the occurrence of each crime, including information about the time, type of crime and its geographic coordinates. This information can also be used to predict future occurrences of crimes. In fact, there are commercial solutions like PredPol (Mohler, 2011) that use actual police data to predict crimes.

Geographic Information Systems have been used to display “hot spots” maps where crimes have occurred in order to analyze relationships between crimes and social and physical environment (Bogomolov, 2014).

However, most applications examine criminal activities and related factors that have already occurred. These retrospective analyses are useful, but the true objective of crime mapping is to produce early warnings across time and space, reporting these results to police crime prevention units.

Attempts to develop predictive models of crime have increased (Groff, 2002), and many of these efforts are in their early stages. Predictive algorithms have been useful for identifying areas to focus preventive interventions. A frequent critic of preventive interventions is that the crime can be displaced (in space or time), thus making interventions ineffective. On the other hand, some studies demonstrate that not 100% of crimes are displaced thus reducing the amount of crime.

Experts in crime prevention agree that crime concentrates in select places called "hotspots" (Sherman, 1989) (Harries, 1999) (Eck, 2001) (Weisburd, 2006) (Ratcliffe, 2006) and there is a term called ‘‘environmental backcloth’’ (Groff, 2007) that defines tendencies of crime concentrations according to the structure and features of certain areas. The combination of different factors can explain the presence or absence of criminal activity by defining opportune or inopportune places by their characteristics. This explains why opportunities for crime are not equally distributed across space and time, and that is why there are useful modeling opportunities for crime forecasting.

Mapping applications in police agencies has been successfully applied in operational policing. However, these applications are restricted to simple density maps based on retrospective analysis of crime events (Groff, 2002). The effect of this approach is taking reactive actions to past events, without considering risk factors or crime opportunities in the future. This approach is based on the hypothesis that a crime will occur in the same place that it did in the past (Kennedy, 2009) (Johnson, 2008).

In (Caplan, 2010) the authors propose that risk terrain modeling should integrate the data of previous events with social, physical and behavioral factors. They suggest that these variables should be combined to reveal a consistent pattern that gives better opportunities to crime.

Usually, risk terrain modeling techniques use approaches similar to raster data models, dividing the layers into grids with equally sized cells (Groff , 2001). According to (Caplan, 2010) the technical approach to develop a risk terrain modeling method can be done by identifying the influencing factors, through meta-analysis or other empirical methods, literature review, professional experience, and practitioner knowledge. In general, the resulting factors included in the model represent a weight that can be applied in presence, absence or intensify other weights depending on the implemented model. Finally, for each cell, the combination of resulting weight generates a raster risk terrain map. The higher the value, the greater the likelihood a crime will be committed in that location in the future.

To evaluate forecasting performance using any forecasting algorithm, we can use standardized metrics. In (Wang, 2012) the authors present two metrics which measure the performance at a particular time: Higher risk percentage (HRP) and True Incident Percentage (TIP).

- HRP represents the percentage of the whole evaluated area, which is considered of high-risk by the model (29). More precisely, this is the percentage of the area that surpasses a threshold value.
- TIP represents the percentage of incidents that actually occurred that took place in the high-risk areas (30).

Both measures must be computed with the same data set and threshold.

$$HRP_{\delta} = \frac{\left|\{s_i | p\left(inci_{s_i,t} = 1\right) > \ \delta\right|}{|\{\ s_i\}|} \tag{29}$$

$$TIP_{\delta} = \frac{|\{inci_{s,t} = 1\ |s_i \subset |\{\ s_i | p\left(inci_{s_i,t} = 1\right) > \ \delta\}\}|}{|\{inci_{s,t} = 1\}|} \tag{30}$$

In (29) and (30) s is the set of locations (cells) for the whole area, δ is a threshold and $p\left(inci_{s_i,t} = 1\right)$ is the probability that an incident will occur in the location s_i at time *t*. As an example, Figure 33 shows the comparison between three different methods: S-T GAM, Spatial GLM and hotspot applied to predict the probability of criminal incidents in Charlottesville from March 2004 to February 2005 using monthly predictions (Wang, 2012). The curve represents the TIP against the HRP values when δ varies from a value for which 0% of the total area is considered a high risk, until a value for which 100% of the area is considered a high risk. The shapes of the curves are stepwise since the covered areas increase or decrease by cells, not continuously. Ideally, many incidents will occur within the high-risk predicted area. Therefore, the curve from a good method should be close to the upper left corner. According to this, S-T GAM is currently one of the best methods.

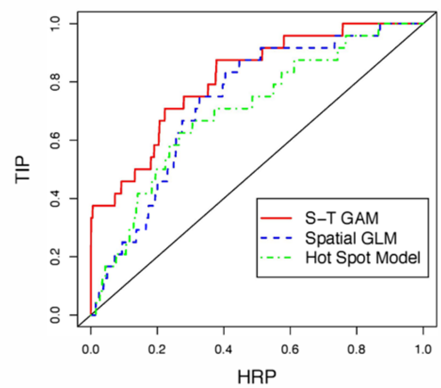


Figure 33. Comparison between S-T GAM and spatial GLM (Wang, 2012)

Crime, as a human activity, is always changing, and developing new behaviors and patterns for specific locations. However, prediction algorithms are based on specific data for a certain place and time (like street widths, shop locations, bus stops, historical data on recent past crimes), which worsens the predicting performance if used for another place or time. Therefore, a forecasting method should include new factors and parameters when used in a different time and/or place scenario. These factors can be specified given a certain location and recent incidents. In this work, I use the Dempster-Shafer theory, allowing users to specify new factors and information providing customized risk terrain modeling for each scenario, providing better forecasting results.

8.1 Dempster-Shafer risk terrain modeling

Using DST, the resulting value will be interpreted as a belief of risk, and it can vary from certainty to plausibility values. The goal of this work is to generate risk maps for specific time and space locations since this is useful information for the police. To accomplish this goal, we must define a set of hypotheses, weights and their variation through time and space.

Based on the police-oriented policing methodology (Eck, 1987) to develop a risk map, we must define the offenders, victims, time and locations of street robbery. According to (Center, s.f.) they are defined as follows:

- Offenders: "The offenders appear to be in their late teens, near 20 years old and mostly males"
- Victims: "Victims who appear to have the money or other valuables, young adults using ATMs alone at night or under the influence of alcohol, victims who

seem unaware of their immediate surroundings when using their cell-phone, or using earphones, etc.".

- Time: Most robberies tend to occur at night; however, some peaks can be explained by the victim's patterns. For example, older adults usually walk on the streets in the morning. Also, those 17 or younger, are attacked between 15:00 and 18:00, which is the school dismissal timeframe. Young adults on paydays during evenings, near bars. Finally, street robbery often occurred on weekends with the mentioned patterns, however during weekdays (Monday-Friday), they occurred patterned with social functions that attract many potential victims to the same places.
- Locations: downtown areas are more crime-ridden than suburbs or residential ones. Frequent robbery locations are parking lots, garages, parks, fields, playgrounds and areas near public transportation in large cities.

The problem-oriented policing methodology also specifies a triangle analysis tool that helps define the relationship between the crime components of a specific type of crime. For street robbery, The Pop center (Center, s.f.) has already developed the following triangle analysis (see Figure 34):

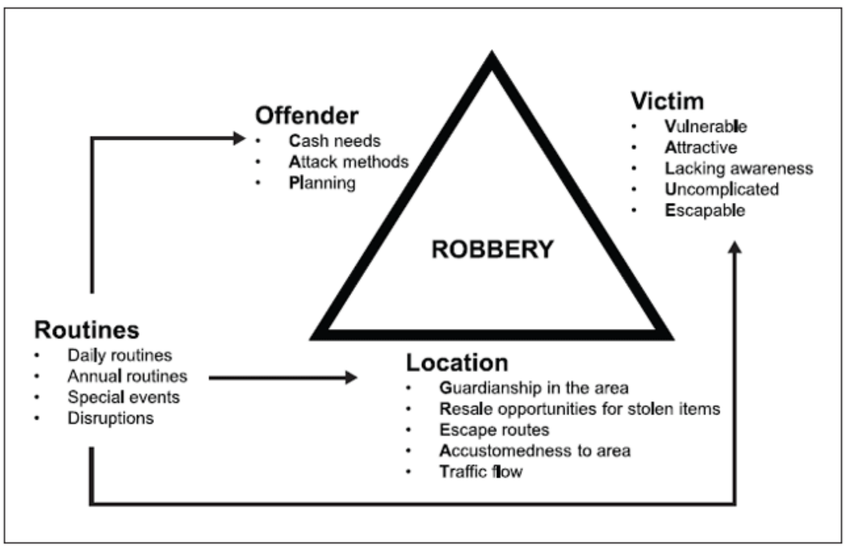


Figure 34. Street Robbery Analysis Triangle (ONK, 2010).

From the problem-oriented policing methodology and the analysis made by the Pop center, I developed a crime forecasting method which is a combination of two types of factors when computing the risk value: Crime opportunity based on physical characteristics and past events in that area.

Considering the physical characteristics, the existing crime analysis literature defines some base factors:

- A lot of youth crime is unsophisticated and unplanned. It is, therefore, more likely to be witnessed than more-sophisticated crimes (sociology.org.uk, s.f.).
- High levels of juvenile petty crime take place in public places (clubs, the street, etc.) where deviance is more likely to be witnessed.

- The lifestyles of young people (they often frequent pubs and clubs) may expose them to situations where criminal behavior is possible/likely (especially violent crimes, joyriding and various forms of petty crime – minor thefts, for example).

Also, according to (Center, s.f.) the street robbery opportunity can be based on the following structures:

- Commercial robbery (e.g., bank robberies, gas stations and convenience stores)
- Pickpocketing
- Vehicle-related robbery (e.g., robbery of armored trucks and taxi drivers, and carjacking)
- Non-stranger street robbery (e.g., drug-related robberies, robberies by prostitutes and robberies by friends, relatives or spouses)
- Home invasions
- Larceny-theft (note that some police agencies may record purse-snatching as larceny-theft)
- Assaults

Using Figure 33, I extracted some crime opportunity factors to be used for defining the hypotheses in the DST framework:

- Shops and amenity places concentrate potential victims of different types.
- Primary and secondary streets provide good escape routes and less risk to the offenders.
- Schools provide better robbery opportunities at certain hours during weekdays.
- ATMs provide opportunities to assault older adults in the morning and near paydays at all day.
- Bars and pubs provide potential victims to offenders at late afternoon on weekdays and night on weekends.
- Bus stops and train stations provide a variable number of victims depending on the use of the transportation network at each location.

Considering the past events, according to (Grove, 2012) "Repeat victimization refers to the repeated criminal victimization of a person, household, place, business, vehicle or other target however defined. Near repeat victimization or near repeats refer to targets with similar characteristics or situations (also virtual repeats). Repeats can be the same or different crime types. It is generally accepted that a small proportion of any population of potential targets experience a vastly disproportionate amount of the crime because they are repeatedly victimized". Criminology also considers that the nearer a crime has occurred in time or space, the bigger the risk that it will occur again. Furthermore, according to (Bowers, 2004) the burglary risk duplicates in properties within 400 meters of a burgled household for up to two months.

This motivates the definition of hypotheses based on time and location of past crimes in the area.

Considering the information presented above, the Dempster-Shafer hypotheses related to physical characteristics of the location are the following:

1. Shops and amenities are extracted from the OpenStreetMap database. Each location with an initial weight of 1/10.
2. Primary and secondary streets are extracted from the OpenStreetMap database. Primary street has 1/5 and secondary 1/10.
3. Schools are also provided by the OpenStreetMap database, each with a weight of 1/5 between 14:00. – 16:00; and 18:00 - 20:00, during weekdays.
4. ATM locations are available in the OpenStreetMap database, each with a weight of 1/20 between 8:00 and 11:00. This includes weight of 1/7 during paydays between 8:00-11:00 and 17:00-22:00.
5. Bars and pub locations are available in the OpenStreetMap database, each with a weight of 1/10 during paydays between 20:00-23:00 during weekdays and 23:00 to 8:00 during weekends (including Friday night).
6. Bus-stops and subway stations with a weight of 1/5 for every 1,000 people using the node in each hour and day. In this study, I use the public transportation network database when available.

The hypotheses related to the near repeat victimization theory are the following:

7. The risk generated by a past event as a weight given by the average of events per square (average events) divided by the number of squares (risk).

$$average\ events = \frac{\sum_{i\,\in squares}\quad events_i}{events};\qquad \text{risk} = \frac{average\ events}{squares}$$

8. A past event duplicates the risk within 200 meters, decreasing to the average during the next two months.
9. An event is most likely to repeat under a similar time period (same hour, same weekday, same day of the month, etc.), the events occurring in each time period will have a complete weight during the same period, decreasing to zero in 4 hours. Each defined period will provide us overlapping hypotheses:
   - Events during weekends will have weight during weekends.
   - Weekday Events during weekends will be have weight during weekends.
   - Events during the Morning will have weight during Morning.
   - Events during the Day will have weight during the Day.
   - Events during the Night will have weight during the Night.

Defining the (initial) weights is an important and time-consuming process. This is done by comparing the prediction results of the algorithm against the available data, taking sample places and data, calibrating the weight according to the samples crime activity. In the next section using automatic weight adjustment methods is discussed.

The SGL equivalent algorithm that allows generating risk terrain is basically the same as previously described. However, the *getCrimeDataByHypo* function returns the union of the crime records and the crime opportunity map data hypotheses:

```
Input: a list of hypotheses and the historical crime records
Result: grid, an array with the belief values for each square of the grid generated for each rectangle of the considered area
/* obtain the hypotheses, the temporal and distance models, and generate a grid for computing the belief value for each cell (quad) of the grid */
hypotheses = getHypos();
temp = getTemporalModel();
dist = getDistanceModel();
quads = genetareGrid();
/*a two-dimensional array for storing the mass for each cell of the grid according to each hypothesis given by the user*/
quadrants = Array();
foreach the hypotheses as hypo do
  //retrieving all crime related data which match the hypothesis
  crimes = getCrimeDataByHypo(hypo);
  foreach the crimes as crime do
    foreach the quads as quad do
      //obtain the mass associated to the hypothesis
      mass = hypo.mass;
      //calculate its contribution to this cell
      mass = mass*temporalEval(datetime,hypo,temp[hypo]);
      mass = mass*distanceEval(quad,crime.location);
      /*adding the contribution of the current crime record to the list of influences the cell (quad) receives from all hypotheses. These values are not simply added, but will be combined later according to Dempster-Schafer´s fuzzy logic rules */
      quadrants[quad][hypo].append(mass );
    end
  end
end

/* combination rules of Dempster-Shafer are applied, thus obtaining a single mass for each hypothesis for each cell
quadrants = combineHypoteses(quadrants);
/* values are normalized between 0 -1 */
quadrants = scaleValues(quadrants, getMin(quadrants), getMax(quadrants));

forall the quadrants as hypolist do
 /*obtaining the mass of the composed hypothesis with the biggest value */
  bighypo = getBiggestHypo(hypolist);
  grid[quad] = bighypo;
end
```

8.2 Forecasting experiments

The results of applying any crime forecasting algorithm is usually represented by a set of cells with associated risk values. In this case, the term "risk belief" must be used because of the Dempster-Shafer theory definition previously described.

Crime forecasting methods can be compared using HRP and TIP metrics. However, the **time period for which we want to make the prediction (couple of hours, a complete day), thresholds** and **weights** for the hypotheses must be defined. These parameters can vary the prediction results or the usefulness of information. In order to deal with this problem, I conducted three different tests:

- The first experiment checks the weights and thresholds selection method, using a large time period with a basic segmentation in a small urban area, obtaining good results on the spatial dimension but imprecise results on time.
- The second is for establishing the most convenient time period. For this I compared the recommendation given by the police according to their needs (8 hours) against what is commonly used, which is 1 month. For this I used data covering a large urban area (50 km2). It also tests an automatic weight adjustment algorithm.
- The third experiment is the one for validating the objective of this work, which presents data about forecasting crime occurrence in 8 large communities of a big city, each community presenting different characteristics.

The cell size was set at 150 meters wide as a recommendation from the police analytics department. Also, in some experiments, the risk maps were evaluated for the same time period as the police surveillance shifts (0:00 to 8:00, 8:00 to 16:00 and 16:00 to 0:00).

8.2.1 Medium scale scenario

In order to test the weights and thresholds parameters, I developed an inquiry experiment in a small area for a long time period. The dataset was obtained from data.world (https://data.world/), it contains the 2017 reports in America and January of 2018. The events hypotheses were supported by the last 3 months of data from 2017 in small area near Austin, Texas (USA). The results were compared to January weekdays as a generalization.

In Figure 34 top-left we see three months of the crime records. Using the same hypotheses, I forecasted the risk of crime occurrence for the weekdays (Monday to Friday) of the following week divided in three time segments: 0:00 to 8:00 (shown at the top right); 8:00 to 16:00 (shown at the bottom-left); 16:00 to 23:59 (shown at the bottom-right).

The forecast is expressed using two types of cells: white ones show places with near 0 belief values, meaning a complete uncertainty about the risk in that places. The red cells show places with some degree of risk belief, going from near 0 (almost white) to 1 (red).

After the risk map is computed, the risk is categorized in 10 levels using Jenks natural breaks (Jiang, 2013); only the five higher levels are considered and shown with color, and the rest is set back to 0.

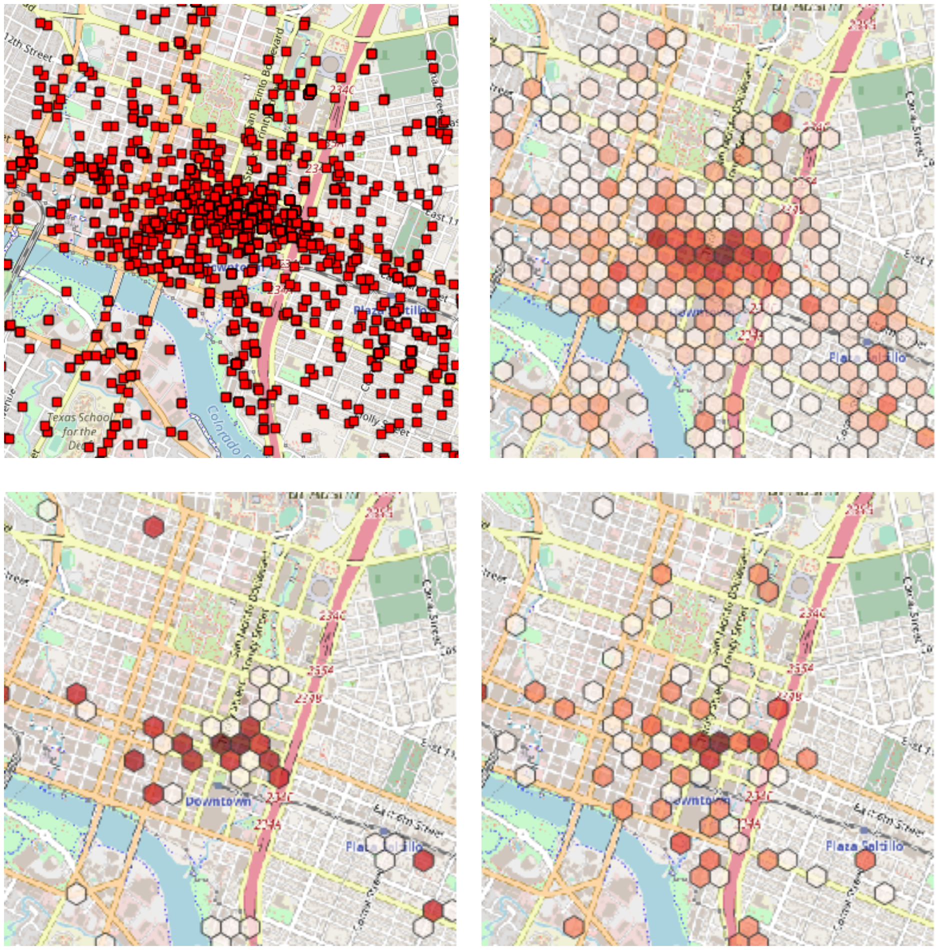


Figure 35. Top left figure shows crime related to assaults, robberies and drugs in Texas (USA) between October and December of 2017. Top right, bottom left and bottom right figures show the result of the crime prediction using our method for a weekday of January.

The forecasted areas were compared with all the crime reports in the same time period during January weekdays. Considering only the 5 upper levels of risk, the following results are:

- 0:00 to 8:00: 61% of the actual crimes occurred inside a high-risk area.
- 8:00 to 16:00: 51% of the actual crimes occurred inside a high-risk area.
- 16:00 to 0:00: 43% of the actual crimes occurred inside a high-risk area.

The spatial forecasting performance of this method was high compared with the results of ST-GAM in Charlottesville (see Figure 35) in the three time-segments. However, these results only predict the location of events over 5-day wide time segments, over a month, excluding weekends.

8.2.2 **Medium scale scenario**

This experiment was conducted using a larger dataset. Due to security concerns, I am not allowed to disclose the location of the results. However, I can share the following characteristics:

- This data was provided by the police department of a country based on reports made by individuals who have been victims of a crime.
- The provided data was gathered from one of the most important urban communities (areas) of a major city of the country.
- Two years of consistent data were used.
- The data describes the following information: burglaries, assaults, shoplifting, drugs, among other types of crimes.

In this experiment I tried to find proper parameters for:

- Time period forecasting precision.
- Weight's calibration method.

To accomplish these objectives, three different sets of experiments in a representative area of a city, of 50km2 and 300,000 habitants were conducted. This area has multiple commercial centers, cinemas, large business areas, subways, and most city services of a developed country big city.

The experiments were designed to compare two forecasting time-precisions: a month vs. a single day. method using automatically improved weights was also tested based on results:

- Experiment 1: single risk map for a whole month using the proposed masses.
- Experiment 2: three risk time intervals per day (0am-8am, 9am-16pm, 17pm-23pm) using the proposed masses.
- Experiment 3: three risk time intervals per day (0am-8am, 9am-16pm, 17pm-23pm) using automatically improved weights.

The improved weights are obtained using a metaheuristic algorithm called Particle Swarm Optimization (Kennedy J., 2011). In a few words, the algorithm consists on having a population (swarm) of candidate solution (set of weights), each particle is moved among the solution space guided by their own best-known position in space and the best-known position of the swarm. When an improved position is discovered this will start to guide the movements. This process is repeated until an acceptable solution is found (or not). Each position is evaluated by running the forecasting algorithm and testing the resulting performance. This method requires a large amount of time and resources; however, the resulting weights can be used until the crime behavior changes in the forecasted area. This means, it is not expected to be run frequently.

In order to provide useful information, all the experiments are limited cover up to 10% of HRP (risk covered area). Figure 36 shows the results. In the figure, HRP and TIP are the percentages of a high-risk area, and the percentage of incidents happened within the high-risk area respectively. According to the Dempster-Shafer theory, some cells have no belief; this means 100% of uncertainty. The curves reach to the last event occurred in a cell with a belief value over 0.

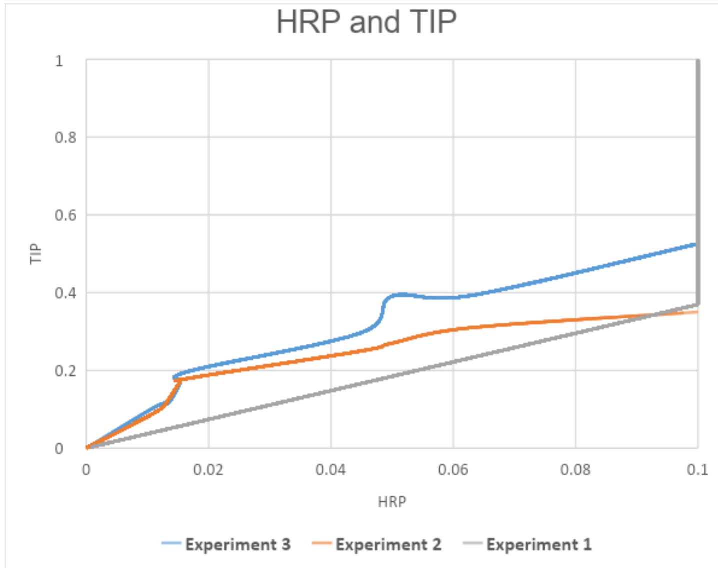


Figure 36. Spatial Dempster- Shafer method performance for crime prediction.

We can see the difference between the Experiment 3 curve and the other ones. Logically, using the right weights should lead to better forecasting results, and the proposed method allows us to accomplish it. However, a significant aggregation, as a monthly aggregation leads to average results. Comparing our results with Figure 32, the Experiment 3 aggregation has a significantly better performance: over 40% of TIP. Experiment 2 provides good results for the riskiest areas, which are probably the easiest to predict, however it ends in a similar way as Experiment 1, where the HRP reaches 10% of the area.

8.2.3 Large scale scenario

The performance of different forecasting algorithms depends on several factors, as the physical characteristic of an area, habitats, the type and size of commercial areas, and many other factors with uncertain characteristics.

In order to test our method under various conditions, a large experiment was designed including location with strong differences in its characteristics. eight areas from a big-city over two months were compared:

- Area A: A large town near a metropolitan city with extensive farming areas. The residential areas can be described as being middle-class.
- Area B: A large and wealthy class district in a metropolitan city, utterly urban with 1º class city services, malls, exclusive shopping areas, subways, restaurants, etc.
- Area C: A small middle-class district in a metropolitan city, entirely urban, however, is mostly residential with essential city services.
- Area D: A medium-size middle-class district in a metropolitan city, thoroughly urban, however, mostly residential with essential city services.

- Area E: A central area of a metropolitan city, with large commercial areas and transportation networks nodes (trains, subways, and buses), a considerable amount of people live in this area, mostly in apartments and older houses.
- Area F: Is a mixed district, with wealthy residential areas in middle-income and poor areas. It is mostly residential.
- Area G: A large residential district (almost a million habitats), with malls, large commercial areas. It is a mix of medium and low-class areas.
- Area H: Is similar to area A, is a large residential area near a metropolitan city. However, it can be described as a low-class district and a dangerous place.

The forecasting was made based on the results of the previous experiments, using the calibrated weight for all the areas, and defining three risk time intervals:

- Time interval 0: 0:00-8:59
- Time interval 1: 9:00-16:59
- Time interval 2: 17:00-23:59

The thresholds used were: 10% of the area, 5% of the area, and 30 cells per area. As previously stated, the cell is a square area of 150 meters width.

The complete table results are shown in Appendix A. Here I am going to discuss the results shown as bar charts in which the average percentage of predicted events (TIP) in each area using three different thresholds.

The hypotheses include at least 2 months of events data for each forecast. In this experiment 3 forecasts per day were made (according to the time intervals given by the police), over two months including all days of the week.

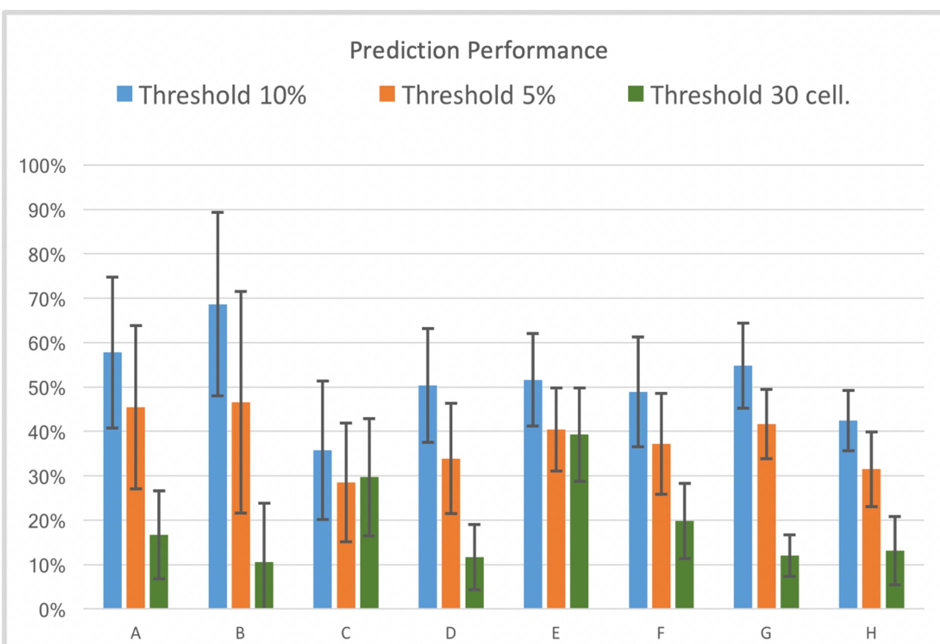


Figure 37. Prediction performance in percentage in different areas and threshold areas.

In Figure 37 the results are shown using 3 different criteria for defining the threshold for an area to be considered risky: the blue bar shows the results when the criteria considered only 10% of the riskiest cells. The orange bar shows the results when considering only 5% of the riskiest cells. The green bar shows results when the criteria considered the 30 most risky cells.

The chart shows similar performance for similar characteristics: A and B are areas with many risk points where crime is distributed among them, obtaining very good forecasting results with 10% of HRP. Residential areas ("D", "F", "G" and "H") also present similar performance, because crime is distributed in similar ways: there are not many crime concentrations points, reducing the forecasting performance. However, there is no difference in residential areas of different size.

Figure 38 shows the same results as used for chart of Figure 37 but they are expressed as the number of crimes predicted in the risky areas. It also includes a fourth bar (in black) which shows the total number of crimes occurred for the forecasted period.

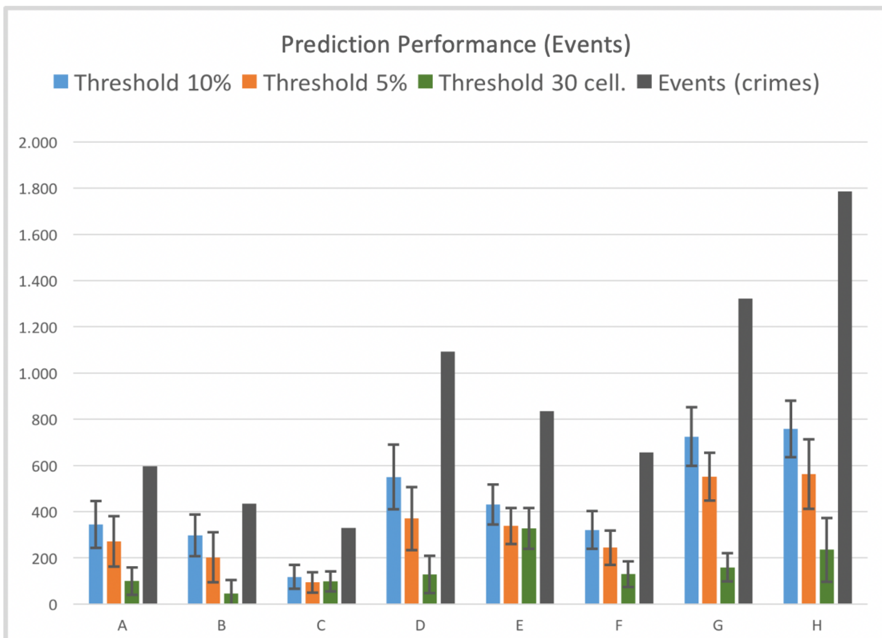


Figure 38. Prediction performance in different areas and thresholds.

Figure 37 shows the real difference in crime occurrence among the areas. From Figure 36 we can see that residential areas ("D", "F", "G" and "H") have similar forecasting results. However, F has a smaller amount of crimes compared to H. From Figure 37, Figure 36 we can conclude two things:

- Commercial areas with high crime occurrences are easier to predict (areas A and B).
- Residential areas have similar forecasting performance, no matter the number of crime occurrence they have.

8.3 **Discussion**

The fundamentals of this chapter were used to build an SDSS in collaboration with the police for the large-scale scenario, furthermore, as a qualitative evaluation I can mention that the resulting SDSS is currently used at a country scale. However, most of the details about this part of the work are confidential.

In a quantitative view, the results presented prove that the method can be used to predict crimes within townships with similar performance to S-T GAM and spatial GLM methods in the worst case. Our method obtains an average of 51% of TIP with 10% of HRP, which is better than the 38% TIP of S-T GAM in a specific area (see Figure 32). However, there is an important trade-off between the covered area and the distribution of crime in it. Higher crime concentrations are easier to predict.

One of the advantages of this method is that the hypotheses are easy to understand, allowing the experts a straightforward process to set and adjust the parameters of their models.

Crime forecasting algorithms are hard to compare with each other, mostly because of the lack of common datasets to test. Forecasting algorithms can have good results in a specific area, but these results may vary in a different location. However, we can note that the most famous crime prediction software (PREPOL (Company, s.f.)) claims that it can forecast 35% of the crimes without specifying the HRP. In this work the method was tested using different locations and data sources. This chapter helps us prove that the application of the method to risk terrain modeling and crime prediction can be a successful test part of the research question 2 (RQ2) by fulfilling the specific goal SG2a of this work: "*Developing a method to generate spatial decision-making scenarios in a flexible and systematic process based on mathematical decision-making theory*".

Chapter 9

Transportation Network Planning Example

Many urban areas are quickly growing. Their decision-making problems are increasingly complex, and better methods to evaluate solutions are needed in order to support this growth (Heilig, 2012).

Many decision problems concerning cities are spatial. A typical spatial problem is to define an area to support a certain requirement or service, e.g., space for a new road, an industry, or a hospital. Furthermore, cities are constantly changing, and they have dynamic problems, like the planning of public transportation services: the routes can be dynamically defined to cope with new requirements and constraints. Some of the decisions may be to find the right location for new bus stations, define a new bus route, or even plan a full transportation network (Yang & Wan, 2009).

According to (Harrison, 2010), a "Smart City" is a city that monitors and integrates data and information of its critical infrastructures, including roads, bridges, tunnels, rails, subways, airports and seaports in order to optimize its resources and maximize services to its citizens (Basso, 2018) (Stylianou, 2019).

At the same time, citizens are using information technologies (IT) to consume and provide data that can be used to support the decision-making process for several cities' requirements. Some of the ITs used by citizens are supported by Cloud Computing Services providing Software as a Service (SaaS). The Software as a Service model of Cloud Computing is often accessed by citizens through mobile applications and web interfaces (Chourabi, 2012). Some of the SaaS services with spatial data properties are for example Google Maps, OpenStreetMaps and WAZE. These services provide geo-localized data in a graphical way, they are free, and they share a singular characteristic: they use crowdsourcing to gather the necessary data to feed the application.

In this Chapter the application of the Dempster-Shafer theory applied to the transportation modeling problem is explored. The goal is to test the hypothesis if the method developed in this thesis is able to make a forecasting which provides useful information about the usage of public transport services using data, which can be obtained from freely available public information. This would serve to make a preliminary study for decision makers for example, when evaluating the implementation of a new bus line.

In order to do this, I develop an application that uses the services provided by Google Maps, OpenStreetMaps, and Waze, to develop a Spatial Decision Support System (SDSS) for transportation network planning, more specifically, the Origin-Destination (OD) evaluation. In order to deal with uncertain data sources, I propose to evaluate de OD using the SDSS method developed in this work.

The following section describes the problem, followed by our proposed approach to the problem.

9.1 Transportation modeling

People's transportation modeling has been largely based on the four-step model (FSM) that was developed in the 1950s (Wegener, 2004). The four-step model is used forecasting transportation demand and it is primarily intended to be used for long term planning or for infrastructure development (Bhat, 1999). It involves four major steps: trip generation, trip distribution, mode choice, and route choice (or assignment).

Trip generation establishes the propensity to travel by estimating how many trips are generated by analyzing possible origins (known as productions) and destinations (known as attractions) separately. For example, how many trips would a residential building produce and how many trips would a shopping center attract? Trips are modeled at different aggregation levels (zone, household, etc.) and sometimes the personal level. The most common models used in trip generation are either category models or regression models combining many socioeconomic and land use related variables. In the next step, trip distribution, the separate origins and destinations are combined to create origin-destination (OD) pairs. The most common way of creating OD pairs is through the use of gravity models, the result of which would yield a matrix of OD pairs. The next step, mode choice, further groups the trips (OD pairs) into different modes of travel: car, public transport, cycling, etc. The effect of time of day is integrated at this stage as well, depending if the data used has a time dependence or not (usually AM peak and PM peak periods or hours). The most common model form used during this process is discrete choice modeling, more specifically logit models. Route choice or route assignment places the generated trips onto the transportation network. Trips are injected into the network according to the OD matrix established in the distribution step and further refined in the mode choice step. Depending on the aggregation level, with zonal being common, household vs zonal, trips are injected into the zone at the centroids and then spread to the road network. Trips are most commonly spread into the road network under the assumption of achieving system equilibrium (at first every user would choose the path of least resistance from origin to destination, and network equilibrium occurs when no user can decrease travel effort by shifting to a new path). Volumes of vehicles on the network as a result of route assignments are calibrated using real life traffic counts (Nagel, 1992).

With the advance of computer processing power, different simulation models have been created that can treat both the demand (trip generations and distributions) and supply (road infrastructure, traffic management actions, incidents, traffic signals, etc.) at different levels of detail. The three main levels of detail are: macroscopic (static, deterministic), mesoscopic (stochastic, dynamic traffic assignment), and microscopic (lane change rules, car following models) (Casas et al., 2011). Often, a mix of aggregate and disaggregate approaches are used to model travel demand (Wu & Mengersen, 2013). Currently, intelligent transportation systems (ITS) are used to manage traffic in real time with sensors placed on the road network, for example (Barceló, 2007).

A public transportation system is typically a complex network. These networks are composed of various transportation lines designed to cooperate and complement an urban scale transportation solution (Guihaire, 2008). Most metropolitan areas having more than 500,000 inhabitants would have at least transportation lines like: railways, highways. Each transportation line is designed to cooperate and complement an urban scale transportation solution (Yang et al., 2009) (Xu & Gao, 2008). The planning of the paths or routes of the new or existing transportation methods are usually based on existing basic data, of the transportation network, volume predictions and the distribution in the network (Castillo, 2008) (Liu, 2010).

When a decision maker chooses a route, the travel time and time reliability are important factors under demand and supply uncertainty. According to (Xu, et al, 2008), modern urban road transportation design is to optimize the system performance, make the traveler arrive in destination conveniently and quickly. Furthermore, when designing an urban route for a new transportation service, the choices must consider the behavior and reliability of transportation network. Another important factor is the OD traffic demand (Castillo, 2008). The OD describes the quantity of traffic demand between a particular line that starts at an origin point and ends in a Destination point during a time period. In order to build a reliable OD, it is necessary to invest an important amount of resources to make large-scale demand studies and surveys.

Therefore, it would be very useful if we could do at least some preliminary evaluations of the new planned route without having to spend too much time and/or money in the process, in order to decide the feasibility of a new route at an early stage. Using the Dempster-Shafer theory allows users to specify new factors and information providing customized demand model for each scenario, providing an approximation to reality. In the following sections I explain how to evaluate the impact of different routes using this method.

9.2 Determining an OD route

The transportation network of Santiago of Chile is composed by bus stations, bus routes, subway stations and routes, shared taxi stations and shared taxi routes. Each OD is composed of a start station and an ending station. A single OD can have multiple sub-ODs on a single route. The design of a public transportation system is a complex task that requires analyzing the public demand for transportation, the traffic demands of the alternative routes for each OD, and the reliability of the OD. In order to define an OD route based on uncertain demand information, I propose to adapt the Belief Map concept to the current scenario (Frez & Baloian, 2013). This concept is based on the Dempster-Shafer Theory. As described in previous chapters, a Belief Map allows us to evaluate a geographical area generating a suitability evaluation on a set of hypotheses that support a solutions proposal.

Based on similar principles, a Belief Route shows the estimated demand of passengers that would like to use a public transportation bus serving that route in its various sections. Each section is delimited by two contiguous bus stops. A Belief Route is composed of 3 basic elements: 1) The origin and destination points; 2) The polyline connecting these two points, which describes the route followed by the bus in order to travel from origin to destination. 3) A set of hypotheses specifying where people demanding the service could be located. Besides estimating the possible demand for the route, in some planning scenarios it is important to estimate the traffic congestion along the whole route in order to evaluate if the traveling time between stops from the origin to the destination is both predictable and it takes reasonable time (Fernández, 2008) (Vuchic, 2002). Using Belief Route and Belief Congestion Route maps, the decision maker can compare various paths and evaluate their ability to satisfy the transportation requirement of the population along the route. Furthermore, the decision maker can also adjust the Origin or Destiny. In order to provide information about the reliability of the route I propose the use about the traffic information from Waze creating a belief value based on historical data. I call the result of this combination as Belief Congestion Route (BCR).

In order to explain the use of the BR and BCR in the decision-making processes, I am going to use a basic example. I want to evaluate two alternative paths for the same OD. In Figure 39 and Figure 40, two different alternatives are shown (A and B respectively). In this example, route A is shorter than B, and the travel time is also shorter according to Google Directions API.

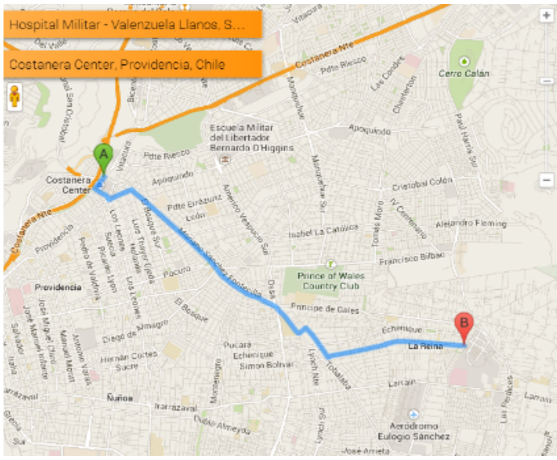


Figure 39. Route 1.

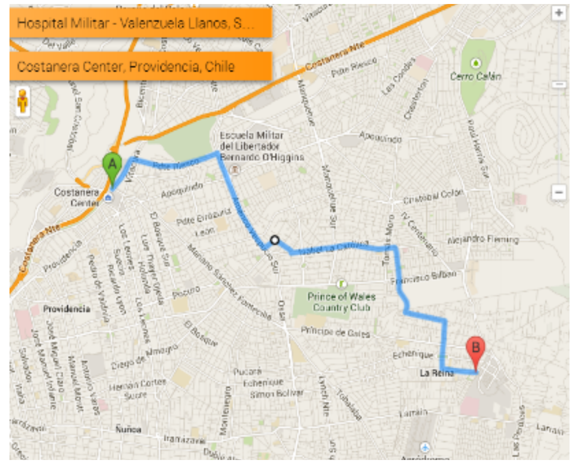


Figure 40. Route 2.

The most critical aspect for the success of the Belief Route and Belief Congestion Route is a correct estimation of the hypotheses, which would generate the "mass" supporting the possibilities of the demand and congestion estimations along the planned route.

In order to test the proposed concept, I am going to use real data from a public transportation system. The testing method is simple: I use real data to evaluate a hypothesis set (used to build the BR) if the prediction generated by the hypothesis set "matches" with the real data I assume the hypotheses hold and thus the generated BRs are valid. In this experiment I will use the following hypotheses:

1. People require more transportation near shopping areas. Weight: 50 %.
2. People require more transportation near amenities. Weight: 50%.

The weights are preliminary guesses, and they will have to be adjusted by comparing the predicted demand with the real data. The validation data will be extracted from the transportation network payment system (Santiago, Chile), which is based on RFID cards. Each bus has a reading device that allows users to pay using the card. When a user starts a trip at any bus stop, the payment is stored with the geolocation of the bus. Using this data, we can quantify the demand of each bus stop of the real route.

The SGL equivalent algorithm that allows generating Belief Routes is the following:

Input: a list of hypotheses and the historical data about path segments records: the path to be evaluated

Result: cuadricula, an array with the belief values for each square of the grid generated for each rectangle of the considered area filtered by the geometrical intersection with the path.

/* obtain the hypotheses, the temporal and distance models, and generate a grid for computing the belief value for each cell (quad) of the grid */

hypotheses = getHypos();

```
temp = getTemporalModel();
path = getPath();
dist = getDistanceModel();
quads = genetareGrid();
/*a two-dimensional array for storing the mass for each cell of the grid according to each hypothesis given by the user*/
quadrants = Array();
forall the hypotheses as hypo do
  //retrieving all features in the area which match the hypothesis
  feats = getFeaturesHypo(hypo);
  forall the feats as feat do
    forall the quads as quad do
      //obtain the mass associated to the hypothesis
      mass = feat.mass;
      /calcutate its contribution to this ce
      mass = mass*temporalEval(datetime,hypo,temp[feat]);
      mass = mass*distanceEval(quad, feat.location);
      quadrants[quad][ feat].append(mass );
    end
  end
end

/* we apply the combination rules of Dempster-Shafer, thus obtaining a single mass for each hypothesis for each cell */
quadrants = combineHypoteses(quadrants);

/* Return list of quadrants that intersects with path */
quadrants = intersects(quadrants,path)
/* values are normalized between 0 -1 */
quadrants = scaleValues(quadrants,getMin(quadrants),getMax(quadrants));

forall the quadrants as hypolist do
 /*obtaining the mass of the composed hypotesis with the biggest value */
  bighypo = getBiggestHypo(hypolist);
  cuadricula[quad] = bighypo;
end
```

Applying the algorithm to route 1 and 2, we can obtain a proposed transportation demand for each path. In Figure 41 and Figure 42, the transportation demand is represented by a BR, according to the OpenStreetMap data and the proposed hypothesis, route B has more demand than route A.

Figure 41. Belief route 1.

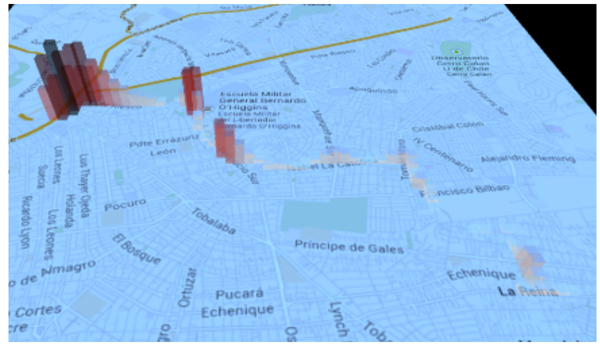


Figure 42. Belief route 2.

The Belief Congestion Routes are computed in a similar way as the Belief Routes, with the difference that now I use a single hypothesis where each report about a traffic jam in Waze adds mass to the rectangle where it was geo-referenced by the reporter. In Figure 43 and Figure 44 the BCR of each route is shown. According to the Waze information of both paths, route 1 has more belief of having congestions or traffic jams, implying less reliability.

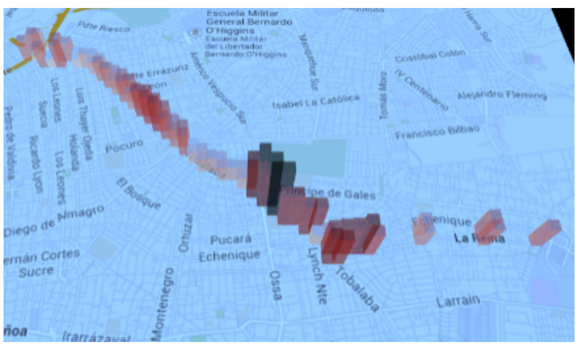


Figure 43. Congestion route 1.

Figure 44. Congestion route 2.

From this example we can note that route 2 has less congestion and more demand than route A. However, route 1 is shorter and the decision will depend on what kind of OD the decision maker is looking for. In order to support the decision, the visual evaluation of the BR and BCR is not enough. An evaluation metrics framework is needed, and it will be part of our future work.

9.3 Belief routes in real world

In this experiment I am going to test if it is possible to estimate the demand without transportation data or surveys. First, given an OD, I will use real bus route with transportation demand data (route 1). Secondly, I will test a set of hypotheses based on the Openstreetmap database, comparing the results with the real data and adjusting weights until there is an acceptable match between both real and estimated demands. Finally, using the same set of hypotheses previously calibrated, I propose a new route (route 2) for the same Origin and Destination points.

I have chosen a large area with high transportation activity in the city of Santiago. This area was selected because it is representative for the city, having many shops in its commercial center, a mainly residential sector with houses and apartments, and an industrial area. The three areas have an important number of subway and bus stations. I selected a single OD with two possible paths. In Figure 45 and Figure 46, two options are shown (routes 1 and 2 respectively).

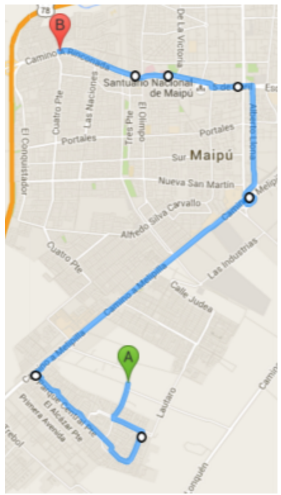


Figure 45. Real bus route, 12km, and 36min trip.

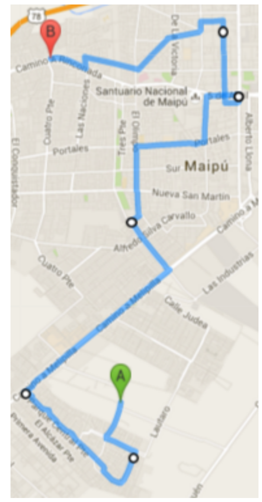


Figure 46. Proposed route, 17km, and 48min trip.

The analyzed area has various types of zones: residential, commercial, city hall and services, subway stations, and industrial areas. The locations of the zone types are shown in Figure 47. I will use the residential areas #1 and #2 as well as the industrial area to analyze the demand, since these are the areas where people start their trips according to the data.

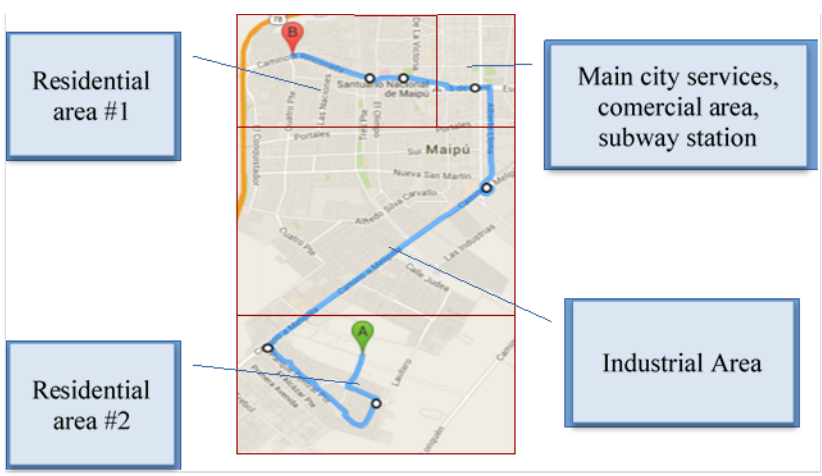


Figure 47. Zone types of the analyzed area.

Figure 48, Figure 49 and Figure 50 show the places where people actually took the bus according to the data provided by the Ministry of Transportation. The colors show the concentration of people: light blue for few people, to red for many people. In residential #1 area, we see two demand hotspots that are distributed among five bus stops. In the Industrial zone, there is only one hotspot at the entrance of a small residential area. The transportation demand in the industrial zone tends to be more equally distributed as in the residential area and it shows only one small section with high demand. Finally, in residential area #2, the demand is distributed among several bus stops at the center of the area.

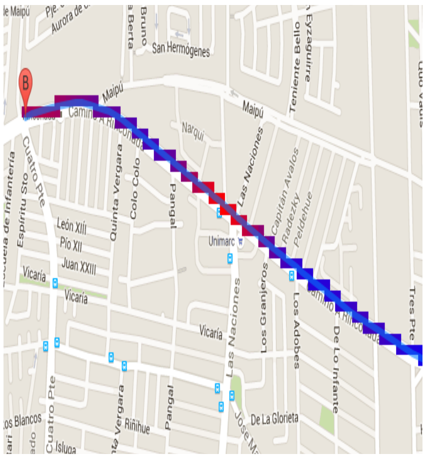


Figure 48. Transportation demand in Residential area #1.

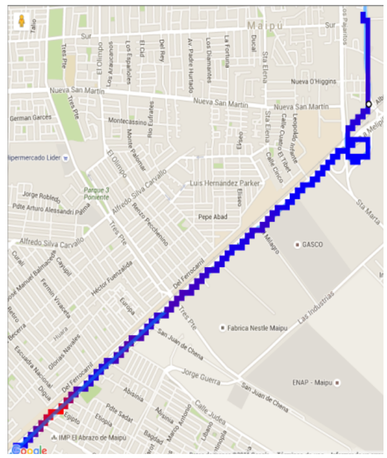


Figure 49. Transportation demand in Industrial area.

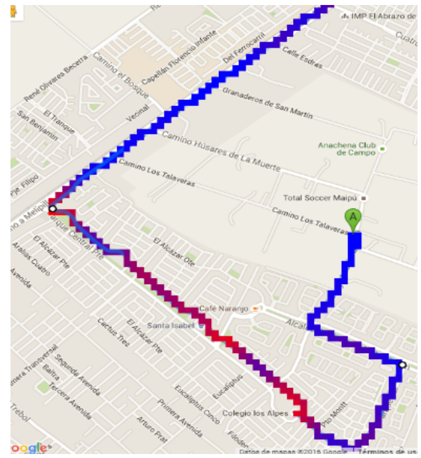


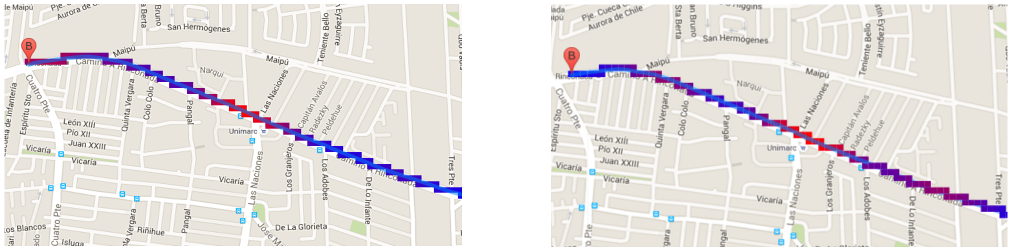
Figure 50. Transportation demand in Residential area #2.

Now I will compare this data against the prediction we obtain when computing the Belief route using DST and information we can get from the web.

In order to calibrate and find the right weighs (local optimum) I used a binary search strategy looking for higher prediction performance in each iteration. This process was applied to both hypotheses, however, due the results, only the shop hypotheses was adjusted.

1. People require more transportation near shopping areas, 26%.
2. People require more transportation near amenities, 50%.

These hypotheses are tuned to comply with the categories used by OpenStreetMap for the type of objects it stores. Apart from streets, we can get information about facilities classified as amenities and shops located in a certain area. Amenities refer to commercial places offering services like cafés, bars, restaurants, schools, universities, libraries, etc., while shops are related to commercial places selling goods like bakeries, convenience stores, supermarkets, medical supplies, etc. Figure 50, Figure 51 and Figure 52 show the real demand vs.. the predicted Belief route for the three selected areas.

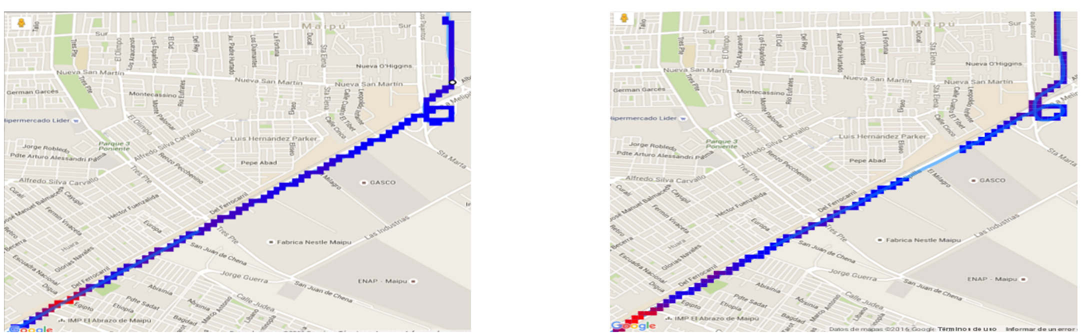


Real Demand Predicted

Figure 51. Real demand vs predicted demand for residential area #1

For residential area #1 (Figure 51), we can see that the prediction was quite accurate for this route section. The Belief Route predicts a high concentration of the demand at the center of the route section, which almost fully matches the real demand. It also predicts a medium demand concentration in almost all the rest of the route, which also matches the real data. The only noticeable difference is that the Belief route predicts a medium-high demand at the end of this route section, which is not backed by real data.

For the industrial area there is also a good matching between the real demand and the belief route: the real data shows a medium demand in almost all the route section except for the last part of the section, where a high demand is shown. The predicted Belief Route also shows a medium demand on almost the whole section except for the same high-level demand at the end. The only difference is that the Belief route predicts low demand at the middle of the section (colored with light blue), which is not backed by real data (Figure 52).

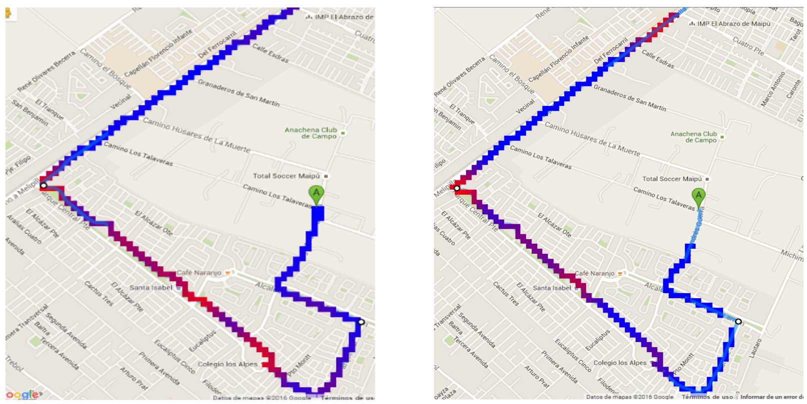


Real Demand Predicted

Figure 52. Real demand vs predicted demand for industrial area.

The prediction for the section going through the residential area #2 presents more differences than the other two areas, but it also shows many matches. The prediction shows a small sub-section with medium-high demand at the beginning of the section whereas the real data shows only medium demand. After that, both figures show a medium demand until the point where the route makes a 90° left turn, where both, the real demand and the belief route, show a high demand. After the turning point there is a small section where the real data shows a medium-low demand whereas the Belief Route shows a medium demand followed by a sub-section where both figures show a medium-high demand. At the end of the section, the Belief route predicts some sections of low demand whereas the real data show medium-low level demand. The explanation may be as follows: the last part of the real route goes through an area with few shops and amenities, but since it is the end of the route (or the beginning, depending on the direction people take the bus), it concentrates more customers coming from the nearby areas. This factor can be considered as an additional hypothesis in future analyses and may be incorporated to the set of hypotheses for obtaining a better prediction.

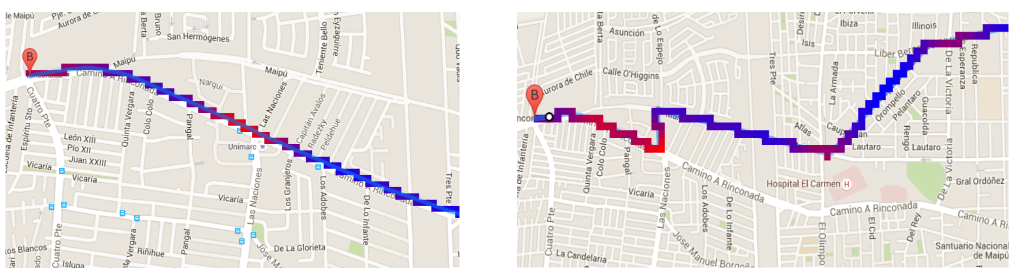
Summarizing, I can say that the Belief Route has a high level of matches with real data and where it does not fully match it, the differences are not large. In fact, the Belief Route never predicts a high demand when the real data shows a low demand, and vice versa. The most difficult place to predict the real demand seems to be the residential area, where there are few shops and/or amenities.



Real Demand Predicted

Figure 53. Real demand vs predicted demand for residential area #2.

Now with the right hypotheses and weights I will compare the real route (route 1) with the proposed route (route 2) using the estimated demand. In Figure 54, we can easily note that for residential area #1, the new route presents a higher demand than the original one. This occurs mainly because the new route passes through areas where more shops and amenities are located.



Original Route with Real Demand Proposed Route with predicted Demand

Figure 54. Original route with real demand vs new proposed route with its predicted demand for residential area #1.

Figure 55. shows the original route passing through the industrial area and the residential area #2 against the new proposed route (shown in a single figure). Here we also clearly note that the proposed route will have higher demand points than the original route (see Figure 55).

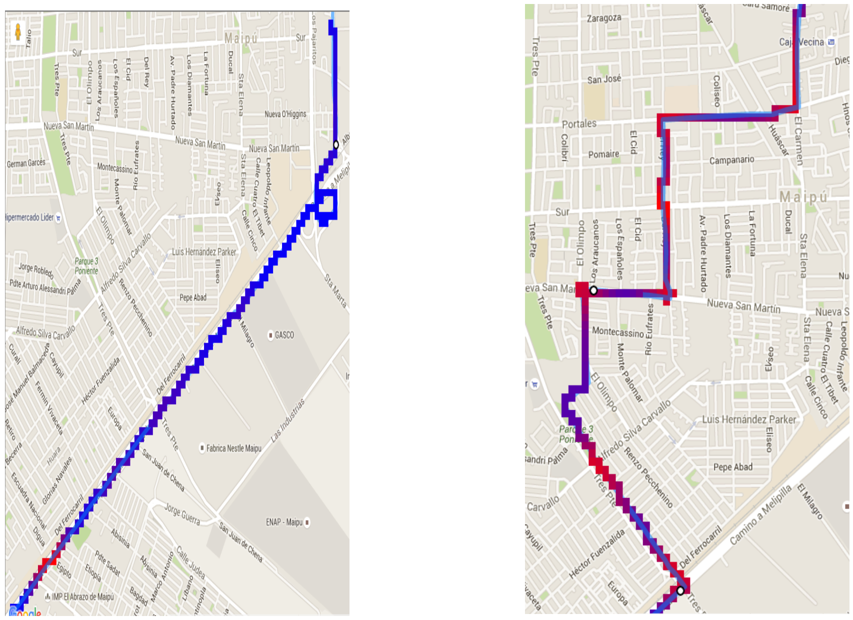


Figure 55. Proposed route on the right, with its predicted demand for industrial area and residential areas vs. real demand of the original route on the left.

In order to have more data about the accuracy of the predictions I conducted a large-scale experiment comprising the whole Santiago region. In a similar way as the previous experiments, the Openstreetmap database was used to generate the belief map using hypotheses considered above. The belief map was calculated on a grid consisting of 1,390 hexagons; each hexagon has an area of 700 square meters. Subsequently, the actual demand map was constructed based on the payment bearings that are georeferenced. I used 80 million markings, corresponding to two full weeks from July 5 until July 18, 2015. The demand map consists of the number of markings that exists in each zone (hexagon). Figure 56 and Figure 57. show the suitability maps for the real and predicted demands. At first glance, it is possible to see that there is a clear relationship between both maps (belief and demand), however this is not true for all cells.

To compare the relationship between both maps, I generated two series of values using the cell associated with each value as a common factor, and then I calculated the Pearson correlation between both series. The correlation between the values associated with each hexagon is 0.53, which indicates a high correlation between the real and the estimated magnitude of demand. However, for cells with less than 5,000 passengers a day, the correlation is 0.09, so it is not possible to use it in areas of low demand. Between 5,000 and 70,000 passengers is 0.50 and between 50,000 and 100,000 passengers, the correlation is 0.73. As a reminder, Pearson's scale ranges from -1.0 to 1.0.

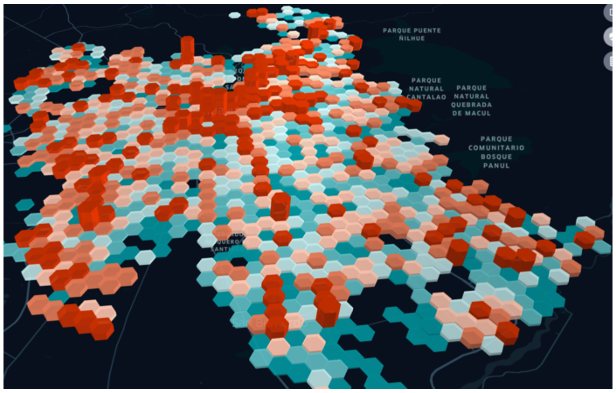


Figure 56. Real demand.

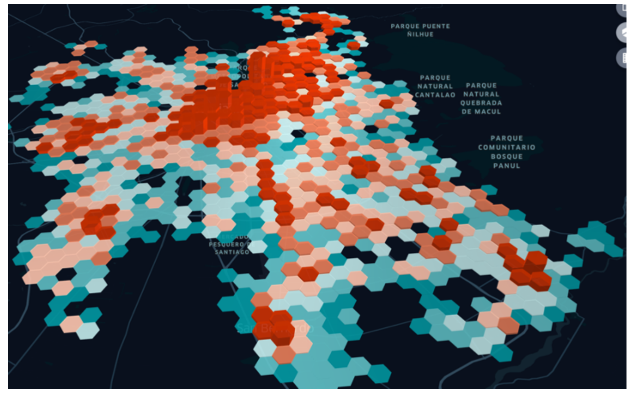


Figure 57. Predicted demand.

9.4 An Application for developing belief routes

Finally, I developed a prototype that allows the users to define an OD pair and a polyline. It also lets the user specify hypotheses for transportation demand modeling, after which it can generate two types of visualization: The demand Belief Route and the Belief Congestion Route. (see Figure 58). The application allows setting a transportation demand hypothesis set compatible with DST. It also allows us to include some model constraints, for example: to avoid schools. After the hypotheses are included, the application allows for choosing the type of 3D map that will be generated and shown: Belief Route or Belief Congestion Route.

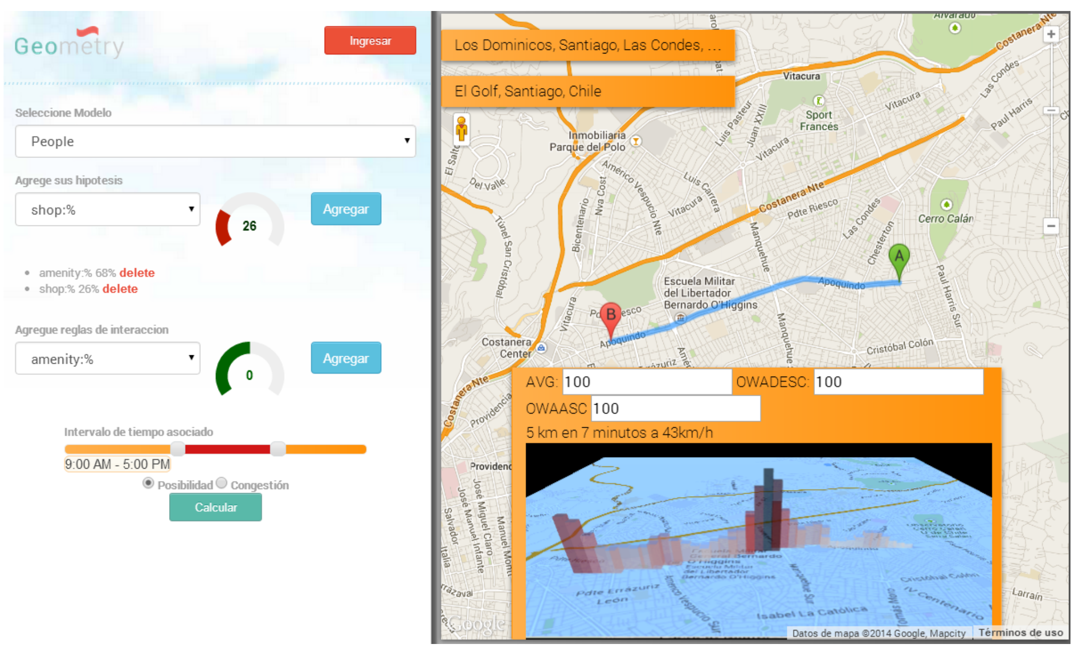


Figure 58. Evaluation of an OD using the developed application.

The application allows setting a transportation demand hypothesis set compatible with the Dempster-Shafer Theory. It also allows including some model restrictions, for example: avoid schools. After the hypotheses are included, the application allows choosing the type of 3D map, which will be generated and shown: BR or BCR.

As seen in Figure 57, the application interface is divided in two sections. The left one is for setting the parameters for computing the Belief Route and the congestion Belief Route. The user can choose one to compute using the toggle buttons at the bottom of this area. At the top portion of this area there are widgets for setting the parameters for computing the Belief Routes. The user should first select the model (Suitability Object), which corresponds to what is going to be predicted. The model should be previously fed to the system as a program module developed in Java-Script and following a pre-defined interface. In the example of the figure the chosen model is called "People". Beneath the model setting widget, we can see the widget used for stating the hypotheses. A pull-down menu displays all attributes available, which can be used to define a hypothesis. The items of this menu are automatically generated according to the installed modules, which download information about locations of facilities and could be useful for predicting the demand. For this work I used only information about objects provided by OpenStreetMap. At the right-hand side of the selected attribute the user can define the weight with which this attribute will contribute to the calculation of the mass by clicking on a circular widget. The light-blue pushdown buttons allow stating additional hypotheses. In the figure we can see the attribute "shop" with a weight of 26%. Below this information, the current and previously defined hypotheses are displayed. In the figure, we can also see that the hypothesis amenity, with a weight of 50% was also stated. Underneath this information we can define attributes for which mass will be decremented or even suppressed if found on the route. For example, I can state that the probability of having people in a park may decrease or the probability of having people on a lake will be zero, even if there are shops and amenities nearby. Finally, there is a widget that allows us to set the window of time for which the prediction will be generated. This is mainly for computing the Belief Congestion Route, which is very sensitive to the time of the day. For this case, data from Waze corresponding to that period of time will be considered.

The right-hand side of the interface shows the map, downloaded from Google Maps, which allows the definition of the route to evaluate and the display of the results.

9.4.1 Discussion

This chapter concludes that it is possible using the existing crowdsourced data to support a transportation network decision-making process. The method uses the Dempster-Shafer Theory to provide a framework to model transportation demand based on maps obtained from Google Maps, information about facilities located near the route from OpenStreetMap and information about the traffic from Waze. With this information, the Belief Route and the Belief Congestion can be computed. The Belief Route predicts the possible demand along an OD route and the Belief Congestion Route predicts the congestion and thus the travel time for a bus on that route. This allows us to make a preliminary evaluation of the route without investing time and resources to perform all the necessary tasks to gather the necessary information for a comprehensive evaluation. In addition, I presented an application that has all necessary functionalities to perform this preliminary evaluation.

To test the validity of this approach I experimented with a real existing route: then compared the prediction of the demand obtained with the Belief Route against the data obtained from the Transportation Ministry about the number of passengers getting in the buses serving that route at each bus stop. The result obtained was that the computed Belief Route predicted the demand quite well. To test the usefulness of this approach, the Belief Route for a new route with the same origin and destination points but traveling through other streets was computed, finding out that this new route would have a greater demand, thus being more efficient with the transportation resources. The Belief Congestion Route could also be computed in order to check if the travel time would have been also acceptable.

Chapter 10

Multiple stakeholders, scenario comparison and combination

This chapter focuses on the SG2b stated in the first chapter of this thesis: "Develop a method to evaluate, compare and work with previously generated scenarios". In a decision support processes, experts must evaluate and compare many scenarios which arise from different hypotheses, for example about where people may be at the time of the emergency and how they will react. This chapter explains this process by helping a hypothetical group of experts, generating, visualizing and comparing the outcomes of the different hypotheses.

To explain the analysis method, I use a well-known study case in Chile: Tsunami evacuation planning.

Chile has the highest level of seismic activity in the world and also has 6,435 kilometers of coast. These conditions generate a complex evacuation scenario under natural disasters like tsunamis. When a large earthquake occurs, the population in coastal areas must go to higher grounds. However, there are some cities in the north of the country with another complication; the tsunami speed is higher than the possible evacuation speed. This is the case for the city of Iquique (see Figure 59).

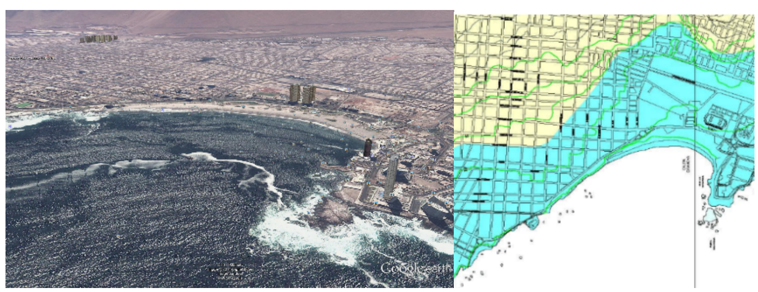


Figure 59. Left Iquique from the sea, Right: Evacuation area.

To make a population evacuation plan, multiple experts and stakeholders must collaborate (Paschke, Boley, Kozlenkov, & Craig, 2007) (Paschke, Kozlenkov, & Boley, 2007). Each one can have different opinions and hypotheses about what the best alternatives are to building an evacuation plan. To simplify the proposed collaboration method, I am going to suppose that they are 5 different experts. Two of them believe that the best evacuation method is for people to go to higher grounds, and the other 3 have different hypotheses: that many people cannot reach the higher ground before the tsunami arrival, so they have to enter tall buildings (vertical evacuation).

In order to evaluate both alternatives, emergency entities use traditional GIS systems, which indicate the number of people, schools and registered information pertinent to the area.

In the tsunami case, the evacuation problem can be classified as an ill-problem, because there is no information about how many people must be evacuated and there is no real knowledge about how much time it could take. An earthquake can occur at different hours of the day and the population in the area changes according to the day and time.

There are some simple cases, such as first dividing the planned output into parts assigned to individual members of the group and then assembling the contributions at a meeting or by a single person. Another simple case is to take turns at improving an initial draft.

The time and belief-based scenarios can provide better information to support the decision-making when there is uncertainty, incomplete information, and/or complex modeling.

Using the Dempster-Shafer Theory, we can build a set of hypotheses that can tell us where people can be. For example, there is an area on the coast of Iquique where there is a high belief of population during the day, due to a high concentration of universities, restaurants, shopping centers, a popular beach (playa Brava), etc. This area is also far from higher grounds (see Figure 60).

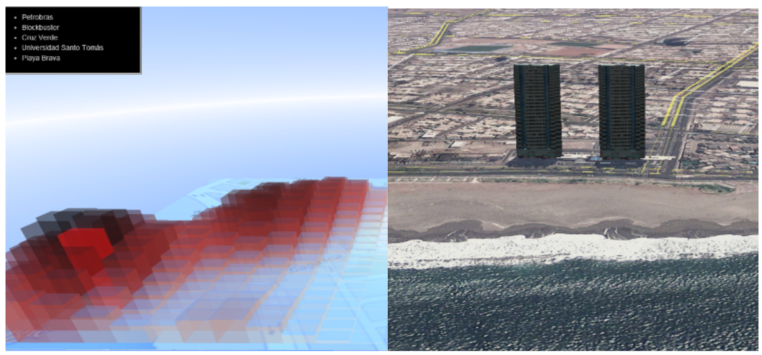


Figure 60. Playa Brava. Left: Population belief. Right: Buildings.

In this area, one proposed solution is a vertical evacuation, this means that people must enter into high buildings. This scenario can be described has an evacuation possibility, meaning that people can take refuge in tall buildings. However, an obvious problem will occur where there are higher concentrations of persons and a few buildings.

In order to address this variable, incomplete information and multiple scenarios, I propose Dempster-Shafer based Smart Geographical Information System (SGIS).

Using the SGIS, the previously mentioned 5 experts can make their own hypotheses evaluation. However, as a result of these processes, they will have 5 different suitability maps at least. For example, the first 2 experts can differ about where people can be at different hours of the day. Also, the other 3 experts can also differ about how tall the buildings must be or what kind construction can actually resist a tsunami.

After each expert builds their own simple scenario, they will have 5 different suitability maps. The next natural following step will group the suitability maps, according to both evacuation scenarios. However, this will not solve the collaboration problem.

Another case can be that each suitability map is based on similar hypotheses, for example, that during the day people are in Commercial areas, Schools, Universities, Libraries, Banks, Bus stations, etc. However, some expert can belief that there will be more people at commercial areas, and others say that they will be in residential areas.

In order to solve the problem, I propose to combine the suitability maps in a hierarchical order. Furthermore, each time a suitability map is combined, this combination must be augmented because it implies the incorporation of the hypotheses set to a final evaluation. Also, part of this arguments can be others suitability maps that validate the provided hypotheses set and the resulting suitability map (see Figure 61).

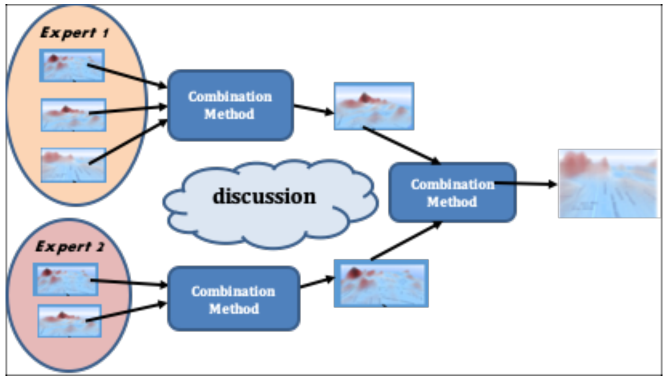


Figure 61. Hierarchical combination process.

The resulting suitability map can be the result of the discussion of each scenario possibility including the expert's hypotheses, known information about the area, and relevant factors that must be focused. However, the result depends on the combination methods.

In section 1.1, I describe the proposed combination methods with some of them designed to merge data, and others to focus on important factors, according to the expert's considerations.

10.1 Merge and compare operators

A complex scenario is the combination of various simple scenarios. Building a complex scenario requires cooperative work between different stakeholders like experts in the particular scenario area and decision makers. In order to provide useful tools for collaborative scenario building for a single area, we must divide the work into two different dimensions: Hypothesis Dimension and Time Dimension (see Figure 62).

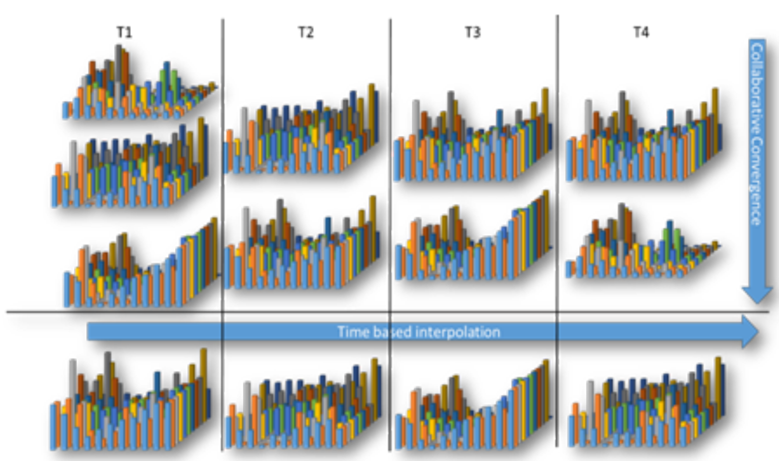


Figure 62. Hypothesis Dimension and Time Dimension: here we see that for the first time instant T1 experts have 3 different hypotheses.

The hypothesis scenario dimension is related to the collaboration between the decision maker and experts. The latest ones have different hypotheses about belief function values for a certain time. For example, one expert will have a certain hypothesis about the number of people at commercial areas for the morning, midday, evening, and night. When combining suitability maps, experts should consider the same time dimension for stating their hypotheses.

In this section, I propose to use five different operators to collaboratively build a scenario based on combinations of suitability maps. In order to simplify and appreciate the differences between the operators, I am going to use three initial scenarios (see Figure 63). Each initial scenario is the result of a specific set of hypotheses.

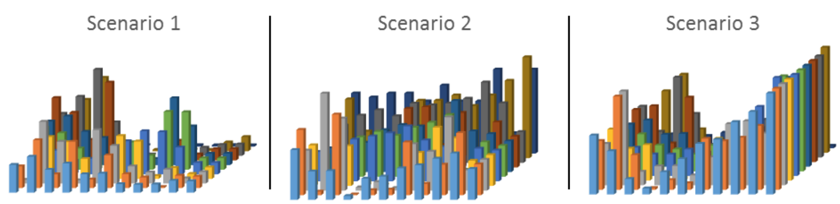


Figure 63. Initial Scenarios.

10.1.1 **Sum operator**

The Sum is probably the simplest operator a decision team should be able to use; it consists of summing the belief value of each scenario for each evaluated location. Graphically, it consists of summing the values of the three bars corresponding to the same cell. Visually, the resulting map shows the normalized sum of the three bars.

This operator can be useful when three independent, but related scenarios must be merged. For example, criminality, transit and street maintenance must be combined to evaluate the governmental resources needed in a general and comparative scope. Using sum, the decision maker can easily identify the need of resources for each location independent of the type of need (see Figure 64).

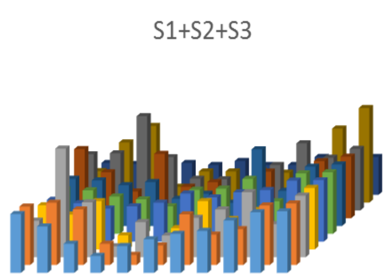


Figure 64. Sum of the three scenarios.

10.1.2 **Average operator**

The average operator is the simple average of the belief degrees of each location in each scenario. The result of this operator is visually similar to *sum* but can be numerically different. For example, if a cell has value 0 for the belief for two of the experts' maps and 100 for another, the sum will be 100, but the average will be 33.3. This operator can be used to find places in which to deploy scarce resources. One example is that of deploying police forces according to criminality (see Figure 65).

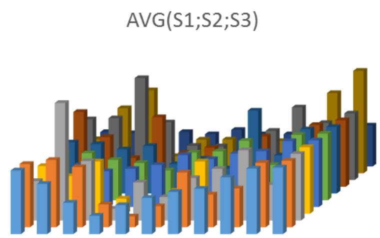


Figure 65. Average of scenarios.

10.1.3 **Subtraction operator**

The Subtraction operator subtracts the belief value of two or more scenarios at each evaluated location. This operator can be useful when it is necessary to evaluate the differences between one scenario and others. For example, if we have a possible flood scenario and a refuge scenario. Using subtract, the decision maker can easily identify the places to refuge with lower flood belief values (see Figure 66).

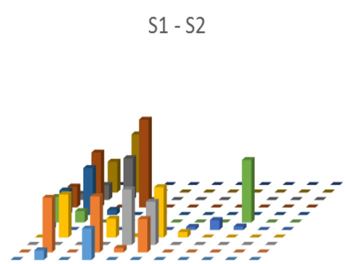


Figure 66. Subtraction of scenarios.

10.2 **Ordered Weight Average (OWA) operator**

An OWA operator is a weighted average given an order (Yager, 1988) (Yager, 2011). The OWA operator has been already used to combine data using the Dempster-Shafer Theory (Merigó, 2009) (Merigó & Casanovas, 2009). There are two basic types of OWA – when the values and weights are ordered from ascending or descending. Assuming that the values are ordered, only two results can be obtained. Given this, I defined 2 operators OWA-ASC and OWA-DESC.

10.2.1 **OWA-DESC**

When using the OWA-DESC operator, values and weights are ordered both in descending order. This combination can emphasize the biggest belief values of each scenario, avoiding that a certain important fact known by one of the experts could be ignored because of the simple average of numbers. For example, if a crime scenario has a large belief degree in a certain location, using average, this information can be mixed with lower degree values of the other scenarios (see Figure 67).

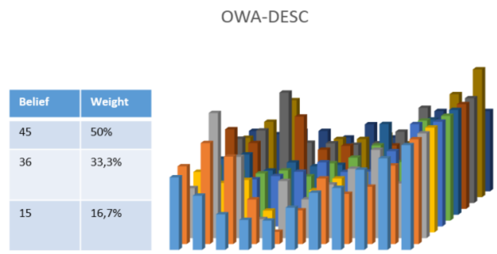


Figure 67. OWA-DESC applied to 3 scenarios.

10.2.2 **OWA-ASC**

When using the OWA-ASC operator, values are ordered in an ascending sequence, and the weights are also ordered in an ascending sequence. This combination emphasizes the belief when the values are constantly high in all scenarios. This operator is similar to average, but it is not susceptible to big, isolated values. It can be applied to allocate specific and limited resources that can support multiple scenarios. It can also be used to identify critical areas (see Figure 68).

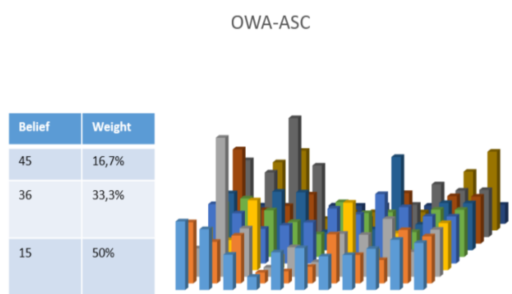


Figure 68. OWA-ASC applied to 3 scenarios.

10.2.3 **Weighted OWA operator**

The weighted OWA (WOWA) operator integrates the weighted average and the OWA operator in the same formulation. Thus, it can represent the importance of the variables and the attitudinal character of the decision maker in the same formulation, under or overestimating the initial data. The main advantage is that it can provide a more complete representation of the information considering any scenario that may occur between the minimum and the maximum.

10.2.4 **Induced OWA operator**

The induced OWA (IOWA) operator is an aggregation operator that follows the methodology of the OWA operator. However, instead of reordering in increasing or decreasing order, it uses order-inducing variables to determine the ordering process of the aggregation. This issue is important because many times the numerical values do not indicate the ordering of the information. The IOWA operator can be used to specify the scenario evaluation order. For example, if we want to order scenarios by their "source quality", it is possible to define an order using the u values in IOWA pairs. However, the order cannot be arranged by an optimal value, because we are working with belief degrees.

10.3 **Operators applied to the transportation network problem**

Based on the previous chapter, I envisage the development of an indicator, which could be used to measure in a more objective way the goodness of the prediction of the Belief Route and the Belief Congestion Route when compared with real data. This indicator can help calibrate the model and choose the hypotheses, which can better predict the demand as well as the travel time. It is also necessary to develop an indicator, which could show the advantage of one route compared with another in terms of satisfying a greater demand, while maintaining acceptable travel times.

I think that this indicator can be built by comparing and interpreting the BR and BCR. These comparisons can be made with two types of OWA operators: AOWA and DOWA. The emphasis of both operators can be interpreted as positive or negative depending on the BR objective. For example, if the decision maker is looking to supply the demand for transportation between two stations, he will look for a low AOWA and a high DOWA. Figure 69. features some possible interpretation of the four combinations of OWA values.

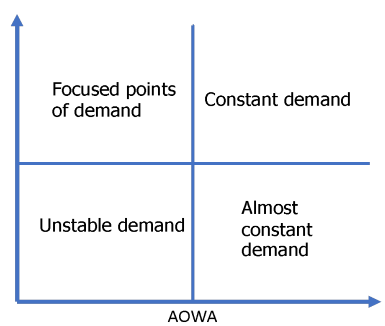


Figure 69. Interpretation of the four combinations of OWA values.

10.4 **Tool Design**

Based on the methodology described above I propose a tool allowing the construction, discussion and combination of suitability maps. The tool supports and implements the collaborative decision model previously explained r by an example in which a team of three experts analyze the tsunami scenario in Iquique. The members of the team have different expertise and/or information on identifying risk zones, people agglomerations and evacuation plans. I present screenshots of the most important stages of the tool with views of workspaces from the point of view of these three experts. The application obtains all the information needed in order to conduct this analysis from open public sources like OpenStreetMap (http://openstreetmap.org), public databases of Chilean ministries and National Office for Emergencies (ONEMI).

The tool implements divergence, argumentation and convergence with three different views. During the Divergence stage each expert creates one or more risk scenarios. A scenario is defined by the problem characteristics and the information an expert has in order to support her/his hypotheses, e.g., a school can be considered an object that can be used for applying a hypothesis like this: “The risk mass in schools during an earthquake is 50% in the morning and 75% in cinemas in the evening and weekends”. The tool extends the theory assuming that the risk mass decreases with the distance to the source location (cinemas and schools) according to a model specified by the expert. Another extension to the theory implemented by the tool is the relation between elements (facilities) and the hypothesis, which is modeled by so called rules: e.g., if the facility is located near a lake the risk mass should not be propagated to the water surface (see Figure 70).

The first view corresponds to the view a user observes when logging into the platform. It shows the ongoing projects in evaluation/discussion, and the scenarios that participants recently generated, which the user has not seen yet (Fig. 70). It also shows the participants of the project, in this case three: Nelson, Jonathan and Alvaro. According to the screenshot, Nelson is the one who is logged in, Jonathan is also online, but Alvaro is not. The screenshot shows there are three ongoing projects “Landslide prevention Chaiten”, “Evacuation planning Iquique”, and “Fire prevention Valparaiso”.

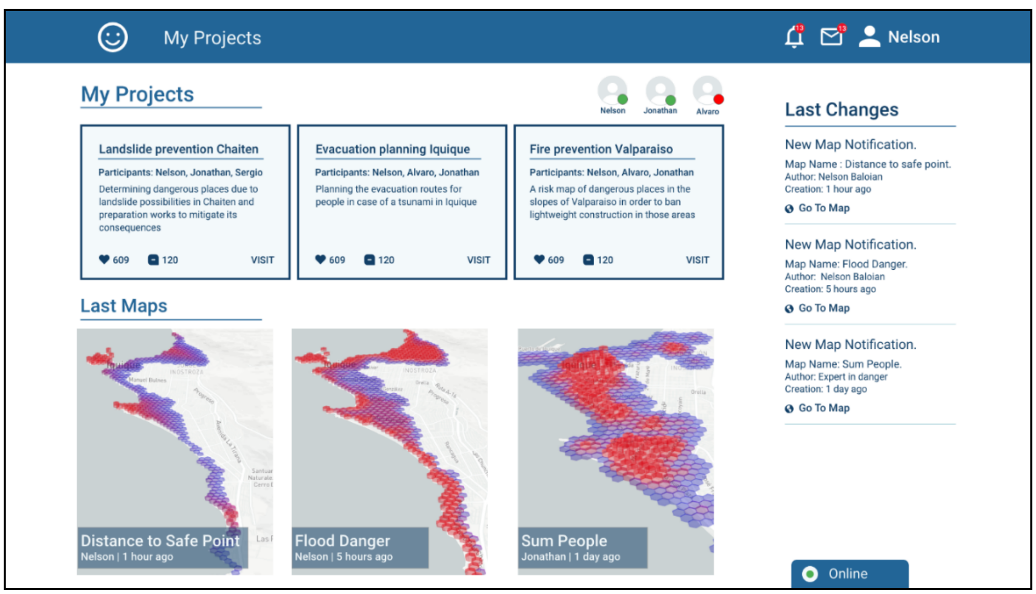


Figure 70. Screenshots of the application showing the available projects for user Nelson.

After selecting a project, the user can see the maps that have been generated for this project sorted by authorship (Figure 71) under the username of the team member who created them.

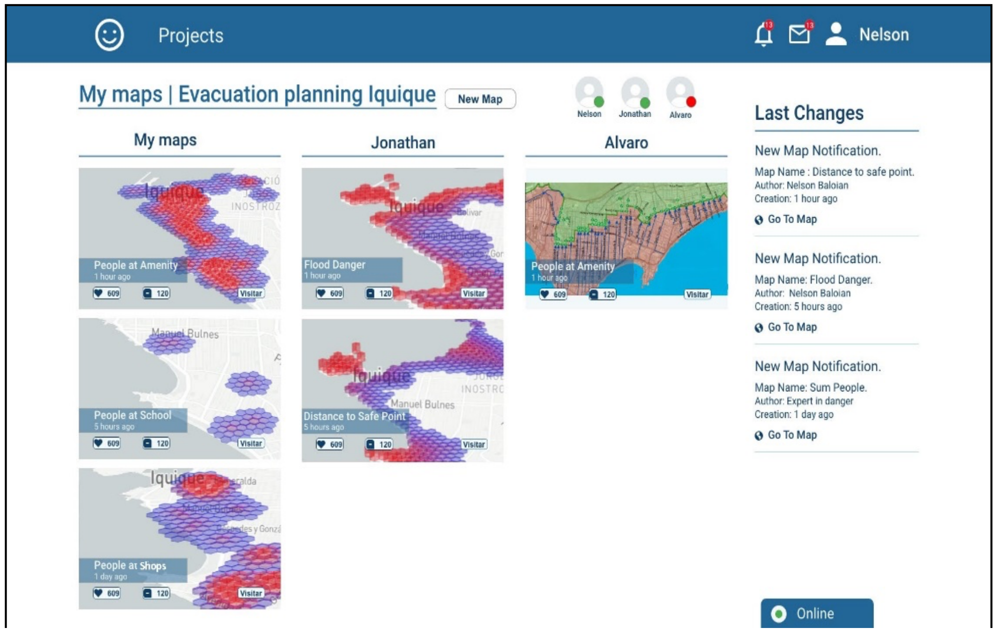


Figure 71. Screenshot showing the user Nelson exploring the maps that have been generated inside the “Evacuation planning Iquique” project.

By clicking the pushbutton labeled “new map” a user can generate a new map from scratch or by combining two or more already generated maps. First, I will explain the generation of a new map from scratch by an example.

Let us suppose that the user Nelson has an expertise in analyzing urban areas where there can be an agglomeration of people during a natural disaster. He has several hypotheses about why people concentrate in certain places and he wants to generate a map showing the number of people in the places near the coastal regions. Using the platform, he can include those hypotheses, giving them a percentage of mass and specifying exceptions that might occur (Figure 72). The platform takes this information and generates a belief map about which places a greater number of people could gather.

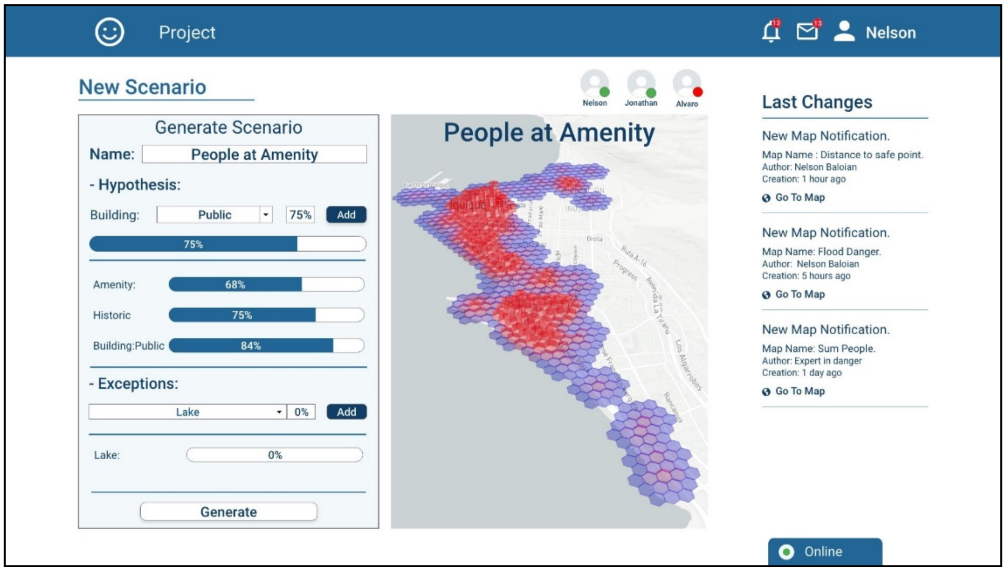


Figure 72. Screenshot of the application when generating a new map by defining hypotheses and their weights.

He thinks that there are usually high concentrations of people near amenities, and near shops and at some hours of the day, near educational buildings. In order to evaluate his hypotheses separately, Nelson generates three scenarios (maps), one for each.

On the other hand, Jonathan has expertise evaluating the level of risk that certain zones may have. He generates two risk scenarios, which evaluate the risk a person being in a certain geographical area of the city may face according to two parameters. The first one depicts the zones in which the risk of being flooded in case of a tsunami is determined according to their altitude (lower altitude means higher risk). In the second one, the risk depends on the distance that would be necessary to cover for a person to get out of the flood zones (longer distances mean greater risk).

Álvaro is an expert in evacuation, he has only uploaded a map with the current evacuation routes, and he must develop a new evacuation plan, which will use the scenarios created by Jonathan and Nelson.

As a team, the three experts decide that in order to generate an evacuation plan, they require only one scenario showing the possible people concentration and another one that shows the risk zones, so both Nelson and Jonathan must combine their scenarios.

To combine the scenarios, the platform provides an interface to use the operators of addition, subtraction, average and ordered weighted averages (OWA). Nelson decides that the right operation for combining his maps should be the addition, since people in the concentration areas will sum up in a real scenario. Figure 73 shows a screenshot of the tool when combining the maps resulting from estimating people in amenities and people at shops with the SUM operation, which is called "Sum People".

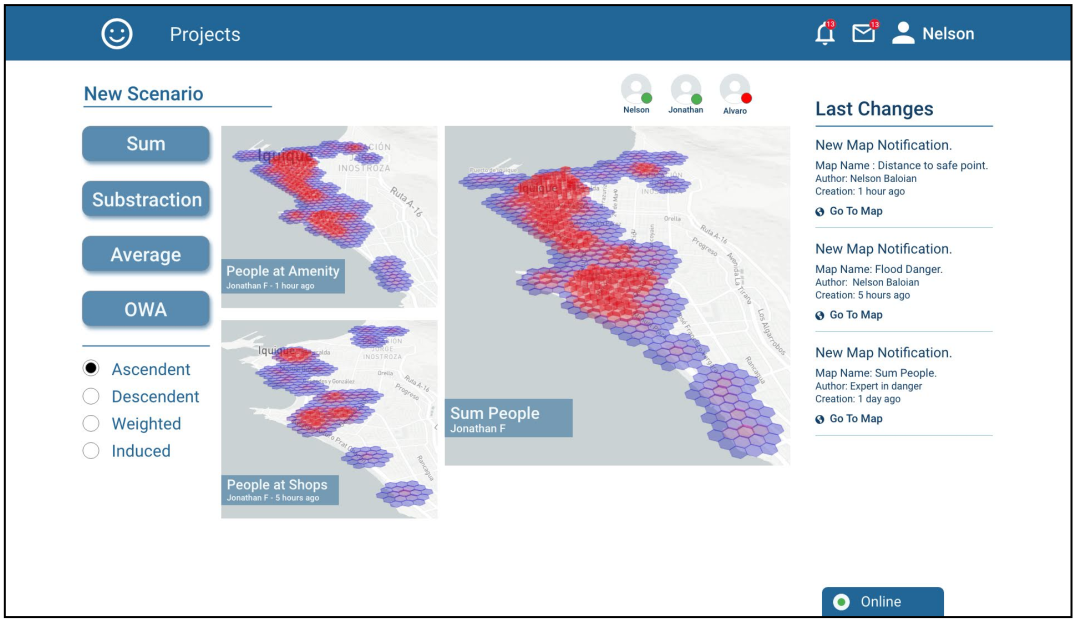


Figure 73. Screenshot of the application when Nelson combines the scenarios generated for people in amenities and people at shops with the sum operator.

Jonathan decides that the most appropriate operation for combining the risk scenarios generated by him was a decreasingly ordered weighted average (OWA.DESC), which gives greater weighting to higher values and lower weighting to lower values. In this manner the high-risk zones in any of the scenarios to be combined are maintained. The generated scenario is called "Danger DOWA " (Figure 74).

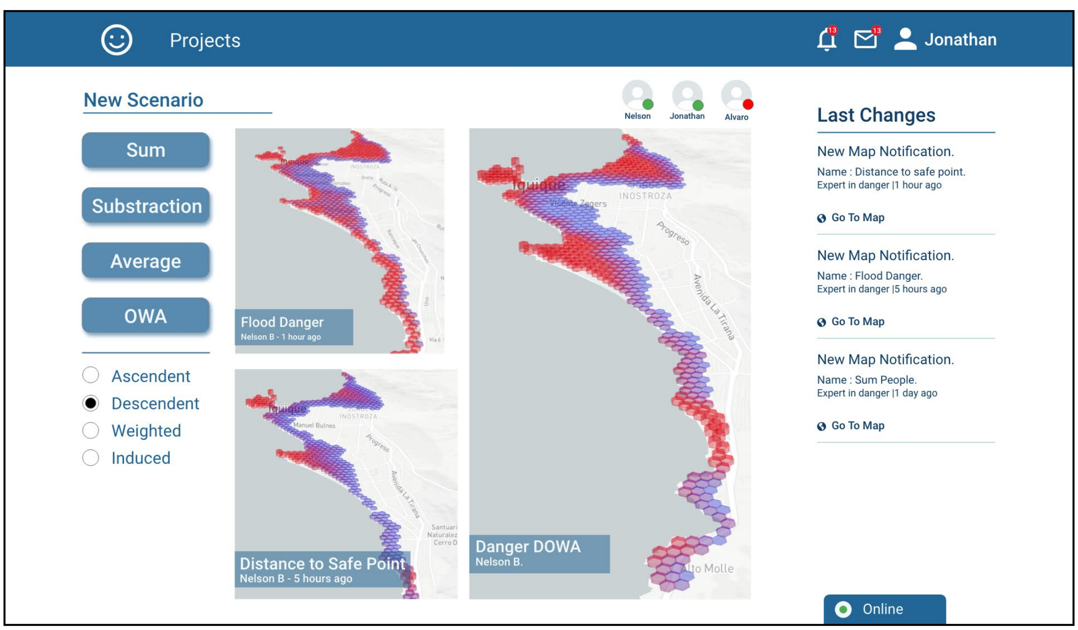


Figure 74. Screenshot showing Jonathan two combined two maps using the OWA-DESC operator.

Finally, Alvaro combines the two scenarios generated by Jonathan and Nelson in a single one called "People in Danger". In order to prioritize areas with high risk and at the same time areas with concentrations of people, he decides to combine using an OWA-ASC operator, being the areas that meet both conditions the most prominent (Figure 75).

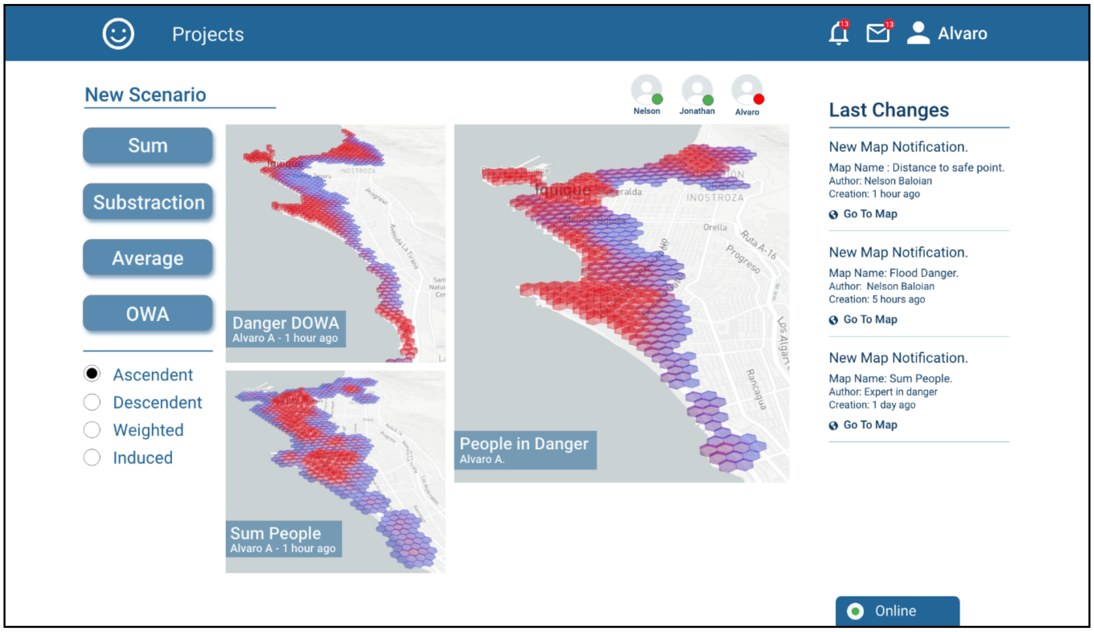


Figure 75. Screenshot showing Alvaro combining the two maps generated by him using the OWA-ASC operator.

The application also implements a tool that can be used for supporting the asynchronous discussion and the pertinence, validity or convenience of using a given scenario for preparing the actions needed to react in case of a disaster and/or plan the rescue and mitigation procedures. This tool consists of creating "Argumentation Objects", e.g., to discuss the need to carry out a differentiated evaluation for night periods. In order to give context to the argument, it is possible to attach previously created scenarios and assign them a discussion category.

The other project participants can review the argument with the attached antecedents (scenarios), discuss it and support the argument based on a voting system (Figure 76).

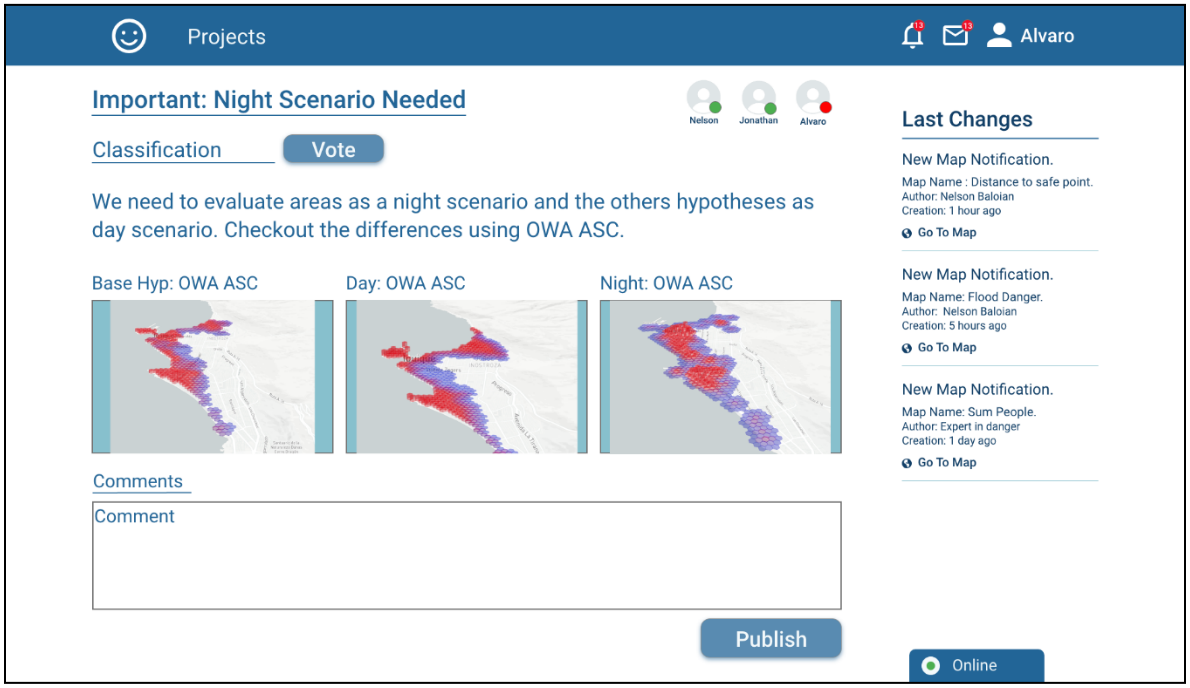


Figure 76. Screenshot showing Alvaro reviewing a scenario.

10.5 Discussion

This chapter presents a method and tool developed based on the theoretical foundations to support the preparedness stage of an emergency situation (Haddow, 2011). I validated this method initially by means of a simulation, detecting the value of developed functionalities, analyzing the case and drawing useful conclusions. However, the method itself must provide the functionalities to formally evaluate a tool designed using the method. According to (Antunes, al, 2012), there are several evaluation techniques for collaborative systems, this work can be classified as a Knowledge-based scenario method, in which the evaluation of these methods emphasizes variables pertaining more to the organization and group than to the individual performance. A Knowledge-based scenario tool can be evaluated using a Scenario-Based Evaluation (SBE) which consists of:

"using semi-structured interviews with users to discover scenarios or detailed descriptions of activities and claims about them. Then, focus groups validate these findings. The frequency and percentage of positive claims help quantify the organizational contributions of the system and the positive and negative claims about existing and envisioned features provide information to aid in the redesign." (Antunes, al, 2012).

The method provides the necessary mechanism to support the execution of a Scenario-Based evaluation. However, user-centric evaluation methods must be used to evaluate specific interactions with the tool.

In this chapter, I could check that if the Dempster-Shafer's theory of plausibility is appropriate for analyzing collaborative scenarios, since it allows multiple stakeholders to apply hypotheses and use uncertain data over geo-referenced information in order to draw conclusions about the necessary action that could be taken in case of a certain type of problem. Furthermore, I started to work in the aggregation of the contributions of various experts, giving the possibility of discussing the plausibility of the hypotheses stated by each expert participating in the preparedness team successfully testing part of the research question 2 (RQ2) by fulfilling the specific goal SG2a of this work: "*Developing a method to evaluate, compare and work collaboratively with previously generated scenarios*".

Chapter 11

Conclusions and Future Work

Collaborative work is relevant in SDSS because of the complex nature of the decision support problems where spatial data are involved (Coggins, Coops, Hilker, & Wulder, 2013) (Tran, O'Neill, & Smith, 2012) (Baloian, Frez, Janser, & Zurita, 2011) (Baloian, Frez, Zurita, & Milrad, 2012). However, the incorporation of collaborative work in SDSS leads us to the following research question:

> *How do we combine spatial data from different sources when each source can have different certainty levels?*

This research question can be separated into a spatial decision-making problem and the data combination problem, formulating two specific research questions:

> ***RQ1**: How do we generate suitability scenarios combining knowledge from various sources in a flexible and systematic process including multiple experts involved in the decision-making process?*

> ***RQ2:** How do we compare or combine two or more scenarios providing a better solution and more information on the problem itself?*

As it is explained in the next section, these research questions lead us to a hard problem to be solved: find a way to represent a hypothesis about how spatial objects interact with a decision support scenario. This particular problem has 2 dimensions: 1) find an appropriate spatial modeling technique for decision support scenarios. 2) Elaborate or adapt a mathematical decision support framework to spatial context.

In order to solve this problem and answer these questions, I started a research, exploring the problem, stablishing specific objectives, components and tests that allows us to develop a conceptual framework to deal with spatial ill-structured problems by supporting the complex steps in SDSS in a flexible but systematic manner. In order to provide an order, the conclusions and findings are described as a summary, showing were and how the research questions and problems were answered and solved.

11.1 Summary and conclusions of the conducted research

I started this research with an exploration of the related topics in order to find the problematic area, and how to improve a DM process in spatial contexts. To this end, I developed a decision support system for general smart-cities planning problems (Frez, Baloian, & Zurita, 2012). From this work, I learned about modeling and technical challenges of spatial decision support systems. This allowed us to learn how important it is to formulate a question in SDSS. For example, if we want to know where to place police officers, the question will be: "what is the likelihood of a crime occurring in this place?" This particular question requires modeling crime behavior. However, the crime modeling must consider complex relations and restrictions between the spatial area and the elements.

In order to understand how to work with georeferenced data, this research describes geographical data representations. I was able to identify two different types of structures depending on the organization methods: R-Tree and Quadtree-based. Both spatial indexes are designed to accomplish different purposes. Most of the analysis methods developed required an association between the elements and a large area. An indexing method based on Quadtree was used as a starting point of this work. However, at the end of this research, Uber's Hexagonal Hierarchical Spatial Index (similar to QuadTree) was used to test large-scale datasets.

As was mentioned, the main problem at this point was find a way to represent a hypothesis about how spatial objects interact with a decision support scenario. This particular problem has 2 dimensions: 1) find an appropriate spatial modeling technique for decision support scenarios. 2) Elaborate or adapt a mathematical decision support framework to spatial context.

In order to solve the mathematical problem, the Dempster-Shafer theory was extended to include spatial applications. For the spatial modelling technique, I combined fuzzy regions (Zhan & Lin,2003) described in Chapter 2 and the Belief Mass Functions (13) from the Dempster-Shafer theory to create spatial representation of hypothesis based on spatial objects. This

The definition of Belief Mass Functions (13) includes space, time and how these functions interact between them. The result was a function with three parameters: distance from the source, timestamp, and interaction interval values.

According to the belief definition, the resulting belief must be low if there is a lack of support for the hypothesis. Also, another important definition was the belief and plausibility meaning in spatial problems. Based on the Mass function parameters and the observed results, I can define plausibility and certainty in spatial context as follows:

- According to the plausibility definition, the function can be described as the combination of the masses of all elements available and discounted by (13).
- The certainty definition function can be described as the suitability mass using only masses of elements inside the evaluation cell and not discounted (or weighted by 1) in space or time.

These definitions are consistent with the Dempster-Shafer theory by the following reason: Dempster probability boundaries (P_*, P^*) I can say that $P_*(T)$ evaluates only the masses of elements inside an area and $P^*(T)$ includes the masses of all elements available. Finally, replacing $P(T)$ by the mass function I can provide a definition of Certainty and Plausibility in space.

$$Certainty = Mass(0)$$

$$Plausibility = Mass(\infty)$$

$$Certainty \leq Mass(d) \leq Plausibility$$

The Mass function always returns a value between belief and plausibility. This result is consistent with the definition of suitability described at the beginning of this work. Moreover, this property makes a mass value comparable with other scenarios only if they were generated using the same Mass function configuration. The configuration is directly related to the DSS problem, and in order to keep similar names, I choose to name the configuration set as *Suitability Object*.

The definition of suitability on this type of SDSS, is based on the definition of $P(T)$ described in (9), and it also asserts the condition $P_*(T) \leq P(T) \leq P^*(T)$, the reason is described in section 4.2. As a result, the $P(T)$ depends on distances, time, probability intervals and model restrictions. The difference between $P_*(T)$ and $P^*(T)$ is given by the suitable object characteristics in space. For example, $P^*(T)$ can be defined using $ms(d) = 1$ and $P_*(T)$ with $ms(d) = 0$.

The Dempster-Shafer theory also requires combination rules in order to "merge" the expert knowledge and for spatial problems I conclude that the *discount and combine method* is appropriate (see section 4.3). Furthermore, as most applications require a suitability map to represent an analysis result, I used the keyword "suitable" to indicate the use of the discount and combine method. However, other combination rules can be used.

In order to make this scenario generation method more "affordable", I developed a Scenario Generation Language (SGL). The SGL is designed to specify scenarios based on the Suitability Object and a location set. It allows the specification of multiple hypotheses with mass support (or expert knowledge). Finally, it provides a model restriction, temporal characteristics, the interaction between the suitability objects and data objects, saving methods, and basic GIS operations as intersections.

The SGL were implemented using existing geodatabases. Nevertheless, the SGL interpreter translates some parts of the expression into spatial predicates to provide compatibility. At this point, the research objectives were accomplished, but without a validation:

1. Adapting and applying belief functions and mathematical decision-making theory in spatial contexts, to combine data with different certainty levels and scenarios.
2. Developing a method to generate spatial decision-making scenarios in a flexible and systematic process to allow the inclusion collaboration between multiple stakeholders in the process.

In order to validate objectives 1 and 2, I developed three SDSS. The first one is designed for crime forecasting. It consists of predicting places with high risk of burglaries, assaults, or other types of crimes that usually occurs in public areas. The results presented prove that the method can be used to predict crimes within townships with similar performance forecasting algorithms as S-T GAM and spatial GLM (Generalized linear models). Our method was compared using standard metrics of the topic (HRP and TIP) obtaining an average of 51% TIP with a 10% HRP, which is better than the 38% TIP of S-T GAM with the same HRP. However, there is an important trade-off between the covered area and the distribution of crime in it. Higher crime concentrations are easier to predict.

One of the advantages of our method is that the hypotheses are easy to understand. This allows the experts a straightforward process to set and adjust the parameters of their models.

Crime forecasting algorithms are hard to compare with each other, mostly because of the lack of common datasets to test the programming code, which is not usually published. Also, forecasting algorithms can have good results in a specific area, but these results may vary in a different location. However, in this work I made an effort to test the method using different locations and data sources. It's important to mention that the critical part of Crime forecasting algorithms is developing a risk terrain model, mostly because it requires expert knowledge about the topic. In order to properly test the algorithm, the risk terrain model was created in collaboration with the police planning and prevention area of the country.

The results of the first application allow us to validate the specific goals 1 and SG2a: **"***Developing a method to generate spatial decision-making scenarios in a flexible and systematic process based on mathematical decision-making theory* **".**

The second application is focused on the public transportation network planning problem, supporting one of the hard questions of the area: how to estimate the transportation demand with incomplete data.

This application uses and combines the existing crowdsourced data to support a transportation network decision-making process. The information used provided by the facilities from OpenStreetMap and information about the traffic from Waze. With this information and the geospatial operation defined in SGL I defined two concepts: The Belief Route and the Belief Congestion. The Belief Route predicts the possible demand along an Origin-Destination route and the Belief Congestion Route predicts the congestion and thus the travel time for a bus on that route. This allows experts to make a preliminary evaluation of the route without investing time and resources to perform all the necessary tasks to gather the necessary information for a comprehensive evaluation. In addition, I presented an application that has all the functionalities necessary to perform this preliminary evaluation.

The application was validated by doing an experiment with a real existing route: I compared the prediction of the demand obtained with the Belief Route against the data obtained from the Transportation Ministry about the number of passengers getting in the buses serving that route in each bus stop. The result obtained was that the computed Belief Route predicted the demand quite well.

To test the usefulness of this approach, the Belief Route for a new route with the same origin and destination points but traveling through other streets was computed, finding out that this new route would have a greater demand, hence being more efficient with the transportation resources, but there are no available data to validate the demand in that route.

In order to have more data about the accuracy of the demand predictions a large-scale experiment comprising the whole Santiago region were conducted. The belief map was calculated on a grid consisting of 1390 hexagons; each hexagon has an area of 700 square meters. I used 80 million markings to build a real demand map using the same hexagons. I then calculated the Pearson correlation between both series. The correlation between the values associated with each hexagon is 0.53, which indicates a high correlation between the real and the estimated magnitude of demand. However, for cells with less than 5,000 passengers, the correlation is 0.09, so it is not possible to use it in areas of low demand. Between 5,000 and 70,000 passengers is 0.50; and between 50,000 and 100,000 passengers, the correlation is 0.73. As a reminder, Pearson's scale ranges from -1.0 to 1.0.

This application helps us test the combination of multiples sources of data and the comparison of different Scenarios which is part of specific goal 2b (SG2b): **"**SG2b: *Developing a method to evaluate, compare and work collaboratively with previously generated scenarios*".

The third application deals with collaborative part (SG2b), by testing the combination of multiples scenarios, following an example of evaluating the evacuation possibilities of a densely populated area in case of the occurrence of a tsunami.

The application was developed based on theoretical foundations to support the preparedness stage of an emergency situation. Using this case, I was able to check if the Dempster-Shafer's theory of plausibility is appropriate for analyzing collaborative scenarios. Nonetheless, for obvious reasons, I did not test the performance of the resulting evacuation scenarios. The application allows multiple stakeholders to apply hypotheses and use uncertain data over geo-referenced information in order to draw conclusions about the necessary action that could be taken in case of a certain type of problem. Furthermore, I started to work in the aggregation of the contributions of various experts using the operators defined in the SGL, giving the possibility of discussing about the plausibility of the hypotheses stated by each expert participating in the preparedness team.

11.2 Future Work

Although the methods presented can be used in several spatial-related problems, they could be further developed in a number of ways:

11.2.1 Automatic SGL statements generation from recurrent characteristics or patterns

Currently without well-qualified decision makers and experts, spatial decision support systems can only provide tools that allow us to analyze the problem faster and easily, but not necessarily in the proper way. However, I already know that there are a large number of spatial problems with recurrent characteristics or patterns. Some of these problems have been solved in some location, and we must ask ourselves if there is a way to determine others location with similar scenarios/conditions. This question motivates the future work; we want to find a way to determinate the SGL statement from a specific problem in a location automatically. If this is possible, we will be able to recommend existing solutions based on other "similar" locations. This would lead to easier modeling tools.

11.2.2 Non-spatial belief relationships

In this work, the belief propagation over space is limited by a distance relation. However, there are some relations that can impact in the belief mass without a specific distance relation. For example, a construction site can cause an increase of trucks on the highway, also increasing the risk of traffic accidents on the same highway at a different location. These kinds of problems can be modeled with the object-to-object relations with a mass transference function. However, the impact will be propagated to the data object geometry in different ways according to the type of relation, and the source of the mass. This would lead to better modeling capacities.

11.2.3 Integration with RDF and GEO-RDF

RDF (Profile, s.f.) is a standard model for data interchange in the web; it permits the merging of data even if the data schema is different. Furthermore, it has a triplet structure that allows linking data to mix structured and semi-structured data across different applications (Brodt, Nicklas & Mitschang, 2010) (Rackham, 2008). GEO-RDF is the standard to represent spatial information using RDF. The features of RDF allowed more complex and precise queries than the proposed WHERE clause in this work. This development will allow us to build complex scenarios merging data from different applications.

Bibliography

Aasman, J. (2008). Unification of geospatial reasoning, temporal logic, & social network analysis in event-based systems. Proceedings of the second international conference on Distributed event-based systems, (pp. 139-145).

Abadi, D., Carney, D., Cetintemel, U., Cherniack, M., Convey, C., Lee, S., Zdonik, S. (2003). Aurora: a new model and architecture for data stream management. The VLDB Journal, 12(2), 120-139.

AbuSabha, R., Peacock, J., & Achterberg, C. (1999). How to make nutrition education more meaningful through facilitated group discussions. Journal of the American Dietetic Association, 99(1), 72-76.

Allen, J. (1991). Time and time again: The many ways to represent time. International Journal of Intelligent Systems, 6(4), 341-355.

Ang, C., & Samet, H. (1990). Node distribution in a PR quadtree. Design and Implementation of Large Spatial Databases, 233-252.

Antunes, P., Sapateiro, C., Zurita, G., & Baloian, N. (2010). Integrating spatial data and decision models in an e-planning tool. In Collaboration and Technology (pp. 97-112). Springer Berlin Heidelberg.

Antunes, P., Herskovic, V., Ochoa, S. F., & Pino, J. A. (2012). Structuring dimensions for collaborative systems evaluation. ACM Computing Surveys (CSUR), 44(2), 8.

Auer, S., Lehmann, J., & Hellmann, S. (2009). LinkedGeoData: Adding a spatial dimension to the Web of Data. The Semantic Web-ISWC 2009, 731-746.

Baban, S. M., & Parry, T. (2001). Developing and applying a GIS-assisted approach to locating wind farms in the UK. Renewable energy, 24(1), 59-71.

Bahl, H. C. (1984). Decision-making theory and DSS design. ACM Sigmis Database, 15(4), 10-14.

Baloian, N., Frez, J., Janser, M., & Zurita, G. (2011, Oct). The Future Role of HTML5 in Mobile Situated Learning Scenarios. In S. Yu (Ed.), 10th World Conference on Mobile and Contextual Learning (mLearn) (pp. 249-257). Beijing, China: Normal University of Beijing.

Baloian, N., Frez, J., Zurita, G., & Milrad, M. (2012, Jul). Supporting Engineering Students Learning Wireless Network Planning Using Mobile, Positioning and Web Technologies. 12th IEEE International Conference on Advanced Learning Technologies (ICALT) (pp. 509-514). Rome, Italy: IEEE Computer Society Press.

Band, L. E., Hwang, T., Hales, T. C., Vose, J., & Ford, C. (2012). Ecosystem processes at the watershed scale: Mapping and modeling ecohydrological controls of landslides. Geomorphology, 137(1), 159-167.

Barceló, J. e. (2007). Online microscopic traffic simulation supports real-time traffic-management strategies. Society for Industrial and Applied Mathematics, 40(9).

Basso, F. B. (2018). Real-time crash prediction in an urban expressway using disaggregated data. Transportation Research Part C: Emerging Technologies (86), 202-219.

Bayer, R. (1997). The universal B-tree for multidimensional indexing: General concepts. Worldwide Computing and Its Applications, 198-209.

Beckmann, N., Kriegel, H., Schneider, R., & Seeger, B. (1990). The R*-tree: an efficient and robust access method for points and rectangles (Vol. 19). ACM.

Bhat, C. R. (1999). Activity-based modeling of travel demand. In Handbook of transportation Science (pp. 35-61). Boston: Springer.

Binaghi, E., Luzi, L., Madella, P., Pergalani, F., & Rampini, A. (1998). Slope instability zonation: a comparison between certainty factor and fuzzy Dempster--Shafer approaches. Natural hazards, 17(1), 77-97.

Bogomolov, A. L. (2014). Once upon a crime: towards crime prediction from demographics and mobile data. In Proceedings of the 16th international conference on multimodal interaction (pp. 427-434). ACM.

Bordogna, G., Pagani, M., & Pasi, G. (2010). Imperfect Multisource Spatial Data Fusion Based on a Local Consensual Dynamics. In Uncertainty Approaches for Spatial Data Modeling and Processing (pp. 79-94). Springer.

Bowers, K. &. (2004). Prospective Hot-Spotting: The Future of Crime Mapping?. British Journal of Criminology 44, 641-658.

Brickley, D. (2006). Basic geo (WGS84 lat/long) vocabulary. URL: http://www. w3. org/2003/01/geo/, W3C Semantic Web Interest Group.

Brickley, D., & Miller, L. (2010). FOAF vocabulary specification 0.98. Namespace Document, 9.

Brodt, A., Nicklas, D., & Mitschang, B. (2010). Deep integration of spatial query processing into native rdf triple stores. Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, (pp. 33-42).

Burggraf, D. (2006). Geography markup language. Data Science Journal, 5(0), 178-204.

Burrough, P. A., McDonnell, R., Burrough, P. A., & McDonnell, R. (1998). Principles of geographical information systems (Vol. 333). Oxford university press Oxford.

Caplan, J. M. (2010). Risk terrain modeling manual: Theoretical framework and technical steps of spatial risk assessment for crime analysis. Rutgers Center on Public Security.

Castillo, E. e. (2008). The observability problem in traffic models: algebraic and topological methods. IEEE Transactions on Intelligent Transportation Systems, 2(9), 275-287.

Castillo, E., Jimenez, P., Menendez, J., & Conejo, A. (2008, June). The Observability Problem in Traffic Models: Algebraic and Topological Methods. Intelligent Transportation Systems, IEEE Transactions on, 9(2), 275-287.

Cats-Baril, W. L. (1987). Decision support systems for ill-structured problems: an empirical study. Decision Sciences,, 18(3), 350-372.

Center, P. (n.d.). Pop Center - street robbery analisis. Retrieved from http://www.popcenter.org/problems/street_robbery/

Cerreta, M., & De Toro, P. (2012). Urbanization suitability maps: a dynamic spatial decision support system for sustainable land use. Earth System Dynamics, 3, 157-171.

Chandrasekaran, S., Cooper, O., Deshpande, A., Franklin, M., Hellerstein, J., Hong, W., . . . Shah, M. (2003). TelegraphCQ: continuous dataflow processing. Proceedings of the 2003 ACM SIGMOD international conference on Management of data, (pp. 668-668).

Checkland, P. (2000). Soft systems methodology: a thirty year retrospective. Systems Research and Behavioral Science, 17, S11--S58.

Chen, H., Wood, M., Linstead, C., & Maltby, E. (2011). Uncertainty analysis in a GIS-based multi-criteria analysis tool for river catchment management. Environmental modeling \& software, 26(4), 395-405.

Chountas, P., Rogova, E., & Atanassov, K. (2010). Expressing Hierarchical Preferences in OLAP Queries. In Uncertainty Approaches for Spatial Data Modeling and Processing (pp. 61-77). Springer.

Chourabi, H. e. (2012). Understanding smart cities: An integrative framework. Hawaii International Conference on.

Coggins, S. B., Coops, N. C., Hilker, T., & Wulder, M. A. (2013). Augmenting forest inventory attributes with geometric optical modeling in support of regional susceptibility assessments to bark beetle infestations. International Journal of Applied Earth Observation and Geoinformation, 21, 444-452.

Company, T. P. (n.d.). Predpol. Retrieved from http://www.predpol.com/

Crossland, M. D., Wynne, B. E., & Perkins, W. C. (1995). Spatial decision support systems: An overview of technology and a test of efficacy. Decision Support Systems, 14(3), 219-235.

Cugola, G., & Margara, A. (2012). Processing flows of information: From data stream to complex event processing. ACM Computing Surveys (CSUR), 44(3), 15.

Date, C., & Darwen, H. (1987). A Guide to the SQL Standard (Vol. 3). Addison-Wesley Reading (Ma) et al.

De Smith, M. J. (2007). Geospatial analysis: a comprehensive guide to principles, techniques and software tools. . Troubador Publishing Ltd.

De Tr'e, G., Dujmovi'c, J., & Van de Weghe, N. (2010). Supporting spatial decision making by means of suitability maps. In Uncertainty Approaches for Spatial Data Modeling and Processing (pp. 9-27). Springer.

De Vreede, G., & Briggs, R. O. (2005). Collaboration engineering: designing repeatable processes for high-value collaborative tasks. System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on, (pp. 17c--17c).

Dean, J. &. (2008). MapReduce: simplified data processing on large clusters. Communications of the ACM, 51(1), 107-113.

Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. The annals of mathematical statistics, 325-339.

Densham, P. J. (1991). Spatial decision support systems. Geographical information systems: Principles and applications, 1, 403-412.

Dimensionally Extended nine-Intersection Model (DE-9IM). (n.d.). Retrieved from https://en.wikipedia.org/wiki/DE-9IM

Eck, J. E., & Spelman, W. (1987). Problem-solving: Problem-oriented policing in Newport News.

Eck, J. E. (2001). Policing and crime event concentration. The process and structure of crime: Criminal events and crime analysis, 9.

Ekanayake, G., & Dayawansa, N. (2003). Land suitability identification for a production forest through GIS techniques. Department of Agricultural Engineering, Faculty of Agriculture, University of Peradeniya, Peradeniya Sri Lanka. Map India Conference.

Erl, T. (2016). Service-oriented architecture: analysis and design for services and microservices. Prentice Hall Press.

FAO, J., & FOODS, M. (2004). Food and Agriculture Organization of the United Nations. Worldwide regulations for mycotoxins in food and feed in 2003.

Fernández, R. M. (2008). Data collection and calibration of passenger service time models for the Transantiago system. European Transport Conference.

Ferretti, V., & Montibeller, G. (2016). Key challenges and meta-choices in designing and applying multi-criteria spatial decision support systems. Decision Support Systems, 84, 41-52.

Fishburn, P. C. (1981). Subjective expected utility: A review of normative theories. Theory and Decision, 13(2), 139-199.

Frez, J., Baloian, N., Pino, J. A., & Zurita, G. (2014, September). Cooperative work for spatial decision making: an emergencies management case. In CYTED-RITOS International Workshop on Groupware (pp. 113-120). Springer, Cham.

Frez, J., Baloian, N., Zurita, G., & Pino, J. A. (2014, May). Dealing with incomplete and uncertain context data in geographic information systems. In Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD) (pp. 129-134). IEEE.

Frez, J., Baloian, N., & Zurita, G. (2012). Software Platform to build geo-collaborative Systems supporting design and planning. Computer Supported Cooperative Work in Design (CSCWD), 2012 IEEE 16th International Conference on, (pp. 350-357).

Frez, J., Baloian, N., & Zurita, G. (2012). Supporting Collaboration for Smarter City Planning. In J. Bravo, D. LÃ³pez-de-IpiÃ±a, & F. Moya (Eds.), Ubiquitous Computing and Ambient Intelligence (Vol. 7656, pp. 338-341). Springer Berlin Heidelberg. Retrieved from http://dx.doi.org/10.1007/978-3-642-35377-2_47

Frez, J., Baloian, N., & Zurita, G. (2012). Supporting collaboration for smarter city planning. In Ubiquitous Computing and Ambient Intelligence (pp. 338-341). Springer.

Frez, J., Baloian, N., & Zurita, G. (2013). Getting serious about integrating decision support mechanisms into Geographic Information Systems. Computer Science and Information Technologies (CSIT), 2013, (pp. 1-11).

Gao, S., Paynter, J., & Sundaram, D. (2004). Flexible support for spatial decision-making. System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on, (pp. 10--pp).

Gemitzi, A., Falalakis, G., Eskioglou, P., & Petalas, C. (2011). Evaluating landslide susceptibility using environmental factors, fuzzy membership functions and GIS. Global Nest Journal, 13, 28-40.

Ghalsasi, S. (2009). Critical success factors for event driven service oriented architecture. Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human, (pp. 1441-1446).

Ghayoumian, J., Ghermezcheshme, B., Feiznia, S., & Noroozi, A. A. (2005). Integrating GIS and DSS for identification of suitable areas for artificial recharge, case study Meimeh Basin, Isfahan, Iran. Environmental Geology, 47(4), 493-500.

Groff, E. R. (2001). Mapping an opportunity surface of residential burglary. Journal of Research in Crime and Delinquency, 3(38), 257-278.

Groff, E. R. (2002). Forecasting the future of predictive crime mapping. Crime Prevention Studies,(13), 29-58.

Groff, E. R. (2007). Simulation for theory testing and experimentation: An example using routine activity theory and street robbery. Journal of Quantitative Criminology, 2(23), 75-103.

Groff, E. R.-5. (2002). Forecasting the future of predictive crime mapping. Crime Prevention Studies. popcenter.

Gross, K., Digate, C., & Lee, E. (1994, #feb#~1). Event-driven rule-based messaging system. Event-driven rule-based messaging system. Google Patents.

Grove, L. E. (2012). reventing repeat victimization: A systematic review. Brottsförebyggande rådet. The Swedish National Council for Crime Prevention.

Guihaire, V. a.-K. (2008). Transit network design and scheduling: A global review. Transportation Research Part A: Policy and Practice, 42(10), 1251-1273.

Güting, R. (1994). An introduction to spatial database systems. The VLDB Journal, 3(4), 357-399.

Güting, R., & Schneider, M. (1995). Realm-based spatial data types: the ROSE algebra. The VLDB Journal, 4(2), 243-286.

Guttman, A. (1984). R-trees: a dynamic index structure for spatial searching (Vol. 14). ACM.

Haddow, G. B. (2011). Introduction to Emergency Management. . Elsevier.

Harley, J., & Britain, G. (1975). Ordnance Survey maps: a descriptive manual. Ordnance Survey Southampton.

Harries, K. A. (1999). Mapping crime: Principle and practice.

Harries, K. D. (1980). Crime and the Environment. Springfield, IL: Thomas.

Harrison, C. e. (2010). Foundations for smarter cities. IBM Journal of Research and Development,, 4(54), 1-16.

Heilig, G. (2012). World urbanization prospects the 2011 revision. United Nations, Department of Economic and Social Affairs (DESA), Population Division, Population Estimates and Projections Section. New York.

Heugens, P. P., & van Oosterhout, J. (2001). To boldly go where no man has gone before: integrating cognitive and physical features in scenario studies. Futures, 33(10), 861-872.

Hwang, J., Balazinska, M., Rasin, A., Cetintemel, U., Stonebraker, M., & Zdonik, S. (2005). High-availability algorithms for distributed stream processing. Data Engineering, 2005. ICDE 2005. Proceedings. 21st International Conference on, (pp. 779-790).

Inagaki, T. (1991). Interdependence between safety-control policy and multiple-sensor schemes via Dempster-Shafer theory. IEEE Transactions on Reliability, 40, 182-188.

Inagaki, T. (1991). Interdependence between safety-control policy and multiple-sensor schemes via Dempster-Shafer theory. Reliability, IEEE Transactions on, 40(2), 182-188.

Jankowski, P., Nyerges, T. L., Smith, A., Moore, T., & Horvath, E. (1997). Spatial group choice: a SDSS tool for collaborative spatial decisionmaking. International journal of geographical information science, 11(6), 577-602.

Jankowski, P. (2018). Behavioral decision theory in spatial decision-making models. In Handbook of Behavioral and Cognitive Geography. Edward Elgar Publishing.

Jiang, B. (2013). Head/tail breaks: A new classification scheme for data with a heavy-tailed distribution. The Professional Geographer , 65(3), 482-494.

Johnson, S. D. (2008). Stable and fluid hotspots of crime: differentiation and identification. Built Environment, 1(34), 32-45.

Keen, P. G. (1987). Decision support systems: the next decade. Decision Support Systems, 3(3), 253-265.

Kennedy, J. (2011). Particle swarm optimization. In Encyclopedia of machine learning (pp. 760-766). Boston: Springer.

Kennedy, L. W., & Van Brunschot, E. G. (2009). The risk in crime. Rowman & Littlefield Publishers.

Kitsuregawa, M., Harada, L., & Takagi, M. (1989). Join strategies on KD-tree indexed relations. Data Engineering, 1989. Proceedings. Fifth International Conference on, (pp. 85-93).

Klein, G. (1997). The recognition-primed decision (RPD) model: Looking back, looking forward. Naturalistic decision making, 285-292.

Klein, M., & Visser, U. (2004). Semantic web challenge 2003. IEEE Intelligent Systems, 19(3), 31-33.

Klyne, G. &. (2002). Date and time on the internet: Timestamps.

Kolas, D. (2008). Supporting spatial semantics with SPARQL. Transactions in GIS, 12, 5-18.

Lassila, O., Swick, R., & others. (1998). Resource description framework (RDF) model and syntax specification.

Laurent, A. (2010). Fuzzy Multidimensional Databases. In Uncertainty Approaches for Spatial Data Modeling and Processing (pp. 43-59). Springer.

les Rackham, C. R. (2008). SPARQL query language for RDF. W3C recommendation, 15.

les Rackham, C. R. (2011). Linked data: the new paradigm for geographic information? jigsaw, 38, 10-12.

Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. Journal of Behavioral Decision Making, 14(5), 331-352.

Liu, S. e. (2010). Modeling and simulation on multi-mode transportation network. . IEEE International Conference on Computer Application and System Modeling.

Liu, S., Chen, S., Liu, Z., & Mao, B. (2010, Oct). Modeling and simulation on multi-mode transportation network. Computer Application and System Modeling (ICCASM), 2010 International Conference on, 7, pp. V7-438-V7-443.

Lopez-Pellicer, F., Silva, M., Chaves, M., Javier Zarazaga-Soria, F., & Muro-Medrano, P. (2010). Geo linked data. Database and Expert Systems Applications, (pp. 495-502).

Madlener, R., Kowalski, K., & Stagl, S. (2007). New ways for the integrated appraisal of national energy scenarios: the case of renewable energy use in Austria. Energy Policy, 35(12), 6060-6074.

Malafant, K., Fordham, D., Uso, J., Brebbia, C., Power, H., & others. (1998). GIS, DSS and integrated scenario modeling frameworks for exploring alternative futures. Advances in ecological sciences. Volume 1: Ecosystems and sustainable development. Proceedings of a conference, Pe{\~n}iscola, Spain, 14-16 October 1997., (pp. 669-678).

Malczewski, J. (2006). GIS-based multicriteria decision analysis: a survey of the literature. International Journal of Geographical Information Science, 20(7), 703-726.

McGuinness, D., Van Harmelen, F., & others. (2004). OWL web ontology language overview. W3C recommendation, 10(2004-03), 10.

Meagher, D. (1982). Geometric modeling using octree encoding. Computer Graphics and Image Processing, 19(2), 129-147.

Merigó, J. M., & Casanovas, M. (2009). Induced aggregation operators in decision making with the Dempster-Shafer belief structure. International Journal of Intelligent Systems, 24(8), 934-954.

Merigó, J. M. (2009). Induced aggregation operators in decision making with the Dempster-Shafer belief structure. International Journal of Intelligent Systems, 8(24), 934-954.

Michelson, B. (2006). Event-driven architecture overview. Patricia Seybold Group, 2.

Mohler G. O., S. M. (2011). Self-Exciting Point Process Modeling of Crime. American Statistical Association, 106 (493), 100-108.

Morris, A. (2003). A framework for modeling uncertainty in spatial databases. Transactions in GIS, 7(1), 83-101.

Morris, A., & Jankowski, P. (2001). Combining fuzzy sets and databases in multiple criteria spatial decision making. In Flexible Query Answering Systems (pp. 103-116). Springer.

Morris, A., Jankowski, P., Bourgeois, B. S., & Petry, F. E. (2010). Decision Support Classification of Geospatial and Regular Objects Using Rough and Fuzzy Sets. In Uncertainty Approaches for Spatial Data Modeling and Processing (pp. 3-8). Springer.

N Baloian, J. F. (2011). The role of HTML5 in mobile situated learning scenarios. Proceedings of the 10th World Conference on Mobile and Contextual Learning, . Beijing, China.

Nagel, K. a. (1992). A cellular automaton model for freeway traffic. In Journal de physique I (pp. 2221-2229).

Natural Resources Management and Environment Department, F. (1976). A framework for land evaluation. (FAO) Retrieved from http://www.fao.org/docrep/x5310e/x5310e00.htm

Nyerges, T. L., & Jankowski, P. (2012). Regional and urban GIS: a decision support approach. Guilford Press.

onk, K. M. (2010). Street robbery. Washington, DC: US Department of Justice, Office of Community Oriented Policing Services,.

openstreetmap. (n.d.). openstreetmap. Retrieved from http://wiki.openstreetmap.org/

Ouellet, C., & Sherk, L. (1967). Woody ornamental plant zonation: III. Suitability mapfor the probable winter survival of ornamental trees and shrubs. Canadian Journal of Plant Science, 47(4), 351-358.

Papoulis, A. &. (2002). Probability, random variables, and stochastic processes. Tata McGraw-Hill Education.

Park, N.-W. (2011). Application of Dempster-Shafer theory of evidence to GIS-based landslide susceptibility analysis. Environmental Earth Sciences, 62(2), 367-376.

Paschke, A., Boley, H., Kozlenkov, A., & Craig, B. (2007). Rule responder: RuleML-based agents for distributed collaboration on the pragmatic web. (RuleML-2008).

Paschke, A., Kozlenkov, A., & Boley, H. (2007). A homogeneous reaction rule language for complex event processing.

Pearson, J. M., & Shim, J. (1995). An empirical investigation into DSS structures and environments. Decision Support Systems, 13(2), 141-158.

Perry, W. L. (2013). Predictive policing: The role of crime forecasting in law enforcement operations. Rand Corporation.

Pop Center. (n.d.). Retrieved from http://www.popcenter.org/

Power, D. J. (2015). Decision support systems. John Wiley & Sons, Ltd.

Profile, G. (n.d.). w3. (GeoRDF) Retrieved from http://www. w3. org/wiki

Ramsey, F. P. (1931). Truth and probability (1926). The foundations of mathematics and other logical essays, 156-198.

Ratcliffe., J. H. (2006). A temporal constraint theory to explain opportunity-based spatial o. Journal of Research in Crime and Delinquency, 3(43), 261-291.

Reed, C., Singh, R., Lake, R., Lieberman, J., & Maron, M. (2006). An introduction to georss: A standards based approach for geo-enabling rss feeds. White Paper OGC.

Reed, S., & Lenat, D. (2002). Mapping ontologies into Cyc. AAAI 2002 Conference Workshop on Ontologies For The Semantic Web, (pp. 1-6).

Ringland, G., & Schwartz, P. P. (1998). Scenario planning: managing for the future. John Wiley & Sons.

Robinson, V. B. (2010). Exploring the sensitivity of fuzzy decision models to landscape information inputs in a spatially explicit individual-based ecological model. In Uncertainty Approaches for Spatial Data Modeling and Processing (pp. 29-42). Springer.

Rocchini, D. (2014). Fuzzy species distribution models: a way to represent plant communities spatially. Journal of Vegetation Science, 25(2), 317-318.

Rosenheck, R. A. (2001). Organizational process: A missing link between research and practice. Psychiatric Services, 52(12), 1607-1612.

Sözer, A., Oğuztüzün, H., & Petry, F. E.& others. (2010). Querying fuzzy spatiotemporal databases: implementation issues. In Uncertainty Approaches for Spatial Data Modeling and Processing (pp. 97-116). Springer.

Saaty, T. L. (1990). How to make a decision: the analytic hierarchy process. European journal of operational research, 48(1), 9-26.

Samet, H. (1984). The quadtree and related hierarchical data structures. ACM Computing Surveys (CSUR), 16(2), 187-260.

Samet, H. (1990). The design and analysis of spatial data structures (Vol. 85). Addison-Wesley Reading MA.

Samet, H. (1995). Spatial data structures. Modern Database Systems, The Object Model, Interoperability and Beyond, 361-385.

Schwartz, P. (1991). The art of the long view. New York: Doubleday.

Seligmann, P., Wijers, G., & Sol, H. (1989). Analyzing the structure of IS methodologies, an alternative approach. Proceedings of the First Dutch Conference on Information Systems, Amersfoort, The Netherlands, EU.

Sellis, T., Roussopoulos, N., & Faloutsos, C. (1987). The r+-tree: A dynamic index for multi-dimensional objects.

Sentz, K. &. (2002). Combination of evidence in Dempster-Shafer theory. 4015. Albuquerque: Sandia National Laboratories.

Shafer, G. (1976). A mathematical theory of evidence (Vol. 1). Princeton university press Princeton.

Sherman, L. W. (1989). Repeat call address policing: The Minneapolis RECAP experiment. Report to the National. Washington, DC: Crime Control.

Shim, J. P. (2002). Past, present, and future of decision support technology. Decision support systems, 33(2), 111-126.

Shim, J. P. (2002). Past, present, and future of decision support technology. Decision support systems, 33(2), 111-126.

Shine, K. P. (2009). The global warming potential—the need for an interdisciplinary retrial. Climatic Change, 96(4), 467-472.

Simon, H. A., Dantzig, G. B., Hogarth, R., Plott, C. R., Raiffa, H., Schelling, T. C., . . . Winter, S. (1987). Decision making and problem solving. Interfaces, 17(5), 11-31.

sociology.org.uk. (n.d.). sociology.org.uk Crime Patterns. Retrieved from http://www.sociology.org.uk/pblsdca.pdf

Stylianou, K. D.-A. (2019). Big Data and Road Safety: A Comprehensive Review. In Mobility Patterns. Big Data and Transport Analytics, 297-343.

Tangestani, M. H., & Moore, F. (2002). The use of Dempster--Shafer model and GIS in integration of geoscientific data for porphyry copper potential mapping, north of Shahr-e-Babak, Iran. International Journal of Applied Earth Observation and Geoinformation, 4(1), 65-74.

Tatbul, N., & Zdonik, S. (2004). Window-aware Load Shedding for Data Streams. Tech. rep., Technical Report CS-04-13, Brown University, Computer Science.

Tatbul, N., & Zdonik, S. (2006). Window-aware load shedding for aggregation queries over data streams. PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, 32, p. 799.

Tran, L. T., O'Neill, R. V., & Smith, E. R. (2012). A watershed-based method for environmental vulnerability assessment with a case study of the Mid-Atlantic region. Environmental Impact Assessment Review, 34, 58-64.

Vatant, B., & Wick, M. (2006). Geonames ontology. Geonames ontology.

Vuchic, V. R. (2002). Urban public transportation systems. Philadelphia: University of Pennsylvania.

Walz, A., Lardelli, C., Behrendt, H., Grêt-Regamey, A., Lundström, C., Kytzia, S., & Bebi, P. (2007). Participatory scenario analysis for integrated regional modeling. Landscape and Urban Planning, 81(1-2), 114-131.

Wang, X. a. (2012). The spatio-temporal modeling for criminal incidents. Security Informatics, 1(1), 2.

Wang, X. a. (2012). The spatio-temporal modeling for criminal incidents. Security Informatics 1.1.

Wegener, M. (2004). Overview of land-use transport models. In Handbook of transport geography and spatial systems 5 (pp. 127-146).

Weigand, H. (2011). The pragmatics of event-driven business processes. Proceedings of the 7th International Conference on Semantic Systems, (pp. 211-218).

Weisburd, D. W. (2006). Does crime just move around the corner? A controlled study of spatial displacement and diffusion of crime control benefits. Criminology, 3(44), 549-592.

Winter-Ebmer., S. R. (2001). Identifying the effect of unemployment on crime. Journal of Law and, 1, 44.

Wong, S. T. (1994). Preference-based decision making for cooperative knowledge-based systems. ACM Transactions on Information Systems (TOIS), 12(4), 407-435.

Wood, L. J., & Dragicevic, S. (2007). GIS-based multi-criteria evaluation and fuzzy sets to identify priority sites for marine protection. Biodiversity and Conservation, 16(9), 2539-2558.

Xu, L. a. (2008). Bi-objective urban road transportation discrete network design problem under demand and supply uncertainty. IEEE International Conference on Automation and Logistics.

Xu, L., & Gao, Z. (2008, Sept). Bi-objective urban road transportation discrete network design problem under demand and supply uncertainty. Automation and Logistics, 2008. ICAL 2008. IEEE International Conference on, (pp. 1951-1955).

Yager, R. R. (1987). On the Dempster-Shafer framework and new combination rules. Information sciences, 41(2), 93-137.

Yager, R. R. (1987). On the Dempster-Shafer framework and new combination rules. Information sciences, 42, 93-137.

Yager, R. R. (1988). On ordered weighted averaging aggregation operators in multi-criteria decision-making. Systems, Man and Cybernetics, IEEE Transactions on, 18, 183-190.

Yager, R. R. (2011). Recent developments in the ordered weighted averaging operators: theory and practice. Springer Science & Business Media.

Yang, L. a. (2009). A Multimodal Composite Transportation Network Model and Topological Relationship Building Algorithm. International Conference on Environmental Science and Information Application Technology.

Yang, L., & Wan, B. (2009, July). A Multimodal Composite Transportation Network Model and Topological Relationship Building Algorithm. Environmental Science and Information Application Technology, 2009. ESIAT 2009. International Conference on, 2, pp. 505-509.

Zadeh, L. (1984). Book review: A mathematical theory of evidence. The Al Magazine.

Zadrożny, S., & Kacprzyk, J. (2010). Bipolar Queries: A Way to Deal with Mandatory and Optional Conditions in Database Querying. In Uncertainty Approaches for Spatial Data Modeling and Processing (pp. 117-132). Springer.

Zhang, L. (1994). Representation, independence, and combination of evidence in the Dempster-Shafer theory. Advances in the Dempster-Shafer theory of evidence, (pp. 51-69).

Zurita, G., Baloian, N., & Frez, J. (2013). Using the cloud to develop applications supporting geo-collaborative Situated Learning. Future Generation Computer Systems.