Fuzzy-based Spatial Knowledge Discovery in Service of Efficient Coastal Erosion Risk Assessment

Amaneh Jadidi[1], Mir Abolfazl Mostafavi[1], and Yvan Bédard[1]

1 Center of Research for Geomatics, Laval University, Quebec City, QC
amaneh.jadidi-mardkheh.1@ulaval.ca , mir-abolfazl.mostafavi@scg.ulaval.ca , yvan.bedard@scg.ulaval.ca

Abstract

The need for effective spatial knowledge discovery in risk assessment increases as the amount of the spatial data which is handled and analyzed for risk assessment has enlarged. Such large variety of spatial data should be integrated from multiple sources presenting different types, semantics, and level of details weather geometrically, temporally, or thematically. Therefore, spatial knowledge discovery has been introduced to help decisions-makers with clear and efficient understanding of such complex cases of data integration. Risk, by itself, is a definition which is given to demonstrate the impact of a physical or human-induced phenomenon causing disasters or harming human-life. Spatial modeling and representing risk through risk zones (as crisp object) bring up the notion of the fiat nature of risk which is based on some interested criteria introduced by stakeholders and experts. In this regard, the issue of information vagueness appears in semantic definition of risk zones as well as geometries. Fiat nature of risk zones adds more complexity on risk assessment and risk zone representation as the conventional GIS methods and analytical information systems support such situations marginally. This paper presents the results of a PhD thesis in this regard entitled as "A Comprehensive Spatial Multidimensional Conceptual Model (SMCM) to Assess the Risk of Coastal Erosion Based on Fuzzy Set Theory". To do so, the fuzzy spatial representation of risk zones and fuzzy multi-scale representation are dealt in order to improve risk assessment process and decision-making. Incorporating Geospatial Business Intelligence (GeoBI) paradigm and fuzzy set theory within Spatial Online Analytical Processing (SOLAP) necessitates developing a framework for designing key elements of a spatial datacube: dimensions, members, hierarchies, measures, and facts based on risk components (hazard, elements at risk and vulnerability). Thereafter, formalizing and adapting fuzzy model in such datacube should be carried out. Finally, some aggregation methods for fuzzy spatial data should be developed to deal with multi-scale representation of risk zones as well as for navigating within such system (roll-up and drill-down operations).

Background and Relevance

Coastal Erosion Risk (CER) is a complex and dynamic process which is the result of several multi-scale and spatiotemporal interactions between hazard and vulnerability of the elements at risk (Alexander, 2000; Blaikie, Cannon, Davis, & Wisner, 2004; Daudé et al., 2009; Varnes, 1984). The inherent complexity of coastal erosion makes assessment of the associated risks complex as well (Cheng, Molenaar, & Stein, 2009). In fact, modeling the erosion, predicting its behavior over time, and estimating its impact on assets and human lives requires processing a large amount of heterogeneous data from multiple sources with different types and levels of details (Jadidi, Mostafavi, Bédard, Long, & Grenier, 2013). Therefore, an information system is required to accommodate and integrate the available heterogeneous data from different sources. This system should also provide users and decision makers with on-the-fly aggregation, analysis, synthesis, and reporting. Therefore, three main issues are conferenced in any Coastal Erosion Risk Assessment (CERA) for any selected region. The first issue is related to the huge amount of data with diverse semantics and from different sources. This introduces uncertainty and information vagueness right at the beginning of a risk assessment process (Bakillah, 2012; Sboui, 2010). The second issue is related to the distinct and multi-scale characteristics of risk based on the needs and interests of the participants and decision-makers (Jadidi, Mostafavi, Bédard, et al., 2013). This emphasizes the need for a risk-aware hierarchical data aggregation method in any CERA process (Cheng et al., 2009; Jadidi, Mostafavi, Bédard, et al., 2013). The third issue is that coastal risk zones have vague semantic definitions and are spatially or temporally uncertain as well as inherently multi-scale. The representation of such zones has been under intense investigations (Cheng et al., 2009; Dilo, By, & Stein, 2007). However, the literature introduces isolated solutions regarding exclusives issues. Indeed, an integrated approach covering all three under one umbrella (in case of CERA) is still missing.

Probabilistic and Possibilistic approaches are the two main approaches which are widely employed to characterize information vagueness associated with risk modeling and representation (Aerts, Goodchild, & Heuvelink, 2003; Choa, Choi, & Kim, 2003; Cowell & Zeng, 2003; Darbra, Eljarrat, & Barceló, 2008; Fisher, Cheng, & Wood, 2007; Kentel & Aral, 2007). The flexibility of the possibilistic approaches such as fuzzy-based model for dealing with information vagueness suggests an efficient solution for spatial representation of risk (Kentel & Aral, 2007). Nevertheless, the idea of fuzzy spatial data model is still a young topic in the Geospatial Business Intelligence (GeoBI) communities. Few efforts have tried to embed the vagueness for spatiotemporal information in SOLAP (Bejaoui, 2009; Siqueira & Ciferri, 2012). But, these efforts are mainly based on extended spatial crisp models. Indeed, none of the cited works use Fuzzy Set Theory explicitly to characterize information vagueness and to integrate it in SOLAP in a systemic way. Moreover, the methods to aggregate spatial measures resulting from uncertain data (i.e. information vagueness) into multiple hierarchical dimensions (and to represent them) are still missing.

Methods and Data

To overcome the stated issues, this research has been carried out in three phases (see Figure 1):

- First, an analytical conceptual framework is proposed to develop a spatial multidimensional conceptual model (SMCM) for CERA. This framework includes four main steps including needs analysis, data inventory, definition of risk components (i.e. hazard, elements at risk, and associated vulnerability index), and finally designing a SMCM which includes identifying analysis dimensions and measures to calculate associated risk (Jadidi, Mostafavi, Bédard, et al., 2013).
- Second, Fuzzy Set Theory is selected to overcome the issue of information vagueness (Dilo, 2006; Kanjilal, Liu, & Schneider, 2010; Molenaar & Cheng, 2000; Pauly & Schneider, 2010; Robinson, 2003; Schneider, 2003a, 2003b). Our proposed approach for CERA is inspired from (Schneider, 2003b). A conceptual framework is then proposed based on Schneider's model consisting of five main steps: (1) Identify hazard, (2) Elaborate vulnerability index, (3) Discretization (grid structures) , (4) Fuzzification, and (5) Fuzzy representation (Jadidi, Mostafavi, & Bédard, 2013). A MATLAB code is developed to perform Fuzzification and Fuzzy representation steps. A membership function is assigned to each cell while the risk value in each cell and for each indicator is assessed using the respective Fuzzy IF-THEN rules extracted from risk components. The membership functions serve to determine the level of risks. After calculating the risk for each indicator, the results are aggregated using an appropriate fuzzy operator (union, intersection, and difference) to calculate the overall risk.
- Third, the proposed fuzzy-based approach is then formalized and adapted for the SMCM which was proposed in phase I (Jadidi, Bédard, & Mostafavi, 2013). This necessitates identifying where the fuzziness happens and how this can be embedded into the database and managed while performing the queries and representing the results. In this regards,

spatial datacube key elements (i.e. spatial level's attributes, spatial levels, spatial members, spatial dimension, spatial hierarchy, spatial measures, and spatial facts) are redefined. Tow concepts of fuzzy partition (a grid-based cell) and fuzzy hierarchy relations are proposed to build such system. Overlay (union, intersection, difference), fusion, and some arithmetic operators are also planned and developed for such system for spatial aggregation purpose.

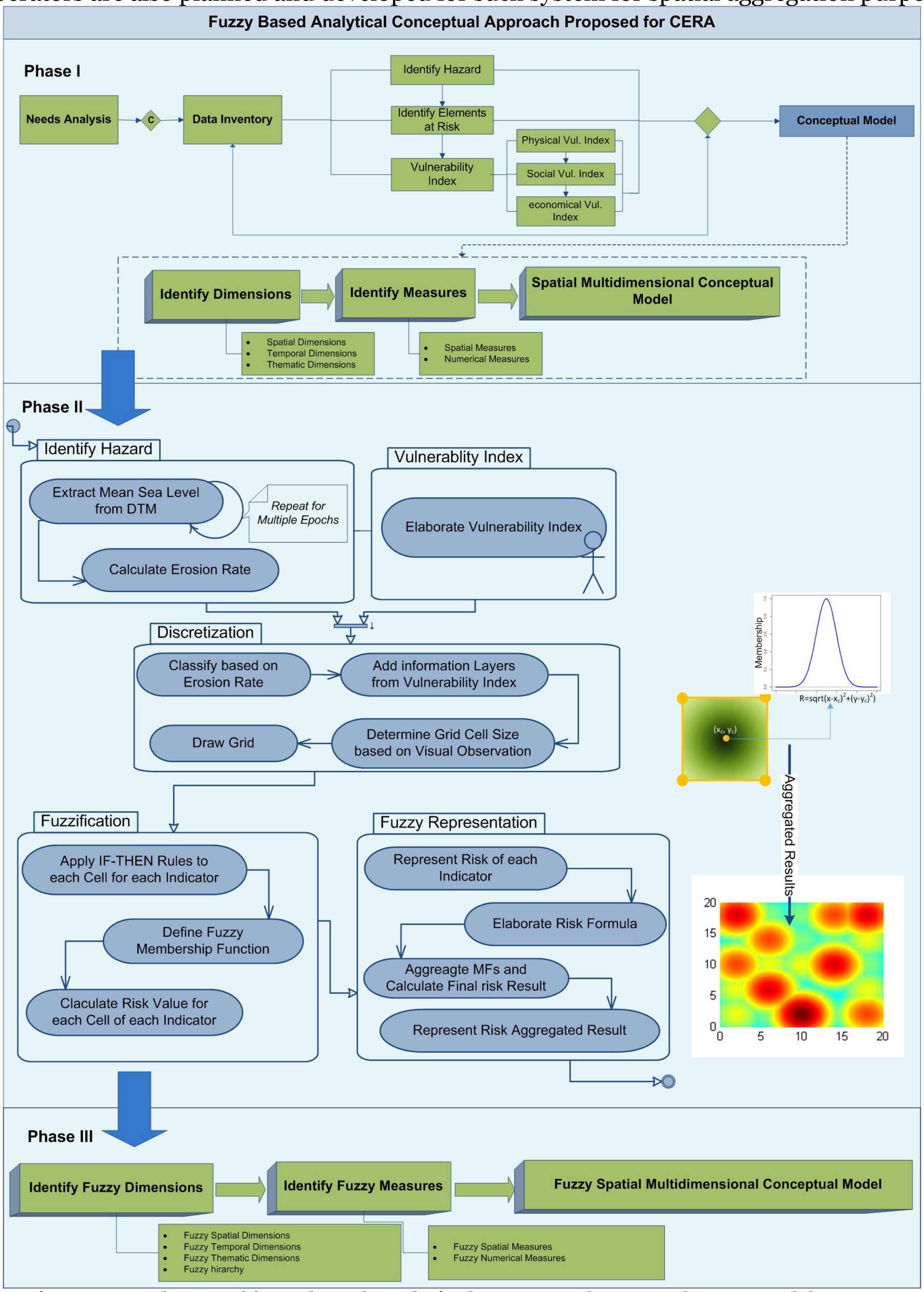

Figure 1: A scheme of fuzzy based analytical conceptual approach proposed for CERA

The region along the coast of the St-Laurence River in Percé, near the tip of the Gaspé Peninsula in Eastern Quebec, Canada is identified as a potential study site to implement and validate the proposed methods. Table 1 presents a list of the datasets and the related parameters that are used for CERA in this case study.

Source	**Vulnerability Indicators (** v_i **)**
LiDAR Data	Slop , DEM, Erosion Rate
Technical and Research Reports	Protection structure, Infrastructure situation, Type of Coastline
Google Open Street Map	Analysis units
Statistic Canada	Population, density of population, economical values
Quebec Prov. Transport Dept.	Road network

Table 1: Data sources and extracted coastal erosion risk parameters (vulnerability indicators), details are provided in (Jadidi, Mostafavi, Bédard, et al., 2013)

Results

As result, a fuzzy SMCM is designed for CERA by following phase I and II (see Figure 1). The CERA model consists of 15 dimensions (two spatial dimensions, one temporal and 12 thematic dimensions) and 13 spatial measures (8 measures with geometry, 5 numeric measures) (Jadidi, Mostafavi, Bédard, et al., 2013). Spatially referenced facts and their respective membership degrees are stored in a fuzzy spatial datacube. Assigning the label of "fuzzy" to a spatial datacube requires at least one dimension of the datacube to be defined based on fuzzy model or fuzzy hierarchy relations. The proposed Fuzzy SMCM is developed based on a star schema model. For the estimation of the measures, a star-query model, which is a common technique in star schema modeling, is proposed in this work. The proposed fuzzy SMCM is adapted for both vector and grid-based information. A Spatial dimension is designed to store in a grid-based structure that is associated to a vector-based data structure (e.g. census division in a spatial dimension). Performing the aggregation on this dimension permits navigating from the grid-based to the vector-based structure and vice versa. The overlay and fusion of fuzzy members use respectively the concept of the intersection and the union of fuzzy objects in fuzzy set theory with a combination of arithmetic operators such as SUM, Average, and Weighted Average. The overlay operator combines two fuzzy partitions to form a new fuzzy partition whilst the fusion operator allows the generalization of a fuzzy partition. Based on the presented fuzzy operators, five aggregation scenarios are investigated to support a fuzzy model in such spatial multidimensional datacube: (1) crisp aggregation onto crisp data, (2) fuzzy aggregation onto crisp data, (3) crisp aggregation onto fuzzy data, (4) fuzzy aggregation onto fuzzy data, and (5) fuzzy aggregation onto mixed of crisp and fuzzy data. The implementation of the proposed fuzzy SMCM is identical to the implementation of typical crisp spatial datacubes. The only difference is in measures calculation where a fuzzy operator, instead of a crisp one, should be applied to respective dimensions, members, and hierarchy relations either crisp or fuzzy data.

Conclusions

An analytical conceptual framework was proposed to overcome the aforementioned limits by accomplishing a comprehensive CERA system through the Geospatial BI paradigm. Fuzzy spatial datacubes are essential to perform more comprehensible knowledge discovery for effective decision-making. A fuzzy-logic-based approach was proposed in this paper to deal information vagueness. This concept was then embedded into a spatial datacube through redefining the spatial datacube elements (dimensions, members, hierarchies, measure and facts) as fuzzy dimensions, fuzzy members, fuzzy hierarchies, fuzzy measures, fuzzy facts, and requisite fuzzy aggregation operators (union, intersection, difference, overlay, and fusion). A

Fuzzy SMCM was developed for CERA through the proposed frameworks. The proposed approach was applied to a study region in Percé, Quebec, Canada for validation purpose. By comparing the results mapped by crisp object model and fuzzy object model, it was demonstrated that the uncertainty which is related to object definition has noticeable influence on the final result.

References

Aerts, J., Goodchild, M. F., & Heuvelink, G. (2003). Accounting for Spatial Uncertainty in Optimization with Spatial Decision Support Systems. *Transactions in GIS*, *7*(2), 211–230.

Alexander, D. (2000). *Confronting Catastrophe: New Perspectives on Natural Disasters* (p. 282). Terra Publishing.

Bakillah, M. (2012). *Real Time Semantic Interoperability in Adhoc Networks of Geospatial Databases : Disaster managment*. Université Laval, Qubec City, Canada.

Bejaoui, L. (2009). *Qualitative topological relationships for objects with possibly vague shapes: imolication on the specification of topological integrity constraint in transactional spatial databases and in spatial data warehouses*. *Geomatics*. Université Laval, Quebec City.

Blaikie, P., Cannon, T., Davis, I., & Wisner, B. (2004). *At Risk Natural Hazards, People's Vulnerability and Disaster* (second., p. 496). New York: Routledge.

Cheng, T., Molenaar, M., & Stein, A. (2009). Fuzzy Approach for Integrated Coastal Zone Management. In Y. X. (Ed.), *Remote Sensing and Geospatial Technologies for Coastal Ecosystem Assessment and Management* (pp. 67–90). Springer Berlin Heidelberg.

Choa, H.-N., Choi, H.-H., & Kim, Y.-B. (2003). A risk assessment methodology for incorporating uncertainties using fuzzy concepts. *Reliability Engineering and System Safety*, *78*, 173–183.

Cowell, P., & Zeng, T. (2003). Integrating uncertainty theories with GIS for modeling coastal hazards of climate change. *Marine Geodesy*, *26*(1-2), 5–18. doi:10.1080/01490410390181261

Darbra, R., Eljarrat, E., & Barceló, D. (2008). How to measure uncertainties in environmental risk assessment. *TrAC Trends in Analytical Chemistry*, *27*(4), 377–385. doi:10.1016/j.trac.2008.02.005

Daudé, E., Provitolo, D., Dubos-Paillard, E., Gaillard, D., Eliot, E., Langlois, P., … Saint-Gérand, T. (2009). Spatial risks and complex systems : methodological perspectives. In Aziz-Alaoui M. A. and Bertelle C. (Ed.), *From System Complexity to Emergent Properties Understanding Complex Systems*. Springer.

Dilo, A. (2006). *Representation of and reasoning with vagueness in spatial information: A system for handling vague objects*. *Geoinformatics*. ITC, Enschede.

Dilo, A., By, R. A. de, & Stein, A. (2007). A system of types and operators for handling vague spatial objects. *International Journal of Geographical Information Science*, *21*(14), 397–426.

Fisher, P., Cheng, T., & Wood, J. (2007). Higher Order Vagueness in Geographical Information: Empirical Geographical Population of Type n Fuzzy Sets. *GeoInformatica*, *11*(3), 311–330. doi:10.1007/s10707-006-0009-5

Jadidi, A., Bédard, Y., & Mostafavi, M. A. (2013). Fuzzy Spatial Datacube for Multi-Scale Coastal Risk Assessment: Towards Fuzzy Spatial Aggregation to Support GeoSemantics Decision Models. *International Journal of Geographical Information Science*, *XX*(XX), Submited.

Jadidi, A., Mostafavi, M. A., & Bédard, Y. (2013). Spatial Representation of Coastal Risk: A Fuzzy Approach for Dealing with Uncertainty. *ISPRS International Journal of Geo-infromation*, *XX*(XX), Submited.

Jadidi, A., Mostafavi, M. A., Bédard, Y., Long, B., & Grenier, E. (2013). Using Geospatial Business Intelligence Paradigm to Design a Multidimensional Conceptual Model for Efficient Coastal Erosion Risk Assessment. *Journal of Coastal Conservation*, *17*(3), 527–543. doi:10.1007/s11852-013-0252-5

Kanjilal, V., Liu, H., & Schneider, M. (2010). Plateau Regions : An Implementation Concept for Fuzzy Regions in Spatial Databases and GIS. In *Proceedings of the 13th International Conference on Information Processing and Management of Uncertainty in KnowledgeBased Systems* (Vol. 6178 LNAI, pp. 624–633). doi:10.1007/978-3-642-14049-5_64

Kentel, E., & Aral, M. M. (2007). Risk tolerance measure for decision-making in fuzzy analysis: a health risk assessment perspective. *Stochastic Environmental Research and Risk Assessment*, *21*(4), 405–417. doi:10.1007/s00477-006-0073-2

Molenaar, M., & Cheng, T. (2000). Fuzzy spatial objects and their dynamics. *ISPRS Journal of Photogrammetry & Remote Sensing*, *55*, 164–175.

Pauly, A., & Schneider, M. (2010). VASA: An algebra for vague spatial data in databases. *Information Systems*, *35*(1), 111–138. doi:10.1016/j.is.2009.05.003

Robinson, V. B. (2003). A Perspective on the Fundamentals of Fuzzy Sets and their Use in Geographic Information Systems. *Transactions in GIS*, *7*(1), 3–30.

Sboui, T. (2010). *A Conceptual Framework and a Risk Managment Approach for the Interoperability between Geospatial Datacubes*. Université Laval.

Schneider, M. (2003a). Vague Spatial Data Types. *Advances in Spatial Databases, Lecture Notes in Computer Science*, *44*, 81–108.

Schneider, M. (2003b). Design and implementation of finite resolution crisp and fuzzy spatial objects. *Data & Knowledge Engineering*, *44*(1), 81–108. doi:10.1016/S0169-023X(02)00131-3

Siqueira, T., & Ciferri, C. de A. (2012). Towards Vague Geographic Data Warehouses. In N. Xiao (Ed.), *GIScience 2012, Lecture Note in Computer Science 7478* (pp. 173–186). Springer.
Varnes, D. J. (1984). *Commission on Landslides and Other Mass-Movements-IAEG Landslide Hazard Zonation: A Review of Principles and Practices*. Paris: UNESCO Press.