

Comparison between Composite Index Solution Surfaces with Fuzzy Composite Index Decision Surfaces

J. Rubén G. Cárdenas

*Risk Management Programme
Understanding and Managing Extremes School
Pavia, Italy
ru.gonzcard@gmail.com*

Àngela Nebot

*Soft Computing research group at Intelligent Data Science
and Artificial Intelligence Research Center
Universitat Politècnica de Catalunya-BarcelonaTech
Barcelona, Spain
angela@cs.upc.edu*

Francisco Mugica

*Soft Computing research group at Intelligent Data Science
and Artificial Intelligence Research Center
Universitat Politècnica de Catalunya-BarcelonaTech
Barcelona, Spain
fmugica@cs.upc.edu*

Abstract—Composite indices are used in many of the traditional approaches to measure risk to natural hazards. However, such indices are often built assuming linear interdependencies between the aggregated components, comprising in this way any realistic representation of the intricate and unseen processes that are behind each component along with their mutual influences when it comes to encapsulate a complex reality in a single construct. In this paper we used a Fuzzy Inference Systems type Mamdani to aggregate physical seismic risk and social vulnerability indicators without assuming linearity. The aggregation is made by establishing rules (if-then type) over the indicators in order to get a fuzzy composite index. We generated solution surfaces representing the outcome of a widely known linear aggregation method and the fuzzy system. Finally we performed a qualitative comparison between both to highlight their differences and possible consequences in terms of risk management

Index Terms—Fuzzy Sets, Risk Management, Seismic Hazard, Seismic Vulnerability, Inference System, Composite Indices

I. INTRODUCTION

The use of composite indices is a well established practice over many field of knowledge. Ranging from simple marketing schemes to natural sciences, a composite index is a suitable way of representing a sort of encapsulated or compressed view of a more complex reality. A composite index is formed by indicators which are in turn raw data that has been collected, organized and presented in order to give information about a status or a condition of interest. Even if indicators can be used solely as independent entities of measure, most of the time they get aggregated to form indices. The aim is to include into a single result a range of variables described by the indicators in order to achieve a model, which covers reality in a more comprehensive manner.

In this way, a composite index may be used to handle intangible process, hidden trends or matters of wider significance than that which is actually measured and present them all as quantifiable entities that can be compared across space and time scales. One of the advantages of composite indices, is their potential to be used as a reference point to assess effects and trends or to be included later on into a more comprehensive or complex model.

A composite index is then the final outcome of method intended to compound a form of influence of all considered indicators to describe a certain attribute. Since an index output will consist most of the time on a single value or number, the normal way to compound indicators influences is by aggregation methods. Once such influences are joined, the index can be considered as a representation of new attribute or dimension that might be poverty, development, commitment, vulnerability, etc. Examples are the Human Developed Index [8], the Commitment to Development Index [13], or the Global Risk index [14].

Generally, there are two main assumptions when aggregation among indicators is about to be performed: (1) they are independent; (2) the influence among indicator follows a linear pattern. A linear aggregation is useful though, just in the case that all indicators share the same measurement units, since they can be added without losing generality as stated in [11].

One major assumption of a linear aggregation scheme is the preference independence between indicators (or, that among indicators there are no synergies), if this assumptions is to be followed, a linear aggregation allows an accurate assessment of the contribution of each indicator separately and allows that such contributions be added to obtain a final score without losing generality.

This research was supported by project TIN2016-79576-R

However, the preference independence assumption seems to be an unrealistic assumption to represent complex processes, for example a social system, where the interrelations among variables are not totally understood and the infinite flux of synergies make that the entire system behaves as a complex system. Opting for a linear aggregation method will cause that an unquantified amount of information be lost and that any composite index resulting of this method will not represent properly the influence and information of its individual indicators as explained in [10].

Using first an artificial dataset we compared the surface solution for the seismic level composite index of the Moncho's equation with the decision surface of a fuzzy inference system based in logic rules and using an artificial dataset. With real data we implemented both indices in Bogota, Colombia performed a detailed analysis in both surfaces in this context to analyze their differences and respective advantages. Section II describes Moncho's equation and its indicators, section III resumes the analysis of the surfaces obtained with artificial data and the implementation in Bogota and section IV highlights the differences between both surfaces when they are adapted to a particular range of indicator values observed in this area.

II. THE MONCHO EQUATION

The use of composite indices is a well established practice over the field of Risk Management [2]. Its primary goal is the definition of a scale of measurement that can be used for different geographical areas. The scope of the measurement can be either the impact of an external agent, such as natural hazard or *ex-ante* conditions of resilience or vulnerability, whether social or structural, over a human settlement. Since the impacts of such agent will not be proportional, the definition of "levels" comes into place. Therefore we can refer to these measurements as "level of risk" or "levels of vulnerability".

In the field of seismic risk, the Moncho's equation as it can be seen in [4] and [5] has been used to produce an integrated seismic risk index by different agencies around the world, such as the probabilistic risk model CAPRA [15] or the Global Earthquake Model of the Eucentre in Italy [3]. Moncho's equation states that the level of seismic risk can be understood as the result of two different dimensions, a physical or structural and a social dimension. Its hypothesis is that any final seismic risk level would be the result of the physical risk, aggravated by social conditions and lack of resilience capacities

Moncho's equation is developed as a composite index. After a careful selection of indicators, importance weights are related to each of them through a hierarchical process, and then a linear aggregation method is implemented.

This model comprises a *Physical Risk* (R_{Ph}), representing the level of risk of structural elements, and an *Aggravation Coefficient* (F) where socioeconomic fragilities and lack of resilience of the context are included. As we said before, the final seismic risk level would be the result of the physical

risk, aggravated by social conditions and lack of resilience capacities.

A calculation of the total risk index is then obtained by a direct application of Equation 1,

$$R_T = R_{Ph} (1 + F) \quad (1)$$

where R_T is the total risk, R_{Ph} is the physical risk and F is a aggravation coefficient.¹

III. NON-LINEAR AGGREGATION USING A FUZZY INFERENCE SYSTEM

In order to compare the performance of the linear aggregation method used in Moncho's equation with a non linear aggregation scheme, we developed a simple Fuzzy Inference Systems type Mamdani (FIS) using the same three dimensions defined previously: physical risk, aggregation and total risk. Our goal was to aggregate different values of physical risk with values of aggravation through the FIS to obtain a final total risk level as an outcome. We defined 5 linguistic labels for the fuzzification of the numerical values of aggravation (very low, low, medium, high, very high) and 6 classes for the physical risk (cero, very low, low, medium, high, very high). The output space (total risk for us) uses 7 linguistic classes (very low, low, medium-low, medium, medium-high, high, very high) all represented by triangular membership functions as well. We used triangular membership functions both for inputs and output in order to represent the gradual change between classes. The FIS final mapping is based on 26 if-then type weightless rules describing the potential interactions of the physical risk and aggravation inputs, as they map the values of a total risk level. Table 1 shows the set of fuzzy rules used in our experiment. These rules are easily interpretable and manageable, and allows a more clear understanding of how a certain total risk level was achieved.

The universe of discourse used in the FIS tries to resemble the numerical range on which Moncho's equation is moving. Since physical risk and aggravation are normalized values in the linear scheme, the universe of discourse of these inputs in the FIS was defined to be inside of the interval [0 1]. For the total risk dimension however, a quick look on Moncho's

¹Both: the Physical Risk Index R_{Ph} and the the Aggravation Coefficient F are estimated through an additive aggregation method, where an analytical hierarchical method is applied to assess weights to each indicator. R_{Ph} and F are then estimated by equation (2) and (3) respectively.

$$R_{Ph} = \sum_{i=1}^p w_{R_{Ph}i} F_{R_{Ph}i} \quad (2)$$

where $F_{R_{Ph}i}$ are physical risk indicators, $w_{R_{Ph}i}$ their respective weights, and p represents the total number of considered indicators.

$$F = \sum_{i=1}^m w_{FSi} F_{FSi} + \sum_{j=1}^n w_{FRj} F_{FRj} \quad (3)$$

where F_{FSi} represent socio-economic fragility characteristics and F_{FRj} the lack of resilience of the exposed context F_{FRj} . While w_{FSi} , w_{FRj} are the assessed weights for each indicator; m and n the total number of indicators for fragility and lack of resilience, respectively. A more detailed information can be found in [6] and [7]

equation can note that the most extreme value of any estimated total risk can only double its original value. Consequently we define the universe of discourse of the FIS output as being within the range [0 2].

TABLE I

RULES DEFINED FOR THE FIS MODEL USED TO ESTIMATE THE TOTAL RISK LEVEL. TR = TOTAL RISK, PhR = PHYSICAL RISK, AG = AGGRAVATION, VH = *very-high*; H = *high*; MH = *medium-high*; ML = *medium-low*; L = *low*; VL = *very-low*

1. If (TR is zero) then (TR is VL)
2. If (Ag is VL) and (PhR is VL) then (TR is VL)
3. If (Ag is L) and (PhR is VL) then (TR is L)
4. If (Ag is M) and (PhR is VL) then (TR is ML)
5. If (Ag is H) and (PhR is VL) then (TR is ML)
6. If (Ag is VL) and (PhR is L) then (TR is L)
7. If (Ag is L) and (PhR is L) then (TR is ML)
8. If (Ag is M) and (PhR is L) then (TR is M)
9. If (Ag is H) and (PhR is L) then (TR is MH)
10. If (Ag is VL) and (PhR is M) then (TR is ML)
11. If (Ag is L) and (PhR is M) then (TR is M)
12. If (Ag is M) and (PhR is M) then (TR is M)
13. If (Ag is H) and (PhR is M) then (TR is MH)
14. If (Ag is VH) and (PhR is M) then (TR is H)
15. If (Ag is VL) and (PhR is H) then (TR is M)
16. If (Ag is L) and (PhR is H) then (TR is VH)
17. If (Ag is M) and (PhR is H) then (TR is VH)
18. If (Ag is H) and (PhR is H) then (TR is H)
19. If (Ag is VH) and (PhR is H) then (TR is VH)
20. If (Ag is VL) and (PhR is VH) then (TR is VH)
21. If (Ag is L) and (PhR is VH) then (TR is VH)
22. If (Ag is M) and (PhR is VH) then (TR is H)
23. If (Ag is H) and (PhR is VH) then (TR is H)
24. If (Ag is VH) and (PhR is VH) then (TR is VH)
25. If (Ag is VL) and (PhR is VL) then (TR is VL)
26. If (Ag is L) and (PhR is VL) then (TR is L)

IV. RESULTS AND COMPARISON

We generated an arbitrary set of 100 numerical values ranging from 0 to 1 taking steps of 0.02 for both inputs and we implemented both models for the complete dataset. Figure 1 shows the resulting surface for Moncho's equation and for the developed FIS respectively.

It is clear how the linearity assumed in Moncho's equation leads towards a smooth surface, but most importantly it is clear how this equation tends to underestimate the influence of the aggravation, especially in the lower part of the surface. This cause that for high values of aggravation, and low values of physical risk, the resulting total risk be low. At the contrary for high values of physical risk and small values of aggravation the opposite is accomplished and high levels of total risk are achieved. En the extreme case, it is noticeably how when aggravation values are always zero, the amount of total risk corresponds to the exact same amount of physical risk for every point.

The fuzzy inference system at contrary, displays a rough surface more suitable to describe the non-linearity response of the simultaneous influence of physical risk and aggravation. In this case, when aggravation values are high and physical risk values are small, the values of total risk are not exclusively small and instead, an important influence of aggravation can

be seen in the final outcome. When aggravation values are small and physical risk values are high, values of total risk are certainly higher than the previous case (maintaining the main assumption of Moncho's equation that total risk is the effect of physical risk aggravated by social vulnerability) but still, the influence of aggravation is not longer underestimated.

As we mentioned before, the structure of the FIS model concede a clear understanding of those exact characteristics of the simulated system that coincided to form a particular total risk level. Figure 2 shows two estimations made by the FIS model and Mocho's equation, respectively. Figure 2 shows part of the set of fuzzy rules that where triggered by an aggravation value of 0.6 and physical risk of 0.25, the same estimation but using the Moncho's equation can be seen at the bottom part. Independently of the difference of total risk values, it is noteworthy to mention how more information can be obtained from the fuzzy approach.

Such values triggered only rules 9 and 10 or a situation where the contribution of the aggravation medium level appear to be the main responsible to generate the final total risk value for low physical risk values. Moncho's equation on the contrary, does not provide this narrative maintaining then, a black box structure. The repercussions for decision making are clear, since it would be possible to know in advance which is the variable and its precise level to improve or decrease in order to achieve a certain level of total risk. The economic or operative efficiency cost of achieving one level or another, could be use in parallel to decide which level is the best option to allocate resources for, once the desired level is known. This could save precious resources to be used in other critical areas.

V. IMPLEMENTATION IN BOGOTA COLOMBIA

We implemented both models in 19 of 20 administrative localities in Bogota, Colombia ² using data reported in [6] and [7]. Such data showed that the non-normalized values of physical risk always tended to be within a range of 0 to 0.3. Therefore we adapt the axis representing physical risk in the fuzzy inference system to have boundaries between zero and 0.5 (and grid steps of 0.005). We then analyzed the response of the model in this range only. The Moncho's model was not modified.

Figure 3 shows the spatial pattern for values of aggravation, physical risk and total risk obtained by means of the fuzzy inference system (up) and the Moncho's equation (down). The distribution pattern of these values reflect the consequence we were describing in previous lines; starting from left to right, it is clear how aggravation and physical risk spatial distribution maps are similar for both models and therefore, a large aggravation area at the southwest part of the city and an important physical risk area at the northern part of the city can be seen. However, when total risk is calculated both models start to differ. For example, the influence of the aggravation area mentioned before cannot be clearly seen if

²In this study we are not taking into account Sumapaz since it corresponds to the rural area of the city

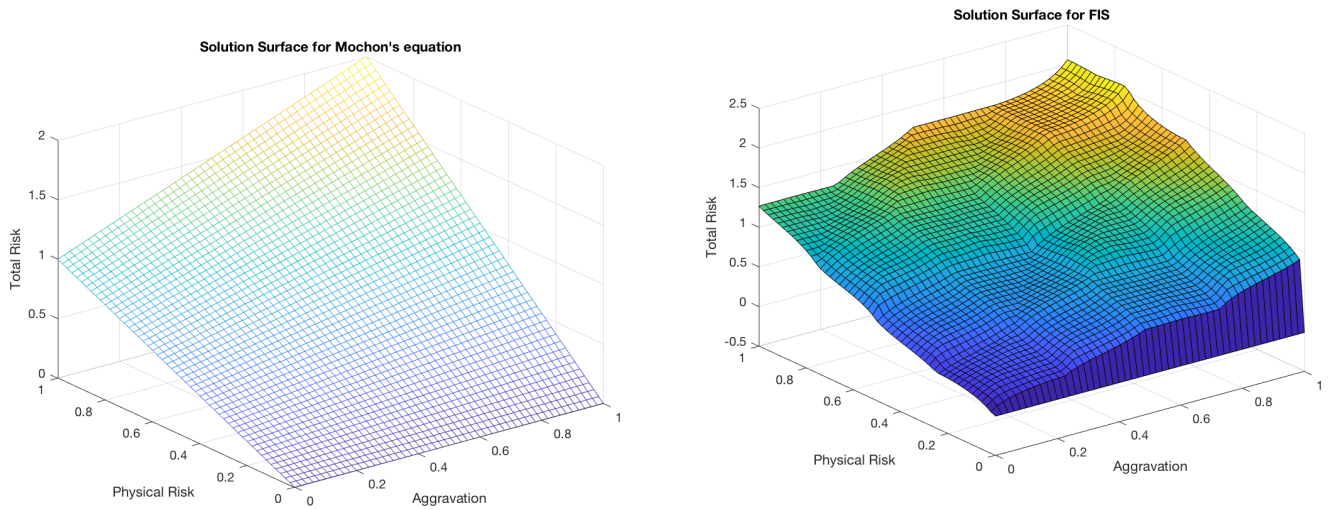


Fig. 1. Solution Surface for Moncho's equation and Fuzzy Decision Surface

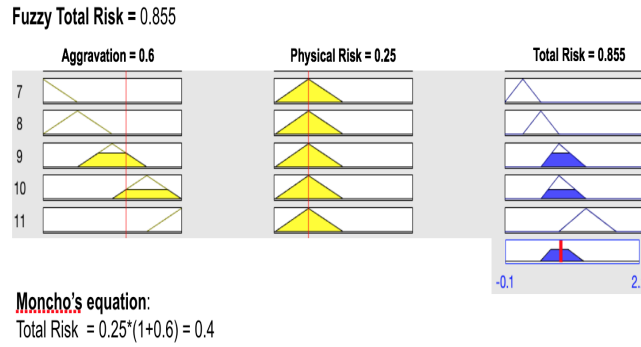


Fig. 2. Triggered Rules

Moncho's equations is used, which translates as the remarkable resembling between the spatial distribution of physical risk and total risk obtained by this method (Figure 3, down).

The fuzzy inference system at the contrary, is able to estimate that this aggravation area does have an influence in the origin of total risk estimated levels, which is reflected in the red colored area of the upper part of Figure 3. As we mentioned, this pattern's difference is largely a consequence of the non-linear aggregation scheme used by the fuzzy system. Under this optic, total risk final level would not depend exclusively on the value (or behavior) of physical risk levels and instead, the influence of a certain aggravation level could be acting more as a driving variable than as an aggravation factor, as it is regarded now in Moncho's equation. A fuzzy approach for the aggregation of indicators of this nature, could allow then a more accurate estimation of the influence that a different indicators or variables may have for the development of further and more complex dynamics.

To observe more clearly how aggravation values where

playing this role, we estimated the difference between the Moncho's solution's surface and the fuzzy decision surface used in this last experiment showed in Figure 4. The aggravation influence over the numerical value of such difference is now clear. We can note that when aggravation values remain low and physical risk values increases up to its maximum, the difference between models is minimal, but if aggravation values increase then the difference, in absolute value, will increase accordingly.

Important consequences in terms of Risk management are clear: if a linear scheme such as Moncho's equation is chosen, then valuable information could be lost, and the influence of a particular component might be underestimated. This translates as a bias for an accurate decision making process for example, to allocate resources since there could be areas reporting both low levels of physical and total risk, but concealing at the same time important aggravation levels. In terms of strategic planning this same bias could cause that those areas presenting high level of social fragility and important levels

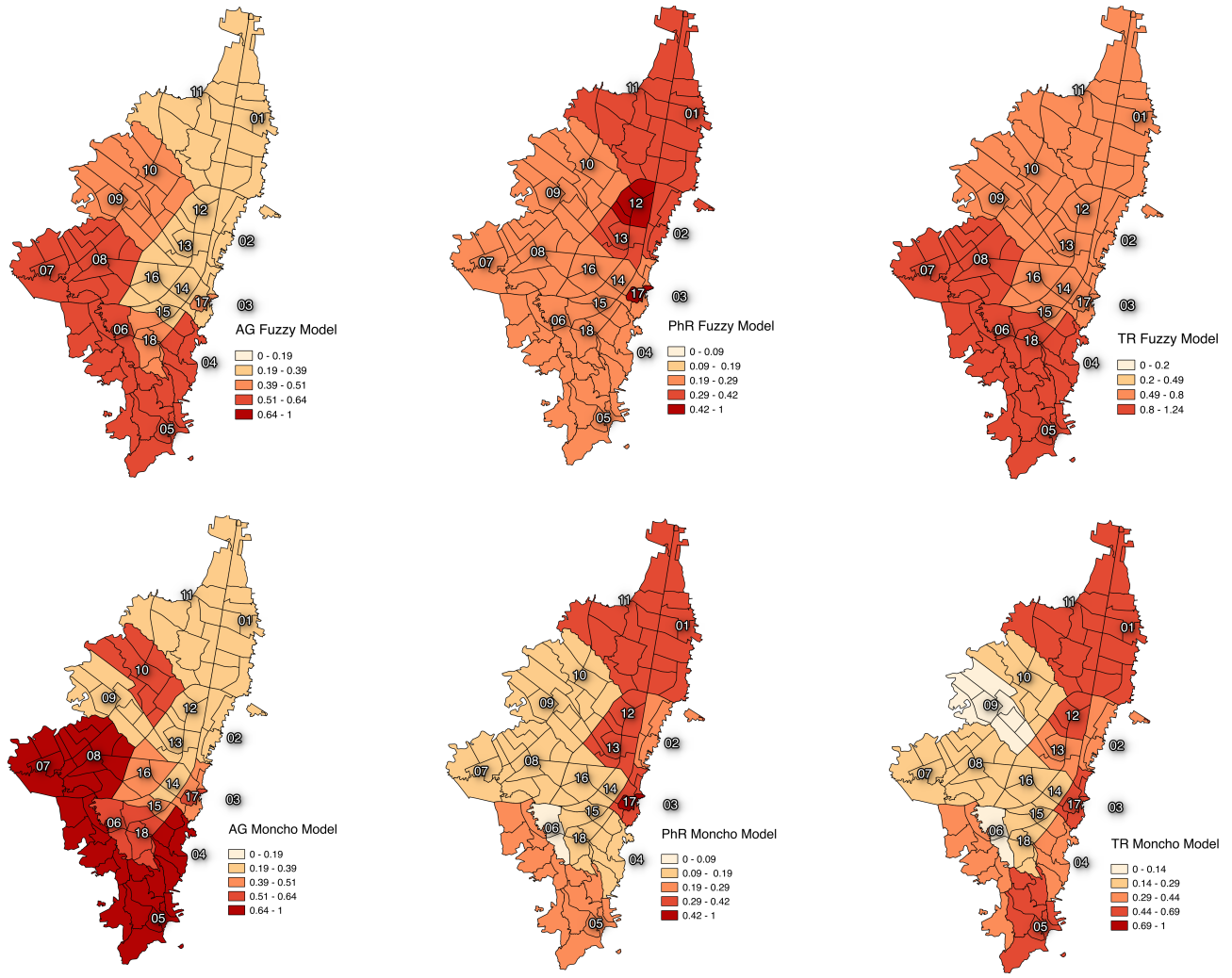


Fig. 3. Seismic risk components obtained for Bogotá city by: (upper part) the Fuzzy Model and (lower part) the Moncho's Model. From left to right: aggravation, physical risk, and total seismic risk. Administrative Localities: (1) Usaquén, (2) Chapinero, (3) Santa Fe, (4) San Cristóbal, (5) Usme, (6) Tunjuelito, (7) Bosa, (8) Ciudad Kennedy, (9) Fontibón, (10) Engativá, (11) Suba, (12) Barrios Unidos, (13) Teusaquillo, (14) Mártires, (15) Antonio Nariño, (16) Puente Aranda, (17) Candelaria, (18) Rafael Uribe, (19) Ciudad Bolívar

of vulnerability not be taken into account, perpetuating in this way a vicious circle that will cause that this aggravations increases with time even more, and more disaster prone areas appear or increase its size. Adopting a fuzzy scheme, let that each of the risk components, physical risk and aggravation, maintain an independent non-linear influence on each other when an aggregation of both is performed allowing that the final output, in this case total risk levels, reflect such non-linearity as well, without assuming that one of the component is the driving variable for the whole process.

VI. CONCLUSIONS

In this paper we showed the difference between a linear methodology and a fuzzy approach to aggregate indicators representing dissimilar realities. We exemplified this difference by considering a composite index that is obtained through the

so called Moncho's equation, a widely used method in the field of seismic risk. This index, labeled as total risk, was compared to the composite index obtained by a fuzzy inference system type Mandami, which maintains the same conceptual elements stated in the Moncho's equation but aggregate their indicators assuming non-linearity. Using a dataset of artificial data we implemented both models and generated their solution surfaces. We showed that the smooth surface representing the possible solutions for Moncho's equation does not accurately represent the complex nature neither of the interrelations between involved concepts nor their final outcome. A fuzzy inference system at the other hand, allows a more realistic representation of the non-linear influences that each risk components may produce to the final risk level. We implemented both methods in Bogotá Colombia to estimate a total seismic risk level. As we discussed, a fuzzy inference system allows a

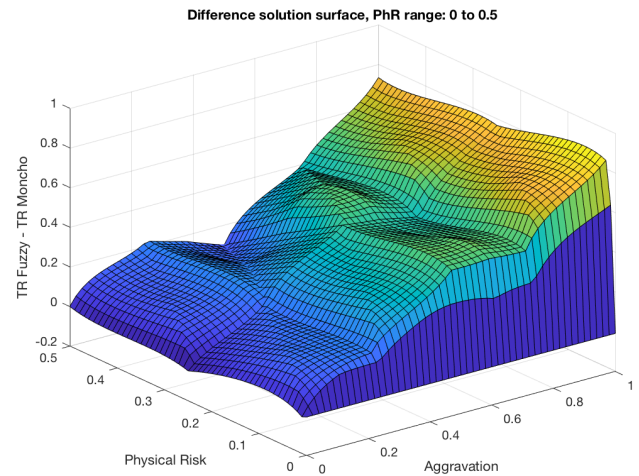
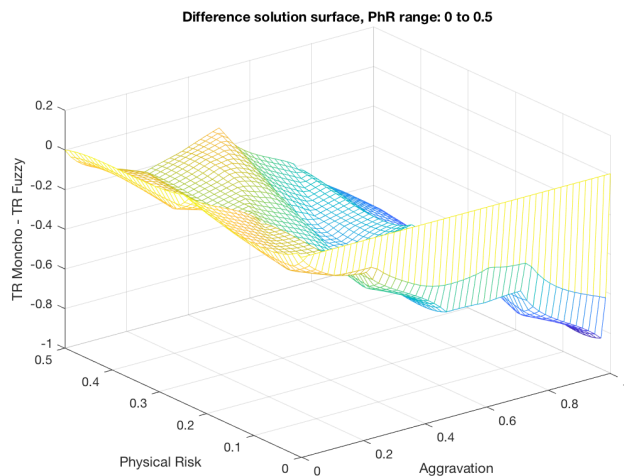


Fig. 4. Difference Solution Surface for Moncho's equation and Fuzzy Decision Surface

more realistic representation of the non-linear influences that each risk components may have to produce a final risk seismic level. Using a simple Mamdani inference system, we produce a surface which does not re-create the submissive dependence of aggravation to physical risk featured in Moncho's equation. In the developed fuzzy system, the aggravation influence is not coerced by the physical risk behavior exclusively and instead, both components are interacting more freely, enabling a more reasonable evolution of these two variables to generate a particular total risk level.

REFERENCES

- [1] J.C. Bertoni, T. Gibbs, M., Hermelin, O. Cardona and A. Lavell. "Understanding and Managing Risk Associated with Natural Hazards: An Integrated Scientific Approach in Latin America and the Caribbean," Science for a Better Life, Vol. 2, pp 88., 2010.
- [2] J. Birkmann, T. Welle "The World Risk Index: An Approach to Assess Risk and Vulnerability on a Global Scale." Journal of Extreme Events, Vol. 2, No. 1 2015.
- [3] C. Burton, B. Khazai, V. Vitor. "Social Vulnerability And Integrated Risk Assessment Within The Global Earthquake Model." DOI:10.13140/2.1.4609.1520, 2014.
- [4] O.D. Cardona, "The need for rethinking the concepts of Vulnerability and Risk from an Holistic Perspective" Mapping Vulnerability: Disasters, Developing and People, Chapter 3, Earthscan Publishers, London, 2003.
- [5] M.L. Carreño, O.D. Cardona, A.H. Barbat. "Disaster risk management performance index", Nat Hazards, 40, pp 1-20 2007
- [6] M.L. Carreño, O.D. Cardona, A.H. Barbat. "New methodology for urban seismic risk assessment from a holistic perspective ", Bull Earthquake Eng, 10, pp 547-565. 2012.
- [7] J.R. González Cárdenas, A. Nebot, F Mugica. "Social Aggravation Estimation to Seismic Hazard using Classical Fuzzy Methods. ", Simulation and Modeling Methodologies, Technologies and Applications. Series: Advances in Intelligent Systems and Computing, Vol. 402, pp 275-293. ISBN: 978-3-319-26469-1. 2016.
- [8] HDRO (Human Development Report Office) "Human Development Report 2018 ", United Nations Development Programme. Retrieved 01 January 2019.
- [9] G.J. Klir, B. Yuan "Fuzzy Sets and Fuzzy Logic: Theory and Applications.", Prentice Hall PTR, ISBN 978-0131011717. 1995
- [10] G. Munda, M. Nardo "Non-Compensatory/Non-Linear Composite Indicators for Ranking Countries: A Defensible Setting.", Applied Economics, 41 pp. 1513-1523. DOI: 10.1080/00036840601019364. 2009.
- [11] M. Nardo, M. Saisana, A. Saltelli, S. Tarantola, A. Hoffman, E. Giovannini "Handbook on Constructing Composite Indicators: Methodology and User Guide. ", OECD Statistics Working Papers, No. 2005/03, OECD Publishing, Paris, <http://dx.doi.org/10.1787/533411815016>.
- [12] P. Peduzzi, P. Dao, 'H. Herold, F. Mouton "Assessing global exposure and vulnerability towards natural hazards: the Disaster Risk Index.", Natural Hazards and Earth System Science, 9(4), pp. 1149-1159.
- [13] L. Lee, A. Kappeli, C. McKee, I. Mitchell, H. Hillebrandt "The Commitment to Development Index: 2018 Centre for Global Development. Retrieved 01 January 2019.
- [14] A. Thow, L. Vernaccini, M. Marin Ferrer, B Doherty. "Global Risk Index", Publications Office of the European Union. ISBN: 978-92-79-80286-7 DOI: 10.2760/754353 (online). 2017
- [15] ERN-AL CAPRA "Comprehensive Approach to Probabilistic Risk Assessment: International Initiative for Risk Management Effectiveness". 2010. available at www.ecapra.org