



---

## 2-FUN

*Full-chain and UNcertainty Approaches for Assessing Health Risks in  
FUture ENvironmental Scenarios*

**FP6 Project-2005-Global-4  
Integrated Project - Contract n°: 036976**

---

### **APPLICATION AND TESTING OF METHODOLOGY FOR RANKING ENVIRONMENTAL STRESSORS (Part 1: Methodological Development)**

Due date of delivery: 31/10/2009

Actual submission date: 28/02/2010

Start date of the project: 01/02/2007

Duration: 48 Months

Lead contractor organisation name for this deliverable: UNIVE

|   |  |
|---|--|
| Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006) |  |
| <b>Dissemination Level</b>  |  |
| <b>PP</b>   | Restricted to other programme participants (including the Commission Services) |

*PROPRIETARY RIGHTS STATEMENT*

*This document contains information, which is proprietary to the 2-FUN Consortium. Neither this document nor the information contained herein shall be used, duplicated or communicated by any means to any third party, in whole or in parts, except with prior written consent of the 2-FUN consortium.*



## Document Information

**Document Name** APPLICATION AND TESTING OF METHODOLOGY FOR RANKING ENVIRONMENTAL STRESSORS (Part 1: Methodological Development)  
**ID** D1.10.doc  
**Revision** Version 4  
**Revision Date** 15/03/2010  
**Authors** Elisa Giubilato, Alex Zabeo, Silvio Giove, Andrea Critto, Antonio Marcomini

## Approvals

|                    | Name              | Company | Date       | Visa              |
|--------------------|-------------------|---------|------------|-------------------|
| <b>Author</b>      | ELISA GIUBILATO   | UNIVE   | 28/02/2010 | Elisa Giubilato   |
| <b>Co-Author</b>   | ALEX ZABEO        | UNIVE   | 28/02/2010 | Alex Zabeo        |
| <b>Co-Author</b>   | SILVIO GIOVE      | UNIVE   | 28/02/2010 | Silvio Giove      |
| <b>Co-Author</b>   | ANDREA CRITTO     | UNIVE   | 28/02/2010 | Andrea Critto     |
| <b>Co-Author</b>   | ANTONIO MARCOMINI | UNIVE   | 28/02/2010 | Antonio Marcomini |
| <b>WP Leader</b>   | ANTONIO MARCOMINI | UNIVE   | 28/02/2010 | Antonio Marcomini |
| <b>Coordinator</b> | FREDERIC BOIS     | INERIS  |            |                   |

## Documents history

| Revision  | Date       | Modification                             | Author                  |
|-----------|------------|--|-------------------------|
| Version 0 | 10/01/2010 | First draft completed                    | E.Giubilato, A. Zabeo   |
| Version 1 | 25/01/2010 | Revised version and comments             | A. Critto, S. Giove     |
| Version 2 | 05/02/2010 | Second draft completed                   | E.Giubilato, A.Zabeo    |
| Version 3 | 10/02/2010 | Revised version and comments             | A. Critto, A. Marcomini |
| Version 4 | 26/02/2010 | Definitive version according to comments | E. Giubilato, A. Zabeo  |



## Table of contents

|   |           |
|---|-----------|
| <b>1. INTRODUCTION .....</b>  | <b>4</b>  |
| <b>2. OBJECTIVES AND MAIN FEATURES OF THE METHODOLOGY .....</b>                         | <b>5</b>  |
| 2.1.Objectives of the methodology.....  | 5         |
| 2.2 Conceptual approach .....   | 7         |
| 2.3 Possible relationships among substances and diseases.....                           | 9         |
| <b>3. BACKGROUND NOTES ON MULTI-CRITERIA DECISION ANALYSIS AND FUZZY APPROACH .....</b> | <b>11</b> |
| 3.1 Multi-Criteria Decision Analysis (MCDA) .....                                       | 12        |
| 3.2 Fuzzy Approach .....  | 13        |
| <b>4. METHODS.....</b>  | <b>14</b> |
| 4.1 Methodological framework .....  | 14        |
| 4.2 PHASE 1: Ranking of chemicals within each Elementary Geographic Unit .....          | 15        |
| 4.2.1 Normalization of criteria values.....   | 17        |
| 4.2.2 Aggregation of normalized values into a unique indicator and weighting .....      | 19        |
| 4.2.3 Final ranking of the chemicals within each EGU .....                              | 22        |
| 4.3. PHASE 2: Ranking of Environmental Chemical Stressors at the Regional Scale .....   | 24        |
| 4.4 PHASE 3: Ranking of the Elementary Geographic Units (EGUs) .....                    | 25        |
| <b>5. CONCLUSIONS .....</b>   | <b>26</b> |
| <b>REFERENCES .....</b>   | <b>27</b> |



## 1. Introduction

Within 2-FUN Work Package 1, entitled “Methodologies for the selection, construction and visualisation of realistic scenarios, accounting for future environmental and societal changes”, the University of Venice (Interdepartmental Centre IDEAS), is charged to achieve the goals set for Task 1.1.

The first phase of activities of this Task is aimed at the development of a methodology for the ranking and selection of scenarios and, in particular, it focused on the development of a methodology for the selection of priority environmental chemical stressors to be further investigated.

The present document illustrates the methodological development of this methodology for the ranking of environmental chemical stressors at the regional scale, allowing also the ranking of priority areas within the region. It is based on a Weight-of-Evidence approach, implemented through a Multi-Criteria Decision Analysis (MCDA) procedure.

In the following chapters the scope and the main features of the methodology will be presented (Chapter 2), followed by a brief introduction on MCDA and Fuzzy Logic (Chapter 3) and the detailed description of the methodological development (Chapter 4).

An additional part of this deliverable (Part 2) will be prepared during the following months and will describe the application and validation of the proposed methodology to an European case-study.



## 2. Objectives and main features of the methodology

### 2.1. Objectives of the methodology

The first phase of activities of Task 1.1 has the specific objective of developing a screening methodology preliminary to a detailed human health risk assessment and aimed at identifying priority environmental chemical stressors and the associated health outcomes to be further investigated.

In complex systems, where multiple stressors interact, many targets are contemporary involved and different health outcomes may be observed, initial assessment efforts have to be addressed to the most relevant scenarios, with the aim of focusing the further risk assessment on the most critical situations (Menzie et al., 2007). This is a “focusing exercise” aimed at guiding assessment efforts and resources towards those scenarios which could correspond to the greatest potential effects on human health.

Typically, at the regional scale different types of environmental and health data may exist for the characterization of the environmental health status. Each chemical stressor (or class of chemical stressors) may be associated, on the basis of available toxicological and/or epidemiological knowledge, to different typologies of diseases identified as possible outcomes of environmental exposure to that chemical. The main objective of the proposed methodology consists in the evaluation of the existing information on environmental contamination and population health status in the region of interest in order to identify which pairs of “chemical – related health outcome” have to be investigated more urgently. Moreover, the methodology will allow the identification of priority areas within the considered region, where further assessment should be focused.

It is important to highlight that the methodology to be developed is not aimed at demonstrate the causal relationship between a selected chemical and an adverse health effect. Rather, starting from the knowledge that environmental exposure to a specific chemical has been identified as a possible cause of a disease, the ranking methodology will try to highlight those cases where there is evidence suggesting a possible impact of environmental contaminants on human health.

This screening phase should provide the adequate information for addressing the efforts of the further risk assessment towards the most critical scenarios, by identifying an appropriate and manageable number of chemical stressors which will be the object of a detailed human health risk assessment in the region of interest (for example, through the application of exposure models, Physiologically-Based Pharmacokinetic models, pathology models such as those developed within 2-FUN project in WP2 and WP3).



According to project purposes, the considered environmental stressors will exclusively be chemical substances (i.e. no physical or biological stressors will be included in the analysis).

The main features of the ranking methodology can be summarized as follows.

- 1) The methodology should be a screening methodology to be applied in a preliminary phase of the assessment, at the regional scale.
- 2) The methodology should make use of the environmental and health data available in the region of interest (environmental and health data from existing monitoring systems, data from previous studies, etc.), and allows a first organization and evaluation of the available information with the purpose of framing and addressing subsequent risk assessment.
- 3) The methodology will be based on the integration of existing information about environmental contamination, population exposure and possible health effects, in order to identify priority environmental chemical stressors. For this purpose, a “Weight-of-Evidence” approach is proposed, which, can be described as “a framework for synthesizing individual Lines of Evidence, using methods that are either qualitative (examining distinguishing attributes) or quantitative (measuring aspects in terms of magnitude) to develop a conclusion regarding questions concerned with the degree of impairment or risk” (Linkov et al., 2009).

The methodological framework that allows the evaluation and integration of data concerning different Lines of Evidence is offered by Multi Criteria Decision Analysis (MCDA) . MCDA tools support the choice among multiple alternatives based on different criteria and provide a systematic approach for handling different types of information overcoming the limitations of unstructured individual or group decision making (Linkov, 2006).

- 4) The ranking methodology will adopt a spatial-oriented approach. The spatial dimension of the assessment will require the identification of spatial entities, called Elementary Geographic Units (EGU), representing the smallest geographical level chosen for data aggregation and allowing the comparison and integration of different types of data (Figure 4). Generally, these EGUs could correspond to administrative areas, such as municipalities or counties, according to the spatial scale of the assessment and the availability of monitoring data.

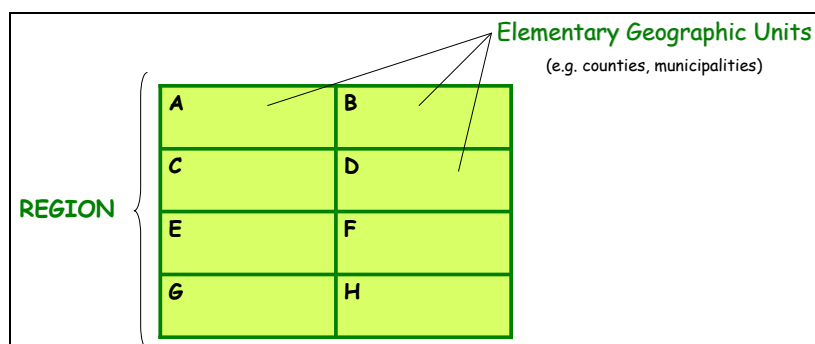


Figure 2. Spatial framework for the assessment.

Each EGU will be characterized by a set of data (i.e., “attributes”) concerning each of the  $n$  environmental chemical stressors and corresponding  $m$  diseases considered in the assessment.

## 2.2 Conceptual approach

In human health risk assessment, the system to be analyzed is constituted by a temporal and spatial sequence of events starting from the release of a chemical in the environment and ending to the development of an adverse effect (health outcome) in human body. For pragmatic purposes, this continuous sequence can be divided into discrete stages (DeWoskin et al., 2007), as illustrated in Figure 3. For each stage, information obtained from experimental, monitoring or modelled data, can be used in the assessment process.

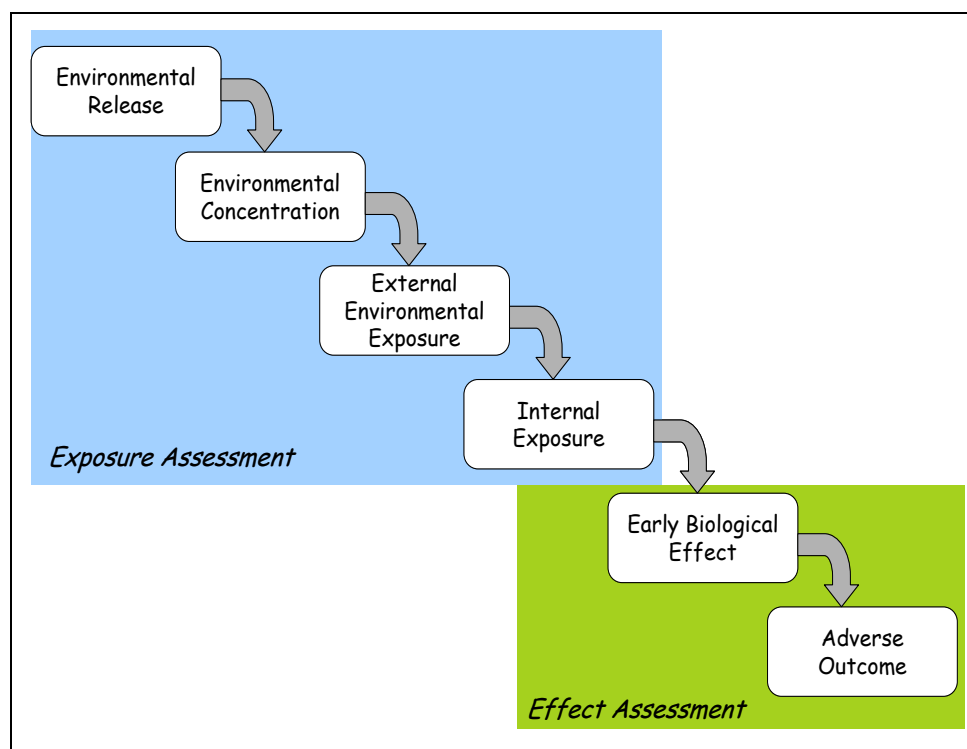


Figure 5. Different steps of the exposure-effects sequence  
(modified from DeWoskin et al., 2007)

The **first step** concerns the release of a chemical substance in the environment as a consequence of industrial activities, transports, accidental releases, etc. The **second step** is related to fate&transport processes which determine the presence of a chemical substance with potential toxic health effect in one or more environmental media (soil, water, air, etc.). The **third step** is related to the evaluation of



the intake, i.e. the quantification of the amount of chemical that daily may enter the body through several exposure routes. This process is also influenced by the vulnerability of the exposed groups (e.g. children are characterized by higher contact rate with contaminated media than adults, which determine a higher intake if compared to adults in the same exposure conditions).

The **fourth step** concerns the process of absorption of the chemical substance into the body, that determines an internal dose measurable by means of adequate biomarkers of exposure.

Once a chemical stressor has entered the human body, it may cause an early effect on the organism (**fifth step**). The effects may be measured as changes (often reversible) in the biological systems by using biomarkers of effects, such as measurement of cholinesterase activity in blood, changes in cytochrome P450 enzyme activity, etc. The chain ends with the manifestation of the full disease state (**sixth step**): at this stage the population health status is measurable by collecting information on the incidence/prevalence of the diseases of interest in a defined area (epidemiological data).

For the development of a disease as a consequence of environmental exposure, each step of the sequence has to be “accomplished”.

The basic idea behind the ranking approach of environmental chemical stressors consists in the evaluation for each chemical and its associated health effects of the existing evidence for each step of the exposure-effect sequence. As stated in paragraph 2.1, the ranking methodology will make use only of existing monitoring data about the region of concern. In particular, for the development of the ranking methodology, 3 steps of the aforementioned sequence are chosen, which are more likely to be covered by monitoring data (Smolders and Schoeters, 2007). These steps can be referred to as 3 Lines of Evidence (LoE), because data related to each of these steps provide information about a specific aspect of the investigated problem. Specifically, the selected Lines of Evidence are:

- 1) LoE “Environmental Contamination”: this LoE provides information about the distribution and quantity of chemical stressors in the environment, taking different matrices into account (e.g. soil, air, surface water, groundwater,...) according to the scope of the analysis and the available data. This information is derived, for example, from existing environmental monitoring data and can be available as single point data or as distribution maps. For evaluating the “entity” of the contamination status, the comparison with reference/threshold values (e.g. Environmental Quality Standards) may be used. This LoE is included in the ranking methodology because it provides information about the potential contamination to which population is exposed (the higher the environmental contamination, the higher the probability that the population it is exposed is).
- 2) LoE “Intake”: this LoE includes data derived from biomonitoring of specific parameters. Biomarkers of exposure may include *measurements of parent compounds, metabolites, or DNA or protein adducts of parent compound and/or metabolites that indicate a direct exposure to the compound of interest* (Ryan et al., 2007). This LoE has been chosen to introduce in the assessment an evaluation of the actual exposure of population to the selected chemicals.



Biomarkers of exposure indeed represent an *integration of exposure from all sources and routes, which provide an important perspective on overall exposure* (Albertini et al., 2006). These data may thus indicate if the considered population living in the contaminated region is actually exposed to specific chemical substances. Some example of quite commonly used exposure biomarkers are concentration of Pb in blood, concentration of Cd in urine, etc. Therefore, this LoE will include data provided by biomonitoring campaigns such as the concentration of chemicals or metabolites in blood and other biological matrices.

- 3) LoE “Observed Effects”: this LoE is aimed at providing information about health effects measured in the population which are suggested to have, according to present toxicological and epidemiological knowledge, an environmental cause. This LoE may include basically epidemiological data. Epidemiological data consist in information about the distribution and/or the frequency of disease/deaths by means of morbidity and mortality data. Health effects may be evaluated as individual-level data or may be aggregated for populations (often identified by administrative units) like those available from local, regional and national surveys. Aggregated data consists in measures of disease frequency and involve the occurrence of new cases or deaths (measure of incidence or mortality) or the presence of existing cases (measures of prevalence) (Morgenstern and Thomas, 1993).

Each LoE could be considered separately: for example, we can compare existing contamination data for several chemicals within a region and define a ranking where chemicals with the highest concentration are on the top. The same concept may apply to other LoEs. Instead of considering these aspects separately we propose to overlap and integrate the different pieces of information provided by the different LoEs for each chemical, through a Multi Criteria Decision Analysis (MCDA) approach (Linkov, 2006), with the aim of achieving a more robust and complete evaluation of available evidence. Chemicals resulting on the top of the ranking list are candidates to become the object of a detailed risk assessment, through the implementation of further monitoring and the use of specific predictive exposure and toxicological models.

The information obtained through the integration of the 3 LoEs will provide the basis for the accomplishment of 3 methodological steps (which will be explained in detail in Paragraph 4.1) leading not only to a ranking of chemicals at local scale and at the regional scale but also to a ranking of Elementary Geographic Units within the considered region.

### **2.3 Possible relationships among substances and diseases**

The identification of the possible causal relationship between a chemical substance and an health outcome should constitute the first step of the assessment. This identification should be based on available toxicological and epidemiological knowledge with the support of expert judgement.

$\underline{\mathbf{S}} = \{S_1, S_2, \dots, S_n\}$  is the set of the  $n$  chemical substances to be taken into account

and

$\underline{D} = \{D_1, D_2, \dots, D_m\}$  is the set of the  $m$  diseases to be considered.

Starting from the known relationships, a scheme illustrating the relationships between chemicals and health outcomes (as illustrated in the example in Figure 4) can be developed.

The arrow exiting from the chemical and entering the health outcome indicates that this chemical has been identified among the possible causes of that pathology.

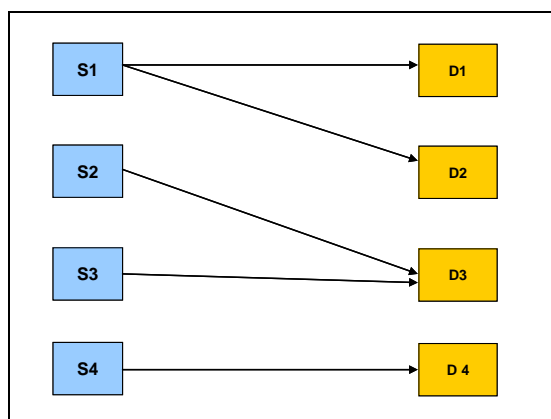


Figure 4. An example of possible relationships among chemical substances ( $S_i$ ) and diseases ( $D_j$ )

The same information can be translated into a matrix composed by  $n$  rows (for the corresponding  $n$  chemicals) and  $m$  columns (for the  $m$  health outcomes).

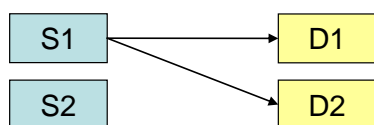
|       | $D_1$ | $D_2$ | $D_3$ | ... | $D_m$ |
|-------|-------|-------|-------|-----|-------|
| $S_1$ | x     | x     |       |     |       |
| $S_2$ |       |       | x     |     |       |
| ...   |       |       |       |     |       |
| $S_n$ |       |       | x     |     | x     |

Figure 5. An example of a ( $n \times m$ ) matrix representing the relationships between chemical substances ( $S_i$ ) and diseases ( $D_j$ ).

It could be useful to highlight that the possible kind of relationships could be traced back to 3 cases (an example of the resulting matrix is reported in each case):

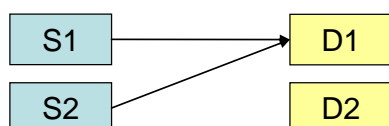


A) the same stressor (Si) is identified as possible cause of two or more pathologies:



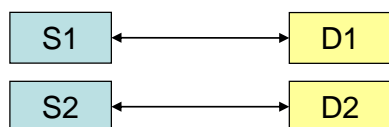
|    | D1 | D2 | D3 | D4 |
|----|----|----|----|----|
| S1 | x  | x  |    |    |
| S2 |    |    |    |    |
| S3 |    |    |    |    |
| S4 |    |    |    |    |

B) two or more stressors are identified as potential causes of the same pathology M<sub>i</sub>:



|    | D1 | D2 | D3 | D4 |
|----|----|----|----|----|
| S1 | x  |    |    |    |
| S2 |    |    |    |    |
| S3 |    |    |    |    |
| S4 | x  |    |    |    |

C) among the considered chemicals, each chemical is identified as possible cause of only one pathology and this pathology at the same time is connected only to that chemical (the relationship is biunique):



|    | D1 | D2 | D3 | D4 |
|----|----|----|----|----|
| S1 | x  |    |    |    |
| S2 |    | x  |    |    |
| S3 |    |    |    |    |
| S4 |    |    |    |    |

In chapter 4 it will be explained how we propose to deal with the described situations along the different methodological phases.

### 3. Background Notes on Multi-Criteria Decision Analysis and Fuzzy Approach

In this chapter, a brief overview will be provided about general principles and characteristics of Multi Criteria Decision Analysis methods and Fuzzy approaches, which are considered useful for a complete understanding of the proposed methodologies.



### 3.1 Multi-Criteria Decision Analysis (MCDA)

MCDA includes a large number of methods for the evaluation and ranking or selection of alternatives that considers all the aspects of a decision problem involving many actors (Giove et al., 2009).

A structural platform common to almost all the decision problems includes the following items:

1. the decision maker (DM). A conceptual figure, a single person, a group of persons or an entity in charge of finding the best solution for the problem under assessment;
2. a set  $A$  of alternatives, in the finite case:  $A = \{a_1, \dots, a_m\}$ , beside whom the DM must choose the best solution;
3. a countable family of criteria or attributes or parameters,  $K = \{k_1, \dots, k_n\}$ . These are aspects of the problem that the DM considers crucial and they also define the alternatives. Criteria can be organized into a hierarchical structure, i.e. a decision tree where the root is the objective function whose leaves are the first-level criteria, each of them split again into second-level criteria (sub-criteria), and so on till the last level, whose terminal leaves are the indicators (or the last level sub-criteria) formed by the available information (data or judgments);
4. an objective or target function (to be optimized) used to score, and in case rank, alternatives, usually an aggregation function;
5. the decision maker's preferences for the different evaluation of the criteria;
6. an algorithmic tool designed to optimize the objective function, considering all the above information

MCDA methods can be categorized into 3 groups (Vinke, 1992): MAUT/MAVT (Multi-Attribute Utility/Value Theory), Outranking and Interactive methods.

The methodology proposed in this report is based on a MAUT/MAVT approach. In Multi-Attribute Utility/Value Theory (MAUT/MAVT), criterion values are first normalized into a common numerical scale by means of a suitable transformation function (or Utility/Value Function). Then criteria are aggregated by a suitable aggregation operator, a function that satisfies a set of rationality axioms.

Using a bottom up approach, this operation is repeated for all the nodes in the decision tree (if the problem is hierarchically structured) for all the alternatives. Each branch or level of the tree may be aggregated to its root by using different aggregation functions based on the criteria relations. At the tree root (the objective) a single numerical value is finally computed, which is the score of the proposed alternatives. The alternatives can then be rated and ranked, since MAUT/MAVT produces a total ordering, and so the best one can be selected.

Many different aggregation operators can be used in the evaluation of a MAVT decision tree. Aggregation operators can be classified, considering their behaviour, in three main classes: minimum operators, mean operators and maximum operators (Calvo, 2002). Aggregation operators can also be



associated to logical operators, especially when dealing with fuzzy sets and fuzzy logic problems (see paragraph 4.3 for a fuzzy logic overview).

The ranking methodology presented in this document makes use of fuzzy measures, therefore aggregation operators which are counterparts of logical operators have been used. More precisely, the methodology implies the use of a conjunction operator. Conjunction operations (denoted by the “AND” word and by the  $\wedge$  symbol) are used when the aim is to highlight situations where all the elements under assessment must be valid at the same time (e.g. my car must be fast AND cheap). In traditional logic, propositions based on conjunction operators are satisfied when all their operands are satisfied, i.e. the proposition is true if and only if all of its operands are true.

Fuzzy logic differs from classical logic because it introduces the concept of different degrees of truth (to be expressed in the range  $[0,1]$ ). In a fuzzy environment the aforementioned definition of conjunction (demanding that all operands are completely true in order to obtain a true result), cannot be applied because the concept of “truth” may assume a value varying between 0 and 1. Conjunction operators, in a fuzzy logic environment, are therefore represented by the set of “minimum operators” which correspond to the family of Triangular-norms (abbreviated in T-Norm) functions (Klement, 2000).

A T-norms is a function  $T:[0,1] \times [0,1] \rightarrow [0,1]$ , having the following properties: commutativity, monotonicity (increasing), associativity, 1 as the neutral element. The T-norm providing the maximum results is the minimum function, while the one providing the lowest result is the product (Klement, 2000). The most notable and commonly used T-norms are the minimum, the probabilistic T-norm, the Lukasiewicz T-norm and the drastic T-norm. Any of the mentioned T-norms can be used to represent conjunction in fuzzy logic, it is important to highlight that each T-norm has its specific properties to be evaluated when choosing the most suitable for the assessed problem.

### [3.2 Fuzzy Approach](#)

Fuzzy logic was first introduced by Lofti Zadeh in 1965 with the aim of providing a means of processing data by extending classical set theory to handle *partial membership* (McKone and Deshpande, 2005).

Fuzzy approaches have the quality to deal with the concept of “partial truth” to quantify uncertainties associated to linguistic variables and allow indeed to translate qualitative and vague information (e.g. qualitative judgement) into numerical reasoning with the aim of incorporating it into a mathematical framework (Chowdury et al., 2009; Li et al., 2007).

Fuzzy logic allows the definition of membership functions (Zadeh, 1965) which correspond to characteristic functions in classical set theory. Membership functions represents the degree to which a

linguistic variable satisfies each term of a Fuzzy set (that is, membership degree). A membership function is a function

$$f: D \rightarrow [0,1]$$

where D is the domain of the fuzzy variable of concern.

The membership function gives the degree of truth of the variable of concern in each of its domain's values.

In Figure 6 three membership functions for three linguistic variables (LOW, MEDIUM, HIGH) for the same term are presented in the same graph. As the term value increases, its membership moves from one variable to the other in a continuous way.

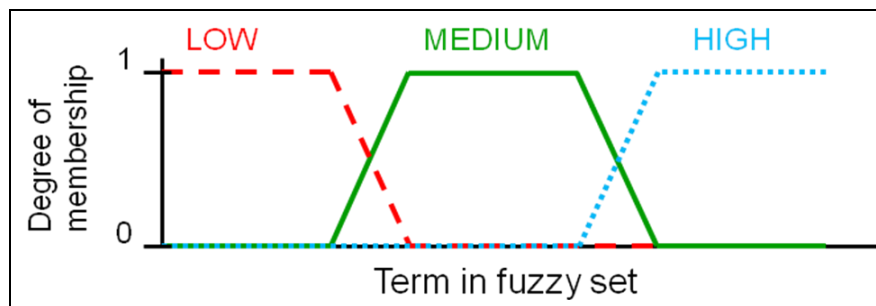


Figure 6. An example of a membership function.

## 4. Methods

In this chapter the methodology proposed for the accomplishment of the goals identified in Chapter 2 will be described in detail. First, in Paragraph 4.1 the methodological framework will be presented, then in the following paragraphs the specific procedures proposed for each step of the methodological framework will be described in detail.

### 4.1 Methodological framework

The proposed ranking methodology will constitute the basis for the development of a “Risk-based Tool for the Regional Ranking of Environmental Chemical Stressors”, a decision-support tool aimed at



helping end-users/decision-makers (e.g. local authorities, environmental and public health agencies, etc.) in the selection of priority environmental chemicals and priority areas to be further investigated.

Three different methodological phases have been identified and will be implemented in the “Risk-based Tool for the Regional Ranking of Environmental Chemical Stressors”: each phase is related to a specific objective and leads to the achievement of a specific ranking as explained in the following paragraphs.

The first phase is aimed at obtaining a ranking of the environmental chemical stressors in each Elementary Geographic Unit (PHASE 1), for example in each county of the region covered by monitoring data.

Afterwards, the results obtained for different EGUs will be integrated and properly treated for achieving a ranking of environmental chemical stressors at the regional level (PHASE 2), to select, within the region, those chemicals to be included in a detailed human health risk assessment (e.g. through implementation of further environmental or health monitoring, use of detailed exposure or toxicological predictive models, etc.).

The last phase of the Ranking Tool is aimed at providing a ranking of the EGUs within the region (PHASE 3), in order to identify and select those EGUs where existing environmental and health monitoring data suggest a situation of possible or actual impacts on population health.

The proposed methodology will be implemented in a dedicated software tool (we will refer to it as the “Risk-based Tool for the Regional Ranking of Environmental Chemical Stressors”) that will allow data storage, application of the different methodological phases of the ranking methodology and delivery of results. The tool will include a geo-database linked to a Geographic Information System (G.I.S.) framework for the spatial visualization of results.

The system will be “fed” with environmental and health monitoring data concerning the region of interest and adequately treated in order to satisfy methodological and software requirements.

#### **4.2 PHASE 1: Ranking of chemicals within each Elementary Geographic Unit**

In this paragraph, the procedure aimed at the ranking of environmental chemical stressors within each Elementary Geographic Unit is described.

First, for each Elementary Geographic Unit (EGU), a set of pairs [chemical, disease] has to be identified, for which environmental and health monitoring data are available.

For each EGU and each pair [chemical, disease] a set A including three type of data (corresponding to the 3 selected Lines-of-Evidence) could be available, that is:

$$A_{i,j} = [C_i, I_i, D_j]$$

where:



$C_i$  = concentration of chemical  $i^{\text{th}}$  in the environmental matrix of interest (e.g. soil, water, air);

$I_i$  = concentration of the chemical  $i^{\text{th}}$  (or its metabolite) in a human biological matrix (e.g. blood, urine);

$D_j$  = indicator of incidence/prevalence of  $j^{\text{th}}$  disease in the EGU;

with

$i = 1, 2, \dots, N$  (number of substances)

$j = 1, 2, \dots, M$  (number of diseases)

Therefore, each EGU will be characterized by a number of sets  $A_{i,j} = [C_i, I_i, D_j]$  which equals the number of pairs [chemical, disease] identified in that unit.

The information provided by the three Lines-of-Evidence (LoEs) has to be merged and evaluated in order to support the decision-maker in the selection of the priority chemicals in the EGU of interest. Different combinations of data from the 3 LoEs may correspond to different situations: the goal is to evaluate these combinations according to the decision-maker rules in order to select those chemicals which results to be related to the most critical situations. This goal requires the aggregation and evaluation of heterogeneous data, a suitable approach to deal with this type of issue is provided by MCDA techniques (Figueira et al., 2005) (see Paragraph 3.1). In fact MCDA supports decision-makers in evaluating and selecting among a number of alternatives based on multiple criteria and adopting systematic analysis that overcomes the limitations of unstructured individual or group decision-making (Linkov et al., 2006).

The methodology for the ranking of environmental chemical stressors within each EGU can be summarized in the following 3 steps:

- 1) Normalization of criteria values;
- 2) Aggregation of normalized values into a unique indicator and weighting to estimate a ranking score for each pair chemical-disease;
- 3) Estimation of a score for each single chemical.

These steps will be described in the following paragraphs and are illustrated in Figure 7.

The results of this first phase (estimate of a score for each single chemical) represent the starting point also for the other two phases (i.e. ranking of chemicals at the regional scale and ranking of the EGUs).

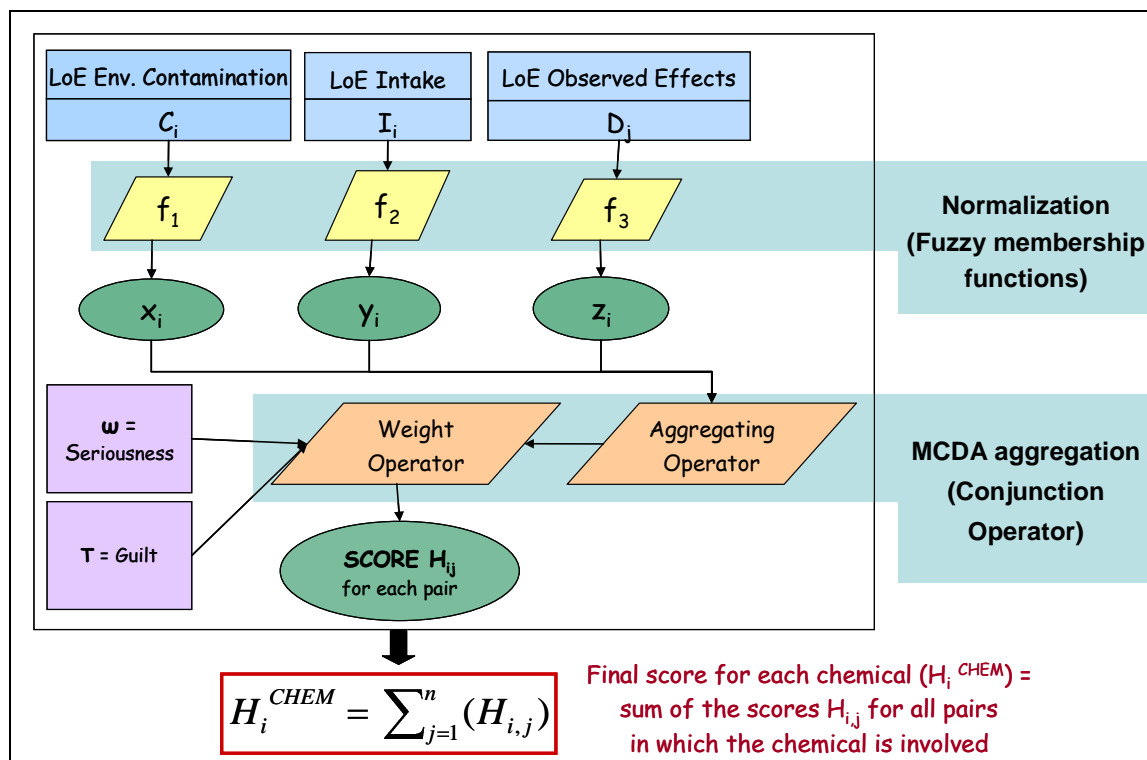


Figure 7. Main steps of the MCDA-based methodology for the ranking of chemicals within each EGU

#### 4.2.1 Normalization of criteria values

The data concerning each LoE have first to be properly transformed (i.e. normalized) into a common numerical scale. To this purpose, an approach based on Fuzzy logic is proposed. Fuzzy logic allows to define membership functions, representing the degree to which a linguistic variable satisfies each term of a Fuzzy set (that is, membership degree) (see paragraph 4.2).

A properly defined function allows then to define the “degree of truth” of a sentence such as “*the lead contamination in soil is high*”. As an example, the classification and evaluation of the contamination value for the chemical  $i^{\text{th}}$  can be achieved by a function like the one represented in Figure 8. On the x-axis are the values that can be assumed by the variable “concentration of the  $i^{\text{th}}$  substance in the environmental matrix” (e.g., mg/kg of dry soil), while on the y-axis (on the scale [0,1], but alternatively also the scale [0%, 100%] can be used) are the values representing the “degree of truth” of the sentence “contamination in the matrix due to the  $i^{\text{th}}$  substance is high”.

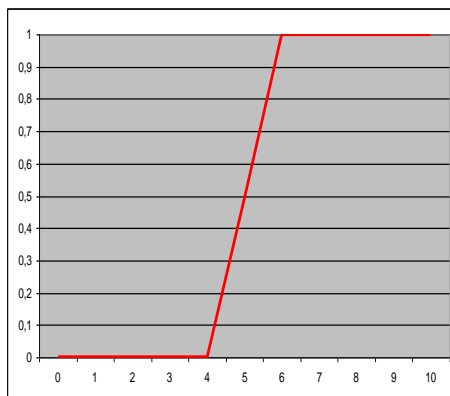


Figure 8. Example of membership function for data of the LoE “Environmental Contamination”.

In order to build this membership function, it is necessary to define (and therefore to ask the expert/decision-maker) two thresholds:

- the threshold below which the degree of truth of the sentence is considered equal to 0 (i.e. *the sentence is false*)
- the threshold above which the degree of truth of the sentence is maximum, i.e. equal to 1, or 100% (i.e. *the sentence is true*).

In the range of values between these two thresholds, the function may have different shapes, but, for the sake of simplicity, we set it as a linear function.

Adopting the same approach, two membership functions, for the LoE “Intake” and the LoE “Observed Effects”, can be defined. In the case of LoE “Intake”, on the x-axis is represented the variable “Concentration of substance  $i^{\text{th}}$  in biological matrix” (e.g. ug/dl of blood), in the case of LoE “Observed Effects” on x-axis a variable representing the incidence or prevalence of a disease (e.g. “number of new cases per year each 10.000 people”) is represented.

By applying the above concepts, a set of three normalized data, expressed on the same scale (i.e., [0,1]), is obtained for each pair [chemical, disease], as follows:

$$A'_{i,j} = [x_i, y_i, z_j]$$

where:

$x_i$  = “degree of truth” of the sentence “the concentration of the chemical substance  $i^{\text{th}}$  is high in the environmental matrix”;

$y_i$  = “degree of truth” of the sentence “intake of the substance  $i^{\text{th}}$  is high”;

$z_j$  = “degree of truth” of the sentence “the incidence/prevalence of the disease  $j^{\text{th}}$  is high”.



#### 4.2.2 Aggregation of normalized values into a unique indicator and weighting

The normalized values must then be aggregated into a unique numerical value. The selection of the aggregation operator/function depends on the relationships among the considered criteria and the meaning/objective of the final indicator to be obtained.

Many aggregation operators exist (see 4.2), for our purposes the conjunction operator ( $\wedge$ ) is selected because our aim is to highlight those situations where all operands (i.e., all three LoEs) are contemporarily verified (not null).

The ranking score for each pair “chemical  $i^{\text{th}}$  and disease  $j^{\text{th}}$ ” can be estimated as the degree of truth of the following statement:

$$H_{i,j} = \left[ (x_i, \delta_x) \wedge g(y_i, \delta_y) \wedge g(z_j, \delta_z) \right] \times T_{i,j}^\alpha \times \omega_j \quad (\text{eq. 1})$$

The ranking score  $H_{i,j}$  related to the  $i^{\text{th}}$  chemical and the  $j^{\text{th}}$  observed effect is thus estimated from the elements illustrated in details in the following paragraph.

First, the ranking score depends on the **normalized values** obtained from each LoE, that are:  $x_i$ ,  $y_i$  and  $z_j$ .

As the decision-maker can decide to assign a different importance to each one of the single criterion (i.e. to each single LoE), there is the possibility to assign a **specific relevance weight**  $\delta_p$  to each of them:

$\delta_x$  = relevance weight assigned to LoE “Source”;

$\delta_y$  = relevance weight assigned to LoE “Intake”;

$\delta_z$  = relevance weight assigned to LoE “Observed Effects”.

For example, the decision-maker could decide to attribute a higher weight to data of LoEs associated to exposure (“Environmental Contamination” and “Intake”) in comparison to the LoE “Observed Effects”.

The relevance parameter  $\delta_p$  can vary on the scale [0,1] and the function  $g_p$  is a simple aggregation function between the normalized value for each LoE and the relevance of that LoE, which determines how much the normalized value has to be considered in the aggregation.

As explained earlier (see paragraph 3.2), the conjunction operator ( $\wedge$ ) in fuzzy environments is expressed as a T-Norm. Among all possible T-Norms we decided to use the minimum function, which is the most widely used function in this kind of elaborations. As shown by the first part of equation 1



(inside the square brackets), a weighted minimum is applied between the LoEs. According to Yager's definition (Yager, 1981) the weighted minimum can be written as follows:

$$\min_{w_1, \dots, w_n}^{\oplus} (x_1, \dots, x_n) = \min_{i=1}^n \left[ \max \left( -w_i, x_i \right) \right]$$

that in our context became:

$$\min \left[ \max(1 - \delta_x, x_i), \max(1 - \delta_y, y_i), \max(1 - \delta_z, z_j) \right]$$

Therefore, it results that the g function can be written as follows:

$$g(x, \delta) = \max(1 - \delta, x)$$

To better understand what the weighted minimum means in practice, the first thing to keep in mind is that 1 is the neutral element of the minimum function. This means that an operand which value is 1 is non influential in the evaluation of the minimum (remember that we are always dealing with numbers in  $[0,1]$ ). Therefore, it also means that if an operand has to be non influential, its value must approach 1. As argument of the g function, the opposite of the weight parameter  $\delta$  is used, so that the higher  $\delta$  the lower its opposite is. Consequently low values of  $\delta$  (i.e. scarce importance for the associated criterion) will likely generate high g function values, which will in turn be non influential in respect to the other g functions results within the weighted minimum equation.

Another element to be included in the aggregating function is the **parameter**  $T_{i,j}^{\alpha}$ .

This parameter allows to define how to face the cases in which two or more chemicals are associated to the same disease  $j^{\text{th}}$ , as illustrated in Figure 9.

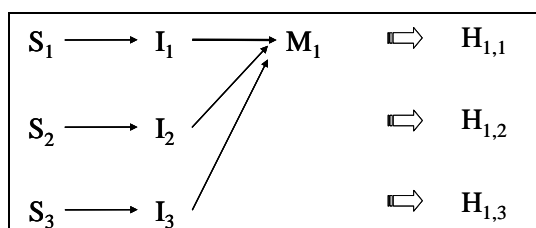


Figure 9. An example of a case where more chemicals ( $S_1, S_2, S_3$ ) are related to the same disease ( $M_1$ ) and generate different scores ( $H_{1,1}, H_{1,2}, H_{1,3}$ ).

In this situation, the decision-maker can decide to evaluate “independently” each set ( $x_i, y_i, z_j$ ), that is, to consider the value of the LoE “Observed Effects” ( $z_i$ ) unchanged for each set. Alternatively, the decision-maker can also decide to “subdivide” the value of  $z_i$  among the involved chemicals, that is, to



attribute a “portion” of  $z_i$  to each set. This choice is determined by his degree of knowledge about the system under analysis and the specific relationship between chemicals and diseases.

To allow the decision-maker to set freely how to deal with this situation, a parameter  $T_{i,j}^\alpha$  is defined and introduced as an additional element in the aggregation phase.

$T_{i,j}^\alpha$  is estimated from a parametric formula defined as follows:

$$T_{i,j}^\alpha = \left[ \frac{E_i \eta_{i,j}}{\sum_{k \in B} E_k \eta_{k,j}} \right]^\alpha$$

where

- B is the set including all chemicals associated to the disease  $j^{\text{th}}$ ;
- $\eta_{i,j}$  is a parameter representing the “probability” that chemical  $i^{\text{th}}$  causes disease  $j^{\text{th}}$  (it is a relative weight, defined on the scale [0, 1] that has to be set a priori by the expert for each pair);
- $E_i$  is an indicator of the “overall evidence of exposure” (because it takes into account information related to both “Environmental Contamination” and “Intake”) and is calculated as follows:

$$E_i = g(x_i, \delta_x) \wedge g(y_i, \delta_y)$$

$\alpha$  is a parameter which allows the decision-maker to define how to deal with the situation where more chemicals are associated to the same disease. Specifically,  $\alpha$  can be:

- 1)  $\alpha = 0$ , (it follows that is  $T_{i,j}^\alpha = 1$ ): in this case the decision-maker is evaluating each set independently and is attributing to each chemical all the evidence of effect ( $z_i$ );
- 2)  $\alpha = 1$ : the decision-maker attributes to each chemical a “partial responsibility” in causing the disease, according to the weights defined a priori;
- 3)  $\alpha \rightarrow \infty$  (it follows that  $T_{i,j} \rightarrow 0$ ): in this case, the higher the value of  $\alpha$ , the higher become the weight given to cases in which only one chemical is associated to a disease (i.e. cases where more chemicals are present tend to be disregarded due to an increasing uncertainty associated to the attribution of effects to one chemicals instead of another one)



As to the decision maker point of view, only few parameters need to be set, which are the importance of each LoE  $\delta$ , the value of  $\alpha$  and the parameter  $\eta$ . All these values can be asked to the user by means of a scale bar. In the case of the parameters  $\delta$  and  $\eta$ , the scale bar will be continuous (in the range [0,1]), identifying the first the degree of importance and the latter the relative weight of the relationship chemical-disease. In the case of the  $\alpha$  parameter the scale bar will be discrete and each point in the scale will be identified by a label which explains the corresponding logic.

The last element to be included for the estimate of the ranking score is the **parameter  $\omega_j$** .

$\omega_j$  is a parameter representing the seriousness of the  $j^{\text{th}}$  disease in terms of potential disability and thus in terms of potential impacts on society due to occurrence of a case of the disease. The decision-maker can therefore decide if further distinguish among different pair [chemical, disease] based not only on the aggregation of information derived from each LoE, but also on the different degrees of seriousness of different diseases.

The “seriousness” of the disease could be evaluated by using expert judgement, or a reference scale such as the system of Disability Weights proposed in the framework of assessment of the global/national environmental burden of diseases (WHO, 2004; Stouthard et al. 2000). This consists in a *set of disease-specific empirical weights to evaluate the level of disability, following standardized methods* (Stouthard et al., 2000). These weights are used to translate measures of incidence of environmental diseases (measured or estimated) into cumulative indicators (e.g. DALYs) in order to quantify the overall burden of disease for a selected population. According to this reference system, and in case integrating also expert judgement, a “seriousness value”  $\omega_j$  could be attributed to each considered disease.

This value  $\omega_j$  is included in the interval [0,1].

As a summary, information to be provided by the decision-maker (i.e. the whole set of values to be attributed to the different parameters included in the methodology) is illustrated in Box 1 at the end of Paragraph 4.2.

#### 4.2.3 Final ranking of the chemicals within each EGU

With the objective of ranking chemicals within the EGU, in the case of a chemical associated to more than one disease, the proposed aggregation procedure is the following:

$$H_i^{CHEM} = \sum_{j=1}^n H_{i,j}$$

that is,  $H_{i,j}$  scores related to the same chemical are summed.

Finally,  $H_i^{\text{CHEM}}$  obtained for each chemical will be compared to obtain the chemicals ranking.

In Figure 10, is reported an example illustrating different possible situations (relationship single chemical-single disease, more chemicals associated to the same disease and more diseases associated to the same chemical). The way of aggregating  $H_{i,j}$  scores is illustrated.

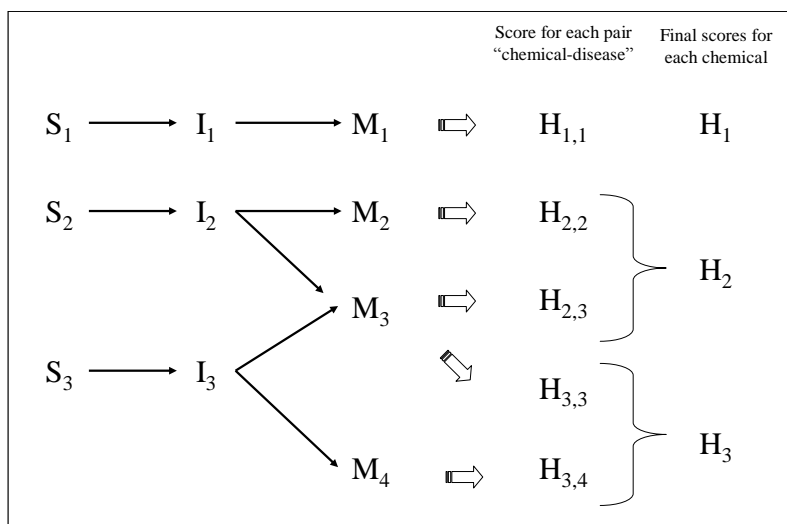


Figure 10. An example of aggregation of scores  $H_{i,j}$  into a single score  $H_i^{\text{CHEM}}$  taking into account different types of relationships between chemicals and diseases.

Because the final objective is to achieve a ranking, the different values to be ranked need to be comparable. This implies that it is necessary to have for each pair [chemical, disease] the whole set of 3 data corresponding to the 3 LoEs (Environmental Contamination, Intake, Observed Effects).

This assumption is included in the system, which will discard all sets where the values for one or more LoEs are missing. It is in charge of the user to provide a dataset as much complete as possible in order to obtain the higher number of results as possible.



**BOX1: Information to be asked to the expert/decision-maker**

1. for each chemical:
  - definition of the membership function for the LoE “Source” (to set the 2 thresholds)
  - definition of the membership functions for the LoE “Intake” (to set the 2 thresholds)
2. for each disease:
  - definition of the membership function for the LoE “Observed Effects” (set the 2 thresholds)
  - definition of the value of parameter  $\omega_j$
3. for each pair [chemical, disease]:
  - definition of the value of parameter  $\eta_{i,j}$  (relative value associated to the relationship chemical  $i$ th and disease  $j$ th in the case of more than one chemical related to one disease)
4. for each LoE:
  - definition value of the relevance weight  $\bar{\delta}_p$  assigned to each LoE/criterion
5. to define how to deal with cases where multiple substances are associated to one disease:
  - definition of the value of  $\alpha$  representing end-user’s approach to deal with uncertainty in causes of diseases.

**4.3. PHASE 2: Ranking of Environmental Chemical Stressors at the Regional Scale**

The objective of this step consists in selecting the priority chemicals at the regional level, after having obtained the ranking of chemicals within each EGU for which data are available.

Within each EGU, a score for each pair [chemical, disease]  $(H_{i,j})_u$  and a final ranking score  $(H_i^{\text{CHEM}})_u$  for each chemical have been estimated in the previous phase of the methodology.

At this point, the criteria chosen to rank the chemical at the regional scale are the following ones:

- the number of EGU in which the chemical is detected;
- the value of the score  $H_{i,j}$  in the  $u^{\text{th}}$  EGU;
- the population  $p_u$  of each EGU.

To include all this information into a unique indicator, the Weighted Sum is chosen, where the scores  $(H_i^{\text{CHEM}})_u$  obtained by a chemical in different EGUs are summed after multiplying each of them by the amount of the population in the corresponding EGU:



$$H_i^{REGIONAL} = \sum_{u=1}^w P_r \times (H_i^{CHEM})_u$$

As values are summed up, the final ranking is influenced by the availability of data for the chemicals of concern within each EGU. Consequently, it is necessary to track the presence of data concerning each chemical.

To face this issue, every score value has attached a “precision percentage value” which is calculated by dividing the number of EGUs ( $En_i$ ) where the chemical of concern has a score value by the total number of EGUs ( $En$ ):

$$P_i^{REGIONAL} = \frac{En_i}{En}$$

#### **4.4 PHASE 3: Ranking of the Elementary Geographic Units (EGUs)**

The objective of this step is to obtain a ranking of the Elementary Geographic Units in order to highlight those areas (counties, provinces, ...) within the region where further investigations can be focused.

The criteria chosen for the evaluation of the “priority score” of each EGU is the sum of all the scores  $H_i^{CHEM}$  estimated within that EGU for all the UC chemicals detected in that EGU. The number of inhabitants of that EGU is considered as well (as a weighting parameter). Therefore:

$$S_u = \left[ \sum_{i=1}^C (H_i^{CHEM}) \right] \times P_u$$



## 5. Conclusions

This report provides a complete description of the methodology proposed for the ranking of environmental chemical stressors at the regional scale, including the conceptual backgrounds (Chapter 3) and the methodological development (Chapter 4). The methodology is being implemented in a “Risk-based Tool for the Regional Ranking of Environmental Chemical Stressors”.

The proposed methodology will be applied to a case-study, chosen within the European context, with the aim of testing and validating in on a real dataset. For this purpose, 2-FUN Case-Study n.3 (Upper Silesia case-study) was initially selected, but data needed for the application were made available only for two chemicals. As it is considered necessary to test the methodology on a more complete dataset, the search for an alternative case-study was also started in the last months, through contacts with European research institutes involved in HEIMTSA and ENVIRISK projects.

The required case-study should have the following characteristics.

For a minimum set of 4 substances, data are required for:

- LoE “Environmental Contamination”, i.e. contamination data from environmental monitoring for one or more environmental matrices;
- LoE “Intake”, i.e. data on exposure biomarkers measured in the population

For a minimum set of 4 disease associable to at least one of selected chemicals, data are needed for:

- - LoE “Observed Effects”, i.e. data on incidence/prevalence of the diseases in the population living in the region.

A second part of this report will therefore be delivered after the application of the methodology to a selected case-study, critically illustrating advantages and possible limitations of the proposed approach.



## References

- Albertini R., Bird R., Doerrer N., Needham L., Robinson S., Sheldon L., Zenick H., 2006. The Use of Biomonitoring Data in Exposure and Human Health Risk Assessment. *Environmental Health Perspectives* 114 (11): 1755-1762.
- Calvo T., Mayor G., Mesiar R., 2002. Aggregation operators: new trends and applications. Physica-Verlag Heidelberg.
- Chowdhury S., Champagne P., McLellan P.J., 2009. Uncertainty Characterization Approaches for Risk Assessment of DBPs in Drinking Water: a Review. *Environmental Management* 90: 1680-1691.
- DeWoskin R., 2007. PBPK models in risk assessment – A focus on chlorophene. *Chemico-Biological Interactions* 166: 352-359.
- Figueira J., Ehrgott M., Greco S., 2005. Multiple Criteria Decision Analysis: State of the Art Surveys, Springer, Berlin.
- Giove S., Brancia A. Satterstrom F.K., Linkov I., Decision Support Systems and Environment: Role of MCDA. In Marcomini A., Suter G.W. II, Critto A., (Eds). Decision Support Systems for Risk Based Management of Contaminated Sites. Springer Verlag (2009), New York.
- IPCS, 1993. Biomarkers and Risk Assessment: Concepts and Principles. Environmental Health Criteria 155. International Programme on Chemical Safety, WHO, Geneva.
- Klement E. P., Mesiar R., Pap E., 2000. Triangular norms. Trends in Logic, Studia Logica Library, Vol.8, Kluwer Academic Publishers.
- Li J., Huang G.H., Zeng G., Maqsood I., Huang Y., 2007. An Integrated Fuzzy-Stochastic Modelling Approach for Risk Assessment of Groundwater Contamination. *Journal of Environmental Management* 82: 173-188.
- Linkov I., Loney D., Cormier S., Satterstrom F. K., Bridges T., 2009. Weight of Evidence evaluation in environmental assessment: Review of qualitative and quantitative approaches. *Science of the Total Environment* 407: 5199-5205.
- Linkov I., Satterstrom F.K., Kiker G., Batchelor C., Bridges T., Ferguson E., 2006. From comparative risk assessment to multi-criteria decision analysis and adaptive management: recent developments and applications. *Environment International* 32: 1072-1093.
- Mckone T.E. and Deshpande A.W., 2005. Can Fuzzy Logic bring Complex Environmental Problem into Focus? *Environmental Science and Technology* 39 (2): 42A–47A.
- Menzie C.A., MacDonell M.M., Mumtaz M., 2007. A Phased Approach for Assessing Combined Effects from Multiple Stressors. *Environmental Health Perspectives* 115 (5): 807-816.
- Morgenstern H. and Thomas D., 1993. Principles of study design in Environmental Epidemiology. *Environmental Health Perspectives Supplements* 101 (S4).
- Ryan P.B., Burke T.A., Cohen Hubal E.A., Cura J.J., McKon T.E., 2007. Using biomarkers to inform cumulative risk assessment. *Environmental Health Perspectives* 115 (5): 833-840.
- Smolders R. and Schoeters G., 2007. Identifying opportunities and gaps for establishing an integrated EDR-triad at a European level. *International Journal of Hygiene and Environmental Health* 210: 253-257.
- Stouthard M.E.A., Essink-Bot M.L., Bonsel G. J., 2000. Disability Weights for Diseases: a Modified Protocol and Results for Western European Region. *European Journal of Public Health* 10 (1): 24-30.
- Vinke P., 1992. Multi-criteria Decision Aid. John Wiley and Sons, Chichester.
- WHO, 2004. Global Burden of Disease 2004 Update: Disability Weights for Diseases and Conditions. World Health Organization, Geneva.
- Yager, R.R., 1981. A new methodology for ordinal multiple aspect decisions based on fuzzy sets, *Decision Sciences* 12, 589-600.
- Zadeh, L.A., 1965. Fuzzy sets. *Information and Control* 8, 338–353.