A PRECAUTIONARY GROUNDWATER POLLUTION POTENTIAL ASSESSMENT AIDED BY A FUZZY LOGIC SYSTEM

E. Cameron[1], G. F. Peloso[2]

[1]GeoStudio, P.zza S. Antonio 15, 23017 Morbegno (SO), Italy.

Tel.: +39 0342 615987, E-mail: geostd@libero.it

[2]Universitàdegli studi di Pavia, Dipartimento di Scienze della Terra, Via Ferrata 1, 27100 Pavia, Italy.

Tel.: +39 0382 505828, E-mail: gpeloso@unipv.it

Abstract

This paper presents a fuzzy logic based approach to data uncertainty management in groundwater pollution potential assessment; a modified parametric model produces a pollution potential score that ensures a degree of groundwater protection which increases with parameters measurement or estimation uncertainty. The model behavior is investigated and the consequences of its use are outlined with respect to risk analysis.

Keywords: fuzzy logic, data uncertainty management, groundwater pollution potential assessment, SINTACS, f-SINTACS

1. PREMISES

Aquifers' pollution potential assessment is an important planning and pollution prevention tool.
Among the assessment methods DRASTIC (Aller et. al. 1987) and SINTACS (Civita and De Maio, Italian Research Council, 1997) assign a partial pollution potential score to each value of the parameters assumed to be relevant (groundwater table depth, net recharge etc.) and produce a total pollution potential score as a weighted sum of the partial ones, with weights that depend on the specific hydrogeological situation.
The actual parameters values are generally not known exactly, mainly because of measurement or estimation errors, and the same apply, therefore, to the true score. This raises the problem of choosing an appropriate pollution potential level for planning and design decisions.
The problem can simply be ignored using only the measure or estimate of each parameter, obtaining directly the final score; it can be argued, however, that because this last is affected by some degree of uncertainty the real pollution potential at a given site may be underestimated, and therefore planning and design decision might not be sufficiently safe for groundwater resources.
Parameter data quality can be increased but time, costs and different technical limitations may not allow to reach the desired level of accuracy.
The model presented in this paper attempts to introduce a safety degree in pollution potential assessment with a given data set; a fuzzy logic system calculates a partial score starting from a parameter measure or best estimate and the mathematically formulated notions of acceptable approximation to its actual value and of negative influence of the parameter itself on pollution potential level.
The model is expected to introduce a quantified and reasonable degree of precaution in pollution potential assessment, reducing the possibility to take planning and design decisions not sufficiently safe for groundwater resources. The underlying precaution principle receives the common interpretation which suggests to take decisions with a safety level directly related to the uncertainty of the information on which they are based; there is no radical or prejudicial action restraining approach and the procedure, which will now be briefly outlined, is mathematically defined even in its subjective elements.
Let's assume that a_P is the actual value of a given parameter P and that m_P is its measure or best estimate; m_P will be thought as the best available approximation of a_P.

Let's suppose now that the higher (or lower) is the value of the parameter *P*, the higher is the related partial score; to obtain a safer pollution potential assessment because of uncertainty, then, a parameter level higher (or lower) than m_P should be used, because the resulting score will be higher than the one calculated using m_P directly. However, due to the above-mentioned interpretation of m_P, numbers increasingly different from this last are less acceptable as approximations of the parameter actual value, so only some degree of deviation is tolerable.

A choice of compromise between two requirements must be made. The model discussed hereafter, derived from SINTACS and called *f*-SINTACS ("f" stands for fuzzy) performs such choice, given some control factors. A key one is the amplitude of the range of numbers considered to some degree acceptable approximations of the actual value a_P of a parameter *P*, provided that such degree is the highest possible for m_P under all conditions; if the amplitude increases then m_P is thought to be a less reliable approximation of a_P, and the model deals with the resulting uncertainty by allowing numbers progressively distant from the former to be considered, to some extent, acceptable approximations of a_P and potentially be used to calculate the partial score. This last, because of the inference process features, will be higher than the one computed when the choice is forced to remain nearer to m_P or to coincide with it, except for special cases like those discussed later.

F-SINTACS and SINTACS produce the same pollution potential score when m_P is considered the only acceptable approximation of a_P; the effect of uncertainty is then removed, and thus it can be quantitatively assessed by comparing the scores computed under the two different conditions.

The conclusions about hydrogeological systems on which SINTACS is based are assumed to be valid; this last, particularly suited for the Mediterranean area, being well experimented in Italy has been considered a good starting point.

This work draws on previous ones regarding DRASTIC (Cameron, E., Peloso, G. F., 2000 and 2001) and, besides discussing a somewhat different assessment method, develops the analysis of the basic concepts and improves the investigation of the model behavior and the consequences of the approach employed. If further tests give positive results a limited software release may be issued.

2. SOME INTRODUCTORY CONCEPTS ABOUT FUZZY SET AND FUZZY LOGIC

A fuzzy set S is described by a function *m* that assigns to an element *x* of a chosen reference set a membership degree $m(x)$ of *x* to S, where $0 \leq m(x) \leq 1$. A zero membership degree, that is $m(x) = 0$, means that *x* does not belong to S while if $m(x) = 1$ then *x* belongs to S as in the ordinary set theory. Values greater than zero and less than one means, freely speaking, that *x* belongs to *S* only to a certain extent.

As an example consider the fuzzy set *high temperature* of Fig. 1 (fuzzy set names will be written in italic type); the function domain is the interval $[0,40]$ whose elements are assumed to be the temperatures, in Celsius degrees, inside a house situated in a temperate climate area.

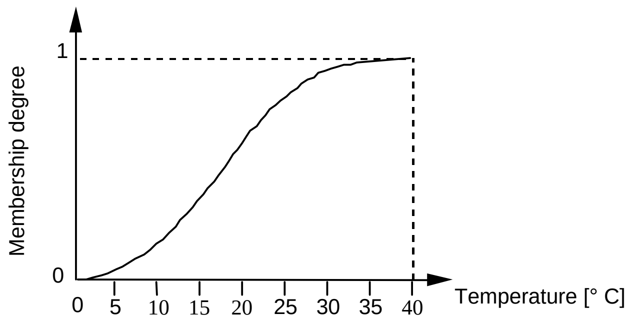

Fig. 1. The fuzzy set *high temperature*

The fuzzy set of Fig. 1 actually provides a mathematical model for the concept of high temperature and to what extent a temperature *T* is compatible with this notion is given by the membership degree $m(T)$ of *T* to the fuzzy set.

From the graph it can be observed that $g(3) = 0$, $g(40) = 1$ and $g(18) \cong 0.5$ meaning that 3, 40 and 18 degrees are, respectively, not compatible, fully compatible and half compatible with the concept of high temperature as represented by the function of Fig. 1.

The function shape is of particular importance from a semantic point of view and like other fuzzy set features depends on the specific problem and the concept to be represented. The choice, usually not univocal, may involve some subjectivity, while the final result, a function, is clearly defined; it cannot be thought as a probability density function because it has a different meaning and, in general, lacks the required properties.

Among the set-theoretical operations that can be performed with fuzzy sets intersection is of special interest for this paper: given two fuzzy sets A and B and the memberships degrees

$m_A(x)$ and $m_B(x)$ of an element *x* to each, then the membership degree of *x* to $A \cap B$ is some convenient function $m_{A\cap B}(x)$ of $m_A(x)$ and $m_B(x)$, where $0 \le m_{A\cap B}(x) \le 1$; for example $m_{A\cap B}(x) = \min[m_A(x), m_B(x)]$ for the standard Zadeh intersection operation or $m_{A\cap B}(x) = m_A(x) \cdot m_B(x)$ for the product one.

A logic based on fuzzy set theory can be developed firstly by asserting that the membership degree $m(x)$ of an element *x* to a fuzzy set S is equal to the truth degree $t(x)$ of the statement "*x* is S".

Thus, considering the preceding example, the statements P_1="[a temperature of] 3 °C is high", P_2="18 °C is high" and P_3="40 °C is high" have a truth degree $t(P_1) = m(3) = 0$, $t(P_2) = m(18) \cong 0.5$ and $t(P_3) = m(40) = 1$ meaning, again loosely speaking, that the first statement is false, the second is half true and the third is true. Phrases in square brackets will be sometimes added for clarity.

If P and Q are, respectively, the statements "*x* is A" and "*y* is B " then a truth degree $t_{P\wedge Q}$ can be assigned to the composite statement "*x* is A AND *y* is B" knowing those, $t(P)$ and $t(Q)$, of P and Q; for example $t_{P\wedge Q} = \min[t(P), t(Q)]$ or $t_{P\wedge Q} = t(P) \cdot t(Q)$ with analogy to fuzzy sets intersection. In the second case, considering the fuzzy set of fig. 1, $m(20) = 0.8$ and $m(30) = 0.9$, so the statement "20° C is high AND 30° C is high" has the truth degree $t(20) \cdot t(30) = m(20) \cdot m(30) = 0.8 \cdot 0.9 = 0.72$.

Other set-theoretical operations and logical operators are also defined but they will not be discussed here.

Fuzzy logic, naturally, includes inference methods; the simple one used by *f*-SINTACS will be exemplified later.

3. INFERENCE STEPS

The core process to assess pollution potential with *f*-SINTACS is the inference of seven weighted partial scores s^*_{Pi} ($1 \le i \le 7$), one for each parameter considered by SINTACS, which are summed to obtain the total one $S^*_T = \sum_{i=1}^{7} s^*_{Pi}$. The partial scores are weighted because they already incorporate, as it will be clear later, the weights (the relevance) assigned to them in pollution potential assessment.

The parameters are: water table depth, net recharge, self purification effect in the zone of aeration, soil type, aquifer type with respect to lithology, hydraulic conductivity and slope of the topographic surface; the corresponding Italian words are used for the acronym SINTACS (as for DRASTIC in English).

The inference algorithm to calculate the partial scores, implemented with Wolfram Research's software Mathematica v. 4.01, will be described for net recharge and comprises the following steps:

1. Retrieval of net recharge measure or best estimate m_{NR}, though as the best approximation of the real value a_{NR}
2. Retrieval of weight w_{NR} attributed to net recharge
3. Choice of the net recharge value $\hat{v}_{NR}$ that is to the highest degree unfavourable for pollution potential and an acceptable approximation of a_{NR}, given the fuzzy set that represents this last notion.
4. Inference of the partial score s^*_{NR} through the rule "IF $\hat{v}_{NR}$ is unfavorably high AND w_{NR} is high THEN s^*_{NR} is high"

The rule (Fox E., 1994) may be read as: "insofar as the net recharge value $\hat{v}_{NR}$ is compatible with the concept of unfavorably high [for pollution potential] and as the weight w_{NR} is compatible with the concept of high, then make the partial score s^*_{NR} compatible with the concept of high".

When the process is executed using only the measure or best estimate m_{NR} of net recharge then, in the third step, $\hat{v}_{NR} = m_{NR}$.

Steps 1 to 4 are repeated for each parameter and the resulting partial pollution potential scores are added to obtain the total one.

To perform the inference it is necessary to provide the fuzzy sets that mathematically represent the concepts of acceptable approximation to net recharge actual value, unfavourably high net recharge, high weight and high partial score; also the inference method and the type of AND operator appearing in the rule, among the different options allowed with fuzzy logic, must be chosen. The first issue will be discussed in the next two paragraphs; the second will be illustrated through the exemplification.

4. THE MODEL FOR THE CONCEPT OF ACCEPTABLE APPROXIMATION TO PARAMETERS ACTUAL VALUES

Because to assess pollution potential *f*-SINTACS normally chooses parameters values different from the measures or estimates that are considered the best approximations of the actual ones, it is necessary to establish how much other options are themselves acceptable as approximations, to remain near enough to the best ones.

It is reasonable to assume that the measure or best estimate m_P of a parameter *P*, being the best approximation to the actual value a_P, fully corresponds to the concept of acceptable approximation of this last, while numbers progressively greater or smaller than m_P are less suitable; thus the fuzzy set that represents the concept of acceptable approximation of a_P should have a maximum of one in m_P and membership degrees gradually decreasing on both sides of it. Since the membership degree of m_P to the fuzzy set is one, so is the truth degree of the statement "m_P is an acceptable approximation of a_P", while for other values *x* the truth degree of the corresponding statement progressively decreases as the difference from m_P increases. It will be assumed that the membership and truth degrees are zero when $x \le m_P + e_{P1}$ or $x \ge m_P + e_{P2}$, where e_{P1} and e_{P2} are two positive values and it can be $e_{P1} \ne e_{P2}$; this means that at the ends or outside the interval $[m_P - e_{P1}, m_P + e_{P2}]$ there are no acceptable approximations of a parameter actual value a_P and therefore no possible model choices.

These notions will be discussed for net recharge.

Let us suppose that the measure or best estimate m_{NR} of net recharge for a given site is 100 mm/year and that e_{NR1}= e_{NR2}= 25 mm/year; the fuzzy set for the concept of acceptable approximation of net recharge actual value is a bell-shaped PI function (Cox E., 1994) which associates a membership degree of one to m_{NR} and whose domain is the interval $[m_{NR} - e_{NR1}, m_{NR} + e_{NR2}] = [100 - 25, 100 + 25] = [75, 125]$ that contains, ends excluded, all the possible model choices. The values m_{NR}, e_{NR1} and e_{NR2} univocally determine the function and thus the fuzzy set, which therefore will be denoted with $\overline{A}_{NRA}(m_{NR}, e_{NR1}, e_{NR2})$, where $\overline{A}_{NRA}$ reminds that the concept represented is of "Acceptable approximation of Net Recharge Actual value". Since m_{NR} = 100 mm/year and e_{NR1}= e_{NR2}= 25 mm/year the fuzzy set $\overline{A}_{NRA}(100,25,25)$ shown in fig 2 is obtained.

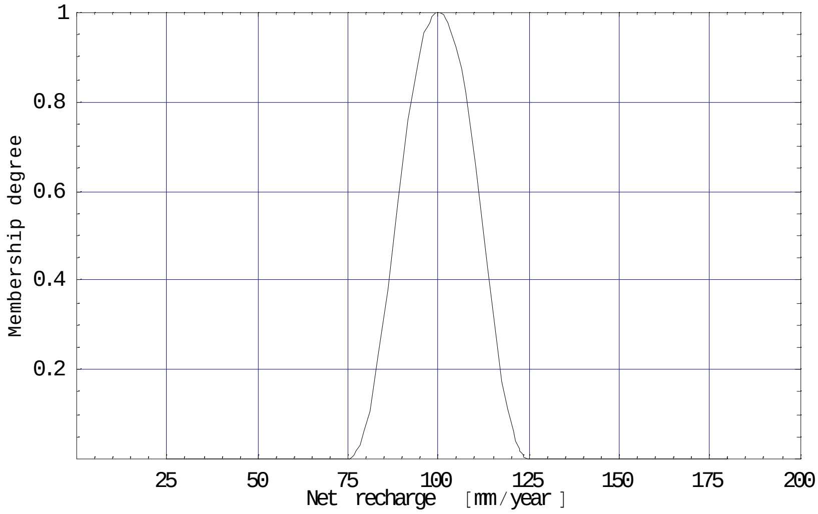

Fig. 2. The fuzzy set $\overline{A}_{NRA}(100,25,25)$

If e_{NR1} or e_{NR2} increase so does the fuzzy set domain amplitude; numbers more distant from the measure or best estimate m_{NR} are considered to some degree acceptable approximations of the net recharge actual value a_{NR}, their previous membership degree to the fuzzy set raises, possibly from 0, and they become potential model choices. This means that m_{NR} still is the best approximation of a_{NR} but is also less reliable, so more values are accepted as approximations of a_{NR} and allowed to be chosen by the model to make pollution potential assessment precautionary enough for the new uncertainty condition.

For m_{NR} = 100 mm/year and e_{NR1}= e_{NR2}= 80 mm/year the fuzzy set $\overline{A}_{NRA}$ becomes that of fig. 3, while an asymmetric example $\overline{A}_{NRA}(100,25,80)$ is shown in fig. 4.

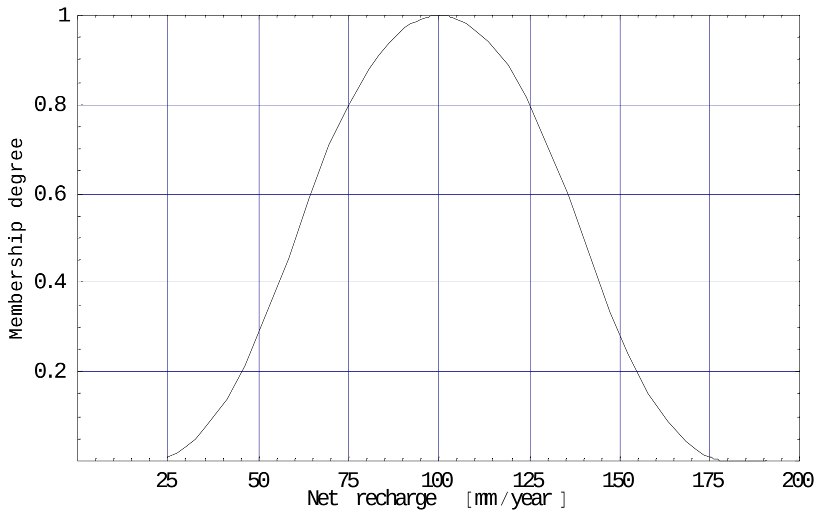

Fig. 3. The fuzzy set $\overline{A}_{NRA}(100,80,80)$

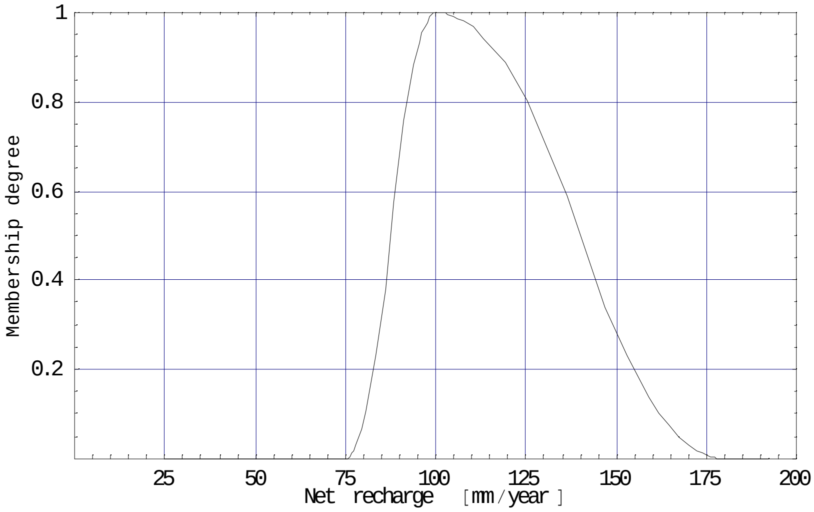

Fig. 4. The fuzzy set $\overline{A}_{NRA}(100,25,80)$

In this case the reliability of m_{NR} as an approximation to a_{NR} "from above" is considered higher then "from below".

5. THE MODELS FOR THE CONCEPTS OF UNFAVOURABLY HIGH NET RECHARGE, HIGH WEIGHT AND HIGH PARTIAL SCORE

The concept of unfavorably high net recharge is represented by the fuzzy set whose function, with domain $[25,550]$, is shown in Fig. 5.

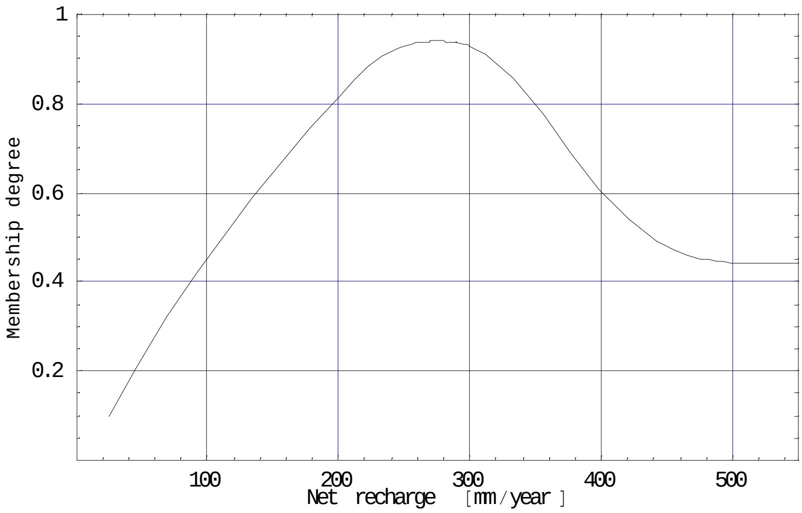

Fig. 5. The fuzzy set *unfavorably high net recharge*

The membership degree $m_1(v_{NR})$ of a given net recharge value v_{NR} to the fuzzy set expresses to what level v_{NR} can be considered unfavorably high for pollution potential, and corresponds to the truth degree $t_1[\mathrm{P}_1(v_{NR})]$ of the statement $\mathrm{P}_1(v_{NR}) =$ "v_{NR} is unfavorably high"; the functional notation is used because of the variable term v_{NR}. From the graph, for example, it can be observed that the truth degree of the statement $\mathrm{P}_1(200) =$ "200 [mm/year] is unfavorably high [for pollution potential]" is $t_1[\mathrm{P}_1(200)] = m_1(200) = 0.82$ while $t_1[\mathrm{P}_1(400)] = m_1(400) = 0.6$.

The fuzzy set shape is that of the curve which in SINTACS directly correlates net recharge values to their respective partial scores; these latter are simply divided by 10 to obtain the membership degrees on the vertical axis. The higher the partial score, and hence the membership degree, the unfavourable the net recharge amount. Shape preservation means that the influence of net recharge on pollution potential in *f*-SINTACS and SINTACS is the same, but the mapping of Fig. 5 mathematically defines a concept for the inference algorithm.

The function values decrease after a maximum, since if an increasing net recharge favours the movement of pollutants toward groundwaters it also causes dilution and dispersion effects that lowers the concentrations; this and other phenomena reduces the adverse effect of net recharge on pollution potential beyond a value of about 280 mm/year. Also the membership degrees ranges from 0.1 to 0.95 and this limits, in *f*-SINTACS and SINTACS alike, the parameter influence on final computation.

The weights w_i (i =1, 2, …,7) attributed by SINTACS to each parameter are normally integer numbers dependent on the hydrogeological situation and the anthropic impact, but each weight can be modified by the user to better adjust the assessment to the specific case, provided that

$\sum_{i=1}^{7} w_i = 26$ and $w_i \le 5$, i =1, 2, …,7.

With *f*-SINTACS how much each weight can be considered high is given by the fuzzy set named *high weight* shown in Fig. 6, a linear function whose domain is the interval $[0, \ ]$

of a discrete set; although fuzzy sets can be defined over discrete domains the choice of a continuous interval allows to treat uncertainty in weight attribution

described later for parameters. This case, however, will not be discussed further.

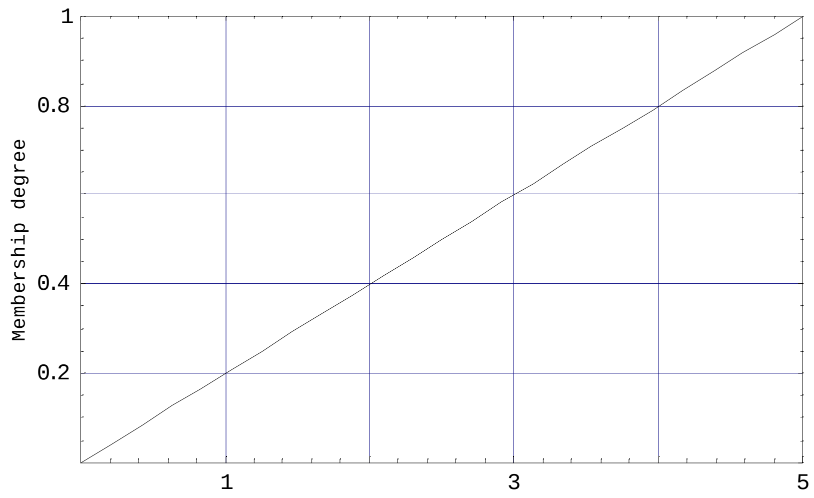

Fig. 6. The fuzzy set *high weight*

The compatibility level of a weight w_i with the concept of high weight is its membership degree $m_2(w_i)$ which corresponds to the truth degree t_2 of the statement $\mathrm{P}_2(w_i) =$ "w_i is high"; thus, for example, $t_2[\mathrm{P}_2(4)] = m_2(4) = 0.8$.

Finally *f*-SINTACS and SINTACS attributes to each parameter a partial score from 0 to 50; like for weights the mathematical model for the concept of high partial score is the linear function of Fig. 7, which can readily be interpreted considering the previous examples; thus the partial score s_i related to the *i*-th parameter has a membership degree $m_3(s_i)$ to the fuzzy set of Fig. 7 and a truth degree $t_3[\mathrm{P}_3(s_i)] = m_3(s_i)$ of the statement $\mathrm{P}_3(s_i) =$ "s_i is high".

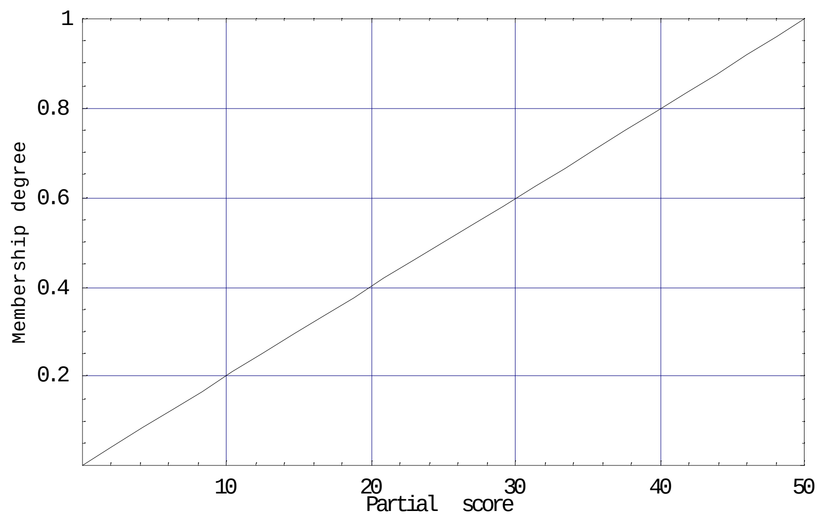

Fig. 7. The fuzzy set *high partial score*

6. PARTIAL SCORE INFERENCE USING ONLY THE MEASURE OR BEST ESTIMATE OF A PARAMETER

The partial score inference performed by *f*-SINTACS when only the measure or best estimate of a parameter is used leads to the same result as SINTACS, except possibly for small differences due essentially to numerical approximations.

The steps described in par. 3 will be exemplified for a net recharge measure or best estimate m_{NR} = 100 mm/year and the weight w_{NR} = 4 commonly assigned to the parameter.

In steps 1 and 2 the values of m_{NR} and w_{NR} are defined as inputs for the process; in step 3 $\hat{v}_{NR}$ is set equal to m_{NR} = 100 mm/year and the process immediately proceeds to step 4, which for clarity will be divided in sub-steps.

Sub-step 4.1. Evaluation of the truth degree t_1 of the statement $\mathrm{P}_1(100) =$ "$\hat{v}_{NR}$ is unfavorably high"; from Fig. 5, since $\hat{v}_{NR}$ = m_{NR} = 100 mm/year, it is $t_1[\mathrm{P}_1(100)] = m_1(100) \cong 0.46$, where $m_1(100)$ is the membership degree of 100 mm/year to the fuzzy set *unfavorably high net recharge*.

Sub-step 4.2. Evaluation of the truth degree t_2 of the statement $\mathrm{P}_2(4) =$ "[a weight of] 4 is high"; from Fig. 6 it is $t_2[\mathrm{P}_2(4)] = m_2(4) = 0.80$, where $m_2(4)$ is the membership degree of 4 to the fuzzy set *high weight*.

Sub-step 4.3. Evaluation of the truth degree *t* of the composite statement $\mathrm{P}_1(100)$ AND $\mathrm{P}_2(4)$= "[a net recharge value of] 100 [mm/year] is unfavorably high AND [a weight of] 4 is high" that forms the rule premise. The chosen AND operator ("product AND") multiplies the two truth degrees to obtain the premise one, ensures a continuous dependence of partial scores on parameters and weight values and makes those of *f*-SINTACS and SINTACS equal when uncertainty is ignored. The composite statement truth degree *t* is $t[\mathrm{P}_1(100)\ \mathrm{AND}\ \mathrm{P}_2(4)] = t_1 \cdot t_2 \cong 0.46 \cdot 0.80 \cong 0.37$.

Sub-step 4.4. Inference of the partial score s^*_{NR}. The method used, called monotonic inference, consists in searching that partial score whose membership degree to the fuzzy set *high partial score* equals the truth degree *t* calculated in the preceding step. This is univocally identified because the mapping of Fig. 7 is one-to-one and, since from sub-step 4.3 it is $t \cong 0.37$ a partial score $s^*_{NR} \cong 18.4$ is obtained.

The entire inference process is depicted in Fig. 8.

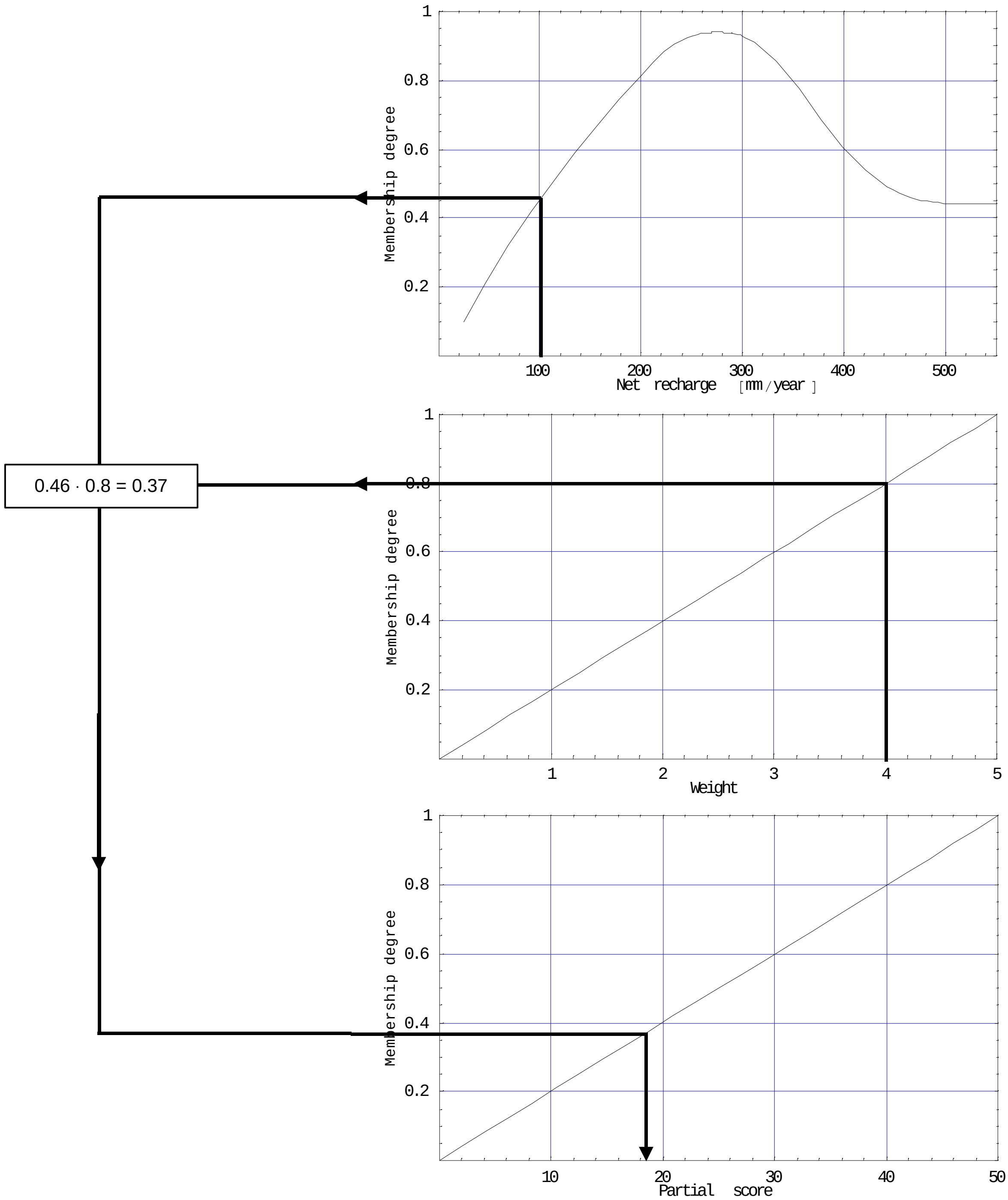

Fig. 8. Net recharge partial score inference when only the measure or best estimate of the parameter is used. Here this last is 100 mm/year and the parameter weight is 4.

7. PARTIAL SCORE INFERENCE CONSIDERING UNCERTAINTY

Because the measure or best estimate m_P of a parameter *P* is only the best approximation to the actual value a_P, and therefore the knowledge of this last is to some extent uncertain, *f*-SINTACS performs a precautionary pollution potential assessment by choosing a parameter value $\hat{v}_p$ that is to the highest degree both an acceptable approximation of a_P and unfavourably high for pollution potential, given the fuzzy set that represents these concepts. The model, in other words, attempts to prevent pollution potential underestimation due to uncertainty by choosing $\hat{v}_p$ so as to obtain a score higher than the one related to m_P, with a choice freedom restricted by the request that $\hat{v}_p$ must be to the highest possible extent a satisfactory approximation of a_P.

The choice is made in inference step 3 (see par. 3) using the fuzzy set representing the concept of acceptable approximation described in par. 4.

Suppose once more that the measure or best estimate of net recharge actual value a_{NR} is m_{NR} = 100 mm/year and that both the numbers outside the interval $[100-25{,}100+25]$ = $[75{,}125]$ and its ends cannot be accepted as approximations of a_{NR}; thus the model for the concept of acceptable approximation to net recharge actual value is given by the fuzzy set $\overline{A}_{NRA}(100{,}25{,}25)$ shown in fig 2 (see again par. 4) and the fuzzy set *unfavorably high net recharge* depicted in fig. 5 (see par. 5) represents the concept implied by the name as previously explained. Since the membership degree of a net recharge value v_{NR} to the intersection between the two fuzzy sets is the minimum between the memberships degrees of v_{NR} to each of them (see par. 2) by superimposing $\overline{A}_{NRA}(100{,}25{,}25)$ and *unfavorably high net recharge* the intersection fuzzy set obtained is that highlighted in Fig. 9 by the thicker line, and the value to which the maximum membership degree corresponds is $\hat{v}_{NR} \cong 113$ mm/year.

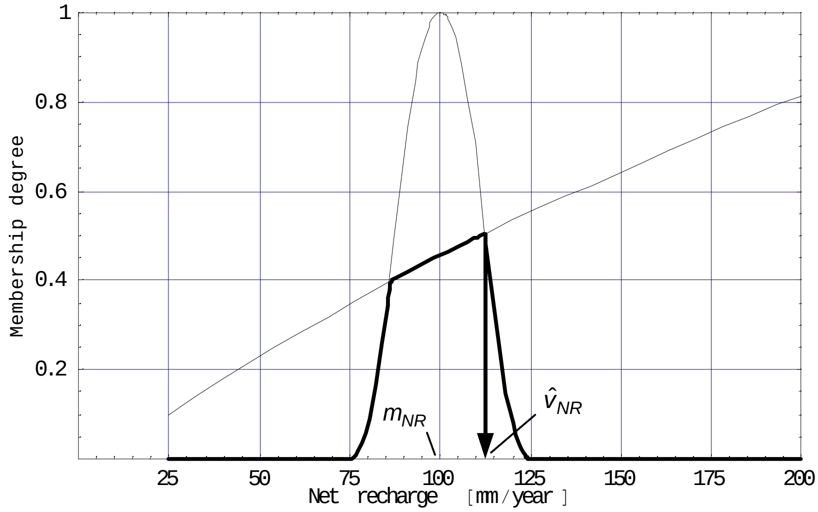

Fig. 9. Intersection between the fuzzy sets $\overline{A}_{NRA}(100,25,25)$ and *unfavorably high net recharge* (thicker line) shown for the interval $[100-25,100+25]=[75,125]$; the value $\hat{v}_{NR}$ to which the maximum membership degree corresponds is about 113.

The quantity $\hat{v}_{NR}$ has an interesting property. Values $x > \hat{v}_{NR}$ that would cause a higher pollution potential score (see fig. 5) appear to be an attractive choice for a precautionary assessment, but are less compatible than $\hat{v}_{NR}$ with the notion of acceptable approximation of net recharge actual value, because their membership degrees to the set $\overline{A}_{NRA}(100,25,25)$ are less than that of $\hat{v}_{NR}$ (see the bell-shaped curve portion that belongs to the intersection fuzzy set in fig. 9). On the other hand if $m_{NR} \le x < \hat{v}_{NR}$ the correspondence of x to the notion of acceptable approximation increases but the resulting partial (and hence total) score would be lower than the one obtained using $\hat{v}_{NR}$ (see fig. 9 and 10) and therefore the assessment would be less safe. Finally values below m_{NR} clearly cannot be accepted whether with SINTACS or *f*-SINTACS. Thus $\hat{v}_{NR}$, that can be regarded as an optimal compromise between acceptability and safeness, is chosen by the model to assess pollution potential. The same conclusion can be reached observing that $\hat{v}_{NR}$ maximizes the truth degree of the statement P(x)="*x* is unfavourably high for pollution potential AND *x* is an acceptable approximation of net recharge actual value" where the AND operator is now the standard Zadeh one described in par. 2.

The inference process, represented in Fig. 10, then proceeds as explained earlier, using $\hat{v}_{NR} \cong 113$ mm/year instead of $m_{NR} = 100$ mm/year in step 4. The membership degree of $\hat{v}_{NR}$ to the fuzzy set *unfavorably high net recharge* increases from 0.46, correspondent to m_{NR}

and calculated in par. 6, to 0.51, while the partial score increases from 18.4 to 20.2 or, rounding to integer numbers, from 18 to 20 with a 11% increment.

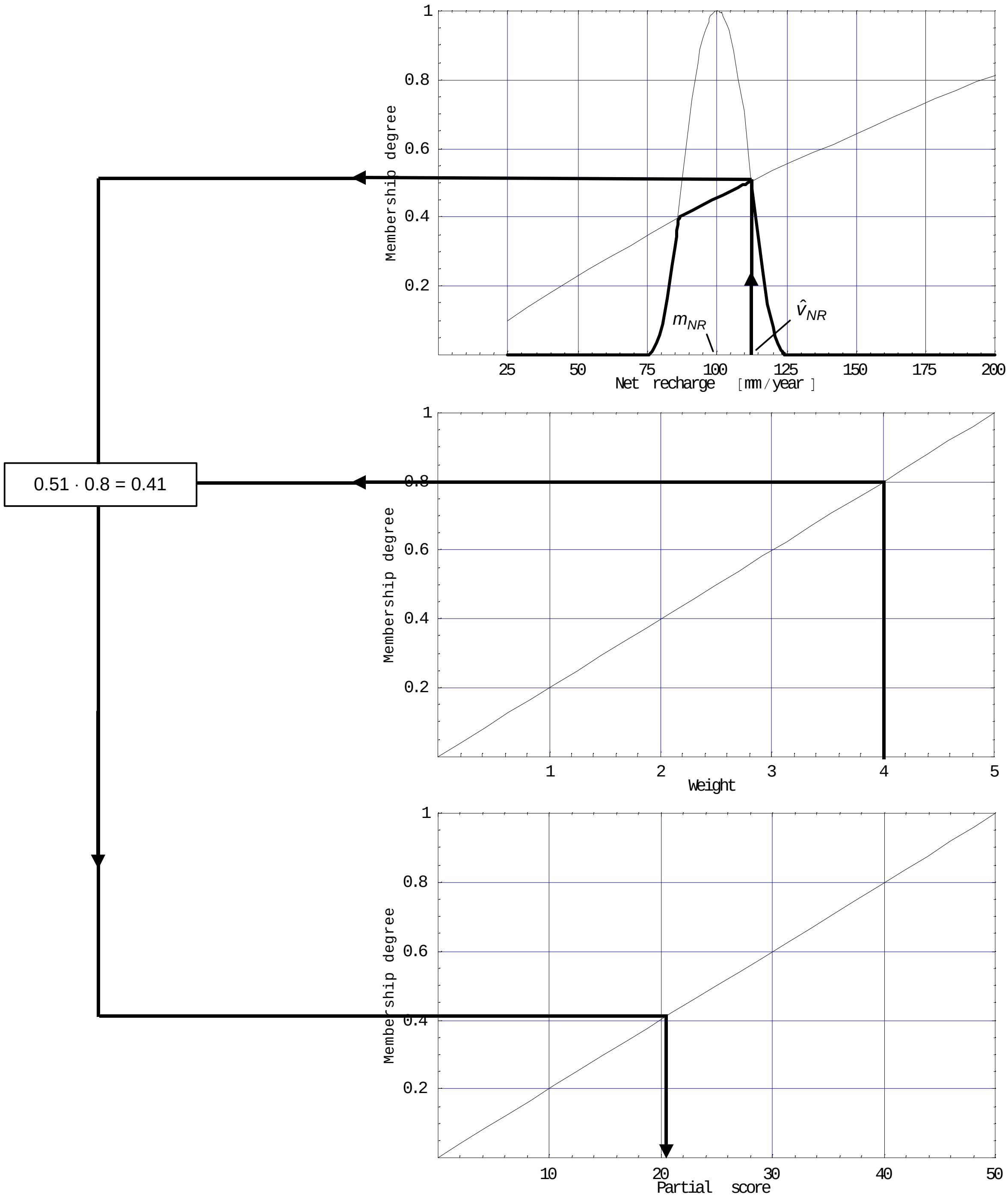

Fig. 10. Net recharge partial score inference using $\hat{v}_{NR}$. Again net recharge measure or best estimate is 100 mm/year and the parameter weight is 4; it is assumed that outside the interval $[100-25,100+25]=[75,125]$ or at its ends there are no acceptable approximation of net recharge actual value.

If the choice tolerance interval amplitude increases because the measure or best estimate m_{NR} becomes less reliable then the assessment model, in favour of safety, is allowed to select a value $\hat{v}'_{NR}$ that is more distant from m_{NR} than $\hat{v}_{NR}$ and causes and additional increment in pollution potential score; for instance if the tolerance interval is now $[100-50{,}100+50]=[50{,}150]$ the situation described in Fig. 9 becomes that of Fig. 11, $\hat{v}'_{NR} \cong 124$ mm/year (instead of $m_{NR} = 100$ mm/year and $\hat{v}_{NR} \cong 113$ mm/year) and the partial score increases again from 20 to 22 with a total percentage of 22%.

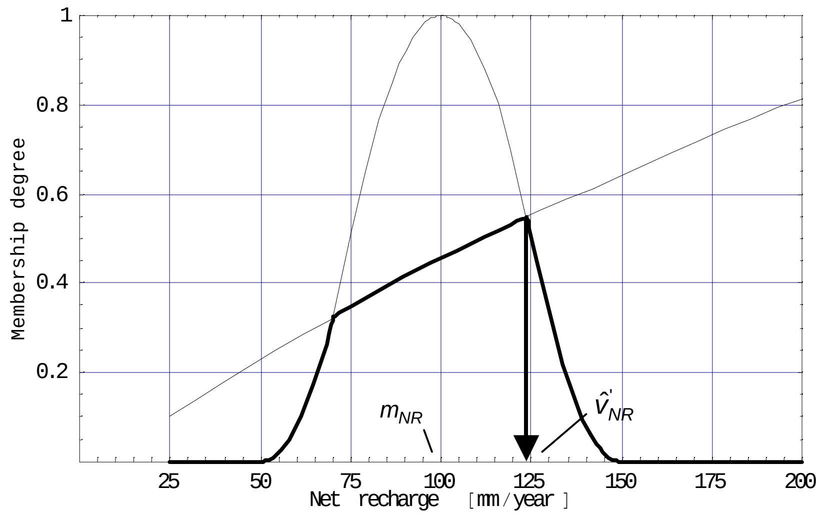

Fig. 11. Intersection between the fuzzy sets $\overline{A}_{NRA}(100{,}50{,}50)$ and *unfavorably high net recharge* (thicker line) shown for the interval $[100-50{,}100+50]=[50{,}150]$; the new value $\hat{v}'_{NR}$ to which the maximum membership degree corresponds is about 124 mm/year and the partial score increases from 20 to 22.

If m_{NR} is greater than the value to which the maximum of the fuzzy set *unfavorably high net recharge* corresponds, about 280 mm/year, then $\hat{v}_{NR} < m_{NR}$, but if the maximum belongs to the fuzzy sets intersection then $\hat{v}_{NR}$ is equal to it even if m_{NR} or the tolerance interval amplitude changes, as it is shown in Fig. 12; the adverse effect of net recharge on pollution potential and the partial score become, in fact, the highest allowed by SINTACS, and hence by *f*-SINTACS, and do not change with uncertainty nor with sufficiently small variations of m_{NR}.
The model choice, moreover, cannot fall outside the permissible values range of a parameter ($[25{,}500]$ for net recharge) but at most corresponds to its upper or lower extreme.

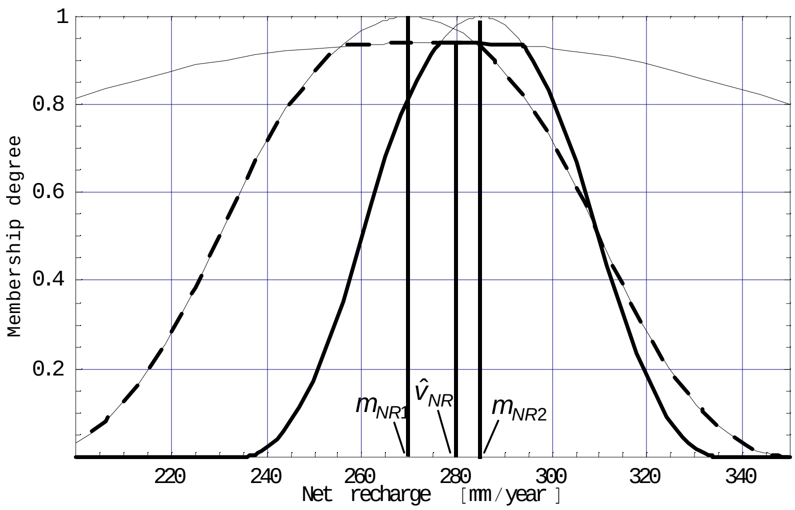

Fig. 12. If the intersection between the fuzzy sets $\overline{A}_{NRA}$ and *unfavorably high net recharge* (thicker line) contains the maximum of the latter then $\hat{v}_{NR}$ is the value, around 280 mm/year, to which the maximum is associated even if the tolerance interval amplitude changes, or m_{NR} does within certain limits

The possibility, simply by repeating the inference, to compare the results obtained ignoring or considering uncertainty as discussed, allows to quantitatively and separately determine its effect and is therefore a key model feature.

8. MODEL ANALYSIS

Decisions involving geological objects, as well as other complex systems, often imply some precaution. Foundations loads, for example, are by design normally less than soil bearing capacity also because the actual value of this latter cannot be exactly assessed due to many uncertainty factor such as soil variability, simplifications in failure phenomena description etc. Not rarely decisions involves a degree of subjectivity, which depends on experience and empirical observations.

Within *f*-SINTACS inference process subjectivity is primarily involved in choosing the functions describing fuzzy sets like $\overline{A}_{NRA}$, that are mathematical models for the concept of acceptable approximations; because the main subjective features are expressed as functions they are unambiguous, and each different proposal can be compared on a quantitative base.

Now let us assume that: 1) the risk R associated to pollution of an underground water resource is a function $R = R(P,H,V)$ of pollution potential P, hazard level H of an action and socio-economical value V of the resource, where $P,H,V \geq 0$; 2) R increases whenever one of the variables also does, if the others remains unchanged: for example $R = P \cdot H \cdot V$; 3) P is an increasing function of pollution potential score and V is constant; 4) it is possible to define an acceptable risk level R_0 such that if $R > R_0$ the action is not undertaken or must cease and vice versa if $R \le R_0$.

Using directly the parameters measures or best estimates the same pollution potential score S_T is obtained by SINTACS and *f*-SINTACS and the corresponding risk level is $R = P(S_T) \cdot H \cdot V$; if uncertainty is considered *f*-SINTACS produces a score S_T^* where, generally, $S_T^* > S_T$ and the corresponding risk level is $R^* = P(S_T^*) \cdot H \cdot V$. Because P is an increasing function of pollution potential score if $S_T^* > S_T$ then $P(S_T^*) > P(S_T)$ and therefore $R^* > R$. To reach the acceptable risk level R_0 using both estimations it must be $R^* = R = R_0$. Since $R^* > R$ and V is constant, the hazard level when the risk is assessed considering uncertainty must be changed to a value H^* so that $R^* = P(S_T^*) \cdot H^* \cdot V = R = P(S_T) \cdot H \cdot V = R_0$ and so, dividing by V and rearranging terms, $H^* = \frac{P(S_T)}{P(S_T^*)} \cdot H$; because $P(S_T^*) > P(S_T)$ then $\frac{P(S_T)}{P(S_T^*)} < 1$ and $H^* < H$. Assessing pollution potential through *f*-SINTACS, thus, leads to a hazard reduction $\Delta H = H - H^*$ required to reach the acceptable risk level and to a strengthened pollution prevention approach because of uncertainty; the related cost $C(\Delta H)$ can be assumed to increase with ΔH . If pollution potential score is already the highest possible then $H = H^*$ and $\Delta H = 0$ regardless of uncertainty.

For the purposes of this discussion other definitions of R and the problems that can be found to specify R_0 or the functions P and, especially, H, V and C can be ignored.

Another path of thoughts can be followed. Under equal conditions if pollution potential score S_{T1}^* is obtained from more reliable parameters measures or best estimates than S_{T2}^*, normally it will be $S_{T1}^* < S_{T2}^*$. The hazard levels required to reach the acceptable risk R_0 are H_1^* and H_2^* such that $R_1^* = P(S_{T1}^*) \cdot H_1^* \cdot V = R_2^* = P(S_{T2}^*) \cdot H_2^* \cdot V = R_0$ and so, again dividing by V and rearranging terms, $H_2^* = \frac{P(S_{T1}^*)}{P(S_{T2}^*)} \cdot H_1^*$; since $S_{T1}^* < S_{T2}^*$ implies $P(S_{T1}^*) < P(S_{T2}^*)$ then $H_2^* < H_1^*$, $\Delta H_2^* = H - H_2^* > \Delta H_1^* = H - H_1^*$ and $C(\Delta H_2^*) > C(\Delta H_1^*)$, where H is the hazard level calculated without uncertainty and $H_1^*, H_2^* < H$.

Accepting a higher degree of uncertainty, then, implies an additional hazard reduction $H_1^* - H_2^*$ to reach the acceptable risk level R_0 with increased costs; this theoretically allows to compare the expenditures required to improve measurement or estimates accuracy or to reduce the action hazard level to help decide which way to proceed to achieve the risk level R_0.

Finally it is also possible to tolerate less accurate measurement or estimates, for example when it is difficult to improve data quality or in preliminary assessment, and be anyway confident that *f*-SINTACS will introduce a degree of precaution in pollution potential assessment that attempts to protect groundwaters from the potentially adverse effect of uncertainty.

The partial score s_P^* obtained by *f*-SINTACS for a given parameter P can be considered a function of its measure or best estimate m_P and the positive values e_{P1} and e_{P2}, which defines the fuzzy sets that are models for the concept of acceptable approximation of parameters real values, while the other fuzzy sets are fixed. It may thus be written $s_P^* = s_P^*(m_P, e_{P1}, e_{P2})$; when $e_{P1} = e_{P2} = 0$ then $s_P^*(m_P, 0, 0) = s_P^*(m_P) = s_P$, the score calculated by SINTACS using m_P only. The partial score $s_{NR}^* = s_{NR}^*(m_{NR}, e_{NR1}, e_{NR2})$ for a net recharge measure or best estimate $100 \le m_{NR} \le 400$ mm/year and for $10 \le e_{NR1} = e_{NR2} \le 70$ mm/year is represented by the surface of Fig. 13.

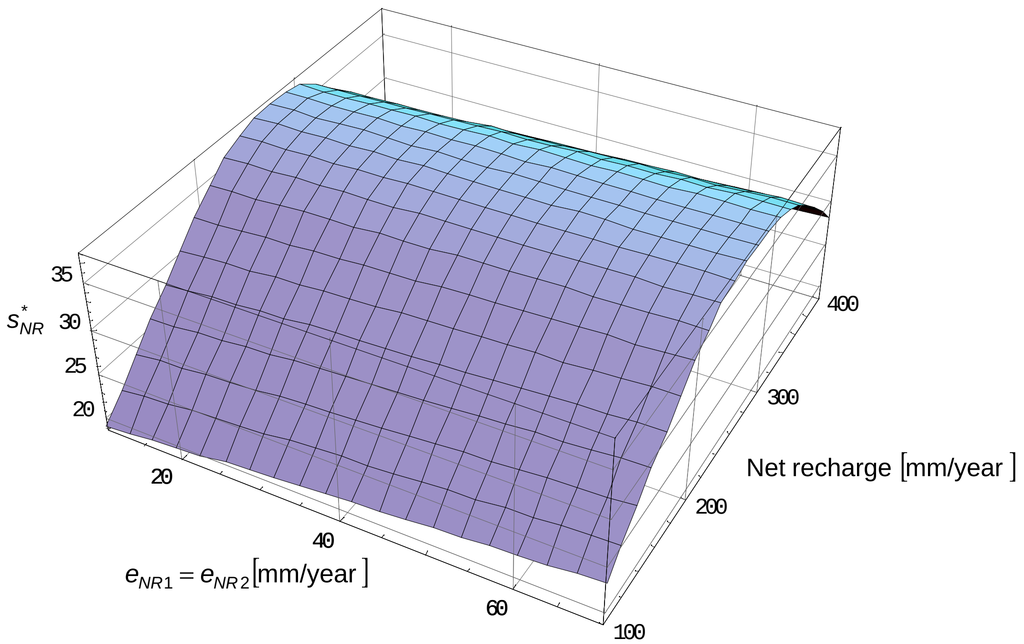

Fig. 13. The partial score $s_{NR}^* = s_{NR}^*(m_{NR}, e_{NR1}, e_{NR2})$ calculated for a net recharge measure or best estimate $100 \le m_{NR} \le 400$ mm/year and for $10 \le e_{NR1} = e_{NR2} \le 70$ mm/year.

For a given m_{NR} the partial score increases when e_{NR1} or e_{NR2} do; this means that the model introduces a degree of precaution in the assessment which intensify with uncertainty, as appropriate. The effect diminishes as m_{NR} approaches the value m_{NR0} to which the maximum of the fuzzy set *unfavourably high net recharge* corresponds, because the adverse influence of net recharge on pollution potential draw nearer to the highest possible and partial score tends to remain constant even if m_{NR}, e_{NR1}or e_{NR2} change (see par. 7); the surface top, thus, becomes progressively horizontal.

The partial score percent increment $s_i = \left(\frac{s^*_{NR} - s_{NR}}{s_{NR}} \right) \cdot 100$, where s_{NR} is the partial score calculated using m_{NR} only, is shown in Fig. 14, again for $100 \le m_{NR} \le 400$ mm/year and $10 \le e_{NR1} = e_{NR2} \le 70$ mm/year.

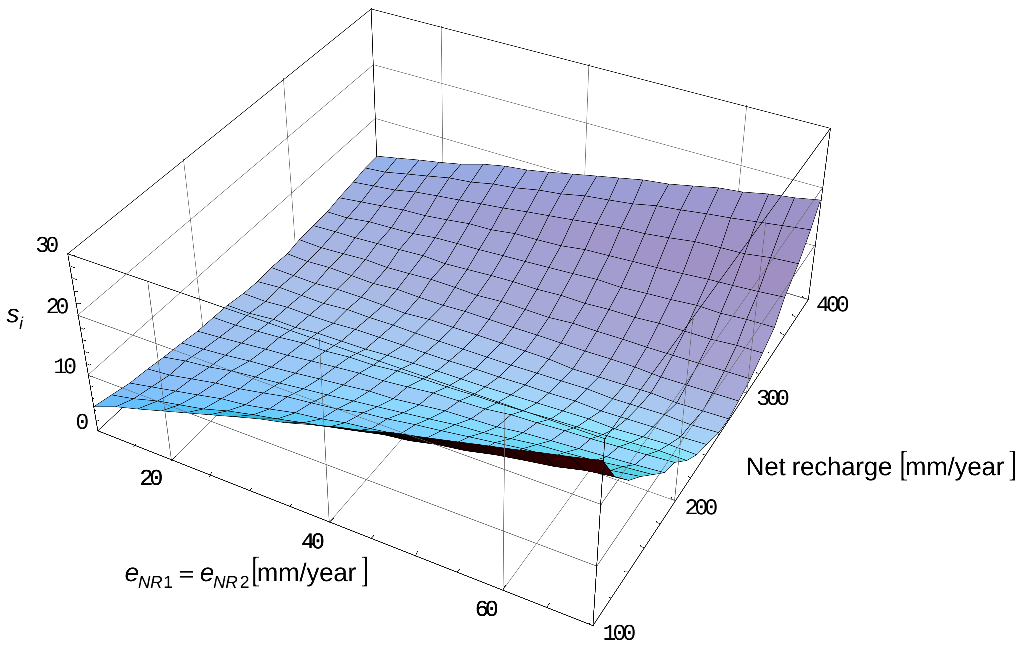

Fig. 14. The partial score percent increment $s_i = \left(\frac{s^*_{NR} - s_{NR}}{s_{NR}} \right) \cdot 100$ calculated for a net recharge measure or best estimate $100 \le m_{NR} \le 400$ mm/year and for $10 \le e_{NR1} = e_{NR2} \le 70$ mm/year.

The partial score percent increment taking uncertainty into account with respect to that obtained without it ranges from 0 to 26% and is already significant (over 10 %) when the precautionary choice of net recharge made by *f*-SINTACS is allowed to be within and interval of amplitude less then ± 25% of the net recharge measure or estimate m_{NR}.

The increment s_i progressively approaches 0 when m_{NR} come close to the value m_{NR0} previously considered, since s^*_{NR} tends to s_{NR} for the reasons already discussed.

The ratio absolute value $r = \left| \frac{s^*_{NR}(m_{NR}+1, e_{NR1}, e_{NR2}) - s^*_{NR}(m_{NR}, e_{NR1}, e_{NR2})}{s^*_{NR}(m_{NR}, e_{NR1}+1, e_{NR2}+1) - s^*_{NR}(m_{NR}, e_{NR1}, e_{NR2})} \right|$ quantify how sensitive is partial score to unit increments of net recharge measure or best estimate with respect to unit increments, on each side, of the length of the interval where the model choice is allowed to fall; the result, shown in fig. 15, ranges approximately between 2 and 6.

As m_{NR} approaches m_{NR0} the score s^*_{NR} becomes progressively independent from e_{NR1} and e_{NR2}, the ratio denominator tends to 0 and, after the appearance of numerical instabilities, r cannot be calculated anymore and the function surface is interrupted.

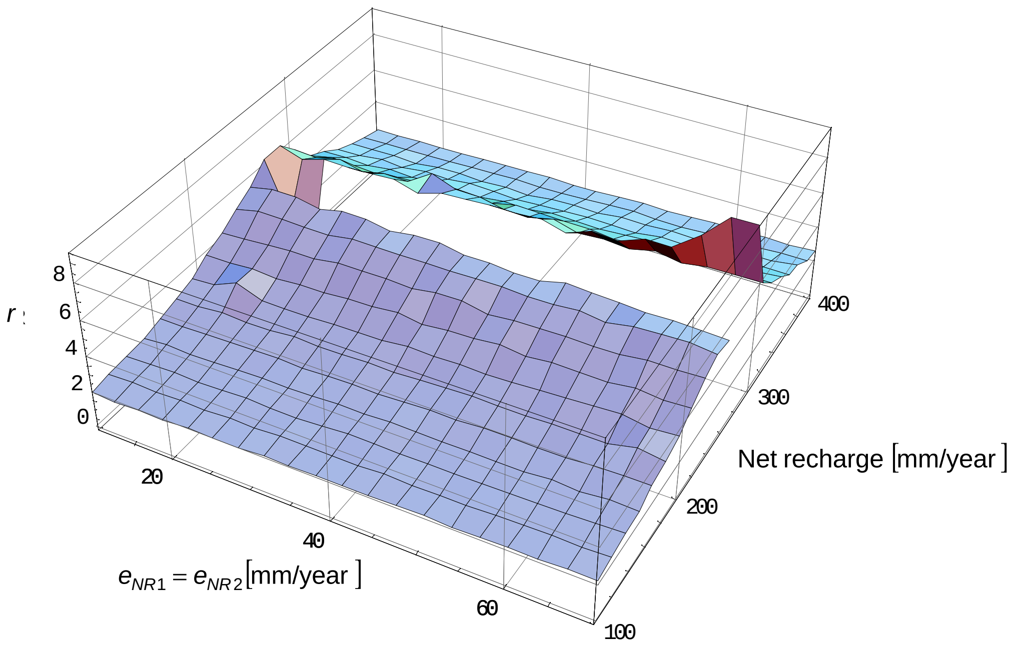

Fig. 15. The ratio absolute value $r = \left| \frac{s^*_{NR}(m_{NR}+1, e_{NR1}, e_{NR2}) - s^*_{NR}(m_{NR}, e_{NR1}, e_{NR2})}{s^*_{NR}(m_{NR}, e_{NR1}+1, e_{NR2}+1) - s^*_{NR}(m_{NR}, e_{NR1}, e_{NR2})} \right|$ calculated for a net recharge measure or best estimate $100 \le m_{NR} \le 400$ mm/year and for $10 \le e_{NR1} = e_{NR2} \le 70$ mm/year.

By comparing fig. 14, 15 and 16 it can be stated that: 1) partial score increases with uncertainty; 2) the increment is never very big even when uncertainty is considerable; 3) the increment is more sensitive to variations of parameters measures or best estimates than of uncertainty, at least when they are both small (or comparable, as it can be supposed considering the functions shapes).

These are all desirable properties. The first ensures a degree of protection induced by *f*-SINTACS that increases with uncertainty, an appropriate behavior given the model purpose. The second allows to expect that the price paid for the additional groundwater protection due to uncertainty will not be excessive. The third suggests that pollution potential score variations, under opportune conditions, depend primarily on those of parameters measures or estimates, which give the fundamental physical description of an hydrogeological system; at the same time if their reliability diminish, the importance of uncertainty in the assessment increases.

The model may suggest to protect, say, site A more than site B if parameters measure or best estimates have a lower adverse effect on pollution potential, but a higher uncertainty, in the former than in the latter; a similar situation is shown for net recharge in fig. 16.

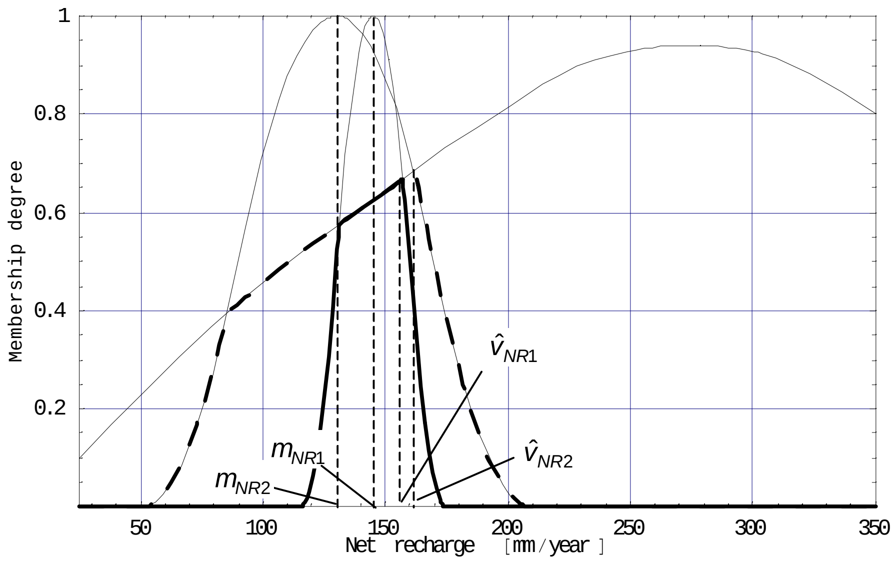

Fig. 16. Effect of differences in net recharge measures or estimation and uncertainty between two sites; it is $m_{NR1} > m_{NR2}$, but the model choices are such that $\hat{v}_{NR1} < \hat{v}_{NR2}$

This is because the model aim is to provide a shield against uncertainty itself, which is strengthened when this latter increases; the situation described, which can be easily detected by comparing pollution potential scores computed with or without uncertainty between any two sites, is expected to occur when there is a rather small difference in measures or best estimates and a marked one in uncertainty (see again fig. 16).

If parameters measures or best estimates in sites A and B lead to the same pollution potential scores and have an equal degree of uncertainty the model may suggest to protect A more than B if the effect of uncertainty can be considered more dangerous in the former than in the latter; this is when the curves portions of the fuzzy sets which express how unfavourably high or low for pollution potential a parameter is, are steeper for the values corresponding to site A than for those of site B, since then a parameter variation can have a higher adverse effect in the former than in the latter and so does uncertainty; see fig. 17.

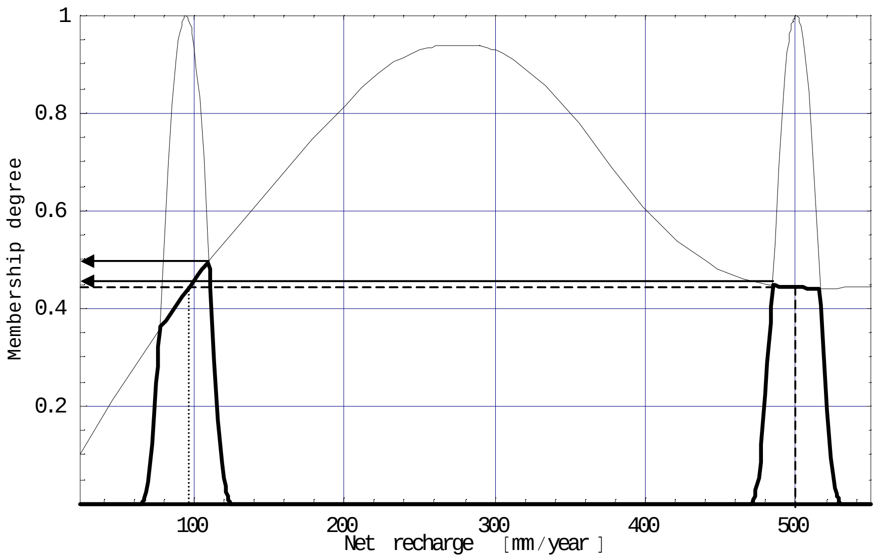

Fig. 17. Effect of parameter curve slope and uncertainty. Two net recharge values (vertical dashed lines) have the same membership degree (horizontal dashed line), and hence lead to the same partial score (see fig. 10); uncertainty is given by the bell-shapep fuzzy sets of equal amplitude. The membership degrees of the intersection maximums are different (arrows) and so are the respective pollution partial scores (see again fig. 10).

However the degrees of uncertainty should in this case be compared to parameters measures or best estimates; in fig. 17 these latter, for example, are about 95 and 500 mm/year while the bell-shaped fuzzy sets amplitude is 30 mm/year on each side in both cases. So, relatively speaking, uncertainty is much more on the first value (a 32% admissible variation) than on the second (6%). Comparing results that, in the sense described, have the same degree of uncertainty should practically eliminate this effect, which is also a consequence of slope changes as the one previously discussed.

As said before a key model feature is the possibility to repeat pollution potential score inference using only parameters measures or best estimates or considering uncertainty, obtaining two pollution potential scores S_T and S_T^* that can be compared; uncertainty effect, thus, can be quantitatively and separately evaluated for example starting from the differences $S_T^* - S_T$ at each site.

Also the model outcome should be sufficiently stable even if, to represent the concept of acceptable approximation, instead of the bell-shaped PI functions similar ones are used, such as the triangular shapes commonly employed in fuzzy set based models.

9. UNCERTAINTY DEFINITION

Because the partial score s_P^* calculated by *f*-SINTACS for a given parameter *P* is a function of its measure or best estimate m_P and the positive values e_{P1} and e_{P2} the question of how to define e_{P1} and e_{P2} arises.

There are no fixed solutions to this problem; e_{P1} and e_{P2} may, for example, be such that outside the interval $[m_P - e_{P1}, m_P + e_{P2}]$ there is a sufficiently low probability (possibly 0) to find the parameter real value, or the choice can be based on different criteria, considering also the availability and reliability of the information required to define e_{P1} and e_{P2} on a quantitative base.

10. FURTHER TESTS AND FINAL REMARKS

The model will be tested further on simulated and real hydrogeological systems, artificially introducing or estimating uncertainty about parameters values, and within decision procedures and cost/benefit schemes; the problem of how to best define e_{P1} and e_{P2} from the features of a given data set will also be analyzed in depth.

F-SINTACS may also contribute to describe, discuss and implement within a scientific framework the so-called “precaution principle” as to the debate around it.

11. REFENCES

1. Aller, L., Bennet, T., Lehr, J. H., Petty, R. J., Hackett, G. (1987). DRASTIC. A standardized system for evaluating groundwater pollution potential using hydrogeological settings. North Wales Water Authority/EPA Serv. EPA 600/2 – 85/018. Washington, 163 pp.
2. Cameron, E., Peloso, G. F. (2001). An application of fuzzy logic to the assessment of acquifers’ pollution potential. Environmental Geology, Vol. 40, num. 11/12, pp. 1305-1315. Revised and translated from the original paper published in Atti Ticinensi di Scienze della Terra, Vol. 41/2000, Ed. New Press. Como, pp. 97-108
3. Civita, M., De Maio, M. (1997). SINTACS. Un sistema parametrico per la valutazione e la cartografia della vulnerabilità degli acquiferi all’inquinamento. Metodologia e automatizzazione. Pitagora Ed. Bologna, 191 pp.
4. Cox, E. (1994). The fuzzy system handbook. Academic Press. London, 607 pp.