Spatial decisions under uncertainty: fuzzy inference in urban water management

C. K. Makropoulos and D. Butler

ABSTRACT

The endogenous complexity and spatial nature of the problems encountered in the urban water management environment present decision-makers with three major problems: (a) in the urban environment, every decision is site-specific, almost on a case-by-case basis, (b) the decision-maker must access, simultaneously, a large amount of information, increasing with rising spatial resolution and (c) the information to be evaluated is heterogeneous, including engineering, economical and social characteristics and constraints. The first two problems indicate that urban water management is an ideal field to develop and use spatial decision support systems (SDSS), while the latter promotes the use of fuzzy inference systems as a key mathematical framework. This research discusses the nature of uncertainty in environmental management in general and urban water management in particular, argues that fuzzy, rule-based, inference systems can be an invaluable tool for uncertainty quantification and presents the relevant elements of a prototype SDSS for urban water management. The examples presented in this paper are based on an application of the SDSS in water demand management.

Key words | fuzzy inference, spatial decision support, uncertainty, urban water

C. K. Makropoulos (corresponding author)
D. Butler
Urban Water Research Group,
Department of Civil and Environmental Engineering,
South Kensington Campus,
Imperial College London
London SW7 2AZ,
UK
Tel.: +44 20 75946018
E-mail: *c.makropoulos@imperial.ac.uk*

INTRODUCTION

Uncertainty in complex environmental systems

Complexity, naturally occurring in real life, is transformed into uncertainty when building an abstraction of the real world: a model. If a system is fairly simple, and therefore little uncertainty is inherent in the associated model, closed-form mathematical expressions can be used to provide insight in the way the system functions. For more complex systems for which significant data exist, model-free methods (such as neural networks) provide a methodology for dealing with uncertainty through learning, based on patterns in available data (Ross 1995). Finally, for more complex systems where scarce data exist and where only ambiguous or imprecise information is available (a widely appreciated fact in urban water management), fuzzy logic provides a framework for handling and formally expressing this uncertainty, assisting the understanding of the system's functions. The imprecision therefore in fuzzy systems is generally high. It is thus evident that the successful application of fuzzy logic, apart from knowledge of the theory involved, is dependent on the correct identification of the type of uncertainty involved in the system's structure. Despic & Simonovic (1997) identify two basic forms of uncertainty in environmental modelling: uncertainty caused by inherent stochastic variability and uncertainty due to a fundamental lack of knowledge (including model structure and model parameters). Intuitively, the second form can be more readily modelled by fuzzy sets, particularly if there is a need for quantification of qualitative criteria and synthesis of non-homogeneous information (technical, economical, political, etc.), where lack of knowledge is coupled with scarce available data and ambiguous cause–effect relationships.

Complexity in designing for a sustainable urban environment

Urban environmental planning is currently burdened with the task of creating a sustainable urban environment

(Vlachos & Braga 2001). But what exactly does this entail? Following the discussion on the urban 'ecological footprint' (Wackernagel & Rees 1996), a sustainable urban environment is one that minimises the city's dual role in a catchment's fragile equilibrium: the role of a sink for resources and a source for pollution. Concentrating on water as a key resource and a pollution transport vehicle at the same time, it can be stated that the two main inputs of water into the urban environment are water supply and rainfall. Strategies for minimising fresh water demand for public as well as private use within the city in the form of Water Demand Management (WDM) can result in minimising water resource consumption, thus attaining the first minimisation goal. The same strategies coupled with minimising rainfall-runoff entering the drainage system (for example, using Sustainable Urban Drainage Systems (SUDS) (e.g. D'Arcy & Frost 2001)) could result in less wastewater generation, improving treatment costs and attaining the second minimisation goal. Waste minimisation by itself, however useful, should also be coupled with the effect of these minimisation strategies on wastewater quality: either directly (through the purification capabilities of methods used for minimisation, e.g. grass swales in rainfall-runoff source control, gully pots viewed as reactors (Memon & Butler 2002)) or indirectly (through the effect the minimised and more concentrated wastewater volumes will have in treatment plants from an engineering and a cost effectiveness point of view).

Conceptually this discussion may be relatively clear, but in practice the identification, application and quantification of the effect of urban water management strategies is a complex problem of a highly spatial nature (Larsen & Gujer 1997; Seder *et al.* 2000). *Each location* within the city boundaries has its own properties and its own set of constraints (social, economic and engineering). The decision-maker in urban areas asks for a higher quality of information, including intelligent decision support for time- and investment-critical planning. Taking into account these site-specific characteristics rather than drawing a black box around the cities results in more realistic and therefore more applicable planning. The environmental planner in the urban environment is thus in need of tools that would be able to address the following issues (Seder *et al.* 2000):

- integrate and coordinate information on a domain-oriented scale;
- support analysis, observation, valuation and forecast of environmental systems and their conditions;
- support decisions as a balance between economic and ecological objectives based on expert knowledge;
- use an information system which is natural to the user and which offers transparency without requiring knowledge of some computer language.

The stated requirements imply that the tools (or decision support systems) should, in principle, integrate knowledge and reasoning as an essential part of the system's functionality while dealing with the inherent ambiguities and uncertainties of any reasoning/decision-making process.

In this discussion we will be dealing with the uncertainty implications of a specific type of urban water management problem: the *object location problem*, which is defined as the determination of optimum locations for facilities in a given geographical area with respect to environmental and economic objectives. Solving an object location problem within the urban environment is a complex task, usually semi-structured, which requires multiple objectives as well as expert judgement to be taken into account (Seder *et al.* 2000). This is a demanding decision-making environment, where optimal planning presupposes a synthesis of heterogeneous information of high spatial resolution to ensure site-specific implementation (Makropoulos *et al.* 1999). We argue that fuzzy inference can assist the decision-maker by meeting all the above-stated requirements, taking into account the uncertainty and ambiguity involved in the decision process, inherent in the urban planning environment.

METHODOLOGICAL AND MATHEMATICAL DEVELOPMENT

Why fuzzy inference?

Mathematical paradigms have been slow in realising that one of the most important ways of conveying information, and therefore knowledge, is natural language (Ross 1995).

Despite its inherent ambiguities, people speaking in the same language have little problem in conveying their thoughts and inferring consequents from antecedents. This is actually the only way to teach and therefore communicate experiences, rules and accumulated knowledge. It is also the only way people arrive at decisions in everyday life, from the simpler to the more complex ones. The uncertainty involved in these decisions due to the imprecise nature of linguistic variables or linguistic rules is something *de facto* acceptable in everyday human practice. Fuzzy logic provides a formal mathematical framework for expressing linguistic variables and rules, and in that context it should be clear that it would also inherit the same ambiguity and imprecision, which follows human reasoning and therefore decision-making in everyday life. This fact should not be considered a drawback but rather an intuitively familiar and thus acceptable fact.

Systems that use IF–THEN rules to represent human knowledge within a fuzzy logic framework are called Fuzzy Inference Systems (FIS) (Mendel 2001) and they are extensively used in our decision support system structure. Two fundamental characteristics of environmental planning, account for this fact: *data scarcity* and the need to take decision-making into account, and simultaneously, *complex quantitative and qualitative criteria*. There are two types of FIS: type-1 and type-2. They both use IF–THEN rules to derive consequents from antecedents (both of which can be linguistic variables). Their main difference is that, in the case of type-2 FIS, the fuzzy membership functions (fmf) of antecedent and/or consequent sets are also fuzzy.

In the following paragraphs, we will briefly discuss type-1 and type-2 FIS and present a modular system architecture for using them to reach spatially sensitive decisions.

Fuzzy inference systems

Type-1 fuzzy inference systems

Type-1 FIS (Figure 1) are widely used in fuzzy control engineering and signal processing applications. Such systems map crisp inputs to crisp outputs and are comprised of four parts: a fuzzifier, a defuzzifier, a set of rules (IF–THEN) and an inference procedure. Once the rules are established, the system can be viewed as a mapping from inputs to outputs and this mapping can be generally expressed as $y = f(x)$ (Mendel 2000). A fuzzy set can be represented by its membership function (which, in the case of type-1 FIS, is a crisp function), which assigns a value of the interval [0,1] to each element x of the universe of discourse. This is the procedure undertaken by the fuzzifier.

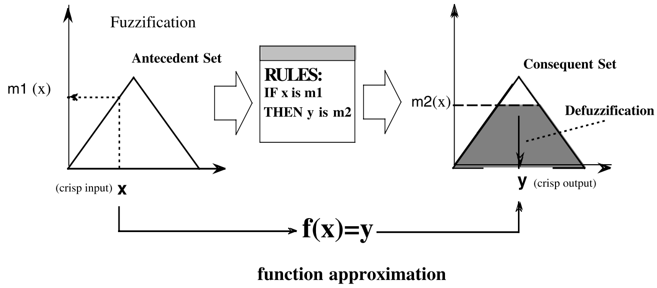

Figure 1 | Schematic of a type-1 Fuzzy Inference System.

The core analytical procedure in any FIS is the *fuzzy inference engine*. A fuzzy inference engine is the way fuzzy logic is applied to combine fuzzy rules into a mapping from input to output fuzzy sets. The general form of the Lth rule in a FIS is: IF x_1 is F_1^L and x_2 is F_2^L and . . . THEN y is G^L, where $F_1^L \ldots F_p^L = A^L$. Therefore the rule can re-expressed as $A^L \to G^L$ or using membership function notation $\mu A^L \to G^L\ (x,y)$. It can be stated that

$$\mu A^L \to G^L\ (x,y) = [T\mu_F(x_i)] \bullet \mu_G(y) \quad (1)$$

where T and • denote a t-norm (i.e. product or minimum). These implications (t-norms) are called *Mamdani* implications.[1] An input of the general form $\mu_A(x)$ of set A will be mapped to the output set B as $\mu_B(y)$, by passing through the rule

$$\mu_B(y) = \sup[\mu_A(x) \bullet \mu A^L \to G^L\ (x,y)] = \sup\{[T\mu_A(x) \bullet \mu_F(x_i)] \bullet \mu_G(y)\} \quad (2)$$

Note that in discreet universes the supremum operation is substituted by a t-conorm (i.e. a maximum operator).

[1]These implications have nothing to do with implication in traditional propositional logic.

Equation (2) is adapted to account for the specific characteristics of the type of FIS used.

Type-1 FIS contain parameters (such as number of rules, number and shape of fmfs, etc.) that can either be pre-defined or can be tuned during a learning process using input–output training pairs, derived from historical records (e.g. using a neurofuzzy procedure).

Type-2 fuzzy inference systems

Both rules used to construct a FIS and membership functions involved in the fuzzification and defuzzification process are usually to a great extent uncertain. According to Mendel (2000), the uncertainty stems from the following reasons:

- There is uncertainty in the antecedents (field data).
- Words included in the rule-based system can mean different things to different people.
- Consequents of the same rule, obtained by polling a group of experts, can be different as the experts will not necessarily agree.

Antecedent and consequent set uncertainties result in uncertain fuzzy membership functions. Type-1 FIS using type-1 fuzzy sets (or ordinary fuzzy sets) cannot capture this uncertainty nor the uncertainty resulting from uncertain rules (caused by uncertain knowledge used to construct these rules). This kind of uncertainty can be better handled by fuzzy sets whose membership function is also a fuzzy set. These fuzzy sets are called *type-2* or *ultra-fuzzy sets* (Graham & Jones 1988). The concept was introduced by Zadeh (1975) as an extension to the concept of ordinary fuzzy sets (type-1). In type-2 sets, each membership value of each element in a set is also a set, in contrast to type-1 fuzzy sets whose membership value is a crisp number. The notion of type-2 fuzzy sets is, to a large extent, analogous to our classic representation of random variables using at least two moments (mean and variance). Type-1 fuzzy membership functions are comparable to calculating only the mean of a probability density function while type-2 take account of the dispersion around the mean (variance). In principle, just as in random uncertainty one can work with higher-order moments, one can work with higher-order fuzzy sets. The complexity, however, will increase rapidly and therefore higher than second-order systems are not used for practical purposes (Mendel 2000). Type-2 fuzzy sets have been studied by a small number of researchers including Dubois & Prade (1978, 1979), Yager (1980), Turksen (1986), Park & Kim (1996) and Mendel (2000, 2001). Figure 2 shows a type-2 fuzzy inference system which uses antecedent and consequent sets whose fmfs values for each point is a crisp set, within the interval [0,1]. Such type-2 sets are called interval type-2 sets and the interval in the secondary membership function can be represented by its left and right end points (upper and lower membership values for point x).

The grey area of the type-2 sets in Figure 2 is a footprint of uncertainty (FOU) for the membership function whose value is uncertain but is known to lie in the interval. The footprint of the uncertainty represents the domain of the secondary membership function. The secondary membership function can be any type-1 fuzzy set (in this example it is an interval set). The structure of a type-2 FIS is very similar to the structure of a type-1 with two important differences. (a) The antecedent and consequent sets are type-2. Note that the triangular shape (seen in Figure 2) is only one of the possible alternatives for fmfs, including Gaussian, bell-shaped, etc. (b) The de-fuzzification procedure is preceded by an analogous operation, termed type-reduction, reducing the result of the inference from a type-2 to a type-1 set. Regardless of the defuzzification procedure used, the type-reduced set is an interval set of the following structure $Y_{\text{Type Reduced}} = [y_{\text{left}}, y_{\text{right}}]$. The main assumption behind type reduction (i.e. finding the centroid of a type-2) is that a type-2 set is comprised of a large number of *embedded type-1 sets*, each associated with a weight which is a function of their secondary membership functions (Mendel 2001).

SDSS architecture

These FISs have been used, in this work, within the context of a spatial decision support system (SDSS) to address a multiple location selection problem in urban water management. The system proposed is based on the loose coupling principle of software integration

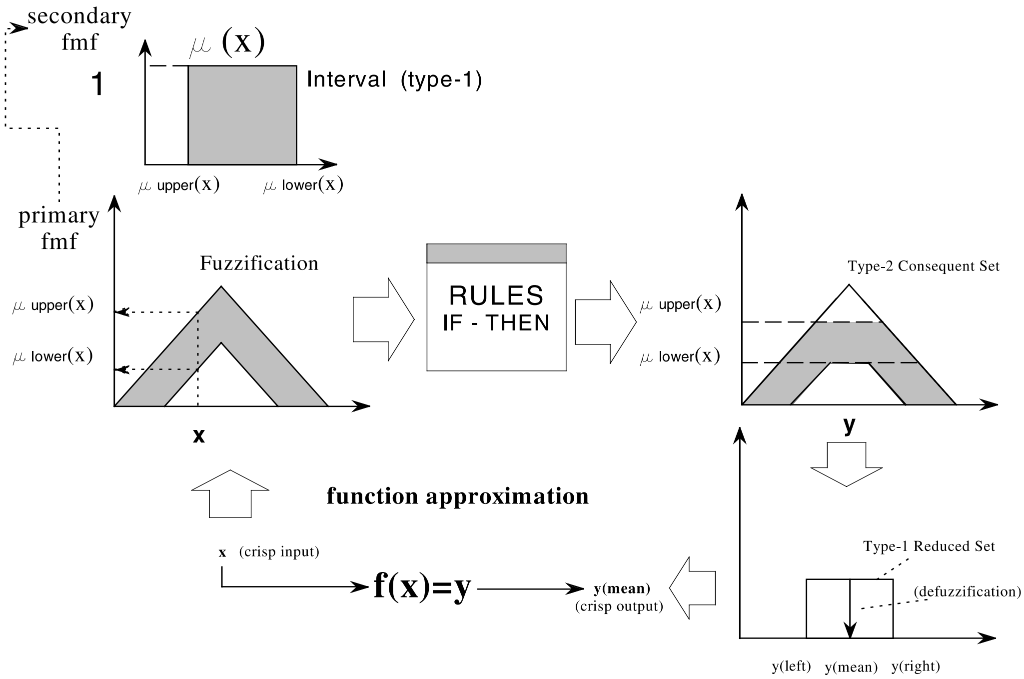

Figure 2 | Schematic of a type-2 Fuzzy Inference System.

(Malczewski 1999), which permits an increased flexibility in components development. A conceptual diagram of the proposed architecture can be seen in Figure 3. This diagram identifies both the structure of the specific system developed as well as the possibilities for modular extensions. In the terminology adopted in the following paragraphs, 'modules' deal with a specific part of the analytical process and are included within an 'application system', which is defined as the analytical engine supporting a specific application in urban water management. One or more 'application systems' are included within the boundaries of the 'overall system', which also includes the core Geographical Information and Database Management Systems as well as a Data Exchange System that allows communication between the different system components.

Each new application system added could interact with the database and the GIS through a general decision-making process (Makropoulos *et al.* 2003) and share information with the rest by exchanging ASCII files through a central analytical engine.

The results of the application system on a specific area of urban water management are presented and discussed in the next paragraphs. The application system supports the reduction of potable water demand by proposing master plans for optimum sitting of water demand, reducing technical measures.

Typically, an 'application system' includes three major modules: fuzzy inference, aggregation and optimisation (see Figure 3). In the first module, the antecedents (characteristic $\boldsymbol{a}$ of a specific location under investigation) are linked with consequents (the suitability $\boldsymbol{s_a}$ for application of a specific strategy to a specific location due to characteristic $\boldsymbol{a}$). Aggregation provides a composite suitability map for all antecedents by using ordered weighted averaging (Yager 1988) and optimisation selects the best combination of technical measures for a given investment scheme based on the suitability maps and the expected impact of the measures in minimising water demand. In this discussion we will concentrate on the issue of using the FISs to capture the uncertainty in locating the water reducing technical measures (leakage reduction through pipe replacement prioritisation, grey water recycling and metering introduction (Foxon *et al.* 2000)) within

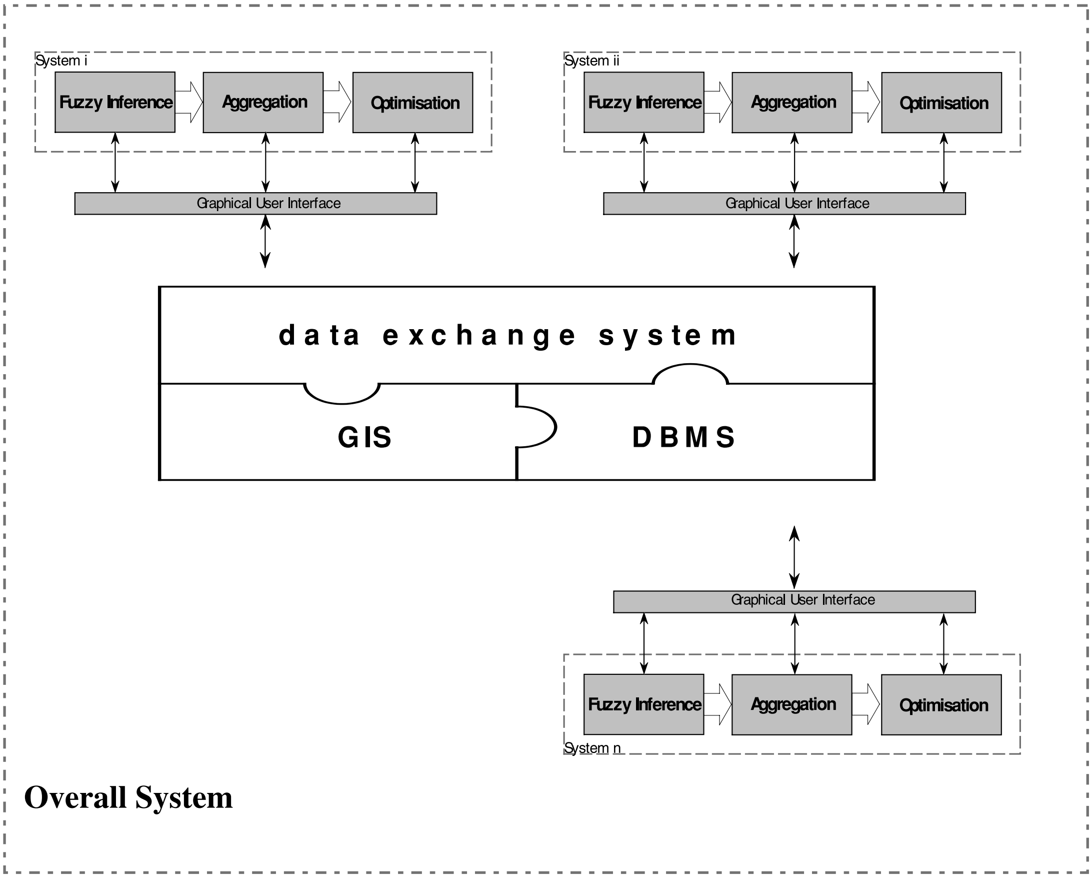

Figure 3 | Modular system architecture.

the city. For a detailed discussion of the decision-making process, as well as on the mathematical development associated with the aggregation and optimisation modules, the reader is referred to Makropoulos *et al.* (2003).

RESULTS PRESENTATION AND DISCUSSION

Presentation of the system and preliminary results on uncertainty quantification

In this study we have utilised fuzzy membership functions of a Gaussian shape. This is purely for computational efficiency purposes because the Gaussian membership function (being closed form) facilitates mathematical computations (integration, differentiation, etc.). The effect of such an assumption is not significant for two reasons:

(a) Due to the complex fuzzification and defuzzification procedures the effect of the actual shape of the input and output fmfs on the crisp result is insignificant (Mendel 2001). The same input–output relationships could be approximated using any of the widely used fmfs (triangular, Gaussian, bell-shaped, etc.) if these functions are correctly tuned.

(b) Data are not sufficient in this case to dictate a specific shape for either input or output fmfs, although in principle, the nature of the problem (the unknown function which the FIS tries to

Table 1 | Attributes of Water Demand Management strategies input to the SDSS

Strategy	Leakage reduction	Metering	Grey water RC
Attributes	1. Soil aggressivity	1. Population density	1. Population density
	2. Age + material	2. Income	2. Income
	3. Traffic Type	3. Education	3. Education
	4. Diameter		4. Renovation status
	5. Distance to fire hydrants		
	6. Max. system pressure		
	7. Pipe density		

approximate) might favour some particular shape.

The attributes taken into account for locating the three technical measures in question can be seen in Table 1. Information on all of these attributes served as input to the FISs and were translated into suitability maps through sets of simple rules such as IF *x Acidic* THEN leakage vulnerability *High*, IF *x Neutral* THEN leakage vulnerability *Low*, IF *x Alkaline* THEN leakage vulnerability *Medium*.

Both input attribute maps and output suitability (or vulnerability[2]) maps were of a raster GIS format. A raster format is, in essence, a map in the form of a matrix, which is (conceptually) created by placing a rectangular mesh on top of the map of interest and using it to extract the information of the map into a matrix with the appropriate number of rows and columns. Each combination of rows and columns of this matrix identifies the location of a *cell*. The information within each cell is considered homogenous and is processed as such. Cells of $30 \times 30\ m^2$ were used in the examples that follow. The term *cell* is used henceforth to describe a specific location within the case study area.

[2]In this paper we use the term 'vulnerability to leakage' interchangeably with the (more general) term 'suitability for application of a leakage reduction strategy'.

Specific facets of uncertainty within the system will now be explored.

Uncertainty in the shape of the fuzzy membership functions

The following examples of assessing the sensitivity of the inference procedure to the parameters of the FIS is from the soil aggressivity attribute (soil pH) of the leakage reduction strategy. A set of simple linguistic rules were modelled as an example:

IF *x* Acidic THEN leakage vulnerability High
IF *x* Neutral THEN leakage vulnerability Low
IF *x* Alkaline THEN leakage vulnerability Medium

The effect of the selection of a fuzzy membership function shape to the output leakage vulnerability map was assessed following the claim that the specific shape is not particularly significant. Figure 4 quantifies the effect of three (commonly used) shapes of antecedent and consequent fmfs: Gaussian, bell-shaped and triangular. Table 2 includes the parameters used to specify the shapes of the fmfs for each FIS.

The results obtained support the relevant literature (e.g. Mendel 2001) claiming that the differences in the

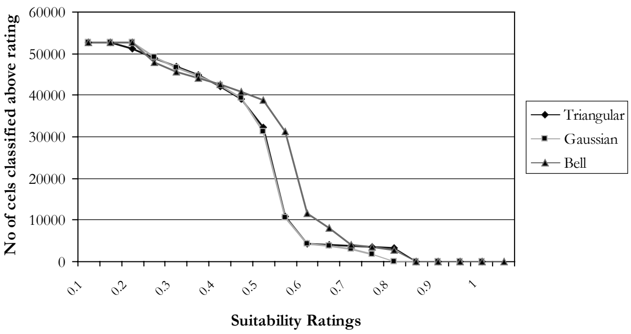

Figure 4 | Differences between FIS using different shapes of fmfs.

analysis caused by the shape of the fmf are not really significant (this seems to be particularly the case between the two most commonly used fmfs: the triangular and the Gaussian). In the absence of specific reasons for the selection of a shape, the designer is therefore to some extent free to select a shape that is more easily handled by the computational tools available. The fact that Gaussian fmfs ensure a smooth transition between membership and non-membership, that they are of closed form and therefore that they are computationally easier to handle and are non-zero for all input values make them the most attractive of the three. As mentioned above, Gaussian fmfs will be used throughout this work. Figure 5 presents the actual outputs of the three FIS as GIS layers of vulnerability to leakage due to soil aggressivity.

Figure 5 is a visual representation of the effect of the linguistic rules and the associated fmfs linking acidic, neutral and alkaline soil pH to pipe vulnerability to leakage. The clear advantage of this visualisation is the ability of the decision-maker to visually quantify the impact of the rules to the result as well as (based on these rules) to visualise the effect of the attribute to the problem in question.

In the example that follows we use the simple form of IF–THEN rules mentioned above in order to investigate the effect of the standard deviation of the fmfs, the standard deviation of the input data (in the case of non-singleton FIS) and the footprint of uncertainty allowed by the type-2 FIS.

Table 2 | Parameters used to define the 3 fmfs of Figures 4 and 5

	Antecedent fmfs	Consequent fmfs
Triangular [*a*,*b*,*c*]:		
$f(x;a,b,c)=\begin{cases}0, & x\leq a\\ \frac{x-a}{b-a} & a\leq x\leq b\\ \frac{c-x}{c-b} & b\leq x\leq c\\ 0, & c\leq x\end{cases}$	Acidic: [− 7, 0, 7] Neutral: [4.5, 7, 9.5] Alkaline [7, 14, 21]	Low: [− 0.47, 0, 0.47] Medium: [0.02, 0.5, 0.98] High: [0.52, 1, 1.47]
Generalised bell [*a*,*b*,*c*]:		
$f(x;a,b,c)=\frac{1}{1+\left\lvert\frac{x-c}{a}\right\rvert^{2b}}$	Acidic: [3.5, 2.5, 0] Neutral: [1.17, 2.5, 7] Alkaline [3.5, 2.5, 14]	Low: [2.36, 2.5, 0] Medium: [2.36, 2.5, 0.5] High: [2.36, 2.5, 1]
Gaussian [σ,*m*]:		
$f(x;\varsigma,m)=e^{-\frac{(x-c)^2}{2\varsigma^2}}$	Acidic: [3, 0] Neutral: [1, 7] Alkaline [3, 14]	Low: [0.2, 0] Medium: [0.2, 0] High: [0.2, 0]

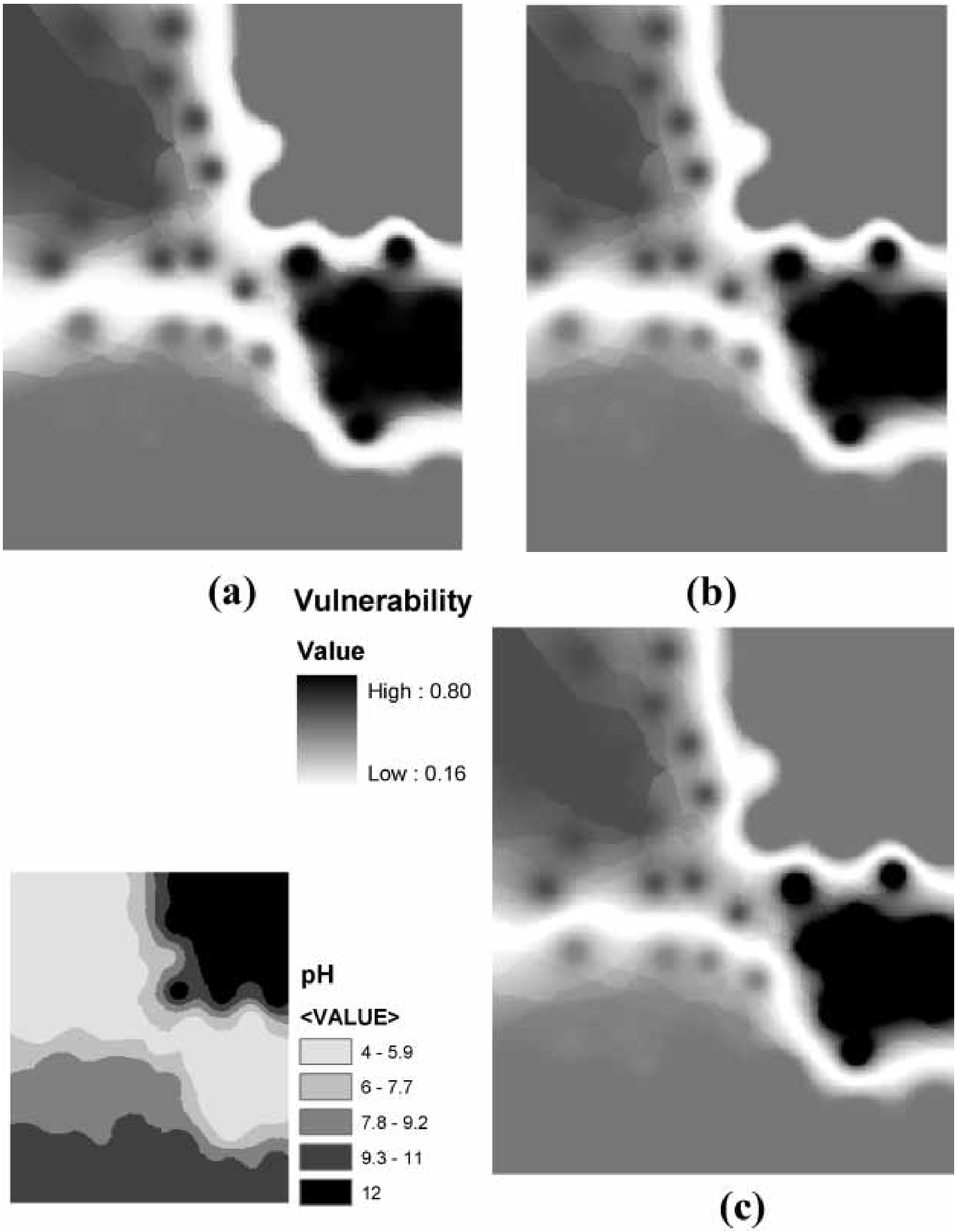

Figure 5 | The soil pH attribute (lower left) and the resulting vulnerability layers from the same FIS using (a) bell-shaped, (b) Gaussian and (c) triangular fmfs.

The three antecedent sets (the linguistic variables: acidic, neutral and alkaline) were modelled with Gaussian fmfs for two sets of parameters (see Table 3) and the resulting suitability ratings can be observed in Figure 6.

The increase in standard deviation, corresponding to an increase in uncertainty as to the relative extents of the sets, has, in this case, one specific effect. It increases the number of input points receiving high scores in all three consequent sets. The reason is that, by increasing the

Table 3 | Parameters for antecedent sets in soil pH

Antecedent sets	Mean	SD (1)	SD (1)	SD (2)
Acidic	2	1	2	3
Neutral	7	1	2	3
Alkaline	10	1	2	3

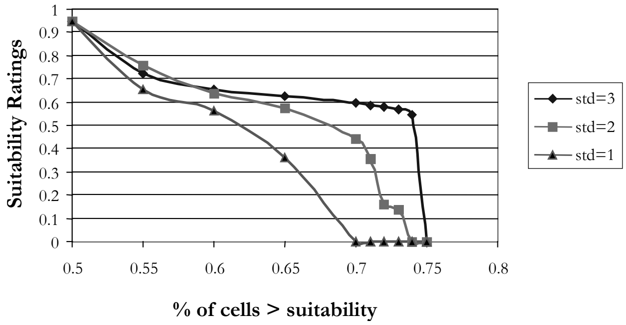

Figure 6 | Cumulative graph of percentage of cells with suitability (vulnerability) ratings over a specified suitability (vulnerability) value.

standard deviation of the antecedent fmfs, there is a tacit acknowledgement of uncertainty as to which set a specific input value belongs. For example, it is not possible to clarify whether an income ***a*** is medium or high. Due to that fact, ***a*** is assigned a high membership in more than one set, and the defuzzification procedure (in this case a centroid defuzzification[3]) results in a more neutral response, with values for all inputs closer to a medium value. This is why the increase is more pronounced in low-to-medium vulnerability ratings while high vulnerability ratings have, in fact, been slightly decreased. The physical meaning of this result is that higher design uncertainty results in more 'indifferent' system-generated suggestions. Figure 6 displays this overall increase in medium-range vulnerability values within the case study area due to the change of the shape of the antecedent Gaussian fmf. Figure 7 displays two layers of leakage vulnerability due to soil aggressivity (pH) corresponding to antecedent fuzzy membership functions' standard deviations of 1 and 2. It also provides an example of the output of the fuzzy inference module of the SDSS. The difference between the two outcomes is generally small (maximum difference≈10%) but its spatial variation provides the decision-maker with a visual quantification of the uncertainty and its effect in site-specific planning.

It should be noted that, due to the parameters used in this example (summarised in Table 3), the resulting layers in Figure 7 are quite different from the ones produced by the inference procedure in Figure 5. This is because the standard deviations of the Gaussian functions used in Figure 7 are not sufficiently large to allow for an adequate overlapping between the fuzzy sets needed to correctly model the inference rules. In this case, pH = 4 is clearly acidic but the set 'acidic' has a mean of 2 and a standard deviation of 1 and 2 for the upper and lower maps of Figure 7, respectively, and therefore pH = 4 is actually receiving a low membership in set 'acidic'. The example serves two purposes: it illustrates the sensitivity of the

[3]In the centroid defuzzification method, the crisp value of the output variable is computed by finding the value of the centre of gravity of the consequent fuzzy set.

Table 4 | Parameters used in the type-2 FIS resulting in Figure 5.19 outputs

Fuzzy membership function	Antecedent sets (population density in inhabitants/km^2)	Consequent sets (suitability to metering introduction)
Low	(mean: [− 20, 20], st.d.: 20)	(mean: [− 0.15, 0.15], st.d.: 0.2)
Medium	(mean: [80, 120], st.d.: 20)	(mean: [0.35, 0.65], st.d.: 0.2)
High	(mean: [180, 220], st.d.: 20)	(mean: [0.85, 1.15], st.d.: 0.2)

system to changes in parameters affecting the shape of the fmfs and stresses the importance of parameters that adequately cover the variables' space and allow for a meaningful (in terms of the rule base) overlapping of both the antecedent and consequent fuzzy sets.

Uncertainty in the data

The proposed approach acknowledges the fact that data input to the FIS are uncertain as well. For example, if pH data are not collected for each and every point in the network but are a result of interpolation from point measurements (as is generally the case), there is considerable uncertainty as to the exact value at any specific point depending, for example, on the interpolation procedure (e.g. spline, kriging, inverse distance weighting) and the distance to a control point (to name but a few possible factors affecting data accuracy). Data uncertainty has been taken into account by treating data not as a single number (a singleton input) but as a fuzzy number, with—in this case—Gaussian membership function, where the mean (and therefore most probable value) is the value of the original data measurement and the standard deviation is specified by the user and applied to all inputs based on a belief in the measurement accuracy (Figure 8). This means in practice that there is a 'most probable' value for the data input (equal to the measurement) but there is also a (lower) possibility of the value being slightly lower or higher.

The analysis quantified differences resulting from non-singleton inputs of varying standard deviations. The results (Figure 9) indicate that, with increasing standard deviation of the input measurements (and thus increasing uncertainty), the numbers of network cells with high vulnerability ratings decrease, while the number of cells with medium vulnerability ratings increase. The values of standard deviation used are 1, 3 and 5. The antecedent set's standard deviation was always equal to 1.

The reason for this effect of increased standard deviation (and thus increased uncertainty) can be observed in the schematic of Figure 10.

The effect of uncertainty in the input data, represented here by increased standard deviation of the input fuzzy number, is that the input value is associated with a progressively larger membership value to the antecedent set (the leftmost set in Figure 10). This is in recognition of the fact that more values are associated with the input value (although with a smaller possibility[4]), and these values are bound to have larger memberships than the 'most probable' one in any antecedent set. The more uncertain a fuzzy number is (i.e. the larger its standard deviation) the larger its membership value is to all antecedent sets. The defuzzification procedure (a centroid defuzzifier) calculates the crisp output as the value associated with the

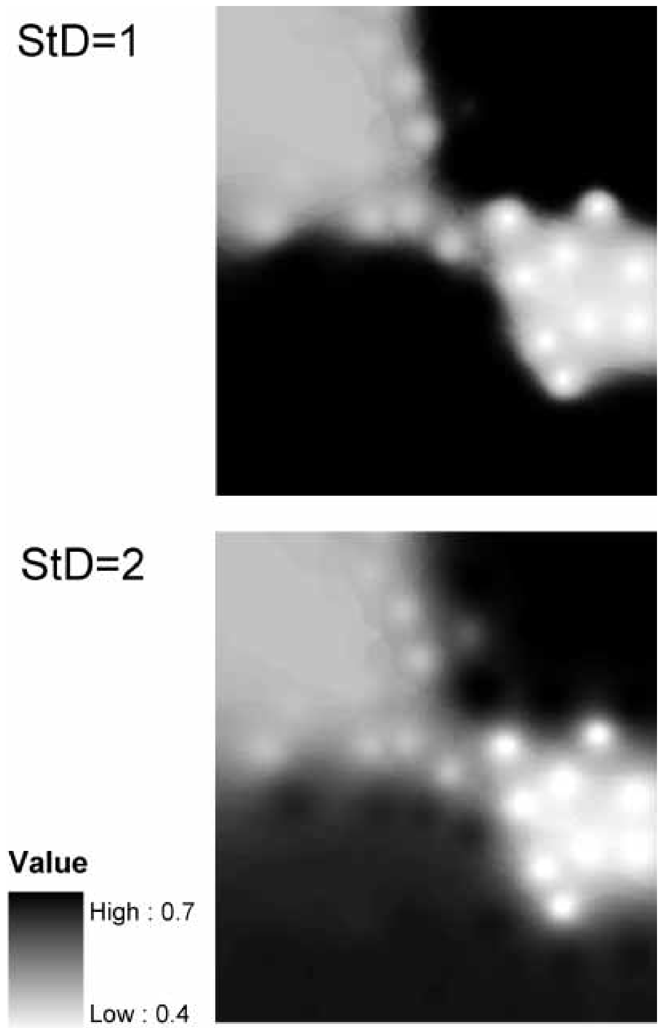

Figure 7 | The soil aggressivity vulnerability layers corresponding to standard deviations (StD) of 1 and 2.

[4]The term *possibility* is used here as the *grade of membership* of an input value to a fuzzy set.

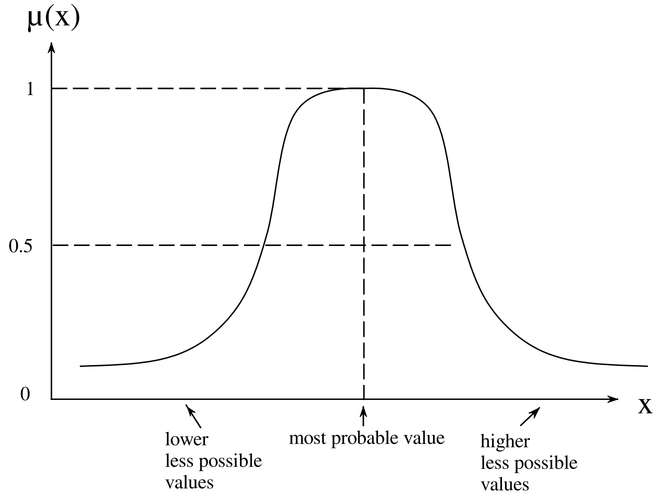

Figure 8 | A fuzzy number quantifying data uncertainty.

centroid of the consequent set resulting from the application of all relevant rules. If the result from the singleton input was a neutral one (i.e. a medium suitability associated with the middle one of three consequent sets) the effect of this increase is not significant. If, however, the result was a small or large suitability, the increase in all sets shifts the new result towards the middle, producing more ‘indifferent’ results, as was to be expected from a physical viewpoint. This increase of medium suitability ratings at the expense of the high suitability ratings explains the shape of Figure 9.

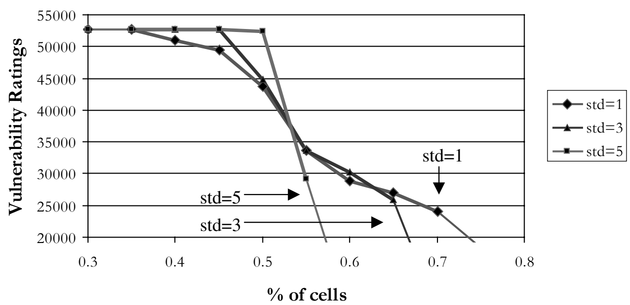

Figure 9 | Cumulative graph displaying the effect of increased data input uncertainty (through changes in standard deviation) to the percentage of cells identified as vulnerable.

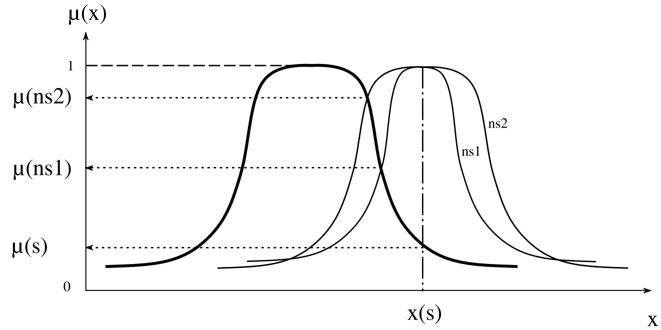

Figure 10 | Fuzzification of two fuzzy numbers (ns1 and ns2) with different standard deviations (SD of ns2>SD of ns1) and one crisp number (*x*(*s*)) which is also the most probable value for the fuzzy numbers.

Uncertainty in the rules

Another interesting example of the SDSS results is the output of the type-2 FISs where the user is able to incorporate his perception of uncertainty in the rules as well. This is programmed by using, as antecedent and consequent set fmfs, Gaussian functions with uncertain means, which are assumed to lie in an interval [m1, m2] and a common standard deviation. The example given below uses the type-2 singleton FIS to assign suitability ratings for metering introduction based on population density. The user specifies the intervals [m1, m2] for both antecedent and consequent fmfs and the standard deviations (one for each of 3 antecedent and 3 consequent fmfs). The output is in the form of 3 maps containing the mean value of the outcome as well as the minimum and maximum values for the same locations. This approach can provide a transparent estimation of the effect of rule uncertainty to the final proposed composite scenario. The analysis (Figure 11) was performed for antecedent fmfs with the parameters shown in Table 4. Figure 11 shows three output suitability layers for the same input [max: (upper map), mean: (centre map), min: (lower map)] of a type-2 FIS, indicating the level of analysis uncertainty associated with rule uncertainty.

The mean (centre map) result coincides with the result obtained by the equivalent type-1 FIS with non-fuzzy antecedent fmfs whose defining parameters (in this case the mean) are given by the mean of the values of the interval presented for each function in Table 4. For example, the ‘medium’ population density fmf of the equivalent type-1 FIS has a mean of 100 and a standard

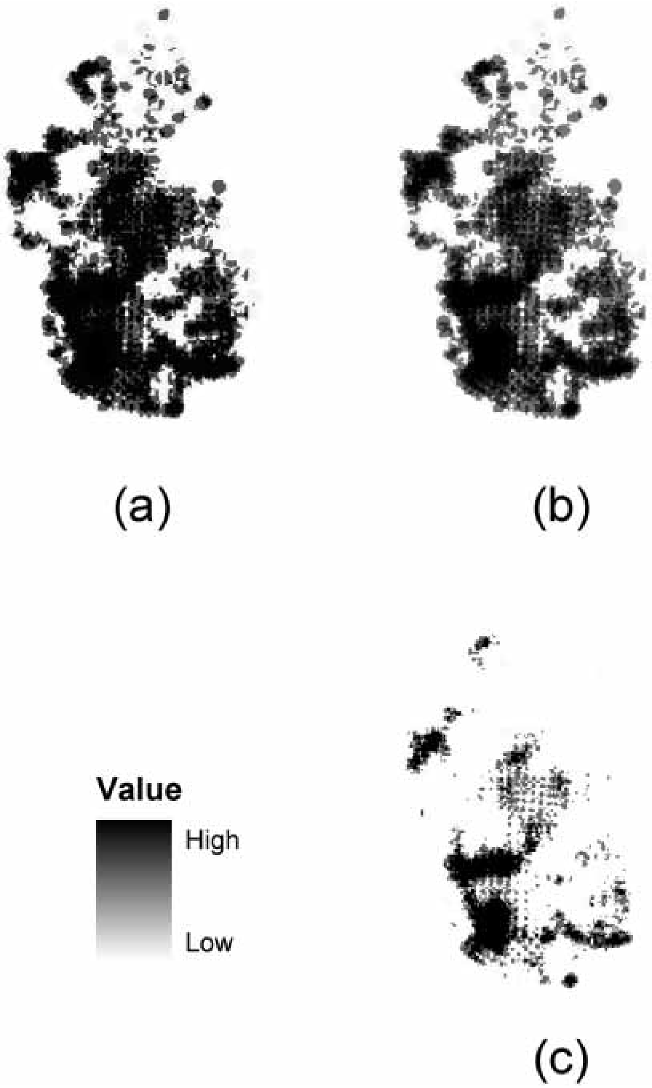

Figure 11 | Three output suitability layers for the same input of a type-2 FIS, indicating the level of analysis uncertainty associated with rule uncertainty [max: (a), mean: (b), min: (c)].

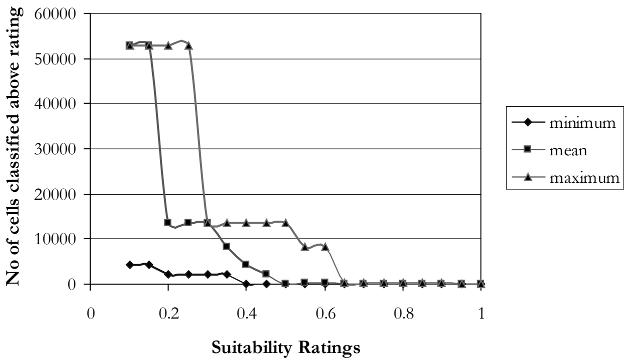

Figure 12 | Cumulative graph comparing the suitability ratings of the three outputs for the case of population density attribute of metering introduction as a WDM strategy.

deviation of 20. The advantage is in the fact that specifying the (rule) uncertainty as an interval in the type-2 FIS allows the result of the analysis to vary accordingly (Figure 11). The visualisation of the uncertainty propagation within the FIS is one of the most interesting characteristics of the use of type-2 FIS within an SDSS.

Figure 12 illustrates the envelope of uncertainty (between the maximum and minimum value) incorporated in the three layers of Figure 11 by identifying the changes in the suitability ratings for households considered for metering introduction. This provides the decision-maker with an understanding of the uncertainty propagation through the decision-making process, thus allowing for more informed decisions. By identifying the elements causing greater variations in the end result, the decision-maker can identify information or knowledge gaps crucial to the analysis and thus better manage the resources for collecting them.

Results discussion

The use of rule based FISs presents a solution to the issue of linking *antecedent* (field data) to *consequent* (suitability for WDM strategy applications), such as linking overhead traffic with leakage vulnerability of the water distribution pipes below, through an unknown function.[5] Using fuzzy inference systems to assign suitability values to spatial data makes use of the function approximation capabilities of these systems and urban water management is a textbook example of a case where the function linking antecedents to consequents is not predetermined. The question of the extent to which a FIS can adequately approximate an unknown function was answered by Wand & Mendel (1992), as well as by Kosko (1992). They

[5]Deterministic relationships for this kind of linking do not exist in the literature. Probabilistic relationships, which can sometimes be found in the literature, are heavily dependent on calibration parameters for specific test data. Most of these relationships, whenever they exist, present few fundamental similarities.

proved (for a number of specific FISs) a universal approximation theorem, in other words that FISs could uniformly approximate any real continuous non-linear function to an arbitrary degree of accuracy. A universal approximation theorem is an existence theorem. It implies that, by using *enough* input and output fuzzy membership functions, *enough* rules, *correctly* tuned, any function can be approximated. It does not, however, explain how to build such a FIS (Mendel 2001). In this case, the design made use of linguistic variables (low, medium, high) with a degree of uncertainty incorporated in type-2 fuzzy membership functions. Linguistic variables and their calculus have been the subject of on-going research ever since Zadeh's 1975 paper on approximate reasoning (Zadeh 1975). In summary it can be stated that although 'words mean different things to different people' (Mendel 2001), there is a general agreement as to their approximate meaning and the degree of uncertainty associated with this difference in meaning can be taken into account. The parameters used in our case to set up the FIS were selected from values proposed in the relevant literature.

This is considered an interesting application of FIS from an environmental point of view and a potentially useful add-in to an SDSS from a decision support point of view. Type-2 non-singleton FIS incorporating data and rule uncertainty in the decision-making process are also considered powerful tools and coupling them with GIS to visualise the result in the form of maps can quantify the spatially variable effect of this uncertainty to urban water management problems of the object location type. Additionally the results indicate that the use of fuzzy numbers instead of crisp data points for input to models in general and decision support systems in particular can effectively address problems of high data uncertainty and ambiguity. The paper *does not advocate the use of type-2 FIS instead of type-1*, but it simply explores their potential in quantifying to a larger extent the uncertainty associated with both data and rules. The premise on which this discussion of type-2 sets was based is that, when there is ambiguity or uncertainty about the exact value of some decisional attribute, we use fuzzy instead of crisp sets. Yet, in type-1 sets, we are asked to determine the fuzzy membership function of that value to some (fuzzy) set exactly, which seems counter-intuitive considering we are not even able to determine its value (Mendel 2001) and thus even less able to determine its membership in a set. In real life, when rules are collected by experts, if we first query the experts about the locations and spreads of the fuzzy sets associated with antecedent and consequent terms, it is very likely that we will get different answers from each expert (Karnik *et al.* 1999). This leads to uncertainty about the locations and spreads of antecedent and consequent fuzzy sets. Such uncertainties can be incorporated into the descriptions of these sets using type-2 membership functions. Naturally, the same problem exists with type-2 sets and the secondary membership function, which is in its turn a crisp function. In principle, to be able to capture uncertainty completely one should work with type-∞ fuzzy sets. This is of course impossible for practical purposes. Higher-order fuzzy sets are more complex and thus a trade-off between complexity and quantified uncertainty has to be reached. This paper argues that type-2 (singleton and non-singleton) FIS are able to incorporate, to some extent, rule uncertainty in the decision-making process while still being applicable in practice for large-scale problems without significant computational burden. The fact that the SDSS used their output in a GIS context to visualise the result in the form of maps assists in the quantification of the spatially variable effect of the uncertainty to urban water management problems of the object location type.

An interesting possibility in designing and using FIS for analytical or decision support purposes is the potential of tuning the FIS parameters (in this case the means and variances of antecedent and consequent sets) by training the original FIS through a neurofuzzy approach. This enables the subjectivity in the design of a FIS to be considerably reduced and partly answers the questions about optimal design put forward above. Such an approach requires input–output training data and changes the parameters of the FIS by a back-propagation procedure, usually a steepest descent algorithm. The training data should take the form of input–output pairs for each inference system (i.e. measurements of input values (i.e. soil pH at point (x,y)) and output leakage vulnerability values of the adjacent water network pipe measured through some user-defined indicator (e.g. corrosion level)). There is, however, a problem associated with this methodology in our particular application, and that is the lack of training data. This lack can be explained by the following two factors. (a) Water company records are scarce (at least in the public domain) and of unspecified reliability. Information on all of the identified attributes for the same network is even scarcer. (b) The suitability rating is not always a measurable quantity. The case of suitability for compulsory metering introduction dependent on the financial level of the occupants is a good example of this problem. This does not mean, however, that the technique is not potentially useful: computerised data of acceptable resolution and accuracy are becoming more and more available and should, in principle, continue to do so in the future and an expert rating could take the place of actual measurements when the rating itself is not a measurable quantity. In this case the FIS would not model reality but the decision-making process of the expert(s). There are, of course, cases where data exist and can be processed to create input–output training pairs, as is the case in pipe prioritisation for leakage reduction. Such a neurofuzzy inference system incorporating additional knowledge from past network records and expert judgement of the system's inference procedure was developed and is currently being tested following the recommendations of Kim & Kassabov (1999), Fenner *et al.* (2000) and Mendel (2001). The authors hope that they will be able to report results shortly in a subsequent publication.

CONCLUSIONS

This paper discussed a mathematical framework for quantifying uncertainty in an SDSS, which can be adapted to a number of urban water management contexts. The use of approximate reasoning (through the use of type-1 and type-2 FIS) is justified by the extent to which linguistic variables have to be used in the planning process when necessary information includes engineering, social and economical constraints. The authors feel that the tools described here are generic enough to allow for similar applications in other fields (Makropoulos & Butler 2001) and that, when there is a level of uncertainty and ambiguity involved in the decision-making process, the use of type-2 fuzzy inference systems coupled with a GIS to capture this uncertainty and present it in a readable form (a map) presents a promising development for environmental planning in general and urban water management in particular.

ACKNOWLEDGEMENT

The authors would like to thank the Hellenic Scholarship Foundation (IKY) for its financial support in the course of this research.

REFERENCES

D'Arcy, B. & Frost, A. 2001 The role of best management practices in alleviating water quality problems associated with diffuse pollution. *The Science of the Total Environment* **265**(1–3), 359–367.

Despic, O. & Simonovic, S. 1997 Methods in evaluating qualitative criteria in water resources multi-criteria decision-making. *Water Resources Research Report* no. 37, Dept. of Civil and Geological Engineering, University of Manitoba.

Dubois, D. & Prade, H. 1978 Operations on fuzzy numbers. *Int. J. Systems Sci.* **9**, 613–626.

Dubois, D. & Prade, H. 1979 Operations in a fuzzy valued logic. *Inform. Control* **43**, 224–240.

Fenner, R., Sweeting, L. & Marriott, M. J. 2000 A new approach for directing proactive sewer maintenance, *Proc. Instn. Civ. Engrs Water & Mar. Engng.* **142**, 67–77.

Foxon, T., Butler, D., Dawes, J., Hutchinson, D., Leach, M., Pearson, P. & Rose, D. 2000 An assessment of water demand management options from a systems approach. *J. Ch. Instn. Wat. & Envir. Mngt.* **14**, 171–178.

Graham, I. & Jones, P. 1988 *Expert Systems: Knowledge, Uncertainty and Decision*. Chapman and Hall, New York.

Karnik, N. N., Mendel, J. M. & Liang, Q. 1999. Type-2 fuzzy logic systems. *IEEE Trans. Fuzzy Syst.* **7**(6), 643–658.

Kim, J. & Kassabov, N. 1999 HyFIS: adaptive neuro-fuzzy inference systems and their application to non-linear dynamic systems. *Neural Networks* **12**, 1301–1319.

Kosko, B. 1992 Fuzzy systems as universal approximators. *IEEE Int. Conf. on Fuzzy Systems*, 1153–1162.

Larsen, T. & Gujer, W. 1997 The concept of sustainable urban water management. *Wat. Sci. Technol.* **35**(9), 3–10.

Makropoulos, C. & Butler, D. 2001 Elements of a decision support system for sustainable urban drainage. *Proc. 1st Nat. Conf. on Sustainable Drainage Systems, Coventry, UK*, pp. 71–80. University of Coventry, Coventry, UK.

Makropoulos, C., Butler, D. & Maksimovic, C. 1999 GIS supported evaluation of source control applicability in urban areas. *Wat. Sci. Technol.* **32**(9), 243–252.

Makropoulos, C., Butler, D. & Maksimovic, C. (2003) A fuzzy logic spatial decision support system for urban water management. *J. Wat. Res. Plann. Mngt.* **129**(1), 69–78.

Malczewski, J. 1999 Spatial multicriteria decision analysis. In *Spatial Multicriteria Decision Making and Analysis: A Geographic Information Sciences Approach* (ed. J. C. Thill), pp. 11–48. Ashgate Publishing, London.

Memon, F. & Butler, D. 2002 Identification and modelling of dry weather processes in gully pots. *Wat. Res.* **36**, 1351–1359.

Mendel, J. 2000 Uncertainty, fuzzy logic and signal processing. *Signal Process.* **80**, 913–933.

Mendel, J. 2001 *Uncertain, Rule-based Fuzzy Logic Systems*. Prentice-Hall, Englewood Cliffs, NJ.

Park, K. & Kim, S. 1996 A note on the fuzzy weighted additive rule. *Fuzzy Sets Syst.* **77**, 315–320.

Ross, T. 1995 *Fuzzy Logic With Engineering Applications*. McGraw-Hill, New York.

Seder, I., Weinkauf, R. & Neumann, T. 2000 Knowledge-based databases and intelligent decision support for environmental management in urban systems. *Comput. Environ. Urban Syst.* **24**, 233–250.

Turksen, I. 1986 Interval valued fuzzy sets based on normal forms. *Fuzzy Sets Syst.* **20**, 191–210.

Vlachos, E. & Braga, B. 2001 The challenge of urban water management. In *Frontiers in Urban Water Management* (ed. C. Maksimovic & J. A. Tejada-Guibert), pp. 1–36. IWA Publishing, London.

Wackernagel, M. & Rees, W. 1996 *Our Ecological Footprint: Reducing Human Impact on Earth*. New Society Publishers, Philadelphia, PA.

Wand, L.-X. & Mendel, J. 1992 Fuzzy basis function, universal approximation and orthogonal least square learning. *IEEE Trans. Neural Networks* **3**, 807–813.

Yager, R. 1980 On measures of fuzziness and negation, part 3: lattices. *Inf. Control* **44**, 236–260.

Yager, R. 1988 On ordered weighted averaging aggregation operators in multicriteria decision-making. *IEEE Trans Syst. Man Cybern.* **18**(1), 183–190.

Zadeh, L. A. 1975 The concept of a linguistic variable and its applications to approximate reasoning. *Inf. Sci.* **8**, 199–249.