Assessment of uncertainties in expert knowledge, illustrated in fuzzy rule-based models

J.A.E.B. Janssena,∗, M.S. Krola, R.M.J. Schielena,b, A.Y. Hoekstraa, J.-L. de Kok a,c

Abstract

The coherence between different aspects in the environmental system leads to a demand for comprehensive models of this system to explore the effects of different management alternatives. Fuzzy logic has been suggested as a means to extend the application domain of environmental modelling from physical relations to expert knowledge. In such applications the expert describes the system in terms of fuzzy variables and inference rules. The result of the fuzzy reasoning process is a numerical output value. In such a model, as in any other, the model context, structure, technical aspects, parameters and inputs may contribute uncertainties to the model output. Analysis of these contributions in a simplified model for agriculture suitability shows how important information about the accuracy of the expert knowledge in relation to the other uncertainties can be provided. A method for the extensive assessment of uncertainties in compositional fuzzy rule-based models is proposed, combining the evaluation of model structure, input and parameter uncertainties. In an example model, each of these three appear to have the potential to dominate aggregated uncertainty, supporting the relevance of an ample uncertainty approach.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

In densely populated delta areas, water management requires balancing of many different interests and user functions. Because of the many different actors, and interaction with the physical environment, governed by many different physical processes, and the need for knowledge from many different areas, the decision making process becomes very complex. To support the decision and policy making process, different tools are utilized. Among these are software tools, where collected data and analytical models serve, for instance, to explore different policy options, analyse real time events, or predict future states of the system configuration. The fact that not all desired information can be described in physical terms may restrict the application of such models. Sometimes experts may be able to provide valuable additional information. In such cases the application of fuzzy rule-based models can be an option (Adriaenssens et al., 2004; Ascough et al., 2008). As in any other environmental modelling approach, it is important to address the uncertainty in the model’s output. This uncertainty assessment is the result of the conceptualization of expert knowledge in a fuzzy rule-based model. We develop a method to assess the different uncertainties which may play a role in this knowledge conceptualization.

A simple hypothetical model is used to illustrate the method.

Uncertainty can be defined as ‘…any departure of the unachievable ideal of complete determinism’ and perceived to be of either an epistemic or a stochastic nature, i.e. either due to a lack of knowledge or due to natural variability in the system (Walker et al., 2003). Lately the notion of ambiguity as a third aspect of uncertainty arose (Brugnach et al., 2007). Ambiguity can be defined as ‘…the simultaneous presence of multiple equally valid frames of knowledge’ (Dewulf et al., 2005). Uncertainty originating from any of these three natures plays a role in river management. This implies an important challenge for modelling for support of strategic river management, namely to adequately address these uncertainties in model outcomes (Clark, 2002; Jakeman and Letcher, 2003; Klauer and Brown, 2004).

Much literature exists describing uncertainty analysis frameworks (for an overview, see e.g. Refsgaard et al., 2007). In general, it is acknowledged that models are simplifications of reality. The process of abstraction of this reality into a software implementation means that elements from reality are omitted, or represented by approximations (see Fig. 1 for a typical example of a representation of the modelling cycle in literature). The process of ongoing abstraction leads to uncertainties in models, additional to those introduced through inputs and parameters. Walker et al. (2003) provide a framework for the description of the uncer-
The objective of this paper is to show how the uncertainty related to using expert knowledge in compositional fuzzy rule-based models can be assessed. We propose and demonstrate a method for the assessment of uncertainties in such models. The outcomes of the uncertainty assessment give an indication of the usefulness of model results, and of the distinctive power of the knowledge in the model.

Linking the uncertainty to the modeler and the fuzzy perspective, we observe that the distinction between epistemic uncertainty (which may include imprecision) and (natural) variability occurs in both. According to Klir and Yuan (1995) fuzzy sets may express two types of uncertainty, namely non-specificity (relating to the size of different alternative sets, and fuzziness (or vagueness, relating to the imprecise boundaries of the fuzzy sets). These interpretations will prove helpful in a later stage of this paper.

2. Methods

For the analysis of uncertainties, the specific characteristic of uncertainties that go along with representing knowledge in fuzzy expert systems are used to interpret classify them using the framework provided by Walker et al. (2003). We apply this to a simple hypothetical model and illustrate the uncertainty propagation through the model.

2.1. Fuzzy expert systems

The impact of different uncertainties on the outcome uncertainty is demonstrated by means of a simple, hypothetical expert system. It is composed of the minimally required components: a knowledge base, an inference engine and a data base. The knowledge base describes the inference rules, derived from experts. The inference engine links these rules to the data from the data base (storing data for each specific task of the expert system), thus resulting in an outcome value.

The rules in the fuzzy knowledge base are generally of the shape 'IF x THEN y', with x and y fuzzy sets. Fuzzy sets are represented by membership functions, describing on the variable domain what the membership – about the distribution – is known, while this is often actually not the case. This may lead to non-conservative uncertainty estimates.

Application of fuzzy set theory is a suitable approach in those cases in which uncertainty is due to incompleteness or imprecision. Its application in environmental modelling became widespread over the past decades, see e.g., Salski (1992), Dorsey and Coover (2003), Adriaenssens et al. (2004), Prato (2005), Van Broekhoven et al. (2006), and Rocchini and Ricotta (2007). Also, uncertainty in fuzzy models is addressed explicitly. Baudrit et al. (2006), for instance, combine stochastic behavior (represented by a probability distribution) and measurement error (without known uncertainty distribution, appropriately described as a fuzzy set). Applications of combined fuzzy and probabilistic uncertainty are found in a.o. Guyonnet et al. (2003), Hall et al. (2007), and Ferraro (2009). Guyonnet et al. (2003) combine Monte Carlo analysis with fuzzy interval analysis and label the result as ‘random fuzzy set’. These applications however focus on the propagation or aggregation of uncertainty in individual fuzzy sets, rather than on application to compositional fuzzy rule base models and the role of uncertainty and its propagation in the different model components. Adriaenssens et al. (2004) touch upon the issue of uncertainty in fuzzy rule-based models, but a comprehensive analysis complying with the perceptions of the environmental modelling community so far fails to materialize. Ascough et al. (2008) specifically state the ample conveyance of uncertainties and its communication as major challenges to research on fuzzy sets.

Fig. 1. The modelling cycle for knowledge production and its steps: delineation of the part of the natural system to be studied, construction of a conceptual model, algorithmic (mathematical) implementation of the conceptual model, implementation of the algorithm in software, calibration of the model parameters, validation of the model results (Kolkman et al., 2005), adapted from Dee (1995).

Fig. 1. The modelling cycle for knowledge production and its steps: delineation of the part of the natural system to be studied, construction of a conceptual model, algorithmic (mathematical) implementation of the conceptual model, implementation of the algorithm in software, calibration of the model parameters, validation of the model results (Kolkman et al., 2005), adapted from Dee (1995).
Fig. 2. Illustration of the fuzzy inference process. The top line depicts the two inputs and a single output; the second and third line show how two input values are partial members of two sets for ‘wet days’ and a single set for ‘dry days’. This leads to two rules to be fired. The implication operator determines the partial membership to the output. These are aggregated into an output surface. The center of area is used to determine the defuzzified output value.

2.2. Uncertainty classification and analysis framework

The basis for the framework for uncertainty analysis is provided in the paper by Walker et al. (2003). It is a very suitable framework because it focuses on uncertainty in model-based decision support. Three different dimensions of uncertainty are distinguished (Walker et al., 2003):

- **Nature**: whether the uncertainty is due to imperfection of our knowledge (epistemic), or due to the inherent variability of the phenomena being described.
- **Level**: where the uncertainty manifests itself along the (continuous) spectrum between deterministic knowledge and total ignorance.
- **Location**: where the uncertainty manifests itself in the components of a model complex: in the context, in the model itself (‘model technical’ or ‘model structure’ uncertainties), in the input, in parameters or in the output.

Some remarks need to be made:

- With regard to the ‘nature’ of uncertainty, ambiguity should also be acknowledged, in accordance with the definition given earlier.
- With regard to the ‘level’ of uncertainty, Walker et al. (2003) use the markers ‘statistical’, ‘scenario’ and ‘recognized ignorance’. We add to that the notion of a qualitative level of uncertainty. This refers to uncertainties which cannot be quantified, but can be described. It is placed between scenario and recognized ignorance.
- With regard to the ‘location’ of uncertainty, the marker ‘output’ denotes the consequence from propagation and aggregation of uncertainties in other locations, rather than a conceptually specific location in itself. Output is therefore discussed as the result of the propagation and aggregation of other uncertainties only.

The resulting analysis framework is depicted in Fig. 3. The location of uncertainty is used as the starting point of the analysis.

2.3. Uncertainty analysis methods

In our analysis of uncertainties, we first describe the impact of separate uncertainties on the model output, and then the aggregated impacts of the combined uncertainties. In this way, the analysis takes the form of a scenario analysis, organized after the location of the uncertainty in the model. The classical approach of sensitivity analysis of models relates to some one of the uncertainties accounted for. The following methods apply to the different uncertainties:

- **Context uncertainty**: The uncertainty in the model context concerns choices made in the step from natural system to conceptual model. Answers to questions such as ‘where do we put the model boundary’ and ‘which input and output variables do we choose’ can be uncertain if there are equally valid alternatives. The uncertainty may be of an epistemic or ambiguous nature. Assumptions or scenario’s are usually used to address these uncertainties.
- **Model structure uncertainty**: can be described as ‘arising from a lack of sufficient understanding of the system that is the subject of the policy analysis, including the behavior of the system and the interrelationships among its elements’ (Walker et al., 2003). It is one of the most difficult uncertainties to address in environmental modelling (Van Asselt and Rotmans, 2002). We here distinguish between two aspects of this uncertainty: the imprecision of knowledge related to the structure of the data on systems’ elements, and the uncertainty in the knowledge on interrelations between elements of the system.

According to the non-specificity as defined by Klir and Yuan (1995), the width of the membership function indicates a lack of knowledge. This is here interpreted as the experts’ inability to connect the different qualitative output states that are distinguished, to precise output values. This interpretation relates to the character of the present application, where fuzzy modelling is applied to represent a modest amount of qualitative information. Following this interpretation, we argue that the size and shape of the output graph, corresponding to a certain combination of input values, reflect an uncertainty in the model structure. The level of this uncertainty is...
'qualitative'. We represent it by the difference \( \delta \) between the centers of area (COA) of the subsets left and right of the original center of area as shown in Fig. 4. This provides a measure of the uncertainty reflected by the width and overlap of membership functions (MFs), following an interpretation that is consistent with using the COA for defuzzification (Janssen et al., 2006). When combined with other uncertainties, the result is comparable to the random fuzzy set (Guyonnet et al., 2003), with this difference that the uncertainty is here directly measured in the fuzzy output graph.

Next, the choice of implication and aggregation operator is considered to contribute to model structure uncertainty. The deviation between outputs obtained with different operators is a measure for this uncertainty, as long as the operators are considered equally valid. The level of this model technical uncertainty is 'scenario'. For the inference procedure there is no equally valid alternative, since Mamdani–Assilian is most suitable for rule-based models based on expert knowledge elicitation (Adriaenssens et al., 2004).

Model technical uncertainty concerns ‘... aspects related to the computer implementation of the model’ (Walker et al., 2003). The model technical uncertainty comprises both software and hardware problems or errors. Analysis of model technical uncertainty would require multiple simultaneous model implementations. This goes beyond the scope of the current study.

Input uncertainty is both uncertainty about ‘... driving external forces that produce changes within the system’ and uncertainty about ‘... the system data that ‘drive’ the model and typically quantify relevant features of the reference system and its behavior’. This uncertainty is considered to be of a stochastic nature, i.e. due to variability in the system (with level marked as ‘stochastic’ in the framework), and was assessed using a Monte Carlo analysis. We run a Monte Carlo analysis on the input, for which we assume a random normal distribution with a standard deviation equaling 20% of the reference value. As an effect size and shape of the output membership functions will vary, and consequently a distribution of COA’s left and right of the original will emerge (see also Janssen et al., 2007).

Parameter uncertainty is uncertainty related to the a priori chosen parameters, described by Walker et al. (2003) as ‘... parameters that may be difficult to identify by calibration and are chosen to be fixed at a certain value that is considered correct. The value of such parameters is associated with uncertainty that must be estimated on the basis of a priori experience’. Parameters determining the shape and size of the membership functions correspond to this location of uncertainty. We acknowledge that if the experts are not so certain about the parameterization of the sets, or if ambiguity exists, a probability distribution of this uncertainty is unlikely to be available. Both are, besides, very likely to occur (Adriaenssens et al., 2004). We therefore assume this uncertainty to be of ‘ambigu-

---

**Table 1**

Parameterization of fuzzy sets for input and output variables. For each set, name, fuzzy set parameters and range of variation for sensitivity analysis are given. Parameter set \([a\ b\ c\ d]\) denotes a trapezoidal fuzzy set with base \([a\ d]\) and top \([b\ c]\).

<table>
<thead>
<tr>
<th># of dry days</th>
<th># of wet days</th>
<th>Agric. suitability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low ([-1\ 0\ 30\ 60]) ±10</td>
<td>Very low ([-1\ 0\ 2.4])</td>
<td>Very bad ([-1\ 00\ 45\ 55]) ±5</td>
</tr>
<tr>
<td>High ([30\ 60\ 70\ 100]) ±5</td>
<td>Low ([2.4\ 6\ 10])</td>
<td>Bad ([45\ 55\ 60\ 70]) ±5</td>
</tr>
<tr>
<td>High ([70\ 100\ 365\ 366]) ±10</td>
<td>Very high ([6.1\ 0\ 15\ 20]) ±2</td>
<td>Average ([60\ 70\ 75\ 85]) ±5</td>
</tr>
<tr>
<td>Very high ([15\ 20\ 65\ 366]) ±2</td>
<td>Very high ([75\ 85\ 90\ 100]) ±5</td>
<td>Good ([90\ 100\ 101\ 110]) ±5</td>
</tr>
<tr>
<td>Very high ([90\ 100\ 101\ 110]) ±5</td>
<td>Very high ([90\ 100\ 101\ 110]) ±5</td>
<td></td>
</tr>
</tbody>
</table>

---

**Table 2**

Input combinations for the 10 cases in the execution of the uncertainty analysis.

<table>
<thead>
<tr>
<th>Case</th>
<th># wet days</th>
<th># dry days</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>25</td>
</tr>
<tr>
<td>6</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>7</td>
<td>17</td>
<td>65</td>
</tr>
<tr>
<td>8</td>
<td>13</td>
<td>75</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>85</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>54</td>
</tr>
</tbody>
</table>

---

**Fig. 4.** Model structure uncertainty; defuzzified value and bandwidth based on COA right minus COA left.

**Fig. 5.** Fuzzy input combinations and resulting output (shaded). The marks indicate the input combinations. They correspond to Table 2 in clockwise direction, starting in the lower left corner.
ous' nature and 'scenario' level, and we run a sensitivity analysis on the parameters. The ranges for the parameters are given in Table 1; in the analysis, resulting parameter combinations were constrained to remain yielding valid membership functions.

Aggregated uncertainty results from all uncertainties above. We assessed it, based on a simultaneous variation of all randomly varied values (parameters and inputs) in two ways. The resulting uncertainty in the range \( \delta \) was evaluated, and compared to the plain model structure uncertainty, next to the resulting uncertainty in the defuzzified COA.

For individual as well as aggregated uncertainties, the above analysis methods yield a quantitative assessment of the consequence for the outcome of the model. Together, it extends common uncertainty analyses and sensitivity analyses, in that it not just involves uncertainties due to uncertain model inputs and parameters.

In the execution of the analysis, using the model in Section 2.3, 10 different input combinations were analysed to illustrate different possible cases (Table 2; Fig. 5). The combinations were chosen to cover a wide range of locations in the different input sets and their overlaps.

Table 3
Fuzzy rule base for the agricultural suitability of a location, depending on the local number of wet and dry days.

<table>
<thead>
<tr>
<th># of wet days</th>
<th># of dry days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Very low</td>
<td>Very good</td>
</tr>
<tr>
<td>Low</td>
<td>Good</td>
</tr>
<tr>
<td>High</td>
<td>Average</td>
</tr>
<tr>
<td>Very high</td>
<td>Bad</td>
</tr>
</tbody>
</table>
2.4. Model description

The conceptual model used in this paper model is essentially a hypothetical simplification of the procedure to assess agriculture suitability in river floodplains, as described by Klijn and De Vries (1997) based on the Dutch HELP procedure (Koerselman, 1987; Werkgroep Cultuurtechnisch Vademecum, 1988). The HELP procedure links soil type and ground water levels to exceed water or water shortage. The decrease in agriculture suitability due to both is next expressed as a percentage of the theoretical maximum yield.

Klijn and De Vries (1997) apply this method specifically to floodplains. They assume:

1. a single soil type in the floodplains;
2. a lowland river;
3. a direct relation between river stage and ground water levels.

Based on these sources we assume that in simple form, the agriculture suitability depends on the number of dry and the number of wet days in this specific area during a long year average year.

When we assume that the rules and sets (Tables 1 and 3) are based on expert opinion, as is likely to be the case in such applications, there is no known distribution of uncertainty around the parameters.

The inputs are supposed to be derived from measured data. The uncertainty in the inputs can, due to the known data distributions, be described in terms of probability distributions.

3. Results

For all uncertainties analysed, the resulting uncertain outputs for the 10 input combinations are depicted as box plots, showing the median, the upper, and the lower quartile in the box. Whiskers indicate the extent of the entire output range.

Model structure uncertainty was assessed for two aspects. Model structure uncertainty from imprecision of knowledge is depicted as the range δ for the model output in Fig. 6a, showing a strong variation between cases. In particular cases 5 and 6 show large uncertainty; this is mainly due to the large non-specificity of the input set ‘Very high’ for the number of wet days. For all cases, the range δ covers over 75% of the uncertainty range. Secondly, the model structure uncertainty from ambiguity in operators choice is depicted in Fig. 6b. The scenario analysis with different operators shows that the uncertainty in the output is small compared to the first model structure uncertainty, but with a similar case-dependency. Again, cases 5 and 6 show the largest uncertainty, Input uncertainty results depicted in Fig. 6c show, the stochastically determined 25–75 percentile boxes and complete ranges of the model outcomes. In the first case random generated inputs may fall outside the variable’s fuzzy range, causing a large number of samples to result in the same output value. Cases 2–6 show a significant uncertainty, of which the 25–75 percentile range covers only a modest share; moreover, the uncertainty range may be asymmetric. In cases 7–10 the outcomes show complete insensitivity to uncertainty in input, indicating that regardless of small variations in the inputs, the values are still mapped to the same output surface, distant from where model output gradients are found. In general, observed uncertainties are partly larger, partly smaller than the model structure uncertainty.

Parameter uncertainty results are shown in Fig. 6d, depicting the sensitivity to the variations in the model parameters. The uncertainties are found to vary with a factor of about 2 between cases. The 25–75 percentile range covers a modest share of this uncertainty. Again, uncertainties are smaller for cases located where model output gradients are low. The magnitudes of the total ranges of this uncertainty per case are very similar to the (first aspect of) model structure uncertainty.

Aggregated uncertainty is given in two ways. The whisker plots for distribution of the upper and lower margins of the range δ for the Monte Carlo simulations is shown in Fig. 6e, illustrating that for some cases, the model structure uncertainty is sensitive to input and parameters. This already could be expected from the comparison of cases is Fig. 6a. The spread in defuzzified COAs in Fig. 6f combines features of all of the earlier figures. The complete uncertainty range for the COAs seem to follow the largest individual uncertainty of the first model structure uncertainty (cases 5 and 6), input uncertainty (case 2) or a combination of model structure, input and parameter uncertainty (cases 1, 3, 4, 7–10). The 25–75 percentile range for the COAs however does not cover the range of uncertainty for cases with large model structure uncertainty.

4. Conclusions and discussions

Description of the uncertainties in model outcomes is considered of paramount importance for the accurate interpretation of these outcomes. This strongly applies to modelled expert knowledge, since it is generally difficult to estimate the uncertainty herein. The method provided in this paper extends the uncertainty framework by Walker et al. (2003) in order to add information on the value of expert knowledge in practical case studies.

Application of this uncertainty framework to a fuzzy rule-based model shows how the uncertainties can be described, where in the model they are located, which considerations to take into account when performing quantitative uncertainty analysis, and how the uncertainties interact with each other. Whereas others have shown that the application of fuzzy sets allows incorporation of non-probabilistic uncertainties, the current application shows how the behavior of fuzzy rule-based models under different uncertainties can be evaluated.

The method, using the δ range to represent the main extent of the fuzzy output, is relatively insensitive to the type of membership function. Also, in the relatively coarse model that was used in this study, outcomes are not very sensitive to the application of different operators. The differences in outcomes between the 10 cases evaluated are largely caused by differences in the non-specificity of the input and output sets relevant to each case, and to the position of the cases in relation to the fuzziness of sets. Non-specificity in particular is a strong factor in model structure uncertainty. Input and parameter uncertainty depend strongly on the inputs for a particular location in relation to the parameters of the model, indicating where large gradients in the output are found.

The findings presented here stress the relevance of an extensive uncertainty analysis on fuzzy rule-based models in general, including model structure uncertainty, and makes the challenges for fuzzy set methodology in Ascough et al. (2008) more concrete. The challenge is of particular relevance because in this type of applications, people may find it difficult to interpret a single defuzzified output value in the light of the underlying sets and rule bases. Aspects of the fuzzy characteristics of output like the δ range may be useful to communicate a measure for the uncertainty covered in the fuzzy output set, next to the degree of membership to the fuzzy output set (Van Broekhoven et al., 2006).

In applications where fuzzy set modelling intends to (after defuzzification) make a simplified representation of an extensive knowledge base (for instance embedded in a model), parameters may be chosen in the modelling process to try and minimize errors in approximation of outputs and output gradients for relevant inputs. In such applications, uncertainty would relate to the uncertainties in the underlying model, combined with approximation
errors introduced by the defuzzified fuzzy model representation, rather than by, e.g. the range \( \delta \) for fuzzy output sets. In such applications, construction of the approximating fuzzy model involves, e.g. model structure elements and parameters to be jointly varied to minimize approximation errors. In that way, uncertainty locations do become interdependent in ways that are not encountered in the present case.

Challenges of fuzzy logic methods in such cases also include fuzzy technical aspects, where fast and accurate defuzzification (Van Broekhoven and De Baets, 2006) may support to make the proposed extensive uncertainty assessment feasible in large fuzzy models.

Larger non-specificity and fuzziness in outcome sets result in larger knowledge uncertainty. The relative contribution of different uncertainties to the total outcome uncertainty may provide a useful lead for uncertainty reduction.

Acknowledgements

This work has been funded by the Cornelis Lely Stichting, project nr. 05.006. Special thanks to Prof. B. De Baets and H. Vernieuwe, PhD, for their valuable advice on current approaches to fuzzy uncertainty.

References


Van Broekhoven, E., De Baets, B., 2006. Fast and accurate center of gravity defuzzification of fuzzy technical aspects, where fast and accurate defuzzification may support to make the proposed extensive uncertainty assessment feasible in large fuzzy models.

This work has been funded by the Cornelis Lely Stichting, project nr. 05.006. Special thanks to Prof. B. De Baets and H. Vernieuwe, PhD, for their valuable advice on current approaches to fuzzy uncertainty.