

An interactive modelling tool to support knowledge elicitation using extreme case models

J.H.A Guillaume^a and B. Fu^a

^a National Centre for Groundwater Research & Training and Fenner School of Environment and Society,
Australian National University, Canberra, Australia
Email: joseph.guillaume@anu.edu.au

Abstract: Knowledge elicitation can be a crucial aspect of modelling. When few data are available, it enables predictions to be made on the basis of expert knowledge. It also provides the opportunity for stakeholders to express their understanding of a system to help assess a model and help ensure that their point of view is accounted for. In this paper, we describe an interactive modelling tool to help express and evaluate stakeholders' knowledge about water requirements of floodplain and wetland vegetation (Figure 1). It aims to maximise the breadth of views to which the user is exposed, and minimise mandatory user input. This helps prompt the user to reflect on their knowledge and empowers them to decide what they feel confident in claiming. This is achieved by automatically generating extreme case models (with different parameter values) for the user to evaluate even before they have given any input. Visualisations of these results prompt the user to provide information that constrains the models. These constraints take the form of key concepts of knowledge about suitability, namely the bounds (e.g. ideally, river red gums require 3-8 months of flooding) and relationship between any two points (e.g. 2 months flooding is better than 1 month flooding). This tool helps to capture uncertainty in elicited knowledge by identifying constraints rather than single models and expecting knowledge to be changeable and evolving. This contrasts with approaches that develop multiple consensus solutions, within which dissenting and novel understandings might be suppressed, and approaches that elicit uncertainty as measurable probabilities or possibilities which are themselves uncertain. Although we use a habitat suitability model as an example, this method is generic and can be used in many other applications eliciting relationships among variables.

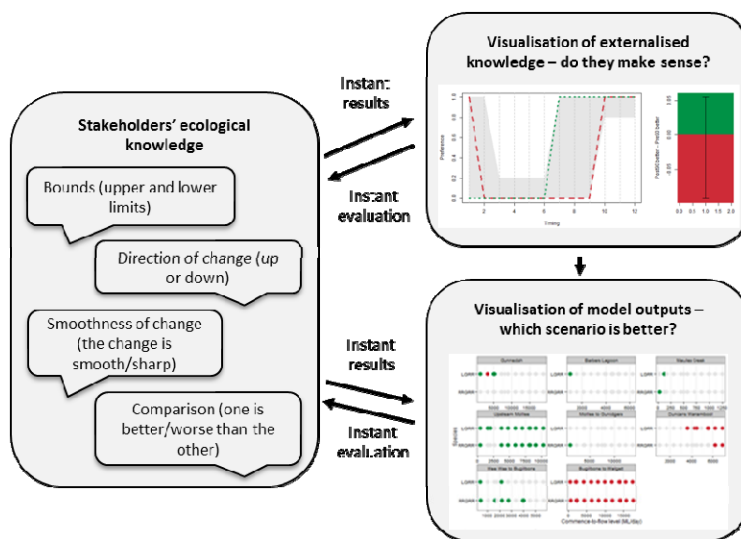


Figure 1: Conceptual framework of the interactive knowledge elicitation tool.

Keywords: Knowledge elicitation, ecological model, habitat suitability, uncertainty

1. INTRODUCTION

Participation of stakeholders is becoming one of the key elements in contemporary natural resource management. Empowerment of stakeholders has several advantages in environmental modelling. For instance, the knowledge and perspectives needed for the management of the social-ecological systems can be effectively captured in the models; it allows greater acceptance of model outcomes through participation and thus easier implementation of policies (Krueger et al., 2012). Even if a model can be built from costly large datasets, knowledge of stakeholders can provide a valuable contribution by assessing the model and helping to fill in knowledge gaps.

Knowledge acquisition is the process of elicitation of knowledge, and the explication and formalisation of that knowledge into a computation model (Cooke, 1994). Knowledge elicitation is a crucial part of knowledge acquisition; however it can be a difficult task. In particular, there may be uncertainty in the user's knowledge such that multiple models (with different parameters or model structures) would be acceptable to them. However, these alternatives may not be captured by techniques that focus on identifying single models, aim to develop consensus or elicit uncertainty as measurable probabilities or possibilities. These methods have in common that they define certain mandatory user inputs. By defining what must be provided, the users' attention is focussed on getting those inputs right, rather than on exploring alternative views that they would accept. This is reinforced by a tendency to see knowledge elicitation as a passive extraction of knowledge to be able to create a model rather than an opportunity for stakeholders to actively reflect on their knowledge. Feedback from model results to stakeholders is therefore often lacking, such that stakeholders may disagree with knowledge and model outputs as captured, hindering the adoption and compliance with decisions based on the model results.

In this paper, we introduce an interactive tool to assist the knowledge elicitation process. We used an ecological model as a case study to demonstrate how the tool can help express and evaluate people's knowledge about water requirements of floodplain and wetland vegetation. This knowledge was used in the ecological model to compare the suitability of different water regimes for the regeneration of vegetation in the Namoi catchment, Australia. Unlike traditional elicitation tools, the focus of this tool is interaction. The model results are instantly produced which provide feedback directly to the stakeholders who can then instantly evaluate and review the existing knowledge.

2. CASE STUDY AREA

The Namoi River Catchment forms part of the Murray Darling Basin and drains an area of approximately 42,000 km² in northern New South Wales, Australia (Figure 2). Rainfall decreases from east to west, with annual average rainfall ranging from 945 mm at Niangala to 480 mm at Walgett in the low lying plains of the west. Flow seasonality and flood frequency have been severely altered by river regulation since 1960.

Much of the Namoi Catchment has been cleared, except for habitat corridors and patches of riverine vegetation (Eco Logical, 2008, 2009). The major streams and rivers are dominated by river oak (*Casuarina cunninghamiana*) and river red gum (*Eucalyptus camaldulensis*). The remnant floodplain vegetation is in better condition than riparian vegetation, and lowland riparian vegetation is in better condition than upland riparian vegetation. Most of the wetlands in Namoi are large in number (1829 natural and 937 artificial wetlands) but small in size and scattered across the floodplain and major tributaries (Eco Logical, 2008).

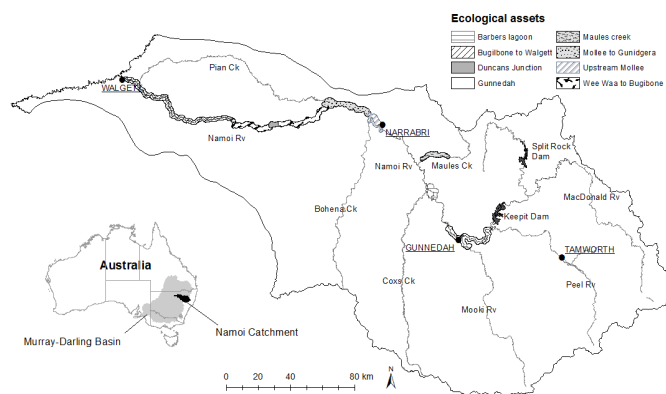


Figure 2. Namoi River catchment, showing asset sections along the river.

3. METHODS

3.1. Habitat suitability model for vegetation regeneration

The ecological model assesses the suitability of stream flow regimes for the regeneration of riparian vegetation in the Namoi Catchment. It identifies characteristics of flooding events (i.e. duration and timing) and generates suitability indices based on species' water requirements. In this paper, we chose two vegetation species: *Eucalyptus camaldulensis* (river red gum) and *Muehlenbeckia florulenta* (lignum). Eight ecological assets were selected (Figure 2). These assets are areas in the catchment that have ecological significance.

The stream flow suitability index for vegetation regeneration was estimated based on weighted average of suitability of flood duration and flood timing:

$$S = w_d D + w_t T$$

where S , D , T are the stream flow suitability index, flood duration index and flood timing index respectively; and w_d and w_t are weights for duration and timing respectively.

The suitability of flood attributes were estimated using preference curves which convert each flood attribute (e.g. flood duration) into a suitability index. An index value ranges from 0 to 1, with zero indicating an unsuitable groundwater regime for regeneration of riparian vegetation and 1 being most suitable. The preference curves were generated from knowledge elicitation which is described in the following section. The flood attributes were generated from daily stream flow time series based on commence-to-flood (CTF) levels, above which a flood or wetting event occurs.

In this paper, we compare mean stream flow suitability indices over two periods: 'Scenario 1' over the period 1975 – 1989 and 'Scenario 2' over the period 1990 – 2005. The model input is historical daily river flow data. The data before 2008 were obtained from PINNEENA 9.2 (Department of Water and Energy, 2008); while more recent flow records (2008-2010) were obtained from the waterinfo website (waterinfo.nsw.gov.au). River gauges were selected based on their proximity to the assets and the completeness of the data within the testing period.

3.2. Knowledge elicitation

The proposed technique contributes to knowledge elicitation through two key points: maximising the breadth of views to which the user is exposed, and minimising mandatory user input. The first is motivated by the idea that exposing the user to alternative views can help prompt new user responses and reflection on their knowledge. Making user input optional rather than mandatory empowers the user to decide what they feel confident in claiming. It avoids pressing the user into making claims with which they are uncomfortable or which the user will later retract as being inaccurate.

In the context of habitat suitability modelling, this is achieved by automatically generating extreme case models for the user to evaluate even before they have given any input. Optimisation is used to identify an extreme suitability preference curve that favours 'Scenario 1' and an extreme suitability preference curve that favours 'Scenario 2', while respecting any information the user has provided. Especially if the user has not provided any input, the curves are likely to be judged to be unrealistic, which prompts the user to provide additional information. Model outputs are also instantaneously generated. These identify which scenario being compared has higher suitability. By comparing the two automatically-generated curves favouring 'Scenario 1' and 'Scenario 2' respectively, it can also identify whether a model output is currently uncertain, which can help guide the user in defining what knowledge still needs to be elicited or further improved. Figure 1 illustrates this basic idea of interaction between expressed knowledge and instantaneously generated visualisation of the knowledge and its outcomes.

User input is elicited in the form of key concepts of knowledge about habitat suitability, namely bounds (e.g. ideally, river red gums require 3-8 months of flooding) and relationship between any two points (e.g. 2 months flooding is better than 1 month flooding). These concepts are translated into mathematical constraints on the preference curve. Rather than eliciting constraints directly in numerical form, it may be preferable for the user to transition through multiple formats (e.g. diagrammatic, verbal, textual, mathematical) to reduce potential information loss (Ford and Sterman, 1998). Examples of the constraints used for modelling suitability of water regime to the regeneration of river red gum are shown in Table 1. These constraints can be explained in textual form. For example, we know that flooding is necessary for river red gum regeneration (0 day flood duration is unsuitable), and prolonged floods (≥ 180 days) start to have detrimental effects.

Table 1: Examples of constraints in flood duration, flood timing and weights derived from knowledge, for modelling suitability of water regime to the regeneration of river red gum.

Attribute	Type	Knowledge in suitability		Constraints on preference value	Possible range for preference curve (grey area)
Flood duration	Bounds	≥ 180 days	Unsuitable	= 0	
		0 days	Unsuitable	= 0	
		30-40 days	Quite suitable	≥ 0.8	
	Direction	0-30 days	Must go up/steady	Direction=1	
		≥ 40 days	Must go down/steady	Direction= -1	
Flood timing	Bounds	Mar - Jun	Not so suitable	≤ 0.2	
		Oct - Dec	Quite suitable	≥ 0.8	
	Direction	Jan - Jun	Must go down/steady	Direction= -1	
		Jul - Dec	Must go up/steady	Direction=1	
Weight	Bounds	Duration and timing are both important		$w_d \geq 0.3$; $w_t \geq 0.3$	NA
	Comparison	Duration is more important than timing		$w_d > w_t$ by at least 0.1	

4. INTERACTIVE WEB APPLICATION

The tool was developed with a prototype interactive web application using the **shiny** package in R (RStudio and Inc., 2013). The web application was designed with an intuitive and user-friendly graphical interface that allows users to easily navigate between the key steps, shifting their focus during the knowledge elicitation process. This is essential as understanding of ecological suitability of a hydrological regime depends on eliciting information about a significant number of species and attributes. The key features of the interface are:

- **Focus on elicitation of one relationship at a time.** The user selects the species and attribute for which they wish to provide information, without making any inputs mandatory. This provides structure to avoid having to address all problems at once and encouraging the user to think about one relationship in depth.
- **Provide visualisations that capture an elicitation-specific result.** The interaction between elicitation and results is kept tight, by avoiding confounding factors in the results wherever possible (Serman, 1994). For example, elicitation of constraints for flood duration of river red gum shows which scenario is better *considering only that attribute*. If the result is uncertain or unexpected, it therefore helps prompt the user to think of additional constraints that might be relevant.
- **Separate diagnosis of necessary knowledge from its elicitation.** Aggregate results are displayed, covering multiple ecological assets, such that 1) they can be quickly scanned to identify automatically generated preference curves that need to be modified and 2) to distinguish cases where existing knowledge is sufficient from cases where additional knowledge elicitation or research is required. As different extreme case models are generated for each asset, the user is prompted with a broad range of very different views.

These features are described in more detail below using the case study examples.

Step 1: Without any knowledge on water requirement of the species elicited, two preference curves are automatically generated that favour each of the two scenarios (Figure 3). The left chart shows the bounds (grey area) which in this case cover the whole chart (i.e. any area is possible), and the two preference curves (green and red dashed lines) which favour each of the two scenarios. The right chart shows the probable range of results on which each scenario is better. The green zone indicates that ‘Scenario 1’ yields a better result in habitat suitability; the red zone indicates where ‘Scenario 2’ is better. In this case because there is no

knowledge elicited and no constraints specified, either scenario can be better (the bar is distributed across both green and red zones).

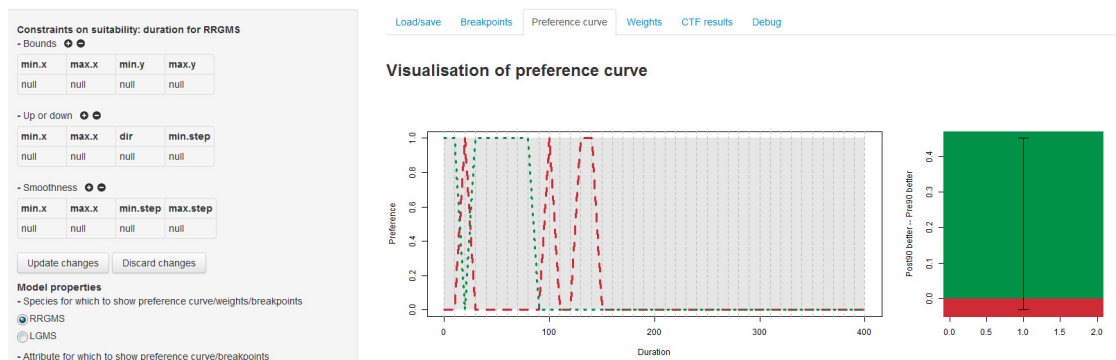


Figure 3: Interactive web interface showing result when no knowledge is elicited.

Step 2: Though it is not mandatory, the user often has knowledge available they wish to express, here with regards to flood duration/timing preference for the regeneration of river red gum. This knowledge is elicited as constraints (e.g. bounds, up/down). The constraints are entered to the left hand side of the interface. The visualisation of the constraints and result (in terms of which scenario is better) for this particular species and attributes is instantly shown on the right hand side of the interface. For example, Figure 4 shows that given current knowledge elicited (i.e. constraints specified) on flood duration and the hydrological inputs for the two scenarios, ‘Scenario 1’ is always better than ‘Scenario 2’ (all green). In contrast, given what we know about preference in flood timing (Figure 5), we are uncertain about which scenario is better (the bar is distributed in red and green zones). These charts help users instantly evaluate constraints and identify if the extreme views captured by the automatically-generated preference curves make sense.

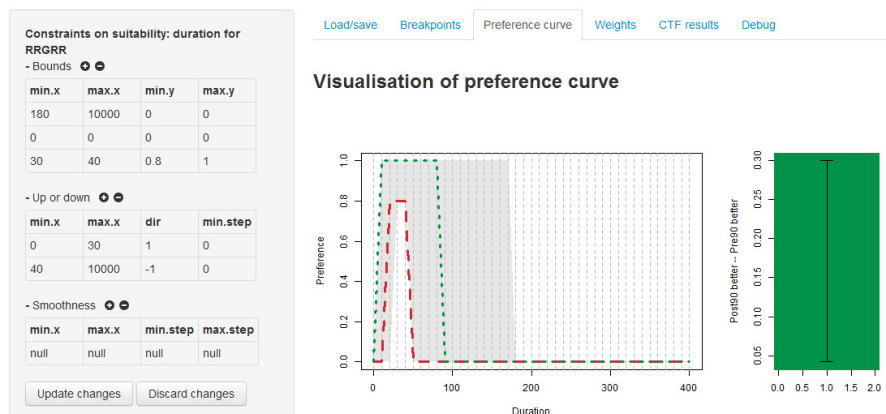


Figure 4: Entering and visualisation on the constraints and result of flood timing for regeneration of river red gum.

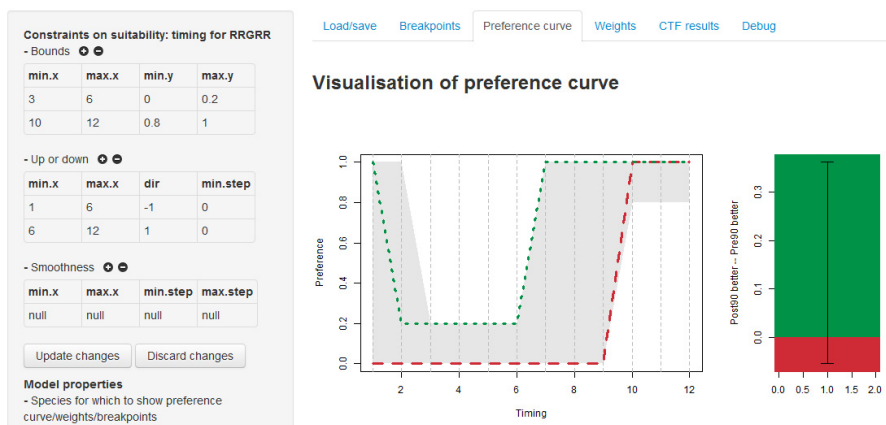


Figure 5: Entering and visualisation on the constraints and result of flood timing for regeneration of river red gum.

Step 3: Similarly, knowledge in the relative importance of the different flood attributes can be elicited. The weights for the extreme cases (which favour each of the two scenarios) are automatically computed and the results provided on which scenario is better based on this information (Figure 6).

Step 4: Examine the summarised results on which scenario is better for all species and locations. For example, the “traffic light diagrams” (Figure 7) shows in green where ‘Scenario 1’ is better or the same (i.e. ‘Scenario 1’ is favourable); in light grey, the result is uncertain (i.e. cannot pick which scenario is more favourable), or in red, ‘Scenario 2’ is better or the same (i.e. ‘Scenario 2’ is favourable). In this case, we found that the results for Upstream Mollee (all green) and Bugilbone to Walgett (all red) are always certain. Thus, little additional knowledge (i.e. having tighter constraints) is needed. However, much higher uncertainties in model outputs are found for other assets, suggesting that the current state of knowledge is insufficient in these cases to conclude which scenario is better. These are the cases on which knowledge elicitation or further research could be focussed, though it is not mandatory for the user to do so. Another example is the overview of all preference curves generated with different species and attributes, allowing users to quickly identify unexpected curves and review constraints (not shown here).

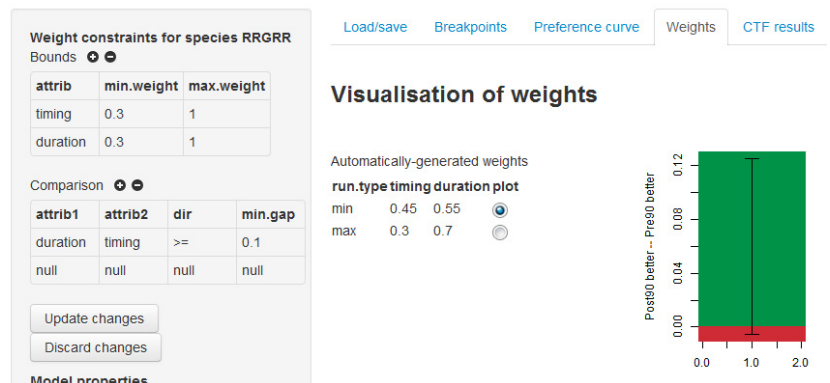


Figure 6: Entering and visualisation of weight constraints.

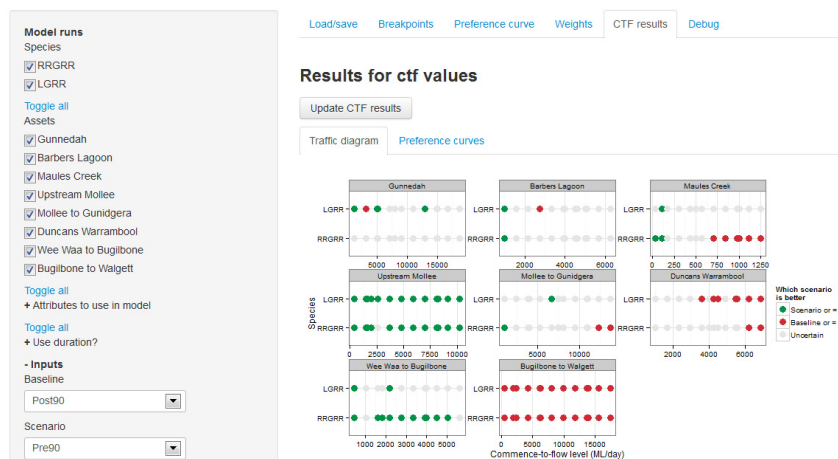


Figure 7: Summarised results for all species and ecological assets using traffic light diagrams.

5. DISCUSSION

The tool presented in this paper differs from existing approaches to elicitation of ecological knowledge in two ways.

- 1) It differs from dominant preference curve approaches that rely on the user drawing the curve by being able to capture and represent the multiple ways of joining two points. Other approaches that elicit multiple curves directly tend to maintain a focus on curves, which favours the generation of (possibly multiple) consensus solutions within which dissenting and novel understandings might be suppressed.
- 2) It differs from typical applications of Bayesian networks or fuzzy logic which focus on uncertainty as measurable probability or possibility, whereas we focus on treating uncertainty by working with constraints and expecting the elicited knowledge to be changeable and evolving.

In domains where i) knowledge is still immature, ii) user views are not well developed or iii) the relationship to be elicited is extremely sensitive to case-specific circumstances, it is important to expect the specific relationship elicited from a user to be changeable and evolving. A solution is to presume values uncertain until shown otherwise and to explore a greater diversity of curves. Eliciting constraints helps to facilitate this approach. Constraints may also be more stable than single curves. They are less likely to change because the

user may be more confident in specifying what cannot happen rather than what will happen. Providing visualisations that trigger changes might also help the user elicited constraints to be identified.

If the knowledge elicited is assumed to be evolving, then an effective knowledge elicitation method should help support the elaboration of knowledge, rather than just extracting it. Methods that help review, evolve and crystallise thinking can improve the user's understanding of their knowledge which can empower them to use and defend it strategically in addition to improving the quality of relationships elicited. This interactive tool-based approach has the advantage of supporting individual reflection even on its own. Knowledge elaboration can also be supported by group-based elicitation. The tool provides the advantage of generating mathematically feasible curves determined by what users have made explicit. It can therefore generate curves that users would have subconsciously filtered out as obviously wrong, and is not limited by the ideas present in the room. The need to express explicit constraints to change the curves generated by the tool also helps focus discussion on *why* the curves need to be changed, rather than *who* has the right or wrong curve.

The work in this paper results from preliminary development. In the future, the tool presented still needs to be integrated into a full knowledge elicitation method, including multiple participants, and still needs to be tested. In particular, testing will need to ensure that the functionality provided is sufficient for the user to be able to express their knowledge. Although we use a habitat suitability model as an example, the principles described could be applied in other applications eliciting relationships between variables.

6. CONCLUSION

This paper introduces an interactive tool to assist the process of eliciting ecological knowledge describing the relationship between variables, in this case the suitability of different hydrological variables for the regeneration of vegetation. The tool's key contributions are to maximise the breadth of views to which the user is exposed and to minimise mandatory input. This is achieved by automatically generating extreme-case models with different parameters for the user to evaluate even before they have given any input, and eliciting user's knowledge in the form of constraints on the relationship. The relationship is used to compare which of two scenarios is better, and this result is also instantly displayed for the user. The emphasis is on empowering the user by providing them direct feedback on possible relationships allowed by the constraints they have provided and therefore enabling them to instantly evaluate and review the effect of the knowledge they have expressed. A prototype interactive web application is demonstrated, which helps manage the user's focus by concentrating on elicitation of one relationship at a time, providing elicitation-specific visualisations of results and separating elicitation from diagnosis of necessary knowledge.

ACKNOWLEDGEMENT

This work was funded by the Australian National Centre for Groundwater Research and Training (NCGRT), and the Cotton Catchment Communities Cooperative Research Centre.

REFERENCES

- Cooke, N.J., 1994. Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies* 41, 801-849.
- Department of Water and Energy, 2008. Continuous Monitoring PINNEENA 9.2 DVD. NSW Department of Water and Energy, Sydney.
- Eco Logical, 2008. Namoi Wetland Assessment and Prioritisation: Draft report prepared for Namoi Catchment Management Authority.
- Eco Logical, 2009. Riverine Vegetation in the Namoi Catchment: an Assessment of Type and Condition: Final report prepared for Cotton Catchment communities CRC and Namoi Catchment Management Authority.
- Ford, D.N., Sterman, J.D., 1998. Expert knowledge elicitation to improve formal and mental models. *System Dynamics Review* 14, 309-340.
- Krueger, T., Page, T., Hubacek, K., Smith, L., Hiscock, K., 2012. The role of expert opinion in environmental modelling. *Environmental Modelling & Software* 36, 4-18.
- RStudio and Inc., 2013. shiny: Web Application Framework for R. R package version 0.6.0.
- Sterman, J.D., 1994. Learning in and about complex systems. *System Dynamics Review* 10, 291-330.