International Environmental Modelling and Software Society (iEMSs)
7th Intl. Congress on Env. Modelling and Software, San Diego, CA, USA,
Daniel P. Ames, Nigel W.T. Quinn and Andrea E. Rizzoli (Eds.)
http://www.iemss.org/society/index.php/iemss-2014-proceedings

Hypothesis Testing for Management: Evolving and Answering Closed Questions Using Multiobjective Visualization

Joseph Kasprzyk[1], **Joseph Guillaume**[2], **Joshua Kollat**[3], **Chris Danilo**[3]

1. University of Colorado Boulder, Civil Environmental and Architectural Engineering, UCB 428, Boulder, CO, 80309, USA (joseph.kasprzyk@colorado.edu)
2. National Centre for Groundwater Research and Training, Fenner School of Environment and Society, Australian National University, Canberra, Australia (joseph.guillaume@anu.edu.au)
3. DecisionVis, LLC, State College, PA, USA (jkollat@decisionvis.com; cdanilo@decisionvis.com)

Abstract: In order to use models to understand deeply uncertain future conditions, managers must be able to pose and test hypotheses about their management problems. In Iterative Closed Question Methodology (ICQM), a series of closed questions are used to structure thinking about hypotheses while looking beyond a problem's existing modeling representation. Our research is exploring how ICQM can contribute to a framework called Many Objective Robust Decision Making (MORDM), which uses multiobjective optimization and ensembles of uncertain future states of the world to create and evaluate robust solutions for environmental management. A visualization software tool; AeroVis, has greatly aided implementation of MORDM, allowing a user to plot tradeoffs between conflicting objectives, "brush" their preferences on plotted and unplotted variables, and view visualizations of solution robustness. This visualization approach provides a rich set of conclusions which is not always well understood (i.e. the user can interpret results that the modeler did not intend). In this presentation, we explore how visualization tools iteratively generate and evaluate management hypotheses and conclusions. We discuss the types of conclusions that can be made from AeroVis MORDM visualizations and walk through experimental examples of how individuals reason with the decision support tool. This illustrates that working within an MORDM framework helps the user consider alternate model assumptions about future inputs, parameters and model structure, supporting the idea that model assumptions can provide useful scenarios for environmental management.

Keywords: uncertainty; visualization; analysis of reasoning

1 INTRODUCTION

Environmental managers often have to make decisions under conditions of deep uncertainty, where stakeholders do not know or cannot agree upon the full suite of risks that affect their system, or their associated probabilities (Lempert 2003). Although quantitative models are often used in these situations, using the models is only a means to an end. Instead, the models must be used to pose and test hypotheses about the management problems. For example, models can be used to test whether a policy would provide a sufficiently reliable water supply in given circumstances, or whether another variation of a policy would perform better by some measure. A series of recent studies (Guillaume and Jakeman, In review; Guillaume and El Sawah, In review) has proposed a methodology called Iterative Closed Question Methodology (ICQM), where a series of closed questions are used to structure thinking about hypotheses while looking beyond a problem's existing modeling representation.

A second important issue in environmental management is how to quantify the performance of portfolios of management strategies. Cost benefit analysis, which is typically used to evaluate plans in the public sector, effectively collapses multiple benefits or goals into a single benefit function. An emerging trend in these systems is to use a multi-objective approach, which treats each objective or goal as independent and provides a tradeoff set of solutions to decision makers and analysts in order to facilitate their decision making. Multi-objective tradeoffs are of higher dimension, requiring the visualization of between three and ten objective functions as well as the associated decision variable values. Therefore, it is an open question on how to best visualize and interpret the results of these types of studies. Specifically, this paper will explore an approach called Many Objective Robust Decision Making (MORDM: Kasprzyk et al. 2013, Herman et al. 2014b), which combines multi-objective optimization, robust decision making, and interactive visualization. A recent MORDM study will be presented at this conference (Herman et al 2014a). MORDM exploits visualizations including three dimensional plots, parallel coordinates, and also the analysis of data using traditional tables of values.

In this paper, we posit that MORDM coupled with interactive visualization provides an implementation of ICQM. In general, our previous work has claimed the effectiveness of interactive, three dimensional visualizations and the multi-objective approach, but we have not explored these issues in the context of providing and answering closed questions and hypotheses. Our paper will focus on reflecting a few case studies that demonstrate the use of reasoning in multi-objective visualization. The results will not seek to compare visualization techniques, but we do hope to provide insight into how multi-objective visualization and decision support can be improved.

2 CONTEXT

2.1 The ICQM Theoretical Approach to Decision Support

IQCM poses a series of closed questions, which are used to structure thinking about hypotheses while looking beyond a problem's existing modeling representation (Guillaume and Jakeman, In review). In an iterative process, an analyst defines a closed question, and agrees upon assumptions on what is plausible. Model scenarios are then identified that satisfy the agreed assumptions, in order to obtain answers to the closed question. Iteration of this process in ICQM helps the analyst further their understanding of the problem. In light of the scenarios generated at the beginning of the process, the analyst can revisit the question, assumptions, and means of identifying model scenarios. The process provides a high level of guidance on how model runs are used, but requires specific methods in order to implement it.

2.2 Practical Implementation of ICQM using MORDM and Visualization

MORDM uses multi-objective optimization and ensembles of future states of the world to create and evaluate robust solutions for environmental management. The approach can be seen as applying ICQM for two reasons. First, MORDM is a technique that can generate new planning alternatives for a user to analyze. Second, by exploring ensembles of multiple states of the world (SOW), MORDM provides scenarios that can contribute to the series of closed questions used in ICQM.

To view examples of the outputs of MORDM, please refer to the published studies of Kasprzyk et al. (2013) and Herman et al. (2014b). The MORDM framework uses the concept of Pareto optimality to develop tradeoff sets of solutions. Informally, a solution is Pareto optimal if no other feasible solution is better in an objective without being degraded in another objective. Tradeoffs can also be presented with plots of solutions' performance in multiple uncertain SOW. Stored results from this process can facilitate iterative hypothesis testing, as described by ICQM. The analyst searches for scenarios, interactively explores a set of closed questions, and filters the results to suit and reflect his assumptions.

The visualizations used in this paper were generated using AeroVis, an interactive, multi-dimensional data visualization software package developed by Penn State University and DecisionVis, LLC (Kollat and Reed 2007, Kasprzyk and Kollat 2014, DecisionVis). The main goal of AeroVis is to facilitate plotting of multivariate data. The plots produced mainly show the conflicts between objective function values, but they can also show decision variable values and other performance data. The user interactively selects variables to show on spatial axes, as well as using the color, size, orientation and transparency of the plotted glyphs. Users can also define bounds on the plot and "brush" their preferences both on plotted and unplotted variables.

3 CASE STUDY

The case study for this work is to develop portfolios for a water supply market (permanent rights, options, and leases) in the Lower Rio Grande Valley (LRGV) of Texas, USA. The study uses a simulation model (Characklis et al. 2006, Kirsch et al. 2009) and multiobjective evolutionary algorithm (MOEA) optimization (see Kasprzyk et al. 2009, Kasprzyk et al. 2012 for details of the MOEA optimization of the LRGV).

The goal of the case study is to find a portfolio that balances market and non-market supply instruments for a single city in the LRGV. The decision variables include the volume of permanent rights specified in the portfolio, an adaptive options contract that allows the city to acquire water later in the year if needed, and a series of risk-based thresholds that control the city's use of water on the market. Objectives quantify the portfolios' performance. These objectives include the cost of acquiring water, the reliability of meeting the city's municipal drinking water demand (as well as other performance criteria including resilience), and measures of efficiency including minimizing the number of leases, minimizing the city's water surplus, and minimizing the cost variability.

Specifically, this study builds from a version of the LRGV presented in Kasprzyk et al. (2013). This study explored two relevant issues for the current work. First, the study explored multiple problem formulations, named I through IV, where I had the fewest number of decision variables and IV was the most complicated set of decision variables. Second, the study was the first application of the MORDM framework to the LRGV where multiple SOW were generated that perturbed the input data to the LRGV simulation model. Each modeling scenario presents an aggregate result from stochastic runs rather than a deterministic story. All SOW are currently considered plausible, but it is expected that scenarios would be missed, and some states of the world included may need to be excluded during the analysis.

4 EXAMPLES OF REASONING

We present two examples of reasoning using visualization with the LRGV case study. Each was first analyzed graphically as a flow of ideas, and then condensed to a table to fit in this paper. The first example is obtained by structuring the results section of a published article (Kasprzyk et al. 2013). The second example is obtained from one of the authors narrating their reasoning during a live exploration of the same dataset. The examples were selected to allow in-depth discussion of contrasting and complementary situations, while using a single case study and analyst, due to practical limitations.

4.1 Published Analysis

The first example is drawn from the results section of an article introducing MORDM (Kasprzyk et al. 2013). The flow of reasoning is shown in abridged form in Table 1. A main chain is shown in black. Branches are shown in grey. The branches either follow up on hypothesized explanations of results or tentatively explore the consequence of a preference. The main chain starts with a prior analysis (Kasprzyk et al. 2012) of a multi-objective plot that suggested that problem formulations III and IV had better expected performance. However, the analysts posited that this conclusion could be altered if different SOW were considered. Percentage deviation in performance was therefore visualized using a parallel coordinate plot. This prompted the reduction of the set of alternatives by "brushing" (or eliminating from view) only the solutions that meet acceptable thresholds of reliability. Minimizing deviation in number of leases allowed the selection of a single solution, which is compared against previously identified solutions in a table and in a multi-objective plot that displays percentage deviation as an indication of robustness. This is used to justify the rejection of previous solutions which lack robustness despite being efficient in a base case scenario. To understand when the solution might perform poorly, threshold values of factors were then identified. The analysts intend for these to be discussed by decision makers. As one example, high losses lead to poor performance. They seem possible, and monitoring of losses is therefore suggested as a mitigating measure.

Table 1. Flow of reasoning for Published Analysis (example 1). Grey indicates branches off the main chain of reasoning

Question	Answer
Which portfolio strategy can best increase the city's reliability while ensuring sufficient water for other regional uses?	Multi-objective plot shows problem formulations III and IV have better expected performance.
Do III and IV allow finer tuning of market use?	Multi-objective plot shows excellent performance in dropped transfers and surplus water.
Which portfolio strategy is best in worst case SOW?	Parallel coordinate plot shows problem formulations I and II have higher deviation in cost and number of leases, lower for reliability.
Do I and II have less market use?	Multi-objective plot shows I and II have more leases and surplus water, for higher reliability.
Which portfolio strategies are retained if deviation in reliability in worst-case SOW must be acceptable?	Brushing identifies solutions with low deviations in cost, number of leases, higher deviations in cost variability.
Which portfolio strategies have fewer conflicts between objectives?	Minimizing variation in number of leases also has low variation in reliability and cost.
Which portfolio strategy is best if the deviation in number of leases is minimized maintaining acceptable reliability?	Parallel coordinate plot identifies one new solution.
Is the new solution better than previous solutions if higher costs are acceptable for greater robustness and ease of implementation?	Previous solutions had good performance in cost, number of leases and surplus water; New solution is simpler; Table shows volumes of rights and options are similar; Multi-objective plot shows previous solutions have higher deviation for critical reliability, but lower for cost.
Will the new solution perform poorly in some circumstances?	E.g., high losses cause poor performance.
Which factors are most important in avoiding poor performance?	E.g., flows, losses and demand affect market use the most.
Will the circumstances in which solutions perform poorly, e.g. high losses, actually occur?	Distribution of losses described are not drastic changes compared to historic data.
Can monitoring help address high losses before they occur?	This is an open question for future management.

4.2 Live Exploration

In the second example, the analyst performed and narrated an exploratory analysis of the LRGV dataset using the AeroVis software, focusing on only formulation IV. The analyst starts simple with a two dimensional plot considering engineering criteria of reliability and critical reliability. They use one plot to try to answer multiple possible questions: looking for trade-offs and looking for solutions that differ significantly from others. In this case, both reliability and critical reliability can be maximized, which is clarified by changing the bounds of the plot. The analyst then adds resilience as an additional objective, which identifies standout solutions of interest. The remainder of the exploration focuses on identifying whether and how these solutions should be eliminated. An initial plot is inconclusive, but a second plot clearly shows the solutions to have high surplus water. The analyst searches for a cause, and finds high volumes of permanent rights and low volumes of options. After checking that this result does not change depending on how the options are used, the analyst returns to other properties of these outliers. They have high cost variability. Following up this result shows increased reliability comes at the expense of higher cost variability. This is verified with another plot, in the process confirming that high reliability comes at the cost of high surplus. The analysis results in placing constraints on surplus and cost variability, allowing a focus on a smaller group of candidate solutions. The analyst ends by commenting that they would continue the analysis by checking the other values of the selected scenarios in a table. If there is anything alarming, they would look for a relationship with other variables.

Table 2. Flow of reasoning for Live Exploration (example 2).

Question	Answer
Is there a trade-off in optimising multiple objectives? Do any solutions differ significantly from others?	A two-variable plot shows both reliability and critical reliability can be maximised.
	A two variable plot with resized bounds shows that critical reliability is always higher within Pareto non-dominated set which can be explained by lower frequency of critical events.
	Overlaying resilience as colour shows resilience tends to increase with critical reliability, but there are some outliers.
Should these standout scenarios be eliminated?	Plot of resilience and reliability shows they have high resilience. This is inconclusive.
	Plot of surplus and reliability shows they have high surplus water. This may be reason to eliminate them.
Are high surplus water scenarios caused by volume of permanent rights?	Plot x=rel, y=surplus, colour=rights shows these scenarios have high levels of rights.
Are high surplus water scenarios caused by options?	Plot x=rel, y=surplus, colour=options low shows these scenarios have low levels of options.
Do options always lead to high surplus?	Plots show that how options are executed on market does not change result.
Should these standout scenarios be eliminated?	Table of values shows that variability in cost is high.
Is cost variability related to other objectives?	Brushing points with high cost variability shows increased reliability comes at the expense of higher cost variability.
	Plot x=surplus, y=reliability, colour=cost variability confirms relationship to cost variability and that high reliability comes at the cost of high surplus.
Should these standout scenarios be eliminated?	(Surplus and cost variability should already both be constrained).

5 DISCUSSION

This discussion summarizes the types of conclusions observed, relates the reasoning processes to the principles of ICQM and considers implications for teaching the use of multi-objective optimization and for its use within management.

5.1 Types of Conclusions

The two examples tackle a variety of questions, and therefore types of conclusions. The first example aims to identify a best solution as a compromise between many objectives, because the utopian solution is not achievable. By emphasizing robustness, it draws conclusions about performance in alternate SOW using a measure of robustness calculated on a set of SOW. If some of the SOW are considered implausible, the measures will however be biased against solutions that perform poorly in those SOW. The analyst therefore also draws conclusions about specific thresholds on variables of SOW in which a solution would perform poorly. As the analysis progresses, conclusions are also drawn that express preferences, both on acceptability of performance and variables to optimize. In parallel, a number of conclusions involve testing hypothesized explanations of observed results, and evaluating the relative importance of factors.

The strategy in the second example identifies a single scenario that stands out in some way and uses its properties to guide a process of decision making by elimination. The emphasis is therefore on coming to a conclusion about the acceptability of the values of each variable. This is however triggered by the identification of standout scenarios. Thanks to the flexibility of AeroVis, this is accompanied by exploration of relationships between variables affected by or driving eliminations of scenarios. As with the first example, a number of conclusions take the form of testing of hypothesized explanations using a variety of plots, particularly about the existence of a trade-offs or relationships between two variables.

5.2 Fit with principles of ICQM

Recall that we posited that MORDM coupled with visualization techniques provides an implementation of the concepts of ICQM, specifically an iterative sequence of evolving and answering closed questions. Each example of reasoning was able to be expressed in this form. It would seem that this arises naturally in multi-objective problems, where any optimization or plot can only show part of the picture at one time. The first example explicitly identifies limitations of each stage of the analysis which moves the analysis forward. The second example starts with broad questions, focuses on outlier solutions and then iterates through each objective to seek to understand and perhaps constrain the analysis further.

In the examples, in addition to a main line of iteration, there are a number of side-branches (e.g. the grey text in Table 1). This matches the expectation in ICQM that the analyst has significant discretion in what changes to make at any point. Several options might therefore be tried before identifying one line to focus on. Some lines of inquiry therefore end because the remaining uncertainty and implications do not warrant being pursued further.

ICQM prescribes the need to eventually consider changes in the question asked (including objectives and decision variables), the assumptions used (e.g. what SOW are plausible) and the method of obtaining results. The first example changes constraints on what solutions are acceptable, and criteria used to optimize the choice, as well as considering different (levels of detail) of decision variables and a set of SOW. It does not however change which SOW are included in the set, though the analysis ends with a hint that this is the next step. The second example focuses on a single solution, varying objectives considered and imposes constraints on acceptable solutions. It does not yet consider what alternate SOW might be plausible, though continuing evaluating variables of the outlier solutions might ultimately evaluate variables related to SOW.

5.3 Implications for multi-objective visualization

The two examples focus on different types of analysis; in other words, the type of analysis in an informal “live” exploration is different than published results. Therefore, it could be argued that the transferability of the results here is limited, and different studies will exhibit different types of reasoning. However, it seems that the analysis of flows of reasoning resulting from visualization can raise issues of interest specifically to teaching and use of multi-objective visualization in management, as explained in the following subsections.

5.3.1 Implications for teaching

A number of guiding principles can be suggested. Both examples started with an initial question in mind, even if the question is quite broad. This suggests the need to come to a dataset at the minimum with the intention of identifying relationships or outliers. The graphical representation of reasoning in particular shows a large number of branches of reasoning, which can seem overwhelming. New users especially should not feel obliged to follow all leads, and should consider it normal to build a coherent argument by reflecting on results rather than during the visualization itself. The analyst’s narrative showed that even in the live exploration example, visualizations were specifically chosen with a hypothesis or question in mind. It is important for students to learn what questions to ask, and to take the time to consider potential next questions (as prescribed by ICQM) rather than quickly jump from one plot to the next. The act of making flows of reasoning explicit also appeared useful to identify strategies that new users could learn.

5.3.2 Implications for management

The examples illustrate the importance of taking a satisficing approach (looking for a solution that is ‘good enough’ rather than strictly best, Simon 1956) when supporting decisions for complex problems. The first example started from a well defined multi-objective problem that had previously been ‘solved’ several times (Characklis et al. 2006; Kasprzyk et al. 2012). The second example took the same dataset and identified further potential changes to preferences. In contrast to the naïve view of hypothesis testing as the last word on a scientific truth, it is best approached as an iterative process that explicitly has no scientifically-determined end. MORDM and visualization appears to provide an effective means of achieving this, but a number of research issues remain. In particular, the flexibility that visualization allows needs to be accompanied by some understanding of prioritization of what questions to ask first. And any decision aiding process, including hypothesis testing for management, needs to be to some extent decoupled from the actual decision making process. Analysis is fundamentally an ongoing process, and it is therefore impossible for management to wait until it finishes.

6 CONCLUSION

Decision making under deep uncertainty is usefully addressed through an iterative process of generating and evaluating hypotheses. This paper presented two examples of an analyst’s reasoning when visualizing model scenarios. Making reasoning explicit allowed identification of a sequence of hypotheses, which could be usefully expressed as closed questions. As a result of the flexibility of multi-objective visualization, a variety of conclusions were made, taking into account alternate model assumptions about future inputs, parameters and model structure. This suggests value in future research by describing how analysts currently use and could better use visualization tools to explore management solutions that are robust to alternate model assumptions.

7 REFERENCES

Characklis, G.W., B.R. Kirsch, J. Ramsey, K.E.M. Dillard and C.T. Kelley (2006). Developing portfolios of water supply transfers. Water Resour. Res. **42**(5): W05403 DOI: 10.1029/2005wr004424.

DecisionVis, LLC. AeroVis Software. http://www.decisionvis.com/

Guillaume, J.H.A, A.J. Jakeman. In review. Iterative Closed Question Modelling for assessing the effect of uncertainty on model predictions.

Guillaume J.H.A., S. El Sawah. In review. Fostering assumption-based stress-test thinking in managing groundwater systems: learning to avoid failures due to basic dynamics.

Herman, J., H. Zeff, P. Reed, and G. Characklis (2014a). “Balancing robustness to future uncertainty between cooperating regional water utilities”. International Environmental Modelling and Software Society Proceedings (this conference).

Herman, J., H. Zeff, P. Reed, and G. Characklis (2014b) “Beyond optimality: Multi-stakeholder robustness tradeoffs for regional water portfolio planning under deep uncertainty.” Water Resources Research (In-Review).

Kasprzyk, J.R., S. Nataraj, P.M. Reed and R.J. Lempert (2013). Many objective robust decision making for complex environmental systems undergoing change. Environmental Modelling & Software **42**: 55-71 DOI: 10.1016/j.envsoft.2012.12.007.

Kasprzyk, J.R., P.M. Reed, G.W. Characklis and B.R. Kirsch (2012). Many-objective de Novo water supply portfolio planning under deep uncertainty. Environmental Modelling & Software **34**: 87-104 DOI: http://dx.doi.org/10.1016/j.envsoft.2011.04.003.

Kasprzyk, J. R., and J. Kollat (2014) “Balancing conflicting management objectives using interactive, three-dimensional visual analytics” International Environmental Modelling and Software Society Proceedings (this conference).

Kollat, J., Reed, P. (2007) “A framework for Visually Interactive Decision-making and Design using Evolutionary Multi-objective Optimization (VIDEO)” Environmental Modelling and Software 22(12), 1691-1704.

Lempert, R. J., Popper, S. W. and Bankes, S. C., 2003. Shaping the next one hundred years: New methods for quantitative, long-term policy analysis, Rand Corporation, USA.

Simon, H.A. (1956). Rational choice and the structure of the environment. Psychological review **63**(2): 129-138.