Masters Proposal

James Saunders

2010-01-18

Contents

1 Introduction

Ecological sustainability is rapidly becoming a greater concern for the global population in general and policy makers specifically. Humanity's increasing population and industrial activities are placing unprecedented strain on the environment. Likewise the global population is becoming increasingly vulnerable to both environmental catastrophes and to the slow degradation of the environment that it is dependent on (Lubchenco, 1998).

This increased vulnerability causes both massive loss of life and negative impacts on quality of life in both developed and developing economies, as well as wreaking massive economic damage on many industries (Lubchenco, 1998).

Humanities impact on the earth is substantial and is rapidly increasing(Vitousek et al., 1997). It has become clear that human activities register on a global ecological scale, making humans effectively on the level of a force of nature(Lubchenco, 1998). In this regard it has become vital to be able to carefully plan our actions with respect to their impact on the earth, as well as enlarge our understanding of pure ecological systems to incorporate social, economic and political effects. We must extend our understanding to deal with social and ecological systems as a 'whole' system. This need has led to the introduction of the term SOCIAL-ECOLOGICAL systems (Peter, 2010).

1.1 Social-Ecological systems

A defining characteristic of social-ecological systems is that they transcend the traditional boundaries of traditional research disciplines. Since they incorporate aspects from many fields of ecological research and social studies it is unlikely that a single research group will have an expert understanding of all the aspects of the system. However the participants and affected parties within social-ecological systems go far beyond researches, scientists and politicians. They include the inhabitants of the system as well as those members of society that make a living off the ecology. As such when dealing with social ecological systems the decision making process must not only be highly trans-disciplinary but also highly participative(Baran and Jantunen, 2004). The great problem facing trans-disciplinary work is the fragmentation of understanding that occurs due to no single member of a research group having a complete view of the system, and due to a lack of a common language and knowledge representation between disciplines.

Social-Ecological systems are COMPLEX SYSTEMS. In that they consist of many interacting subsystems giving rise to non-linear emergent behaviour. Further Social-Ecological systems exhibit the properties of a self-organising system and thus may be considered COMPLEX ADAPTIVE SYSTEMS Cowan et al. (1994), Levin (1999).

Complex adaptive systems are inherently difficult to model, especially those in the social-ecological domain. This is due, in part, to the modeler being caught between two approaches. The first being the reductionist approach of trying to model all the individual agents in the system, however the inherent non-linearity of the system means that even the smallest deviations between model and system can cause huge emergent differences. The second approach is to try take a qualitative holistic approach, however this lacks quantifiable rigor. The challenge is then to integrate these two seemingly contradictory approaches. Wu and Marceau (2002)

Further global social, economic and ecological systems are becoming increasingly interconnected. This makes decisions made in one area and sector affect the sustainability of a host of other systems (Folke et al., 2005). For example legislation passed by the US department of energy to promote the use of bio-fuels prompted fears of decreased food security, because land previously used for food production has been re-purposed for energy production and corn prices increased (Runge and Senauer, 2007). This very simple example illustrates how decisions made in different organisations and countries can have broad ranging effects outside of their traditional domains.

1.2 The difficulty of decision making in Social-Ecological systems

Decision makers are faced with a multitude of problems when faced with making Social-Ecological decisions. The domain is often not well understood, and there is no single expert opinion on which to base the decisions. The link between actions and effects are obscure. Small interventions in the systems can have disproportionately large effects, both in and outside of the system. Never the less, decisions have to made, in fact it is increasingly important that active intervention does take take place. Thus decision makers are forced to make often ill-informed decisions, often with severe consequences.

1.3 The importance of monitoring in decision making

Whenever decisions are made they are based upon some presumed relationship between the factors in the system in question. In simple systems these relationships are clear and can be relied upon. However in complex systems the the relationships between factors are difficult to determine and multi-faceted. Inherent in modeling a complex system to the point where we can reason about it is a process of simplification, where information is lost. Due to this all models of complex systems have a degree of uncertainty, as well as a likely hood that they are incorrect and not fully representative of the systemRichardson (2003).

Monitoring complex systems is important to for 3 reasons: Firstly to continue checking that the assumptions underpinning the decision were correct; secondly to check that the system has not evolved in some way that requires an evolution of the previous decision; and lastly to check that the decision is having the desired effect. Governance can only be effective if its rules and interventions evolve as the system in governs does(Dietz et al., 2003).

1.4 Increasing amounts of available data

Large amounts of high quality data are becoming more and more accessible. Organisations such as the Long Term Ecological Research network (LTER) and the SAEON (South African Earth Observation Network) extract large quantities of data that is accessible to decision makers and researchers. The awakening realization of the importance of understanding climate change has lead many governments to invest in new monitoring capacity. Added to this is the increased pressure on governments and scientific institutions to make data more accessible to the public.

This increasing amount of raw data must be put to use, and harnessed to increase our decision making ability(Lubchenco, 1998). However rather than aiding the decision making process, information overload is hindering our ability to make sense of the data. In order to extract value from this data we need to be able to aggregate and represent it in a way that is useful to decision makers.

2 Related Work

2.1 Complexity Theory

2.2 Bayesian Networks for modeling social ecological systems

A Baysian Network is a statistical model that takes the form of a Directed-Acyclic Graph (DAG) where the nodes represent variables in the social-ecological system and the links represent a causal relationship between two variables. Each variable (or node) is separated into several possible states, these can be numerical (0-10, 10-20, 20-30) or qualitative, (healthy, unhealthy). The relationship between a node and its parents (the nodes that affect it through a causal link) is defined in a Conditional Probability Table (CPT), which relates possible combinations of parents nodes' states to the states of the child node (Pearl and Shafer, 1988). These tables can be populated automatically through an equation or historical data; or manually by an expert in the field (Peter, 2010). Once a bayesian network has been setup it can queried by setting the distribution of probablitities of the states of a subset of the nodes, which will then propogate to show the likely states of the remaining nodes, this is know as a "what-if" query(Potgieter, 2004). An example of a Bayesian network, showing the graphical structure is shown in Figure 1. Bayesian networks have been used extensively in many fields, especially medical diagnosis, financial modelling and machine learning.

Bayesian networks have been used as a tool to model social-ecological systems(Baran and Jantunen, 2004, Borsuk et al., 2004). They exhibit many desirable features for modelling complex systems. Most notably they are:

- Ability to create and change subsystems without changing the entire network (Borsuk et al., 2004).
- Simple graphical structure which makes participatory design more accesable (Baran and Jantunen, 2004).
- Ability to do scenario analysis through "what-if" queries, which allows for the interogation of different courses of action for decison makers (Peter, 2010).
- Includes uncertainty in the modeling predictions, giving decision makers an idea of the certainty of the represented knowledge.

These models were used to inform water regulations in the European union (Baran and Jantunen, 2004)and North Carolina in the US (Borsuk et al., 2004).

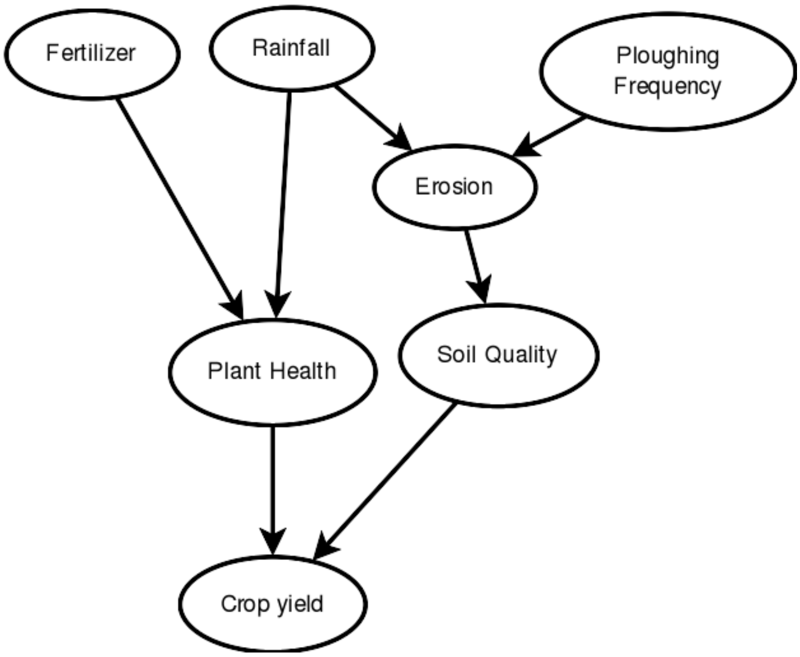

Figure 1: An example of a Bayesian network showing the relationships between variables of an agricultural model.

2.3 Bayesian Participatory-Based Decision Analysis

- BPDA (Bayesian Participatory-Based Decision Analysis) architecture developed to assist in these problems
  - General Modeling framework requiring only assumptions of causality and conditionality.
  - Uses graphical causal maps and Bayesian networks to aggregate all the knowledge about the system into a single artifact, unifying participants understanding of the inter-relationships and causal links of the system.
  - Aggregates both information from both soft-approaches (General Complexity, Narratives, Expert opinion, work-shopping) and hard-approaches (Restricted Complexity, Statistical and mathematical models)

3 Core Contribution

The crux of the proposed masters lies in turning the BPDA models into a monitoring framework by linking it up to real-time or near real-time data-sources. The purpose of this monitoring framework is interrogate the correctness and ultimately the value of the models. This is done with a mindset based on the end use of such models, which is ultimately to assist decision makers in complex social-ecological problems.

Central to the reasoning behind the approach taken is the understanding that a model is inherently incorrect. As stated by Richardson (2003):

> “...the model is no more than a rough and ready caricature, or metaphor, of reality. As such the knowledge contained in the model should be regarded with a healthy skepticism, seeing it as a limited source of understanding.”

Given this understanding that models are inherently flawed, and that our interest in modeling social-ecological systems is primarily to aid decision makers, the proposed masters intends create a real-time monitoring system that:

- Interrogates the validity of the model.
- Interrogates the validity of decisions made based on the model.

3.1 Interrogating the validity of the model

Fundamentally a model should represent reality to a satisfactory degree to allow reasoning about the system. A model that does not represent reality, or rather is consistently inaccurate in its prediction about reality, is of little use. Thus the first and most important monitoring task is to check that over time the model is generally correct. Obviously since Bayesian models are probabilistic in nature checking their accuracy requires a large amount of data. Further more the phrase accurate needs to be rigorously defined in a quantitative and qualitative sense.

Undoubtedly over time there will be statistical discrepancies between the model and reality. These discrepancies can be of a few very different types:

- Minor inaccuracies in the model
- Major flaw in the initial model
- A symptom of some evolution that has occurred within the system being modeled.

3.1.1 Dealing with minor discrepancies

Minor tweaking of the Conditional Probability Tables with in the model is to be expected, when only minor discrepancies are detected the information gathered should be incorporated into the model, thus making it more accurate. This would seem to be especially useful in the cases where subjective expert opinion was used to generate the probability tables for a set of variables, as it would allow for the relationships to be accurately quantified.

3.1.2 Detecting a major problem with the initial model

It is possible ,and often likely, that their are major inaccuracies in the initially designed model, in that it does not, and never did, reflect the reality of the system being modeled. In this case it is imperative to detect this and as early as possible. Once detected It would be desirable to be able to localize the parts of the model that are the least correct. This approach is possible due to the separated nature of subsystems within a Bayesian network. Major errors in the model can also occur if an important variable has been left out of the model. In this case it would be interesting to try to see if some sort of statistical regularity can detect the presence of a hidden variable, in order to alert the system modelers to the possibility. Once an error is detected ideally the sub-systems would be re-modeled, and then incorporated back into the framework and used to review the decision making procedure.

3.1.3 Detecting and dealing with system evolutions

Complex systems have the capacity and propensity to evolve. They can often rapidly move between multiple stable states. When this occurs it is likely that the model will no longer be accurate, however this is not because the model was incorrect, simply that it modeled the system in a previous state. Differentiating this from the case where there were initial errors in the model is vital. It would allow the monitoring framework to alert relevant parties that a phase shift has taken place. This would potentially lead to updating the model, but with the added information of why it is no longer reflective of reality.

Further if the stable states of an environment are known, or over time have been observed, the monitoring framework could switch models to use the model that best describes the new state of the system. Further the evolution of the system will now have been carefully monitored, showing how the relationships between variables changes, and which variables and relationships become more dominant.

3.2 Interrogating the validity of decisions

As previously stated decisions are made based on an assumption that some cause-and-effect relationship exists within the system. A decision is only useful so long as the relationship it is based on is correct and stays correct. To this end the most important role of the monitoring system is to monitor the correctness and continuing validity of the assumptions upon which a decision is based.

A necessary step in monitoring decision assumptions is to develop some way to codify these assumptions with relation to the model. This codification should also have a corresponding visualization if appropriate.

When a codified assumption is found to no longer be true, or is not found to be true with an acceptable probability, the monitoring framework should alert decision makers and modelers, allowing them to take further action. This will tie into the general interrogation of the validity of the model as well as the detection of system evolution.

4 Methodology and Outline of Activites

5 Forseeable Risks

5.1 Data collection and time-scales

In any project where large amounts of data is required, gaining access to enough data sources becomes a problem. Further more, bayesian machine learning requires data at standard spatio-temporal scale. In order to use all the available data sources some of the data will have to be preprocessed or interpolated, which in itself can bias the system. The masters programme will not cover a significant length of time in terms of social-ecological systems. This may make it difficult to get meaningful results for some of the more complicated matters surrounding system evolution. This may require that simulations are used, which is a much less rigorous test of the system and theory.

5.2 Reasoning about model inaccuracies

Much of the automated reasoning in the proposed system revolves around trying to quantify how much the model does not know. This raises the problem of not being able to bench mark in real systems, since we cannot know what we don't know. As stated by Taleb (2008):

> "We don't have the luxury of sitting down to read equation that governs the universe; we just observe data and make assumption about what the real process might be and "calibrate" by adjusting our equation in accordance with additional information"

This raises obvious problems about judging the correctness of a prediction outside of simulated data.

5.3 Codification of decisions and assumptions

The codification of knowledge and intention is a subtle and complicate affair which can become and endless amount of possible work. The challenge here is to balance the need for a semantically meaningful representation with a simple, practical and achievable framework. This aspect should not dominate the other aspects.

5.4 Software complexity

The software underlying the BDPA framework is complex (as is most machine learning systems) is complex and multifaceted. There is a risk that too much research time is sacrificed.

6 Benefit of Research

The pinnacle of this line of research would be to have real-time updating causal models of complex systems, that would allow decision makers to accurately reason about complex systems (and specifically social-ecological systems). This would allow for greater accuracy and positive impact of interventions as well as traceability on the reasoning behind decisions, and post decision analysis of the effectiveness of the intervention.

[Talk more about the importance of knowing what you dont know]

However that point is most likely far in the future, but this research is a necessary stepping stone to getting there. More immediately it would be nice if a simple tool could be developed from the research to help in situations similar to the case study.

References

Eric Baran and Teemu Jantunen. Stakeholder consultation for bayesian decision support systems in environmental management. In *Regional Conference on Ecological and Environmental Modeling*, 2004.

M.E. Borsuk, C.A. Stow, and K.H. Reckhow. A bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecological Modelling*, 173 (2-3):219–239, 2004.

G.A. Cowan, D. Pines, and D. Meltzer. *Complexity: Metaphors, Models, and Reality*. Perseus Books, 1994.

Thomas Dietz, Elinor Ostrom, and Paul Stern. The struggle to govern the commons. *Science*, 302:1907 – 1912, 2003.

C. Folke, T. Hahn, P. Olsson, and J. Norberg. Adaptive governance of social-ecological systems. *Annual Review of Environment and Resources*, 30:441–473, 2005.

S.A. Levin. *Fragile Dominions: Complexity and the Commons*. Perseus Books, 1999.

Jane Lubchenco. Entering the century of the environment: A new social contract for science. *Science*, 279:491–497, 1998.

J. Pearl and G. Shafer. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann San Mateo, CA, 1988.

Camaren Peter. *Bayesian Participatory-Based Decision Analysis: An Evolutionary, Adaptive Formalism for Integrated Analysisof Complex Challenges to Social-Ecological Systems Sustainability*. PhD thesis, University of Cape Town, 2010.

Anet Potgieter. *The Engineering of Emergence in Complex Adaptive Systems*. PhD thesis, University of Pretoria, 2004.

K.A. Richardson. On the limits of bottom-up computer simulation: Towards a nonlinear modelling culture [electronic version]. In *Proceedings of the 36th Hawaii International Conference on System Sciences, IEEE*, 2003.

C. Ford Runge and Benjamin Senauer. How biofuels could starve the poor. *Foreign affiars [Electronic Version]*, May/June, 2007.

Nassim Nicholas Taleb. *The Black Swan*. Penguin Books, 2008.

Peter Vitousek, Harold Mooney, Jane Lubchenco, and Jerry Melillo. Human domination of earth's ecosystems. *Science*, 277:494–499, 1997.

Jianguo Wu and Danielle Marceau. Modeling complex ecological systems: an introduction. *Ecological Modelling*, 153(1-2):1 – 6, 2002. ISSN 0304-3800. doi: DOI:10.1016/S0304-3800(01)00498-7. URL `http://www.sciencedirect.com/science/article/B6VBS-44W9MRB-1/2/d2656f481bed6ae88582363e81983720`.