Jul 1st, 12:00 AM

Possible Courses: Multi-Objective Modelling and Decision Support Using a Bayesian Network Approximation to a Nonpoint Source Pollution Model

David Swayne

Jie Shi

Possible Courses: Multi-Objective Modelling and Decision Support Using a Bayesian Network Approximation to a Nonpoint Source Pollution Model

David Swayne, **Jie Shi**
University of Guelph, Computing research Laboratory for the Environment,
Guelph, N1G2W1, CANADA, dswayne@uoguelph.ca

Abstract: Modelling systems frequently work in a single domain, such as physical or chemical process modelling, hydrology or combinations, to simulate process in nature such as pollution transport or the production of food or manufactured goods. Side-effects of agro-industrial processes, or gains / losses from production enterprises are separately modelled without the ability to examine trade-offs or alternatives. Multi-objective modeling attempts to combine "apples and oranges" through decision theoretical principles. Such treatments can couple production and waste systems to quantify the economic cost of remediation. We demonstrate such an application, from the data acquisition, model calibration through to the hypothesis testing, for a nonpoint source pollution model together with a yield / energy / revenue model based on corn / grain / meadow rotations typically found in Southern Ontario, Canada, using realistic economic data obtained from agricultural operations similar to those found in this region.

Keywords: *multicriteria decision support, Bayesian networks, nonpoint source pollution*

1. Introduction

Agricultural nonpoint source pollution (AGNPSP) has been increasingly recognized as a major contributor to the declining quality of lakes and rivers in Canada (and elsewhere). Many modeling systems have been constructed to mimic the transport and fate of agricultural chemicals and nutrients in surface waters. They have been very successful, even quantitatively, but the problems persist. Our research aims to couple the AGNPSP problem to the economic side of agriculture, to develop a realistic estimate of the costs inherent in environmentally sustainable agricultural practices.

In order to do this, we have to have a set of models and data for several different problem areas. For the moment we will consider the example of crop production instead of food or dairy animals. Some aspects of those studies may yield to pollution models that might be considered to be more like point source. These issues may eventually be incorporated into standard models, but for our demonstration project we restricted ourselves to standard crops that are usually grown in Southern Ontario, Canada, using standard means of production and a limited suite of conservation practices. Our inputs of yield and crop pricing were based on historical yield data, and in some cases on generating a distribution of prices for the particular commodities that was based on ranges containing historical prices. Our erosion and sediment data were obtained from experimental work in Canada, particularly on the GAMES model [Rudra 1986]. Our model strategy does not depend on the internal mechanics of the model chosen, i.e. the methodologies of model construction would not change if a new model were substituted that was capable of generating similar data.

2. Background

Sustainable development, is defined as economic development that meets the needs of the present without compromising the ability of future generations to meet their needs [Hermanides 1987; Cornelissen 2000]. In other words, a proper environmental policy objective should consider three dimensions: economic, social and environmental aspects. Conflict exists between them. There is no way to satisfy all of the criteria. For agriculture land use management policy, the two major paradigms are: yield-oriented policy (conventional, production - driven), and environment - oriented policy (environmentally - driven) [Huylenbroeck 1996]. Recent research has investigated long-term sustainable land use management [Bessembinder 1996; El-Swaify 1996]. Most of the work is on the land use arrangement/planning, or production analysis. It usually uses linear programming, logistic regression analysis, or dynamic programming techniques to optimize a criteria function [Bessembinder 1996; Hipel 1996] for resolving conflict. The trade-off process used to solve the conflicts between economic and environmental objectives in a rural planning project [Huylenbroeck 1996], is described as having three steps: 1) separate aggregation of the economic and ecological criteria; 2) visualization of trade-offs; and, 3) discrete compromise analysis to support the final choice

A substantial amount of recent research has been carried out to develop decision support tools for the management of agro-forestry resources. Among these tools the Multiple Criteria Decision-Making (MCDM) approach plays a prominent role [Marangon 1998]. Most of these tools perform decision analysis based on calculating the result by assigning weights/scores [Parton, 1996; Yakowitz 1996; Noghin 1997; Wang 1998; Tan 1998; Bots 2000; Costa 2001; Janssen, 2001] to the alternative criteria attributes, or by assigning a probabilistic proportion to a decision tree structure [Gratch 1995; Warburton 1998]. Some have linked their Expert System with a Geographical Information System (GIS) database [Abu-Zeid 1996; Crosetto 2000; Rao 2001]. Since the 1980s, a probability-utility based approach using Influence Diagrams (Decision Networks) [Tatman 1990] has become accepted as an efficient alternative for many classes of models [Howard 1984; Shachter 1986].

Influence Diagrams are a class of graphical modelling paradigm that can represent probabilistic inference and decision analysis models. The reasons that an Influence Diagram is an effective modelling framework for a diverse array of problems involving probability are: a) it captures both the structural and qualitative aspects of the decision problem and serves as the framework for an efficient quantitative analysis of the problem; b) it allows efficient representation and exploitation of the conditional independence in a decision model; and, c) it has proven to be an effective tool for not only communicating decision models among decision analysts and decision makers, but also for communicating between the analyst and the computer.

3. Criteria

Watershed pollution comes from many sources. We emphasize agriculture land use activities because much of the contamination of surface waters is due to nonpoint Source (NPS) pollution. According to the USEPA [Osmond 1996], approximately 60% of the total NPS pollution load on assessed surface waters is due to agricultural runoff. The primary agricultural pollutants are sediment, nutrients and pesticides. For decision procedures these issues are the set of available options, the criteria and the uncertainty on the outcomes of each option. Each watershed is unique in its physical characteristics, land uses, water resources, socioeconomic status, and public concerns.

Generally speaking, decision - making involves the need to evaluate a finite number of possible choices (alternatives/candidates) based on a finite number of attributes (criteria). In our research, the decision alternatives are seven cropping tillage methods (Table 1) arranged into nine multi-year scenarios. The criteria attributes are soil erosion rate, sediment yield (local and total in the whole watershed), operating net revenue (price*yield – energy, labour and capital cost). These latter (cost) criteria estimates are defined as ranges (where, for example the energy inputs for a particular strategy are based on estimates from the literature, and sub-divided in our model to reflect variance within a particular range, and increased for no-till versus normal cropping.

Our construction combines environmental pollutant transport models with economic (crop yield, expense, revenue) models to investigate a sustainable tillage system in order to encourage the adoption of improved management systems. A user interface was developed to provide the communication tool which links the decision maker’s interaction with the graphic probability model. The application enables the user to view the impact of parameters (such as sediment yield, erosion rate) on decision alternative scenarios. It can also aid the user in making long term soil productivity predictions.

4. Belief networks

A *belief network* (also called *Bayesian network,* or *probabilistic network*) is a graphical presentation of probability combined with mathematical inference calculation. It is used to represent dependencies between random variables. Each variable represented as *node*, is connected by directed links, represented as *arrows* or *arcs*, with conditional probability table (CPT) values assigned to the variables making up a belief network. The nodes in a belief network are called *chance nodes*. Chance nodes represent uncertain events or variables. They can be a continuous or discrete random variable, or a set of events. A d*eterministic node* is a special case of chance node, which operates deterministically on other nodes. The *arrows* are the directed links between variables (nodes) and this direction represents the conditional dependent relationship of these nodes. For example, an arrow entering a chance node means that the author's probability assignment represented by the chance node is conditional on the node at the other end of the arrow (its input).

5. Influence Diagrams

Influence Diagrams (ID) are the extension of belief networks. In addition to nodes for representing random variables provided by belief networks, Influence Diagram also provides *decision nodes* for modeling alternatives and *utility nodes* for the utility evaluation. A decision node indicates a decision facing the decision maker—similar to decision nodes in decision trees. An arrow entering a decision node means that the author's decision is made with knowledge or the outcome of the uncertain quantity at the other end of the arrow. The utility nodes represent the utility function of the decision maker. Utilities are associated with each of the possible outcomes of the decision problem modelled by the Influence Diagram. Influence Diagrams are useful and powerful tool for modelling a decision problem. They can be used to model both simple decision problems (only one decision node) and sequential decision problems (more than one decision nodes and utility nodes). The later is also known as *dynamic decision modelling*. In this case, the next decision always depends on the previous decision or states.

While at Guelph, Dorner [2000] built a belief network based on simulation data of running the GAMES model. The nodes and the relationships between the nodes were obtained from GAMES. Based on the universal soil loss equation, GAMES requires as its input: a segmentation of the watershed into individual homogeneous plots of land based on slope and aspect ratio, area of plot, soil type, so-called cropping factor (the erodibility of soil for a particular crop – essentially a land use factor), and precipitation information for the region. The plots are connected in a dendritic drainage network, from which the individual model components in the GAMES model inherit their interconnection (e.g. Figure 1). The conditional probability distributions for the probability model came from Monte Carlo simulations using GAMES. The model provides results at both field and watershed scales and examines soil loss and sediment transport on a seasonal basis. Figure 2 depicts a partial integration of GAMES into the combined model, for a particular cell or plot X from Figure 1.

Figure 1. Approximate layout of the STRAT watershed. The actual layout has 471 nodes each representing a homogeneous plot.

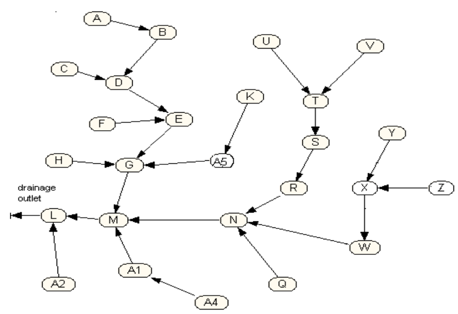
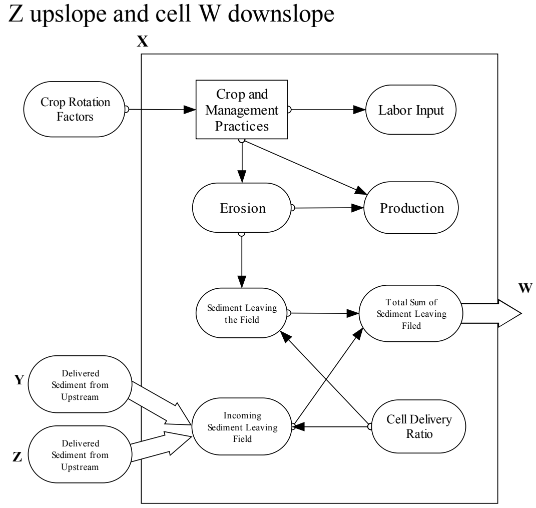


Figure 2. GAMES layout of the plot X in Figure 1(above at lower right), with cells Y and Z upslope and cell W downslope



The watershed examined in this the project was Stratford Avon Upper Watershed (Figure 1), which was part of a comprehensive study for the Great Lakes Pollution from Land Use Activities (PLUARG) in the late 1970s [Wall 1978; Wall 1979]. STRAT is an upland watershed where erosion rates vary significantly over the entire watershed. It has an area of 537 hectares with land slopes up to 9% [Dickinson 1990]. Most of the surface material is comprised of sandy hills.

The STRAT watershed is rural, primarily in cropland. The primary rotations are hay(meadow) / grain / corn, grain / corn or hay / grain. The soil types are loamy soils or clay loam.

6. Utility

In order to determine the desirability of an outcome to the decision maker, a value is assigned to each outcome. The term '*utility*' is used in sense of the quality of being useful. The following equation is used to calculate the expected utility $EU(A|E)$ of action A given evidence E.

$$EU(A|E) = \\ = \Sigma_i \, (P(Result_i(A)|E,\ Do(A)) * U(Result_i(A)\ ))$$

the extended formula for expected utility after new evidence E_j is:

$$EU(A|E,E_j) = \\ = \Sigma_i \, (P(Result_i(A)|E, Do(A),E_j) * U(Result_i(A)\ ))$$

$Result_i(A)$ are the possible outcome states after executing a nondeterministic action A.
$U(S)$ denotes the utility of state S. $Do(A)$ is the proposition that action A is executed in the current state. An action A will have possible outcome states for $Result_i(A)$, where the index i ranges over the different outcomes. .
Maximum Expected Utility (MEU) is the fundamental idea of decision theory. This means that a decision is rational if, and only if, it chooses the action that yields the highest expected utility averaged over all the possible outcomes of the action.

$$MEU(A|E) = \\ \max_A \Sigma_i (P(Result_i(A)|E,Do(A)) * U(Result_i(A)\ ))$$

7. Experiments

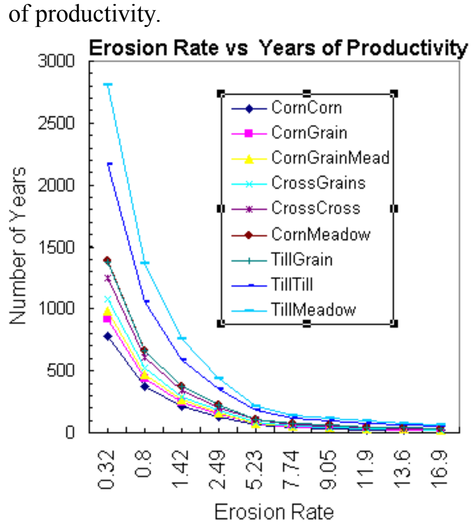
The principal experiments involving **environmental effects** are based on nine yearly crop tillage management rotation scenarios (Table 1) We analyzed the relationships between erosion on various management strategies with the long-term economic income. Briefly, CORN is the North American name for maize using standard non-conservation procedures, GRAINS are cereal crops (eg. wheat), conservation tillage (CONS_TILL) is the technique of planting without plowing, CORN_X_SLOPE refers to corn rows contouring the slope to minimize erosion, and MEADOW refers to leaving a field untilled for a season.
We conducted experiments on **economic**, **environmental** and **integrated** models. The analysis is broadly divided into three parts: local short-term policy; local long-term policy; and, whole watershed policy analysis. We have produced a prodigious set of results, so only a few examples are discussed here.

Table 1. Crop Rotation Scenarios

Scenario Number	Scenario Name	Abbreviation
#1	CORN-CORN	CornCorn
#2	CORN-GRAINS	CornGrains
#3	CORN-MEADOW	CornMeadow
#4	CORN_X_SLOPE-GRAINS	CrossGrains
#5	CORN_X_SLOPE-CORN_X_SLOPE	CrossCross
#6	CORN_CONS_TILL-CORN_CONS_TILL	TillTill
#7	CORN_CONS_TILL-GRAINS	TillGrains
#8	CORN_GRAINS-MEADOW	CornGrainMeadow
#9	CORN_CONS_TILL-MEADOW	TillMeadow

For the first experiment, we chose to analyze the results for ten cells from the network selected for a range of to their erodibility. All 471 cells were involved, but we extracted results for the selected cell numbers for detailed examination. We ran nine alternative tillage management scenarios.
The TillTill management practice produces the lowest erosion rate for continuous cropping. If the erosion rate is in excess of a tolerable rate, however, even good policy cannot change the injury to the field. This result verified Pierce *et al*.[1983] that a field with an erodibility higher than 5.0 ton/ha should be put into fallow for recovery. Otherwise the injury is unrecoverable.

Figure 3. Experiment 2, total number of years of productivity.



The next set of experiments were more long-term, and looked at earning potential over long stretches, as well as years of soil productivity left in the fields tested. Figure 3 shows one part of that result, productive years remaining in the field under continuation of the scenarios.

The third set of experiments, on the whole watershed, explored *point policy* versus *range policy* options. For point policy, erosion rates were set, and the system in the field and upslope was adjusted to meet the policy objective. In the range policy options, erosion rates for several fields were left in an acceptable range, and the system made decisions only when the permissible range was exceeded. We determined that the point policy produced more satisfactory results, in general, when compared with the range policy.

The **economic effects** of the various policy options are difficult to represent in the limited space of this paper. In general, field conditions of low erosion represent opportunities for economical production. Therefore the conversion cost for environmental conservation practices is very high, when compared to the same effort for plots with high erosion potential. This follows common sense, when more gain is realized from less effort.

Immediate economic effects are evaluated by noting the difference in yield (translating into revenue less expenses). Long-term economic costs and benefits are calculated from pursuing policy year-to-year to extinction, calculating total differences over a range of revenue models. This latter analysis is still problematic, since the discounting into the future is anyone's guess.

8. Conclusions

A decision tree does not have the necessary structure to contain all the probabilistic information needed to answer value-of-information questions. The joint probability distribution of all the uncertain variables is needed and this is more easily and naturally captured in Influence Diagram. The efficiency and speed using the ID is unchallenged. In our specific problem, the decision network had 12 nodes for each sub-network (field) and a total of over 471 sub-networks for a whole watershed. The utility function had 5 variables and each variable had 3 to 20 attributes (discrete variables) or value ranges (continuous variables). The CPT (conditional probability table) of the utility had 181,440 entries.

If we use a traditional optimization algorithm, it would be a long computation. With ID, the user can interact directly and adjust the complex parameters to get a better decision based on the specific criteria preference.

Research using weighted multicriteria methods reaches conflicting conclusions regarding the alternatives' effects on the economic returns and the environment. Heilman *et al*, [1996] shows that no-till tillage can improve economic returns to the farmer and have positive off-site benefits. But Prato and Fulcher [1996] report that no-till tillage can reduce the sediment yield but produces the lowest net return among all the alternatives. Our result suggests that the conservation tillage management might be a suitable choice of the alternatives for a long-term policy, because it satisfies both environmental and farm's needs. Our analysis indicates that conservation tillage management technique has a limit to potential benefits. If the field's soil condition is too bad (if the erosion rate is already in excess of an appropriate range), there is no technique available to help the field recover from the damage. The tolerance erosion rate in our simulations is 7.5 ton/ha. This agrees with the reported result by Pierce [1983]. If conventional tillage management is used, the field can only last for less than 100 years of cropping with top soil depth 38.5 cm. But with conservation tillage management the profitable years can last more than three times that.

Our research suggests also that we can build a comprehensive model that contains most of the relevant environmental and economic factors for environmentally responsible decision-making.

Acknowledgements

The authors are grateful for the hard work of the session organizers and the referees. The generous support of the Canadian Natural Sciences and Engineering Research Council is acknowledged.

References

Abu-Zeid, K.M., 1996. A multicriteria decision support system for evaluating cropping pattern strategies in Egypt. Proc. First Inter. Conf. on Multiple Objective Decision Support Systems (MODSS): 105-120.

Bessembinder, J.J.E., M.K.V. Ittersum, R. Rabbinge and L.O. Fresco, 1996. Long-term explorations of sustainable land use. MODSS: 437-445.

Bots, P.W.G., Hukshof, J.A.M., 2000. Designing multi-criteria decision analysis processes for priority setting in health policy. J. Multi-Crit. Decis. Anal. 9:56-75

Cornelissen, A.M.G., J. van den Berg, W.J. Koops, M. Grossman and H.M.J. Udo, 2000. Assessment of sustainable development: a novel approach using fuzzy set theory. ERIM Report Series Research in Management: 1-17.

Costa, C.A.B. 2001. The use of multi-criteria decision analysis to support the search for less conflicting policy options in a multi-actor

context: case study. J. Multi-Crit. Decis. Anal. 10: 111-125.
Crosetto, M., S. Trarantola, and A. Saltelli, 2000. Sensitivity and uncertainty analysis in spatial modeling based on GIS. Agriculture, Ecosystems and Environment 81: 71-79
Dickinson, W.T., R.P. Rudra and G.J. Wall, 1990. Targeting remedial measures to control non-point source pollution. Water Resources Bulletin, 26(3):499-507.
Dorner, S., D.A. Swayne, R.P. Rudra, C. Pal and C. Newald, 2001. Integrating parametric uncertainty and modeling results into an advisory system for watershed management. Elsevier. advances in environmental research. 5(4): 445-451.
El-Swaify, S.A. & D.S. Yakowitz, 1996. Environmental Management: Concepts, Approaches, and Applications MODSS: 3-7.
Gratch, J.M., 1995. Sequential inductive learning. *In*: Report (University of Illinois at Urbana-Champaign. Dept. of Computer Science) UIUCDCS-R-95-1908. Urbana, Ill.
Heilman, P., J.L. Leonard, D.S. Yakowitz. 1996. Quantifying economic incentives needed for control of nonpoint source pollution in agriculture. MODSS: 513-526.
Hermanides ,G. and P. Nijkamp, 1998. Multicriteria evaluation of sustainable agricultural land use: a case studey of Lesvos. *In*: Multicriteria Analysis for Land-Use Management. Kluwer: 61-78.
Hipel, K.W., D.M. Kilgour, L.Fang and X. Peng, 1996. Using the decision support system GMCR for resolving conflict in resources management. MODSS'96: 23-46.
Howard, R.A. and Matheson, J.E., 1984. Influence Diagrams. Readings on the Principles and Applications of Decision Analysis. Menlo Park, CA: Strategic Decision Group. Vol. II: 719-762
Huylenbroeck, G.V., 1996. A multicriteria approach for trade-off analysis between economic and environmental objectives in rural planning. MODSS: 419-435.
Janssen, R., 2001. On the use of multi-criteria analysis in environmental impact assessment in the Netherlands. J. Multi-Crit. Decis. Anal. 10:101-109.
Marangon, F. & Tempesta, T., 1998. Rural landscape and economic results of the farm; a multi-objective approach. *In*: Multicriteria Analysis for Land-Use Management.. Kluwer:49-60.
Noghin, V., 1997. Relative importance of criteria: a quantitative approach. J. Multi-Crit. Decis. Anal. 6:355-363
Parton. K.A and A. Nissapa, 1996. Use of goal programming in aquaculture policy decision making in southern Thailand. MODSS 1996, pp 641-653.
Pierce, F.J., W.E. Larson, R.H. Dowdy, W.A.P. Graham, 1983. Productivity of soils: Assessing long term changes due to soil erosion. J. Soil Water Cons. 38(1):39-44.
Prato, Tony, Chris Fulcher, Shunxiang Wu, and Jian Ma. Multiple-Objective Decision Making for Agroecosystem Management. *Agricultural and Resource Economics Review* 25: 2 (1996): 200-212.
Rao, K.S. and R. Pant, 2001. Land use dynamics and landscape change pattern in a typical micro watershed in the mid elevation zone of central Himalaya, India. Decision Sciences Web Journals. Agriculture, Ecosystems & Environment. 86:(2) 113-124.
Rudra, P.R., W.T. Dickinson, D.J. Clark and G.J. Wall. 1986. Games – A Screening Model of Soil Erosion and Fluvial Sedimentation in Agricultural Watersheds. Canadian Water Resources Journal, 11(4):58-71.
Shachter, R. 1986. Evaluating influence diagrams. Operations Research, 43: 871-882.
Swayne, D.A, 2003. Putting the decision in decision support system. Proc. IEMSs: 45-52. A. Rizzoli and A.J. Jakeman (eds).
Tan, Y.S. and N.M. Fraser, 1998. The modified star graph and the petal diagram: two new visual aids for discrete alternative multicriteria decision making. J. Multi-Crit. Decis. Anal. 7: 20-33
Tatman, J.A. and Shachter R.D., 1990. Dynamic Programming and Influence diagrams. IEEE Transactions on System, Man and Cybernetics, 20(2):365-379
Wall, G.J., L.J.P.V. Vliet, and W.T. Dickinson, 1978. International reference group on great lakes pollution from land use activities: Contribution of sediment to the Great Lakes from agricultural activities in Ontario. IJC technical report re: PLUARG, International Joint Commission, Windsor, Ontario.
Wall, G.J., W.T. Dickinson and L.P.J. van Vliet, 1979. Agricultural sources of fluvial suspended sediments. Progress in Water Technology, 11(6):481-499
Wang, M. and Yang, J., 1998. A multi-criterion experimental comparison of three multi-attribute weight measurement methods. J. Multi-Crit. Decis. Anal. 7: 340-350
Warburton, A., 1998. The decision tree polytope and its application to sequential decision problem. J. Multi-Crit. Decis. Anal. 7: 330-339
Yakowitz, D.S. A Multiattribute Tool for Decision Support: Ranking a Finite Number of Alternatives. MODSS: 205-215.