A Bayesian network for investigating the decline in fish catch in Switzerland

Mark E. Borsuk, Patricia Burkhardt-Holm, and Peter Reichert
Swiss Federal Institute for Environmental Science and Technology (EAWAG)
8600 Dübendorf, Switzerland
(mark.borsuk@eawag.ch)

Abstract: Catches of brown trout have decreased about 50% in many rivers and streams in Switzerland in the past 15 years. Additionally, the health status of numerous brown trout populations has been assessed to be impaired. In order to evaluate the causes for these phenomena, a nationwide interdisciplinary project named "Fischnetz" was launched in 1999. Twelve hypotheses for the fish population declines were proposed and laboratory and field research projects were initiated to investigate these suggested causes. To apply the results of these investigations to the task of discerning the relative causal importance of each of the hypotheses, a Bayesian probability network is being developed. The development of a "Bayes net" begins with eliciting mental models about the cause-and-effect relationships among system variables from subject-matter experts. Represented as a graphical network, these models imply a set of assumptions about the conditional dependencies among the variables, which simplifies the problem of working with imprecise knowledge. Hard-to-derive joint probability distributions are replaced by a set of conditional distributions, which can be characterized using either: (1) experimental investigation, (2) collected field data, (3) process-based models, or (4) elicited expert opinion. Such information, available as a result of the "Fischnetz" research program and from the scientific literature, will be integrated into the network, thus quantitatively summarizing all relevant information. The quantified network will then be used to assess the historical causal importance of anthropogenic changes, as well as predict the effect of proposed management actions. Analyses will be carried out for individual streams using site-specific information as evidence to update less specific prior beliefs. The results can be used form the basis for preliminary management and to prioritize future research projects based on their ability to reduce uncertainty in model-based assessments. In this paper, a first prototype of the network is presented and the methodology for its construction and application is discussed.

Keywords: causal attribution; probabilistic modelling; expert elicitation; decision support

1. INTRODUCTION

Several indications over recent years have suggested that fish populations in many Swiss rivers and streams have experienced serious declines. Annual catch records of anglers indicate a decrease of up to 50 % since the 1980s [Friedl 1999]. These declines are especially apparent in the more anthropogenically impacted midlands and northern regions of Switzerland. Of the most commonly caught fish species, brown trout *(Salmo trutta fario)*, grayling *(Thymallus thymallus)*, and nase (*Chondrostoma nasus*) are those with the greatest declines [Frick et al. 1998]. In parallel with the indications of decreasing fish catch, fish health monitoring since the 1980s has produced evidence of an impaired health status of native species. Brown trout with both macroscopic visible lesions and histopathological tissue alterations have been observed in a number of Swiss streams [Wahli et al. 1998, Bernet et al. 2000]. The causes of the widespread health problems and decreased abundance of fish in Swiss rivers are not readily apparent.

In January 1998, representatives of the cantonal fisheries administrations, the Swiss federal environmental administration, and several research institutions, met to discuss the observed problems in Swiss fish populations. As a result of this meeting, a nationwide research network named *"Fischnetz" ("Netzwerk Fischrückgang Schweiz")* was initiated [Burkhardt-Holm et al. 2002].

Among the goals of the Fischnetz project are the following:

- to collect and evaluate available, but scattered, data on the status of Swiss rivers and on fish catches, fish health, and fish populations,
- to improve communication and coordination of relevant individual research activities in various Swiss universities, research institutions, and cantonal and private organizations,
- to initiate new research activities wherever significant gaps in information are identified, and
- to use the resulting research findings to identify the most important causes of the present situation and consider opportunities for improvement.

To achieve these integrative aims, a method is required to combine quantitative data and qualitative knowledge into a coherent analytical framework. We have found causal Bayesian networks [Pearl 1988] to be one of the most promising methods for performing such types of ecological assessment and prediction [Borsuk et al. 2002]. These models focus attention on cause-and-effect relationships of direct scientific or policy relevance and then represent the effects of remaining influences with probabilistic expressions. In this paper, we describe Bayesian networks and their relevance for the Fischnetz project. A preliminary causal model is then presented based on twelve working hypotheses currently being used to organize the Fischnetz research. This network is then expanded using results from the published literature and causal mental models elicited from Fischnetz project coordinators. The assumptions underlying this network allow the complex chain linking anthropogenic causes to ecological effects to be factored into an articulated sequence of conditional relationships. Each of these relationships can then be quantified independently using an approach suitable for the type and scale of information available. Although quantification has not yet occurred, we describe anticipated methods, results, and suggested uses of the final integrated model.

2. CAUSAL BAYESIAN NETWORKS

Fundamental to developing and using Bayesian networks is viewing the model as a graph. In the graph, rounded nodes represent important system variables, and an arrow from one node to another indicates a dependent relationship between the corresponding variables. Such networks can be easily drawn using conventional scientific notions of cause-and-effect. The interesting point that is made explicit in the graph is the conditional independence implied by the *absence* of connecting arrows. These independencies allow each relationship indicated by the *presence* of an arrow to be quantified separately, perhaps based on disparate forms of information [Reckhow 1999]. Quantification of these relationships consists of parameterizing conditional probability distributions that reflect the aggregate response of each variable to changes in its "up-arrow" predecessor, together with the uncertainty in that response.

Conditional probability relationships may be based on either: (1) experimental investigation, (2) collected field data, (3) process-based models, or (4) elicited expert judgment. Observational field data that consist of precise measurements of the variable or relationship of interest is likely to be the most useful and least controversial, form of information. Unfortunately, appropriate and sufficient data may not always exist. Experimental evidence may fill this gap, but concerns may arise regarding the applicability of this information to the natural, uncontrolled system, and appropriate experimental data may also be limited. As a consequence, the elicited judgment of scientific experts may be required to quantify some of the probabilistic relationships. Established techniques exist for performing these elicitations [e.g. Morgan and Henrion 1990, Meyer and Booker 1991], and help to assure accurate and honest assessments.

Once all relationships in a network are quantified, probabilistic predictions of model endpoints can be generated conditional on certain values for "up-arrow" causal variables. These predicted endpoint probabilities, and the relative change in probabilities between alternative scenarios, convey the magnitude of expected system response to historical changes or proposed management while accounting for predictive uncertainties. In addition to prediction, probability networks can be used to perform probabilistic inference when observations of certain model variables are made. Inference is the process of probabilistically estimating the value of all other variables (or distributional parameters) in the network given the values for the observed variables. Inference is useful for updating probabilities based on new observation and for assessing the likely cause of an observed event (fish declines, for example) when data on causal variables is not available. These tasks are particularly valuable when additional monitoring is likely to occur concurrent with the management effort. Statistical inference involves the use of Bayes theorem, thus the term *Bayesian* network (which also reflects the use of subjective probabilities, another defining trait of Bayesian analysis).

3. GRAPHICAL NETWORK FOR FISH CATCH DECLINE

The participants in the Fischnetz project have formulated twelve hypotheses on possible causes of the observed decline in fish catch [Burkhardt-Holm et al. 2002], as follows:

1. The decline is due to more than one of the factors that follow, with each factor having a different significance depending on the geographical region involved.
2. The fish population is suffering from reproductive failure of adult fish.
3. The fish population is suffering from reduced recruitment of young stages.
4. The health and fitness of fish is impaired.
5. Chemical pollution (both nutrients and synthetic compounds) is having harmful effects.
6. Poor morphological quality and longitudinal connectivity is having an impact.
7. Fine sediments are at harmful levels.
8. There is an insufficient amount or quality of food available.
9. Fisheries stocking and management has had negative impacts.
10. Excessive removal by birds or anglers, or changed angler behavior, is responsible for apparent declines.
11. Water temperature patterns have changed in harmful way.
12. Hydrological regime and sediment transport have been detrimentally modified.

These hypotheses include cause-and-effect relationships at multiple levels, so there is some degree of overlap and interaction among them. These dependencies can be made explicit by using the graphical representation of Bayesian networks (Figure 1). This representation makes it clear that these hypotheses do not relate to separate causes, but rather to an inter-dependent set of factors operating at various points in the causal chain. Therefore, unless the decline in fish catch can be attributed entirely to excessive fish removal (an unlikely conclusion, given simultaneous declines in fish health), then hypothesis #1, stating that multiple factors are involved is undoubtedly correct. Of course the question then becomes, “What is the relative causal importance of each factor?”

Quantitatively assigning causal attribution and identifying suitable management actions requires a more refined network than that shown in Figure 1. Specifically, a sufficient number of intermediate variables must be included so that the impacts of specific causes are separately identifiable from

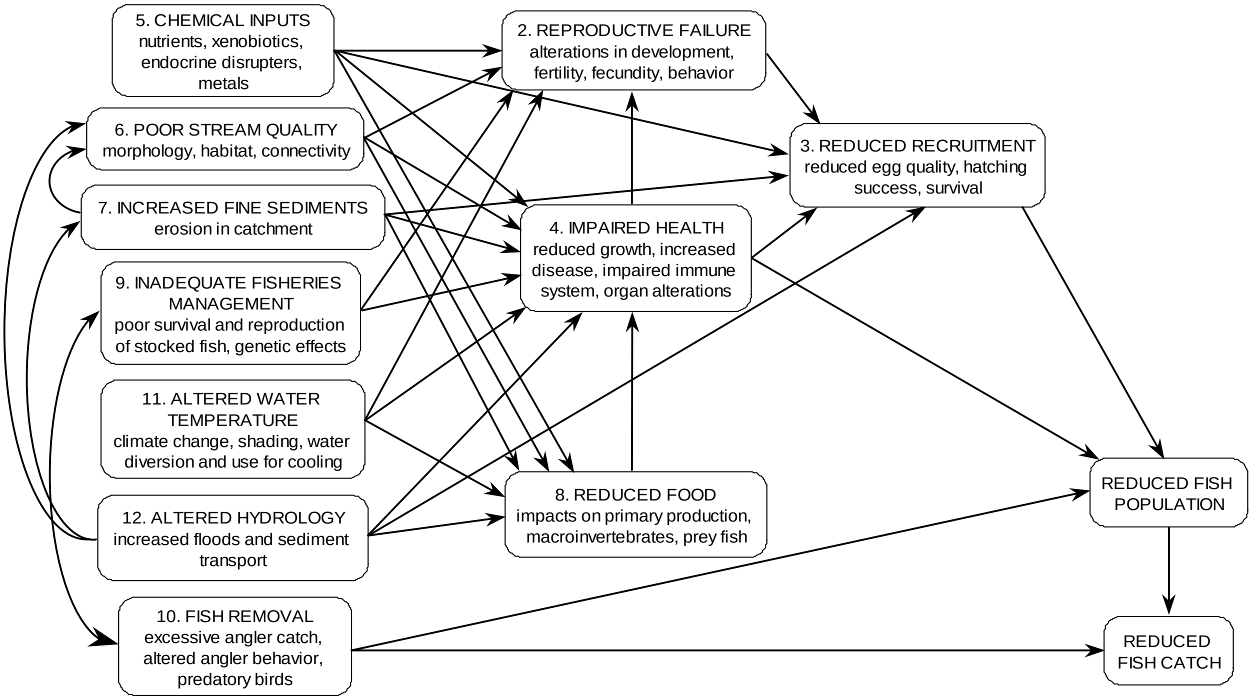

Figure 1. Graphical model of the relationships among causal hypotheses and effects. Numbers correspond to the numbered hypotheses in the text.

data [Pearl 2000]. Additionally, further decomposition of the causal chain may make it easier for experts to apply their knowledge of specific processes to the assessment of conditional probabilities [Morgan and Henrion 1990]. Using a combination of a published network describing the general factors determining the population viability of resident salmonids [Lee and Rieman 1997, Lee et al. 2000] and the expert opinion of Fischnetz project coordinators, a more detailed causal network was developed for application to Swiss streams (Figure 2).

As discussed above, arrows between nodes in a Bayesian network represent conditional dependencies, generally interpreted as causal relationships. For example, in Figure 2 reproduction (middle, left) is shown as being dependent on fecundity and spawning/incubation success, which, in turn, are causally dependent on the variables chemical water quality and health status, and the variables water quality, stream flow, sediment load, and bed structure, respectively. The relationship between these variables raises an interesting point that is made explicit in the graph: once the values for spawning/incubation and fecundity are known, the antecedent variables are not needed to estimate reproduction. Thus, quantification of separate parts of the causal chain can occur independently using suitable data or expert opinion. This aspect of Bayesian networks significantly facilitates their use for integrating the results of multi-team, multi-disciplinary research projects such as Fischnetz.

The construction of the graphical network shown in Figure 2 represents the current state of development of the modeling component of the Fischnetz project. Anticipated future progress and results are discussed in the next section.

4. ANTICIPATED MODEL DEVELOPMENT AND RESULTS

With the basic structure of the model determined (Figure 2), attention turns to the development of the conditional probabilities characterizing the dependencies among variables. These relationships should be sufficiently general so that they can be applied to the range of conditions found in Swiss streams (as described by the outermost, or “marginal”, nodes). A probabilistic, age-structured population model, recast as a Bayesian network, has been developed by researchers at the U.S. Forest Service [Lee et al. 2000] and can be used to relate the nodes that are shaded in Figure 2. However, the inputs to this model (the outermost row of shaded nodes) must

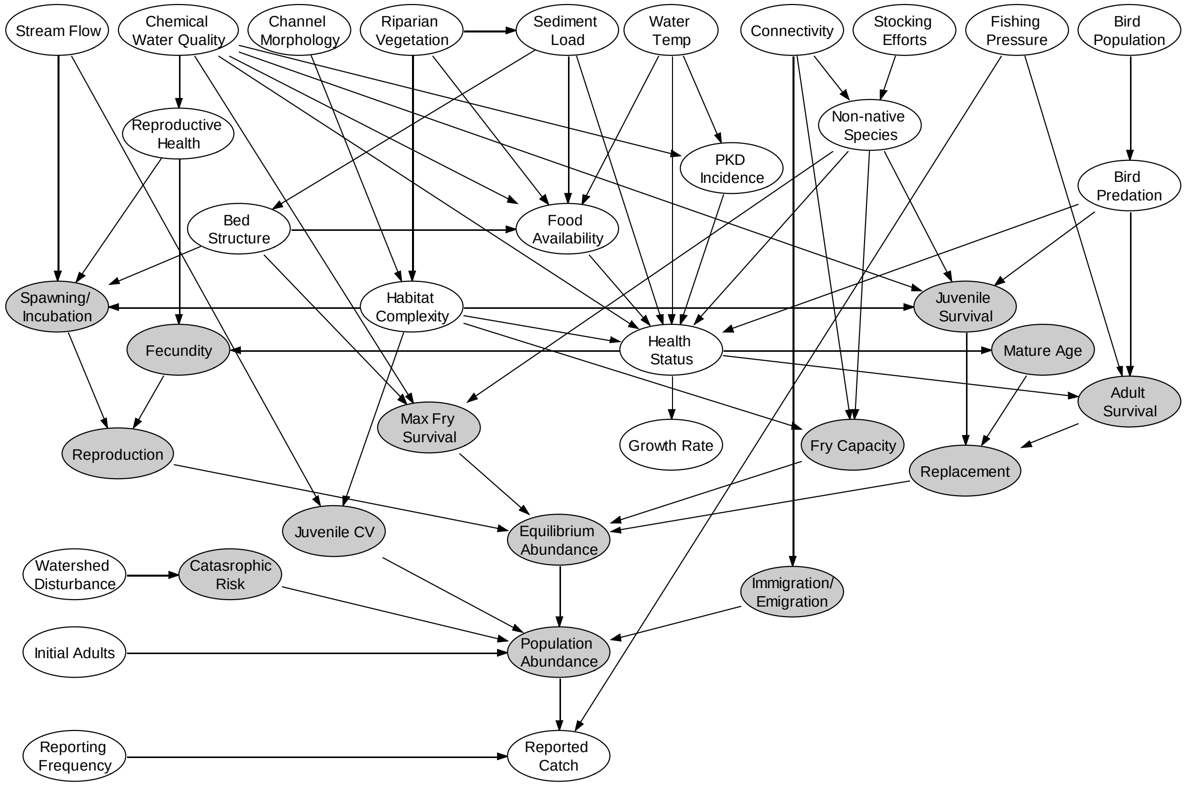

Figure 2. Preliminary causal Bayesian network used to predict fish catch. Shaded nodes indicate variables included in the population model of Lee and Rieman [1997] and Lee et al. [2000].

still be related to the environmental and anthropogenic factors that represent the root causes of fish catch decline, and which may differ across streams. This is the point in the analysis where the data and experience resulting from the Fischnetz projects are expected to be most valuable. For example, a recent subproject of Fischnetz investigated whether exposure to polluted river water may led to increased incidence of organ alterations and infectious disease [Schmidt-Posthaus et al. 2001]. We anticipate that the results of this study will be used to characterize the relationship between water quality and health status shown in Figure 2. Other relationships, such as that between bed structure and fry survival, may not have been directly investigated as part of Fischnetz. However, scientists within the Fischnetz project possess a great deal of knowledge gained from field and laboratory experience and literature review. This knowledge can be quantified using methods of expert elicitation and used to generate preliminary estimates of model relationships, which can later be revised as appropriate data are collected. In fact, by including estimates of uncertainty, scientists are able to account for many factors driving variability that may be difficult or impossible to address empirically, as well as communicate the degree of confidence they have in their assessments. For this reason, we expect formally assessed subjective probabilities to be a major component of our network.

Once all conditional probabilities in the Bayesian network are characterized, predictive analyses can be performed for each fish population of interest. This is done by assigning appropriate values to each marginal variable describing local watershed and stream conditions and anthropogenic influences. Probability distributions for all the intermediate variables and the model endpoint, reported fish catch, are then computed using the entire set of conditional relationships. If marginal variables are sequentially set to values corresponding to present and historical conditions, then the difference in endpoint values between these two scenarios conveys the historical causal importance of the factors that have changed. The magnitude of this importance depends on both: (1) the magnitude of change experienced by the causal factor and (2) the net sensitivity of the endpoint to changes in this factor. Similar comparative analyses can be performed to assess the impact of proposed management actions; the scenario representing historical conditions is simply replaced with one representing future "improved" conditions, and comparisons are again made with the present situation.

We expect to apply the model to the full set of midland stream reaches to assess the broad-scale importance of the various causes. To obtain the data required for such an extensive analysis, we will rely on the participation of cantonal authorities familiar with local stream conditions. We anticipate sending out a questionnaire or holding a series of meetings to determine appropriate values for the model inputs. GIS-based data obtained from the federal environmental administration may also be used in this process. In locations that have been more intensely monitored, data may exist to characterize intermediate, as well as marginal, variables. Inclusion of this information in the network can be expected to reduce predictive uncertainty by shortening the causal chain leading to final effects.

Despite our best efforts at data gathering, quantitative data may not be available to characterize all marginal nodes for all stream reaches. In these cases, the values of these variables can be represented by probability distributions based on the range of values observed in similar streams for which data do exist. Alternately, marginal distributions can be elicited from scientists or authorities familiar with conditions in the stream and its watershed. Values might also be inferred from observations of intermediate variables, if available, using the process of Bayesian inference mentioned above. These streams can then be targeted for future data collection, if the preliminary analysis suggests that they are impaired.

5. CONCLUSIONS

Causal Bayesian networks are well-suited for the integrated modelling necessary to assess ecological consequences potentially caused by many factors. The graphical model explicitly represents cause-and-effect assumptions between system variables that may be obscured under other modeling approaches. This representation allows biologists, hydrologists, chemists, and ecotoxicologists, to all see how their research contributes to a quantitative synthesis. The flexibility of the method allows for multiple forms of information to be used to quantify model relationships, including formally assessed expert opinion when quantitative data are lacking. The probabilistic nature of the predictions promotes risk-based decision making and facilitates prioritization of future research and monitoring efforts. The result is an integrative, transferable model that can be used effectively within the existing environmental management process.

6. REFERENCES

Bernet, D., H. Schmidt, T. Wahli, and P. Burkhardt-Holm. 2000. Effects of waste water on fish health: an integrated approach to biomarker responses in brown trout (Salmo trutta L.). Journal of Aquatic Ecosystem Stress and Recovery **8**:143-151.

Borsuk, M. E., C. A. Stow, and K. H. Reckhow. 2002. Ecological prediction using causal Bayesian networks: A case study of eutrophication management in the Neuse River estuary. In preparation.

Burkhardt-Holm, P., A. Peter, and H. Segner. 2002. Project Fishnet: A balance between analysis and synthesis. Aquatic Sciences **6**: In press.

Frick, E., D. Nowak, C. Reust, and P. Burkhardt-Holm. 1998. Der Fischrückgang in den schweizerischen Fliessgewässern. Gas Wasser Abwasser **4**:261-264.

Friedl, C. 1999. Fischfangrückgang in schweizerischen Fliessgewässern. Swiss Federal Office of Environment, Forests and Landscape, Bern.

Lee, D. C., and B. E. Rieman. 1997. Population viability assessment of salmonids by using probabilistic networks. North American Journal of Fisheries Management **17**:1144-1157.

Lee, D. C., B. E. Rieman, and W. Thompson. 2000. Bayesian Viability Assessment Module (BayVAM): A tool for investigating population dynamics and relative viability of resident and anadromous salmonids. USDA Forest Service, Boise, ID.

Meyer, M., and J. Booker. 1991. Eliciting and Analyzing Expert Judgment: A Practical Guide. Academic Press, London.

Morgan, M. G., and M. Henrion. 1990. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press, Cambridge.

Pearl, J. 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo, CA.

Pearl, J. 2000. Causality: Models, Reasoning, and Inference. Cambridge University Press, Cambridge, UK.

Reckhow, K. H. 1999. Water quality prediction and probability network models. Canadian Journal of Fisheries and Aquatic Sciences **56**:1150-1158.

Schmidt-Posthaus, H., D. Bernet, T. Wahli, W. Meier, and P. Burkhardt-Holm. 2001. Morphological organ alterations and infectious diseases in brown trout (Salmo trutta) and rainbow trout (Oncorhynchus mykiss) exposed to polluted water. Diseases of Aquatic Organisms **44**:161-170.

Wahli, T., W. Meier, H. Segner, and P. Burkhardt-Holm. 1998. Immunohistochemical detection of vitellogenin in male brown trout of Swiss rivers. Histochemical Journal **30**:753-758.