BAYESIAN NETWORK MODELING OF FISH FARM BENTHIC INFLUENCE - A DESIGN SCIENCE STUDY IN ENVIRONMENTAL MANAGEMENT

Bjørnar Tessem, Dept. of Information Science and Media Studies, University of Bergen, bjørnar.tessem@uib.no

Abstract. Fish farming in Norwegian waters is a heavily regulated activity where the aim of regulation is to reduce the environmental impact of fish farms. The most prominent environmental consequences of fish farming are escapes and organic waste from feces and feeding. The Norwegian Directorate of Fisheries (DoF) is responsible for the supervision of salmon production at the fish farms through data collection and analyses and, based on this, the implementation of measures on the actors in the business. In this paper, we present the initial development of a Dynamic Bayesian Network model describing the dynamics of salmon feeding and production, and how they influence levels of organic enrichment at the individual facilities. The aim of this modelling process is to develop a model supporting decision making regarding how to focus resources and activities within DoF. However, as there is a lack of data available from the fish farms, the research also has the purpose of helping to identify which types of data DoF must require from the fish farming industry to ensure efficient governance. The project is thus a design science project where the design and construction of a computational model for decision making also helps to identify needs in the organization, influences stakeholders, and potentially provides a more efficient use of resources to both the government and the fish farming business.

Keywords: Bayesian networks, fish farming, benthic organic enrichment, design science, information systems

1 INTRODUCTION

Fish farming in Norway and other countries is a commercial activity struggling with pollution and other environmental impacts. The Directorate of Fisheries (DoF)\(^1\) is one of three public institutions that regulate the activities of the Norwegian fish farming industry. One central responsibility is the reduction of the environmental effects of open net pens, which have become the standard production platform throughout the industry. The main effects of this type of fish farming are twofold: the genetic impact of escaped farmed salmon and trout on native populations, and the organic enrichment of the sea floor resulting from feeding and feces from the farmed fish. In this article we focus on the establishment of new technological solutions to help in the surveillance of fish farm’s organic foot prints.

Bayesian networks (BN) is a technology that has been helpful in many instances of environmental modeling. A review of their application is presented in Aguileira et al. (2011). Of particular interest is the work by Giles (2008) in modelling the benthic (sea bottom ecology) impact of fish farming. It has also been used by DoF before, as Bayesian networks have been used to guide inspections in the management of wild fish resources (Tessem et al, 2009, Tessem, 2013). BN become particularly relevant for this domain because of the challenge that the data collected in fish farming contains variables that are measured imprecisely and with several sources of uncertainty. Simultaneously there exists substantial knowledge regarding the interactions between the pollutants from fish farms and the environment. This combination of well established knowledge in an area with uncertainty in the information sources opens for the use of Dynamic Bayesian Networks (DBN), as models of how a system of probabilistic variables changes over time. Data from the modeled system are entered into the model, and the model uses these data to provide probabilities for non-measured variables through mathematically well-founded algorithms.

The goal of the research presented here is twofold. First, it aims at building BN models that will help DoF to identify or predict the fish farms that have or will have problematic levels of organic enrichment (i.e. at

\(^1\) http://www.fiskeridir.no/english
the level where it classifies as pollution). These models should build on existing knowledge of how organic waste influences the benthic environment, opening for the use of dynamically entered data about production and organic enrichment measurements. Second, such models will help DoF to identify which types of data that should be collected from the actors in fish farming. These data should, in contrast to how they are reported today, be delivered in machine-readable form, with the aim of being able to run the Bayesian network models in an automatic manner.

This research fits into the design science research (DSR) perspective within information systems research (Hevner and Chatterjee, 2010). The aim is to construct information system artifacts for an organization and evaluate those in a rigorous manner. The goal is to obtain more knowledge on how such artifacts can be built and used, but also to assess the impact and value to the organization of such artifacts. A future evaluation strategy would involve the analysis of interviews with experts from DoF, the fish farming industry, environmental consultancy groups, and biologists, combined with experiments comparing system predictions to measurements of actual environmental status. The DSR contributions would include design insights of knowledge-based uncertainty models to support decision making in environmental governance, as well as an assessment of evaluation methods for the knowledge models used in a decision support system where quantitative data may be scarce and/or uncertain.

The next section provides an introduction to the fish farming industry in Norway and the problems associated with benthic organic enrichment. We also present the Bayesian network technology employed in the modeling part of the project. The following section describes the modeling process and the results as part of a design science process, including initial evaluation steps. We conclude with a discussion of the project’s status so far and further work to be done.

2 BACKGROUND

2.1 FISH FARMING AND DOF

The level of organic enrichment of fish farm facilities (localities is the term used in public documents) is normally measured through the use of “MOM-B tests” (Hansen et al., 2001). The MOM (Modelling-Ongrowing fish farms-Monitoring) programme consists of three tests: MOM-A, which is a simple measurement of sedimentation rate, MOM-B, which is a more thorough sediment investigation directly beneath the fish farm, and MOM-C, which is a comprehensive investigation on the effect of fish farms on the benthic fauna in a larger area around the fish farm. MOM-A tests are not reported to DoF, and MOM-C are rarely performed due to their cost and focus on long-term and medium-distance effects. MOM-B tests are conducted most often, normally about once a year in production periods. They are comprised of grabs of benthic substance at the localities, collected manually from small boats. Each grab contents are analyzed chemically and perceptually to provide a coarse assessment of the benthic status. This data collection method brings with it several sources of error:

1. The grabs may hit rocks or underwater knolls that rise up from surrounding polluted areas. These grabs will often contain no material and will therefore be counted as no organic enrichment according to the MOM-B specification. This is particularly problematic in Norway as the sea bottom is normally very rugged.

2. The required number of grabs in one test is 10, which means that large parts of the bottom area are not covered, in particular for large localities. Furthermore, the grabs are not done at the same location from one MOM-B test to the next.

3. These tests are expensive and are for many fish farms not done in a systematic manner, often with more than a year between each test.

4. The perceptual analysis is based on individual interpretation and will vary considerably from person to person. One possible source of error is that the analyst may want to give a “positive” interpretation of the grabs to ensure that the locality owner is satisfied and will buy analysis services again.

The results of a MOM-B test are graded on a qualitative scale of 1 to 4. 1 indicates a very good status, 2 indicates a good situation, 3 means bad, and 4 translates to very bad. The final scores are a conversion
from real valued scores (indices) ranging from 0.0 to 5.0 on the chemical and perceptual scores. Normally a fish farm is required to take actions if a MOM-B test shows levels 3 or 4 of organic enrichment.

To reduce levels of organic enrichment, all facilities are obliged to have a period of fallowing after a production period. During this period (and also during production) benthic animals feed on the organic matter, returning the organic matter from the benthic to the sea water ecosystem. At the same time, ocean currents disburse the organic enrichment over a larger area while no new waste is produced, in a sense refreshing the locality so that new production can be initiated. Fish farmers consider fallowing a useful practice as it ensures better water quality for the fish, resulting in improved fish health.

As mentioned, DoF does not receive continuously updated information about the benthic status of the localities, which could help to identify localities with particular problems. But when MOM-B tests are received, and they show high levels of pollution at a fish farm, DoF will take actions, according to standard practices. Such actions includes requiring the locality to do a new MOM-B test, stopping production (very rare), or enforcing a prolonged period of fallowing. Still, many fish farms reach high levels of pollution during production, and this is not reported due to the lack of conducted MOM-B tests. This implies that these fish farms are actually producing more than permitted. It is in DoF’s mandate to focus their efforts to improve on the practice of those fish farms. But for the DoF and its field inspectors it is, in this situation of missing status data, difficult to identify the fish farms that may be the problematic ones, and further to initiate actions in order to reduce dangers of pollution.

It is in this context DoF recognizes the need for an improved system for assessing the status of benthic organic enrichment of the fish farm localities around Norway. The chosen approach is to develop technology that is able to use data collected from the fish farms and perform automated analyses of the data to produce an updated assessment, supported by a knowledge-based model of the ecological system. Currently, some of the data sources are easily accessible from databases, e.g. results of MOM-B tests, estimated biomass of the localities each month, amount of feeding each month, and weather data. Fish farms are also required to deliver production plans, indicating planned levels of bio-mass and feeding for the next production period. Other data, including details from the MOM-B tests and currents measurements at the locality are available from pdf documents or digital copies of paper documents. These are non-standardized formats and therefore difficult to extract data from. The goal is that the technology shall use data that are more easily processable than those found in pdf documents, and also to include new data that in the future will be required from the fish farmers to be delivered in an easily processable format. This will allow DoF to estimate the ecological status, and further suggest actions needed to ensure less risks of pollution. The aim is to avoid an increase in costly and unnecessary MOM-B tests, but still have control over how the localities develop regarding benthic organic enrichment. In line with the openness policies of Norwegian government, there should be a possibility for the fish farms to run the knowledge-based models, perhaps experimenting with production plans, in order to gain a better understanding of how their locality’s benthic status might develop.

DoF has some experience with similar kinds of problems and has implemented solutions to guide inspections in the management of wild fish resources (Tessm et al. 2009, Tessem, 2013), through the application of Bayesian Networks (BN). As an introductory measure to address the issues surrounding fish farm benthic organic enrichment, DoF aims to use Bayesian networks also in this case as a fundamental knowledge-based technology for their new system, and integrate BN with other software for controlling the benthic status. The tools developed should support DoF’s processes on

- Identifying localities with a high probability for unacceptable pollution
- Estimating the length of fallowing periods
- Assessing the potential for higher production levels
- Identifying localities with unexpected status (as reported in MOM-B tests)

This project fits well within a design science approach (Hevner and Chatterjee, 2010) as it focuses on issues in a particular organization, aims to provide innovative design of information system solutions, focuses on evaluating the technology developed, and contributes to new regulations for data collection in the industry. The validated findings may have consequences for aquaculture and regulatory practices in the industry directly, and has potentially some general relevance to design science in information systems.
2.2 BAYESIAN NETWORKS

Bayesian networks (Korb and Nicholson, 2010) is a technology developed within the artificial intelligence community for the purpose of handling uncertainty in knowledge based systems (or expert systems). A Bayesian network models a domain of interest with a collection of probabilistic variables \((X_1, \ldots, X_n)\). As data (evidence) are entered for some of the evidence variables \(E\), probabilistic dependencies are used to update probabilities for an unobserved variable \(X\). These are obtained from the general use of Bayes rule, formulated as

\[
P(X = x_j | E) = \frac{P(E | X = x_j)P(X = x_j)}{\sum_i P(X = x_i | E)}
\]

where \(x_i\) are the possible values for \(X\). The denominator is in effect only normalizing the posterior probability distribution of \(X\) so it sums up to 1. The reason for usefulness of Bayes’ rule is that we very often will have knowledge about effects of a particular state \(X=x_j\), \(P(E|X=x_j)\), and the prior probabilities \(P(X=x_j)\). But what we observe in practice, for instance, as a medical doctor, are the effects \((E)\) or symptoms of \(X\). Thus, this rule allows us to obtain a posterior probability distribution for \(X\) given incoming observed effect (evidence).

In principle, a Bayesian network needs the same number of probabilities as in a joint probability distribution table for the variable collection with \(n\) variables, i.e., \(O(2^n)\), which normally would make computation with Bayesian networks exponential and in practice intractable for large collections of variables. However, with the use of known or assumed independencies among variables, the need for initial conditional probabilities can be reduced and computation becomes manageable. This is a result of the chain rule of probability, which says that

\[
P(X_1, X_2, \ldots, X_n) = P(X_1 | X_2, \ldots, X_n)P(X_2 | X_3, \ldots, X_n) \ldots P(X_{n-1} | X_n)P(X_n)
\]

With an ordering of the variables exploiting known conditional independencies the probability collected for each variable can be simplified to \(P(X_i | X_{i+1}, \ldots, X_n) = P(X_i | dep(X_i))\) where \(dep(X_i)\) is the set of variables that \(X_i\) is directly conditionally dependent of. From this ordering and the belonging probabilities we construct a directed acyclic graph that contains all the variables of the domain as nodes, and for each node add an incoming edge from \(X_j\) if \(X_j\) in \(dep(X_i)\). In addition, for each node we add the conditional dependency probability (CDP) table \(P(X_i | dep(X_i))\).

The reason for the success of Bayesian Networks is that in many problem domains, experts are able to identify independencies among variables, normally by using knowledge about causality, and thus reducing complexity of the graphs significantly, making computation and storage tractable. Experts may also provide numerical values for the CDP tables. Eliciting expert knowledge is the normal way of constructing such graphs (which are the Bayesian networks described here), but it is also possible to use machine learning algorithms, both to learn the graph’s structure and CDPs (Heckerman, 1999) from data. The simplified probability models that the constructed graphs represent makes propagation of evidence with Bayes’ rule less complex, and efficient procedures for doing this have been devised. Figure 1 shows as an example of a BN a simple model with four variables and the CDPs associated with the variables.
Fig. 1. An example Bayesian Network model for benthic enrichment in fish farms and the conditional probability tables. (Drawn with the Hugin Bayesian network tool)

2.3 DYNAMIC BAYESIAN NETWORKS

Bayesian networks are also applicable to model dynamic processes that run over time. In this approach one models time slices of a domain. Instances of these time slices represent the state of some system at a particular point in time, and full models contain copies of the nodes in the time slice model, one set of copies for each time step. Each time slice contains a set of output nodes, which contain information that influences the status of the next time slice. Output nodes are then mapped to input nodes of the next time step. These input nodes have the same possible outcome values as the output nodes, and evidence propagated through these input node variables contribute to the distributions of the other variables in the next time slice. As for ordinary Bayesian Networks, parameters and structure for DBN can be learned (Murphy, 2002).

Fig. 2. A dynamic Bayesian network with two time slices

2.4 PROCESS FOR BAYESIAN NETWORK MODELING

Chen and Pollino (2012) describe what they consider good practices in Bayesian Network modeling in environmental domains. Among others, they list 6 bullet points to be satisfied for the use of Bayesian networks:

- Domain involves high levels of uncertainty
- Limited/incomplete data on key system variables
- Requiring both quantitative and qualitative information
- Integrating several system components
- Requiring stakeholder participation in the modeling process
- Relationships between variables are non-linear and complex
The domain of this research, fish farm benthic organic enrichment, satisfies all of these requirements, and hence the choice of DBN seems to be a viable attempt to produce better models for this domain. It is particularly important that a Bayesian approach is able to combine both collected data, and the application of expert knowledge in parts of the domain where we have less data or data with low precision. This is in contrast to modeling techniques where you only rely on measured data to draw statistical conclusions about system states (frequentist models), making explicit knowledge about the domain less relevant. Chen and Pollino (2012) also suggest a process for developing the BN, the steps being

1. Build a conceptual model from existing knowledge
2. Establish the structure for the model
3. Establish model parameters (CDP tables) by discretization of continuous variables, and eliciting knowledge from experts and/or data. Assess the possibility for non-discrete variables
4. Reflect about uncertainty and lack of precision in model
5. Evaluate of the model

Developing Bayesian networks is a highly iterative process, and models may be subject to significant changes during development, including being maintained during production. The iterative approach of design science, with continuous and repeated improvement of artifacts, and following evaluations thus matches well with Chen and Pollino’s (2012) process for developing Bayesian networks. In the next section of the article is described how this practice is followed and results in two iterations of a DBN design.

3 MODEL BUILDING

The model building process has gone through two main iterations with modifications and improvements. The model as it stands today is a result of modifications resulting from a first iteration which was then commented on in meetings with experts at DoF. The current model is a DBN model and is shown in Figure 3. It contains several variables that we do not have simple access to data for today, and for these variables, procedures for automated data collection must be established. The model is built using Hugin’s tool for BN modeling, and tests and machine learning has been enabled by using Hugin’s Java interface.

![Dynamic Bayesian network model for benthic state under a fish farm locality](https://www.hugin.com)

Fig. 3. Dynamic Bayesian network model for benthic state under a fish farm locality

---

2 [http://www.hugin.com](http://www.hugin.com)
3.1 ESTABLISHING CONCEPTS

Before we describe the first modeling attempts, it is necessary to describe what happens in a fish farm when it comes to feeding and the effect of organic waste. Salmon are typically placed in open net pens where each contains thousands of individual fish. Normally there are about 10-15 pens at a locality. The fish are fed daily, and eat nearly all of the food offered. However, the salmon also produce feces, which sink to the bottom of the sea, and leads to organic enrichment. On the seafloor there are animals (like Polychaetes) that feed on this, circulating the organic material back into the ecosystem again. However, when levels of organic waste rise too high, these animals are killed by lethal levels of certain compounds (mainly H$_2$S) that develop in the pollution, and the capacity of the locality to restitute from high level of organic enrichment is thereby reduced.

As a first attempt at building a core concept base for the BN, we studied Giles (2008), a review article on studies of the benthic (sea bottom environment) impact of fish farming. Giles has a collected a number of scientific articles on this topic and built a BN that models the effect of production parameters and natural locality parameters on chemical measurements (like sulphide, ammonium, acidity, etc). The input parameters are Fish Density, Farm volume, Water Depth, Max Current, and Average Current. Giles’ model is a central input to our modelling process. However, in discussion with local biologists and other experts, it came up that weather conditions, like an occasional storm that causes resuspension of the polluting material, also have an impact. In addition, a variable like Bottom Current influences the continuous resuspension of organic material. Resuspension is considered to reduce the level of organic enrichment, as it brings oxygen to the polluted areas and distributes the organic matter over a larger area, thus diluting the organic waste. On the other hand, variation in the direction of the current will counter this effect, shifting the organic matter back and forth, and lead to less oxygen-rich water. This again will reduce the transformation of the organic matter and may also lead to lower productivity in the fish farm itself.

A significant issue with Giles’ model is that it is a static model. It does not involve all the dynamic processes that enable the benthic environment to restitute itself, and hence it must be modified in order to fit the purpose and data we have for this project.

3.2 THE FIRST MODEL

We did consider modifying Giles’ model to a dynamic model that could use data from MOM-B tests as well as production and site data. However, at this stage the only data we have easily accessible from the MOM-B-tests are the summarized MOM-B values, which are essentially a qualitative aggregated value. Some chemical parameters are measured, but not as many as in Giles’ model. And the perceptual assessment done by the tester also impacts the final MOM-B value, but is not included in Giles’ model. So there is a mismatch in the data needed by Giles’ model, the data we have available, and finally the lack of dynamics in Giles’ model.

Also, the detailed data of MOM-B tests are only available in non-standardized pdf files, both the chemical measurements and the perceptual. Due to this fact, and the fact that the final assessment is so uncertain (for reasons explained above), it was decided that the first Bayesian network models should only focus on the aggregated MOM-B score and consider this score to be a highly uncertain variable that is influenced by added organic waste, resuspension and ecological recirculation.

As we are here considering a dynamic model, we also need the previous state of the benthic environment, the weather data from the previous month, monthly feeding and biomass, and certain static variables. So, as output node we use the benthic status in one time slice, this again functioning as input to the initial benthic state in the next time slice. Because of the monthly reporting of production data we found that a time slice interval of one month would be suitable. The DBN is shown in figure 4. Notice the auxiliary variables like Effect storm, Effect waste, and Currents. Such auxiliaries are used in many real world applications of BN, as they reduce the complexity of computing, and aggregate the results of many variables into one single effect on a critical variable.

When we compare this model to Giles’ model we have introduced dynamics by having benthic state being impacted by previous month’s benthic state. In addition, we include the idea of resuspension in the
model. Giles’ model is in fact a more elaborate version of the part of our model pointing to the node Effect waste.

Setting the parameters of the first model was done manually using assessments of how much each variable and its value affect the child variables. After the structural modeling and manual parameter initialization, the machine learning algorithm provided by Hugin was run in order to improve values. However, we only noticed a slight change, most likely because we have very few MOM-B tests, so we still lack data to be able to run satisfactory machine learning processes.

Fig. 4. The Bayesian network model resulting from the first round of domain modeling

It is also worth noticing that Bayesian Networks in practice enforce the use of discrete variables, i.e., variables with a finite number of outcomes (except when variables are known to be normally distributed). This leads to the need for a transformation of continuous variables like biomass and feeding into discrete levels. The number that marks the borders between levels is an important parameter when using data to run models and should be set with care, and is a research topic in itself within the Bayesian network community (Dougherty et al. 1995).

The strategy for discretization of Feeding and Biomass chosen here has been to identify values that essentially represent the borders for any given locality based on their production level. We are currently operating with three levels (high, middle, low) for these variables, with borderlines set at intervals of one-third of maximum levels of historical production values. The other (static) variables are discretized using distinct values that are known to be significant in the domain. The approach to discretization in BN is a challenging problem, requiring particular attention, and is an area where we will need to experiment and use ideas from other domains. A promising approach is the dynamic discretization approach by Neil et al. (2007), although this is not yet available in the Hugin tool.

3.3 A SECOND MODEL

The initial organic enrichment model was presented locally at DoF during two meetings as well as in continuous discussions with members of the staff. Their comments were collected and used as input to a new model.

The process led to a restructuring of the model, where a MOM-B test result for a particular month is predicted from three variables: previous MOM-B level, added organic waste, and locality effect (as seen in Figure 3). Added organic waste is computed from biomass, feeding and currents. We also became aware that it is possible to use a numerical, computational model for added organic waste, building on existing results from previous biological research (Stigebrandt et al., 2004). The introduction of locality effect as an aggregated auxiliary was the result of the identification of new variables that influence the dynamics. For instance, if a locality is in a fjord, it is likely to exhibit lower transformation of organic enrichment due to lower oxygen levels in the water. Bottom topography is also introduced as a factor; drop-offs below the locality are considered good, whereas a rugged seafloor with pits that can become potential pollution pools is not good. The second model is the one presented in Figure 3 above.
A run of the Bayesian model for a particular fish farm is found in Figure 5, where we use the production and weather data for that locality in the simulation. It runs from February 2009 to January 2014 and shows an estimated index value for the MOM-B tests. The index from MOM-B tests is actually a real number ranging from 0.0 to 5.0 based on the chemical and perceptual assessments, and is found in the written reports. As our model work with probabilities for the qualitative levels for the benthic state, (1, 2, 3, and 4) each of these are converted into a corresponding numerical index value $mom-es$; taking values 0.6 for 1, 1.6 for 2, 2.6 for 3 and 3.6 for 4. From the computed probability distribution on the benthic state at a time, we estimate a MOM-B index at time $t$ to be the expectation value of $mom-es$. The curve in Figure 5 shows the simulated MOM-B indices. The horizontal lines in the figure show the boundaries for categories of indices. Indices lower than 1.1 gives MOM-B status 1, and then we use steps of size 1.0 to 2.1 for status category 2, etc. The results reported in real MOM-B tests are 0.17 in February 2009, 1.61 in January 2012, 1.42 in August 2012, and 1.31 in August 2013. Notice the model’s increase in levels during production periods, and the return to a healthier state in following periods. The fish farming site, according to the Bayesian network model, was in benthic state 3 in a short period in 2012, which, if true, should have resulted in reduced production. The data points reported, however, was never at this level.

### 3.4 IDENTIFICATION OF DATA NEEDS

Many of the variables are not available in easily machine-readable formats, although they can be collected manually. However, it is the goal of DoF to collect these data from the fish farms, or automatically collect them from currently enforced procedures. Most of the data that are lacking reporting procedures for are static, i.e., they do not change over time, like currents data, the locality in a fjord, depth, etc. As a consequence, the extra costs of enforcing a new reporting regime are not high since much of this data has already been collected. A summary of non-auxiliary variables required and the approach to collecting data for these variables is given in table 1.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Strategy for getting data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fjord</td>
<td>Find locality in vann-nett.no. This is a web service that defines water category at geolocation</td>
</tr>
<tr>
<td>Storm</td>
<td>Use nearby weather stations and get information from met.no, a web service that has weather data for Norway.</td>
</tr>
</tbody>
</table>
### Variable Strategy for getting data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Strategy for getting data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topography, Shallow, Bottom current,</td>
<td>Data that the fish farm should be required to report in machine-readable form when applying for licensing. Meaningful categories need to be defined for the purpose.</td>
</tr>
<tr>
<td>Bottom structure, Current variation,</td>
<td></td>
</tr>
<tr>
<td>Current strength</td>
<td></td>
</tr>
<tr>
<td>Biomass, Feeding</td>
<td>Data that is already reported on a monthly basis</td>
</tr>
<tr>
<td>Benthic state</td>
<td>This is now measured and reported as the aggregate results of MOM-B tests. This is sufficient for the models presented here. In the future, all data from MOM-B tests should be reported in machine readable form, both chemical tests and perceptual assessments. That would allow for a richer assessment of benthic state.</td>
</tr>
</tbody>
</table>

### 3.5 DISCUSSION

So far the design process has led to the Bayesian models presented in this section. They have been tested for several localities with results similar those found in Figure 5. Further, machine learning algorithms have been run, so far with little improvement in simulation results. The current model is the result of a qualitative evaluation of the first model with specialists at DoF’s Aquaculture and Coastal Management Department and led to a richer model that is conceptually easier to understand. The modeling process will be an ongoing activity and new models with new variables and assessments of how discretization should be realized, will potentially influence the value of using these models in practice.

To obtain more complete input for the models we will also include interviews with biology researchers who have extensive experience with fish farming. In order to include decision makers in government and industry we also aim to develop software tools that run the models and provide access to the simulation results without requiring users to understand the underlying BN model. Empirical evaluations of such tools will provide us with a better understanding of the potential of such tools in government and the industry.

There are challenges with machine learning for these models. We do not have large amounts of data on MOM-B tests for any given locality at DoF; tests are reported about once a year for each locality. Machine learning with so much missing data has its problems, and is a significant topic for machine learning research. However, it is known that a couple of the large fish farming companies in Norway have done extensive MOM-B testing at some of their localities without reporting them to DoF. These data could improve the result of machine learning, and steps have been taken to gain access to these data. It is a kind of a paradox in this discussion that it is the issue of lacking data that we are trying to overcome through this project. It should motivate a search for other methods for finding the probabilities of the CDP tables of the Bayesian network.

The completed research project may also have a more general relevance to design science beyond the specific models developed. First, it aims to show how a knowledge-based technique for decision support can be applied in an uncertain domain where scientific knowledge is particularly important (here the scientific knowledge area is benthic environment). Further, we combine this scientific model with regulatory practice into a computational system, which is a topic where DSR results are hard to find. Second, evaluation practices in this project will involve a combination of scientist assessments and quantitative assessments of the precision of the Bayesian models on the one hand, and government and fish practitioners evaluating the decision support system as a whole. There are activities regarding how evaluation in DSR could be put into a more systematic framework (Venable at al., 2012), but evaluation practice has gathered less interest in the DSR community. A more complex evaluation regime like the one planned in this project will not be standard in design science research, but will be required in our domain, and will con-
tribute to insights about how such evaluations can be conducted, and how they should influence repeated design iterations.

4 CONCLUSION

This paper reports on the current status of a design science project in the management of natural resources. The result so far is a designed model that has gone through rudimentary assessments by experts. We are still at a proof of concept level in terms of evaluation, and have verified that Bayesian networks can have a role in the management of the natural environment, and in particular in the public governance of fish farming localities. Through its development we have also gained a better understanding of the kind of data needed for running models of this highly dynamic system.

When it comes to aspects of the model itself that remain to be handled, a suitable discretization for static variables needs to be established. Dynamic variables also need to be discretized and collected regularly to be able to run the models for a longer period. This can potentially be done by dynamic discretization. A significant challenge is the small sets of available data from the essential MOM-B test that measures the state of the system. For governance purposes and the industry, the frequency of MOM-B tests seems to be adequate. However, for machine learning it is difficult to obtain reliable results for model parameters with the current test frequency. Construction of decision support systems based on the DBN model, and evaluation of prototypes of these systems are central activities that remain in this project.

REFERENCES


