Developing Decision Support with Bayesian Networks in Fisheries Surveillance

Bjørnar Tessem*[,1], Joachim Keyser[2], Paul Johe[3] and Sigbjørn Ulvatn[3]

[1]*Department of Information Science and Media Studies, University of Bergen, Postbox 7802, NO-5020 Bergen, Norway*

[2]*Department of Informatics, University of Bergen, Postbox 7803, NO-5020 Bergen, Norway*

[3]*Directorate of Fisheries, Postbox 185 Sentrum, NO-5804 Bergen, Norway*

Abstract: An application of Bayesian networks for decision support in surveillance of mackerel fisheries in the North Sea is presented. The application area involves assessment of each fishing vessel's risk for acting irregularly with regard to release of caught fish, and then combining these risk assessments to suggest vessels on which to concentrate. The knowledge underlying the Bayesian model developed is acquired through extensive communication with the domain experts. Following the approach of information systems design research, we have developed a software tool and an understanding of the organisation's needs in order to effectively realise use of the tool. The prototyping development process used is model-centric in the sense that development of the Bayesian networks drives the other activities in the project, such as establishing data sources and overall system requirements. Through further investigations and development of the Bayesian modelling approach presented here, including applying it to other risk areas, we have the potential to make both the operational practices in Norwegian fisheries surveillance more efficient as well as to improve the quality of their operational decisions. In our continuing efforts we will apply, evaluate and further improve on the work processes established during the development of this prototype.

INTRODUCTION

The use of Bayesian networks and their variants [1, 2] has been a successful approach to handling uncertainty in quite a few decision support applications over the last decades [3]. Much of the progress of this methodology has been motivated by properties of practical applications and has resulted in methods for handling loops in the networks, machine learning techniques for structure and parameters of the networks, and also decision-making approaches based on Bayesian networks.

According to Gupta *et al*. [4], "Decision-making Support Systems (DMSS) are Information Systems designed to interactively support all phases of a user's decision-making process." As this type of system (often under the name of Decision Support Systems (DSS)) has spread in business and other organisations, we have seen a wide variety of research ranging from technological to organisational issues. The architecture of these types of systems has been a main topic in the research, and as early as 1980 Sprague [5] described the architecture of DSS systems in terms of three layers: a database layer, which handles the organisation of data used in the decision making; a model layer, which handles the aggregation and analysis of data; and a user presentation/interaction layer, which gives a decision maker access to the results from the computational model as well as the background data.

*Address correspondence to this author at the Department of Information Science and Media Studies, University of Bergen, Postbox 7802, NO-5020 Bergen, Norway; Tel: (+47) 555 84 103; Fax: (+47)555 89 149; E-mail: bjornar.tessem@uib.no

Gachet and Haettenschwieler [6] point out that so far no commonly agreed upon development methodology for DMSS exists, and show how development methods, to varying degrees, are founded on systems engineering versus decision-making support perspectives. As an alternative to the existing approaches, they suggest a tripartite model for development of DMSSs, with components for general system functionality (container), and knowledge base (content), with a kernel that bridges container and content into a working system. They suggest that this tripartite model is a useful structure that can guide development processes.

In this paper, we present the methods that were used to develop a prototype version of a DMSS that applies Bayesian network models for the use in a complex decision-making problem within the Norwegian Directorate of Fisheries (NDF). As it is a goal for NDF to be able to use its data resources more extensively and, in particular, to support the operational activities of the inspectors, tools are needed for automatic data analysis and presentation, potentially in the form of DMSS. The work we present here is the result of a process involving knowledge engineering of a Bayesian model as well as developing the DMSS itself.

During the project, we have followed a design research approach as summarized by Hevner *et al*. [7]. This is a constructive approach to information systems research, and describes a research method where the researchers construct a new type of information system following an exploratory approach and, from this process, identify knowledge of importance to the IS community. This knowledge may consist of solutions to technical and organisational problems handled through the approach. For example, as we progressed through the NDF project, we handled issues involving computational complexity, structural modelling of the Bayesian

1874-1339/09 2009 Bentham Open

network, building tools for testing the models, identifying need for data sources, and working with the domain experts for the continuous improvement of the models. Knowledge that may have a more general applicability is contained in the evolution of the development process model for our particular DMSS.

We start the presentation with short introductions to Bayesian networks. We then continue with a presentation of the Norwegian Directorate of Fisheries (NDF), following up with a more elaborate description of the design research approach. The activities of the research process are then presented in sections about model building, the simulation tool, and evaluations of the software. We then discuss some of the contributions to Bayesian network knowledge engineering and to DMSS development, before we conclude.

BAYESIAN NETWORKS

Bayesian networks are an approach for handling uncertain knowledge that developed in the artificial intelligence community beginning in the middle of the 1980s [8]. The approach is based on the use of Bayes' rule, which in its simplest form, can be stated as:

$$P(h \mid e) = \frac{P(e \mid h)P(h)}{P(e)}$$

where $P(h|e)$ represents the probability of some hypothesis h given observed evidence e. $P(h)$ represents the a priori probability of h given no evidence at all. We have corresponding meanings for the other expressions in the formula.

The application of independence assumptions among variables is also central. Using this approach, one is able to create a graphical representation of the dependencies among the variables of a domain and to further exploit this graphical structure to make computation of probabilities for the variables more efficient. A thorough explanation can be found in [8] or [9].

A very simplistic model from the fisheries domain that illustrates this concept is given in Fig. (**1**). Assume that we would like to find the probability that a fishing boat has dumped fish into the sea after having brought it on board. This is illegal in Norwegian fishing zones, because it effectively kills off the fish population and thus reduces its potential to remain at a sustainable level.

The model contains four variables: **Dumping** (D) which represents the probability for the fishing boat having, in fact, dumped fish; **Floating fish** (F) which represent the probability for having dead fish floating in the water; **Norwegian** (N) which represents the nationality of the fishing boat; and **Large catch (L)** which represents the size of the last catch that the fishing boat brought on board. The graphical model shown in Fig. (**1**) indicates that there is an independency between **Floating fish** and the **Norwegian** and **Large catch** variables, given the value of the **Dumping** variable. At the same time, this structure also means that the two variables **Large catch** and **Norwegian** are dependent, if we are able to learn something about **Dumping** or **Floating fish**, but not if we do not have this type of evidence (conditional dependency). The arrows (edges) in the model normally indicate direct causal relationships between variables.

On all nodes with incoming edges, we assign probabilities for each of the possible values, given all combination of values for the parents. For all nodes without parents, we assign a priori probabilities for each of the possible values for the variable. The probabilities are given in Fig. (**1**). From the formulas and the model parameters, we compute the a priori probability P(D = yes) (i.e., when we have no evidence) to 0.29.

If we now observe floating fish, then we have an observation that changes the posterior probabilities for all of the other variables, including the one that is of primary interest, namely P(D|F=yes). By application of the algorithms (which are based on the repeated application of Bayes' rule and the dependency relations) for computation of probabilities in Bayesian networks, we get a new probability distribution for all of the other variables in the network. For the technicalities of this computation, refer to [2] or [9]. The value P(D|F=yes) now becomes 0.87.

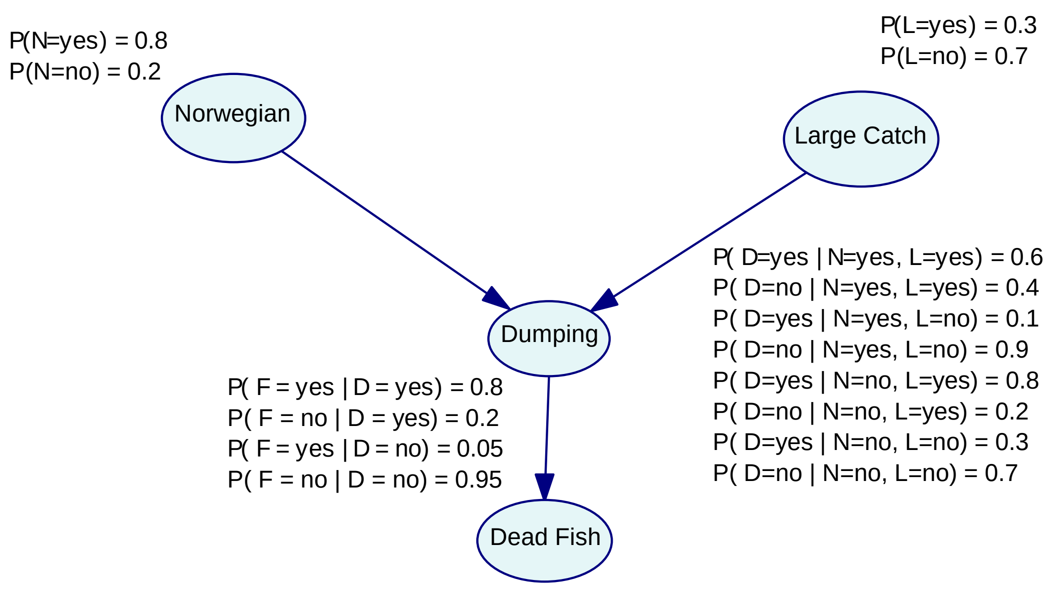

Fig. (1). A simplistic model for probability of fish dumping (Drawn in the GeNIe tool (http://genie.sis.pitt.edu/)).

In the research literature, we find advanced algorithms for computing updated probabilities in networks that are multiply connected; i.e., where there are several paths between certain node pairs [8]. Approaches for decision-making through the modelling and use of influence diagrams [10], in combination with the Bayesian modelling, also exist. As the number of probabilities that we need to specify may become very large, techniques have been developed for reducing this number, for instance, through the use of noisy or-gates [11]. We also find research on the knowledge engineering process in the literature, to establish both structure and the conditional probabilities (see, for instance, [9]). Lastly, we mention that machine learning techniques are extensively used to learn both structure and probabilities for a domain (see, for instance, [8]).

THE NORWEGIAN DIRECTORATE OF FISHERIES AND THE ELORV PROJECT

The Norwegian Directorate of Fisheries (NDF) conducts the public management of marine resources and the marine environment in Norwegian areas. Fisheries are an important contributor to the Norwegian economy, and the total nominal value of wild fish caught in Norwegian waters amounts to about 12 billion Norwegian Kroner (NOK)(1.5 billion Euros) per year (2007). The export value is about 18 billion NOK. Excluded from these numbers is the value of the salmon fish farming industry.

Historically, we know that fish populations are sensitive to over-fishing and that, to maintain a sustainable population, most of the marine species are quota regulated. Hence, there is a need for governmental control of the resources. Within the NDF, the Control Section has the mandate to set the premises for inspections of vessels and their catches when on the sea, and of landings of fish as they are delivered on shore. They cooperate with the Norwegian Coast Guard, which is a part of the Norwegian navy, on surveillance of the fishing vessels during their operation and inspections on the sea. They also cooperate with six production organisations that manage all sales of commercially caught fish in Norway.

Selecting vessels and on-shore landing facilities to control has, until now, mainly been based on the local inspectors' up-to-date knowledge of how the fisheries develop and of the participating actors and their history. This opens up potential for a lot of subjective judgement, and the reasons for the choices of objects to inspect have not always been clear.

NDF maintains several large data sources with information pertinent to the assessment regarding which inspection objects to select. It has hourly reports on the large fishing vessels' geographic locations, reports of their catches, their quotas, sales of the catches, results of previous inspections, ownership relations, tips received about irregularities, and other sources. The inspectors have access to these data, but the data organisation is too complex for the inspectors to be able to analyse them effectively in the continuously changing domain of their work.

In this context, the ELORV (Electronic Operational Risk Valuation) project is a development project governed by the Control Section that allows use of all of these data sources to help inspectors to do a more objective assessment of which objects to inspect. The hope is that the automated analysis of up-to-date information found in the databases will give a more correct picture of the risk locations and will help to identify objects that have not been as easily seen using the earlier, manual approaches. In this project, maintaining and organising the data sources are considered important tasks, as are presentation technologies. Nevertheless, in the end, a technology for risk assessment must be chosen, and the choice has fallen on Bayesian networks as a proven, robust technology for knowledge-based reasoning with uncertain information. It has also been used previously in similar domains, such as crime risk analysis [12] and terrorism risk assessment [13].

The first stages of the project then became the assessment of the usefulness of Bayesian networks in this context. The questions asked were:

- Is it actually possible to model fisheries and the behaviour of the actors in fisheries with Bayesian networks?
- Is it feasible to connect the Bayesian network models to the extensive data sources of NDF?
- Is it possible to construct a meaningful decision support system for the inspectors based on the Bayesian models?

These questions match well with the simple architecture proposed by Sprague in 1980 [5], as they relate to the three system parts: data repository, model, and user interface.

As a first exploratory domain for the use of Bayesian networks, we chose the mackerel (*Scomber scombrus*) fisheries, which are conducted in the Norwegian parts of the North Sea every fall (August-November). More than 100 fishing vessels of different categories participate in this type of fishing; many come from Norway but vessels from other European countries also represent a significant part of the fishing fleet. Within mackerel fisheries there are several *risk areas*, i.e., possible irregular behaviours on behalf of the fishing vessels or land facilities. Included in these are dumping of caught fish, underreporting of catches, rewriting of catches to other species, and so on. The risk area chosen for the first Bayesian models was release of caught mackerel. *Catch release* further in the presentation is therefore a short term representing releasing mackerel that has been caught within the fishing gear (purse seine) but not taken on board the fishing vessel. Research has shown that a large portion of mackerel caught within a purse seine will be stressed and will die within few hours or days after the release, if not already dead [14]. Thus, the current Norwegian regulations say that all caught mackerel should be brought on board. It is the fishing vessel crew's responsibility to ensure that this is done. If the fishing vessel does not have the capacity, the mackerel not brought on board should be estimated and withdrawn from the quota, or could be given to other neighbouring vessels.

On a longer term, a central goal within ELORV is to create risk models for most of the regulated fisheries and risk areas, in order to support both the inspectors on the Coast Guard ships and the land-based inspectors. It is the intention that the experiences from the exploration and modelling of the mackerel fisheries should be used in modelling of the other areas. For instance, we see a clear potential that the initial, confirmed models from particular risk areas and fisheries will be directly usable, with minor modifications, in new contexts. It should also give us insights into how we, in general, realise the connection between a Bayesian network and the databases, as well as how the models may be utilized in different decision-making contexts.

A DESIGN RESEARCH APPROACH

Design research [7] is an approach within information systems research that focuses on how information technology can improve an organisation's performance. The steps are

1. to identify the problem area on which you want to focus within the organisation
2. to identify potential technological approaches
3. to work with the organisation to explore a chosen technological approach
4. to design and implement a technological solution
5. to evaluate the solution with regard to the organisation's goals
6. to analyse the design process to identify new insights

These steps may again be repeated several times. As a researcher, you will thus work iteratively, to develop and present general knowledge on the applicability of a particular technology in the chosen case and in similar cases.

In this paper, we present the work with the mackerel catch release risk area within ELORV as a design research project. The goal for this first exploratory part of the ELORV project is to establish knowledge about how to work with Bayesian models within the domain, and to obtain a first working model for a risk area. To do this, we have chosen a prototyping approach with a model-centric perspective, as the knowledge model developments fully guided the construction of the underlying database connection and the top-level decision-making support tool.

KNOWLEDGE MODELLING

In ELORV, Bayesian networks were chosen as a candidate technology because of their mathematically sound approach to uncertainty management. They also represent an intuitive approach to causality, which makes it easy to apply in modelling of experts' knowledge, and they have proven their worth in many application areas [9]. The graphical representation also seems to be useful for both experts and knowledge engineers.

The initial work towards an agreed-upon Bayesian network model went on for more than half a year and involved about 20 persons. In the end, we were able to present a model with 39 variables for each active fishing vessel. Of these, 23 variables were to be computed from NDF's databases, and are considered evidence represented as *evidence nodes*. Another 15 nodes are auxiliary nodes, i.e., representing variables used to aggregate information from evidence into one single variable. Finally, we have one target node that gives the probability of catch release for each single fishing vessel. The final model is given in Fig. (**2**).

The modelling process started in a rather large meeting that included two of the authors as knowledge engineers, a

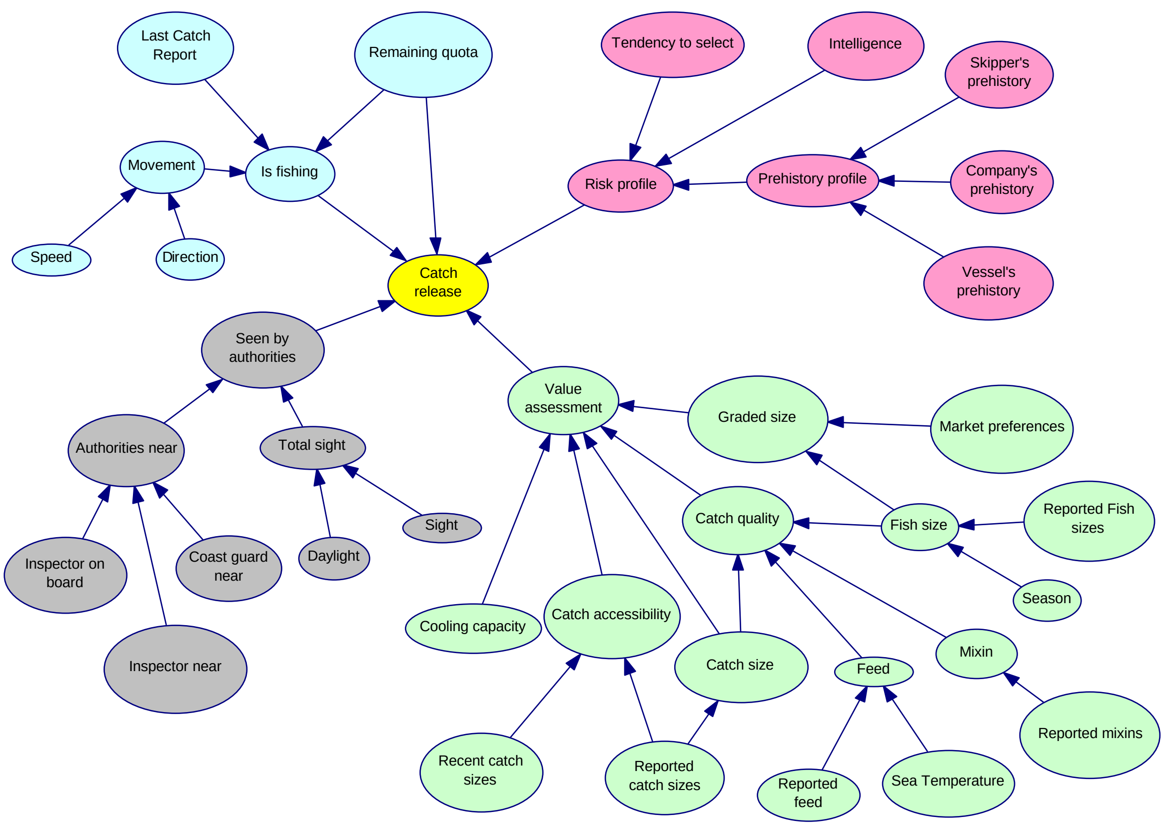

Fig. (2). The final Bayesian network diagram for Catch release.

project manager and administrative staff from NDF as well as experts on different aspects of NDF's work, including experts on control, statistics and regulations. In addition, we had representatives from the in-house ICT department. Important contributions were provided by field inspectors from both the Coast Guard and the land based inspection forces. It was at this time that we decided to focus on the mackerel fisheries for the exploratory work.

During a brain-storming session, we identified several possible risk areas on which to focus, such as catch release, dumping, rewriting of caught fish into other species, and over-fishing quota. We also identified many of the variables that influence the risks for different illegal or irregular activities. These include physical distance of the Coast Guard to a fishing vessel, remaining quota for the vessel, weather conditions, and fish quality. Fish quality again depends on its content of *Feed* (different zooplankton species can decrease value of the fish if eaten by the fish), size, and mix-in of horse mackerel (*Trachurus trachurus*) in the catch.

After the meeting, we started modelling the structure of a Bayesian network using the identified variables as well as including variables representing many of the identified risk areas. We used the GeNIe tool (http://genie.sis.pitt.edu/) to perform the modelling. The model was then presented again to a somewhat reduced expert group and was discussed. Corrections were suggested and we discussed the meaning of the different variables in terms of the type of information we were able to access. For instance, how do we estimate the quality of fish that a vessel may catch? To solve this, we identified attributes in the entries from the database of catch reports. The approach is to select those entries originating from the vicinity of the vessel's current location in the fishing fields. These entries are reports containing information about catch size, fish size, content of feed, mix-ins, and so on. Aggregating and automatically analysing these types of reports thus gives us good assessments of what to expect for the variables indicating fish quality. Similarly, we identified which sources we could use for computing values for many of the other evidence nodes.

At this second meeting, it was also decided to get data from a week in October-November 2007 in order to have realistic data for testing the model as it developed. These data included inspection data, weather data, positioning data, and catch report data, and from these, we were able compute values for many of the relevant variables in the Bayesian model.

A modified model was developed and, in a new meeting, we tried to work with eliciting the experts' probability assessments. The experts were uncomfortable with this and found it difficult. Therefore, we chose to follow an approach similar to Daniels *et al.* [13] where we, as knowledge engineers, interpret the qualitative relations among the variables and convert these into numerical probability values. We also found it hard to handle the complexity created from including many risk areas into one single model, so we decided to focus on one single risk area, namely irregular release of caught mackerel, i.e., catch release. We continued revising the model, also including parameters reflecting the qualitative probability relations indicated by the experts in the meetings.

Quite early in this process, it became evident for us that we needed to create the Bayesian models so that they were modelling one single vessel's risk at a particular point in time. This way, we were able to use the same parameters for each vessel, whereas the evidence given by assigning values to variables would be found in the databases using the vessel as an input parameter to the database search. This object-oriented approach to modelling is described in [15]. The independence assumption significantly reduces the computation complexity with which we would have to cope. By skipping these independence assumptions between vessels, we would have to cope with a single Bayesian network with more than 2000 nodes, whereas we now have to handle a 39 node network instantiated for each fishing vessel.

At this stage, the model consisted of two sub models, one handling the probability of the fishing vessel being observed by the coast guard, and the other summing up the knowledge we had in the data about the fishing vessel's propensity to be involved in irregular activities, computed from the inspection database at NDF. We also had plans to include nodes handling the quality of fish, but it had not yet been included. During simulations of the model (described below), we now observed that boats moving with high speed towards land, or moving as if they intended to go fishing for herring further north, would be considered risky objects. To handle this, we added a sub model for handling the probability that a boat is actually fishing mackerel.

As a last step in the development of the model, we started to work with variables relating to the quality of the fish caught. If the quality is not considered good by the fishermen, they will be more likely to release the fish in order to try to get a better catch. The experts considered this to be a correct assessment, but more important is how valuable the fishermen view the catch to be. For instance, large catches are not good because they take too long to load and unload and this reduces fish quality. On the other hand, small catches are not considered good unless you can immediately get a new haul that would fill the boat up to an optimal level. The fishermen always consider catches of the largest fish to be very valuable, as they will get a significantly higher price per kilo for the largest fish even though other quality factors, like feed and mix-ins, score negatively. These complex relations between catch size, fish size, fish quality, and market price was not really in evidence for us, until we had short stand-up meeting with an experienced inspector and former fisherman. This short meeting, in fact, had a very large influence on how we constructed the sub model for the fishermen's value assessment of a catch, which we also assume directly influences the probability of catch release. The value assessment sub model is given in Fig. (**3**).

The parameters given in the models are essentially based on our assessment of the qualitative relations given by the experts. It is a goal for us to improve these by the use of machine learning techniques, but, at this stage of the modelling work, the experts are fairly well satisfied with the risk assessments given by the model, and the model has already been put into test use on live data. This implies that we should run the model every hour using the last position data and last catch report data. Positions are updated every hour, and catch reports are available for the system almost immediately as they are sent from the fishing vessel.

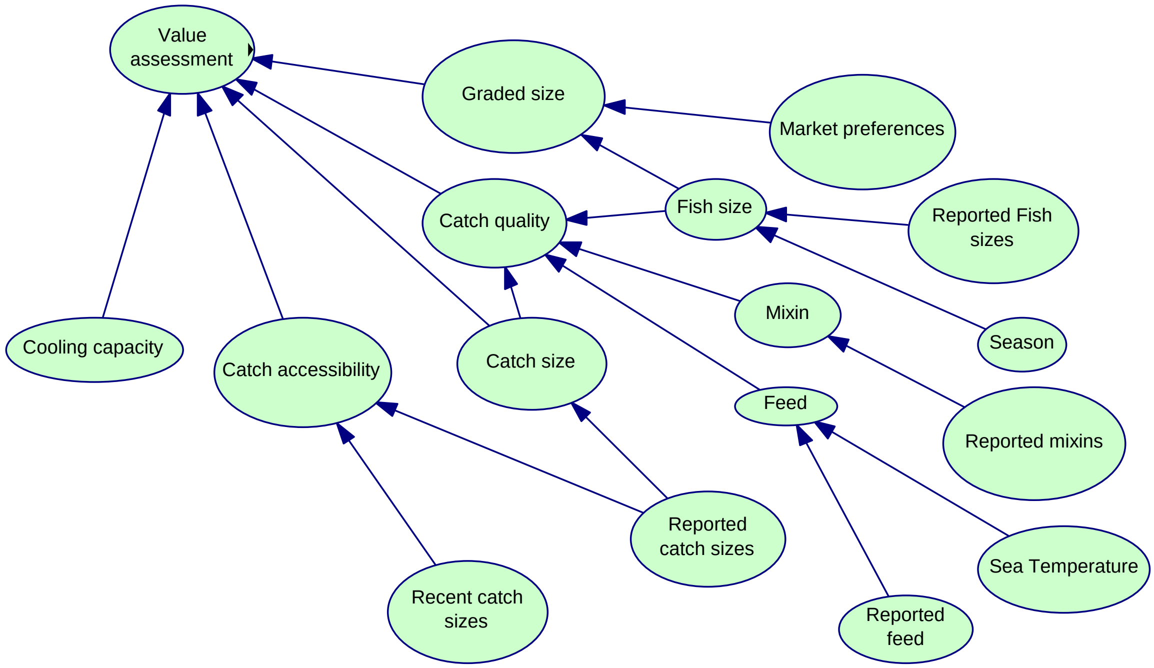

Fig. (3). The Value Assessment sub model.

Another aspect of the model is the decision about what values we can give to each variable and how the information in the databases is interpreted in terms of these values. This has also been an exploratory process, and this is based so far on an *ad hoc* assessment of the nature of the suitable values, also confirmed by experts. Approaches to do this discretization of variables [16] exist, but so far, we have chosen to adhere to the values established during the knowledge engineering process.

Impact on Organisational Issues

The development activities also led us to focus on some organisational issues. Firstly, it led to increased focus on obtaining continuously updated information on sales of catches from the production organisations. This information is of central importance not only to this project, but also for many of the other activities at NDF. Secondly, it has led to the start of a process in the Norwegian Coast Guard on the automation of their geographical positioning reporting. Today this is done manually every sixth hour through radio. The fishing vessels, in contrast, report their positions automatically every hour. The coast guard ship positions are of significant importance to the risk assessments.

Lastly, it has also started a process on improving the information systems that help handling intelligence from the public and inspectors about irregular activities. This intelligence needs to go through a better process of quality assurance, and it needs to be categorised so that it more easily can be used by automatic tools such as the one presented here.

SIMULATIONS OF THE MODEL

As previously described, it was decided quite early in the modelling process to extract data from NDF's databases to be able to run simulations for testing of the model. The data were picked from a week in fall 2007 when the mackerel fishing season was close to its end, but still very intense. The data sources were the following:

- Inspections database: NDF's own database storing data on inspections of fishing vessels and land facilities. It includes also any irregularities observed and the consequences for the actor.
- Location databases: The coast guard ships reported their location every six hours, and the Norwegian fishing vessel positions are automatically registered every hour, including information about speed and direction.
- Weather database: containing information on wind, visibility, air and sea temperature, and precipitation from the Norwegian meteorological services.
- Catch report database: All fishing vessels are obliged to report their catch of mackerel as soon as they are heading to land for delivery. This contains information about catch size, fish size, amount of mix-in, and amount of feed.
- Inspector on board data: We also had a list of vessels that have inspectors on board permanently throughout the fishing season.

From these data, we were able to compute 18 out of the 23 evidence variables in our final model. For the rest of the variables, we will have to establish data sources and also computation strategies in the future. This is not only a technological problem, but also an organisational issue, which will be discussed below.

EVALUATION

The evaluation of the Bayesian model and the supporting tools went on throughout the whole development process. This continuous evaluation had significant effects on how the final model developed. Furthermore, it also has informed us about how the final decision support tool will be presented to the users in the Coast Guard.

Simulations and Visualizations as Evaluation Tools

To support our own evaluation through simulations, we built a visualisation tool based on the OpenMap open source toolkit for visualisation of maps (http://openmap.bbn.com/), which indicated the positions of the fishing vessels and the Coast Guards vessels on a map of the North Sea. One of the first results of the visualisation was that we had to reduce the number of vessels to be shown on this map, as a very large proportion of the vessels would be fishing herring in Northern Norway instead of mackerel, and many of them would also be in harbours. Therefore, we introduced an option for filtering away vessels that we definitely knew would not be fishing mackerel. Nevertheless, there would still be a collection of vessels that would not have high risks associated with them, because of a high likelihood that they would not be fishing mackerel, even if they were located in the relevant parts of the North Sea. As a result, we included an assessment into the model nodes for the probability that, in fact, a vessel was actively searching for mackerel or in the state of fishing it. The sub model for this is given in Fig. (**4**).

A snapshot from the visualisation tool is given in Fig. (**5**). What we show in the visualisation are the most risky fishing vessels (red spots), less risky fishing vessels (green spots), Coast Guards vessels (black spots), and we also indicate blue areas where there is a possibility of high risk reduction if a Coast Guard ship is present. This is done by applying a simple clustering algorithm to find clusters of fishing vessels. Such clusters of fishing vessels indicate large schools of mackerel in the area, and the presence of the Coast Guard around such clusters significantly reduces the risk of irregular release of caught mackerel for all vessels in the cluster. In the figure, we also see a *tool tip* (with Norwegian text) indicating name and computed risk for one of the fishing vessels.

At first, we tried to implement an approach for suggesting the best course to take for each coast guard ship in the sea, in order to reduce the total risk in an optimal way and furthermore, to indicate this in the map. However, to do this involves a complex, stochastic optimization problem, where we have to take into consideration the possible movements of the fishing vessels and the likelihood that they would catch fish in the near future, combined with the time it would take for the coast guard to get to a suggested location. The quality of these kinds of decisions will not be good enough and, in reality, will interfere with other tactical considerations taken by the Coast Guard's operational centre.

The visualisation tool used during development is not the one that will be used in production, as NDF already has a well-developed map tool for showing the location of fishing vessels in Norwegian waters. In any case, the development process uncovered how we can present risk information in this map. In the presentation, we need tool tips (small pop up windows connected to objects in a graphical user interface) for vessels, giving information about the vessel's name, and, if the user is interested, values for the evidence used in the risk evaluation for the particular vessel. This is now a requirement in the integration of our solutions with the existing map tool.

Expert Evaluations

As we finalized the Bayesian model, it was sent to experienced inspectors in the Coast Guard for evaluation. During this process, further suggestions for improvement of the model were given by one of the experts. For instance, we came to know that fishing vessels out in the fishing fields, but with no mackerel quota left, would almost with certainty be fishing horse mackerel. Thus, the risk of illegal release of mackerel would be practically zero. Horse mackerel is not a quota-regulated species, and is not considered to be under particularly high fishing pressure, so NDF and the ELORV project will not focus on this fishery. Some adjustments of the probabilities, as well as naming of variables, were also the result of this activity.

DISCUSSION

The presentation of this development process documents several useful findings that are relevant both for development of decision-making support systems in general, and particularly for the knowledge engineering process pertaining to Bayesian networks as applied in DMSS.

The knowledge engineering process started with many stakeholders, and with a collection of knowledge that gave a good overview on fisheries, and mackerel in particular. As we continued the work, the group of people with relevant

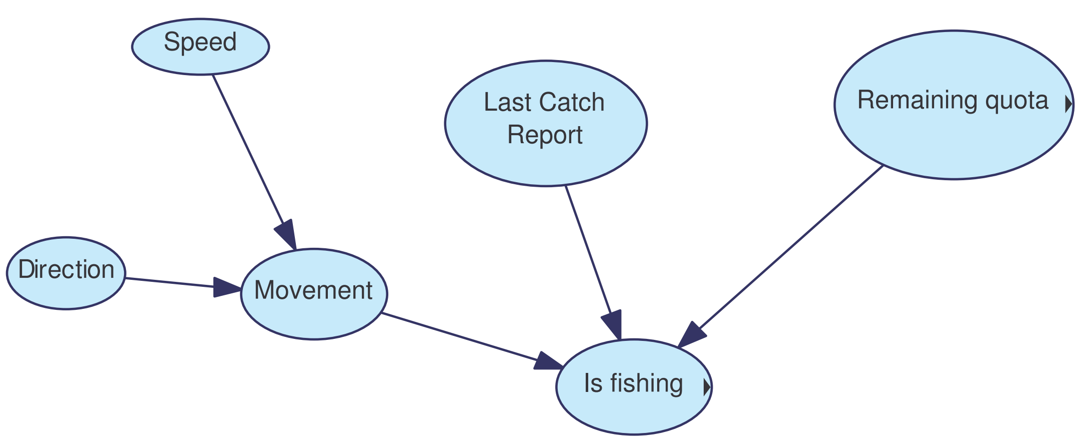

Fig. (4). The Is Fishing sub model.

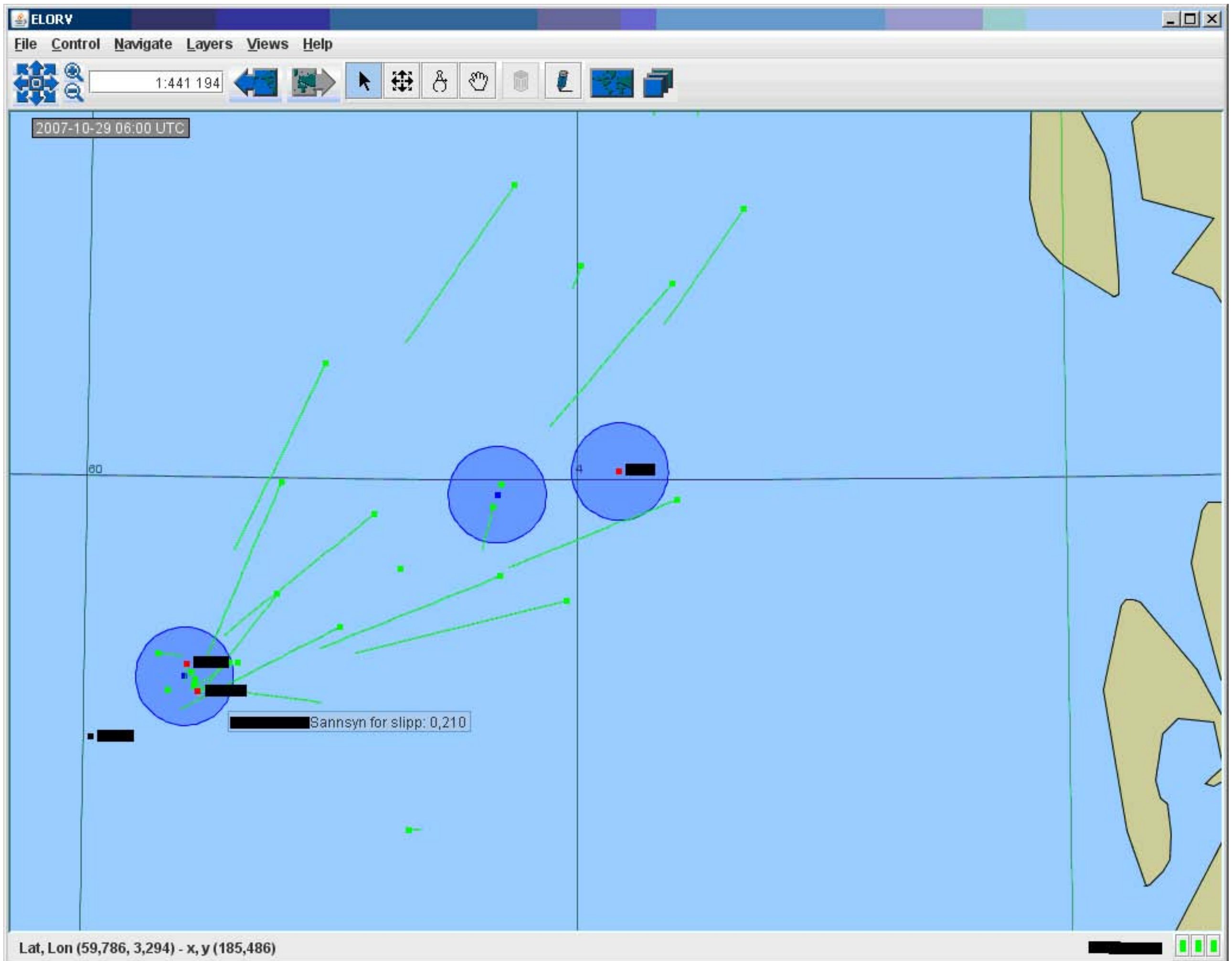

Fig. (5). A snapshot of the visualization tool. Vessel names and identifications are hidden.

knowledge gradually diminished to the most specialized experts. In particular, we involved future users from the Norwegian Coast Guard. At the same time, the opportunity for the knowledge engineers to work with the modelling process from offices within the NDF organisation made knowledge and experts easily accessible. This has led to an incremental and modular development of the Bayesian model as well as the supporting software. In many ways, the model that developed really takes into consideration many aspects of mackerel fisheries that would be difficult for one single expert to maintain control over. Integrating these aspects into the model will hopefully help the decision-makers to use a wider perspective when making their decisions.

Two considerations were essential when constructing the Bayesian model. First of all was computational feasibility, but the match between the model and the experts' view of the problem domain was also important. To handle computability issues, we ended up with a modularisation strategy where we focused on constructing the model for one risk area only, and then instantiating the model with one fishing vessel for each instance. Assuming independence among fishing vessels further reduces computational complexity. In addition, the identification of sub models, each of which addresses different aspects of the problem area, not only reduces computational complexity, but also makes the model more understandable for the experts. This object-oriented approach to modelling has significantly helped in the modelling process.

The quality of the structure of the Bayesian network is very much confirmed by the experts. However, the parameters (the numbers) and the discretization of the variables could most likely be improved by applying data mining techniques. It is, however, not possible to apply the EM algorithm [17] directly for learning the parameters, as we, in this risk area (catch release), are modelling something that is hardly ever *observed* in situ. We know that catch release does happen, based on observations of large amounts of dead mackerel on the ocean bottom. However, it is very difficult to observe if you are not present at the exact moment that it happens. Mackerel does not float by itself as do many other fish species, so you will not be able to observe dead mackerel if released. Also, no inspections or data analysis can give us direct proof that a particular vessel has released mackerel, but we believe that through advanced data analysis, we may find situations where catch release most likely has happened (or not happened). Thus, we hope to construct a data set that would give us a validation/calibration of the current parameters. Model validation is an area for further research in this project.

If we take a look at modern proposals for DMSS system development, the tripartite model for the development process model, as described by Gachet and Haetteschwieler [6], has its strengths when the organisation's decision-making tasks are well understood and established. This is definitely a candidate approach, though it is still very abstractly formulated, for further development of the ELORV system when the current prototyping stages are finished. However, it really did not fit well with the initial goals of this project, which were to establish knowledge about how Bayesian networks can be used in surveillance of fisheries. Our approach for the quick development of prototypes, in fact, matches better with a middle-out perspective, as described by Hurst *et al*. [18] and Spragues three-layer model [5]. As a middle layer, we have the Bayesian network model; on the top, we have the surrounding system including user interfaces; and at the bottom, we find the database services. The construction of the middle-layer model drives the process, since both the surrounding system (including the organisation) and database services are significantly influenced as a result of the knowledge model building.

Combining our observations about development process with existing knowledge about DMSS development, a result for the ELORV project is the work process model to be used and elaborated further in work with other risk areas and fisheries. Thinking similarly to Laskey *et al.* [19], the process is iterative and incremental in the Bayesian network model construction itself. However, as the model is developed, it also drives the development of the rest of the system. To summarize, an iteration of this model-centric development approach is as follows:

1. Communication with experts. In the beginning of the development work, this will be done on a large scale in meetings to get an overview. Later, this can be done with fewer experts, and in the last stages, single experts can be interviewed to give their opinions on specific areas of the Bayesian model.
2. Development of the Bayesian model, focusing on a single or few essential aspects of the risk area, and identifying sub models for aggregating information pertaining to this aspect.
3. Evaluation of the model through simulations. This can be done on real data extracted from previous fishing seasons (years). This helps to debug the models, and to verify on a coarse scale whether the knowledge engineers' understanding of the risk area is correct.
4. Evaluation of the model through expert assessments. Experts can critique the model, its structure, and its parameters by playing with the present model in a suitable tool.
5. Simultaneous development of the support tool with visualization and data presentation.
6. Evaluation of the support tool through simulations and expert assessments.
7. Identification of data sources needed, assessment of existing data sources, and discussion of actions to be taken by the organisation to satisfy needs for data and improved data quality.
8. Identification of other organisational consequences.
9. Long and short term planning.

As a tenth activity, in the future, we will most likely have to add steps for handling machine learning and validation of model parameters.

The research presented here is a case study, and it follows that the approaches developed may have limited applicability in general situations. Still, the case confirms that Bayesian networks are a suitable tool for assessing risk in a dynamic context, in particular if one uses effective modularization techniques to handle the combined influence of all objects and data involved in the assessment. In addition, the suggested development process may be a candidate in situations where knowledge engineers have little knowledge about the applicability of a particular knowledge representation approach. In addition, we verify that system prototyping may be used not only to establish requirements for a system, but also to change and possibly improve an organisation's practice.

CONCLUSION

In this paper, we have documented the ways by which we have been able to develop a prototype support tool for decision making in fisheries surveillance. We have followed a design research approach, which has been successful, as we as developers have been able to construct a system, based on Bayesian networks, that initially seems to have a high degree of acceptance among the users. The system's use value has been enhanced through the efforts we took to make the models become computationally efficient through the use of modularisation techniques. Users have been active in providing input as well as in the continuous evaluation of the Bayesian model and the software, and thus contributed significantly. In addition, the development process has had some influence on NDF as an organisation, as it has made evident the tasks to which the organisation has to assign more work resources. A contribution to the ELORV project is the development process model, which will be used as a guideline in further development of the system.

The resulting model is constructed mainly from analysis of discussions about the qualitative relations between variables, and the confirmation of the parameters from real data is still lacking. The model's parameters are, in fact, hard to verify, as is the chosen discretization of the model's variables. Consequently, these are issues for further research, and using data mining and other approaches for these problems will most likely lead to an even better support tool.

Further applications of the same strategy for building Bayesian models and tool building will be used for risk assessments in other fisheries and for other risk areas. Based on the initial success, we believe that our approach can be used successfully in the further development of the ELORV project.

REFERENCES

[1] R. E. Neapolitan, *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*. New York: John Wiley & Sons, Inc. 1990.

[2] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann Publishers Inc., 1988.

[3] O. Pourret, P. Naïm, and B. Marcot, Eds., *Bayesian Networks: A Practical Guide to Applications*, Chichester, England: John Wiley & Sons Ltd., 2008.

[4] J.N.D. Gupta, G.A. Forgionne, and T. M. Mora, Eds., *Intelligent Decision-making Support Systems.* London: Springer-Verlag, 2006.

[5] R. H. Sprague, A framework for the development of decision support systems, *MIS Quarterly*, Vol. 4, No. 4, pp. 1-26, Dec. 1980.

[6] A. Gachet, and P. Haettenschwieler, Development processes of intelligent decision-making support systems: review and perspective. In: J.N.D. Gupta, G.A. Forgionne, and T. M. Mora, Eds., *Intelligent Decision-making Support Systems.* London: Springer-Verlag, 2006, pp. 97-121.

[7] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design Science in Information Systems Research", *MIS Quarterly*, Vol. 28, No. 1, pp.75-106, 2004,

[8] R. G. Cowell, S. L. Lauritzen, A. P. David, and D. J. Spiegelhalter, *Probabilistic Networks and Expert Systems*. 2nd ed., New York: Springer-Verlag Inc., 2007.

[9] K. Korb, and A. E. Nicholson, *Bayesian Artificial Intelligence,* Boca Raton, FL: CRC Press Inc., 2003

[10] R. D. Shachter, "Evaluating influence diagrams," *Operations Research,* Vol. 34, No. 6, pp. 871-882, Dec. 1986.

[11] M. A. Shwe, B. Middleton, D. E. Heckerman, M. Henrion, E. J. Horvitz, H. P. Lehmann, and G. F. Cooper, "Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base," *Methods of Information in Medicine,* 30, pp. 241-255, 1991.

[12] R. Boondao, Crime risk factors analysis, In: O. Pourret, P. Naïm, and B. Marcot, Eds., *Bayesian Networks: A Practical Guide to Applications*, Chichester, England: John Wiley & Sons Ltd., 2008, pp. 73-85.

[13] D.C. Daniels, L.W. Hudson, K.B. Laskey, S.M. Mahoney, B.S. Ware, and E.J. Wright, Terrorism Risk Management, In: O. Pourret, P. Naïm, and B. Marcot, Eds., *Bayesian Networks: A Practical Guide to Applications*, Chichester, England: John Wiley & Sons Ltd., 2008, pp. 239-262.

[14] I. Huse, J. Saltskår, and A.V. Soldal, Overleving av makrell som er trengt i not (Survival of mackerel crowded in purse seine), *Fisken og Havet,* no. 1-2008. Bergen, Norway: Institute of Marine Research, 2008. In Norwegian.

[15] D. Koller, and A. Pfeffer, "Object-oriented Bayesian networks," In *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann Publishers, 1997, pp. 302-313.

[16] S. Monti, and G. Cooper, "A multivariate discretization method for learning Bayesian networks from data," In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI 98)*, Madison, WI. 1998, pp. 404-413.

[17] S. L. Lauritzen, "The EM algorithm for graphical association models with missing data," *Computational Statistics and Data Analysis*, Vol. 19, No. 2, pp. 191-201, Feb. 1995.

[18] E.G. Hurst, D.N. Ness, T.J. Gambino, and T.H. Johnson, Growing DSS: a flexible, evolutionary approach. In: J.L. Bennett, Ed., *Building Decision Support Systems*, Reading, MA: Addison-Wesley, 1983, pp. 111-132.

[19] K. B. Laskey, and S. M. Mahoney, "Network Engineering for Agile Belief Network Models," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 12, No. 4, pp. 487-498, Jul. 2000.

Received: December 01, 2008 Revised: March 06, 2009 Accepted: March 08, 2009