TRANSNAV
http://www.transnav.eu
the International Journal
on Marine Navigation
and Safety of Sea Transportation
Volume 14
Number 3
September 2020

DOI: 10.12716/1001.14.03.09

Using Bayesian Networks to Model Competence of Lifeboat Coxswains

R. Billard
Virtual Marine, St. John's, Newfoundland, Canada

J. Smith, M. Masharraf & B. Veitch
Memorial University of Newfoundland, St. John's, Newfoundland, Canada

ABSTRACT: The assessment of lifeboat coxswain performance in operational scenarios representing offshore emergencies has been prohibitive due to risk. For this reason, human performance in plausible emergencies is difficult to predict due to the limited data that is available. The advent of lifeboat simulation provides a means to practice in weather conditions representative of an offshore emergency. In this paper, we present a methodology to create probabilistic models to study this new problem space using Bayesian Networks (BNs) to formulate a model of competence. We combine expert input and simulator data to create a BN model of the competence of slow-speed maneuvering (SSM). We demonstrate how the model is improved using data collected in an experiment designed to measure performance of coxswains in an emergency scenario. We illustrate how this model can be used to predict performance and diagnose background information about the student. The methodology demonstrates the use of simulation and probabilistic methods to increase domain awareness where limited data is available. We discuss how the methodology can be applied to improve predictions and adapt training using machine learning.

1 INTRODUCTION

Lifeboat training is normally performed in controlled conditions to minimize the risk to trainees and equipment. Trainees are given limited or no opportunity to practice skills in operational scenarios that represent offshore emergencies. For this reason, human performance in emergencies is difficult to predict due to the limited data that is available. Forecasts of coxswains' skill transfer to real-life operational scenarios have relied on experts' opinion. Even so, there is limited information on how much skills learned in lifeboat training transfer to adverse weather conditions. The modelling of human performance in harsh environments has not been possible due to the scarcity of human performance data.

With the advent of lifeboat simulator technology, it is now possible for trainees to practice in weather conditions typical of their location of operation and to apply their skills in realistic emergency scenarios. Simulation provides the possibility to apply knowledge in applications in highly contextualized environments that are representative of plausible emergencies. Research has shown that practice in realistic scenarios helps development of mental models to improve performance (Klein, 2008). The study of human performance using simulation is evident in other operations including flight (McClernon et al. 2011), medical (Stefandis et al. 2007) and marine (Sellberg, 2017) training. Lifeboat training data can now be collected to assess the amount of practice needed to acquire skills and to evaluate how skills learned in practice transfer to new scenarios (Billard, 2019).

Data collected from a lifeboat simulator allow us to assess performance on tasks that were prohibitive to do, even in calm water training. This new data can be used to model learning and skill acquisition using probabilistic methods. We can study the interaction between tasks using Bayesian Networks (BN) to derive models of student competence (Millán and Pérez De-la-Cruz, 2002). These models can be used to study the relationship between training factors and to examine how practice on related tasks impacts performance. Due to scarcity of human performance data, initial models of competence can be formed with expert input (Groth et al., 2014). Performance data collected from simulator studies can provide evidence to inform models of trainee competence and validate their predictive accuracy. Bayesian methods have been used to model performance on lifeboat launch and manoeuvring tasks in initial training in calm weather conditions (Billard et al., 2020). Similar approaches can be applied to model performance in more adverse weather conditions.

In this paper, we present a methodology to form probabilistic models of human performance that can be used to study this new problem space. We use a BN to define a model of the competence of slow-speed maneuvering (SSM) based on tasks performed in adverse weather conditions during an offshore emergency. The model is derived from a combination of expert prediction and data collected from an experimental study.

The methodology is used to investigate the following research goals:

- how to formulate a BN model of competency using knowledge of task type and available performance measures; and,
- how to combine expert knowledge and data collected from simulator exercises to improve the model's predictive accuracy.

We evaluate the model using available data sets from a simulator study on lifeboat coxswain performance. We demonstrate how this model can be used to 1) predict performance as trainees practice skills in simulator scenarios, and 2) diagnose background information about the student.

The paper presents an approach that is relevant to training providers and researchers. We discuss how to apply the methodology and resultant models to study performance, improve expert assumptions, and extend to training applications where new data sets are being created. The models can be used to improve training programs, adapt training exercises to individual needs, and investigate human performance in new scenarios.

2 BACKGROUND

2.1 *Competence – Slow Speed Maneuvering*

We demonstrate the methodology of creating a BN model of competence using evidence captured in an experiment designed to study lifeboat training.

We must first frame our definition of competence considering our research goals and the objective measures that can be made. The concept of competence is a diverse topic that has diverse definitions. For our purposes, we consider how competence is normally measured in marine training through completion of demonstrable tasks specific to learning objectives (IMO 2014, STCW 2010). We consider competence the "existence of learnable cognitive abilities and skills which are needed for problem solving" as identified in research on skill acquisition (Weinert, 2001). We assume that completing tasks of a similar cognitive or physical skill form demonstrates competence.

We construct a model of competence for the skill of Slow Speed Maneuvering (SSM), as demonstrated by the ability to complete tasks related to stopping a lifeboat next to an object in the water. It is expected that trained lifeboat operators have this required competence to perform in an emergency. The completion of tasks in an emergency scenario can include stopping next to a number of objects including a life raft, a person in the water (PIW), a small vessel for transfer of personnel, or a large vessel for securing the lifeboat for recovery. All tasks considered under the competence of SSM require a similar application of skills and similar performance measures.

We assume there is a relationship between the SSM tasks based on the type of skill needed to perform the task. The maneuvering and stopping of a lifeboat is primarily a physical task and requires application of psychomotor skills to control the lifeboat, including manipulation of lifeboat throttle, steering, and making visual observations. There are also cognitive skills, including deciding angles of approach and judging distance from a target object. Practice on SSM tasks within a practice scenario is expected to improve performance on related SSM tasks based on the similarity of the tasks and type of skill that is applied.

2.2 *Simulator exercise and experiment*

We use data collected from a simulator scenario to formulate our model and provide evidence that can be used to inform and evaluate our methodology.

Data was taken from an experiment that used a lifeboat simulator to study skill acquisition and transfer in lifeboat coxswains. The experiment was designed to evaluate how skills acquired in different training programs transferred to a plausible emergency event that required the launch and maneuvering of a lifeboat in weather conditions typical of offshore operations. Participants completed training using different approaches over a year long period and then participated in a new simulator exercise for assessment purposes. The assessment scenario included a combination of launch tasks and on-water tasks. Details of the scenario are provided in Figure 1. Additional details on the experimental test plan and simulator used in the study can be found in Billard et al. (2019).

In real scenarios or in simulator exercises, SSM tasks form a part of the whole training exercise. Other tasks may need to be completed, including inspecting the lifeboat, launching the lifeboat, and navigating the lifeboat. These tasks require application of different skills and have different measures, as described in previous research (Billard et al. 2018, Billard et al. 2020). As such, these tasks are not related to competence of SSM and are excluded from the BN model creation as practice on these tasks is predicted to not affect SSM competence.

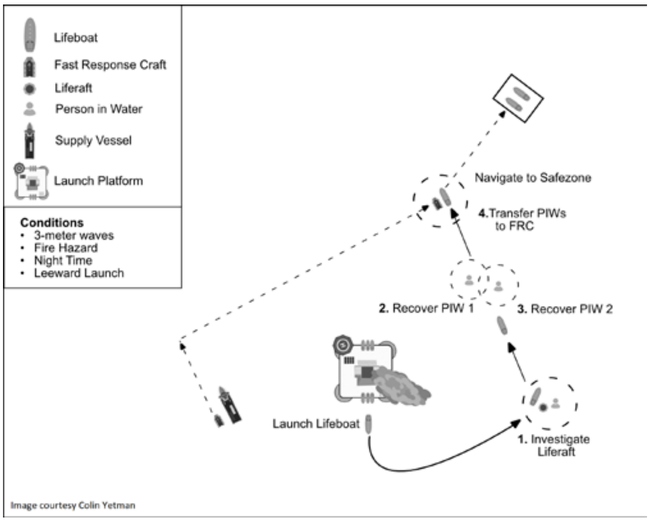

Figure 1. Simulator assessment scenario with SSM tasks

The data collected from the assessment scenario provided evidence to evaluate SSM competence modelled in a BN. The scenario contained 4 slow speed maneuvering tasks including, in order, stopping next to a Life Raft for inspection (LR), picking up two persons in the water (PIW1, PIW2), and stopping next to a Fast Rescue Craft (FRC) for transfer of personnel. These tasks provide evidence for the assessment of the SSM competence.

All participants completed the scenario at least two times and data was collected for the maneuvering tasks for each attempt. Tasks were completed in the same order with each attempt. A total of 39 participants completed the study.

2.2.1 *Measuring Performance*

The rubric used to define completion of the SSM task was derived from recognized training standards and is based on expected performance identified by Subject Matter Experts (SMEs). Each task requires approaching an object from a preferred direction, stopping close to the target, and maintaining a stopping speed. The specific parameters used to measure success differed slightly for each task (i.e. light contact with a vessel is acceptable for coming alongside a vessel, but not allowed for a PIW). Table 1 provides an outline of task objectives and the corresponding measures used in the simulator exercise. Completion of tasks was based on several simultaneous measures captured by the simulator, each of which had to be performed correctly to be considered a successful completion. Additional details on the scoring measures and rubric has been presented previously (Billard et al. 2018).

2.2.2 *Bayesian Network Modelling*

Bayesian Networks (BN) use a graphical structure to represent the relationship between several random variables as represented in a directed acyclic graph (DAG). A sample BN DAG is provided in Figure 1. Nodes (a,b,c,d,e) represent the variables and arcs (arrows) represent the probabilistic relationship between the variables. Bayesian inference algorithms create a relationship between latent variables, which are inferred, based on the state of observed variables.

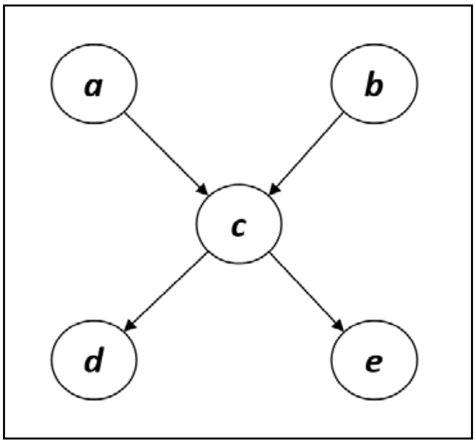

Figure 2. Sample Bayesian network DAG

Table 1. Slow speed manueuvering competence tasks

Task Identifier	Task Description	Task Objective	Measures
LR	Stop at a Life Raft	Approach a static object accounting for wind and wave direction. Use a speed to allow stopping. Stop close to Life Raft (2-3 boat lengths) and maintain position	direction of approach speed at stop time stopped
PIW	Recover a Person in the Water (PIW)	Approach a drifting PIW accounting for wind and waves to minimize chance of contact. Use a speed to allow stopping. Stop close enough to PIW to allow pickup and maintain position in waves	contact speed heading at stop number of attempts
FRC	Come Alongside a Fast Response Craft (FRC)	Approach a FRC accounting for wind and wave direction. Use a speed to allow stopping. Stop close to vessel (less than 0.5 meters) and at an angle to allow personnel transfer and maintain position.	

Building a BN includes the following steps:

1 Defining the variables that are being studied, both latent and observable, creating the nodes of the BN.
2 Defining the relationships between variables using arcs. The arcs represent a causal influence between the variables. Variables in the network that are not graphically connected are conditionally independent of each other (i.e. a and b are conditionally independent).
3 For each of the variables, defining the probability conditions with parent variables through Conditional Probability Tables (CPTs). The probabilities can be learned from real data or defined by experts.

Detailed description of BNs and how they are created is provided in other literature (de Clerk et al., 2013, Millán et al., 2010).

Creating a BN to use observable evidence to study an inherent competence has applications in training frameworks including Intelligent Tutoring Systems (ITS) (Millán and Perez-De-La-Cruz., 2002, Käser et al. 2017) and Evidence Centered Design (ECD) (Mislevy et al., 2004). In these frameworks, the BN forms a model of the competency that is being investigated (the student model) and identifies the relationships to the performance measures (the evidence) in the practice scenario (the activity). The relationships form a construct of competence, a latent variable, that can be measured through the collection of performance data, an observable variable.

In our case, we use the observable completion of SSM tasks to quantify the latent variable of SSM competence using evidence collected through a simulation study.

3 METHODOLOGY

We use a BN methodology to model competence and predict the performance of lifeboat operators as they apply skills learned in training to a new scenario. We create a BN model using observable measures from a simulation scenario designed to evaluate coxswain performance in a plausible emergency. We use a combination of expert prediction and simulator data to create and revise our model. The methodology creates a student model of SSM competence that can be used for the prediction of performance on tasks and the diagnostic study of causal relationships between model variables.

The steps in the methodology include the following, as outlined in figure 3:

1 Defining a generic BN student model of competence - based on completion of tasks that are considered similar in the type of skill applied
2 Characterizing the BN model as a SSM competence student model - based on the evidence gathered in a simulator practice exercise
3 Creating the initial CPTs of the model nodes based on expert estimates
4 Refining the CPTs based on experimental data - using the simulator experimental data to tune the model parameters
5 Validating the model accuracy for predictive and diagnostic use cases using simulator data

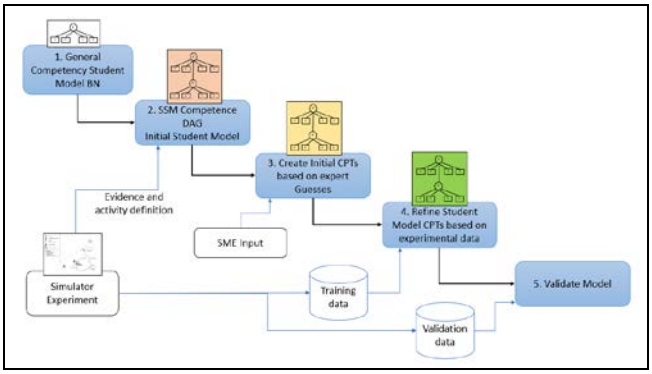

Figure 3. Methodology of creating and validating a SMM competence Bayesian network

We perform two validation cases to show how the BN model can be applied and how the model changes with new data or variables. We first demonstrate how the predictive accuracy of the model changes as the methodology is applied. We evaluate the predictive accuracy of the model first formed with expert estimates and then re-evaluate the predictive accuracy after data have been used to refine the CPTs. We then present an example of how new variables can be added to the model and show how the model can be applied to diagnose the relationship between the new variable and observable evidence. The validation of models is discussed in Section 4.

3.1 *Step 1 - Defining a generic BN student model of competence*

We first describe the types of variables and relationship assumptions for the BN student model.

We assume a latent variable of competence (C) and relate to task evidence nodes (Ei), which can be measured or observed in a scenario. The tasks are related by the type of skills needed to complete the tasks successfully.

To create the DAG, we assume a structure where observable evidence of completing tasks changes the probability of the competence, as described in previous research (Millán and Pérez De-la-Cruz, 2002). The generic model is presented in Figure 4. In the model structure, we assume a causal relationship where the latent variable *(C)* causes the evidence E_1, E_2, E_3, … E_i. In this relationship, evidence about mastering a task changes the probability of the latent parent. Consequently, evidence about mastering *C* changes the probability of its children (E_i) and evidence about mastering a task affects the probability of mastering the rest of the tasks on the same level. This models assumes conditional independence of the E_i given *C* (for each $i = 1,\ldots n$). In this DAG, the CPT parameters that need to be identified are the prior probability of the competence, *P(C)*, and the conditional probabilities of the evidence nodes $\{P(E_i|C), i=1, \ldots n\}$

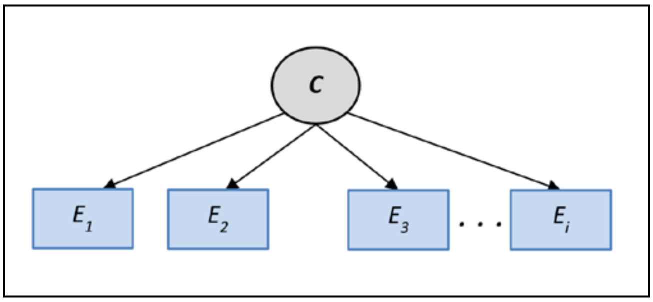

Figure 4. Competence model BN DAG

3.2 *Step 2 – Characterizing the BN Competence Model as a SSM competence student model*

We design the BN model to match the activity, in this case the slow speed maneuvering exercises performed in the simulator study.

Figure 5 shows the DAG for the experimental study consisting of two scenarios, each having 4 evidence nodes. In the simulator study, the trainee practiced the same scenario twice, creating two sets of evidential nodes, as the trainee completed the same tasks with each attempt. As an input of evidence in the BN, the task was either considered to be completed (Yes) or not completed (No) based on the performance requirements set by SMEs to measure successful completion of task.

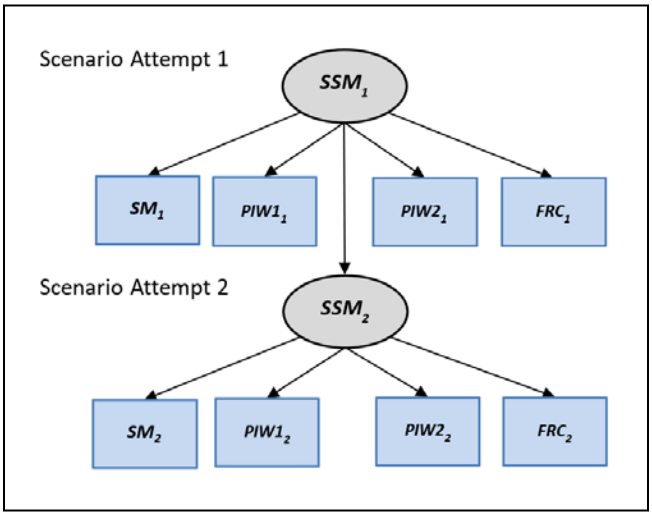

Figure 5. Bayesian network DAG – Simulator assessment scenario

The structure of the model assumes a learning effect with tasks practiced in a training session consisting of multiple simulation exercises. We use a dynamic model indicating the trainee's competence can be measured with each simulator exercise attempt. We define a relationship between the measure of competence in the first attempt (*SSM_1*) and the measure of competence on the second attempt (*SSM_2*). The relationship assumes the measure of competence in the first attempt impacts the probability of the second attempt through a defined *CPT* $\{P(SSM_2|SSM_1)\}$. Based on the similarity of the task types it is expected that practice on any of the task types can improve the performance on other tasks, including future attempts at the same task using the same scenario.

3.3 *Step 3 – Creating initial CPTs based on expert estimates*

The structure of the BN requires the definition of CPTs including the prior probabilities of the SSM competence and the conditional probability of completing the evidence nodes (tasks) given the competence.

For each of the tasks, we make predictions on the relationship between having the *SSM* competence and the ability to complete tasks. As defined in modelling of human performance (Millán et al, 2002), we use estimates of slip and guess to define the conditional probabilities. In our context, a slip is the probability of not being able to complete the task successfully despite having the competence. The probability of completing the task successfully when having the competence $\{P(Task_i|SSM_i)\}$ is therefore $1 - s$, where s is the slip factor. A guess (*g*) is the probability of completing the tasks successfully without having the competence. The CPTs require definition of the probability of completing the task whilst having the competence (1 - s) and the probability of completing the task while not having the competence (*g*).

We estimate the CPT parameters for each of the evidence nodes and the conditional probabilities for each of the competence variables. The probabilities of slip and guess were estimated by SMEs and took into the account the following:

1. The participants in the study had received initial training and refreshed skills over a one-year period. It was expected that some participants had acquired enough skill to achieve competence.
2. The simulator scenario in the study had not been practiced before and had challenging weather conditions (moderate sea states). These factors impact the probability of completing tasks that had been practiced in previous training events in less adverse weather.
3. The task of stopping next to a PIW is more difficult to complete than stopping next to a life raft or stopping next to an FRC (Billard et al. 2020). We assume the probability of a slip is higher and the probability of a guess is lower for the PIW task.
4. The performance of tasks in the simulator, either successfully or unsuccessfully, is considered practice. Competence is expected to increase as the scenario is repeated. The probability of slip on tasks is expected to reduce and the probability of a guess is expected to increase.

In considering the type of task and the environmental conditions, SMEs estimated that there is a reasonable chance of slip given the difficulty of the task and the expectation that people could make errors despite having the competence. The irregularity of wind, wave, and propulsion forces create some variability in performance. Environmental forces could have a sudden negative impact (i.e. causing the vessel to overshoot position) resulting in slip. The environmental forces can also increase the chance of success of an inexperienced driver (e.g. helping slow and stop a vessel that is approaching too fast) creating a successful guess.

Table 2 provides a breakdown of the probabilities used in the BN. These are considered an initial estimate of the probabilities based on an expert prediction. The assumed initial probability of having the competence of SSM is estimated to be 60%, and increases in probability in the second scenario. For the evidence nodes, the probability of a successful completion of task is assumed to be lower for tasks that are more difficult. The assumed probability of completing LR and FRC tasks was assumed to be 70%. The probability of completing the PIW task was estimated as 60% due to the increase in slip factor as the task is more challenging. Similarly, the assumed probability of a guess for the tasks of LR and FRC was assumed to be 30% and the estimated probability of a guess for the PIW task was estimated as 20%. To account for the effect of practice, the SSM competence is expected to increase for the second scenario. The assumed probability of a successful completion for each task was increased by an increment of 10% and the guess rate for each task was also assumed to increase by an increment of 10%.

These estimates are an initial guess of expected outcomes provided by subject matter experts. The estimates are based on expert prediction as they could not be derived from data. The next step in the methodology uses experimental data to refine the CPTs used in the BN.

3.4 *Step 4 – Refine CPTs based on experimental data*

The BN model was created in modelling software, GeNIe, developed by Decision Systems Laboratory of the University of Pittsburgh. The DAG was based on the relationship diagram provided in Section 3.2, and the probabilities outlined in Section 3.3 were used to create the CPTs for each of the nodes.

Data were collected in a simulator exercise, with evidence collected for each of the 39 participants who completed the two scenarios. The data set was split randomly into two groups: a learning data set and a validation data set. One set of the data (19 records) was used to adjust the parameters of the BN (the learning data) model and the second data set (20 records) was used to predict the accuracy of the model (the validation data).

Conducting parameter learning in the Bayesian Network is often termed training the BN. In this exercise, the parameters of the BN CPTs are adjusted in an effort to match the BN model predictions to the outcomes of the learning data set. This exercise is performed in the GeNIe modelling software, which uses an EM algorithm to learn parameters from data (Dempster, 1977). In our use case, we start training the BN with the probabilities set by the experts. As we have a small data set, we assume a low level of confidence in the parameters (20%) to allow the parameters to be flexible to change.

We are now able to make comparisons between the original BN model, based on expert predictions, and the updated model, trained with experimental data.

4 VALIDATION CASES

4.1 *Validation Case 1 - Evaluating model predictive capability using task evidence*

The validation data set is used measure the predictive accuracy of the BNs. The initial models developed by expert prediction and the trained models are applied to a new data set (the validation data) to compare each model's predicted outcomes with evidence provided in the data set.

Two validation steps are performed to show how the methodology resulted in an improved BN model:

1 Testing the predictive accuracy of the BN with initial expert predictions of CPT – this step evaluates the suitability of the probabilities estimated by the SMEs.
2 Testing the predictive accuracy of the BN after using the simulation data – this validation shows the impact of using additional simulator data to revise the model parameters.

The validation demonstrates the use of BN for prediction, as the model attempts to identify the most likely occurrence of the evidence nodes. For each of the validation exercises we consider the model's ability to predict the outcome of the final two tasks in the simulation exercise (PIW22 and FRC22). These two evidence nodes are selected as they are the last two tasks performed in the simulator exercise. Performance on these tasks is expected to be more likely a result of competence gained through practice than due to a random slip or guess. We compare the predicted outcome of the evidence nodes from the BN model to the actual outcome from the data set.

Table 2. Inputs to BN - Expert estimates

Scenario Attempt 1				
$P(SSM_1)$	60.0%			
$\boldsymbol{SSM_1}$	$P(LR_1 \vert SSM_1)$	$P(PIW1_1 \vert SSM_1)$	$P(PIW2_1 \vert SSM_1)$	$P(FRC_1 \vert SSM_1)$
Y (1 - s)	70.0%	60.0%	60.0%	70.0%
N (g)	30.0%	20.0%	20.0%	30.0%
Scenario Attempt 2				
$\boldsymbol{SSM_1}$	$P(SMM_2 \vert SSM_1)$			
Y (1 - s)	70.0%			
N (g)	30.0%			
$\boldsymbol{SSM_2}$	$P(LR_2 \vert SSM_2)$	$P(PIW1_2 \vert SSM_2)$	$P(PIW2_2 \vert SSM_1)$	$P(FRC_2 \vert SSM_2)$
Y (1 - s)	80.0%	70.0%	70.0%	80.0%
N (g)	40.0%	30.0%	30.0%	40.0%

A benchmark comparison is made with a BN that uses a uniform distribution for initial CPT parameters for all latent and observable nodes. We use this BN to make a comparison with a model that is formed with no expert input and driven only by available data. This approach disregards the expert predictions and assumes an equal probability (50%) for completing or not completing tasks, and related slip and guess probabilities. The parameters are adjusted using the same learning data and using the same learning algorithm as in the expert prediction.

Table 3 shows the differences in prediction accuracy of the BN models that were investigated. The Table indicates the number of times the model and validation set had a common outcome on successful completion of task (Yes) or when tasks were not successfully completed (No) for the 20 records in the set. The predictive accuracy of the BN based on expert guesses was 75%, indicating the expert informed probabilities were reasonable. The predictive accuracy of the model increased slightly to 78% when trained with experimental data. The approach of using expert input showed a much higher predictive accuracy than a model trained from uniform parameters. This outcome suggests that the expert guess was needed to generate a suitable model given the amount of available data.

Table 3. BN model predictions and comparisons

	Initial Expert Estimate	Expert Estimate Trained	Uniform Trained
Overall	75% (30/40)	78% (31/40)	48% (19/40)
PIW22			
Combined	80% (16/20)	80% (16/20)	50% (10/20)
Yes	80% (8/10)	80% (8/10)	0% (0/10)
No	80% (8/10)	80% (8/10)	100% (10/9)
FRC2			
Combined	70% (14/20)	75% (15/20)	45% (9/20)
Yes	100% (11/11)	73% (8/11)	0% (0/11)
No	33% (3/9)	78% (7/9)	100% (9/9)

The method also allows us to investigate how the data set changed the BN CPTs from the initial expert estimates. These changes provide insights on the predicted competence and task difficulty, as a refinement to the estimates initially made by the SMEs. Table 4 presents the change in CPT from the initial estimates provided in Table 2. The outcomes show the initial probability of SSM competence (SSM1) was lowered by 13%, indicating the initial estimate of competence was too high. The outcomes also show that most of the probability parameters for successful PIW pickup for each attempt had to be lowered, suggesting this task was more difficult than predicted. The probabilities for stopping at a life raft were increased for each attempt.

Given the limited amount of data that is available, it is difficult to make conclusive remarks about the final probabilities of the BN model. Additional data are expected to further change the CPTs and increase the predictive accuracy of the BNs.

4.2 *Validation Case 2 – Investigate diagnostic causal relationship of background training*

In this section we discuss how the BN can be used as a diagnostic tool and identify causes given a set of observations. We incorporate additional information about the test participants and show how the model can be used to associate performance to the new information. We introduce a new evidence node, Background Training (BT), to indicate whether the participants received hands-on training during their regular practice prior to performing the simulator exercise. Participants who received hands-on training in regular practice sessions were more likely to be able to complete on-water tasks compared to those who did not (Billard et al. 2019). This information is known for all participants who completed the simulator scenario and the related validation data sets. 26 of 39 participants received hands-on training; 13 did not.

The updated BN for this model is provided in Figure 6. The BT node is introduced and forms a causal relationship having an influence on the starting competence of the trainee (SSM1).

We again define the conditional probabilities for the influence of training on competence using an expert estimate as there were no existing data available. It is assumed that those who received hands-on training had a higher probability of having the competence, but not greater than 60% as training had not been received in the weather conditions used in the assessment scenario. It was assumed the participants who had not received hands-on training had a lower probability of having the competence, having not received any scenario-based practice. The probability of having received initial training was set to 50%, making the initial probability random. This allows the model to predict the causal affect based on the evidence nodes from the simulator experiments and inherent relationships. Table 5 shows the new CPT values defined in the BN.

Table 4. Change in BN probabilities – trained model

Scenario Attempt 1				
$P(SSM_1)$	47% (-13%)			
$\boldsymbol{SSM}_1$	$P(LR_1 \vert SSM_1)$	$P(PIW1_1 \vert SSM_1)$	$P(PIW2_1 \vert SSM_1)$	$P(FRC_1 \vert SSM_1)$
Y (s)	76.1% (+ 6.1%)	57.4% (- 2.6%)	50.1% (-9.9%)	63.7% (- 6.3%)
N (g)	41.5% (+11.5%)	16.6% (- 3.4%)	13.4% (- 6.6%)	23.8% (- 6.2%)
Scenario Attempt 2				
$\boldsymbol{SSM}_1$	$P(SMM_2 \vert SSM_1)$			
Y (1 - s)	67.7% (- 2.3%)			
N (g)	*25.6% (- 4.4%)*			
$\boldsymbol{SSM}_2$	$P(LR_2 \vert SSM_2)$	$P(PIW1_2 \vert SSM_2)$	$P(PIW2_2 \vert SSM_1)$	$P(FRC_2 \vert SSM_2)$
Y (1 - s)	83.8% (+ 3.8%)	69.3% (- 0.7%)	70.4% (+ 0.4%)	81.2% (+ 1.2%)
N (g)	48.4% (+ 8.4%)	26.4% (- 3.6%)	28.6% (- 1.4%)	32.1% (+ 2.1%)

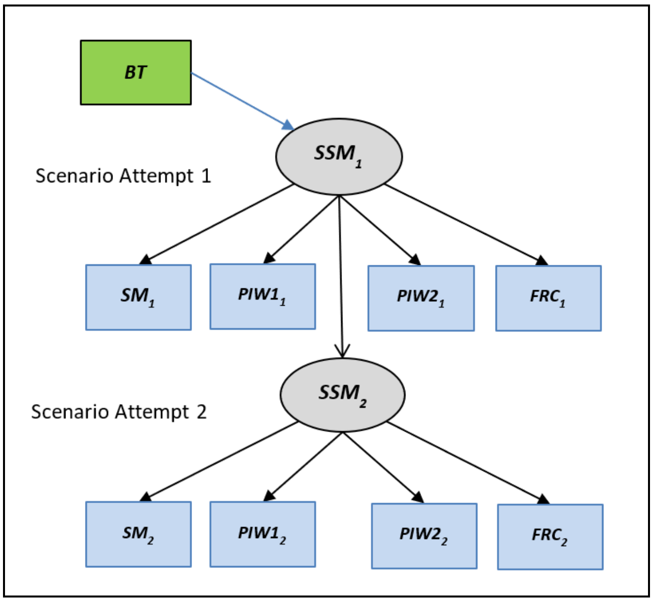

Figure 6. BN with training evidence introduced

Table 5. Background training (BT) conditional probabilities

P(BT)	50%
BT	$P(SSM_1 \mid Training)$
Y (1-s)	60%
N (g)	40%

We perform a similar validation procedure outlined in section 4.1. We compare the BN model prediction of BT to the evidence from the validation data set. The evidence in this case is knowledge of the trainee's background in terms of having received hands-on training (Yes) or not (No).

Table 6 indicates the model correctly guessed if background training had been received for 65% of the records in the data set. This outcome suggests that additional data or a revised estimate is needed to refine the model and increase the predictive accuracy for this evidence node. As highlighted in Table 7, the conditional probabilities of having the SSM1 competence decreased for both cases (with or without having received background training) when data were used to train the model. These changes in probability can be used to refine the expert estimate or initial CPT for new data sets.

Table 6. Diagnostic accuracy – background training

	Expert Estimate Trained
BT	
Overall	65% (13/20)
Yes	54% (7/13)
No	86% (6/7)

Table 7. Change in SSM1 CPTs

BT	$P(SSM_1 \mid Training)$
Y (1-s)	55.4% (-4.6%)
N (g)	35.3% (-4.7%)

5 DISCUSSION

The methodology in this paper presents an approach to use available information and background expert experience to create probabilistic models of human performance in scenarios for which there is limited available data. This approach can be applied to training applications where the desire is to investigate how observable measures of performance impact skills acquisition and competence. We chose lifeboat coxswain training as the use of simulation has extended training capabilities, and data from new scenarios are available to study this problem area.

We presented a method to develop a student model of lifeboat competence that integrates expert prediction and evidence from a simulator experiment. We derived the BN model for SSM competence using a framework that has been applied in ITS and ECD to use observable evidence from a simulation assessment to design the model. We demonstrated how the BN model can be used to predict performance and diagnose causal relationships, illustrating how the model can be applied to investigate relationships between latent and observable variables.

The validation examples indicate that embedding expertise in the model can result in a high initial predictive accuracy, despite using a small data set. The model's predictive accuracy was further increased as simulator data were used to inform the BN probabilities. This outcome indicates that domain knowledge is valuable in initializing probabilistic models in cases where there is limited data. It is expected that the model's predictive accuracy would improve further if the CPTs are trained with a large data set derived from user performance data.

The scalability of the BN model is a strength that can be further explored. We presented a model of lifeboat coxswain competence that is very narrow (a single competence) and derived from a scenario with fixed weather and tasks. For this study, the modelling of competency is specific to the environmental conditions used in the scenario. In a training program involving multiple practice exercises, the number and order of task types can be varied, and the level of difficulty can change with environmental conditions (i.e. increase in wave height or wind, day or night). The probabilities are expected to be different in scenarios that are easier or more difficult. Additional background information can also be considered, including time between training events and student training experience. The relationship between other competencies can also be established (e.g. practice in maintaining heading seakeeping exercises may improve control of the vessel in SSM).

Figure 7 shows an example of how the BN could be expanded to explore causal relationships between variables as more information on the student is known and as evidence is gathered through a training program. These BNs can become complex as they form a detailed model of student competence. These models can be used to investigate factors that affect performance while gaining insights on human performance limitations.

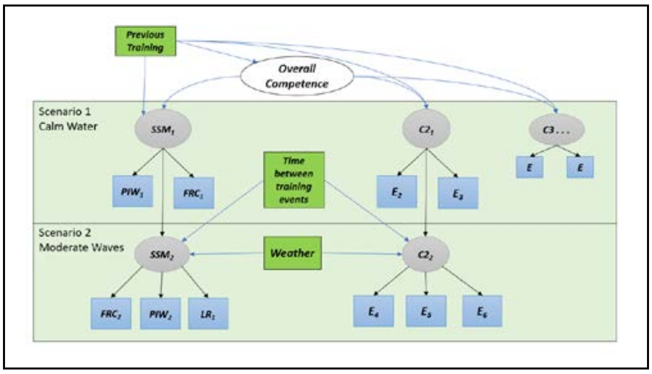

Figure 7. Sample BN with expanded relationships representing a lifeboat training program

The formation of a student model using BNs offers additional means to apply probabilistic models to improve training. We have presented a model to study performance based solely on assessment of task performance (i.e. was the task completed successfully or not). The model can be expanded to investigate the specific behaviours performed by the participant in completing the task to study which actions result in the highest probability of success. This type of model tracing is possible given the measures identified in the rubric. The outcomes can be used to model novice and expert performance as inputs to ITS (Millán et al, 2011). The probabilistic modelling of the BN can be integrated with machine learning algorithms to build adaptive training applications to customize training material to an individual's strengths and weaknesses based on evidence gathered in training.

To conclude the discussion, we make four recommendations to researchers who wish to use the methodology to study human performance and training for situations that have limited data. First, we advise the student model to be built as early as practicable to allow for the student BN to be informed with evidence that will be collected. This approach will allow for alignment between the student model with research objectives, and scenarios can be designed to study relationships of interest. Second, we recommend a balance of expert and data-driven input in the probabilistic models. As demonstrated, the modelling of CPTs using expert input can provide a model with suitable predictive accuracy. In cases where data are being collected for scenarios with limited initial data, the expert prediction is a guess. Probabilistic models derived from large data sets are expected to have a higher predictive accuracy. We also suggest that users consider the extended uses of relationship modelling of the BN approach. The BN models can be restructured, and new variables added (latent or observable) to investigate causal relationships and influence of new information. Finally, we suggest the use of simulation to perform assessments and collect data for situations that are normally prohibitive due to risk. Simulation scenarios extend studies to new operating conditions and provide a consistent measure of performance. Digital measures from a simulator exercise can input directly into probabilistic models such as BNs to apply machine learning and adapt training in real time.

ACKNOWLEDGEMENTS

We thank Petroleum Research Newfoundland and Labrador and the Industrial Research Assistance Program of the National Research Council who sponsored the study. The authors acknowledge with gratitude the support of the NSERC/Husky Energy Industrial Research Chair in Safety at Sea.

REFERENCES

Billard, R., Smith, J.J.E. (2018). Using simulation to assess performance in emergency lifeboat launches. Proceedings, e Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC). Paper number 19179.

Billard, R., Smith, J., Veitch B., (2019) Assessing lifeboat coxswain training Alternatives using a simulator. The Journal of Navigation, Published online by Cambridge University Press: 19 September 2019.

Billard, R., Musharraf, M., Smith, J., Veitch B., (2020), Using Bayesian methods and simulator data to model lifeboat coxswain performance. WMU Journal of Maritime Affairs. Published May 2020. https://doi.org/10.1007/s13437-020-00204-0

de Klerk, S., Veldkamp, B.P., Eggen, T., (2015). Psychometric analysis of the performance data of simulation-based assessment: A systematic review and a Bayesian network example. Computers & Education 85 (2015), 23-34.

Dempster, A.P., Laird, N.M., Rubin, D.B. (1977), Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 39, No. 1. (1977), pp.1-38.

Groth K., Smith, C., Swiler, L. (2014). A Bayesian method for using simulator data to enhance human error probabilities assigned by existing HRA methods. Reliability and System Safety 128 (2014), 32-40

International Maritime Organization., & International Conference on Training and Certification of Seafarers (2010). STCW including 2010 Manila Amendments, 2017 Edition.

International Maritime Organization. (2014). International Convention for the Safety of Life at Sea (SOLAS), Consolidated Edition. London: International Maritime Organization.

Käser, T., Klingler, S., Schwing, A., Gross, M. (2017). Dynamic Bayesian Networks for student modeling. IEEE Transactions on Learning Technologies, Vol. 10, No. 4. Oct.-Dec. 1 2017.

Klein, G., (2008), Naturalistic decision making. Human Factors: The Journal of Human Factors and Ergonomic Society, 50(3), 456-460.

McClernon, C. K., McCauley, M. E., O'Connor, P. E., & Warm, J. S. (2011). Stress training improves performance during a stressful flight. Human Factors: The Journal of the Human Factors and Ergonomics Society, 53(3), 207-218.

Millán, E., Perez-De-La-Cruz, J.L., (2002). A Bayesian diagnostic algorithm for student modeling and its evaluation. User Modeling and User-Adapted Interaction 12: 281-330, Kluwer Academic Publishers, Netherlands

Millán , E., Loboda, T., Perez-de-la-Cruz, J.L. (2010). Bayesian networks for student model engineering. Computers and Education, 55, 1663-1683

Mislevy, R. J., Almond, R. G., & Lukas, J. (2004). A brief introduction to evidence-centered design. CSE technical Report. Los Angeles: The National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Retrieved from http://www.cse.ucla.edu/products/reports/r632.pdf.

Sellberg, C. (2017). Simulators in bridge operations training and assessment: a systematic review and qualitative synthesis. WMU Journal of Maritime Affairs, 16(2), 247-263.

Stefanidis, D., Korndorffer, J.R., Markley, S., Sierra, R., Heniford, B.T., & Scott, D.J. (2007). Closing the gap in operative performance between novices and experts: does harder mean better for laparoscopic simulator training? Journal of the American College of Surgeons, 205(2), 307-313.

Weinert, F. E. (2001): Competencies and Key Competencies: Educational Perspective. International Encyclopedia of the Social and Behavioral Sciences, vol. 4, Elsevier, 2433–2436.