SyRRA2018-001

Exploration of methods for using SACADA data to estimate HEPs: Final Report

Katrina M. Groth, Reuel Smith, and Ramin Moradi

December, 2018

Systems Risk and Reliability Analysis Lab
Center for Risk and Reliability
University of Maryland
College Park, MD 20742

SYRRA

CENTER FOR RISK AND RELIABILITY

Exploration of methods for using SACADA data to estimate HEPs: Final Report

SyRRA2018-001

December 2018

Katrina M. Groth, Reuel Smith, and Ramin Moradi

Department of Mechanical Engineering

Systems Risk and Reliability Analysis Lab

4298 Campus Dr.

University of Maryland

College Park, MD 20742

Abstract

This report provides summary of the project “Exploration of methods for using SACADA data to estimate HEPs”. The goal of the project was to conduct exploratory research on how to use the U.S. Nuclear Regulatory Commission’s SACADA (Scenario, Authoring, Characterization, and Debriefing Application) database to develop an algorithm for estimating human error probabilities (HEPs). The approach used by the University of Maryland SyRRA lab uses a combination of Bayesian statistical methods and Bayesian Network models to conduct data analysis on SACADA data and to construct hybrid models informed by both data and engineering models. The end results provided various algorithms for mapping and binning SACADA data to be used within HEP estimation, and demonstrated a variety of options which create a framework for streamlined updating of HEPs as the amount and variety of SACADA data increases. This report summarizes the project's major accomplishments, and gathers the abstracts and references for the publication submissions and reports that were prepared as part of this work.

Introduction

Human Reliability Analysis (HRA) is the aspect of Probabilistic Risk Assessment (PRA) that is concerned with systematically identifying and analyzing the causes and consequences of human errors. There are numerous HRA methods that provide guidance for determining the human error probability (HEP), which is the (conditional) probability of a human failure event (HFE), given the context of various parameters and contextual factors which are often called performance influencing factors (PIFs).

A critical challenge for the field of HRA is the need for traceable, data-informed models which provide a defensible basis for risk-informed decision making. Currently the NRC is pursing data collection through the recently developed SACADA framework and database. SACADA provides a common basis and structure for HRA data collection, analysis, and exchange. Currently the structure of the SACADA database is final. However, the SACADA database is continually growing in content. According to recent information from NRC, SACADA currently contains the results from approximately 35 scenarios with results coming from both published international simulator experiments and non-published operator training activities. At this stage, an important research question emerges: how can the SACADA data be used to improve HEP estimation for HRA?

This report describe the results of a project "Exploration of methods for using SACADA data to estimate HEPs". The goal of the work conducted by the University of Maryland SyRRA lab is to use of Bayesian statistical methods and Bayesian Network models to conduct data analysis on SACADA data and to construct hybrid models informed by both data and engineering models. The end results provided various algorithms for mapping and binning SACADA data to be used within HEP estimation, and demonstrated a variety of options which create a framework for streamlined updating of HEPs as the amount and variety of SACADA data increases.

The project produced the following results: (1) An algorithm for using SACADA data to develop a qualitative and quantitative basis for estimating human error probabilities (HEPs) and (2) documentation of literature findings, methodology, approach, and findings in presentations and reports.

Summary of Work

The primary objective the proposed work was to explore approaches and methods for using the SACADA context elements and data to enhance the estimation of HEPs. To achieve this objective, the SyRRA group proposed to use Bayesian statistical methods and Bayesian Networks to enable the use of qualitative and quantitative information from SACADA. Bayesian methods were

selected because of their ability to formally combine various types of data for use in predictive model building as well as in developing numerical estimates for human error probabilities (HEPs). A key benefit of the Bayesian approach for HRA is that it enables various amounts of simulator data to be used to enhance the technical basis of HRA methods, and furthermore, it is an iterative process that can be used to continually improve HRA as the data content of SACADA expands [1].

There are multiple activities which were conducted in support of this goal. Activities included:

- A literature review and familiarization with SACADA structure and data types

- Developing a mapping that relates elements of SACADA (including the contextual factors, the scenario elements, and the performance outcomes) onto various HRA concepts (including human error probability (HEP), performance shaping factors (PSFs)).
- Running mathematical and statistical analyses designed to visualize, categorize, and cluster the data and parameterize models.
- Development of a demonstration or materials to effectively communicate the methods and results

The first step was to develop the framework for the analysis; this is documented in a conference paper [2] presented at the PSAM conference and enclosed in this report as Appendix A. A second activity involved development of the full algorithm; this has been documented in a draft journal paper which is enclosed in this report as Appendix B.

Key challenges were encountered in enabling SACADA data to be used within HEP quantification across multiple performance contexts and multiple levels of abstraction. It is necessary to map multiple data sources (SACADA, cognitive literature, existing HRA methods) onto a single, consistent terminology framework. However, each source of data contains up to 200 relevant variables or data collection elements, and completing this mapping process is recommended as follow-on work.

As part of the demonstration development, we sought to develop a MATLAB demo of the algorithm. However, we ran into some limitations in available prior information, the need to map over 200 variables, and code modernization needs of the Bayes Net Toolbox (BNT) [3] for MATLAB [4]. We envision that the Matlab demo would contain a series of MATLAB routines to extract and catalog the necessary states as well as tally their frequency according to their SACADA situational factors. Further based on the SACADA states, the code also extracts the state associated PIFs and their characteristics based on the initial state-to-PIF map that was designed for this project. BNs for each MCF would be developed and implemented via the BNT. However because the BNT is no longer maintained by the original creators, several modifications would be necessary in order to update the package to be compatible with the current version of Matlab. Development of a Matlab demo is recommended as part of follow-on work.

Appendix C contains information gathered to be used in a demonstration of the algorithm. Appendix C1 contains a table of published, expert-elicited probabilities for specific PSFs used in the SPAR-H model [5, 6]). These could be used to generate prior probabilities for the Pr(PSF), but would require mapping from the variables in SPAR-H onto the variables in the causal models.

Appendix C2 contains graphical models which were developed in GeNie [7], based on a study by Whaley et. al [8]. We created models for the in macro cognitive function (MCF) "failure of detecting and noticing" to demonstrate the concept. For a full-scale model, it would be necessary to model the additional MCFs in the future [9, 8]. The setup for this involved writing fifteen associated causal graphical models relating the PIFs (from the hierarchy defined by Groth and Mosleh [10] (see Appendix C3). The current graphical models assume PIF independence; this simplification was necessary to enable meeting project objectives within the project timeframe; an important next step is to revise the BNs to include PIF-to-PIF dependencies to correctly capture the cognitive literature via the algorithm from [11] which is incorporated in our algorithm.

Finally, the SACADA states that were provided in the data spreadsheets required a correlation map that defined the states as PIFs and PIF characteristics defined by Groth and Mosleh [10] and by Chang et. al [12]. This was an extensive mapping project where the SyRRA group had to define PIFs and characteristics for over 200 separate SACADA states where a draft is presented in Appendix C4.

Publications & Presentations

- Final copy of PSAM conference paper "A framework for using SACADA to enhance the qualitative and quantitative basis of HRA" by Katrina Groth [2] (Enclosed in this report as Appendix A).

  K. Groth, "A framework for using SACADA to enhance the qualitative and quantitative basis of HRA," in *Probabilistic Safety Assessment and Management PSAM 14*, September 17-21, 2018, Los Angeles, CA, 2018.
- Dr. Groth organized a special session on HRA data analysis at PSAM 14. The session consisted of 2 paper sessions (comprising 8 technical papers) and 1 panel session consisting of 5 invited speakers with Dr. Groth serving as moderator.
- A draft of a journal paper "A hybrid approach to HRA using simulator data, causal models, and cognitive science" by Katrina Groth, Reuel Smith, and Ramin Moradi [9] (Enclosed in this report as Appendix B).

  K. M. Groth, R. C. Smith and R. Moradi, "A hybrid approach to HRA using simulator data, causal models, and cognitive science," *In Production,* 2018.
- (Presentation) K. Groth "Exploration of methods for using SACADA data to Bayesian update HEPs" at the U. S. Nuclear Regulatory Commission Human Reliability Analysis Data Workshop, March 15, 2018.

A framework for using SACADA to enhance the qualitative and quantitative basis of HRA

Katrina M. Groth*

Systems Risk and Reliability Analysis Lab, Center for Risk and Reliability, Department of Mechanical Engineering, University of Maryland, College Park, MD

Abstract: The purpose of this research is to explore ways to use Bayesian methods with data from the US Nuclear Regulatory Commission's (NRC) Scenario, Authoring, Characterization, and Debriefing Application (SACADA) system. SACADA is a database designed to enable collection of nuclear power plant (NPP) control room simulator and crew training data to improve both operator training and human reliability analysis (HRA). This paper presents a framework to use SACADA data and causal modeling to enhance the qualitative and quantitative aspects of HRA. The framework is a multi-faceted approach involving causal models as well as multiple sources and types of data. Elements of the framework include a comprehensive set of performance influencing factors, crew failure modes, Bayesian Network causal models, Bayesian parameter updating, and temporal modeling. This paper also outlines a path forward for developing the framework to enhancing the technical basis of HRA and enabling streamlined use of SACADA data as the volume and variety of data increases.

Keywords: HRA data, Bayesian updating, SACADA, Bayesian Networks

1. INTRODUCTION

A critical challenge for the field of human reliability analysis (HRA) is the need for traceable, data-informed models that provide a defensible basis for risk-informed decision making. Currently the U.S. Nuclear Regulatory Commission (NRC) is pursing data collection through the recently developed Scenario, Authoring, Characterization, and Debriefing Application (SACADA) framework and database [1]. The SACADA database is one of several international data collection activities focused on collecting human performance data from nuclear power plant (NPP) control room simulator scenarios. It also offers a common basis and structure for HRA data collection, analysis, and exchange. SACADA is actively being populated with data and, in parallel, the initial data is being analyzed to provide insight into the development and use of SACADA [2].

As SACADA and similar databases become more mature, an important research question emerges: *how can this data be used to improve HRA?* As a first step toward answering that question, the U.S. NRC asked three teams to develop methods for using SACADA data to quantify the probability of a human failure event associated with a given performance context. In HRA, this quantity is called the human error probability (HEP), which is a conditional probability with the conditioning factors representing the context of performance in terms of performance influencing factors (PIFs), also called performance shaping factors (PSFs). This paper is one outcome of the analysis being conducted by the University of Maryland. The methods developed by the two other teams are also presented in papers at this conference [3], [4].

This paper defines a framework for using SACADA data to improve both the qualitative and quantitative basis of HRA using causal Bayesian Networks and Bayesian parameter updating. The SACADA data is described in Section 2 of this paper. Section 3 describes the approach to development of the framework. The proposed framework is described in Section 4.

* kgroth@umd.edu

2. DESCRIPTION OF SACADA DATA

2.1. Database structure

The SACADA data taxonomy is described in detail in [1] and the current state of development is also described in another paper in this conference [2]. SACADA collects data about multiple facets of crew and system (machine) performance in NPP control rooms. SACADA provides a detailed structure for collecting information about the characteristics of a simulator scenario, the plant conditions, a task-level breakdown of activities involved in responding to conditions in those scenarios, crew roles, and the performance outcome (aggregated at the task-level) for multiple crews that have performed the scenario. SACADA provides a common set of elements for capturing the context of the scenario (which is generally the same across all crews), and a set of performance factors that are used for debriefing crews when performance is considered less-than-satisfactory.†

Each scenario starts with a plant initial condition and contains one or more plant malfunctions that are pre-programmed to occur in the simulator during the exercise. These scenarios are designed before the crews are run on the scenario, and multiple crews run each scenario.

Each scenario is decomposed into a series of tasks or *training objective elements (TOE),* each of which represents one activity that the crews must complete to respond to the specified plant condition or malfunction. The TOE is the basic data unit for SACADA. In general, there are several TOEs involved in the response to each malfunction in a scenario. Each TOE is associated with one of five macrocognitive functions: monitoring/detecting; diagnosis; response planning/decision making; manipulation/execution; and communication/coordination.‡ Example TOEs include things like "ensure the charging cooling pump 1A is in service," "announce transition to procedure [number]" and "monitor [system X] and identify increasing trend in pressure." From an HRA perspective, these TOEs resemble tasks or sub-tasks. Many HRA methods are designed to be used at the scenario or event level, and thus would include multiple TOEs.

For each TOE, the data collection team characterizes the context of the scenario using situational factors (SFs), which are similar to the HRA concepts of PIFs. There are approximately 29 SFs in SACADA, although only a subset are used for each TOE depending on which type of macrocognitive function is associated with the TOE. Some of these factors are rated as one of two states (present/absent), some have up to four states, and some represent a summation of multiple constituent factors rated on the two point (present/absent) scale, as described in [1]. The SFs document the context of each TOE, and they do not change depending on which crew is running the scenario. The outcome of the crew performance for each TOE is ranked on a four-point scale (ordered from best to worst performance, where "SAT" is an abbreviation of the word satisfactory): SAT+, SAT, SATΔ, UNSAT.

For crews that receive a score of SATΔ or UNSAT, a second worksheet is completed to capture the causes of the degraded performance. The worksheet contains approximately 21 performance factors (PFs) that are also similar to PIFs. As with SFs, only a subset of the PFs are used depending on which macrocognitive functions are involved. Several of the PF factors are rated as being in one of two states (present/absent), some have up to six states, and some represent a summation of multiple factors rated on the two point (present/absent) scale as described in [1]. These PFs are used to describe the reasons for errors (or near misses) for the crew on a specific TOE.

2.2. Current data

† Satisfactory performance is defined from a training perspective. Use of this data for HRA purposes requires some additional considerations, which are described later in this paper.
‡ In SACADA, this is defined as "external communication," meaning communication beyond the crew or team. It is worth noting that this may be a narrower interpretation of the macrocognitive concept of "communication," which also involves the concept of teamwork and within-team communication.

This section provides summary information about the current SACADA data set provided for descriptive purposes. It is important to caution against misinterpretation of the summary information provided in this section. Because of the causal nature of the underlying factors, the author cautions that **no statistics of conclusions about HRA or human performance should be drawn from these descriptive data, because they are aggregated across contexts and because not all combinations of contexts are represented in the current data.** Some contexts may be over- or under-represented in the training data when compared to the contexts experienced in real operational events.

As of July 2017, SACADA contains data from 86 simulator scenarios with results coming from both published international simulator experiments and non-published operator training activities. Within these 86 scenarios, there are 329 malfunctions and a total of 2,155 TOEs. Each scenario was performed by several crews, and on average each TOE was performed by 12 crews. In total, the current SACADA database contains 26,153 crew-TOEs (this number represents the sum-product of the 2,155 TOEs and the variable number of crews that attempted each TOE). Of the 2,155 TOEs, 149 TOEs had one or more crews with a rating of UNSAT and 219 TOEs with a rating of SATΔ. Of the 26,153 crew-TOE data points, there were 209 scores of UNSAT and 261 scores of SATΔ.

3. APPROACH TO DEVELOPMENT OF THE FRAMEWORK

After initial analysis of the SACADA data and review of existing HRA needs, the next step was to consider the desirable factors of the framework for using this data to enhance HRA. The approach to development of the proposed method was based on the desirable characteristics of advanced HRA methods outlined by many HRA studies (e.g., [5]). In particular, it was decided that during the first stage of this work, the method should meet the following criteria:

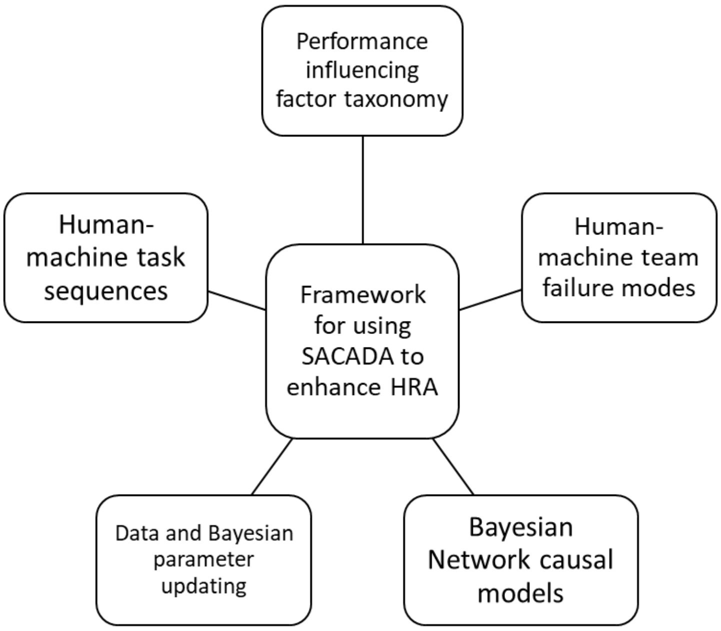
1. The proposed method should be based on a causal model of human-machine performance.[§] That model should be rooted in both cognitive science and systems science.
2. The structure of the model should provide explicit representation of the causal factors that affect human-machine performance.
3. The method should support the quantitative and qualitative aspects of HRA as a part of probabilistic risk assessment (PRA), including quantification of HEPs.
4. The method should provide a framework that is both data-informed and model-informed.
5. The method should be flexible enough to accommodate changes in SACADA structure as SACADA is developed further.
6. The method should be able to incorporate additional data, models, and information to address contexts or factors that are not represented in SACADA.
7. The framework should be capable of providing quantitative insights that can be used to help the data collection teams improve human performance (e.g., via training).

4. PROPOSED FRAMEWORK

The main elements that are included in the proposed framework are illustrated in Figure 1. These elements are a comprehensive set of PIFs, human failure modes, Bayesian Network causal models, Bayesian parameter updating, and temporal evolutional of performance. Each of these elements is described in more detail in this section.

[§] I use the term "human-machine performance" to emphasize that there is an important role of both machines and humans in the concept of human reliability.

Figure 1: Main elements of the proposed framework for using SACADA to enhance HRA



4.1 A comprehensive set of PIFs

The first element in the proposed framework is a comprehensive taxonomy of PIFs [6]. The taxonomy provides a consistent vocabulary and structure for combining data and information from multiple sources and at multiple levels of detail in a transparent and repeatable way. The Groth and Mosleh taxonomy also provides non-overlapping, orthogonal set of causal factors, meaning that each PIF is defined uniquely. The term "orthogonal" is used to indicate that, while the factors are uniquely defined, they may not be independent in a statistical or causal sense.

4.2 Human-machine team failure modes

This aspect of the framework involves having a defined set of failure modes for the human-machine team. These capture the ways that the human-machine teams could fail in responding to a condition or malfunction. The term "human-machine failure mode" is designed to reflect that the HRA concept of a human failure event (HFE) involves a contribution from both the human response and the machine response. In the proposed framework, each of these failure modes would be represented in a causal model as described in Section 4.3.

The definition of these human-machine team failures modes could be achieved via multiple approaches. One option is to use the failure modes or failure-mode-identification approaches defined by the existing HRA methods, which acknowledge both a cognitive and a machine element (e.g., methods such as IDA **[7]**, IDAC **[8]** IDHEAS **[9]**, or PHOENIX **[10]** and HRA research activities **[11]–[15]**). Another option is to use first-principles reliability techniques (e.g. by doing a "human-machine" FMEA or HAZOP with consideration of macrocognitive functions and machine functions). Another approach would be to define one failure mode for each of the five macrocognitive functions used in SACADA.

4.3 Bayesian Network (BN) causal models

In the proposed framework, BN causal models are used to capture the detailed causal pathways and interdependencies among PIFs. In this framework, BNs are used because of their ability to model cause-and-effect relationships and the use of probabilistic inference. Furthermore, BNs provide the ability to reason about any variable in the model, enabling quantitative insights relevant to improving performance.

This approach will allow all relevant PIFs (both observable and unobservable) to be used in model development. In addition, it enables the explicit inclusion of data collection elements within the model structure, and these data collection elements could enable using data from different versions of SACADA.

These causal relationships would be initially developed from cognitive science and systems science, and could eventually be developed or validated by using SACADA data. The BN model will be developed by creating an explicit map between PIFs (first element described above) and the human-machine team failure modes (second element described above). This could be accomplished by using the causal mapping approach developed and illustrated in the work of Zwirglmaier, Straub, and Groth **[16]**. The starting point for development of this causal mapping has been developed as part of NRC's work on defining a cognitive basis for HRA **[17], [18]** and in the IDAC model **[19]**. The size of the resulting BN models could be reduced (e.g., to facilitate quantification) by using node reduction algorithms **[16]**. A second approach would be to follow the approach of **[20]** and use factor analysis clustering, or structure learning algorithms techniques directly on the SACADA data.

4.4 Data and Bayesian parameter updating

The quantitative parameters of the BN models would be populated using multiple sources of data. Prior information on the relationship between the PIFs and the human-machine failure modes could be defined by using an existing HRA method (e.g., as illustrated in **[21]** using SPAR-H **[22]**) or other HRA databases (e.g., **[23]–[27]**). Prior information on the PIF prior probabilities could be assembled from a variety of published sources in cognitive science and HRA, or via formal expert elicitation with HRA experts. The SACADA data would be used as information to Bayesian update multiple parameters within the model, using a Bayesian updating approach **[28]**.

There are several reasons to include this aspect in the framework. First, it allows the causal model to be populated with appropriate information from multiple sources, including data. It also enables the model to include variables and information which are unlikely to be represented in the data (e.g., control room environment). This also provides a way to address the reality that some factors will be over- (or under-) weighted in the training contexts (e.g., high task complexity).

4.5 Human-machine task sequences

The final aspect of the framework involves modeling the sequential aspects of human-machine activities associated with the response to a malfunction. This addresses the need to treat an HFE as the outcome of a process involving several sequential activities or tasks involving different macrocognitive functions. This notion that human failure involves a series of activities and that there is dependency between HFEs has been acknowledged in even the earliest HRA methods **[29]**. It is considered in depth in simulation-based methods which explicitly model sequences of activities or subtasks involved in PRA event (e.g., **[7], [8], [30], [31]**). This sequential and semi-temporal aspect of performance needs to be modeled explicitly in order to reflect the fact that that failure is a process, not a single event.

SACADA provides the first opportunity to use data inform this process. This opportunity arises as a result of several coupled aspects of the data. The first aspect is the detailed consideration of TOEs at the level of macrocognitive tasks (rather than at the higher event level used in many HRA methods). The second is the alignment of these TOEs with responses to specific malfunctions (more akin to the event level in HRA), and the third is that the PIFs are collected at the TOE level. This provides new potential to transform the treatment of sequential and temporal aspects of the PIFs and the activities that comprise an HRA event. A mechanism for modeling this dependency would be through the use of Dynamic Bayesian Networks, which were first introduced in **[12], [32], [33]** and further expanded within the HUNTER framework **[30], [31], [34]**.

5. CONCLUSION

The SACADA database provides a unique opportunity to enhance the foundations of HRA. This work provides a new framework for enhancing the foundations of HRA by using SACADA data together with scientific information and new modeling approaches. The next steps of this work involve further developing each aspect of the proposed framework. The framework proposed in this work will enable a path toward an HRA vision that is both model-based and data-informed, enhancing the technical basis of HRA and enabling streamlined use of SACADA data as the volume and variety of SACADA data increases.

Acknowledgements

This research was partially supported by the U.S. Nuclear Regulatory Commission (NRC). The content is solely the responsibility of the authors and does not represent the official views of the NRC.

References

[1] Y. J. Chang *et al.*, "The SACADA database for human reliability and human performance," *Reliability Engineering & System Safety*, vol. 125, no. 0, pp. 117–133, 2014.
[2] Y. H. J. Chang and C. Franklin, "SACADA Data for HEP Estimates," presented at the Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 14), Los Angeles, CA, 2018.
[3] P. Nelson and R. Grantom, "Methodology for Determination of Human Error Probabilities from Simulator Sourced Data," presented at the Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 14), Los Angeles, CA, 2018.
[4] M. A. Azarm, I. S. Kim, C. Marks, and F. Azarm, "Analyses methods and pilot applications of SACADA Database," presented at the Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 14), Los Angeles, CA, 2018.
[5] A. Mosleh and Y. H. Chang, "Model-based human reliability analysis: prospects and requirements," *Reliability Engineering and System Safety*, vol. 83, no. 2, pp. 241–253, Feb. 2004.
[6] K. M. Groth and A. Mosleh, "A data-informed PIF hierarchy for model-based Human Reliability Analysis," *Reliability Engineering and System Safety*, vol. 108, pp. 154–174, Dec. 2012.
[7] S. H. Shen, C. Smidts, and A. Mosleh, "IDA: a cognitive model of nuclear power plant operator response during abnormal conditions," PhD Thesis, University of Maryland at College Park, 1994.
[8] Y. H. J. Chang and A. Mosleh, "Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents: Part 1 - Overview of the IDAC Model," *Reliability Engineering and System Safety*, vol. 92, no. 8, pp. 997–1013, Aug. 2007.
[9] J. Xing, G. Parry, M. Presley, J. Forester, S. Hendrickson, and V. Dang, "An Integrated Human Event Analysis System (IDHEAS) for Nuclear Power Plant Internal Events At-Power Application," NUREG-2199, Volume 1, May 2017.
[10] N. J. Ekanem, A. Mosleh, and S.-H. Shen, "Phoenix – A model-based Human Reliability Analysis methodology: Qualitative Analysis Procedure," *Reliability Engineering & System Safety*, vol. 145, pp. 301–315, Jan. 2016.
[11] J. Oxstrand, D. L. Kelly, S.-H. Shen, A. Mosleh, and K. M. Groth, "A Model-Based Approach to HRA: Qualitative Analysis Methodology," in *Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 11)*, Helsinki, Finland, 2012.
[12] A. Mosleh, S.-H. Shen, D. L. Kelly, J. H. Oxstrand, and K. Groth, "A Model-Based Human Reliability Analysis Methodology," in *Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 11)*, Helsinki, Finland, 2012.
[13] K. M. Groth, S.-H. Shen, J. Oxstrand, A. Mosleh, and D. Kelly, "A Model-Based Approach to HRA: Example Application and Quantitative Analysis," in *Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 11)*, Helsinki, Finland, 2012.

[14] J. Park, Y. Kim, and W. Jung, “A framework to estimate HEPs from the full-scope simulations of NPPs: unsafe act definition, identification, and quantification,” Korea Atomic Energy Research Institute (KAERI), Daejeon, Korea, KAERI/TR-6401/2016, Jun. 2016.
[15] Y. Kim, J. Park, and W. Jung, “A classification scheme of erroneous behaviors for human error probability estimations based on simulator data,” *Reliability Engineering & System Safety*, vol. 163, pp. 1–13, 2017.
[16] K. Zwirglmaier, D. Straub, and K. M. Groth, “Capturing cognitive causal paths in human reliability analysis with Bayesian network models,” *Reliability Engineering & System Safety*, vol. 158, pp. 117–129, Feb. 2017.
[17] A. M. Whaley *et al.*, “Building a Psychological Foundation for Human Reliability Analysis,” US Nuclear Regulatory Commission, Washington DC, NUREG-2114, Aug. 2012.
[18] A. M. Whaley *et al.*, “Cognitive Basis for Human Reliability Analysis,” US Nuclear Regulatory Commission, Washington DC, NUREG-2114, Jan. 2016.
[19] Y. H. J. Chang and A. Mosleh, “Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents. Part 4: IDAC causal model of operator problem-solving response,” *Reliability Engineering and System Safety*, vol. 92, no. 8, pp. 1061–1075, 2007.
[20] K. M. Groth and A. Mosleh, “Deriving causal Bayesian networks from human reliability analysis data: A methodology and example model,” *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 226, no. 4, pp. 361–379, Aug. 2012.
[21] K. M. Groth and L. P. Swiler, “Bridging the gap between HRA research and HRA practice: A Bayesian Network version of SPAR-H,” *Reliability Engineering and System Safety*, vol. 115, pp. 33–42, Jul. 2013.
[22] D. Gertman, H. Blackman, J. Marble, J. Byers, and C. Smith, “The SPAR-H Human Reliability Analysis Method,” US Nuclear Regulatory Commission, Washington DC, NUREG/CR-6883, 2005.
[23] J. Park and W. Jung, “A database for human performance under simulated emergencies of nuclear power plants,” *Nuclear Engineering and Technology*, vol. 37, no. 5, p. 491, 2005.
[24] J. Park and W. Jung, “OPERA–a human performance database under simulated emergencies of nuclear power plants,” *Reliability Engineering and System Safety*, vol. 92, no. 4, pp. 503–519, 2007.
[25] A. B. Skjerve and A. Bye, Eds., *Simulator-based Human Factors Studies Across 25 Years: The History of the Halden Man-Machine Laboratory*. London: Springer, 2010.
[26] S. Y. Choi and W. Jung, “Qualitative human event analysis with simulator data by using HuRAM+ and HERA,” in *Proceedings of the European Society for Reliability Annual Meeting (ESREL 2013)*, Amsterdam, 2013.
[27] B. Hallbert *et al.*, “Human Events Repository Analysis (HERA) System Overview,” US Nuclear Regulatory Commission, Washington DC, NUREG/CR-6903, Vol. 1, 2006.
[28] K. M. Groth, C. L. Smith, and L. P. Swiler, “A Bayesian method for using simulator data to enhance human error probabilities assigned by existing HRA methods,” *Reliability Engineering & System Safety*, vol. 128, pp. 32–40, 2014.
[29] A. D. Swain and H. E. Guttmann, “Handbook of Human Reliability Analysis with emphasis on nuclear power plant applications,” US Nuclear Regulatory Commission, Washington DC, NUREG/CR-1278, 1983.
[30] R. Boring *et al.*, “Human Unimodel for Nuclear Technology to Enhance Reliability (HUNTER): A Framework for Computation-based Human Reliability Analysis,” in *Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 13)*, Seoul, Korea, 2016.
[31] R. Boring, D. Mandelli, J. Joe, C. Smith, and K. Groth, “A Research Roadmap for Computation-Based Human Reliability Analysis,” Idaho National Laboratory, Idaho Falls, ID, INL/EXT-15-36051, Jul. 2015.
[32] K. M. Groth, “A Data-Informed Model of Performance Shaping Factors for Use in Human Reliability Analysis,” PhD Thesis, University of Maryland, College Park, MD, 2009.

[33] N. J. Ekanem and A. Mosleh, “Human failure event dependency modeling and quantification: A Bayesian network approach,” in *Proceedings of the European Society for Reliability Annual Meeting (ESREL 2013)*, Amsterdam, 2013.
[34] R. Boring *et al.*, “Integration of Human Reliability Analysis Models into the Simulation-Based Framework for the Risk-Informed Safety Margin Characterization Toolkit,” Idaho National Laboratory, Idaho Falls, ID, INL/EXT-16-39015, Jun. 2016.

A hybrid approach to HRA using simulator data, causal models, and cognitive science

Katrina M. Groth, Reuel Smith, Ramin Moradi

[a]Systems Risk and Reliability Analysis Lab (SyRRA), Center for Risk and Reliability, University of Maryland, College Park, MD, 20742, USA

Abstract

In this paper we define a methodology for using a multiple types of data to advance the field of Human Reliability Analysis (HRA). The methodology uses causal models built from and parameterized by a combination of cognitive literature, existing HRA methods, simulator data, and expert elicitation. The main elements of the framework include a comprehensive set of performance influencing factors, human-machine failure modes, Bayesian Network causal models, and Bayesian parameter updating. The methodology enhances both the qualitative and the quantitative basis of HRA, adding significant scientific depth and technical traceability to the highly complicated problem of modeling human-machine performance in complex systems.

Keywords: Bayesian networks, human reliability analysis, macro cognitive function, HRA data, Bayesian updating

1. Introduction

The need for data-informed and traceable models which enable risk-informed decision making is one of the main challenges in the field of Human Reliability Analysis (HRA).[1] In this regard, several international organizations are pursing data collection through comprehensive control room simulator studies. SACADA [2], OPERA [3], and HURAM+ [4] are among several HRA databases developed from nuclear power plant (NPP) control room simulator studies conducted under the auspices of the U.S. NRC, international research organizations, and partner utilities, which provide a framework for HRA data collection, analysis and exchange. SACADA is continuously being used to generate new data while the formerly produced data is being thoroughly analyzed for developing and optimizing the use of SACADA [5].

As these databases become more mature, an important research question emerges: *how can this data be used to improve HRA*? As the first step towards answering this question, we explore how to use multiple elements of this simulator data together with causal models to enhance HRA.

The rest of this paper is organized as follows. Section 2 describes the approach to development of the framework. SACADA data structures and its embedded features are described in Section 3 of this paper while Section 4 discusses further elements of the framework. The algorithm used for developing this framework is explained systematically in Section 5. Finally, this paper ends with summary and conclusions in Section 6.

2. Approach to the Development of the Framework

After initial familiarization with the available data, the first step was to consider the desirable factors of the framework for using this data to enhance HRA. The approach to development of the proposed method was based on the desirable characteristics of advanced HRA methods outlined by many HRA studies (e.g., [6]). In particular, it was decided that during the first stage of this work, the method should meet the following criteria:

1. The proposed method should be based on a causal model of human-machine performance.[2]
2. The structure of the model should provide explicit representation of the causal factors that affect human-machine performance based on both cognitive science and systems science.
3. The method should support the quantitative and qualitative aspects of HRA as a part of probabilistic risk assessment (PRA), including quantification of HEPs.
4. The method should provide a framework that is both data-informed and model-informed.
5. The method should be flexible enough to accommodate changes in database structure
6. The method should be able to incorporate additional data, models, and information to address contexts or factors that are not represented in the simulator data sources
7. The framework should be capable of providing quantitative insights that can be used to help the data collection teams improve human performance (e.g., via training).

Email addresses: kgroth@umd.edu (Katrina M. Groth), smithrc@umd.edu (Reuel Smith), raminmrd@umd.edu (Ramin Moradi)

[1]This paper is significantly extended version of a paper presented at PSAM 14 [1]

[2]We use the term “human-machine performance” to emphasize that there is an important role of both machines and humans in the concept of human reliability.

3. Description of SACADA Data

3.1. Database Structure

Chang et. al [2] has described the SACADA data taxonomy thoroughly. The current state of development is also described in Chang and Franklin [5]. SACADA data consists of recorded data about multiple facets of crew and system (machine) performance in NPP control rooms. A detailed structure for collecting information about the plant conditions, the characteristics of a simulator scenario, a task-level classification of activities involved in responding to conditions in those scenarios, crew roles, and the performance outcome (aggregated at the task-level) for multiple crews that have performed the scenario is provided by SACADA. SACADA provides a common set of elements for capturing the context of the scenario (which is generally the same across all crews), and a set of performance factors that are used for debriefing crews when performance is considered less-than-satisfactory.[3]

These scenarios are designed before the crews are run on the scenario, and multiple crews run each scenario. Each scenario begins at a plant initial condition and contains one or more plant malfunctions that are purposefully programmed to occur during the exercise. Each scenario has a correct response that can be decomposed into a series of tasks or training objective elements (TOE), each of which represents one activity that the crews must complete to respond to the specified plant condition or malfunction. Each TOE is associated with one of five macro cognitive functions: monitoring/detecting; diagnosis; response planning/decision making; manipulation/execution; and communication/coordination.[4] Example TOEs include things like "ensure the charging cooling pump 1A is in service," "announce transition to procedure [number]" and "monitor [system X] and identify increasing trend in pressure." From an HRA perspective, these TOEs resemble tasks or sub-tasks. Many HRA methods are designed to be used at the scenario or event level, and thus would include multiple TOEs.

The data collection team characterizes the context of each scenario using situational factors (SFs) (which are similar to the HRA concepts of PIFs), for each TOE. The SFs document the context of each TOE, and they remain the same no matter which crew is running the scenario. In SACADA, There are approximately 29 SFs in SACADA. Some of these factors are rated as one of two states (present/absent), some have up to four states, and some represent a summation of multiple constituent factors rated on the two point (present/absent) scale, as described in Chang et. al [2]. A subset of these SFs are assessed for each TOE, and which subset is assessed is defined by the type of Macro Cognitive function associated with the TOE.

The outcome of the crew performance for each TOE is ranked on a four-point scale (ordered from best to worst performance, where "SAT" is an abbreviation of the word satisfactory) [2]:

- *SAT+*: Outstanding crew performance
- *SAT*: Satisfactory crew performance that meets performance requirements
- *SAT*Δ: Unsatisfactory crew performance where performance requirements are met but with deficiencies
- *UNSAT*: Unsatisfactory crew performance where performance requirements are not met

When a crew receives a score of SATΔ or UNSAT, to capture the causes of a degraded performance a second worksheet is completed. The worksheet contains approximately 21 performance factors (PFs) that are also similar to PIFs. As with SFs, only a subset of the PFs are used depending on which macro cognitive functions are involved. Several of the PF factors are rated as being in one of two states (present/absent), some have up to six states, and some represent a summation of multiple factors rated on the two point (present/absent) scale as described in [2]. These PFs are used to describe the reasons for errors (or near misses) for the crew on a specific TOE.

3.2. Current Data

This section provides summary information about the current SACADA data set provided for descriptive purposes. Because of the causal nature of the underlying factors, the author cautions that **no statistics of conclusions about HRA or human performance should be drawn from these descriptive data, because they are aggregated across contexts and because not all combinations of contexts are represented in the current data**. Some contexts may be over- or under-represented in the training data when compared to the contexts experienced in real operational events.

As of July 2017, SACADA contains data from 86 simulator scenarios with results coming from both published international simulator experiments and non-published operator training activities. Within these 86 scenarios, there are 329 malfunctions and a total of 2,155 TOEs. Each scenario was performed by several crews, and on average each TOE was performed by 12 crews. In total, the current SACADA database contains 26,153 crew-TOEs (this number represents the sumproduct of the 2,155 TOEs and the variable number of crews that attempted each TOE). Of the 2,155 TOEs, 149 TOEs had one or more crews with a rating of UNSAT and 219 TOEs with a rating of SATΔ. Of the 26,153 crew-TOE data points, there were 209 scores of UNSAT and 261 scores of SATΔ.

4. Elements of the framework

The main elements of the proposed framework were defined in the conference paper by Groth [1] (including a comprehensive set of PIFs, human failure modes, Bayesian Network causal models, Bayesian parameter updating, and temporal evolution of performance) are illustrated in Figure 1.

[3] Satisfactory performance is defined from a training perspective. Use of this data for HRA purposes requires some additional considerations, which are described later in this paper.

[4] In SACADA, this is defined as "external communication," meaning communication beyond the crew or team. It is worth noting that this may be a narrower interpretation of the macro cognitive concept of "communication," which also involves the concept of teamwork and within-team communication.

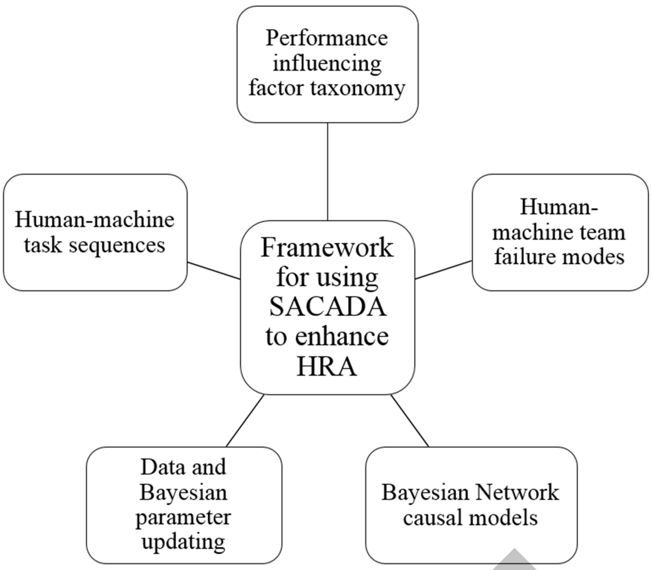


Figure 1: The main elements of the proposed framework for using SACADA to enhance HRA.

4.1. A comprehensive set of PIFs

The first element in the proposed framework is a comprehensive taxonomy of PIFs, such as the one from Groth and Mosleh [7]. The taxonomy provides a consistent vocabulary and structure for combining data and information from multiple sources and at multiple levels of detail in a transparent and repeatable way. The taxonomy also provides non-overlapping, orthogonal set of causal factors, meaning that each PIF is defined uniquely. The term "orthogonal" is used to indicate that, while the factors are uniquely defined, they may not be independent in a statistical or causal sense.

The data elements from SACADA and other databases would be mapped onto this taxonomy. The use of such a taxonomy is important because it provides a comprehensive list of PIFs and clear definitions. By contrast, PIFs collected in most HRA data sources are limited to the PIFs that can be observed or collected in that performance context.

4.2. A set of human-machine team failure modes

This aspect of the framework involves having a defined set of failure modes for the human-machine team. These capture the ways that the human-machine teams could fail in responding to a condition or malfunction. The term "human-machine failure mode" is designed to reflect that the HRA concept of a human failure event (HFE) involves a contribution from both the human response and the machine response. The definition of these human-machine team failures modes could be achieved via multiple approaches. One option is to use the failure modes or failure-mode-identification approaches defined by the existing HRA methods, which acknowledge both a cognitive and a machine element (e.g., methods such as IDA [8], IDAC [9], IDHEAS [10], or PHOENIX [11, 12] and HRA research activities [13, 14, 15, 16, 17]). In this paper, we define one failure mode for each of the five macro cognitive functions defined in [18, 19]: (1) detecting and noticing, (2) understanding and sense-making, (3) decision-making, (4) action, and (5) teamwork. In Whaley, for each macro cognitive function, the team identified proximate causes for why the cognitive function may fail, cognitive mechanisms underlying the failures, and factors that influence the cognitive mechanisms and may lead to human performance errors[18, 19].

Another option is to use first-principles reliability techniques (e.g. by doing a "human-machine" FMEA with consideration of macro cognitive functions and machine functions).

4.3. Bayesian Network (BN) causal models

In the proposed framework, BN causal models are used to capture the detailed causal pathways and interdependencies among PIFs. In this framework, BNs are used because of their ability to model cause-and-effect relationships and the use of probabilistic inference. Furthermore, BNs provide the ability to reason about any variable in the model, enabling quantitative insights relevant to improving performance.

This approach will allow all relevant PIFs (both observable and unobservable) to be used in model development. In addition, it enables the explicit inclusion of data collection elements within the model structure, and these data collection elements could enable using data from different versions of SACADA.

These causal relationships would be initially developed from cognitive science and systems science, and could eventually be developed or validated by using SACADA data. The BN model will be developed by creating an explicit map between PIFs (first element described above) and the human-machine team failure modes (second element described above). This is accomplished by using the causal mapping approach developed and illustrated in [20]. The starting point for development of this causal mapping has been developed as part of NRCs work on defining a cognitive basis for HRA [18, 19] and in the IDAC model [21]. The size of the resulting BN models could be reduced (e.g., to facilitate quantification) by using node reduction algorithms [20]. A second approach would be to follow the approach of [22] and use factor analysis clustering, or structure learning algorithms techniques directly on the SACADA data.

4.4. Data and Bayesian parameter updating

The quantitative parameters of the BN models would be populated using multiple sources of data. Prior information on the relationship between the PIFs and the human-machine failure modes could be defined by using an existing HRA method (e.g., as illustrated in [23] using SPAR-H [24]) or other HRA databases (e.g., [3, 4, 25, 26, 27]). Prior information on the PIF prior probabilities could be assembled from a variety of published sources in cognitive science and HRA, or via formal expert elicitation with HRA experts such as in [27]. The SACADA data would be used as information to Bayesian update multiple parameters within the model, using a Bayesian updating approach [28].

There are several reasons to include this aspect in the framework. First, it allows the causal model to be populated with

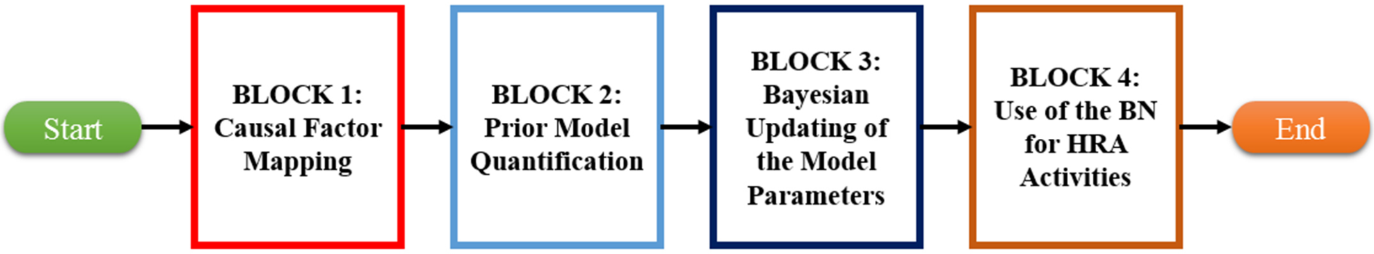


Figure 2: Block diagram depicting the algorithm that executes the steps outlined in the original framework.

appropriate information from multiple sources, including data. It also enables the model to include variables and information which are unlikely to be represented in the data (e.g., control room environment). This also provides a way to address the reality that some factors will be over- (or under-) weighted in the training contexts (e.g., high task complexity).

4.5. Human-machine task sequences

The final aspect as well as the next step of the framework involves modeling the sequential aspects of human-machine activities associated with the response to a malfunction. This addresses the need to treat an HFE as the outcome of a process involving several sequential activities or tasks involving different macro cognitive functions. This notion that human failure involves a series of activities and that there is dependency between HFEs has been acknowledged in even the earliest HRA methods [29]. It is considered in depth in simulation-based methods which explicitly model sequences of activities or subtasks involved in PRA event (e.g., [8, 9, 30, 31, 32]). This sequential and semi-temporal aspect of performance needs to be modeled explicitly in order to reflect the fact that that failure is a process, not a single event.

SACADA provides the first opportunity to use data inform this process. This opportunity arises as a result of several coupled aspects of the data. The first aspect is the detailed consideration of TOEs at the level of macro cognitive tasks (rather than at the higher event level used in many HRA methods). The second is the alignment of these TOEs with responses to specific malfunctions (more akin to the event level in HRA), and the third is that the PIFs are collected at the TOE level. This provides new potential to transform the treatment of sequential and temporal aspects of the PIFs and the activities that comprise an HRA event. A mechanism for modeling this dependency would be through the use of Dynamic Bayesian Networks, which were introduced in [33, 14, 34], and further expanded within the HUNTER framework [30, 31, 32].

5. Overview of Algorithm

A simplified outline of the algorithm is presented in Figure 2. At a high-level, the algorithm involves four main activities: causal factor mapping (BN structure development), prior model quantification (parameterization of the BN), Bayesian updating of the model parameters, and use of the BN for HRA activities.

The algorithm includes the following steps:

- Block 1: Causal Factor Mapping (BN structure development)
  - Create a causal map of the relationship between the PIFs, failure mechanisms, proximate causes of failure, and MCFs.
  - Simplify BN structure using node reduction
- Block 2: Prior model quantification (BN parameterization)
  - Map existing HRA method PIFs to PIF taxonomy (see Table 1)
  - Use existing HRA method to get priors of the probability of MCF error $Pr(MCFerror|PIFs)$
  - Expert elicit priors for PIFs, $Pr(PIFs)$
- Block 3: Bayesian update model parameters
  - Map simulation data source variables to PIFs
  - Use simulation data source to update probability of MCF error $Pr(MCFerror|PIFs)$
  - (Optional) Perform Bayesian update of the PIF probabilities $Pr(PIFs)$
- Block 4: Use the BN for HRA activities.

This algorithm uses an existing HRA method and expert elicitation as the basis for determining the priors parameters for the quantitative model. Similarly, the algorithm then uses the simulator data source as the data used to update the model.

5.1. Block 1: Causal Factor Mapping

Block 1 entails developing a full mapping of the causal factors which are relevant to modeling human failure events.

This step of the algorithm draws a graphical model between several specific elements: the failure of an MCF (which is similar to the concept of a failure mode in PRA), the proximate causes of failure (PC), the failure mechanisms (FM), and the

Table 1: Proposed PIF taxonomy for use in HRA data collection and causal models [7].

Organization-based	Team-based	Person-based	Situation/ stressor-based	Machine-based
Training program	Communication	Attention	External environment	HSI
Availability	Availability	To task	Conditioning events	Input
Quality	Quality	To surroundings	Task load	Output
Corrective action program	Direct supervision	Physical & psychological abilities	Time load	System response
Availability	Leadership	Alertness	Other loads	
Quality	Team coordination	Fatigue	Non-task	
Other programs	Team cohesion	Impairment	Passive information	
Availability	Role awareness	Sensory limits	Task complexity	
Quality		Physical attributes	Cognitive	
Safety culture		Other	Execution	
Management activities		Knowledge/ experience	Stress	
Staffing		Skills	Perceived situation	
Scheduling		Bias	Severity	
Workplace adequacy		Familiarity with situation	Urgency	
Resources		Morale/ motivation/ attitude	Perceived decision	
Procedures			Responsibility	
Availability			Impact	
Quality			Personal	
Tools			Plant	
Availability			Society	
Quality				
Necessary information				
Availability				
Quality				

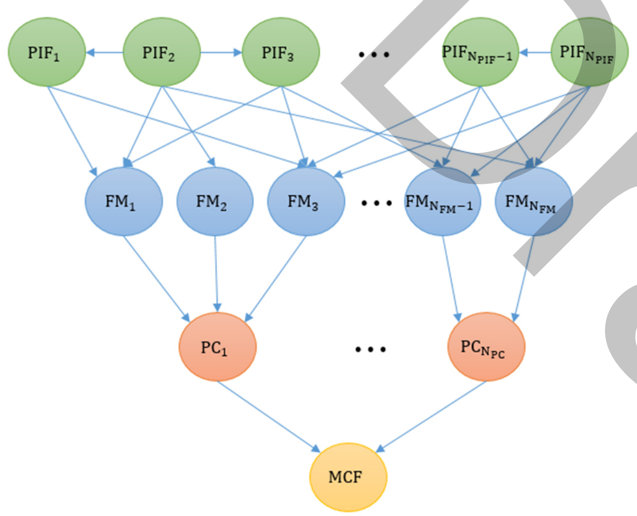


Figure 3: General structure of a Bayesian Network consisting of a causal structure linking PIFs (green) directly to PIFs and to failure mechanisms (FM) (blue), which are directly linked to proximate causes of failure (PC) (orange), which directly link to a macro cognitive function (MCF) (yellow). The MCFs could furthermore be directly linked to a node, defining the probability of an HFE.

associated PIFs [5]. Figure 3 illustrates the form of these relationships as the directed acyclic graph portion of a Bayesian Network (BN; also called Bayesian Belief Network, BBN).

In general, the PIF nodes will be discretized into a small number states of the PIF (e.g., high/medium/low, adequate/inadequate). The FMs and PCs will be binary nodes with states (true/false) The MCF node could be a binary (true/false) node, or could be a value node representing the probability of a failure of the MCF.

Whaley et. al. conducted an extensive literature review which defined five MCFs: (1) detecting and noticing, (2) understanding and sense making, (3) decision making, (4) action, and (5) team coordination [18, 19]. The Whaley work also developed associated proximate causes of failure (PC), failure mechanisms (FM), for each MCF based on published cognitive literature. The relationships between MCF, PCs, and FMs should be directly captured in the causal model. In addition, the relationship between FMs and PIFs, and between PIFs, must also be captured explicitly within the model.[6]

The target node of the BN is the failure of an MCF; this is characterized as the probability of an MCF failure conditional on the state of its parent notes $Pr(MCFerror|parents)$. The next layer of the model, the parents of MCFerror, indicates that failure of an MCF can be caused by one or several different proximate causes of failure; this causal relationship is directly mapped from [18, 19]. Similarly, each proximate cause of failure is a result of a set of failure mechanisms, as indicated in [18, 19].

The final layer of the model is the set of PIFs which causally influence the occurrence of the failure mechanisms. This layer of the model includes causal arcs from PIFs to FMs, and causal arcs which capture interdependence between PIFs. These relationships are partially discussed in the Whaley work, but are incomplete since the focus of the Whaley et al. work was to identify PIFs relevant to FM rather than to create a model of the PIFs and FMs. Identification of the interdependencies among the PIFs and the relationship between FMs and sets of PIFs requires going to the original cognitive literature cited in Whaley.

[5] Note that the terminology of PCs and FMs used in this paper is chosen to be consistent with [19]. If PCs are not used, one could directly link FMs to the failure of MCFs

[6] These relationships are partially discussed in [19], but are incomplete in [19] since the focus of the work in [19] was to identify PIFs relevant to each FM rather than to create a model of the PIFs and FMs. Identification of the interdependencies among the PIFs and the relationship between FMs and sets of PIFs requires going to the original cognitive literature cited in [19].

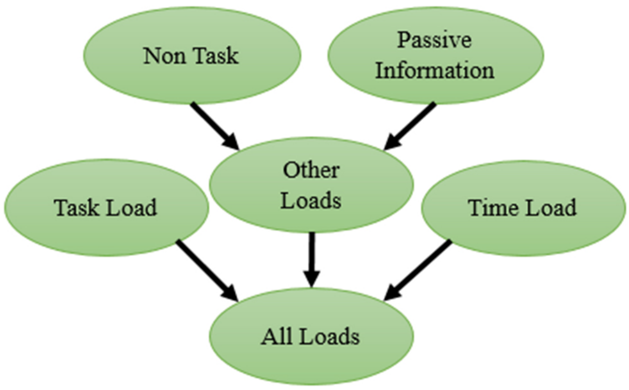


Figure 4: The Bayesian network characterizing the relationship between three different types of loads defined by Groth and Mosleh [7].

Thus, there are several aspects involved in developing this final layer.

The first step is to conduct direct causal mapping of PIF and FM relationships, which can be extracted from cognitive literature. The work by [20] illustrates the approach for directly mapping the causal relationships and causal pathways which are present in cognitive literature, conducted by reviewing the PIFs identified by Whaley for each FM, and furthermore reviewing each literature source cited in [18, 19].

The second step involves capturing causal relationships between PIFs at different levels of detail or depth within the hierarchy. Several PIF nodes as defined by Whaley [18, 19] require some expansion to capture specific aspects of the PIF. For example, Whaley [18, 19] uses the PIF "Load" which could refer to several different types of loads. Following the Groth and Mosleh PIF taxonomy [7], the concept of "Load" would be represented as the BN outlined in Figure 4. This two step process can be followed to fully expand the outer layers of the BN for each MCF.

Once the full causal structure is developed for an MCF, the next step is to conduct node reduction to simplify the BN structure for easier quantification. To do this, we apply a node reduction an algorithm designed by Shachter [35] as described by [20]. The algorithm allows for simplification of the BN structure in such a way that the dependence and independence assumptions are not changed [36]. The three rules of the node reduction algorithm are the following:

- Two nodes Z_1 and Z_2 may have their link (or arc) directions reversed if node Z_1 has the same parents as Z_2 and if the reversal doesn't cause the BN to become cyclical.

- Nodes without child nodes or evidence are referred to as *barren nodes* and can therefore be removed from the BN without changing the overall BN logic.

- A chance node may be removed if it precedes a value node of interest and *only* a value node of interest; the value node inherits the conditional predecessors of the removed node.

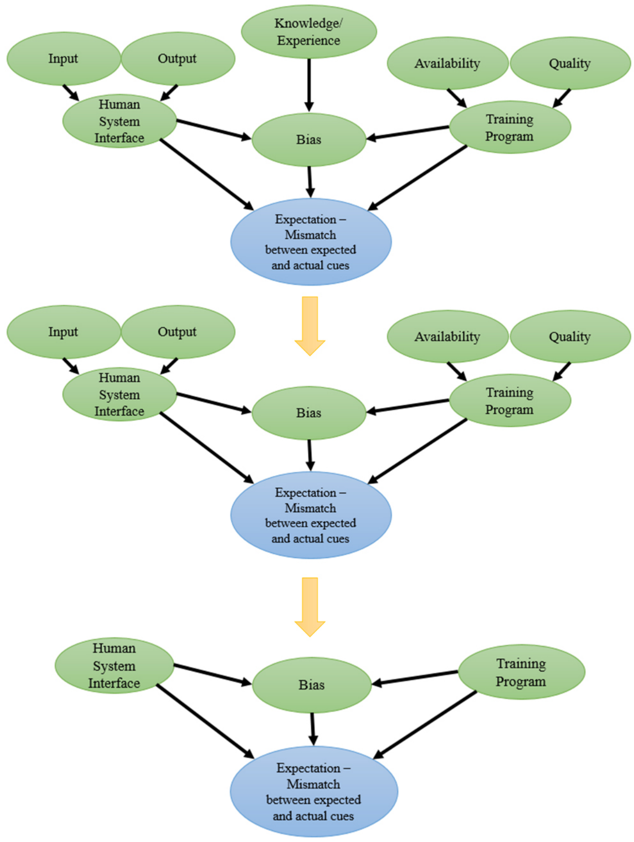


Figure 5: An example of node reduction on the BN for "Expectation Mismatch between expected and actual cues." [18, 19] By node reduction of the top BN, the bottom BN may be attained.

Figure 5 presents a very simple example of causal mapping and subsequent node reduction using one of the failure mechanisms: "Expectation Mismatch between expected and actual cues." This failure mechanism describes a scenario where a plant operator does not detect a particular cue (i.e., an indicator state) because that the operator is expecting to see a different cue.

The PIFs identified for this FM in [18, 19], are *HSI*, *training program*, and *knowledge and experience*. Each of these directly or indirectly influences the FM. Further review of the literature cited in [18, 19] illustrates an important additional PIF, *bias*, and a causal relationship between this *bias* and the other PIFs identified by Whaley. This causal relationship among PIFs is identified by [37] who noted that an operator is able to visually scan a panel more quickly if they have seen it before (e.g. *training program* and *knowledge and experience*). This indicates that both the operator *knowledge and experience* and the *training* supports the development mental models (biases) of indicator configuration. A further causal relationship is found in the work of Nikolic et. al. [38], who noted that the expectation (a type of *bias*) of a particular indicator signal will affect the likelihood that the signal will get the operators attention. This expectation is due to the operators *knowledge or experience* of panel signals and due to the configuration of the display (the *HSI*). This source is used to create the direct link from *bias* to the FM, and further supports the links from *knowledge and experience* and *HSI* to *bias*.

The relationship between *HSI Input*, *HSI Output*, and the *HSI node* and similar relationships with parents of *Training* are directed based on the the decomposition of these PIFs into additional levels of detail as illustrated in the PIF hierarchy (Table 1).

A next step in the process would be to perform a similar cognitive literature review and mapping to establish the outer layers of the BN for the other FMs in the model; for brevity this is not illustrated in this paper.

To apply node reduction to this structure, one can first reverse the direction of several of the node links one by one. Reversing the link from *bias* to *knowledge and experience* leaves the *knowledge and experience* node barren, and thus it can be removed, as illustrated in Figure 5. We could apply further node reduction to remove the *HSI Input*, *HSI Output*, *Availability* and *Quality*, fully absorbing these nodes into the "typical" PIFs of *HSI* and *training*. However, we could also choose to stop at this level because these relationships are straightforward to quantify (see Block 2).

The resulting node-reduced BN would then be used as a simplified structure to be quantified in Block 2.

5.2. Block 2: Prior model quantification (BN parameterization)

Block 2 of the algorithm involves assignment of the prior values for the conditional and marginal probabilities of all the nodes in the BN structured we defined in Block 1. These prior probabilities represent the probabilities prior to being Bayesian updated using HRA data.

In a BN, all nodes, Z_i, are specified by a conditional probability distribution given their parents, $Pr(Z_i|pa(Z_i))$. For nodes without parents, this reduces to the marginal distribution $Pr(Z_i)$. We restrict ourselves to BNs with discrete random variables, which are described by their conditional probability mass function (PMF), which are summarized in conditional probability tables (CPTs).

The BN factorizes the joint probability of all of the nodes into a product of conditional probabilities.

$$Pr(Z_1, Z_2, Z_3, ..) = \prod_i Pr(Z_i|pa(Z_i))$$

In the BN structure defined in Block 1, we have six different types of probabilistic relationships to quantify:

$$Pr(PIF_x) \tag{1}$$

$$Pr(PIF_x|pa(PIF_x)) \tag{2}$$

$$Pr(FM_k|pa(FM_k)) \tag{3}$$

$$Pr(PC_j|pa(PC_j)) \tag{4}$$

$$Pr(MCF_i|pa(MCF_i)) \tag{5}$$

$$Pr(HFE|MCFs) \tag{6}$$

where $pa(PIF)$ and $pa(FM)$ are a subset of the PIFs, $pa(PC)$ are a subset of the FMs, $pa(MCF)$ are all PCs defined for that MCF, and $pa(HFE)$ are all of the MCFs.

Equation 1 requires information about the (marginal) probabilities of the root node PIFs. The prior probabilities for these nodes can be directly obtained by expert elicitation, e.g., as in [39].

Equation 2 denotes the probabilities of PIFs which have parent nodes; in this case the parents are other PIFs. These relationships can also be directly obtained by expert elicitation. In addition, these relationships could be quantified simply using an OR or a Noisy-OR relationship.

Equations 3, 4, 5, 6 may be quantified in one of several ways depending on the availability of prior information. Option 1 assumes that there is relevant existing HRA information that can be used to quantify the relationships between PCs, FMs, and MCFs. Options 2, 3, and 4 assume that this information does not exist.

Option 1. Expert elicitation or cognitive-based HRA methods. It is possible to directly expert elicit the priors that define the arcs from PIFs to FMs, from FMs to PCs, and from PCs to MCFs, and from MCFs to HFEs. A better option is to leverage the probabilities from the two existing HRA methods which use the concepts of FMs, PCs, and MCFs: PHOENIX [12] and IDHEAS [10], or via the cognitively focused HRA method ATHEANA [40]. These methods would need to be augmented with additional expert elicitation to account for multiple PIFs which were not included in either PHOENIX or IDHEAs, and to account for a small number of FMs and PCs which were omitted from one or both methods.

Option 2. OR Logic, or a deterministic probability assignment, can be used to efficiently and quickly define the relationships between a set of parent nodes $Z_{i=1}, \cdots, Z_{i=n}$ and a child node X via a truth table composed of 0s and 1s. Essentially, if any parent is true, the probability of the target state of the child node is 1. For example, if a FM_1 occurs, its child PC_1 would also occur. This logic provides a logarithmic reduction in the number of parameters which must be specified. This deterministic relationship is readily justifiable for the relationship between FMs and PCs, between PCs and MCFs, and between MCFs and HFE because of the way these concepts were defined in [19]. Furthermore, this approach is consistent with the way that these relationships are quantified in both PHOENIX and IDHEAS.

Option 3. NOISY-OR logic. offers a similar conditional gate called a noisy-OR gate [41, 42], which has also been extended to multi-state nodes. A noisy-OR gate is an extension of an OR-gate that relates a set of parent nodes $Z'_{i=1}, \cdots, Z'_{i=n}$ to a child node X. The noisy-OR gate is further defined by an inhibitor (or node suppression) probability q_i for each node Z_i. Quantification of a noisy-OR gate is governed by a strict true/false scenario for the output node where the probability of the child node being true ($X = 1$) given that a single parent is true is $1 - q_z$; for multi-parent nodes this becomes $1 - q_i^t$ where t denotes the number of parents in a true state.

For example, With respect to the conditional PMFs for PC_j, the quantification is as follows:

$$Pr(PC_j = 0|FM_1, FM_2, \ldots, FM_{n_{FM}}) = \prod_{k=1}^{n_{FM}} q_k^{fm_k}$$
$$Pr(PC_j = 1|FM_1, FM_2, \ldots, FM_{n_{FM}}) = 1 - \prod_{k=1}^{n_{FM}} q_k^{fm_k} \quad (7)$$

As with option 2, this approach can be readily justified for the FM to PC, PC to MCF, and MCF to HFE relationships. This approach also provides additional potential for quantifying PIF to FM relationships.

Option 4. Additional node reduction can be performed to further simplify the BN structure for quantification beyond what was demonstrated in 5. In Figure 3 we applied reduction to only the PIF nodes. However, we can further reduce the full conceptual BN model (Figure 3) by applying the third rule of node reduction stating that *a chance node may be removed if it precedes a value node of interest and only a value node of interest*. In this case, we would implement the MCF node as a value node, and as such all of the PC nodes may be removed based on this rule. The FM nodes may also be removed by application of this same node reduction rule. This reduces the general BN structure to a much simplified structure outlined in Figure 6 that only consists of PIF nodes and the MCF node of interest. The relationship between MCF and HFE would be additive.

Clearly, this node-reduced structure directly resembles the vast majority of HRA methods in that it directly relates the PIFs to specific types of human errors.

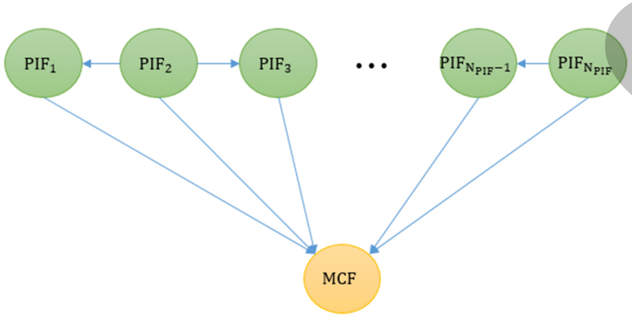


Figure 6: Simplified structure of a Bayesian Network based on a complete node reduction the full BN (Figure 3) that removes all FM and PC nodes. The simplified BN consists only of the series of N_{PIF} PIFs and a macro cognitive function error (MCF).

With this node quantification, the remaining relationships (i.e., the PIF to MCF relationships) can be quantified in a straightforward manner by using an existing HRA method. Several options for an existing HRA exist including CREAM [43], CBDT [44], THERP [29], HEART [45], and SPAR-H [24]. A second option is to define priors directly from existing HRA databases (e.g. [3, 4, 24, 25, 26, 27]).

To do this, the PIFs of an existing HRA method or HRA data source would be mapped onto the same PIF taxonomy used in the previous step. Table 2 presents an example of this mapping where the PIFs of the SPAR-H HRA method ((1) available time, (2) stress/stressors, (3) complexity, (4) experience and training, (5) procedures, (6) ergonomics/HMI, (7) fitness for duty, and (8) work processes) are matched with the PIFs from the proposed PIF taxonomy (Table 1). Since many existing HRAs are made up of a limited number of PIFs, this mapping procedure may need to be augmented with expert information. For example, *bias* is a PIF whose description does not readily match any of the PIFs listed under SPAR-H [7]. Thorough revision of supplementary HRAs is necessary to find matches for all PIFs of a full taxonomy.

Table 2: The eight SPAR-H PIFs [24] as they match with the full PIF taxonomy as defined by Groth and Mosleh [7].

SPAR-H PIF	Full PIF Taxonomy Equivalent(s) [7]
Available Time	Task load Time load
Stress/Stressors	Stress Perceived situation Severity Urgency Perceived decision Responsibility Impact Personal Plant Society
Complexity	Conditioning events Task complexity Cognitive Execution System response
Experience/Training	Training program Availability Quality Knowledge/experience Familiarity with situation
Procedures	Resources Procedures Availability Quality
Ergonomics/HMI	HSI Input Output
Fitness for Duty	Physical & psychological abilities Alertness Fatigue Impairment
Work Processes	Safety culture Direct supervision Leadership Morale/motivation/attitude

Upon completion of the PIF mapping step, the MCF error conditional on the PIFs $Pr(MCF|PIF_1, PIF_2, \ldots, PIF_{n_{PIF}})$ (where $MCF = 1$ if there is error and $MCF = 0$ if there is no error), is quantified directly by application of the HRA method, as illustrated in [23].

5.3. Block 3: Bayesian update model parameters

The third block of the algorithm involves performing Bayesian updating on all of the probability nodes in the BN. Using a Bayesian updating process allows us to incorporate multiple sources of data in the assignment of probabilities. The Bayesian updating process for HRA methods has been described by [28].

As stated in Section 4.4, there are several possible data sources which could be used in this algorithm. The first task in implementing the data is to first map the data variables and states (i.e., the PIFs, SFs, and causal factors) to the PIFs that are defined in the selected taxonomy; this step is similar to the mapping performed for prior data sources in Block 2. In SACADA, the SFs have multiple states which together characterize the scenario context. Appendix A provides a complete example of mapping the SACADA SF states to the PIF taxonomy in Table 1 based on the detailed variable definitions provided by both original sources.

As with all mapping procedures in this algorithm, the data source-to-PIF mapping requires a careful consideration of which PIFs are captured by the variables used in the data source. For example, Appendix A Table A.3 cites three SACADA states under the *workload* significant factor (normal, concurrent demands, and multiple concurrent demands). All states fall under the *task load* PIF by definition, however the characteristics would be of *lower* to *higher* load in the given state order. Additionally, the *multiple concurrent demands* state would be given an additional PIF stress due to the increased demands on the operator. Therefore it is important to clarify any and all ambiguity in the definitions of the data states before mapping begins. Several revisions and reviews may be necessary to build a map that accurately accounts for the range of PIFs that may be associated with the data.

After mapping is concluded, the next step in this block of the algorithm is to use the simulation source data D to update prior values placed on the probabilistic relationships defined in Equations (1-6) in Block 2. For example, SACADA can be used to update probability of MCF error $Pr(MCF|D, PIF_1, PIF_2, \ldots, PIF_{n_{PIF}})$, and in some cases may be used to update the interdependencies between specific PIFs. Note that SACADA provides joint probabilities of multiple PIFs and outcomes; translating this into CPTs can be obtained by applying the definition of conditional probability.

The process in Block 3 would be repeated for each data source. If the BN is implemented in software such as GeNiE [46] or the Bayes Net Toolbox [47, 48], then this step can be facilitated substantially.

5.4. Block 4: Use the BN for HRA activities

The fourth and final block indicates that the use of the BN is a separate concept from the development of the BN therein. Using the BN for HRA activities would proceed similarly to existing uses of BNS, such as those described in the review by Mkrtchyan et al. [49].

6. Conclusion

The algorithm presented in this paper provides the first comprehensive methodology for fusing cognitive literature, existing HRA models, and HRA data from multiple sources. The algorithm enhances both the qualitative and quantitative basis of the field of HRA. We demonstrated how to combine existing HRA models with a causal understanding of failure (that is, cognitive information akin to "physics of failure" information regarding human/machine teams) in order to deal with inherent data limitations. The combination requires a complex data fusion procedure and the design of a BN structure which defines the relationship between PIFs, FMs, PCs, and MCFs; which have both been defined in this paper by way of a complex BN node quantification method. This adds credibility and traceability to the HRA models and thus makes them more useful in a NPP situation. The second finding is the newfound potential that the SACADA database has as a useful source of HRA data and as a means of providing a temporal evolution of human error and PIFs. States defined by the SACADA database can be readily mapped to HRA PIFs and PIF characteristic based on a complete taxonomy of PIFs. This in addition to the quality of the SACADA database makes probabilities such as $Pr(FM|PIF_1, PIF_2, \cdots, PIF_{n_{PIF}})$ readily quantifiable which aids in the implementation of the algorithm described in this paper. The framework and HRA modeling algorithm proposed in this work will enable a path toward an HRA vision that is both model-based and data-informed, enhancing the technical basis of HRA and enabling streamlined use of HRA databases as the volume and variety of HRA data increases.

Acknowledgements

The authors of this article would like to thank the Y. James Chang of the United States Nuclear Regulatory Commission (NRC) for support throughout this research.

References

[1] K. M. Groth, A framework for using SACADA to enhance the qualitative and quantitative basis of HRA, in: Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 14), 2018.

[2] Y. H. J. Chang, D. Bley, L. Criscione, B. Kirwan, A. Mosleh, T. Madary, R. Nowell, R. Richards, E. M. Roth, S. Sieben, A. Zoulis, The {SACADA} database for human reliability and human performance, Reliability Engineering & System Safety 125 (0) (2014) 117–133, special issue of selected articles from {ESREL} 2012. doi:http://dx.doi.org/10.1016/j.ress.2013.07.014.

[3] J. Park, W. Jung, OPERA a human performance database under simulated emergencies of nuclear power plants, Reliability Engineering and System Safety 92 (4) (2007) 503–519. doi:10.1016/j.ress.2006.01.007.

[4] S. Y. Choi, W. Jung, Qualitative human event analysis with simulator data by using HuRAM+ and HERA, in: Proceedings of the European Society for Reliability Annual Meeting (ESREL 2013), 2013.

[5] Y. H. J. Chang, C. Franklin, SACADA data for HEP estimates, in: Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 14), 2018.

[6] A. Mosleh, Y. H. J. Chang, Model-based human reliability analysis: prospects and requirements, Reliability Engineering and System Safety 83 (2) (2004) 241–253. doi:doi:10.1016/j.ress.2003.09.014.

[7] K. M. Groth, A. Mosleh, A data-informed PIF hierarchy for model-based human reliability analysis, Reliability Engineering and System Safety 108 (2012) 154–174. doi:10.1016/j.ress.2012.08.006.

[8] S. H. Shen, C. Smidts, A. Mosleh, IDA: a cognitive model of nuclear power plant operator response during abnormal conditions, Ph.D. thesis, University of Maryland at College Park (1994).

[9] Y. H. J. Chang, A. Mosleh, Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents: Part 1 - overview of the IDAC model, Reliability Engineering and System Safety 92 (8) (2007) 997–1013.

[10] J. Xing, G. Parry, M. Presley, J. Forester, S. Hendrickson, V. Dang, An integrated human event analysis system (IDHEAS) for nuclear power plant internal events at-power application, Report, NUREG-2199, Vol 1 (2017).

[11] N. J. Ekanem, A. Mosleh, S.-H. Shen, PHOENIX a model-based human reliability analysis methodology: Qualitative analysis procedure, Reliability Engineering and System Safety 145 (2016) 301–315.

[12] N. J. Ekanem, A. Mosleh, Phoenix a model-based human reliability analysis methodology: Quantitative analysis procedure and data base, in: Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 12), 2014.

[13] J. Oxstrand, D. L. Kelly, S.-H. Shen, A. Mosleh, K. M. Groth, A model-based approach to hra: Qualitative analysis methodology, in: Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 11), 2012.

[14] A. Mosleh, S.-H. Shen, D. L. Kelly, J. H. Oxstrand, K. M. Groth, A model-based human reliability analysis methodology, in: Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 11), 2012.

[15] K. M. Groth, S.-H. Shen, J. Oxstrand, A. Mosleh, D. Kelly, A model-based approach to HRA: Example application and quantitative analysis, in: Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 11), 2012.

[16] J. Park, Y. Kim, W. Jung, A framework to estimate HEPs from the full-scope simulations of NPPs: unsafe act definition, identification, and quantification, Tech. Rep. KAERI/TR-6401/2016, Korea Atomic Energy Research Institute (KAERI) (6 2016).

[17] Y. Kim, J. Park, W. Jung, A classification scheme of erroneous behaviors for human error probability estimations based on simulator data, Reliability Engineering and System Safety 163 (2017) 1–13.

[18] A. M. Whaley, J. Xing, R. L. Boring, S. M. Hendrickson, J. C. Joe, K. L. Le Blanc, E. Lois, Draft report building a psychological foundation for human reliability analysis, Draft version of NUREG-2114, US Nuclear Regulatory Commission, Washington DC (8 2012).

[19] A. M. Whaley, J. Xing, R. L. Boring, S. M. L. . Hendrickson, J. C. Joe, K. L. Le Blanc, S. L. Morrow, Cognitive basis for human reliability analysis, NUREG-2114, US Nuclear Regulatory Commission (1 2016).

[20] K. Zwirglmaier, D. Straub, K. M. Groth, Capturing cognitive causal paths in human reliability analysis with Bayesian network models, Reliability Engineering & System Safety 158 (2017) 117–129. doi:10.1016/j.ress.2016.10.010.

[21] Y. H. J. Chang, A. Mosleh, Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents. part 4: IDAC causal model of operator problem-solving response, Reliability Engineering and System Safety 92 (8) (2007) 1061–1075. doi:10.1016/j.ress.2006.05.011.

[22] K. M. Groth, A. Mosleh, Deriving causal Bayesian networks from human reliability analysis data: A methodology and example model, Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 226 (4) (2012) 361–379. doi:10.1177/1748006X11428107.

[23] K. M. Groth, L. P. Swiler, Bridging the gap between HRA research and HRA practice: A Bayesian network version of SPAR-H, Reliability Engineering and System Safety 115 (2013) 33–42. doi:10.1016/j.ress.2013.02.015.

[24] D. Gertman, H. Blackman, J. Marble, J. Byers, C. Smith, The SPAR-H Human Reliability Analysis method, NUREG/CR-6883, US Nuclear Regulatory Commission (2005).

[25] J. Park, W. Jung, A database for human performance under simulated emergencies of nuclear power plants, Nuclear Engineering and Technology 37 (5) (2005) 491–502.

[26] Simulator-based Human Factors Studies Across 25 Years: The History of the Halden Man-Machine Laboratory, Springer, London, 2010.

[27] B. Hallbert, A. Whaley, R. Boring, P. McCabe, Y. H. J. Chang, Human event repository and analysis (HERA): The HERA coding manual and quality assurance, NUREG/CR-6903, Vol. 2, US Nuclear Regulatory Commission (2007).

[28] K. M. Groth, C. L. Smith, L. P. Swiler, A Bayesian method for using simulator data to enhance human error probabilities assigned by existing HRA methods, Reliability Engineering & System Safety 128 (2014) 32–40. doi:doi:10.1016/j.ress.2014.03.010.

[29] A. D. Swain, H. E. Guttmann, Handbook of Human Reliability Analysis with emphasis on nuclear power plant applications, NUREG/CR-1278, US Nuclear Regulatory Commission, Washington DC (1983).

[30] R. Boring, D. Mandelli, M. Rasmussen, S. Herberger, T. Ulrich, K. M. Groth, C. L. Smith, Human unimodel for nuclear technology to enhance reliability (HUNTER): A framework for computation-based human reliability analysis, in: Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 13), 2016.

[31] R. Boring, D. Mandelli, J. Joe, C. Smith, K. Groth, A research roadmap for computation-based human reliability analysis, INL/EXT-15-36051, Idaho National Laboratory, Idaho Falls, ID (7 2015).

[32] R. Boring, D. Mandelli, M. Rasmussen, S. Herberger, T. Ulrich, K. M. Groth, C. Smith, Integration of human reliability analysis models into the simulation-based framework for the risk-informed safety margin characterization toolkit, INL/EXT-16-39015, Idaho National Laboratory, Idaho Falls, ID (6 2016).

[33] K. M. Groth, A data-informed model of performance shaping factors for use in human reliability analysis, Ph.D. thesis, University of Maryland, College Park, MD (2009).
URL `http://drum.lib.umd.edu//handle/1903/9975`

[34] N. J. Ekanem, A. Mosleh, Human failure event dependency modeling and quantification: A Bayesian network approach, in: Proceedings of the European Society for Reliability Annual Meeting (ESREL 2013), Amsterdam, 2013.

[35] R. D. Shachter, Evaluating influence diagrams, Operations Research 34 (6) (1986) 871–882.

[36] D. Straub, A. Der Kiureghian, Bayesian network enhanced with structural reliability methods: Methodology, Journal of Engineering Mechanics 136 (10) (2010) 1258–1258.

[37] W. Van Zoest, A. Lleras, A. Kiingstone, J. T. Enns, In sight, out of mind: The role of eye movements in the rapid resumption of visual search, Perception & Psychophysics 69 (7) (2007) 1204–1217.

[38] M. I. Nikolic, J. M. Orr, N. B. Sarter, Why pilots miss the green box: How display context undermines attention capture, International Journal of Aviation Psychology 14 (1) (2004) 39–52.

[39] B. Hallbert, A. Kolaczkowski, The employment of empirical data and Bayesian methods in human reliability analysis: A feasibility study, NUREG/CR-6949, US Nuclear Regulatory Commission, Washington DC (2007).

[40] S. E. Cooper, A. M. Ramey-Smith, J. Wreathall, G. W. Parry, D. C. Bley, W. J. Luckas, J. H. Taylor, M. T. Barriere, A technique for human error analysis ATHEANA, NUREG/CR-6350, US Nuclear Regulatory Commission, Washington DC (1996). doi:10.2172/249298.

[41] S. Srinivas, A generalization of the noisy-or model, in: Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence, 1993.

[42] F. J. Díez, Parameter adjustment in bayes networks. the generalized noisy or-gate, in: Proceedings of the Ninth Conference Conference on Uncertainty in Artificial Intelligence, Washington, DC.

[43] E. Hollnagel, Cognitive Reliability and Error Analysis Method (CREAM), Elsevier, New York, 1998.

[44] G. W. Parry, A. J. Spurgin, P. Moieni, A. Beare, An approach to the analysis of operator actions in probabilistic risk assessment, Tech. rep., EPRI (1992).

[45] J. Williams, HEART: a proposed method for assessing and reducing human error, in: 9th Advances in reliability technology symposium, 1986.

[46] L. BayesFusion, GeNIe Modeler (2018).

[47] Mathworks, MATLAB 2018a (2018).

[48] K. Murphy, Bayes Net Toolbox (1997-2002).

[49] L. Mkrtchyan, L. Podofillini, V. N. Dang, Bayesian belief networks for human reliability analysis: A review of applications and gaps, Reliability Engineering & System Safety 139 (2015) 1–16.

Appendix A. SACADA-to-PIF Mapping

This appendix outlines the results of a preliminary mapping of the SACADA states to the PIFs states for use in the algorithm. The detailed taxonomy and definitions of the SACADA states is given in Chang et. al. [2] while the PIF taxonomy is given by Groth and Mosleh [7]. Each PIF is given a characteristic in terms of its relationship to one of the discrete states of the PIF (e.g., *good/nominal/poor/bad*, *high/medium/low*, or *not applicable*. In [2], the situational factors are associated with different macrocognitive functions, and a set of overarching factors which affect the performance of control room operators in all macrocognitive functions. The mapping for *Overarching* *factors* is presented in Table A.3. The mapping for *Monitoring and Detecting* is separated into Table A.4 for alarm detection and monitoring, and Table A.5 for indicator detection and monitoring. The mapping for *Understanding and Diagnosis* is presented in Table A.6. The tables for *Response Planning* and for *Manipulation* are presented in Table A.7 and Table A.8 respectively. The SFs for communication are presented in Table A.9. In some cases, SACADA SFs were irrelevant to determining PIFs; these are indicated by N/A.

Table A.3: Mapping of the SACADA SFs for "Overarching factors" onto the PIF taxonomy.

SF Subgroup	SF State	PIF	PIF State
Workload	Normal	Task Load	Low
	Concurrent Demands	Task Load	High
	Multiple Concurrent Demands	Task Load	High
		Stress	High
Time Criticality	Expansive Time Available	Time Load	Low
		Perceived Situation Urgency	Low
	Nominal Time Available	Task Load	Low
	Barely Adequate Time Available	Time Load	High
		Perceived Situation Urgency	High
Extent of Communications Required	Nominal Communication	Communication Availability	High
		Loads Non-task	High
	Extensive Onsite Communication	Communication Availability	High
		Loads Non-task	High
	Extensive Communication Within the Control Room	Communication Availability	High
		Loads Non-task	High
Other Demands/Factors	Non-Standard	Other Loads	High
	Noisy Background	External Environment	High
	Coordination	Loads Non-task	High
	Communicator Unavailable	Communication Availability	High
	Multiple Demands	Task Load	High
	Memory Demands	Task Load	High

Table A.4: Mapping of the SACADA SFs for "detecting an alarm" onto the the PIF taxonomy.

SF Subgroup	SF State	PIF	PIF State
Detection Mode	Self-Revealing	HSI Output	Good
	Procedure Directed Check	Procedures Availability	Good
		Task Complexity Cognitive	Good
	Procedure Directed Monitoring	Procedures Availability	Good
		Task Complexity Cognitive	Good
	Awareness/ Inspection	Task Complexity Cognitive	Medium
Status of Alarm Board	Dark	Loads Passive information	Good
	Busy	Loads Passive information	Normal
	Overloaded	Loads Passive information	Bad
		Task Complexity	Bad
Expectation of Alarm/Indication Change	Expected	System Response	Good
	Not Expected	System Response	Bad

Table A.5: Mapping of the SACADA SFs for "detecting or monitoring an indicator" onto the PIF taxonomy.

SF Subgroup	SF State	PIF	PIF State
Detection Mode	Procedure Directed Check	Procedures Availability	Good
		Task Complexity Cognitive	Good
	Knowledge-Driven Monitoring	Task Complexity Cognitive	Medium
	Procedure-Directed Monitoring	Procedures Availability	Good
		Task Complexity Cognitive	Good
	Awareness/Inspection	Task Complexity Cognitive	Medium
Individual Indicator	Slight Change	HSI Output	Poor/Bad
	Distinct Change	HSI Output	Good
Mimics/Display etc.	No Mimics	HSI Output	Poor
	Small Indications	HSI Output	Bad
	Similar Displays	HSI Output	Bad

Table A.6: Mapping of the SACADA SFs for "understanding or diagnosis" onto the PIF taxonomy.

SF Subgroup	SF State	PIF	PIF State
Diagnosis Basis	Procedure	Procedures Availability	Good/Bad
	Skill	Procedures Availability	Good/Bad
		Skills	Good/Bad
	Knowledge	Procedures Availability	Good/Bad
		Knowledge/Experience	Good/Bad
Familiarity	Standard	Familiarity with Situation	Good
	Novel	Familiarity with Situation	Poor
	Anomaly	Procedures Quality	Poor
		Familiarity with Situation	Bad
Outcome	Procedure-Based Activity	N/A	
	Skill-Based Behavior	N/A	
	Knowledge-Based Behavior	N/A	
Information Integration	Timing of Information	Necessary Information Availability	Poor
	Ambiguous Information	Necessary Information Quality	Poor
	Integration Required	Necessary Information Quality	Poor
		Task Complexity Cognitive	Poor
Diagnosis Information Specificity	Specific	Necessary Information Quality	Good
		Task Complexity Cognitive	Good
	Not Specific	Necessary Information Quality	Poor
Information Quality	Missing Information	Necessary Information Availability	Poor
	Misleading Information	Necessary Information Quality	Poor
	Conflicting Information	Necessary Information Quality	Poor

Table A.7: Mapping of the SACADA SFs for "response planning or decision making" onto the PIF taxonomy.

SF Subgroup	SF State	PIF	PIF State
Decision Basis	Procedure	Procedures	Good/Bad
	Skill	Skills	Good/Bad
	Knowledge	Knowledge/Experience	Good/Bad
Familiarity	Standard	Familiarity with Situation	Good
	Adaptation Required	Familiarity with Situation	Bad
	Anomaly	Procedures Quality	Poor
		Familiarity with Situation	Bad
Uncertainty	Clear	Procedures Quality	Good
		Necessary Information Quality	Good
	Uncertain	Procedures Quality	Poor
		Necessary Information Quality	Poor
	Competing Priorities	Attention to Task	Poor
		Attention to Surroundings	Poor
		Task Complexity	High
		Perceived Situation Urgency	High
	Conflicting Guidance	Procedures Quality	Poor
		Communication Quality	Poor
		Direct Supervision Leadership	Poor
Outcome	Procedure-Based Activity	N/A	
	Skill-Based Behavior	N/A	
	Knowledge-Based Behavior	N/A	

Table A.8: Mapping of the SACADA SFs for "manipulation" onto the PIF taxonomy defined.

SF Subgroup	SF State	PIF	PIF State
Type of Action	Simple and Distinct	Task Complexity Execution	Low
	Order	Task Complexity Execution	Medium
	Maintaining	Task Complexity Execution	Medium
Location	Main or Auxiliary Control Board	HSI Input	Good
	Back Control Panels	HSI Input	Bad
Guidance	Procedure	Task Complexity Execution	Low
	Skill of the Craft (Non-Faulted Hardware)	Task Complexity Execution	Low
	STAR (Faulted Hardware)	Task Complexity Execution	Medium
Recoverability	Immediately Recoverable	System Response; Perceived Decision Impact	Low
	Recoverable With Significant Efforts	System Response; Perceived Decision Impact	High
	Unrecoverable	Conditioning Event; Perceived Decision Impact	High
Additional Factors	Unintuitive Plant Response	System Response	Bad
	Unintuitive Controls	HSI Input	Bad
	Additional Mental Effort Required	Task Load; Task Complexity	Bad
	Inadequate Feedback	Necessary Information (Availability, Quality)	Bad
	Similar Controls	HSI Input	Bad

Table A.9: Mapping of the SACADA SFs for "communication" onto the PIF taxonomy.

SF Subgroup	SF State	PIF	PIF State
Communication Driver	Specifically Procedure Directed	Communication Availability	High
	Not Specifically Driven	Communication Availability	Low
Direction of Communication	From Booth	N/A	
	To Booth	N/A	
	Public Address Announcement	N/A	
	Other	N/A	

Appendix C: Tables and Figures

Appendix C1. Table of Elicit Priors for PIFs

The SyRRA group selected the following table based on a report by the NRC depicting a database of PIF probabilities under the SPAR-H methodology [5, 6].

	Inadequate	Barely Adequate	Nominal	Extra	Expansive
Available Time	1.00×10^{-6}	0.159	0.683	0.136	0.023
	Nominal	**High**	**Extreme**		
Stress/Stressors	0.841	0.136	0.023		
	Nominal	**Moderate**	**High**		
Complexity	0.5	0.341	0.159		
	Low	**Nominal**	**High**		
Experience and Training	0.333	0.333	0.333		
	Not Available	**Incomplete**	**Available but poor**	**Nominal**	
Procedures	0.05	0.2	0.3	0.45	
	Missing/misleading	**Poor**	**Nominal**	**Good**	
Ergonomics/HMI	0.023	0.136	0.683	0.159	
	Unfit	**Degraded Fitness**	**Nominal**		
Fitness for Duty	1.00×10^{-6}	0.159	0.841		
	Poor	**Nominal**	**Good**		
Work Processes	0.023	0.819	0.159		

Appendix C2. Figures Depicting Failure of Macro Cognitive Function Detection and Noticing

The following models are based on the failure of macro cognitive function "Detecting and Noticing" as defined by Whaley et. al [8]. It is important to note that these models neglect interdependency among PIFs and must be revised to capture this dependency before further use.

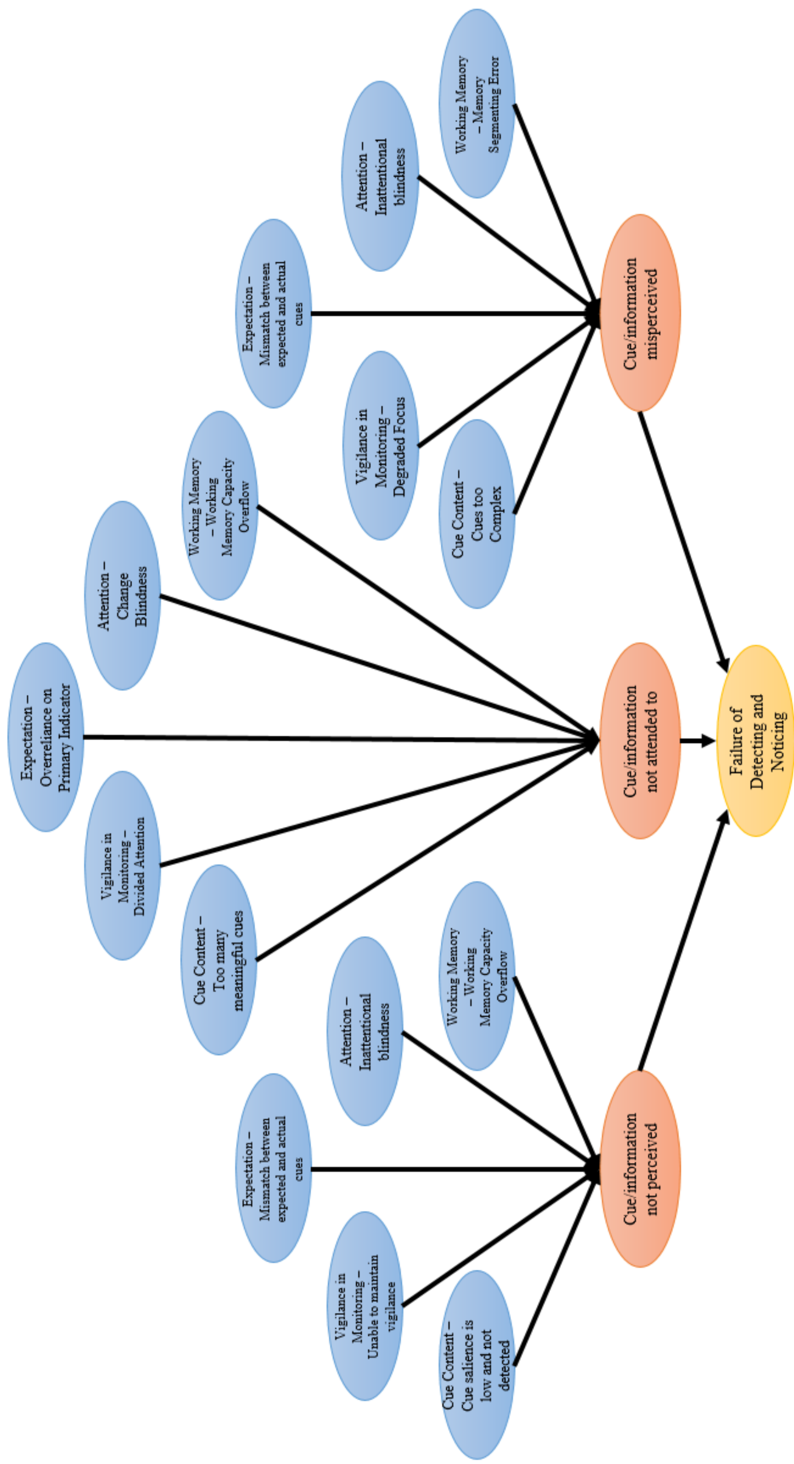


Figure 1: The BBN structure defining the relationship between the failure of Detecting and Noticing, its three proximate causes of failure, and their failure mechanisms

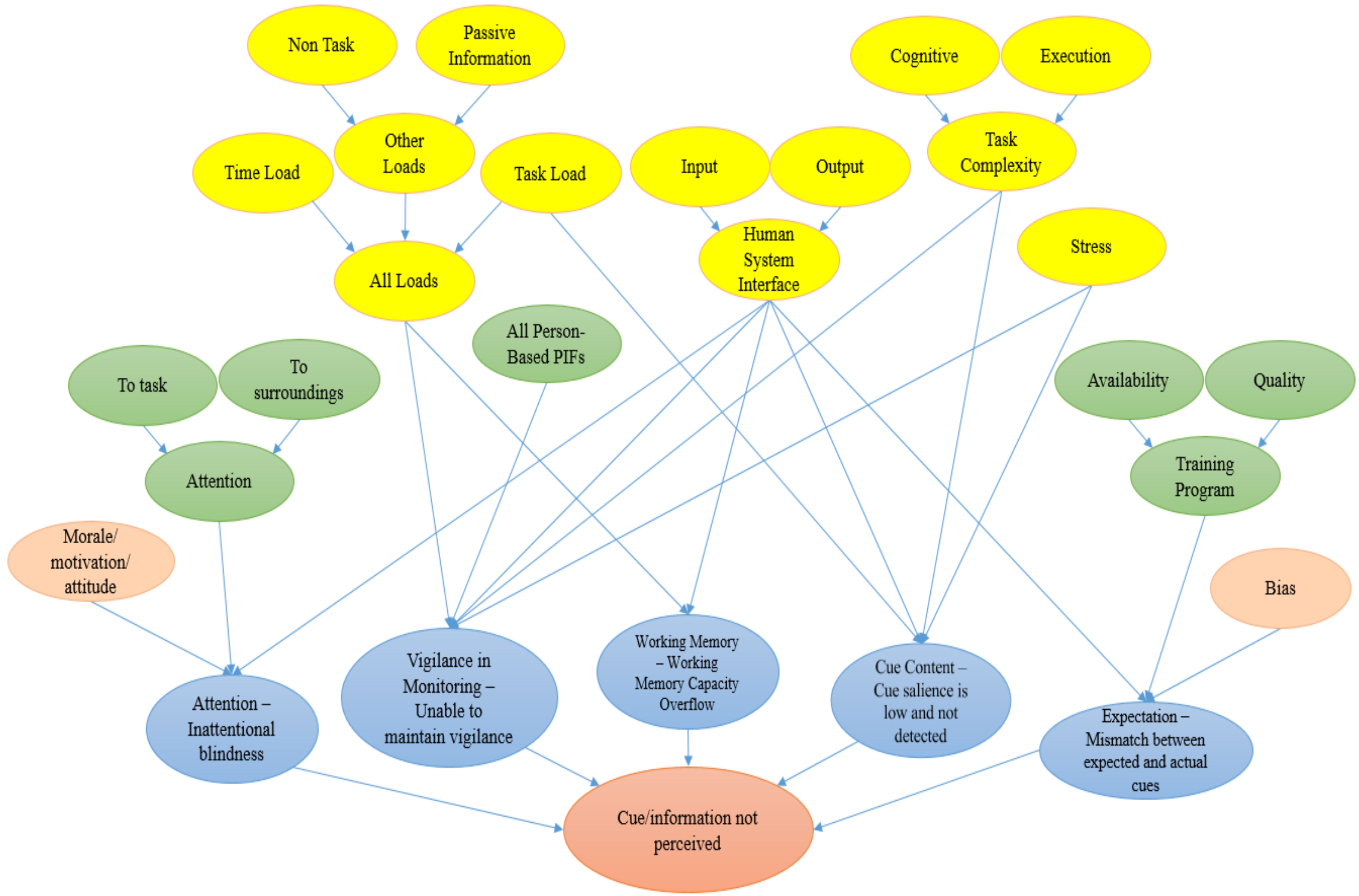


Figure 2: The BBN outlining the proximate cause of failure: Cue/Information not Perceived

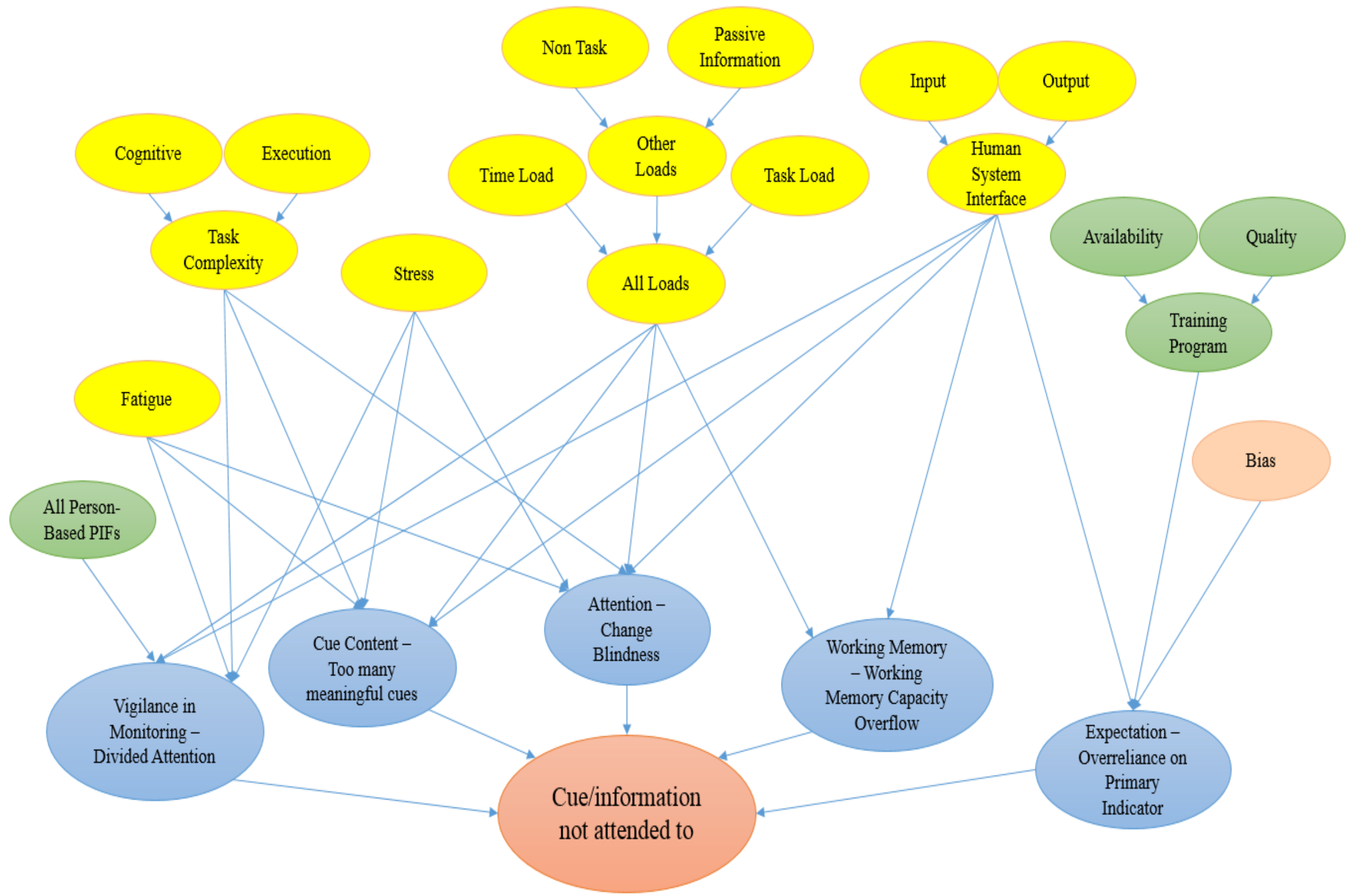


Figure 3: The BBN outlining the proximate cause of failure: Cue/Information not Attended to

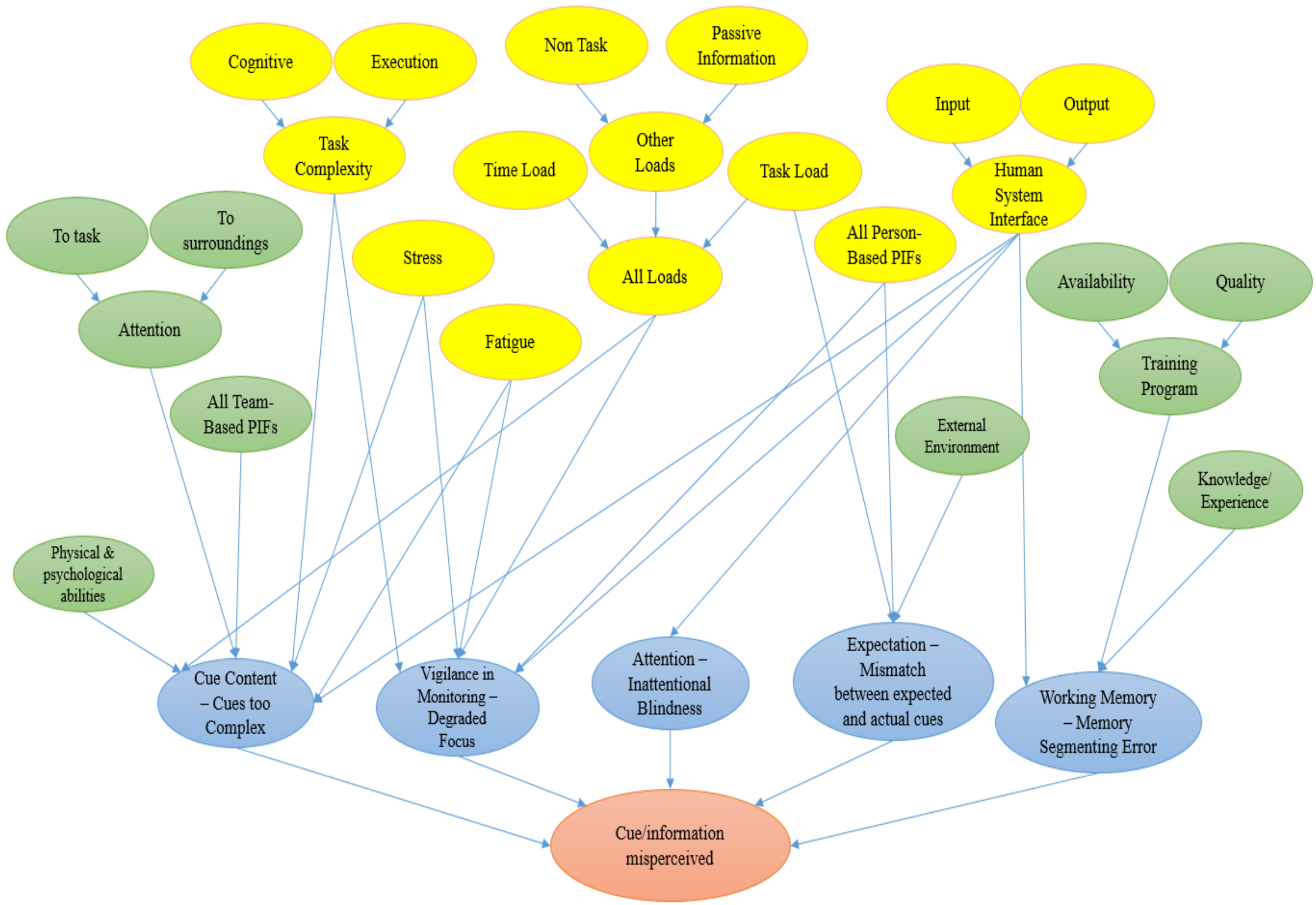


Figure 4: The BBN outlining the proximate cause of failure: Cue/Information not Misperceived

Appendix C3. PIF Taxonomy for Project

The BBNs that the SyRRA group designed in Appendix C2 stick to the following PIF taxonomy defined by Groth and Mosleh [10].

Organization-based	**Team-based**	**Person-based**	**Situation/stressor-based**	**Machine-based**
Training program	Communication	Attention	External environment	HSI
Availability	Availability	To task	Conditioning events	Input
Quality	Quality	To surroundings	Task load	Output
Corrective action program	Direct supervision	Physical & psychological abilities	Time load	System response
Availability	Leadership	Alertness	Other loads	
Quality	Team coordination	Fatigue	Non-task	
Other programs	Team cohesion	Impairment	Passive information	
Availability	Role awareness	Sensory limits	Task complexity	
Quality		Physical attributes	Cognitive	
Safety culture		Other	Execution	
Management activities		Knowledge/experience	Stress	
Staffing		Skills	Perceived situation	
Scheduling		Bias	Severity	
Workplace adequacy		Familiarity with situation	Urgency	
Resources		Morale/motivation/attitude	Perceived decision	
Procedures			Responsibility	
Availability			Impact	
Quality			Personal	
Tools			Plant	
Availability			Society	
Quality				
Necessary information				
Availability				
Quality				

Appendix C4. Example of SACADA State-to-PIF Mapping

The following table shows the draft of SACADA state-to-PIF and characteristic mapping for demonstration.

SACADA Situational Factor		SACADA States	PSF (Groth, 2012)	Characteristics
Importance		1:Other	Perceived Situation Severity Perceived Situation Urgency	NA NA
		2:Significant	Perceived Situation Severity Perceived Situation Urgency	NA NA
		3:Safety Significant	Perceived Situation Severity Perceived Situation Urgency	NA NA
		4:Critical	Perceived Situation Severity Perceived Situation Urgency	NA NA
Cognitive Type		1:Monitoring/Detection	N/A	NA
		2:Diagnosis & Response Planning	N/A	NA
		3:Manipulation	N/A	NA
		4:External Communication	N/A	NA
Monitoring/Detection	Detection Type	1:Alarm	HSI Output	Good
		2:Status Tile	HSI Output	Good
		3:Meter	HSI Output	Good
		4:Indication Light	HSI Output	Good
		5:Flag	HSI Output	Good
		6:Computer	HSI Output	Good
		7:Other	HSI Output	Good
Alarms/Status Tile	Detection Mode	1:Self-Revealing	HSI Output	Good
		2:Procedure Directed Check	Resources Procedures Availability Resources Procedures Quality	Good Good
		3:Procedure Directed Monitoring	Resources Procedures Availability Resources Procedures Quality	Good Good
		4:Awareness/Inspection	Attention to Surroundings	Good
	Status of Alarm Board	1:Dark	Other Loads Passive information	Normal
		2:Busy	Other Loads Passive information	Normal
		3:Overloaded	Other Loads Passive information Task Complexity	Bad Bad
	Expectation of Alarm/Indication Change	1:Expected	System Response	Good
		2:Not Expected	System Response	Bad
		3:Not Applicable	N/A	NA
Meter/Light/Flag	Detection Mode	1:Procedure Directed Check	Resources Procedures Availability Resources Procedures Quality	Good Good
		2:Knowledge-Driven Monitoring	Knowledge/Experience	Good
		3:Procedure-Directed Monitoring	Resources Procedures Availability Resources Procedures Quality	Good Good
		4:Awareness/Inspection	Attention to Surroundings	Good
	Individual Indicator	1:Slight Change	HSI Output	Poor/Bad
		2:Distinct Change	HSI Output	Good
	Mimics/Display etc.	1:No Mimics	HSI Output	Poor
		2:Small Indications	HSI Output	Bad
		3:Similar Displays	HSI Output	Bad
		1:Primarily Diagnosis	N/A	NA

Diagnosis and Response Planning	Diagnosis or Response Planning	2:Primarily Response Planning/Decision Making	N/A	NA
Diagnosis	Diagnosis Basis	1:Procedure	Resources Procedures Availability	Good/Bad
		2:Skill	Resources Procedures Availability Skills	Good/Bad Good/Bad
		3:Knowledge	Resources Procedures Availability Knowledge/Experience	Good/Bad Good/Bad
	Familiarity	1:Standard	Familiarity with Situation	Good
		2:Novel	Familiarity with Situation	Good
		3:Anomaly	Resources Procedures Quality Familiarity with Situation	Poor Bad
	Outcome	1:Procedure-Based Activity	Resources Procedures	Good
		2:Skill-Based Behavior	Skills	Good
		3:Knowledge-Based Behavior	Knowledge/Experience	Good
	Information Integration	1:Timing of Information	Resources Necessary Information Availability	Poor
		2:Ambiguous Information	Resources Necessary Information Quality	Poor
		3:Integration Required	Resources Necessary Information Quality Task Complexity Cognitive	Poor Poor
	Diagnosis Information Specificity	1:Specific	Resources Necessary Information Quality Task Complexity Cognitive	Good Poor
		2:Not Specific	Resources Necessary Information Quality	Poor
		3:Not Applicable	N/A	NA
	Information Quality	1:Missing Information	Resources Necessary Information Availability	Poor
		2:Misleading Information	Resources Necessary Information Quality	Poor
		3:Conflicting Information	Resources Necessary Information Quality	Poor
Response Planning/Decision Making	Decision Basis	1:Procedure	Resources: Procedures	Good/Bad
		2:Skill	Skills	Good/Bad
		3:Knowledge	Knowledge/Experience	Good/Bad
	Familiarity	1:Standard	Familiarity with Situation	Good
		2:Adaptation Required	Familiarity with Situation	Bad
		3:Anomaly	Resources Procedures Quality Familiarity with Situation	Poor Bad
	Uncertainty	1:Clear	Resources Procedures Quality Resources Necessary Information Quality	Good Good
		2:Uncertain	Resources Procedures Quality Resources Necessary Information Quality	Poor Poor
		3:Competing Priorities	Attention to Task Attention to Surroundings Task Complexity Perceived Situation Urgency	Poor Poor High High
		4:Conflicting Guidance	Resources Procedures Quality Communication Quality Direct Supervision Leadership	Poor Poor Poor
	Outcome	1:Procedure-Based Activity	Resources Procedures	Good
		2:Skill-Based Behavior	Skills	Good
		3:Knowledge-Based Behavior	Knowledge/Experience	Good

Manipulation	Type of Action	1:Simple and Distinct	Task Complexity Execution	Low
		2:Order	Task Complexity Execution	Medium
		3:Maintaining	Task Complexity Execution	High
	Location	1:Main or Auxiliary Control Board	HSI Input	Good
		2:Back Control Panels	HSI Input	Bad
	Guidance	1:Procedure	Resources Procedures	Good
		2:Skill of the Craft (Non-Faulted Hardware)	Skills	Good/Bad
		3:STAR (Faulted Hardware)	Conditioning Events	Bad
	Recoverability	1:Immediately Recoverable	System Response Perceived Decision Personal Impact	Low Low
		2:Recoverable With Significant Efforts	System Response Perceived Decision Personal Impact	High High
		3:Unrecoverable	Perceived Decision Personal Impact	High
	Additional Factors	1:Unintuitive Plant Response	System Response	Bad
		2:Unintuitive Controls	HSI Input	Bad
		3:Additional Mental Effort Required	Task Load	Bad
		4:Inadequate Feedback	Resources Necessary Information Availability Resources Necessary Information Quality System Response	Bad Bad Bad
		5:Similar Controls	HSI Input	Bad
Communication Between Crew and Simulator Booth	Communication Driver	1:Specifically Procedure Directed	Communication Quality	Medium
		2:Not Specifically Driven	Communication Quality	Medium
	Direction of Communication	1:From Booth	Communication Quality	Medium
		2:To Booth	Communication Quality	Medium
		3:Public Address Announcement	Communication Quality	Medium
		4:Other	Communication Quality	Medium
Overarching Issues	Workload	1:Normal	Task Load	Low
		2:Concurrent Demands	Task Load	High
		3:Multiple Concurrent Demands	Task Load Stress	High High
	Time Criticality	1:Expansive Time Available	Time Load Perceived Situation Urgency	Low Low
		2:Nominal Time Available	Task Load	Low
		3:Barely Adequate Time Available	Time Load Perceived Situation Urgency	High High
	Extent of Communications Required	1:Nominal Communication	Communication Availability Other Loads Non-task	High High
		2:Extensive Onsite Communication	Communication Availability Other Loads Non-task	High High
		3:Extensive Communication Within the Control Room	Communication Availability Other Loads Non-task	High High
	Other Demands/Factors	1:Non-Standard	Other Loads	High
		2:Noisy Background	External Environment	High
		3:Coordination	Other Loads Non-task	High
		4:Communicator Unavailable	Communication Availability	High
		5:Multiple Demands	Task Load	High
		6:Memory Demands	Task Load	High

References

[1] K. M. Groth, C. L. Smith and L. P. Swiler, "A Bayesian method for using simulator data to enhance human error probabilities assigned by existing HRA methods," *Reliability Engineering & System Safety,* vol. 128, p. 32–40, 2014.

[2] K. Groth, "A framework for using SACADA to enhance the qualitative and quantitative basis of HRA," in *Probabilistic Safety Assessment and Management PSAM 14*, Los Angeles, CA, 2018.

[3] K. Murphy, *Bayes Net Toolbox,* 1997-2002.

[4] Mathworks, *MATLAB R2018a,* 2018.

[5] D. Gertman, H. Blackman, J. Marble and C. Smith, "The SPAR-H Human Reliability Analysis method. Technical Report NUREG/CR-6883," US Nuclear Regulatory Commission, Washington DC, 2005.

[6] B. Hallbert, A. Kolaczkowski and E. Lois, "The Employment of Empirical Data and Bayesian Methods in Human Reliability Analysis: A Feasibility Study," Idaho National Laboratory, NUREG/CR-6949, Washington DC, 2007.

[7] BayesFusion, LLC, *GeNIe Modeler,* 2018.

[8] A. M. Whaley, J. Xing, R. L. Boring, S. M. L. Hendrickson, J. C. Joe and K. L. Le Blanc, "Building a Psychological Foundation for Human Reliability Analysis," Office of Nuclear Regulatory Research, 2012.

[9] K. M. Groth, R. C. Smith and R. Moradi, "A hybrid approach to HRA using simulator data, causal models, and cognitive science," *Submitted to Reliability Engineering and System Safety,* 2018.

[10] K. Groth and A. Mosleh, "A data-informed PIF hierarchy for model-based Human Reliability Analysis," *Reliability Engineering and System Safety,* vol. 108, p. 154–174, 2012.

[11] K. Zwirglmaier, D. Straub and G. K. M, "Capturing cognitive causal paths in human reliability analysis with Bayesian network models," *Reliability Engineering & System Safety,* vol. 158, p. 117–129, 2017.

[12] Y. J. Chang, D. Bley, L. Criscione, B. Kirwan, A. Mosleh, T. Madary, R. Nowell, R. Richards, E. M. Roth, S. Sieben and A. Zoulis, "The SACADA database for human reliability and human performance," *Reliability Engineering & System Safety,* vol. 125, p. 117–133, 2014.