Attempt to predict human error probability in different industry sectors using data from major accidents and Bayesian networks

C. Morais*[1a,b], R. Moura[b,a], M. Beer[c,d] and E. Patelli[a]

[a] University Institute for Risk and Uncertainty, University of Liverpool, United Kingdom
[b] National Agency for Petroleum, Natural Gas and Biofuels (ANP), Brazil
[c] Institute for Risk and Reliability, Leibniz University Hannover, Germany
[c] School of Civil Engineering & Shanghai Institute of Disaster Prevention and Relief, Tongji University, China

Abstract: Looking into aviation, nuclear power generation, oil & gas and chemical industries, one can notice their interaction between organisational factors, technological systems and humans – the so-called complex socio-technical systems.
To prevent accidents from occurring, engineers carry out safety analyses, and to calculate the likelihood of some scenarios they have to know the failure rates. It is easy to understand that components' failure rates are evaluated differently from human's failure rate. This subject is called Human Reliability Analysis (HRA), and it should be analysed ideally through the cooperation between engineers, psychologists and sociologists. Bayesian network is a probabilistic methodology that allows these three professional groups to better communicate through its intuitive graphical representation of the conditional probabilities.
This paper presents a Bayesian model of a dataset of major accidents from different industrial sectors, instead of using scenario simulators and expert elicitation.
The steps required to construct a model are presented together with tools for the assessment of the conditional probability and the model validation. The proposed approach allows to calculate the Human Error Probabilities as outputs of the model.

Keywords: Human Reliability Analysis, Human Error Probability, Bayesian network, major accidents dataset.

1. INTRODUCTION

To illustrate the industrial need for Human Reliability Analysis HRA and how research can contribute to the area providing more consistent Human Error Probabilities, this paper starts with the following example: Imagine a team is designing a new engineered system (e.g. a chemical industry) where an operator has to open an equipment door only after its internal pressure drops. The pressure during that equipment operation is high enough to cause a fatality, so the operator has to wait to open the equipment door at the right moment by observing a pressure gauge.

During one of the QRA (quantitative risk assessment) meetings, after identifying the hazard, the team has to know if the risk level of this operation meets the risk criteria of their organisation (or the safety regulator). If not, they have to recommend additional safety barriers.

To assess the overall risk level of this operation, one has to account for equipment and human failure. That is because, for an operator to open the equipment door at the wrong moment, one of the two following failures have to happen before: the operator failing to observe the pressure gauge or the pressure gauge displaying a false measure.

The pressure gauge supplier has informed its failure rate. How does a team should assess the human failure rate? This number is usually called Human Error Probability (HEP), and there are different ways to obtain it.

[1] caroline.morais@liverpool.ac.uk

1.1. State of the art for obtaining human error probabilities (HEP)

A HEP must ideally be obtained by observing someone's errors and knowing how many times he/she performs an action – correctly or not, which is described in the Equation (1) as the opportunities for error.

$$HumanErrorProbalitity = \frac{Numberofobservederrors}{Numberofopportunitiesforerror} \quad (1)$$

Ideally, these numbers should be extracted from real operations' observation. However, this is a difficult task as one should observe all the operational phase of an industrial installation. Furthermore, even the errors that did not lead to recordable events should be accounted – such accidents or near-misses (i.e. events with the potential for undesirable consequences [1]).

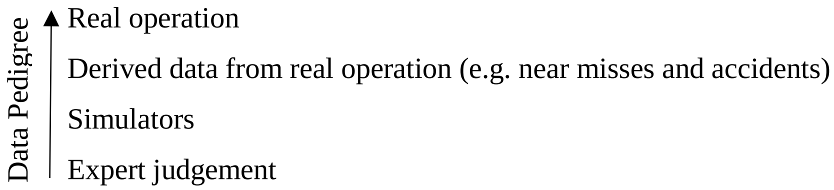
To tackle this problem, some methods were developed to quantify those HEPs, among two other objectives (i.e. identifying the errors and their consequences, and discuss ways of reducing the likelihood of errors or remediation of those errors in the system). These methods are called Human Reliability Analysis (HRA).

HRA has started to be developed in the 1960's and is enforced by different regulators around the world. There are at least thirty nine HRA methods developed, but few are recognised and accepted by the safety regulators [2], and fewer are used in practice by organisations. Quantitative HRA techniques generally fall into two categories: those using a database and those using expert opinion [3]. Although even "data-based" methods tend to rely to some extent on expert opinion, the core data used is from the real operation or from simulators.

As summarised in Figure 1, the data considered to have better quality pedigree, giving better HEP estimates, are the ones closer to the real operation. HEPs based on expert opinion alone is not advisable, as experts potentially bring uncertainty and bias to the HEPs estimates. Simulator's data are not considered with the same quality as those from the real operation because, during simulation events, the operators are somewhat prepared to an event to happen, and possibly less concerned with production goals than reality [3]. Although research is being developed on creating correction factors to account these effects [3], [4], [5] e [6], simulators' data are still not considered as noble as the derived data from the real operation.

Figure 1. Data quality pedigree for calculating Human Error Probability

$Huma$



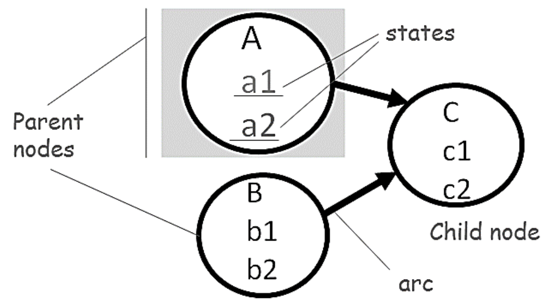
For this reason, much research is being conducted using operationally derived data from the real operation, as near misses and accidents occurred in industrial installations [7], [8].

1.2. Bayesian network as a tool to gather data for HEPs

The relationships between the parameters described in the methodology section of this paper were modelled into a Bayesian network (BN), known as a systematic way of learning from experience and incorporating new evidence (deterministic or probabilistic).

BNs can be defined as statistical models used to represent probability distributions, providing combined probability distribution associated with an event and exploiting information about the existing conditional dependencies [9]. BNs can be represented by acyclic graphs, where nodes are connected to each other by arcs (Figure 2). Child nodes must have a causality relationship with each parent node. Figure 2 is an example of a graphic representation for the conditional probability equations below (Equations 2 and 3).

Figure 2 - Directed acyclic graphs typical of a Bayesian network



P(C=c1 | A=a1,B=b1) (2)

P(C=c2 | A=a1,B=b1) = 1-P(C=c1 | A=a1,B=b1) (3)

The number of combinations to consider to generate a child's node conditional probability is two (a pair of combinations) to the power of the number of states of the parent nodes (2^states of the parent nodes). All these possible combinations are usually accounted into Conditional Probability Tables (CPT), as shown in Table 1.

Table 1 – Example of Conditional Probability Table for the simplified BN of Figure 2

A	State 1		State 2	
B	State 1	State 2	State 1	State 2
State 1 of C	P(C=c1 | A=a1,B=b1)	P(C=c1 | A=a1,B=b2)	P(C=c1 | A=a2,B=b1)	P(C=c1 | A=a2,B=b2)
State 2 of C	P(C=c2 | A=a1,B=b1) Or 1-P(C=c1 | A=a1,B=b1)	P(C=c2 | A=a1,B=b2) Or 1- P(C=c1 | A=a1,B=b2)	P(C=c2 | A=a2,B=b1) Or 1- P(C=c1 | A=a2,B=b1)	P(C=c2 | A=a2,B=b2) Or 1-P(C=c1 | A=a2,B=b2)

With BN it is possible to combine different sources of information and make HRAs compatible with Probabilistic Safety Assessments, due to its probabilistic representation of uncertainty [10].

2. DATA USED IN THE MODEL – MAJOR ACCIDENTS DATASET

The dataset used in this work has been obtained from the analysis of 238 major accident reports from different industrial sectors using the same framework, with the intention to optimise the learning from cross-sector accidents [11]. The framework used was the classification scheme adapted from the Cognitive reliability and error analysis method (CREAM) [12].

The dataset, named MATA-D, the Multi-attribute Technological Accidents Dataset [11] contains the relevant information from the accident reports, condensed into a table with the numbers zero and one. The presence of factors that could have contributed to an accident (the so-called Performance Shaping Factors, PSFs) was accounted into the dataset as the number one, as well as indications of workers' cognitive functions and actions executed that contributed to the accidents. When there was no evidence of an organisational, technological and person-related factor, the number zero was inserted. Tables 2 and 3 relate the PSFs, errors of cognition and execution used to create the dataset. To have a full description and meaning of each PSF, error of cognition and execution, see [12].

Table 2. Performance shaping factors used as a framework to create the MATA-D dataset [11], [12], [13]

Organisational factors	Technological Factors	Person Related Factors
Communication failure	*Equipment failure*	Memory failure
Missing information	Software fault	Fear
Maintenance failure	*Inadequate procedure*	*Distraction*
Inadequate quality control	Access limitations	Fatigue
Management problem	Ambiguous information	Performance variability
Design failure	*Incomplete information*	Inattention
Inadequate task allocation	Access problems	Physiological stress
Social pressure	Mislabeling	Psychological stress
Insufficient skills		Functional impairment
Insufficient knowledge		Cognitive style
Temperature		*Cognitive bias*
Sound		
Humidity		
Illumination		
Other		
Adverse ambient conditions		
Excessive demand		
Inadequate workplace lay-out		
Inadequate team support		
Irregular working hours		

Table 3. Errors of cognition and execution used as a framework to create the MATA-D dataset [11], [12], [13]

Errors of cognition		*Errors of execution*
Observation	*Observation missed*	*Wrong time*
	False observation	*Wrong type*
	Wrong identification	*Wrong place*
Interpretation	*Faulty diagnosis*	Wrong object
	Wrong reasoning	
	Decision error	
	Delayed interpretation	
	Incorrect prediction	
Planning	*Inadequate plan*	
	Priority error	

3. METHODOLOGY – MODELLING THE DATA

3.1. The Bayesian model

Previous works had already used the Bayesian network to model Human Performance under organisational, technological and person-related factors for different purposes, as classified and investigated by [10].

The following procedure was used to create the model: 1st select the nodes and their states, 2nd develop the BN structure, 3rd asses the conditional probability table, and 4th verification and validation step [10]. Figure 3 presents a summary of what has been made to achieve each step. For more details, and simplifications considered in the model, see [13]. The preliminary results of the validation step are discussed in the present paper, in the results' section.

Figure 3 – Summarised process to create and evaluate the BN model

Selection of nodes and states

- The parent nodes used in the model are the same PSFs used to classify the accident dataset.
- The child nodes are errors of cognition and errors of execution, also the same from the dataset.
- The states of the nodes have been defined as the ‘presence’ or ‘absence’ of the PSFs and the human errors reported in the reports (in the dataset as '0' and '1'). For some nodes, it was also created a state named "state of ignorance", when no data was available for a set of combinations.
- Note that, at a previous model [13], the errors of cognition were also parent nodes of errors of execution. However, it have been decided to remove the links between cognitive and execution errors in the model used in this paper. That's because there were many accidents that didn't have a cognitive error preceding an execution, and this lack of combinations was preventing the software to do the inference to the CPT whenever the probabilities to the states were both equal to zero. Although the lack of combinations was solved adding a state named "state of ignorance" for the affected nodes, this new state was propagated to the CPTs of the next child nodes, causing more impossible combinations and more imprecision - so it was decided to remove the links between those nodes.

Structure

- Parent nodes were linked by arrows to the child nodes.
- The connections between the nodes were proposed based on relations between factors and human errors identified on [14].
- Simplifications to the structure had to be made because the algorithm used do not support an elevated number of conections to one child node. The simplification was also helpful to decrease the number of non-exhistent combinations between the PSFs. The simplifications were applied not only to the connections but also to the number of nodes.
- The nodes were restricted to the factors and human errors considered significant for the ocurrence of major accidents. The selection was made through the aplication of an algorithm aplied on [14], named SOM (self-organising maps).

Conditional Probability Table (CPT)

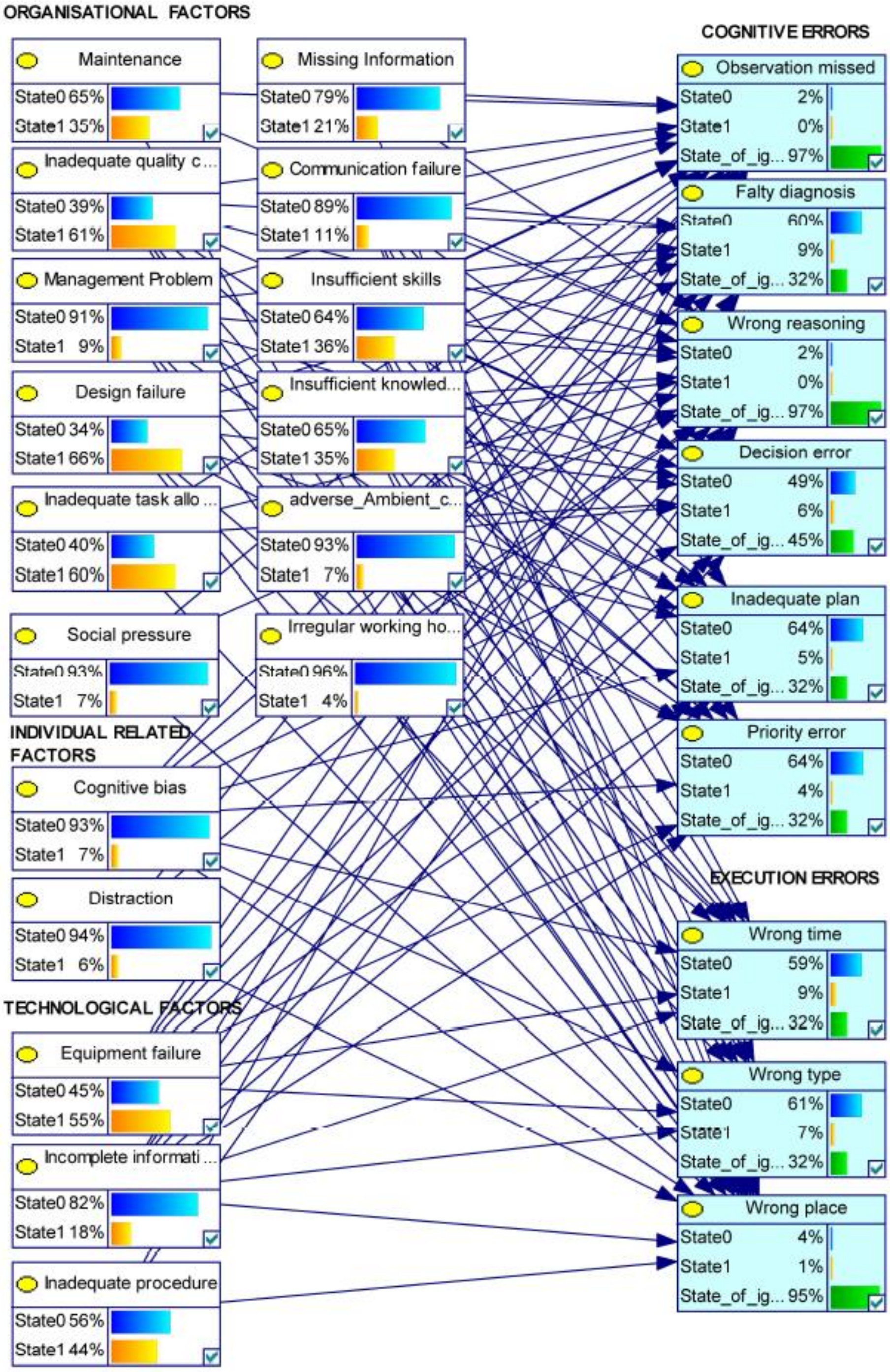
- The prior probabilities to input the CPTs for each node were also obtained from the dataset MATA-D. They have been obtained by calculating how many times a specific combination of factors and errors have occurred, divided by the total number of accidents of the dataset.
- Software MATLAB was used to calculate the probabilities of each possible combination between PSFs and errors.
- Software GeNIe Modeler, for academic use [15], was used to create the model and calculate, and to calculate the posterior probabilities, using the clustering algorithm embedded in the software. The node type used was ‘chance – general’. Other software containing Bayesian network toolboxes can be used for the same purpose, e.g. Cossan-X [16], Netica [17] and Uninet [18].
- The child node with the highest number of combinations had nineteen parent nodes directly linked to it. That means 524,288 combinations (two to the power of nineteen) inside the node's CPT.

Verification & validation

- Verification: to verify if the model behaved according to its specifications, some scenarios were created, changing some set of the PSFs to its extremes (sets of parent nodes were assumed to be 0 and 1). After that, the posterior probabilities of the human errors' nodes were calculated again.
- Some discussions about the sensitivity of each error for each PSF can be evaluated from results obtained at the verification step.
- Validation: The first attempt to validate the model was made comparing the BN results against HEPs described at [12], which the author informed to have extracted from a variety of sources, mainly [19], [20], [21] and [22]. These preliminar results are discussed in the present paper (see Results' section)

The Bayesian model of the major accident dataset is presented in Figure 4. The model inputs are the number of times each parameter (PSFs, cognitive functions and human actions) was observed in a dataset of major accidents. The model outputs are the human failure probabilities, as the results presented in the next section.

Figure 4. Bayesian model of PSFs and errors of cognition and execution using MATA-D



4. RESULTS

4.1. Human Error Probabilities (HEPs) from major accidents dataset

After building the model and inserting the prior probabilities of parent and child nodes (through their conditional probability tables), the marginal probability distributions were calculated using Genie software embedded algorithm. The results are presented in Table 4.

Table 4. Results of Human Error Probabilities (HEP) from the Bayesian model

Cognitive Error Probability		Execution Error Probability	
Observation		Wrong time	8.99×10^{-2}
Observation missed	3.91×10^{-3}	Wrong type	7.36×10^{-2}
Interpretation		Wrong place	1.09×10^{-2}
Faulty diagnosis	8.61×10^{-2}		
Wrong reasoning	4.95×10^{-4}		
Decision error	6.12×10^{-2}		
Planning			
Inadequate plan	4.81×10^{-2}		
Priority error	3.81×10^{-2}		

4.2. The validity of the HEPs found

To validate a model, one should test if the system does what is supposed to do in the real world: if the outputs have a good correlation to 'real world data' [3]. In other words, we should ask ourselves: "Did we built the right system?" [23].

Considering the data quality criteria proposed in Figure 1, there is an understanding that a model should be tested against another kind of real operation or derived data [3].

The first attempt to validate the model was made comparing the BN results against human error probabilities described at [12], which the author informed to have extracted from a variety of sources, mainly [19], [20], [21] and [22]. According to him, data sources for behaviours such as *observation* and *execution* were relatively well established at that time (1998). In the other hand, the author declared that *interpretation* and *planning* behaviours were mostly based on expert judgements. The results of this first validation attempt can be seen at Table 5. Note that not all the cognitive and execution errors used for the model were presented at this table, only the ones that the author on [12] proposed a HEP reference.

Table 5 – Comparing HEPs from present model with other sources [12] (that compiled from [19]-[22])

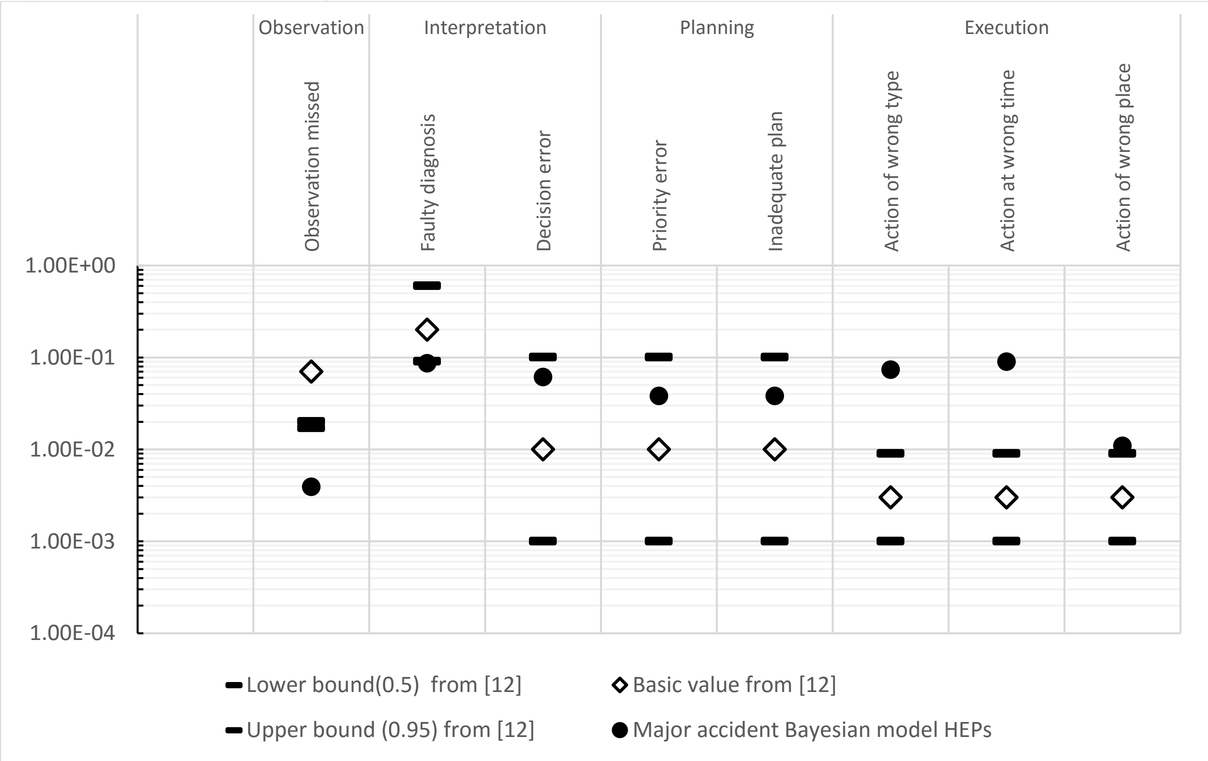
Cognitive functions and human actions	Generic failure type	Lower bound(0.5) from [12]	Basic value from [12]	Upper bound (0.95) from [12]	Major accident Bayesian model HEPs
Observation	Observation not made	2.00×10^{-2}	7.00×10^{-2}	1.70×10^{-2}	3.91×10^{-3}
Interpretation	Faulty diagnosis	9.00×10^{-2}	2.00×10^{-1}	6.00×10^{-1}	8.61×10^{-2}
	Decision error	1.00×10^{-3}	1.00×10^{-2}	1.00×10^{-1}	6.12×10^{-2}
Planning	Priority error	1.00×10^{-3}	1.00×10^{-2}	1.00×10^{-1}	3.81×10^{-2}
	Inadequate plan	1.00×10^{-3}	1.00×10^{-2}	1.00×10^{-1}	3.81×10^{-2}
Execution	Action of wrong type	1.00×10^{-3}	3.00×10^{-3}	9.00×10^{-3}	7.36×10^{-2}
	Action at wrong time	1.00×10^{-3}	3.00×10^{-3}	9.00×10^{-3}	8.99×10^{-2}
	Action of wrong place (or out of sequence)	1.00×10^{-3}	3.00×10^{-3}	9.00×10^{-3}	1.09×10^{-2}

Figure 5 provides a better visualisation of the model results, showing that the results for the cognitive functions of interpretation and planning are closest to the value proposed by literature and most of them inside the boundaries also proposed by literature.

For the cognitive function of ‘Observation’, the HEP obtained is below the ‘lower bound’– showing a tendency of this result to be optimistic. This means that the results suggest a smaller probability of occurrence of a human failure than those compiled by [12]. It is not desirable in HEPs to have a high degree of optimism, as it can lead to under-estimated risk predictions [3]. However, a better understanding of the data used in [12] is still needed, as the basic value proposed is over the upper bound of the same literature source.

On the other hand, all the execution errors obtained from the BN model were greater than the higher bound proposed by the literature. In other words they show a pessimistic tendency, being the cost of pessimism being not desired as well, as it can lead to slightly over-designed plants [3].

Figure 5 – Graph comparing HEPs from the BN model with HEPs from literature [12]



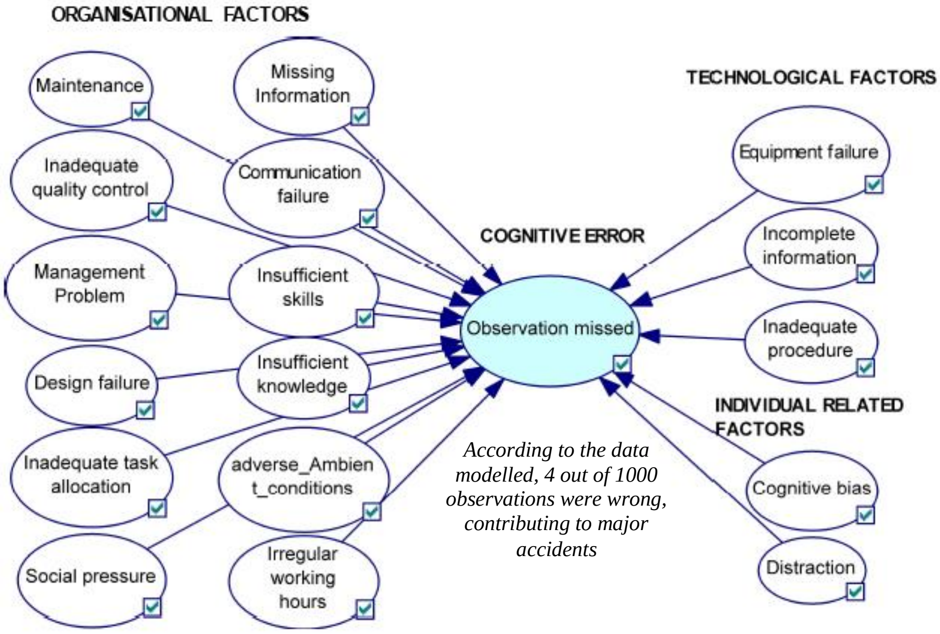
Apart from the apparent mistake on the literature for the Observation cognitive function data, better understanding of the all the data used in [12] is still needed, as it is possible that the HEP generated could serve to validate data from [12] and not the opposite. Furthermore, it has to be investigated if the HEPs found by the model could be used to validate some HRA methodologies.

Together with a better understanding of the degree of optimism and pessimism, a complete validation is yet to be pursued and presented in future work. Other criteria, established at [3] may be considered, i.e. the presence of a predictive relationship (usually a correlation) and precision (agreement with HEPs within a factor between 3 and 10).

4.3. Example of application of the HEP results

To explain one of the ways these results can be applied in industry practice, the example of the introduction section can be used. Figure 6 presents the BN model with only the cognitive function and human action considered in the example: the action of *opening the equipment door at the right moment* can be described as “wrong time” and *reading a pressure gauge* can be vulnerable to “wrong observation”.
The model result for “wrong observation” show that from one thousand observations conducted, only four have the potential to be wrong (HEP of wrong observation is 3.91×10^{-3}).

Figure 6 – Example of HEPs use in an industrial context



5. CONCLUSION

Quantified Human Reliability Analysis (HRA) is useful to check if one organisation or regulator risk criteria are met. One of the aims of the HRA is to quantify the likelihood of failure, the so-called Human Error Probability (HEP). Accurate and realistic HEPs are important to help decision-makers to prioritise which risks to tackle and to intercede in the factors that impact human performance.

As there is imprecise information of the number of opportunities of errors over a hypothetical operational lifetime of a system to generate an ideal HEP, it was chosen to use a probabilistic tool named Bayesian network (BN).

The BN model of human performance proposed in this paper uses the same framework of a dataset of major accidents to achieve probability estimates for errors of cognition and execution. The basic aspects of the model (nodes, states, structure and conditional probability table) were all extracted from the MATA-D dataset [11] and [14], avoiding expert judgement. The Bayesian model inputs are the number of times each parameter (performance shaping factors - PSFs) was observed in a dataset of major accidents, and the outputs are the human error probabilities (cognitive and execution errors**).**

This is not the first time a dataset of incident events derived from real operation is used to find HEPs (e.g. [7], [8]). However, the previous publicly available works were focused on near-misses events – whereas the present one is based on a dataset of major accidents. Although the approach seems promising for its data quality, as investigation reports of major accidents have the potential to uncover more factors that trigger a human error then near misses reports [1], the results obtained suggest caution before use.

Not all the HEPs found are inside the range proposed on existing HEPs from other sources [12]. All the execution errors (wrong time, type and place) have a higher probability of occurrence if compared to what has been practised, and one cognitive error (observation missed) had shown smaller probability. Kirwan [3] defends that if techniques are imprecise, it is desirable that they are pessimistic rather than optimistic, as it will compromise the cost of over-designed plants rather their safety.
A possible interpretation of these results is that the other methods being compared in [12] had analysed general HEPs, not necessarily considering the consequences of errors, whereas the present results show an indication (and probability estimation) of errors causing major accidents.

For this reason, it is suggested that future development of the model should include a broader investigation into the verification and validation requisites of the model – including the discussion if the HEP generated from this model should be validated or better serve to validate other HRA methods or HEP results.

Acknowledgements

Thanks to ANP, the Brazilian Oil & Gas Regulator, and EPSRC-"Quantification and Management of Risk & Uncertainty in Complex Systems & Environments." (EP/L015927/1), for being supportive of the research of Human Reliability Analysis and Human Factors.

References

[1] Center for chemical Process Safety – CCPS. *"Guidelines for Risk Based Process Safety"*, John Wiley & Sons, Inc., New Jersey, (2007).
[2] J. Bell and J. Holroyd. *"Review of human reliability assessment methods",* Health and Safety Laboratory (2009).
[3] B. Kirwan. *"Validation of human reliability assessment techniques: part 1—validation issues",* Safety Science, 27(1), pp.25-41, (1997).
[4] J. Forester. "*The US HRA Empirical Study: Assessment of HRA Method Predictions Against Operating Crew Performance on a US Nuclear Power Plant Simulator"*. United States Nuclear Regulatory Commission, Office of Nuclear Regulatory Research (2016)
[5] J. Forester et al. *"The International HRA Empirical Study–Final Report–Lessons Learned from Comparing HRA Methods Predictions to HAMMLAB Simulator Data".* HPR-373 OECD Halden Reactor Project, Norway, (2013)
[6] Y. Kim et al. *"A classification scheme of erroneous behaviors for human error probability estimations based on simulator data*", Reliability Engineering & System Safety, 163, pp.1-13, (2017)
[7] W. Preischl and M. Hellmich. *"Human error probabilities from operational experience of German nuclear power plants",* Reliability Engineering & System Safety, 109, pp.150-159, (2013)
[8] 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,226(4), pp.361-379, (2012).
[9] S. Tolo et al. *"Bayesian Network Approach for Risk Assessment of a Spent Nuclear Fuel Pond",* Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management, pp. 598-607. (2014)
[10] L. Mkrtchyan et al. *"Bayesian belief networks for human reliability analysis: A review of applications and gaps"*, Reliability engineering & system safety, 139, pp.1-16, (2015).
[11] R. Moura. *"Learning from past accidents to improve system design"*, Safety Science, 84, pp. 37-45, (2016)
[12] E. Hollnagel. *"Cognitive reliability and error analysis method"*, Elsevier, 1998.
[13] C. Morais et al. *"Human reliability analysis – accounting for human actions and external factors through the project life cycle",* Proceedings of the European Safety and Reliability Conference (2018) (*in press*)
[14] R. Moura et al. *"Learning from major accidents: graphical representation and analysis of multi-at-tribute events to enhance risk communication*", Learning from Incidents Special Issue of the Safety Science Journal, 99, pp. 58-70 (2017)
[15] BayesFusion, LLC. GeNIe Modeler. http://www.bayesfu-sion.com/, Accessed: 30 November 2017.
[16] E. Patelli et al. "OpenCossan 2.0: an efficient computational toolbox for risk, reliability and resilience analysis", Proceedings of the joint ICVRAM ISUMA UNCERTAINTIES conference, (2018) http://icvramisuma2018.org/cd/web/PDF/ICVRAMISUMA2018-0022.PDF
[17] Norsys Software Corp. "Sensitivity to Findings". Netica Software Documentation. (2006) https://www.norsys.com. Accessed: 13 January 2018.
[18] R. M.,Cooke et al. *"Continuous/Discrete non-parametric Bayesian belief nets with UNICORN and UNINET"*, Technical report, TU Delft. (2007). http://www.lighttwist.net/wp/uninet , Accessed: 13 January 2018.

[19] A. N. Beare et al. "*A simulator-based study of human errors in nuclear power plant control room tasks* (NUREG/CR-3309)". U. S.Nuclear Regulatory Commission, 1983, Washington.
[20] D. I. Gertman & H. S. Blackman "*Human reliability & safety analysis data handbook"* John Wiley & Sons Inc., 1994, New York.
[21] A. D. Swain and H. E. Guttman.. "*Handbook of human reliability analysis with emphasis on nuclearpower plant applications (*NUREG CR-1278)*".* NRC, 1983, Washington.
[22] J. C.Williams. *"Human reliability data – The state of the art and the possibilities"*. Proceedings Reliability, Vol. 1, June 14-16, (1989).
[23] F.G. Jentsch." *Verification and Validation of Complex Systems: Human Factors Issues*". Springer, 1993, Berlin.