CHEMICAL ENGINEERING TRANSACTIONS

VOL. 77, 2019

A publication of AIDIC

The Italian Association
of Chemical Engineering
Online at www.cetjournal.it

Guest Editors: Genserik Reniers, Bruno Fabiano
Copyright © 2019, AIDIC Servizi S.r.l.
ISBN 978-88-95608-74-7; **ISSN** 2283-9216

DOI: 10.3303/CET1977127

How Big Data & Analytics Can Improve Process and Plant Safety and Become an Indispensable Tool for Risk Management

Pankaj Goel[a,b], Hans Pasman[a]*, Aniruddha Datta[b]

[a] Mary Kay O'Connor Process Safety Center, Artie McFerrin Department of Chemical Engineering,
[b] Department of Electrical and Computer Engineering Texas A&M University, College Station, Texas 77843-3122
hjpasman@gmail.com

With the advances in digitization, Information Technology (IT), and connected devices, data are becoming plentiful. And with the past 30 years of developments of Artificial Intelligence tools leading to great enhancements in dealing with various levels and types of uncertainty, much has become tangible, where in the past it used to remain vague and fuzzy. Tools like neural networks can distil information from datasets, while probabilistic methods can characterize randomness. Bayesian causation networks enable finding critical pathways and help to design and monitor effective safeguards, while Petri nets enable analysis of time-critical events. Interval analysis, Dempster-Shafer theory, and fuzzy logic can assist in delimiting uncertainty in measurement results and expert judgment. System dynamics modeling and Functional resonance analysis may unravel interactively degrading processes. All this can improve understanding about communication lines and mechanisms of interactions within a plant socio-technical system, and the influences on achievement and performance. This will result in reformed work processes, manufacturing conditions and help in identifying abnormal trends. Therefore, while planning and prediction are based on observational evidence and trends, the new technologies will be a strong support for management, in recognizing and evaluating risks, including safety risks. Although applications of big data and analytics are still young, nevertheless in process control and reliability prediction of equipment a few achievements have already been demonstrated. However, much more is possible. For example, in the case of process safety performance indicators, lagging indicators are usually available but the techniques may stimulate the recording of the more important leading indicators for the prediction of safety and culture trend in a company in relation to its economic health. The paper will present more details on the methods and an example of dynamic risk mapping.

1. Introduction

With the advances in technology over past four decades, process plants use different control systems such as Programmable Logic Controllers (PLC), Distributed Control System (DCS), and Supervisory Control and Data Acquisition systems (SCADA) for monitoring and controlling the plant operations (Goel et al., 2017b). At the same time with development in IT, communication methods and connected devices, process plants are producing incredible amounts of data in different forms stored in 'data lakes' (data warehouses). This requires new and innovative approaches and methods to create Business Intelligence and actionable insights. The industry can get significant benefits with the use of intelligent systems and big data analytics methods. Several attributes such as volume, variety, velocity, value, veracity, variability, and valence characterize Big Data (Goel et al, 2017a). Figure 1a shows different data collected during process plant operations. Static data means data or reports generated over a period and remains fixed for a considerable amount of time while dynamic data means data, which changes with time and are continuous. Structured data refers to data in a table or specific report formats, while unstructured refers to data primarily expressed as text. The collected data is usually the raw data and requires pre-processing, cleaning and analysis to derive the expected information for decision making. Figure 1b highlights the various data analysis types such as descriptive,

Paper Received: 7 December 2018; Revised: 18 May 2019; Accepted: 23 June 2019

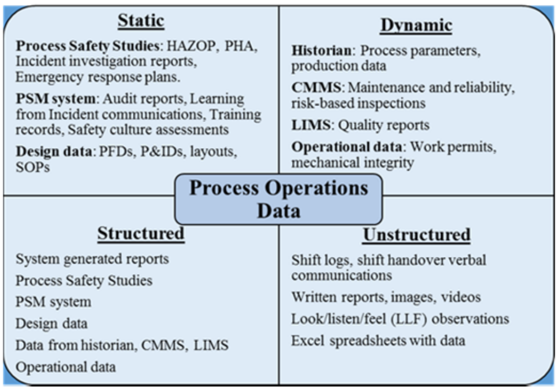
Please cite this article as: Goel P., Pasman H., Datta A., 2019, How Big Data & Analytics can improve process and plant safety and become an indispensable tool for risk management, Chemical Engineering Transactions, 77, 757-762 DOI:10.3303/CET1977127

diagnostic, predictive and prescriptive and relation between analysis, decision making and human input. Human input is highest at descriptive analysis level, and lowest during the prescriptive analysis.

2. Data processing methods (analytics)

There are many methods available by which based on information in the form of data containing uncertainty a prediction can be made for a situation. Apart from conventional statistics, in a sequence of increasing uncertainty handling and decreasing quantitative character, we shall briefly consider here Bayesian causation networks (BN), Interval analysis, Dempster-Shafer theory, fuzzy logic, and functional resonance analysis.

In case a cause-effect structure can be derived and cause probabilities, e.g., on failures in discrete or continuous form are given, a Bayesian network (BN) of nodes representing the variables and directed edges the causal links is the most obvious choice. For practitioners, Fenton and Neil (2012) give a clear and applications-oriented description of BN; Kjaerulff and Madsen (2008) delve deeper. Both prediction of an effect probability and inference of a most probable cause based on observations is possible. Drawback is that feedback loops are not allowed. The power of the Bayes approach is that prior information can be updated with new evidence to a posterior distribution.



Static

Process Safety Studies: HAZOP, PHA, Incident investigation reports, Emergency response plans.

PSM system: Audit reports, Learning from Incident communications, Training records, Safety culture assessments

Design data: PFDs, P&IDs, layouts, SOPs

Dynamic

Historian: Process parameters, production data

CMMS: Maintenance and reliability, risk-based inspections

LIMS: Quality reports

Operational data: Work permits, mechanical integrity

Process Operations Data

Structured

System generated reports
Process Safety Studies
PSM system
Design data
Data from historian, CMMS, LIMS
Operational data

Unstructured

Shift logs, shift handover verbal communications
Written reports, images, videos
Look/listen/feel (LLF) observations
Excel spreadsheets with data

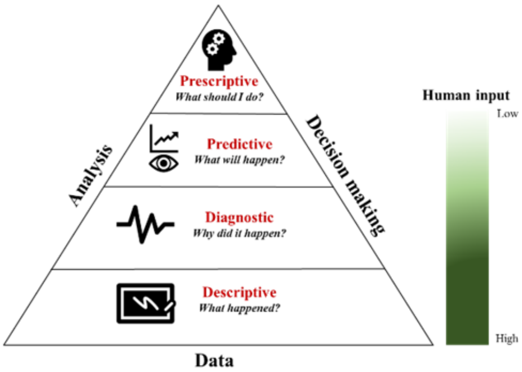


Figure 1a: Process operations data (Goel et al., 2017a)   *Figure 1b: Data analysis types and actions (PFDs: Process Flow Diagram; P&IDs: Piping & Instrumentation Diagram; SOPs: Standard Operating Procedure; CMMS: Centralized Maintenance Management System; LIMS: Laboratory Information Management System)*

Interval analysis is applied when the value bounds are certain but information within that range is vague. Alefeld and Mayer (2000) treat the interval arithmetic. In risk assessment it is used to express and handle imprecise probability, typical for epistemic uncertainty (lack of knowledge). Combination with random (aleatory) information can be shown in a p-box plot, e.g., as a fuzzy cumulative normal distribution with mean and/or standard deviation within bounds, see also Choudhary et al. (2016). Expert probability estimates form a category of imprecise probability. The Dempster-Shafer (DS) belief approach (Shafer 1976, Dempster 2008) is applicable. The subjective expert answers for the probability of event occurrence and non-occurrence need not to sum to 1 as probability theory requires. Therefore, one distinguishes in the interval [0, 1] three sections separated by belief and plausibility. The interval to belief is often taken about what the expert is at least sure of, and that to plausibility as highest estimate bound; the third part represents ignorance. The expert is assigned a reliability value. To combine the answers of more than one expert with different reliability the DS rule has been developed. Sentz and Ferson (2002) tied the DS approach and that of p-box plots together. Certa et al. (2017) applied the DS rule to combine expert risk estimates as part of failure mode effects and criticality analysis (FMECA).

Zadeh (1965) designed fuzzy sets and logic to deal with vague information. Among many others, Wierman (2010) described the approach. There are many applications. If a variable value or object description cannot be characterized sharply, one can indicate a center and left and right extremes (fuzzification), though. From the extremes to the center-point membership increases, hence from 0 at the extremes to 1 at the center; linear or curved membership functions can be defined. If variables have to be combined the logic in the form of IF-THEN-ELSE rules applies. Executing the logic is called inference. Results can be defuzzified either to a centroid of a fuzzy set (Mamdani model) or a constant or function (Sugeno model). The approach is heavily applied in control systems because in complex systems fuzzy approach outweighs precision. More recently type-2 fuzzy set has evolved and in particular the interval version of it. If experts first independently of each other grade a value linguistically (e.g., high, medium, low) or as index, then independently indicate on a continuous scale an interval for the grade, mathematical treatment as developed by Liu and Mendel (2008) merges the information to an objective result, facilitating decision-making.

Last but not least is the Functional Resonance Analysis Method (FRAM) developed by Hollnagel (2017). It serves to analyze variability in a socio-technical system (STS) and to determine causal structures for scenarios. An STS is a hierarchical organizational structure of layers connected by communication lines, controlling a technical process. Lack of transparency in such complex system blurs causation. FRAM nodes describe each a system function and are modeled as hexagons with at the vertices contacts for Input (I); Output (O); Resources (R) consumed; Constraints/controls (C); Preconditions (P); and Time (T). By not detailing the process inside a node but connecting appropriate vertices of different functions FRAM supports causation thinking.

3. Applications

One approach that would benefit from developments sketched in the Introduction is the Process Resilience Analysis Framework (PRAF) developed by Jain et al. (2017). This method of ultimate resort includes an integrated method using process plant data, simulation and optimization approach to find the operating region bounds in which a plant can operate efficiently and safely (Jain et al., 2018a). This approach relies on integration of technical and social factors in the process plant under study (Jain et al., 2018b and c), and it assumes reliable dynamic risk assessment for decision making. However now, with the availability of data streams and analytical methods dynamic and operational risk management can be made much more effective. In the following sections we shall give an example.

3.1 Dynamic Risk mapping

Facilities have various subsystems and/or components that have complex interactions, which result in changing operations environment. This affects the risk profile of the facilities and hence it is important to study the emergent behavior of these interactions within the complex systems. So far, the body of literature that is concerned with dynamic risk profiles due to emergent behavior of complex process systems using big data analytics is small. In this section, a systematic methodology is described and developed. For this purpose, the process unit system is reproduced as a system of layers as illustrated in Figure 2. Based on this system of layers, a dynamic risk profile is obtained by the incorporation of the wealth of data generated in the facility from various sources such as historic information, Centralized Maintenance Management System (CMMS), operational data, and Process Safety Management (PSM) system in the form of indicators (Jain et al., 2018b). With the real plant data, the risk could be assessed also applying contributions from safety culture survey data, audit reports and more.

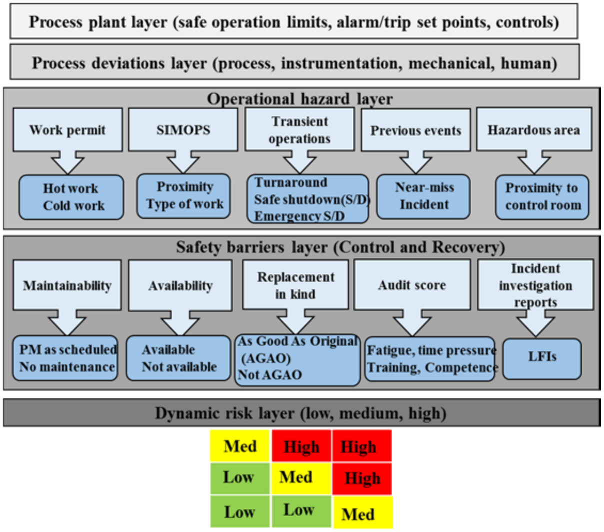

Process plant layer (safe operation limits, alarm/trip set points, controls)
Process deviations layer (process, instrumentation, mechanical, human)
Operational hazard layer
Work permit
SIMOPS
Transient operations
Previous events
Hazardous area
Hot work Cold work
Proximity Type of work
Turnaround Safe shutdown(S/D) Emergency S/D
Near-miss Incident
Proximity to control room
Safety barriers layer (Control and Recovery)
Maintainability
Availability
Replacement in kind
Audit score
Incident investigation reports
PM as scheduled No maintenance
Available Not available
As Good As Original (AGAO) Not AGAO
Fatigue, time pressure Training, Competence
LFIs
Dynamic risk layer (low, medium, high)
Med High High
Low Med High
Low Low Med

Figure 2: Dynamic risk mapping layers; the blue boxes will receive the data streams for the parameters determining the risk (PM: Predictive Maintenance; LFIs: Learning From Incidents), from Goel et al. (2017a)

The dynamic risk evaluation involves different steps similar to a Layer of Protection Analysis (LOPA) study. As illustrated in Figure 4, the following is a step-wise methodology that involves layer-wise analysis from plant layer to safeguards layer to calculate the final risk as low, medium or high as indicated in the matrix of Figure 2.

Step 1 Scenario identification: To define a scenario in details applying basic fault and event tree. Fault tree analysis helps to identify the initiating and basic events leading to the top event. Event tree analysis supports the identification of safety barriers in place to prevent and mitigate the consequence. F_1 (see Figure 4, layer 2) is evaluated from the scenario analysis in the form of initiating scenario probability leading to a risk of major consequence. Depending on the scenario, this follows different combinations of AND/OR gate calculations.

Step 2 Plant operations assessment: This step deals with identification of the dynamic factors based on the operational hazard layer. These could be from issued work permits, ongoing SIMOPS (simultaneous operations), transient operations, previous events, and hazardous area classification. Outcome of this step is Operations Hazards Factor F_2 acting as an additional factor leading to increased event probability: in the conventional case it is not considered, in the dynamic case $F2 \leq 1$. Contributions to F_2 by various operational activities are time-averaged, composed as AND gates, while the smaller the value the larger the effect.

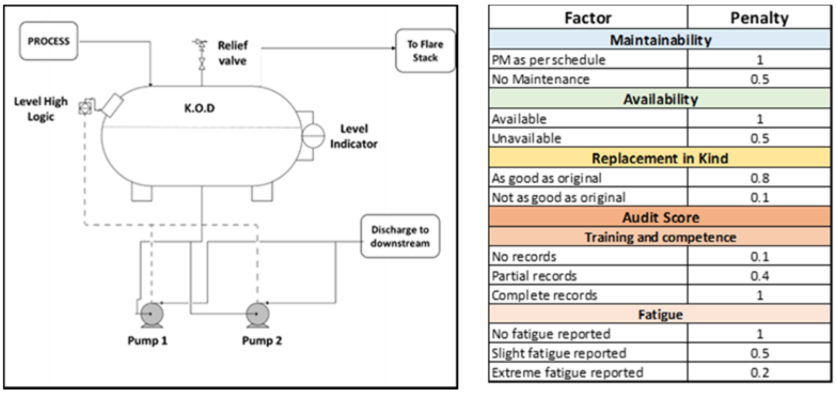
Step 3 Barrier health assessment: This step is a combination of identifying the existing control and recovery barriers available for the scenario and assessment of their health, based on the conditions of items from the safety barriers layer. Here, F_3 is evaluated after dividing the probability of failure on demand (PFD) of each protection layer, assumed independent of the others (IPLs), by a corresponding penalty factor. The penalty factors are determined based on indicators of maintainability, availability, replacement and audit (see the right side of Figure 3). F_3 is derived from the product of penalty factors adapted PFDs (LOPA approach).

Step 4 Calculation: The final step is to calculate the risk of a major consequence occurring from the collected operations data. The proposed method follows LOPA approach, incorporating additional factors based on the data from dynamic operations. Equation (1) is used to calculate risk of a major consequence as shown below:

$$R = F_1 * F_3/F_2 \tag{1}$$

3.2 Example case study

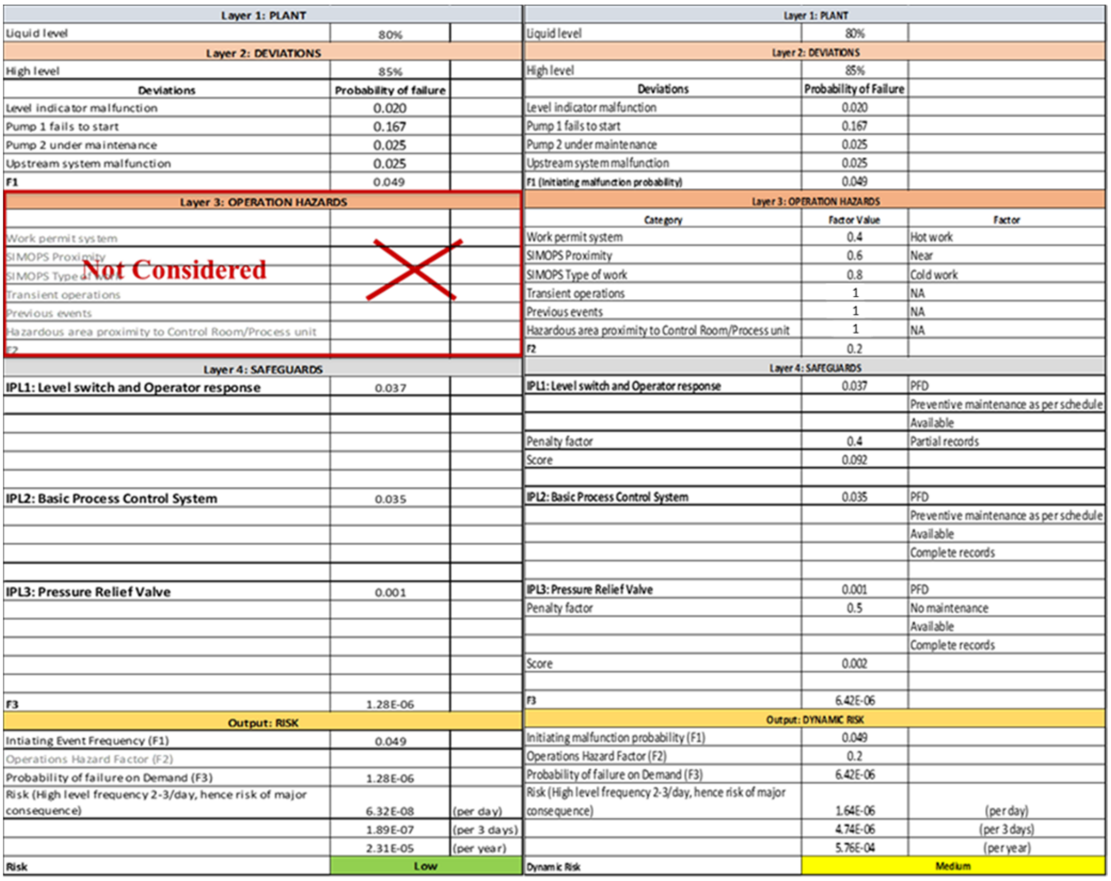
An accident scenario is considered to analyze and map the dynamic risk profile. This type of dynamic risk profile analysis would support more informed operational decisions, improved maintenance plans, work execution strategies, and overall safer and more reliable operations. The way data mining is performed is as follows: At any moment in time discrete parameter values (true [1] or false [0]) will be read by the risk calculation module at a suitable time frame sequence. Beside the parameter values inputs to the risk calculation module are user defined weights for the fourth layer parameters expressing the degree of effectiveness of the relevant parameter.



Factor	Penalty
Maintainability	
PM as per schedule	1
No Maintenance	0.5
Availability	
Available	1
Unavailable	0.5
Replacement in Kind	
As good as original	0.8
Not as good as original	0.1
Audit Score	
Training and competence	
No records	0.1
Partial records	0.4
Complete records	1
Fatigue	
No fatigue reported	1
Slight fatigue reported	0.5
Extreme fatigue reported	0.2

Figure 3: Left: K.O. drum with piping (Talebberrouane et al., 2016); Right: Relevant penalty scores

The example scenario (Figure 3) concerns a knock out drum (K.O.D.) which includes a level switch and a level transmitter indicator. During the normal operation the process stream is captured in the K.O.D. The liquid from the process stream is discharged with the help of pumps as soon as the level reaches a set point measured by the level switch. High level occurs 2 to 3 times per day. If at high level increase continues, a hazard situation of a major risk event is due to liquid discharge to the flare stack causing liquid-carryover and spreading of fire or even explosion. For the purpose of the study, we assume that level indication may malfunction, that of the two pumps in the process stream one is under maintenance and the other may fail to start, and that the upstream process may be under upset condition (isolation valve fails). In this case, the following three barriers are available: Barrier 1: High level switch and Basic Process Control System (BPCS) cutting off flow to K.O.D. with an operator response; Barrier 2: Operator checking that BPCS is working; and Barrier 3: Pressure Relief Valve connected to the vent line. In conventional risk assessment analysts do not explicitly consider various variables related to human and organizational factors, nor do they consider changes in the conditions and in input data. The latter, such as component failure data may have been determined over the years in the plant but are often estimates based on information from elsewhere. The effect of correlation and dependencies are usually ignored. Even for this simple scenario variants are imaginable, which may worsen the situation, such as the sticking of the pressure relief valve. Anyhow, for this simplified example, in the static QRA maintenance influences, which can appear as issuing work permits, and nearby maintenance SIMOPS, which can be a threat to the plant, or other events are not considered (F_2 is 1 in layer 3 of the left table of Figure 4). Hence, due to ignoring operational hazards and health or robustness of barriers the calculated risk seems *Low* (rounded value 2.10^{-5}/yr).


Not Considered

Layer 1: PLANT		
Liquid level	80%	
Layer 2: DEVIATIONS		
High level	85%	
Deviations	**Probability of failure**	
Level indicator malfunction	0.020	
Pump 1 fails to start	0.167	
Pump 2 under maintenance	0.025	
Upstream system malfunction	0.025	
F1	0.049	
Layer 3: OPERATION HAZARDS		
Work permit system		
SIMOPS Proximity		
SIMOPS Type of work		
Transient operations		
Previous events		
Hazardous area proximity to Control Room/Process unit		
F2		
Layer 4: SAFEGUARDS		
IPL1: Level switch and Operator response	0.037	
IPL2: Basic Process Control System	0.035	
IPL3: Pressure Relief Valve	0.001	
F3	1.28E-06	
Output: RISK		
Intiating Event Frequency (F1)	0.049	
Operations Hazard Factor (F2)		
Probability of failure on Demand (F3)	1.28E-06	
Risk (High level frequency 2-3/day, hence risk of major consequence)	6.32E-08	(per day)
	1.89E-07	(per 3 days)
	2.31E-05	(per year)
Risk	Low	

Layer 1: PLANT		
Liquid level	80%	
Layer 2: DEVIATIONS		
High level	85%	
Deviations	**Probability of Failure**	
Level indicator malfunction	0.020	
Pump 1 fails to start	0.167	
Pump 2 under maintenance	0.025	
Upstream system malfunction	0.025	
F1 (Initiating malfunction probability)	0.049	
Layer 3: OPERATION HAZARDS		
Category	**Factor Value**	**Factor**
Work permit system	0.4	Hot work
SIMOPS Proximity	0.6	Near
SIMOPS Type of work	0.8	Cold work
Transient operations	1	NA
Previous events	1	NA
Hazardous area proximity to Control Room/Process unit	1	NA
F2	0.2	
Layer 4: SAFEGUARDS		
IPL1: Level switch and Operator response	0.037	PFD
		Preventive maintenance as per schedule
		Available
Penalty factor	0.4	Partial records
Score	0.092	
IPL2: Basic Process Control System	0.035	PFD
		Preventive maintenance as per schedule
		Available
		Complete records
IPL3: Pressure Relief Valve	0.001	PFD
Penalty factor	0.5	No maintenance
		Available
		Complete records
Score	0.002	
F3	6.42E-06	
Output: DYNAMIC RISK		
Initiating malfunction probability (F1)	0.049	
Operations Hazard Factor (F2)	0.2	
Probability of failure on Demand (F3)	6.42E-06	
Risk (High level frequency 2-3/day, hence risk of major consequence)	1.64E-06	(per day)
	4.74E-06	(per 3 days)
	5.76E-04	(per year)
Dynamic Risk	Medium	

Figure 4: Left: Conventional risk analysis approach result; Right: Dynamic risk mapping result (NA is not active).

However, the dynamic risk mapping approach developed in this study is using data from the plant informing us on various parameters for the operations (layer 3), such as whether hot work occurs. Also, results of health of barriers by maintenance inspection and testing results (layer 4) can be monitored. If needed this can be followed by repair, or *e.g.* replacement with a similar instrument, hence confirming availability or not. In case activities are on, hazard values for different operations are assigned based on experience, for example, a value of 0.4 for hot work. For these values expert estimates can be used applying methods described in Section 2. In the course of time updates may be established. This way we get a different value of risk depending upon the actual daily operations in the plant. In this example scenario the value of risk at a certain time and given conditions is calculated to be *Medium* (rounded value 6.10^{-4}/yr). Hence, we can see that with the help of dynamic risk mapping by considering more realistic scenarios and failure values we have a very different and more realistic value of risk. This risk value is not constant and may change depending on various key scenarios during the plant operations. In reality even more factors can be taken into account. Data on the reliability of safety critical components as the pressure relief valve or the level switch can be collected and by making use of Bayesian update the values over time made more realistic. The limited volume of this paper does not enable a more detailed description of how a risk module for this purpose is built, but it is obvious that for each HAZID based scenario a cause-effect structure, basically a bowtie, will be built in the form of a Bayesian or Petri network. For the future, additional indicators can be included, such as pipe vibration, sudden gas concentration, or power usage. As Albalawi et al. (2017) contend, it is even thinkable to pick-up from a safety-Lyapunov-based economic mode predictive controller a signal of a large process disturbance. And if an event occurs such risk module system may help to detect quicker where the cause of a disturbance may be found. The system should be able to handle risk during transient exposed people concentrations and turn-around operations as the latter are prone to accident. With Big Data & Analytics a real-time risk dashboard comes under reach.

4. Conclusions

In this study, the authors established a systemic methodology to determine the dynamic risk profile of process plants. An easy, friendly, and excel-based prototype user interface has been developed for this methodology. This approach utilizes data that is already recorded in the process plant system and can provide real time risk profiles. The developed method was demonstrated using an example case study of high liquid level scenario in a K.O.D and comparing it to the conventional method of risk analysis.

Acknowledgments

We thank late Dr M. Sam Mannan for his endeavour and stimuli to advance process safety and risk assessment research for a safer world and to encourage this work.

References

Albalawi, F., H. Durand, P.G. Christofides. 2017, Distributed Economic Model Predictive Control for Operational Safety of Nonlinear Processes, AIChE Journal, 63 (8), 3404-3418.

Alefeld, G. and Mayer G., 2000, Interval analysis: theory and applications, Journal of computational and applied mathematics, 121 (1-2), 421-464.

Certa, A., Hopps F., Inghilleri R. and La Fata C.M., 2017, A Dempster-Shafer Theory-based approach to the Failure Mode, Effects and Criticality Analysis (FMECA) under epistemic uncertainty: application to the propulsion system of a fishing vessel, Reliability Engineering and System Safety, 159, 69-79.

Choudhary, A., Voyles I.T., Roy C.J., Oberkampf W.L. and Patil M., 2016, Probability Bounds Analysis Applied to the Sandia Verification and Validation Challenge Problem, ASME Journal of Verification, Validation and Uncertainty Quantification, 1 (1), 011003 1-13.

Dempster, A.P., 2008, Upper and lower probabilities induced by a multivalued mapping. Classic Works of the Dempster-Shafer Theory of Belief Functions, Springer, 57-72.

Fenton, N. and Neil M., 2012, Risk assessment and decision analysis with Bayesian networks, CRC Press.

Goel, P., Datta A. and Mannan M.S., 2017a, Application of big data analytics in process safety and risk management. Big Data (Big Data), 2017 IEEE International Conference on, IEEE.

Goel, P., Datta A. and Mannan M.S., 2017b, Industrial alarm systems: Challenges and opportunities, Journal of Loss Prevention in the Process Industries, 50, 23-36.

Hollnagel, E., 2017, FRAM: the functional resonance analysis method: modelling complex socio-technical systems, CRC Press.

Jain, P., Pasman H.J., Waldram S.P., Rogers W.J. and Mannan M.S., 2017, Did we learn about risk control since Seveso? Yes, we surely did, but is it enough? An historical brief and problem analysis, Journal of Loss Prevention in the Process Industries, 49, 5-17.

Jain, P., Pasman H.J., Waldram S., Pistikopoulos E. and. Mannan M.S, 2018a, Process Resilience Analysis Framework (PRAF): A systems approach for improved risk and safety management, Journal of Loss Prevention in the Process Industries, 5, 61-73.

Jain, P., Mentzer R. and Mannan M.S., 2018b, Resilience metrics for improved process-risk decision making: survey, analysis and application, Safety science, 108, 13-28.

Jain, P., Rogers W.J., Pasman H.J. and Mannan M.S., 2018c, A resilience-based integrated process systems hazard analysis (RIPSHA) approach: Part II management system layer, Process Safety and Environmental Protection, 118, 115-124.

Kjaerulff, U.B. and Madsen A.L., 2008, Bayesian networks and influence diagrams, Springer Science+ Business Media, 200, 114.

Liu, F. and Mendel J.M., 2008, Encoding words into interval type-2 fuzzy sets using an interval approach, IEEE transactions on fuzzy systems, 16 (6), 1503-1521.

Sentz, K. and Ferson S., 2002, Combination of evidence in Dempster-Shafer theory, Citeseer.

Shafer, G.,1976, A mathematical theory of evidence, Princeton University press.

Talebberrouane, M., Khan F. and Lounis Z., 2016, Availability analysis of safety critical systems using advanced fault tree and stochastic Petri net formalisms, J. Loss Prevention Process Industries, 44, 193-203.

Wierman, M.J., 2010, An introduction to the mathematics of uncertainty, Creighton University, 149-150.

Zadeh, L.A., 1965, Information and control, Fuzzy sets, 8 (3), 338-353.