

Improving Process Safety and Product Quality using Large Databases

Ankur Pariyani,^a Warren Seider,^a Ulku Oktem,^b Masoud Soroush^c

^a*Department of Chemical and Biomolecular Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA; pariyani@seas.upenn.edu, seider@seas.upenn.edu*

^b*Risk Management and Decision Processes Center, Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA; oktem@wharton.upenn.edu*

^c*Department of Chemical and Biological Engineering, Drexel University, Philadelphia, PA 19104, USA; masoud.soroush@drexel.edu*

Abstract

This paper introduces a novel modeling and statistical framework (based on Bayesian theory) that utilizes extensive distributed control system and emergency shutdown databases, to perform thorough risk and vulnerability assessment of chemical/petrochemical plants. Quality variables are utilized, in addition to safety (or process) variables, to enhance both process *safety* and product *quality*. To effectively achieve these objectives, new concepts of *abnormal events* and *upset states* are defined, which permit the identification of near-miss events from the databases. The databases for a fluid catalytic cracking unit at a major petroleum refinery are used to demonstrate the application and performance of the techniques introduced herein. The results show that with the novel utilization of near-miss data, one can perform robust risk calculations using both product-quality and safety data.

Keywords: Process safety, product quality, risk assessment, Bayesian theory, chemical process industries.

1. Introduction

In the chemical process industries (CPIs), the extent of human and financial losses due to incidents is staggering – the U.S. Chemical Safety and Hazard Investigation Board website [1] lists about 65 serious accidents that occurred over the past decade, with their consequences and key technical findings. Based on the severity levels, incidents can be broadly classified as *accidents* or *near-miss consequences*. Though accidents have low probabilities of occurrence, they have high severities, often accompanied by on-site and/or off-site major impacts. Near-miss consequences, on the other hand, have much higher probabilities of occurrence, but have limited, or sometimes no, impact. Recent studies have demonstrated the importance of identifying near-miss consequences to predict the probability of accidents [2, 3], and reporting, monitoring, and investigating them to reduce the potential of accidents [4-6].

Because the CPIs have been adapting plants with minimal design changes to produce higher-quality products at higher production rates, the rate of reporting of near-miss consequences has increased in recent years, with more companies seeking to improve their reporting and investigation of incidents [7, 8]. Several industry-standard software packages, which perform quantitative risk assessment using incident (or consequence) databases, are widely used. In addition, several papers and books ([2, 3, 9-11]) have

proposed the analysis of incident databases using fault-trees, hazard and operability (HAZOP) studies, failure mode and effects analysis (FMEA) and Bayesian theory, among other approaches, to gain predictive insights concerning incidents. However, because most chemical processes have hundreds of variables to monitor their dynamics (especially large-scale continuous processes), in our views, a significant amount of *precursor* information, on unsafe conditions, is overlooked and un-utilized as it is immersed in their large dynamic databases. The dynamic data associated with their distributed control systems (DCSs) and emergency shutdown (ESD) systems, which help human operators assess and control plant performance, especially in the face of potential *safety* and *quality* problems, contain real-time information on the progression of disturbances (or special-causes) and the performance of their *safety and/or quality systems* (barriers to protect the process from abnormal behavior). To our knowledge, existing techniques do not adequately utilize this information. In this paper, new methods are presented that utilize this information to assess the risk levels and predict the probability of the occurrence of incidents.

Furthermore, in this paper, quality variables are utilized, in addition to safety (or process) variables, to enhance both plant *safety* and product *quality*. This approach improves the methods for integrating safety and quality management systems introduced over the past two decades [12, 13].

2. Model Development

Modern chemical processing units are equipped with distributed control systems (DCSs) and emergency shutdown (ESD) systems to assure safe operation and high product-quality performance. The DCSs involve controller elements distributed throughout the units, with central servers that issue controlling actions. This allows the operators (human + machine) to control the variables well within their defined operating envelopes to optimize the profitability, safety, quality, and flexibility of the units. Depending upon the type of measurements, variables are defined as *process* or *quality variables*. The former are online measurements that track the dynamics of the process (for example, temperatures, pressures, flow rates and their rates of change, etc.) whereas, the latter are often estimated measurements (using mechanistic and/or statistical models) related to the *quality* of the products (for example, viscosity, density, average molecular weight, etc.).

Based on their sensitivity and importance, variables are classified into two categories: primary and secondary variables. Primary, or key, variables are most crucial for the *safety* of the process and are associated with the ESD system. Whenever these variables move beyond their ESD limits, emergency shutdowns or ‘trips’ are triggered shortly after a small time-delay. Secondary variables, on the other hand, are not associated with the ESD. For large-scale processes, typically 150-300 operating variables are monitored; however, only a small percentage (<10%) are associated with the ESD.

The control chart of any primary variable is divided into four zones, beginning with its green-belt zone (normal operation), during which the process variable lies within acceptable limits. When the variable moves beyond these limits, into its yellow-belt zones, high/low alarms are triggered. When it moves beyond the limits of its yellow-belt zones, into its orange-belt zones, high-high/low-low alarms are triggered. The border between its orange- and red-belt zones is the threshold limit for the triggering of

the ESD system. *Abnormal events* begin when process (or quality) variables move from their green-belt zones to their yellow-, orange-, or red-belt zones, triggering alarms. Clearly, these departures can be interpreted as precursors to undesirable consequences or accidents, when safety and/or quality systems fail to maintain normal operation. Consequently, in this paper, abnormal events, for variables that return to their green-belt zones, are recognized as *near-miss events*, which could have propagated to incidents. As a result, vast amounts of near-miss information become available for dynamic risk assessment.

Depending upon their criticality, abnormal events are classified in three categories: *least-critical abnormal events* that cross the high/low alarm thresholds, but do not cross the high-high/low-low alarm thresholds; *moderately-critical abnormal events* that cross the high-high/low-low alarm thresholds, but do not cross the ESD thresholds; and *most-critical abnormal events* that cross the ESD thresholds. Secondary variables don't have red-belt zones, and consequently, most-critical abnormal events cannot occur.

2.1. Dynamic Databases: DCS and ESD Logs

The use of incident databases only by the industry-standard software packages restricts their abilities to achieve a high degree of predictive accuracy of the frequencies and consequences of incidents. The current lack of dynamic analyses to identify and target near-miss events has contributed to many serious accidents over the last decade [1]. Had there been systematic procedures for analyzing dynamic data identifying near-miss events and the performance of their safety and/or quality systems, it seems clear that a large fraction of these accidents (and shutdowns) would have been avoided by alerting the plant-management well in advance.

This work relies primarily on the efficient extraction of knowledge from *dynamic databases*, namely DCS and ESD databases. These databases are inherently heterogeneous and describe the state of a plant instantaneously or over a period of time. To our knowledge, the utilization of dynamic databases has not previously been achieved as prior analyses [2, 3, 9-11] have focused on the usage of incident (consequence) databases. Typically, the DCS database contains abnormal event data; that is, alarm identity tags for the variables, alarm types (low, high, high-high, etc.), times at which the variables cross the alarm thresholds (in both directions), variable priorities, etc. Its associated ESD database, of greater consequence, contains trip event data, timer-alert data, etc. As discussed earlier, the framework utilizes the DCS and ESD databases for the FCCU at a major petroleum refinery over an extended period. The DCS databases are vast, with 5,000-10,000 alarm entries recorded every day, equaling 500-1,000 abnormal events.

2.2. Safety and/or Quality Systems

A safety and quality management structure of any process handles abnormal events with various *safety and/or quality systems (SQSs)*, which are components of the DCS, the human operators, and the ESD system. Five SQSs are defined, as follows: a) the basic process control system (BPCS) – SQS¹, b) the operator (machine + human) corrective actions, Level I – SQS², c) the operator (machine + human) corrective actions, Level II – SQS³, d) the override controller – SQS⁴, and e) the emergency shutdown (ESD) system – SQS⁵. They are usually activated sequentially and the corrective actions become more rigorous down the sequence. The actions of the automatic safety and/or

quality systems influence the actions of human operators and this causative relationship is modeled using *copulas* (discussed next).

2.3. Industrial FCCU as Case Study

The FCCU studied herein converts low-value, heavy oil into lighter and more valuable products and processes more than 250,000 barrels of oil per day. In summary, for the study period, a total of 2,036 abnormal events occurred for the primary variables (1,527 associated with process variables, 509 with quality variables). Variables, equipped with all 5 safety and/or quality systems, experienced 1,857 abnormal events (1,720 least-critical, 21 moderately-critical and 116 most-critical abnormal events, with 2 leading to 2 ESDs). Those variables not equipped with high-high/low-low alarms and an override controller, experienced 179 abnormal events (176 least-critical and 3 most-critical leading to 3 ESDs).

Using the event-tree approach presented in [2], Table 1 is obtained, which is used as likelihood data for a Bayesian simulation of the entire study period.

Table 1. Failure (F) and Success (S) Counts for Safety and/or Quality Systems

I	II		III		IV		V	
Abnormal events	S	F	S	F	S	F	S	F
2,036	1,896	140	21	116	114	2	5	0

3. Bayesian Simulation

In our previous work, Bayesian analysis was used for dynamic risk analysis [2, 14, 15]. Assumed prior distributions were updated dynamically with data (incident consequences), to yield mean failure and incident probabilities. Instead, herein, the complete posterior distributions are computed. Furthermore, the failure counts of safety and quality system 1 are modeled using a Poisson distribution (likelihood) and the failure rate (θ_1) using a Gamma distribution. The failure probabilities of the other four safety systems ($\theta_j, j = 2, \dots, 5$) are modeled using Beta distributions and their failure and success counts are modeled using Binomial distributions (likelihood).

The joint failure probability distribution of the failure rate and probabilities of safety and/or quality systems is given as:

$$p(\theta_1, \theta_2, \dots, \theta_{N_s}, \mathbf{W} | Data) \propto \underbrace{(\theta_1)^{n_1} e^{-\theta_1} \prod_{j=2}^{N_s} (\theta_j)^{K_j} (1-\theta_j)^{L_j}}_{\text{Likelihood}} \times \underbrace{p(\theta_1, \theta_2, \dots, \theta_{N_s} | \mathbf{W})}_{\text{Prior}} \times p(\mathbf{W})$$

where, \mathbf{W} is the Spearman rank correlation matrix between the failure rate and probabilities of safety and/or quality systems, and N_s ($=5$) is the number of safety and/or quality systems. The copulas (c), which are multivariate functions, are used to model the joint probability distribution of the random variables, as a function of one-dimensional, marginal, cumulative distribution functions and their correlations [16, 17]. Primarily, they are tools for modeling the dependences of the random variables. Therefore, the joint prior distribution of the failure rate and probabilities, conditional on the correlation matrix is given by:

$$p(\theta_1, \theta_2, \dots, \theta_{N_s} | \mathbf{W}) \propto c(F_1(\theta_1), F_2(\theta_2), \dots, F_{N_s}(\theta_{N_s})) \times \prod_{k=1}^{N_s} p(\theta_k)$$

The simulation results (posterior distributions of failure rate/probabilities and their correlation matrix) are obtained using the random-walk, multiple-block, Metropolis-Hastings Algorithm [18] – with Gamma and Beta distributions as *proposal* distributions for the failure probabilities of the safety and/or quality systems, and the Inverse-Wishart distribution for the correlation matrix, and their means at the previous iteration values (random-walk distributions). The Cuadras and Augé copula [19] is used to model the dependences among the failure probabilities.

3.1. Results and Key Findings

The histograms of the marginal posterior failure rate and probabilities of safety and/or quality systems (θ) associated with the primary variables for the studied unit, based on the likelihood data in Table 1 and calculated using Cuadras and Augé copula, are presented in Figure 1.

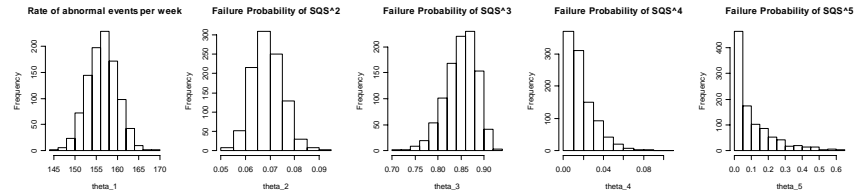


Figure 1. Marginal posterior distributions for θ using Cuadras and Augé copula

Next, the mean failure rate/probabilities are presented in Table 2 when the quality data in the DCS and ESD databases are not included in the Bayesian analysis. These are compared with the results when both the safety and quality data are included.

Table 2. Comparison of the mean failure probabilities and probabilities of occurrence of incidents – with and without quality data

<i>Mean</i>	θ_1 (Rate/week)	θ_2	θ_3	θ_4	θ_5	<i>Prob. of ESD*</i>	<i>Prob. of Accident*</i>
Safety data only	119.17	0.085	0.869	0.025	0.14	1.58E-3	2.58E-4
Safety and quality data	156.36	0.068	0.838	0.024	0.145	1.1E-3	1.9E-4

* Given an abnormal event

Clearly, Table 2 shows that the exclusion of quality data results in the overestimation of the failure probabilities of the safety/quality systems, θ_2 , θ_3 , θ_4 , and θ_5 , the probabilities of an emergency shutdown, and the occurrence of an accident. The quality data complement the safety data to yield more near-miss information, as well as more realistic and reliable results.

Figure 1 presents the performances of the SQSs and the probabilities of the occurrence of shutdowns and accidents for the study period and the associated uncertainties. As expected, when data points are limited, the variance increases. The main advantage of

using a copula model is that it allows the sharing of information (through correlations), embedded in the dataset. These and other results show that the means of the posterior distributions for safety systems I-III, calculated using different copulas, do not differ significantly due to the availability of many data points. However, for safety systems IV and V, with limited data points, the choice of copula, and consequently, the amount of information transmitted, has a significant impact. This is being examined closely in our current research. In addition, this framework calculates finite posterior distributions of the correlations between the failure rate/probabilities and demonstrates that the safety and quality systems are well correlated in practice.

4. Concluding Remarks

This paper presents an overview of new techniques to: (i) utilize the vast dynamic databases recorded in the CPIs for dynamic risk analysis, and (ii) assess and enhance process safety and product quality in a unified way. To effectively achieve these objectives, new concepts of *abnormal events* and *upset states* are introduced, which permit the identification of near-miss events from the databases. A combined modeling and statistical framework (based on Bayesian theory) is developed to obtain thorough and robust risk estimates (with associated uncertainties) using the DCS and ESD system databases. As the failure probabilities of SQSs increase, the recognition of near-misses increases – alerting operators and management to consider corrective actions; e.g., improved (1) DCS configurations and tuning, (2) operator training, (3) operating regimes, (4) process designs, and (5) alarm system configurations. The technique presented herein improves upon existing techniques by accounting for the near-misses experienced by individual variables, thereby leading to improved risk estimates. More specifics are provided in recent papers submitted for publication.

5. References

- [1] U.S. Chemical Safety and Hazard Investigation Board, <http://www.csb.gov/>.
- [2] A. Meel and W.D. Seider, Chem. Eng. Sci., 61 (2006) 7036.
- [3] W. Yi and V.M. Bier, Management Science, 44 (1998) S257.
- [4] J.R. Phimister, U. Oktem, P.R. Kleindorfer, and H. Kunreuther, Risk Anal., 23 (2003) 445.
- [5] I. Rosenthal, P.R. Kleindorfer and M.P. Elliott, Proc. Safety Prog., 25 (2006) 135.
- [6] S. Jones, C. Kirchsteiger, and W. Bjerke, J. of Loss Prev. in the Process Ind., 12 (1999) 59.
- [7] UNEP/ILO/WHO International Programme on Chemical Safety, <http://www.who.int/ipcs/emergencies/identifying/en/index.html>.
- [8] K. Rasmussen, The experience with Major Accident Reporting System from 1984 to 1993. European Commission, Joint Research Center, EUR 16341 EN, 1996.
- [9] J.M. Santamará Ramiro and P.A. Braña Aísa, Risk Analysis and Reduction in the Chemical Process Industry, Blackie Academic and Professional, New York, 1998.
- [10] J. Steinbach, Safety Assessment for Chemical Processes, Wiley-VCH, Weinheim, 1999.
- [11] L.M. Morrison, J. Hazard. Mater., 111 (2004), 161.
- [12] S.G. Herrero, M.A.M. Saldana, M.A.M. Del Campo, D.O. Ritzel, J. Saf. Res., 33 (2002) 1.
- [13] M.M. Williamsen, Prof. Safety, 50(2005), 41.
- [14] A. Meel, L.M. O'Neill, W.D. Seider, U. Oktem, and N. Keren, J. of Loss Preven. in Proc. Ind., 20 (2007) 113.
- [15] A. Meel and W.D. Seider, Comput. Chem. Eng., 32 (2008), 827.
- [16] R.B. Nelsen, An Introduction to Copulas, Lecture Notes in Statistics, Springer, 1999.
- [17] P. Embrechts, F. Lindskog, and A. McNeil, Handbook of Heavy Tailed Distributions in Finance, Elsevier, 2003, 331-385.
- [18] A. Gelman, J.B. Carlin, H.S. Stern and D.B. Rubin (Second Edition), Bayesian Data Analysis, Chapman & Hall/CRC, 2004.
- [19] C.M. Cuadras, and J. Augé, Commun. in Statistics-Theory and Methods, A10(1981), 339.