Available online at www.sciencedirect.com
ScienceDirect
Journal of Loss Prevention in the Process Industries 20 (2007) 113–127
Journal of Loss Prevention in the process industries
www.elsevier.com/locate/jlp

Operational risk assessment of chemical industries by exploiting accident databases

A. Meel[a], L.M. O'Neill[a], J.H. Levin[a], W.D. Seider[a,*], U. Oktem[b], N. Keren[c]

[a]*Department of Chemical and Biomolecular Engineering, University of Pennsylvania, Philadelphia, PA 19104-6393, USA*
[b]*Risk Management and Decision Processes Center, Wharton School, University of Pennsylvania, Philadelphia, PA 19104-6340, USA*
[c]*Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, IA 50011-3130, USA*

Received 11 June 2006; received in revised form 17 October 2006; accepted 18 October 2006

Abstract

Accident databases (NRC, RMP, and others) contain records of incidents (e.g., releases and spills) that have occurred in the USA chemical plants during recent years. For various chemical industries, [Kleindorfer, P. R., Belke, J. C., Elliott, M. R., Lee, K., Lowe, R. A., & Feldman, H. I. (2003). Accident epidemiology and the US chemical industry: Accident history and worst-case data from RMP*Info. *Risk Analysis*, *23*(5), 865–881.] summarize the accident frequencies and severities in the RMP*Info database. Also, [Anand, S., Keren, N., Tretter, M. J., Wang, Y., O'Connor, T. M., & Mannan, M. S. (2006). Harnessing data mining to explore incident databases. *Journal of Hazardous Material*, *130*, 33–41.] use data mining to analyze the NRC database for Harris County, Texas.

Classical statistical approaches are ineffective for low frequency, high consequence events because of their rarity. Given this information limitation, this paper uses Bayesian theory to forecast incident frequencies, their relevant causes, equipment involved, and their consequences, in specific chemical plants. Systematic analyses of the databases also help to avoid future accidents, thereby reducing the risk.

More specifically, this paper presents dynamic analyses of incidents in the NRC database. The NRC database is exploited to model the rate of occurrence of incidents in various chemical and petrochemical companies using Bayesian theory. Probability density distributions are formulated for their causes (e.g., equipment failures, operator errors, etc.), and associated equipment items utilized within a particular industry. Bayesian techniques provide posterior estimates of the cause and equipment-failure probabilities. Cross-validation techniques are used for checking the modeling, validation, and prediction accuracies. Differences in the plant- and chemical-specific predictions with the overall predictions are demonstrated. Furthermore, extreme value theory is used for consequence modeling of rare events by formulating distributions for events over a threshold value. Finally, the fast-Fourier transform is used to estimate the capital at risk within an industry utilizing the *frequency* and *loss-severity* distributions.

Keywords: Risk; Frequency modeling; Consequence modeling; Abnormal events; Chemical plants

Abbreviations: Companies A, B, C, D, E, F, G—A, B, C, D, E, F, G; Basic indicator approach, BIA; Capital at risk, CaR; Center for Chemical Process Safety (AIChE), CCPS; Equipment failure, EF; Environmental Protection Agency, EPA; Extreme value theory, EVT; Fast-Fourier transform, FFT; Heat transfer units, HT; Identically and independently distributed, iid; Inverse fast-Fourier transform, IFFT; Internal measurement approach, IMA; Loss distribution approach, LDA; Markov-chain Monte Carlo, MCMC; Major Accident Reporting System, MARS; National Response Center, NRC; Others, O; Operator error, OE; Occupational Safety and Health Administration, OSHA; Process Safety Incident Database, PSID; Process safety management, PSM; Process units, PU; Process vessels, PV; Quantile-quantile, Q-Q; Risk Management Plan, RMP; Standardized approach, SA; Storage vessel, SV; Transfer line, TL

*Corresponding author. Tel.: +1 215 898 7953.
E-mail address: seider@seas.upenn.edu (W.D. Seider).

0950-4230/$ - see front matter
doi:10.1016/j.jlp.2006.10.003

Nomenclature

a,b	parameters of *Beta* prior probability distribution
a_i, b_i	parameters of prior probability distribution of cause i for an incident
d_1, d_2, d_3	cumulative number of incidents of causes EF, OE, and O at the end of each year
e_i	probability of involvement of equipment type i
$E(\mu\|Data)$	expected posterior mean of μ
$E(q\|Data)$	expected posterior mean of q
$E(y)$	expected value of number of incidents in a year
$E[y_i\|y_{-i}]$	expected value of prediction of incident in year i based on incidents in y_{-i}
$f(e_i)$	prior probability distribution of involvement of equipment i for an incident
$f(x_i\|Data)$	posterior probability distribution of involvement of equipment i conditional upon *Data*
$f(x_i)$	prior probability distribution of cause i for an incident
$f(x_i\|Data)$	posterior probability distribution of cause i conditional upon *Data*
f_l	discrete *loss-severity* distribution function
$f_z(Z)$	discrete probability distribution function of total loss
$F_u(y)$	cumulative probability distribution for distribution of losses, l, over threshold u
$G(l)$	*Generalized Pareto* distribution of losses
l	loss associated with an incident
$M_i+N_i+O_i$	cumulative number of incidents associated with equipment i at the end of each year
n_p	number of points desired in *total loss* distribution
$N_{\mathrm{C/P}}$	number of incidents associated with compressors and pumps
N_{d}	amount of damage, \$
N_{e}	number of evacuations
N_{EF}	number of incidents associated with equipment failures
N_{f}	number of fatalities
N_{h}	number of hospitalizations
N_{HT}	number of incidents associated with heat-transfer equipment items
N_{i}	number of injuries
N_{OE}	number of incidents associated with operator errors
N_{PU}	number of incidents associated with process units
N_{SV}	number of incidents associated with storage vessels
N_{t}	number of years
N_{TL}	number of incidents associated with transfer-line equipment
N_{total}	total number of incidents
N_{U}	number of incidents associated with unknown causes
$p(\lambda)$	prior distribution of λ
$p(\lambda\|Data)$	posterior distribution of λ given *Data*
$p(q\|Data)$	marginal posterior distribution of q given *Data*
$p(\mu\|Data)$	marginal posterior distribution of μ given *Data*
P_N	probability generating function of the frequency of events, N
p_i, q_i	parameters of prior probability distribution of involvement of equipment i in an incident
q	parameter of the *Negative Binomial* distribution
s	total number of incidents in N_{t} years
u	threshold value of l for *loss-severity* distribution
$V(y)$	variance of number of incidents per year
w_{d}	dimensionless damage measure
w_{e}	dollar amount per evacuation, \$
w_{f}	dollar amount per fatality, \$
w_{h}	dollar amount per hospitalization, \$
w_{i}	dollar amount per injury, \$
x_1, x_2, x_3	probabilities of causes EF, OE, and O for an incident
y_i	number of incidents in year i
z_i	predictive score for incidents in year i
Z	total annual loss for a company

Greek

α, β	parameters for *Gamma* density distribution function
$Beta(a, b)$	*Beta* density distribution with parameters a and b
ϕ_l	characteristic function of the *loss-severity* distribution
ϕ_Z	characteristic function of *total loss* distribution
λ	average annual number of incidents
λ_{B}	average annual number of incidents for company B with losses greater than u
λ_{F}	average annual number of incidents for company F with losses greater than u
μ	parameter of the *Negative Binomial* distribution
ξ, β	parameters of the *generalized Pareto* distribution
$Gamma(\alpha, \beta)$	*Gamma* distribution with parameters α and β

Subscript

i	year counter
n	year vector