

Catastrophic oil spill analysis

Z.-G. Ji, W. Johnson & G. Wikel
Bureau of Ocean Energy Management, USA

Abstract

The Deepwater Horizon oil spill (DWH) in 2010 in the Gulf of Mexico is the largest accidental marine oil spill in the history of the petroleum industry. After DWH, a key question was asked: what is the likelihood that a similar catastrophic spill will happen again? The extreme value theory (EVT) has been widely used in studying rare events, including hurricane damage, stock market crash, insurance claim, flooding, earthquake, etc. In this paper, the EVT is applied to analyse oil spills in the U.S. outer continental shelf (OCS). Based on the 49 years (1964–2012) of OCS oil spill data, the EVT is capable of describing the oil spills well. The return period of a catastrophic oil spill in the OCS areas is estimated to be 165 years, with a 95% confidence interval between 41 years and more than 500 years. Findings in this study are very useful to oil spill risk assessment, contingency planning, and environmental impact statement on oil exploration, development, and production.

Keywords: catastrophic oil spill, Deepwater Horizon, extreme value theory, risk analysis.

1 Introduction

The Deepwater Horizon oil spill (DWH) began on April 20, 2010 in the Gulf of Mexico on the BP-operated Macondo Prospect. The explosion and sinking of the Deepwater Horizon oil rig killed 11 workers. The exploration well flowed for 87 days, until it was capped on July 15, 2010. The total discharge is estimated at 4.9 million barrels (bbl) [1]. DWH is the largest accidental marine oil spill in the history of the petroleum industry.

After DWH, a key question was asked: what is the likelihood that a similar catastrophic spill will happen again? An answer to this question is essential to oil spill risk assessment, contingency planning, and environmental impact statement on oil exploration, development, and production. In recent years, statistical



theory of extreme values, especially the extreme value theory (EVT), have been widely used in studying rare events, including hurricane damage, stock market crash, insurance claim, flooding, earthquake, sea level rise, etc. [2–4]. In this paper, the EVT is applied to analyze oil spills in the U.S. outer continental shelf (OCS) in the past 49 years (1964–2012). The objective is to estimate the likelihood of catastrophic oil spills in the OCS areas.

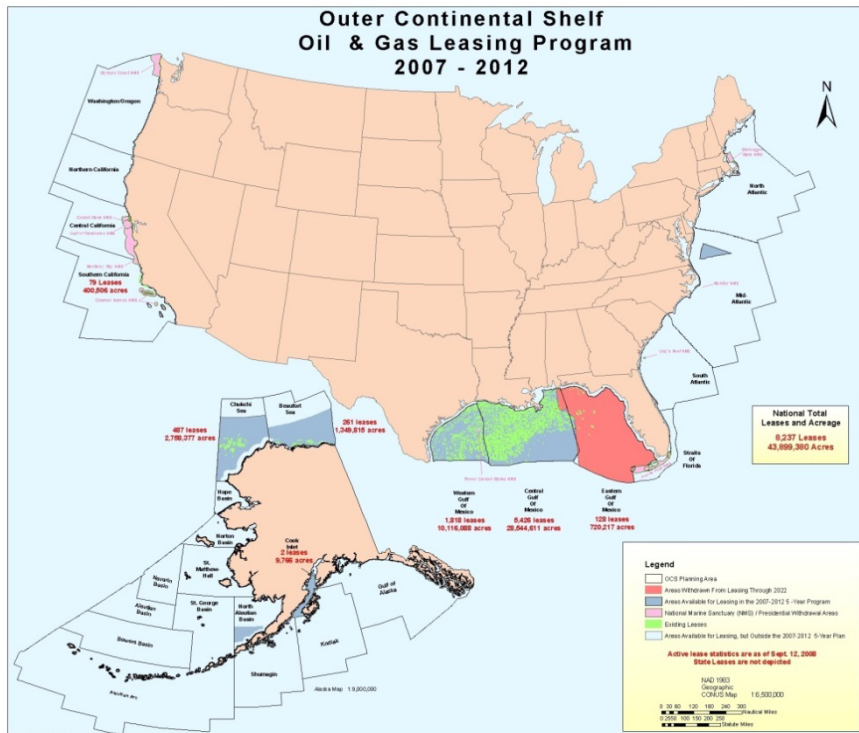


Figure 1: Outer Continental Shelf (OCS) planning areas.

The term “Outer Continental Shelf” is a legal term created by Federal statute and is distinct from the geographic term “continental shelf”. Legally, the OCS comprises that part of the submerged lands, subsoil, and seabed, lying between the seaward extent of the States’ jurisdiction and the seaward extent of Federal jurisdiction, which is generally 3 geographic miles seaward from the State’s coast line to about 200 to 300 nautical miles offshore. The majority of OCS activity occurs in the Gulf of Mexico, the Pacific Ocean, and off the coast of Alaska (Figure 1). The western and central portions of the northern Gulf of Mexico (GOM) make up one of the world’s major oil and gas producing areas, and have been a steady and reliable source of crude oil and natural gas for more than 50 years. The need to balance the value of these resources against the potential for environmental damage is an important concern [5].

EVT has gone a long way since its beginnings with Fisher and Tippet [6]. EVT provides a solid mathematical basis and no other reliable alternative has been suggested. It provides models relevant for the assessment of rare events, even outside the range of previous observations. Extreme value methodology is now commonly used in a wide range of fields, such as economic damage [7, 8], finance [2, 3], earth sciences [4, 9, 10], traffic prediction [11], large oil spills [12], etc. This paper focuses on how the statistical theory of extreme values can be applied to improve the oil spill risk analysis. A statistical modeling approach based on the generalized extreme value distributions is advocated for assessing the risk of catastrophic oil spills.

2 Mathematical approach

The Oil Spill and Risk Analysis (OSRA) model [13, 14] is used to estimate the probability of oil spills in a specific lease area, to calculate oil spill trajectories from selected spill launch points (i.e., places that a spill is assumed to occur), and to determine the probability that an environmental resource or coastline segment might be affected by oil released from the selected launch point. One part of the OSRA model addresses the probability of oil-spill occurrence. The probability is based on the past oil spill incidents in the proposed leasing areas, and the past oil spill probability is then projected into the future spill occurrence rate according to the estimated oil production rate and transportation scenarios. To date, the information from OSRA results is often being used in three different types of documents: (1) environmental assessment and environmental impact statement reports on OCS oil exploration, development, and production; (2) oil spill response plans submitted by industry; and (3) environmental reports submitted by industry.

A generalized extreme value (GEV) family of functions describes the distribution of the block maximum of a set of data, which are a series of independently and identically distributed (shortly i.i.d.) observations [15]. The cumulative distribution function (CDF) of the GEV is given by:

$$F(x; \mu, \sigma, \gamma) = \begin{cases} \exp\{-[1 + \gamma(x - \mu)/\sigma]^{-1/\gamma}\}, & 1 + \gamma(x - \mu)/\sigma > 0, \gamma \neq 0 \\ \exp\{-\exp[-(x - \mu)/\sigma]\}, & \gamma = 0 \end{cases} \quad (1)$$

The GEV distribution has three parameters: 1) the location parameter (μ) specifies the center of the distribution; 2) the scale parameter (σ) determines the size of deviations about the location parameter; and 3) the shape parameter (γ) governs how rapidly the upper tail decays. These three parameters can be estimated by fitting the GEV distribution to data. In this study, Eq. (1) is used to estimate the largest annual oil spill in the OCS areas. The behavior of the annual maximum is then approximately described by one of the GEV family of distributions given by the CDF $F(x; \mu, \sigma, \gamma)$.

In recent years, the development of software for statistical extremes has been rapid, particularly in the open source environment [16]. The statistical software language and environment called R [17] has been widely used

(<http://www.r-project.org/>). It is open-source and freely available without proprietary licensing requirements. Further, R is the software that contains extensive utilities for modeling extreme values. In this study, the R statistical package was utilized to estimate location, scale, and shape parameters using the ML method in fitting the annual maxima of oil spills.

3 Model results

Figure 2 shows the time series of number of large oil spills (more than 1,000 bbl) derived from the OCS data for 49 years from 1964 to 2012. In this study, the GEV distribution of Eq. (1) is fitted by the ML method [15] directly to the annual maxima of the OCS oil spills from 1964 to 2012. The R software is used to carry out the statistical analysis without any manual adjustment to the parameters. The resulting parameters are that the location parameter is 2.442 (which is equivalent to 277 bbl); the scale parameter is 0.839; and the shape parameter is -0.073. In order to assess the model fit graphically, a probability–probability plot (PP plot) (Figure 3) and a quantile–quantile plot (QQ plot) (Figure 4) are used for the transformed data set.

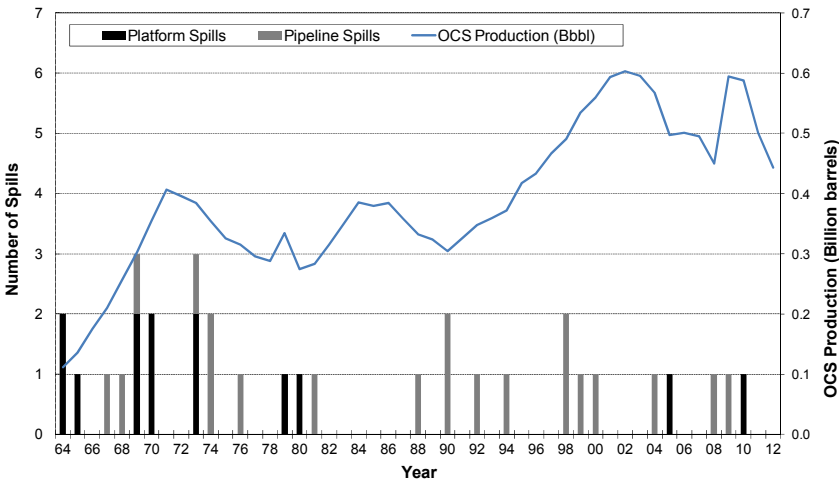


Figure 2: Large OCS spills ($\geq 1,000$ bbl) (1964–2012).

Figure 3 assesses the distribution fit by a PP plot for the estimated GEV parameters. In Figure 3, the empirical distribution of oil spill probability based on the historical annual maxima data is plotted against the modeled GEV distribution, in which the line of equality indicates perfect fit. Figure 3 presents the probability (or percentile) of the empirical distribution for each quantile value, against the modeled probability of the GEV for the same value. For example, the value of 1456 bbl is the 50th percentile of the empirical distribution

and the 48th percentile (approximately) of the GEV. A “perfect” fit would result in the line of equality. How close the PP plot is to this line is a measure of goodness of fit.

Figure 4 presents the distribution fit by a QQ plot for the estimated GEV parameters. In Figure 4, the quantiles of the modeled GEV distribution are plotted against the empirical quantiles based on the historical data. Again, the line of equality indicates perfect fit. The QQ plot shows the historical data values against the equivalent percentiles for the modeled GEV. For example, there are 49 annual maxima in the OCS oil spill dataset. The value of 1456 bbl is equivalent to the median (50th percentile) in the empirical distribution. This value is plotted against the 50th percentile of the modeled GEV distribution, which is 1374 bbl. Both Figures 3 and 4 indicate that the fits are well and follow the line of equality reasonably.

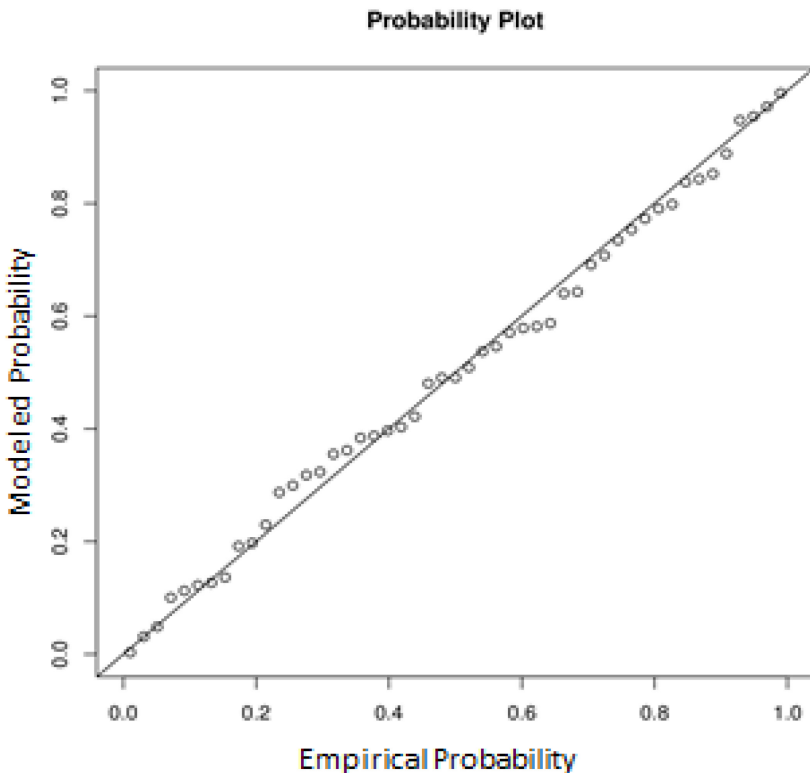


Figure 3: PP plot for fit of GEV distribution to annual oil spill maxima (line of equality indicates perfect fit).

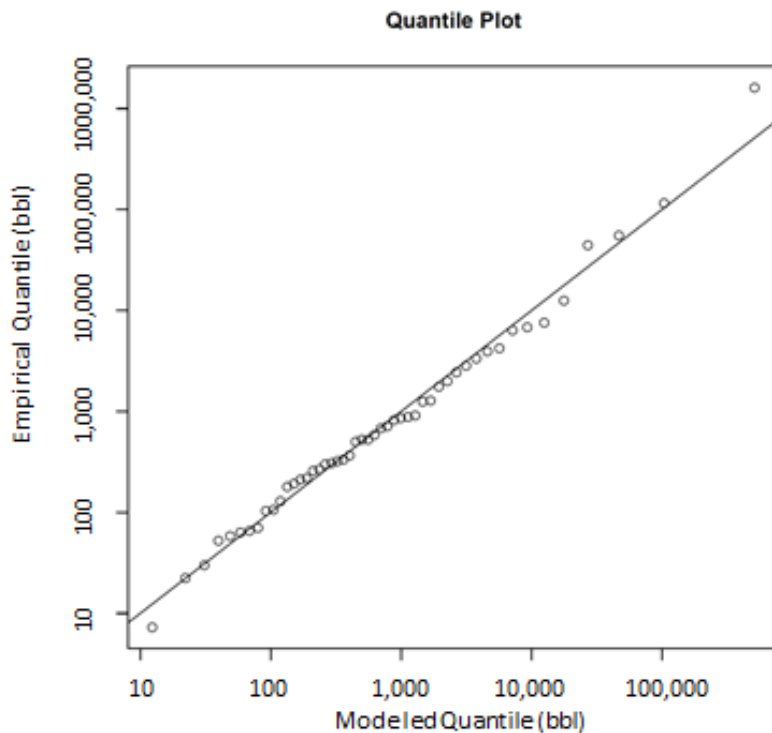


Figure 4: QQ plot for fit of GEV distribution to annual oil spill maxima (line of equality indicates perfect fit).

The return level (or return value) of a random variable is the quantile value which is exceeded, on average, once in a period of time (called return period). For example, return period (such as 100-year flood) based on extreme precipitation (i.e., certain return value) is commonly used to assess the capacity of drainage systems. Figure 5 gives return levels under different return periods. The dots are calculated based on the observations and the curve is from the GEV distribution calculated using the ML method with the R software. Figure 5 shows that the return levels for 1-year, 10-year, 100-year return periods are 277, 15,849, and 562,341 bbl, respectively. The return level of one million bbl (i.e. the catastrophic oil spill) corresponds to a return period of 165 years. Therefore, based on the past 49 years of OCS oil spill records, the GEV distribution indicates that a catastrophic oil spill in the OCS areas is most likely to occur with a period of 165 years. Figure 6 gives the density distributions from the model (solid line) and the data (dashed line). It is evident that the model matches the data results reasonably well.

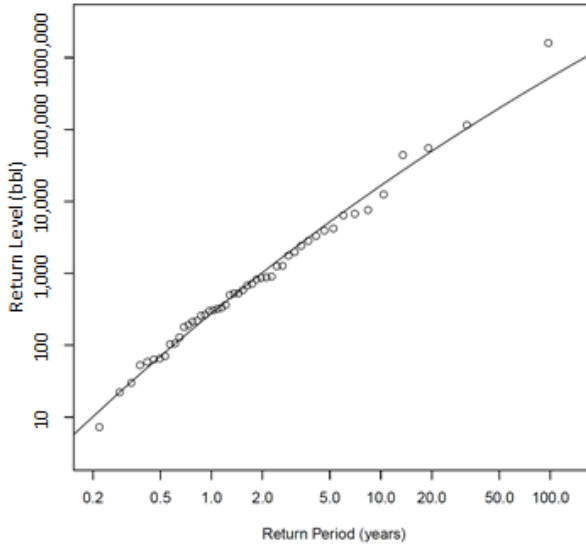


Figure 5: Return levels under different return periods. The dots are calculated based on the observations and the curve is from the GEV distribution calculated using the ML method with the R software.

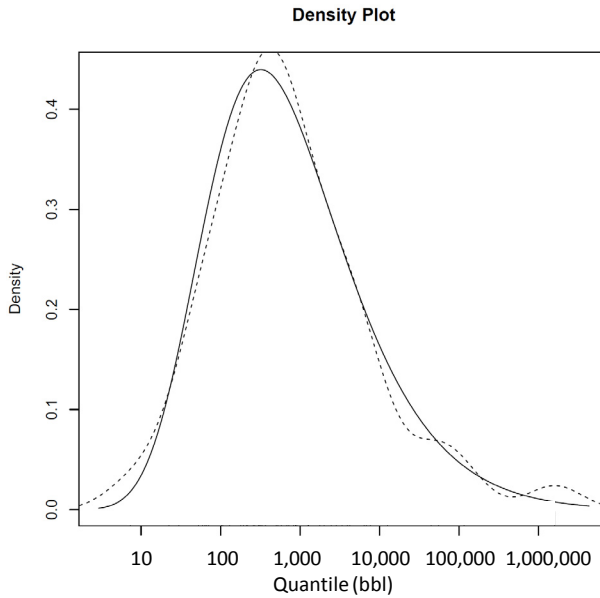


Figure 6: Density distributions from the model (solid line) and the data (dashed line).

4 Summary and discussions

This paper intended to answer this key question: what is the likelihood that a catastrophic oil spill similar to DWH will happen again? An answer to this question is essential to oil spill risk assessment, contingency planning, and environmental impact statement on oil exploration, development, and production. In the past 10 years, statistical theory of extreme values has been widely used in extreme event analyses in a variety of areas, including science, engineering, finance, and insurance. In this paper, the generalized extreme value distributions are applied to analyze oil spills in the OCS in the past 49 years (1964–2012).

The blocks method from the EVT is applied to analyze the OCS oil spill data. Based on the 49 years of OCS oil spill data, the GEV distributions are capable of describing the oil spills in the OCS areas well. The return period of a catastrophic oil spill in the OCS areas is estimated to be 165 years, with a 95% confidence interval between 41 years and more than 500 years.

Although the fundamental probabilistic theory of extreme values has been well developed for a long time, the statistical modeling of extremes remains a subject of active research. From the perspective of oil spill risk analysis, much remains to be done. Oil spill risks vary from regions to regions and are composed of many components. The impact of oil production is not included in this study. The seasonal variation of oil spill risk is another interesting and important topic. Future research challenges include the statistical modeling of complex extreme events, such as considering oil production, seasonality, and spatial dependence in the statistical modeling of catastrophic oil spill. More accurate methods in statistical analysis, which take into account multiple influencing factors, might be needed to consider these effects.

References

- [1] USCG, On Scene Coordinator Report on Deepwater Horizon Oil Spill (Report), 2011. http://www.uscg.mil/foia/docs/dwh/fosc_dwh_report.pdf.
- [2] Sanders, D.E.A., The Modelling of Extreme Events. *British Actuarial Journal*, 11, pp 519–557. doi:10.1017/S1357321700003251, 2005.
- [3] Fasen, V., Klüppelberg, C. & Menzel, A., Quantifying extreme events. *Risk-A Multidisciplinary Introduction*, 2012.
- [4] Katz, R.W., Parlange, M.B., & Naveau, P., Statistics of extremes in hydrology. *Adv Water Resour* 25:1287–1304, 2002.
- [5] Ji, Z.-G., Use of physical sciences in support of environmental management, *Environ. Manag.*, 34(2), 159–169, 2004.
- [6] Fisher, R.A. & Tippett, L.H.C., Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philos. Soc.*, 24:180–190, 1928.
- [7] McNeil, A., Estimating the tails of loss severity distributions using extreme value theory, *ASTIN Bulletin*, 27, 117–137, 1997.



- [8] Jagger, T.H., Elsner, J.B., & Saunders, M.A., Forecasting US insured hurricane losses. In: Diaz HF, Murnane RJ (eds) *Climate extremes and society*. Cambridge University Press, Cambridge, pp 189–208, 2008.
- [9] Dorland, C., Tol, R.S.J., & Palutikof, J.P., Vulnerability of the Netherlands and Northwest Europe to storm damage under climate change. *Clim Change* 43:513–535, 1999.
- [10] Katz, R.W., Statistics of extremes in climate change. *Climatic Change*, 71–76:100, 2010.
- [11] Zheng, L., Ismail, K. & Meng, X., Freeway safety estimation using extreme value theory approaches: A comparative study, *Accident Analysis & Prevention*, Volume 62, January 2014, Pages 32–41, 2014.
- [12] Eckle, P., Burgheer, P. & Michaux, E., Risk of large oil spills: a statistical analysis in the aftermath of Deepwater Horizon. *Environmental Science and Technology*, 46: 13002 – 13008, 2012.
- [13] Ji, Z.-G., Johnson, W. R., Marshall, C. F., Rainey, G. B. & Lear, E. M., Oil-spill risk analysis: Gulf of Mexico outer continental shelf (OCS) lease sales, central planning area and western planning area, 2003–2007, and Gulfwide OCS program, 2003–2042, in *OCS Report 2002–032*, Minerals Management Service, Herndon, VA, 2002.
- [14] Ji, Z.-G., Johnson, W.R. & Li, Z., Oil Spill Risk Analysis Model and Its Application to the *Deepwater Horizon* Oil Spill Using Historical Current and Wind Data, in *Monitoring and Modeling the Deepwater Horizon Oil Spill: A Record-Breaking Enterprise* (eds Y. Liu, A. Macfadyen, Z.-G. Ji and R. H. Weisberg), American Geophysical Union, Washington, D. C. doi: 10.1029/2011GM001117, 2013.
- [15] Reiss, R.-D. & Thomas, M., *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. Birkhauser, Basel, 2nd edition, 2001.
- [16] Gilleland, E., Ribatet, M. & Stephenson, A.G., A software review for extreme value analysis, *Extreme*, 16:103–119, 2013.
- [17] R Development Core Team: *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, 2012.