Comparative assessment of oil spill risks for different regions and facility types

Phase 2 Project Report

Peter Burgherr, Matteo Spada, Anna Kalinina

Technology Assessment Group
Laboratory for Energy Systems Analysis
Paul Scherrer Institut (PSI)

December 2014 (rev.&ext.)

Commissioned by BP International Ltd.

Executive Summary

Background and motivation of this study

The current study was motivated by an article published by Eckle et al. (2012a) addressing the issue if an accident such as the **Deepwater Horizon (DWH) oil spill in 2010** should have been considered by relevant stakeholders (e.g. industry, authorities) as well as the general public or not. Right after the accident occurred different opinions and statements were made including the notion that DWH has been a black swan, and thus could not have been anticipated.

This is of course also an interesting discussion within the scientific realm; however, terms like black swans can serve as metaphors to help explaining extreme events to a broader public, but an **objective, fact-based, transparent and to the extent possible quantitative risk assessment** is still required **to support decision-making and policy formulation processes** in a manner that current state-of-the-art of science is adequately represented, and can be incorporated by different stakeholders.

Comparative risk assessment in a nutshell

It is a well-known fact that sufficient, timely and uninterrupted supply of **energy is an essential prerequisite for modern human society** and its manifold complex and interconnected activities, relying on and generating a multitude of services. However, **no energy technology is absolutely risk free**, and both accidental events affecting safety and intentional attacks affecting security can occur during all stages of an energy chain (Hirschberg et al., 1998). Furthermore, various stakeholders may exhibit **different risk perspectives**, depending on their background, expectations and objectives.

Therefore, risk assessment is an important method to determine the risk given a specific setting, and not necessarily only to determine the absolute risk. For example, in a societal context, risk assessment is often based on an individually weighted combination of objective as well as subjective factors, which can be evaluated using quantitative risk metrics or are just amenable in a qualitative manner. Regardless of the perspective, applying a **consistent method of assessing risk is crucial when comparing risk among stakeholder perspectives.**

Scope, objectives and structure of the study

The Technology Assessment (TA) group of the Laboratory for Energy Systems Analysis (LEA) at the Paul Scherrer Institute (PSI) carried out the current study on oil spill risks for BP International Ltd.

This report starts with an **introduction including scope and objectives** of the study, followed by a **general overview on comparative risk assessment** of energy technologies. The **methodological part** is kept relatively compact because this part was already included in detail in the intermediate report of project phase 1. Therefore, only methods relevant for the understanding of the results presented here are included, which also avoids unnecessary repetition of previously presented topics and issues. The **results** chapter comprises five major parts, namely (1) determination of **severity threshold**, (2) **distribution fitting** for **spill frequency**, (3) distribution fitting for **spill severity**, (4) **extreme value analysis** for worst case spill scenarios, and (5) **regionalized spill risk indicators** for selected infrastructure types. While topics 1 to 4 heavily rely on **Bayesian modeling**, topic 5 addressing regionalized spill risk is closely linked to the use of a **Geographic Information System (GIS) and geo-statistical methods**. Finally, **conclusions and recommendations for future research** are provided in a concise format to summarize the essence and key findings of the performed risk assessment that relied on scientific state-of-the-art. In addition to the actual report, the study includes two more deliverables in electronic format. First, the optimized and **documented R-code used for statistical analyses** is provided, which should allow BP to directly use or adapt the methods for their own internal purposes. Second, a number of **GIS layers** were compiled to ensure a proper documentation of the geo-referenced spill data used in the geo-statistical analyses.

Methodological developments

The current study aimed to **combine classical risk assessment methods** (e.g. aggregated indicators, frequency consequence curves) **with advanced Bayesian modeling** (e.g. threshold determination, distribution fitting, extreme value analysis) **and GIS-based, geo-statistical analyses** has a large potential to describe new relationships and previously undiscovered patterns in historical accident data that could not be detected without such a **combined and multi-layer approach**.

A major methodological development undertaken within this study comprises the **treatment of extreme event**s (or worst case spill accidents). Generally, this is evaluated by fitting standard fat-tail distributions (e.g. Generalized Pareto) to the data or using Extreme Value Theory (EVT) methods such as the block maxima or peaks over threshold (POT). The outputs of these methods are often simply summarized by reporting a recurrence period and confidence interval the worst case event under consideration. The **Bayesian modeling approach** proposed and implemented in this study has several advantages, namely (1) it explicitly addresses **aleatory and epistemic uncertainty**; (2) it is **based on Bayes theorem** and thus previous knowledge can be represented in the form of the prior distribution; (3) it can be **easily updated if new information becomes available**, (4) it can be **estimated for the quantile of interest**, and (5) it is scalable, i.e. it is **applicable to global and regional data**.

Key findings and main conclusions

The **determination of a severity threshold** is a key prerequisite in comparative risk assessment. On the one hand it provides a formal procedure to test for which minimal threshold a **dataset can be considered sufficiently complete** to assume a consistent coverage of accidental events. On the other hand threshold analysis is an **important aspect in extreme value analysis** because the tail behavior is expected to exhibit different properties from the rest of the distribution. However, threshold identification is normally not a purely mathematical procedure because a variety of subjective, stakeholder-specific factors need to be taken into account as well in decision-making and policy formulation processes. In this study the lowest of the three tested thresholds (200t, 700 t, 1000 t) was chosen to define the data sets for subsequent analyses.

Different distributions were fitted to spill frequency and severity to find the best distribution model. To select the best model an information criterion, i.e. BIC, was used because they have several advantages compared to classical goodness of fit tests. The **Negative Binominal distribution (NBI) fitted best to spill frequency** data both overall and for different infrastructure types. The NBI has the advantage that it is more general than the Poisson distribution, and thus often can fit better, which is why it is rather commonly used in the analysis of accident statistics, insurance and other fields. **For spill severity, the lognormal distribution fitted best**; this is in good accordance with other studies in the literature. However, quite often it is also the case that other distributions (e.g Generalized Pareto, Weibull, Reverse Gumbel) can also describe the data sufficiently adequate, and then it is the question if one selects only one distribution using a decision criterion like the BIC or takes into account the contributions from different distributions by means of Bayesian Model Averaging (BMA).

Extreme value analysis was performed for both the global and regional levels to examine if the worst case spill events should be expected or not. **At the global level, historically observed maximum spills should be expected at the 1% quantile** for the different infrastructure types. The only exception was for pipeline spills, but this does not mean that the historically most extreme event could not happen again, but rather that it has a lower frequency than the 1% quantile. **In the case of platform spills the uncertainty interval was much larger than for any other infrastructure type**, which partially reflects the limited accident data fulfilling the severity threshold. Furthermore, the platform spill results reported here are in good accordance with other published studies (Eckle et al., 2012a; Ji et al., 2014).

Calculation of **regionalized spill risk indicators** was performed **for selected offshore and onshore infrastructure types**. In the case of offshore spills the analysis was restricted to ship spills because the data set for platform spills was too small to calculate robust regional indicators. Fore onshore spills pipelines, refineries, and storage facilities were considered.

For **offshore ship spills, normalized risk indicators for 16 distinct maritime regions** were calculated that subsequently were further aggregated in a combined risk score including accident rate, spill rate and maximum consequences. In a second step, frequency consequence (FN) curves were constructed to complement these highly aggregated results. **Overall, risk indicators indicated that the Mediterranean Sea, Black Sea, Caspian Sea, NE Atlantic, Persian Gulf and Caribbean Sea are among the regions with the highest spill risk.** Additionally, **FN-curves showed that spill risk in Baltic Sea, Gulf of Mexico and Yellow Sea can also be significant**, although the worst case consequences appear to be more limited than in other regions.

Two different sets of regionalized risk indicators were computed for onshore spills. On the one hand an evaluation of 10 onshore regions was conducted, similar to the approach for offshore spills. On the other hand four distinct country clusters were defined that were assumed to differ in their mode of operation. As for offshore spills, risk indicators were complemented by FN-curves. Furthermore, region-specific spill risk was compared against hotspots of BP activities to identify company-relevant areas of concern.

The evaluation of 10 global onshore regions indicated that "Brazil & Colombia", "Africa", "Middle East", "Russian Federation" concerning pipelines, and "USA & Mexico" in terms of accident rates had the highest spill risk. Since BP maintains significant activities in all of these regions (or at least specific parts within them), this clearly demonstrates that adequate risk assessment and management strategies are inevitable to ensure best practices and good governance. The region "Other Asia Pacific" showed also an increased spill risk, but with regard to BP activities in Indonesia and Australia, no accidental spills with very large consequences occurred.

When looking at the four country clusters, "USA" was identified as a region with high accident rates, which could reflect a different industry and regulatory approach to operational hazard. However, **this does not necessarily imply higher spill rats or more frequent extreme events**. "Russia" was also identified as an area of concern, although differences were found between infrastructure types.

In summary, this study has shown that both detailed global and regional oil spill risk indicators provide valuable insights, and thus can serve as objective inputs for risk management and prioritization, strategy development and policy formulation.

Lessons learnt and recommendations for future research

A **comprehensive and consistent collection of accident data** is of utmost importance to ensure an **objective, transparent and quantitative risk assessment** of accidental oil spills. **PSI's database ENSAD provides a strong foundation for this type of analysis**. However, it is necessary to continuously update the database both in content and scope to keep up with the changing and newly emerging needs of different stakeholders such as industry, insurance, authorities and regulators, etc. Furthermore, accident data should be complemented by additional facility-specific characteristics, and detailed background data required for indicator normalization. The current study has clearly shown that this second aspect should not be underestimated because not all data are readily and freely available.

The current study analyzed only spill data from ENSAD, but the novel **methodological developments could also be applied to different well-known spill dataset** covering different region and infrastructure types (e.g. OGP, ITOPF, IOPCF, EMSA, USGS, etc.)

The availability of a limited number of historical accident data for certain regions, infrastructure types or activities is also often a challenge. While the issue of "big data" and their adequate mining has recently achieved a lot of attention, the **analysis of "tiny data" has been largely neglected**. However, several approaches have been discussed in the scientific literature to address this problem. For example **Approximate Bayesian Computation (ABC)** provides a promising approach for fitting models that are intuitive and only require the specification of a generative model.

Future research areas for further methodological and analytical improvements include (1) **Bayesian Model Averaging (BMA)** to avoid partially subjective decisions of the best fitting distribution model; (2) joint geo-statistical analysis of **spatio-temporal patterns**; (3) calculation of **regional risk indicators at higher spatial resolution** (e.g. local level) or for individual locations; (4) **combining** results of **Life Cycle Assessment (LCA)** from normal operation **and** accidental effects addressed by **Comparative Risk Assessment (CRA)**; and (5) **scenario-based analysis to take into account potential** changes in the distribution, frequency and severity of natural hazards due to **climate change** that could particularly affect oil infrastructures, and thus should be considered in designing future strategies and infrastructure portfolios.