Analysis of PHMSA Spill Data for Pipeline Spill Risk Analysis
Philip Carpenter¹, Morgan Henrie², Yoshihiro Okamoto III, Paul Liddell³

¹ Great Sky River Enterprises LLC
² MH Consulting, Inc.
³ Alyeska Pipeline Service Company

ABSTRACT

Analysis of the operating risk due to hazardous liquid pipeline spills is a challenging task. One reason for this is the fact that the spill probability and the magnitude of the release are both difficult to quantify for any particular pipeline. One source of comprehensive information regarding spills in the United States is the Department of Transportation (DOT) Pipeline and Hazardous Materials Safety Administration (PHMSA) hazardous incident database. This database is a mandatory compendium of spill incidents dating from 1968, and relies on self-reporting on the part of pipeline operators. We use this database to calculate historical average spill incident rates for various designated spill causes, and demonstrate a methodology using a tail risk analysis of the limited release orifice information supplied by the database to calculate a more complete aperture probability distribution for those causes. The incident rates developed by this analysis are then adapted to obtain more reliable incident rates for a particular pipeline by using the PHMSA results as a prior estimate and implementing a Bayesian posterior update based on historical sparse spill information for that system. We conclude by showing how the information obtained can be used in conjunction with pipeline hydraulic, leak detection performance, logistic response, and other models to perform a probabilistic cost-based risk analysis for the pipeline system.

INTRODUCTION AND BACKGROUND

Hazardous Liquid Pipeline Spill Risk

The economic infrastructure of the entire world is heavily dependent on the existence of pipelines to transport commodities of various kinds. The commodities are required to support various energy consumption needs as well as water, chemical processing, food product, and other gas or liquid products. In 2014, an estimated 2,175,000 miles (3,500,000 km) of pipeline were installed in some 120 countries worldwide. A full 65% of this mileage is in the United States, followed by Russia (8%), and Canada (3%). [1]

Of particular interest is that portion of pipeline systems that transport hazardous liquids. Transport of hazardous liquids by pipeline in the U.S. is governed by 49 Code of Federal Regulations (CFR) Part 195 [2]. These regulations address pipeline transport of petroleum, petroleum products, anhydrous ammonia, ethanol or liquid carbon dioxide. In 2014, there were roughly 200,000 miles (321868 km) of hazardous liquid pipelines in the United States. These pipelines transported some 16.2 billion barrels (2.58 billion m³) of crude oil and petroleum products, and the quantity of transported commodity has continued to grow in the interim [3].

Since, by definition, the commodities addressed by [2] are hazardous, inadvertent spillage of commodities is a significant public concern. This is for good reason. In the United States, between 2010 and 2015, the number of pipeline spills averaged 393 per year. The average annual volumetric spillage over this period was 83,648 BBls (13,300 m³) for all U.S. pipelines, with annual costs of about $357,000,000/year [4]. This is a significant impact to both the industry and the public alike.

To that end, the combined efforts by both government and industry to mitigate and/or prevent liquid commodity spills are significant. These include regulatory enforcement activities, system design modifications, pipeline inspections of various kinds, leak detection system installations, and implementation of spill response procedures. Equipment and personnel dedicated to these efforts have increased accordingly.

It is often difficult to know with any rigor how much any particular mitigation effort will have in terms of reducing the
spill risk for a particular pipeline. For this paper, we will assume that risk consists of two parts: (1) a negative consequence of some kind coupled with (2) a probability value associated with the negative consequence. In the case of a hazardous liquid spill, the negative consequence might be expressed in physical terms such as the total discharge volume. Most of us would agree, however, that something more fundamental to the operational bottom line has greater importance. To that end, we will use the reported total spill remediation cost as the fundamental measurement of negative consequence in this paper.

The Problem of Data

A major roadblock in terms of evaluating spill risk is the lack of useful and quantifiable data. It is virtually impossible to predict the magnitude of a spill in terms of event frequency, spill volume, financial and social costs, and other factors in the absence of supporting data. It is also not possible to calculate the risk of a spill for a particular pipeline based solely on easily obtainable design and operating factors alone. A spill is generally an unanticipated incident, and because of this, it is essentially a random or stochastic event. This implies that to analyze it properly, we will need to employ (1) the language of statistical analysis combined with (2) a lot of supporting data.

The best existing compendium of spill-related statistical information in the United States is the Department of Transportation (DOT) compendium of spill incidents compiled by the Pipeline and Hazardous Materials Safety Administration (PHMSA) (available at www.phmsa.dot.gov). This is a detailed database of spill incidents dating back to as far as 1968, and it has proved to be of immense value for this analysis. That said, the data available is incomplete in many ways, and required significant processing to be of value for the purposes of this analysis.

The Complexity Problem

Another significant issue is the sheer intricacy of models designed to predict the spill risk. Let’s assume that the risk is proportional to the volume of the spilled commodity. The risk for a particular pipeline would therefore be the product of a number of important factors. Among these are:

- The probability of a leak
- The volumetric leak rate
- The time required to detect the leak
- The time required to control the leak
- The volume of the spill
- The dependence of the spill cost on the volume
- The dependence of the spill cost on other factors such as spill location

It is important to note that each one of these factors can be difficult to analyze or simulate. For example, the leak rate is a function of the leak aperture size combined with the local pipeline pressure, the commodity properties, and (if below ground) the properties of the soil outside the pipe. Modeling this is nontrivial, especially if the leak is below ground.

![Figure 1: Leak Detection for Various Detection Methods](image)

In addition, the leak may be detected through multiple means. The leak could be detected by a Computational Pipeline Model (CPM) [4], or by some other leak detection system (LDS). Alternately, the discharge might be detected by pipeline operations personnel, the pipeline controller, a member of the public, or by some other means. Figure 1 shows the statistics for detection of pipeline spills through various means over the period 2010 - 2016. [5]

Note that at about 6% of the total, CPM detected spills constitute only a tiny fraction of the total spills over this period. Even accounting for the fact that about 35% of spills did not have a method of detection identified, it is unlikely that a redistribution of the spill detection rates on a weighted basis would raise the CPM detection to much above 10%. Virtually all other spill detections fall into some sub-category of “detections by direct observation.” (It is tempting to conclude from this figure that leak detection systems do not add much value. This is not necessarily the case. We will return to this issue later in this paper.)

Based on this, we would require different predictive models to determine the detection time for the leak. Finally, the leak will continue to flow until it is brought under control, or limits out due to hydraulic constraints or the closing of isolation valves.

Finally, in addition to the proposed dependence of the spill cost on the escaped commodity volume, other factors may come into play as well. Among these could be the location of the spill, i.e., onto water or land, in a populated vs. unpopulated area, into a high consequence area (HCA), etc.

Risk-Adjusted Cost and Tail Risk

Most risk analysis approaches determine the risk-adjusted cost value $RV_E$ of an event by multiplying the probability of the event $P_E$ by the cost of the event $C_E$.
The PHMSA incident rate can be developed via a relatively straightforward analysis of the PHMSA data. As noted above, it is difficult to obtain the data needed to perform an informed spill risk analysis. In the following sections we describe the process of extracting useful information from the PHMSA database regarding spill incidence rate and leak aperture size, and then applying it to a specific pipeline.

General PHMSA Data Analysis Approach

As previously noted, a prime issue is the absolute lack of leak flow rate data for any of the reported incidents. Unfortunately, most CPM systems are characterized by performance curves that provide leak rate to detection time relationships. From this relationship one can derive an estimated spill volume as a product of time-of-detect and leak rate. Knowledge of the spill volume (which is provided by PHMSA) is not useful for this risk analysis task; however, the spill volume does contribute to the spill remediation cost.

This analysis handles the identified issues as follows:

1. The average pipeline spill volume impact is assumed to be described as a combination of (a) an incident rate that can be applied as a Poisson probability distribution on an incident/mile/year basis plus (b) a one-sided leak rate distribution (i.e., it only applies for positive leak rates) with heavy tails.
2. The PHMSA incident rate can be developed via a relatively straightforward analysis of the PHMSA data.
3. The PHMSA database provides no leak rate information. It does, however, provide discharge aperture size information for about 5% of all cases. We assume that the leak rate can be obtained if the pipeline pressure and leak aperture size are known. The distribution of orifice sizes is assumed to be probabilistically described by a one-sided, heavy-tailed distribution, such as a Pareto, lognormal or Weibull distribution.

4. The PHMSA analysis will describe the spill incidence and magnitude for an “average” pipeline, which may not be representative of a particular pipeline. For this analysis, therefore, we developed a method by which average data could be adapted to a particular pipeline. This is described below.

5. Finally, and as noted previously, it is necessary to use the obtained spill incident rate and aperture size distribution as inputs to a process that determines the spill rate, the time to detect the spill, the time to respond, and the resulting cost. These other components of a spill risk analysis are not available in the PHMSA data, and require other models. We have not spent a significant amount of effort developing these other models. To confirm the usefulness of the PHMSA results, however, we developed a very minimalistic risk model, which we describe at the end of this document.

Description of PHMSA Databases

The PHMSA hazardous liquid pipeline databases are currently provided to the public in the form of tab-delimited text files. Utilization of these files usually involves reformatting them to a comma-separated (.csv) format, suitable for processing via Microsoft Excel, for example. While this analysis was performed using Excel, it is quite practical to import the datasets into other databases such as SQL (Structured Query Language). The PHMSA databases contain specific information for all hazardous liquid pipeline operator-reported spills dating back to 1968.

The requirement for completion of these databases originates in CFR 191.5. In 1969, this particular set of regulations was modified to require pipeline operators to submit pipeline spill “incident reports” to PHMSA. Since establishment of the original reporting requirements, there have been four major and distinct periods of hazardous liquid pipeline incident reporting requirements. Each subsequent set of data correspond to regulatory reporting changes:

- 1968 to 1986
- 1986 to 2002
- 2002 to 2010
- 2010 to present (2018)

Over the years, the definition of what constitutes an incident and the volume of reportable information required on said incident has changed substantially. The original regulations, dating to October 4, 1969, required “...the reporting of any failure in a pipeline system...in which there is a release of the commodity transported resulting in any of the following [specific to hazardous liquid pipelines]:

1) Explosion or fire not intentionally set by the carrier.
2) Loss of 50 or more barrels of liquid.
3) Death of any person
4) Bodily harm to any person resulting in one or more of the following:
   a) Loss of consciousness
   b) Necessity to carry the person from the scene
   c) Necessity for medical treatment
   d) Disability which prevents the discharge of normal duties or the pursuit of normal activities beyond the day of the accident
5) Property damage of at least $1,000 to other than the carriers’ facilities, based upon actual cost or reliable estimates.”

Note that 50 BBLs is roughly 8 m³.

Data reporting demands have changed substantially in line with these periods. In particular, current regulatory reporting requirements (since 2002) modify the previous list by requiring that incident reports be filed when there is a:

1) Release of 5 gallons (versus loss of 50 or more barrels previously) if the release is:
   a) Not otherwise reportable under this section;
   b) Not one described in 195.52(a)(4) which is a release which “resulted in pollution of any stream, river, lake, reservoir, or other similar body of water that violated applicable water quality standards, caused a discoloration of the surface of the water or adjoining shoreline, or deposited a sludge or emulsion beneath the surface of the water or upon adjoining shorelines”.
   c) Confined to company property or pipeline right-of-way, and
   d) Cleaned up promptly

This regulatory change reduced the reportable volume from roughly 8 to 0.02 m³.

As regulatory requirements have changed, the numbers of reportable incidents and quantity of reporting data entry fields have significantly increased. Figure 3 summarizes the number of reported incidents between the years of 1998 and 2017. Note the dramatic increase in incidents starting in 2002. As shown, the overall yearly incident average from 1990 to 2002 jumped from an average of 185.9 incidents per year with standard deviation of 33.4 to 391 incidents/year between 2002 and 2017 with a 42.7 incident per year standard deviation. This increase in reported incidents is a result of reporting requirement changes. Predominately, the increase can be related to the
reduction in spill volume from 50 barrels to 5 gallons. Intuitively one expects more small spills to occur than large ones.

![PHMSA Pipeline Incidents: 1998-2017](image)

**Figure 3: PHMSA Pipeline Incidents 1998 - 2017**

Further, as the reporting requirements have changed, PHMSA has also modified the standard reporting form. Figure 4 provides a chronological view of the number of data fields contained in the PHMSA reporting database. From the original database formation through 2001, the number of data fields remained almost constant at the 62-64 count. As shown, the number of data fields increased by a factor of 4 in 2002, and by 2.25 (from the 2002 level) in 2010. From the first database the number of today’s data fields has increased by a factor of 9.46 (587 data fields), or roughly an order of magnitude.

In analyzing the difference between the various database sets one sees an ever-increasing refinement in details required. As an example, the pre-1986 database required the reporting entity to identify the incident location by state, county, and city. The current database reporting scheme includes changes in incident location reporting mandates to include latitude, longitude, pipeline/facility name, site milepost/valve station name, pipeline segment name, and location type.

Changes in 2010 also prescribe input as to association with HCAs, onshore/offshore spill category, type of incident – to include delineation as a rupture, corrosion, operator error, equipment failure, etc. and a greatly expanded section on remediation requirements, recent pipeline/facility testing performed, and third-party involvement. It is notable that many of the new fields are often not filled in, probably because they do not apply to most incidents.

![PHMSA Incident Database Data Field Count](image)

**Figure 4: PHMSA Incident Database Data Field Count**

Other 2010 reporting requirements include the requirement to provide estimates as to the leak/spill defect or aperture size for mechanical punctures, and similar information for ruptures, both in terms of length, width, and orientation. Note that only about 5% of all reported spills have such orifice information provided. Based on the distribution of equivalent areas for ruptures and mechanical punctures, it appears that reporting of size information for aperture size tails off for equivalent diameters (i.e., the diameter of a round hole having the same area as the PHMSA incident aperture) of about 1-inch. Presumably, the 95% of remaining spill incidents for which hole or aperture size information is not provided correspond to apertures smaller than this.

We also note that a significant fraction of spills occur via release paths corresponding to pump and valve seals, valves opened due to error, and failed connections. For purposes of this analysis, we will regard these as equivalent holes. In addition, there is a class of spill (“overfill or overflow”) that does not correspond to a hole. We will discuss these again later in this document.

Additional useful information provided by the post-2010 database includes the spill volume at the time of detection and the site pressure. This data is available for nearly all reported spills. The PHMSA database also includes the cost associated with remediation of the spill. However, the database does not require the reporting entity to provide an estimate of leak rate. The fact that this leak rate information is not included is probably due to the difficulty of estimating this information. Note, however, that most (though not all) computational-based leak detection systems utilize flow rate and not spill volume to characterize detection time.

In summary, the older PHMSA incident databases (prior to 2010) can be characterized as minimalistic. They provide basic information about where the incident occurred, commodity involved, basic location information, and minimal corrosion data. The current database provides greater incident location information granularity, prior testing, investigation efforts, and resulting spill size, location, and impacts to the environment,
HCAs, and so forth.

**2010 – Present PHMSA Spill Incident Analysis**

The key components of the PHMSA 2010 – Present database structure that are applicable to this analysis are shown in Figure 5. A small number of spills in the database are offshore incidents. Since the case pipeline for this analysis is an onshore line, these incidents were excluded from the analysis.

![Figure 5: Key PHMSA Data Analysis Components](Image)

Note that spills can be assigned by location as either **Facility** or **ROW** (Right-of-Way), and we maintained this discrimination in this work. Additional information regarding the INCIDENT_AREA_TYPE (aboveground, from a tank, transition area, belowground, etc.) was not used.

It is important to recognize that the database classifies leaks according to:

- **RELEASE_TYPE**
- **CAUSE**

The CAUSE field is further sub-categorized as one of the following:

1. Corrosion failure
   - Equipment failure
   - Excavation damage
   - Incorrect operation
   - Material failure
   - Natural force damage
   - Other outside force damage
   - Other incident cause

It is worth noting that the Other Outside Force Damage field is further sub-classified via the PHMSA CAUSE_DETAILS field. This sub-categorization was not used.

Spills are also categorized, for each CAUSE, by their manifestation via the PHMSA RELEASE_TYPE field as one of the following:

1. Leak
2. Mechanical puncture
3. Rupture
4. Overfill/overflow
5. Other

It is important to recognize that the RELEASE_TYPE category indicates how the spill manifests itself, and technically has nothing to do with the incident cause. Thus, although most corrosion CAUSE cases manifest themselves as PHMSA RELEASE_TYPE category *Leak*, there are still a number of corrosion CAUSE cases in the PHMSA database that manifest as PHMSA RELEASE_TYPE Mechanical Puncture or Rupture.

PHMSA further subcategorizes the Leak category into a number of other components (*Connection Failure*, *Crack*, *Pinhole*, etc.). These RELEASE_TYPE subcategories were not used in this analysis.

With respect to performing risk analysis, the PHMSA database offers no information as to leak rate, and only a limited amount of information regarding aperture size. Dimensional information (from which we presumably could, with additional information, calculate the leak rate) is supplied for PHMSA RELEASE_TYPE Mechanical Puncture and Rupture. Note that LEAK_TYPE Pinhole is characterized by PHMSA as having an effective diameter of 1 inch or less, but no further specific dimensional information is provided for Pinhole in the database. Dimensional data are also not supplied for any other LEAK_TYPE category. Despite the sparseness of the supplied dimensional information, this data was used to support analysis of the distribution in aperture sizes, as will be discussed later.

This top-level analysis indicates that:

1. Pipeline mileage in the U.S. increased gradually at an average rate of about 2%/year over this period.
2. There are about 160 times more events onshore than offshore. This almost certainly reflects the much greater proportion of onshore to offshore pipeline miles in the United States. Note that it was not possible to obtain supporting information regarding onshore and offshore mileage in the U.S. The fact that the offshore incidents
were a small fraction of the total, and because the pipelines for this analysis was onshore, the offshore results were discarded from further analysis.

3. The onshore average pipeline incident rate is 0.001871 incidents/mile/year (0.001163 incidents/km/year) with a standard deviation of 0.0004257 incidents/mile/year (0.000265 incidents/km/year). This rate was fairly constant on a year-to-year basis.

4. The incident rate is much higher on a per mile basis for facilities than it is for ROWs.

5. The blend of incident causes is substantially different for facilities when compared to spill causes on ROWs.

Figure 6 provides a graphical view of Table 3 yearly incident counts for all pipeline facilities and ROWs. Note that counts and rates for 2018 have been adjusted to reflect the fact that only six months of data are available for 2018. As noted above, the facilities incident counts are considerably higher than the counts obtained on ROWs. Note the relative statistical consistency from year-to-year.

Figure 6: Aggregate Pipeline Annual Facility and ROW Incident Counts

Note: 2018 rate annualized based on six months of PHMSA data.

Figure 7 below shows the year-to-year trends of facility spills classified according to the PHMSA Spill CAUSE field normalized on an incident per year per mile basis.

Note that the vast majority of incidents at facilities are caused by Equipment Failures of one kind or another, followed by Corrosion Failure and then Incorrect Operation. This probably reflects the fact that pump stations and terminals are rife with pumps, valves, tanks, and other active components. Note that Excavation Damage is the least significant cause of incidents at facilities.

Interpretation of the facility incident rates might be a bit more useful if we could interpret the per-mile rates on a per installation basis. Unfortunately, this would have required some knowledge of the miles of installed pipeline for various diameters as well as the number of facilities per mile, again segregated by pipe diameter. Unfortunately, this data was not located in the time available for this assessment.

Figure 7: Pipeline Facility Normalized Incident Rate Trends by PHMSA Cause

Note: 2018 rate annualized based on six months of PHMSA data.

Similarly, Figure 8 shows the year-to-year trends of spill incidents on aggregate U.S. ROWs based on incident cause. In this case, the leading cause was Corrosion Failure, followed by Material Failure and Equipment Failure. Unlike the situation in facilities, however (and perhaps not surprisingly), Excavation Damage is a fairly significant contributor in fourth place. Note that the decrease in rates for 2018 is believed to be an artifact of the annualization process, resulting from the fact that many pipelines file incident reports well after the incident date, and therefore not significant.

Figure 8: Pipeline ROW Normalized Incident Rate Trends by PHMSA Cause

Note: 2018 rate annualized based on six months of PHMSA data.

Based on the analysis, incident rates were obtained for all combinations of PHMSA Spill CAUSE and RELEASE_TYPE over the period 2010 to the middle of 2018. These are shown for all spills, facility-only spills and ROW-only spills in Table 1, Table 2, and Table 3 on page 20.
Aperture Analysis Overview

As noted previously, performance of a pipeline-specific risk analysis requires that we specify not only the spill/unintended release incidence rate but also determine the leak rate as well. Ultimately, the spill costs have a spill volume dependency (which we will discuss later). The spill size in turn is the product of the leak rate and the combined (1) time to detect (a function of the leak detection methodology) plus (2) the additional time required to stem the leak (a function of time to stem the leak such as pipeline shutdown, gravitational effects, pipeline logistic spill response, and so forth). We will return to these issues in the last part of the document.

Certain issues arise when attempting to extract leak rate information from the PHMSA data:

- The leak rate associated with spills is not a component of the PHMSA data sets;
- Related information, such as pressure at the spill site is included for most cases; and
- Leak orifice information (aperture length, width, and, in some cases, orientation) is included for the PHMSA Release Types Mechanical Puncture and Rupture. These cases constitute only about 5% of all spills.

The leak rate can, in principle, be calculated if the orifice size is known by using orifice and other pressure-flow relations. We’ll take this as a given. However, only 5% of our PHMSA data points actually have dimensional information. Therefore, our process must be capable of generating a distribution based on the limited data set. We use the following process to determine the aperture distribution:

1. The aperture area \( A_L \) is assumed to be distributed as one-sided probability distributions that are functions of the PHMSA CAUSE. This is necessarily true, since the leak areas can only be positive. Consequently, two-sided distributions (such as a Gaussian or Normal distribution) cannot be used.
2. The distributions must permit fat tails. This is because the hole sizes range over several orders of magnitude: the largest hole in the PHMSA data is roughly 40 ft\(^2\) (3.72 m\(^2\)) while the smallest is \( 3.5 \times 10^{-5} \) ft\(^2\) (3.25 x 10\(^{-6}\) m\(^2\)). It is highly likely that a significant proportion of the unlabeled data are smaller than this latter value.
3. Since only 5% of the data have corresponding measurements of the aperture, the process must provide an estimate of the distribution based on this limited data set.

Based on these concerns and constraints, we selected probability distributions of the Pareto, Lognormal and Weibull type for consideration in this analysis. These one-sided distributions are all capable of modeling heavy tails, and so will also generate relatively rare large spill apertures with reliable probabilities. They also will tend to cluster most leaks at the very small end of the scale, which matches not only our intuition, but the behavior of the data (i.e., only a small fraction of the leak apertures are large). The assumed distributions have their various pros and cons, many of which became clear as part of the analysis. These are discussed in the next section.

The selected distributions, their probability density functions (PDFs) and their case-specific constants required to set the probabilistic fit are described in Table 4. Note that the parameter \( x \) in this table should be taken to be the aperture area \( A_L \).

It’s worth noting that the cumulative distribution function (CDF) in Table 4 is the integration of the probability distribution function, and always takes a value running from zero (on the left side of any chart on which it is plotted) to 1.0 (on the right). Thus, the CDF(\( A_L \)) is always a measure of the fraction of leak aperture areas that are less than \( A_L \). Similarly, the Complementary Cumulative Distribution Function CCDF(\( A_L \)) is the fraction of leak aperture areas that are greater than \( A_L \).

We noted that only 5% of the total points are available to fit these distributions. If we wish to obtain leak aperture size distributions according to the PHMSA spill CAUSE, this situation may be worse, since the amount of data with attributed sizes varies by CAUSE. To address this, we performed a censored data analysis. This analysis assumes that we have a sample of the orifice sizes above a certain cutoff value and know the number of samples but not the actual measurements below the cutoff value. It is only possible to do this because all PHMSA Spill CAUSE categories include examples of RELEASE_TYPEs Mechanical Puncture and Rupture, for which aperture measurements are provided.

Thus, we implicitly assume that the probability distributions for all spill causes can be extrapolated to small aperture sizes based on the fit of the distribution to the large aperture sizes, even if the RELEASE_TYPE for the apertures without dimensional information is not Mechanical Puncture and Rupture. This is not unreasonable, since these RELEASE_TYPEs are shared among all of the PHMSA CAUSEs.

The censored data analysis utilized a combination of Python programs coupled with the PyMC3 probabilistic programming library [8]. The PyMC3 programming language allows one to perform nonlinear regressions and infer probabilistic data fits based on prior assumptions about the constants defining the function as well as the data available to support the fit. A Python script was created for each of the three distributions used for the analysis. The cutoff value for each PHMSA spill CAUSE censored data case was estimated by observing the point at which the observed data tailed off. Any data points smaller than this value were set to be censored or not observed when the runs were made.

Aperture Analysis Results

Censored leak aperture analysis was completed for all PHMSA CAUSEs, with the constants for all probability fits shown in Table 5 on page 21. Charts showing the Pareto, Lognormal and Weibull probability distributions for all spill causes are
provided by Figure 9 through Figure 18. All distributions are provided in the form of complementary cumulative distribution functions, since these functions can easily be plotted on logarithmically scaled axes, and also provide an easy way of visually assessing the quality of the fit when compared to the data.

In all cases, the amount of observed data was dwarfed by the unobserved samples. Since there was no absolute lower cutoff for the observed samples, we made a case-by-case assessment for this cutoff, dependent of the PHMSA Spill CAUSE being assessed. Any samples below these points that had observations were assigned as unobserved data points.

Figure 9: Leak Aperture Complementary CDFs: All Data

Figure 10: Leak Aperture Complementary CDFs: Equipment Failure

Figure 11: Leak Aperture Complementary CDFs: Incorrect Operation

Figure 12: Leak Aperture Complementary CDFs: Other Outside Force Damage

Figure 13: Leak Aperture Complementary CDFs: Corrosion Failure
The distributions discovered by the censored data analysis often extend to very low aperture sizes that would result in extremely small leak rates, and very small cumulative volumes when allowed to accumulate over reasonable detection times. This is illustrated by Table 6 on page 21. This table was calculated by assuming leak rates of various sizes, and then used a simple orifice model to determine the equivalent aperture size, the number of days to spill one cubic foot of commodity, as well as the spill volumes that would accrue over a 24-hour period. The assumed internal pressure was 500 pounds per square inch (PSI) (3.48 MPa) with an orifice coefficient of 0.75. For an actual buried pipeline, added soil resistance might result in smaller flow rates.

As we can see, the volume spilled over the course of 24 hours becomes insignificant as the effective aperture size becomes less than roughly $10^{-11}$ to $10^{-10}$ ft$^2$ (a hole roughly a thousandth of an inch in diameter, or about $10^{-12}$ to $10^{-11}$ m$^2$) and approaches unobservable for smaller hole sizes. It is our
assumption that more realistic and certainly more conservative risk analysis will be obtained if the leak apertures obtained from the probabilistic fits defined in Table 6 and by the constants obtained in the next section are limited to a minimum of $10^{-11}$ to $10^{-10} \text{ft}^2$ ($10^{-12}$ to $10^{-11} \text{m}^2$). Note that this minimum aperture size may be tuned to higher values when applied to the risk model to meet minimum volume levels seen in the PHMSA data.

As a final note, overflow is a special case of spill because it does not occur as the result of a pressurized release via a leak orifice. Spills from tanks usually occur as the result of a failure of control mechanisms designed to prevent overflow of tank flows, and are not usually modeled as pressure releases from an orifice. The PHMSA database provides no information that would allow us to determine anything related to a flow rate. Because of this, we recommend that such releases be modeled by determining the flow-to-tankage probability distribution for each of the tanks in the case analysis pipeline system.

## Case Study Pipeline Data Analysis

The PHMSA data analysis produces statistical parameters that apply to the aggregate of all U.S. pipeline systems. For incident rates in particular, this means that the results are averages across all pipelines. However, any particular pipeline may have an incident rate that is greater or less than the aggregate result. Some pipelines may experience high rates of corrosion as a result of age or line-specific operating approaches. Others may be buried deeply and have low rates of excavation damage, and so forth. Most operators would like to be able to adapt the analysis to their particular pipeline.

Ideally, of course, one would already have a good idea of what the incident rate might be for one’s own pipeline by examining the number of spills over the operational history. For a pipeline with a short operating history, however, the number of incidents may be insufficient to provide an incident rate with a high level of confidence.

The approach used in this analysis was to (1) collect all applicable spill results for the pipeline used in this analysis, and (2) use the results in conjunction with the PHMSA results to calculate a Bayesian posterior incident rate applicable only to the pipeline used in this analysis.

### TAPS Data Collection

TAPS spills dating from pipeline startup in 1977 to the middle of 2018 were collected from the PHMSA databases and other sources. The data sources include PHMSA, historical documents, and Alyeska Pipeline Service Company (Alyeska) internal incident data base. Alyeska has been the operator of TAPS throughout the pipeline’s operating history. All incidents contained within these various sources which met the PHMSA reporting requirement of 5 gallons or more were included. Any spills contained in independent databases which did not meet the current PHMSA reporting requirements were excluded.

The initial analysis determined that there was no significant variation in the reporting of ROW spill incidents over the entire period of operation between 1977 and 2018, however, the analysis clearly indicated that TAPS facility incidents increased significantly in 2002, reflecting the overall general increase in PHMSA reporting after that date.

Each TAPS incident was attributed (if not already addressed in the PHMSA database) by PHMSA CAUSE. Raw TAPS incident rates were calculated for each PHMSA CAUSE. Rates for ROW spills were obtained for the entire period of TAPS operation, while facility spills utilized only the period from 2002 on. The two rates were then summed to develop an overall raw rate for each PHMSA CAUSE.

### Bayesian Posterior Incident Analysis

Although the raw spill rates were useful, they could be misleading, especially given the sparseness of the data. In particular, there were no incidents in the entire operation period for certain PHMSA CAUSEs. Adoption of a zero incident rate for these cases could give a misleadingly low estimate of the true incident rate. Similarly, assuming that one incident would apply over the operating period (i.e., a spill is assumed to occur immediately before or after the data collection period) could yield a high estimate. In addition, these methods do not support an estimate of the uncertainty of the result.

To achieve a more statistically justified result, we developed a more refined method that uses the PHMSA mean incident rates as a Bayesian prior estimate of the TAPS incident rate mean and standard deviation, and then calculates the Bayesian posterior for these values. To calculate the posterior value for the spill incident rate in various Spill CAUSE categories, we assume that spills occur randomly in time via the Poisson distribution. The probability mass function (PMF) for this distribution is:

$$P_K = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

**Equation 4: Poisson Distribution**

Parameter $\lambda$ in this equation is the mean incidence rate in incidents/unit time (units 1/time), $t$ is the time interval of interest, and $k$ is an integer $k = 0, 1, 2, 3...$. This equation calculates the probability of seeing a number of events $k$ over an observation period $t$, given an expected average rate $\lambda$. Note
In Bayesian inference theory, we typically assume that the parameters of a probability distribution (such as the mean and variance for a Gaussian distribution) are themselves generated from other probability distributions. The prior distribution for the parameters can be virtually any appropriate probability distribution. Obtaining the posterior generally involves a messy numerical integration over all of the possible values of the prior distribution function parameters given a set of observations.

However, if the posterior probability distributions $p(\theta|x)$ (where $\theta$ is the desired set of parameters and $x$ is the set of observations) are in the same probability distribution family as the prior probability distribution $p(\theta)$, the prior and posterior are both called conjugate distributions, and the prior is called a conjugate prior for the likelihood function. The conjugate prior for the rate parameter $\lambda$ of the Poisson distribution is the gamma probability distribution:

$$ f_\lambda = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta \lambda} = \text{Gamma}(\alpha, \beta) $$

**Equation 5: Gamma Distribution**

where $\alpha$, $\beta$ are the shape and rate parameters for the distribution and $\Gamma$ is the Gamma function. Thus, $\lambda$ is probabilistically distributed as:

$$ f_\lambda = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta \lambda} = \text{Gamma}(\alpha, \beta) $$

**Equation 6: Gamma Distribution of the Poisson Incident Rate**

The goal would be to calculate the change in parameters $\alpha$ and $\beta$ derived from the PHMSA data as a result of observing the actual spills in TAPS.

The nice thing about using the conjugate prior, however, is that it allows us to express the posterior distribution in closed form. Per [9], the posterior distribution for the Poisson distribution rate parameter $\lambda$, assuming a Gamma distribution prior with parameters $\alpha$ and $\beta$ is given by:

$$ f_\lambda = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta \lambda} = \text{Gamma}(\alpha, \beta) $$

**Equation 7: Posterior Distribution of the Poisson Incident Rate**

In this equation, $i_p$ is the posterior value (i.e., the presumed valued for our particular pipeline) of the rate parameter (in our case, the incidence rate) and $k_i$ are the number of observations or incidents for each period $i$ in the set of $n$ time periods, each of duration $t$. We assume that mean priors for the rate are provided by the PHMSA analysis described above, however, this previous analysis only provides a mean incident rate. This is not the same as the prior values for parameters $\alpha$ and $\beta$. To address this, we note from [10] that the Gamma distribution mean $\mu$ and standard deviation $\sigma$ are related to the distribution parameters $\alpha$ and $\beta$ by:

$$ \mu = \frac{\alpha}{\beta} $$

**Equation 8: Gamma Distribution Mean**

and:

$$ \sigma^2 = \frac{\alpha}{\beta^2} $$

**Equation 9: Gamma Distribution Variance**

Therefore:

$$ \alpha = \frac{\mu^2}{\sigma^2} $$

**Equation 10: Gamma $\alpha$ as Function of Mean and Variance**

And:

$$ \beta = \frac{\mu}{\sigma^2} $$

**Equation 11: Gamma $\beta$ as Function of Mean and Variance**

We now make a couple of reasonable engineering assumptions:

- The PHMSA incident rates shown in Table 1, Table 2, and Table 3 are reasonable estimates for the prior mean spill incident rate $\mu$, since they represent averages obtained over thousands of real spill incidents on actual pipelines.
- Obtaining the prior variance $\sigma^2$ for the incident rate requires a bit more thought. Since TAPS is an actual pipeline, it should be clear that its incident rate could diverge substantially from the PHMSA mean rate. From an engineering point of view, it is reasonable to assume that the specific pipeline incident rate is within shouting distance of the PHMSA rate. We assume that ‘shouting distance’ means to the same order of magnitude. We therefore use the PHMSA rate as a relatively “informed” prior for the TAPS rate, so that $\sigma \approx \mu$.

On this assumption, the prior distribution for the TAPS Poisson rate parameter for each spill CAUSE is given by:

$$ f_\lambda = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta \lambda} = \text{Gamma} \left( 1, \frac{1}{\mu_{PHMSA}} \right) $$

**Equation 12: Prior Distribution of Poisson Rate as Gamma Function of PHMSA Incident Rate**

where $\mu_{PHMSA}$ is the spill incident rate based on the PHMSA analysis. In accordance with Equation 7, therefore, the distribution of the posterior value of the Poisson rate parameter based on the TAPS observations is given by:
Posterior Analysis Results

The analysis described in the last section produces a TAPS-specific posterior or revised estimate using the PHMSA analysis as a prior distribution for the spill incident rate in combination with available pipeline-specific (in this case, TAPS) incidence information. In general, we can expect that if no pipeline-specific data is available the posterior will reflect the PHMSA prior. As more pipeline data are made available, the posterior probability distribution will come more and more to reflect the raw pipeline incident rate.

An example of the prior and posterior distributions obtained via this analysis is shown in Figure 19 for a line wide analysis of all TAPS ROW spills between 1977 and 2018. Note that the moderately informative PHMSA prior (if it was completely uninformative, it would be a uniform distribution) has been replaced by a more peaked distribution with an obviously smaller variance. The average incident rate of 0.00040 ROW spills/mile/year (0.000249 spills/km/year) is slightly smaller than the PHMSA prior of 0.0004363 spills/mile/year (0.000271 spills/km/year).

Nevertheless, the Bayesian posterior using the approach described in the last section is 0.0000423 corrosion spills/mile/year (2.63e-5 spills/km/year). This value, though significantly reduced to less than a third of the PHMSA prior of 0.000134 spills/mile/year (8.33e-5 spills/km/year), is still clearly non-zero. This is because the PHMSA prior is low enough that observing zero spills over a roughly 41-year operating period still has a relatively high probability. Consequently, the posterior rate, though reduced in magnitude, is still not zero. Had we observed zero spills over a much longer period, it would be reduced even more.

Posterior Analysis of Aperture Distributions

A Bayesian posterior analysis of the spill aperture distributions vis-à-vis TAPS was not attempted because (1) the number of TAPS data points would be too sparse to provide any meaningful revision of the PHMSA priors, and (2) only a very limited number of the TAPS data points were accompanied by usable aperture sizes. For the time being, the PHMSA results from Table 5 are the last word on this topic.

Spill Risk Analysis

Probabilistic Spill Risk Modeling

Knowing the spill incident probability or rate and the probability distribution for the associated leak aperture is useful, but insufficient for performing a spill risk analysis. This is because of the following:

- The spill risk can be quantified in different ways: spilled volume, total response time, total cost, or by other means.
- The information obtained and provided in this analysis is not sufficient to calculate any of these. It must be
supplemented by additional calculations, which implies that other models will be required.

- Many of the parameters in the required models are not known with certainty. As was the case with the spill aperture models previously discussed, the required models are fundamentally probabilistic in nature.

For example, to calculate the spilled volume, we will need to know the probability of experiencing a spill as well as the size of the aperture. This information is supplied by the analysis previously described. In addition to this information, however, we would need to have models for the following:

1. The expected line pressure, which drives the leak.
2. A model that provides the flow rate through the orifice. Note that different models will apply for above and below ground pipe.
3. Leak detection models that address detection of leaks that are missed by the CPM system, and instead detected by operating company personnel, members of the public, and other observers. A significant fraction of spills are identified this way.
4. Remote isolation models, (i.e., models that describe the amount of pipeline that can be isolated by closing valves).
5. Shutdown spill models that calculate the volume lost once the pipeline remote valves are closed.
6. Logistic models that address time to get to site and provide further isolation of the pipeline (i.e., installation of stopples and clamps).
7. A model that relates the spill response cost to the spill volume.
8. Models that address residual uncertainties not addressed by the spill volume uncertainty. These might include the impact of whether or not the spill was in an HCA, impact of the spill being over water, impact of the landowners or stakeholder directly impacted by the spill, impact of being in a populated vs. unpopulated area, and others.

An example of a complete spill risk analysis is shown schematically in Figure 21.

Figure 22 provides a simplified example of the internal probabilistic calculations for the first three modules. Note that the model consists of user-specified constants, deterministic nodes, and probabilistic nodes. In line with the model types summarized previously in Table 4, the leak aperture area in the leak generator module is probabilistically generated by a Weibull model. The parameters for the Weibull aperture model are user-specified. The hydraulic module consists of a very simple Gaussian probabilistic node with user-specified mean and standard deviation plus a constant density. Finally, the leak flow module utilizes a simple orifice relationship, with the orifice coefficient generated probabilistically. The leak flow output from this module would then be fed into the leak detection system model, in line with Figure 21.

Figure 22: Internal Probabilistic Risk Model Components

All of the models shown in Figure 22 are very simple, and could be significantly enhanced. For example, the hydraulic module could be enhanced so that the pressure would be a function of location, in line with hydraulic modeling principles. The leak flow module could be modified so that the orifice coefficient is a function of whether the pipe is above or below ground, with a more sophisticated ground flow model.

Solution of the probabilistic spill risk model utilizes a Monte Carlo simulation approach. For each case run, the model is executed using the user-specified constant inputs. Deterministic nodes take the inputs supplied to them via the user-specified inputs, or outputs provided by other nodes. Probabilistic nodes generate random outputs according to the values of their inputs and their probability distributions. Since each case is only one sample, cases are repeated until a well-defined output distribution is obtained.
A Proof-of-Principle Spill Risk Model

To test the reasonableness of the Monte Carlo risk analysis approach, we constructed a very simple proof-of-principle/toy risk model for a hypothetical pipeline. We assumed a 100-mile-long (161 km), 12-inch-diameter (305 mm) pipe with a centerline burial of 3.5 feet (1.1 m). The components of the risk model were defined as follows:

1) The leak generation model utilized a Weibull fit for the aperture size assuming parameters for all spills causes in line with Table 5. Since handling of the incident rate is relatively easy, we did not explore this aspect of the risk analysis, but note that for a PHMSA incident rate of 1.949E-03 spill incidents/mile/year (1.21E-03 spill incidents/km/year, in line with Table 3), we would expect that the 100-mile-long pipeline would experience one spill every 5 years on average.

2) In line with the PHMSA analysis, 25% of all spills were assumed to occur on the pipeline ROW.

3) The pipeline pressure was assumed to be normally distributed with a mean value of 550 pounds per square inch gauge (PSIG) (3.79 MPa) and a standard deviation of 300 PSIG (2.07 MPa). Any pressures generated with a value less than zero were arbitrarily set to 0.1 PSIG (0.0007 MPa).

4) Leak rates were calculated as follows:

   a) Leaks inside of facilities were modeled using standard orifice relations.

   b) ROW leaks were assumed to be below ground. To enable this, our very simple ROW leak rate model assumed:

      i) A model such that any leaks of sufficient pressure and aperture size would probabilistically cause a catastrophic failure of the soil above the pipe. This model was inspired by the Jeffrey model used to simulate deformation and geologic stresses near a subsurface magma-filled cavity [11], but was enhanced as follows.

         (1) The soil above the pipe can experience tensile failure if the aperture and pipeline pressure are high enough.

         (2) In the event of a soil failure, the commodity is assumed to flow relatively unimpeded by the surrounding soil, and utilizes an orifice relationship. The orifice coefficient is probabilistically generated.

         (3) If there is no soil failure, the flow rate is a function of the soil resistance based on a porous soil Darcy flow model [12].

5) Leaks can either be detected by a CPM system, or as a result of being detected by people in the area:

   a) The CPM detection model utilized a simple API (American Pipeline Institute) 1149 model [7] assuming an upstream and downstream flow meter combo with an accuracy of 0.4%. Packing volumes assumed a pressure uncertainty of 10 PSI (0.07 MPa) and a thermal uncertainty of 4 °F (2.2 °C). Leaks at rates less than the root mean square of the meter accuracies will never be detected by the CPM system.

   b) We also assumed an alternate detection mechanism by external observers. The probabilistic detection model assumes a Poisson distribution, a mean visit rate, and a probability of detection for each visit. We assumed a (very) arbitrary mean visit rate of 2.5 times per day, and a probability of detection for each visit of 65%. This was based on a very simple model assuming visual detection on a 10-foot (3-meter) radius by randomly moving observers. The visit rate is order-of-magnitude at best, based on typical U.S. population density, and reasonable assumptions regarding the fraction of the population in transit as pedestrians at any time.

   c) The spill is probabilistically detected by either the observer or CPM model, whichever occurs first.

   d) ROW spills that are detected by external observers always have a minimum or occult spilled volume that lognormally correlates with the soil porosity and the depth of burial.

6) The maximum spill size following detection of the spill is lognormally distributed.

7) The operating company response time is likewise lognormally distributed.

8) The total spilled volume is the sum of (1) any occult spill volume (if it is a ROW spill) plus (2) additional volumes based on time-to-detect plus (3) volumes spilled while responding to the spill.

9) Cost as a function of spilled volume is based on a probabilistic spill cost model previously developed from PHMSA analysis and shown in Figure 23. See [5].
Note that on the assumption that a CPM system is installed to our 12-inch-diameter proof-off-principle pipeline, the largest spill analyzed (with a tail probability of $10^{-4}$) will be on the order of $300$MM. Sensitivity analysis performed with multiple 10,000 sample runs yielded maximum costs for this probability generally between $200$MM and $800$MM, although a few outliers outside this range were observed. It is likely that runs with much larger total samples would refine these estimates.

Now refer to Figure 25. This chart from [5] shows the CCDF of actual spill remediation costs for PHMSA data on all U.S. pipelines between 2010 and 2015. Note the strong similarities between the simulation cost on our proof-off-principle pipeline and actual simulation costs obtained from the PHMSA data. Finally, referring back to Figure 24, note the significant increase in the tail of the cost distribution to much higher values if CPM is not installed. The maximum cost value observed for this comparison curve was more than ten times the size of the largest cost for the distribution calculated when CPM detection was observed.

Figure 23: PHMSA Spill Costs vs. Spilled Commodity Volume

It is important to note that only the aperture distribution model and the cost vs. volume model shown in Figure 23 originated from the analysis of the PHMSA data. All other inputs were independently obtained based on the models described above and engineering analysis designed to develop reasonable and representative parameters consistent with the assumed models.

Proof-of-Principle Spill Risk Results

The risk model described above was run using 10,000 Monte Carlo iterations. A minimal amount of tuning of the minimum aperture size was performed to ensure that the minimum observed volumes at facilities was consistent with the PHMSA data. Other than selecting parameters that appeared a priori to be sensible (i.e., selecting reasonable soil permeabilities and porosities, external observer sampling rates, etc.), no other significant tuning was performed. Typical simulation results are shown in the form of the inverse complementary cumulative distribution functions for spill costs with and without CPM installed in Figure 24.

Figure 24: Proof-of-Principle Spill Analysis Cost Distribution


Figure 25: PHMSA Spill Cost Distribution

The following are also worth noting:

- The risk-adjusted cost for the CPM-enabled case from Figure 24 is $503,000. Eliminating CPM detection of leaks raised the risk-adjusted cost to $12.6MM.
- Despite this, only 8% of leaks in the risk-based simulation were detected by the simulated CPM system. This compares favorably with the 6% of leaks detected by a leak detection system in Figure 1.

Despite these apparent agreements between the risk model and the CPM data, at this time we feel that we need to temper any enthusiasm for comparisons between the risk model and the PHMSA cost distributions because:

1. The risk model example was targeted to a single generic pipeline example while the PHMSA data represents the cost for all pipelines in the U.S. Consequently, the examples are not comparable.
2. A full 35% of all leaks in the PHMSA data were not attributed by cause. See Figure 1. Consequently, comparisons between the model and the PHMSA data may not be on a consistent basis.
3. Furthermore, the PHMSA data indicates that only about 31% of pipelines experiencing leaks had CPM systems installed. Our simulation assumed that all pipelines have CPM installed.
4. That said, many pipelines have over/short oil accounting systems capable of alarming or notifying the controller of a system imbalance. It is also notable that the pipeline controller detects roughly 2% to 3% of spills. Given that over/short systems and the pipeline controller all work off the same information available to the CPM system, this brings to the fore the question of just what constitutes a pipeline CPM system. In that sense, it is possible that many more pipelines technically have either CPM or its functional equivalent installed than are indicated in the PHMSA database.
5. The risk model we employed is, at this time, necessarily simplistic in nature, and simply does not have the explanatory power that a more detailed simulator might supply. It also has not been seriously tested or vetted beyond the few test runs summarized here. More work is required to assess this approach.

At this time, therefore, it is really not possible to draw strong conclusions based on the risk model PHMSA cost comparisons performed and summarized here. The following assessments, therefore, are suggestive only:

• The analysis indicates that the actual fraction of spills detected by CPM systems may be fairly low even when CPM implementation is at a very high level.
• Despite this, CPM systems may have very high value in terms of significantly reducing the risk-adjusted cost of pipeline commodity leaks because they rapidly detect potentially high-volume spills in a much shorter time than would be possible via external observation.
• The overall simulation results appear to indicate that the combination of combining analysis of PHMSA data with well-thought-out engineering and process models has potential to provide useful probabilistic models of pipeline spill risk.

CONCLUSIONS AND FUTURE WORK

Our analysis has demonstrated that it is possible to extract useful statistical information from the PHMSA pipeline spill/incident databases, properly adjust those results so that they are applied to a specific pipeline, and then use the results as components of a probabilistic spill risk model that can both enable users to evaluate those risks and perform what-if analyses designed to evaluate optimum approaches to minimize risk-adjusted costs for their pipelines. Future work that we are contemplating would include:

• Enhancement of the various models to improve pipeline-specific calculations. In particular, it would be straightforward to implement better hydraulic modeling designed to provide better assessment of pipeline pressures, modification of those pressures induced by the leak itself, and site-specific calculation of maximum spill volumes based on spacing between valves.
• Improvement of the below-ground leak rate model.
• Introduction of a true, site-specific logistic model designed to address the operator response time.
• Enhancement of the external observer spill detection model to address detections by operating company personnel would be highly useful.
• We very much like having the ability to import actual tested leak detection performance curves of the types described in [13], [14], [15].
• It would also be advantageous to have a much better understanding of how to properly model spill detections by members of the public.
• The capability of analyzing the risk impact of other leak detection types, such as rarefaction wave and external systems is also desirable.
• Further analysis of the PHMSA data (and other data sources) to enhance the spill cost model shown in Figure 23. This model has a relatively low coefficient of determination of 0.31, indicating that other factors beyond the simple spill volume are also important in determining the spill cost. These may include water vs land spills, facility vs ROW sites, non-HCA vs HCA spills, and other factors, such as the number of people impacted by the spill, the state or postal district where the spill occurs, and so forth.
• Finally, the ability to enhance the risk analysis by providing additional statistical outputs and charts (e.g., volumetric distributions, risk profiles as functions of milepost), analyzing cumulative spill risk over various periods of time, and exporting outputs to external models would add considerable value.

REFERENCES

7. “Pipeline Variable Uncertainties and their Effect on Leak Detectability,” API 1149, November 1993
8. https://docs.pymc.io/

**Author Biography**

Philip Carpenter is an engineering consultant with over 35 years of experience in the areas of pipeline hydraulics, pipeline operation and control, real time systems, statistical analysis, and numerical simulation of physical processes. His company, Great Sky River Enterprises, LLC, provides services in these areas to the oil and gas industry. He has a B.S. in Aerospace Engineering and an M.S. in Engineering Science, both from the State University of New York at Buffalo, and is a registered Professional Engineer in the Commonwealth of Pennsylvania.

Morgan Henrie is the owner of MH Consulting, a firm providing national and international support to the petroleum industry. He has been involved in designing, evaluating, and implementing Supervisory Control and Data Acquisition systems, telecommunications infrastructures, and leak detection systems for more than 20 years; and has a PhD in Engineering Management and Systems Science from Old Dominion University.

Yoshi Okamoto II is an automation engineer at Alyeska Pipeline Service Company in Anchorage, Alaska. He has 5 years of experience in pipeline leak detection. His primary duties are supporting the leak detection program and Supervisory Control and Data Acquisition system for the Trans Alaska Pipeline System. Yoshihiro has a B.S. in Electrical Engineering from the University of Alaska Anchorage.

Paul Liddell is an automation engineer at Alyeska Pipeline Service Company in Anchorage, Alaska. He has over 45 years of experience as a software engineer/developer; 40 years in Supervisory Control and Data Acquisition systems development, integration and support; and 20 years in leak detection support. Paul majored in Computer Science at Washington State University.
# TABLES

### Table 1: 2010 - 2018 PHMSA Facility Spill Incident Rate (Incidents/Mile/Year)

<table>
<thead>
<tr>
<th>Leak Causes</th>
<th>Leak Cause Overall Rate</th>
<th>Mechanical Puncture</th>
<th>Rupture</th>
<th>Overfill / Overflow</th>
<th>Other</th>
<th>Other Cause Overall Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrosion Failure</td>
<td>2.506E-04</td>
<td>0.000E+00</td>
<td>4.102E-06</td>
<td>0.000E+00</td>
<td>3.913E-06</td>
<td>2.587E-04</td>
</tr>
<tr>
<td>Excavation Damage</td>
<td>3.398E-06</td>
<td>6.877E-06</td>
<td>5.326E-07</td>
<td>5.326E-07</td>
<td>3.325E-06</td>
<td>1.467E-05</td>
</tr>
<tr>
<td>Incorrect Operation</td>
<td>1.212E-04</td>
<td>0.000E+00</td>
<td>4.601E-06</td>
<td>6.724E-05</td>
<td>6.204E-05</td>
<td>2.551E-04</td>
</tr>
<tr>
<td>Material Failure</td>
<td>3.604E-05</td>
<td>0.000E+00</td>
<td>1.779E-06</td>
<td>0.000E+00</td>
<td>2.309E-06</td>
<td>4.013E-05</td>
</tr>
<tr>
<td>Natural Force Damage</td>
<td>4.902E-05</td>
<td>5.326E-07</td>
<td>1.101E-06</td>
<td>7.634E-06</td>
<td>1.812E-05</td>
<td>7.641E-05</td>
</tr>
<tr>
<td>Other Outside Force Damage</td>
<td>5.987E-06</td>
<td>1.143E-06</td>
<td>0.000E+00</td>
<td>5.326E-07</td>
<td>9.077E-06</td>
<td>1.674E-05</td>
</tr>
<tr>
<td>Other Cause</td>
<td>2.498E-05</td>
<td>5.967E-07</td>
<td>0.000E+00</td>
<td>4.374E-06</td>
<td>1.670E-05</td>
<td>4.665E-05</td>
</tr>
<tr>
<td>Release Type Overall Rate</td>
<td>1.175E-03</td>
<td>1.090E-05</td>
<td>2.091E-05</td>
<td>1.352E-04</td>
<td>1.710E-04</td>
<td>1.513E-03</td>
</tr>
</tbody>
</table>

### Table 2: 2010 - 2018 PHMSA Right-of-Way Spill Incident Rate (Incidents/Mile/Year)

<table>
<thead>
<tr>
<th>Leak Causes</th>
<th>Leak Cause Overall Rate</th>
<th>Mechanical Puncture</th>
<th>Rupture</th>
<th>Overfill / Overflow</th>
<th>Other</th>
<th>Other Cause Overall Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrosion Failure</td>
<td>1.292E-04</td>
<td>0.000E+00</td>
<td>6.218E-06</td>
<td>1.221E-06</td>
<td>1.707E-06</td>
<td>1.384E-04</td>
</tr>
<tr>
<td>Equipment Failure</td>
<td>6.560E-05</td>
<td>0.000E+00</td>
<td>5.561E-07</td>
<td>5.238E-07</td>
<td>3.775E-06</td>
<td>7.046E-05</td>
</tr>
<tr>
<td>Excavation Damage</td>
<td>5.942E-06</td>
<td>4.201E-05</td>
<td>3.883E-06</td>
<td>6.105E-07</td>
<td>1.715E-06</td>
<td>5.416E-05</td>
</tr>
<tr>
<td>Incorrect Operation</td>
<td>1.436E-05</td>
<td>0.000E+00</td>
<td>1.153E-06</td>
<td>0.000E+00</td>
<td>4.972E-06</td>
<td>2.049E-05</td>
</tr>
<tr>
<td>Material Failure</td>
<td>8.655E-05</td>
<td>5.238E-07</td>
<td>1.202E-05</td>
<td>1.832E-06</td>
<td>1.174E-06</td>
<td>1.021E-04</td>
</tr>
<tr>
<td>Natural Force Damage</td>
<td>8.752E-06</td>
<td>5.775E-07</td>
<td>1.628E-06</td>
<td>0.000E+00</td>
<td>6.052E-07</td>
<td>1.156E-05</td>
</tr>
<tr>
<td>Other Outside Force Damage</td>
<td>1.054E-05</td>
<td>4.885E-06</td>
<td>5.561E-07</td>
<td>0.000E+00</td>
<td>6.139E-06</td>
<td>2.212E-05</td>
</tr>
<tr>
<td>Other Cause</td>
<td>1.161E-05</td>
<td>5.775E-07</td>
<td>1.039E-06</td>
<td>0.000E+00</td>
<td>3.853E-06</td>
<td>1.708E-05</td>
</tr>
<tr>
<td>Release Type Overall Rate</td>
<td>3.326E-04</td>
<td>4.858E-05</td>
<td>2.706E-05</td>
<td>4.187E-06</td>
<td>2.394E-05</td>
<td>4.363E-04</td>
</tr>
</tbody>
</table>
### Table 3: 2010 - 2018 PHMSA Spill Incident Rate (All Spills, Incidents/Mile/Year)

<table>
<thead>
<tr>
<th>Leak Causes</th>
<th>Corrosion Failure</th>
<th>Equipment Failure</th>
<th>Excavation Damage</th>
<th>Incorrect Operation</th>
<th>Material Failure</th>
<th>Natural Force Damage</th>
<th>Other Outside Force Damage</th>
<th>Other Cause</th>
<th>Release Type Overall Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.000E+00</td>
<td>1.754E-06</td>
<td>4.889E-05</td>
<td>0.000E+00</td>
<td>5.238E-07</td>
<td>1.110E-06</td>
<td>6.028E-06</td>
<td>1.174E-06</td>
<td>5.948E-05</td>
</tr>
<tr>
<td></td>
<td>1.032E-05</td>
<td>9.348E-06</td>
<td>4.416E-06</td>
<td>5.753E-06</td>
<td>1.380E-05</td>
<td>2.729E-06</td>
<td>5.561E-07</td>
<td>1.039E-06</td>
<td>4.797E-05</td>
</tr>
<tr>
<td></td>
<td>1.221E-06</td>
<td>5.541E-05</td>
<td>1.143E-06</td>
<td>6.724E-05</td>
<td>1.830E-05</td>
<td>7.634E-06</td>
<td>5.326E-07</td>
<td>4.374E-06</td>
<td>1.394E-05</td>
</tr>
<tr>
<td></td>
<td>5.620E-06</td>
<td>5.928E-05</td>
<td>5.040E-06</td>
<td>6.702E-05</td>
<td>3.483E-06</td>
<td>1.873E-05</td>
<td>1.522E-05</td>
<td>2.055E-05</td>
<td>1.949E-04</td>
</tr>
</tbody>
</table>

### Table 4: Leak Aperture Probability Distribution Functions

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pareto w/constants $x_m, \alpha$</td>
<td>$f_X = \frac{\alpha x^\alpha}{x^{\alpha+1}}$</td>
<td>$F_X = 1 - \left(\frac{x_m}{x}\right)^\alpha$</td>
<td>$1 - F_X = \left(\frac{x_m}{x}\right)^\alpha$</td>
</tr>
<tr>
<td>Lognormal w/constants $\mu, \sigma$</td>
<td>$f_X = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$</td>
<td>$F_X = \frac{1}{2} \left(1 + \text{erf}\left[\frac{\ln(x) - \mu}{\sqrt{2}\sigma}\right]\right)$</td>
<td>$1 - F_X = \frac{1}{2} - \text{erf}\left[\frac{\ln(x) - \mu}{\sqrt{2}\sigma}\right]$</td>
</tr>
<tr>
<td>Weibull w/constants $\lambda, k_W$</td>
<td>$f_X = \frac{k_W}{\lambda} \left(\frac{x}{\lambda}\right)^{k_W-1} e^{-\left(\frac{x}{\lambda}\right)^{k_W}}$</td>
<td>$F_X = 1 - e^{-\left(\frac{x}{\lambda}\right)^{k_W}}$</td>
<td>$1 - F_X = e^{-\left(\frac{x}{\lambda}\right)^{k_W}}$</td>
</tr>
<tr>
<td>ID</td>
<td>Cause/Scope</td>
<td>Pareto Distribution</td>
<td>Lognormal Distribution</td>
</tr>
<tr>
<td>-----</td>
<td>--------------------------------</td>
<td>---------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$x_m$</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>0</td>
<td>ALL CAUSES</td>
<td>2.44468E-05</td>
<td>-0.493502</td>
</tr>
<tr>
<td>1</td>
<td>EQUIPMENT FAILURE</td>
<td>1.65444E-09</td>
<td>-0.386346</td>
</tr>
<tr>
<td>2</td>
<td>INCORRECT OPERATION</td>
<td>0.000262349</td>
<td>-0.658119</td>
</tr>
<tr>
<td>3</td>
<td>OTHER OUTSIDE FORCE DAMAGE</td>
<td>8.50041E-06</td>
<td>-0.350085</td>
</tr>
<tr>
<td>4</td>
<td>CORROSION FAILURE</td>
<td>1.83526E-05</td>
<td>-0.502306</td>
</tr>
<tr>
<td>5</td>
<td>OTHER INCIDENT CAUSE</td>
<td>6.24705E-12</td>
<td>-0.152804</td>
</tr>
<tr>
<td>6</td>
<td>MATERIAL FAILURE OF PIPE OR WELD</td>
<td>0.000206174</td>
<td>-0.42301</td>
</tr>
<tr>
<td>7</td>
<td>EXCAVATION DAMAGE</td>
<td>0.002710301</td>
<td>-0.514958</td>
</tr>
<tr>
<td>8</td>
<td>NATURAL FORCE DAMAGE</td>
<td>2.31333E-07</td>
<td>-0.296569</td>
</tr>
<tr>
<td>9</td>
<td>ALL EXTERNAL FORCE DAMAGE (3+7+8)</td>
<td>0.000992775</td>
<td>-0.5773</td>
</tr>
<tr>
<td>10</td>
<td>ALL MATERIAL FAILURES (4+6)</td>
<td>4.8051E-05</td>
<td>-0.46894</td>
</tr>
</tbody>
</table>

Table 6: Equivalent Spill Volumes Relative to Leak Rate and Aperture

<table>
<thead>
<tr>
<th>Leak Rate (CFS)</th>
<th>Aperture Area (ft^2)</th>
<th>Aperture Diameter (ft)</th>
<th>Aperture Diameter (in)</th>
<th>Days Required to Spill 1.0 ft^3</th>
<th>Spill Volume After One Day (ft^3)</th>
<th>Spill Volume After One Day (BBL)</th>
<th>Equivalent Linear Dimension after One Day (inches)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100.0</td>
<td>0.45623</td>
<td>0.76216</td>
<td>9.145913</td>
<td>1.15741E-07</td>
<td>8640000</td>
<td>1538845.7</td>
<td>2449.29</td>
</tr>
<tr>
<td>10.0</td>
<td>0.04562</td>
<td>0.24102</td>
<td>2.892191</td>
<td>1.15741E-06</td>
<td>864000</td>
<td>153884.57</td>
<td>1137.73</td>
</tr>
<tr>
<td>1.0</td>
<td>0.00456</td>
<td>0.07622</td>
<td>0.914591</td>
<td>1.15741E-05</td>
<td>8640</td>
<td>15388.457</td>
<td>528.49</td>
</tr>
<tr>
<td>1.0E-01</td>
<td>0.00046</td>
<td>0.02410</td>
<td>0.289219</td>
<td>0.0001158</td>
<td>864</td>
<td>1538.8457</td>
<td>245.49</td>
</tr>
<tr>
<td>1.0E-02</td>
<td>4.5623E-05</td>
<td>0.00762</td>
<td>0.091459</td>
<td>0.0011574</td>
<td>864</td>
<td>153.88457</td>
<td>114.04</td>
</tr>
<tr>
<td>1.0E-03</td>
<td>4.5623E-06</td>
<td>0.00241</td>
<td>0.028922</td>
<td>0.0115741</td>
<td>86.4</td>
<td>15.388457</td>
<td>52.97</td>
</tr>
<tr>
<td>1.0E-04</td>
<td>4.5623E-07</td>
<td>0.00076</td>
<td>0.009146</td>
<td>0.1157407</td>
<td>8.64</td>
<td>1.5388457</td>
<td>24.61</td>
</tr>
<tr>
<td>1.0E-05</td>
<td>4.5623E-08</td>
<td>0.00024</td>
<td>0.002892</td>
<td>1.1574074</td>
<td>0.864</td>
<td>0.15388457</td>
<td>11.43</td>
</tr>
<tr>
<td>1.0E-06</td>
<td>4.5623E-09</td>
<td>7.62159E-05</td>
<td>0.000915</td>
<td>11.574074</td>
<td>0.0864</td>
<td>0.015388457</td>
<td>5.31</td>
</tr>
<tr>
<td>1.0E-07</td>
<td>4.5623E-10</td>
<td>2.41016E-05</td>
<td>0.000289</td>
<td>115.74074</td>
<td>0.00864</td>
<td>0.001538846</td>
<td>2.47</td>
</tr>
<tr>
<td>1.0E-08</td>
<td>4.5623E-11</td>
<td>7.62159E-06</td>
<td>9.14591E-05</td>
<td>115.74074</td>
<td>0.000864</td>
<td>0.000153885</td>
<td>1.15</td>
</tr>
</tbody>
</table>

Note: 500 PSI pressure differential assumed