Puncture Failure Size Probability Distribution for CO_2 Pipelines[1]

Jiahuan Yi, Sergey Martynov and Haroun Mahgerefteh*

Department of Chemical Engineering, University College London, London WC1E 7JE, United Kingdom

*Corresponding author:

E-mail address: h.mahgerefteh@ucl.ac.uk (Haroun Mahgerefteh)

Abstract

The safe operation of pressurised CO_2 pipelines is key to the success of Carbon Capture and Storage as a viable means for tackling global warming. As such, the prediction of their puncture size in the event of pipeline failure and how it compares with that of hydrocarbon pipelines are fundamentally important questions that must be resolved for the subsequent pipe risk assessment and mitigation planning. The above requires the use of sufficient failure statistics to derive the corresponding puncture size probability distribution. Nevertheless, this presents a significant challenge for CO_2 pipelines given their relatively small number currently in operation. In order to address such a challenge, this paper presents the development and application of a robust statistical analytical

[1] *The manuscript accepted for publication in International Journal of Greenhouse Gas Control. The final version was published in Volume 125, May 2023, 103889.* https://doi.org/10.1016/j.ijggc.2023.103889

technique for the confident prediction of the puncture failure size probability distribution for CO_2 pipelines in the absence of sufficient failure statistics. The above involves fitting statistical distributions to the historical puncture size failure data using the Maximum Likelihood Estimator (MLE). The minimum acceptable sample size sufficient for acquiring a reliable MLE is determined by calculating the mean squared error of the MLE as a function of the data sample size. To improve the estimation confidence given the scarcity of historical failure data, the bootstrapping method is employed to obtain the 95% confidence interval of the MLE. The application of the above technique to pressurised CO_2 and hydrocarbon pipelines indicates that as compared to the latter, CO_2 pipelines are most likely to experience smaller puncture size failures (e.g. $\leq$ 50 mm), thus resulting in smaller magnitude but more prolonged releases. This directly impacts the preventive and emergency response planning required especially in the case of buried CO_2 pipelines where small leaks can remain undetected for long periods.

Keywords: Carbon Capture and Storage (CCS), CO_2 pipelines, High-pressure pipeline safety, Pipeline failure probability distribution

1. Introduction

The intensive use of fossil fuels has resulted in excessive CO_2 emissions worldwide, leading to global warming. Carbon Capture and Storage (CCS), involving the capture of CO_2 from fossil fuel power plants and industrial operations such as cement and steel making for the subsequent long-term geological storage, is widely recognised as a key player in addressing the above issue. According to the Global CCS Institute (GCCSI, 2020), the amount of captured CO_2 needed for the energy sector to achieve net zero emission will increase significantly by hundred-fold in the next three decades, reaching around 5.6 Gt in 2050.

An essential part of the CCS chain involves the large-scale transportation of the captured CO_2 to the storage site. High-pressure pipelines are widely considered or already being employed as the most practical and economical transport option (Mahgerefteh et al., 2012). In Norway, for example, Sleipner and Snøhvit projects each inject ca. 1 million tonnes of CO_2 per year into saline aquifers (GCCSI, 2021; Ringrose, 2018) employing sub-sea pipelines. Several CCS projects are being developed connecting onshore capture facilities to offshore geological storage locations. These include Northern Lights (Norway), Porthos and Athos (Netherlands), ERVIA (Ireland) and ACORN (UK). Many plan to commence operation well before 2030, operating at the order of 1 million tonnes of captured CO_2 per year (Moe et al., 2020) using high-pressure (80 to 110 bar) pipelines with diameters typically ranging from 100 to 600 mm. Taking advantage of the economies of scale, many CCS projects comprise industrial clusters, connecting several major emission sources using common pipeline transportation network and storage infrastructures. Major examples in the UK include

Humber Zero, Net Zero Teesside and HyNet (GCCSI, 2021).

Given the above, the global demand for high-pressure CO_2 pipelines is expected to increase substantially, with estimates ranging from 95,000 to 500,000 km in length by 2050 (IEA, 2010).

In the case of densely populated areas, such as most regions in Europe, pipeline routing through or near densely populated areas may become inevitable (Cosham and Eiber, 2008; Koornneef et al., 2010; Vitali et al., 2022). This poses a risk in the event of an accidental release. At CO_2 concentration of 10% v/v, an exposed individual would lapse into unconsciousness in 1 min. At above 20% v/v, the gas is instantaneously fatal (Pohanish and Greene, 1996). The ability of CO_2 to collect in depressions in the land, in basements and in other low-lying areas such as valleys near the pipeline route, presents a significant hazard if leaks continue undetected (Barrie et al., 2004). Hydrocarbons such as natural gas will eventually ignite or explode in such areas if, and when, conditions are right, but CO_2 can remain undetected for a very long time (Mahgerefteh et al., 2008). Additionally, the captured CO_2 will be usually mixed with potentially toxic impurities such as H_2S at low concentrations (ca. 50 ppm; Jensen et al., 2014) whose natural dispersion might be impeded by the dense CO_2 vapour layer close to the ground, further increasing the hazard.

There are several other hazards associated with the accidental release of CO_2. It can act as an ignition source for nearby combustible materials due to friction induced static discharge. In 1953, such an incident resulted in 29 fatalities (Barrie et al., 2004). CO_2 also reacts with water to form carbonic acid leading to the corrosion of carbon steel pipelines (Nesic, 2012). Supercritical CO_2, widely considered to be the most economical state for pipeline transportation is a powerful solvent giving possible toxic contamination and sealing problems (Connolly and Cusco, 2007).

The accidental release of CO_2 following a pipeline puncture and the subsequent Joule Thomson expansion cooling may result in the pipe wall temperature dropping below its ductile to brittle transition temperature, giving rise to the risk of the initial puncture transforming into a running brittle fracture (Mahgerefteh and Atti, 2006).

If the CO_2 temperature drops below its triple point (-56.6 °C at 0.518 MPa; Angus et al., 1973) during its rapid expansion, for example during emergency depressurisation, the resulting solid CO_2 may cause vent valve blockage. Finally, the unusually high saturation pressure of CO_2 reduces the pipeline's resistance to long running ductile fractures (Aursand et al., 2013).

Given the above hazards, along with the extensive projected use of CO_2 pipelines as part of the CCS chain, it is clear the risks associated with their operation must be reliably predicted in order to implement appropriate mitigation steps to reduce such risks to as low as reasonably practicable.

An important part of the risk assessment process for pressurised pipelines involves calculating the probability of loss of containment events. Such information is in turn employed to estimate the individual and societal risk levels (Goodfellow et al., 2012) forming the basis for appropriate control and emergency mitigation planning.

To this end, many studies (see for example Duncan and Wang, 2014; EGIG, 2018; Lyons et al., 2020) have focused on collecting pipeline failure statistics and estimating the corresponding frequencies for given failure modes such leak, puncture, or Full Bore Rupture (FBR). Among these modes, the through-wall punctures, formed often as a result of corrosion or external interference, are found to be by far the most frequent (Lydell, 2000). Given that the puncture size directly affects the failure consequences and hence the subsequent determination of the appropriate control and mitigation measures, a reliable technique for estimating the puncture size failure frequency must be established. The efficacy of such techniques (see for example Duncan and Wang, 2014) is largely dependent on ensuring that a ‘sufficiently’ large number of real incident data points are available to be representative. This is however problematic in the case of CO_2 pipelines given their relatively low number. As of 2021, there were only ca. 8000 km of CO_2 pipelines; almost entirely for enhanced oil recovery and mostly located in the United States (IEA, 2021). This compares with over 4.6 million km of pressurised hydrocarbon pipelines crossing the globe (CIA, 2021).

Accordingly, given the above, it is imperative that i) a sufficiently large sample size is taken to determine the generic puncture frequency and ii) a rigorous methodology is employed to determine the corresponding puncture size probability distribution. Another important issue to address is how such a probability distribution for CO_2 pipelines compares to that of hydrocarbon pipelines.

Several methodologies to determine the generic puncture frequency for CO_2 pipelines have been reported in the open literature; the most popular being those based on using natural gas pipelines data as a proxy. CO_2 and natural gas pipelines are similar in their construction materials, fabrication techniques and failure mechanisms (Barrie et al., 2005; Duncan and Wang, 2014).

To obtain the puncture size probability distribution, a histogram using existing pipeline failure data is constructed by first segmenting the entire range of puncture sizes into a series of intervals (bins) and then counting how many values fall into each bin. Duncan and Wang (2014) employed the above technique to approximate the puncture diameter occurrence probability distribution for CO_2 pipelines using the incident data from the Pipeline and Hazardous Material Safety Administration (PHMSA) database. In their study, puncture diameters ranging between 0 and 380 mm were divided into 6 bins. The analysis showed that the most prevailing puncture diameters were between 50 to 100 mm, whereas medium-sized punctures (150 to 200 mm) had the lowest probability of occurrence.

The resulting histogram can be parameterised and extended to a smooth probability distribution function. The validity of such functions is largely dependent on the size of the sample employed to derive the underlying histograms. Despite their usefulness, methods to reliably handle 'small' samples sizes are not well established.

Given the above limitation, in most risk assessment studies for pressurised pipelines, the puncture size is usually assumed to be a discrete variable as opposed to continuous variable as is the case in reality. In many of these studies only a limited number of representative puncture sizes are used to cover the whole size spectrum and as a result the predicted failure risk levels can only present rough estimations, rendering the subsequent strategies for risk mitigation uncertain.

Medina et al. (2012), for example used only two representative puncture sizes of 10 and 40 mm and FBR to calculate the expected cost of pipeline failure consequences for a risk-based optimisation of emergency shut down valve spacing for on-shore pipelines. Rusin and Stolecka (2015) on the other hand, used the same approach to calculate the frequency of the various failure modes for CO_2 pipelines for optimising inline emergency isolation valve spacing. The through-wall failure was simply assumed to be either puncture or rupture with the ratio of puncture/rupture occurance probabilty taken as 9:1.

Considering the above limitations, this paper presents the development of a statistical analytical technique for determining reliable puncture failure size probability distribution for pressurised CO_2 pipelines using available historical failure data. The above involves a) using the Maximum Likelihood Estimator (MLE) to fit statistical distribution functions to historical failure data for estimating the unknown fitting parameters that characterise these statistical distribution functions, and b) performing a Monte Carlo simulation test to assess the quality (statistical significance) of the MLE based on the data sample size. When the MLE 'quality' is low, a bootstrapping method, which can artificially inflate the sample size, is employed to calculate the MLE confidence intervals.

The paper proceeds as follows. Section 2 commences with a brief introduction, filtering and processing of the pipelines failure historical data used for this study, followed by the description of the methodology employed to obtain a credible probability distribution of the puncture size. In Section 3, the 'quality' of MLE is first evaluated based on Monte Carlo simulation tests involving calculating the corresponding mean squared error of the MLE. Next, puncture size probability distributions derived from the filtered and processed historical failure data alongside the recommended fitting parameters for CO_2 pipelines are presented and compared against those for natural gas and crude oil pipelines. Conclusions and suggestions for future work are presented in Section 4.

2. Methodology

2.1 Data review

Several bodies collecting and publishing the failure statistics for CO_2 and hydrocarbon pipelines exist (see for example Concawe, 2011; EGIG, 2018; PHMSA, 2020), but few provide detailed information on the size of through-wall puncture. This study adopts the Pipeline and Hazardous Material Safety Administration (PHMSA) database where such information is available. The puncture size, assumed to be oval, is expressed in terms of Equivalent Puncture Diameter (EPD) given by Koch (2008):

$$\mathrm{EPD} = 1.55\frac{A^{0.625}}{P^{0.25}} \tag{1}$$

where A and P are respectively the oval puncture cross-section area and perimeter calculated based on the circumferential and longitudinal lengths of the puncture recorded in the database.

The PHMSA database holds data on the loss of containment incidents for federal and state-regulated CO_2 and hydrocarbon pipelines operating in the US since 1970s. Whilst the focus is on CO_2 pipelines, this study also examines natural gas and crude oil pipelines punctures size failure probability distributions for comparison purposes as compared to those for CO_2 pipelines.

In much of the databases spanning over 50 years, the records are of varying quality and level of detail for the various incidents. So, it is necessary to review and filter such data before use. The PHMSA updates its reporting criteria for pipeline incidents every 10 to 20 years for the past 5 decades. This study employs the data since 2010, when the reporting criteria were last updated. The raw data are publicly available from the PHMSA database (PHMSA, 2020). From 2010 to present, 6495 loss of containment incidents have been recorded but not all are relevant for this study for the following reasons.

First, a large proportion of the loss of containment incidents are reported for leaks from pipeline auxiliary equipment (e.g. relief valves, compressors, connectors) rather than those from the pipeline itself. Second, the puncture size information is reported for selected incidents only. Taking account of the above limitations leads to a remaining total of 1906 useful EPD data points employed in the current work.

2.2 Statistical distribution models

The probability distribution of a continuous variable is often expressed as the Cumulative Distribution Function (CDF). In this study, we employ the Weibull and lognormal distributions as the potential statistical functions to represent the CDF of the puncture size. Both functions are widely used in reliability engineering for the assessment of pipeline failures (see for example Chaplin, 2015; Goodfellow et al., 2012). Other possible distributions such as the gamma and exponential distributions have been used to a much lesser extent and hence are not considered here. The CDF of the Weibull distribution takes the form:

$$F(x; \alpha, \beta) = 1 - e^{-\left(\frac{x}{\alpha}\right)^{\beta}} \qquad x \geq 0 \tag{2}$$

where α and β are respectively the scale and shape parameters of the Weibull distribution. x and F respectively represent the random variable and CDF.

The CDF of the lognormal distribution is given by:

$$F(x; \mu, \sigma) = \frac{1}{2} \operatorname{erfc}\left(-\frac{\ln x - \mu}{\sigma\sqrt{2}}\right) \tag{3}$$

where μ and σ are respectively the mean and standard deviation of the variable, x. erfc, on the other hand, refers to the complementary error function which is defined as:

$$\operatorname{erfc}(x) = -\frac{2}{\sqrt{\pi}} \int_{x}^{\infty} e^{-t^2} dt \tag{4}$$

Typical examples of Weibull and lognormal CDFs are respectively shown in Figures 1 and 2.

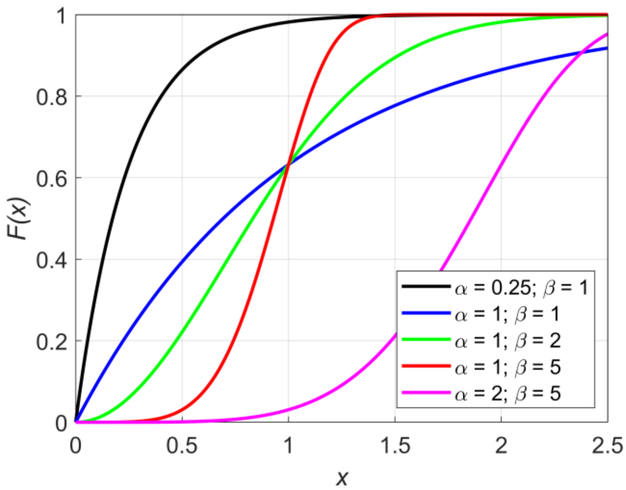

Figure 1. Typical examples of Weibull CDF showing the variation of the CDF against the random variable.

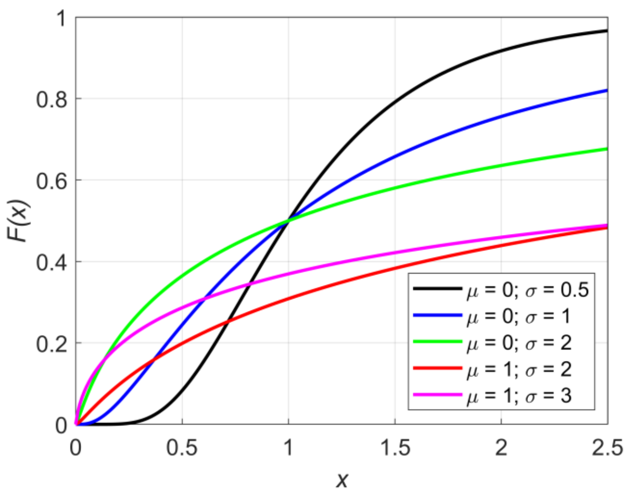

Figure 2. Typical examples of lognormal CDF showing the variation of the CDF against the random variable.

2.3 Distribution fitting

The selected puncture size sample data from the PHMSA database (PHMSA, 2020) described in Section 2.1 are fitted to both Weibull and lognormal distributions described in Section 2.2 to acquire the fitting parameters. The Maximum Likelihood Estimator (MLE) is used for this purpose to best characterise the probability distribution of the sample data. The MLE is a widely adopted method for estimating the parameters of an assumed probability distribution for a given set of observed data, by finding the parameter values that will most likely generate the observed data. Mathematically, the MLE can be defined as:

$$\hat{\theta} \stackrel{\text{def}}{=} \underset{\theta}{\operatorname{argmax}}\, L(\theta; X) \tag{5}$$

where θ is the unknown true parameter characterising the assumed probability distribution and $\hat{\theta}$ refers to the MLE of θ. X denotes the data sample that contains n observations (x_1, x_2, …, x_n) of the data population. L, on the other hand, is called the likelihood function which calculates the product of the probability densities of each value in X and is mathematically expressed as:

$$L(\theta; X) = \prod_{i=1}^{n} f(x_i; \theta) \tag{6}$$

In essence, the process of maximum likelihood estimation is to find the estimator that maximises the likelihood function, Eq. (6).

According to Ginos (2009), the MLE is among the most dependable statistical estimators for parameter estimation. Some appealing features of the MLE include it being consistent, efficient and asymptotically normal (Long and Freese, 2006). However, these properties have been only proven to hold as the number of data being used in the estimation process approaches infinity (Ji, 2020). This is an issue in the case of CO_2 pipeline failures, where relatively small sample sizes are available thus limiting the applicability of MLE for the present study. Given this, whether the sample size can suffice for a high-quality MLE needs to be determined.

Eliason (1993) suggested that a sample size of more than 60 is usually large enough for estimating no more than 5 parameters using MLE. Long and Freese (2006) on the other hand, suggested that it is risky to use MLE with sample sizes smaller than 100, while sample sizes over 500 seem adequate. However, most of the literature dealing with MLE do not provide specific sample size guidelines. In general, there are no rules of thumb, and the appropriate sizes heavily depend on the question at hand.

In this study, to determine the appropriate sample size, the quality of MLE is assessed by examining the mean squared error which is the averaged square difference between the estimated and the actual values (Ryan, 2007). The use of mean squared error is very common in the study of MLE (see for example Ginos, 2009; Nielsen, 2011), and it is considered an excellent general-purpose error metric for numerical predictions (Neill and Hashemi, 2018). Mathematically, the mean squared error of the MLE, $\hat{\theta}$ to an unknown parameter, θ is defined as the addition of the variance and bias squared:

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Variance}(\hat{\theta}) + \mathrm{Bias}^2(\hat{\theta}, \theta) \tag{7}$$

where MSE denotes the mean squared error and the variance and bias are respectively given by:

$$\mathrm{Variance}(\hat{\theta}) = E\left[\left(\hat{\theta} - E[\hat{\theta}]\right)^2\right] \tag{8}$$

$$\mathrm{Bias}(\hat{\theta}, \theta) = E[\hat{\theta}] - \theta \tag{9}$$

where E denotes the expected value.

In the present study, we perform Monte Carlo simulation tests to investigate the quality of MLE based on computing the mean squared errors for different sample sizes. The tests involve, i) determining the Weibull and lognormal distribution parameters and sample sizes being tested; ii) for a given sample size, N, calculating the corresponding MLE using N data randomly sampled from the Weibull and lognormal distributions determined in step i); iii) repeating step ii) for a sufficiently large number of times (typically over 1,000 times) and computing the corresponding mean squared error.

2.4 Determination of probability distribution

Following the above Monte Carlo simulation tests, the sample size sufficing for a high-quality Maximum Likelihood Estimator (MLE) is obtained. For sufficiently large samples, the resulting high-quality MLE can be used with confidence to characterise the probability distribution of the puncture size. However, given that two distribution models (i.e. Weibull and lognormal) are employed in this work, the one-sample Kolmogorov-Smirnov (K-S) goodness-of-fit test (Kolmogorov, 1933) involving comparing the sample data with the predictions of both models is further employed to determine which model provides a statistically better fit representing the sample population. The test process involves, i) specifying a null hypothesis; ii) computing the K-S statistic and critical value at a chosen significance level and iii) accepting the null hypothesis if the K-S statistic is smaller than the critical value or rejecting the null hypothesis if otherwise.

The K-S statistic is computed based on quantifying the greatest vertical distance between the empirical CDF (the sample data) and the CDF of the reference distribution, that is (Conover, 1999):

$$D = \sup_x |F_n(x) - F(x)| \tag{10}$$

where D denotes the K-S statistic. $\sup$ stands for supremum which means the greatest. $F_n(x)$, on the other hand, is the empirical CDF for n ordered sample data points ($x_1 < x_2 < \ldots < x_n$), which is a step function jumping up by $1/n$ at each of the n data points.

The critical value, on the other hand, is usually determined using a K-S test critical value table, which can be easily obtained from several literature, such as Massey Jr (1951). In particular, for $n \geq 40$, the critical value is computed based on a specific equation depending on the chosen significance level. In the current study, a significance level of 0.01 is chosen for the K-S test and the corresponding equation for calculating the critical value is given by:

$$\text{Critical value} = \frac{1.63}{\sqrt{n}} \tag{11}$$

To deal with the small sample sizes, the bootstrapping method, which can artificially inflate the sample size by random sampling with replacement is employed to calculate the MLE confidence interval. The methodology was first introduced by Efron (1979) for making inferences from data without making strong distributional assumptions, and was later employed by many authors for enhancing the confidence in using MLE for small samples (see for example Tsagkanos, 2008; Wei and Li, 2019). Unlike the case for sufficiently large samples where a single value of the MLE is acquired, the bootstrapping process produces a range of values where the MLE is expected to lie. It should be noted that for small samples which may not be statistically representative of the population being considered, the aforementioned K-S test cannot accurately reflect the goodness-of-fit between the model predictions and the data. Given this, the probability distributions derived based on either Weibull or lognormal models are considered statistically valid for the purpose of this study if the bootstrapping method is employed.

The bootstrapping process comprises the following steps. First, the bootstrap samples are generated. This involves resampling the original data sample with replacement to create a resampled dataset (also known as a bootstrap sample) that have the same size as the original sample. Second, the MLE of each bootstrap sample is computed based on Eqs. (5) and (6). Third, the above first and second steps are repeated for a sufficiently large number of times to obtain a distribution for the possible values of the MLE. Fourth, the MLE confidence interval is calculated based on the distribution obtained from the third step. Several options including the normal approximation method, percentile method, bias-corrected method etc. can be adopted to calculate the MLE confidence interval. In this study, the percentile method, which is considered suitable for small samples (Wei and Li, 2019), is employed. The MLE confidence interval based on the percentile method can be given as follows (Jung et al., 2019):

$$[\hat{\theta}_{\text{lower limit}}, \hat{\theta}_{\text{upper limit}}] = [\hat{\theta}_j, \hat{\theta}_k] \tag{12}$$

where j and k respectively refer to the jth and kth quantiles of the collection of the possible MLE values ordered from lowest to highest. Here, j and k are respectively:

$$j = \frac{s}{2} \times B \tag{13}$$

$$k = \left(1 - \frac{s}{2}\right) \times B \tag{14}$$

where s is the level of significance and B is the number of bootstrap samples generated in the bootstrapping process.

3. Results and discussion

3.1 Monte Carlo simulation setup

In this section, four tests following the Monte Carlo simulation steps described in

Section 2.3 are performed. Based on investigating several assumed distributions, tests 1 to 4 respectively examine the MLE quality as a function of sample size for Weibull scale parameter, Weibull shape parameter, lognormal mean, and lognormal standard deviation. Each test examines three pairs of parameters varying the value of the tested parameter whereas fixing that of the non-tested parameter. The corresponding parameter values for tests 1 to 4 are given in Table 1. The investigated values are selected based on the fact that small pipeline punctures are far more frequent than large ruptures (Lydell, 2000).

For each pair of examined parameter values, the mean squared error of the MLE is calculated for a wide range of sample sizes, $N =$ 10, 20, …, 100, 200, …, 500. The following details how the Monte Carlo simulation steps described in Section 2.3 are implemented for a given N.

First, N data are randomly selected from the distribution characterised by the examined value pair using a random value generator. Second, using the selected N data, the MLE to the examined parameter is computed based on Eqs. (5) and (6). Third, in order to accurately approximate the mean squared error of the examined MLE, 10,000 MLEs to the examined parameters are generated by repeating the above steps. The mean squared error of these MLEs is then computed based on Eq. (7).

The above process is executed for each investigated sample size. The resulting mean squared errors are then plotted against the corresponding N. As such, a figure showing the variation of the MLE mean squared error as a function of the sample size is obtained.

Table 1. Summary of the Weibull and lognormal distribution parameter values examined in the four Monte Carlo simulation tests for investigating the quality of MLE based on the sample size.

Test no.	**Tested parameter**	**Non-tested parameter**
1	Weibull scale parameter, α=1, 1.5, 2	Weibull shape parameter, β=2
2	Weibull shape parameter, β=1.5, 2, 2.5	Weibull scale parameter, α=1
3	Lognormal mean, μ=0, 0.5, 1	Lognormal standard deviation, σ=1
4	Lognormal standard deviation, σ=1, 1.5, 2	Lognormal mean, μ= 0.25

3.2 Monte Carlo simulation results

Figure 3 presents the simulation results for tests 1 to 4 described in Table 1, respectively showing the variations of the MLE mean squared error as a function of the sample size, N for the Weibull scale parameter, Weibull shape parameter, lognormal mean, and lognormal standard deviation. Figure 4 on the other hand shows the same results plotted in logarithmic scale to aid visualisation.

As may be observed from both Figures 3 and 4, three distinct regions describing the behaviour of the mean squared error variation with sample size may be identified. Initially, when the sample size is smaller than ca. 100, the mean squared error drops significantly indicating that the MLE quality is highly sensitive to the sample size and therefore the MLE should be used with caution in this region. At sample sizes between 100 to 200, the rate of decrease in mean squared error slows down, indicating that using samples with more than 100 data points will substantially improve the MLE quality. In the third region where the sample size surpasses 200, the rate of decrease in mean squared error further slows down, meaning that further increasing the sample size provides limited improvement in the MLE quality.

The above indicates that the minimum acceptable sample size sufficing for acquiring a reliable MLE is at least 100 while with ideally more than 200 data points, sufficiently reliable statistical representation of the puncture size data population may be obtained.

Figure 4 provides a closer look at the mean square error for large sample sizes. Here, for all tested parameters, the mean squared error drops almost linearly when the sample size is increased from 100 to 200, suggesting minimal marginal increase in the MLE quality. When the sample size exceeds 200, the mean squared error tend to converge between the 0.001 to 0.01 range, again indicating that further increasing the sample size provides limited improvement in the MLE quality. This further strengthens the conclusion drawn from Figure 3 that 100 is the minimum acceptable sample size sufficing for acquiring a reliable MLE while more than 200 is ideal.

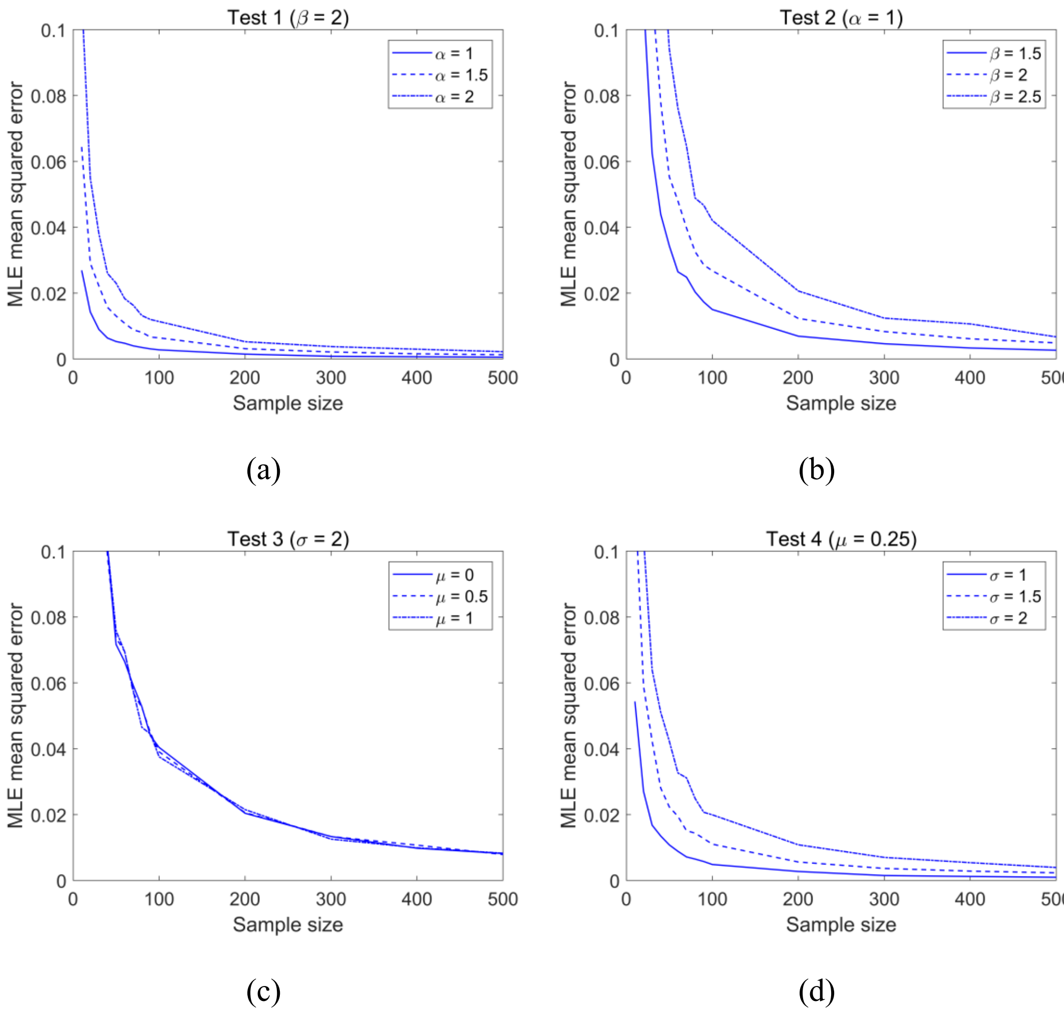

Figure 3. Simulation results for tests 1 to 4 described in Table 1 showing the variations of the MLE mean squared error as a function of the sample size, N for the Weibull scale parameter (a), Weibull shape parameter (b), lognormal mean (c), and lognormal standard deviation (d).

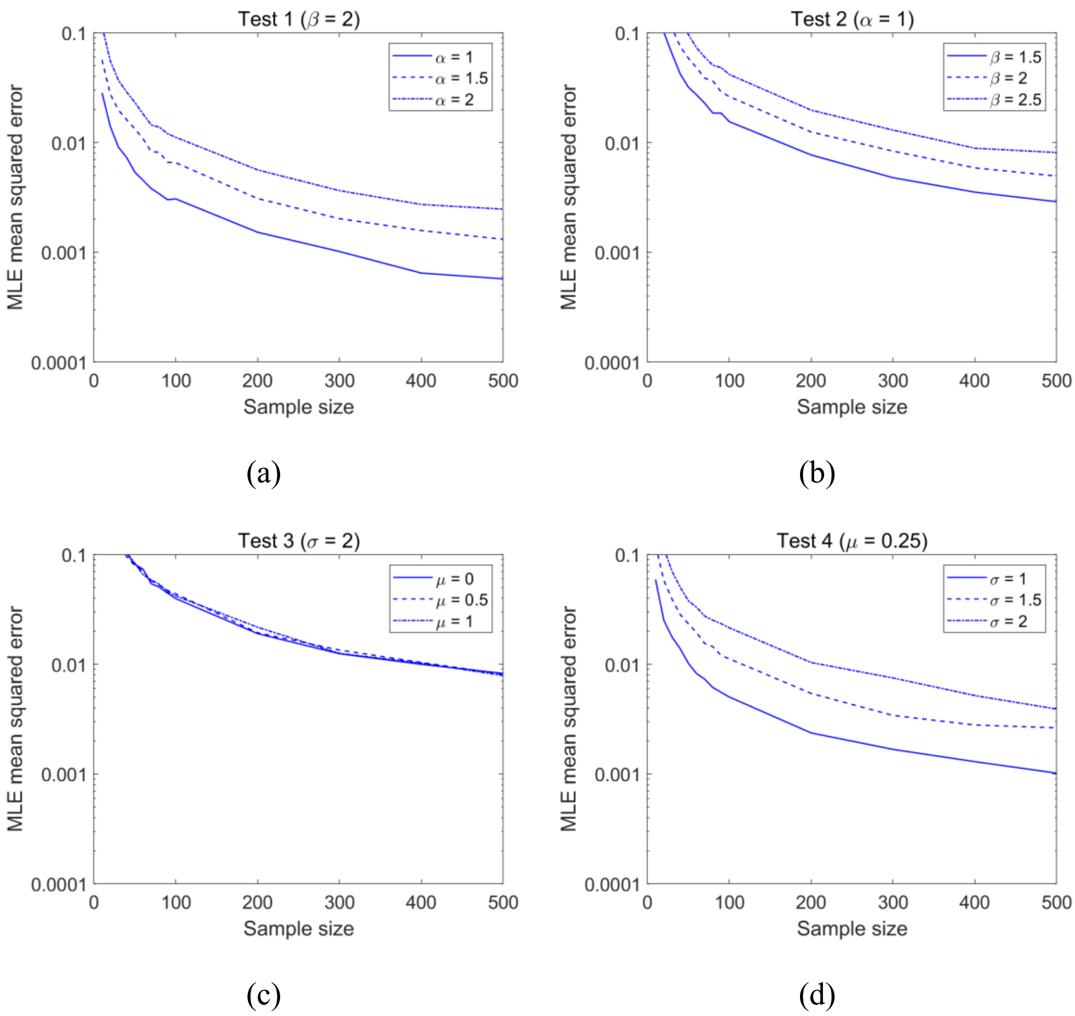

Figure 4. Simulation results for tests 1 to 4 presented in Table 1 showing the logarithmic variations of the MLE mean squared error as a function of the sample size, N for the Weibull scale parameter (a), Weibull shape parameter (b), lognormal mean (c), and lognormal standard deviation (d).

3.3 Probability distribution results

The following section presents the application of the methodology presented in Section 2 to obtain the probability distribution of the equivalent puncture diameter data from the PHMSA database (PHMSA, 2020). Here, 1456 data points (see Section 2.1) are divided into three pipeline inventories, covering natural gas, crude oil and CO_2. The corresponding data counts, and parameter estimation methods determined based on the results of the Monte Carlo simulation tests (see Section 3.2) are summarised in Table 2.

Table 2. Failure counts summary from the PHMSA database and parameter estimation methods employed for deriving the probability distributions of the equivalent puncture diameter for natural gas, crude oil and CO_2 pipelines.

Pipe fluid	**Failure count**	**Parameter estimation method**
Natural gas	1072	MLE with K-S test
Crude oil	816	MLE with K-S test
CO_2	18	MLE with bootstrapping

As can be observed from Table 2, the failure counts for both natural gas and crude oil pipelines exceed the minimum acceptable sample size (i.e. 100) for acquiring a reliable MLE, as concluded in Section 3.2. As a result, their puncture size probability distribution parameters can be estimated confidently using MLE and therefore a further bootstrapping step is not necessary. For CO_2 pipelines on the other hand, the corresponding failure count of 18 is far less than the 100 threshold and therefore the bootstrapping technique is employed to enhance the MLE confidence.

3.3.1 Natural gas pipelines

Figure 5 presents the comparison of the variation of the cumulative failure probability versus equivalent puncture diameter for the field data against the predictions by the Weibull and lognormal Cumulative Distribution Functions (CDFs) for natural gas pipelines. The parameters for the Weibull and lognormal CDFs and the corresponding K-S test results including the null hypotheses, K-S test statistics and critical values are summarised in Table 3. The critical value used for accepting or rejecting the null hypothesis in the K-S test (see Section 2.4) is calculated using Eq. (11).

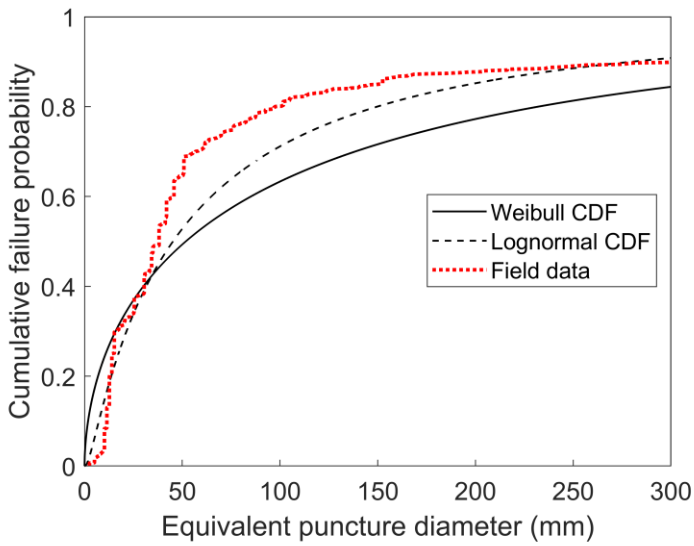

Figure 5. Comparison of the variation of the cumulative failure probability versus equivalent puncture diameter for the field data (data points) against the predictions by the Weibull (solid line) and lognormal (dashed line) CDFs for natural gas pipelines.

Table 3. Summary of the parameters of the predicted Weibull and lognormal CDFs and the corresponding K-S test results including the null hypotheses, K-S test statistics and critical values for natural gas pipelines.

Distribution	Parameter	Value
Weibull	Scale parameter, α	99.475
	Shape parameter, β	0.562
	Null hypothesis	"The data are from a Weibull distribution."
	K-S test statistic	0.213
	Critical value	0.053
Lognormal	Mean, μ	3.812
	Standard deviation, σ	1.422
	Null hypothesis	"The data are from a lognormal distribution."
	K-S test statistic	0.173
	Critical value	0.053

The K-S test results in Table 3 show that the K-S statistics for both Weibull and lognormal CDFs are greater than their critical values, meaning the null hypotheses are rejected (see Section 2.4). This indicates that both CDFs are not statistically good fits to the field data. However, visually, the lognormal CDF appears to be a better fit overall as it more closely mirrors the recorded data throughout. The lognormal CDF can hence be recommended to represent the probability distribution of the equivalent puncture diameter for natural gas pipelines, with the lognormal mean and standard deviation respectively being 3.812 and 1.422.

As can be observed from the recommended lognormal CDF in Figure 5, as the equivalent puncture diameter increases, the rate of increase in the cumulative failure probability generally slows down. The steepest rise is observed when the equivalent puncture diameter increases from 0 to ca. 100 mm, meaning that relatively smaller punctures have a higher probability of occurrence for natural gas pipeline failures as compared to catastrophic ruptures. Specifically, around 70% of such failures are in the form of punctures smaller than 100 m. On the other hand, punctures smaller than 50 mm account for ca. 45% of the failures while ruptures which have an equivalent puncture diameter equal to or over 150 mm, only accounts for 20%.

3.3.2 Crude oil pipelines

Figure 6 shows the comparison of the variation of the cumulative failure probability versus equivalent puncture diameter for the field data against the predictions by Weibull and lognormal CDFs for crude oil pipelines. The parameters of the Weibull and lognormal CDFs and the corresponding K-S test results including the null hypotheses, K-S test statistics and critical values are summarised in Table 4. The critical value used for accepting or rejecting the null hypothesis in the K-S test (see Section 2.4) is calculated using Eq. (11).

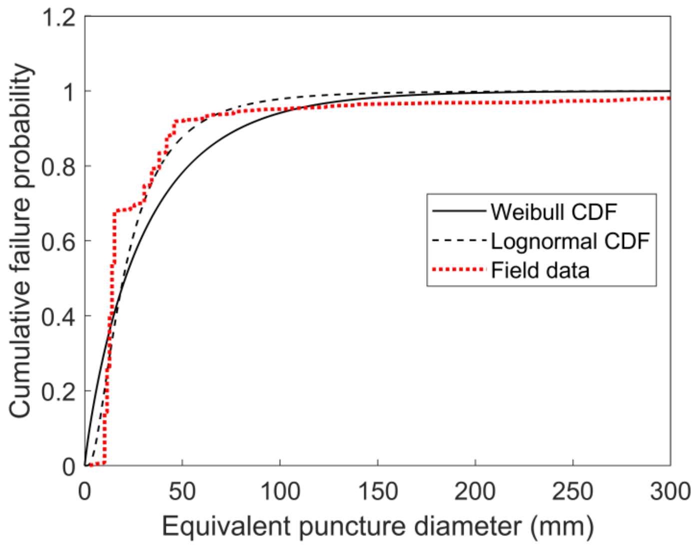

Figure 6. Comparison of the variation of the cumulative failure probability versus equivalent puncture diameter for the field data (data points) against the predictions by Weibull (solid line) and lognormal (dashed line) CDFs for crude oil pipelines.

Table 4. Summary of the parameters of the predicted Weibull and lognormal CDFs and the corresponding K-S test results including the null hypotheses, K-S test statistics and critical values for crude oil pipelines.

Distribution	**Parameter**	**Value**
	Scale parameter, α	0.11831.259
	Shape parameter, β	0.897
Weibull	Null hypothesis	"The data are from a Weibull distribution."
	K-S test statistic	0.284
	Critical value	0.073
	Mean, μ	2.984
	Standard deviation, σ	0.801
Lognormal	Null hypothesis	"The data are from a lognormal distribution."
	K-S test statistic	0.065
	Critical value	0.073

As can be seen from Figure 6, in general both Weibull and lognormal data show good agreement with the field data. However, the K-S test results in Table 4 suggest differently. The K-S statistic for the test of Weibull distribution is greater than the critical value while that for the test of lognormal distribution is otherwise smaller. This means that the null hypothesis for Weibull distribution is rejected while that for lognormal distribution can be accepted. The above indicates that the data are more likely to be drawn from the lognormal CDF. Visually, the lognormal CDF more closely follows the field data covering the most prevalent pipeline failures (equivalent puncture diameter < 100 mm, accounting for ca. 90% of the failures), whence best represents the probability distribution of the equivalent puncture diameter for crude oil pipelines, with the lognormal mean and standard deviation respectively being 2.984 and 0.801.

Comparing to the recommended CDF for natural gas pipelines (see Figure 5), it is obvious that the equivalent puncture diameter is more concentrated at smaller values (< 100 mm) in the corresponding CDF for crude oil pipelines, indicating that small punctures are more frequent in crude oil pipelines. This may be attributed to the fact that in the PHMSA database records, more natural gas pipeline failures are initiated by mechanisms (e.g. excavations and natural forces) that are more likely to result in catastrophic failures.

3.3.3 CO_2 pipelines

In the case of CO_2 pipelines, only 18 equivalent puncture diameter data (see Table 2) are available; far less than the 100 sample size threshold required for obtaining a reliable Maximum Likelihood Estimator (MLE) (see Section 3.2). The bootstrapping method described in Section 2.4 is therefore employed to calculate the MLE confidence interval. To specify, the bootstrapping process which involves creating resampled datasets that have the same size as the original one, is first implemented to generate a number of bootstrap samples for CO_2 pipelines, each containing 18 equivalent puncture diameter data points. The resampling is performed with replacement and thus the resulting bootstrap samples may or may not be identical to the original dataset. To ensure that a sufficiently large number of the possible bootstrap samples are accounted for, 20,000 iterations of the resampling of the original dataset are carried out, corresponding to 20,000 bootstrap samples. Once all the bootstrap samples are obtained, the data points for each bootstrap sample are fitted to both Weibull and lognormal distributions following the distribution fitting process based on Eqs. (5) and (6). The resulting 20,000 MLEs are then segmented into 50 equi-distance bins by their values to generate a distribution of the possible values of the MLE. Using the distribution, the corresponding confidence interval can then be calculated based on Eq. (12). The chosen level of significance, s for obtaining the confidence interval (see Eqs. (13) and (14)) is 0.05, corresponding to 95% confidence.

Figure 7 demonstrates an example bootstrapping result for the MLE of the lognormal mean, μ. As may be observed, the possible values of MLE are normally distributed, varying from ca. 2.5 to 3.5. The arithmetic mean of the MLE (i.e. the mean of the resulting normal distribution) is 2.927. The lower and upper bounds of the 95% confidence interval covering the majority of the possible values are respectively 2.659 and 3.227. Similar distribution of the MLE is also observed for the Weibull scale & shape parameters and the lognormal standard deviation. The corresponding arithmetic means and 95% confidence intervals of the MLEs for these investigated parameters are summarised in Table 5.

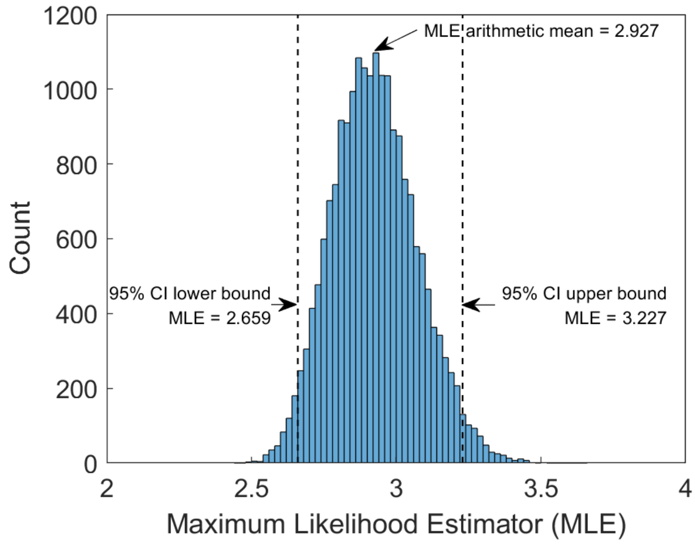

Figure 7. The bootstrapping result for the lognormal mean, μ, showing the Confidence Interval (CI) and arithmetic mean of the MLE.

Table 5. Summary of the arithmetic means and 95% confidence intervals (dashed lines) of the MLEs for the Weibull scale & shape parameters and the lognormal mean & standard deviation.

Distribution	**Parameter**	**95% confidence interval**	**Arithmetic mean**
Weibull	Scale parameter, α	[17.808, 36.976]	26.217
	Shape parameter, β	[1.171, 2.247]	1.559
Lognormal	Mean, μ	[2.659, 3.227]	2.927
	Standard deviation, σ	[0.354, 0.791]	0.593

The confidence intervals summarised in Table 5 essentially represent the tolerable uncertainties in the prediction of the MLE. As such, any CDF characterised by the MLE in these intervals can be used with confidence to represent the probability distribution of the equivalent puncture diameter for CO_2 pipelines. Figure 8 presents the resulting Weibull and lognormal CDF ranges derived from these intervals showing the variation of the cumulative failure probability versus equivalent puncture diameter for CO_2 pipelines. The lower and upper bounds of the ranges are noted in the figure. The corresponding recommended CDFs for natural gas and crude oil pipelines respectively obtained in Sections 3.2.1 and 3.2.2 are also presented for comparison purposes.

As discussed in Section 2.4, given that the sample size involved in deriving the above results is small, the K-S test cannot accurately reflect the goodness-of-fit between the model predictions and the data and hence, both Weibull and lognormal predictions are considered statistically valid for the purpose of this study.

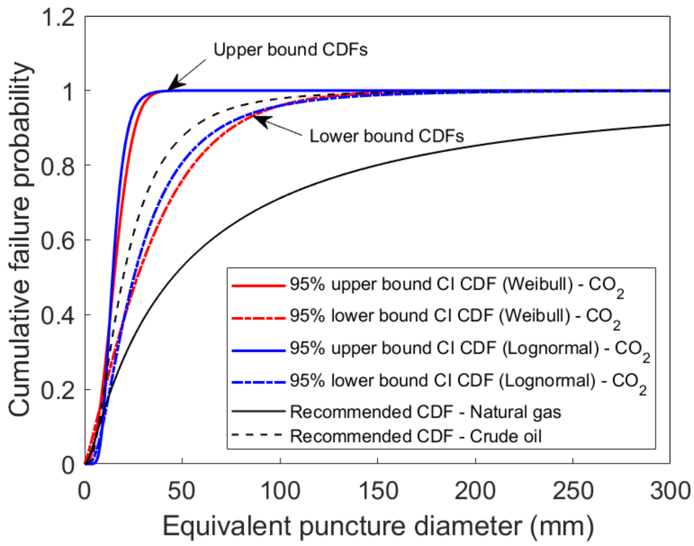

Figure 8. Weibull and lognormal CDF ranges derived from the Confidence Intervals (CIs) summarised in Table 5 showing the variation of the cumulative failure probability versus equivalent puncture diameter for CO_2 pipelines. The corresponding recommended CDFs for natural gas and crude oil pipelines respectively obtained in Sections 3.2.1 and 3.2.2 are also presented for comparison purposes.

As can be observed from both lower and upper bound CDFs in Figure 8, it is estimated that punctures smaller than 50 mm account for ca. 80% to 99% of the failures for CO_2 pipelines. In comparison, only ca. 45% of the failures for natural gas pipelines are in the form of punctures smaller 50 mm. The corresponding number for crude oil pipelines is ca. 90% which is lower than the upper bound value of the estimated range for CO_2 pipelines. The above suggests that small punctures are generally more common in CO_2 pipelines. This may be attributed to the fact that in the PHMSA database records, a major proportion of the CO_2 pipeline failures are resulted from corrosions which are more likely to initiate small but continuous releases rather than catastrophic ruptures. The presence of even small amounts of water (ca. > 650 ppm; Connell, 2005) as a common impurity in CO_2 pipelines also makes them more prone to corrosion. It should be noted that the above conclusions are drawn based on the data currently held in the PHMSA database. The pipelines from which these data were extracted were mostly constructed before 2010, some even dating back to several decades ago when the pipe construction criteria were different from current ones. While corrosion may remain a major failure cause of future CO_2 pipelines due to the presence of corrosive impurities (e.g. water, H_2S) in the CO_2 stream, other failure mechanisms such as external interference and ground movement etc. which are likely to cause more catastrophic releases should not be ignored. With the expected growing number of deployed CO_2 pipelines, failures due to such mechanisms may become increasingly dominant. In addition, the continued improvement in pipeline design standards, as well as cathodic protection techniques will render failures resulting from corrosion and material defects less probable. The above may change the puncture size profile for future CO_2 pipeline failures. Nevertheless, the technique developed here can serve as a powerful tool complimented with the growing wealth of failure data.

Using Figure 8, decision makers can select from a range of credible CDFs based on their subjective preferences to represent the failure probability distribution of the equivalent puncture diameter for the design or/and risk assessment of CO_2 pipelines. For example, the lower bound CDF can be taken as the worst-case scenario CDF for determining the risks associated with CO_2 pipeline failures, as it represents the highest probability occurrence of larger puncture sizes among the possible CDFs. It should be however noted that although the upper bound CDF is considered statistically valid, in practice, it is more reasonable to use the CDFs closer to the lower bound as they cover a wider range of puncture sizes, providing greater safety margins for quantitative risk assessment.

4. Conclusions

The development and application of a statistical analytical technique for determining puncture size probability distribution for pressurised CO_2 pipelines was presented.

A particularly important feature was addressing the pressing dilemma of the relatively small pool of the recorded historical data available for CO_2 pipelines to ascertain a reasonable prediction of their failure risks in the context of CCS operations and how these are compared against those for hydrocarbon pipelines.

The methodology involved fitting the Weibull and lognormal distributions to the puncture diameter data obtained from the PHMSA database using the Maximum likelihood Estimator (MLE) method in conjunction with either Kolmogorov-Smirnov

(K-S) test if the sample size was deemed sufficiently large, or, bootstrapping if not. Whether the sample size being employed was statistically representative was determined by calculating the corresponding MLE quality using Monte Carlo simulation. Using the above method, the puncture diameter probability distribution, expressed as the Cumulative Distribution Function (CDF) was determined for CO_2 pipelines and compared against those for hydrocarbon (natural gas and crude oil) pipelines.

Remarkably, the results obtained indicated that as compared to the hydrocarbon pipelines, CO_2 pipelines are more likely to experience smaller puncture size failures (at least 80% of the failures being in the form of punctures smaller 50 mm), thus resulting in smaller magnitude but more prolonged releases. This directly impacts the preventive and emergency response planning required especially in the case of buried CO_2 pipelines, where small leaks can remain undetected for long periods.

Furthermore, despite being a continuous highly random variable, in practice, the through-wall puncture size in pressurised pipelines is often taken as a discrete variable meaning that only limited range of representative puncture sizes are selected thus compromising the validity of the subsequent quantitative risk assessment performed. The present study fills this important gap by introducing a method for accurately predicting the puncture size failure probability distribution by treating it as a continuous function.

In addition to calculating the puncture size profile governing the subsequent risk following pipeline failure, the proposed method can fully capture the randomness of the through-wall puncture size. This provides a reliable prediction of the expected puncture size following pipeline failure, thus enabling a more accurate evaluation of the failure risk. This is particularly important for pipelines where the failure statistics data available are sparse, such as those for CO_2 pipelines.

It should be noted that, in the present study, the statistical significance of the samples employed to derive the distributions is only assessed based on the sample size. Whether the sample ideally covers, for example, a sufficiently wide range of operating conditions, or the entire range of puncture size is not investigated. Given this, future work should focus on investigating the sample statistical significance based on other sample features, such as the ‘sample quality’.

Acknowledgement

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 884418. The work reflects only the authors’ views, and the European Union is not liable for any use that may be made of the information contained therein.

Data availability statement

The raw data used for this research have been made available at doi:10.5522/04/22015280.

References

Angus, S., Armstrong, B., de Reuck, K.M., 1973. International Thermodynamic Tables of the Fluid State: Carbon Dioxide. Pergamon Press.

Aursand, E., Aursand, P., Berstad, T., Dørum, C., Hammer, M., Munkejord, S.T., Nordhagen, H.O., 2013. CO_2 pipeline integrity: A coupled fluid - structure model using a reference equation of state for CO_2. Energy Procedia 37, 3113–3122. https://doi.org/10.1016/j.egypro.2013.06.197

Barrie, J., Brown, K., Hatcher, P., Schellhase, H., 2005. Carbon dioxide pipelines - A preliminary review of design and risks, in: Greenhouse Gas Control Technologies 7. Elsevier, pp. 315–320. https://doi.org/10.1016/B978-008044704-9/50032-X

Barrie, J., Brown, K., Hatcher, P.R., Schellhase, H.U., 2004. Carbon dioxide pipelines: A preliminary review of design and risks, in: Proceedings of the 7th International Conference on Greenhouse Gas Control Technologies. Vancouver.

Brown, S., Martynov, S., Mahgerefteh, H., Fairweather, M., Woolley, R.M., Wareing, C.J., Falle, S.A.E.G., Rutters, H., Niemi, A., Zhang, Y.C., Chen, S., Besnebat, J., Shah, N., Dowell, N. mac, Proust, C., Farret, R., Economou, I.G., Tsangaris, D.M., Boulougouris, G.C., van Wittenberghe, J., 2014. CO2QUEST: Techno-economic assessment of CO_2 quality effect on its storage and transport. Energy Procedia 63, 2622–2629. https://doi.org/10.1016/j.egypro.2014.11.284

Chaplin, Z., 2015. Data updates to HSE's PIPeline INtegrity model(PIPIN). Buxton.

CIA, 2021. Pipelines [WWW Document]. The World Factbook. URL https://www.cia.gov/the-world-factbook/field/pipelines/ (accessed 6.16.21).

Concawe, 2011. Oil pipelines management group's special task force on oil pipeline spillages (OP/STF-1). Brussels.

Connell, D.P., 2005. Carbon dioxide capture options for large point sources in the midwestern United States: An assessment of candidate technologies. Final report.

Connolly, S., Cusco, L., 2007. Hazards from high pressure carbon dioxide releases during carbon dioxide sequestration processes, in: IChemE Symposium Series No.153. pp. 1–5.

Conover, W.J., 1999. Practical Nonparametric Statistics, 3rd ed. John Wiley & Sons, New York.

Cosham, A., Eiber, R.J., 2008. Fracture propagation in CO_2 pipelines. Journal of Pipeline Engineering 7.

Duncan, I.J., Wang, H., 2014. Estimating the likelihood of pipeline failure in CO_2 transmission pipelines: New insights on risks of carbon capture and storage. International Journal of Greenhouse Gas Control 21, 49–60. https://doi.org/10.1016/j.ijggc.2013.11.005

Efron, B., 1979. Bootstrap methods: Another look at the jackknife. Ann Stat 7, 1–26.

EGIG, 2018. Gas pipeline incidents: 10th report of the European gas incident data group (period 1970-2016).

Eliason, S., 1993. Maximum likelihood estimation: Logic and practice, Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-096. Sage, Newbury Park, CA.

GCCSI, 2021. Global status of CCS 2021.

GCCSI, 2020. Global status of CCS 2020.

Ginos, B.F., 2009. Parameter estimation for the lognormal distribution. Brigham Young University.

Goodfellow, G., Turner, S., Haswell, J., Espiner, R., 2012. An update to the UKOPA pipeline damage distributions, in: International Pipeline Conference. American Society of Mechanical Engineers, pp. 541–547. https://doi.org/10.1115/IPC2012-90247

IEA, 2021. Net zero by 2050: A roadmap for the global energy sector.

IEA, 2010. Energy technology perspectives 2010: Scenarios and strategies to 2050. Paris.

International Energy Agency (IEA), 2021. About CCUS. Paris.

Jensen, M., Schlasner, S., Sorensen, J., Hamling, J., 2014. Subtask 2.19 - Operational flexibility of CO_2 transport and storage. Grand Forks, ND.

Ji, Q., 2020. Foundation and basics, in: Probabilistic Graphical Models for Computer Vision. Elsevier, pp. 11–29. https://doi.org/https://doi.org/10.1016/B978-0-12-803467-5.00007-1

Jung, K., Lee, J., Gupta, V., Cho, G., 2019. Comparison of bootstrap confidence interval methods for GSCA using a Monte Carlo simulation. Front Psychol 10, 2215. https://doi.org/https://doi.org/10.3389/fpsyg.2019.02215

Koch, P., 2008. Equivalent diameters of rectangular and oval ducts. Building Services Engineering Research and Technology 29, 341–347.

Kolmogorov, A.N., 1933. Sulla determinazione empirica di una lgge di distribuzione. Giornale dell'Instituto Italiano degli Attuari 4, 83–91.

Koornneef, J., Spruijt, M., Molag, M., Ramírez, A., Turkenburga, W., Faaij, A., 2010. Quantitative risk assessment of CO_2 transport by pipelines - A review of uncertainties and their impacts. J Hazard Mater 177, 12–27.

Kruse, H., Tekiela, M., 1996. Calculating the consequences of a CO_2 pipeline rupture. Energy Convers Manag 37, 1013–1018. https://doi.org/10.1016/0196-8904(95)00291-X

Long, J.S., Freese, J., 2006. Regression Models for Categorical Dependent Variables using Stata. Stata Press.

Lydell, B.O.Y., 2000. Pipe failure probability-the Thomas paper revisited. Reliab Eng Syst Saf 68, 207–217. https://doi.org/10.1016/S0951-8320(00)00016-8

Lyons, C.J., Goodfellow, G.D., Haswell, J. v, 2020. UKOPA pipeline product loss incidents and faults report (1962-2018). Ambergate, UK.

Mahgerefteh, H., Atti, O., 2006. Modeling low - temperature - induced failure of pressurized pipelines. AIChE Journal 52, 1248–1256. https://doi.org/10.1002/aic.10719

Mahgerefteh, H., Brown, S., Denton, G., 2012. Modelling the impact of stream impurities on ductile fractures in CO_2 pipelines. Chem Eng Sci 74, 200–210. https://doi.org/10.1016/j.ces.2012.02.037

Mahgerefteh, H., Denton, G., Rykov, Y., 2008. Pressurised CO_2 pipeline rupture, in: IChemE Symposium Series No.154.

Mahgerefteh, H., Jalali, N., Fernandez, M.I., 2011. When does a vessel become a pipe AIChE Journal 57, 3305–3314. https://doi.org/10.1002/aic.12541

Massey Jr, F.J., 1951. The Kolmogorov-Smirnov test for goodness of fit. J Am Stat Assoc 46, 68–78.

Medina, H., Arnaldos, J., Casal, J., Bonvicini, S., Cozzani, V., 2012. Risk-based optimization of the design of on-shore pipeline shutdown systems. J Loss Prev Process Ind 25, 489–493. https://doi.org/10.1016/j.jlp.2011.12.005

Moe, A.M., Dugstad, A., Benrath, D., Jukes, E., Anderson, E., Catalanotti, E., Durusut, E., Neele, F., Grunert, F., Mahgerefteh, H., Gazendam, J., Barnett, J., Hammer, M., Span, R., Brown, S., Munkejord, S.T., Weber, V., 2020. A trans - European CO_2 transportation infrastructure for CCUS: Opportunities & challenges.

Neill, S.P., Hashemi, M.R., 2018. Ocean modelling for resource characterization, in: Fundamentals of Ocean Renewable Energy. Academic Press, pp. 193–235.

Nesic, S., 2012. Effects of multiphase flow on internal CO_2 corrosion of mild steel pipelines. Energy & Fuels 26, 4098–4111. https://doi.org/10.1021/ef3002795

Nielsen, M.A., 2011. Parameter estimation for the two-parameter Weibull distribution.

PHMSA, 2020. Source data [WWW Document]. URL https://www.phmsa.dot.gov/data-and-statistics/pipeline/source-data

Pohanish, R.P., Greene, S.A., 1996. Hazardous Materials Handbook. Van Nostrand Reinhold, New York.

Ringrose, P.S., 2018. The CCS hub in Norway: Some insights from 22 years of saline aquifer storage. Energy Procedia 146, 166–172. https://doi.org/10.1016/j.egypro.2018.07.021

Rusin, A., Stolecka, K., 2015. Reducing the risk level for pipelines transporting carbon dioxide and hydrogen by means of optimal safety valves spacing. J Loss Prev Process Ind 33, 77–87. https://doi.org/10.1016/j.jlp.2014.11.013

Ryan, T.P., 2007. Modern Engineering Statistics. John Wiley & Sons.

Tsagkanos, A., 2008. The bootstrap maximum likelihood estimator: The case of logit.

Applied Financial Economics Letters 4, 209–212. https://doi.org/https://doi.org/10.1080/17446540701604309

Vitali, M., Zuliani, C., Corvaro, F., Marchetti, B., Tallone, F., 2022. Statistical analysis of incidents on onshore CO_2 pipelines based on PHMSA database. J Loss Prev Process Ind 77, 104799. https://doi.org/10.1016/j.jlp.2022.104799

Wei, S., Li, N., 2019. Bootstrap estimation for Weibull distribution parameters based on small sample and censored condition. Statistics & Decision 34–37.