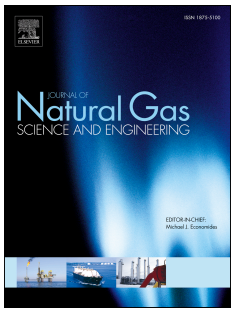
Accepted Manuscript



Crater formation by the rupture of underground natural gas pipelines: A probabilistic-based model

Rafael Amaya-Gómez, J. Giovanni Ramírez-Camacho, Elsa Pastor, Joaquim Casal, Felipe Muñoz

PII: S1875-5100(18)30126-4

DOI: 10.1016/j.jngse.2018.03.011

Reference: JNGSE 2502

To appear in: *Journal of Natural Gas Science and Engineering*

Received Date: 28 December 2017

Revised Date: 1 March 2018

Accepted Date: 14 March 2018

Please cite this article as: Amaya-Gómez, R., Ramírez-Camacho, J.G., Pastor, E., Casal, J., Muñoz, F., Crater formation by the rupture of underground natural gas pipelines: A probabilistic-based model, *Journal of Natural Gas Science & Engineering* (2018), doi: 10.1016/j.jngse.2018.03.011.

- A model to predict crater dimensions given a LOC in underground pipelines is proposed
- The model implements 57 real accidents of natural gas underground pipelines
- The approach proposes worst, mean and less severe scenarios to support decision-making
- Some applications in Domino effect scenarios and Right-of-Way distances were discussed

Crater formation by the rupture of underground natural gas pipelines: A probabilistic-based model

Rafael Amaya-Gómez[a,*], J.Giovanni Ramírez-Camacho[b], Elsa Pastor[b], Joaquim Casal[b], Felipe Muñoz[a]

[a]*Chemical Engineering Department, Universidad de los Andes, Cra 1E No. 19A-40, Bogotá, Colombia*
[b]*Centre for Technological Risk Studies (CERTEC), Barcelona East School of Engineering (EEBE), Universitat Politècnica de Catalunya, BarcelonaTech (UPC), Eduard Maristany 10-14, 08019-Barcelona, Catalonia, Spain*

Abstract

Parallel and crossing pipelines are frequently implemented due to land-use restrictions and their ease of operation and maintenance. Given the proximity of these pipelines and the hazardousness of the substances they transport, an eventual Loss of Containment (LOC) in a parallel/crossing corridor can lead to a domino effect that should be considered in Quantitative Risk Analysis (QRA). For underground pipelines, this LOC is accompanied by a formation of a crater, which can uncover adjacent pipelines triggering a domino effect scenario to take place. This paper aims to develop a model to predict feasible crater dimensions (i.e., width and depth) from a LOC in underground natural gas pipelines using operational and structural parameters. For this purpose, a recent review of 57 underground natural gas pipeline accidents were considered in a probability-based approach once data was processed. This approach initially predicts the Width-to-Depth crater ratio (WD) using a multivariate regression. Then feasible crater dimensions were determined using the regression prediction interval and the width-depth joint probability function, which is approximated with a Gaussian copula. This approach proposes a worst, mean and less severe scenarios to support decision-making processes regarding parallel or crossing underground natural gas pipelines with a LOC. Besides the identification of domino effect scenarios, this information can be used to support pipeline segmentation for risk analysis or even to support Right-of-Way (ROW) definition during pipeline installation.

Keywords: Crater formation, Crater model, Underground pipeline, Domino effect, Natural gas.

1. Introduction

Pipelines are the safest and most cost-effective means of transporting hydrocarbons such as natural gas over long distances. These pipelines are frequently installed in a parallel or crossing distribution, and they include other services such as water distribution or electric wiring (Casal et al., 1995; Majid & Mohsin, 2013). This is due to land-use restrictions and the fact that pipelines distributed this way make operation and maintenance easier (Shi et al., 2012). However, these

*Corresponding author, Tel. (+57-1) 3394949 Ext.3095
Email address: r.amaya29@uniandes.edu.co (Rafael Amaya-Gómez)

systems undergo different degradation processes related to their operation and surrounding conditions that can lead to a possible Loss of Containment (LOC), which in turn, can severely affect the population or environment around the pipelines. Because of the proximity of the parallel and crossing pipelines, a LOC can trigger a domino effect of accidents in the adjacent pipelines, thus aggravating the effects of the initial event (Ramírez-Camacho et al., 2015). Therefore, the domino effect should be included in the Quantitative Risk Analysis (QRA) of pipelines to reduce uncertainty in the risk calculation.

For underground pipelines, a crater is formed because of LOC, which can expose adjacent pipelines to a possible domino effect event. In the case of natural gas pipelines, this crater can be caused by the sudden and catastrophic rupture of a high-pressure pipeline (e.g., due to a mechanical failure), or by the explosive ignition of the gas accumulated from a leak in the pipeline; for instance, due to weld cracks, corrosion defect or an accidental perforation. Once the crater is formed by the rupture of the initial pipe, or source pipe (SP), the likelihood of a domino effect event depends on whether or not a second pipe, or -target pipe (TP), is within the crater's boundaries. If the TP was not exposed in the initial rupture, then the TP would not suffer any damages because of the protecting soil that surrounds it. If the TP lies within the crater dimensions (totally or partially), it can fail because of two main scenarios. The first one is related to the overpressure from the SP explosion; this could affect the TP totally or partially, leading to its failure. The other scenario is associated with a thermal failure when the initial rupture did not affect the TP. This thermal failure occurs due to the impingement of a fire that would spring from the gas leaking from the SP. Based on the aforementioned scenarios, given a LOC, the probability of a pipeline being exposed should be associated with the crater dimensions and the spatial distribution of the parallel/crossing pipeline corridor.

To prevent such domino effect scenarios, safety distances between parallel and crossing pipelines and their surroundings have been proposed in standard or code practices such as that reported for natural gas and water parallel pipelines by the Energy Commission of Malaysia (clearance of > 300 mm) or the ASMEB31.8 (clearance of > 6 in for an underground structure) (Mohsin et al., 2014; Shi et al., 2012). Other approaches like Silva et al. (2016) suggest a10 m separation for underground pipelines based on an analysis of historical accidents, and PEMEX (2009) recommend a minimum separation of 1 m in the same ditch. Nevertheless, as was pointed out in Alileche et al. (2015), safety distances in domino effect scenarios depend on the escalating effects from the primary system, so these safety distances should actually consider the minimum separation at which escalating events are avoided. In the case of natural gas, the possible final events are Vapour Cloud Explosion (VCE), Flash Fire, and Jet Fire (Ramírez-Camacho et al., 2017), so the safety distance would be delimited by: the extension of the flame envelope (Flash Fire); the flame length and its direction (Jet Fire); or the overpressure associated with the explosion energy (VCE) (Alileche et al., 2015). Consequently, further analysis is required to estimate a safety distance as the approaches reported in Haklar (1997), Haklar & Dresnack (1999), Sklavounos & Rigas (2006), Mohsin et al. (2014) show for Jet Fire scenarios, which are the most probable events (Bubbico et al., 2016).

Based on the aforementioned, to estimate safety distances in parallel or crossing underground pipelines, whether or not the crater formed by a LOC can expose the other pipe should be determined. Once this information is known, approaches like Haklar (1997); Haklar & Dresnack (1999) can be implemented to support decision-making in domino effect scenarios. Currently, the work aiming to predict the crater dimensions from a pipeline LOC is limited. Therefore, the central problem here is to create a model that adequately predicts the crater dimensions. There are some approaches such as the Gasunie, Battelle, Advantica or the NEN 3651 models (Leis

et al., 2002; Acton et al., 2010; NSI, 2012) that were developed based on experimental results which depend on soil properties from the pipeline location. There are other approaches that use TNT explosion models based on a set of experiments or FEM (Finite Elements Modeling) (Ambrosini et al., 2002; Ambrosini & Luccioni, 2006; Krishna et al., 2016). Finally, there is an approach called the Accident-based model recently proposed by Silva et al. (2016) which implements a polynomial regression on 17 underground accidents.

The aforementioned models have some limitations that may prevent their use or affect the final predictions. The Gasunie, Battelle, Advantica, NEN 3651, and the Accident-based models all focus on crater width predictions, but only the Gasunie and Batelle models incorporate crater depth calculations. This prevents the Advantica, NEN3651, and Accident-based approaches from being eligible for use in crossing pipelines. These five models are deterministic and they require a soil classification, which affects their predictions. Hence, they have important sources of uncertainty. Regarding the TNT models, these approaches may significantly differ from the real crater dimensions because of the ”hydraulic mining hose” described by Peekema (2013). According to Peekema (2013), the final crater dimensions will be greater because of the fluid pushed out from the broken pipeline lightly scours the soil. Therefore, TNT models would underestimate the real crater dimensions, representing a threat for crossing or parallel corridors. To the best of the authors' knowledge, no other model that deals with crater formation caused by ruptures or leakages of underground pipelines appears to be publicly available.

Bearing in mind these limitations, the objective of this paper is to propose a probability-based approach to model the most likely crater dimensions using information about design, operation, and installation of a pipeline, which is known. For this purpose, a historical review of 90 natural gas accidents was considered (Ramírez-Camacho et al., 2017) focusing on parameters such as the pipeline wall thickness, diameter, burial depth, and operating pressure. After data processing, 57 of these 90 records are implemented in a multivariate regression to predict the Width-to-Depth crater ratio (WD) and determine its prediction interval. Using the WD prediction interval and the joint width-depth probability density function, which is obtained from a Gaussian copula, a maximum and mean criterion were proposed to determine the less severe, mean and worst scenarios for the crater dimensions.

This paper is organized as follows: Section 2 reviews the available pipeline crater models. Section 3 presents the probability-based approach. Section 4 discusses the main results related to the crater predicted dimensions, and Section 5 presents some concluding remarks.

2. Review of pipeline crater approaches

This section describes the five available models which seek to predict the crater dimensions due to a failure of a buried pipeline: i) Gasunie, ii) Batelle, iii) NEN 3651, iv) Advantica and v) Accident-based model. The assumptions and limitations of each model are described below; a summary of each model is presented in Table 1. This table illustrates the type of information that each model implements. Among the information reported, there are structural parameters (e.g., pipeline diameter), operation parameters, soil parameters, installation parameters (e.g., burial depth), properties of the transporting fluid (e.g., density), and the output from each model.

2.1. Gasunie model

The Gasunie model has three main assumptions. They are that: i) the ends of pipes are disconnected because of the rupture; ii) the crater is initially formed by removing the soil near

Table 1: Summary of models to predict crater dimensions

Model	Approach	Struct.	Op.	Soil	Installation	Gas	Output	Ref.
Gasunie	Experimental	✓	✗	✓	✓	✗	W&D	(Leis et al., 2002)
Batelle	Experimental	✓	✗	✓	✓	✓	W&D	(Leis et al., 2002)
Advantica	Experimental	✓	✓	✓	✗	✗	W	(Acton et al., 2010)
NEN3651	Experimental	✓	✓	✗	✓	✗	W	(NSI, 2012)
Accident-based	Historical analysis	✓	✓	✓	✓	✓	W	(Silva et al., 2016)
Proposed model	Probabilistic	✓	✓	✗	✓	✓	W&D	-

W: Width, W&D: Width and Depth, Struct.:Structural parameters, Op.: Operating parameters

the pipeline and then the axial length is increased due to soil erosion by the fluid; and iii) the soil has homogeneous properties (Silva et al., 2016). This model was proposed by the Delft Hydraulics Laboratory and N. V. Nederlandse Gasunie Company, considering an elliptical cross-sectional as the one illustrated in Fig. 1. In this case, D is the crater depth, α_1 is the crater angle at the ground surface, α_2 is the crater angle at $D/2$, a and b are the ellipse shape parameters, and W is the crater width.

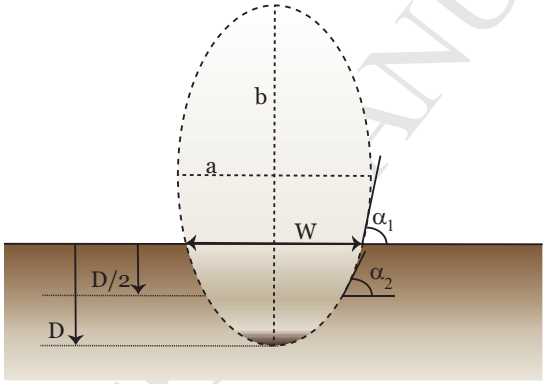


Figure 1: Gasunie model Scheme.

If the rupture occurs at the top of the pipeline, this model assumes that the depth of the crater is independent of the soil type. The crater depth is then determined based on the pipeline diameter D_p and the cover depth D_c as $D = D_p + D_c$, where the cover depth corresponds with the distance from the ground level to the top of the pipeline.

In case of a guillotine rupture, the soil type and moisture affect the depth of the crater (Leis et al., 2002); therefore, a soil parameter ν –reported for different soils in Leis et al. (2002) and Luo et al. (2009)– and a function $R(\nu)$ are implemented as follows:

$$R(\nu) = 0.28 + 0.62(5 - \nu) - 0.07(25 - \nu^2) \tag{1}$$

The crater depth is then determined as follows:

$$D = \begin{cases} 4.3D_p + D_c, & \text{if } \nu \leq 0.6 \\ \dfrac{R(\nu)D_p}{0.3} + D_c, & \text{if } 0.6 < \nu < 2 \\ 2.2D_p + D_c, & \text{if } \nu \geq 2 \end{cases} \tag{2}$$

The crater angles can be determined by the soil parameter ν as follows:

$$\alpha_1 = \tan^{-1}(\nu + 1)$$
$$\alpha_2 = \tan^{-1}\left[\left(\frac{2.8 + 0.5\nu}{10}(\nu + 1)\right)\right] \tag{3}$$

These angles are used to determine the ellipse shape parameters (i.e., a and b) by solving the following equations simultaneously:

$$\tan\alpha_1 = \frac{b}{a}\sqrt{\left(\frac{b}{b-D}\right)^2 - 1}, \quad \tan\alpha_2 = \frac{b}{a}\sqrt{\left(\frac{b}{b-0.5D}\right)^2 - 1} \tag{4}$$

Finally, the width of the crater is determined as follows:

$$W = 2a\sqrt{1 - \frac{(b-D)^2}{b^2}} \tag{5}$$

As pointed out by Luo et al. (2009), the Gasunie model is based on empirical correlations that omit the pipeline operating pressure. So, this model could under or overestimate the crater dimensions depending on high or low-pressure pipelines.

2.2. Batelle model

The Batelle model is an improvement of the Gasunie model in which the width of the crater is determined as in a chemical explosion (Silva et al., 2016). This model assumes that a guillotine rupture caused the crater and that it has two cross-sections, one of them following that reported in the Gasunie model to determine the crater depth. The width of the crater is determined by:

$$W = 2\sqrt{\frac{D_p\left(D_c + \frac{D_p}{2}\right)}{u_{kr}}u_x - \left(D_c + \frac{D_p}{2}\right)^2} \tag{6}$$

where u_{kr} is the critical gas velocity and u_x is the velocity of the explosive gases. The velocity of the explosive gases is obtained from the gas (ρ) and soil (ρ_{soil}) densities, the speed of sound (c) and the specific heat ratio (γ) as follows:

$$u_x = \sqrt{\frac{\rho c^2}{3\rho_{soil}(\gamma^2 - 1)}} \tag{7}$$

The critical gas velocity u_{kr} is commonly taken as 2.54 *m/s* (Leis et al., 2002); however, Silva et al. (2016) have observed that an underestimation of the crater width can be prevented if a mean critical velocity of 1.8542 m/s is used instead. This model enhances the dimensions obtained from the Gasunie model by including the specific heat ratio, the critical velocity at which the soil can be removed, and the gas and soil densities. Nevertheless, this model uses the same qualitative soil classification as in the Gasunie model for its depth calculation, which is an important source of uncertainty.

Table 2: Crater width regressions for the Advantica Model reported by Silva et al. (2016)

Soil type	Pressure (bars)	Diameter (in)	Regression
Sandy soil	20	*	$W = 0.3999D_p + 5.469$
	40	≤ 12.8	$W = -10^{-14}D_p + 10.875$
		> 12.8	$W = 0.3934D_p + 5.7275$
	60	≤ 24.0	$W = 0.0278D_p + 14.6060$
		> 24.0	$W = 0.3927D_p + 5.80$
	80-150	*	$W = 0.39999D_p + 5.4695$
Clay soil	20	≤ 36.1	$W = 0.0237D_p + 6.0135$
		> 36.1	$W = 0.093D_p + 3.4989$
	40	≤ 36.0	$W = 0.0258D_p + 5.9839$
		> 36.0	$W = 0.1445D_p + 1.6881$
	60	≤ 24.0	$W = 0.0237D_p + 5.9989$
		> 24.0	$W = 0.2437D_p + 0.5545$
	80	*	$W = 0.3148D_p + 0.1522$
	100	*	$W = 0.3710D_p + 0.0842$
	150	≤ 12.6	$W = -0.0075D_p + 5.5811$
		> 12.6	$W = 0.3562D_p + 1$
Mixed soil	≤ 80	≤ 36.1	$W = 0.0244D_p + 10.276$
		> 36.1	$W = 0.1946D_p + 4.0742$

*Any diameter

2.3. Advantica model

This model is based on experimental results obtained from Acton et al. (2010) in which twelve natural gas releases were reported with pressures ranging from 20 to 150 bars and diameter between 25 to 80 mm in adjacent buried pipelines. Considering this information, Silva et al. (2016) determined linear regressions regarding the crater width shown in Table 2.

2.4. NEN 3651 & Accident-based model

The NEN 3651 was developed by the Nederlands Normalisatie Instituut to provide requirements for elements reported in NEN 3650. The available information about this standard is limited and only the radius of the corresponding crater for a guillotine rupture is reported, which includes the internal pressure of the pipeline in bars. The equivalent width of the crater is as follows:

$$W = 2\sqrt{0.64(D_p^3P)^{2/3} + 0.65(D_p^3P)^{1/3}) - 0.83D_c^2} \tag{8}$$

Finally, the Accident-based model proposed by Silva et al. (2016) uses 17 reported accidents to obtain a multivariate polynomial regression of the width of the crater. This regression considers parameters reported in the previous models such as the specific heat ratio, the density of the soil, and the operating pressure:

$$W = 40.795 + 0.382D_p - 0.068P + 4.844D_c - 10.069\gamma - 0.020\rho_{soil} \tag{9}$$

3. Probabilistic-based approach based on natural gas accidents

The methodology is divided in three phases (Fig. 2). Initially, the records obtained in Ramírez-Camacho et al. (2017) are processed. These records are then used to predict the WD ratio. Finally, the feasible crater dimensions are determined using the width-depth joint probability and a maximum or mean value approach depending on the pipeline diameter. Each of these phases will be described in detail below.

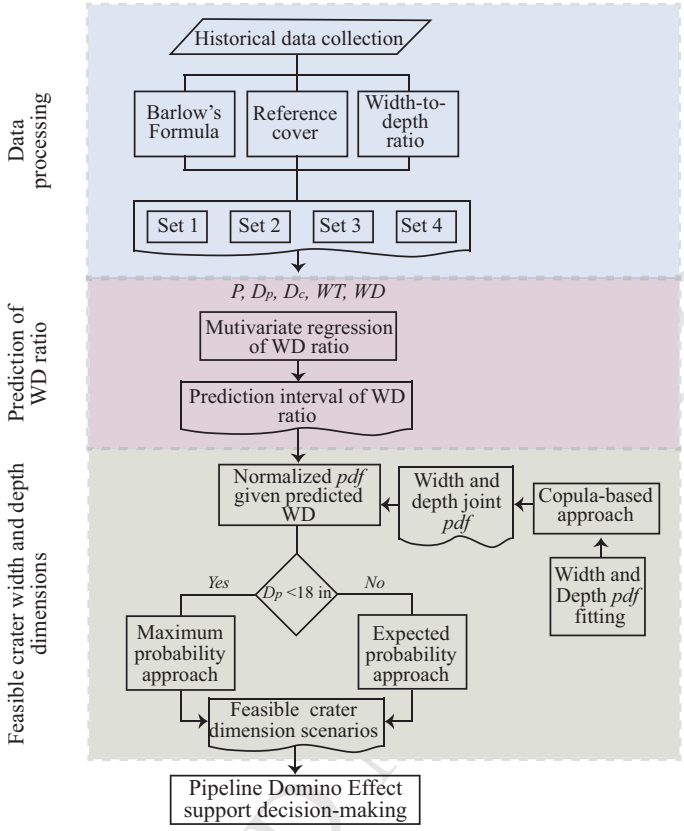


Figure 2: Overall methodology

3.1. Data processing

Some of the records gathered did not include information about all features; for instance, 29 records did not report the width or the depth of the crater, and another 9 were included based on their crater-approximated dimensions. In fact, only 28 of the 90 records included information regarding the pipeline diameter, wall thickness, operating pressure, burial depth, and the crater dimensions (width and depth). Therefore, three different strategies (Barlow's formula, a reference cover and a WD ratio) were considered to complete the *a priori* unavailable information.

Barlow's formula was implemented to obtain an approximation of the pipeline's internal pressure or the wall thickness. This formula associates the pipeline's internal pressure P, the allowable stress, which is assumed to be 72% of the yield strength σ_Y (Zhao et al., 2012), the wall thickness WT, and the outside diameter D_p as follows (Stewart, 2016, p.834):

$$P = \frac{2 \cdot (72\%\sigma_Y) \cdot WT}{D_p} \tag{10}$$

From 11 records that did not include the operating pressure or the wall thickness, only 3 of them could be included with Barlow's formula. This is because 8 of 11 records did not provide information regarding the pipe material; thus, their yield strengths were unknown. This strategy

could not gather more records because 24 other records omitted the pipeline operating pressure and the wall thickness. So, this formula could not be implemented to approximate the missing information.

A reference cover depth was implemented to avoid the use of the burial depth, which is a feature that 40 records did not report. In this paper, a minimum cover depth of 3 ft. (0.9144 m) was implemented, which is commonly reported for transmission pipelines (ASME, 2002, 2004) despite that the cover depths depend on several variables such as the soil properties and pipeline route. Therefore, the cover depths were translated to this reference considering that the difference between the burial and the cover depths is the pipeline diameter (Fig. 3). With the burial depth of the pipeline defined as b_d and the pipeline diameter as D_p, the cover depth is calculated as $D_c = b_d - D_p$. Considering the difference between this cover depth and the reference $D_c^r = 0.9144 - D_c$, the adjusted depth with this reference cover is given by $D^* = D + D_c^r$. It was assumed that those records that did not report their cover depth have a reference cover of 0.9144 m. This is not a strong assumption considering that the cover depth can be as many as two times the pipeline diameter (Mokhtari & Alavi, 2015). The records reported herein have a mean relation of $1.8D_p$ and more than 70% reported a cover depth less or equal to $2D_p$. Nonetheless, a parallel evaluation was implemented without using this transformation to determine if this assumption affects the final results significantly.

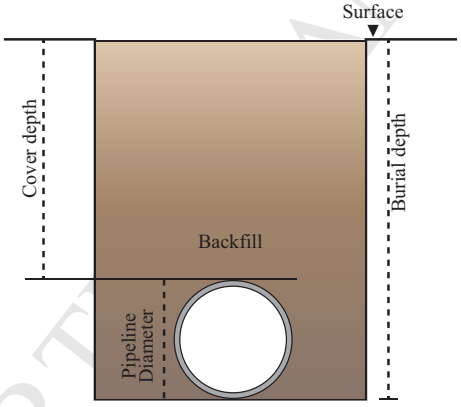


Figure 3: Cover and burial depth scheme

Finally, a WD crater ratio was used to complete the missing information in the 29 records that did not report the crater's width or depth, bearing in mind the assumption that the crater was symmetrical. For this purpose, the models reported in Section 2 could be implemented; nonetheless, there are several unknown parameters such as the soil parameters (Leis et al., 2002) that prevent these models from being taken into account. The available information about the WD ratio is scarce and only historical records of pipeline ruptures are reported. For instance, McGillivray & Wilday (2009) reviewed the dimensions of the craters caused by historical gas pipeline ruptures in the UK and found that the crater width varies from 1.7 to 33 m, whereas the crater depth ranged from 1.7 to 7.6 m. They reported a total of 22 records where their WD ratios have a mean value of 3.5 and a median near to 3.33. These results suggest that the width and depth of the crater have a ratio near to 3:1 (i.e., $W/D = 3$), so we assumed this WD ratio to complete these 29 records. As in the cover depth assumption, a parallel assessment was carried out without this WD relation to determine if the final results are affected considerably. For the

sake of simplicity, these sets will be referred to from here on as in Table 3.

Table 3: Dataset Classification

Dataset	Reference Cover Depth	WD Relation
Set 1	No	No
Set 2	Yes	No
Set 3	No	Yes
Set 4	Yes	Yes

3.2. Prediction of WD ratio

3.2.1. Multivariate regression approach

Based on the records obtained from the data processing, a multivariate parametric regression is proposed to model the WD ratio. Additionally, a logarithmic regression was used in this approach because some features, such as the WD, had wider ranges (i.e., greater than 10:1). These features follow nonlinear performances, and the logarithmic scale helped stabilize the variance (Maindonald & Braun, 2010). Recall, that if Y denotes the response and X_i the variables evaluated, the logarithmic regression would have the general form of Eq. 11, which is equivalent to a general power relation. Here β_0 is the intercept, β_i are the regression coefficients obtained from a least squares approach and ϵ is the associated error.

$$\log Y = \beta_0 + \sum_i \beta_i \log X_i + \epsilon \tag{11}$$

Variables are selected to determine which predictors (i.e., features) better described the WD final response. For this purpose, a mixed approach that combines forward and backward selection was considered. This selection starts with no variables and then variables are added one by one to provide a better fit. If the p-value of one of the variables rises above a given threshold, then the variable is removed. This procedure is repeated until all variables in the model have an adequate p-value and the external variables, if included in the model, would affect the regression capabilities (James et al., 2013).

The model fit was initially determined based on two well-known numerical measures: the Residual Standard Error (RSE) and the adjusted R^2 statistic. The first one is an estimate of the standard deviation of the error based on the Residual Sum of Squares (RSS), and the latter is a measure of the linear relationship between the variables and the response. Besides these numerical measures, the Akaike and Bayesian Information Criteria (AIC and BIC) were implemented. These selection criteria are numerical values that are by themselves meaningless, but are used to compare competing models or regressions and they are commonly used instead of other tests (see for instance Posada & Buckley (2004)). AIC ranks the information loss and the unaccountable information, whereas BIC is based on a log likelihood function incorporating the deviance of the model fit and penalizing for additional parameters (Stancescu, 2014). Both criteria prefer those models with the lowest results. Additional information about these selection criteria can be found elsewhere (see for instance Konishi & Kitagawa (2008)).

Set $k = p + 1$ where p denotes the number of fitted parameters and let n be the sample size for the regression. For small samples (i.e., $n/k < 40$) the modified version of AIC, which is calculated based on p by the RSS, is recommended (Symonds & Moussalli, 2011). The modified versions of the AIC and BIC implemented are presented below:

$$AIC_c = \left\{ n \left[\log\left(\frac{RSS}{n}\right) \right] + 2k \right\} + n + n\log(2\pi) + \frac{2k(k+1)}{n-k-1}$$
$$BIC_c = n \left[\log\left(\frac{RSS}{n}\right) \right] + k\log n + n + n\log(2\pi) \tag{12}$$

To evaluate the robustness of the regressions, their linearity, independence, homoscedasticity, and normality assumptions (Yan & Su, 2009) were evaluated using the regression diagnostic plots (i.e. residual and quantile-quantile (QQ) plots). The linearity assumption indicates that the response and the predictors have a linear, additive relationship; the independence assumption establishes statistical independence of the regression errors; the homoscedasticity assures a constant variance of the errors; and the normality assumption means that these errors are normally distributed. Additionally, an outlier and leverage diagnosis was implemented to evaluate if any register should be removed. The first identifies records that may differ from the bulk of the data evaluated, whereas a leverage observation has an unusual value in the independent variables (i.e., predictor). To identify possible outliers, the studentized residuals plot was considered, which corresponds with the residual errors divided by their standard error. According to James et al. (2013), the observations whose studentized residuals are greater than 3 (absolutely) are possible outliers. Regarding the leverage diagnosis, the leverage statistic h_i (between $1/n$ and 1) was calculated to identify observations with higher leverage. For this purpose, the leverage plot included in the diagnostic plots was considered.

3.2.2. WD Prediction interval

Based on the regression coefficients, the confidence interval for the coefficients and the prediction interval were determined. According to Rencher & Schaalje (2008), the $100(1-\alpha)\%$ confidence interval of the *jth* regression coefficient β_j is given as follows:

$$\tilde{\beta}_j \pm t_{\alpha/2,n-k-1} RSE \sqrt{g_{jj}} \tag{13}$$

where $t_{\alpha/2,n-k-1}$ is the Student's t inverse cumulative distribution at $\alpha/2$ with $n-k-1$ degrees of freedom, $\sqrt{g_{jj}}$ is the jth diagonal-element of $(X'X)^{-1}$ and $RSE = \sqrt{\sum_i^n \frac{(y_i - \hat{y}_i)^2}{(n-k-1)}}$, where y_i is the *ith* reported WD and $\hat{y}_i$ is the *ith* predicted WD. Here X is a $n \times k$ matrix that corresponds with the fitted observations for each of the k predictors.

Consider the future response $\tilde{y}_o$, which is obtained from the regression coefficients evaluated at a $(k+1)$ column vector X_o with predictors $1, x_0, x_1, \ldots, x_k$. So, the prediction interval is given by (Rencher & Schaalje, 2008):

$$\tilde{y}_o \pm t_{\alpha/2,n-k-1} RSE \sqrt{1 + X_o'(X'X)^{-1}X_o} \tag{14}$$

In this paper, the prediction intervals were determined using the *predict* function in the open source R-project.

3.3. Feasible crater dimensions

3.3.1. Width and depth fitting probabilities

Initially, the width and depth probability density functions (*pdf*) were determined based on the information gathered from the historical review. These density functions can be determined

by Goodness of Fit tests such as Chi-Square, Kolmogorov Smirnov or Anderson Darling. However, these approaches usually depend on predefined parameter values and not on estimated parameters from the available data (Stancescu, 2014). Therefore, an information criteria approach was implemented instead, where the number of parameters of the fitted distribution is considered to select the simple and consistent model. In this paper, a BIC was used following the Sheppard (2012) function, and a copula was implemented to approximate the width-depth joint *pdf* based on the width and depth marginal *pdf*s.

3.3.2. Copula-based approach

Overall, copulas are functions that couple multivariate distributions from their marginal distributions. These functions have uniform one-dimensional margins, and they are invariant under any monotone, increasing transformations on the marginal distributions (Montes-Iturrizaga & Heredia-Zavoni, 2015). These functions represent a parametric approach for modeling the dependent structure in joint distributions of random variables (Escarela & Hernández, 2009) as shown in several areas such as finance, stock market, or even mortality models (Frees & Valdez, 1998). Consider the vector $\mathbb{X}$ of m random variables with marginal distribution functions $F_i(x_i), i = 1, \ldots, m$. Let the set of transformations $U_i = F_i(x_i)$ define a dependent, uniformly distributed vector of m random variables $\mathbb{U}$ on $[0, 1]^m$. Based on Sklar's theorem, if the marginal distributions $F_i(x_i)$ are continuous, then the $\mathbb{X}$ joint probability distribution function can be determined as follows (Montes-Iturrizaga & Heredia-Zavoni, 2015):

$$F(\mathbb{X}) = C(F_1(x_1), \ldots, F_m(x_m)) = C(U_1, \ldots, U_m) \tag{15}$$

where $C(\mathbb{U}) = C(U_1, U_2, \ldots, U_m)$ is the copula distribution. Moreover, the joint *pdf* of $\mathbb{X}$ can be determined using the marginal *pdf* of all the random variables ($f_i(x_i)$ for $i = 1, \ldots, m$) and the copula density function $c(F_1(x_1), \ldots, F_m(x_m))$ as:

$$f(\mathbb{X}) = c(F_1(x_1), \ldots, F_m(x_m)) \prod_i^m f_i(x_i) \tag{16}$$

A detailed introduction to the theory and copula types can be found in Nelsen (2007); Escarela & Hernández (2009); Montes-Iturrizaga & Heredia-Zavoni (2015).

To determine the width-depth joint distribution, the Gaussian copula was used due to its similarity to the bivariate Gaussian distribution. This copula follows this general form:

$$C_r(v_1, v_2) = \Phi_2[\Phi^{-1}(v_1), \Phi^{-1}(v_2)], (v_1, v_2)^T \in (0, 1)^2 \tag{17}$$

where $\Phi^{-1}(\cdot)$ is the inverse normal cumulative distribution and $\Phi_2(\cdot, \cdot)$ is the joint distribution function of a bivariate Gaussian distribution with mean $\mathbb{O}$ and correlation matrix $\mathcal{R}$. Because this matrix is normalized, it is associated with the Pearson correlation coefficients matrix. Therefore, it is used to incorporate the dependence between the width and depth random variables. For this purpose, the measures most implemented at invariant scales are the population versions of Kendall and Spearman. They correspond to a form of dependence known as concordance, which allows a reliable estimation when the copula is assumed to belong a specific parametric family such as the one presented in this case (Escarela & Hernández, 2009).

3.3.3. Normalized pdf given the predicted WD ratio

Assume that the WD ratio is determined from the above multivariate regression and the width is taken as $W = WD \cdot D$. Then, the width-depth joint *pdf* from Eq. 16 is given by:

$$f_{W,D}(WD \cdot d, d) = c(F_D(d), F_W(WD \cdot d))\, f_D(d)\, f_W(WD \cdot d) \tag{18}$$

where F_D (F_W) and f_D (f_W) are the distribution and *pdf* of the crater depth (width). Note that this joint *pdf* follows paths as in Fig. 4 because of the WD linear relationship.

Consider the prediction interval in Eq.14 and denote the lower, fit and upper WD as WD_{lwr}, WD_{fit} and WD_{fit}, respectively. To determine the feasible dimensions, three preliminary scenarios are proposed for a pipeline diameter less than 18 in:

- **Upper scenario:** This scenario uses the upper limit of the WD prediction interval in the joint *pdf* to select the most probable crater depth: $D^* = \max_d \left\{ f_{W,D}(WD_{upr} \cdot d, d) \right\}$. The width of the crater is then calculated using this WD ratio.
- **Lower scenario:** This scenario uses the lower limit of the WD prediction interval in the joint *pdf* to select the most probable crater depth: $D^* = \max_d \{ f_{W,D}(WD_{lwr} \cdot d, d) \}$. The width of the crater is then calculated using this WD ratio.
- **Mean scenario:** This scenario uses the fitted WD value in the joint *pdf* to select the most probable crater depth: $D^* = \max_d \left\{ f_{W,D}(WD_{fit} \cdot d, d) \right\}$. The width of the crater is then calculated using this WD ratio.

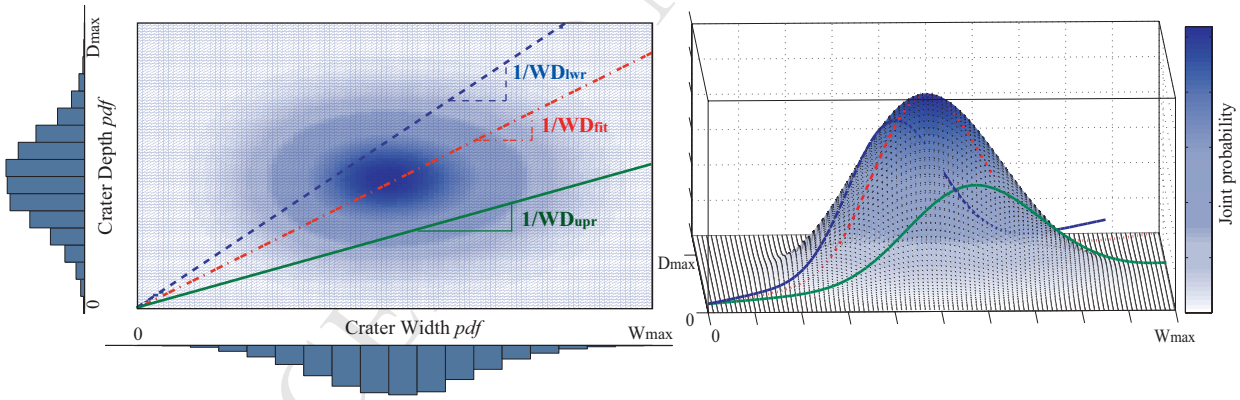


Figure 4: Scheme of the three preliminary scenarios

Fig. 4 shows the linear relation of the upper, lower and fit WD ratio, and the width-depth joint probability schematically. The three lines represent the *pdf* obtained from Eq.18 using these three ratios. Note that a higher WD ratio may involve a lower feasible depth in comparison with the fit and lower confidence limit because of the shape of the joint *pdf*.

In the case of a pipeline whose diameter is greater or equal to 18 inches, these scenarios change slightly. Instead of using a maximum criterion to determine the feasible dimensions, we consider the expected value of the normalized *pdf*. Algorithm 1 describes the main procedure used to obtain the feasible crater dimensions using the maximum or the expected approaches. In

this case, the input parameters are the WD ratio, a partition parameter n_p, the maximum width W_{max} and the pipeline diameter D_p. Here n_p and W_{max} are chosen from the final user.

Algorithm 1 Feasible crater dimensions

Input: WD ratio, n_p (partition), $W_{\max}$, D_p (pipeline diameter).
Output Feasible dimensions of W and D^*

Set $w_{eval} := [0.001 : W_{\max}]$ {Vector with n_p equally spaced points from 0.001 to $W_{\max}$}
Set $d_{eval} := w_{eval}/WD$ {Calculate the equivalent depth for each width in w_{eval}}
Initialize the vector $f_{W,D} = [0, \ldots, 0]$ {With n_p points}
for $i = 1$ to n_p **do**
    $f_{W,D} = c(F_D(d_{eval}(i)), F_W(WD \cdot d_{eval}(i)))\, f_D(d_{eval}(i))\, f_W(WD \cdot d_{eval}(i))$. {See Eq. 18}
end for
if $D_p < 18in$ **then**
    $ind_{\max}$ =Find(max($f_{W,D}$))
    $W = w_{eval}(indi_{\max})$ {The width with the maximum $f_{W,D}$}
    $D^* = d_{eval}(indi_{\max})$ {The adjusted depth with the maximum $f_{W,D}$}
else
    $f_{W,D} = f_{W,D}/\text{Area}(f_{W,D}, d_{eval})$. {Obtained the normalized *pdf*, dividing by the area of $f_{W,D}$ with d_{eval}}
    $D^* = \mathbb{E}[f_{W,D}]$ {The adjusted depth is the mean of $f_{W,D}$ with d_{eval})}
    $W = D^* \cdot WD$ {Calculate the width with the WD ratio}
end if

Finally, if the burial depth b_d is reported, the corresponding crater depth is obtained using $D = D^* + D_c - 0.9144$, where D^* is the depth obtained from the maximum/expected value criterion and D_c is the cover depth calculated as $D_c = b_d - D_p$. Recall that if the cover depth is not reported, it is assumed to be the reference cover of 0.9144, so the crater depth obtained follows $D = D^*$.

As an alternative analysis of the preliminary scenarios mentioned before, the decision-maker can select the WD within the prediction interval whose probability is the highest (using the *pdf* of the WD ratio, just mentioned) or the following scenarios to approximate the crater dimensions:

- **Less Severe scenario:** This scenario implements the lowest width and depth dimensions obtained from the Lower, Upper and Mean fits.
- **The most likely scenario:** This scenario implements the width and depth dimensions from the Lower, Mean or Upper fits with the higher probability on the joint *pdf* (See Fig. 4).
- **Worst scenario:** This scenario implements the greatest width and depth dimensions obtained from the Lower, Upper and Mean fits.

The aforementioned scenarios were considered with the initial Lower, Mean and Upper fits.

4. Results and Discussion

4.1. Data processing

4.1.1. Summary results of the proposed datasets

To evaluate if the data process affected the predictor variables significantly, the mean, coefficient of variation (CoV) and the variable range were used for each dataset described in Table 3. To show their lack of information, Table 4 illustrates these results, including the total records from each variable and dataset.

Table 4: Basic statistics from the datasets.

Dataset	Feature	Diameter (in)	Wall thickness (in)	Operating pressure (bar)	Crater dimension	
					Width (m)	Depth (m)
Set 1 (30 Records)	Total records	57	41	47	61	61
	Mean	26.00	0.34	57.56	11.43	4.50
	CoV	38.37%	36.35%	27.03%	56.62%	61.21%
	Range	[6-56]	[0.071-0.752]	[17.2-92.4]	[1.5-33]	[0.6-18.3]
Set 2 (41 Records)	Total records	57	41	47	61	61
	Mean	26.00	0.34	57.56	11.43	4.33
	CoV	38.37%	36.35%	27.03%	56.62%	51.08%
	Range	[6-56]	[0.071-0.752]	[17.2-92.4]	[1.5-33]	[0.6-10.88]
Set 3 (40 Records)	Total records	84	57	65	89	89
	Mean	26.12	0.34	56.13	10.91	4.16
	CoV	37.19%	32.40%	26.12%	58.63%	62.14%
	Range	[6-56]	[0.071-0.752]	[17.2-92.4]	[1.5-33]	[0.6-18.3]
Set 4 (57 Records)	Total records	84	57	65	89	89
	Mean	26.12	0.34	56.13	10.91	4.09
	CoV	37.19%	32.40%	26.12%	58.63%	53.48%
	Range	[6-56]	[0.071-0.752]	[17.2-92.4]	[1.5-33]	[0.6-10.88]

These four datasets are depicted in Fig. 5. From the initial 90 accidents, 30 records were obtained without considering the reference cover depth nor the data processing for the WD relation. Once the reference cover (Set 2) or the WD relation (Set 3) was implemented from these records, 11 and 10 new records were processed respectively. Finally, when both approaches were implemented, 6 new records (apart from Set 2 and 3 records) become available for the regression analysis.

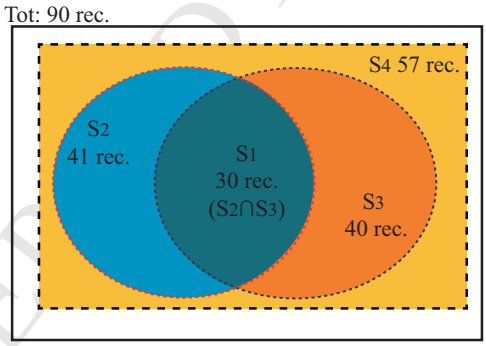


Figure 5: Datasets distribution scheme.

Overall, the differences among these datasets are not significant; the mean crater depth obtained in Set 4 decreases 9% from the initial data of Set 1. Regarding the dispersion of the datasets, the CoV of the diameter, wall thickness and the crater depth have a mean decrease of 3.44%, whereas the width CoV increases around 2.01%. Finally, the obtained ranges from each dataset remain almost the same. These results suggest that the data processing did not provide an important bias in this approach; on the contrary, it provided almost double that of the initial available records for the multivariate regression approach.

4.1.2. Width and depth fitting pdfs

Based on the BIC approach, the Width and Depth distributions for each dataset were fitted to Gamma distributions. Table 5 shows their fitting parameters. Note that the shape and scale parameters do not change among each dataset significantly. In fact, even though the correlation coefficient increases when the WD relation is implemented, this is not drastic given that near 30% of the initial WD dataset is included with this approximation.

Table 5: Width and Depth crater fitting BIC parameters.

Dataset	Feature	Distribution	Shape parameter	Scale parameter	Correlation coefficient
Set 1	Width	Gamma	3.0645	3.7308	0.4350
	Depth	Gamma	3.2347	1.3917	
Set 2	Width	Gamma	3.0645	3.7308	0.4387
	Depth	Gamma	3.8699	1.1178	
Set 3	Width	Gamma	2.8884	3.7765	0.5556
	Depth	Gamma	3.0618	1.3585	
Set 4	Width	Gamma	2.8884	3.7765	0.5788
	Depth	Gamma	3.4049	1.2006	

The probability density function of the Width and Depth from each dataset is shown in Fig. 6, which illustrates that the distributions of the datasets do not present an important change among them. Set 1 and Set 4 have almost the same shape with a difference near to 0.02 around a depth of 2 m. Regarding to the width distributions, the *pdf* obtained from each dataset match almost entirely. The aforementioned distributions and the results from Table 4 suggest that these assumptions will not severely affect the final predictions of the regression. In contrast, the multivariate approach will obtain more reliable results due to the increase of the sample size. Based on these results, the records from Set 4 were employed for the rest of the proposed approach.

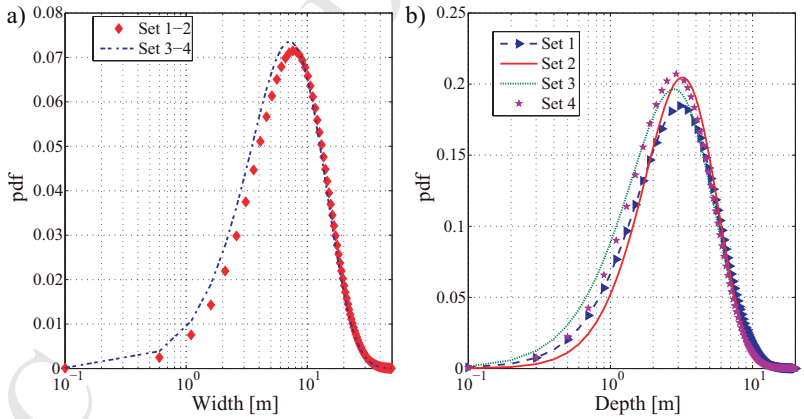


Figure 6: a) Width and b) Depth probability density function of each dataset.

4.2. Prediction of WD ratio

Initially, the mixed selection method was implemented in R-project to select the variables that best describe the WD regression. The results of this selection model indicate that the WD ratio is mainly approximated with the operating pressure and the pipeline diameter. Recall that

the mixed selection compares each regression model using its p-value by adding variables one by one to provide a better fit for the regression model. Table 6 shows the best multivariate regressions using these variables after an outlier with a $WD = 9$ was removed. Note that these regressions only illustrate the variables implemented. They include possible interactions among their predictors, and they neglect the intercept parameter in all cases (i.e., β_0 in Eq. 11).

Table 6: Best multivariate regressions from Set 4.

Formula	RSE	R^2-adj	AIC_c	BIC_c	p-Value
$\log(WD) \sim \log(P)$	0.4226	0.8474	65.4377	69.4884	2.48E-24
$\log(WD) \sim \log(P) + \log(D_p)$	0.4183	0.8505	65.2646	71.3407	1.94E-23
$\log(WD) \sim \log(P) + \log(D_p) + \log(WT)$	0.4155	0.8525	65.4581	73.5595	1.19E-22
$\log(WD) \sim \log(P) + \log(D_p) + \log(P)\log(D_p)$	0.4123	0.8548	64.5987	72.7002	7.95E-23
$\log(WD) \sim \log(P) + \log(WT) + \log(D_p) + \log(WT)\log(D_p)$	0.4191	0.8499	67.3638	77.4905	1.30E-21
$\log(WD) \sim \log(P) + \log(WT) + \log(D_p) + \log(WT)\log(D_p) + \log(WT)\log(P) + \log(P)\log(D_p) + \log(P)\log(WT)\log(D_p)$	0.4272	0.8440	72.1959	88.3987	4.80E-19

P: Operating pressure [bar], WT: Wall thickness [in], D_p: Diameter [in], WD: Width-to-Depth ratio

The best regression is the one with a highest R^2-adjusted and lowest RSE, *AICc*, *BICc* and p-Value. These results initially suggest that the operating pressure could determine an adequate regression model (i.e., $\log(WD) \sim \log(P)$) because the RSE and the R^2-adjusted were almost the same as the other models. However, their information criteria give better results than the other models. Nevertheless, AIC and BIC penalize the models that include more predictors, so further information is required for this selection. Consequently, the regression model of $\log(WD) \sim \log(P) + \log(D_p) + \log(P)\log(D_p)$ was evaluated because it obtained better results in RSE, R^2-adjusted and AIC. This is due to the relaxation of the additive assumption, which is associated with the interaction parameter between $\log(P)$ and $\log(D_p)$ (James et al., 2013). This relaxation can be shown using Eq. 11 as follows:

$$\begin{aligned} \log(WD) &= \beta_1 \log(P) + \beta_2 \log(D_p) + \beta_3 \log(P)\log(D_p) + \epsilon \\ &= \left[\beta_1 + \beta_3 \log(D_p)\right] \log(P) + \beta_2 \log(D_p) + \epsilon \\ &= \hat{\beta}_1 \log(P) + \beta_2 \log(D_p) + \epsilon \end{aligned} \tag{19}$$

where $\hat{\beta}_1 = \beta_1 + \beta_3 \log(D_p)$. Since $\hat{\beta}_1$ changes with $\log(D_p)$, the effect of $\log(P)$ is no longer constant. Indeed, adjusting $\log(D_p)$ will change the impact of $\log(P)$ on $\log(WD)$ (James et al., 2013).

Considering this regression, the main assumptions were evaluated based on their diagnostic plot shown in Fig.7. A pattern that is slightly affected by some records (red line) is shown in Fig. 7a, but overall it has a flat tendency confirming the linearity assumption. The QQ-plot, which is illustrated in Fig. 7b, indicates that the residuals of the regression are normally distributed because the records lie almost on the diagonal of the QQ-plot. Regarding homoscedasticity and the independence assumptions, note that in Fig. 7c the residuals are mostly equally spread. This result suggests that the variance does not change drastically along the fitted values and the data are independent. Finally, Fig. 7d depicts the evaluation of the leverage. Note in this figure that some records obtain a high leverage that may affect the regression performance, so it was assessed whether it was beneficial to remove these data points, considering the sample size. For this purpose, an outlier and leverage diagnoses were implemented with the studentized residuals and Residuals-Leverage plots. Recall that the observations whose studentized residuals are greater than 3 (absolutely) are possible outliers. For this regression, the studentized residuals

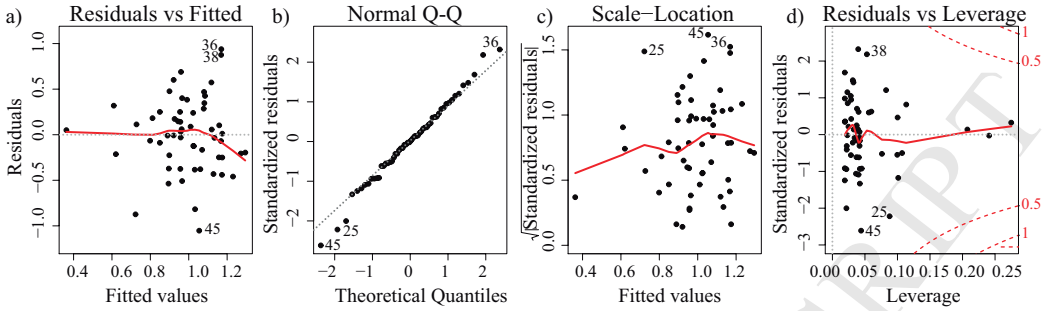


Figure 7: Diagnostic plot $\log(WD) \sim \log(P) + \log(D_p) + \log(P)\log(D_p)$.

obtained range from -2.77 to 2.43, so there is not strong evidence for the presence of outliers. The leverage diagnosis obtained results near to 0.25, which indicate that the model could neglect this leverage and maintain the records with a high leverage. In conclusion, the main assumptions are satisfied and this regression mainly describes the dataset.

Table 7: Best multivariate regressions from Set 4.

Predictor	Coefficient	Confidence interval	
		2.5%	97.5%
$\log(D_p)$	-0.1648	-0.5449	0.2152
$\log(P)$	0.0026	-0.0034	0.0085
$\log(D_p)\log(P)$	0.1156	-0.1510	0.3822

Table 7 shows the coefficients obtained and their confidence intervals following Eq. 13 and $\alpha = 5\%$. The prediction interval in Eq. 14 was determined using $X_o' = \left[\log(D_p), \log(P), \log(D_p)\log(P)\right]$ and $\tilde{y}_o = \log(WD)$, where the latter is obtained from the coefficients reported in Table 7 using Eq. 19. Note that the predictor vector X_o' does not include the first entry associated with 1 because this regression omits the intercept. Considering the evaluated predictors X_o', this prediction interval is given as follows:

$$\tilde{y}_o \pm 0.8273 \sqrt{1 + X_o' \begin{bmatrix} 0.55094 & -0.09003 & -0.10852 \\ -0.09003 & 0.11263 & -0.01187 \\ -0.10852 & -0.01187 & 0.03043 \end{bmatrix} X_o} \tag{20}$$

The aforementioned results allow decision makers to predict the WD ratio of a crater with a reference cover of 0.9144 m based on the operating pressure, the pipeline diameter, and the product of both parameters.

4.3. Feasible crater dimensions

4.3.1. Copula-based approach

Considering the correlation coefficient of Set 4 (Table 5), the Kendall correlation coefficient was 0.7890. This coefficient serves as the off-diagonal entries in the correlation matrix ($\mathcal{R}$) of the bivariate Gaussian distribution in the copula construction. Based on the definition of the copula

density and the *pdf* of the multivariate Gaussian, this copula density function is described as follows (see Blumentritt (2011)):

$$c(\mathbb{U}) = |\mathcal{R}|^{-0.5} \exp\left\{-\frac{\left[\Phi^{-1}(U_1),\ldots,\Phi^{-1}(U_d)\right](\mathcal{R}^{-1}-\mathbb{I}_d)\left[\Phi^{-1}(U_1),\ldots,\Phi^{-1}(U_d)\right]^T}{2}\right\} \tag{21}$$

where $|\mathcal{R}|$ denotes the determinant of $\mathcal{R}$, $\mathbb{U} \in [0,1]^d$ and $\mathbb{I}_d$ is the d-identity matrix. Then, considering that the joint *pdf* of a random vector is obtained from the copula density function, and that the marginal densities of each variable (see Eq. 16), the width-depth joint density distribution is given by:

$$f_{W,D}(w,d) = |\mathcal{R}|^{-0.5} \exp\left\{-\frac{\left[\Phi^{-1}(F_W(w)),\Phi^{-1}(F_D(d))\right](\mathcal{R}^{-1}-\mathbb{I}_2)\begin{bmatrix}\Phi^{-1}(F_W(w))\\ \Phi^{-1}(F_D(d))\end{bmatrix}}{2}\right\} f_W(w) f_D(d) \tag{22}$$

where F_D and F_W are the cumulative Gamma distributions of the width and depth; f_D and f_W are the *pdf* of the width and depth; and Φ^{-1} is the inverse normal cumulative distribution, the latter can be determined from the error function as: $\Phi^{-1}(p) = \sqrt{2}\text{erf}^{-1}(2p-1)$, where $p \in (0,1)$.

4.3.2. Feasible width and depth of craters - Case studies

To illustrate how the feasible width and depth of the crater can be calculated, let us consider the following real accidents: Brunswick (GA, USA) in 2002 and Warren (MN, USA) in 2014. The first occurred on November 30, 2002 in a pipeline distributing natural gas and it created a 3 m long, 3 m wide, 1.5 m deep crater. The second accident took place on May 26, 2014 in a natural gas transmission pipeline and led to the evacuation of 10 families and a crater which was 9.1 m long, 9.1 m wide and 4.6 m deep. The relevant information from these accidents is summarized in Table 8. Note that because the burial depth b_d is not reported in both cases, the adjusted depth D^* coincides with the reported depth. For more information about these accidents, the reader may refer to the information reported in Ramírez-Camacho et al. (2017).

Table 8: Brunswick and Warren accidents summary.

Case	D_p (in)	P (bar)	b_d (m)	Crater's Width (m)	Crater's Depth (m)	D^* (m)	WD
Brunswick	8	17.2	-	3.0	1.5	1.5	2.00
Warren	24	56.9	-	9.1	4.6	4.6	1.98

To calculate the fitted WD ratio and its prediction interval (for both accidents), implement Eq. 19 and Eq. 20. Evaluate these equations at the pipeline diameter D_p and the operating pressure P from Table 8, and the regression coefficients reported in Table 7. The result of the initial prediction was a WD ratio of 1.417 for Brunswick's accident and 2.641 for Warren's accident. The prediction intervals obtained were [0.568-3.537] for the Brunswick accident and [1.146-6.087] for the Warren accident. Note that in both cases, the real WDs were within these prediction intervals.

Based on the prediction interval and the regression results, let us describe the process to approximate the crater feasible dimensions using Algorithm 1. Consider the prediction interval

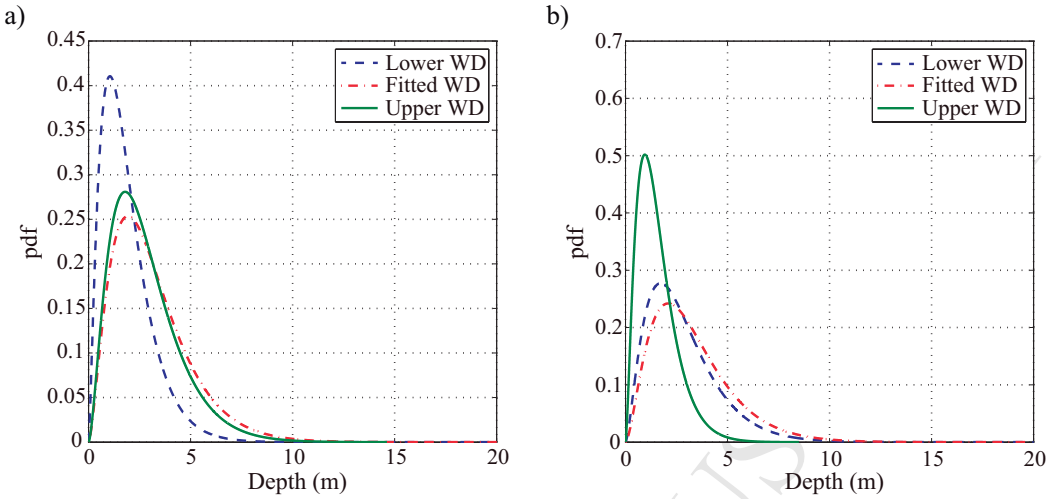


Figure 8: Depths' *pdf* with the regression results for the a) Brunswick accident and b) Warren accident

results (WD_{lwr}, WD_{fit} and WD_{uwr} ratios) and the D_p reported in Table 8. For these cases, take a W_{max} of a 99.9% percentile for the fitted Gamma distribution (See Table 5) and a partition number of $n_p = 10,000$. The joint *pdf* that follows the WD relation as in Eq. 18 (i.e., $f_{W,D}(WD \cdot d, d)$) is calculated from Eq. 22. The normalized *pdf* is then obtained from the quotient between $f_{W,D}$ and its area derived using a trapezoidal numerical integration. This normalized *pdf* for the lower, mean and upper fitted WD results in both accidents are illustrated in Fig. 8. Finally, the feasible dimensions are obtained using the maximum criterion for the Brunswick accident and the expected value criterion for the Warren accident, where the latter was determined using a trapezoidal rule integration. The crater dimensions obtained are presented in Table 9.

Table 9: Feasible dimensions of the craters.

Case	Lower Fit		Mean Fit		Upper Fit	
	W	D	W	D	W	D
Brunswick	0.594	1.046	2.734	1.929	6.361	1.798
Warren	3.222	2.811	8.632	3.268	9.453	1.553

For the Brunswick accident, the Less Severe scenario corresponds with the Lower Fit, the Worst scenario uses the width reported in the Upper Fit, and the depth uses the one from the Mean Fit. For the Warren accident, the Less Severe scenario considers the width from the Lower Fit and the depth from the Upper Fit, whereas the Worst scenario considers the width from the Upper Fit and the depth from the Mean Fit. In both cases, the Most Likely scenario coincides with the Mean Fit. These scenarios are illustrated in Fig. 9, where it can be shown that the reported crater dimensions are somewhere between the Most Likely and the Worst scenarios. Additionally, note that both the Less Severe scenarios have reported dimensions that are significantly smaller than the other scenarios and the dimensions reported; hence, the other scenarios could be considered instead.

Considering the limited amount of available data, these scenarios give an adequate prediction of the crater dimensions. Note that the Worst scenario of the Brunswick accident includes the

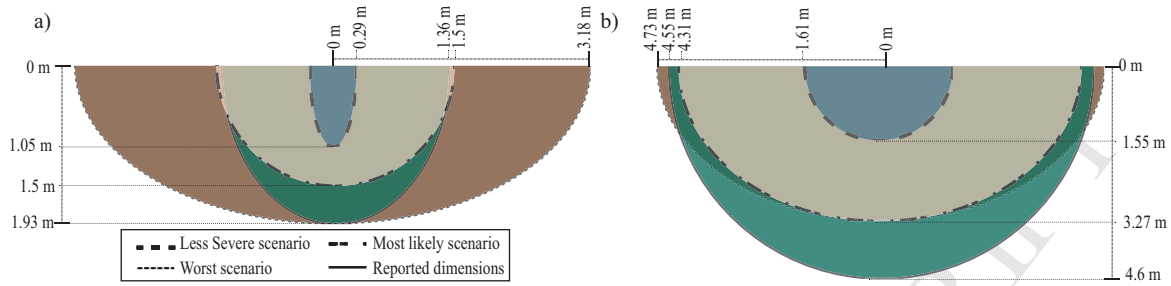


Figure 9: Crater Scenarios for the a) Brunswick and b) Warren accidents

real crater dimensions, whereas for the Worst scenario of the Warren accident predicts the real crater dimensions almost entirely. To illustrate the model prediction capabilities, the dimensions reported from 9 additional accidents were compared with those obtained with the probabilistic model in the Lower, Mean and Upper fits (Table 10). The results indicate that the real width and depth dimensions lie almost entirely within these three fits.

Table 10: Examples of the prediction of the preliminary scenarios.

D_p	P	b_d	Real	Real	D^*	WD	Predicted WD			Lower Fit		Mean Fit		Upper Fit	
(in)	(bar)	(m)	W (m)	D (m)	(m)		Lower	Mean	Upper	W	D^*	W	D^*	W	D^*
6	70.0	-	2.0	2.00	2.00	1.000	0.667	1.814	4.938	0.792	1.188	3.836	2.115	6.295	1.275
16	27.6	-	2.4	3.00	3.00	0.800	0.767	1.850	4.461	1.011	1.319	3.931	2.126	6.417	1.439
16	55.2	-	4.6	1.53	1.53	3.000	0.989	2.314	5.414	1.552	1.569	4.994	2.158	6.076	1.122
18	58.9	1.1	5.2	2.80	3.07	1.693	1.052	2.450	5.705	2.829	2.689	8.152	3.328	9.672	1.695
20	46.9	0.6	6.1	2.03	2.86	2.136	1.009	2.337	5.412	2.653	2.630	7.833	3.352	9.828	1.816
24	54.6	1.0	9.0	4.50	5.02	1.791	1.129	2.601	5.995	3.150	2.790	8.538	3.282	9.506	1.586
24	54.8	1.8	10.0	2.20	1.92	5.198	1.131	2.605	6.003	3.156	2.792	8.546	3.281	9.502	1.583
30	71.4	0.9	7.6	3.00	3.78	2.012	1.332	3.091	7.174	4.016	3.015	9.466	3.062	8.788	1.225
30	69.4	1.8	9.1	1.80	1.68	5.428	1.320	3.057	7.082	3.962	3.022	9.415	3.080	8.844	1.249

D_p: Pipeline diameter, P: Internal pressure, b_d: Burial depth, D^*: Adjusted depth, W: Predicted width

However, for the sake of a deeper comparison, consider the Most Likely crater dimensions that are associated with the mean fit, and their 25% and 75% percentiles from the quotient between $f_{W,D}$ and its area. Figure10 shows the real *vs*. Most Likely crater dimensions with these percentiles. This figure demonstrates that crater dimensions can be predicted using this Mean fit range with an error smaller than 2 m, here the greatest error was found in a 18-inch pipeline. Considering that the Maximum Probability approach (for $D_p < 18$in) produces shorter predictions than the Expected probability approach (for $D_p \geq 18$in), these two approaches could be implemented at the same time for this pipeline diameter with the aim of reducing the prediction range. Nevertheless, the differences between the real dimensions and the Mean fit range suggest adequate prediction performance of the model taking into account the limited amount of available records and that properties of the soil were not considered.

Finally, the probabilistic model was compared with some of the models reviewed in Section 2 noting that these models depend on soil properties that usually are not included in incident reports, or that are subject to significant uncertainty. However, considering some of the cases discussed in Silva et al. (2016), our probabilistic predictions were comparable with the majority of the models reviewed in Section 2. This comparison is depicted in Fig. 11, which shows suit-

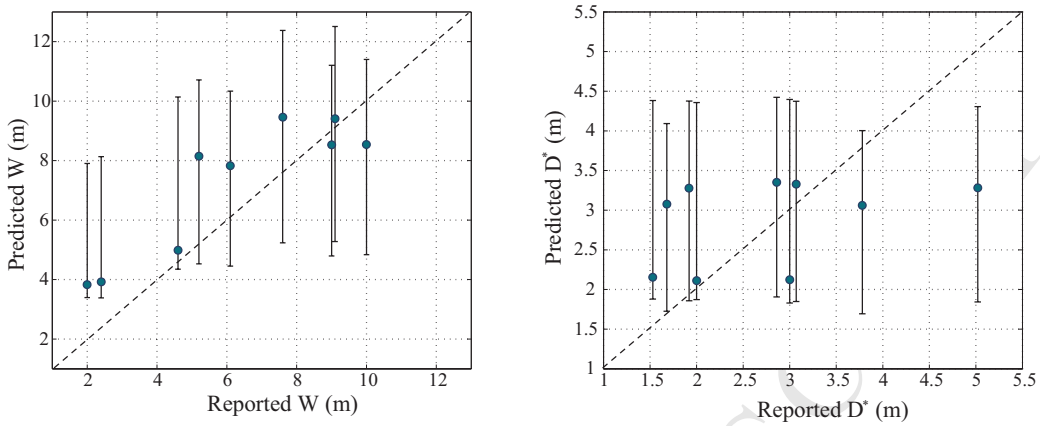


Figure 10: Real vs. predicted dimensions

able accuracy for the probabilistic model in almost all cases, with the exception of the accident with a 20-inch, 15 m wide pipeline. This accident occurs in a location with a ”Salty Clay Loam” soil according to Silva et al. (2016), which may affect the model predictions. The advantage of the proposed model is that only design parameters must be implemented (i.e., P, b_d and D_p). On the contrary, the methods described in Section 2 need additional parameters that are subjected to significant uncertainty (e.g., soil properties and crater angles) or are not well-documented publicly (NEN 3651 model).

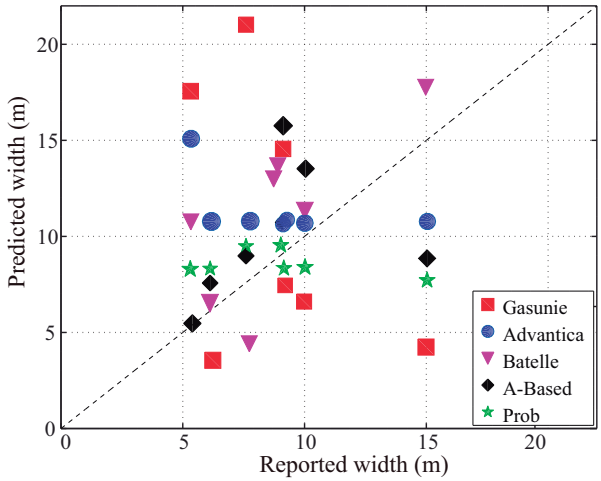


Figure 11: Comparison of the width prediction capabilities of available approaches

4.3.3. Some applications

The importance of the crater dimensions lies in their possibility to evaluate a domino effect scenario in parallel pipelines as well as the possibility to support decision-making processes in the construction and design of parallel corridors. The formation of a crater can expose adjacent pipelines and trigger a domino effect due to the initial overpressure or by a thermal failure given a Jet Fire in the source pipe. For the construction/design of parallel corridors, the crater dimensions can be used to define pipelines Right-of-Way (ROW) or to propose safe distances between parallel pipelines to avoid the events that can escalate to a domino effect, for instance, using Jet Fire models as in Haklar (1997); Haklar & Dresnack (1999); Sklavounos & Rigas (2006); Mohsin et al. (2014). Additionally, other applications such as the adjustment of the view factor from a Jet Fire model, which considers the crater obstacle, can be also considered. This adjustment would prevent an overestimation of the received heat radiation of an object on the ground level.

For the evaluation of domino effect, let us consider two real accidents involving parallel pipelines with natural gas. The first occurred in Rapid City (Canada) in 1995, where a 42-inch gas main pipeline failed due to stress corrosion cracking. The crater formed by this rupture uncovers an adjacent 36-inch gas main pipeline that failed because of the heat overload from a fire on the source pipeline. The second accident took place in Buick (Canada) in 2012, where a rupture and fire occurred in a 16-inch gas-gathering pipeline due to a hook crack. The rupture exposed a parallel 6.625-inch gas-gathering pipeline, which after 25 minutes failed because of flame impingement. The parameters from these pipelines are summarized in Table 11.

Table 11: Operating and structural parameters from domino effect accidents (TSB, 1997, 2013b)

Case	Source pipeline			Target pipeline		
	D_p (in)	b_d (m)	P (bar)	D_p (in)	b_d (m)	P (bar)
Rapid City	42	1.5	60.7	36	1.5	60.7
Buick	16	0.5	66.6	6.625	0.5	8.7

D_p: Pipeline diameter, b_d: Burial depth and P:Operating pressure

The predicted worst scenario crater for the source pipeline in the Rapid City accident is illustrated in Fig. 12. The width and depth for the source pipe's crater are 10.04 m and 5.50 m, respectively. Note that if the crater were assumed to be symmetrical, the distance from the crater to the target pipe would be around 1.9 m. Nevertheless, this distance could be shorter because of the flame direction and the "hydraulic mining hose" described by Peekema (2013). Note also that if the crater were not assumed to be symmetrical, then it would be oriented to the target pipe, so it is likely that the initial rupture would uncovers the 36-inch main pipeline. Assuming that the target pipeline fails due to the fire, the predicted crater of this pipe would have a width of 9.91 m and a depth of 5.30 m. As a result, the equivalent crater width and depth of the accident are around 17 m and 5 m, respectively, which do not differ from the real dimensions significantly. According to the report, the width was 23 m and the depth was around 5 m, which is an interesting result taking into account the limited amount of available information. Other approaches like the Advantica or the Accident-based models obtained width results of about 12.23 and 16.5 m, which are farther for, the reported wide[1].

As in the Rapid City accident, Fig.13 shows the predicted crater with the worst scenario for the source pipe in the accident near Buick. This figure indicates that a rupture in the source

[1]Gasunie and Batelle model require a further classification of the soil that may affect the final result, so these approaches were omitted.

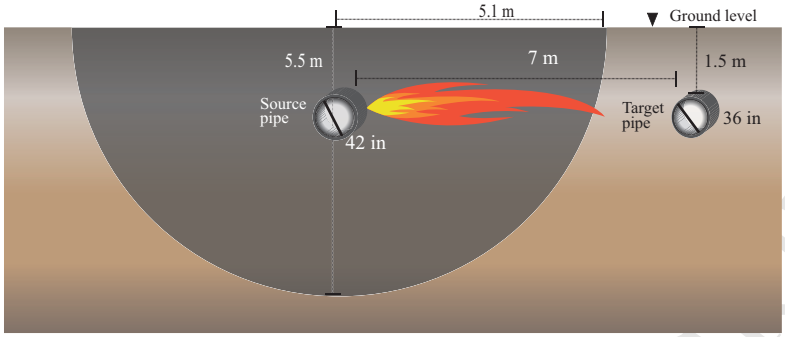


Figure 12: Predicted crater of the Rapid City accident

pipe certainly would uncover the 6.625-inch gathering line because the difference between the crater and the target pipe is around 0.01 m. As in the previous accident, a fire is oriented to the target pipe that can reduce this distance. Following a similar procedure as in the Rapid City accident, the equivalent crater dimensions would be around 8.71 m wide and 3.42 m deep. These dimensions are greater than the reported crater that is 7.6 m wide and 1.1 m of deep, but it is closer if the Advantica model is implemented with a width of 10.6 m.

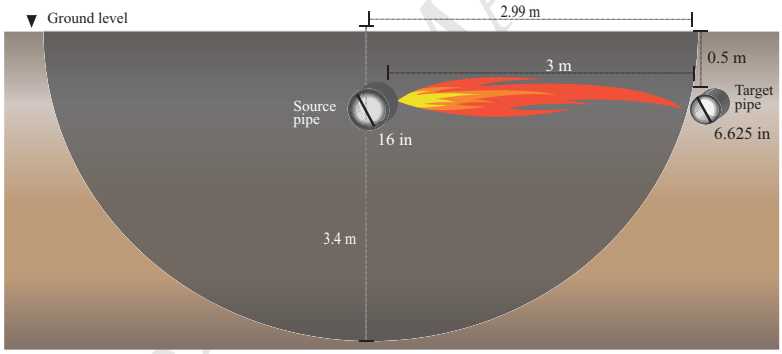


Figure 13: Predicted crater of the Buick accident

Finally, there are potential applications for the use ROW definition. According to the guidelines for parallel construction of pipelines of INGAA (2008), federal agencies in the United States such as the Federal Regulatory Commission, require the implementation of an existing ROW for the construction of new interstate natural gas pipelines. Therefore, parallel pipelines can be located in a same ditch with a minimum distance between them of 1 m. Note that these ROW ditch lengths could depend on the pipeline diameter: i) 10 m for a pipeline with a diameter less than 8 inches; ii) 13 m for a pipeline whose diameter is between 10 to 18 inches; and iii) 15 m for a pipeline with a diameter between 20 to 36 inches (PEMEX, 2009). Now, if two parallel pipelines whose diameters are in the last category (20-36 inches) are in the same ditch, they would be separated by at least 1 m and as much 14 m. Suppose this separation is about 7-8 m (as in the Rapid City accident), then the examples and the predicted width dimensions suggest that a domino effect would take place. For this scenario, a safety distance of about 10m

could be suggested like in Silva et al. (2016) by using only an extension of the crater width; nevertheless, this scenario is likely to be followed by a jet fire (Bubbico et al., 2016), so further calculations that are outside of the scope of this paper should be implemented. For more details see Ramírez-Camacho et al. (2015); Alileche et al. (2015); Haklar (1997); Haklar & Dresnack (1999); Sklavounos & Rigas (2006); Mohsin et al. (2014). Nonetheless, the probabilistic-based approach proposed in this paper could be implemented to support ROW definition by evaluating the exposure possibility of two parallel or crossing pipelines.

5. Conclusions

This paper presents a model to predict the feasible crater dimensions (i.e., width and depth) produced by an underground natural gas (NG) pipeline given a LOC. This model was developed based on a recent review of 90 real accidents of underground NG pipelines(Ramírez-Camacho et al., 2017) in a probability-based approach with a maximum or expected criterion. The feasible dimensions obtained with this model allow a decision maker to evaluate the possibility of exposure of a parallel or cross pipeline, which in turn can be used to identify a domino effect scenario and to evaluate Right-of-Way (ROW) corridors for parallel or crossing routes.

The probabilistic approach was developed based on a multivariate regression and a Gaussian copula. The multivariate regression predicts the width-to-depth ratio (WD) of the crater based on a trained set of 57 over 90 reported accidents after a gathering process, whereas the copula approximates the width-depth joint *pdf*. Based on a prediction interval of the multivariate regression and the joint width-depth density, three scenarios were proposed: Less Severe, Most Likely and Worst scenarios.

The model has been tested for the accidents in Brunswick (GA, USA) in 2012 and Warren (MN, USA) in 2014 and obtained results near to those reported. The greatest difference was around 1.3 m for the Warren's accident depth, which could be attributed to a non-reported cover depth that is greater than the cover reference used. Results from 9 additional cases and a comparison with other models that predict the crater width indicate that the model has adequately prediction capabilities considering the limit amount of available records. In addition, the Rapid City (Canada, 1995) and Buick (Canada, 2012) accidents were used to illustrate that these pipelines were very likely to trigger a domino effect scenario because of the proximity between their parallel pipelines, the crater dimensions obtained, and the jet flame direction relative to the target pipes.

This paper seeks to support decision-making processes regarding NG pipelines due to crater formation. However, the readers should bear in mind that this approach does not implement information regarding soil properties, pipeline route, and crater angles because none of these parameters are usually reported.

Acknowledgements

R. Amaya-Gómez thanks the National Department of Science, Technology and Innovation of Colombia for the PhD scholarship (COLCIENCIAS Grant No. 727, 2015).

J.G. Ramírez-Camacho thanks the Mexican National Council of Science and Technology (CONACyT) for the Ph.D. scholarship and the Universidad de Los Andes in Bogotá for the support given to this research.

Nomenclature

α	Significance level	D^*	Adjusted depth
α_1, α_2	Crater's angle at the ground surface and at $D/2$ for the Gasunie model	D_c	Cover depth
β_i	Regression coefficients	D_p	Pipeline diameter
ϵ	Regression residual error	$F_i(\cdot)$	Distribution of the i^{th} variable
γ	Gas specific heat ratio	$f_i(\cdot)$	Density function of the i^{th} variable
$\mathbb{I}_d$	The d-identity matrix	$f_{W,D}$	Joint width-depth density function
$\mathbb{U}$	Uniformly distributed random vector	n	Sample size for the regression
ν	Soil parameter for the Gasunie model	n_p	Partition parameter
$\Phi^{-1}(\cdot)$	Inverse normal cumulative distribution	P	Operating pressure
$\Phi_2(\cdot,\cdot)$	Bivariate Gaussian distribution with mean $\mathbb{O}$ and correlation matrix $\mathcal{R}$	ROW	Right-of-Way definition
ρ, ρ_{soil}	Gas and soil densities	$t_{\alpha/2,n-k-1}$	Student's t inverse cumulative distribution at $\alpha/2$ with $n-k-1$ degrees of freedom
σ_Y	Yield strength	u_x	Velocity of the explosive gases for the Batelle model
$\text{erf}^{-1}(\cdot)$	Inverse error function	u_{kr}	Critical gas velocity for the Batelle model
$\tilde{\beta}_j$	Predicted regression coefficient	W	Crater's width
$\tilde{y}_o$	Prediction of the future regression response	W_{max}	Maximum evaluated width
a, b	Ellipse shape parameters for the Gasunie model	WD	Width-to-depth ratio
b_d	Burial depth	WT	Pipeline wall thickness
$C(\cdot,\cdot)$	Copula distribution function	X	Fitted observations for the k predictors
$c(\cdot,\cdot)$	Copula density function	X_o	Predictors to be evaluated
D	Crater's depth		

References

Acton, M., Jackson, N., & Jager, E. (2010). Development of Guidelines for Parallel Pipelines. In *2010 8th International Pipeline Conference* (pp. 485–495). Calgary, Alberta, Canada. doi:10.1115/IPC2010-31287.

Alileche, N., Cozzani, V., Reniers, G., & Estel, L. (2015). Thresholds for domino effects and safety distances in the process industry: A review of approaches and regulations. *Reliability Engineering & System Safety, 143*, 74 – 84. doi:https://doi.org/10.1016/j.ress.2015.04.007.

Ambrosini, R., & Luccioni, B. (2006). Craters Produced by Explosions on the Soil Surface. *Journal of Applied Mechanics, 73*, 890–900. doi:10.1115/1.2173283.

Ambrosini, R., Luccioni, B., Danesi, R., Riera, J., & Rocha, M. (2002). Size of craters produced by explosive charges on or above the ground surface. *Shock Waves, 12*, 69–78. doi:10.1007/s00193-002-0136-3.

ASME (2002). *ASME B31.4-2002: Pipeline Transportation Systems for Liquid Hydrocarbons and Other Liquids*. Technical Report American Society of Mechanical Engineers New York. Page 41.

ASME (2004). *ASME B31.8-2003: Gas Transmission and Distribution Piping Systems*. Technical Report American Society of Mechanical Engineers New York. Page 33.

BAM (2009). *On the risks of transporting liquid and gaseous fuels in pipelines*. Technical Report Federal Institute for Materials Research and Testing Berlin, Germany. Research Report 289.

Blumentritt, T. (2011). *On Copula Density Estimation and Measures of Multivariate Association*. Ph.D. thesis University of Cologne. Published in Eul-Band 171, 2012.

Bubbico, R., Carbone, F., Ramírez-Camacho, J., Pastor, E., & Casal, J. (2016). Conditional probabilities of post-release events for hazardous materials pipelines. *Process Safety and Environmental Protection, 104, Part A*, 95 – 110. doi:https://doi.org/10.1016/j.psep.2016.08.011.

Casal, J., Cisteró, J., & Tort, J. (1995). Some risks of energy distribution. Analysis of an explosion involving electrical and gas networks. *Loss Prevention and Safety Promotion in the Process Industries, 1*, 565–575.

DOT (2002). *Corrective Action Order CPF No. 3-2002-1003-H*. Technical Report Department of Transportation Washington D.C., USA.

DOT (2008). *Rupture of Houstonia 200-Line operated by Panhandle Eastern Pipeline Company, LP, Pilot Grove, Missouri, August 25, 2008*. Technical Report Department of Transportation Washington D.C., USA. Pipeline Failure Investigation Report No. 20090030-5319.

DOT (2009). *Corrective Action Order CPF No. 3-2009-1009-H*. Technical Report Department of Transportation Washington D.C., USA.

Escarela, G., & Hernández, A. (2009). Modelado de parejas aleatorias usando cópulas. *Revista Colombiana de Estadística*, *32*, 33 – 58.

Exponent (2015). *PG&E Line 118B Fresno In-Service Rupture Analysis*. Technical Report Exponent, Inc. Failure Analysis Associates California, USA. Doc. No. 1502991.000-7356.

Frees, E., & Valdez, E. (1998). Understanding Relationships Using Copulas. *North American Actuarial Journal*, *2*, 1–25. doi:10.1080/10920277.1998.10595667.

Haklar, J. (1997). *Analysis of Safe Separation Distances from Natural Gas Transmission Pipelines*. Ph.D. thesis New Jersey Institute of Technology.

Haklar, J., & Dresnack, R. (1999). Safe Separation Distances from Natural Gas Transmission Pipelines. *Journal of Pipeline Safety*, *1*, 3–20.

HInt Dossier (2005). Gas Pipeline Explosion at Ghislenghien, Belgium. https://www.iab-atex.nl. Data accessed: July 1 2015.

HSE (2000). *Report on a study of international pipeline accidents*. Technical Report Health and Safety Executive Liverpool, UK. Contract Research Report 294/2000.

HSE (2002). *Report on a second study of pipeline accidents using the Health and Safety Executives risk assessment programs MISHAP and PIPERS*. Technical Report Health and Safety Executive Liverpool, UK. Research Report 036.

INGAA (2008). *Guidelines for Parallel Construction of Pipelines*. Technical Report F-2008-05 (Interstate Natural Gas Association of America.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics. Springer New York.

Konishi, S., & Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*. Springer Series in Statistics. Springer.

Krishna, S., Krishnamurthy, R., & Gao, M. (2016). Integrity of Buried Gas Pipeline Subjected to an Adjacent Pipe Rupture Event. In *International Pipelines Conference and Exposition 2016*. Calgary, Canada. doi:10.1115/IPC2016-64511.

Leis, B. N., Pimputkar, S. M., & Ghadiali, N. D. (2002). *Line rupture and the spacing of parallel lines*. Technical Report Contract PR-3-9604 Pipeline Research Council International.

Luo, Z., Wang, X., & Yang, H. (2009). Analytic study based on failure effects model of adjacent oil pipeline. In *2009 International Conference on Management Science and Engineering* (pp. 271–276). Moscow, Russia. doi:10.1109/ICMSE.2009.5317455.

Maindonald, J., & Braun, W. (2010). *Data Analysis and Graphics Using R: An Example-Based Approach*. Cambridge Series in Statistical and Probabilistic Mathematics (3rd ed.). Cambridge University Press.

Majid, Z., & Mohsin, R. (2013). Multiple failures of API 5L X42 natural gas pipeline. *Engineering Failure Analysis*, *31*, 421 – 429. doi:http://dx.doi.org/10.1016/j.engfailanal.2013.02.012.

McGillivray, A., & Wilday, J. (2009). *Comparison of risks from carbon dioxide and natural gas pipelines*. Technical Report RR749 Health and Safety Executive.

MHIDAS (2007). Major Hazardous Incident Data Service. AEA technology. London: HSE-Health and Safety Executive. U.K.

Mishra, K., & Wehrstedt, K.-D. (2015). Underground gas pipeline explosion and fire: CFD based assessment of foreseeability. *Journal of Natural Gas Science and Engineering*, *24*, 526 – 542. doi:https://doi.org/10.1016/j.jngse.2015.04.010.

Mohsin, R., Majid, Z., & Yusof, M. (2014). Safety distance between underground natural gas and water pipeline facilities. *Reliability Engineering & System Safety*, *131*, 53 – 60. doi:https://doi.org/10.1016/j.ress.2014.06.008.

Mokhtari, M., & Alavi, A. (2015). A parametric study on the mechanical performance of buried {X65} steel pipelines under subsurface detonation. *Archives of Civil and Mechanical Engineering*, *15*, 668 – 679. doi:http://doi.org/10.1016/j.acme.2014.12.013.

Montes-Iturrizaga, R., & Heredia-Zavoni, E. (2015). Environmental contours using copulas. *Applied Ocean Research*, *52*, 125 – 139. doi:http://doi.org/10.1016/j.apor.2015.05.007.

Nelsen, R. (2007). *An Introduction to Copulas*. Springer Series in Statistics (2nd ed.). Springer New York.

NSI (2012). *NEN 3651: Additional Requirements for Pipelines in or Nearby Important Public Works*. Technical Report Netherlands Standarddization Institute.

NTSB (2003). *Natural Gas Pipeline Rupture and Fire Near Carlsbad, New Mexico, August 19 2000*. Technical Report

National Transportation Safety Board Washington, USA. Pipeline Accident Report NTSB/PAR-03/01.

NTSB (2011). *Natural Gas Transmission Pipeline Rupture and Fire, San Bruno, California, September 9, 2010*. Technical Report National Transportation Safety Board Washington, USA. Pipeline Accident Report NTSB/PAR-11/01.

NTSB (2013). *Rupture of Florida Gas Transmission Pipeline and Release of Natural Gas, Near Palm City Florida, May 4, 2009*. Technical Report National Transportation Safety Board Washington, USA. Pipeline Accident Brief NTSB/PAB-13/01.

NTSB (2014). *Columbia Gas Transmission Corporation Pipeline Rupture, Sissonville, West Virginia, December 11, 2012*. Technical Report National Transportation Safety Board Washington, USA. Pipeline Accident Report NTSB/PAR-14/01.

Peekema, R. (2013). Causes of natural gas pipeline explosive ruptures. *Journal of Pipeline Systems Engineering and Practice*, *4*, 74–80. doi:10.1061/(ASCE)PS.1949-1204.0000116.

PEMEX (2009). *Diseño, construcción, inspección y mantenimiento de ductos terrestres para transporte y recolección de hidrocarburos*. Technical Report NRF-030-PEMEX-2009 Petróleos Mexicanos.

PHMSA (2008). *Corrective Action Order CPF No. 1-2008-1004H*. Technical Report Pipeline and Hazardous Materials Safety Administration Washington, USA. Pipeline Accident Report NTSB/PAR-14/01.

PHMSA (2009). *Corrective Action Order CPF No. 4-2009-1021H*. Technical Report Pipeline and Hazardous Materials Safety Administration Washington, USA. Pipeline Accident Report NTSB/PAR-14/01.

PHMSA (2011a). *Corrective Action Order CPF No. 2-2011-1011H*. Technical Report Pipeline and Hazardous Materials Safety Administration Washington, USA.

PHMSA (2011b). *Corrective Action Order CPF No. 3-2011-1018H*. Technical Report Pipeline and Hazardous Materials Safety Administration Washington, USA.

PHMSA (2011c). *Tennessee Gas Pipeline Co. - Internal Corrosion, East Bernard, Texas. December 8, 2010*. Technical Report Pipeline and Hazardous Materials Safety Administration Washington, USA. Failure Investigation Report.

PHMSA (2011d). *Tennessee Gas Pipeline Co. - Wrinkle Bend Failure, Natchitoches Parish, LA. November 30, 2010*. Technical Report Pipeline and Hazardous Materials Safety Administration Washington, USA. Failure Investigation Report.

PHMSA (2012a). *Corrective Action Order CPF No. 4-2012-1011H*. Technical Report Pipeline and Hazardous Materials Safety Administration Washington, USA.

PHMSA (2012b). *Southern Star Central Rupture, Overpressure in KS. June 30, 2005*. Technical Report Pipeline and Hazardous Materials Safety Administration Washington, USA. Failure Investigation Report.

PHMSA (2012c). *TransCanada/Bison Pipeline Natural Gas Transmission Release near Gillette, WY. July 20, 2011*. Technical Report Pipeline and Hazardous Materials Safety Administration Washington, USA. Failure Investigation Report.

PHMSA (2013a). *Corrective Action Order CPF No. 4-2013-1016H*. Technical Report Pipeline and Hazardous Materials Safety Administration Washington, USA.

PHMSA (2013b). *Tennessee Gas Pipeline Line 100-1, Batesville, Mississippi. November 21, 2011*. Technical Report Pipeline and Hazardous Materials Safety Administration Washington, USA. Failure Investigation Report.

PHMSA (2014). *Notice of Proposed Safety Order CPF No. 3-2014-1008S*. Technical Report Pipeline and Hazardous Materials Safety Administration Washington, USA.

PHMSA (2015a). *Corrective Action Order CPF No. 4-2015-1009H*. Technical Report Pipeline and Hazardous Materials Safety Administration Washington, USA.

PHMSA (2015b). *Notice of Proposed Safety Order CPF No. 2-2015-1001S*. Technical Report Pipeline and Hazardous Materials Safety Administration Washington, USA.

Posada, D., & Buckley, T. (2004). Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. *Systematic Biology*, *53*, 793–808. doi:10.1080/10635150490522304.

Ramírez-Camacho, J., Amaya-Gómez, R., Muñoz, F., Pastor, E., & Casal, J. (2017). Historical analysis and conditional probabilities of accidents involving the formation of a crater by the explosive rupture of buried natural gas pipelines. Unpublished.

Ramírez-Camacho, J., Pastor, E., Casal, J., Amaya-Gómez, R., & Muñoz Giraldo, F. (2015). Analysis of domino effect in pipelines. *Journal of Hazardous Materials*, *298*, 210 – 220. doi:http://doi.org/10.1016/j.jhazmat.2015.05.033.

Rencher, A., & Schaalje, G. (2008). *Linear Models in Statistics*. (2nd ed.). Wiley.

Sheppard, M. (2012). ALLFITDIST: Fit all valid parametric probability distributions to data. MATLAB Central File Exchange.

Shi, H., Zhang, W., Yu, Z., & Huang, L. (2012). Design of Parallel Pipelines in Second West-East Natural Gas Pipeline Project. In *2012 9th International Pipeline Conference* (pp. 139–148). Calgary Alberta, Canada. doi:10.1115/IPC2012-90268.

Silva, E., Nele, M., Frutuoso e Melo, P., & Könözsy, L. (2016). Underground parallel pipelines domino effect: An

analysis based on pipeline crater models and historical accidents. *Journal of Loss Prevention in the Process Industries*, *43*, 315 – 331. doi:http://doi.org/10.1016/j.jlp.2016.05.031.

Sklavounos, S., & Rigas, F. (2006). Estimation of safety distances in the vicinity of fuel gas pipelines. *Journal of Loss Prevention in the Process Industries*, *19*, 24 – 31. doi:https://doi.org/10.1016/j.jlp.2005.05.002.

Stancescu, D. (2014). Fitting Distributions to Dose Data. https://www.cdc.gov/niosh/ocas/pdfs/dps/dc-fitdist050114.pdf.

Stewart, M. (2016). *Surface Production Operations* volume III: Facility Piping and Pipeline Systems. Boston: Gulf Professional Publishing. doi:https://doi.org/10.1016/B978-1-85617-808-2.09992-2.

Symonds, M., & Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, *65*, 13–21. doi:10.1007/s00265-010-1037-6.

TSB (1997). *TransCanada PipeLines Limited Line 100-3, 914-millimetre (36-inch) Main Line Kilometre Post Main Line Valve 30-3 + 0.245 kilometres Line 100-4, 1,067-millimetre (42-inch) Main Line Kilometre Post Main Line Valve 30-4 + 0.220 kilometres, Rapid City, Manitoba, 29 July 1995*. Technical Report Transportation Safety Board Canada. Commodity Pipeline Occurrence Report P95H0036.

TSB (2009). *Natural gas pipeline rupture TransCanada Pipeline Inc. 914-Millimeter-Diameter Pipeline Line 2 MLV 107-2 + 6.031 KM Near Englehart, Ontario, 12 September 2009*. Technical Report Transportation Safety Board Canada. Pipeline Investigation Report P09H0074.

TSB (2011). *TransCanada PipeLines Limited 914.4 -Millimetre-Diameter Pipeline Line 100-2 MLV 76-2 + 09.76 KM, Beardmore, Ontario, 19 February 2011*. Technical Report Transportation Safety Board Canada. Pipeline Investigation Report P11H0011.

TSB (2013a). *TransCanada PipeLines Limited (NOVA Gas Transmission Ltd.) North Central Corridor Loop (Buffalo Creek West Section) Chainage 27+996 m near Fort McMurray, Alberta, 17 October 2013*. Technical Report Transportation Safety Board Canada. Pipeline Investigation Report P13H0107.

TSB (2013b). *Westcoast Energy Inc., carrying on business as Spectra Energy Transmission Nig Creek Pipeline, Kilometre Post 1.93 near Buick, British Columbia, 28 June 2012*. Technical Report Transportation Safety Board Canada. Pipeline Investigation Report P12H0105.

TSB (2014). *Natural gas pipeline rupture TransCanada Pipe Lines Limited Line 400-1 at Mainline Valve Site 402 near Otterburne, Manitoba, 25 January 2014*. Technical Report Transportation Safety Board Canada. Pipeline Investigation Report P14H0011.

Yan, X., & Su, X. (2009). *Linear Regression Analysis: Theory and Computing*. World Scientific Publishing Company Pte Limited.

Zhao, X., Zhang, G., Luo, J., & Zhang, H. (2012). Impact of improving design factor over 0.72 on the safety and reliability of gas pipelines and feasibility justification. *Chinese Journal of Mechanical Engineering*, *25*, 166–172. doi:10.3901/CJME.2012.01.166.

Appendix A. Dataset

Table A.12 shows the processed records from each dataset in Fig. 5. This table includes the corresponding ID from the review in Ramírez-Camacho et al. (2017) and the initial source.

Table A.12: Records obtained from each dataset

	Set	ID	Date	Place		D(in)	t(m)	P(bar)	D^*(m)	W(m)	WD	REF
S1	4		04/03/1965	Natchitoches, LA, USA	24	0.250	54.6	5.024	9.0	1.791		(HSE, 2002; MHIDAS, 2007)
S1	7		02/03/1974	Monroe, LA, USA	30	0.438	56.0	7.326	9.1	1.242		(HSE, 2002)
S1	8		15/03/1974	Farmington, NM, USA	12	0.250	34.9	3.459	5.2	1.503		(HSE, 2002)
S1	9		21/05/1974	Meridian, MS, USA	6.6	0.071	21.1	1.982	3.0	1.514		(MHIDAS, 2007)
S1	10		09/06/1974	Bealeton, VA, USA	30	0.312	50.5	2.776	11.0	3.962		(HSE, 2002)
S1	11		09/08/1976	Cartwright, LA, USA	20	0.250	54.1	3.872	7.6	1.963		(HSE, 2002)
S1	14		04/11/1982	Hudson, IA, USA	20	0.281	57.7	3.272	15.0	4.584		(BAM, 2009; HSE, 2002)

Table A.12: continued from previous page

Set	ID	Date	Place		D(in)	t(m)	P(bar)	D^*(m)	W(m)	WD	REF
S1	15	25/03/1984	Erlangen-Eltersdorf, Germany	28	0.276	67.5	6.626	30.0	4.528		(HSE, 2000)
S1	17	25/11/1984	Jackson, LA, USA	30	0.311	71.4	3.776	7.6	2.012		(HSE, 2002)
S1	19	27/04/1985	Beaumont, KY, USA	30	0.469	69.7	3.576	11.6	3.243		(HSE, 2000)
S1	21	21/02/1986	Lancaster, KY, USA	30	0.375	69.4	1.676	9.1	5.428		(HSE, 2000)
S1	23	06/06/1990	Marionville, ON, Canada	12.7	0.252	47.0	1.737	1.5	0.864		(HSE, 2002)
S1	26	15/07/1992	Potter, ON, Canada	36	0.360	69.0	5.400	13.6	2.519		(HSE, 2002; MHIDAS, 2007)
S1	27	22/12/1993	Palaceknowe, Moffat, Scotland	36	0.752	48.0	2.829	10.0	3.535		(MHIDAS, 2007; BAM, 2009)
S1	30	23/03/1994	Edison, NJ, USA	36	0.675	68.2	2.429	20.0	8.235		(HSE, 2000)
S1	31	23/07/1994	Latchford, ON, Canada	36	0.360	69.0	4.900	16.0	3.265		(HSE, 2000)
S1	33	29/07/1995	Rapid City, MB, Canada	42	0.371	60.7	2.981	23.0	7.715		(TSB, 1997)
S1	34	15/04/1996	St. Norbert, MB, Canada	34	0.500	50.0	5.478	13.5	2.464		(HSE, 2000)
S1	36	19/08/2000	Carlsbad, NM, USA	30	0.335	47.0	5.576	15.5	2.780		(NTSB, 2003)
S1	44	23/03/2003	Eaton, CO, USA	24	0.250	56.0	3.824	15.0	3.923		(BAM, 2009)
S1	45	01/05/2003	Pierce County, WA, USA	26	0.281	43.6	4.275	6.1	1.427		(MHIDAS, 2007)
S1	48	30/07/2004	Ghislenghien, Belgium	40	0.512	80.0	4.830	14.0	2.898		(BAM, 2009)
S1	54	25/08/2008	Pilot Grove, MO, USA	24	0.281	55.2	1.924	10.0	5.198		(DOT, 2008)
S1	57	04/05/2009	Palm City, FL, USA	18	0.250	58.9	3.072	5.2	1.693		(NTSB, 2013)
S1	66	20/07/2011	Gillette, WY, USA	30	0.438	92.4	3.476	7.5	2.157		(PHMSA, 2012c)
S1	69	21/11/2011	Batesville, MS, USA	24	0.250	51.6	2.424	23.8	9.818		(PHMSA, 2013b)
S1	74	28/06/2012	Buick, BC, Canada	16	0.250	66.6	1.921	7.6	3.957		(TSB, 2013b)
S1	77	08/10/2013	Harper County, OK, USA	30	0.344	55.8	4.776	9.1	1.905		(PHMSA, 2013a)
S1	85	27/06/2014	East Godavari, AP, India	18	0.133	44.1	3.372	7.0	2.076		(Mishra & Wehrstedt, 2015)
S1	89	17/04/2015	Fresno, CA, USA	12	0.254	88.3	2.319	5.6	2.415		(Exponent, 2015)
S2	18	10/03/1985	Ignace, ON, Canada	36	0.360	66.5	3.000	10.6	3.533		(HSE, 2002; BAM, 2009)
S2	24	15/01/1991	Cochrane, ON, Canada	30	0.360	63.1	7.000	33.0	4.714		(HSE, 2002)
S2	25	08/12/1991	Cardinal, ON, Canada	20	0.252	63.4	2.700	9.0	3.333		(HSE, 2002)
S2	40	15/03/2002	Iron County, MI, USA	36	0.375	51.7	9.100	9.1	1.000		(DOT, 2002)
S2	43	02/02/2003	near Viola, IL,USA	24	0.312	56.0	7.600	12.0	1.579		(MHIDAS, 2007; BAM, 2009)
S2	55	14/09/2008	Appomattox, VA, USA	30	0.344	55.1	4.600	11.3	2.457		(PHMSA, 2008)
S2	68	16/11/2011	Glouster, OH, USA	36	0.344	52.5	4.600	9.1	1.978		(PHMSA, 2011b)
S2	70	03/12/2011	Marengo County, AL, USA	36	0.331	54.8	4.300	16.8	3.907		(PHMSA, 2011a)

Table A.12: continued from previous page

Set	ID	Date	Place		*D*(in)	*t*(m)	*P*(bar)	*D**(m)	*W*(m)	WD	REF
S2	75	11/12/2012	Sissonville, WV, USA	20	0.281	64.1	4.300	10.9	2.535		(NTSB, 2014)
S2	78	17/10/2013	Fort McMurray, AB, Canada	36	0.465	92.0	5.000	15.0	3.000		(TSB, 2013a)
S2	80	25/01/2014	Otterburne, MB, Canada	30	0.370	63.3	3.000	12.5	4.167		(TSB, 2014)
S3	5	22/02/1973	Austin, TX, USA	10.8	0.373	36.9	1.205	3.1	2.530		(HSE, 2002)
S3	46	13/12/2003	Toledo, WA, USA	26	0.281	35.0	5.708	15.4	2.698		(HInt Dossier, 2005)
S3	50	30/06/2005	Douglas County, KS, USA	20	0.312	47.0	2.856	6.1	2.136		(PHMSA, 2012b)
S3	60	12/09/2009	near Englehart, ON, Canada	36	0.400	68.7	2.933	6.1	2.080		(TSB, 2009)
S3	61	05/11/2009	Bushland, TX, USA	24	0.250	53.0	4.324	12.9	2.983		(PHMSA, 2009)
S3	62	09/09/2010	San Bruno, CA, USA	30	0.375	25.9	3.410	7.9	2.317		(NTSB, 2011)
S3	63	30/11/2010	Natchitoches Parish, LA, USA	30	0.312	46.3	1.710	4.6	2.690		(PHMSA, 2011d)
S3	64	08/12/2010	East Bernard, TX, USA	24	0.500	49.6	3.057	7.6	2.486		(PHMSA, 2011c)
S3	65	19/02/2011	Beardmore, ON, Canada	36	0.360	66.2	5.233	13.0	2.484		(TSB, 2011)
S3	79	29/11/2013	Houstonia, MO, USA	30	0.312	61.6	3.843	9.5	2.472		(PHMSA, 2014)
S4	20	20/08/1985	Lowther, ON, Canada	36	0.360	67.9	4.900	14.7	3.000		(HSE, 2002)
S4	22	02/03/1986	Callander, ON, Canada	36	0.360	62.6	4.000	12.0	3.000		(HSE, 2002)
S4	58	05/05/2009	Rockville, IN, USA	24	0.312	54.6	2.833	8.5	3.000		(DOT, 2009)
S4	73	06/06/2012	Laketon, TX, USA	26	0.250	47.4	1.667	5.0	3.000		(PHMSA, 2012a)
S4	87	14/01/2015	Brandon, MS, USA	30	0.375	52.9	2.667	8.0	3.000		(PHMSA, 2015b)
S4	90	03/08/2015	Falfurrias, TX, USA	16	0.250	57.0	3.033	9.1	3.000		(PHMSA, 2015a)