

STATISTICAL CALIBRATION OF PIPELINE IN-LINE INSPECTION DATA

J. M. Hallen¹, F. Caleyó¹, L. Alfonso¹, J. L. González¹ and E. Pérez-Baruch²

¹ ESIQIE-IPN, México D.F., México; ² PEMEX-PEP-RS, Tabasco, México

Abstract: This paper describes a statistical methodology for the calibration of in-line inspection (ILI) tools which is based on the comparison of the ILI readings with field-measurements results. The systematic and random errors that affect the ILI and the field tools are estimated and from this information, an unbiased estimation of the true depth of the defects detected by the ILI tool is produced. The influence of the number of field verifications on the reliability of the calibration process is addressed. The methodology is tested through Monte Carlo simulations and illustrated using a real-life ILI dataset produced by an UT tool.

Introduction: Today in-line inspection (ILI) tools used to detect, locate and size pipeline anomalies such as dents, cracks and corrosion metal loss are based on the magnetic flux leakage (MFL) and ultrasonic (UT) principles [1]. The information provided by ILI tools consists of geometrical data describing each detected anomaly, *e.g.* its length, depth, width and orientation. This information is affected by built-in measurement errors, both systematic and random, that have to be considered when the ILI data is used to conduct integrity studies and fitness-for-purpose investigations.

Pipeline integrity analysts and ILI vendors are now aware of the fact that in assessing the severity of a defect, the key issue is how accurately its geometry has been measured. In this sense, the two important parameters that ILI vendors provide are the probability of detection (POD) and the sizing accuracy of the tool.

The sizing accuracy associated with an ILI tool is quoted as an accuracy level and a percent confidence. The depth sizing accuracy associated with corrosion metal loss of today high resolution (HR) MFL tools is typically claimed to be $\pm 10\%$ of the pipe wall thickness (wt) with a confidence level of 80%. In the case of extra high resolution (XHR) MFL tools, the sizing accuracy is $\pm 5\%$ wt at 80% confidence. On the other hand, today HR and XHR UT tools are claimed to have a sizing accuracy of ± 0.6 mm and ± 0.3 mm at a confidence level of 80%, respectively [1].

Pipeline operators assess the accuracy of a pipeline inspection through the statistical comparison of the metal loss sizes predicted by the ILI tool with the sizes obtained through field inspections.

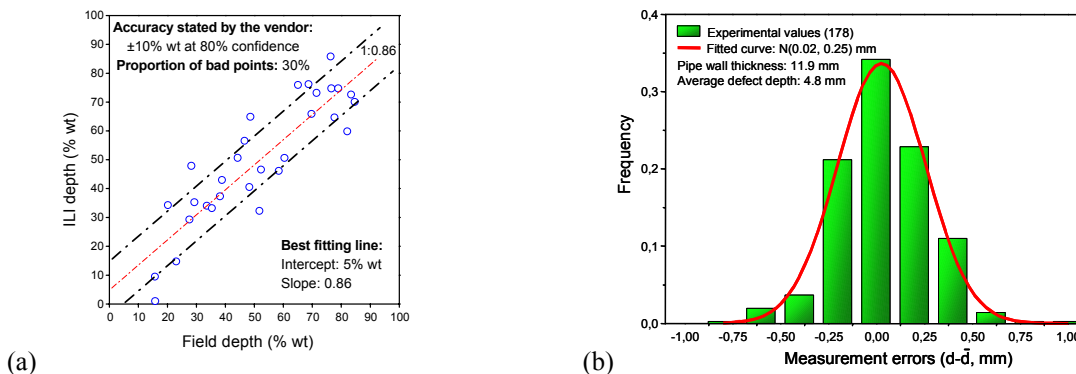


Figure 1 (a) Typical plot of ILI depth readings against field depth measurements. (b) Distribution of measured depths obtained for a single internal metal loss using a portable UT flaw detector.

If the accuracy quoted for the ILI tool is achieved during the inspection, then the ILI vs. Field depths plot will be fitted to a straight line with slope=1.0 and about 80% of the comparison points will fall within the sizing tolerance imposed by the confidence level quoted for the tool. Many times, the ILI data is affected by systematic errors in the form of a constant bias (additive error) and/or in the form a non-constant bias (multiplicative error). This latter case is illustrated in Fig. 1a for a MFL tool with a sizing accuracy of $\pm 10\%$ wt at 80% confidence.

Not seldom, researchers assume that the field tool has no errors and estimate the slope of the best fitted line in Fig. 1a using the ordinary least squares (OLS) regression model. However, it is widely recognized that field instruments, e.g. pit depth gages, portable UT flaw detectors, laser scanners and bar bridging systems, have significant measurement errors. Fig. 1b shows the distribution of 178 depth measurements conducted by an experienced field crew on an internal metal loss using a portable UT flaw detector (0.025 mm resolution).

From these results, it seems justified to postulated that the field instruments have measurement errors that distribute following a normal probability function with mean ≈ 0 (no bias) and variance that strongly depends on the tool type. For internal metal loss, the 80% tolerance for a single depth reading is about ± 0.3 mm while for external metal loss, this tolerance increases to ± 0.5 mm for pit gages with a resolution of 0.4 mm (± 1 mm for pit gages with 0.8 mm resolution). Taking into account the sizing accuracy claimed for the ILI tools, it can be concluded that in some particular situations, the measurement errors of the field tool are comparable or even larger than those of the ILI tool. This situation will arise more likely for ILI runs conducted using XHR tools. If the errors of the field instrument are accounted for, then the OLS regression model is not longer valid to estimate the slope of the best-fit line in Fig. 1a. In this situation, the accuracy and precision of the ILI tool are better estimated using error-in-variables (EIV) models [2]. The classical EIV methods allow to deal with the problem of estimating the slope of the best-fit line in Fig. 1a when both the field and the ILI tools are affected by errors and the ratio of the error variances or one of these variances is known.

However, in many practical cases, the analyst faces the challenge of estimating simultaneously the parameters of the best-fit line and the variance of the errors that affect both tools. This problem has been addressed in recent publications [3-5] which focus on the estimation of the variance of the measurement errors of the ILI and field tools using methods available in the literature such as the Grubbs and Jaech-CELE estimators [6]. A new Bayesian method, capable to overcome some of the limitations of these estimators, has been also proposed [5]. Nevertheless, the bias between the ILI depth readings and the field measurements is assumed in these works to be constant so that the not uncommon situation in which the comparison of ILI with field results obeys a non-constant bias model (Fig. 1a) has not been addressed yet. In addition, consistent procedures for the statistical calibration of the ILI tools are still missing in the literature.

The statistical calibration model: Calibration is the process whereby the scale of a measuring tool is determined on the basis of an informative or calibration experiment. Since the field measurements have error, the calibration of the ILI tool is a comparative calibration process. The calibration experiment follows the model [2]:

$$d_{ILI} = \alpha_{IF} + \beta_{IF}d_{True} + \varepsilon_{ILI}, \quad d_{Field} = d_{True} + \varepsilon_{Field} \quad \text{and} \quad (\varepsilon_{Field}, \varepsilon_{ILI})^T \sim NI(0, \text{diag}(\sigma_{Field}, \sigma_{ILI})) \quad (1)$$

where d_{True} is the true defect depth, d_{ILI} is the ILI depth, d_{Field} is the field depth, α_{IF} and β_{IF} are the intercept and slope associated with the non-constant bias of the ILI tool and ε_{ILI} and ε_{Field} are the random errors associated with the ILI and field tools, respectively. This errors are assumed to be uncorrelated with d_{True} and distributed normally and independently ($\sim NI$) with mean=0 (Fig. 1b) and variances σ_{ILI}^2 (ILI tool) and σ_{Field}^2 (field tool). The variable d_{True} will be treated as fixed so that model (1) will be considered as a functional relationship.

On the other hand, the prediction stage of the calibration process is modeled using [2]:

$$\bar{d}_{True} = \bar{\xi}_{FI} + \bar{\eta}_{FI}d_{ILI} \quad \text{with} \quad \bar{V}(\bar{d}_{True}) \sim f(\bar{\sigma}_{ILI}^2, \bar{\sigma}_{Field}^2, \bar{\sigma}_{\eta}^2) \quad (2)$$

where \bar{d}_{True} is the estimator of the defect true depth d_{True} , $\bar{v}(\bar{d}_{\text{True}})$ is the estimated variance of the predicted true depth and $\bar{\xi}_{\text{FI}}$ and $\bar{\eta}_{\text{FI}}$ are the estimator of the intercept and slope of the calibration line, respectively.

It is noted that $\bar{v}(\bar{d}_{\text{True}})$ depends not only on the variance of the errors of the ILI tool but also on the variance of the calibration model. Therefore, it will be always greater than $\bar{\sigma}_{\text{ILI}}^2$ because of the additional (model) error introduced into the analysis during the estimation of the calibration line.

The sampling distribution (d_{Field} , d_{ILI}) does not allow to identify the model (1) because it is not possible to find an unique relationship between the unknown population parameters and the corresponding estimated parameters [2]. Additional information is required in order to produce consistent estimators of d_{True} , α_{IF} , β_{IF} , σ_{ILI} and σ_{Field} . The classical EIV procedures are capable to solve the measurement model (1) only when the ratio of the measurement error variances $\delta = \sigma_{\text{ILI}}^2 / \sigma_{\text{Field}}^2$ or one of these variances is known [2]. However, within the context of the statistical calibration of the ILI tool, none of these specification are available, so that the classical EIV methods can not be used. Indeed, the sizing accuracy of the ILI tool needs to be corroborated.

In the next sections, a new methodology capable to solve models (1) and (2) consistently will be described and illustrated using both Monte Carlo simulations and a real-life case study.

Results and discussion: Figure 3 shows a flowchart describing the calibration methodology proposed in this work. Each one of the stages in this figure will be outlined in the following sections.

Estimation of the systematic measurement errors: The slope and intercept of the comparison plot are estimate using the Wald's grouping method in which β_{IF} is found by partitioning the data into two subsets and passing a straight line through the mean points of these subsets. The Wald's estimator of the slope and intercept of the fitted line is [8]:

$$\bar{\beta}_{\text{IF}} = \frac{\langle d_{\text{ILI}} \rangle_2 - \langle d_{\text{ILI}} \rangle_1}{\langle d_{\text{Field}} \rangle_2 - \langle d_{\text{Field}} \rangle_1} \quad \text{and} \quad \bar{\alpha}_{\text{IF}} = \langle d_{\text{ILI}} \rangle - \bar{\beta}_{\text{IF}} \langle d_{\text{Field}} \rangle \quad (3)$$

where $\langle d_{\text{ILI}} \rangle_i$ and $\langle d_{\text{Field}} \rangle_i$ are the mean values of the d_{ILI} and d_{Field} readings in the two data subsets ($i = 1, 2$) and $\langle d_{\text{ILI}} \rangle$ and $\langle d_{\text{Field}} \rangle$ are the mean values of these readings for the entire dataset.

This estimator is consistent with the true slope in model (1) if the grouping is independent of the errors and the means of the true values in each groups remain different as the number of observations approaches infinity [1,8].

In this work, a modification of the classical Wald's estimator is introduced to guarantee that the above conditions are satisfied even if the field readings show large measurement errors. When it is expected that $\delta = \sigma_{\text{ILI}}^2 / \sigma_{\text{Field}}^2 \geq 1$, the median of the field readings is used to group the sampling data. Conversely, under the assumption that $\delta < 1.0$, the grouping is done using the median of the ILI readings. It is noted that, in contrast to the classical EIV method, only an estimator of the order of the ratio $\sigma_{\text{ILI}}^2 / \sigma_{\text{Field}}^2$ is required in this modified Wald (M-Wald) method. Table 1 lists the expected range for δ (δ^*) assumed in this work for typical ILI and field tools. These predictions can be modified to consider other field tools such as laser scanners and bridging bar systems.

Monte Carlo simulations were used to evaluate the performance of the outlined M-Wald approach with respect to that of the classical EIV method with δ known. Figure 4 shows the results of this comparison when $\delta^* < 1.0$ for 2000 samples of size 30 created using the population parameters given in this figure. The performance of the M-Wald estimator is similar to that of the classical EIV estimator yet it requires only the information given in Table 1 for δ^* and not the exact value of δ .

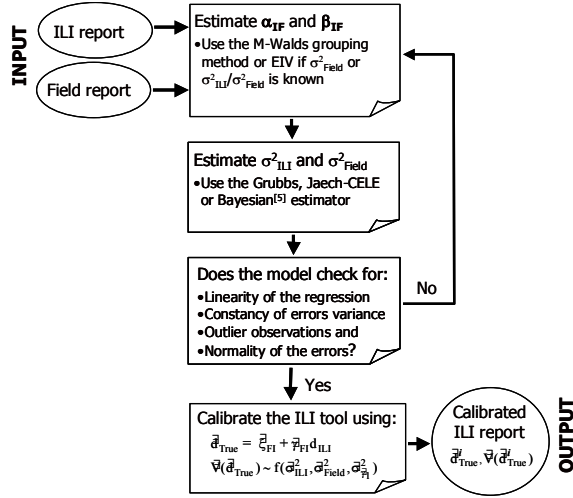


Figure 3. The statistical calibration framework.

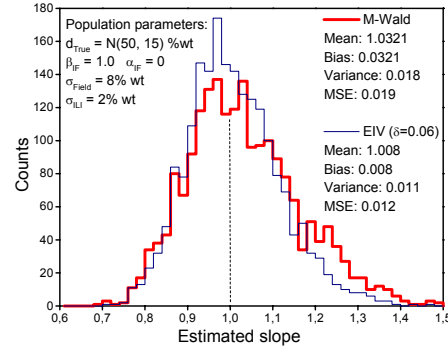


Figure 4. Monte Carlo estimations produced by the M-Wald and EIV methods when $\delta^* < 1$.

Table 1. Expected ranges for $\delta = \sigma_{ILI}^2 / \sigma_{Field}^2$ for typical ILI and field tools.

		ILI inspection tool			
		HR MFL	HR UT	XHR MFL	XHR UT
Defect type and field inspection tool	Internal (UT, 0.025 mm)	$\delta^* > 1.0$	$\delta^* > 1.0$	$\delta^* > 1.0$	$\delta^* \approx 1.0$
	External (Pit gage, 0.4 mm)	$\delta^* > 1.0$	$\delta^* \approx 1.0$	$\delta^* \approx 1.0$	$\delta^* > 1.0$
	External (Pit gage, 0.8 mm)	$\delta^* \approx 1.0$	$\delta^* \approx 1.0$	$\delta^* < 1.0$	$\delta^* < 1.0$

Estimation of the random measurement errors: Several methods are on hand to estimate the variance of the measurement errors [2-6]. For the two instrument case, one reading by each, the classical estimators are those proposed by Grubbs and Jaech [6]. For the non-constant bias model, the Grubbs estimator is:

$$\hat{\sigma}_{Field}^2 = mxx - mxy / \bar{\beta}_{IF} \text{ and } \hat{\sigma}_{ILI}^2 = myy - mxy \bar{\beta}_{IF} \quad (4)$$

In many practical situations, this estimator produces negative error variances. In such cases, the constrained expected likelihood (CELE) estimator proposed by Jaech [6] can be used. A modification of the Jaech's estimator is proposed here in order to account for the situations where the non-constant bias model applies:

$$\hat{\sigma}_{Field}^2 = SI_0 I^{-1}, \quad \hat{\sigma}_{ILI}^2 = S - \hat{\sigma}_{Field}^2, \quad S = mxx + myy - mxy(1 + \bar{\beta}_{IF}^2) \bar{\beta}_{IF}^{-1}, \quad (5)$$

$$I_0 = \int_0^1 xf(x)dx, \quad I_1 = \int_0^1 f(x)dx \text{ and } f(x) = [mxx(1-x)^2 + x^2myy + x mxy[2-x(1+\bar{\beta}_{IF}^2)]] \bar{\beta}_{IF}^{-1}$$

where x is a dummy variable. As in (4), $\bar{\beta}_{IF}$ is assumed to be 1.0 if the constant bias model is used.

The superiority of the M-Wald+M-Jaech estimators over the estimators available in the literature becomes more evident in the situations where $\beta_{IF} \neq 1.0$. For example, consider a typical situation where a XHR MFL inspection is performed on a 11mm wt pipeline and field verifications are conducted for 30 external metal loss using a 0.4 mm (1/64") pit gage. In this case, $\delta^* \approx 1.0$ since $\sigma_{Field} \approx 4\%$ wt and $\sigma_{ILI} \approx 5\%$ wt. If additionally, it is supposed that the ILI tool underestimated the defect depth and this can be modeled with $\beta_{IF} = 0.75$, then the non-constant bias model seems more adequate to deal with this statistical comparison.

This situation was studied using Carlo simulations for 2000 samples of size 30. Expression (5) was used to estimate the measurement errors for both the constant ($\beta_{IF}=1.0$) and non-constant bias approximations (Fig. 4). For both tools, the non-constant bias model (M-Wald+M-Jeach) allows to estimate the measurement error with a smaller bias. Although the constant bias model produces a smaller variance in the estimation of σ^2_{ILI} , the bias in this estimation doubles that predicted by the non-constant bias solution.

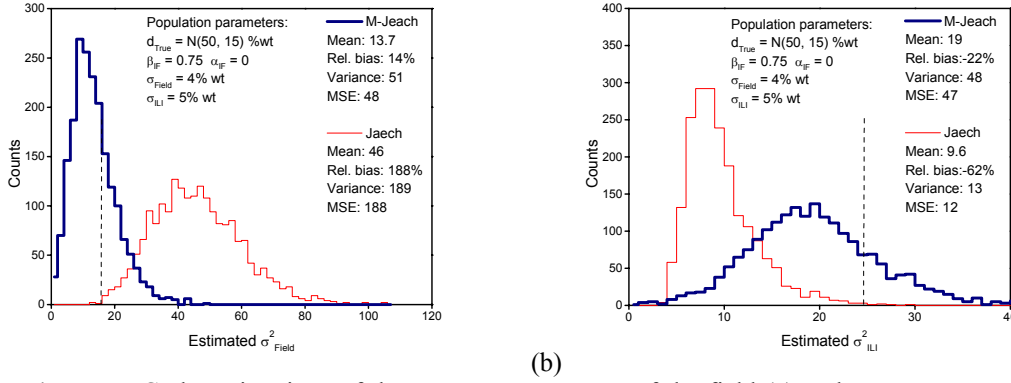


Figure 4. Monte Carlo estimations of the measurement errors of the field (a) and ILI (b) tools under the constant (Jeach) and non-constant bias (M-Jeach) assumptions.

Model checks: In this stage, the solution proposed for model (1) is inspected in order to detect faulty conditions such as non-linearity in the regression, lack of variance homogeneity, outlier observations and non-normality of the errors. First, the true defect depths and the residuals \bar{r}_i from the fitted EIV model are estimated as [2]:

$$\bar{d}_{True} = d_{Field} + \bar{r}_i \frac{\bar{\beta}_{IF} \bar{\sigma}_{Field}^2}{\bar{\beta}_{IF}^2 \bar{\sigma}_{Field}^2 + \bar{\sigma}_{ILI}^2}, \quad \bar{r}_i = d_{ILI} - \bar{\alpha}_{IF} - \bar{\beta}_{IF} d_{Field} \quad \text{and} \quad \bar{V}(\bar{d}_{True}) = \bar{\sigma}_{Field}^2 - \frac{\bar{\beta}_{IF}^2 \bar{\sigma}_{Field}^4}{\bar{\beta}_{IF}^2 \bar{\sigma}_{Field}^2 + \bar{\sigma}_{ILI}^2} \quad (4)$$

Then, the estimated residuals are plotted against the estimated true depths in order to check the linearity in the regression and the variance homogeneity postulated in model (1). The normality of the errors is investigated by plotting the ordered residuals against the expected value of the normal order statistic for a sample of the same size. The outlier observations are identified using the MM-estimate [8] which shows a high breakdown point and an excellent efficiency when the errors have normal distribution.

Calibration of the ILI tool: The goal of the prediction stage of the calibration process is the estimation of true value of the metal loss penetration from the ILI readings. The calibration parameters are estimated as [2]:

$$\bar{\varphi}_{FI} = \begin{cases} \frac{R}{S} \frac{(m_{xx} - \bar{\sigma}_{Field}^2) m_{xy} + 2(n-1)^{-1} m_{xy} \bar{\sigma}_{Field}^2}{m_{xy}^2 + (n-1)^{-1} (m_{xx} m_{yy} - m_{xy}^2)} & \text{if } \bar{\kappa} > 1.0 \\ \frac{R}{T} \frac{m_{xy} m_{xy}^{-1}}{m_{xy} m_{xy}^{-1}} & \text{if } \bar{\kappa} \leq 1.0, \end{cases} \quad (5)$$

$$\bar{\zeta}_{FI} = \langle d_{Field} \rangle - \bar{\varphi}_{FI} \langle d_{ILI} \rangle \quad \text{and} \quad \bar{\kappa} = (m_{xx} - m_{yy}^{-1} m_{xy}^2) / \bar{\sigma}_{Field}^2$$

If the calibration experiment is carried out using a sample of size n , then the estimator of the true defect depth for the $(n+1)$ th ILI reading and the estimated variance of this estimator are [2]:

$$\begin{aligned} \bar{d}_{\text{True}}^{n+1} &= \bar{\zeta}_{\text{FI}} + \bar{\gamma}_{\text{FI}} d_{\text{ILI}}^{n+1}, \quad \bar{V}(\bar{d}_{\text{True}}^{n+1}) = (n-1)^{-1} s_{\text{bb}} + \bar{V}(\bar{\gamma}_{\text{FI}})[(d_{\text{Field}}^{n+1} - \langle d_{\text{Field}} \rangle)^2 - \bar{\sigma}_{\text{ILI}}^2] + \bar{\gamma}_{\text{FI}}^2 \bar{\sigma}_{\text{ILI}}^2, \\ \bar{V}(\bar{\gamma}_{\text{FI}}) &= (n-1)^{-1} m\bar{x}\bar{x}^{-2} [(m\bar{x}\bar{x} + \bar{\sigma}_{\text{Field}}^2) s_{\text{bb}} + \bar{\sigma}_{\text{Field}}^2] \bar{\gamma}_{\text{FI}}^2, \\ s_{\text{bb}} &= (n-1)^{-1} \sum_{i=1}^n [(d_{\text{Field}}^i - \langle d_{\text{Field}} \rangle) - \bar{\gamma}_{\text{FI}}(d_{\text{ILI}}^i - \langle d_{\text{ILI}} \rangle)]^2 \text{ and } m\bar{x}\bar{x} = \begin{cases} m\bar{x}\bar{x} - \bar{\sigma}_{\text{Field}}^2 & \text{if } \bar{\lambda} > 1 \\ m\bar{x}\bar{y}^2 m\bar{y}\bar{y}^{-1} & \text{otherwise} \end{cases} \end{aligned} \quad (6)$$

The estimator of d_{True} is claimed to be unbiased under the fixed model [2]. The variance $\bar{V}(\bar{d}_{\text{True}}^{n+1})$ is larger than $\bar{\sigma}_{\text{ILI}}^2$ because it depends not only on the measurement errors but also on the errors introduced into the analysis by the calibration procedure. If γ_{FI} is known, then the best estimator of $\bar{V}(\bar{d}_{\text{True}}^{n+1})$ is $\gamma_{\text{FI}} \bar{\sigma}_{\text{Field}}^2$.

To illustrate the calibration computations, the typical ILI-Field comparison presented in the previous section is considered again, but this time the ILI readings are also used as input to predict the true defect depths. Figure 5b shows the regression of the estimated true depths on the actual true depths.

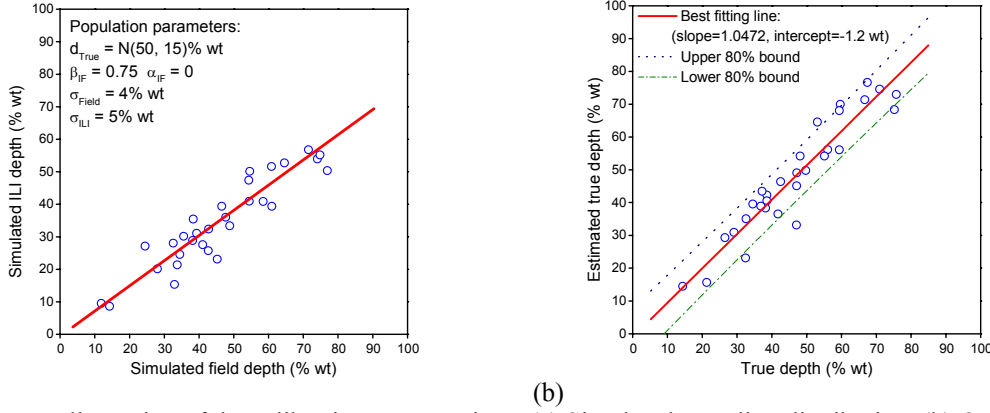


Figure 5. Illustration of the calibration computations. (a) Simulated sampling distribution. (b) OLS regression of the estimated true depth on the actual true depths.

The results presented in this figure show that expressions (5) and (6) estimate consistently both the calibration line and the true values of the defect depths. The error in the determination of $\bar{\gamma}_{\text{FI}}$, relative to the population parameter γ_{FI} is about 3% while the OLS regression of \bar{d}_{True} on d_{True} produces a slope very close to 1.0. The model variance of the OLS regression is 5.7% wt. This value is very close to the average variance estimated for $\bar{d}_{\text{True}}^{n+1}$ (5.5% wt) which results larger than the variance predicted for the ILI tool (4 % wt) as expected.

On the other hand, the number of points that fall outside the 80% tolerance bounds ($CI_{80} = \pm 1.28 \gamma_{\text{FI}} \bar{\sigma}_{\text{Field}}^2$) is 5. On the assumption that the number of points that fall within CI_{80} follows a binomial distribution with $n=30$ and $P=0.8$, the confidence in rejecting this experiment is 40%, which confirms the validity of these computations.

These results were confirmed using 2000 Monte Carlo simulations. The calibration slope γ_{FI} , the slope β_{ols} of the OLS regression of \bar{d}_{True} on d_{True} and the model variance of this regression were estimated to be, respectively, $\langle \bar{\gamma}_{\text{FI}} \rangle = 1.338 \pm 0.006$, $\langle \beta_{\text{ols}} \rangle = 1.004 \pm 0.002$ and $\langle \bar{\sigma}_{\text{ols}}^2 \rangle = 38.3 \pm 0.6$ at 95% confidence. In addition, the confidence in rejecting the calibration experiments was 76% or less, 80% of time.

A generalization of the ILI tool rejection criteria: When numerous field verifications are done, the tolerable number of bad points n_{bad} in the ILI vs. Field depths plot can be predicted under the assumption that the successful verifications show a binomial distribution. A success is defined as

a point that fall within the accuracy tolerance quoted for the ILI tool. If p_s denotes the confidence level used to establish this tolerance and p_{rej} the confidence required to reject the inspection, then n_{bad} can be found using $n_{bad}=n-\text{quant}[\text{BinPDF}(n, p_s), 1-p_{rej}]$, where n is the total number of points, BinPDF is the binomial distribution and $\text{quant}(f,p)$ is to the p -quantile of the f probability function. This expression assumes that the field tool shows no errors.

If the errors of the field instrument are considered, then the value of p_s must be modified to reflect the effect of σ_{Field}^2 on n_{bad} . The new confidence level p_s^* to be used in the expression for n_{bad} is $p_s^* = \text{erf}[\text{inverf}(p_s)\{1+1/\delta'\}^{-1/2}]$, where $\delta' = \sigma_{ILI}^2 / \sigma_{Field}^2$, being σ_{ILI}^2 the variance quoted for the ILI tool (note the tolerance range Δ_θ associated to a given confidence level θ can be found using $\Delta_\theta = \pm \text{inverf}(\theta)\{2(\sigma_{Field}^2 + \sigma_{ILI}^2)\}^{1/2}$).

Obviously, $p_s^* < p_s$ when $\delta'^{-1} \neq 0$ since the errors of the field tool increase the scatter of the comparison points relative to the scatter obtained when $\sigma_{Field}=0$. Accordingly, a larger number of bad points can be tolerate if the errors of the field tool are taken into account. This is shown in Fig. 6 where the dependence of n_{bad} on n is plotted for $\delta'^{-1}=0$, $\delta'^{-1} > 1$ and $\delta'^{-1} < 1$ assuming that the confidence in rejecting the ILI tool is 80%.

Influence of the number of verifications: The influence of the number of field verifications on the reliability of the proposed calibration approach was assessed using Monte Carlo simulations for three different populations: $\{d_{True} \approx N(50,15)\%$, $\beta_{IF}=1.0$, $\alpha_{IF}=0\}$ with $(\sigma_{Field}, \sigma_{ILI})=(3, 8)$, $(4, 5)$ and $(8, 2)\%$ wt. In each case, 2000 experiments were performed for different sample lengths ($n=10, 20, 30, 50$ and 1000). The mean square error $MSE_{\bar{d}}$ associated with the estimated variances $\bar{d}(\bar{d}_{True}^{n+1})$ was computed in each experiment (Fig. 7).

As Fig. 7 shows, the optimum number of field verifications to be conducted is about 30. If the sample length drops below this figure, the estimation errors increase significantly. On the other hand, the quality of the calibration does not increase considerably if the number of field verifications is larger than 30. In addition, Fig. 7 suggests that the most accurate estimations are produced when the errors of the ILI and field tools are similar.

A real-life case study: A 36" OD, 11 mm wt, oil pipeline was inspected using a HR UT ILI tool with 80% tolerance of ± 0.6 mm ($\pm 5.4\%$ wt). A total of 829 external and 101 internal metal loss were detected, located and sized. To calibrate the ILI tool, 47 external metal loss were measured at dig sites using a pit gage with a 0.4 mm resolution. The plot of the ILI readings against the field measurements is shown in Fig. 8a. A first computation cycle allowed to identify point A as an outlier observation. In a second run, the solution listed in Fig. 8a satisfied all the assumptions in model (1). No reasons were found to reject the linearity for the EIV regression model (1) or the normality of the errors (the K-S statistic for the test of normality produced a significance level of 0.11).

The underestimation associated with this ILI run was modeled through a non-constant bias with $\bar{\beta}_{IF} = 0.913$ and $\bar{\alpha}_{IF} = -1.3\%$ wt. The ratio of the estimated measurement error variances was found to be close to 1.0. This value agrees with the predictions given for δ in Table 1. In contrast, in the constant bias solution (relative bias of -3.3% wt), this ratio increased to 3.6 which strongly disagrees with the expected value for δ . This could be misinterpreted as meaning that the sizing accuracy of the ILI tool is much better than that claimed by the vendor. Based on the evidences provided by Fig. 8, the reasonable conclusion is that the ILI tool performs as quoted with respect to the random measurement errors whilst a non-constant bias affects its readings.

The rejection criteria discussed before can be applied to find the degree of confidence with which the ILI run can be rejected. Assuming that the constant-bias model applies and $\sigma_{Field}=0$, the number of bad points at 80% confidence ($\pm 5.4\%$ wt) is 13 (Fig. 8c). This means that the confidence in rejecting the ILI inspection is as high as 87% when this simple model is used. In contrast, if the non-constant bias solution is assumed and the outlier observation A is dropped, the

generalized rejection criteria gives a 80% tolerance of $\pm 6.7\%$ wt which reduces the number of bad points to 6. This time, the confidence in rejecting the ILI run is only 8%.

Once the population parameters in model (1) are consistently estimated, the prediction of the true depth can be carried out for the rest of the defects in the ILI report. In this example, the calibration parameter were estimated to be $\bar{\mu}_{FI} = 1.103$ and $\bar{\alpha}_{FI} = 1.4\%$ wt while the average value of the estimated variance of the predicted true depths was determined to be 3.5% wt. Therefore, the true defect depth associated with each ILI reading was calculated using $\bar{d}_{True} = 1.4 + 1.103d_{ILI}$ and a variance of $12.3(\% \text{ wt})^2$ was assigned to each one of the predicted true depths.

Finally, Fig. 8c shows the distributions of the ILI readings and the calibrated defect depths for the 782 external metal loss found in the inspected pipeline (for the sake of simplicity they are not corrected for the POD factor).

In agreement with the previous results, the distribution of the predicted true depths is shifted to the right as a result of the underestimation produced by the ILI tool and shows a larger spread than the ILI readings.

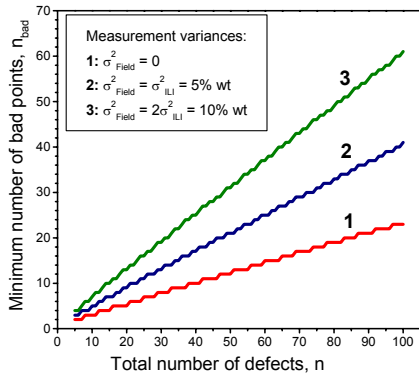


Figure 6. Generalization of the ILI tool rejecting criteria (80% confidence).

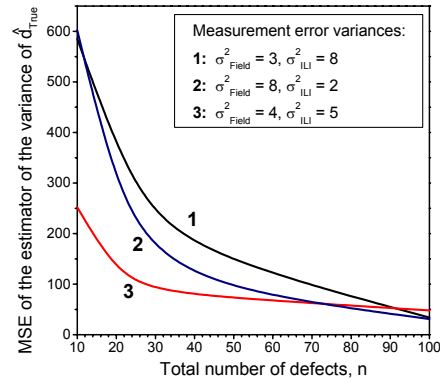


Figure 7. Dependence of the mean square error of the variance estimation on the number of defects.

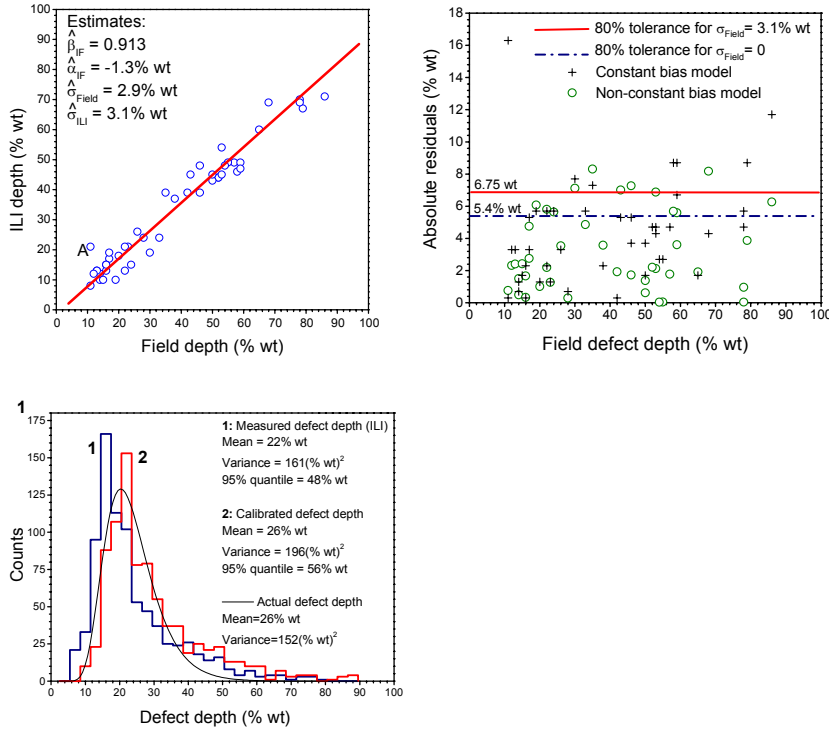


Figure 8. Results of the application of the calibration framework to a real life case study. (a) Sampling distribution. (b) Tool rejection criteria. (c) Measured, calibrated and actual depth distributions of the defect population.

The best fitted probability distribution for these histograms were $\text{LogN}_{\text{ILI}}(22, 13)$ and $\text{LogN}_{\text{True}}(25, 14)$, where $\text{LogN}(\mu_{\text{fit}}, \sigma_{\text{fit}})$ refers to the log-normal distribution with mean μ_{fit} and standard deviation σ_{fit} .

The way these results are used depends on the approach used to perform the probabilistic risk assessment of the pipeline. For instance, suppose that the failure probability of a pipeline segment is to be computed based on a “typical” defect whose depth attribute is defined through the distribution of the depths of all the defects in the pipeline. In such a case, the depth distribution to be used is one of the same kind of that that best fit the ILI readings, yet with variance $\sigma_{\text{fit}}^2 - \sigma_{\text{ILI}}^2$ and mean determined by the measurement model selected. Fig. 8c shows the actual defect depth distribution predicted from the ILI readings assuming a non-constant bias model.

On the other hand, if the failure probability of the segment is to be computed using defect attributes that are defined separately for each defect, then the depth value and variance to be assigned to it are those predicted using (6), *i.e.* $\bar{a}_{\text{True}}^{n+1}$ and $\bar{v}(\bar{a}_{\text{True}}^{n+1})$. Compared with the previous “distribution” approach, this “direct measurement” method is much more accurate in predicting the segment failure probability since it allows to take into account the most critical defects in the tail of the measured and calibrated depth distributions [9].

Conclusions: A new statistical methodology has been developed for the calibration of MFL and UT ILI tools from field verifications. In contrast to the methods so far available in the literature, the methodology proposed and successfully tested here is capable to estimate the measurement errors of the ILI and field tools for both the constant and the non-constant bias EIV model. The information required to identify the EIV model that describes the comparison of the ILI depth readings with the field measurements can be easily derived from the sizing accuracy claimed for the ILI and field tools. New ILI tool rejecting criteria have been proposed and it has been shown

that to reject an ILI run, the measurement errors of the field tool have to be taken into account as they play a key role in computing the number of tolerable bad points in the ILI-Field depth plot. Following the results obtained in this work, the optimum number of field verifications to be conducted is about 30. For this two tools -one measurement each- case, the most accurate results are obtained when both tools have similar errors.

References:

- [1] J. Tiratsoo (Ed.). Pipeline Pigging & Integrity Technology (2003). Clarion Tech. Publishing.
- [2] W. A. Fuller. Measurement Error Models (1987). John Wiley & Sons, Inc. New York.
- [3] A. Bhatia, T. Morrison, N. S. Mangat, Proc. IPC-1998. ASME International. Vol. 1; pp 315.
- [4] T. B. Morrison, N. S. Mangat, G. Desjardins, A. Bhatia. Proc. IPC-2000. ASME International. Vol. 2; pp 839.
- [5] R. G. Worthingham, T. B. Morrison, N. S. Mangat, G. J. Desjardins. Proc. IPC-2002. Paper 27263.
- [6] J. L. Jaech. Statistical Analysis of Measurement Errors (1985). John Wiley & Sons, Inc. New York.
- [8] A. Wald. Annals of Mathematical Statistic (1940). 11; 284.
- [8] V. J. Yohai VJ. The Annals of Statistic (1987) 15; 642.
- [9] F. Caleyó, J. L. González, J. M. Hallen. Int. Journal Pressure Vessels & Piping (2002). 79;77.