Reliability of Flaw Size Calculation based on Magnetic Flux Leakage Inspection of Pipelines

K. REBER, NDT Systens & Services, Stutensee, Germany
A. BELANGER, Tuboscope Pipeline Services, Houston, U.S.A.

Abstract. While data acquisition for MFL-Inspection is not an especially complex method, the off-line analysis and defect size reconstruction requires more sophisticated methods. It has been established that a full reconstruction of the flaw geometry is rarely possible. However the approaches that exist differ by the degree of reliability. The experimental side is represented here by data of intelligent pigs obtained by in-line inspection of pipelines. The problem with in-line inspection is, that the number of flaws usually does not allow for an in-depth treatment of every single item. Lengthy iterative or 3D-FEM solutions are not practical. Methods that have earlier been developed for magnetic flux leakage have more commonly been applied to crack-like features in component testing. Sherbinin [1] has given formulas for the stray flux expected on crack-like features depending on crack opening and depth. In in-line inspection this type of flaw is rarely found due to the small crack opening of typical pipeline cracks. For corrosion-type flaws the profile of the flaw is too irregular to apply closed form formulas. In the in-line inspection industry methods have thus been established, that quickly convert characteristic signal features into measures for flaw depth, flaw length and flaw width. Deconvolution methods with a transfer function are not very common. In this work we compare the efficiency of a method based on a simple signal parameter conversion to a neural network based method. In both cases the parameters are set by actual tool measurements obtained by pulling an inspection tool through a pipe with known artificial defects. It is discussed how the methods perform when interpolating and extrapolating to flaw sizes, that are not previously tested.

1. Introduction

Magnetic flux leakage (MFL) is one of the oldest but at the same time most widely used methods for the inspection of pipelines in operation. While for component testing other methods such as eddy current and ultrasonic testing have become the method of choice, the in-line inspection of pipelines using intelligent pigs continues to be based on MFL for various reasons. Intelligent pigs are tools that operate completely self sustained. No interaction with testing personnel is possible. At the same time the ambient conditions in high pressure natural gas transportation pipelines are quite hostile for subtle measurement methods. In its technical requirements MFL is a quite simple methods. Hall sensors scan a magnetic field, the power to magnetize the steel is delivered by permanent magnets and the pumps that propel the medium and the tool. Hence, power consumption is small. The involved field strength is quite high, the sensors do not need to be especially sensitive compared to other measurements of magnetic stray fields [2]. If the pipe wall is magnetized to a sufficiently high level, influences of rolling texture, residual stresses, material hardening and steel grade can be neglected. Of course at lower magnetization levels, these influences become measurable and can subsequently be used for the quantization of these material parameters [3]. For a repeatable measurement of metal loss type features, however,

a high level of magnetization is known to be important [4]. The improvement of MFL inspection pigs has, among other optimization measures, been focusing on reaching a rather high level of magnetization. A magnetizing field level of 10 kA/m should today be considered a requirement for reliable MFL inspections. For a typical pipeline steel this would result in a magnetization level of 1.8 T [5].

The sophistication of the MFL testing methods lies in the image reconstruction. Since the measured value is nothing but a magnetic field, a mapping of magnetic field onto flaw geometry is required. Since this derivative is inverse compared to traditional mechanical design tasks, the problem is called an inverse problem. Inverse problems in NDE have been discussed in detail. Several methods have been described. Their application to specific techniques and data is still an area of active research and everyday performance improvement.

2. Overview of methods

For further discussion the inverse problem solving methods shall be subdivided into several categories. We can distinguish heuristic methods that neglect the underlying physical phenomenon and physical models that use electromagnetic theory for the solution. The first group can further be divided into simple calibration methods that map the signal using signal processing methods and use an analytical regression method. The second group uses more advanced methods of regression like neural networks. For the physical models we can distinguish between direct inversion methods and iterative approaches that use a forward solution.

2.1 Calibration

This is the most widely used method and at the same time the most simple one. Artificial defects are put into a pipe. They are either drilled, ground or made by electric discharge. The set of defects should comprehend all the shapes that are expected, i.e. deeper and shallower ones as well as internal and external defects. They need to have a well defined size such that an actual length, width and depth can be determined. Areas of general corrosion are thus not suitable for the calibration. For the validation, of course, this type of defect should be considered. It is also typical for MFL that the flux signals strongly depend on the orientation of the defect against the magnetization axis. Often 100 sample defects and more are manufactured. Since the MFL-signals easily overlap, a certain spacing between the defects is required.

The intelligent pig is then pulled through the pipe section with these artificial defects. The signals are recorded and calibration curves of signal versus defect size are set up. An example on how this could be done is shown in **Figure 1**.The benefit of this method is, that a full scale system test is performed at the same time. Often these tests are repeated regularly to check the performance of the tool. If the magnetic circuit has changed for some reason (wear, redesign, etc) this would become clear and a recalibration or refurbishment could be done.

2.2 Neural Networks

Obviously the problem of mapping the MFL-signal onto the actual defect size is a regression problem. We have some bins given by the artificial defects and need to find the defects geometries for all other defects with a somehow similar but different signal. Since the mapping is given by an a priori unknown function and any closed form equations are usually falling short of the underlying complexity of the problem, other means of regression are considered. Upda et al.[6] have proposed to use neural networks for regression.

Different types of neural networks can be considered. In the learning process that network is set up by minimizing the difference between the output of the artificial defects

and their corresponding actual size. For the prediction the signal of an unknown defect is fed into the network and the output delivers an estimated defect size.

2.3 Direct Inversion

The signals received from the MFL-measurement is considered to be a convolution of the actual defect shape and a transfer function. The nature of the transfer function is unknown. However, the results from direct measurement can be used to determine a possible transfer function. Naturally the lack of uniqueness typical for inverse problems does not fully determine the actual transfer function. However, the determination of the defect shape and some heuristic assumption on the defect would reduce the size determination to a deconvolution problem.

2.4 Iterative Inversion

There are several models to calculate the MFL signal based an the actual geometry. For instance, the faces of the defect are modeled with magnetic dipoles of various size and dipole moment to describe the magnetic charge density. The stray field can be calculated on the dipole field distribution. One of the most popular forward methods is finite element modeling. A certain starting geometry is assumed. The forward method is used to calculate the expected MFL-signal. The geometry is then iteratively adapted to generate a field distribution that best fits to the actually measured signal.

3. Performed Tests

An intelligent pig has been sent through a test loop pipe section with artificial defects. The data of altogether 340 defects were recorded. The data of the magnetic stray field values has been extracted for further signal processing. The set has been divided into a training set to describe the forward problem and a validation set to check the performance. The training set comprised 251 samples, the validation set 30 samples. The rest has been discarded. The defects were all external and ranged from a depth of 10 % to 75 % and a size of 6 mm × 6 mm to 76 mm × 76 mm. The aspect ratio ranged from 1 to 10. This set has been used to test different methods

3.1 Parameter Extraction

For a calibration curve an analytic equation is assumed. Among the tree parameters to be calculated, the depth is the most interesting one. The length and the width of defects is usually a more direct conversion of the measured parameter. For the flaw depth the relation between the signal parameter and the actual depth value is very complex. Usually a large number of parameters are extracted from the Flux leakage signal. For this demonstration we want to restrict ourselves to four parameters. The signal amplitude, the width of the signal, the length of the signal and the background value. For the calculation of the signal parameters, the matrix of magnetic field levels is extracted from the tool data. The upper left part of **Figure 1** shows a display of the matrix of magnetic field levels for a 50 mm × 50 mm artificial flaw. The magnetic field extends over a larger area as compared to the actual position of the flaw. The other pictures in **Figure 1** show sample calculations on how a parameters could be extracted from the signal, that describe the MFL-signal. The methods to calculate the parameters will have to be checked on the actual inspection tool. Tools that measure other field components than the axial component will show different signal shapes, which will then require a completely different parameter extraction routine. The number of parameters is a priori not limited to the ones shown. The wall position, i.e. internal, external or mid-wall flaw is also important information. Often the nominal wall thickness is also used as an input. The problem with MFL measurement is that the actual local wall thickness (untouched by corrosion) is not directly measured; however, signal strength and size strongly depend on the actual wall thickness, as this is an integral part of

the magnetic circuit. The background value yields information that is somewhat related to the wall thickness. A histogram of all field values is shown in the lower left part of **Figure 1**. The most measured value, i.e. the maximum in the histogram, corresponds to the light blue color in the contour-scan.

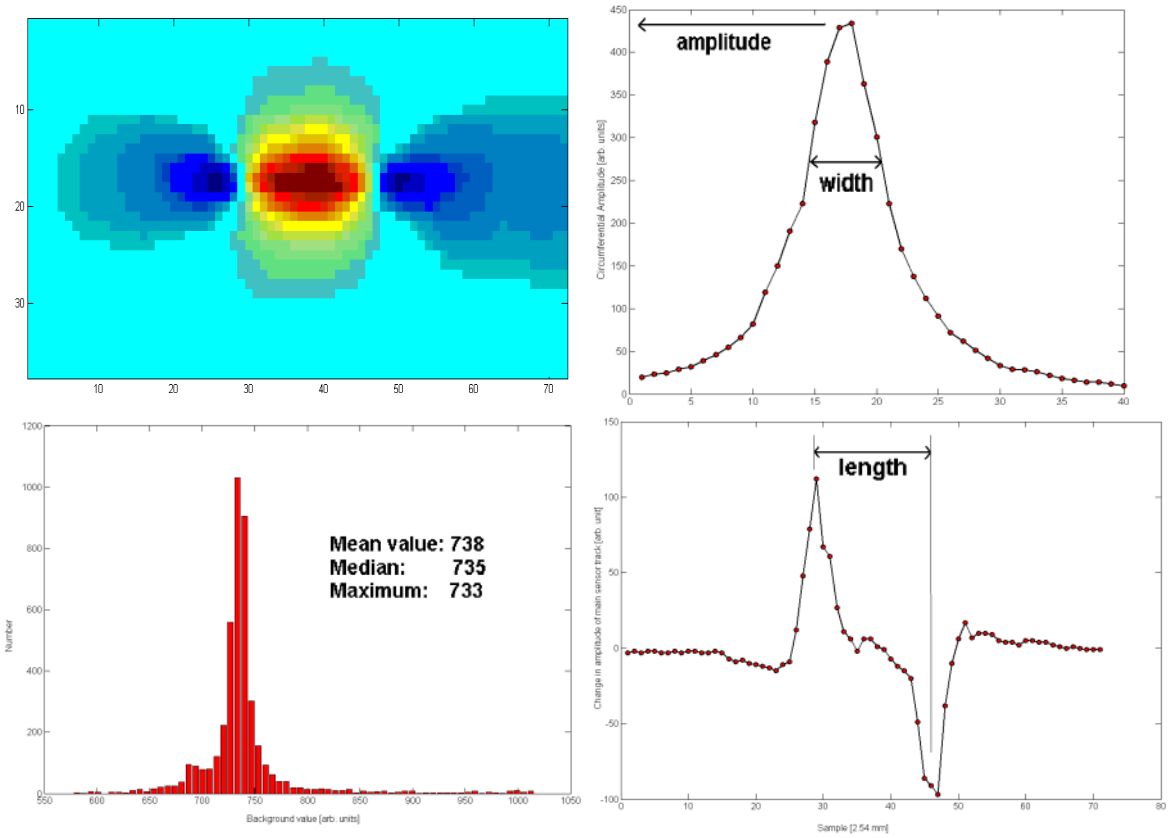

Figure 1: Extraction of signal parameter

The relation between wall thickness and background value is further investigated. The right side of **Figure 2** shows the dependence of the background value on the wall thickness. Thicker walls show lower background value, because more flux is guided by the steel wall. The lower curve shows the amplitude of a 25 mm pit with 50% depth versus the wall thickness. It is seen that the amplitude (signal strength) depends on the wall thickness in a similar manner.

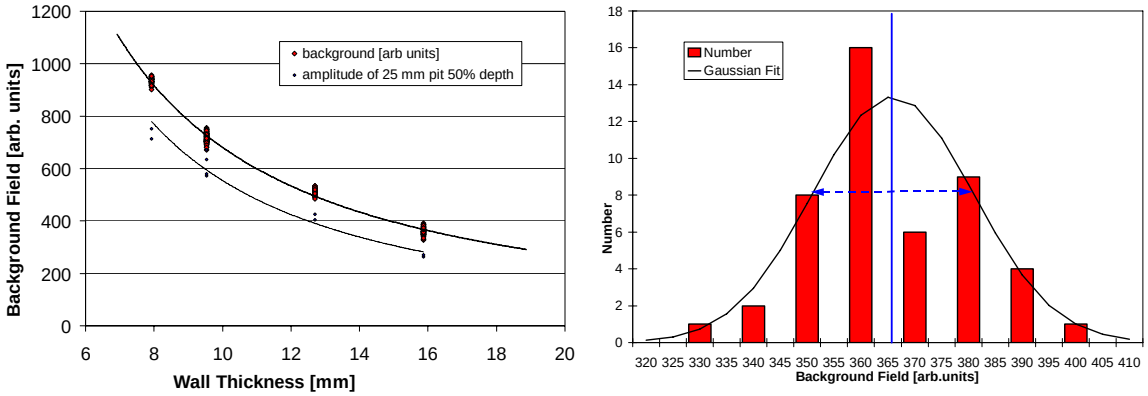

Figure 2: Relation between wall thickness and background value.

The right of **Figure 2** shows a histogram of background values measured in a 15.9 mm wall. The average is found at 366 units, which can be converted into A/m using the Hall sensor sensitivity. The standard deviation is 15 units corresponding to a wall thickness uncertainty of 0.5 mm. It should be mentioned, however, that at the higher speeds usually encountered in real inspections (this data was from a test loop), the background value will vary strongly with the speed. In general the accuracy of the wall thickness measurement should be taken as about ± 1 mm.

Several other tool dependent parameters affect the MFL-signal. The generated flux strongly depends on the employed tool. The sensor lift-off changes the signal strength and the sensor position within the magnetic unit will also influence the signal. These tool dependencies should cancel out if calibration and field measurements are done under the same conditions. Any change of these conditions would require a recalibration. For instance an internal coating acts like a sensor lift-off.

3.2 Calibration

With the signal parameters calculated, the depth is then derived with a calibration curve. Since the dependence is highly non-linear, a relationship is guessed and then adapted to best fit the input data. The relationship could be implemented as a power series. For a demonstration we take the very simple ansatz

$$dep_{predict} = \left(\frac{width}{length}\right)^{a} \frac{amplitude^{b}}{background^{c}} . \qquad (1)$$

The depth is proportional to a fractional power of the amplitude normalized by the background value. The relation is scaled by the aspect ratio. The parameter a, b and c are then optimized by minimizing

$$\sum_{i=1}^{n} (dep^{i}{}_{predict} - dep^{i}_{train})^{2} . \qquad (2)$$

The result is shown in **Figure 3**. For this trial all defects with a width less than 10 mm have been excluded. This size is where the regression is most problematic. Usually this size is stated to be outside of the capabilities of the inspection tool.

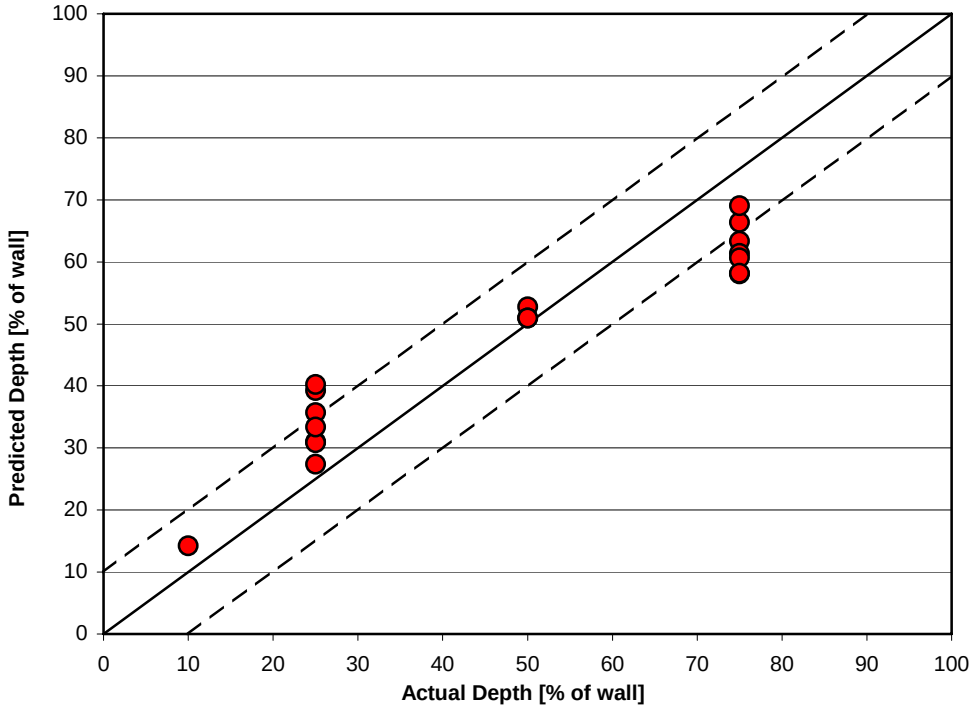

Figure 3: Validation results for a simple parameter regression

3.3 General regression

A way to perform general regression without prior knowledge about the relation of the input to the output is a so-called general regression neural network (GRNN)[7]. Depth prediction is given by

$$dep_{predict} = \frac{\sum_{i=1}^{n} dep_{train}^{i} \exp\left(-\frac{\Delta_i}{\sigma}\right)}{\sum_{i=1}^{n} \exp\left(-\frac{\Delta_i}{\sigma}\right)} \tag{3}$$

where i counts through the samples, σ is a generalization factor, and Δ_i is given by

$$\Delta_i = \sum_{j=1}^{m} w_j (x_j - x_j^i)^2 \,. \tag{4}$$

To describe the function of the GRNN in words, the method finds those samples where the features vector (parameter vector) is close to the input vector. The resulting depth value is the weighted depth of the closest samples. The factors w_j can be used to account for the importance of different features. In the above case only four features were used, all of which are rather important for a correct sizing of the depth. The advantage of this method over the calibration method is that we have a measure for the reliability of the result. With a large value in the denominator of equation (3), we know that we have a very similar training sample and it is wise to assume that the depth is also close. Of course there is no guarantee that the mapping of the features vector onto depth is unique. The result on the sample validation set is shown in **Figure 4**. Another advantage is that we do not have a training process. Still we need to choose some parameters, like σ, wisely, i.e. with regard to the output on a validation set.

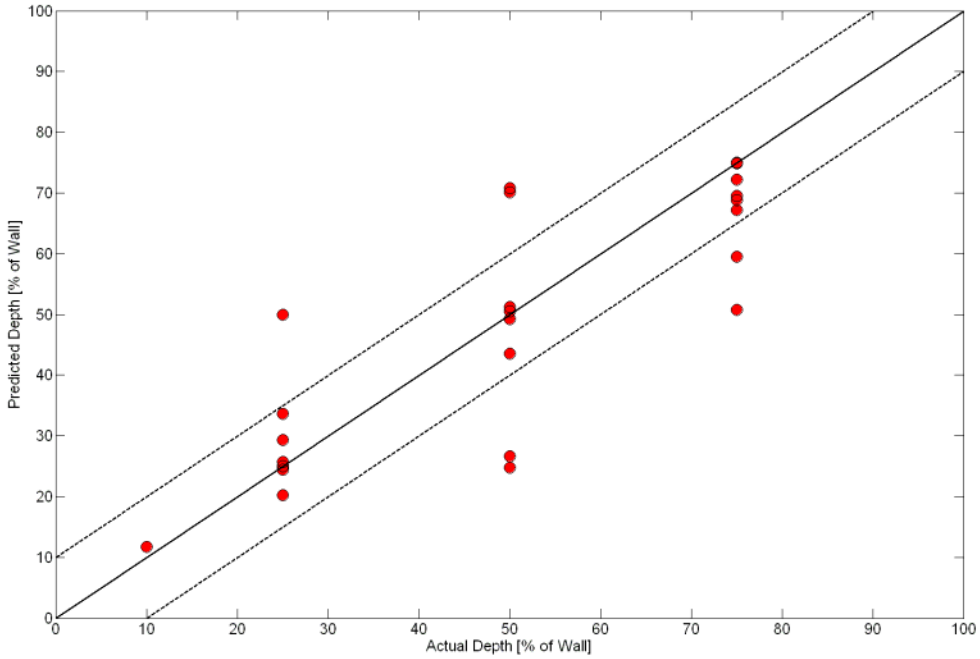

Figure 4: Result of depth prediction with the GRNN method.

One problem is immanent of the GRNN method. It is difficult to assess the method's ability for generalisation. It is seen in **Figure 4** that many samples are matched almost exactly, while some are much more off as compared to the calibration method. If a sample in the validation set is also found in the training data it is not difficult to identify the corresponding one and get the exact depth. However, for other verification data the prediction will have to be based on much different samples. This problem of overfitting is present in any regression method. In this method, however, the minimum error in a validation set usually is a model that is already overfit.

3.4 Multi layer Perceptron

Yet a different method for regression is given by other neural networks, the most popular of which are the multi-layer perceptron (MLP) and the radial basis function network (RBF). It has already been proposed by Upda et al.[3] that this method could be used for the depth estimation of MFL-data. Some pigging vendors use this method [8], but little details have been published. For details about MLP-regression networks refer to the widely available textbooks on neural networks e.g.[9]. In principle, information about the MFL-signal, like the features-vector, is fed into the input nodes of the network. The input nodes are connected to the output via one or several hidden layers. The output in this case would again be the defect depth. The layers are connected with weights, that are adjusted such that the output error on a training set is minimized. In this manner MLP-neural networks can also be used as a universal regression method.

In this demonstration the four already mentioned signal parameters have been used as input values. An MLP has been trained using a four-node hidden layer. With the network trained the validation set is fed through the network. A result is shown in **Figure 5**. It can be seen that apart from some 75% defects the prediction is rather well. Most of the deep defects that are underestimated are not within the stated specification of the minimum size. Removing these defects from the validation set, the stated confidence level of 80% is achieved. In principle it could be envisaged that the output is a three node layer yielding depth, length, and width. In practical experience it turned out that a separate MLP for all of these figures is more accurate.

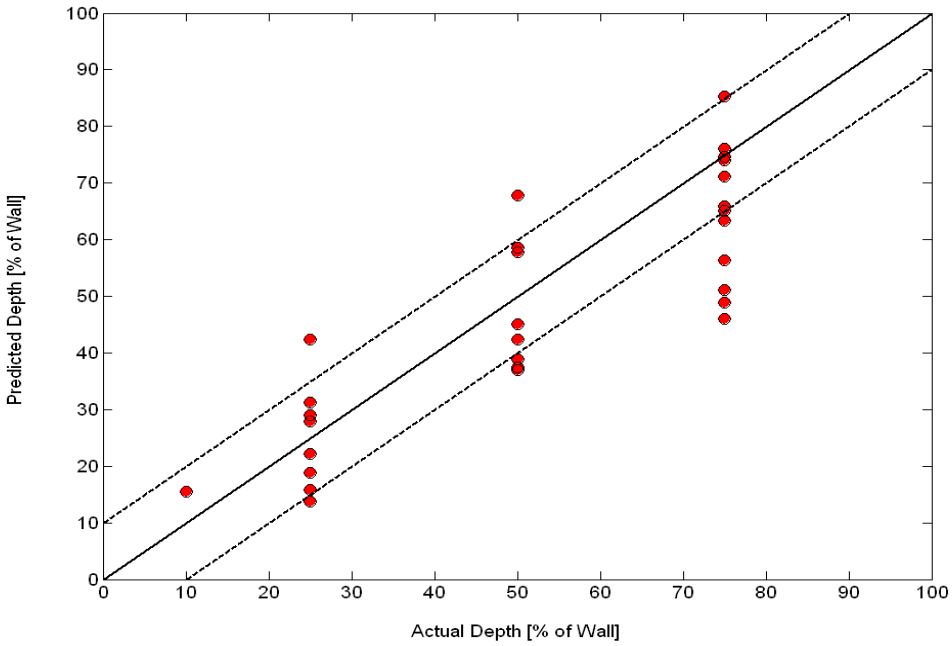

Figure 5: Verification result of a Multi-Layer Perceptron

The ability to generalize for an MLP is determined by the number of training iterations. In textbooks it is usually stated, that training should be stopped, when the error on the validation set starts to increase. As a matter of experience, the curve of the validation error versus the training iterations is not convex, i.e. local minima may occur. It is recommended, that the training process is repeated many times until a "feeling" for the typical development of the validation error is obtained.

3.5 Extrapolation

For all of the above mentioned methods, the error on a given validation set is not the sole consideration for the assessment of performance. The problem of generalization and overfitting has already been discussed. Related to this is the problem of extrapolation. The training vectors form a convex hull in the feature vector space. As long as the depth prediction stays within this convex hull, there is a fair control over what the method predicts. If the vector is outside of this range, the methods differ considerably by how they respond. The effect can easily be demonstrated by predicting depth values in excess of the 75% found in the training set. If a defect with 90% wall loss is encountered, the training data has to be extrapolated. The calibration method will perform reasonably if the number of parameters is limited. The GRNN method will find the 75% defects to be the closest and size accordingly. The MLP method may react unpredictably if some overfitting has taken place. Beyond the training data, the approximated function may become very steep. For all of the methods, the sizing equations should be sufficiently smooth to allow some extrapolations; but since this is difficult to control, a trick can be applied. Some manual training data can be generated by extrapolating the signals manually, making up a features vector, and then adding this to the training set. In this manner some "zero depth" and some through wall defects (100%) can be generated. For instance, it is reasonable to assume an amplitude of zero units for a depth of zero percent. The blue diamonds in **Figure 6** show actual measured amplitudes for defects with 10 mm length and 60 mm width. The red diamonds show data for extrapolated depth values. Also length and width parameters should be extrapolated by plotting them versus actual depth. The background values are of course not affected by the change in depth.

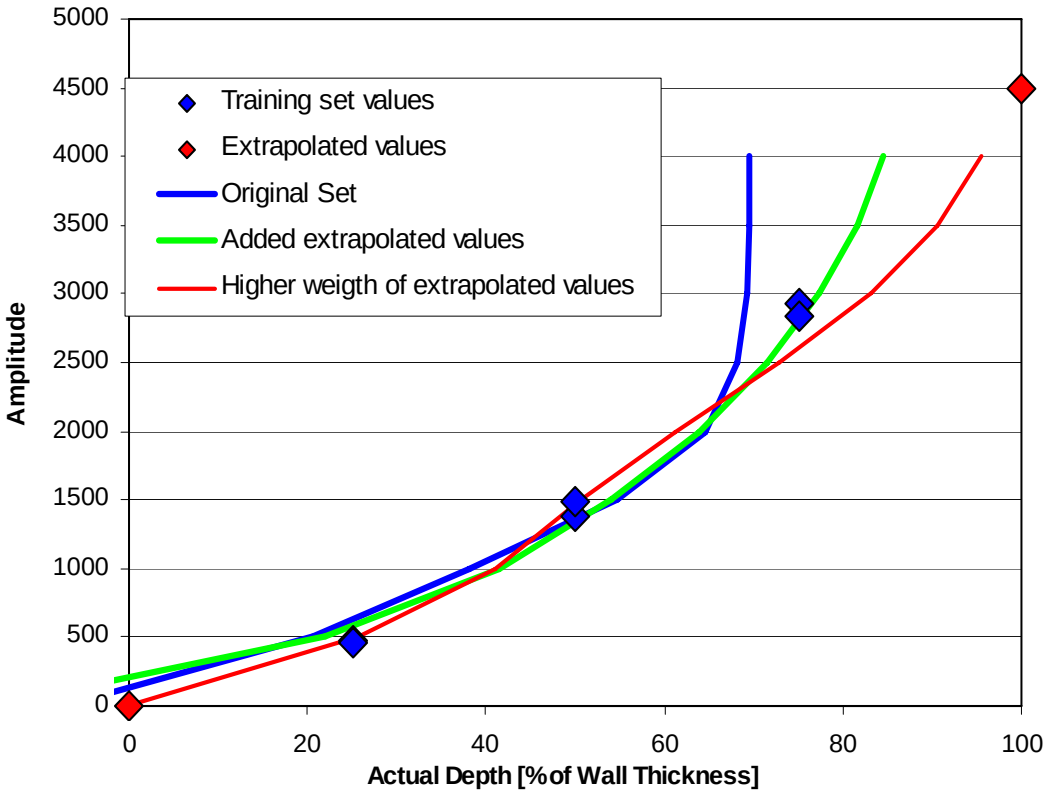

Figure 6: Extrapolation of the input for a 10 mm × 60 mm circumferential defect

The lines in **Figure 6** show the regression curve of an MLP for different training sets. In other words, several features vectors with different amplitude values have been forward fed through the MLP. This amplitude is then plotted versus the output depth. We see how the blue curve fails to correctly size deep defects. For the green curve some extrapolated "fake" defects have been added. An improvement is visible. For the red curve several 100%-defects with the same feature vector have been added. In the optimization process this puts a higher weight on getting the values right especially for deep defects. We see that in the latter case 90% defects are likely to be sized correctly.

4. Other Issues

The above shown tests only show a starting point for the exact depth sizing. To get more accurate results, more influencing factors have to be taken into account. For instance, the inspection tool movement will lead to eddy current generation in the pipe wall. This affects the obtained MFL signals [10]. To account for this, either a speed independent signal is deduced [11], or the speed is taken as an extra input into the sizing method (e.g. extra input node for the MLP method).

Another influence that is difficult to account for are changes in the magnetic material parameter or in the magnetic history. In both cases the magnetization level is different for the same size of a defect. To minimize the effect, a magnetization as high as possible is required. If we assume the saturation magnetization of steel to be rather constant (2.1 T) the influence of different material properties should start to diminish at high magnetization levels. The effect of premagnetization has been observed. If the pipe is inspected for the first time and has never been magnetized, the amplitude levels are up to 20% higher as compared to inspection after many inspection runs. Since this is an effect that is difficult to take into account and is often unknown, a magnetized pipe should be assumed. In case of doubt, a magnetization run could be performed.

One issue with increasing importance is the problem of comparing recent inspection results with earlier inspection results. It has been demonstrated that this is easily done for ultrasonic inspection data [12]. The above shown sizing methods make it obvious that small changes in the set-up of the sizing parameter will influence the depth result in a systematic manner. Any retraining of the MLP would lead to slightly different values of the depth. If the training is done properly and overfitting is avoided, the change will be small and negligible for the single defect; however, it will affect all defects in the same manner. If two inspection runs use different sizing parameter, we will see this systematic change in depth. It will easily be mistaken for a corrosion growth in the case of depth increase. The fact that the sizing model has changed between inspection runs is almost inevitable. These methods are constantly reviewed and improved. For the GRNN, for instance, any new verification result could be added and the reliability is improved.

As a solution for the run comparison, it has been proposed that the MFL signals strength is normalized. Then old and new MFL signals are fed through the same sizing method and the results are compared for possible corrosion growth [13]. Of course the normalization procedure can hardly take into account the non-linearity of the magnetization problem. it can only linearly scale the magnetic fields levels with respect to magnetic responses of items that are clearly unchanged (like girth welds).

5. Conclusion

It has been demonstrated that many factors influence MFL defect signals and that a defect size reconstruction is not straight forward. The sizing method applied in practice today often takes into account the complexity of the problem by a multidimensional non-linear regression model. This does not require a priori knowledge about the physical nature, but rather relies on sample measurement or FEM results. A reliable defect sizing is achieved. Problems like defects of unexpected size, speed effects or run comparisons need special solutions.

References

[1] Calculation of the magnetostatic field of surface defects. I. Field topography of the defect models, N.N. Zatsepin, V.E. Shcherbinin, Defektoskopiya, **5**, pp.50-59, (1966)

[2] SQUID System for Magnetic Inspection of Prestressed Tendons on Concrete Bridges, H.-J. Krause, . Wolf, W. Glaas, E. Zimmermann, M.I. Faley, G. Sawade, G. Neudert , U. Gampe, J. Krieger, 15th WCNDT,Rome, 2000

[3] D.E. Farrell, J.H. Tripp, P.E. Zanzucchi, J.W. Harris, G.M. Brittenahm, W.A. Muir, Magnetic Measurement of Human Iron Stores, IEEE, Trans. Mag. **16(5),** (1980), pp.818-823

[4] Belanger A. "Managing HIC-Affected Pipelines Using Multiple-Technology Hard-Spot Tools." Conference Proceedings Pipeline Pigging, Integrity assessment & Repair 2004.

[5] Magnetisation as a key parameter of magnetic flux leakage pigs for pipeline inspection, H.J.M. Jansen, P.B.J. van de Camp, M. Geerdink, Insight **36(9)**, (1994), pp.672-677

[6] Magnetic and other physical properties of X52, X60, X70 and X80 grade line pipe steels, F. Richter, steel research 60, (1989), No.**9**, pp.417-424

[7] The application of artificial intelligence to the analysis of magnetic-flux leakage data, S. Udpa, P. Porter, The Pipeline Pigging Conference, Houston, February 13th-16th 1995

[8] Neural Systems for Control, C. Schäffner, D. Schröder, p.235, Acad.Press 1997

[9] Tri-axial Sensors and 3-dimensional magnetic modeling of defects combine to improve defect sizing from magnetic flux leakage signals, S. Cholowsky, S. Westwood, NACE 2004

[10] Pattern Recognition with neural networks in C++, A.S. Pandya, R.B. Macy, CRC Press, Boca Raton, 1995

[11] Numerical simulations on electromagnetic NDT at high speed, Y. Li, G.Y. Tian, S. Ward, Insight, vol.**48(2)**, (2006), pp.103-108

[12] Invariance Transformations for Magnetic Flux Leakage Signals, S. Mandayam, L. Udpa, S.S. Udpa, W. Lord, IEEE Trans. Mag., Vol. **32(3)**, (1996), pp.1577-1580
[13] Advantages of high resolution ultrasonic in-line inspection tools regarding run comparisons and integrity assessment of pipelines, paper # 05162, K. Reber, M. Beller, N. Uzelac and O.A. Barbian, NACE 2004, Houston
[14] Dawson,S.J.;Race, J.M.;Krishnamurthy,R.;and Peet, S. Pipeline Corrosion Management, Corrosion 2001, NACE, Houston paper#1627