# Analyses of risk estimates: how to begin

by W. Kent Muhlbauer, WKM Consulting, Austin, Texas, US

As risk assessment methods have progressed, the amount of data available to operators has increased dramatically. This article explores how this information can be best used without becoming overwhelming, with a focus on the current top four analysis techniques.

Modern risk assessment uses and generates huge amounts of data. With today's inspection opportunities, coupled with advanced risk assessment, there are often dozens – if not hundreds – of pieces of information assigned to every millimetre of pipeline.

That data is used to generate risk estimates. Then, even when those risk estimates are summarised, there are often still hundreds of values per kilometre of pipeline. So there is a lot of data going into and coming out of the risk assessment, reflecting the numerous real world risk considerations that accompany every pipeline.

When confronted by large amounts of data, a type of analysis paralysis can set in. This 'inaction reflex' can take either of two forms: we freeze, because we don't know where to start, or we get so caught up in the analyses that we never act. Let's tackle the first here.

First, let's recognise that having lots of data is a good thing. Far from viewing the hundreds of thousands of bits of information accompanying modern risk assessment as leading to burdensome work, it should be viewed as the goldmine that it is. The information brings countless opportunities for increased understanding and more efficient management of risk.

It requires only the application of a few common analyses tools to begin reaping the benefits. Complex, detailed statistical analyses could be useful, but are certainly not required.

Simple to understand and apply techniques will quickly and painlessly yield knowledge from the information. With this knowledge, we are on a path to efficient risk management.

Here are the current top four analysis techniques, presented in the typical order of application.

## HISTOGRAM ANALYSIS

Histogram analysis is a good first choice in understanding any large data set. Beginning with highest level risk estimates – expected loss (EL), frequency of failure (FoF), consequence of failure (CoF) – and moving through to next tier estimates – FoF from third party damages, FoF from external corrosion, etc. – we gain increasing insights into the characteristics of the systems represented by the data.

Unlike some other visual support tools – like matrices and bowties – histograms are truly accurate and efficient representations of risk, communicating information far beyond what the underlying numbers carry. I will address this topic in a future *Pipelines International* column, titled 'Seduced by graphics: the myth of managing risk by images'.

Histograms provide knowledge of the behaviour of data sets. As an indicator of the underlying distribution of entire populations, the histogram generated from a subset of the population can be used to generate an equation to predict future values.

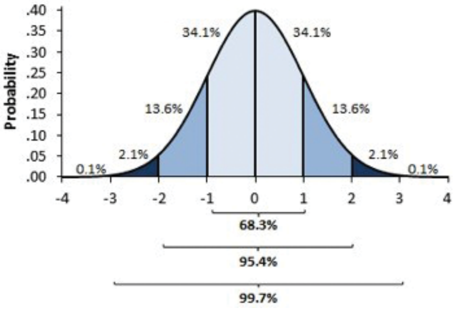


**FIGURE 1:** A histogram showing a bell distribution.

At the least, the histogram provides understanding and context for measures of central tendency (mean, median and mode) and dispersion (range, standard deviation, etc.). For example, if the histogram suggests a normal distribution (a bell curve, as seen in Figure 1), then much can be immediately inferred, such as the average = median = mode and that 68 per cent of all data will fall ±1 standard deviation from the mean.

Sometimes data – including risk estimates – groups in distinct subpopulations, as seen in Figure 2. Characteristics such as population

W. Kent Muhlbauer is a regular columnist for *Pipelines International*, author of *Pipeline Risk Assessment: the Definitive Approach and its Role in Risk Management*, and presenter of the Advanced Pipeline Risk Management course run internationally with Clarion Technical Conferences and Tiratsoo Technical.

For more information, visit **www.pipelinerisk.net** density, pipe wall thickness, depth of cover, operating pressures and more could be the underlying inputs that generate these subgroups. Overlaying these characteristics with histograms of risk estimates should confirm suspected relationships between such factors and risk levels, or prompt additional investigation when such suspicions are not confirmed.

Much information is conveyed at a glance when histograms are compared. Consider the five pipelines of varying lengths and rates of risk shown in Figure 3. A narrative or table of values would have difficulty in displaying as much information as efficiently.

## GROUPINGS

As a good next step, a tabulation of 'buckets' of key risk variables lets us better understand both the combinations of risk discriminators and their impacts. Seeing these characteristics of the pipelines – i.e. what proportions of the system(s) falls into which combinations – sets the stage for better understanding the risk assessment results.

Which combinations lead to higher risk? Is this consistent with the subject matter expert's knowledge of causation? Are the differences consistent with the current understanding of the underlying science and engineering?

As an example of a grouping analyses, consider a natural gas pipeline with a single diameter and maximum allowable operating pressure, but a variety of pipe installation dates, wall thicknesses, and integrity assessment types and dates.

Each of these should logically have an impact on risk and we should be interested in how much variability each characteristic contributes and how risk estimates change with each changing characteristic. For instance, from a simple database query, we may learn that a pipeline section has four different wall thicknesses, two different years of installation, two types of inline inspections (ILIs) used in three different years, resulting in 48 combinations. Each combination has a different risk implication and the risk for each combination can be quantified.

At a glance, an understanding can be gained of the system characteristics that may have taken the long-term operator's employee years to learn. Additionally, important aspects of risk that are associated with each combination can now be seen.

A simpler example given in Table 1 shows combinations of installation dates, nominal wall thicknesses and integrity assessment ages that help determine the components' wall thicknesses today – termed 'effective available wall thickness'. In general, more recent and more accurate confirmation of 'no damage' leads to higher effective available wall thickness.

We quickly see how much of this pipeline falls into each combination and what effective wall thickness is associated with that combination, to be used for subsequent risk estimations.

## CORRELATIONS

Graphs – as illustrated in Figure 4 – or calculations of how certain risk estimates change in reaction to others is a more advanced analyses opportunity. Correlations or lack of correlations should be explainable – if they are not, then error checking is in order.

Some correlations can be simple: how does

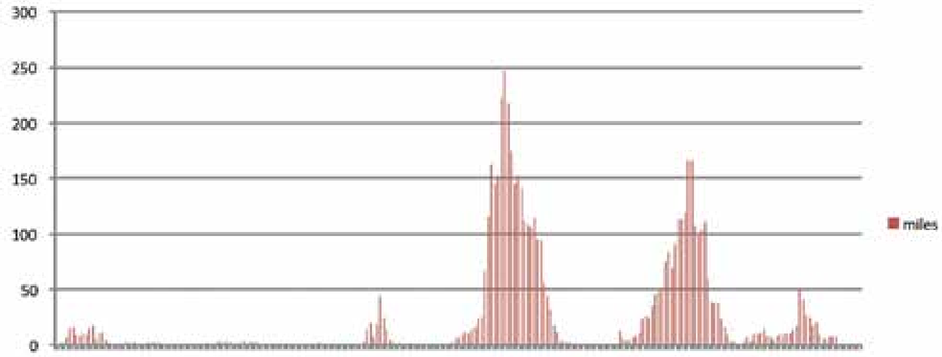


**FIGURE 2:** A histogram showing data forming two distinct subgroups.

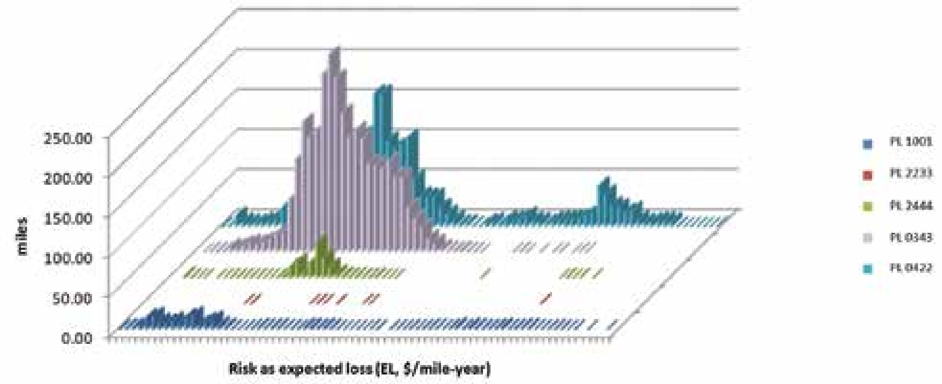


**FIGURE 3:** A histogram comparing five pipelines and their rates of risk.

| INSTALL DATE | SEG COUNT | LENGTH | NOM WALL | INSP AGE NDE | INSP AGE ILI ML | TEST AGE | EFF AVAILABLE WALL THICKNESS |
|---|---|---|---|---|---|---|---|
| Year | Count | Miles | Inches | Years | Years | Years | Inches |
| 1940 | 143,866 | 27.25 | 0.199 | 78 | 4 | 1 | 0.157 |
| 1940 | 116 | 0.02 | 0.199 | 78 | 4 | 18 | 0.122 |
| 1940 | 9 | 0.002 | 0.219 | 78 | 4 | 1 | 0.208 |
| 2016 | 45,377 | 8.59 | 0.219 | 2 | NA | 2 | 0.168 |
| 2016 | 4,120 | 0.78 | 0.25 | 2 | NA | 2 | 0.194 |
| 2016 | 2,001 | 0.38 | 0.322 | 2 | NA | 2 | 0.257 |

**TABLE 1:** A tabulation of key risk variables.

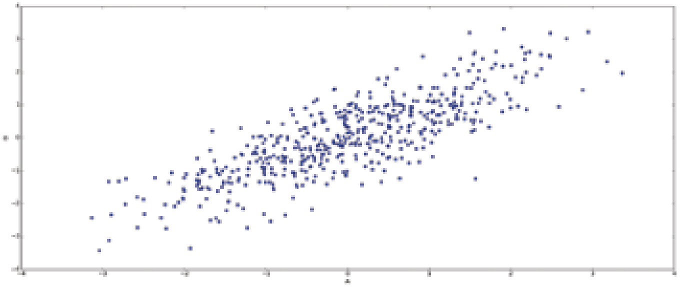

FIGURE 4: A correlation plot.

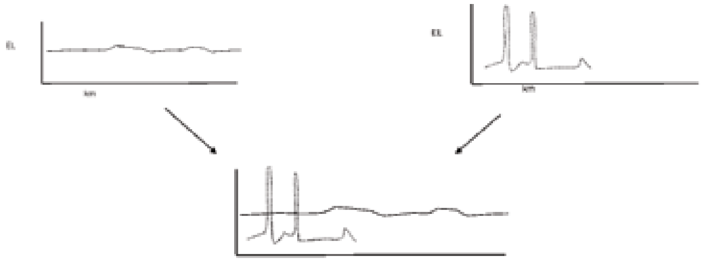

FIGURE 5: Profiles and comparison of two pipelines with the same overall total risk level.

outside force damage potential react to changes in depth of cover? Others can be more complex: how does effective wall thickness react to ILI date and technology type?

## PROFILES

Plotting changes in any risk estimate along a pipeline route is the first step in risk management. We must identify risk peaks/valleys and distinguish systemic risk issues from localised issues, rate-of-risk versus total risk and more, before we can effectively understand, much less manage, the risk.

Let's illustrate this with an example. Say we have two different pipeline segments' risk assessment results. These pipelines carry different types of hydrocarbons, are different lengths, operate at different pressures and are located far from each other.

However, by pure coincidence, these two segments have exactly the same total risk level (EL in $/year). By that single measure, the pipelines appear equivalent and might be candidates for identical risk management strategies. However, that notion is quickly dispelled once we create profiles of each. We plot each as EL in $/km-yr versus length, which can then be compared as show in Figure 5.

This immediately shows that, despite the numerical equivalent of total EL, these two pipelines have dramatically different localised risk levels and variability in risk levels. They clearly would not be effectively managed by the same risk reduction measures, regardless of the underlying drivers of their respective risks.

Until there is a basic understanding of the patterns within the section of interest, we cannot even begin to diagnose risk issues and propose risk reduction measures. P