

Pipeline Risk Modeling – How Much Data do I Need?

W. Bryce P.Eng., Dr. K. Oliphant, P.Eng., JANA Corporation

Abstract

There is more and more discussion in distribution and transmission pipeline risk modeling of quantitative risk modeling approaches. The most common concern expressed by those examining quantitative modelling approaches is having enough data to feed these models, with a common assumption that the data needs are well beyond the currently available asset data. This often results in adopting softer scoring or index modeling approaches as data collection projects are implemented. This paper demonstrates that, while data is certainly a key component of driving to accuracy and specificity in any modelling approach, the focus on data often leads to the adoption of risk modelling approaches that ignore fundamental risk modelling requirements, leading to completely fallacious model outputs.

The Move to Quantitative Risk Modeling and the Question of Data

As DIMP and TIMP programs evolve, there is growing discussion in the natural gas industry of moving to more robust quantitative modeling approaches. This naturally leads to discussions on the data requirements of these modeling approaches with a common assumption that current data is not sufficient to fuel the data needs of these approaches. This sometimes also leads to the interim adoption of scoring or index modeling approaches while data collection initiatives are undertaken on the presumption that an Index type scoring model will accommodate the presumed lack of data. While data is critical to any risk modeling approach, the more fundamental concern is adopting the correct risk modelling approach. As shown in the discussion that follows, with the correct risk modeling approach effective risk projections can be made even in the face of significant data gaps or data quality issues.

The Importance of the Right Structure

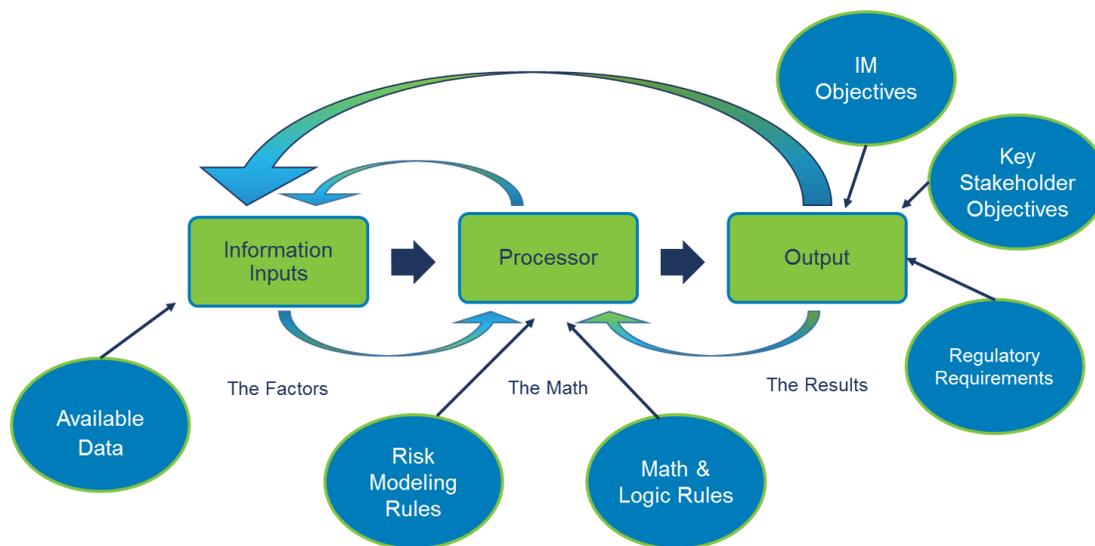
The critical component of having a useful risk modelling approach is having the right structure for the modeling approach. Regardless of the available data, if the right structure for the risk modeling approach is not used the risk model outputs will be useless. We are all familiar with GIGO – garbage in means garbage out. The issue is more fundamental than this though. Even if perfect data is available, if the data are not processed correctly there will still be garbage out.

So how can it be ensured that the right risk modeling structure is employed – ensure that the data that is available are processed correctly? The first step is to consider the requirements that are imposed on the risk modeling approach by external factors. As shown in **Figure 1**, every model is composed of essentially three components: 1. the model inputs (the factors or data), 2. the processor (where the data is manipulated), and 3. the model outputs.

For each of these three components there are requirements imposed on the risk modeling approach by the objectives of the integrity management program and there are basic risk modeling rules and math and logic rules (imposed by the rules of mathematics). These requirements apply regardless of the type of risk modeling approach (i.e. Quantitative or Index). Unfortunately they are often overlooked, or rationalized away, in favor of adopting a modeling approach that can handle the presumed lack of data.

For example, Key Stakeholder Objectives impose requirements on the risk model outputs (e.g. if there are corporate objectives around health and safety, health and safety measures need to be part of the model output). The objectives of the integrity management program also impose constraints on the model outputs (e.g. financial optimization of integrity management programs requires risk model output to be in financial terms (or risk units) and for the risk benefit of mitigations to be captured in the risk models in financial terms (or risk units)). Regulatory requirements (e.g. address all threats) and evolving regulator expectations (e.g. the move to quantitative approaches to risk) also impose requirements on the structure of the risk modeling approach.

Figure 1: Requirements of a Risk Modeling



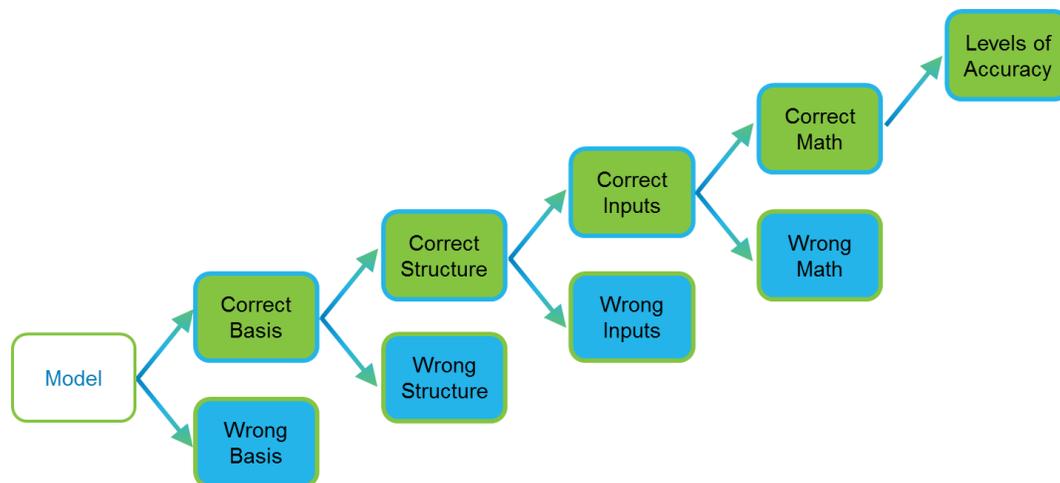
There are similar requirements on the 'Processor' (or how we do the math) that are imposed by the 'rules' of risk modeling and math and logic rules. As discussed in more detail in the section **Why Many Risk Models Fail**, a simple grade four math rule is violated by many risk modeling structures (particularly Index type approaches) that results in completely false risk model outputs.

The available data is, of course, also a critical component in terms of how the risk modelling approach is structured as it dictates the inputs we have for modeling. It is, however, only one consideration out of many and can be addressed with careful modeling structure. In fact, the very absence of data is a key justification to move to quantitative risk modeling – not the reverse.

As a simple historical example of how we can make reasonable projections with quantitative models, even in the face of imperfect data, when we get the structure and approach right, we look back 2500

years to the Greek philosopher Eratosthenes. By knowing the angles of shadows cast in two cities during the summer solstice and by performing the right calculations using his knowledge of geometry and the distance between the two cities, Eratosthenes was able to develop a model to estimate the circumference of the earth to within 16% of the actual value (well in advance of the advent of calculators and Google Maps). **Figure 2** demonstrates how he was able to do this and how we can apply similar logic to our risk modelling approaches for natural gas pipelines.

Figure 2: The Path to Functional Risk Models



First Eratosthenes based his model on the correct basis – he assumed the Earth was a sphere (at the time many thought it was flat and there were even a few proponents proclaiming it was a rectangle). In structuring an effective risk modelling approach we can achieve the same by understanding the mechanisms driving potential failure and the possible consequences of failure.

Second, based on his knowledge of geometry he used the correct structure for his model, knowing that based on the angles of the shadows the distance between the two cities represented 1/50 of a circle. Again, a fundamental understanding of the mechanisms underlying pipeline failure and the external requirements imposed on the modeling approach enable us to get the right structure.

He then got reasonable (note: not perfect) input data, reportedly paying someone to pace out the distance between the two cities.

Finally, he did his math correctly. As a result, he was able to model the circumference of the earth to within 16%. Without the right basis, the correct structure, reasonable data inputs and doing the math right, we most likely never would have heard of Eratosthenes. Because he did get those correct we are talking about his level of accuracy. The same applies in our approach to risk modelling. We need to first ensure we have the fundamentals right and then worry about how improving the accuracy of our data will drive to greater modelling accuracy. Going back to Eratosthenes, in 2014, his exact same approach was used with more accurate estimates for the distance between the two cities to yield model estimates that were within 0.15% of the Earth’s actual circumference.

Why Many Risk Models Fail

Unfortunately, many risk modelling approaches do not get these fundamentals right. While there are many breakdowns, the discussion that follows focuses on one particularly common error, observed especially in many Index or Scoring type risk modelling approaches and ‘Quantitative’ Models that take indices and slap a veneer of the quantitative over them (though also observed in some quantitative models).

Figure 3 provides the standard definition of risk as the product of probability of failure and the consequences of that failure. The key in terms of this discussion is the multiplication sign.

Figure 3: The Definition of Risk



Figure 4 provides a high level summary of the number of PHMSA reported incident per 10,000 gas release for gas distribution systems by threat type. If the data is normalized to corrosion risk it is seen that there is a range of up to 85 times difference in the potential of a leak becoming a PHMSA reportable incident among the threat categories with a corresponding difference in potential consequences.

Figure 4: PHMSA Reported Incidents per 10,000 Gas Releases

	Incidents per 10,000 Releases	Relative to Corrosion
Material / Component Failure Risk	0.9	4.5
Corrosion Risk	0.2	1
Natural Forces Risk	1.8	9
Excavation Damage Risk	1.9	9.5
Other Outside Force Damage Risk	17	85
Incorrect Operations Risk	4.9	24.5

The table is part of a larger diagram where 'Overall Risk Asset' is connected to a plus sign (+) which then branches into the six risk categories listed in the table.

This highlights that all leaks do not have the same potential consequences. Returning to the definition of risk in **Figure 3** as the probability of failure times the consequences of failure and going

back to grade four math, where we know that we need to multiply before we add, means that we need to address the probabilities and consequences for each threat, in fact each sub-threat, we consider in our risk models independently. This fundamental rule of math, however, is violated by many risk model structures, particularly most Index or Scoring type approaches. The result, as demonstrated below, is that risk modelling approaches that do not take this simple rule into account produce incorrect, in fact completely wrong, outputs on both an absolute and relative basis.

Table 1 provides a simple comparison of the risk for two gas distribution mains. Whether we use probability of failure estimates or risk ranks, if we sum the probabilities of failure and consequences for different threats and then multiple those sums, as has been observed in many Index type models for gas pipelines examined, we would predict that Main 2 is at higher risk than Main 1.

Table 1: Incorrect Risk Comparison of Two Main Segments

Threat	PoF Rank Main 1	CoF Rank Main 1	Risk Rank Main 1	PoF Rank Main 2	CoF Rank Main 2	Risk Rank Main 2
Corrosion	0.1	1.7		0.5	1.7	
Excavation Damage	0.5	6.0		0.2	6.0	
Natural Forces	0.4	4.0		0.3	4.0	
Material/Welds	0.1	1.2		0.6	1.2	
Total Sum	1.1	12.9	14.19	1.6	12.9	20.64

Table 2 looks at the same data when we properly multiple the probability of failure by the consequences for each threat separately, as dictated by the math rule that we must multiply before we add. When the math is done correctly Main 1 is actually seen to be higher risk than Main 2. While it is a simple example, it highlights that risk models that are not properly structured simply produce wrong results, not 16% off like Erasthenes but wrong. Though it may seem very obvious, this is a common error that has been observed in many pipeline risk modeling approaches.

Table 2: Correct Risk Comparison of Two Main Segments

Threat	PoF Rank Main 1	CoF Rank Main 1	Risk Rank Main 1	PoF Rank Main 2	CoF Rank Main 2	Risk Rank Main 2
Corrosion	0.1	1.7	0.05	0.5	1.7	0.85
Excavation Damage	0.5	6.0	3	0.2	6.0	1.2
Natural Forces	0.4	4.0	1.6	0.3	4.0	1.2
Material/Welds	0.1	1.2	0.06	0.6	1.2	0.72
Total Sum	1.1	12.9	14.19	1.6	12.9	20.64
Average	0.275	3.225	0.89	1.075	3.225	3.47
True Risk	-	-	4.71	-	-	3.97

The Data is not the Key Issue

While data certainly is a key component of any risk modelling approach, data concerns have unfortunately led to the adoption of risk modeling approaches, such as Index or Scoring methods, that are in many cases adopted without consideration of the fundamental requirements of the risk modeling approach. The primary focus in selecting a risk modelling approach needs to be on getting these fundamentals correct in the context of the available asset data. With solid fundamentals, the risk models can then be evolved in concert with data collection initiatives to drive to greater model accuracy and specificity.