How Bayesian Networks can contribute to Pipeline Integrity Management

Helio Alves, Luiz Oliveira
DNV GL Oil & Gas, Brazil

François Ayello, Gerry Koch, Andrea N. Sánchez, Narasi Sridhar
DNV GL Strategic Research & Innovation, U.S.A.

1. INTRODUCTION

Risk based approach has been recommended as a preferred alternative for Integrity Management of Pipelines. For both onshore and subsea (or submarine) pipeline Integrity Management standards, managing the risk related to the pipeline system threats is essential for maintaining the integrity of the pipeline system. It is been recognized in the industry that proper management of the threats taking into account the probabilities of failure and the consequences of those events can result in an efficient tool to avoid early accidents and reduced time life, in addition to reducing risk associated costs, which can be very high for offshore assets.

Risk based Pipeline Integrity Management takes into account:

- Identification of threats and failure modes
- Estimation of probabilities of failure (PoF)
- Estimation of consequences of failure (CoF)
- Estimation of Risk level (PoF x CoF)

The condition of an existing asset consist of its complete history, starting from design, fabrication, commissioning, past operation conditions until the present time. One of the main questions is “how long the asset can be safely operated?”, the answer to it depends on estimating the future conditions and behavior of the individual sections of the asset. Numerical models have been developed to estimate corrosion rates, stresses in corroded components subject to internal pressure, loads in pipelines under free spans, etc. However, most of those models are either prescriptive or follow empirical and semi-empirical approaches that may be valid only under specific conditions resulting in large uncertainties when extrapolating results in time.. Field inspection methods have also uncertainties and human error involved, both issues recognized by service companies and operators. Data from other similar assets can be used, but with limitations construction-related practices, environmental and operating conditions may be significantly different. Hence, risk estimations represented by a single number do not convey the large uncertainties that are involved in those calculations. Managing uncertainty is a key element in the assessment of the integrity of industrial assets.

Bayesian networks have demonstrated to be a feasible approach to assess corrosion threats in a pipeline (as referred in section 3 below). Bayesian network is a mathematical approach based on Bayes theorem that represents that cause-consequence relationship that exists between the variables of a complex system in the form of conditional probabilities. The Bayesian network approach integrates various sources of knowledge (models, field data and subject matter expert opinion) into one framework. The probabilistic nature of Bayesian networks allows the inclusion of the uncertainties associated from each variable.

2. INTEGRITY MANAGEMENT OF PIPELINES

Bayesian Networks can be used for the assessment of probability of events that can lead to pipeline failure and how this information can be used in an Integrity Management perspective. Pipelines are subject to many different threats, as described in Pipeline Integrity management documents and standards [1-3]. It is usual to consider that pipeline integrity is associated with structural/containment functions, as stated in DNV RP-F116 (2015). Hence, the threats related to those functions can be assessed in terms of the influence in the pipeline integrity condition based on the significance of the Probability of Failure (PoF) associated to the mentioned functions, as a component of the total Risk. In a quantitative approach for Risk Assessment, the methodology for assessing the PoF can be more or less representative of the real situation, depending on the approach used and the reliability of the information used.
Integrity Management starts during early stages of design, and shall follow the project until abandonment (inclusive). After the takeover of the project by operations, Integrity must be maintained during operational phase. Risk Based Integrity Management has been intensively used, and is a well-accepted method in many industries. Hence, managing integrity is normally closely related to managing risk. Risk can be defined by its components Probability of Failure and Consequence of Failure. In this paper, a probabilistic approach for defining the PoF component of Risk is presented, where Bayesian Network approach is presented. This approach has the characteristic of permitting the use of any existing information as a valid input to PoF, and take into account the uncertainties and beliefs associated with the existing knowledge, model or data available.

3. THE BAYESIAN NETWORK MODELLING APPROACH

3.1 The Bayes Theorem

The relationship between frequency and probability is defined by the well-known Bernoulli's limit theorem:

$$p(|f(x) - p(x)| > \mathcal{E}) \rightarrow 0 \text{ as } N \rightarrow \infty \qquad (1)$$

In this equation, the frequency, $f(x)$, of a population of data approaches the probability, $p(x)$, of that same data as the number of trials approaches infinity. Stated another way, frequency is a measurable quantity based on repeated observations, whereas probability represents the degree of belief or confidence in the measured frequency, also referred to as probability of frequency by Kaplan and Garrick [4]. Some authors refer to probability as simply a degree of belief in an event. This view stems from the idea that not all phenomena can be repeated in a controlled manner to derive statistical distributions. This is especially true of complex systems. Therefore, probability can be assigned to the strength of an expert's belief about an event and can then later be corrected using repeated observations. This is at the heart of Bayes theorem and it is often referred to as belief network. These two perspectives can be combined – where possible statistical distributions are derived through the use of mechanistic models that are in turn based on experimental data with their associated uncertainties, but we can also include direct probability distributions representing the degree of belief of an expert in a given observation. These two streams of probabilities are linked in a Bayesian network that can be updated through laboratory or field observations

The Bayes Theorem states that posterior probability of an event (i.e. probability of the event after an observation is made) is related to the prior probability of the event (i.e. before the observation is made) through the probability of observing the event and the conditional probability of observation given the event occurred, as given by Equation 2:

$$\boldsymbol{P}(\boldsymbol{Ai}|\mathrm{B}) = \frac{\boldsymbol{P}(\boldsymbol{B}|\boldsymbol{Ai}) * \boldsymbol{P}(\boldsymbol{Ai})}{\sum_{j=1}^{n} \boldsymbol{P}(\boldsymbol{B}|\boldsymbol{Aj}) * \boldsymbol{P}(\boldsymbol{Aj})} \qquad (2)$$

where $P(A_i \mid B)$ is the posterior probability of an Event Ai given the observation B, $P(A_i)$ is the prior probability of the event A_i before the observation B, and $P(B \mid A_i)$ is called the likelihood function and is the probability of the observation B given that event A_i occurred. The denominator in Equation 2 is called the probability of observation and is the sum of all the conditional probabilities of B given events, A_j multiplied by the probabilities of A_j. The prior and posterior probabilities can also be considered as "cause" and "consequence" of a process. The term "consequence" in this sense should not be confused with consequence of a risk model. It is strictly a relation between two events in a process, one leading to the other. An example would be the presence of water in a pipeline leading to corrosion.

Bayesian networks use Bayesian inference (Equation 2) on a larger scale. Bayesian networks allow calculating the probability of not one but numerous interconnected parameters. Bayesian networks are often represented graphically where many random variables are connected by cause-consequence dependencies. This approach is particularly useful to perform risk assessment of corrodible systems, because of the ability to consider a great number of events that can lead to failure. Since, Bayes' Theorem shows what is and isn't evidence and also describes the strength of the evidence, a succession of Bayesian inferences will provide the most probable scenarios leading to failure, quantify the certainty of the scenarios, and provide a mathematical framework to reduce uncertainties through observation (*i.e.* data gathering). The main limitation of Bayesian networks is that in order to solve the succession of Bayesian inferences numerically only directed acyclic graph can be used. In other words, the strength of the cause-consequence relationships must be preserved and no feed-back loops are allowed between causes and consequences to alter the strength of these relationships. However, knowledge of a consequence can be used to update the probability of a cause without altering the strength of their relationship. The heart of the Bayesian network model is the derivation of the conditional probability tables or $P(B \mid A_j)$ from fundamental models and subject matter experts.

Pipelines are exposed to a number of risk factors or threats, such as intrusion by third parties, fabrication defects, cyclic loading, and external and internal corrosion. The intrusion and fabrication-related factors may be considered to be essentially time-invariant factors, although some parameters that affect intrusion probability (e.g. population density) change over time. The probabilities of these failure modes may be estimated from the probabilities (i.e. frequencies) of related factors, such as surrounding population centers, vehicle traffic, type of welding used to manufacture the pipeline, etc. The probabilities of time-dependent factors, corrosion, fatigue and stress corrosion cracking, have to be assessed using a number of models. The next section shows an example applied to internal corrosion.

3.2 The Internal Corrosion Example

The solution selected to manage and present the various events leading to pipeline failure in a graphical form is Bayesian network. Bayesian networks are probabilistic graphical models that encode probabilistic causal relationships between variables of interest. Therefore, every event shown in a Bayesian network is linked by cause-consequence relationships, additionally every relationship is quantified. Bayesian networks offer many advantages over other types of graphical models, which will be described later in the paper.

Figure 3 shows a high level example of a Bayesian network for internal corrosion damage assessment in pipelines [5]. In this model, events that lead directly or indirectly to corrosion damage and failure are linked using cause-consequence relationships. The model allows a relatively easy determination of the chain of events that could lead to pipeline failure. Although the network shown in Figure 3 is quite simple, it can become rather complex, hence the implementation of the model in a numerical application is recommended. In a real application, the main nodes could represent, for instance, uniform corrosion rate, localized corrosion rate, erosion rate, microbiologically influenced corrosion rate, all contributing to develop and grow a flaw size (depth and length) and resulting pipeline remaining strength or probability of failure.

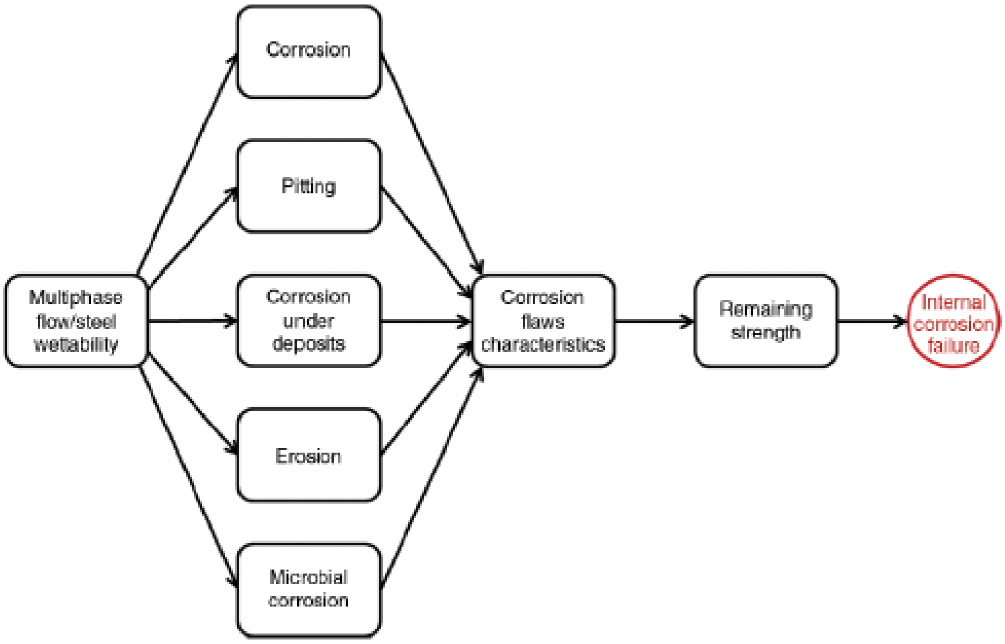

Figure 3. Sketch view of a basic network created for pipelines internal corrosion damage assessment.

Data uncertainty is of major concern in every type of risk assessment. A model that can deal with data uncertainty is crucial for a good a reliable pipeline corrosion risk assessment. Data uncertainty is usually dealt with by using Monte Carlo simulations, however in order to effectively utilize the Monte Carlo method, it is necessary to have all models (e.g. flow, uniform corrosion, localized corrosion, erosion, microbiologically influenced corrosion, etc.) run together in one single framework. Rarely all parameters that have an effect on corrosion are found in the same framework, which makes it difficult to carry the effect of data uncertainty from one model to the next. Bayesian networks do not have this problem. To illustrate this, a zoom on a real network for Internal Corrosion [6] is shown in Figure 4. This figure shows the part of the corrosion model that calculates the uninhibited corrosion rate (corrosion inhibition, flow and steel wettability are not shown here for simplicity reasons). First, the figure shows that each event is not represented by a number but a probability density function that describes the relative likelihood for each variable to take on any given value, and no state, no matter how unlikely, is overlooked. A red bar shows a node that has a certain state (e.g. in Figure 4, H_2S concentration is known to be zero), blue bars show nodes with uncertain information (e.g. the temperature is known to be below 60°C, but the exact value is unknown) and green bars show unknown variables (Fe^{2+} and O_2 concentrations are unknown, hence they are represented by flat distributions). Consequently, because of the uncertainty in the data, the uniform uninhibited corrosion rate is known with uncertainty.

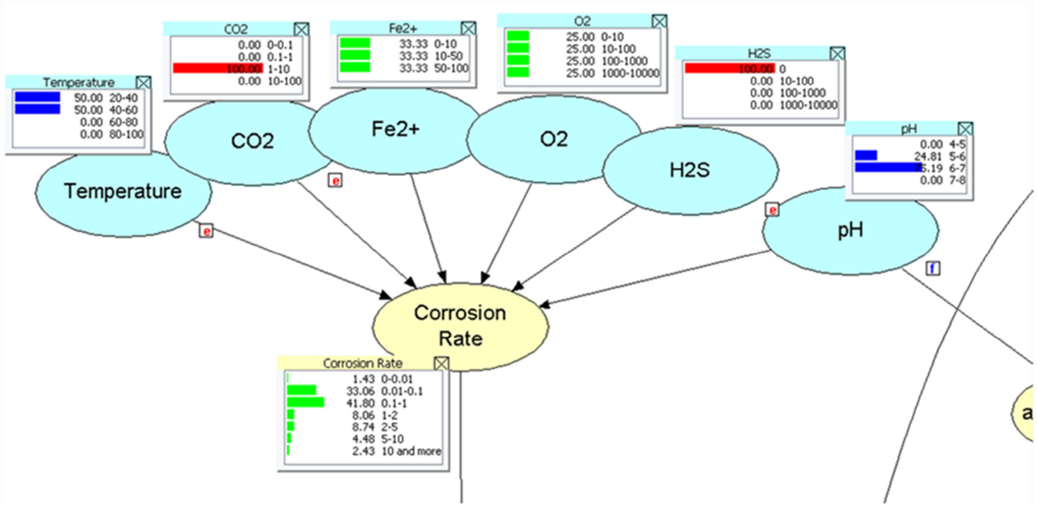

Figure 4. Detailed view of partial network showing influence of parameters (Temperature in Celsius, CO_2 partial pressure in Bar, Fe^{2+} concentration in ppm, O_2 in ppb, H_2S in ppm) in the corrosion rate assessment (in mm/year).

Some examples of the Bayesian calculation for Internal Corrosion of pipelines demonstrated in Figure 4 are shown in Tables 1 to 3, from software Hugin. In these tables the green bars represent unknown and the red bars represent known variables. There are not really any inputs and outputs in the conventional sense of modeling, but only known and unknown variables, where the known variables have an effect on the probability of the unknown variables.

In Table 1, all variables are unknown. The uniform uninhibited corrosion rate distribution is flat and the corrosion rate could be anything, as low as 0 and as high as 10 mm/year and higher. This should be expected as no inputs are known; however, this example demonstrates that the model can still run even with no data.

Table 1. Calculation with unknown data

T		CO2		Fe2+		O2		H2S		pH		Cor. Rate	
C	%	Bar	%	ppm	%	ppb	%	ppm	%	-	%	mm/year	%
20-40	25	0-0.1	25	0-10	33	0-10	25	0-10	25	4-5	25	0-0.01	3
40-60	25	0.1-1	25	10-50	33	10-100	25	10-100	25	5-6	25	0.01-0.1	15
60-80	25	1-10	25	50-100	33	100-1000	25	100-1000	25	6-7	25	0.1-1	43
80-100	25	10-100	25			1000-10000	25	1000-10000	25	7-8	25	1-2	18
												2-5	8
												5-10	4
												>10	9

In Table 2, all input parameters are known, hence the uniform uninhibited corrosion rate is also known with a high degree of certainty.

Table 2. Calculation with all known data

T		CO2		Fe2+		O2		H2S		pH		Cor. Rate	
C	%	Bar	%	ppm	%	ppb	%	ppm	%	-	%	mm/year	%
20-40	100	0-0.1	0	0-10	100	0-10	0	0-10	0	4-5	0	0-0.01	0
40-60	0	0.1-1	0	10-50	0	10-100	100	10-100	0	5-6	100	0.01-0.1	0
60-80	0	1-10	100	50-100	0	100-1000	0	100-1000	0	6-7	0	0.1-1	99.5
80-100	0	10-100	0			1000-10000	0	1000-10000	100	7-8	0	1-2	0.5
												2-5	0
												5-10	0
												>10	0

Table 3 shows more realistic conditions where some data is known (i.e. temperature, CO_2 partial pressure and pH), some data is uncertain (O_2 concentration is lower than 100ppb, H_2S concentration is lower than 100ppm, but the exact values for O_2 and H_2S are unknown), and some data is completely unknown (Fe^{2+} concentration). Consequently the uniform uninhibited corrosion rate is known with uncertainty and is probably between 0.1 and 1 mm/year but could be as high as 5-10 mm/year with 7% probability.

Table 3. Combination of known (red), uncertain (blue) and unknown green) data

T		CO2		Fe2+		O2		H2S		pH		Cor. Rate	
C	%	Bar	%	ppm	%	ppb	%	ppm	%	-	%	mm/year	%
20-40	100	0-0.1	0	0-10	33	0-10	50	0-10	50	4-5	0	0-0.01	0
40-60	0	0.1-1	0	10-50	33	10-100	50	10-100	50	5-6	100	0.01-0.1	0
60-80	0	1-10	100	50-100	33	100-1000	0	100-1000	0	6-7	0	0.1-1	45
80-100	0	10-100	0			1000-10000	0	1000-10000	0	7-8	0	1-2	21
												2-5	27
												5-10	7
												>10	0

Depending on the rest of the model (i.e. corrosion inhibition, flow, wettability of the steel) this high uninhibited corrosion rate value might be acceptable or not. If the high uniform uninhibited corrosion rate is unacceptable, for example because of the lack of corrosion inhibition or the presence of high water–cut, then a sensitivity analysis on the model can help prioritize what data should be gathered in order to reduce the uncertainty on the uniform uninhibited corrosion rate. It should be noted that such sensitivity analysis is not generic, because the sensitivity analysis will take into account what is known of the system, when a parameter changes, so do the results of the sensitivity analysis.

As mentioned above, it is unlikely that every aspect of corrosion can be modeled with a high degree of accuracy. Thus, a methodology that assesses corrosion damage would have to depend on many different models such that uncertainties in the corrosion prediction can reduce. Because models (i.e. knowledge) improve and modeling software changes over time, the creation of one unified corrosion threat assessment methodology would require dynamic tools that can be easily updated. Bayesian networks provide an answer to both problems by:

a. combining different sources of knowledge (different models & different software products), and
b. creating a methodology that can be easily updated as new knowledge becomes available.

Similar approach can be applied also to external corrosion [7] and other threats like stress corrosion cracking, geological and geotechnical events and third party intervention.

3.3 Information Sources

The Bayesian Network approach can be applied with different sources of information.

Physics based models

One of the most reliable ways to derive conditional probability tables is to use fully tested and recognized physics based models because these represent our understanding of the underlying phenomena and recognized models can be assumed to have been peer-tested. Physics based models (such as multiphase-flow or corrosion rate models in the following examples) are run multiple times over all sets of possible inputs in a Monte Carlo fashion. It is recognized that a specific phenomenon, such as CO_2 corrosion, may be represented by a number of models that may produce different calculated results for the same input parameters. In the Bayesian network construct, multiple models can be run multiple times and results can be combined in the conditional probability tables, using weighting functions for different models. This is very useful as trust is increased in the areas where all models provide the same values and justified doubt emerges in the areas where models diverge. If the veracity of different models for the same phenomenon is not known *a priori*, equal weighting functions can be applied that can then be corrected later through observations.

Expert’s knowledge

Quantification of the causal relationships is possible using expert's knowledge. This is necessary as there are many mechanisms of pipeline failure with no reliable mechanistic model. Stress-corrosion cracking and microbiologically influenced corrosion are examples of complex phenomena where many overlapping mechanisms may operate and a detailed model of the complete phenomenon is difficult to achieve. Yet, some experts understand parts of these problems quite well and this knowledge should not be discarded because it is difficult to quantify. This knowledge is added scenario by scenario in usually smaller conditional probability tables. Conditional probability tables derived from expert's knowledge have larger uncertainties and this uncertainty is carried by the Bayesian network model all the way to the final results.

Field data

Field data can be used to populate the conditional probability tables. Every instance for each input set is counted and used to generate columns of the conditional probability tables. Using field data has many drawbacks. First, field data usually does not cover all possible sets of input parameters and when it does some combination of inputs have many measurements while others have few. Second, field data varies from field to field and estimation of the uncertainties associated to the data is difficult. And most importantly, while field experts are easily challenged, field data is rarely put into context, providing a false sense of security.

If an analytical or a numerical model is available to describe a phenomenon, then probabilistic analysis of this phenomenon can be performed using distributed inputs in to this model. If several such phenomena exist and they can be integrated into a numerical model, probabilistic analysis can be conducted by repeatedly running such an integrated model using a probabilistic driver, such as the Monte Carlo method, each run being called a realization. In such a case, the Bayesian method provides no advantage, and could indeed be less rigorous. However, in complex systems where different phenomena are connected in diverse ways and cannot be described by an integrated model, Bayesian networks provide a rigorous mathematical method to combine different types of probabilistic knowledge in order to make informed decisions. The other major advantage of Bayesian models is the reversibility of the Bayesian inference as shown in Equation. 2. Unlike other models where there are inputs and outputs, in Bayesian models there are only unknown and known probabilities. If two events are linked, then knowing the probability of one event improves the knowledge of the probability of the other event. Finally, Bayesian networks are graphical models, making the visualization of complex chains of events easy to understand, unveiling the probable mechanisms of failure. A Bayesian network model developed for corrosion helps the user understand corrosion phenomena and implicitly suggests ways to control or combat corrosion.

4. CONCLUSIONS

Uncertainties management is a key aspect in an integrity management program. Uncertain data and missing information can impose a large error and misinterpretation when assessing the risk of a pipeline.

Bayesian networks provide a mathematical framework to extract knowledge form different models under the form of conditional probability tables. The combined models can be used to provide more reliable information than a single model would.

Bayesian networks are particularly well suited to deal with uncertainty. They do not have inputs and outputs but known and unknown parameters. Known parameters combined with the knowledge contained in the conditional probability tables can help deduce the state of the unknown parameters.

Another great advantage of Bayesian networks is their ability to incorporate new evidence (new knowledge), such as the results of a new pipeline inspection, and reassess the forecast results taking into account the new evidence. This ability represents a key factor to the managers in charge of the pipeline integrity as they can update their integrity program and actions always based on the most recent knowledge of the conditions of the pipeline.

One situation where Bayesian networks can be advantageous is for unpiggable pipelines, where the lack of inspection data can be compensated by other information available.

5. REFERÊNCIAS

[1] DNV RP-F116, Integrity management of Submarine Pipeline Systems, Det Norske Veritas AS, February 2015, Norway.

[2] ASME B31.8S, Managing System integrity of Gas Pipelines, The American Society of mechanical Engineers, 2012, USA.

[3] API 1160, Managing System Integrity for Hazardous Liquid Pipelines, American Petroleum Institute, 2013, USA.

[4] KAPLAN, S., GARRICK, B.J., “On the Quantitative Definition of Risk”, Risk Analysis, 1 (1), p.11-27, 1981.

[5] AYELLO, F., JAIN, S., SRIDHAR, N., KOCH, G.H., “Quantitative Assessment of Corrosion Probability – A Bayesian Network Approach”, *Corrosion*, Vol. 70, No. 11, NACE International, November 2014.

[6] AYELLO, F., SRIDHAR, N., KOCH, G.H., KHARE, V., AL METHEN, A.W., SHABBIR, S., “Internal Corrosion Threat Assessment of Pipelines Using Bayesian Networks” , CORROSION 2014 paper 3851, NACE International, 2013, San Antonio, TX, USA.

[7] JAIN, S., SANCHEZ, A., GUAN, S., WU, S., AYELLO, S., SRIDHAR, N., FENG, Q., YAN, B., “Probabilistic Assessment of External Corrosion Rates in Buried Oil and Gas Pipelines”, CORROSION 2015 paper 5529, NACE International, 2015, Dallas, TX, USA.