21, rue d'Artois, F-75008 PARIS
http : //www.cigre.org

CIGRE US National Committee
2017 Grid of the Future Symposium

Using a Transformer Asset Health Index to Generate a Probability of Failure

C. SCHNEIDER, R. CORNELL
American Electric Power
USA

T. MCGRAIL, G. M. KENNEDY
Doble Engineering
USA

K. KOPECHANSKI
CEATI International Inc.
Canada

SUMMARY

Industry interest in power transformer failures and failure rates has led to much discussion of the actual definition of 'failure', and the applicability of statistics associated with published failure studies. It seems that the electric supply industry has a passion for condition assessment of assets. Power transformers in particular, which are indexed, classified, and ranked so that operational planning and response for large capital intensive assets can be effective, and long term financial plans can be justified. This paper will look at sources of variability in the generation of asset health indices from available raw data, and the possibility of deriving a useful probability of failure from an asset health index. Variability stems from the initial measurement, through encoding systems to classify data, and the subsequent combination of encoded data to a final health index. It is possible to relate the raw data through analytics to a probability of failure – depending on the approach used to obtain and encode the data, and the inclusion at the outset of relevant and justifiable timescales. The steps for such an approach are outlined.

KEYWORDS

Asset health index, probability of failure, condition assessment, health assessment

cmschneider@aep.com

INTRODUCTION

There is growing interest in 'asset health indices' (AHI) in the electric supply industry. The motivation is often to obtain an overview of asset performance, to indicate likely candidates for strategic planning for repair, replacement and refurbishment (1, 2, 3). In parallel, many systems in use identify candidates for maintenance or short term intervention. This range of possible timescales for action can mean there is disappointment from a particular approach. The index may not address the requirements. One solution is to ensure that we have a very clear definition of the problem to be solved. We must know what question are we going to try to answer with a single AHI? Further, how will the AHI bring value to the organization? What will each number mean? Is there a relationship between the raw data, any encoding of that data into scaled parameters, the timescales associated with the different AHI's, and a clear and auditable justification for an AHI and an action (4, 5)? If not, the resulting AHI may not perform its functions well.

It is fairly easy to develop a system for calculation of an AHI. It is possible to generate a transformer AHI systems based solely on dissolved gas analysis (DGA) of the oil in each transformer. This is not a 'bad' approach. In fact, it is a reasonable way to start. We can expect to have a reasonable volume of DGA data available, and the industry (as a whole) has studied the relationship between DGA and transformer condition.

There should be no surprises in an AHI score. It encapsulates the data we have, and what we already know in terms of analyses based on standards, guides, and heuristics. The only surprises to be expected are those that reflect a sudden change in available data, and the consequent change in the AHI. The size of the change in AHI should also have a meaning for what the change represents.

Practical AHI systems may take many pieces of raw data, apply numerous coding and functions to develop component scores, and then collate results to a final AHI. Again, this is not 'bad' – but has to be considered in terms of whether it is useful in answering a question (6). An analogy may help. If the tire pressure monitor in a car gives the average pressure of the 4 tires – it may be said to give an indication of the overall health of the 4 tires. However, tires do not usually fail 'on average'; they fail 'in particular'. An individual tire may need to be addressed with some urgency – an urgency which is lost when we look at the average pressure. AHI systems which weight different parameters are a form of averaging – any sense of urgency can be rapidly lost in the analyses. Here is an example of a practical system where this occurs. Combining and weighting data from tires, transmission, cooling, steering, engine, etc. loses focus on any one aspect of the car. The resulting system may not retain monotonicity. That is, the AHI index may seem to be 'better' for a transformer which is in worse condition, and requiring more urgent intervention.

Care must be taken in developing an index: if we need a means to identify maintenance intervention in the short term and replacement in the longer term, we may need two different indices. The development of an AHI as a technical effort is not necessarily complex. However, at the onset, the development should include the relationship between data, symptoms, failure modes and timescales. If the AHI is not based on those relationships, and retained through the analytic compilation of component scores and a subsequent AHI, then it will be difficult to derive any timescale for action with any degree of confidence – even for relative ranking purposes.

DATA AND STATISTICS

In making a measurement of some value, our measurement technique and measurement system will provide sources of both systematic and random errors. The result of the measurement is only an 'estimate' of the true value (7). Numerous measurements of the same value will provide a 'distribution' around the actual value, often in the form of a Normal (a.k.a. Gaussian) distribution, symmetrical about a 'true' value. The spread of the distribution is characterized by the standard deviation, the confidence interval, and the error. In Figure 1, a measurement of 25 units is characterized by an error of +/-10%, with a confidence interval of 90%, and a resulting standard deviation of 1.52 units. The vertical axis indicates the probability that the result is at a particular x-axis value.

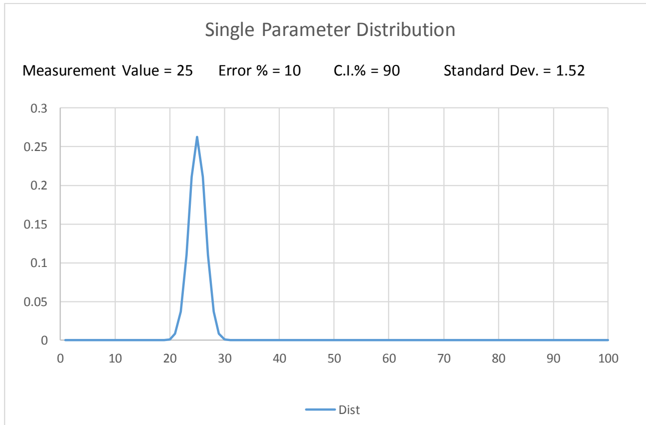

Figure 1
Parameter Measurement with Normal Distribution

The data in Figure 1 shows that there is uncertainty built into our measurements. Any system which attempts to encode the data into a scale (say 1-5 for condition), will be subject to error due to the precision of the original measurement. Table 1 shows possible category, or coding boundaries for the measurement in Figure 1.

Table 1
Category or Coding Limits for a 0-100 Measurement

Category/Code	Lower Limit	Upper Limit
1	0	20
2	20	40
3	40	60
4	60	80
5	80	100

Overlaying the category boundaries on a new measurement, which lies near a boundary, shows the likelihood that the 'true' result is one or other of the categories, Figure 2.

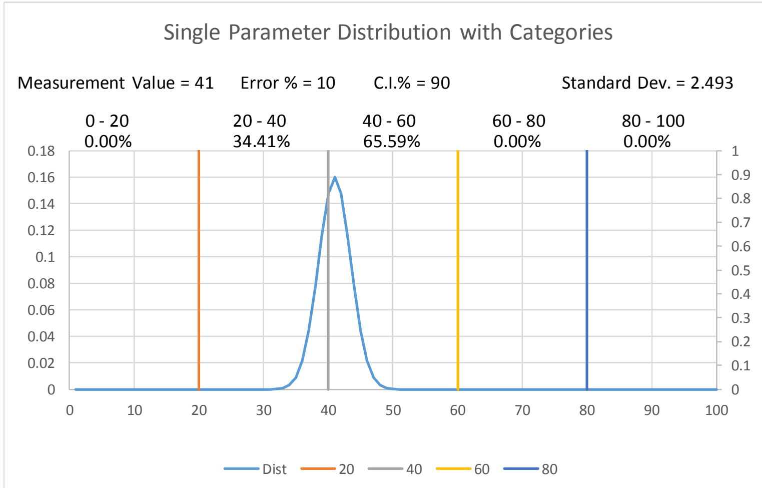

Figure 2
Parameter Measurement with Normal Distribution

The resulting category – the condition code for this parameter – is almost twice as likely to be code 3 (40-60) than code 2 (20-40). But the encoding is not definitive, as the raw data is not definitive.

As an example, consider a car tire at 28 psi. How do we relate that measurement to the probability of failure of the tire? The probability of failure of the car? Without extensive experiments and many actual failures, our data will always be uncertain, and the relationship with a probability of failure even more uncertain. In Figure 3, two examples of relationships between measured parameters and failure rates are shown: a linear version and a logistic. The logistic is possibly more realistic, as it is bound by an upper limit of 100%.

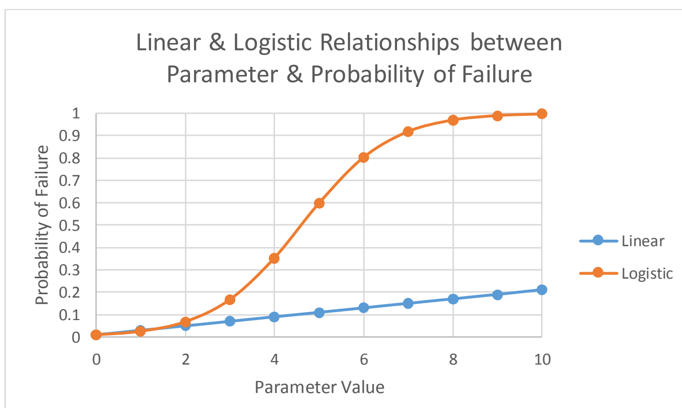

Figure 3
Linear and Logistic Relationships for Measured Parameter and Probability of Failure

But the relationship between parameter and probability of failure cannot be developed without extensive test results. In the case of tire pressures, there would need to pump a lot of tires to the appropriate pressure, and then failures recorded in practice. An effort to do this was undertaken for some DGA results.

CIGRE Technical Brochure 296 (8) summarizes DGA data from oil samples taken from the bottom of the main tank of a population of transformers. The samples were taken ‘shortly before or after’ a fault, so assumed to correspond to the fault. The timescale of ‘shortly’ is not defined. Charts were developed linking gas concentration and probability of failure in service, as shown in Figure 4 for acetylene concentrations.

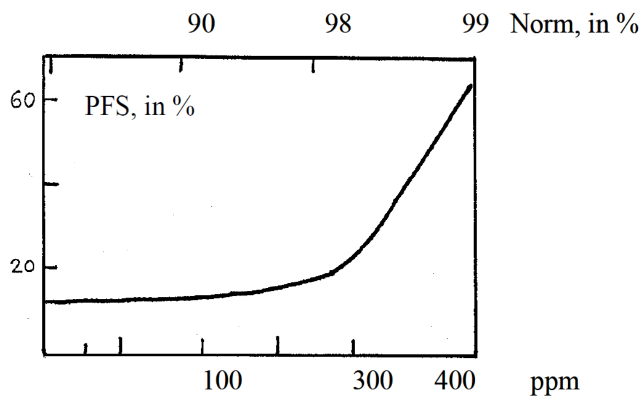

Figure 4
CIGRE TB 296: Acetylene Concentration and Probability of Failure (Copyright CIGRE 2006)

The shape of the curve in Figure 4 is somewhat logistic – but is it expected that the probability of failure below concentrations of 100 ppm acetylene to be constant at about 12%? The applicability of the results may be limited to the originating organization. How then, is it possible to generate data for a wide range of assets? Transformers are often treated as fungible assets but are, in general, unique and predominantly handmade devices. They may have families, but they are not all interchangeable.

CIGRE TB 642 is a transformer reliability survey covering 56 utilities in 21 countries, and including more than 950 known transformer failures in generation, transmission, and distribution (9). The data was collated and analyzed, with one note: "All populations show a low hazard rate and no distinct bathtub curve character". This is interesting in that aging transformers do not seem to correlate with an increasing likelihood of failure. If we use age as an input to an asset health analysis, we may be responding more to prejudice than fact.

Context is also very important (5, 10, 11, 12, 13). If the loss angle (dissipation factor or power factor) is measured for a transformer main winding, how is a result of 0.75% interpreted? For some manufacturers, this result would be higher than expected, but within the lower 75th percentile. For another manufacturer, such a value may put the result in the top 5th percentile. What are we to make of the distribution? We know, intuitively, that a higher power factor is an indication of insulation deterioration, but what does it mean probability of failure? What is the failure mode? Does our raw data lead to an indication of what failure mode is in operation, and how long it will take to reach culmination?

There is much variability of raw data, including the distribution around a true value, the subsequent lack of definitiveness in encoding a parameter to a code or category, and the paucity of data linking a parameter or category to an actual probability of failure. We can see that any asset health index is likely to be far removed from any real probability. As a note, if the encoded categories do not have a timescale associated with them, and probability of failure within that timescale, then any subsequent analysis is going to be fuzzy, at best.

FURTHER ANALYSIS

Figure 5 shows the IEEE C57.104 condition codes for a variety of dissolved gases used in transformer DGA analyses (14).

Status	Dissolved key gas concentration limits [μL/L (ppm)[a]]							
	Hydrogen (H_2)	Methane (CH_4)	Acetylene (C_2H_2)	Ethylene (C_2H_4)	Ethane (C_2H_6)	Carbon monoxide (CO)	Carbon dioxide (CO_2)	TDCG[b]
Condition 1	100	120	1	50	65	350	2 500	720
Condition 2	101–700	121–400	2–9	51–100	66–100	351–570	2 500–4 000	721–1920
Condition 3	701–1800	401–1000	10–35	101–200	101–150	571–1400	4 001–10 000	1921–4630
Condition 4	>1800	>1000	>35	>200	>150	>1400	>10 000	>4630

Figure 5
IEEE C57.104 Dissolved Gas Analysis Codes (Copyright IEEE)

The codes in figure 5 are based on distribution data of multiple populations of transformers, and anecdotal/actual evidence of failures – to allow for increasing condition code to reflect poorer health. But none of the codes have any indication of what action must be taken, or how soon an action should be taken – apart from TDCG, which recommends resampling in given timescales.

The data available for development of an AHI must be encoded so that an overall health index may be produced. The different codes must be calibrated. For example, data encoded as 3 for Hydrogen, must have the same sense of urgency as a 3 for vibration. That sense of urgency may be recorded as a timescale for action. The action may be different for different data sources, but the urgency must be calibrated, or we will find that some 3's are more urgent than others. Combining such uncalibrated data will be impossible.

For example, in Figure 6, multiple data sources have been collated and coded based on predefined rules. This gives a score for several factors. The codes are based on: 1= good and 5= bad. The individual factors are then combined by simple addition.

Factor	Trf 1	Trf 2	Trf 3
DGA Main Tank Score	2	1	1
Dielectric Score	1	1	1
Thermal Score	2	1	1
Mechanical Score	3	4	1
Oil Score	1	1	1
DGA LTC Tank Score	3	1	5
Operational Score	2	3	3
Design/manufacturer Score	1	4	1
Subject Matter Expert Score	3	1	2
Sum	**18**	**17**	**16**
Normalized Sum (%)	40.0	37.8	35.6

Figure 6
Multiple Factor Combination via Simple Addition

Simple addition is a version of a weighting scheme where all of the encoded values are weighted identically (usually with the weight set to a value of 1). The system in Figure 6 is calibrated, so there is the same timescale for action for all 2's, and a different (but consistent) timescale for all other codes. Which transformer in Figure 6 is most urgent? Transformer number 3 has a score of 5, the most urgent condition code, for one factor. The simple sum hides this fact. Normalizing the sum to a percent of a maximum possible score does not help the interpretation. The combination process – the weighted sum -- has lost the sense of urgency and, thus, the link to probability of failure.

Using different weightings does not necessarily help with prioritization, due to the fact that we could have a small change in a heavily weighted score – producing the same overall effect as a large change in a lightly weighted score. The link between condition, 'cause' and resulting health score 'effect' is broken.

In addition, when available data changes and a parameter value moves across a boundary, the precision of the measurement means that there is uncertainty, and possibly significant uncertainty, in where the 'true' measurement lies. The resulting code is uncertain, and the resulting health score is uncertain.

There are systems which help urgent data stand out. Log scales put more emphasis on higher codes. In Figure 7, the linear 1-5 of the data in figure 6 is replaced with an exponential/log style score.

Linear	Log
0	0
1	1
2	3
3	10
4	30
5	100

Factor	Trf 1	Trf 2	Trf 3
DGA Main Tank Score	3	1	1
Dielectric Score	1	1	1
Thermal Score	3	1	1
Mechanical Score	10	30	1
Oil Score	1	1	1
DGA LTC Tank Score	10	1	100
Operational Score	3	10	10
Design/manufacturer Score	1	30	1
Subject Matter Expert Score	10	1	3
Sum	**42**	**76**	**119**
Normalized Sum %)	4.7	8.4	13.2

Figure 7
Multiple Factor Combination via Log Scale and Simple Addition

The individual codes are calibrated so that all 3's still have the same time scale for action. A score of 100 is most urgent. The normalized sum becomes unimportant, as we are looking for any simple addition sum where the total exceeds 100. In such a case, a contributory score may be at 100 and, thus, most urgent. The system described here is roughly log3, so that three contributory scores of a particular level are almost equivalent to one score at the next level up.

The log system retains the sense of urgency needed to indicate assets in poorer health, and is more likely to fail.

EXAMPLE OF A LOG AHI SYSTEM IN USE

There are many asset health systems that take a set of raw data, and then manipulate that data into a final score – without referencing failure modes or timescales. Such systems may have some relative ranking capability, but this is not guaranteed. The ranking may not be monotonic with the most urgent cases – not necessarily having the worst score.

The system described here is in practical use in a utility. Figure 8 shows an extract from a table of transformers evaluated using a log3 scale approach, with transformer identifiers removed. The table is ranked by "Overall Condition – Now"

			Overall Condition			Core and Windings			Oil		OLTC	Exterior
Design	sign	Year	Now	Mitigated	Possible Ir	Dielectri	Thermal	Mechanic	Ageing	Contamina		
A04a	32	1965	221	213	8	100	100	1	13	10	3	10
E11b	32	1959	170	103	68	30	60	1	190	10	10	10
G02b	104	1994	170	135	35	30	60		36	100		1
E11b	32	1959	154	143	11	30	100	1	23	10	10	3
D07	12	1965	152	126	26	60	60	1	70	10	1	
M01	5	1957	151	94	57	30	60	1	160	10	3	10
H02	111	1971	147	100	47	3	60		140			3
P21	104	1972	144	139	5	1	3	100	13		1	10
C04	32	1968	138	85	54	10	60	1	140	30	1	
H07a	12	1964	133	107	26	1	100	1	70	10	3	
A04b	102	1967	132	106	26	10	60	1	70	10	3	
P06a	131	1967	131	107	24	1	60	1	63	10	1	10
H07a	12	1964	129	106	23	1	100		63	10	1	
E11a	102	1955	129	105	24	10	60	1	70		1	10
H07a	12	1966	129	106	23	1	100		63	10	1	
E11b	32	1959	129	107	22	30	60	3	43	30	10	
F08	120	1956	124	105	19	3	60	1	50	10	1	
L05	111	1962	122	99	23	1	60		63	10	1	
A04b	102	1967	122	96	26	1	60		70	10	3	
A10	3	1960	122	106	16	100	3	1	40	10	1	

Figure 8
Logarithmic approach to transformer analysis

The logarithmic approach is applied to a range of factors which are grouped together:

- Core and Windings: evaluated via dielectric, thermal, and mechanical factors
- Oil: evaluated through ageing, contaminants.
- Manufacturer and design
- OLTC and other factors

Note that the table in Figure 8, in reality, extends well to the right – with other factors and components evaluated. The initial score calculated is a 'Now' value for Overall Condition. In addition, each transformer is analyzed for likely failure modes operating and what can be done to mitigate them (resulting in a 'Mitigated Condition'), and an improvement. Intervention can then be planned by looking at risk and cost benefit for a particular intervention.

The log scale makes sure that urgent cases are at the top of the list. Each transformer is analyzed individually to identify family issues. An assessment would evaluate whether the failure mode is likely to be benign or involve tank rupture and possible substantial collateral damage. These consequence analyses are included in the risk analysis.

Transformers are then grouped into an action plan, with those which: need attention 'immediately'; on the replacement list for the next 2 years; 2-5 years, 5-15 years, and those seen as being good for the foreseeable future. This broad brush approach allows for volume analysis of transformer replacement – a strategic goal of the AHI at the outset – with a timescale for action. Urgent cases are dealt with as maintainable items.

This log approach is predicated on the AHI being used for long term replacement. A separate analysis is performed to review bushings, tap changers, and other maintainable items. This is equivalent to addressing car tire issues separately to the car itself. It is understood that the consequence of a tire failure may be the replacement of the car, but a new car is not required just because one tire looks to be in bad condition.

A probability of failure may be calculated indirectly. There ae timescales associated with intervention required based on the values of the raw data – timescales which are calibrated across the whole data set. Higher scores have a more urgent need and a shorter timescale. These have a higher probability of failure. How high? Difficult to say based on the math and discussion we have seen in this paper. But by preserving the sense of urgency through calibrated analyses and log scales, the relative probability is also preserved, whatever that might be. This leads to questions:

What is an acceptable probability of failure over the next year for a particular transformer?

How about over the next week? How about during peak summer load? How about during the next storm?

These values are not easy to either assess or calculate, and may be subject to organizational concerns over ‘admitting’ that the probability of failure is not zero.

LINKING ASSET HEALTH INDEX AND PROBABILITY OF FAILURE

If timescales and/or probabilities are not put in at the ‘front end’ of the AHI effort, and related to raw data and failure modes, it will be difficult to generate a reasonable probability of failure at the ‘far end’.

There is an approach where, if the AHI retains relative urgency, we can map the ranked list of transformers through to historical failure rates, and predict which units are most likely to fail within a given timescale. This is, in fact, a very appealing approach – as it’s based on empirical data in terms of failures, and can be justified in that populations are unlikely to be varying rapidly. It does require that future years will look somewhat like previous years, and that the population of transformers does not change too rapidly, year on year.

It is possible to build probability of failure into an AHI from the outset. Caveats discussed in this paper include data precision, poor relationship between parameter values and failure rates, conflation of imprecision through data encoding techniques (1-5, A-D etc), further conflation through component score combinations and weightings, and overestimation of the accuracy of the final result. We have to be careful to check candidates against known failure modes. What did other, similar, units do? Is the diagnosis bad, the prognosis good? Ranking is the start of an asset health review – not the end.

In any AHI system which generates a score, there needs to be an understanding of the accuracy or precision of the final score. If it comes out as 4.8 on a scale of 1 to 10, and is accepted as the health of the transformer, remember that 4.8 is an estimate – and has a degree of uncertainty. The problem is that, if the AHI is not monotonic in relation to urgency, the final ranking is not monotonic. A higher/worse score may not reflect the urgency in any degree (see Figure 6). Without understanding the source of the AHI, making claims about the ranking can be extremely misleading. To put it another way: will the top units in an AHI list definitely be the ones that fail over the next few years?

The ‘true’ probability of failure will collate condition-based data within operational and external parameters. The act of maintenance can lead to conditions which lead to failure.

AN APPROACH BASED ON THE PRINCIPLE OF EQUIVALENT PROBABILITY

The approach identified here is based on actions and timescales associated with raw or derived data values or parameters. In a very simple example, a value is measured for a parameter and an action associated with it depending on a binary view. The data is encoded as indicating ’good’ or indicating ‘bad’. The data could be hydrogen level in a DGA analysis, or a power factor, or a derived value of interest.

For each encoded condition, there is an associated action and a timescale. For example:

- ‘good’: re-measure and review in one year
- ‘bad’ : perform further testing and evaluation within 1 month

Please note the values and timescales are purely indicative.

The overall historic and expected failure rate for the asset is 0.5%, with the expectation for a ‘good’ asset that it has a 0.5% failure probability in the year-long timescale. Then for the ‘bad’ asset, it is possible to state that it is has an equivalent probability in the timescale identified (0.5% in one month in this instance). The annual equivalent is not 6%, which is the plain multiplication, but 5.8%, based on the math of probability analysis. Table 2 summarizes the approach.

Table 2

Equivalent Annual Probabilities

Code	Hydrogen	Timescale	Action	PoF	Annual Equivalent
Good	<100	1 year	Resample	0.5%	0.5%
Bad	>=100	1 month	Replace	0.5%	5.8%

The approach can be extended to multiple parameters, with multiple codes (for example: good, fair, poor, bad), and multiple components. It can be applied to numeric condition codes: 1 through 10. There is no need to have a log or a linear scale, but all encodings will have the same equivalent probability.

It is useful to have equivalence based on time applied to failure modes – not just an individual analytic which encodes data. This requires collating data at the failure mode level - and maintaining the calibrated timescales. It should be noted that having more contributory data to collate for a given failure mode will improve precision – confirming the urgency of the situation. Collated scores for a failure mode should reflect the urgency. Having more data which is 'good' does not ameliorate the 'bad' data, unless we can redefine what the relevance of a component to the failure mode is.

The overall approach is summarized as follows:

1. Identify the assets of interest
2. For each asset class: identify the components (or asset subsystems, or whatever they are called in the asset system)
3. For each component: identify high level failure modes (a simple RCM analysis will suffice)
4. For available data: identify analytics (simple, standard, *ad hoc*) which indicate a failure mode in operation
5. Score each analytic with a consistent and calibrated timescale (of appropriate values for the application), with each code/category labelled and assigned a PoF
6. For each analytic, identify the relevance to each of the failure modes
7. Collate analytics for each failure mode to score the failure mode – action & timescale – based on probability equivalence and calibrated codes
8. Collate Failure modes for each component, and score the component, based on probability equivalence and calibrated codes
9. Collate failure modes for each component, and score the asset, based on probability equivalence and calibrated codes

The result is a set of scores which are calibrated and consistent across the assets of interest. The approach has not yet been fully implemented in any practical systems, but at least one practical system follows the main detail – permitting calibrated probabilities based on timescales – and applies those to the assets of interest.

SUMMARY

It is unlikely that we will soon have a purely scientific and mathematically rigorous approach to Asset Health Indices, which begins with data and relates to failure modes directly – with extensive statistics for measurable parameters and consequent failure. -. Using ranking and looking for relatively poor condition units is valuable, but needs to be done with further consideration of the fleet and operational environment.

A clear statement of the role and application of an AHI helps identify what data to include in the generation of the AHI. The process by which an AHI is generated removes information from the final statistic, and the value of the final statistic must be clear – is it useful?

Consistent calibration of encoded data, based on equivalent probabilities, help allow for justifiable AHI's which are: traceable, refer to probabilities entered at the outset, and are consistent across the analysis from data to AHI.

ACKNOWLEDGEMENT

This paper is the result of project T163700-30/113: "*Translating the Health Index into Probability of Failure"*, undertaken for and funded by the electrical utility Members of CEATI's Station Equipment Asset Management (SEAM) Program.

BIBLIOGRAPHY

[1] T. McGrail,"Practical Steps to Improve Asset Management Applications", TrafoTech Conference, New Delhi, India, 2016
[2] T. McGrail, D. Angell & K. Elkinson, "Asset Management - Data, Analysis & Decisions", Electric Light & Power magazine, Volume 18, Issue 7, July 2013
[3] Tony McGrail, "Asset Management: Data and Decisions", IEEE International Smart Grid Conference, Perth Western Australia, 2011
[4] T. McGrail, K. Elkinson, "Substation Asset Life Decisions: Replacement v Refurbishment", Minnesota Power System Conference (MIPSYCON), Minneapolis, USA, 2013
[5] T. McGrail, K. Elkinson, R. Heywood, "The 3 C's of Condition Monitoring: Control, Context, Conclusion", SEA Marconi "MyTransfo" Conference, Turin, Italy, 2016
[6] T. McGrail, "Connecting Asset Health Indices, Asset Probability of Failure and Risk", CIGRE Canada Technical Conference, Vancouver, Canada, 2016
[7] S. Stigler, "The 7 Pillars of Statistical Wisdom", Harvard University Press, 2016
[8] "Recent Developments in DGA Interpretation", CIGRE Technical Brochure 296, Pub. 2006
[9] "Transformer Reliability Survey", CIGRE Technical Brochure 642, Pub. 2015
[10] T. McGrail, R. Heywood, "Intelligent Transformers: Monitoring, smart grid integration and asset management", Coil Winding Insulation and Electrical Manufacturing Exhibition, Berlin, Germany, 2014
[11] T. McGrail, K. Elkinson, M. Lawrence et al, "Aspects of Power Transformer Asset Management", IEEE PES T&D Conference, Orlando, USA, 2012
[12] K. Wyper, G. MacKay, TransGrid Australia & T. McGrail "Condition Monitoring in the Real World" 80th International Conference of Doble Clients, Boston, USA, 2013
[13] T. McGrail, "Asset Health Indices: What's the Point?", TJH2B TechCon, 2016
[14] The Interpretation of Dissolved Gas Analyses of Oil Filled Transformers", IEEE C57.104 , IEEE, 2011