Data Engineering Pipeline for High-Voltage Transformer Asset Management

Yutao Zheng

Melinda Hodkiewicz
Department of Mechanical Engineering
University of Western Australia

Adriano Polpo de Campus
Mathematics & Statistics
University of Western Australia

Desmond McEwan & John Nazimek
CEED Client: Water Corporation

Abstract

The Water Corporation has a large asset base which includes hundreds of high voltage transforers. In order to allocate capital funding appropriately, there is a need to indentigy when transformers requir replacements. In the absence of failure data, other methods have to be developed to risk rank the transformers for prioritsation for renewals planning.

This project develops a pipeline to aggregate data from maintenance and condition monitoring data over the past 22-years. The resulting data frame can be used to answer queries relevant to asset managers. Examples include "what is the reliability event implied by the work order?", "did the work order record an end of life event correctly?", and "was preventative work completed as planned?". Answers to these questions are important input for reliability assessment and data quality improvement. The pipeline will produce data to feed into the electrical asset class plan which will include a PowerBI dashboard.

1. Introduction

The Water Corporation owns and operates over seven hundred high voltage transformers in their plant across Western Australia. By definition in AS/NZS 3000:2018, "high voltage (HV)" is exceeding 1kV for alternating current. To facilitate improved renewals planning, the Water Corporation intends to develop more reliable estimates of the remaining useful life (RUL) for their HV transformers. The Water Corporaiton has spent approximately $136k per year on average maintaining and repairing transformers over the last 22 years. In the current practice, estimation of transformer life is based on age, industry norms and engineering judgement. Integration of data from condition monitoring reports and historical maintenance records is time-consuming as this data can be difficult to locate and collate digitally. This leads to over conservative decision-making for renewal and disposal.

Transformers play a key role in power distribution system. Figure 1 depicts a typical power distribution system. Only those elements inside the dash line box are applicable to the Water Corporation's assets base.

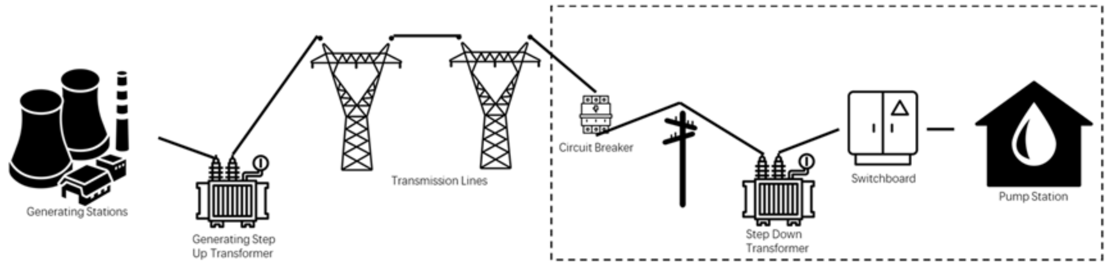

Figure 1 A Typical Power Distribution System. The current flow is 3-phase alternating current which is represented by a simplified single line connecting each element. Only the step down transformer is the scope of this research.

Transformers have a proven record of high reliability with a failure rate is approximately 1% (Martin et.al., 2017). A Weibull distribution on transformer life data from approximately 6000 transformers in period of 2000-2015) suggests there are two distributions, with a change in distribution occuring at approximately 20 years, which separates the failures due to infant mortality and age-related failure modes (Martin et al. (2017).

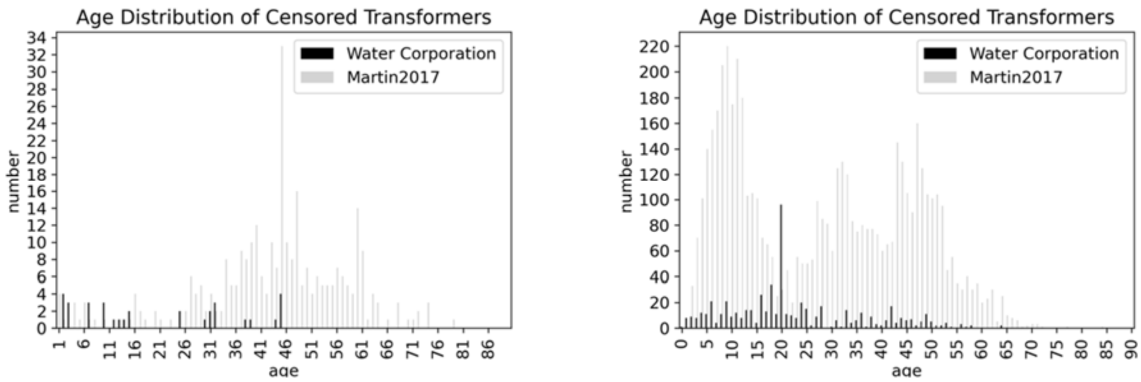

Figure 2 Age Distribution of Censored (Left) and Retired (Right) Power Transformer. Data is from the Water Corporation from 1999-2020 and Martin et al. (2017). Censored transformers are in operation at this moment, of which the start up year is known but the end of life is unknown.

Figure 2 shows a histogram of the life data in the Martin study (in light), the current age of the Water Corporation operational transformers and end-of-life age of the WC decommissioned transformers (in dark). Significant differences between Martin's and Water Corporation's datasets are apparent indicating that the parameters from the Weibull distribution used in Martin's research may not be appropriate for Water Corporation's datasets. Other factors besides age, such as condition monitoring data and operating environment, should be taken into consideration to evaluate Water Corporation transformer reliability.

The specific aim of this study is to develop a repeatable and transparent process which constructs reliability performance measures of HV transformers owned and operated by the Water Corporation. The scope of this project includes:

- Collection and organisation of transformer characteristics, work order and condition monitoring data for the last 22 years (1999 – 2020).
- Perform data analysis on available data to describe the life distribution of selected HV power transformers.
- Assessment of existing and potential condition monitoring data and maintenance strategies.

- Development of data for the Water Corporation's asset management planning dashboard for HV transformers.

2. Process

The construction of the data pipeline begins with understanding the database used in the Water Corporation. There are three databases involved in the process of data acquisition for HV transformers: SAP, Nexus and HV equipment register sheet. SAP includes a computerised maintenance management system which records asset information such as transformer functional location and maintenance work orders. Nexus is the Water Corporations Document Management System, which includes storage of oil analysis reports. The HV equipment register sheet is in development and aims to record characteristic data such as rating voltage/power and manufacturing information. Data is extracted from these three databases as csv files and imported into Jupyter Notebook for analysis using Python code.

All assumptions made in processing are documented explicitly in the code. The following table demonstrates how rules used to infer the reliability event implied by the work order are documented. This clarity of documentation for processing is not possible using Excel.

Event Name	Degree of Impact on Transformer Life	Criteria
End of Life	It is the end of the transformer life.	The work orders contain keywords of "replace, install and remove" and any possible notation or misspelling related to them (case insensitive);
		The actual cost of the work orders is greater than $5000;
		The actual time spend on this work orders is greater than 0;
		The work order is classified as corrective
Repair	Major impact on transformer reliability	The work orders contain keywords of "repair" and any possible notation or misspelling related to them (case insensitive);
		The actual cost of the work orders is greater than $100;
		The actual time spend on this work orders is greater than 1h
		The work order is classified as corrective.
Mainten-ance	Minor impact on transformer reliability	The work orders contain keywords of "service/inspection" and any possible notation or misspelling related to them (case insensitive);
		The actual cost of the work orders is greater than $100;
		The actual time spend on this work orders is greater than 1h
		The work order is classified as preventative.
Oil Analysis	Minor impact on transformer reliability	The work order has been associated to an oil analysis report
Unsure	No impact on transformer reliability	The work orders contain keywords of "inspection/test/reset/restart" and any possible notation or misspelling related to them (case insensitive)
		The actual cost of the work orders is greater than $100;
		The actual time spend on this work orders is greater than 0;

Table 1 Rule-Based Method to infer specific reliability events described in work order records

Data from Harris Dam pump station, Munster sewage pump station, Beenyup wastewater recovery facility and Anaconda Dr wasterwater pump station were selected for analysis as they have each had a transformer failure recently.

3. Preliminary Results and Discussion

The first step of this project was to collect, merge and clean the transformer characteristics, work order and condition monitoring data from the three databases. This is shown in in Figure 3. The characteristic data is collected from the HV equipment register sheet and SAP, includes the transformer functional location, description, user status, location and start up date. The transformer maintenance data from SAP includes the type of work order (preventative or corrective), the cost, the completed task code and the actual start and finish time. The condition monitoring data had to be extracted by hand from oil analysis reports stored in Nexus and includes moisture, average breakdown voltage and results of dissolved gas analysis. Beacause functional location is the key common identifier for transformer inforamtion, work order and condition monitoring data, it is utilized to merge the work order and condition monitoring data.

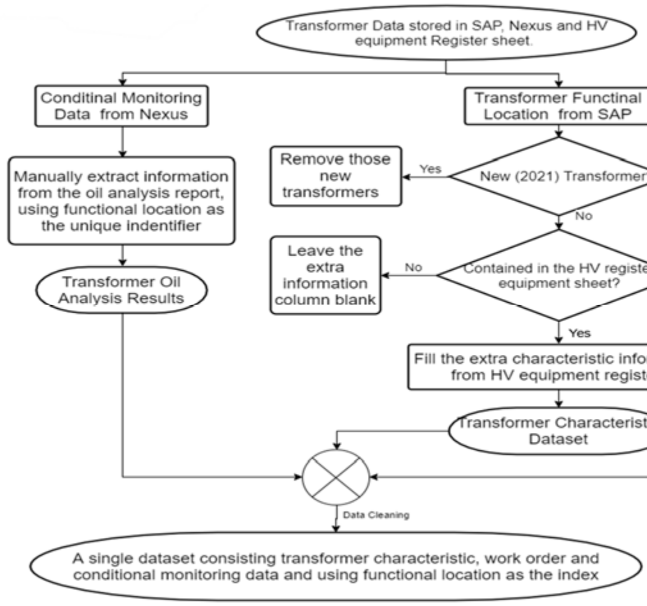

Figure 3 Pipeline for Data Acqusition

Extracted data from SAP and NEXUS needs cleansing before further processing and analysis. The resulting merged data is in a single data frame on which rules can be run, as well as performing querying and analysis. The pipeline is shown in Figure 4. Firstly, codes specific to SAP such as PM01 and PM02 are replaced by “corrective”, PM03 is replaced by “capital” and PM04 is replaced by “preventative”. The date and time data in “txt” format is converted into “datelike” format so Jupyter Notebook can operate calculation on those date and time data. Based on some assumptions, the time spent on each work order can be estimated.

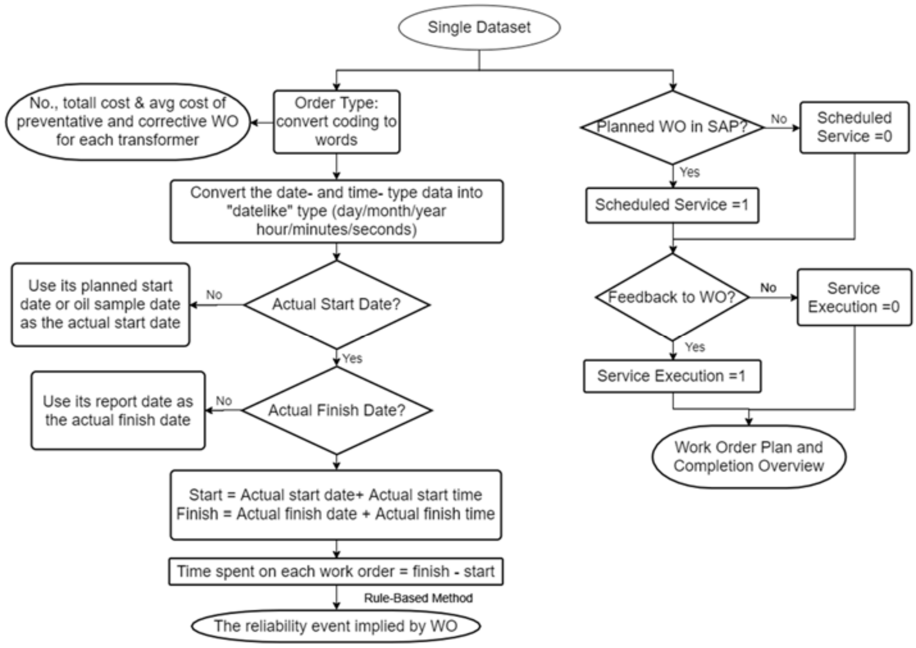

Figure 4 Pipeline for Data Processing and Analysis

Applying the rule-based method to the data from 4 selected plants, there are 215 out of 375 (57%) work orders implying a reliability event. Those unmarked records are not removed from the datasets but kept for futher validation. The numbers of each event are 2 EOL, 27 repair, 104 maintenance, 44 oil analysis and 38 unsure. These results are reasonable where maintenance should be the majority event and end of life and repair is relatively rare. However, the number of end of life events does not match the reality. In fact, there should be at least 4 end of life events because the 4 selected plants have had at least 4 transformer failures in the past.

“Schedule service” and “service execution” have been introduced and evaluated as shown in figure 4. Those two values help whether the scheduled work order was closed properly by completing the feedback reports. There are 49 out of 375 (13%) planned work orders without their feedback so the time spent on actual work is unknown. There are 42 out of 375 (11.2%) missing their planned work order but having their feedback as the oil analysis reports. In other words, there are 42 of 44 oil analysis reports that exist but those oil analysis events are not recorded in the transformer operation data, i.e. in SAP.

Figure 5 shows the order cost distribution for the transformers from the four selected plants. Most of the work orders are located in the range from \$0 to \$500, and only a few work orders (8 in this case) are above \$10,000.

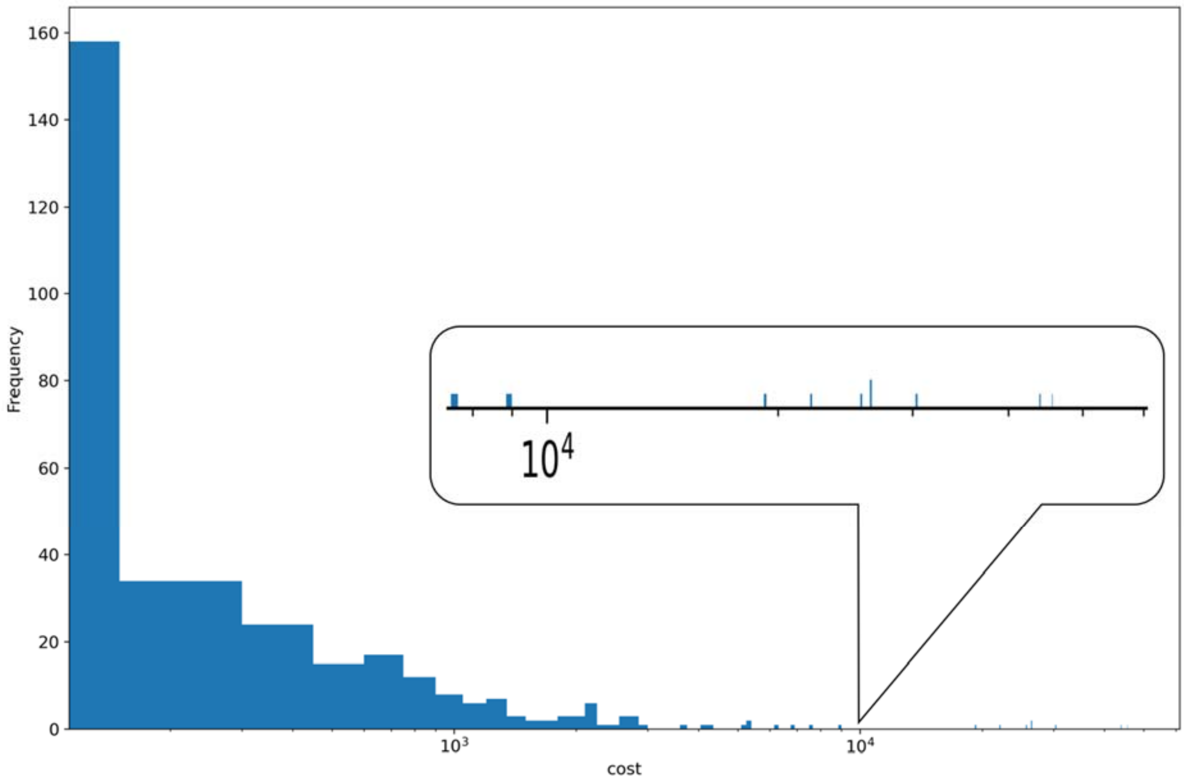

Figure 5 HV Transformer Work Order Cost Distribution

4. Conclusions and Future Work

This work suggests that the Weibull analysis parameters for transformer life estimation produced in Martin’s research are not appropriate for estimating a quantitative reliability performance measure for high-voltage transformers owned and operated by the Water Corporation. The Water Corporation transformer population has experienced very few failures and so statistical analysis of life data on its own has too much uncertainty. To reduce this uncertainty this work develops a pipeline to bring other fortuitous data to the analysis process, namely information on whether scheduled maintenance was completed (or not) and historical oil analysis data. The contributions of this work are 1) the pipeline to collect data from multiple sources into a single data frame of clean data, and 2) the development of an explicit rule set for processing the data that is transparent and replicable. The pipeline will be used to present data in PowerBI for use by the Asset Planning team. Because there are lacking of failures in the data set, the application of traditional frequentist reliability analysis is not sensible.

Future works should focus on 1) automatic extraction from the oil analysis reports. The extraction was done manually in this project but can be done semi-automatically by using advanced image processing techniques which can extract information from PDF file, and 2) the application of Bayesian methods incorporating expert prior knowledge and covariates from oil analyis and compliance with scheduled maintenance activities, to improve relative risk ranking estimates for the Water Corporation's transformer asset fleet.

5. Acknowledgements

The author would like to express my deepest appreciation to Prof. Melinda Hodkiewicz, Assoc. Prof. Adriano Polpe de Campus, Mr. Desmond McEwan and Mr. John Nazimek, the In-Service Assets Metro Team and Asset Performance Team at the Water Corporation and to Dr. Jeremy Leggoe and Ms Amanda Bolt from the CEED office for organizing this wonderful program.

6. References

Martin, D., Marks, J., Saha, T., Krause, O., Russell, G., & Alibegovic-Memisevic, A. (2017). On the development of power transformer failure models: An Australian case study. 2017 IEEE Power & Energy Society General Meeting, 1–5. https://doi.org/10.1109/PESGM.2017.8274571