SUBMARINE CABLE FAULTS IN THE ERA OF ARTIFICIAL INTELLIGENCE: FORECASTING, PREVENTION AND MITIGATION

Federico Grasso (Elettra Tlc S.p.A.)
Email: f.grasso@elettratlc.it

Elettra Tlc S.p.A, Deposito Cavi Sottomarini, Molo di Mezzogiorno snc, 95121 Catania

Abstract: Submarine cable faults do not occur entirely at random. Historical data show a correlation between the cause of damage and the geographical zone where it happened or the period of the year. Sometimes, even the severity of the fault can be related to other factors. Excluding unpredictable events that can happen anywhere anytime (earthquakes, landslides, turbidity currents), it is possible to outline a predictive model of risk that affects submarine cables for any given region.
The author fed two different open-source machine-learning algorithms (Random Forest and XGBoost) with nearly one decade of cable faults (more than 1500 events) in order to obtain the probability of a fault and its severity occurring at a given location in the future by correlating together a multitude of data, such as cable's RPL, cable type, burial status, water depth, time of the year, known fishing areas and others. Together with its potential, limitations and bias of this approach are also discussed.

1. SUBMARINE CABLE FAULTS: AN OVERVIEW

The density of network cables on the ocean bottom has made cable protection and network planning crucial to the uninterrupted service of undersea communication. Understanding these events and the resulting faults underlies the purpose of many fault studies [1].

Submarine cable faults do not occur entirely at random. Historical data show a correlation between the cause of the damage and the zone of seabed where it occurred. Faults due to anchoring or trawling, for example, are definitely more probable at shallow depths. [2]

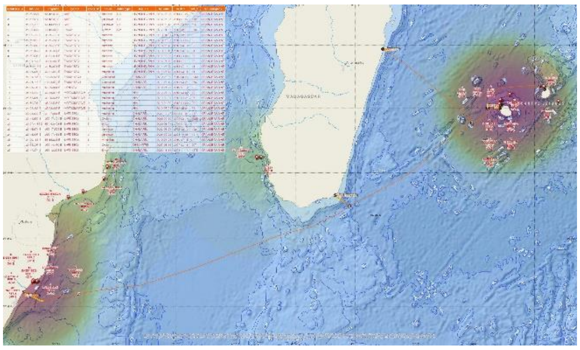

Figure 1: Cable faults from 2000 to 2017 in southern Indian Ocean.

Figure 1 shows a plot of all faults occurred to submarine cables in the southern portion of the Indian Ocean between the Mascarene Islands and South Africa between 2000 and 2017. The concentration of faults is highlighted in shades of red. It is apparent that almost all faults are contained within the 500m water depth area, as confirmed by other studies [3].

2. THE ALGORITHM

The Random Forest algorithm was specifically chosen because it allows to rank the importance of variables in a regression or classification problem in a natural way. In order to forecast the occurrence of a cable fault the following variables have been used:

Training set: this file contains the data of almost two decades of cable faults (location, severity, cable's burial status, water depth and so on) and has been used to train the algorithm with real-world data.

Cable RPL: The Route Position List of the cable together with the water depth associated with each point on the list and the armouring of the cable.

Severity: Each fault in the training set has been associated with a degree of severity, ranging from 0 (no fault) to 2 (cable cut or very severe fault).

Test set: a list of faults happened in latest year, which have not been included in the training set, and which have been used to test the validity of the forecasts predicted by the algorithm.

Random Forest has then been implemented in Python code with sci-kit learning tools and pandas.

3. BIAS AND LIMITATIONS

The method of work described in this paper is clearly biased: by using only data coming from actual repairs (which represent a subset of all faults), not all faults are used to train the algorithm, but only those severe enough to require a repair. We know for a fact that some minor loss events on a cable are not targeted for a repair as long as they remain within an acceptable power budget threshold. This means that a cable fault has occurred (even though a minor one) but we haven't accounted for it unless a repair is conducted. However, this source of bias can be almost entirely mitigated by using data coming from the operator's logs, extracted from the submarine line terminal equipment (SLTE). The benefit coming from this kind of approach is twofold: on one hand it allows to account for all types of cable "problems", on the other hand it gives us a much richer sets of data, well-correlated both in space and in time.

There is still another limitation. Sometimes the cause of the fault cannot be conclusively determined. This prevents the algorithm to attribute a statistical weight based on the cause of the damage. For example, this may lead to attribute the same weight both to a strongly-correlated event (for example damage due to fishing activity) and to a random, entirely uncorrelated one (earthquakes and other natural hazards) unless the two events can't be told apart. Unfortunately, there is no way to mitigate this limitation.

4. CONCLUSIONS

The scope of the algorithm presented is not to divine the future. It is aimed at building a reliable risk model which affects cables on a given area. These results can be used by cable operators to accurately assess the risks for their cables and to act upon them by taking the necessary countermeasures. It is mainly intended as a proof of concept, to demonstrate that machine-learning analysis on cable faults can be applied. The method presented is biased due to the very nature of the data used but this bias can be easily overcome by using real-time data coming from operators instead of public-domain data about cable repairs. These real-time data are easily collected and will allow much more precise forecast and analysis.

5. REFERENCES

[1] Kordahi ME, Stix R, Rapp RJ, Sheridan S, Lucas G, Wilson S, Perratt B, "Global trends in submarine cable systems faults", SubOptic 2016.
[2] Carter L, Gavey R, Talling PJ and Liu JT, "Insights into Submarine Geohazards from

Breaks in Subsea Telecommunications Cables", Oceanography 27(2):58-67, 2014.
[3] Kordahi ME, Shapiro S and Lucas G, "Trends in Submarine Cable System Faults", SubOptic 2007, May 2007, Baltimore, MD, USA.

