INCIPIENT FAULT PREDICTION IN POWER QUALITY MONITORING

Volker HOFFMANN
SINTEF AS – Norway
volker.hoffmann@sintef.no

Katarzyna MICHAŁOWSKA
SINTEF AS – Norway
katarzyna.michalowska@sintef.no

Christian ANDRESEN
SINTEF Energy Research AS – Norway
christian.andresen@sintef.no

Bendik Nybakk TORSÆTER
SINTEF Energy Research AS – Norway
bendik.torsater@sintef.no

ABSTRACT

European and global power grids are moving towards a Smart Grid architecture. Supporting this, advanced measurement equipment such as PQAs and PMUs are being deployed. These generate vast amounts of data upon which machine learning models capable of forecasting incipient faults can be built. We use live measurements from nine PQA nodes in the Norwegian grid to predict incipient interruptions, voltage dips, and earth faults. After training ensembles of gradient boosted decision trees on spectral decompositions of cycle-by-cycle voltage measurements, we evaluate their predictive performance. We find that interruptions are easiest to predict (95 % true positive, 20 % false positives). Earth faults and voltage dips are more challenging. Our models outperform naïve classifiers. We have explored forecast horizons of up to 40 seconds, but we have indications that forecast horizons of at least a few minutes are feasible.

INTRODUCTION

Changes in how electric energy is generated, consumed, distributed, and transmitted are beginning to challenge the reliability of the European power grid [1,2,3]. At the same time, advances in instrumentation, communications, and data analysis are driving novel solutions to address these challenges [2,4]. Real-time grid monitoring is unlocking a torrent of data driving development of new methods for fault detection, fault localisation, and self-healing [5,11].

Of particular interest in this contribution is the data becoming available through deployment of high-resolution monitoring instrumentation, such as *Power Quality Analysers* (PQAs) and *Phasor Measurement Units* (PMUs). These instruments continuously record voltage, phase, and power data as well as tag detected events (faults). When combining this data with machine learning (ML) methods, advances in detection, characterisation, and prediction of faults are becoming possible [3,5,6,12].

As of the end of 2018, existing work on predicting faults based on PQA data is sparse. Authors report successful use of clustering methods and multi-hidden Markov models to predict disturbances [6,7,8]. On short time horizons, LSTMs have also been successfully deployed to predict disturbances [9].

Continuing along this track, this contribution demonstrates the feasibility of predicting a range of incipient events from Norwegian PQA data.

The structure of the contribution is as follows. We begin by summarising the sources of data with some emphasis on how the PQA infrastructure and data processing and archiving pipelines look like. This is followed by a description of the actual data. We then describe the machine learning models and their evaluation metrics. Before concluding, we evaluate and compare our models with respect to classification performance.

DATA SOURCES

Our objective is to export PQA measurements that precede a fault to train a machine learning model to predict such faults on previously unseen data. Doing so requires a mature data collection, archival, and retrieval solution, which we now describe prior to characterising the data.

PQA Infrastructure & Data Pipeline

Norwegian grid companies must install PQAs in at least one critical node in their concession area. Most are installed in the high- (HV) and medium-voltage (MV) grid. They measure parameters relating to voltage quality, such as voltage level, frequency and harmonic distortion.

We analyse data from *Elspec* PQAs, which continuously sample voltage and current waveform at rates of up to 50 kS/s with data being compressed where appropriate. They can therefore efficiently store and reproduce continuous waveforms for long periods. *Elspec* devices collect many events and disturbances every year, and some nodes in the Norwegian grid have collected data continuously for over 15 years. A 10-year time-series from an Elspec node in a 22 kV MV grid is shown in Figure 1.
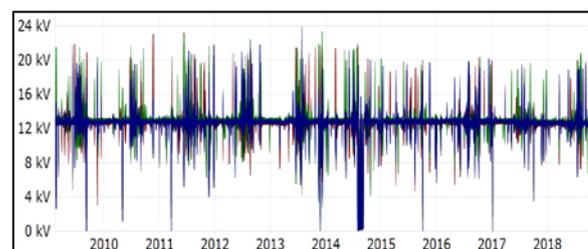

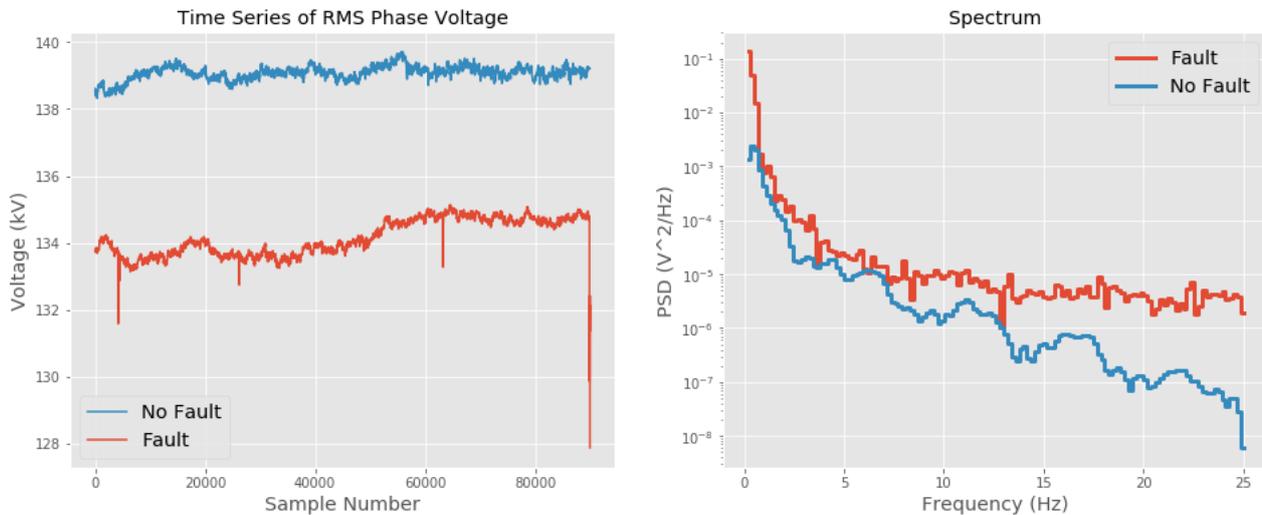**Figure 1**: 10 years of RMS phase voltage (22 kV MV grid).

Figure 2: Illustration of the feature extraction procedure. *Left*: Single-phase cycle-by-cycle RMS voltage for an observation without a recorded fault (blue) and an observation with a recorded fault (red, note the voltage drop at the right edge). *Right*: The corresponding (smoothened) Fourier spectrum in terms of the power spectral density. The two operating conditions have spectra that differ significantly above 10 Hz, which is what the classifier picks up on. Note that this is a simplified example.

The software *Automatisk hendelsesanalyse – AHA* (Eng. "Automatic event analysis") is used to generate lists of fault events and disturbances based on the measurement data from PQAs [10]. The tool sweeps time-series of power quality (PQ) data from selected nodes to detect different types of events. The tool can identify and classify interruptions, earth faults, voltage dips, and rapid voltage changes. These are annotated with event type, start time and end time for each event. This information is used to generate time-series for various voltage quality parameters, with a user-customised duration and resolution for each event.

Time-series used in this contribution have been generated automatically based on the lists of faults from AHA. In our pipeline, the user can select what kind of faults to load time-series for, as well as specify the number of time-series without faults.

Data Used In This Work

We exported 7616 time-series (*observations*) from nine PQA nodes in the Norwegian HV and MV grid. The selection covers voltage levels from 22 kV to 132 kV and each observation contains 1800 seconds (30 minutes) of data. Exports contain cycle-by-cycle RMS voltages for the three phase-to-ground and phase-to-phase voltages. Each observation contains 540'000 data points for a total of 40 GB of data.

There are 4101 observations without faults, 1944 with recorded voltage dips (< 90 % of nominal voltage), 133 with interruptions, and 1438 with earth faults. We exclude observations with problematic data (e.g., missing data on one or more phases) and retain 4101, 1940, 132, and 1433 observations (no faults, voltage dips, interruptions, and earth faults). Planned interruptions were not removed.

METHODS & MODELS

Classification problems in machine learning generally go through three steps. First, the features the classifier learns on must be selected. This is called feature extraction (or selection). Second, the selected algorithm is trained. Third, the predictive performance is evaluated.

We use an ensemble of *gradient boosted decision trees* (a variant of random forests) to predict whether a fault is incipient in each time-series. As input features, we use the smoothened power spectral density of the three phase-to-ground and phase-to-phase RMS voltages. The output that the model attempts to match is whether a fault occurred in a given time-series. Classification performance is assessed by comparing receiver operating characteristic (ROC) curves as well as the area under these curves (AUC). We use (stratified) five-fold cross-validation. We attempt to predict each fault class independently, i.e., we have three binary classification problems instead of a single multi-class problem. In each run, the data is balanced 50/50 between time-series with and without faults by undersampling the majority class.

Figure 2 illustrates the first step of the modelling, where we extract features from the time-series for the ML algorithm to train on. Features are the power spectral densities at different frequencies. The example compares the smoothened spectra of a time-series without a recorded fault to one with a sudden voltage dip at the end of the measurement period. While the differences in spectrum is principally what the classifiers will pick up on during training, we note that there are many different ways of calculating spectra. In particular, we must select (i) over which time interval, (ii) at which time-slice, and (iii) into
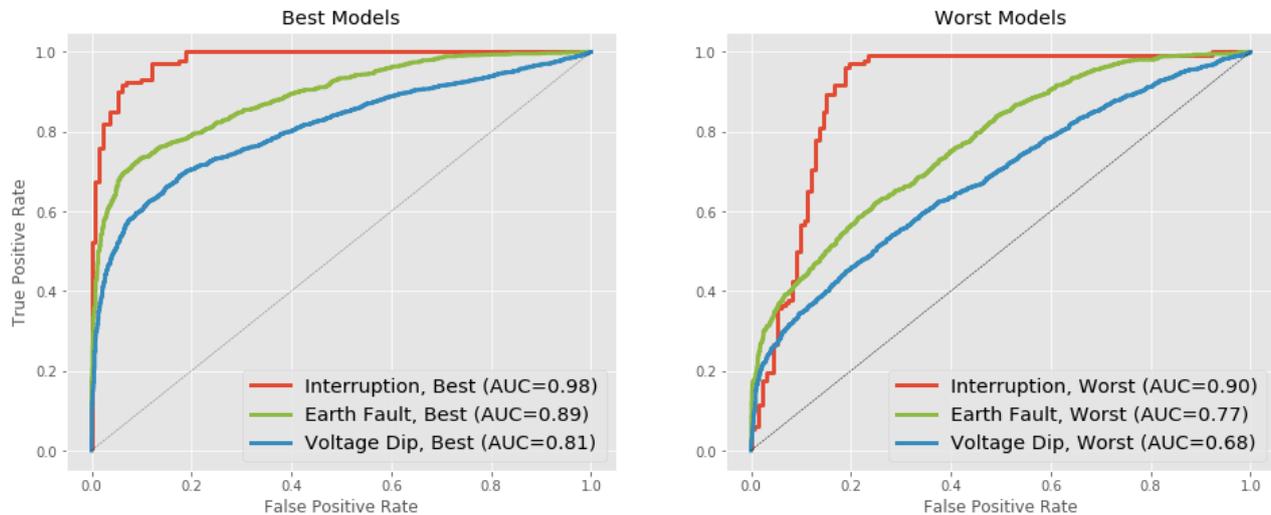
Figure 3: ROC curves for the best (*left*) and worst (*right*) performing models for each fault type. The diagonal line is a naïve classifier, which tosses a coin to determine whether a fault is incipient or not. The AUC for a naïve classifier is 0.5.

how many frequency components we process the time-series. The second parameter is particularly important, because it sets the forecasting horizon. There are also additional parameters that we have not explored here to limit scope (e.g., how to smooth the power spectral density). To understand model performance, we evaluate the classifier over a range of parameter choices, cf. Table 1. In total, we evaluate 264 combinations.

We use the *XGBoost* implementation of gradient boosted trees with default hyperparameters [13]. The smoothened spectrum is calculated using the Welch method [14], with a top hat window function and segment lengths twice the number of frequency bins.

During inference, the machine learning models accept the spectral decomposition of a time-series as input and return a probability of a given incipient fault. To translate a probability into a binary decision ("fault" or "no fault"), a threshold is set. Setting a high threshold only labels faults in which the classifier is confident (low false negative rate), but this comes at the cost of potentially missing a large number of faults (low true positive rate). Conversely, a low threshold will detect more faults (high true positive rate), but at the cost of classifying nominal conditions as faults (high false negative rate). By sliding the threshold, true and false positive rates can be traded off. To compare different models across all thresholds, we can calculate

true and false positive rates across thresholds. This results in a ROC curves, summarised by the AUC. Models with larger AUC have more skill and naïve (random) classifier has AUC of 0.5.

If we hold out some observations with known results, we can use these to evaluate classification performance. By varying which observations to hold out, we can further harden the classifier against overfitting. We vary the split five times for cross-validation. The split is stratified so that the class balance is maintained.

MODEL EVALUATION

Figure 3 shows the ROC curves for the best- and worst-performing models for each type of fault. We note:

1. Both best- and worst-performing models are most successful at predicting incipient interruptions. They are less capable of predicting earth faults and least capable of predicting voltage dips.

2. The best model can predict 95 % of interruptions at a false positive rate of 20 % at a threshold of 0.35 (to limit scope, we do not show rates across thresholds). The best earth fault prediction only reaches this level at a false positive rate of 70 % (threshold of 0.2). No model achieves this performance for voltage dips without avoiding false positive rates exceeding 95 %.

Table 1: Overview of parameters that have been varied for each model. In total, 11 * 6 * 4 = 264 models were evaluated.

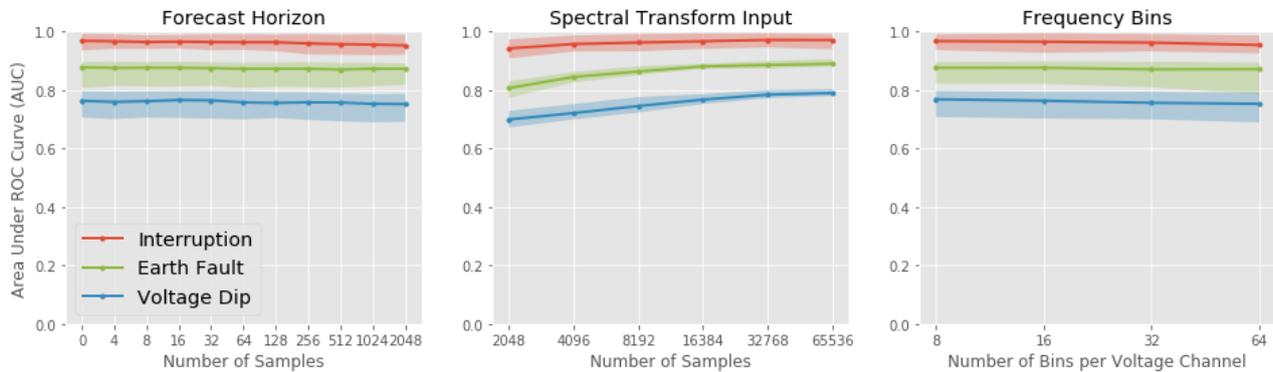| Parameter	In Units Of	Explored Values
Forecast Horizon	Number of Samples	0, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048
	Seconds	0, <1, <1, <1, <1, 1, 2.5, 5, 10, 20, 40
Spectral Transform Input	Number of Samples	2048, 4096, 8192, 16384, 32768, 65536
	Seconds	40, 80, 160, 320, 640, 1280
Frequency Bins	Number of Bins	8, 16, 32, 64

Figure 4: Model performance (expressed as AUC; the area under the ROC curve) depending on forecast horizon (*left*), length of the time-series going into the spectral transform (*middle*), and number of frequency bins (per voltage pair) used in the spectral decomposition (*right*). We distinguish between fault types (see legend). As we explore three model parameters in total, each panel marginalises (by indicating the 10 to 90 percentile range) over the two parameters.

3. Even the worst-performing model for interruptions predicts 95 % of interruptions at a false positive rate of 20 % (using threshold of 0.2). This is much better than the models for earth faults and voltage dips.

Overall, these first results are encouraging, especially for the prediction of full interruptions. While earth faults and voltage dips are more difficult to predict, even the worst models outperform a naïve coin toss. However, an operational deployment will want to balance true and false positives rates. From a model performance perspective, this can be done by selecting the threshold corresponding to the point of the ROC curve that is closest to the upper left corner. For the best models (and the worst model in the case of interruptions), this point is comparatively easy to locate. For the worst models, visual selection of this point is ambiguous. Trading off true and false positives is therefore more nuanced.

But how exactly do the models depend on the parameters of the feature extraction, i.e. the forecast horizon, the length of the signal for the spectral decomposition, and the number of frequency bins? Figure 4 shows model performance as each parameter is varied while marginalising over the other two. We find that:

1. Again, predictive performance for interruptions is best, performance for earth faults is intermediate, and voltage dips are the most difficult to predict.

2. The performance for each fault type is nearly constant as the forecast horizon increases. The relative difference of the median AUC between the shortest (0 s) and longest (40 s) forecast horizons is only 1.6 %, 0.5 %, and 1.5 % (service interruption, earth fault, and voltage dip, respectively). This suggests that the horizon can be increased beyond 40 seconds.

3. Increasing the number of samples used to calculate the spectrum increases the predictive performance. The improvement is largest for voltage dips (13 % relative

difference in median AUC between using fewest and most samples) and earth faults (10 % improvement), but small for outages (3 % improvement).

4. Increasing the number of frequency bins used as features weakly decreases performance. Comparing 8 to 64 bins, the median AUC decreases by 1.4 %, 0.5 %, and 2 % (interruptions, earth faults, and voltage dips, respectively).

5. When marginalising over the two variables not shown in each panel, the difference between the median and the 10/90 percentiles of the computed AUCs is usually between 5 % and 10 %. This suggests that the trends are robust and are only weakly complicated by higher order interactions.

The most important finding is that predictive performance does not decrease drastically over the explored forecast horizons. This suggests that extending the horizon to at least a few minutes is feasible. As this is a most critical requirement for operationalisation, we will explore this in future work.

The improved performance from increasing the number of samples is likely related to the increased frequency resolution during the spectrum calculation. While the improvement is welcome, it must be balanced by the (linearly) increased computational load. The performance decrease when using more frequency bins is likely caused by overfitting, which manifests as decreased out-of-sample performance (recall that we use six voltage pairs, so that 64 frequency bins means using 384 features).

CONCLUSION

We explored how well gradient boosted decisions trees can predict three different types of fault events (interruptions, voltage dips, and earth faults) based on cycle-by-cycle phase-to-phase and phase-to-ground voltages in the HV and MV grid. After describing our data pipeline, as well as

the method for generating user-specified datasets based on power quality data, we presented results obtained from a set of 7616 observations from nine measurement nodes in the Norwegian power grid.

While service interruptions are easiest to predict, earth faults and voltage dips are more challenging. The best-performing models can be tuned to successfully predict 95 % of faults at a false positive rate of 20 %. Depending on operational requirements, these metrics can be traded off against one another.

While we have only explored forecast horizons of up to 2048 samples (~40 seconds), results suggest that horizons on the order of at least a few minutes are possible. Along with more complex models and larger amounts of training data, we plan to explore this in future work.

ACKNOWLEDGEMENTS

REFERENCES

[1] IEA, 2017, "World Energy Outlook 2017"

[2] N. IqtiyaniIlham, M. Hasanuzzaman, M. Hosenuzzamana, 2016, "European smart grid prospects, policies, and challenges," *Renewable and Sustainable Energy Reviews,* vol. 67, 776—790.

[3] C. A. Andresen, B. N. Torsæter, H. Haugdal, K. Uhlen, 2018, "Fault Detection and Prediction in Smart Grids", *2018 IEEE 9th International Workshop on Applied Measurements for Power Systems (AMPS)*, Bologna.

[4] S. M. Muyeen and S. Rahman, 2017, "Control and Security Challenges for the Smart Grid", *Institution of Engineering and Technology.*

[5] P. Gopakumar, M. J. B. Reddy and D. K. Mohanta, 2015, "Transmission line fault detection and localisation methodology using PMU measurements," *IET Generation, Transmission & Distribution,* vol. 9, 1033-1042.

[6] Z. Yuxun, A. Reza, S. J. Costas, 2018, "Partial Knowledge Data-Driven Fvent Detection for Power Distribution Networks", *IEE Trans. Smart Grid,* vol. 9, 5152-5162

[7] K. Manivinnan, C. L. Benner, B. D. Russel, J. A. Wischkaemper, 2017, "Automatic identification, clustering and reporting of recurrent faults in electric distribution feeders," *19th International Conference on Intelligent System Application to Power Systems (ISAP)* , San Antonio.

[8] F. Xiao, Q. Ai, 2018, "Data-Driven Multi-Hidden Markov Model-Based Power Quality Disturbance Prediction That Incorporates Weather Conditions," *IEEE Trans. Power Syst.* vol. 34, 402—412.

[9] E. Balouji , I. Y.H. Gu, M. H. J. Bollen, A. Bagheri, M. Nazari, 2018, "LSTM-based deep learning method with application to voltage dip classification", *18th International Conference on Harmonics and Quality of Power (ICHQP)*, Ljubljana.

[10] H. Kirkeby, 2017, "Automatisk hendelsesanalyse," *Informasjonsteknologi og elektroteknikk - det digitale energiskiftet: NEF teknisk møte 2017*, Trondheim.

[11] B. D. Russell, R. M. Cheney, T. L. Anthony, C. L. Brenner, C. F. Wallis, W. E. Muston, 2009, "Reliability improvement of distribution feeders through real-time, intelligent monitoring", *Power & Energy Society General Meeting, 2009. PES '09. IEEE*, Calgary.

[12] A. S. Masoum, N. Hashemnia, A. Abu-Siada, M. A. S. Masoum, S. M. Islam, 2017, "Online transformer internal fault detection based on instantaneous voltage and current meassurements considering impact of harmonics", *IEEE Trans. Power Deliv.,* vol. 32, 587—598.

[13] T. Che, C. Guestrin, 2016, "XGBoost: A Scalable Tree Boosting System", *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* ACM, 785—794.

[14] P. Welch, 1967, "The use of the fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms", *IEEE Trans. Audio Electroacoust.* vol. 15, 70-73.