The value of multiple data sources in machine learning models for power system event prediction

Volker Hoffmann
Software & Service Innovation
SINTEF AS
Oslo, Norway
volker.hoffmann@sintef.no

Jonatan Ralf Axel Klemets
Dept. of Energy Systems
SINTEF Energy Research
Trondheim, Norway
jonatan.klemets@sintef.no

Bendik Nybakk Torsæter
Dept. of Energy Systems
SINTEF Energy Research
Trondheim, Norway
bendik.torsater@sintef.no

Gjert H. Rosenlund
Dept. of Energy Systems
SINTEF Energy Research
Trondheim, Norway
gjert.rosenlund@sintef.no

Christian A. Andresen
Dept. of Energy Systems
SINTEF Energy Research
Trondheim, Norway
christian.andresen@sintef.no

***Abstract*—We describe a method for assessing the value of additional data sources used in the prediction of unwanted events (voltage dips, earth faults) in the power system. Using this method, machine learning models for event prediction using (combinations of) different data sources are developed. The value of each data source is the improvement in model performance it brings. In addition, feature importance is retrieved using SHapley Additive exPlanations (SHAP). The methodology is applied to models that predict faults based on power quality and weather data. We find that models that combine sources outperform models using either in isolation. They predict ground faults and voltage dips with AUCs (Area Under Curve) of 0.74 and 0.80, respectively. Meteorological data appears more valuable than power quality data and the most important features are dew point, month of the year, and the power spectral density at 4.7 Hz.**

***Index Terms*—Machine Learning, Power system, Fault Prediction, Predictive Models, Multiple Data Sources**

I. Introduction

A. Motivation & Background

As society is being electrified, the reliability of the power system is becoming increasingly important. Power system interruptions and severe power quality (PQ) disturbances have a large impact and are harmful to all parts of society. While grid operators work hard to keep reliability high, events occur regularly due to phenomena such as component degradation, vegetation, and inclement weather. Presently, operators have limited possibilities to predict events and perform mitigating actions and better systems for event prediction are required.

Literature suggests that many unwanted events are caused by a combinations of rather than single factors [1, 2, 3]. This paper presents a methodology for combining data from multiple sources[1] and evaluate the added value quantitatively.

[1]By *data source*, we mean a measuring instrument or data forecasts. Each source (a power quality analyzer, a weather forecast) provides data for different *variables* (e.g. current, voltage, temperature, humidity).

B. Relevant Literature

Recently, fault prediction in power systems has risen in interest [3, 4, 5, 6]. Hitherto, most attempts are based on synthetic data [7], but ongoing roll-outs of advanced sensing equipment now enable modelling based on real data[8].

In [9, 10, 11], predictive models based on machine learning methods have been developed based on PQ data measured in the field. A data-driven prediction system for predicting line trip faults in power systems was proposed in [5], where long short-term memory (LSTM) together with support vector machine (SVM) was used to develop a prediction model. The prediction model was trained using historical field data from a substation in China, consisting of measurements of the voltages, current, and active power.

Weather data has also been used for fault prediction. For example, [12] explores the impact of weather seasonality of PQ disturbances and [13] predicts weather-related outages in the electrical grid. The latter uses three flavours of neural networks and meteorological data (wind-speed, air pressure, and temperature) and finds that all methods give comparable results. Weather observations occurring within a few hours of faults are most critical and earlier data does not improve predictions significantly. The authors do, however, emphasize the need for oversampling due to imbalanced data.

Most existing literature is focused on using either a single variable, or a combination of variables from a single source. In [5], models that combine measurements of different variables from the same source (voltage, current, active power) are compared to a model using only current. The combined model significantly outperformed the current-only model. In [4], authors show the benefits of combining parameters from two different data sources. Here, PQ and weather data were combined to predict PQ disturbances. However, the performance of a single data source was not evaluated.

978-1-7281-7660-4/21/$31.00 ©2021 IEEE

This is the accepted version of an article published in 2021 International Conference on Smart Energy Systems and Technologies - SEST
10.1109/SEST50973.2021.9543226

C. Contributions and Structure

With the increasing use of advanced instrumentation and retrofitted sensors, the amount of accessible data is ever-expanding. Data sources include operator-specific sources (e.g., SCADA, maintenance logs), external sources (e.g., weather and satellite data), and in-situ retrofits (e.g., wireless sensors directly attached to equipment). See also Table I for a non-exhaustive list with some references.

Integrating additional data sources almost always requires additional (and potentially brittle and poorly integrated) IT infrastructure. Simply adding all available sources is, therefore, (a) unlikely to be effective in terms of costs and benefits, (b) may not be necessary if most value is originating from a few sources, (c) lead to models that generalize poorly due to overfitting [20]. Therefore, it is important to have a pragmatic, robust, and uncomplicated approach for assessing the value added by new data sources.

The core contributions of this paper is thus two-fold:

1) A methodology to assess the value added by using several data *sources* during training of machine learning models as well as an objective metric for *value*.
2) An example utilizing field-data following this methodology to show how the fusion of power quality and meteorological improve the predictive capability of the models.

The paper has the following structure. Section II opens with an illustrative example of combining data sources. In section III, the proposed method for assessing the value of sources is presented. The results from applying this method to a number of sources are presented in section IV. In section V, the results and potential impact on future models are discussed.

II. Motivating Example

We begin with a motivating example (to which we return later). Fig. 1 shows distributions of events recorded at a single node in the Norwegian 22 kV distribution grid. One of the most basic analyses one can perform is to calculate the distribution of events over time – without taking into account even the event type. Doing this, we find (top panel) that events are not uniformly distributed over the year, but clump into the summer and winter months. If we now add information (admittedly from the same data source) about the type of event (middle panel), we discover that the two lumps actually correspond to different types of events. Suppose, we then add another data source (meteorological data) to the analysis (bottom panel). In that case, we discover the clumping of events in the summer and winter months may be driven by correlations between air temperature and event incidence. Clearly, there is validity in combining multiple data sources.

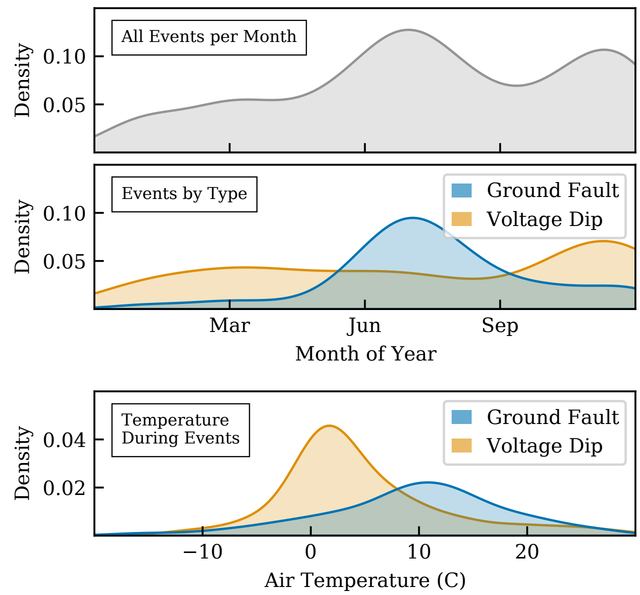

Fig. 1: Distribution of different events in the grid and their dependence on month of the year and air temperature. Distributions (i.e., the probability density functions; PDFs) are derived from kernel density estimates (KDEs) over the available samples. *Top*: Total distribution of all events over the 12 months of a year. The number of events peak in summer and winter. *Middle*: As the panel above, but grouped by event type. Ground faults peaks in the summer while voltage dips in the winter. *Bottom*: Distribution of air temperature during event occurrence, grouped by event type. Voltage dips cluster around the freezing point and ground faults occur at higher temperatures.

TABLE I: Overview of different types of data sources that could be used for predicting faults and other unwanted events in the grid. While the list is non-exhaustive, it is hoped that it does give a flavour of the diversity of available data.

Data Type	Data Source	Ref.
Power Quality	Power Quality Analyzers (PQA), Phasor Measurement Units (PMU), Fault Recorders	[14]
Weather	Weather Observations & Forecasts, Lightning	[15]
End-User	Advanced Metering Systems (AMS), Battery Manag. Sys. (BMS), Household Data	[16]
Grid Sensors	Transformer Temperatures, Line Sensors	[17]
Grid Topology	Grid Config., Power Flows, Load Distribution	[18]
Market	Price Data, Bidding Curves	[19]

Ultimately, we wish to develop machine learning models to predict events in the grid. The *value* of each data source is then the improvement of predictive performance it brings.

III. Methodology

We propose a pragmatic approach to assess the value added by combining data sources. Given two sources A and B, we (i) train and evaluate a machine learning model on dataset A, (ii) train and evaluate the same model (using identical hyper-parameters) on dataset B, (iii) train and evaluate the model on the intersection of dataset A and B, and (iv) ask the model to explain the features it is most sensitive to. The notion of *value* added by a data source is then the marginal improvement in the performance metric of the machine learning model.

We illustrate this approach by three datasets: (i) event logs, (ii) time-series of power quality measurements, and (iii) time-series of weather observations. We now describe data sources, pre-processing, models, evaluation scheme, and techniques used to explaination model predictions.

This is the accepted version of an article published in 2021 International Conference on Smart Energy Systems and Technologies - SEST
10.1109/SEST50973.2021.9543226

A. Data Sources

We use (a) event logs from a single site in the Norwegian 22 kV distribution grid, (b) sequences of phase-to-ground voltages, and (c) historical weather observations from the nearest observing site. Weather data is sourced through the Norwegian Meteorological Institute[2]. Event logs and voltage data are obtained from power quality instrumentation deployed in the grid [8]. We use data from 2008 to 2020. Event logs indicate the event type (earth fault, voltage dip, interruption), the phase(s) involved, and a timestamp.

Voltage sequences and weather conditions during the hours containing events are the positive samples for our models. The negative samples (”no event”) are voltage sequences and weather conditions at random times (between 2008 and 2020).

B. Features & Forecast Horizon

We use cycle-by-cycle RMS voltage sequences to calculate averaged spectra using the Welch method. Spectra are calculate on the sum of the three line voltage. The DC component is discarded and the (base-10) logarithm is calculated. We use 20 minute long sequences. For the positive (”event”) samples, sequences are cut-off 2.5 seconds before the actual event. The spectrum is calculated with 64 frequency bins and covers a spectral range $0 \leq f \leq 25$ Hz for a resolution of $\Delta f = 25/64 \sim 0.39$ Hz. The power spectral density at these frequencies is the first set of features.

For the weather conditions, we use 16 variables as features (cf. appendix A). Observations are made hourly, so we select weather conditions corresponding to the hour during which the event occurs (positive samples) or the voltage sequence begins (negative samples). Missing observations are forward-filled.

The final feature is the month. It is taken from timestamps of the event or the voltage sequence (positive/negative samples).

The forecast horizon is determined by the features. Models using only the month have an infinite horizon. Models using weather are limited by the forecast availability (usually 72 hours). Models using power quality data are limited to time on to the order of the cut-off (2.5 seconds).

C. Pre-Processing, Modelling, & Tuning

The XGBOOST implementation of gradient boosted decision trees [21] is used to (initially) train three types of models to predict the probability of an event. First, models are trained only on voltage sequences. Second, models are trained only on meteorological data. Third, models are trained on voltage sequences and meteorological data. Main hyperparameters for XGBOOST were a maximum tree depth of three, a binary logistic objective function, the logloss evaluation metric, and 64 learners. Other parameters remained at their default values.

Taken together, pre-processing and training introduces dozens of parameters that can be tuned for predictive performance (e.g., voltage sequence length, model hyperparameter, additional weather observations). However, our objective is to demonstrate the value of data sources. As such, no parameters were tuned and we use what we consider sane defaults.

[2]https://frost.met.no/index.html

D. Model Evaluation & Balancing

Model performance is evaluated using three-fold cross-validation.[3] Datasets are artificially balanced so that curves of receiver operating characteristics (ROC) and their summary metric (the area under the curve; AUC) can be used to compare performance. These metrics compare model performance across decision thresholds.

Statistically, the incidence of ground faults and voltage dips is roughly one every 405 and 286 hours (17 and 12 days), respectively. As such, the datasets are heavily imbalanced and aggressive balancing during model development can result in models with limited real-world applicability. However, proper treatment of imbalanced datasets requires metrics that are more complicated to reason about. To not detract from the core point (the relative value of different sources), we have chosen to present balanced datasets and use ROC and AUC.[4]

E. Feature Importance

Once trained, machine learning models can be asked to explain how they arrived at a particular prediction. We use SHapley Additive exPlanations (SHAP, [22]) to quantify the contribution of each feature value to each final prediction. For a single prediction, the model can – for example – explain how much a particular value of wind speed pushed the model to predict an above average risk for a voltage drop. By aggregating over features and predictions, the model can also explain the overall contribution of features. In other words, we can find out (a) what features influence the final predictions the most, and (b) in which way feature values push the final prediction. For example, are higher or lower dew points correlated with the increased risk of earth faults?

IV. Results

We now pick up the exploratory analysis from Section II, describe the performance of models using weather and voltage sequences, show how to better control for seasonality, and close with analyzing the most important features.

A. Exploratory Analysis

In the period of 2008 to 2020, a total of 2117 events occur at the site under consideration – 1195 ground faults, 913 voltage dips, and 9 service interruptions. Most (71 percent) events are strongly clustered in time [23], so we reject all duplicates that occur within the same hour. This leaves 614 events – 253 ground faults, 358 voltage dips, and 3 service interruptions. We ignore service interruptions to avoid small number statistics.

Fig. 1 shows distributions of events with respect to time (top and middle panel) and air temperature (bottom). We find that events occur most frequently in either the summer or winter and that ground faults dominate in the summer while voltage

[3]During initial testing, three-fold cross-validation was selected as trade-off between the high-variance and computational intensity of Leave-One-Out schemes and the biasing dangers of a simple training/validation split.

[4]We have tested and reproduced our core results (the ordering of the value of our sources) on a moderately imbalanced (3.5x) dataset using more appropriate metrics (precision, recall, and average precision).

This is the accepted version of an article published in 2021 International Conference on Smart Energy Systems and Technologies - SEST
10.1109/SEST50973.2021.9543226

dips dominate in the winter. The incidence of ground faults in the summer also increases more than the incidence of voltage dips does in the winter. Their concentration (how far the peak of the distribution is above the baseline) also differs.

These patterns suggest correlations with weather conditions, so we expect that attempts at modelling (predicting) events is improved by adding meteorological data. However, there are many meteorological variables and – during initial data exploration – it remains unclear which help most during modelling. Interactions with the most obvious variable – air temperature – was used as an initial assumption. Indeed, Fig. 1 reveals that air temperature during voltage dips and ground faults follow different distributions.

This analysis suggests that meteorological variables can improve the predictive capability of models. However, it cannot answer whether voltage dips are caused by cold weather or just happen to occur more often in the winter. This is an example of *correlation does not imply causation*.

B. Model Performance

Nevertheless, correlations can be exploited without a deep understanding of the underlying causality if they help in improving the ultimate bottom-line – model performance. If they do, there is a value in adding meteorological parameters, even if they serve only as proxy variables. We now illustrative this by way of model performance.

Fig. 2 and Table II show the predictive performance of the models trained to predict ground faults. Models using only power quality data tend to perform poorly with mean AUC being ~ 0.57 during three-fold cross-validation. Models using only meteorological data perform much better with mean AUC ~ 0.70. Models that combine voltage and meteorological data perform best (mean AUC ~ 0.71). Additionally, models using only meteorological data and those using both meteorological and voltage data perform almost equal at low and high decision thresholds. At intermediate decision thresholds (0.25 to 0.75), models combining sources outperform single-source models.

We have also trained models to predict voltage dips and find identical trends (not shown). In line with [24], models asked to predict voltage dips perform better than those used to predict ground faults. Their baseline performance has a mean AUC of 0.71 (voltage data) and 0.79 (meteorological data). Combing both data sources increases the mean AUC to ~ 0.80. Trends across the ROC curve are identical to ground fault prediction – performance improves at intermediate thresholds.

C. Controlling for Seasonality

Thus far, we have relied on seasonality being implicitly encoded in the meteorological variables. However, using 16 variables to encode seasonality is both heavy-handed and potentially obscures more direct relationships between events and weather. To better account for seasonality, we add a new feature – the month of the year. This makes for (i) one new model using only the month to predict events, and (ii) modifications to the existing models to include the month, cf. Table II.

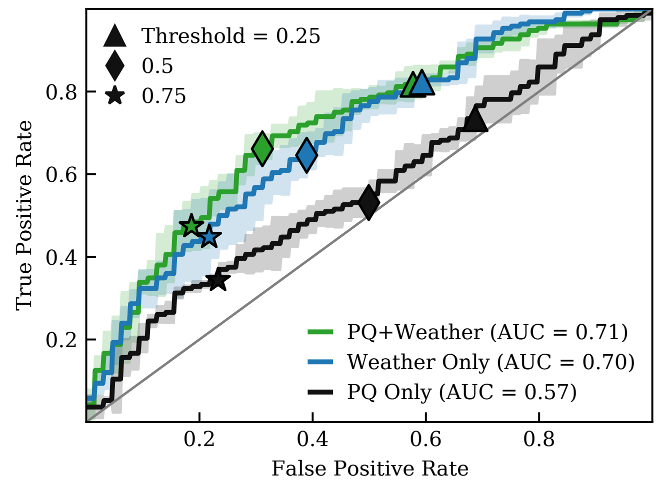

Fig. 2: Receiver operating characteristic (ROC) curves for three machine learning models. One model uses only power quality data (black), one only meteorological data (blue), and one power quality and weather data (green). The diagonal is the performance of a naïve model (coin toss). Models were evaluated using three-fold cross-validation. We show the mean (lines) and the mean $\pm$ standard deviation (shaded) across folds. Additionally, three decision thresholds are indicated (per model). The model combining weather and voltage data outperforms both others at intermediate thresholds and performs similar to the model using only weather at low and high thresholds.

We make four major observations. Firstly, the model using only the month performs 25 per cent better than the model using only power quality. It also more stable (AUC varies less across folds). It performs approximately equal as the model using weather data (their difference is smaller than AUC variation), but is more stable (AUC varies by 0.03 vs 0.01). Second, adding the month to the model using weather data (minimally) decreases the average performance (mean AUC decreases from 0.71 to 0.70) while at the same time stabilizing it. Third, combining the month with power quality data does not increase model performance as much as adding meteorological data (18 vs 25 per cent difference). However, models are more stable (AUC varies by 0.01 vs 0.02). Fourth, adding the month to a model using power quality and meteorological data improves performance as well as stability.

TABLE II: Model performance for different feature sets. We show mean and standard deviation of the AUC across the folds. Models tagged with * include the month of the year.

Feature Set	AUC (Mean $\pm$ Std)
Only Power Quality Data	0.57 ± 0.03
Only Weather Data	0.70 ± 0.03
Only Month (of the Year)*	0.71 ± 0.01
Weather Data, Month*	0.70 ± 0.03
Power Quality Data, Month*	0.67 ± 0.01
Power Quality Data, Weather Data	0.71 ± 0.02
Power Quality Data, Weather Data, Month*	0.74 ± 0.01

This is the accepted version of an article published in 2021 International Conference on Smart Energy Systems and Technologies - SEST
10.1109/SEST50973.2021.9543226

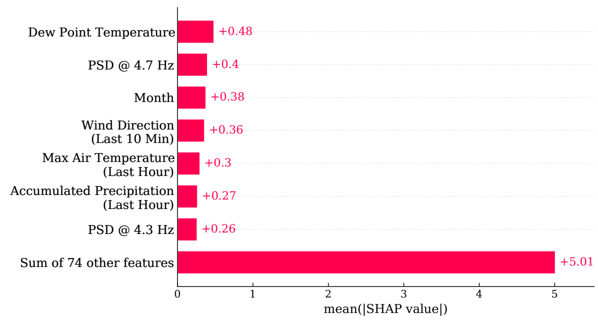

Fig. 3: The seven most important features in terms of their mean absolute contribution (the SHAP value) to all predictions. Features are ranked top to bottom. The remaining 74 features account for – on average – a contribution of $5.01/74 \sim 0.07$. Meteorological features are self-explanatory, month is the month of the year, and PSD the power spectral density at a particular frequency.

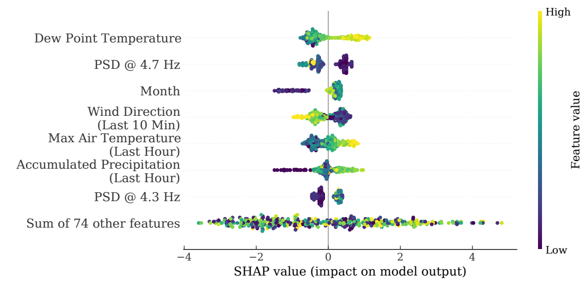

Fig. 4: Beeswarm plot of the seven most important features. Each dot corresponds to a single prediction and shows how the feature influenced the final prediction. Negative SHAP values mean that the feature value pushed the model towards predicting a lower event probability (and vice versa). Colours indicate the value of the feature during the prediction (relative to its average). Dots clump where there are many predictions. Meteorological features are self-explanatory, month is the month of the year, and PSD the power spectral density.

D. Feature Importance

Having understood differences in model performance, we now try to identify the features most important for predictions. Fig. 3 shows the seven features with the highest mean absolute contribution towards the final prediction (as well as the total contribution of the remaining features) for the ground fault predicting model. The three most important features are from unique feature sets each – the dew point temperature during the last hour, the power spectral density at 4.7 Hz, and the month of the year. This backs up our findings that models combining data sources outperform models that do not.

Additionally, the importance between the most and second most relevant features are within 20 per cent of one another (even the fifth is within 60 per cent). This suggests that no single feature dominates over all others, implying that the relations between features and events are nuanced.

By plotting the relative contribution of the value of each feature for every prediction the model makes, we can gain insight into these nuances (Fig. 4). We find that (a) for most predictions, the top seven features do not have SHAP values exceed ± 1, (b) of the 74 remaining features, there are predictions with SHAP values in excess ± 2, although (c) most are concentrated between -2 and 2. Further, (d) most predictions have low dew point temperatures (which push models away from predicting an event), (e) when dew point temperatures are above average, they almost always increase the probability of event occurring. We further find that (f) wind direction has a consistent effect (westerly winds correlate with events, easterly winds do not). Additionally, (g) the power spectral density at 4.7 Hz is distributed into two clumps of nearly identical colour. This suggests that the value of the feature itself is not a strong driver of the prediction, but that it interacts with other features. Finally, (h) in most predictions, the month increases event probability irrespective of its numerical value. This indicates interaction with other features – suggesting that the month can indeed help models disentangle seasonal variations from more direct relationship between events and weather conditions.

V. Discussion

Having shown how multiple data sources combine to produce superior models, we now calculate their value. Recall that the value gain is our metric of choice – AUC. To produce a robust ranking, we begin with a naïve model, proceed to the worst single source model, and then add the worst of the remaining possible source combinations, until we arrive at the model combining all sources. In our case, we begin with (i) the coin toss (AUC = 0.5), followed by (ii) the worst single source model (power quality, AUC $\sim$ 0.57), (iii) the worst (remaining) two-source model (power quality, month, AUC ~ 0.67), and (iv) the three-source model (power quality, month, weather, AUC ~ 0.74). This means that power quality, weather, and the metadata of the event logs (month) have a value of $0.57 - 0.5 = 0.07$, $0.67 - 0.57 = 0.1$, and $0.74 - 0.67 = 0.07$, respectively. Note that stabilizing effect of month is not accounted for in the quantitative ranking.

The large value of weather data is reflected in the feature relevance – top features are dominated by meteorological variables. If the most important features were, for example, related to voltages, there would be no gain in the addition of weather data. That said, even if meteorological features have the largest impact on the predictions, this does not mean that they are the underlying cause of the events. Instead, they may simply correlate with the underlying cause and act as a proxy variable. The month of the year appears to control for this as models without (not shown) rely on the maximum hourly temperature, which correlates with the month.

If we repeat the analysis to predict voltage drops (not shown), we find value of power quality, the month of the year, and the weather data to be 0.08, 0.14, and 0.2. In contrast to earth fault models, using only weather data performs much better than using only the month (AUC ~ 0.79 vs 0.64). This suggest that voltage drops depend more strongly on weather conditions than earth faults. In fact, this is what Fig. 1 suggests – voltage dips are spread out more over the year, but also lump more strongly around the freezing point.

This is the accepted version of an article published in 2021 International Conference on Smart Energy Systems and Technologies - SEST
10.1109/SEST50973.2021.9543226

VI. Conclusion and Further work

We have presented a methodology to quantify the value added when combining multiple (independent) data sources for developing machine learning models that predict unwanted events in power systems. The notion of *value* was linked to the marginal improvement in model performance when adding a data source. We also show how feature importance can be used to assess which parameters that drive model performance.

We applied the methodology to the prediction of ground faults and voltage dips based on event logs, power quality data, and meteorological data. We find that meteorological data is more valuable the power quality data and that much of that value is derived by its implicit seasonality. If we control for this seasonality (by adding an explicitly seasonal variable; the month of the year), models perform even better – likely because it helps the models to focus on non-seasonal correlations between weather conditions and events. By applying the same methodology to voltage drops and ground faults, we find that ground faults tend to correlate more with (unknown) seasonal effects whereas voltage drops correlate more with certain weather conditions.

In the future, a systematic exploration of various data sources (cf. Table I) may result in models capable of forecasting events with performance sufficient for operational deployment. Beyond roles in decision support, models and their explanatory power may also be applied to analyse root causes (or at least correlations) of past, ongoing, or even future failures (in a simulation scenario). We expect the most powerful models to result from combining sources that sample different physical processes.

Acknowledgment

The authors gratefully acknowledge the Research Council of Norway and industry partners for support (project 268193/E20 EarlyWarn), and also the industrial partners for contributing with knowledge and experience. The authors would also like to thank Torfinn Tyvold (NTNU) for supplying event lists.

References

[1] M. Doostan and B. H. Chowdhury, "Power distribution system fault cause analysis by using association rule mining," *Electric Power Systems Research*, vol. 152, pp. 140–147, 2017.

[2] Statnett, "Yearly fault statistics." https://www.statnett.no/for-aktorer-i-kraftbransjen/systemansvaret/arsrapporter-fra-feilanalyse/a. On-line, accessed 30- Mar- 2021.

[3] J. Hare, X. Shi, S. Gupta, and A. Bazzi, "Fault diagnostics in smart micro-grids: A survey," *Renewable and Sustainable Energy Reviews*, vol. 60, pp. 1114–1124, 2016.

[4] F. Xiao and Q. Ai, "Data-Driven Multi-Hidden Markov Model-Based Power Quality Disturbance Prediction That Incorporates Weather Conditions," *IEEE Transactions on Power Systems*, vol. 34, pp. 402–412, 2018.

[5] S. Zhang, Y. Wang, M. Liu, and Z. Bao, "Data-Based Line Trip Fault Prediction in Power Systems Using LSTM Networks and SVM," *IEEE Access*, vol. 6, pp. 7675–7686, 2018.

[6] M. Viswavandya, S. Patel, and K. Sahoo, "Analysis and comparison of machine learning approaches for transmission line fault prediction in power systems," *Journal of Research in Engineering and Applied Sciences*, vol. 06, pp. 24–31, 2021.

[7] T. Weckesser, H. Jóhannsson, and T. V. Cutsem, "Early prediction of transient voltage sags caused by rotor swings," in *IEEE PES General Meeting Conference, National Harbor, MD, USA*, 2014.

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

[9] K. W. Høiem, V. Santi, B. N. Torsæter, H. Langseth, A. C. A., and G. H. Rosenlund, "Comparative Study of Event Prediction in Power Grids using Supervised Machine Learning Methods," in *in 2020 International Conference on Smart Energy Systems and Technologies (SEST), Istanbul, Turkey*, 2020.

[10] G. H. Rosenlund, K. W. Høiem, B. N. Torsæter, and C. A. Andresen, "Clustering and Dimensionality-reduction Techniques Applied on Power Quality Measurement Data," in *in 2020 International Conference on Smart Energy Systems and Technologies (SEST), Istanbul, Turkey*, 2020.

[11] G. W. Chang, Y.-H. Hong, and G.-Y. Li, "Stochastic prediction of voltage sags in a large transmission system," *IEEE Transactions on Power Delivery*, vol. 34, pp. 1785–1794, 2019.

[12] K. Michałowska, V. Hoffmann, and C. Andresen, "Impact of seasonal weather on forecasting of power quality disturbances in distribution grids," in *2020 International Conference on Smart Energy Systems and Technologies (SEST), Istanbul, Turkey*, 2020.

[13] C. Brester, H. Niska, R. Ciszek, and M. Kolehmainen, "Weather-based Fault Prediction in Electricity Networks with Artificial Neural Networks," in *2020 IEEE Congress on Evolutionary Computation (CEC)*, IEEE, 7 2020.

[14] ENTSO-E, "Digital fault recorders." https://new.siemens.com/global/en/products/energy/energy-automation-and-smart-grid/protection-relays-and-control/general-protection/phasor-measurement-unit-pmu.htm. On-line, accessed 30- Mar- 2021.

[15] WORLD METEOROLOGICAL ORGANIZATION, "World weather information service." https://worldweather.wmo.int/en/home.html. On-line, accessed 30- Mar- 2021.

[16] Y. Yang, Q. Ye, L. J. Tung, M. Greenleaf, and H. Li, "Integrated size and energy management design of battery storage to enhance grid integration of large-scale pv power plants," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 1, pp. 394–402, 2018.

[17] W. D. A. G. Hillary, K. L. I. M. P. B. Jayarathna, L. I. Ranasinghe, S. M. B. P. Samarakoon, N. M. T. N. Rathnayake, J. R. Lucas, and R. Samarasinghe, "A tool for estimating remaining life time of a power transformer," in *2017 Moratuwa Engineering Research Conference (MERCon)*, pp. 373–378, 2017.

[18] ENTSO-E, "Entso-e transparency platform." https://transparency.entsoe.eu/. On-line, accessed 30-Mar-2021.

[19] Nordpool , "Easy-to-integrate power market data." https://www.nordpoolgroup.com/services/power-market-data-services/. On-line, accessed 30-Mar-2021.

[20] N. Dessì and B. Pes, "Similarity of feature selection methods: An empirical study across data intensive classification tasks," *Expert Systems with Applications*, vol. 42, no. 10, pp. 4632–4642, 2015.

[21] T. Chen and C. Guestrin, "Xgboost," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug 2016.

[22] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), vol. 30, Curran Associates, Inc., 2017.

[23] T. S. Tyvold, B. N. Torsæter, A. C. A., and V. Hoffmann, "Impact of the Temporal Distribution of Faults on Prediction of Voltage Anomalies in the Power Grid," in *in 2020 International Conference on Smart Energy Systems and Technologies (SEST), Istanbul, Turkey*, 2020.

[24] V. Hoffmann, K. Michałowska, B. Torsæter, and C. Andresen, "Incipient Fault Prediction in Power Quality Monitoring," in *CIRED 2019 Conference, Madrid, Spain*, 2019.

Appendix A: Meteorological Variables

Models used the following meteorological measurements: (a) dew point (measured on the hour), (b) relative humidity (on the hour), (c) accumulated precipitation (last hour), (d,e,f) air temperature (on the hour, min/max last hour), (g-h) wind speed (max last hour, mean last 10 minutes before the hour), (i) wind direction (mean last 10 minutes before the hour), (j) wind speed of gust (max last hour), (k) wind direction of gust (max last hour), (l) duration of precipitation, (m) surface snow thickness, (n-o) time of max wind speed and wind gust (last hour). See https://frost.met.no/elementtable for details.

This is the accepted version of an article published in 2021 International Conference on Smart Energy Systems and Technologies - SEST
10.1109/SEST50973.2021.9543226