Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Chair of Intelligent Maintenance Systems

CH-8093 Zürich

Prof. Dr. Olga Fink
Stefano-Franscini-Platz 5

8093 Zürich
Schweiz
ofink@ethz.ch

Eidgenössisches Departement für Umwelt,
Verkehr, Energie und Kommunikation UVEK
Bundesamt für Verkehr BAV
Abteilung Politik
Sektion Direktionsgeschäfte
Mühlestrasse 6, 3063 Ittigen,

Zürich, 18.1.2022

Progress Report «Project INITIATE»

1 Introduction

This report provides an overview of the progress made in project INITIATE. This project is a research project funded by the fonds for railway infrastructure "Bahninfrastrukturfonds" BIF of 2014. In this project, ETH and SBB work together to improve the detection and prediction of critical conditions in railway power network systems by developing intelligent algorithms for control system data.

Figure 1: Power grids sites from which the generated data of case study I and II originates.

In June 2021, a group of domain experts from different fields collected and rated various case studies as candidates for investigation. The domain experts estimated the highest overall benefits of all case studies as well as immediate benefits if selected as the first case study. Based on these scores, two studies were selected to be the most promising and, thus, were processed first: a case study on state estimation and fault detection from power grid sensors in local areas such as power plants with connected substations (**case 1**) and a case study on hydro-power plant efficiency estimation from pressure and flow sensors (**case 2**). Two workgroups were formed to study the different cases, led by Philipp Wenk from SBB's side and Olga Fink from ETH's side:

- Project Lead: Philipp Wenk and (SBB) Olga Fink (ETH)
    - Study case 1: state estimation and fault detection from power grid sensors
        - Robert Strietzel (SBB)
        - Raffael Theiler (ETH)
    - Study case 2: hydro-power plant efficiency estimation
        - Nima Riahi (SBB)
        - Mengjie Zhao (ETH)

Figure 1 shows that both studies are limited to data from one location/site. Case 1 focuses on EMS (energy management system) data from *Etzelwerk* and substations, case 2 on data from *Vernayaz*. The scope was selected according to the proposal: WP1/Year1: Development of physics-guided neural networks for selected grid components for early fault detection and detection of systematic faults, demonstrated on at least one hydropower plant.

2 Case Study 1: State Estimation and Fault Detection from Power Grid Sensors

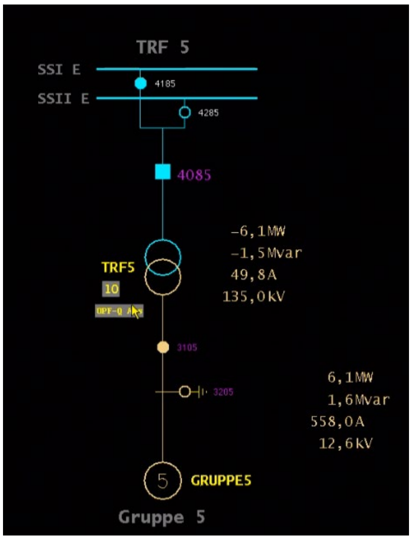

Figure 2 Generator group 5 of Etzelwerk in EMS with online sensor data. PowerFactory is based on a similar schematic overview.

The purpose of this study is to detect, understand and analyze faults of different severities that affect the power grid. For that, we use prior knowledge such as the grid schematics, electrical simulations and analyze time series from power grid sensors. Additionally, we use the EMS event protocol that lists state changes of components such as the electrical switches or the voltage regulation. A separate incident log of manual intervention was extracted on site. However, this dataset turned out to be sparse (~4 events for 6 months) and, therefore, not suitable for learning-based approaches. For this case study, all data originates from *Etzelwerk* and connected substations.

The key idea of this case study is that if we had a trustworthy and robust state estimator, we could reliably detect faults and anomalies by analyzing the difference of the current power plant state to the predicted state. With this study, we aim to respond to the shortcomings of the current situation. On one hand, the simulations currently have deficits. The offline PowerFactory simulation is not accurate for all regions of the power grid, and it makes no use of EMS sensor data yet, which is available in the online EMS-integrated simulations (see figure 2), but those simulations are crude. On the other hand, the information that is collected with the recorded EMS sensor data is not used to its full potential because the quality of the sensor data is typically difficult to validate due to the accessibility of the sensors which poses a major challenge and limitation. Currently, there is only a very limited systematic sensor validation process integrated in EMS. We hypothesize that the combined signals can be used for an improved self-validation as well as to refine a power grid state estimation if we find an appropriate validation strategy.

These hypotheses allow to formulate a goal. With this case study, we aim to find a data driven model that learns from simulation, prior knowledge and data such that the computed result better reflects the true state of the power grid. Additionally, the learned model should make a statement about how trustworthy a reading from either a sensor or the simulation is, given the neighboring sensors. More precisely, we pursue objective 2 of the proposal, that is to increase the overall data quality by detecting sensor misalignment, drift or biases in the signals. Similar algorithms could be employed to detect wrong parameters in the model.

The first step of this case study was to set up all the technical components. ETH was introduced to the detailed PowerFactory simulation that is available at SBB. We exported a subset of the simulation, containing *Etzelwerk* and the surrounding substations for offline usage at ETH and resolved all license-related issues with PowerFactory. To collect data, in a first step, connectivity to EMS was established. A refined EMS export script allows now to extract time series for the Python data preprocessing pipeline which cleans up irregularities and aligns the time series of currently ~100 sensors (usually measuring P, Q, I, V).

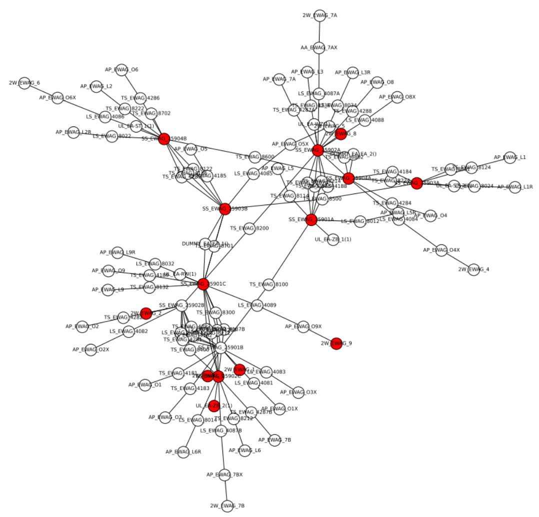

Figure 3 Schemata graph of all generator groups of Etzelwerk and the surrounding substations extracted from PowerFactory to Python. Nodes with attached data sources from sensors are highlighted in red.

To apply PowerFactory more effectively, we aimed to learn from a different project at SBB, which also makes use of PowerFactory in a similar fashion. We could modify and reuse parts of the code base of that project for our purpose. Using the provided library, we were able to achieve our first objectives:

- To simulate a state estimation in PowerFactory with data from the EMS system conceptually. This is called a quasi-dynamic simulation.
- To extract schemata and metadata as well as the state of components during simulation from PowerFactory as a prerequisite for our data driven model.

A promising class of machine learning models for this study is graph neural networks (GNNs). GNNs are neural models that capture the dependencies in graphs via message passing between the nodes. They are an excellent choice to process data from power grids where the nodes are components such as transformers or generators and the bus bars or transmission lines represent the edges. Figure 3 shows such a graph of all electrical components connected to *Etzelwerk* that we plan to use in a GNN. The graph was extracted from the schematical drawings in PowerFactory by iterating over the components and the red nodes (devices) are linked to the EMS data sources.

In preliminary experiments, we noticed that the simulation speed of PowerFactory runs at best in real time which would lead to months of processing EMS data in the simulator for state estimation. Hence, we are motivated to replace the simulator with a GNN.

3 Case Study 2: Hydro Plant Efficiency Estimation

3.1 The aim of the study

Around 60 percent of Switzerland's domestic electricity production comes from hydropower. Understanding and improving the efficiency of hydropower plants is essential for Switzerland's energy policy in order to guarantee a secure, economical and ecological energy supply. An accurate efficiency estimation enables

- highly efficient and reliable power generation as well as optimized operation, scheduling of machines with nozzle control optimization
- early detection of failures and optimized maintenance revision with continuous monitoring

This case study aims to provide a better efficiency estimation by answering the following questions:

- Which factors affect hydropower efficiency and are feasible to be considered in this case study?
- How precise are the sensor measurements and what are their influences on efficiency estimation?
- How to ensure a reproducible efficiency estimation given the same operational conditions?
- How do generating units influence each other in terms of efficiency when multiple units are operating?

3.2 The description of facilities

This case study is conducted as a pilot case for the *Vernayaz* hydro power plant owned and operated by SBB. The schema of the hydro plant is shown in Figure 4 and an overview of the operational specifications is given in Table 1. The power plant, which was commissioned in 1928, produces 240 GWh annually for the Swiss Rail Network and is equipped with three machines (G1, G2, G3), each consisting of a dual-jet Pelton turbine with a horizontal shaft and a synchronous single-phase generator. Details of different operating units are listed in Table 2.

Table 1. Overview of Vernayaz hydro plant operating statistics.

Maximum waterfall height	646.25 m
Nominal flow	17 m3/s
Nominal rotation speed	501 rpm
Utility frequency	16.7 Hz
Total apparent power	107 MW

Table 2. Vernayaz hydro plant turbine unit details.

	G1	G2/G3
Maximum power	21 MW	37.4 MW
Nominal flow	3.94 m3/s	6.7 m3/s
Average wheel diameter	2.02 m	2.00 m
Apparent power	27 MW	40 MW
Commission year	1974	1989

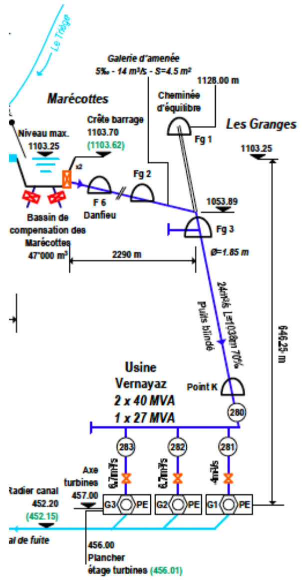

Figure 4 Hydro schemata of hydro plant Vernayaz

3.3 Statistical Analysis and Methodology

The statistical analysis is carried out based on the EMS operating data from 2020 and 2021 and cross-compared with the efficiency study from 2016 by SBB. These data consist of generated power, flow rates, pressure, water level, speed of rotation, all recorded at 1-second sampling intervals.

Hydro efficiency is defined as the ratio of the total electrical power produced to the theoretical hydraulic power of the water passing through the turbine. The latter is determined by the hydraulic head (proportional to the net water fall height, determined from the turbine-level pressure) and the volumetric flowrate.

The measurement accuracy of the above-mentioned sensor values plays a key role in the efficiency assessment. Based on the previous analysis, the hydraulic measurements showed unreliable signals. This case study further confirms that there is a delay of 1-8 seconds between the electrical sensor measurements and the flowrate measurements. With the sensor calibration, a more accurate efficiency estimate is obtained. The current study shows that the net hydraulic head is 1-2% higher than the gross hydraulic head, which implies that there is a systematic measurement error in pressure and/or water head. SBB hydraulic experts have been informed about this and are actively searching for an explanation.

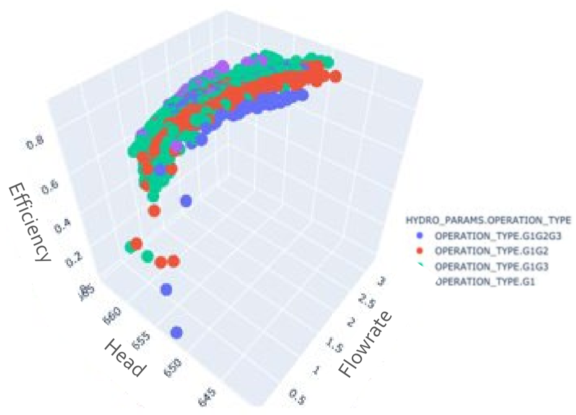

Figure 5 Efficiency assessment of G1 under different operating conditions

The hydro efficiency is assessed by analyzing the operating data for each group for the entire available time period. Based on the expert experience at SBB, a steady-state filter is required to filter out the oscillations in the signals in order to avoid transient phenomena that would invalidate certain assumptions in the formula. In the current stage of the study, a simple hourly averaging filter is applied, which will be replaced by a suitable steady-state filter in the future.

The current case study shows that the efficiency of an individual operating unit can be strongly influenced by the interaction with other operating units. As shown in Figure 5, the efficiency of a single operating unit differs from that under the same operating condition with multiple operating units. This corroborates other findings in the literature. In the future, further investigations will be conducted to better model and understand the interactions between different operating units. A promising approach for this purpose is Graph Neural Networks.

In addition to the operation of other units, the number of active nozzles also affects efficiency. This information is not included in the EMS and needs to be extracted from the regularization strategy. Analyzing the effect of the number of active nozzles and the interaction between operating units will be the focus of the next step.

4 Outlook

In the past five months, we gained essential insights from the data of case study one and two. Here, we provide a general outlook for the studies based on those initial insights. Currently, we are on track for the completion of the foreseen yearly work packages as presented in the proposal. While those work packages are the main objective, it is still a possibility to investigate other case studies that are currently available as ideas, to derive new case studies from the current insights or to scale successful algorithms to different power plants and grids.

Case study 1

With this first evaluation completed, the next steps will be to use the graph of Figure 3 together with the connected data sources (red) to train a data driven GNN model. Ideally, we learn the difference between simulation and data that we take from the real world to optimally predict the power plant state which can then be used to fulfill the main objective, to reason about credibility of sensor data based on the surrounding information which could highlight "interesting" faults. Interesting here means faults that are not directly obvious to an observer that uses the EMS failure detection algorithms that are already in place. This case study is ongoing and will be included in the next bi-annual report.

Case study 2

After preliminary data analysis, the next step will be to further refine efficiency resolution by considering the number of active nozzles as well as apply GNN to better understand the impact of the interaction between operating units on efficiency. This case study will be finalized by the end of March 2022. In the future, the same efficiency estimation strategy can be applied to assess hydro efficiency of other hydro plants as well as implemented to monitor the hydropower performance live. When efficiency monitoring is available, further downstream fault detection tasks can be performed using machine learning methods.