Real-time Health Management: Outcomes of the DIGI-Hydro Project

J. Schmidt, E. Doujak, C. Königswieser, A. Langer, and A. Schübl

Abstract: The research project DIGI-Hydro aimed to equip existing hydropower machines in Austria with sensors and evaluate how operators can judge the machine's state based on sensor outputs. The joint effort of researchers from different domains and companies led to novel research results, including a better understanding of hydropower plant operation. We summarize our project results in three parts: (1) data handling, (2) visual data analysis, and (3) plant damage characteristics.

1 Introduction

Hydropower remains a key pillar of the energy supply in Austria, which generates around 60 percent of its electricity from hydroelectric power plants [1]. A few hundred large-scale hydropower plants and thousands of smaller ones are in operation, some of which have been running for over 40 years. As such, most of these hydropower plants were planned and built long before significant developments in sensor acquisition, data evaluation, and data analysis could occur [2]. The high age of the plants is considered a substantial factor in why digitization in hydropower is progressing only slowly. However, recent efforts toward a digitalized operation of hydropower plants, particularly the creation of digital twins, have sparked optimism about the future of the industry. Digital twins, virtual representations or replicas of physical objects or systems created using real-time data from sensors, allow users to try out different scenarios virtually and to install preventive measures like predictive maintenance [4].

Physical models and simulations are considered an essential part of the creation process of digital twins [5]. From the physical twin, data sets are recorded, and control and scheduling commands are fed back (enabling supervision and operation). This data allows for creating numerical models and physical test beds to obtain the required simulation capability. As an essential step towards a **digital representation**, solid, scalable, and easy-to-access pipelines for data preparation and analysis must be installed first. However, collecting enough data of the required quality poses new research challenges. In the case of hydropower plants, researchers need to record high-frequency data to capture all possible flow phenomena in the data. Recording data over several weeks and months is required to capture many situations (e.g., heavy rain, snowmelt, drought). Physics-based models can also only be proven accurate when validated with real-world data. Again, data collection is considered an essential part of this process. Even worse, not all physical effects can be pictured in the simulation results. Due to these limitations, validation of physics-based models, nowadays, is conducted mainly for parts where data is significant and obtainable [6].

Based on the aim to work toward creating a digital twin, the **DIGI-Hydro research project** (funded by the FFG Funding Service in Austria) was started in 2021 and is now, after four years, coming to an end. DIGI-Hydro was geared towards injecting new momentum into research on digitalization in hydropower, thus laying the foundations for developing new concepts for digitally operating hydropower plants and assessing their condition automatically. Two research organizations (TU Wien and VRVis Zentrum für Virtual Reality und Visualisierung Forschungs-GmbH) and two companies (HAKOM Timeseries GmbH and Vibroconcept GmbH) explored how operators can integrate digital technologies into hydropower plant operation planning. We equipped existing hydropower machines in Austria with sensors and evaluated how operators can judge the machine's state based on sensor outputs. Sensor data helped us to assess how damaging processes within the machine can be measured and which sensors are needed to achieve this. Within the four years of the project, we faced several challenges common to digitalization projects. One of the main challenges was managing the large amounts of raw sensor data (i.e., several Terabytes), not only to store and backup it but also to analyze it and make sense of it. The raw sensor data was used to validate the flow and mechanical simulations of the hydraulic runner of the machine unit, leading to new insights into fatigue issues of different operational points. One of the core developments of DIGI-Hydro was the creation of a characteristic damage factor map of the investigated hydropower plant, which provided a detailed base for the lifetime investigation of a runner when operating the hydropower plant under different premises. In all stages, the data analysis was supported by the visual inspection of the data, where we, similar to the data handling and storage, had to find ways around dealing with the large amount of collected data.

In this paper, we **summarize the project results** generated by DIGI-Hydro. We divide the results into three parts: (1) data handling, (2) visual data analysis, and (3) plant damage characteristics. For data handling, we outline our data collection and storage pipeline. For visual data analysis, we present results for visually inspecting the data. For plant damage characteristics, we present novel results for calculating damage characteristics maps and the possibility for real-time lifetime estimation.

2 Project Results

The joint effort of researchers from different domains and companies in different fields led to many results, including a better understanding of hydropower plant operation. All project results were based on data collection, providing data-driven insights into the operation of different types of power plants.

2.1 Data Handling

Throughout the project [7], we recorded and analyzed data from two different sources:

- **Operation data**: This type of data, recorded by the plant operators, described the current state of operation of the whole power plant. It contained:
  - Plant operating data (35 channels): Measurement values describing the current operation state of the whole power plant and include, for example, machine performance, rotation speed, pressure, and efficiency.

   - Machine diagnostic data (12 channels): Measurement values, such as shaft and case vibration, were included.
- **Sensor data**: This type of data was recorded by sensors installed during the DIGI-Hydro project; therefore, this type of data has only been recorded temporarily. Installed sensors:
   - Validation data (15 channels): Validation sensors measured different parameters within the machine, such as water pressure, acceleration, and water level.
   - Strain gauge measurements (2 channels): Strain gauge sensors measured strain along the turbine blades.
   - Vibration data (1 channel): Vibration sensors measured possible building vibrations.

Operators delivered operation data in several junks as comma-separated values (CSV) files. We used the PULSE system to record the sensor data. The PULSE system kept the data in binary files that were not directly readable by our analysis tools. To ensure efficient data storage and retrieval, we decided on the Apache Parquet file format, which stores relatively large amounts of data concisely and compactly.

Since data security was important when collecting data, we only stored it within a secure network at TU Wien. Operation data recorded by the power plant operators (machine diagnostics and operation data) could be transferred directly to TU Wien via secure file transfer. The data of sensors installed during the DIGI-Hydro project was recorded on-site and stored in a local network-attached storage (NAS). When necessary, we replaced this local storage unit at regular intervals, and the data was brought to Vienna and transferred to a server at the TU Wien. The data flow and storage are illustrated in Fig. 1.

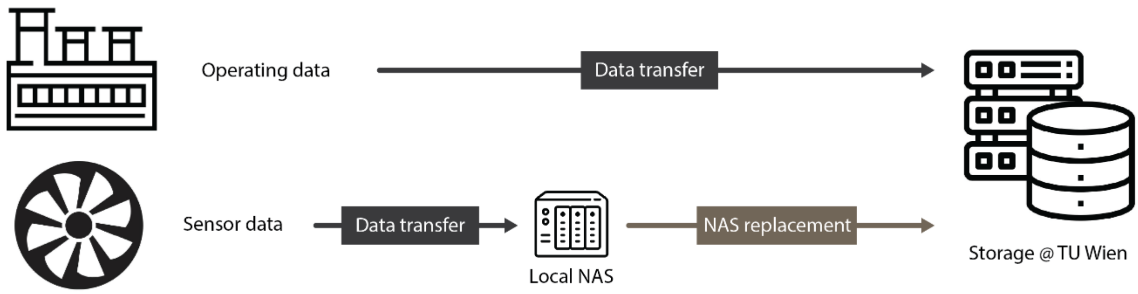

Fig. 1: Data flow in DIGI-Hydro. We collected operation data from the hydropower plan operators which was transferred directly to TU Wien. The recorded sensor data was first stored at a local network-attached storage (NAS) inside the hydropower plant. The NAS was replaced at regular intervals and brought to Vienna, where the data was transferred to the server storage at TU Wien.

2.2 Visual Data Analysis

From the beginning of the project, visual access to the data was considered essential to get an overview of the information collected and to form first hypotheses about the observed phenomena. We relied on three main visualization approaches, which we repeatedly used throughout the project: (i) time series plots, (ii) heat maps, and (iii) correlation matrices.

Time series plots, as clear and informative line-based representations of sensor data, provided us with a detailed analysis of the temporal changes in sensor data over time. In Fig. 2, a time series plot illustrates the machine's operation in spring 2022, with the machine's starting and stopping phases visible.

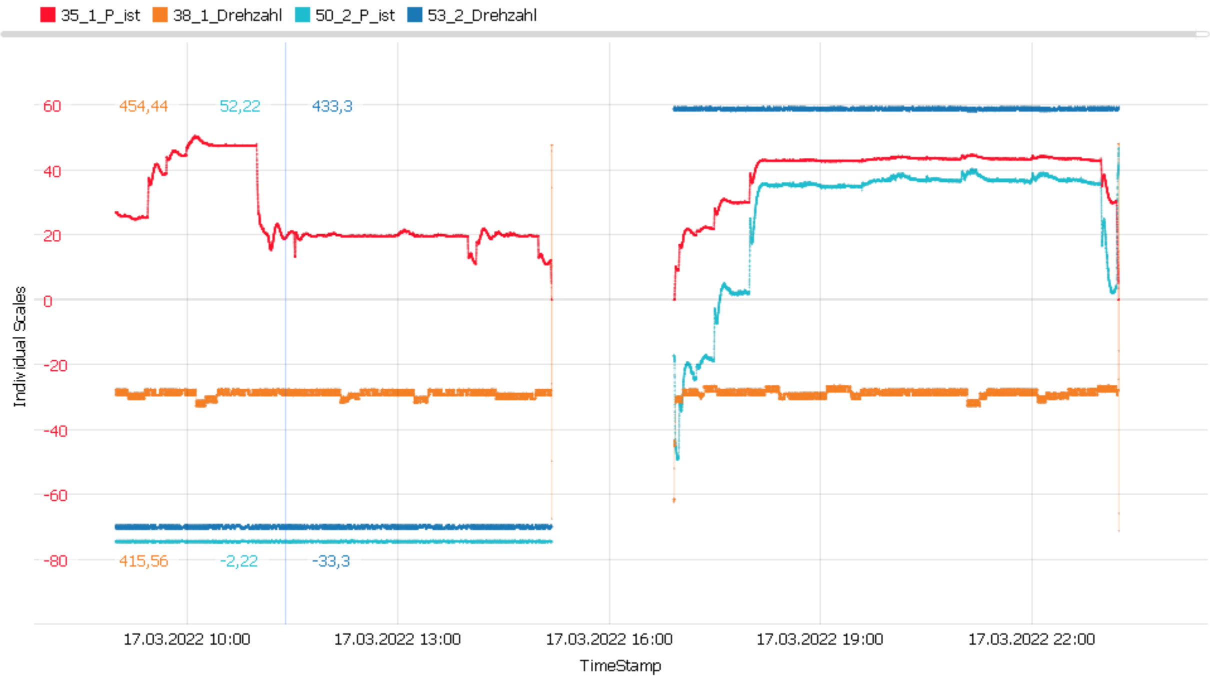

Fig. 2: In this time series plot the machine speed of a power plant for both machines is shown (orange = machine 1, blue = machine 2). While at first only machine 1 was operating, machine 2 was started later on.

Handling large amounts of data also requires a special treatment of the data for visualization. Due to the volume of data, we decided to use an aggregated representation of the aggregated data. We extracted the periods from the sensor data where the machine was in operation, which reduced the data volume by 60% and made it easier to process the data. We calculated the parameters *Q* and *H_Net* from the reduced sensor data and defined a subdivision into sub-ranges for both parameters.

Based on this subdivision, we extracted and visualized a heat map of the time spent in specific operational stages (Fig. 3). The color inside represents the number of seconds spent in this operating state. This representation could identify two operating modes, representing single and dual-machine operations. In the dual-machine operation (left-hand band), two hotspots represent the operational stages primarily used during the operation.

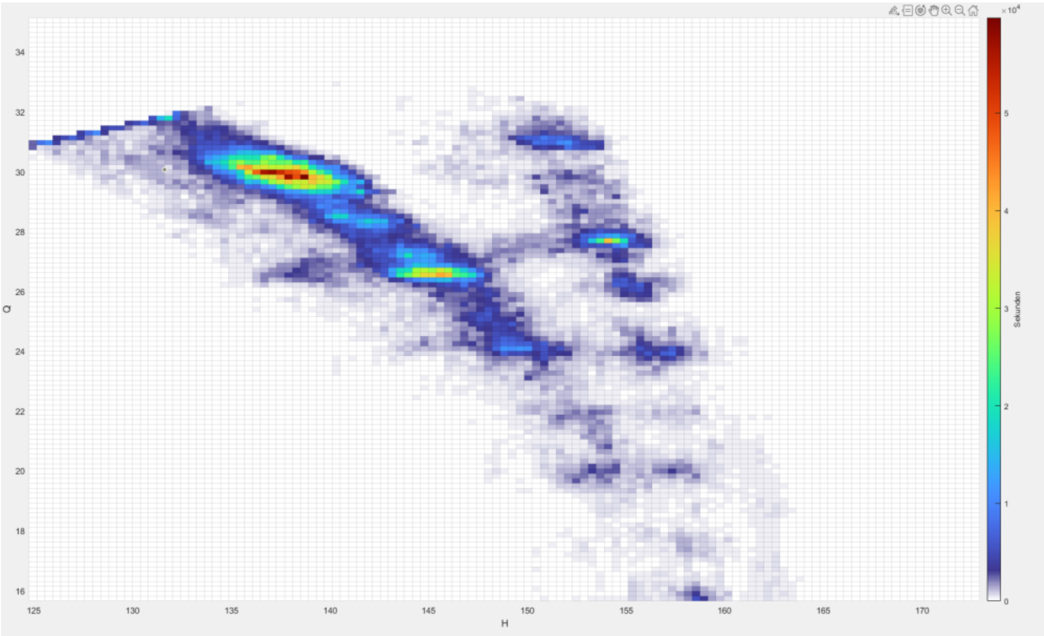

Fig. 3: The representation of the operating states as a heat map clearly shows the differences between single- and dual-machine operation of one of the power plants.

Within certain operational stages, we were further interested in the correlation of the sensor signals. Correlations between all sensor signals are calculated and visualized in a correlation matrix. The sensors are arranged in a matrix, and a rectangle is placed at every grid point. The color of the grid rectangle indicates the correlation. This way we could quickly identify strong correlations (both positive and negative). The correlation matrix for a specific operational stage is shown in Fig. 4.

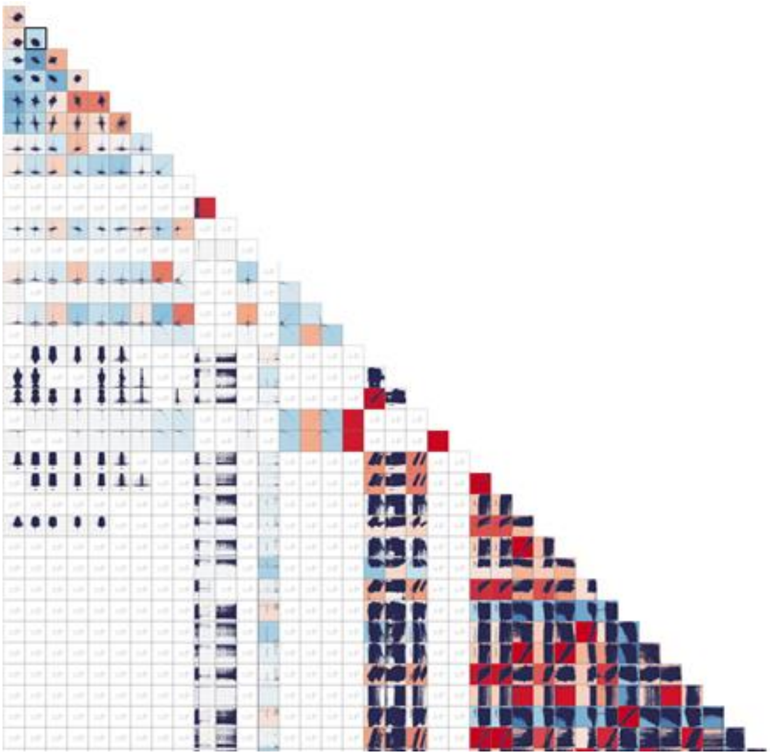

Fig. 4: In a correlation matrix it is possible to identify strong positive and negative correlations between sensor values in specific operational stages.

2.3 Plant Damage Characteristics

The raw sensor data was the primary source for designing and calculating different simulation scenarios [8]. Determining the influencing variables on the operation of a hydropower plant was carried out according to a multi-level approach [9]. On the one hand, we created a simulation model in computational fluid dynamics (CFD) and compared it with the measurements. On the other hand, we examined various models and simulations to optimize the existing simulation model. The multi-level approach was intended to ensure the quality of the simulation.

As shown in Fig. 5, a virtual image of the machine unit machine was created as a 3D computer-aided design (CAD) model. The entire machine set extends from the spiral inlet to the suction pipe outlet. The CAD model was used to determine the water volumes as the basis for the flow simulation. The simulation process breaks the water volume down into more minor volume elements. This process is known as meshing and results in a structure of finite-volume elements. Meshing is very time-consuming and takes up most of the simulation effort.

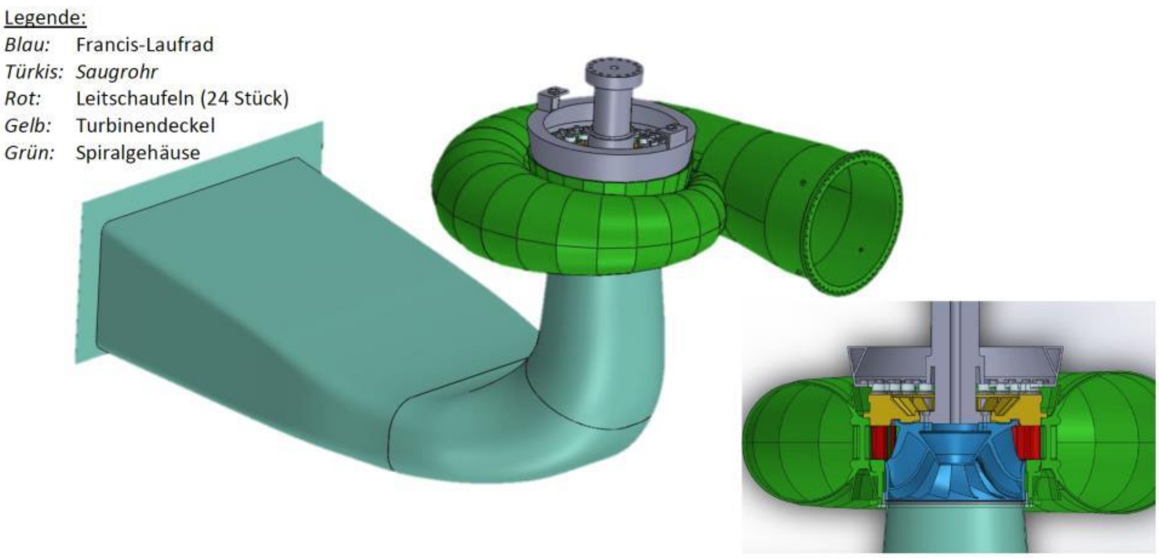

Fig. 5: 3D CAD model of the real machine unit.

Great attention had to be paid to the mesh to obtain meaningful and high-quality simulation results. The choice of operating points was made in coordination with the heat map (as seen in Fig. 3), and we ensured that the operating points used most frequently were also simulated. Over 60 operating points were evenly distributed over the machine characteristics field (*Kennfeld*). We used over 1.5 million CPU hours at the Vienna Scientific Cluster (VSC) to simulate the 60 operating points. After completing the simulations, we used a specially developed program to approximate a surface by the previously selected 60 discrete points. This new, digitized machine characteristics field (*Kennfeld*), shown in Fig. 6, shows the machine's efficiency at the respective operating points [10].

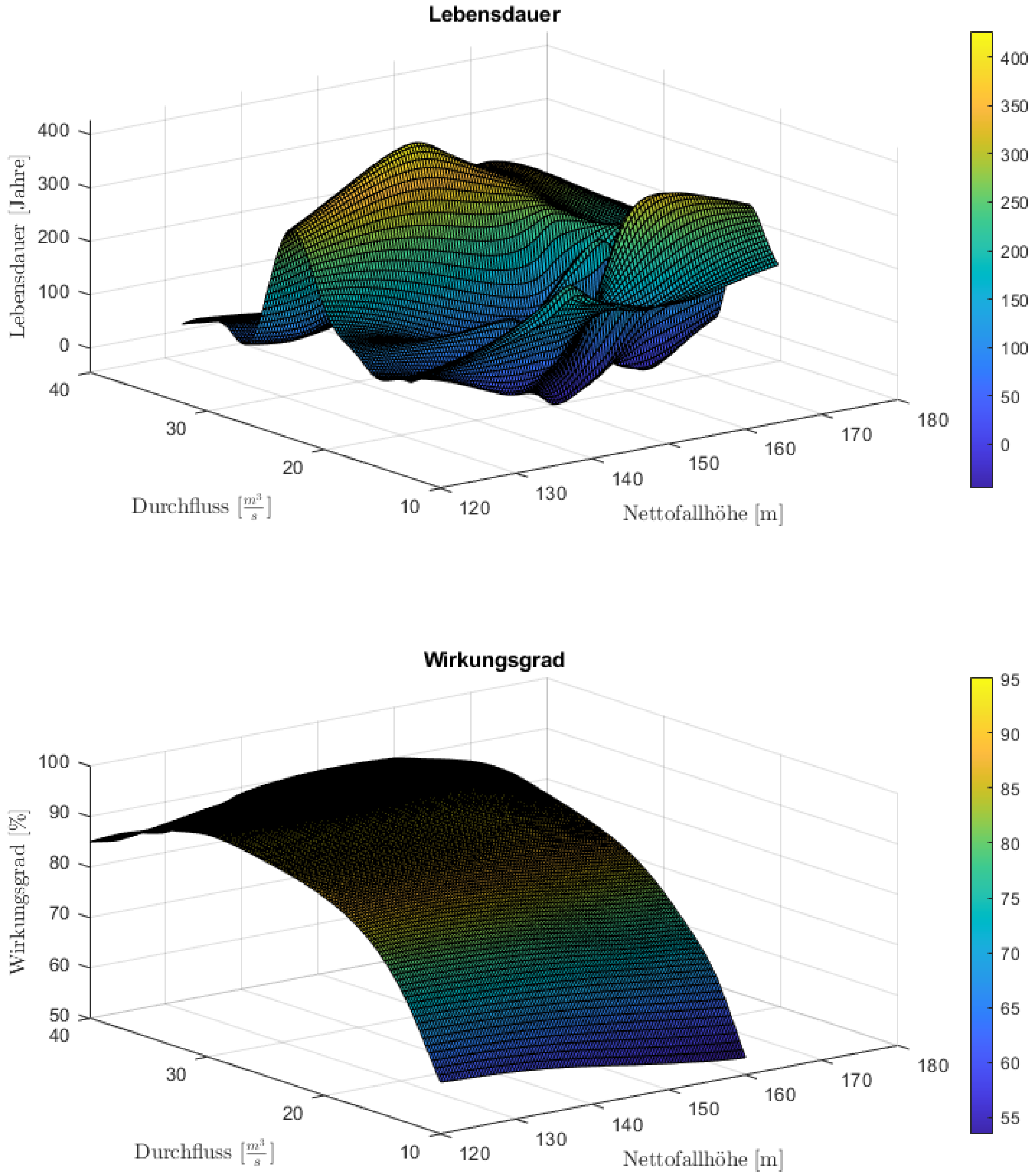

Fig. 6: Digitized and simulated efficiency map of the turbine. The upper diagram represents the calculated damage on a certain operational stage. The lower diagram represents the efficiency of the plant.

The flow simulation results provided the basis for the turbine runner's finite element calculation and the machine's service life calculation. The entire CFD calculation resulted in a data volume of around 54TB. The resulting natural frequencies and modes (natural modes of vibration) are essential for assessing the vibration behavior during operation. We considered the water surrounding the impeller by including the vibrating water mass to ensure that the results correspond as closely as possible to reality.

3 Real-Time Health Management

The project results described in Section 2, data handling, plant damage calculation, and visual data analysis, prepared the basis for providing a real-time health management application for a hydropower plant.

The current plant damage can be calculated based on the current operational stage (given by *Q* and *H_Net*). The accumulated damage can also be calculated and monitored over time, according to the different operational stages.

Fig. 7 shows the visual interface of the health management tool. On the left side, the plant's heat map shows the operational stages in which the power plant was run most often. We can also see the two options of one-machine and two-machine operation. A yellow dot with a red outline shows the current machine operation point. Based on the current machine operation, the current damage can be calculated (bar chart on the right side). The damage values accumulate over time and show the current state of the machine (line plot on the right side).

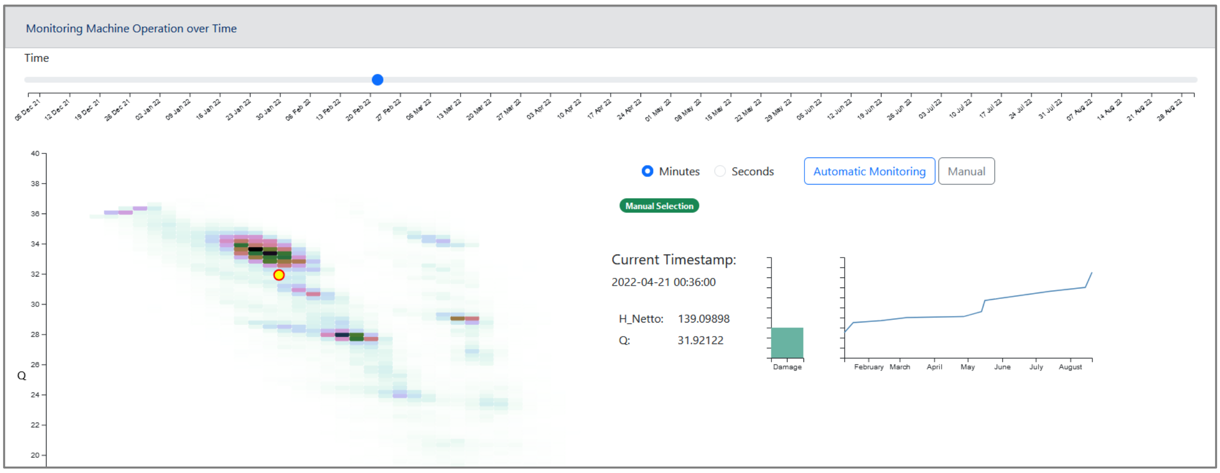

Fig. 7: Visual interface of the real-time health management application implemented in the DIGI-Hydro project. On the left side, the heatmap representing frequent operational stages is shown. On the right side, the damage caused by the current operational point is represented.

4 Lessons Learned

In this Section, we summarize our lessons we learned in the course of the DIGI-Hydro research project.

4.1 Data Storage

One of the most challenging tasks we identified in working with the data was handling and storing the large amounts of data we generated. In DIGI-Hydro, we decided to permanently record the sensor data, independent of the current machines' state. The raw sensor data eventually summarized up to a high amount of data daily, as we recorded more than 60 sensor channels. For the nine months of the last measurement period (December 2021 - August 2022), we collected 75 TB of data - about 232 GB per day and about 9.7 GB per hour. In our setup, we collected sensor data on-site.

Otherwise, it would have been necessary to transfer the data over the network to TU Wien, and with about 9.7 GB per hour, this would not have been possible. As a lesson learned, we recommend processing data locally at the power plants and aggregating information before transferring it to external data storage.

4.2 Data Analysis

Like the data storage problem, the vast amount of data made it challenging to analyze the entire data holistically. For analysis, we followed the approach of splitting the data into smaller junks that could then be analyzed separately. For the split, we used information from the operational data to concentrate on interesting timespans. Using predefined ideas and theories about the data, we could proceed with confirmatory analysis on smaller data junks, which made it easier for us.

4.3 Data Fusion

In the DIGI-Hydro project, we mainly concentrated on sensor data. In the future, fusing data from diverse data sources will be needed to create a more comprehensive understanding of plant operations. Data fusion will involve data from sensors, simulations, physical models, weather, climate, and environmental sources to provide a more accurate and holistic view of the system’s performance.

4.4 Data Visualization

Before starting to implement data visualization approaches we searched for existing solutions for analyzing such large amounts of sensor data. Unfortunately, existing visual analytics applications have reached their limits in serving as a proper application for exploring the data. Several frequently used commercial applications for visual analytics have already been evaluated, and the comparison shows that most applications reach their limits when loading and visualizing a dataset with 50 GB [11]. Increasingly, techniques for showing large time series data have been implemented and proposed in visualization research [12]. Using these tools, however, requires programming knowledge and skills in either Python or JavaScript.

5 Conclusion

In conclusion, the DIGI-Hydro research project has successfully advanced the digitalization of hydropower plants by integrating sensor technology, data analysis, and simulation methods. Through the project, significant progress was made in handling large volumes of sensor data, exploring techniques for visual data analysis, and assessing plant damage characteristics. Based on these efforts, we were able to created and implement a real-time health monitoring application, enabling more accurate lifetime predictions and improved maintenance strategies. The outcomes lay the groundwork for future digital twin applications in hydropower, offering a promising path toward more efficient and reliable plant operations.

References

[1] Erneuerbare Energie Österreich (2024). *Wasserkraft*. https://www.erneuerbare-energie.at/wasserkraft

[2] J. Giesecke, E. Mosonyi, & S. Heimerl (2009). *Wasserkraftanlagen: Planung, Bau und Betrieb*. Springer Science & Business Media.

[3] M. Ersan & E. Irmak (2024). *Development and Integration of a Digital Twin Model for a Real Hydroelectric Power Plant*. Sensors 24(13):4174.

[4] N. Nikolakis, K. Alexopoulos, E. Xanthakis, & G. Chryssolouris (2019). *The digital twin implementation for linking the virtual representation of human-based production tasks to their physical counterpart in the factory-floor*. International Journal of Computer Integrated Manufacturing 32:1-12.

[5] M. Zhang, F. Tao, B. Huang, A. Liu, L. Wang, N. Anwer, & A. Y. C. Nee (2021). *Digital twin data: methods and key technologies*. Digital Twin 1:2.

[6] D. J. Waag, K. Worden, R. J. Barthorpe, & P. Gardner (2020). *Digital Twins: State-of-the-Art and Future Directions for Modeling and Simulation in Engineering Dynamics Applications*. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering.

[7] J. Schmidt, E. Doujak, R. Wickert, & G. Beyerle (2022). *Enabling Big Data Analysis for Digitalisation in Hydropower*. Proceedings of the 21st International Seminar on Hydropower Plants (Viennahydro ’22), Nov. 9-11, Vienna, Austria.

[8] E. Doujak, S. Stadler, G. Fillinger, F. Haller, M. Maier, A. Nocker, J. Gaßner, & J. Unterluggauer (2022). *Fatigue Strength Analysis of a Prototype Francis Turbine in a Multilevel Lifetime Assessment Procedure Part I: Background, Theory and Assessment Procedure Development*. Energies 15:1148.

[9] E. Doujak, J. Unterluggauer, G. Fillinger, A. Nocker, F. Haller, M. Maier, & S. Stadler (2022). *Fatigue Strength Analysis of a Prototype Francis Turbine in a Multilevel Lifetime Assessment Procedure Part II: Method Application and Numerical Investigation*. Energies 15:1165.

[10] E. Doujak, A. Maly, J. Unterluggauer, F. Haller, M. Maier, C. Blasbichler, & S. Stadler (2022). *Fatigue Strength Analysis of a Prototype Francis Turbine in a Multilevel Lifetime Assessment Procedure Part III: Instrumentation and Prototype Site Measurement*. Energies 16:6072.

[11] M. Behrisch, D. Streeb, F. Stoffel, D. Seebacher, B. Matejek, S. H. Weber, S. Mittelstaedt, H.-P. Pfister, and D. Keim (2018). *Commercial Visual Analytics Systems-Advances in the Big Data Analytics Field*. IEEE Transactions on Visualization and Computer Graphics 25(10):3011-3031.

[12] J. Van Der Donckt, J. Van der Donckt, E. Deprost, & S. Van Hoecke (2022). *Plotly-Resampler: Effective Visual Analytics for Large Time Series*. In Proceedings of IEEE Visualization and Visual Analytics, VIS ’22, pp. 21-25.

Authors

Dipl.-Ing. Dr. Johanna SCHMIDT
VRVis Zentrum für Virtual Reality und Visualisierung Forschungs-GmbH
Donau-City-Straße 11, 1220 Vienna, AUSTRIA
Phone: +43-1-9089892-302
E-mail: johanna.schmidt@vrvis.at

Johanna Schmidt is the head of the research group *Visual Analytics* at the VRVis Zentrum für Virtual Reality und Visualisierung Forschungs-GmbH. Her main research interests include the visual analysis of time-dependent and heterogeneous data and the integration of visualization in data science workflows. She acquired her PhD in Visual Computing at TU Wien in 2016. Johanna is leading and participating in several application-oriented research projects.

Dipl.-Ing. Dr. Eduard DOUJAK
TU Wien
Institute of Energy Systems and Thermodynamics
Research Unit of Fluid-Flow Machinery
Getreidemarkt 9/302, 1060 Vienna, AUSTRIA
Phone: +43-1-58801-302400
E-mail: eduard.doujak@tuwien.ac.at
Eduard Doujak graduated in mechanical engineering from the TU Wien and now works as an assistant professor at the TU Wien since 1995. He is involved in the High Head Hydraulic Machines and Plants investigations with special consideration to pumped storage systems as well as in Uprating and Refurbishment of Hydropower Plants. He is coordinating and leading several projects on digitalization of hydraulic systems. In this matter he holds national and international patents and received the national INVENTUM price for the second best invention in 2011. He is member of several Advisory Committees of international conferences.

Constantin KÖNIGSWIESER, BSc
TU Wien
Institute of Energy Systems and Thermodynamics
Research Unit of Fluid-Flow Machinery
Getreidemarkt 9/302, 1060 Vienna, AUSTRIA
Phone: +43-1-58801-302500
E-mail: constantin.koenigswieser@tuwien.ac.at

Constantin Königswieser holds a bachelor degree in mechanical engineering. He finished his bachelor studies at the TU Wien and now works as a student assistant at the TU Wien. His research interests include industrial and mechanical engineering, CAD, finite element analysis, and solid mechanics.

Dr. Anastasia LANGER
HAKOM Time Series GmbH
Lemböckgasse 61, 1230 Vienna, AUSTRIA

Phone: +43-1-8157980-112
E-mail: anastasia.langer@hakom.at

Anastasia studied in the field of geology/geochemistry and received her PhD with a thesis on experimental and thermodynamic modeling of processes in the Earth's interior. In addition to experience in thermodynamic modeling of phase equilibria, she has also gained experience in the development of statistical and rule-based models for the energy industry, e.g., risk models for geothermal energy, oil and gas production, and forecasting models for electricity and gas generation and consumption. Anastasia Langer now works as a Senior Analyst at HAKOM Time Series GmbH.

DI Alfred SCHÜBL
VibroConcept GmbH
Dreyhausenstraße 11/3 1140 Wien, AUSTRIA
Phone: +43-17862607-24
E-mail: schuebl@vibroconcept.at

DI Alfred Schübl finished his studies at the Institute of Communications Engineering at the Vienna University of Technology. Alfred now has more than 25 years of experience in the vibration monitoring of hydropower machines. The results of this work have been presented in several publications at international conferences.