

Deliverable 2.4 Public Dataset Representing Geothermal Operation

Version March 2026



Authors**Company:**

TNO

From:Jonah Poort
Zeryab Alam
Pejman Shoeibi Omrani
Ryvo Octaviano**Email:**jonah.poort@tno.nl
pejman.shoeibiomrani@tno.nl
ryvo.octaviano@tno.nl**Project coordinator:**Laura Precupanu
laura.precupanu@tno.nl**Subject:**

Publicly available dataset generated by the models to represent geothermal production

Reviewers**Company:**

WEP

From:

Arturs Blinovs

Email:arturs.blinovs@we-p.nl**Date:**

11-03-2026

This project is executed with funding of the National Growth Fund of the Ministry of Economic Affairs, implemented by the Netherlands Enterprise Agency. The specific funding for project *Data gedreven optimalisatie van aardwarmte systemen* (NWN10005) concerns the funding rules 2022.

Executive summary

This document presents Deliverable D2.4 – Representative Dataset, produced within the Nieuwe Warmte Nu! (NWN) program to support the development and validation of the GEMINI digital twin for deep geothermal systems in the Netherlands. The dataset provides a publicly available, synthetic yet realistic representation of the operation of a typical low-enthalpy geothermal doublet. It enables users who may not have access to real operational data to evaluate, test, and familiarize themselves with GEMINI's core functionalities.

To construct this representative dataset, internal TNO-developed models were used to model a standard geothermal doublet including essential primary-side components such as the production and injection reservoirs and wells, an electrical submersible pump (ESP), a separator, filters, the central heat exchanger, and booster (injection) pump. The system was simulated over a six-month period at 15-minute resolution. Variables were allowed to vary stochastically using a Bernoulli-Gaussian process, ensuring realistic operational fluctuations in pressures, temperatures, flow rates, and pump frequencies.

In addition to the baseline dataset representing normal operations, two classes of anomalous events were modelled and the data generated from those scenarios are published as a scenario reflect degradation mechanisms. These anomalies included heat exchanger fouling and filter clogging. These anomalies allow users to explore fault signatures, validate monitoring strategies, and test predictive maintenance workflows within the digital twin environment.

This document also explains how the dataset can be imported into GEMINI software and be linked to the simulation models. Together, the dataset and accompanying guidance form a reusable template for interacting with geothermal digital twin models and provide an accessible foundation for testing, training, and demonstration purposes.

Introduction

The integration of Digital Twin technology into a low-enthalpy geothermal plant represents a significant advancement in real-time monitoring, predictive maintenance, and overall operational efficiency. The focus of the Nieuwe Warmte Nu! (NWN) project is to develop and implement a Digital Twin for deep geothermal plants (doublets). This will enable monitoring and simulation of plant operations, allowing for performance optimization, optimized maintenance planning, and risk mitigation.

Deliverable D2.4 of the project consists of a dataset that will be made publicly available and is representative of geothermal operation in the Netherlands. This dataset is published to provide those interested in the GEMINI digital twin with a synthetic yet realistic minimum set of data required to be able to test the basic functionality of the digital twin, even if they have no data of their own to which the digital twin can be connected. In addition, it allows for an interested party to explore the capabilities of the GEMINI digital twin without having to already go through the more involved step connecting it to their own real-time databases.

This document functions as a reference document, and contains a high-level description of the dataset as well as a description of the template project in GEMINI that this data was generated for. The data can be downloaded as two .csv files from the same source as this document. One file, named “GEMINI_Clean_17280datapoints.csv” contains simulated data corresponding to regular operating conditions, while the file “GEMINI_Anomalous_17280datapoints.csv” corresponds to data in which simulated anomalies have been included.

Description of the modelled system

The representative dataset was developed for a basic single geothermal doublet. The example system consists of the main blocks to represent the typical primary side of such a geothermal doublet, and consists of the following units:

- The reservoir (modelled as the near wellbore area around the producer and injector well).
- The production well.
- An ESP.
- A degasser/separator unit.
- A filter system.
- A heat exchanger.
- A second filter system.
- A booster pump.
- The injection well.

This system is schematically represented (along with the simulated variables) in Figure 1. For the simulated data, the secondary system including gas dryer, gas boiler, combined heat and power (CHP), and grid gas have been left out.

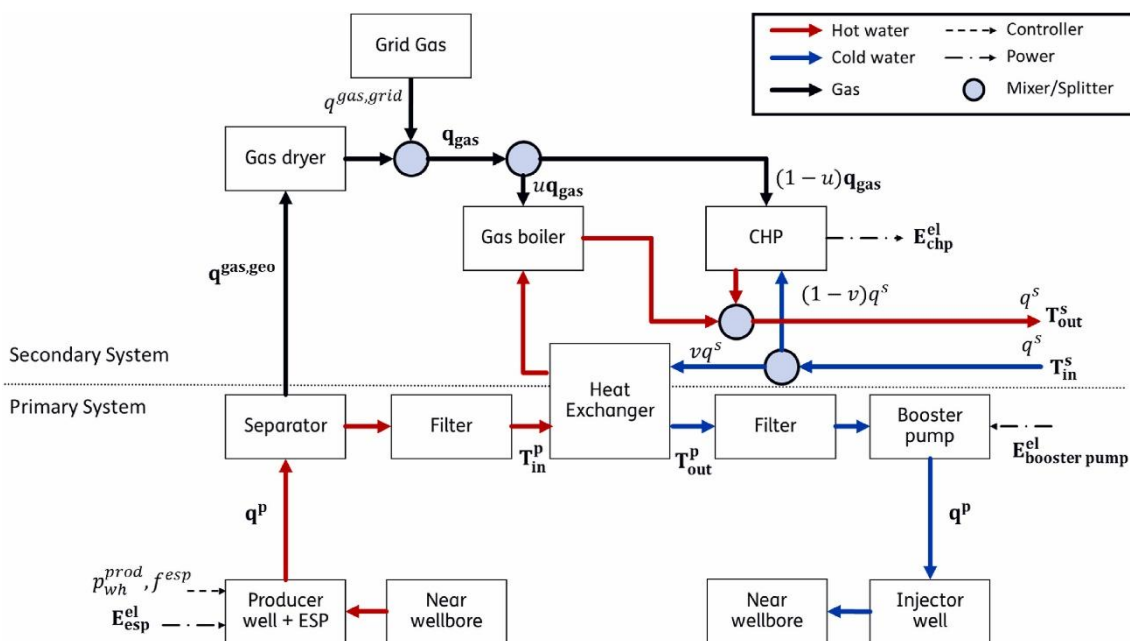


Figure 1 Flow diagram of the units included in the project template including simulated variables. Reproduced from (Shoeibi Omrani, et al. 2026) licensed under CC BY 4.0.

To allow a new user of the GEMINI digital twin to test the system with the data provided alongside this report, the system as described above has also been implemented as a template project within the digital twin. This template is shown in Figure 2, and can be accessed in the digital twin by loading the “geothermal example” project that comes default with GEMINI, or by clicking the “Template” button in the “Diagram Builder” application (Figure 3).

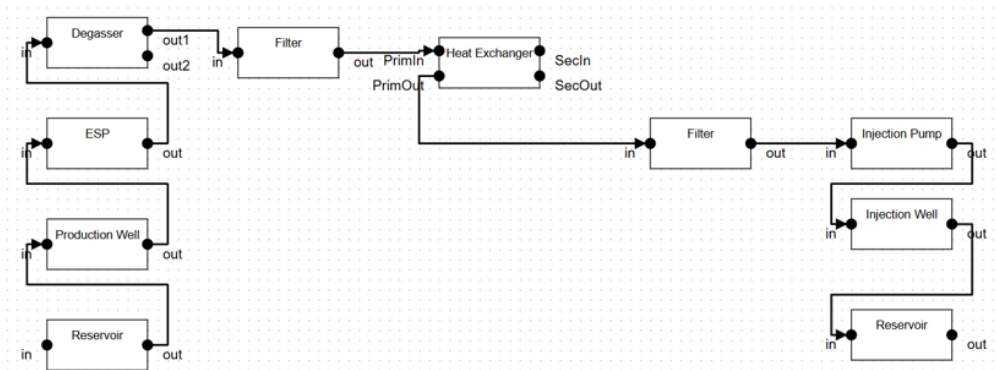


Figure 2 Schematic block diagram view of the template project as created in the system diagram builder of the GEMINI digital twin.

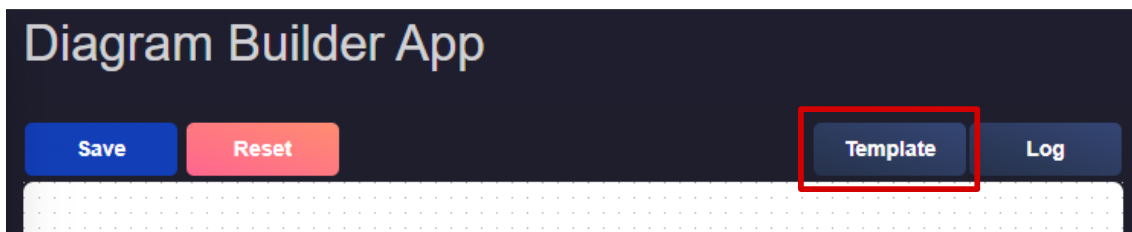


Figure 3 Screenshot of the "Diagram Builder" application in the GEMINI digital twin, highlighting the "Template" button used to create the example system from Figure 2.

Description of the generated data

The example data was generated using models internally developed at TNO. They are described in Annex I of (Shoeibi Omrani, et al. 2026).

The modelled system is described through 31 variables describing the temperatures, pressures, pump power draws, and produced power in the system. As mentioned in the system description, the modelled system only includes the primary side of the geothermal doublet. As such, no combined heat and power (CHP) and gas boiler are included in the calculations. All gas filtered out at the degasser is assumed to all be directed to the CHP (although its functionalities are not modeled), and all secondary side water flow goes through the heat exchanger. In addition, it is assumed that there is no heat loss to the ambient environment at the top side, and heat exchange only takes place in the heat exchanger itself. Furthermore, pressure drops in pipes connecting surface equipment is also neglected.

The system uses the following fixed boundary conditions:

- The reservoir productivity index of the producer is set to 20, while that of the injection reservoir is set to 3.4 (in order to simulate injection pressures seen in Dutch geothermal sites).
- The production reservoir pressure is set to 265 bar, while that of the injection reservoir is set to 255 bar.
- The ambient temperature is fixed to 20 °C.

If the simulated variables would be kept constant, the resulting data would also be constant for the entirety of the simulation period, leading to uninteresting and unrealistic values not representative of real operation. Therefore, in order to simulate a stable, but more realistic period with natural variations, the Bernoulli-Gaussian process was used to randomly introduce changes into the system. In this process, every simulated minute, there is a chance for a number of variables to fluctuate slightly. These variables include:

- Producer wellhead pressure ("Pwh_prod"): there is a 0.01% chance per minute to increase or decrease the variable by a value drawn from a normal (Gaussian) distribution with a mean of 0 and standard deviation of 0.0005. This value is clipped to 5–10 bar. This models rare pressure-related interventions with minimal magnitude.
- ESP frequency ("ESP_freq"): there is a 0.002% chance per minute to increase or decrease the variable by a value drawn from a uniform distribution between 3 and 5. This value is clipped to 55–70 Hz, and models very rare operator interventions within typical ESP operating limits.
- Flow rate at the secondary side of the heat exchanger. Secondary ("Qs"): there is a 0.03% chance per minute to increase or decrease the variable by a value drawn from a Gaussian distribution with a mean of 0 and standard deviation of 0.1. This value is clipped to 145–165 m³/h, reflecting occasional tuning actions with small but realistic magnitude.
- Inlet temperature ("T_s_in"): there is a 0.2% chance per minute to increase or decrease by a value drawn from a normal distribution with a mean of 0 and a standard deviation of 0.045. This value is clipped to 20–30°C in order to capture infrequent thermal adjustments while enforcing realistic bounds.

These variables are plotted in Figure 4.

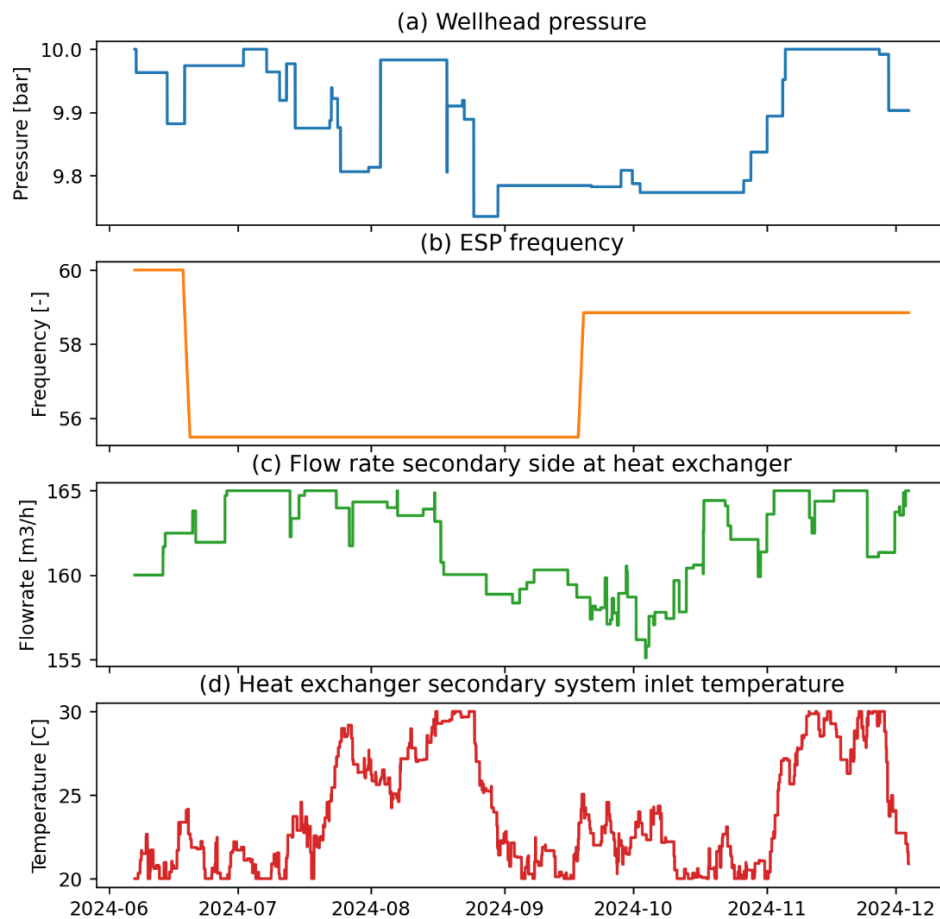


Figure 4 Plot of the four variables varied during the simulation of the geothermal data. Respectively wellhead pressure (a), ESP frequency (b), secondary heat exchanger side flowrate (c) and inlet temperature (d).

The changes of the above four variables in turn affect the other variables simulated in the system. In addition to the prior described four variables, the models provide the following additional variables:

- The heat exchanger flowrate on the production side ("Qp"), as well as its inlet and outlet pressures, and pressure drop over the heat exchanger ("HEX_Pin", "HEX_Pout", "HEX_PDrop").
- The temperatures at the inlet and outlet of the production side of the heat exchanger ("T_p_in" and "T_p_out"), as well as the outlet temperature of the heat exchanger at the secondary side ("T_s_out").
- The ESP intake and discharge pressures ("ESP_Pin" and "ESP_Pout"), its head ("ESP_Head"), and its power draw ("ESP_PowerDraw").
- The separator inlet and outlet pressures ("Sep_Pin" and "Sep_Pout").
- The inlet and outlet pressures of filter A (before the heat exchanger) and filter B (after the heat exchanger), as well as their pressure drops ("FilterA_Pin", "FilterA_Pout", "FilterA_PDrop", and "FilterB_Pin", "FilterB_Pout", "FilterB_PDrop").
- The injector pump inlet and discharge pressures, and its power draw ("InjectorPump_Pin", "InjectorPump_Pout", "InjectorPump_PowerDraw").
- The thermal power generated by the geothermal doublet ("Thermal_PowerGen").
- Status indicators for the ESP ("esp_status"), heat exchanger ("hex_status"), and both filters ("filterA_status", "filterB_status"). These indicators show whether the corresponding unit is in a simulated degradation mode or experiencing normal behavior.

All data was generated for a period of six months at a time interval of 15 minutes. Figure 5, Figure 6, and Figure 7 show a selection of the simulated variables, focusing on the primary side heat generation at the heat exchanger, the ESP, and the injector pump respectively. As can be seen, the variations introduced in the variables plotted in Figure 4 have effects throughout the rest of the system. For instance, the ESP

frequency drop causes a subsequent drop in primary side flowrate, which, combined with an increase in the heat exchanger secondary side inlet temperature, causes two significant periods in which the system produces a lot less heat than on average (one around August and one around November, see Figure 5). This same frequency drop also causes considerable changes to the injector pump discharge pressure (Figure 7), which, due to the drop in primary side flowrate, needs to inject at lower pressure.

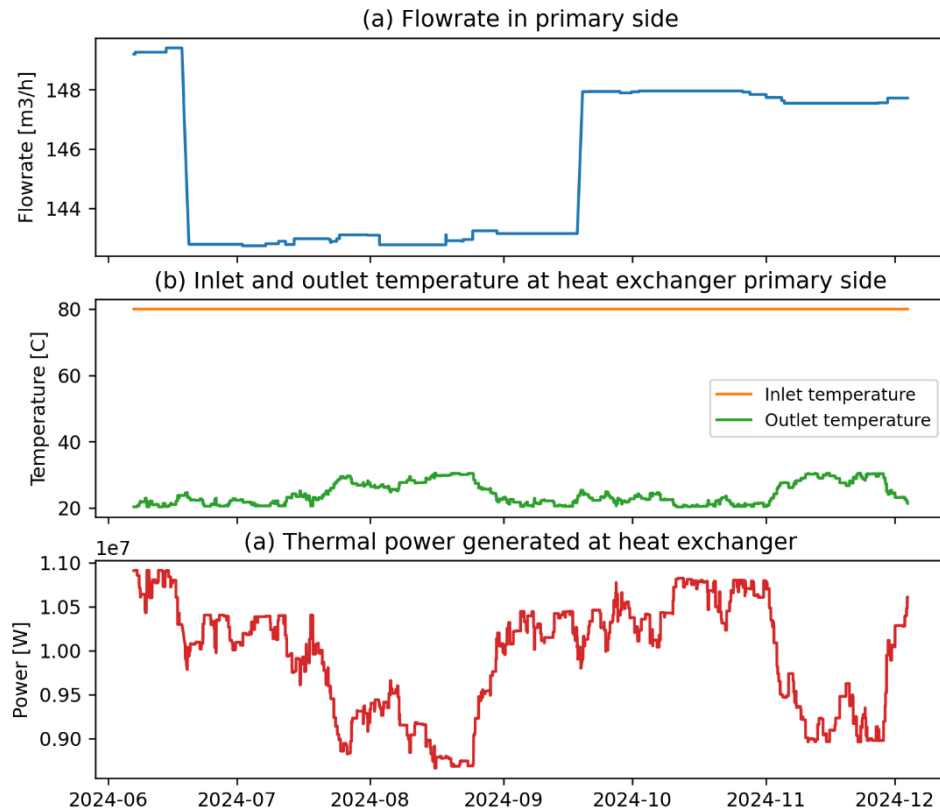


Figure 5 Simulated flowrate in the primary side of the doublet (a), inlet and outlet temperatures at the primary heat exchanger side (b), and the thermal power generated in the heat exchanger (c)

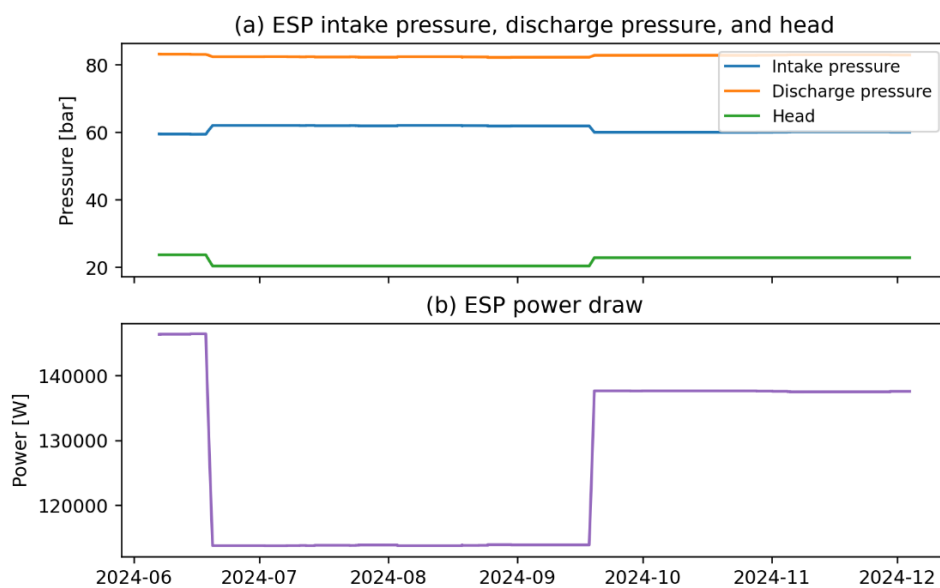


Figure 6 Simulated intake pressure, discharge pressure, head (a), and power draw (b) of the ESP.

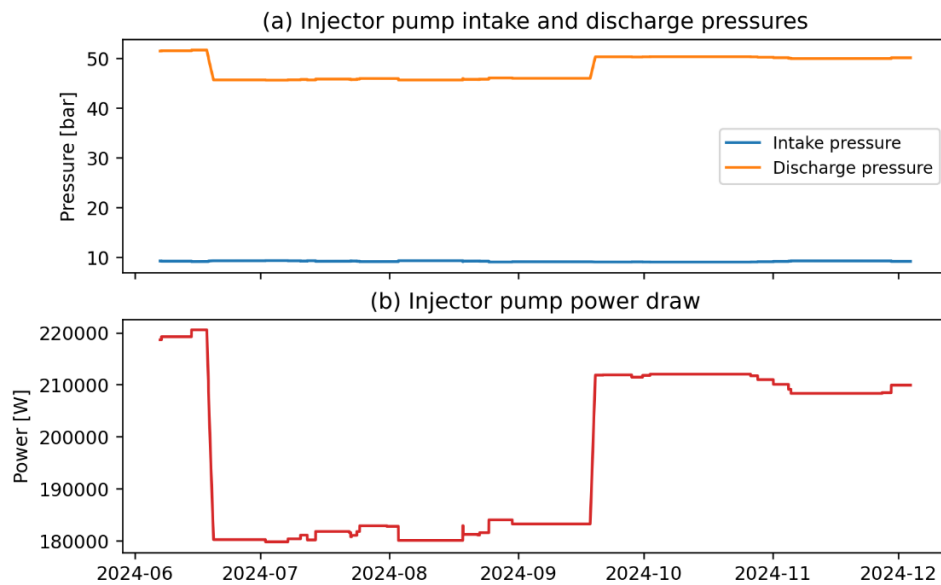


Figure 7 Simulated intake and discharge pressure (a), and power draw (b) of the injector pump.

Data with anomalies

To provide additional data to the user that reflects operation outside of the stable norm, additional data was generated in which production anomalies were artificially simulated by slowly increasing resistances in two of the units in the example doublet. These anomalies correspond to the physical phenomena of heat exchanger fouling and filter clogging (in both filters). These anomalies were simulated using random Bernoulli-Gaussian processes as increased resistances to the flows through these units, resulting in an increase in the pressure drop over each system. A comparison of the regular and anomalous data for the heat exchanger (showing one anomaly) and filter A (showing four subsequent anomalies) are shown in Figure 8 and Figure 9 respectively.

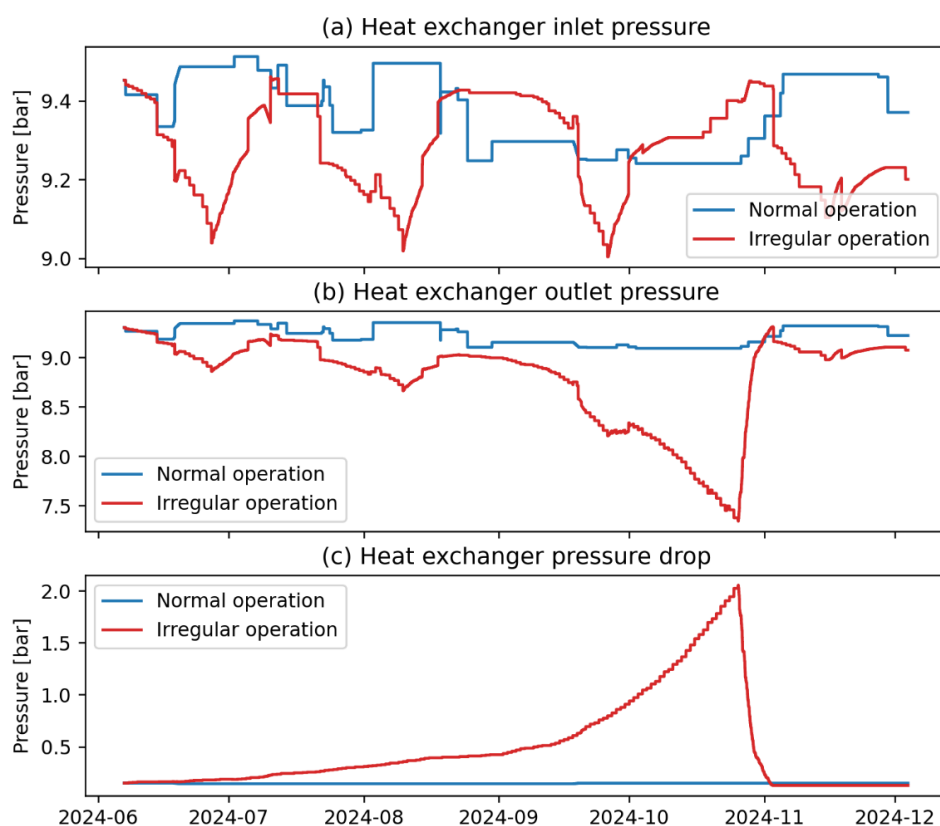


Figure 8 Comparison plots between regular (blue) and anomalous (red) simulated data for the heat exchanger inlet pressure (a), outlet pressure (b), and pressure drop (c).

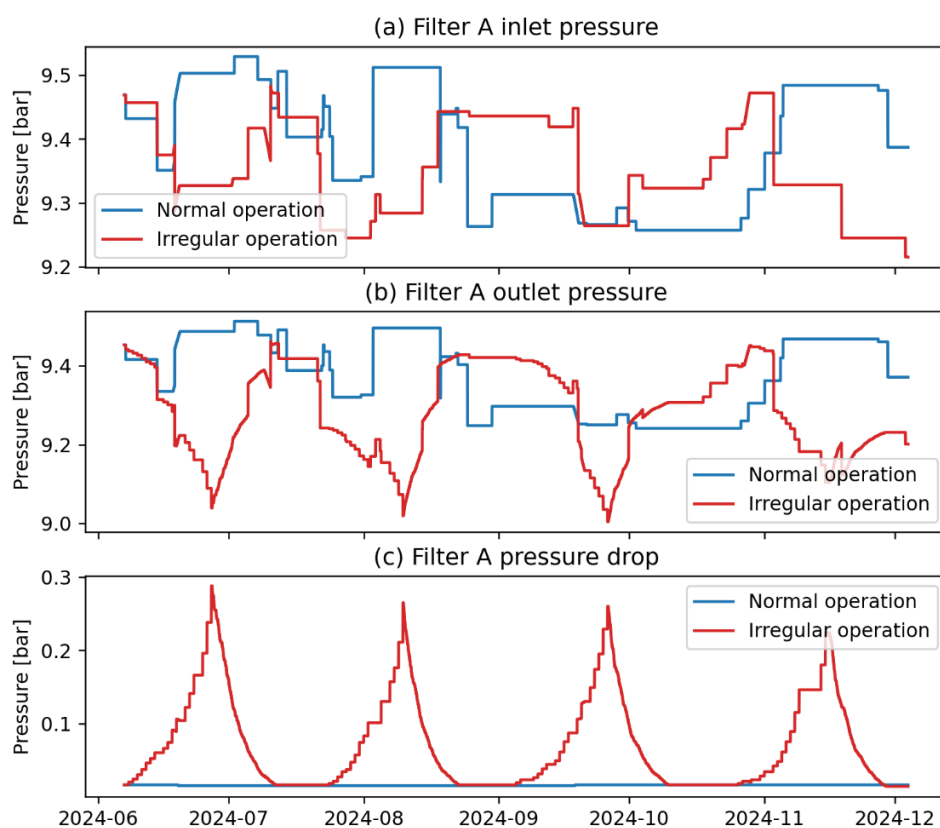


Figure 9 Comparison plots between regular (blue) and anomalous (red) simulated data for filter A inlet pressure (a), outlet pressure (b), and pressure drop (c).

Using the data in the GEMINI digital twin

The simulated data as described in the previous section can be used in GEMINI digital twin by uploading the .csv files manually within GEMINI. This can be done through the “Data Management” application, in the “Tag Manager” menu under “Manual Data Import”. In this menu, one can click the “Choose file” button, which opens a local file explorer window through which the user can navigate to the local .csv file with the simulated data. Once selected, the data can be uploaded to GEMINI using the “Upload Data” button. Figure 10 shows a screenshot of the GEMINI digital twin that highlights the above menus to navigate to manual data upload to the system.

It should be noted that the manual importing and uploading of data using a CSV file is only available when no connection to an external database (e.g., AvevaDB) is in place. This is done to prevent conflicts between manually uploaded data and data coming directly from an external database. Please ensure that the correct settings are configured in the plant configuration before uploading data.

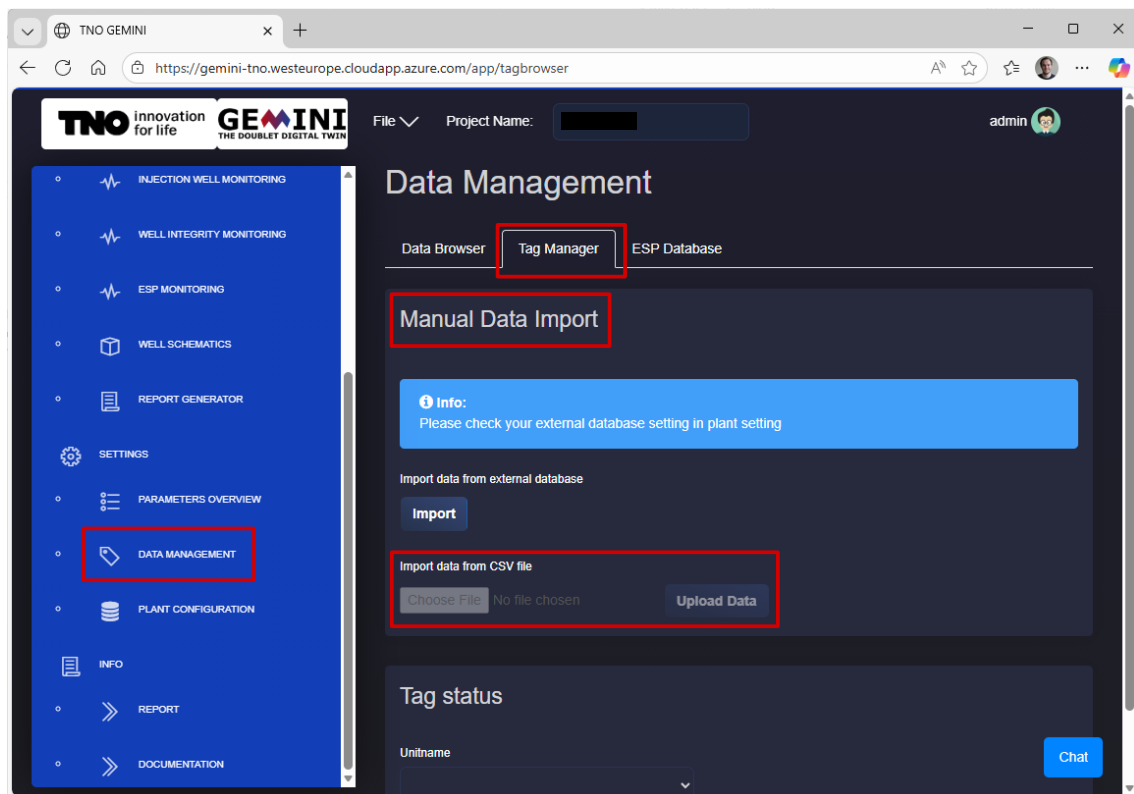


Figure 10 Screenshot of the manual data uploading menu within the GEMINI digital twin.

Once the data has been uploaded to the system, the tags (variables) need to be connected to the right unit variables in order for them to be used by the underlying models, modules, and applications in the GEMINI digital twin. This is done through the “Diagram Builder” application. By right-clicking on a unit in the system diagram as shown in Figure 2, one can open the unit’s corresponding component parameter menu (Figure 11). In this menu, one can navigate to the “Tagnames” sub-menu in order to provide which variable name in the uploaded .csv datafile refers to the corresponding variable name as used by the GEMINI digital twin system (e.g. the inlet pressure of the ESP is called “ESP_Pin” in the csv file).

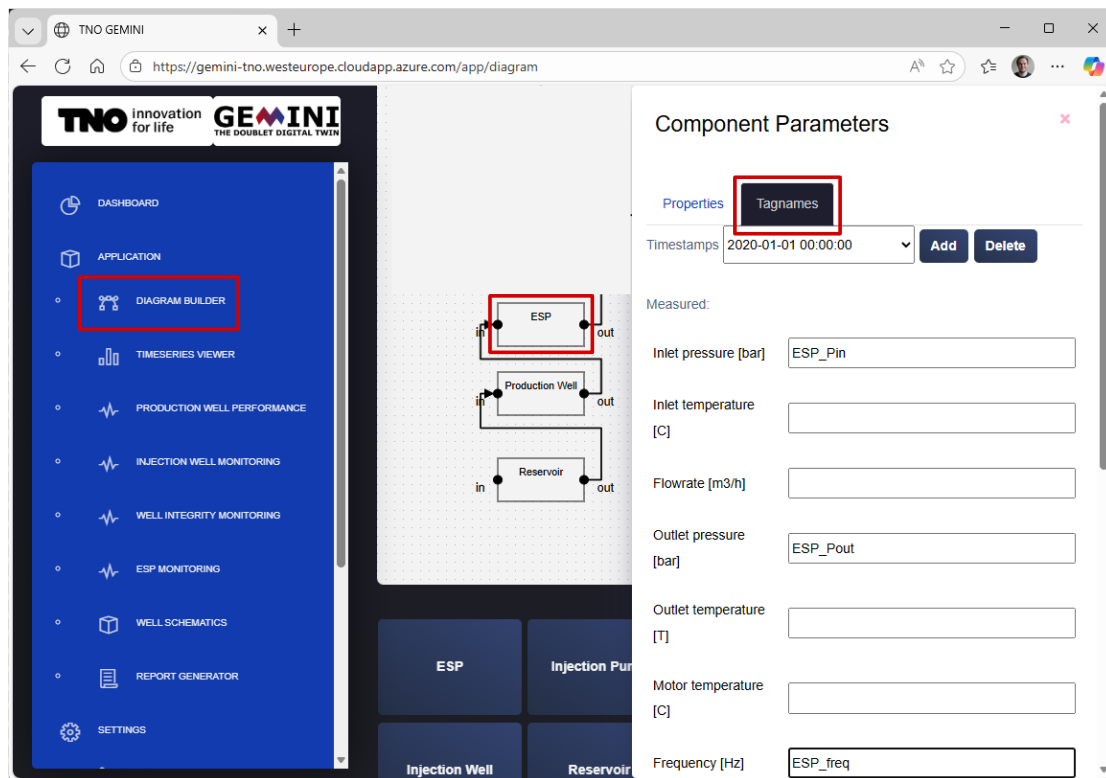


Figure 11 Screenshot of the "Diagram Builder" application within the GEMINI digital twin showing the "Tagnames" sub-menu of the "Component Parameters" menu of the highlighted ESP unit.

Finally, once the tagnames have been connected to their internal variable names, they can be used in the various applications within GEMINI and plotted and viewed in the "Timeseries Viewer" application (see Figure 12).

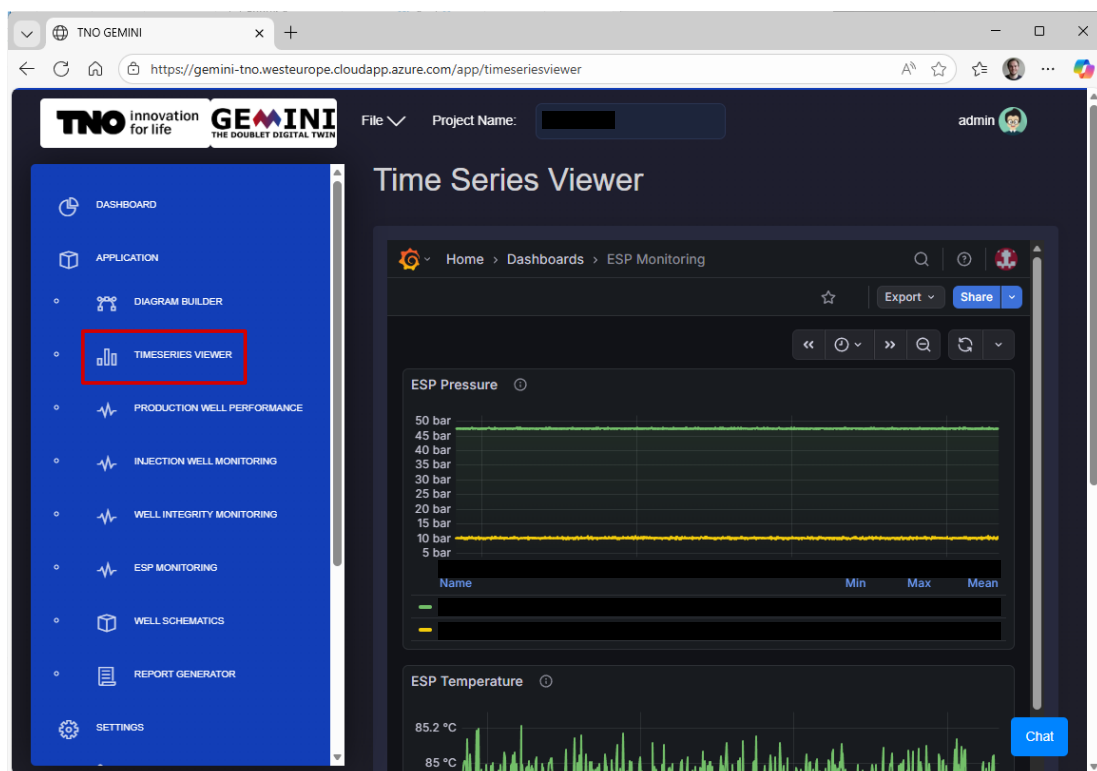


Figure 12 Screenshot of the "Timeseries Viewer" application through which uploaded data can be viewed in the GEMINI digital twin.

Conclusions

This deliverable provides a comprehensive representative dataset designed to support users of the GEMINI digital twin in exploring and testing its functionalities under normal and anomalous operating conditions. By simulating a primary-side geothermal doublet and generating six months of synthetic yet realistic operational data, the dataset offers an accessible option for those interested in the digital twin to interact with the GEMINI environment even in the absence of real operational measurements.

The dataset incorporates stochastic variations in key variables to emulate natural operational fluctuations, ensuring that the resulting time-series resemble more natural fluctuating data rather than constant, unchanging values. In addition, two types of degradation and fault scenarios were simulated: heat-exchanger fouling, and filter clogging. This enables users to investigate characteristic fault signatures, test detection strategies, and evaluate predictive maintenance workflows within GEMINI.

Beyond data generation, the document outlines the workflow for importing, mapping, and visualizing the dataset within the GEMINI digital twin, including the use of the Data Management interface, the Diagram Builder, and the Timeseries Viewer. This ensures that new users can readily integrate the dataset into GEMINI's monitoring and modelling environment and start using it.

Together, the representative dataset and accompanying instructions form a reusable and extensible template, lowering the barrier of entry for organizations and researchers interested in geothermal digital twin technology. They provide a robust basis for building familiarity with GEMINI's capabilities.

References

Shoeibi Omrani, Pejman, Paul J.P. Egberts, Huub H.M. Rijnaarts, and Shahab Shariat Torbaghan. 2026. "Geothermal plant operation and control under demand uncertainties." *Renewable Energy* 257.