PROCEEDINGS, 48th Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 6-8, 2023
SGP-TR-224

Deploying Digital Twins for Geothermal Operations with the GOOML Framework

Iraklis Konstantopoulos[1], Paul Siratovich[1], Grant Buster[2], Nicole Taverna[2], Jon Weers[2], Andrea Blair[1], Jay Huggins[2], Christine Siega[3], Warren Mannington[3], Alex Urgel[3], Johnathan Cen[3], Jaime Quinao[4], Robbie Watt[4], John Akerley[5]

[1]Upflow NZ, PO Box 61, Taupō 3330, New Zealand
[2]National Renewable Energy Laboratory, USA
[3]Contact Enegy Ltd, New Zealand
[4]Ngati Tuwharetoa Geothermal Assets Ltd, New Zealand
[5]Ormat Technologies, Inc., USA

hercules@upflow.nz

Keywords: field optimization, digital twins, plant operations, algorithms, machine learning, Wairakei, strategy, modelling

We present an overview of, and applications with GOOML, a geothermal operations optimisation framework based on machine learning and components modelling under constraints informed by physics. The framework was developed using real-world operations, with data from fields of various types (e.g., brine and dry steam) and configurations (e.g., single plant, binary, direct heat) to develop digital twins. GOOML aims to increase the output of steam fields all over the world by allowing operators to run myriad scenarios simulating changes to their fields. In doing so we remove the need to experiment in the physical world, therefore foregoing significant costs associated with developing new operational strategies. During the research phase, GOOML has suggested strategies that can bring about increases in electricity generation of order 1%. In the next phase, we hope to deploy this system at steamfields around the world.

1. INTRODUCTION

It may be an understatement to describe real-world geothermal operations as complex. While the mathematics involved in understanding and characterising individual components are not particularly onerous, it is the complexity of fields and the inability to comprehensively measure conditions at every point within them that makes optimising them so challenging.

Consider, for example, the Wairakei steam field (located in New Zealand), its decades of development and multiple end-uses, covering multiple electricity generation stations and various industrial heat applications. Without a digital framework, manually coming up with a bird's eye view—or even a snapshot of operations—for such a system is a daunting task. Any of the dozens of wells can be readily characterised, as can any one of the stations. But the interplay of connections between wells, stations, and all the machinery between them is too complex for a human to compute on the fly.

Historically, engineers have relied on 'historians' (database tools) and spreadsheets to plan and optimise operations. Extricating time-series data from historians can be a time-consuming task and engineers often do not have access to technical staff in information technology or database administration who could automate data flows and quality assurance. In extreme cases this forces engineers to repeat the data extraction process for every analysis.

With the data extracted and prepared, spreadsheets are typically employed to analyse data for tasks such as shut planning, strategy, development, and optimisation. Spreadsheet applications are not designed for highly dimensional tasks like these and tend to multiply the repetition factor. In data operations, repetition is a massive issue, as it opens scope for human error in data entry and greatly slows down processes; as such it is generally avoided. The software ('tech') sector, in recent times, has seen an explosion of investment in the infrastructure and practitioners of data engineering and data science. In the energy sector, we have not kept up with this trend, owing partly to antiquated information systems that disallow the interoperability of data and also to the culture of outsourcing tasks that generally fall under the umbrella of "IT" (which reinforces the skills gap for automation tasks).

In past proceedings our team has presented GOOML (Buster et al., 2021a, Taverna et al., 2022), a framework that aims to automate all the tedious parts of this process and therefore allow engineers to work on tasks that require human brains—that is, decision making. GOOML is not designed to provide a theoretical framework, but a workbench that helps operators make concrete decisions about how to run their field. In this paper we present the current state of GOOML, which is transitioning from research to commercial application.

The framework aims to optimise an entire steam field in a single simulation environment. It represents a technological leap that allows operations staff to move away from time-consuming and laborious methods. By computing at a scale and complexity not achievable by a human operator, GOOML delegates repetitive tasks to a smart machine; what would take months of work now takes mere minutes. It also moves analysis and prediction away from the current deterministic setup and into the stochastic domain.

In the following section we will describe the team's approach in developing a codebase and assorted tools alongside field and plant operators to ensure we build something useful in the real world. In Section 3 we will provide some examples based on a hypothetical steam field developed as a GOOML demonstration. And in Section 4 we will present our plans for bringing this extraordinary capability to operations desks around the world.

2. DEVELOPMENT

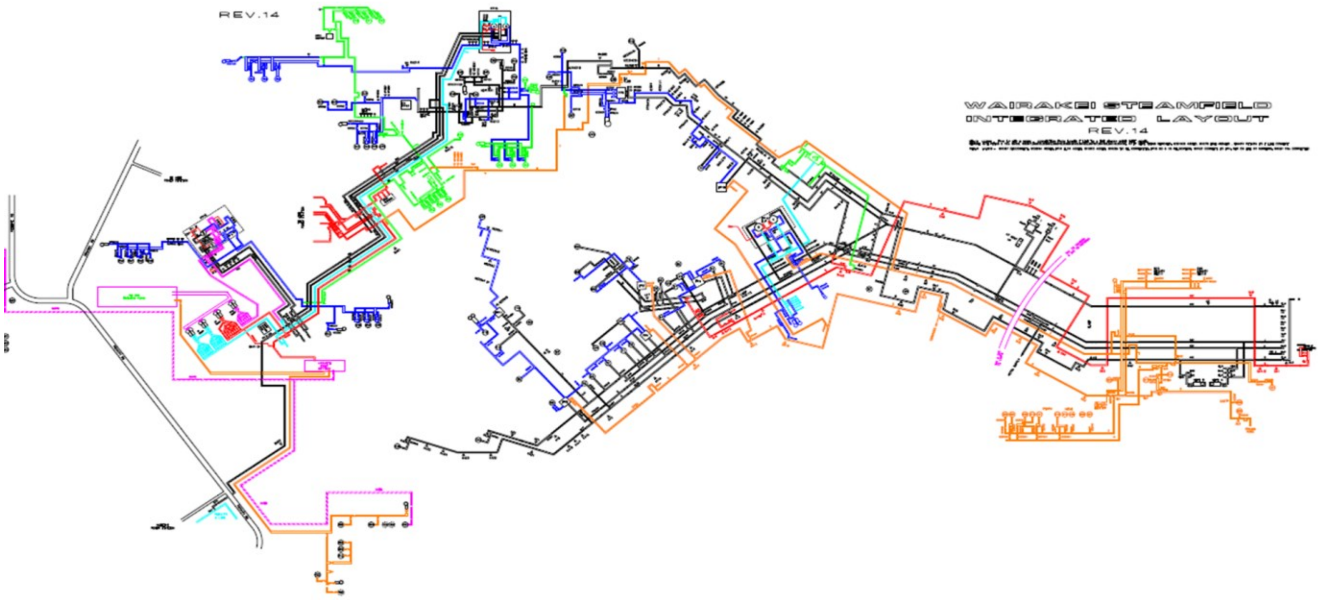

Figure 1: The layout of the Wairakei (Contact Energy) steam field, the basis on which we have developed the GOOML digital framework. This diagram illustrates the complexity of a steam field with dozens of wells, kilometres of pipelines, and several power generating stations.

The development of this project relied heavily on the interaction between lab researchers (physicists and programmers) and power plant staff (engineers and operators). This collaborative, iterative approach was aimed first at understanding the limitations of current processes employed around the world. This investigation revealed a grounding reality check: fields are nowhere near perfectly instrumented, either because their construction predates the digital age, or because extensive instrumentation has not been seen to return on the investment by the companies that operate geothermal fields and power plants.

Before we even get to analysis, we therefore have to address the lack of data and the variety of data quality issues. The foundation of this whole project is, in fact, the data curation suite (Taverna et al., 2022) that aims at cleaning data and preparing time-series and configuration tables for ingestion. This data suite is configured and specialised for the data pipelines present at each steam field. The end-product is a dataset adapted to a shared GOOML data standard, which can be readily ingested by the code.

With clean data at hand, we were able to focus on modelling. The component-based systems model that we describe in detail in Buster et al. (2021a) took apart the whole steam field and focussed on one subsystem (or ”component” in software language) at a time. Without getting lost in whole-system reconnaissance we were able to characterise the properties of mechanical components and encode them using models of varying complexity. For example, wells are treated through regression models (simple statistical learning), but flash plants, turbine generators, and binary plants are modelled with neural networks (deep learning).

With each type of machinery characterised individually using real-world data, we can synthesise an overarching system model. Where this approach extends previous efforts in this arena is in eschewing the theoretical approach of idealised components and instead hybridising *data-driven* thermodynamics with a component-based systems model. That is, machine learning is only applied once we have set guardrails through thermodynamics and data cleaning.

2.1 Experimental setup: the three test fields

We partnered with three geothermal plant operators to establish the research project: Wairakei (Contact Energy Limited), Kawerau (Ngati Tuwharetoa Geothermal Assets, NTGA), and McGinness Hills (Ormat Technologies Inc.). Each of these steam fields has unique attributes; training and validating models on a broad base of features helps generalise and avoid statistical biases by overfitting on certain features.

The current Wairakei (shorthand: WRK) field is the outcome of many generations of power technology, which yields a very complex and interconnected steam field. Multiple connections allow fluid sharing between generation using flash technology at two separate Wairakei Stations (WRK A and B) and Te Mihi Station, the Poihipi Road dry steam plant, and a bottoming Ormat binary unit that processes brine, separated from the Wairakei steam field wells. Fluid is currently re-injected into the reservoir and some fluid is disposed of into the Waikato river. Similarly, the condensing cycle for Wairakei A&B uses the river for cooling and has a significant bioreactor installed to help reduce the amount of sulphur flowing into the river. The multiple generations of technology at Wairakei provide an ideal (though complicated) testbed for mapping optimisation solutions. This, in particular, makes the steam field a good candidate for exploration of digital optimisation.

McGinness Hills (shorthand: MGH) is relatively simple in nature, featuring single-phase pumped wells that feed dedicated organic Rankine cycle (ORC) units. There is no fluid sharing among the three units (MGH 1–3) such that an optimisation exploration would focus on brine and heat delivery.

Kawerau (shorthand: KAW) is a multi-faceted operation that delivers steam to industrial users as well as feeding electricity generation. Finding the balance between steam provision and power production is a key tenet of the operation. Similar to WRK, KAW is built on multiple decades of infrastructure, such that instrumentation and data fidelity is as fit for purpose as was available and financially feasible at the time of commissioning. Some wells at the complex are decades old and newer wells have only been brought into service since 2020. This varying age and mix of end-uses for fuel makes for a highly interconnected system with many pressure-fields that pose serious challenges to building an idealised digital twin.

2.2 Innovations in data handling to enable machine learning

Two decisions that contributed greatly to the success of this project were: to ensure adherence to modern and rigorous techniques of data cleansing, gap-filling and quality assurance (a.k.a. wrangling) and, a design that allowed us to apply machine learning (ML)—an obvious course since the research project was funded as an ML application for Geothermal operations.

The underlying code is written entirely in Python, a very widely used language and one with a relatively low barrier to entry. Even so, the intended user does not interact with (nor are they required to be entirely comfortable with) pure code. Instead, we developed two frameworks to set up and interact with a digital twin.

Real world data are typically not ready for analysis. Before we can even begin to set up the digital twin, we need to curate data. Toward this end we built a data standard and pipeline, described in full in Taverna et al. (2022). Through a series of generic and field-specific translators and cleaners, this pipeline brings real-world data into a state of conformity. Picking up from this point on, the main GOOML code can focus on physics and statistics, rather than data fixes and editing.

After the data are prepared, a steam field is configured through a plain-text interface, in the JavaScript Object Notation (JSON) standard. In this step we take the engineering diagram of a field (e.g., the WRK setup shown in Fig. 1) and map everything into a JSON 'component'. For example, a well carrying steam and brine named W241 will feature as a component named W241 of type TwoPhaseWell. We can now either provide well equations in this single JSON configuration file, or store the equation in a separate file (for book keeping) and simply 'point' at it (that is, define its location on the filesystem). We also specify the series of connections between components, and map components to their associated data streams. The latter is done in separate configuration files since different input data is used for different modelling purposes (i.e., historical versus forecast modelling). This system of field configuration aims at capturing all the complexity while making it accessible for both human and machine perusal.

The following code block shows an example JSON configuration file for a hypothetical steam field:

```
{
  "components": {
    "w1": {
      "eqns": "w1.json",
      "type": "TwoPhaseWell"
    },
    "w2": {
      "eqns": "w2.json",
      "type": "TwoPhaseWell"
    },
    "vlv_w1": {
      "type": "valve"
    },
    "vlv_w2": {
      "type": "valve"
    },
    "join_fp1": {
      "type": "join_junction"
    },
    "fp1": {
      "type": "flashplant",
      "fn_dims": "fp1.json"
    },
    "silencer1": {
```

```
        "type": "SilencerJunction"
      }
      "tg1": {
        "type": "TurbineGenerator",
        "tf_model": null
      },
      "join_tg": {
        "type": "join_junction"
      },
      "vlv_tg1": {
        "type": "Valve"
      }
    },
    "connections": {
      "w1": "vlv_w1",
      "w2": "vlv_w2",
      "vlv_w1": "join_fp1",
      "vlv_w2": "join_fp1",
      "join_fp1": "fp1",
      "fp1.vapor_out": "join_tg",
      "fp1.liquid_out": "silencer1",
      "vlv_tg1": "tg1"
    },
    "data": null
}
```


The second framework we developed was a set of ‘drivers’ in the Jupyter interactive environment for the user to leverage the field configuration and the codebase. This enables our ideal users or testers to interact with GOOML without ever having to write a line of Production Python code, or writing complex engineering processes to clean the data.

The overarching rationale is: Physics and Thermodynamics lead, Statistics and ML execute, and the user never has to write code, just interact with the existing source code. With the complement of code, configuration, and driver scripts forming a digital twin, we empower the end user to model and explore a steam field. By automating the preparation phases we obfuscate operationalising the digital twin, so that users can jump straight to getting insights from the data.

2.3 From steam field setup to software design

By validating results with site and optimisation engineers, we endeavour to imbue the code with their trust. The outcome is a digital workbench for risk-free experimentation, an environment where engineers can focus on ‘what-if’ scenarios without spending hours wrangling and cleaning data.

At the culmination of the research phase, GOOML has suggested actionable strategies at WRK to boost electricity generation. First it suggested an ambitious strategy that would add 18 MWe, but which exceeded the threshold in daily mass take set by the environmental consent. Following on this, and with the consent limit explicitly added, GOOML suggested a new strategy to boost generation by 2 MWe, representing a valuable 1% gain in total field output.

GOOML suggested this strategy in the risk-free ML workbench without requiring any additional machinery, just the existing plant hardware and machinery. This stands to bring a considerable efficiency gain with no capital expenditure and modest staff time to execute. In the future, commercial phase real-world tests and feedback will prove some of these AI-suggested gains.

3. CURRENT STATE EXAMPLES

In the previous section we described the JSON architecture for encoding the physical setup of a field as a software configuration. In Taverna et al. (2022) we describe the suite of data curation tools that bring data to a common GOOML standard.

To demonstrate the capability of the code and the current interface without revealing proprietary information from real operating power plants, we prepared a hypothetical steam field named *Kahunanui* (shorthand: KHN), which covers most features of the software. It draws from five two-phase wells to feed mass through three flash plants through a complex network, which ultimately feeds four turbine generators. Three of the generators form an ‘A’ station along with a binary plant, and the fourth generator forms a ‘B’ station on its own. We also added a cross junction that allows the operator to divert mass from one side of the steam field to the other (A to B).

Figure 2 illustrates the KHN field configuration through our pressure solver: this provides a point-in-time solution of the model (in essence, a prediction) of well-head pressures and mass flows throughout the system. Mass flow and pressure are just two of the measurements this GOOML digital twin can provide in such an illustration.

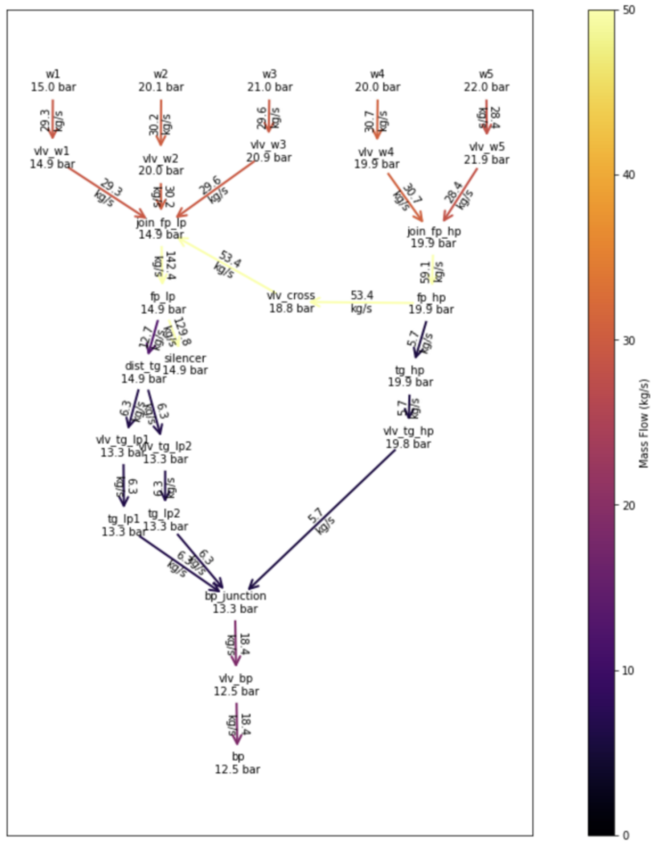

Figure 2: A point-in-time solution for the pressure model for the synthetic Kahunanui steam field. This digital twin allows us to make changes to a single or multiple components and predict the effect on the rest of the field. We apply a combination of thermodynamic principles and statistical methodology to produce this result.

The use case for the pressure solver is for a field operator—a site or optimisation engineer, a development manager, an executive, or anyone—to run a suite of scenarios. What if we have to lower the pressure in w3? How do we plan for an outage of w5? What if we divert all available mass to the A station?

Naturally, getting to this point does require setup:

1. First we process data feeds. In the case of KHN these are provided as synthetic time series, being realistic data with added noise. In a real-world scenario we would work with data, IT, and engineering teams to obtain time series, and then map and remediate data quality issues.

2. Then we need to configure the steam field. Starting with the engineering diagram we would set up the JSON configuration, one component at a time, and map their interconnections. KHN is imagined into a JSON config, but a real plant might require people deployed on the actual field recording physical tags on well heads.

3. Finally, we train component models. We train the simple regression and deep learning models described in the previous section through Jupyter notebooks. This makes a fully connected, interactive, physics-informed environment available, where we can manipulate any and all parameters.

These three steps complete the digital twin at a point in time; no change needs to be made unless a component changes in the field. Periodic re-training of models is important in the geothermal context, where mineral deposits in pipes change performance through mass-flow and enthalpy changes, but the digital twin will be usable for as long as the field remains physically the same.

In the case where the field does change, GOOML makes it simple to update the digital twin. We alter the JSON configuration to reflect changes in components, connection, or data streams. For example, assume that well *w1* in the hypothetical configuration presented above were to be permanently shut down on 24 March 2023; and that, after discovering a mistake, liquid from *fp1* were to retroactively be piped to an imagined plywood factory named *timber world*, rather than a silencer. With + and - denoting additions and removals from the configuration (mimicking git ‘diff’ notation), we have:

```
+ "timber_world": {
+     "type": "join_junction"
+ }

+ "w1_layout_change": {
+     "change_dates": [
+         "20230324"
+     ],
+     "type": "PlantLayoutChange"
+ }

- "fp1.liquid_out": "silencer1",
+ "fp1.liquid_out": "timber_world",
```

Notice that we did not altogether remove well w1 from the configuration. Instead we used the plant layout feature so that we can still represent w1 in historical data. After this layout change the data feed would set the well’s ‘condition’ to shut. With this new config, we can simply re-initialise the system to start training new models.

Note also that we configure timber world as a join_junction because GOOML is not currently integrated with direct use applications, and instead sees them as junctions that mass flows into.

Once we are comfortable that we have appropriately captured the data trends, the thermodynamic performance of the components and the physical constraints of the system, we are free to experiment in our digital system. As we have set up the system to be changeable and trainable, when the time comes to retrain, we simply amend the JSON config with new decline rates, enthalpy trends or other pertinent information.

4. FUTURE

With the research phase of GOOML complete, we are building a software package to take this codebase from a lab facility to a ubiquitous digital experimental apparatus. While it was developed using data from real sites, GOOML is not exclusively tailored to WRK, MGH, or KAW; it can be deployed to any geothermal site in the world.

This is the goal of the commercial phase that is now beginning. Working alongside the engineering teams with our commercial partners is what made the research phase a success, so we plan to work with a development partner (or partners) in order to maintain this strong link to plant operations. We are now conducting user interviews so we can shape a minimum viable product.

The digital twin approach we have taken with GOOML offers a valuable new mode of operation for the geothermal industry. It allows operators to run countless ‘what if’ scenarios entirely in the digital realm, without the need for risky and costly physical interventions. The scope of this software is wide reaching, especially given that it can run on a regular laptop or desktop machine—no need for supercomputers or cloud infrastructure.

In the immediate term, we plan to make the existing set of features available through a software suite and a graphical user interface. We will continue to follow the principle of co-development with engineers to stay close to our intended user base. Our aim is to provide a solution that is configured once and then just works time after time with only minor reconfiguration required when a plant’s layout changes.

Once we have a product available we can focus on adding some features requested by members of our user network, e.g., the impact on pressure and mass flow of elevation changes.

Meanwhile, we will work toward making the code that underpins this research openly available. Along those lines, we have posted the data that comprise our Kahunanui functional steamfield to the Geothermal Data Repository (Taverna et al., 2023; also see Buster et al., 2021b for an earlier synthetic field configuration named Big Kahuna).

5. SUMMARY

We have provided an overview of the development, deployment, current and future state of GOOML, a framework for geothermal optimisation using machine learning, thermodynamics, and component-based systems modelling.

GOOML is the result of an international, public-private collaboration between Upflow New Zealand, the US National Renewable Energy Lab, and three commercial partners in the two countries: Contact Energy Limited, Ngati Tuwharetoa Geothermal Assets

Limited, and Ormat Technologies, Inc. We hope to see this level of international exchange in research, operational know-how, and funding continue in the already closely-knit geothermal community.

The project is crossing from the research to the commercialisation phase. We plan to build a simple-to-use software product for anyone on the Geothermal planning and operations chain. Just as the research phase, exemplified by Buster et al. (2021a) and Taverna et al. (2022), brought to life an amazing facility for lab work, we hope future software development will bring GOOML to the operations room.

ACKNOWLEDGEMENTS

We express our immense gratitude to our commercial partners, Contact Energy Limited, Ngati Tuwharetoa Geothermal Assets Limited, and Ormat Technologies, Inc., for their support of this project. Without their data, staff, and institutional backing, this project would have not been possible.

This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Geothermal Technology Office Award Number DE-EE0008766. This work was authored by Upflow, Ltd. and the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308 with funding provided by U.S. Department of Energy Office of Energy Efficiency and Renewable Energy (EERE) Geothermal Technologies Office (GTO). The views expressed in the article do not necessarily represent the views of the DOE or the United States Government. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid- up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for United States Government purposes.

REFERENCES

Buster, G., Siratovich, P., Taverna, N., Rossol, M., Weers, J., Blair, A., Huggins, J., Siega, C., Mannington, W., Urgel, A., Cen, J., Quinao, J., Watt, R., and Akerley, J. A New Modeling Framework for Geothermal Operational Optimization with Machine Learning (GOOML). Energies (2021), 14, 6852.

Buster, G., Taverna, N., Rossol, M., Weers, J., Siratovich, P., Blair, A., and Huggins, J. 2021. "GOOML Big Kahuna Forecast Modeling and Genetic Optimization Files". United States. https://dx.doi.org/10.15121/1812319. https://gdr.openei.org/submissions/1314.

Taverna, N., Buster, G., Konstantopoulos, I., Weers, J., Siratovich, P., and Blair, A. 2023, submitted. "GOOML Kahunanui Example Jupyter Notebooks". United States. https://gdr.openei.org/submissions/1466.

Taverna, N., Buster, G., Huggins, J., Rossol, M., Siratovich, P., Weers, J., Blair, A., Siega, C., Mannington, W., Urgel, A., Cen, J., Quinao, J., Watt, R., and Akerley, J. Data Curation for Machine Learning Applied to Geothermal Power Plant Operational Data for GOOML: Geothermal Operational Optimization with Machine Learning. Proceedings 47th Stanford Geothermal Workshop. United States