PROCEEDINGS, 51st Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 9-11, 2026
SGP-TR-230

Neural Network–Based Short Term Forecasting of Water Levels in a Low-Temperature Geothermal Field

Halldora Gudmundsdottir[1,2] and Roland N. Horne[1]

Department of Energy Science & Engineering, Stanford University, Stanford, California, USA[1]

Reykjavik Energy, Reykjavik, Iceland[2]

halldora@stanford.edu

Keywords: low-temperature, field study, water level, time series analysis, neural networks

ABSTRACT

Low-temperature geothermal fields are a critical component of district heating systems in Iceland, where short term fluctuations in heat demand and reservoir conditions can pose operational challenges. In such systems, the ability to reliably predict short term water level behavior in production wells is essential for ensuring secure heat delivery, particularly during periods of extreme weather.

This study explores neural network–based time series models for short term prediction of water levels in a low-temperature geothermal reservoir. Using operational data from the Laugaland geothermal field in Iceland, multilayer perceptron (MLP) models are developed to predict water level dynamics based on production and injection rates. Model performance is evaluated under direct and recursive forecasting scenarios, with emphasis on robustness, uncertainty, and operational relevance.

The results show that MLP models incorporating lagged inputs and target feedback accurately reproduce short term water level behavior and provide stable predictions across multiple initializations. Recursive predictions reveal increasing uncertainty with longer predictions horizons, and a scenario-based application illustrates how the models can be used to assess the risk of water levels approaching critical pump depth under sustained high-demand conditions. While not intended as a replacement for physics-based reservoir models, the findings suggest that neural network time series models can provide practical and computationally efficient decision support for short term management of low-temperature geothermal district heating systems.

1. INTRODUCTION

Low-temperature geothermal resources play a critical role in Iceland's district heating systems and are a cornerstone of the country's energy infrastructure. These systems offer reliable, low-emission energy, but their long-term sustainability depends on careful reservoir management and the ability to respond to operational uncertainty.

Developing and maintaining low-temperature geothermal fields presents several challenges. Unlike high-temperature reservoirs, these systems often exhibit lower productivity and greater sensitivity to pressure drawdown. Successful drilling is not guaranteed, even after extensive exploration, and additional production wells are frequently required to meet growing demand or compensate for declining performance. Moreover, short-term demand for geothermal heat is strongly influenced by weather conditions, which cannot be anticipated with certainty. Harsh winters can lead to extreme production demands, while mild winters may mask underlying reservoir stress. These fluctuations place operational strain on district heating systems and increase the risk of water levels approaching pump depth in production wells.

The Laugaland low-temperature geothermal field in Iceland exemplifies these challenges. Despite efforts to drill additional successful wells, production capacity remains limited, and during cold winters the water level in key wells can approach critical operational thresholds. In such cases, the ability to predict short-term reservoir behavior, particularly water level changes, is essential. Reliable predictions allow operators to anticipate worst case scenarios, estimate deliverable heat to consumers, and determine when contingency measures must be activated. These predictions are also valuable for utility companies managing multiple district heating systems and infrastructure projects, where informed prioritization of resources is necessary.

Conventional reservoir modeling approaches can provide valuable insights into long-term behavior but often face limitations when applied to short-term operational forecasting. These methods typically require detailed geological characterization and calibration, which may be impractical for rapid decision-making or in data-limited low-temperature fields. As a result, there is growing interest in alternative, data-driven approaches that can complement traditional modeling by offering flexible, fast, and computationally efficient predictions.

In this study, we investigate the use of data-driven models to predict short term water level behavior in a low-temperature geothermal field, with a focus on operational relevance rather than detailed physical interpretation. The primary objective is to assess whether such models can provide reliable short term forecasts that support decision making under uncertain demand conditions. We present results from the Laugaland field and demonstrate how these predictions can be used to evaluate risk, inform operational planning, and support sustainable management of district heating systems. While the findings are not intended to be definitive, they highlight the potential of data-driven approaches as practical tools for short term geothermal reservoir management and outline directions for ongoing and future work.

2. METHODOLOGY

In this paper, we investigate neural network-based time series prediction models for predicting short term water level behavior in a low temperature geothermal field. Multilayer perceptron models are developed to map inputs, production and injection rates, to water level response in a production well. All models are trained on historical operational data and evaluated on held-out test sets using standard regression metrics. Both direct one-step-ahead and recursive predictions are conducted, and model robustness is evaluated through multiple independent training runs with different initializations.

2.1 Learning-based Prediction Methods

In its simplest form a neural network, often referred to as a Multilayer Perceptron (MLP), consists of an input layer, a hidden layer and an output layer. A network is typically classified as a deep learning model when it contains more than one hidden layer (Goodfellow, Bengio, & Courville, 2016). Figure 1 illustrates a feedforward neural network with a single hidden layer, which is denoted here as a two-layer network, since the input layer is not counted.

Each layer comprises neurons that perform linear transformations followed by nonlinear activation functions. Most neural networks are fully connected, meaning that each neuron is connected to all neurons in adjacent layers, while no connections exist between neurons within the same layer. The influence between connected neurons is governed by weights, which may be positive (amplifying) or negative (dampening), with larger magnitudes indicating stronger influence.

For the network shown in Figure 1a, the mapping from an input vector $\boldsymbol{x} = [x_1, x_2, x_3]$ to an output vector $\hat{\boldsymbol{y}} = [\hat{y}_1, \hat{y}_2]$ can be expressed as

$$\boldsymbol{z}^{[1]} = \boldsymbol{W}^{[1]}\boldsymbol{x} + \boldsymbol{b}^{[1]}, \tag{1}$$

$$\boldsymbol{a}^{[1]} = g^{[1]}\left(\boldsymbol{z}^{[1]}\right), \tag{2}$$

$$\boldsymbol{z}^{[2]} = \boldsymbol{W}^{[2]}\boldsymbol{a}^{[1]} + \boldsymbol{b}^{[2]}, \tag{3}$$

$$\boldsymbol{a}^{[2]} = g^{[2]}\left(\boldsymbol{z}^{[2]}\right) = \hat{\boldsymbol{y}}, \tag{4}$$

where $\boldsymbol{W}^{[l]}$ and $\boldsymbol{b}^{[l]}$ denote the weight matrix and bias vector of layer l, $g^{[l]}$ corresponds to the activation function. Training the neural network involves optimizing the parameters $\boldsymbol{W}^{[l]}$ and $\boldsymbol{b}^{[l]}$ to minimize a cost function that quantifies the discrepancy between predicted and observed outputs across all training samples. This optimization is typically performed using iterative gradient-based methods.

In this work, neural networks are used to map time series of injection flow rates to time series of production data. To account for temporal dependencies inherent in time-series data, we augment the network inputs with time-lagged variables and target feedback. Target feedback refers to including previous output values as part of the input to the model. For example, when using one time lag and target feedback, the input vector at time t can be written as $\boldsymbol{x} = [x_1(t-1), x_1(t), x_2(t-1), x_2(t), x_3(t-1), x_3(t), y_1(t-1), y_2(t-1)]$ with the corresponding output $\hat{\boldsymbol{y}} = [\hat{y}_1(t-1), \hat{y}_2(t-1)]$. This network structure is illustrated schematically in Figure 2.

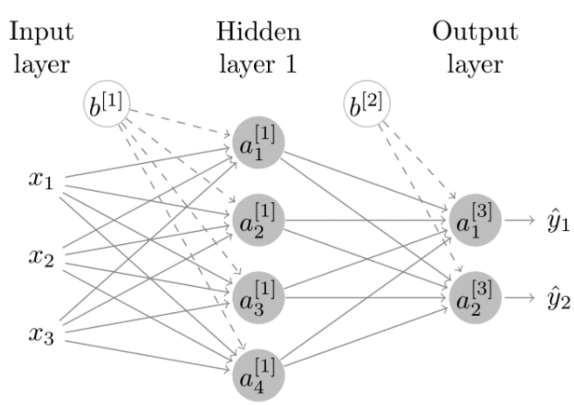

Figure 1: A 2-layer neural network with one hidden layer. Nodes represent neurons and arrows the connection between neurons.

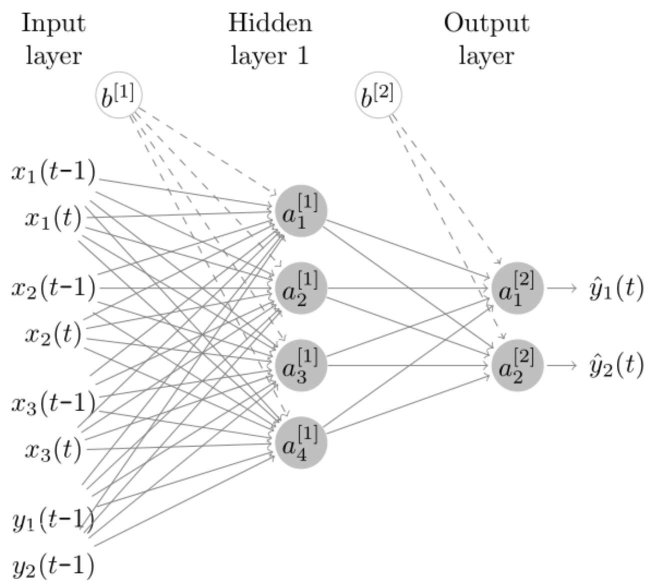

Figure 2: A 2-layer neural network with one hidden layer. The network models time series data, with one time lag and target feedback added to the input.

2.2 Performance Metrics

Prediction accuracy is evaluated on a held-out test set using standard regression metrics. We report the root mean squared error (RMSE), which is scale-dependent and penalizes large errors, the weighted absolute percentage error (WAPE), which provides a unit-free measure for comparison across operating conditions, and the coefficient of determination R^2, which measures the proportion of variance in the observations explained by the model (Dodge, 2008):

$$RSME = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y^{(i)} - \hat{y}^{(i)})^2}, \tag{5}$$

$$WAPE = \frac{\sum_{i=1}^{m}|y^{(i)} - \hat{y}^{(i)}|}{\sum_{i=1}^{m}|y^{(i)}|}, \tag{6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{m}\left(y^{(i)} - \hat{y}^{(i)}\right)^2}{\sum_{i=1}^{m}(y^{(i)} - \bar{y}^{(i)})^2}, \tag{7}$$

3. LAUGALAND CASE STUDY

3.1 Data Description

This study focuses on the Laugaland geothermal field within the Rangárveita district heating system in southern Iceland. The Rangárveita district heating system combines geothermal water from the Laugaland and Kaldárholt fields to supply heat to the service area. Laugaland is a relatively small, low-permeability reservoir where production induced drawdown is pronounced, and operational limits are controlled by water level relative to pump intake.

The Laugaland field includes two closely spaced production wells, LL-04 and LL-06, and one injection well, GN-01 (Figure 3). LL-04 and LL-06 are separated by approximately 35 m, while GN-01 is located about 200 m from LL-04, indicating a strong hydraulic connection within the main productive zone. Production from LL-04 began in the early 1980s, and sustained drawdown associated with increasing heating demand led to the initiation of injection into GN-01 around 2000 to support pressure recovery. Continued demand growth and limitations in production capacity from LL-04 motivated the drilling of LL-06 in 2017 (Kristinsson 2023 and Oskarsson et al., 2020).

The modeling objective is to predict the produced water level in well LL-06 as a function of, production from LL-06, production from the hydraulically connected well LL-04, and injection into GN-01 (input series shown in Figure 4). These variables represent the dominant operational controls on reservoir pressure in the Laugaland system.

The analysis uses daily-averaged operational data covering the period from the beginning of 2020 through the end of 2025. The dataset includes production and injection rates and corresponding water level measurements. Measurement uncertainty and occasional data gaps are present, but the dataset captures both short-term operational responses and longer-term reservoir behavior, making it suitable for data-driven prediction of water level dynamics in well LL-06. The dataset was divided chronologically into training, validation, and test subsets using a 60/20/20 split.

Figure 3: Close-up view of the Laugaland geothermal field showing the spatial arrangement of the production wells LL-04 and LL-06 and the injection well GN-01, illustrating their close hydraulic proximity (courtesy of Reykjavík Energy).

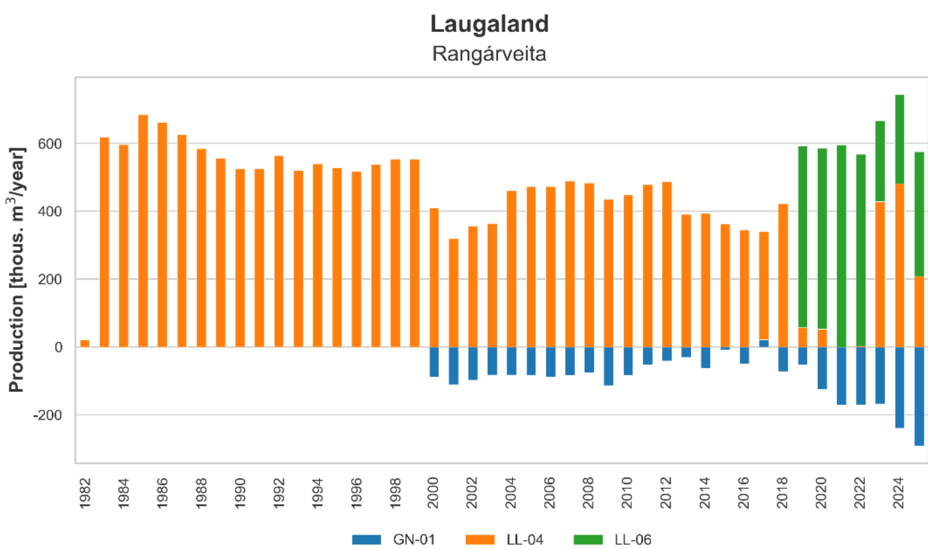

Figure 4: Total annual production and injection from LL-04, LL-06, and GN-01 (courtesy of Reykjavík Energy).

3.2 Predictions Using Neural Networks

This section evaluates the performance of multilayer perceptron (MLP) models under three prediction scenarios: (i) direct one-step-ahead prediction without target feedback, (ii) direct one-step-ahead prediction with target feedback, and (iii) recursive multistep prediction. Models are trained on the training part of the time series and evaluated on a held-out test set, as described in Section 1.3. A validation split is used for coarse hyperparameter tuning.

Hyperparameter tuning is intentionally kept simple and limited to a small grid search over number of hidden layers, neurons, and input lags. The objective is not to identify the optimal architecture, but rather to assess the sensitivity of prediction performance to architectural and temporal design choices.

Neural networks are inherently stochastic, and the model performances vary depending on the initialization of the network weights. To assess the robustness of each MLP configuration, all models are trained and evaluated on multiple independent random initializations. Performance metrics are summarized using P10, P50, and P90 statistics.

3.2.1 MLP without target feedback

Figure 5 shows one-step ahead predictions on the training and test set produced by a multiplayer perceptron (MLP) model with five hidden layers and 50 neuron per layer. The model uses two lagged time steps of the input variables as additional inputs, while no target feedback from the output is included.

The MLP model captures the overall temporal structure of the water level time series and reproduces the cyclic behavior observed in the data. Specifically, the timing of the major peaks and drawdowns is generally well aligned with the observations, indicating that network successfully learns the temporal dependencies from the lagged inputs alone. However, some differences are observed during periods of rapid change and extreme values, where the model tends to underestimate the magnitude of deep drawdowns and high peaks.

Table 1 summarizes the one-step-ahead test performance across multiple random initializations using P10, P50, and P90 statistics. The relatively wide spread across initializations indicates moderate predictive skill but limited robustness for this configuration.

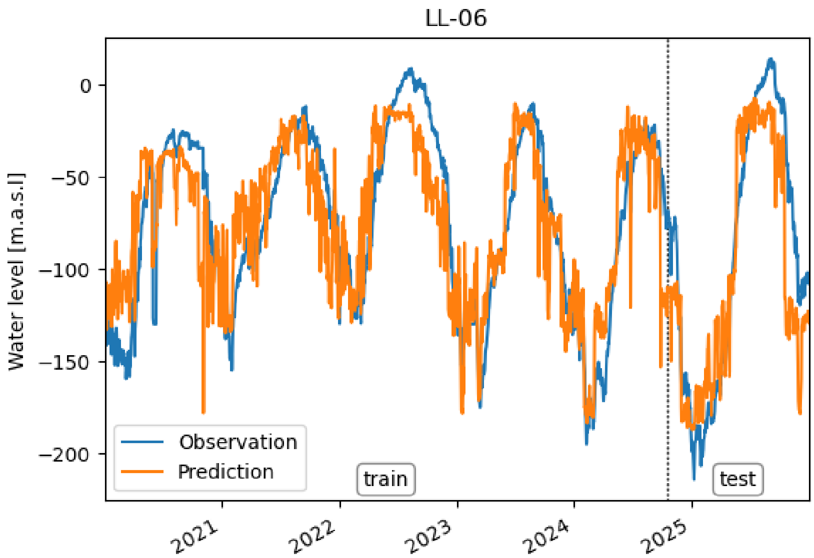

Figure 5: One-step-ahead predictions for the water level in well LL-06 produced by an MLP with five hidden layers and 50 neurons per layer, using two lagged input time steps and no target feedback. Observed and predicted water levels are shown for the training and test periods, separated by the vertical dashed line.

3.2.2 MLP with target feedback

Figure 6 presents one-step-ahead predictions produced by an MLP with three hidden layers and 50 neurons per layer, using two lagged input time steps and incorporating target feedback. As in the previous configuration, predictions are generated non-recursively, but the inclusion of previous output values provides additional temporal context.

Compared to the MLP without target feedback, this configuration gives a greatly improved fit to the observed data. The predicted water level closely tracks both the timing and the magnitudes of peaks and drawdowns. These qualitative improvements are represented quantitatively in Table 1, where MLP with target feedback achieves near perfect median performance and a very narrow P10-P90 range across initializations. This indicates both substantially improved accuracy and robustness. Despite using fewer hidden layers, the model outperforms the deeper network, highlighting that the architectural choices related to temporal information could be more influential than increased network depth.

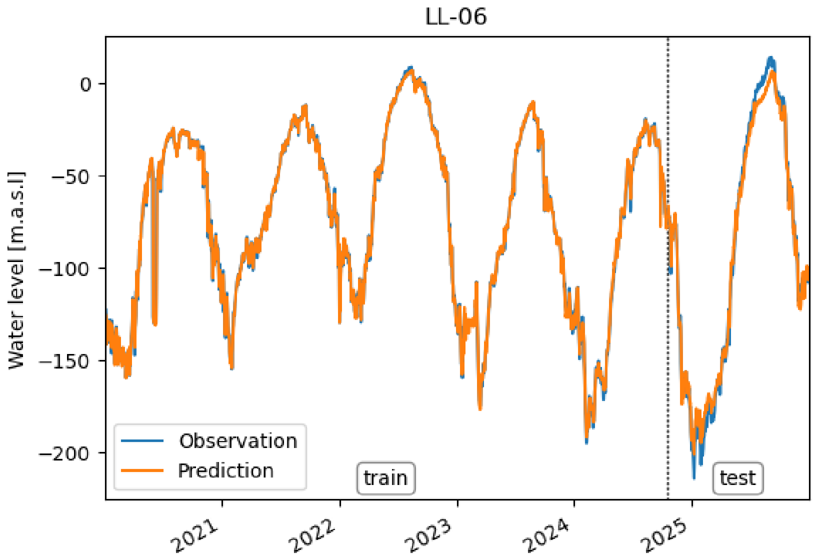

Figure 6: One-step-ahead predictions for the water level in well LL-06 produced by an MLP with three hidden layers and 50 neurons per layer, using two lagged input time steps and incorporating target feedback. The vertical dashed line indicates the separation between training and test data.

<u>3.2.3 Recursive MLP predictions</u>

Figure 7 shows recursive multistep predictions on the test set generated using the same MLP configuration as in the previous experiment, but with predictions iteratively fed back as inputs to generate predictions beyond the first time step. Recursive predictions give a more realistic view on the performance skill of the model, as practical use requires predictions to be generated without access to future observations.

While the model continues to reproduce the cyclic structure of the water level time series, prediction accuracy decreases as the prediction horizon increases. Deviations between predicted and observed values grow over time, and peak magnitudes and drawdowns are underestimated.

Table 2 summarizes recursive prediction performance for different horizons using P10, P50, and P90 statistics. As the prediction horizon increases, median performance generally decreases and variability across initializations increases. This behavior is characteristic of recursive predictions, where small one-step errors accumulate and propagate through subsequent predictions.

An apparent increase in median R^2 at the six month horizon compared to shorter horizons reflects the nature of the metric rather than improved absolute accuracy. Over longer horizons, the water level time series is dominated by low frequency seasonal variability, which is the recursive MLP captures reasonably well. As a result, a substantial fraction of the variance is explained despite large absolute errors, as seen by the corresponding increase in RMSE and WAPE. This highlights the importance of interpreting R^2 alongside scale dependent error metrics.

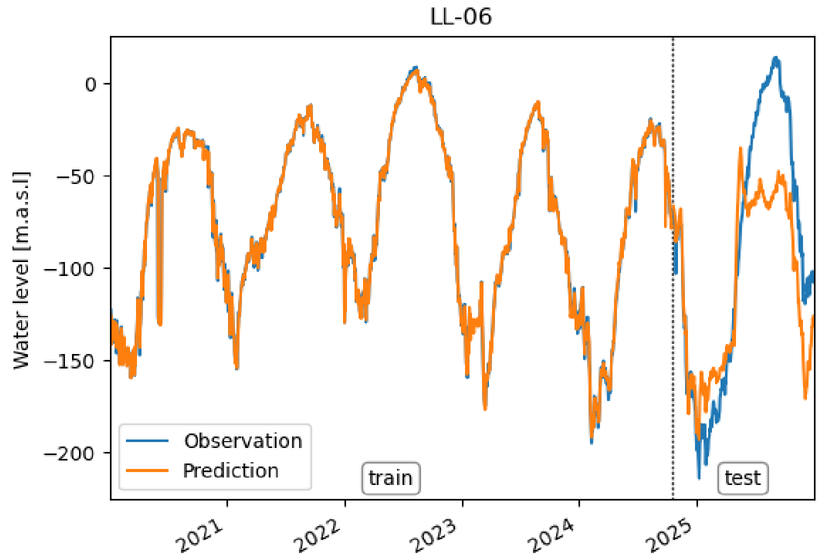

Figure 7: Recursive multistep predictions for the water level in well LL-06 generated using the MLP with target feedback. Predictions are iteratively fed back as inputs to generate forecasts beyond the first time step, illustrating performance degradation with increasing prediction horizon. The vertical dashed line marks the start of the test period.

Table 1: One-step-ahead prediction performance for well LL-06 on the test set. Results are reported as P10 / P50 / P90 statistics across multiple independent random initializations for the best MLP configurations with and without target feedback.

	R^2	RMSE (m.a.s.l.)	WAPE (%)
MLP	0.71 / **0.76** / 0.80	31.3 / **34.5** / 37.4	27.0 / **31.2** / 34.3
MLP w/target feed	0.990 / **0.992** / 0.997	3.6 / **6.0** / 6.7	3.0 / **4.9** / 5.9

Table 2: Recursive multistep prediction performance for well LL-06 at different forecast horizons. Results are summarized using P10 / P50 / P90 statistics across multiple independent random initializations for the best MLP configuration with target feedback.

	R^2	RMSE (m.a.s.l.)	WAPE (%)
7 days	-1.3/ **0.52** / 0.79	1.2 / **1.8** / 4.0	1.3 / **2.1** / 4.7
30 days	-1.1/ **0.39** / 0.76	4.5 / **7.2** / 9.7	3.9 / **7.6** / 10.5
6 months	-0.7/ **0.73** / 0.88	13.4 / **20.3** / 51.3	7.6 / **10.5** / 26.3

3.2.4 Operational scenario predictions

While the preceding sections focus on predictive accuracy, one of the main practical values of these models lies in their ability to support short term decision making under specific operating conditions. In practice, operators are often less concerned with average model performance and more interested with scenario-based questions such as whether water levels may reach critical operational limits during sustained hot water demand.

To illustrate this application, we consider a high demand scenario motivated by the lowest observed water level during the study period. In January 2025, the water level in LL-06 declined to about -220 m.a.s.l., raising concerns about whether operational preparations, such as delivery cutback, might become necessary. Using the trained MLP with target feedback, recursive forecasts were generated under the assumption that production and injection rates from LL-04, LL-06, and GN-01 persist at the levels observed in January 2025 and remain constant throughout the prediction horizon. This persistence assumption represents a high stress scenario corresponding to continued heat demand and the objective is to investigate when a critical water level threshold is reached.

The operational threshold is defined relative to the pump intake depth, but the pump is located approximately 265 m.a.s.l. and a conservative critical level of 240 m.a.s.l. is used to represent the safety margin for operational purposes. Across the ensemble of recursive forecasts, the time required for the predicted water level to reach the critical threshold varies between model initializations. Under the assumed sustained demand scenario, the MLP model configuration indicates that the critical level could be reached within few days (P10 = 3.1 days, P50 = 4.8 days, P90 = 6.6 days).

Although these results are dependent on the scenario and not intended as deterministic predictions, they demonstrate how neural network–based forecasting models can be used to translate current operating conditions into actionable time windows for mitigation measures, such as reducing delivery temperature or adjusting demand.

4. CONCLUSIONS

This study demonstrated the potential of neural network-based time series models for short term prediction of water levels in a low-temperature geothermal reservoir. Using production data from the Laugaland field in Iceland, MLP models were shown to accurately reproduce short term water level dynamics when temporal context was included through lagged inputs and target feedback. An evaluation across multiple model initializations showed both high predictive accuracy for short horizons and increasing uncertainty with recursive predictions.

A scenario-based application showed how recursive ensemble predictions can be used to support decisions regarding the risk of water level approaching critical operational thresholds under sustained demand conditions. Rather than providing deterministic predictions, this application highlights the value and potential of neural network models for exploring plausible future scenarios and their associated uncertainty.

This study is not intended to be exhaustive, and multiple extensions are possible. Future work will include systematic comparison with classical time series prediction methods and physics-based reservoir models, as well as evaluation of more advanced learning-based approaches, including deeper neural networks and alternative machine-learning techniques. Additional work will also explore other operational variables. In particular, ongoing cooling trends observed in production wells at Laugaland motivate future studies of coupled water level and temperature prediction, accounting for the combined effects of production induced cold water recharge and nearby injection.

ACKNOWLEDGEMENTS

We acknowledge financial support from the Energy Resources Engineering Department at Stanford University and Landsvirkjun, the National Power Company of Iceland. We also thank Reykjavík Energy and Veitur for access to data from the Laugand geothermal field.

REFERENCES

Dodge, Y. The Concise Encyclopedia of Statistics. Springer, New York, 2008.

Brownlee, J. Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python, v1.3 (2018).

Goodfellow, I., Bengio, Y., & Courville, A. Deep Learning. MIT Press. Retrieved from http://www.deeplearningbook.org (2016).

Hyndman, R.J. and Athanasopoulos, G. Forecasting: Principles and Practice, 3rd ed. OTexts, Melbourne, Australia (2018).

Kristinsson, S. G. Laugaland í Holtum - Hitastigulsboranir árið 2023 [Laugaland - Temperature gradient drilling 2023]. Rep. ISOR-23046. Iceland Energy Authority (2023).

Oskarsson, F., Ingimarsson, H., Kristinsson, S. G., Hjartarson, A., and Thorbergsson, A. Rangárveitur – Mat á orkuöflunarkostum vegna stækkunar veitunnar [Rangarveita District Heating - Assessment of energy supply options for the expansion of the utility]. Rep. ISOR-2020/041. Iceland GeoSurvey (ISOR), (2020).

Shumway, R. H., and Stoffer, D. S. Time Series Analysis and Its Applications: With R Examples, 3rd ed. Springer (2011).