Proceedings World Geothermal Congress 2026
Calgary, Canada, June 8 – 11, 2026

Field Validation Of A Machine Learning Model For Real-Time Prediction Of Geothermal Well Output Using Differential Pressure Measurements

Agata Rostran Largaespada[12]*, Egill Júliusson[3], Hilmar Már Einarsson[4] and María Sigríður Guðjónsdóttir[1]
Reykjavik University[1]*, GRÓ-GTP[2], Artic Green Energy[3], Landsvirkjun[4]
agatar@ru.is

Keywords: Two-phase flow, Machine learning, Real-time monitoring, Geothermal Field validation.

INTRODUCTION

Accurate real-time estimation of geothermal well output mass flow rate and enthalpy remains a major challenge in two-phase geothermal wells. Conventional measurement techniques, such as separator tests, Tracer Flow Tests (TFT) (Hirzt et al., 1993), and Liquid Phase Tracer and Orifice Plate Differential Pressure Method developed and field-tested by Landsvirkjun and Kemia in 2011 (Hauksson, 2011), provide reliable results but are intrusive, costly, and infrequent. These limitations restrict their use for continuous monitoring or real-time operational control.

Differential pressure (DP) measurements across orifice plates have been explored as a non-intrusive alternative for continuous well monitoring. By analyzing the pressure drop and recovery across an orifice plate, it is possible to infer flow characteristics indirectly. However, estimating steam quality (x), enthalpy and flow rate from DP readings is complex due to multiphase flow behavior and site-specific conditions (James, 1965; Lin, 1982; Helbig and Zarrouk, 2012).

To address these challenges, Juliusson et al. (2023) developed a machine learning framework using differential pressure (DP) sensor data from geothermal wells at Theistareykir, achieving promising results for predicting steam quality. Building upon that work, the present study extends the methodology by comparing four machine learning workflows that integrate different clustering, feature selection, and regression techniques.

Among the tested configurations, the best-performing workflow integrated Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Recursive Feature Elimination with Cross-Validation (RFECV), and Random Forest Regression (RFR). This combination achieved a Root Mean Squared Error (RMSE) of 0.010 and a Mean Absolute Percentage Error (MAPE) of 0.56%, confirming its capability to accurately predict steam quality using DP sensor inputs.

The next step involves field validation of the trained model at the Theistareykir and Krafla geothermal fields to assess its performance under real operating conditions. This phase aims to demonstrate the proof of concept for real-time geothermal well monitoring using differential pressure measurements and machine learning, which will serve as proof of concept for continuous, non-intrusive well monitoring using machine learning.

1 METHODS

1.1 Machine learning Framework

A machine learning (ML) framework was developed to estimate steam quality directly from differential pressure (DP) measurements across an orifice plate installed at geothermal wellheads. The model uses pressure and geometric inputs to infer two-phase flow behavior in geothermal wells.

The workflow combines feature extraction, clustering, and regression to model nonlinear relationships between flow parameters. Four configurations were evaluated, each integrating different combinations of:

- Clustering: K-means (MacQueen, 1967) and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (Ester et al., 1996) were tested to group similar operational states and remove outliers.
- Feature selection and reduction: Principal Component Analysis (PCA) (Jolliffe, 2002) and Recursive Feature Elimination with Cross-Validation (RFECV) (Awad and Fraihat, 2023) were applied to identify the most relevant variables and reduce dataset complexity while preserving variance.
- Regression model: Random Forest Regression (RFR) (Breiman,2001) was employed to capture nonlinear relationships between differential pressures and steam quality. Model hyperparameters such

as the number of trees, depth, and sample splits were tuned using GridSearchCV with five-fold cross-validation to balance accuracy and generalization.

The configuration integrating DBSCAN, RFECV, and RFR achieved the best predictive performance, confirming its suitability for real-time geothermal well monitoring due to its high accuracy and computational efficiency.

1.2 Field setup and Input parameters

Field data for this study were obtained from a series of two-phase flow measurement conducted by Landsvirkjun between 2018 and 2023 at the Bjarnarflag and Theistareykir geothermal fields. These experiments provided the reference dataset used to train and later validate the machine learning (ML) model. Several measurement techniques were applied, including the separator method, and the Liquid Phase Tracer and Orifice Plate Differential Pressure Method (Hauksson, 2011). The separator method measures geothermal output by physically separating the two-phase flow into steam and water. Steam flow is determined from differential pressure (DP) measurements across an orifice plate or other gas flow meters, while the liquid phase is quantified using a weir box or similar device (Helbig and Zarrouk, 2012). These measurements provide direct estimates of steam and water flow rates for calculating total mass flow and enthalpy.

The combined tracer–DP method integrates chemical tracer injection with DP measurements to enable continuous flow characterization without interrupting well operation. In this setup, a pump injects sodium fluorescein dye upstream, which dissolves exclusively in the liquid phase. Downstream pressures (P_1 and P_2) are recorded during each test run, and liquid samples are collected for concentration analysis. The samples represent only the liquid phase at the operating pressure (P_1), even though the discharge becomes two-phase at atmospheric conditions, hence the use of a cooling coil for condensation before analysis.

The experimental setup used for this study is based on a three-tap differential pressure (DP) orifice plate installed at geothermal wellheads. The configuration includes three pressure sensors positioned upstream and downstream of the orifice to capture pressure variations across the restriction (Steven et al., 2018). The system geometry is defined by the pipe diameter (D), the orifice diameter (d), and their ratio (β = d/D), which influences flow characteristics and pressure recovery (Figure 1). These sensors continuously record differential pressure values used to derive key flow parameters for machine learning (ML) modeling.

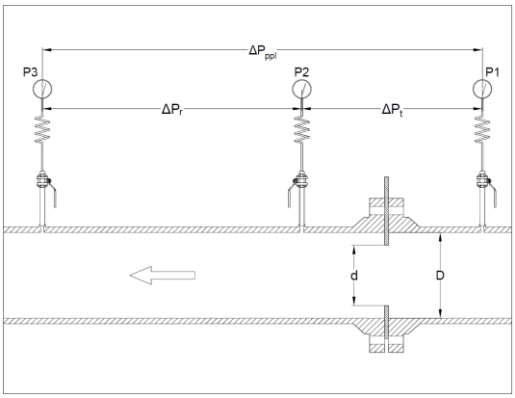


Figure 1 Three-tap differential pressure over an orifice plate setup showing pressure tap locations and geometric parameters (D, d).

From the three measured pressures, the following differential pressures are defined:

- $\Delta P_{12} = P_1 - P_2 \quad (Traditional\ pressure\ loss, \Delta P_{\mathrm{t}})$
- $\Delta P_{32} = P_3 - P_2 \quad (Recovery\ pressure, \Delta P_{\mathrm{r}})$
- $\Delta P_{13} = P_1 - P_3 \quad (Permanent\ pressure\ loss, \Delta P_{\mathrm{ppl}})$

These differential pressures are used to compute additional non-dimensional parameters that characterize two-phase flow behavior:

$$PLR = \Delta P_{\mathrm{ppl}} / \Delta P_{\mathrm{t}} \quad , \quad PRR = \Delta P_{\mathrm{r}} / \Delta P_{\mathrm{t}} \quad , \quad RPR = \Delta P_{\mathrm{ppl}} / \Delta P_{\mathrm{r}}$$

In addition, the density ratio (DR) between steam and water is defined as:

$$DR = \rho_{steam} / \rho_{water}$$

are the densities of steam and water, respectively, determined from thermodynamic properties at the measured pressures and temperatures.

The input features for the ML model therefore include upstream pressure (P_1), differential pressures ($\Delta P_{12}, \Delta P_{13}, \Delta P_{32}$), geometric parameters (D, d, β), derived ratios (PLR, PRR, RPR, DR). These variables constitute the experimental measurements (EM) obtained from the DP meter setup and serve as the independent inputs to the machine learning model, while steam quality (x) is the target variable predicted by the trained Random Forest Regression model.

The reference measurements (RM) correspond to conventional techniques such as separator method, TFT or the Liquid Phase Tracer and Orifice Plate Differential Pressure Method, which provide benchmark values for enthalpy (h) and mass flow rate (ṁ) under known pressure conditions. The relationship between the reference measurements and the differential pressure (DP)-based experimental can be expressed as:

$$f(h, \dot{m}|P)_{RM} \approx g(P_1, \Delta P_{12}, \Delta P_{13}, \Delta P_{32})_{EM}$$

where $f(h, \dot{m} \mid P)_{RM}$ denotes the functional relationship between enthalpy and mass flow rate for a given pressure P, and $g(P_1, \Delta P_{12}, \Delta P_{13}, \Delta P_{32})_{EM}$, represents the predictive model trained on the experimental DP signals.

The experiments covered typical geothermal operating conditions, enthalpies of 1000–2800 kJ/kg, and mass flow rates of 5–35 kg/s. These ranges represent the operating conditions covered by the experimental data and therefore define the limits of the model's training and testing domain.

1.3 Evaluation metrics

Model performance will be assessed using the Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) between predicted and reference steam quality values. These metrics quantify prediction accuracy and provide a basis for comparison during field validation, where new data from the Theistareykir and Krafla geothermal fields will be used to evaluate real-time model performance and transferability.

2 RESULTS AND DISCUSSION

To determine the optimal workflow for predicting steam quality (x) from differential pressure (DP) measurements, four machine learning configurations were compared. Each combined different clustering and feature selection strategies, while Random Forest Regression (RFR) served as the predictive model. The dataset included more than 18,000 data points collected between 2019 and 2023 from geothermal wells at Theistareykir and Bjarnarflag, where measurements were obtained using the separator method and the Landsvirkjun Liquid Phase Tracer and Orifice Plate Differential Pressure Method. The tests covered a broad range of two-phase flow conditions. Table 1 summarizes the performance metrics for each configuration, measured using the RMSE and MAPE.

Table 1 Summary of the performance metrics for each configuration, evaluated using the Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE).

Configuration	Workflow Summary	RMSE	MAPE (%)
Case 1	K-means + Targeted Features + RFR	0.044	4.05
Case 2	K-means + SelectKBest + RFR	0.032	3.29
Case 3	DBSCAN + PCA + RFR	0.014	0.71
Case 4	DBSCAN + RFECV + RFR (Selected Model)	0.010	0.56

The best-performing workflow (Case 4) combined DBSCAN for noise filtering and Recursive Feature Elimination with Cross-Validation (RFECV) for optimized feature selection. This configuration identified a streamlined and physically meaningful feature set (P_1, ΔP_{12}, ΔP_{32}, d, DR, RPR) while maintaining high predictive accuracy even when trained on only 25 % of the total dataset. This demonstrates its computational efficiency and robustness for real-time applications.

Figure 2 shows the measured versus predicted steam quality for Case 4, where data points align tightly along the 1:1 diagonal, confirming strong predictive agreement between model outputs and reference measurements.

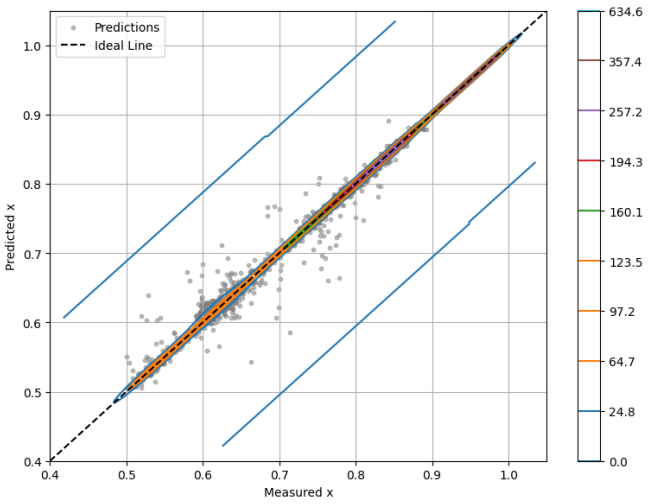


Figure 2 Measured vs. predicted steam quality (x) for Case 4 using DBSCAN + RFECV + RFR. Contours represent data density, with most predictions clustering along the $x_{measured} = x_{predicted}$ line. These are results from the full dataset (100%) and gray dots representing individual predictions.

The model accurately captures two-phase flow behavior over typical geothermal operating ranges, with upstream pressures (P_1), up to 15 bar, enthalpies of 1000–2800 kJ/kg, and mass flow rates of 5–35 kg/s. Its precision and generalizability confirm readiness for real-time deployment, enabling continuous estimation of steam quality from DP sensor inputs without interrupting well operation.

Field validation will extend this work to other geothermal fields, testing the model on previously unseen wells within the same operational range. This phase will provide the final proof of concept for non-intrusive, machine-learning-based real-time geothermal well monitoring.

3 PLANNED FIELD VALIDATION

Field validation planned for 2025–2026 will be conducted at the Theistareykir and Krafla geothermal fields to verify model performance under real operating conditions. These wells were not part of the original training dataset, allowing the model to be tested on new field data while remaining within the trained operational range.

The trained machine learning (ML) model will be integrated with existing three-tap differential pressure (DP) orifice plate sensors through a live data acquisition system, enabling real-time prediction of steam quality and well output. In this setup, DP signals from the wellhead will be continuously recorded and preprocessed to remove noise, outliers or missing values. The cleaned data will then be processed by the trained model (Case 4), using the input parameters described in section 2 to predict steam quality. From these predictions, enthalpy can be derived using thermodynamic relationships while the total mass flow can be obtained by applying established two-phase flow correlations together with DP measurements.

Model predictions will be compared with reference values obtained from conventional methods, including the separator method and the Landsvirkjun Liquid Phase Tracer and Orifice Plate Differential Pressure Method, to evaluate real time accuracy and responsiveness. Model performance will be quantified using RMSE and MAPE.

During field testing, the following objectives will be evaluated:

- Model validation: Comparison of real-time predictions of steam quality (x) and derived parameters (enthalpy, mass flow rate) with reference measurements.
- System performance: Assessment of data acquisition reliability, model response time, and stability of predictions under varying flow regimes.
- Operational integration: Evaluation of model deployment feasibility for long-term, continuous well monitoring.

Successful validation will demonstrate the model's ability to predict geothermal well output non-intrusively and continuously, confirming its potential as a real-time monitoring tool for geothermal resource management and optimization.

4 CONCLUSIONS

This study developed and optimized a machine learning (ML) framework for predicting steam quality in geothermal wells using differential pressure (DP) measurements across an orifice plate. Four modeling workflows were tested, combining various clustering and feature-selection methods. The best-performing configuration integrating DBSCAN for noise filtering, Recursive Feature Elimination with Cross-Validation (RFECV) for feature selection, and Random Forest Regression (RFR) achieved a Root Mean Squared Error (RMSE) of 0.010 and a Mean Absolute Percentage Error (MAPE) of 0.56%, demonstrating high predictive accuracy and computational efficiency.

The upcoming field validation at the Theistareykir and Krafla geothermal fields will verify model performance under real operating conditions. Live DP data will be processed in real time and compared against measurements obtained from conventional well-testing methods to confirm the model's accuracy, responsiveness, and long-term stability.

By combining machine learning with non-intrusive DP sensing, this approach provides a cost-effective and scalable solution for continuous geothermal well monitoring. Successful field validation will mark an important step toward data-driven, real-time optimization of geothermal resource management and utilization.

Acknowledgements

The authors gratefully acknowledge the support of the GRÓ Geothermal Training Programme and Landsvirkjun for data access and collaboration. Special thanks to Karl Emil Sveinsson and Helgi Alfreðsson for their contributions to experimental setup and field data collection.

REFERENCES

Awad, M., & Fraihat, S.: Recursive Feature Elimination with Cross-Validation with Decision Tree: Feature Selection Method for Machine Learning-Based Intrusion Detection Systems. *Journal of Sensor and Actuator Networks*, *12*(5), (2023), 67. https://doi.org/10.3390/jsan12050067

Breiman, L.: Random Forests. *Machine Learning*, *45*(1), (2001), 5–32. https://doi.org/10.1023/A:1010933404324

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, (1996), 226–231.

Hauksson, T.: *Landsvirkjun, Kröflustöð. Afkastamælingar borhola með þynningaraðferð og tvífasa mæliblendu. Innleiðing aðferða*), (No. LV-2011/018), Landsvirkjun, (2011).

Hirtz, P., Lovekin, J., Copp, J., Buck, C., and Adams, M.: Enthalpy and mass flowrate measurements for two-phase geothermal production by tracer dilution techniques. *Geothermics*, 22, (1993), 193–200.Helbig, S., and Zarrouk, S.J.: Measuring Two-Phase Flow in Geothermal Pipelines Using Sharp-Edge Orifice Plates. Geothermics, 44, (2012), 52–64.

James, R.: Metering of Steam-Water Two-Phase Flow by Sharp-Edged Orifices. *Proceedings of the Institution of Mechanical Engineers*, *180*(1), (1965), 549–572. https://doi.org/10.1243/PIME_PROC_1965_180_038_02

Jolliffe, I.T.: Principal Component Analysis (2nd edn), Springer-Verlag, (2002).

Juliusson, E., Sveinsson, K. E., Þórhallsson, R., & Steven, R.: Real-Time Monitoring of Geothermal Wells using the Dual-Differential Pressure Method. *Proceedings World Geothermal Congress Beijing, China,* (2023).

Lin, Z. H.: Two-phase flow measurements with sharp-edged orifices. *International Journal of Multiphase Flow*, *8*(6), (1982), 683–693. https://doi.org/10.1016/0301-9322(82)90071-4

MacQueen, J. B.: Some Methods for Classification and Analysis of Multivariate Observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, University of California Press, (1967), 281–297.

Steven, R., Shugart, C., & Kutty, R.: *Orifice Meter Multiphase Wet Gas Flow Performance – The Pressure Loss Ratio Solution to the 'Ill-Posed' Problem*. North Sea Flow Measurement Workshop, (2018), 22-24 October.