Investigating the Variance in Activity of the Lady Knox Geyser, Waiōtapu Thermal Wonderland, NZ

Buna R. Rachman[1*], Sadiq J. Zarrouk[1], Katherine Luketina[2],
,Jim McLeod[2], Claire Kotze[2], and Andreas W. Kempa-Liehr[1]

[1]Department of Engineering Science, The University of Auckland, Private Bag, Auckland 92019, New Zealand

[2] Waikato Regional Council, 160 Ward St, Hamilton 3204, New Zealand

brac892@aucklanduni.ac.nz

Keywords: *geothermal springs, geysers, geothermal monitoring, geothermal resource management, Waiōtapu, machine learning, feature engineering, time-series analytics*

ABSTRACT

The Lady Knox Geyser is within the tourist complex of Waiōtapu Thermal Wonderland (WTW), and is soap-induced to erupt for the entertainment of tourists. Currently, the height and duration of the eruption vary. The Thermal Park Management deems the eruption duration to have become shorter. The WTW management and the Waikato Regional Council would like to have the variability investigated. Understanding Lady Knox Geyser's behaviour is part of the landowners, the Iwi of Ngati Tahu – Ngati Whaoa, commitment to Kaitiakitanga (Guardianship, Conservation, and Sustainability). This study examines hot-spring temperature time series data from the Lady Knox Geyser. The geyser shows a distinct peak around the same time every day, as it was induced by the Park Rangers using soap to trigger the eruption. During its quiescent period, the spring temperature mean at several metres depth is 103°C.

Detailed scrutiny of data shows that a different pattern of temperature data is distinguished between a short eruption duration and a long one. However, some eruptions do not abide by this distinct pattern. The correlation between peak spring temperature and weather data was analysed to determine whether there are any correlations between eruption duration and intrinsic factors and/or external perturbations. The peak spring temperature and air pressure are correlated with Lady Knox Geyser's eruption duration during the 3 months the geyser was investigated. In addition to correlations, time-series spring temperature datasets were used to build a machine-learning model that would predict the duration of eruption events in the future. A reasonable model would allow the WTW Management to know the duration of the eruption before it is induced and to help decide any actions to preserve the Lady Knox Geyser better.

1. INTRODUCTION

The subsurface geology of the Lady Knox Geyser is deeply connected to the volcanic activity that has shaped New Zealand's landscape. As part of the Taupō Volcanic Zone (TVZ), the geyser and Waiōtapu Thermal Wonderland (WTW) feature striking geothermal manifestations, including hot springs, mud pools, and geysers (Migoń and Pijet-Migoń, 2017). Dowling (2011) defines geotourism as tourism that fosters appreciation and understanding of geological features and processes. In this context, the storytelling by park rangers—delivered before each induced eruption—enhances public engagement with geothermal science (Migoń and Pijet-Migoń, 2017). The site also demonstrates effective geotourism management by offering well-integrated visitor amenities, such as a sheltered amphitheatre, while maintaining the natural landscape (Migoń and Pijet-Migoń, 2016). Finally, the unique soap-induced eruption and the dramatic spectacle of the Lady Knox Geyser captivate visitors, as reflected in photographs in Figure 1.

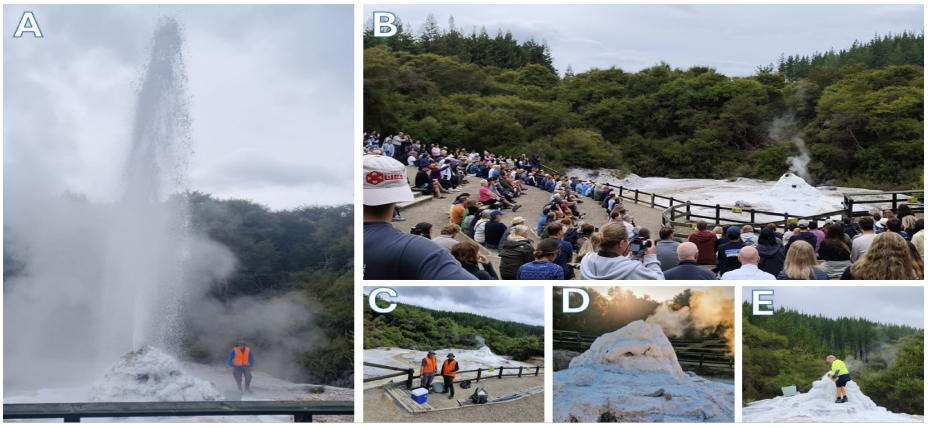

Figure 1: Picture of Lady Knox Geyser taken during the research project. (A) Lady Knox Eruption. (B) Tourists at the WTW amphitheatre. (C) Data collection with the Waikato Regional Council. (D) Dormant stage post-eruption. (E) The WTW ranger inserts soap into the geyser vent.

To protect geothermal features, partnerships between Māori-owned trusts and the local government are essential (Dowling, 2011). The Waikato Regional Council (WRC) monitors the Lady Knox Geyser and other protected geothermal springs (Nikrou et al., 2013); they are obligated to the sustainable management of regional geothermal features based on the Resource Management Act 1991 (Luketina, 2012). A map of the geothermal systems within the TVZ monitored by the WRC can be seen in Figure 2. The Lady Knox Geyser is a surface feature within the Waikite-Waiōtapu-Waimangu geothermal system that is high-temperature, dual-phase, and liquid-dominated.

Lately, the management team of WTW has been concerned with the reduction of eruption duration and height of Lady Knox Geyser, which is one of the main attractions for tourists, besides the world-famous Champagne Pool, the most photographed geothermal feature in New Zealand. Despite three decades of geothermal monitoring, no specific investigation has been done at Lady Knox Geyser to understand this variable behaviour. This research, a collaboration between WRC and the Geothermal Institute, attempts to answer the concern. This study aims to answer the following research questions:

1. What variances were observed in Lady Knox Geyser's activity?
2. Is there any relation between the intrinsic factors and external drivers, on the Lady Knox Geyser's eruption behaviour?
3. Can a Machine Learning model be developed to process time-series features and predict eruption cycle durations?

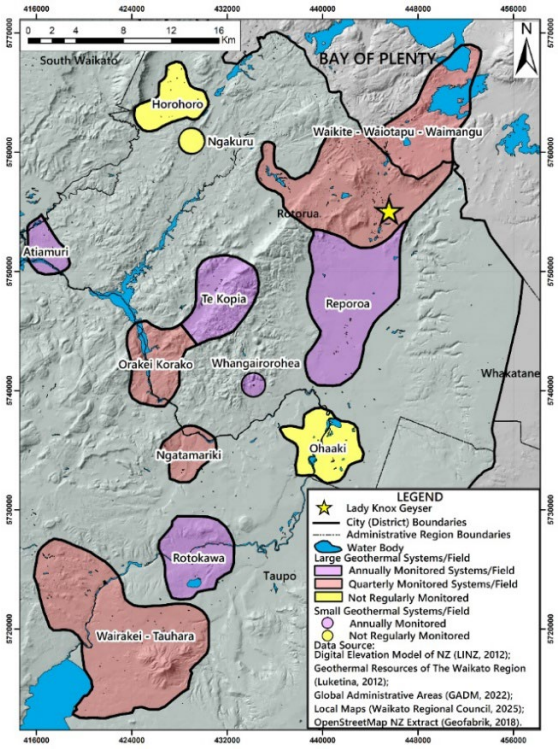

Figure 2: Map of geothermal systems of the TVZ area monitored by WRC, modified from (Lebe, 2021).

2. DATA COLLECTION

The primary field data for this research is spring temperature measurements. A total of three thermocouples were deployed for this research.

2.1 Hot Spring Temperature Measurements

Spring temperature was recorded using thermocouples connected to a waterproof high-temperature data logger. With a 32,000-point memory capacity, three data periods were collected from the Lady Knox Geyser between September and December 2024. Each period spans 22 days, with readings every minute, offering higher data density than (Nikrou, Newson, McKibbin, et al., 2013) who recorded every two minutes for up to 45 days. A sketch is provided in Figure 3 to illustrate the field data collection setup.

2.2 Weather Data Collection

The weather data, which consists of rainfall, air pressure, air temperature, and wind speed, was retrieved from the National Institute of Water and Atmospheric Research (NIWA) public database (NIWA, 2025). It is important to select the station nearest to the Lady Knox Geyser. The weather station chosen was Rotorua EWS because it was only 30.7 km from the Lady Knox Geyser. The second closest weather station is 37.6 km away.

3. DATA AND DISCUSSION

3.1 Overall geyser Temperature Dataset

Lady Knox Geyser Spring temperature dataset was examined throughout each period of 22 days to have a general idea of the cyclic time-series data. The overall periods are plotted in Figure 4. It illustrates periods of relatively stable temperature followed by daily peaks of temperature readings, which indicate when the Lady Knox Geyser is induced for an eruption using soap at 10:15 am every day. A total of 66 eruptions were recorded from all three periods. The distinguished and repetitive cycle has an average amplitude of ~1.5 °C, except for several anomalies that will be discussed below.

3.2 Daily Eruption Patterns and Anomalies

A typical daily eruption profile, shown in Figure 5, includes a quiescent phase with stable temperature, a sharp peak indicating eruption, and a post-eruption cooldown. Then, the post-eruption phase shows a stepwise temperature drop rather than a smooth, curved decline. This temperature drop profile is suspected to be an artefact of the instrumentation. Specifically, a characteristic of the analogue to digital converter (ADC) mechanism built into the logger. In physical space this cooldown will be a smooth decline. A different data logger with a higher resolutions ADC will likely capture this smooth decline.

This study aims to understand the reduction in visible eruption duration. The eruption cycle—used to assess this duration—is defined by the thermocouple data: it starts at the temperature rise (t_1) and ends when the temperature falls below a threshold (t_2). The difference ($\Delta t = t_2 – t_1$) represents the eruption cycle duration. However, this is not the "visible eruption", rather it is the defined detected eruption. The visible eruption will start when spillage from the vent starts until it stops, which is measured by the Splash Zone thermocouples.

ISSN 2703-4275

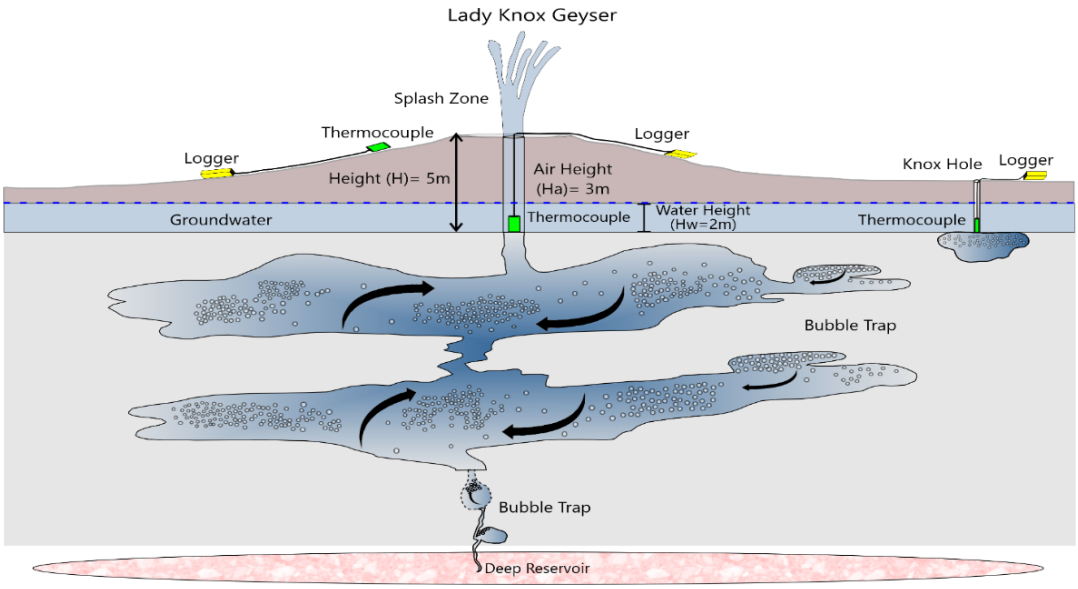

Figure 3: Illustration of deploying thermocouples and loggers to The Lady Knox Geyser. Theoretical subsurface aquifer chambers modified from (Lloyd, 1975), bubble traps modified from (Munoz-Saez et al., 2015).

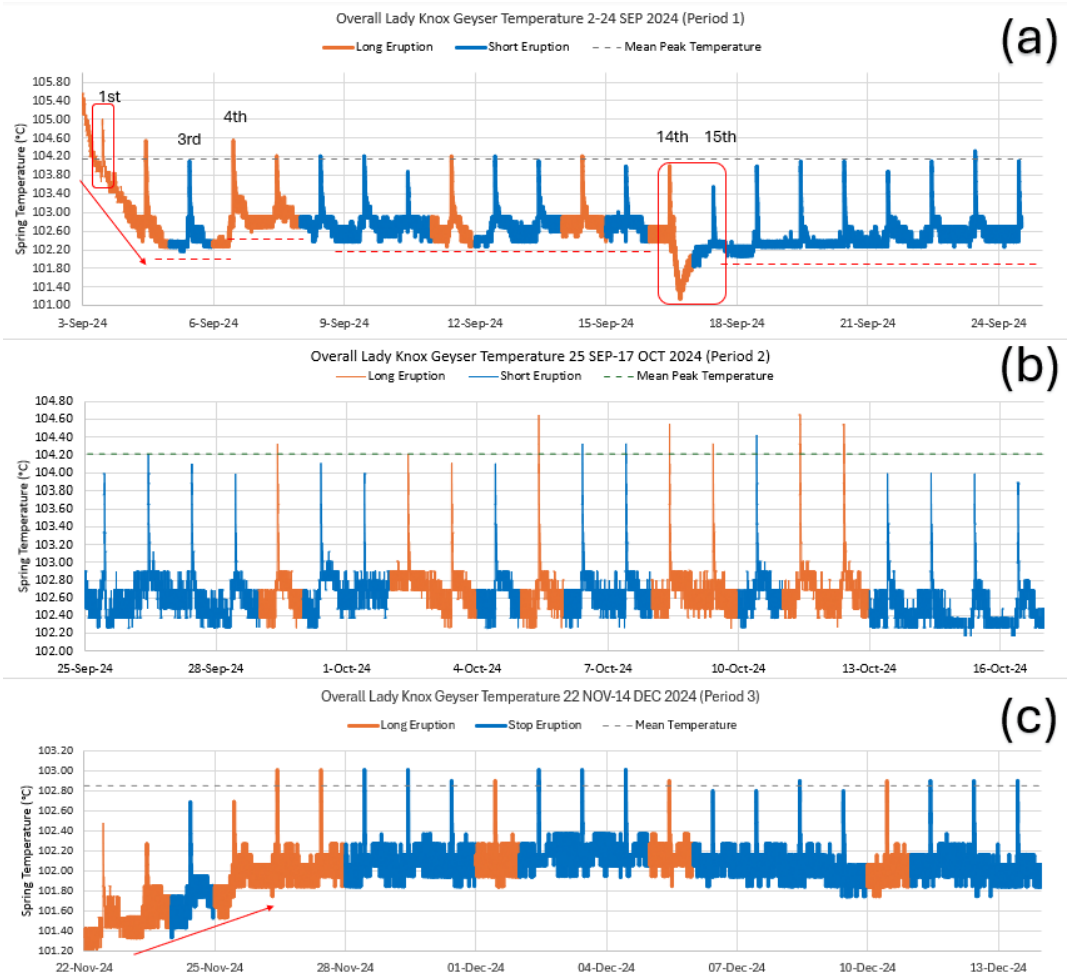

Figure 4: The overall temperature time-series plots for Lady Knox Geyser. (a) Period 1. (b) Period 2. (c) Period 3 . Orange curves represent the long eruptions, and blue curves are the short eruptions.

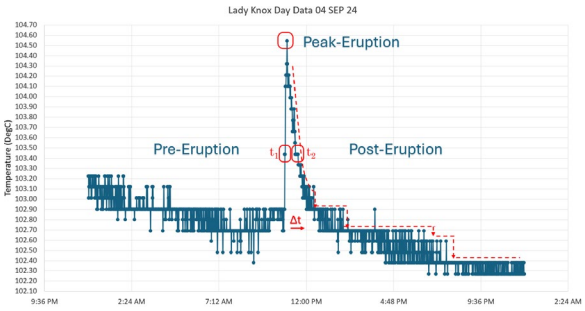

Figure 5: Daily eruption profile of the Lady Knox Geyser on 04 Sep 2024, which consists of a dormant pre-eruption phase, an eruption peak, and a quiescent post-eruption period. Variables t_1 and t_2 are marked in the plot. Δt represents the eruption cycle duration or difference between t_1 and t_2

Having examined three periods at the Lady Knox Geyser, a common pattern between the two groups of data could be distinguished. The groups are based on the eruption cycle duration (Δt), which was calculated by manually recording and subtracting the start-stop times for a total of 66 eruptions. The eruption cycle durations could be divided into long eruptions and short eruptions. Long eruptions have an extended Δt than the average eruption cycle duration in the dataset, whereas short eruptions are less than the mean Δt. A summary of the geyser eruption cycle durations for the three periods is presented in Table 1 as follows.

Table 1. Summary of the Lady Knox Geyser Eruption Cycle Duration from all three periods.

Period (Date)	Mean Eruption Cycle Duration (min)	Total Long Eruptions	Total Short Eruptions
1st (2-24 SEP 24)	48	8	14
2nd (24 SEP – 17 OCT 24)	47	8	14
3rd (21 NOV-13 DEC 24)	27	9	13

Table 1 shows that short eruptions outnumber long ones, with 41 (62%) short eruptions and 25 (38%) long eruptions. The mean of eruption cycle duration in Periods 1 and 2 is 47–48 minutes but drops to just 27 minutes in Period 3. While the cause is unclear, it may relate to seasonal weather. Periods 1 and 2 were recorded in Spring, whereas Period 3 was taken in Summer. Reduced aquifer recharge from less rainfall during drier months could explain shorter eruptions.

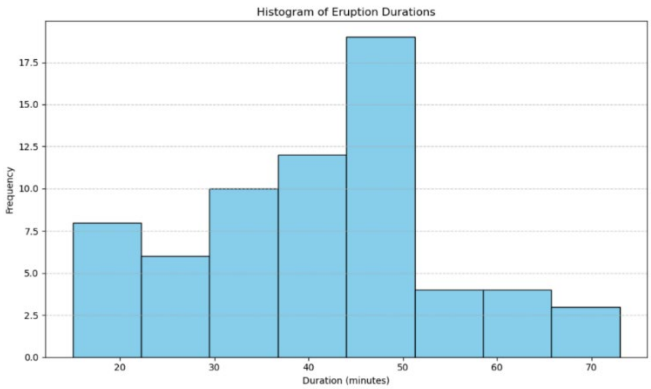

Figure 6: Histogram for all periods

Figure 7 shows that most eruption durations cluster around 40–50 minutes, with a sharp decline in frequency beyond 50 minutes, indicating that longer eruptions are rare. The distribution has at least a weak bimodality with a less pronounced peak at 20 minutes, accounting to shorter durations. A key observation from the 66 recorded eruptions is a distinct pre-eruption pattern: long eruptions typically follow a stable temperature profile, while short eruptions arepreceded by a temperature rise a few hours before the event. This comparison is illustrated in Figure 7.

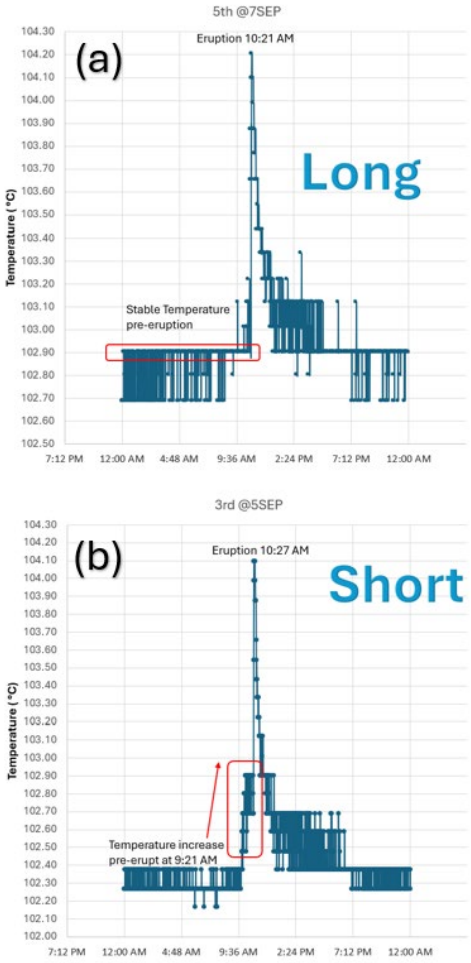

Figure 7: Temperature Profile Pattern Comparison between (a) Long and (b) Short Eruptions.

The long-short eruption pattern was largely consistent in Periods 1 (2–24 Sep) and 2 (24 Sep–17 Oct), with 82% of eruptions following the pattern. Long-duration eruptions (e.g., Sep 7, 11, 14) showed stable pre-eruption temperatures, such as 102.9 °C on Sep 7th (Figure 7a), with a 50-minute eruption duration cycle. In contrast, short-duration eruptions (e.g., Sep 5, 6, 9,10) were preceded by a step-like temperature increase. For instance, on Sep 5 (Figure 7b), the temperature rose from 102.4 °C to 102.9 °C, with a 36-minute eruption duration cycle. However, this pattern was less evident in Period 3 (21 Nov–13 Dec), where only 4 out of 22 eruptions followed this pattern distinction.

A clear anomaly in the 1st Period is the 14th Eruption on 16 September 2024, refer to Figure 4(a), which recorded the lowest post-eruption resting temperature at 101.15 °C—unlike other eruptions, which remained above 102 °C. As shown in the daily profile on September 16th (Figure 8), after the main peak at 104 °C, the temperature dropped as expected but then unexpectedly rose again, suggesting a possible secondary eruption during the refill phase. This was the only event with a temperature increase immediately after an eruption. The anomaly could also be a pre-play, preparing for a natural eruption.

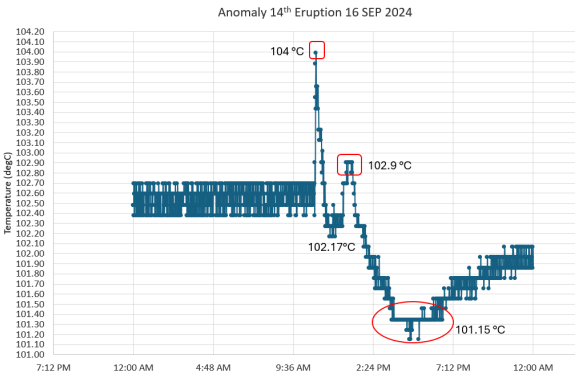

Figure 8: Eruption Anomaly on 16 September 2024.

3.3 Data Correlations with External Drivers

This section evaluates how external drivers (e.g., weather) and intrinsic factors (e.g., peak spring temperature) affect Lady Knox Geyser's eruption cycle duration. As shown in Figure 9. Periods 1 and 2 exhibit similar value ranges, with a mean spring temperature of 102.65 °C, while Period 3 is slightly lower at 102 °C. Air pressure and air temperature also differ: Periods 1 and 2 have a mean of 980 hPa and 10 °C, while Period 3—collected during summer (late Nov–Dec)—has a mean of 982 hPa and 16 °C. This seasonal variation could explain the deviation in Period 3. A data gap exists between 18 October and 20 November 2024. Initially, only Periods 1 and 2 were planned for analysis, but Period 3 was later added to assess seasonal effects.

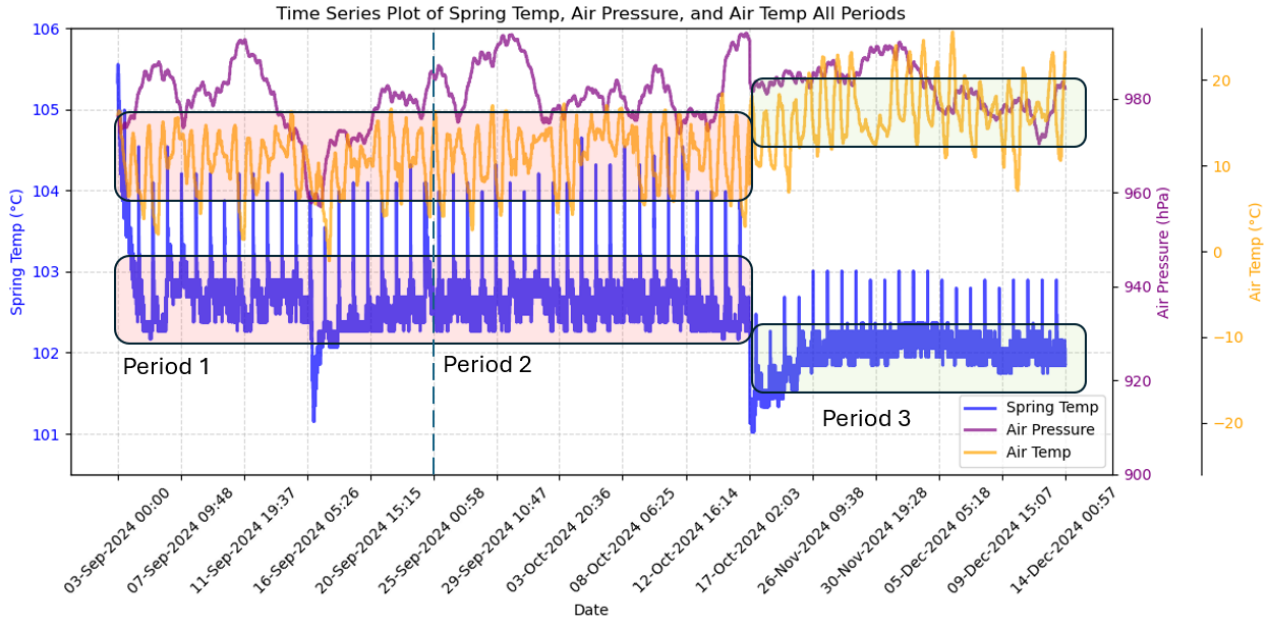

Figure 9: Lady Knox Geyser Spring temperatures are plotted with air pressure and air temperature for all three periods from September to mid-December 2024

ISSN 2703-4275

Lower air pressure may facilitate faster groundwater inflow, sustaining eruptions, similar to how atmospheric loading affects water recharges described by Ingebritsen and Rojstaczer (1993). While previous studies have linked barometric pressure to eruption frequency and the interval between eruptions (Rinehart, 1972; Hurwitz et al. 2008), few have examined its impact on eruption durations. To explore this relationship further, Figure 10 presents cross plots of eruption cycle duration and air pressure across all three periods combined. The combined period showed a weak negative correlation (r = -0.2469).

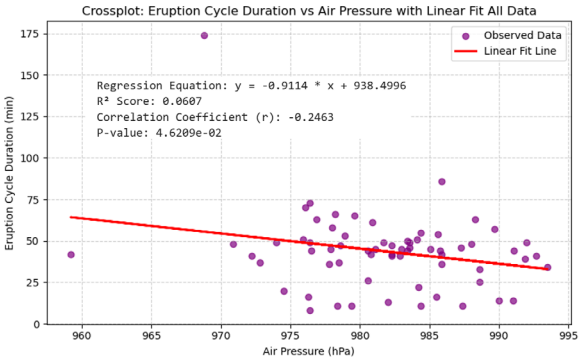

Figure 10: Lady Knox Geyser Eruption Cycle Duration plotted against air pressure for All Periods

Correlation analysis was performed with other weather variables, namely Air Temperature and Wind Speed, to investigate whether they affect the duration of the eruption cycle. Hurwitz et al., (2008) observed that Air temperature influenced the geyser eruption interval at Yellowstone, albeit localised to Aurum Geyser only. (Marler,1954) reported that the Interval Between Eruptions (IBE) became more sparse in windy conditions. Table 2 below summarises the correlation outputs.

Table 2 shows that the combined periods have moderate inverse correlations between air temperature and eruption duration (r = -0.45). This suggests cooler air may accelerate heat loss from erupting fluids, shortening eruptions. Wind speed showed no statistically significant correlations from the eruption duration across all periods because P-values > 0.05; one factor may be that the cone-shaped vent shields the geyser from windy influences, and that it is situated in a more sheltered location than the weather station.

3.4 Data Correlations with Intrinsic Factors

This sub-section explores the relationship between peak eruption temperature and eruption cycle duration depicted in Figure 11. Since geyser eruptions are driven by heat-to-kinetic energy conversion (Hurwitz and Manga, 2017), it is hypothesised that higher peak temperatures may reflect greater thermal energy, potentially extending eruption duration. The combined period yields moderate correlation (r = 0.5483), a significant p-value (1×10^{-6}), and R^2 (0.3006), that is stil considered as a weak model fit per Cohen's (1988) guidelines, refer to Table 2.

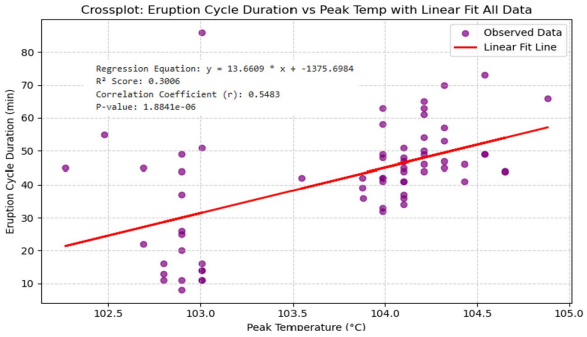

Figure 11: Lady Knox Geyser Eruption Cycle Duration plotted against Peak Temperature for All Periods

Table 2. Correlation Outputs of Eruption Cycle Duration Data against Weather Factors and Peak Spring Temperature.

Variable	Period	r	P-Value	R^2
Air Pressure	1,2,3	-0.25	0.046	0.06
Air Temperature	1,2,3	-0.45	1.34×10^{-4}	0.21
Wind Speed	1,2,3	0.07	0.57	0.005
Peak Spring Temperature	1,2,3	0.55	1×10^{-6}	0.30

4. MACHINE LEARNING MODEL

4.1 Machine Learning Methods

To prepare for the Predictive Machine Learning model , the datasets underwent filtering and labelling. A systematic time-series feature engineering (STSFE) method was applied using the Python package "tsfresh" (Christ et al., 2018). The task—predicting Lady Knox Geyser's eruption cycle duration—was framed as a time-series regression problem, as the target variable is continuous (eruption duration measured in minutes). A supervised learning framework was used to train and test both linear and ensemble models, including Random Forest and Gradient Boosting. The workflow, inspired by Abrasaldo et al. (2023) is summarised in Figure 12.

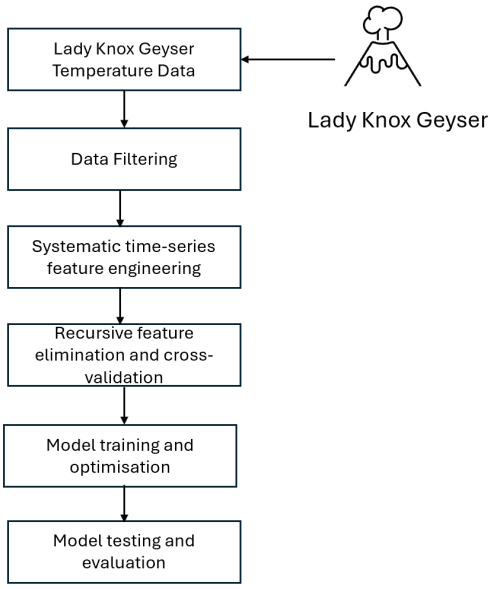

Figure 12: STSFE Machine Learning workflow to predict the eruption cycle duration of the Lady Knox Geyser modified from (Abrasaldo et al., 2023)

4.2 Data Preparation and Labelling

Raw Geyser Spring Temperature data that covered 3 periods from 2-24 September, 25 September – 17 October, and 21 November – 14 December 2024 were extracted and filtered for this study. There are missing data from 18 October to 20 November due to no thermocouples being deployed, as explained, but no imputation of missing values was done. Post-eruption data were excluded from the analyses. This study applies a supervised Machine Learning workflow that requires the labelling of eruptions, identifying the pre-eruption window, and defining the Y target.

The time-series input was divided into $N_w = 66$ windows of time length t_w, with each window corresponding to one eruption. Initially, 6 hours of pre-eruption data were used, but this yielded no statistically significant time-series features. After reducing the window to 3 hours, statistically significant time-series features could be extracted. Each pre-eruption data window was assigned a target value, namely the geyser cycle duration in minutes. These durations were stored in a vector $\vec{y}$. Its dimension corresponds to the number of windows N_w (Abrasaldo et al., 2023). Figure 13 illustrates the pre-eruption window (t_w) and eruption events from September 4th-6th 2024.

4.3 Time-series feature engineering

The automatic time series feature extraction is performed on each window using "tsfresh" to extract a total of K=777 time-series features from each pre-eruption window. The extracted features include distribution characteristics, autocorrelation metrics, properties of linear regressors, energy, entropy, and stochasticity metrics (Abrasaldo et al., 2023). The time-series features comprise a matrix consisting of N_w rows and K columns. Each column charts the change of a feature over the analysis periods, and each row characterises the pre-eruption period at a given time. This process is illustrated in Figure 14.

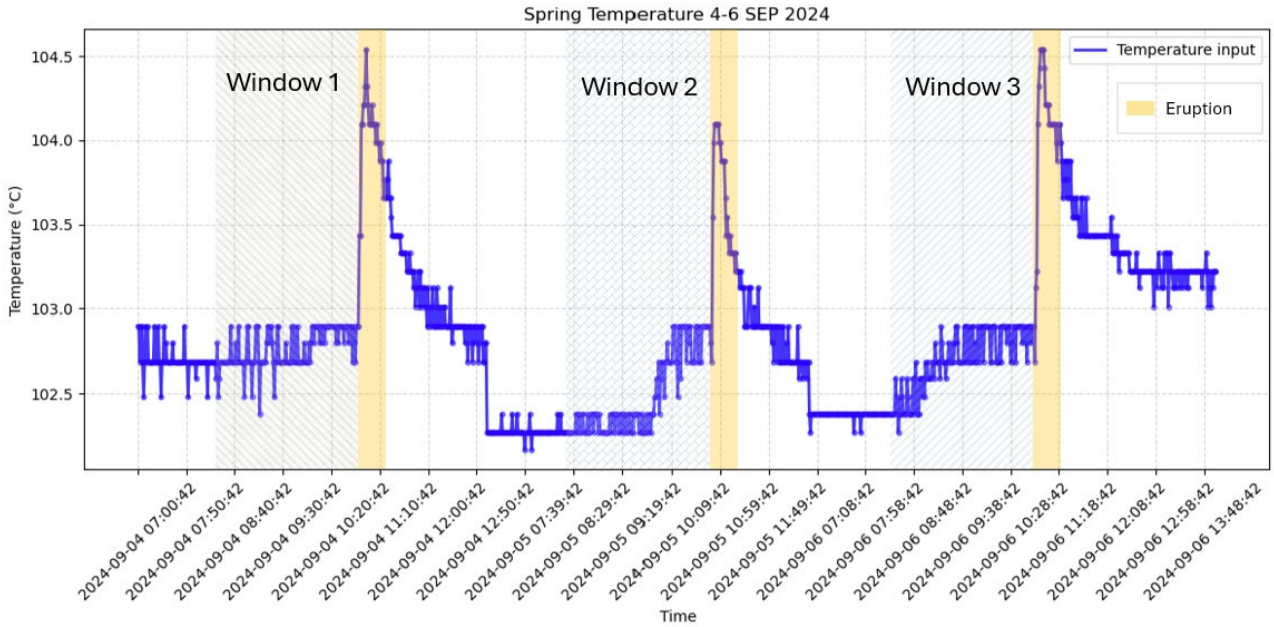

Figure 13: Time-series feature extraction process; windows are the temporal pattern prior to each eruption event

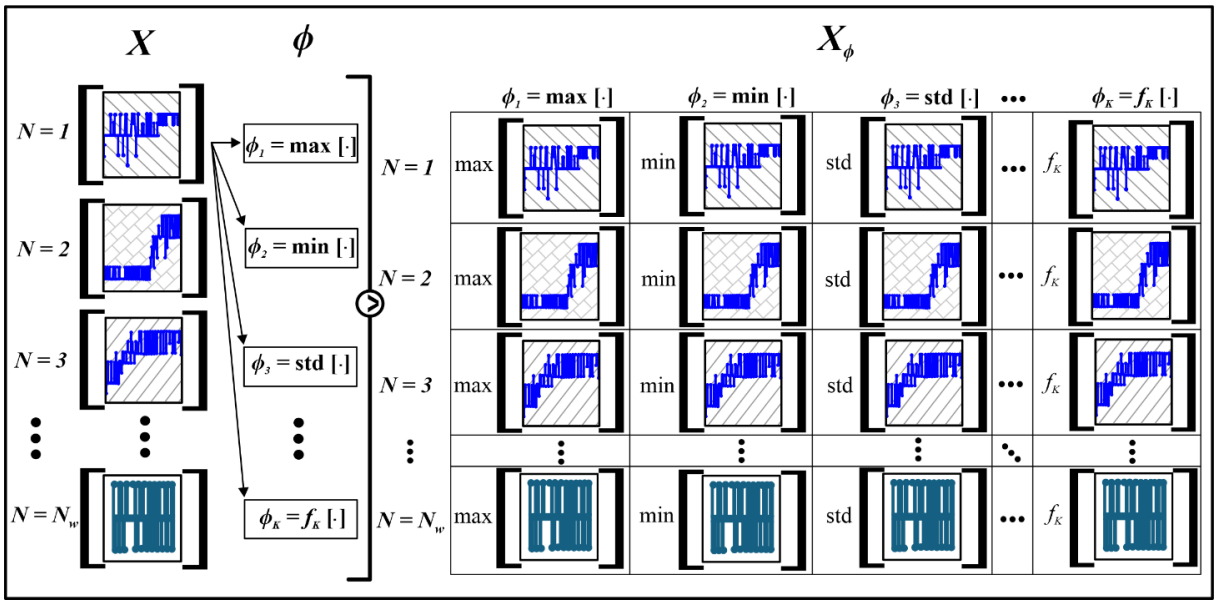

Figure 14: N_w x K Feature matrix utilised for building the predictive model, modified from (Abrasaldo et al., 2023). N_w, each pre-eruption window, is passed into K different time-series computed features

4.4 Feature-Selection

Time-series features that are extracted from a dataset to describe a complex phenomenon may risk having high degrees of collinearity, which indicates two features that have a strong linear relationship, or multicollinearity when the regression model contains several extracted features that are dependent on one another (Shrestha, 2020). Multicollinearity can distort results by making it difficult to assess the individual importance of predictors in a statistical model. (Harrell, 2015). A step-by-step feature selection process filtered the extracted time-series features to leave only those that are non-collinear and relevant to predict the target. The sequence of the feature selection process and the defined matrix variables of this study are illustrated in Figure 15.

The feature selection process begins with univariate feature selection as implemented by the "tsfresh" (Christ et al., 2018). The "tsfresh.feature_selection.relevance" module evaluates the significance of extracted features by calculating p-values using statistical tests such as Kendall's tau for numerical targets (Christ et al., 2023). These p-values are then subjected to the Benjamini–Yekutieli procedure (Benjamini and Yekutieli, 2001) to control the false discovery rate and determine which features to retain. In this study, reducing the pre-eruption window from 6 to 3 hours enabled the selection of 72 statistically significant time-series features.

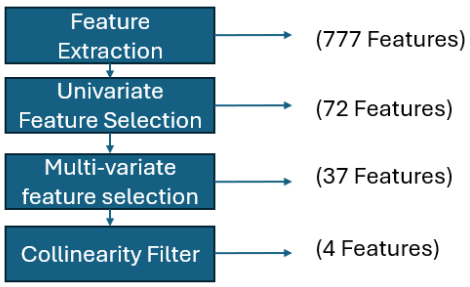

Figure 15: Feature Selection Process and defined matrix variables for the Lady Knox Geyser study.

The next step in the feature selection process is to minimise the risk of including collinear features in the final feature matrix. A multivariate feature selection approach was implemented using the Recursive Feature Elimination with Cross-Validation (RFECV) algorithm from scikit-learn (Pedregosa et al., 2011). This process performs RFE in a cross-validation loop to find the optimal number of features. RFECV recursively removes the least important features as determined by an external estimator, in this study, a Random Forest regressor. The RFECV setting was configured with 10-fold cross-validation (cv=10) and eliminated 5 features per iteration (step=5). This process resulted in 37 features.

The final step of the feature selection process involved removing highly collinear features from the dataset. A Pearson correlation matrix was computed, and features with correlation coefficients above 0.75 were dropped to reduce redundancy. This threshold, though arbitrary, is commonly used to indicate strong collinearity. As a result, 33 of the 37 features selected by RFECV were removed, leaving 4 final features, as listed in Table 3 with descriptions adapted from (Christ et al., 2025).

Feature importance was then assessed using a linear regression model, revealing that Temperature__fft_aggregated__aggtype_"centroid" had the highest importance score (110.65), indicating it was the strongest predictor of the target variable. The feature importance scores are in Figure 16.

Table 3. Explanation of best time-series features from the Feature Selection Process (Christ et al., 2025) .

Time-series property	*Tsfresh feature name*	p-value	Description
Aggregated type "centroid"	Temperature__fft_aggregated__aggtype_"centroid"	175 x 10^{-4}	Returns the spectral centroid (mean), variance, skew, and kurtosis of the absolute fourier transform spectrum.
Coefficient attribute "abs"	Temperature__fft_coefficient__attr_"abs"__coeff_43	208 x 10^{-4}	Calculates the fourier coefficients of the one-dimensional discrete Fourier Transform for real input by the Fast Fourier Transformation algorithm
Linear trend intercept	Temperature__agg_linear_trend__attr_"intercept"__ chunk_len_10__f_agg_"max"	9 x 10^{-6}	Calculate a linear least-squared regression for values of the time-series that were aggregated over chunks versus the sequence from 0 up to the number of chunks minus one.
Coefficient attribute "abs"	Temperature__fft_coefficient__attr_"abs"__coeff_45	396 x 10^{-4}	Calculates the fourier coefficients of the one-dimensional discrete Fourier Transform for real input by the Fast Fourier Transformation algorithm

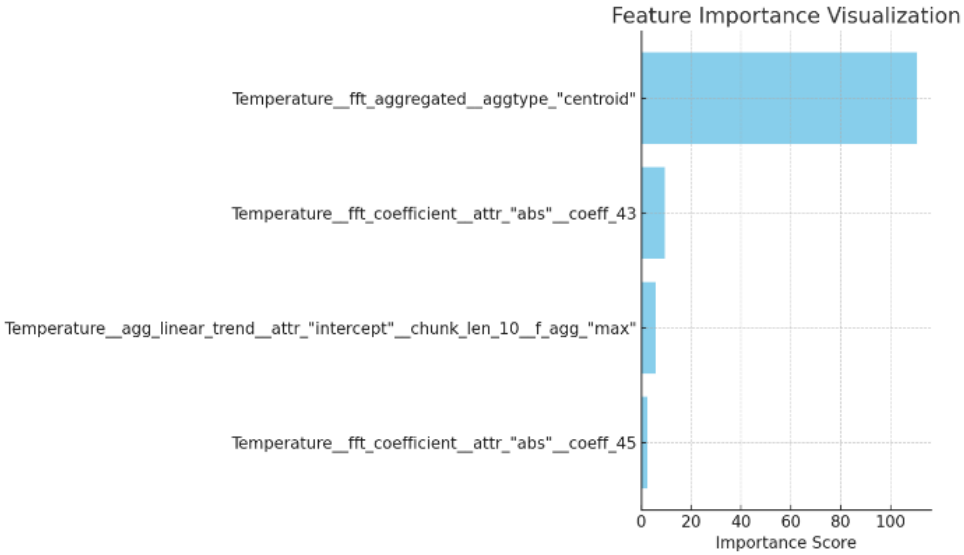

Figure 16: Feature Importance results computed from a Linear Regression Model

4.5 Regression model performance

The filtered time-series feature matrix was used as input for regression models to predict geyser eruption duration. A hold-out dataset consisting of 9 eruptions was isolated so that the model was trained using just 57 eruptions. A linear regression model was initially selected to be trained and then tested against cross-validation prediction results. The output was a poor model fit. The red dots refer to hold-out data, and the grey dots are training data (Figure 17).

The next analysis is to perform a reliable and robust assessment of the performance of different machine learning models. This comparison utilises "cross_val_predict" results from training data (57 eruptions) and indicates that the Ridge Model performed best with the least Mean Absolute Error (MAE) of 18.88 minutes. The last step is to evaluate regression models against hold-out data that was initially set aside to assess the performance of the model on unseen data. The hold-out data evaluation also consistently shows that the Ridge Model has the best performance with an MAE of 9.78 minutes , or 20% error based on the average eruption duration of 48 minutes. The assessment of Machine Learning Models can be viewed in Figure 18.

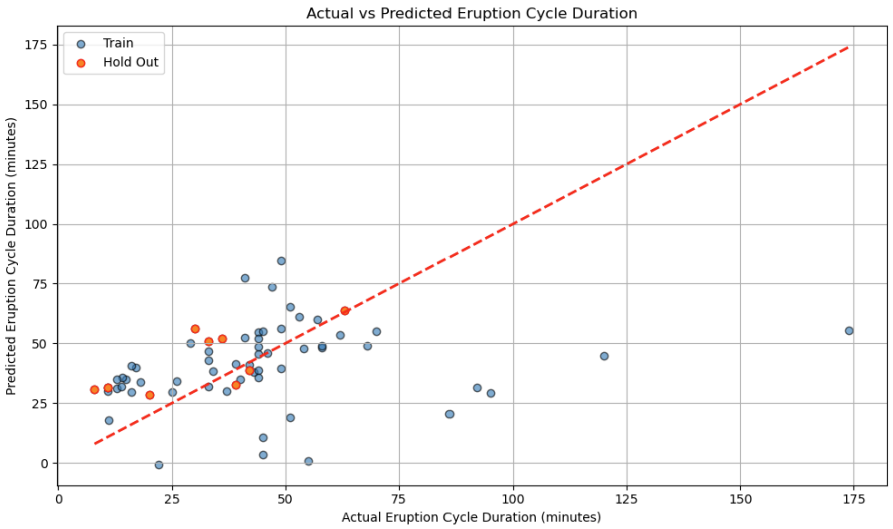

Figure 17: Actual vs Predicted Data from "cross_val_predict" results

Figure 18: Evaluation of Machine Learning Models with Hold-Out Data

5. CONCLUSION

The Lady Knox Geyser is both a natural landmark and a site of cultural significance for the Iwi within the Waiōtapu Thermal Wonderland (WTW), New Zealand. As a key tourist attraction alongside the iconic Champagne Pool, its daily induced eruptions offer a rare opportunity to witness the geyser's complex dynamics in a controlled setting. As a monitored geothermal feature under the Waikato Regional Council (WRC), it provides an important case study at the intersection of natural system behaviour, public engagement, and resource management within the Waikato Region and broader Taupo Volcanic Zone (TVZ).

This study set out to investigate the perceived shortening of eruption durations and the variability in geyser behaviour. Meaningful progress was achieved because this project represents the first time-series temperature analysis performed at Lady Knox Geyser. Moreover, this was the first study that attempted to build a Machine Learning model to predict eruption cycle duration at the geyser. The report concludes by revisiting the initial research questions from Section 1 and summarising our findings.

1. Across three periods totalling 90,000+ data points and 66 eruptions, Lady Knox Geyser's eruption cycles can be categorised into short and long events. Short eruptions dominated, accounting 62% (42 eruptions).
2. The intrinsic factor analysed was the peak spring temperature. The strongest relationship for peak temperature and eruption cycle duration was found among the combined periods, yielding a moderately positive correlation ($r = 0.5483$, $p = 1 \times 10^{-6}$, $R^2 = 0.3$). Among external pressure, air temperature showed a potential negative correlation ($r = -0.45$, $p = 1.34 \times 10^{-4}$, $R^2 = 0.21$), although it is still considered a weak model fit. Furthermore, air pressure showed a weaker correlation, and wind speed had none.
3. A Machine Learning model was developed using Systematic Time-Series Feature Engineering (STSFE) with spring temperature as the sole input. Among Machine Learning models tested, the Ridge Model consistently performed best during the initial model evaluation phase and when tested against hold-out data with the least MAE of 9.78 minutes or 20% error from the average eruption duration of 48 minutes.

Based on these promising results, our research will continue with the development of more generalising models, e.g. by integrating the feature selection into a temporal cross-validation.

ACKNOWLEDGEMENTS

This research project was possible through the support of the New Zealand Ministry of Business, Innovation and Employment (MBIE) through the Reversing Carbon Emissions in Geothermal Energy Industry – Template for Emissions – Intensive Industries project funds (UOAX2211). The authors also thank the WTW landowners, the Iwi of Ngati Tahu – Ngati Whaoa, and the WTW concessionaires for allowing access to the Lady Knox Geyser during field data collection.

REFERENCES

Abrasaldo PMB, Zarrouk SJ, Kempa-Liehr AW. 2023. Characterising and predicting low discharge pressure events in less permeable geothermal production wells. Geothermics. 112. https://doi.org/10.1016/j.geothermics.2023.102756

Akiba T, Sano S, Yanase T, Ohta T, Koyama M. 2019. Optuna: A Next-generation Hyperparameter Optimization Framework. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. [place unknown]: Association for Computing Machinery; p. 2623–2631. https://doi.org/10.1145/3292500.3330701

Benjamini Y, Yekutieli D. 2001. The Control of the False Discovery Rate In Multiple Testing Under Dependency. The Annals of Statistics. 29(4):1165–1188.

Christ M, Braun N, Neuffer J, Kempa-Liehr AW. 2018. Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package). Neurocomputing. 307:72–77. https://doi.org/10.1016/j.neucom.2018.03.067

Christ M, Braun N, Neuffer J, Kempa-Liehr AW. 2023. tsfresh.feature_selection package [Internet]. [accessed 2025 Jun 25]. https://tsfresh.readthedocs.io/en/stable/api/tsfresh.feature_selection.html

Christ M, Braun N, Neuffer J, Kempa-Liehr AW. 2025. Overview on extracted features. tsfresh documentation [Internet]. [accessed 2025 Jun 25]. https://tsfresh.readthedocs.io/en/latest/text/list_of_features.html

Cohen J. 1988. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. New York: Lawrence Erlbaum Associates.

Dowling RK. 2011. Geotourism's Global Growth. Geoheritage. 3(1):1–13. https://doi.org/10.1007/s12371-010-0024-7

Harrell FE. 2015. Regression Modeling Strategies With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis [Internet]. 2nd ed. New York: Springer. http://www.springer.com/series/692

Hurwitz S, Kumar A, Taylor R, Heasler H. 2008. Climate-induced variations of geyser periodicity in Yellowstone National Park, USA. The Geological Society of America. 36(6):451–454. https://doi.org/10.1130/G24723A.1

Hurwitz S, Manga M. 2017. The Fascinating and Complex Dynamics of Geyser Eruptions [Internet]. https://doi.org/10.1146/annurev-earth-063016

Ingebritsen SE, Rojstaczer SA. 1993. Controls on Geyser Periodicity. Science (1979). 262(5135):889–892.

Lebe J. 2021. Geothermal features annual monitoring report - June 2021. Waikato Regional Council Technical Report 2021/17 [Internet]. www.waikatoregion.govt.nz

Lloyd. E.F. 1975. Geology of Whakarewarewa Hot Springs. Rotorua: New Zealand Geological Survey.

Luketina K. 2012. The Waikato Regional geothermal resource. Waikato Regional Council Technical Report 2012/10. Hamilton. [Internet]. www.waikatoregion.govt.nz

Marler GD. 1954. Does the Cold of Winter Affect the Thermal Intensity of the Hot Springs in Yellowstone Park. Am J Sci. 252:38–54.

Migoń P, Pijet-Migoń E. 2016. Geoconservation and tourism at geothermal sites – lessons learnt from the Taupo Volcanic Zone, New Zealand. In: Proceedings of the Geologists' Association. Vol. 127. p. 413–421. https://doi.org/10.1016/j.pgeola.2016.04.002

Migoń P, Pijet-Migoń E. 2017. Interpreting Geoheritage at New Zealand's Geothermal Tourist Sites—Systematic Explanation Versus Storytelling. Geoheritage. 9(1):83–95. https://doi.org/10.1007/s12371-016-0185-0

Munoz-Saez C, Namiki A, Manga M. 2015. Geyser eruption intervals and interactions: Examples from El Tatio, Atacama, Chile. J Geophys Res Solid Earth. 120(11):7490–7507. https://doi.org/10.1002/2015JB012364

Nikrou P, Newson J, Mckibbin R, Luketina K. 2013. Geothermal Spring Temperature Analysis. In: 35 th New Zealand Geothermal Workshop:2013 Proceedings. [Internet]. https://www.researchgate.net/publication/273003259

NIWA. 2025. Climate Station daily/hourly data [Internet]. [accessed 2025 May 24]. https://niwa.co.nz/climate-and-weather/climate-data/national-climate-database/climate-station-dailyhourly-data

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research [Internet]. 12:2825–2830. http://scikit-learn.sourceforge.net.

Rinehart JS. 1972. Fluctuations in Geyser Activity Caused by Variations in Earth Tidal Forces, Barometric Pressure, and Tectonic Stresses. J Geophys Res. 77(2):342–350. https://doi.org/10.1029/jb077i002p00342

Shrestha N. 2020. Detecting Multicollinearity in Regression Analysis. Am J Appl Math Stat. 8(2):39–42. https://doi.org/10.12691/ajams-8-2-1