Modeling Cooling Tower Fan Speed Using Symbolic Regression

Mohamed S. Kandil[1], Gary Chang[2] and J.J. McArthur[1]

[1]Dept. Architectural Science, Toronto Metropolitan University, Canada
[2]H.H. Angus & Associates, Canada
mkandil@torontomu.ca, garychang@hhangus.com, jjmcarthur@torontomu.ca

Abstract

Chilled water-cooling systems are essential for maintaining optimal building temperatures in commercial applications, such as hospitals, where precise temperature control is critical for patient care. Accurately estimating cooling tower fan speed is vital for energy efficiency and operational optimization. Traditional physics-based models often require iterative solving techniques, such as Newton-Raphson, leading to high computational costs and slow performance. This paper presents an alternative approach using symbolic regression to develop a data-efficient, explicit model for predicting cooling tower fan speed. Trained on just 72 hours of data, the model was validated on three separate periods: the first week immediately after training, the following week, and another week approximately two months later, achieving R^2 values of 0.92, 0.87, and 0.83, respectively. The inference speed is nearly instantaneous, as the model eliminates iterative calculations, making it well-suited for real-time applications. Using commonly available system parameters, the model enhances both computational efficiency and energy management. Results demonstrate its potential to improve HVAC system performance, reduce energy consumption, and support sustainable building operations, offering a scalable solution for practical deployment in commercial cooling systems.

Keywords –

HVAC, Cooling Tower, Simulation, Symbolic Regression

1 Introduction

Efficient cooling is essential for maintaining optimal environmental conditions in commercial buildings, particularly in facilities like hospitals where temperature control is critical for patient care [1]. Cooling towers, as integral components of chilled water-cooling systems, play a crucial role in rejecting heat from the system. Accurate estimation of the cooling tower's fan speed is vital, as it serves as a proxy for estimating power consumption through the application of the affinity laws, which describe how fan power scales with changes in fan speed. This relationship can be leveraged in simulations to optimize control variables for the condenser water loop, thereby enhancing the cooling system's performance [2]. By balancing the power consumption between chillers and cooling towers, it is possible to minimize the overall energy consumption of the chilled water system, leading to improved operational efficiency and reduced operational costs [3, 4, 5].

Traditionally, physics-based models (also known as gray-box models), such as Merkel-based models, have been employed to simulate cooling tower performance [6]. These models rely on a combination of fundamental physical principles and empirical data, providing robust options for system analysis and power estimation. However, they do not offer a direct output for fan speed, necessitating the use of iterative search methods like the Newton-Raphson technique to solve for fan speed [2]. This requirement for iterative solving introduces complexity and computational inefficiencies, rendering these models less suitable for real-time applications where rapid and responsive control is necessary.

In contrast, data-driven models (or black-box models) present a promising alternative by utilizing machine learning techniques to learn the relationship between input parameters—such as temperature, flow rate, and environmental conditions—and output variables like fan speed directly from data. While data-driven models can achieve high accuracy, they often suffer from limitations in generalization capability, particularly when operating outside the range of the training data. Additionally, these models typically require large datasets for effective training, which can limit their applicability in scenarios where data availability is constrained or where dynamic, real-time system responses are needed [7].

The motivation for this research lies in overcoming the limitations of both gray-box and black-box approaches by developing an interpretable and data-efficient model using symbolic regression (SR). Symbolic regression provides a unique methodology that automatically discovers mathematical expressions to describe the relationships between input and output variables [8]. Unlike traditional machine learning methods, SR generates transparent, closed-form mathematical models that are both interpretable and computationally efficient [9]. This makes SR particularly suitable for real-time control strategies, as the resulting models eliminate the need for iterative numerical solutions required by gray-box models while avoiding the black-box nature and data dependency of conventional machine learning techniques.

2 Objectives of the Study

The primary objectives of this research are as follows:

1. Develop a data-efficient explicit model for cooling tower fan speed using symbolic regression, incorporating key input variables such as wet-bulb temperature, inlet and outlet water temperatures, and water flow rate, while minimizing the need for large datasets.
2. Improve simulation speed and computational efficiency by providing an explicit, direct model for fan speed estimation, thereby eliminating the need for iterative solutions required by traditional physics-based models and enabling faster real-time decision-making.
3. Leverage symbolic regression to create an interpretable and generalizable data-driven model that performs well with limited data and can be adopted in real-time applications, offering a practical and computationally efficient alternative to black-box models that typically require large, extensive datasets.

By addressing these objectives, this research aims to present a data-efficient, computationally efficient, and interpretable approach for estimating fan speed in chilled water-cooling systems. This advancement has the potential to enhance energy optimization, improve real-time control mechanisms, and offer a practical solution for systems with limited data, thereby contributing to more sustainable and cost-effective building operations.

3 Literature Review

Cooling towers are critical components of HVAC systems, responsible for rejecting heat from processes or buildings to the atmosphere. Their operational efficiency and energy consumption depend significantly on the control of fan speed, which in turn determines the leaving water temperature. Accurately predicting this temperature is essential for improving energy efficiency and ensuring optimal performance [3, 10, 11]. To model the leaving water temperature, three primary methods have been developed over the years: the Merkel method, the Poppe method, and the enthalpy-driven effectiveness-NTU method [6]. These models have served as the foundation for cooling tower analysis and control strategies.

The Merkel method, first proposed in 1925, simplifies the cooling process by assuming a constant Lewis factor (the ratio of sensible to latent heat transfer coefficients) and neglecting the evaporation losses of water. It integrates the heat and mass transfer equations along the height of the cooling tower to estimate the leaving water temperature [6]. While this approach has been widely adopted for its simplicity and ease of implementation, it fails to capture the complex interactions between heat and mass transfer that occur in real operating conditions, particularly under varying ambient and operational parameters.

The Poppe method improves upon Merkel's assumptions by considering the simultaneous and coupled heat and mass transfer processes inside the cooling tower. Unlike Merkel's method, the Poppe approach does not assume a constant Lewis factor and accounts for variations in the air and water enthalpy. This method provides higher accuracy in predicting the leaving water temperature, particularly in scenarios where Merkel's assumptions lead to discrepancies [12]. However, the Poppe model requires solving a system of coupled nonlinear differential equations, which makes it computationally intensive and less suited for real-time control.

The enthalpy-driven effectiveness-NTU method provides a more flexible and systematic framework for predicting the leaving water temperature. By introducing the concept of heat exchanger effectiveness, this method evaluates cooling tower performance using enthalpy as the driving force. It allows for improved accuracy in capturing cooling tower dynamics under varying conditions compared to traditional methods [2]. Despite its advantages, the enthalpy-driven method still involves solving nonlinear relationships, which can pose challenges for real-time optimization and control strategies. Another, simpler model was proposed based on Merkel’s theory and the effectiveness-NTU method. This model was developed through energy balance and heat-mass transfer analysis [13].

For a closed-loop cooling tower leaving water temperature control strategy, the fan speed must be set to its optimal value to maintain the optimal leaving water temperature. However, if the desired leaving water temperature setpoint deviates from this optimal condition, determining the appropriate fan speed requires solving the governing equation numerically. Equation (1) illustrates this relationship [2]:

$$T_{CWS} = f_T(T_{CWR}, T_{WB}, M_{CW}, SPD, K_d'A) \quad (1)$$

where T_{CWS} is the leaving water temperature, T_{CWR} is the entering water temperature, T_{WB} is the ambient air wet bulb temperature, M_{CW} is the water mass flow rate, SPD is the fan speed, and $K_d'A$ the design moisture transfer coefficient.

Solving for SPD under these conditions requires iterative numerical techniques such as the Bisection method or Newton-Raphson method, which are widely used for solving nonlinear equations. These methods, while effective, are computationally demanding and time-consuming, especially for real-time applications where fast control responses are essential [2, 3]. Similarly, for closed-loop approach temperature control, the fan speed must also be determined iteratively if the approach temperature setpoint differs from the optimal value. These computational requirements can lead to delays in control responses, limiting their feasibility for real-time applications [3]. Thus, while the Merkel, Poppe, and enthalpy-driven NTU methods provide robust frameworks for cooling tower performance prediction, their reliance on computationally intensive numerical solvers introduces significant overhead during closed-loop control.

To overcome the computational challenges associated with traditional models, data-driven approaches have gained increasing attention. Data-driven methods utilize operational data to identify complex system dynamics and provide real-time predictive models. Among these approaches, symbolic regression (SR) has emerged as a powerful technique for deriving explicit and interpretable mathematical expressions from data [8, 9]. SR identifies relationships between variables without relying on predefined model structures, enabling the discovery of closed-form equations that can directly describe the behavior of cooling towers, including fan speed and leaving water temperature.

Although symbolic regression has been applied successfully in other areas of HVAC systems, such as building thermodynamics [14] and gray-box modeling of air conditioning systems [15], its application to cooling tower modeling remains underexplored. [14] used SR to model building energy dynamics and demonstrated a 16.1% reduction in peak power through a predictive control framework. [15] applied SR for gray-box modeling of unitary air conditioning systems, achieving high accuracy in predicting cooling capacity and coefficient of performance (COP). [16] further showcased the potential of SR for indoor temperature forecasting, reducing model complexity while improving computational efficiency. These studies highlight SR's ability to capture nonlinear dynamics effectively while providing explicit and interpretable solutions, making it a suitable candidate for real-time applications.

In this research, we address the limitations of traditional cooling tower models and propose a symbolic regression-based model to predict cooling tower fan speed as a function of the leaving water temperature, entering water temperature, water mass flow rate, and wet bulb temperature. Unlike traditional methods, which require iterative numerical solutions to determine fan speed for a given setpoint, our approach formulates an explicit and interpretable model derived directly from operational data.

The proposed symbolic regression (SR) model eliminates the need for computationally intensive techniques, such as the Bisection method or Newton-Raphson iterations, which are typically employed to solve the governing nonlinear equations. By generating closed-form expressions, the SR model enables faster and more efficient inference, making it particularly well-suited for closed-loop control strategies. This capability allows for real-time adjustments of fan speed to maintain desired leaving water temperature or approach temperature setpoints. Furthermore, the resulting SR model is computationally efficient and highly interpretable, providing clear relationships between key operational variables—leaving and entering water temperatures, wet bulb temperature, and water mass flow rate. This facilitates real-time optimization of cooling tower performance, overcoming the bottlenecks inherent to traditional numerical approaches while ensuring robust and accurate control of cooling tower operations.

4 Methodology

4.1 Case study and Data Collection

The primary case study for this research is a hospital located in the Greater Vancouver area, within ASHRAE climate zone 4. This hospital is one of the busiest in the region, providing comprehensive care across multiple specialties, including trauma, cardiac, neurosurgery, high-risk obstetrics, neonatal intensive care, and acute mental health services. The diverse range of services necessitates a robust and efficient cooling system to maintain optimal environmental conditions, which are critical for patient care. The hospital's central cooling plant, comprising chillers and cooling towers, is responsible for meeting the facility's substantial cooling demands.

Data for this study were collected from the central cooling system of the hospital over a five-month period, from May 1, 2024, to September 30, 2024. However, only 72 hours of data, specifically from July 8th to July

10th, were utilized for training the models (see Figure 1). This subset consists of 288 data points, sampled at 15-minute intervals, capturing the system's dynamic behaviour. Representing approximately 2% of the total collected data, these 72 hours were carefully selected to include a wide range of operational conditions, such as fluctuations in outdoor temperature, humidity, and cooling demand. This ensures a diverse and representative dataset that effectively captures the cooling system's performance across different operating states.

The collected variables include leaving water temperature (T_{CWS}, in °C), entering water temperature (T_{CWR}, in °C), condenser water mass flow rate (M_{CW}, in L/s), dry-bulb outdoor air temperature (OAT, in °C), relative humidity (RH, in %), and cooling tower fan speed (SPD, in %). The fan speed during this period varied, with a lower bound of 33.3%, and its operational range was crucial for evaluating the system's performance.

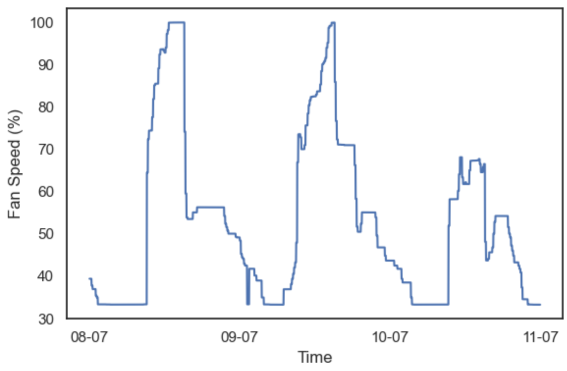

Figure 1: Cooling tower fan speed dataset used for model training

4.2 Data Pre-processing and Feature Engineering

To ensure the quality and relevance of the data for modeling, several pre-processing and feature engineering steps were undertaken. Wet-Bulb Temperature was calculated using both Relative Humidity and Air Temperature based on the formula provided by [17]. This additional feature was included to better capture the thermodynamic conditions affecting the cooling tower performance.

The raw dataset underwent a series of filtering criteria to remove unreliable or irrelevant measurements. First, instances where the fan speed (SPD) was below 33.3% were discarded to eliminate periods of suboptimal operation. Additionally, data points with a wet-bulb temperature below 10°C were excluded to focus the analysis on conditions where cooling demands are more significant. Measurements where the condenser water supply temperature (T_{CWS}) exceeded the condenser water return temperature (T_{CWR}) were also removed, as such scenarios are physically implausible and likely indicative of sensor errors or operational issues.

Further, any records where the condenser water mass flow rate (M_{CW}) fell below 50 L/s were filtered out to ensure the dataset only included stable and efficient operating conditions. Finally, rows containing NaN values in any of the input or target variables were discarded to maintain data integrity and avoid computational issues during model training.

After the pre-processing steps, a total of 283 data points remained, reflecting the impact of the filtering criteria in removing unreliable or irrelevant data. This cleaned dataset was subsequently split into training and testing subsets, with an 80-20% division, respectively. The split was performed without shuffling to preserve the temporal sequence of the data, which is critical given the time-series nature of the dataset. For modeling, the fan speed (SPD) was designated as the target variable, while the remaining variables were selected as the input features. This structured and pre-processed dataset provided a robust foundation for the subsequent symbolic regression modeling and analysis.

4.3 Symbolic Regression Modeling with PySR

The modeling process was conducted using Python, leveraging the PySR package [18] for symbolic regression. Symbolic regression is a data-driven approach that searches the space of mathematical expressions to find models that best fit the data. Unlike traditional regression methods that require predefined forms of the model, symbolic regression explores a wide variety of possible equations, offering both accuracy and interpretability [8].

PySR employs a genetic algorithm-based approach to evolve mathematical expressions, optimizing them to minimize the error between the predicted and actual values. The underlying computational engine utilized Julia, a high-performance programming language, to enhance the efficiency and scalability of the symbolic regression process. This integration allows for rapid evaluation of numerous candidate models, facilitating the discovery of optimal or near-optimal symbolic representations of the target variable.

In this study, the symbolic regression was configured with specific hyper-parameters to guide the model generation process. The binary operators included in the search space were addition (+), multiplication (*), subtraction (-), division (/), and exponentiation (^). The loss function was set to L2DistLoss(), which corresponds to the mean squared error between the predicted and actual fan speeds. Model selection was based on accuracy to ensure that the generated models achieved high predictive performance.

To accelerate the training process, several performance settings were enabled. The multithreading setting was activated to utilize multiple CPU cores, improving computational efficiency. The fast_cycle setting was set to False, which prevents the batch over population subsamples algorithm, designed to speed up the cycles by approximately 15%, but at the risk of being less algorithmically efficient. The turbo setting was also False, as it refers to the experimental use of LoopVectorization.jl, which speeds up the search evaluation but may not support certain operators and does not work with 16-bit precision floats. Finally, the bumper setting was turned off; this experimental feature uses Bumper.jl to speed up the search evaluation, though it also does not support 16-bit precision floats. These configurations were carefully chosen to strike a balance between computational efficiency and model accuracy, ensuring that the symbolic regression process remained robust, scalable, and suitable for generating interpretable models for cooling tower fan speed prediction.

4.4 Model Training and Validation

The training phase involved fitting the symbolic regression model to the training dataset using the specified hyper-parameters and performance settings. The model was trained to identify the mathematical relationship between the input features (T_{CWS}, T_{CWR}, T_{WB}, and M_{CW}) and the target variable (SPD). This relationship can be mathematically represented in the following general form:

$$SPD = f_{SPD}(T_{CWS}, T_{CWR}, T_{WB}, M_{CW}) \tag{2}$$

Once the model was generated, its performance was evaluated using two key metrics: R-squared (R^2) to measure the proportion of variance in the target variable that is predictable from the input features and Root Mean Squared Error (RMSE) to provide an absolute measure of the prediction error in the same units as the target variable.

To assess the model’s generalizability, validation was performed on an independent dataset comprising two weeks of subsequent data. This validation step was crucial to ensure that the symbolic regression model maintained its accuracy and reliability when applied to new, unseen data. The performance metrics on the validation set provided an indication of the model's robustness and its potential applicability in real-world cooling system optimization.

Overall, the methodology combines rigorous data collection and preprocessing with advanced symbolic regression techniques to develop an accurate and interpretable model for predicting cooling tower fan speed. The use of PySR and its efficient computational backend facilitates the discovery of optimal mathematical models that can enhance the understanding and control of cooling system operations.

5 Results

5.1 Symbolic Regression Model

The best-performing model derived through symbolic regression is represented by the following mathematical expression:

$$\begin{aligned} SPD = {} & 0.707M_{cw} + \frac{0.707M_{cw}}{T_{WB}} - \frac{13.2M_{cw}}{T_{cwr}} \\ & + {T_{cwr}}^{\frac{0.764T_{WB}}{T_{cwr}-9.71}} - 13.5 \\ & - {M_{cw}}^{-0.878M_{cw}} \end{aligned} \tag{3}$$

This equation provides a direct relationship between cooling tower fan speed and the key input features: wet-bulb temperature, condenser water supply temperature, condenser water return temperature, and mass flow rate of condenser water. The model is interpretable and captures the nonlinear interactions among the input variables.

5.2 Performance Evaluation on Test Data

The model's performance was first evaluated on the designated test dataset, comprising 20% of the data selected from the initial 72-hour training period. The following metrics were obtained:

- R^2 score: 0.8881
- RMSE: 3.9325

These results indicate that the model captures a significant portion of the variability in the fan speed while maintaining low prediction error. The R^2 score of 0.89 demonstrates high predictive accuracy, while the RMSE of approximately 3.93 reflects the model's strong ability to estimate fan speed with minimal deviation from the actual values.

5.3 Performance Evaluation on Validation Data

To assess the model's generalization capability, it was validated on an independent dataset across three separate periods: the first week after training (July 11th to 18th), the following week (July 18th to 25th), and an additional week nearly two months later (September 1st to 8th), demonstrating its credibility and robustness.

For the week of July 11th to 18th, 2024, the model achieved an R^2 of 0.92 and an RMSE of 3.76, demonstrating its strong predictive accuracy and robustness in estimating fan speed. The time-series plot (Figure 2) illustrates a close alignment between actual and predicted values, while the scatter plot (Figure 3) confirms a strong correlation, with most points closely following the diagonal reference line, indicating minimal prediction error.

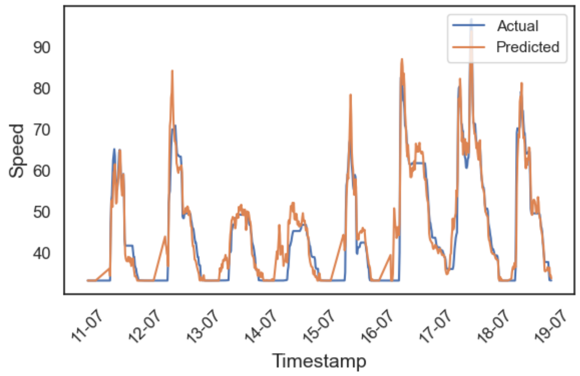

Figure 2 Time series plot of actual versus prediction values for first week of validation data

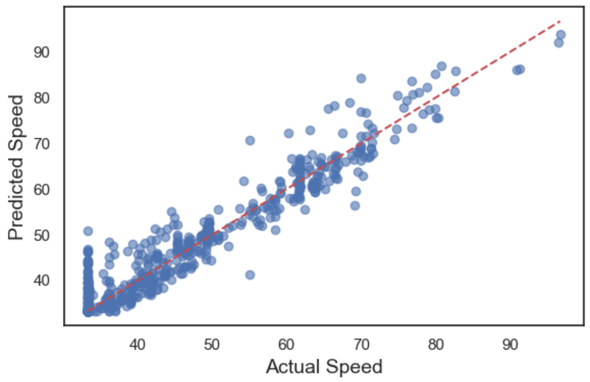

Figure 3 Scatter plot of actual versus prediction values for first week of validation data

For the week of July 18th to 25th, 2024, the model achieved an R^2 of 0.87 and an RMSE of 3.82, maintaining strong predictive performance and consistency in estimating fan speed. The time-series plot (Figure 4) illustrates a close match between actual and predicted values, while the scatter plot (Figure 5) confirms a strong correlation, with most points aligning along the diagonal reference line. The slight decline in R^2 suggests minor operational variations, yet the model continues to generalize well, reinforcing its reliability in diverse conditions.

To evaluate the performance and generalizability of the model, it was further validated on an independent dataset covering the week of September 1st to 8th, 2024, achieving an R^2 of 0.83 and an RMSE of 6.20. While the slight decline in R^2 suggests some operational variations, the model remains effective in capturing key system dynamics. The time-series plot (Figure 6) shows strong alignment between actual and predicted values, while the scatter plot (Figure 7) confirms a consistent correlation, reinforcing the model's reliability in predicting fan speed under varying conditions.

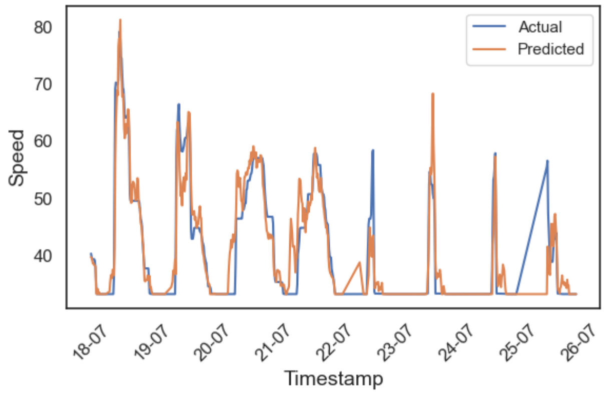

Figure 4 Time series plot of actual versus prediction values for second week of validation data

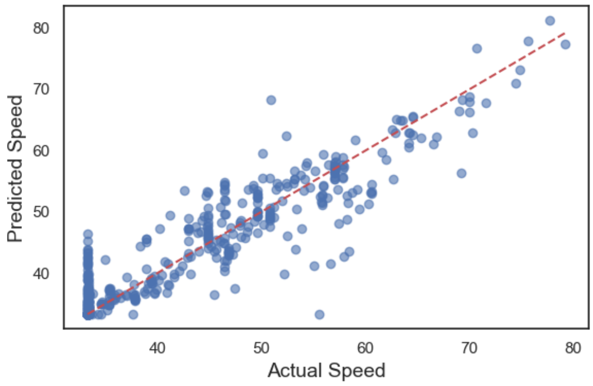

Figure 5 Scatter plot of actual versus prediction values for second week of validation data

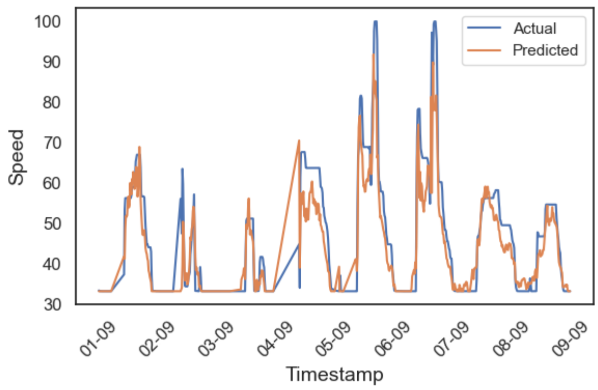

Figure 6 Time series plot of actual versus prediction values for third week of validation data

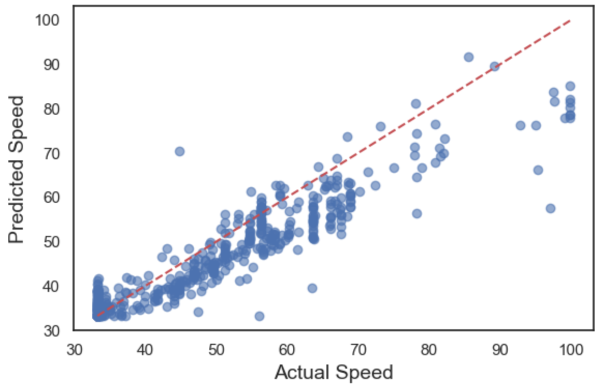

Figure 7 Scatter plot of actual versus prediction values for third week of validation data

6 Discussion

On the test dataset (20% of the training period), the model achieved an R^2 score of 0.8881 and an RMSE of 3.9325, capturing nearly 89% of the variance in fan speed with minimal prediction error. Further validation confirmed its generalizability. During the first validation week (July 11–18, 2024), the model achieved an R^2 of 0.9206 and an RMSE of 3.7566, demonstrating strong agreement between predicted and actual values. For the second validation week (July 18–25, 2024), the model maintained an R^2 of 0.8676 and an RMSE of 3.8246, despite slight variations likely caused by operational changes. A third validation nearly two months later (September 1–8, 2024) resulted in an R^2 of 0.8349 and an RMSE of 6.1975. While the R^2 declined slightly over time, the model remained effective, confirming its robustness under changing conditions.

The symbolic regression model eliminates the need for iterative numerical methods, such as Newton-Raphson, which are computationally intensive and impractical for real-time applications. By generating explicit mathematical expressions, the model enables fast inference, allowing real-time adjustments to cooling tower fan speed to maintain desired operating conditions.

Beyond computational efficiency, the model supports practical energy optimization strategies for chilled water plants, where balancing chiller and cooling tower operation is critical to minimizing total energy consumption. Cooling tower fans consume energy, but their operation influences chiller efficiency; optimizing only one component in isolation may lead to suboptimal system performance. The proposed model can be integrated into real-time optimization frameworks to determine the best trade-off between chiller and cooling tower energy use, reducing overall power consumption while ensuring adequate cooling.

With its fast inference capability, the model enables real-time control adjustments based on changing conditions, preventing excessive cooling tower fan speeds that could increase energy use unnecessarily. Its explicit equations provide a direct relationship between key operational parameters, making it suitable for predictive control applications where chiller load and cooling tower operation are jointly optimized for efficiency.

The model's interpretability also facilitates decentralized control, where cooling tower fan speed is adjusted at the local level rather than relying on a central controller. This reduces communication delays and computational overhead while improving responsiveness to system dynamics. In large-scale systems with multiple cooling towers, localized optimization enhances overall performance, leading to energy savings and improved reliability.

By integrating the model into smart HVAC control platforms, facilities can automate real-time energy management strategies to optimize cooling performance while reducing costs. This scalable approach makes the model applicable to hospitals, commercial buildings, and industrial facilities where precise control and efficiency are essential.

In summary, the symbolic regression model provides a practical, real-time solution for cooling tower control. By balancing chiller and cooling tower energy use, it enhances system-wide efficiency, offering a valuable tool for intelligent HVAC management in energy-conscious applications.

7 Conclusions

This study demonstrated the effectiveness of symbolic regression for predicting cooling tower fan speed in central chilled-water systems. The SR model achieved high accuracy, strong generalization, and computational efficiency across test and validation datasets. By producing explicit and interpretable mathematical models, it overcomes the limitations of traditional physics-based approaches and data-hungry black-box techniques.

The proposed model offers a practical and scalable solution for real-time optimization and decentralized control of cooling systems, facilitating improved energy efficiency and operational performance. Future work will focus on enhancing the model's adaptability to dynamic system changes and integrating it into broader energy management frameworks to further optimize large-scale cooling operations.

Acknowledgments

This research was funded by the Natural Science and Engineering Research Council (Canada) [ALLRP-580958-22] and H. H. Angus and Associates Consulting Engineers Ltd.

Bibliography

[1] D. Schurk, "Conditioning for the Environment of Critical Care: Hospital Operating Rooms," *ASHRAE Journal*, vol. 61, no. 10, pp. 16-24, 2019.

[2] G. Wang, S. Ghoddousi, L. Ding and L. Song, "Investigation of Different Cooling Tower Fan Control Strategies Using Actual Chiller Cop and Calibrated Cooling Tower and Fan Models," *Energy and Buildings*, vol. 277, p. 112585, 2022.

[3] J. Liao, X. Xie, H. Nemer, D. E. Claridge and C. H. Culp, "A simplified methodology to optimize the cooling tower approach temperature control schedule in a cooling system," *Energy Conversion and Management*, vol. 199, p. 111950, 2019.

[4] K. Ma, M. Liu and J. Zhang, "Online optimization method of cooling water system based on the heat transfer model for cooling tower," *Energy*, vol. 231, p. 120896, 2021.

[5] K. Zaw, K. Matsuoka and T. Poh, "A real-time operational optimization of commercial district cooling systems: A practical application," *Energy and Buildings*, vol. 297, p. 113434, 2023.

[6] J. C. Kloppers and D. G. Kröger, "Cooling Tower Performance Evaluation: Merkel, Poppe, and e-NTU Methods of Analysis," *J. Eng. Gas Turbines Power*, vol. 127, no. 1, pp. 1-7, 2005.

[7] S. Ikeda and T. Nagai, "A novel optimization method combining metaheuristics and machine learning for daily optimal operations in building energy and storage systems," *Applied Energy*, vol. 289, p. 116716, 2021.

[8] G. Kronberger, B. Burlacu, M. Kommenda, S. M. Winkler and M. Affenzeller, Symbolic Regression, CRC Press, 2024.

[9] D. Angelis, F. Sofos and T. E. Karakasidis, "Artificial Intelligence in Physical Sciences: Symbolic Regression Trends and Perspectives," *Archives of Computational Methods in Engineering*, vol. 30, p. 3845–3865, 2023.

[10] K. Singh and R. Das, "Exergy optimization of cooling tower for HGSHP and HVAC applications," *Energy Conversion and Management*, vol. 136, pp. 418-430, 2017.

[11] G. F. Cortinovis, J. L. Paiva, T. W. Song and J. M. Pinto, "A systemic approach for optimal cooling tower operation," *Energy Conversion and Management*, vol. 50, no. 9, p. 2200–2209, 2009.

[12] D. R. Baker and H. A. Shryock, "A Comprehensive Approach to the Analysis of Cooling Tower Performance," *Journal of Heat Transfer*, vol. 83, no. 3, pp. 339-349, 1961.

[13] G.-Y. Jin, W.-J. Cai, L. Lu, E. L. Lee and A. Chiang, "A simplified modeling of mechanical cooling tower for control and optimization of HVAC systems," *Energy Conversion and Management*, vol. 48, no. 2, pp. 355-365, 2007.

[14] Y. Ozawa, D. Zhao, D. Watari, I. Taniguchi, T. Suzuki, Y. Shimoda and T. Onoye, "Data-driven HVAC Control Using Symbolic Regression: Design and Implementation," in *2023 IEEE Power & Energy Society General Meeting (PESGM)*, Orlando, FL, USA, 2023.

[15] S. Yousaf, C. R. Bradshaw, R. Kamalapurkar and O. San, "A gray-box model for unitary air conditioners developed with symbolic regression," *International Journal of Refrigeration*, vol. 168, pp. 696-707, 2024.

[16] D. Zhao, Z. Chen, Z. Li, Y. Yuan and I. Taniguchi, "Improving Building Temperature Forecasting: A Data-driven Approach with System Scenario Clustering," in *2024 IEEE Power & Energy Society General Meeting (PESGM)*, Seattle, WA, USA, 2024.

[17] R. Stull, "Wet-Bulb Temperature from Relative Humidity and Air Temperature," *Journal of Applied Meteorology and Climatology*, vol. 50, no. 11, p. 2267–2269, 2011.

[18] M. Cranmer, "Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl," *arXiv preprint arXiv:2305.01582*, 2023.