Machine learning-based power prediction for the icebreaker Henry Larsen using vessel motions and environmental data in open water conditions

Chittamuru, Samarasimha Reddy; Matthew, Hamilton; Akinturk, Ayan; Kennedy, Allison; Barnes, Joshua; Singh, Balsher

Machine Learning-Based Power Prediction for the Icebreaker Henry Larsen Using Vessel Motions and Environmental Data in Open Water Conditions

Samarasimha Reddy Chittamuru
Ocean, Coastal, and River Engineering
National Research Council
St. John's, Canada
samarasimhareddy.chittamuru@nrc-cnrc.gc.ca

Hamilton Matthew
Department of Computer Science
Memorial University of Newfoundland
St. John's, Canada
mhamilton@mun.ca

Ayhan Akinturk
Ocean, Coastal, and River Engineering
National Research Council
St. John's, Canada
ayhan.akinturk@nrc-cnrc.gc.ca

Allison Kennedy
Ocean, Coastal, and River Engineering
National Research Council
St. John's, Canada
allison.kennedy@nrc-cnrc.gc.ca

Joshua Barnes
Ocean, Coastal, and River Engineering
National Research Council
St. John's, Canada
joshua.barnes@nrc-cnrc.gc.ca

Balsher Singh
Ocean, Coastal, and River Engineering
National Research Council
St. John's, Canada
balsher.singh@nrc-cnrc.gc.ca

***Abstract*—Accurate power prediction for a ship vessels is critical for improving operational efficiency and reducing fuel consumption, hence, shipborne emissions. This paper presents a machine learning approach for predicting the required power for ship at a given operational / environmental conditions using time series data that combines vessel telemetry, including speed and motions, with environmental data from ECMWF (European Centre for Medium-Range Weather Forecasts). Focusing on open water conditions to avoid ice interference, we apply ensemble methods, specifically Random Forest and XGBoost (Extreme Gradient Boost), and evaluate their performance using Time Series Split and Block Time Series Split techniques to handle temporal dependencies. Our models are assessed using Root Mean Squared Error, Mean Absolute Error, and R-squared. In this paper, the Canadian Coast Guard Vessel is used as a case study. The results demonstrate the effectiveness of machine learning in predicting vessel power and highlight the importance of selecting appropriate data splitting strategies to prevent data leakage.**

***Index Terms*—Machine Learning, Power Prediction, Time series split, Ice breaker vessel motion, Data leakage.**

I. Introduction

Power prediction for ships serves multiple purposes, including improving operational efficiency, optimizing fuel consumption, and calculating emissions. With the increasing focus on reducing greenhouse gas (GHG) emissions, especially under the guidelines set by the International Maritime Organization (IMO), accurate power prediction has become essential for both operational and environmental reasons. The IMO continues to push for significant reductions in GHG emissions from ships, with its current strategy targeting a 40% reduction in CO_2 emissions by 2030 and 70% by 2050, compared to 2008 levels. These stringent environmental requirements, set by the Marine Environment Protection Committee (MEPC), highlight the need for accurate methods to evaluate ship performance [1]. Power prediction is crucial because it directly influences the calculation of key performance metrics, such as engine load, fuel consumption, and emissions. By accurately predicting the propulsion power, ship operators can estimate fuel consumption and monitor emissions more effectively, ensuring compliance with IMO regulations while optimizing operational performance.

In this paper, we develop a machine learning framework for predicting the power consumption of the Canadian Coast Guard Icebreaker CCGS *Henry Larsen* under open water (no ice) conditions. We leverage time series data that combines detailed vessel telemetry, including speed and motion metrics, with environmental data sourced from the ERA5 (fifth-generation European Centre for Medium-Range Weather Forecasts reanalysis for the global climate and weather) hindcast model [2], as well as wind data collected from the ship's onboard sensors. We apply machine learning models to capture the relationship between operational and environmental factors and evaluate their performance.

II. DATA

A. Sources

The dataset used in this study is composed of two primary data sources: vessel telemetry data collected directly from the icebreaker and environmental data sourced from the ERA5 reanalysis model. The vessel telemetry data includes key operational metrics such as speed over ground (SOG), heading (HDT), wind speed and direction collected from the vessel's OpDAQ^{TM} system, which has a sampling frequency of 1Hz, and vessel motions, including pitch, roll, and accelerations (X, Y, and Z directions) from the MTi-300 (IMU sensor) with a sampling frequency of 20Hz [3].

To complement the vessel's telemetry, external environmental data was sourced from the ERA5 reanalysis model, which provides hourly estimates of wind speed, wind direction, wave direction, and swell height. This environmental data plays a crucial role in understanding the external conditions that influence the vessel's power consumption.

As shown in Table I, the dataset is categorized into three main groups: Ship Parameters, Motion Parameters, and Environmental Parameters, each containing relevant features for the power prediction task.

TABLE I
CATEGORIZATION OF PARAMETERS

Category	Features
Ship Parameters	Speed Over Ground (SOG), Heading (HDT), Wind Speed, Wind Direction, Power
Motion Parameters	Pitch, Roll, X, Y, Z Accelerations, Pitch Rate, Roll Rate
Environmental Parameters	Mean Direction of Total Swell, Mean Period of Total Swell, Significant Height of Total Swell, Mean Wave Direction, Mean Wave Period, Peak Wave Period

B. Data Cleaning and Preperation

The data used in this study follows the NMEA (National Marine Electronics Association) standards [4], meaning that each channel includes data quality codes and references to indicate the reliability of the recorded measurements. As part of the data cleaning process, only data with the highest quality indicators were retained to ensure the integrity of the analysis. This filtering ensures that only the most accurate and reliable telemetry data is used for predicting power consumption. The dataset spans from 2020 to 2023, covering several years of vessel operations.

Since the icebreaker operates in various modes, including icebreaking and open water navigation, it is essential to account for different operational conditions. However, due to the absence of direct information on the vessel's operational mode, predicting power across all modes would introduce unnecessary complexity and uncertainty. To mitigate this issue, we focused on open water conditions by filtering the dataset to include only high speed region, where the vessel's Speed Over Ground (SOG) exceeds 12 knots. This speed threshold reasonably ensures that the ship is not operating in icebreaking mode and is considered to be in open water, which simplifies the power prediction task by excluding ice-related operations.

In addition to the vessel's operational data, environmental data from the ERA5 reanalysis model was incorporated. The ERA5 data was geospatially matched to the vessel's location using the nearest grid point without interpolation. The ERA5 grid has a spatial resolution of approximately 31 km (0.25° x 0.25°) [2], providing global atmospheric and surface conditions. Key environmental variables such as wind speed, wind direction, wave direction, and swell height were integrated from ERA5 to complement the vessel telemetry data.

C. Feauture Engineering

To enhance the predictive performance of the machine learning models, two key features were incorporated from the vessel's operational and environmental data.

- Relative Directions: The relative direction of environmental factors, such as waves and swell, in relation to the vessel's heading was calculated. These relative directions provide essential information about how external forces influence power consumption under varying conditions.
- Categorical Binning of Directions: Inspired by traditional nautical terms that describe the relationship between a vessel's course and incoming waves, such as Beam sea, Bow sea, Head sea, Quartering sea, and Overtaking (following) sea, the relative wave and swell directions were categorized into 30-degree intervals. This classification captures the different impacts of environmental forces depending on their direction relative to the vessel's heading, helping the models better interpret how sea conditions influence power consumption.

Additionally, to understand how each feature impacts power consumption, a correlation analysis was performed between Power and other available features. As shown in Figure 1, this analysis highlights the most influential parameters for power consumption prediction. Such insights allow us to focus on features that are strongly correlated with power, improving model performance.

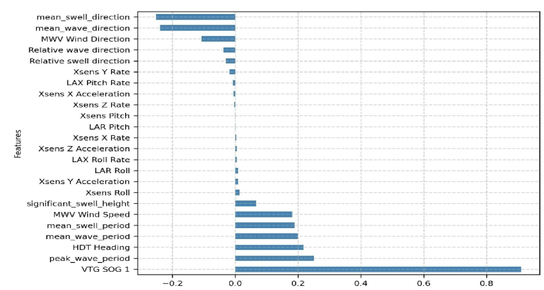

Fig. 1. Correlation of Power with Other Features

III. METHODOLOGY

Predicting the power consumption of an icebreaker such as the Henry Larsen requires understanding the intricate interplay between ship operations, motion dynamics, and external environmental conditions. To capture these relationships effectively, we categorized the data into three primary groups: Ship Parameters, Motion Parameters, and Environmental Parameters, as shown in Table I. From these categories, we created three distinct scenarios—Scenario A, Scenario B, and Scenario C—each designed to explore how different combinations of operational, motion, and environmental factors influence power consumption. The feature combinations for these scenarios are detailed in Table II.

TABLE II
SCENARIO-BASED FEATURE COMBINATIONS

Scenario	Features
Scenario A: Operational Data + Environmental Data	Ship Parameters (SOG, Wind Speed, Wind Direction) + Environmental Parameters (Wave, Swell Data)
Scenario B: Operational + Motion Data	Ship Parameters + Motion Parameters (Pitch, Roll, Accelerations, Pitch Rate, Roll Rate)
Scenario C: Full Feature Set	Combination of Ship Parameters, Motion Parameters, and Environmental Parameters (Wave and Swell Data from ERA5)

A. Machine Learning Models

Initial experimentation using scikit-learn [5] with traditional machine learning models, such as linear regression and decision trees, did not yield satisfactory results for predicting the complex and dynamic power consumption of the CCGS Henry Larsen. Given the intricate relationships between the vessel's operations, motion, and environmental conditions, these simpler models failed to capture the complex underlying patterns.

As a result, we opted for ensemble learning methods, specifically Random Forest [6] and XGBoost [7], which are better suited for handling non-linear interactions and high-dimensional datasets:

- Random Forest is an ensemble method that constructs multiple decision trees during training and merges their outputs to improve prediction accuracy. It is particularly effective for handling datasets with mixed feature types (categorical and continuous) and provides inherent feature importance rankings, helping interpret model decisions.
- XGBoost is a high-performance gradient-boosting algorithm. It builds decision trees sequentially, optimizing each tree to correct the errors of the previous one. XGBoost is known for its scalability, ability to handle large datasets, and regularization features, which help prevent overfitting.

Both models were selected for their ability to model complex feature interactions and their success in time-series forecasting tasks. The models were applied to the three feature scenarios (A, B, and C) to determine the most effective combination of features for predicting power consumption.

B. Addressing Data Leakage

Given the time-series nature of the dataset, preventing data leakage was crucial. We implemented two validation techniques tailored specifically for time-series data to preserve its sequential integrity [8]:

- Time Series Cross-Validation (TSCV): This method maintains the temporal order of the data, ensuring that the model is trained on past data and validated on future data. In each fold, the training set is expanded while validation is conducted on the subsequent time segment.
- Block Time Series Cross-Validation (BTSCV): This technique introduces a gap between training and validation sets to account for potential autocorrelation in the data. BTSCV is more conservative, ensuring no overlap between training and validation, thus mitigating data leakage risks.

Figure 2 illustrates the structural differences between TSCV and BTSCV folds. The key distinction lies in the margin introduced in BTSCV, which further ensures temporal dependencies are excluded from validation sets. This dual approach allows us to rigorously evaluate the models, preserving the time-series patterns and ensuring robust performance assessments.

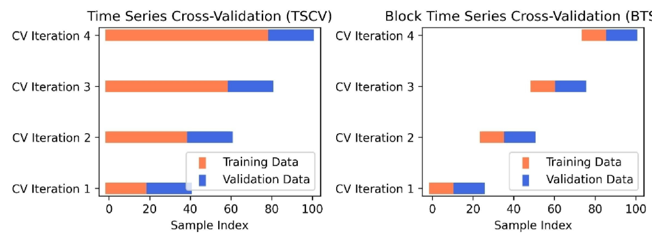

Fig. 2. Time Series and Block Time Series Cross-Validation Splits

C. Hyperparameter Tuning and Model Optimization

Given the complexity of the data and the time constraints of model training, we employed RandomizedSearchCV for hyperparameter tuning in conjunction with Time Series Cross-Validation (TSCV) and Block Time Series Cross-Validation (BTSCV). RandomizedSearchCV was chosen over traditional grid search due to its efficiency in exploring a wider hyperparameter space. Unlike grid search, which exhaustively evaluates all possible hyperparameter combinations, RandomizedSearchCV samples from the hyperparameter distribution, reducing computational cost while yielding highly optimized models.

Figure 3 shows the hyperparameter grid for both Random Forest and XGBoost. By using TSCV and BTSCV, the temporal order of the data is preserved, preventing data leakage and improving model reliability.

	Parameters	Random Forest	XGBoost
0	n_estimators	100, 200, 300, 400, 500	100, 200, 300, 400, 500
1	max_depth	5, 10, 15, 20, 25	3, 5, 7, 10
2	min_samples_split	2, 5, 10	-
3	min_samples_leaf	1, 2, 4	-
4	bootstrap	True, False	-
5	learning_rate	-	0.01, 0.05, 0.1
6	colsample_bytree	-	0.6, 0.7, 0.8, 1.0
7	subsample	-	0.6, 0.7, 0.8, 0.9

Fig. 3. Hyperparameter Grid for Ensemble Models

D. Cross Validation and Model Evaluation

To ensure reliable evaluation and prevent overfitting [9], we applied TSCV and BTSCV, both essential for preserving the temporal structure of the data in time-series analysis. TSCV incrementally expanded the training data with each fold, ensuring future data was never used to train the model, effectively capturing temporal dependencies and evolving patterns. BTSCV further improved evaluation by introducing a blocking gap between training and validation sets, preventing overlap and mitigating the effects of autocorrelation, making it a more conservative and robust approach for highly correlated time-series data.

For model evaluation, we used several key metrics to assess the performance of our machine learning models [10]: Root Mean Squared Error (RMSE), R-squared (R^2), and Adjusted R-squared (Adj R^2). RMSE was chosen for its ability to quantify the average magnitude of error, giving insight into the typical prediction error of the model in terms of power consumption. R^2 measured the proportion of variance in the target variable that could be explained by the model, providing a straightforward indication of model accuracy.

We also used Adjusted R^2 to account for the number of predictors in the model, offering a more accurate reflection of model performance, especially when dealing with multiple features. This metric is particularly important in models with a large number of variables, as it adjusts for the potential inflation of R^2 when irrelevant features are included.

Cross-validation using TSCV and BTSCV further ensured robust evaluation. By evaluating model performance across multiple folds, we were able to identify how well the models generalize to unseen data and prevent overfitting. The use of cross-validation metrics such as mean RMSE and mean R^2 across folds helped provide a comprehensive understanding of the model's predictive power and consistency. This approach allowed us to evaluate each model's performance under realistic conditions, ensuring the reliability of our predictions.

IV. Results

The models were trained and evaluated according to the outlined methodology, ensuring robust performance metrics across time-series cross-validation. Each model's evaluation metrics represent the average of multiple cross-validation folds: four folds for TSCV and three for BTSCV, chosen to account for temporal dependencies within the dataset.

A. Adjusted R^2 Comparison

Adjusted R^2 was chosen as the primary evaluation metric for model comparison because it accounts for model complexity while measuring how well the model generalizes to unseen data. Unlike standard R^2, adjusted R^2 penalizes models with excessive features, making it particularly valuable in this study, where various parameter combinations are tested across models. This metric ensures that the selected model demonstrates both accuracy and parsimony, reflecting true predictive performance without overfitting.

The heatmap in Figure 4 illustrates the average adjusted R^2 values across cross-validation folds (four for TSCV and three for BTSCV), allowing for a fair comparison of model performances. XGBoost generally exhibits superior adjusted R^2 scores in Scenarios A and C, particularly under TSCV, indicating consistent predictive accuracy across these conditions. Random Forest performance varies significantly, achieving its highest scores in Scenario B under BTSV, but its performance diverges more under other conditions, as seen in Scenario A's BTSV results. Overall, XGBoost with TSCV consistently maintains higher adjusted R^2 scores, suggesting it may be more effective for power prediction tasks.

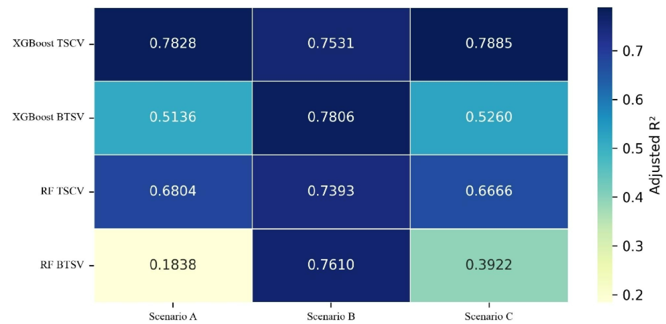

Fig. 4. Heatmap Comparing Average Adjusted R^2 Values Across Models

B. Optimized Model

The results of Scenarios A, B, and C using the XGBoost TSCV model revealed that the adjusted R^2 scores for these scenarios were close to each other. This suggests that including all features in Scenario C may not necessarily yield better predictions, as some features might add noise rather than improve accuracy. In data-intensive models, adding redundant or less relevant features can lead to overfitting, where the model becomes sensitive to fluctuations rather than generalizable patterns. This indicates a need to carefully select features to balance predictive power and model complexity.

To address this, we devised a strategy to determine the optimal number of features [11] that would improve prediction accuracy. Using the top gain scores from the XGBoost TSCV model in Scenario C, we ranked the features by importance. Figure 5 displays the gain-based feature importance plot,

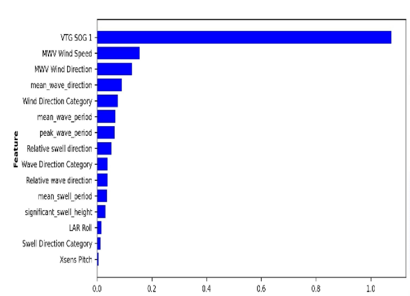

Fig. 5. Top Feature Importance (Gain) for Optimal Model Selection

where speed over ground (SOG), wind speed, and direction emerge as the most impactful predictors, underscoring their significance in the prediction model.

Building on this feature ranking, we experimented with models using varying numbers of features, from the top 5 to the top 15, to find the optimal configuration. For each subset, we performed model tuning and evaluation to identify the number of features that maximized adjusted R^2. Figure 6 illustrates this relationship, highlighting the peak performance achieved with 11 features. Notably, the adjusted R^2 of the 11-feature model surpassed that of Scenario C, confirming that this streamlined selection of top features yielded a more accurate and robust model. Additionally, Scenarios A, B, and C are plotted on the chart as comparison points, showing that the optimized 11-feature model outperforms these configurations.

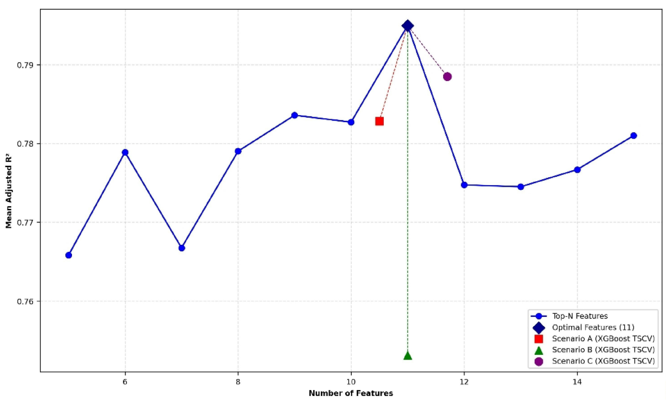

Fig. 6. Adjusted R^2 Comparison for Models with Top-N Features

In conclusion, by selecting only the top 11 features, we created an optimized model that enhances prediction accuracy while minimizing unnecessary complexity. This optimized model not only improves generalizability by reducing noise from irrelevant features but also achieves comparable or better performance than the original, more complex configurations of Scenarios A, B, and C.

C. Model Performance Limitations

The environmental data used in the model, sourced primarily from ERA5 reanalysis, is limited to hourly estimates. While this provides essential context on environmental conditions, the temporal resolution may not capture rapid changes in dynamic open water conditions. Additionally, the absence of ocean currents data limits the accuracy of ship speed through water (STW), relying instead on speed over ground (SOG). Including STW would enable a more precise estimation of vessel resistance, particularly in challenging sea states, affecting power predictions.

Although the dataset was filtered to focus on open water conditions, instances may still exist where similar conditions correspond to different operational profiles. For example, conditions tagged as open water might sometimes include towing operations or other power-intensive activities, resulting in power consumption values that deviate from standard navigation profiles. This operational variability can introduce noise into the model, complicating pattern detection.

D. Future Work

Further enhancement of model performance could be achieved by incorporating higher-resolution ocean currents data, enabling the calculation of speed through water (STW) to improve environmental accuracy. Additionally, a refined classification of operational profiles would allow the model to better distinguish power requirements across various mission types. Advanced machine learning techniques, such as neural networks and hybrid models, may further improve predictive accuracy by capturing non-linear data relationships. These improvements will help the model adapt to complex conditions in open water, supporting efficient power management for icebreaker operations.

V. ACKNOWLEDGMENT

The authors gratefully acknowledge the Canadian Coast Guard for its substantial support in providing critical vessel telemetry data, without which this research would not have been possible.

REFERENCES

[1] IMO, "Marine Environment Protection Committee (MEPC) 77, 22-26 November 2021," 2021.

[2] H. Hersbach, B. Bell, P. Berrisford, et al., "The ERA5 global reanalysis," Q J R Meteorol Soc, vol. 146, pp. 1999–2049, 2020, doi:10.1002/qj.3803.

[3] A. Zambon, L. Moro, J. Brown, A. Kennedy, D. Oldford, "A measurement system to monitor propulsion performance and ice-induced shaftline dynamic response of icebreakers," J. mar. sci. eng., vol. 10, pp. 522, 2022, doi: 10.3390/jmse10040522.

[4] K. Betke, "The NMEA 0183 Protocol," pp. 14, 2001.

[5] F. Pedregosa, et al., "Scikit-learn: Machine learning in Python," JMLR, vol. 12, pp. 2825-2830, 2011.

[6] L. Breiamn, "Random Forests," Machine Learning, vol. 45, pp. 5-32, 2001, doi: 10.1023/A:1010933404324.

[7] C. Tianqi, C. Guestrin, "Xgboost: A scalable tree boosting system," Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 785-794, 2016.

[8] RJ. Hyndman, "Forecasting: principles and practice," OTexts, 2018.

[9] B. Ghojogh, M. Crowley, "The Theory Behind Overfitting, Cross Validation, Regularization, Bagging, and Boosting: Tutorial," arXiv preprint arXiv, 2019, doi: 10.48550/arXiv.1905.12787.

[10] D. Chicco, M.J. Warrens, G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," peerj computer science, vol. 7 , 2021.

[11] I. Guyon, A. Elisseeff, "An introduction to variable and feature selection," JMLR, vol. 3, pp. 1157-1182, 2003.