Large Scale Wave Energy Farms and Predictive Power Outputs

Brody Jones
A&O SCI C111 Machine Learning for Physical Sciences
School of Engineering
University of Los Angeles California
brodyjones@g.ucla.edu

December 6, 2024

Abstract

Wave energy farms, which harness the mechanical motion of ocean waves to produce electricity through arrays of three-tether Wave Energy Converters (WECs), represent a promising avenue in the pursuit of sustainable and diverse clean energy sources. However, predicting the power output of large-scale wave energy installations remains a challenging regression problem due to the complex, nonlinear hydrodynamic interactions between buoys and the ocean environment. Traditional hydrodynamic simulations are computationally costly, making machine learning (ML) approaches appealing for rapid and efficient power prediction.

In this study, we evaluate multiple ML models, including Decision Trees, Random Forests, Gradient Boosting regressors, Support Vector Regression (SVR), and K-Nearest Neighbors (KNN), to predict both total and individual buoy power outputs based on the spatial configuration of WECs. We demonstrate that ensemble-based methods, especially when fine-tuned via systematic hyperparameter optimization, outperform simpler models and significantly reduce prediction errors. This improvement in predictive accuracy has critical implications: accurate and fast predictions of WEC farm output can guide optimal buoy placements, inform maintenance schedules, and support real-time operational decisions, ultimately increasing the economic viability and reliability of wave energy farms.

Introduction

Wave energy represents a burgeoning frontier in renewable energy, with the potential to provide a steady, predictable source of clean electricity. Unlike wind or solar, wave energy sources are often more consistent and can produce power day and night. These farms have shown indications of achieving high energy density per unit area and a high level of predictability. However, large-scale wave energy farms, composed of numerous Wave Energy Converters (WECs), face complexities that challenge straightforward predictive modeling. Predicting the power output of these farms involves understanding intricate hydrodynamic interactions, which traditionally require computationally expensive simulations. By applying machine learning (ML) techniques to forecast power outputs—both at the total farm level and for individual buoys—it is possible to reduce this computational burden and streamline decision-making regarding buoy placement, maintenance schedules, and long-term operational strategies.

Background and Motivation

Wave energy farms rely on arrays of WECs, each converting the mechanical motion of waves into electrical energy. The relative positioning of buoys, the intensity and directionality of incoming waves, and the complex fluid-structure interactions all are being actively researched to understand their influence on the overall power output. Compared to well-established renewables like solar or wind, wave energy has received less focus and standardization. Hydrodynamic calculations and finite-element modeling, while accurate, are slow and resource-intensive. This uncovers the motivation to explore data-driven predictive models: faster predictions can facilitate optimization, help stakeholders identify cost-effective configurations, and support adaptive operation plans that respond dynamically to changing ocean conditions. Ultimately, building accurate predictive models advances the reliability and commercial viability of wave energy farms.

Methodology

1. Dataset

The dataset used in this research captures the positions of 49 WECs from a large-scale farm in Perth, Australia. It also recorded each buoy's corresponding power output, and an aggregated total power output for the entire farm configuration. The data also includes environmental and positional features: two-dimensional coordinates for each WEC, derived inter-buoy distances, and measured power outputs. The dataset included 149 features over 30,000 instances, and thus was sufficiently dense for large scale optimization, as it was used by its original research team. Unequipped with the computing power for such a dense problem, we opted to trim a 1000 instance train-test split for the more lightweight regression problem. Our research aimed to understand the effectiveness of smaller, faster ML models on this complex problem. Features were preprocessed and scaled when necessary. Some experiments involved engineering new features, such as inter-buoy distances, to help the models capture spatial relationships between WECs.

2. **Preliminary ML Models**

We used five familiar machine learning models to try to predict both total-farm power output and individual buoy power output:

- Decision Tree Regression:
    - A hierarchical tree based decision model that splits on optimized features
    - Nonparametric, supervised learning model that is easy to interpret
    - Prone to overfitting, but offers a good starting point
- K-Nearest Neighbors Regression:
    - Classifies new data points by referencing closest existing data points
    - Nonparametric, supervised learning model based in geometry
    - Easy to implement due to “lazy learning” but relatively simple in structure
- Support Vector Regression:
    - Kernel based approach that attempts to replicate non-linear relationships
    - Good at predicting continuous result values using the most impactful data points (support vectors)
    - Transforms into a higher dimensional space, but may be computationally expensive and less accurate when applied in poor contexts
- Random Forest Regression:
    - Utilizes an ensemble of decision trees to reduce variance and often outperforms single decision trees
    - Random feature selection and ensemble learning allow for overfitting prevention and more robust versatility
    - Can handle large datasets efficiently with parallel computation
- Gradient Boosting Models:
    - Iterative ensemble method that builds new trees to correct the errors of previous ones
    - Weak learning, loss function, and an additive model approach
    - Gradient optimization and residual learning helps to focus the direction of model training
    - Hyperparameter tuning allows for state-of-the-art performance

3. **Model Training and Hyperparameter Training**

All models were initially trained and tested with as many default parameters as possible. We used pre-built models from the SciKit Learn Library, and tested predictive capability for both individual buoy and total power output. The performance of each model was evaluated using mean squared error. Notably, the ensemble models were able to capture the complex interactions of our wave energy farm. This probed further fine-tuning on the parameter based models, Random Forest and Gradient Boosting. We employed Randomized Search CV in order to find the ideal hyperparameters for each of these models. We tested ranges for max_depth, learning_rate, n_estimators for both models, as well as regularization terms for XGBoost (reg_alpha, reg_lambda, gamma, etc). The goal is to find parameters that minimize our mean squared error loss function without overfitting.

4. **Computational Platform**

All modeling and tuning were conducted on Google Colab, a cloud-based environment offering GPUs and TPUs. While this platform provided accessibility and convenience, memory and runtime limitations constrained the size of the hyperparameter search spaces and the complexity of models that could be trained simultaneously. This required sound choice of parameter ranges and sample sizes, balancing computational feasibility with the need for experimentation.

Results and Discussion

1. **Performance of Base vs. Ensemble Models**

Initial runs revealed that simple models like Decision Trees and KNN produced relatively high MSE values for total farm output prediction (on the order of billions). SVR performed even worse, likely due to inadequate kernel parameterization and large data size. Ensemble methods—Random Forest and especially Gradient Boosting—delivered substantially lower error rates. For total power output, baseline XGBoost reduced the MSE from over 2.44e9 (Random Forest baseline) to about 2.36e9. Similarly, for individual buoy predictions, boosting methods achieved an MSE of about 2.40e7, far outperforming simpler methods (Random Forest — 4.10e7, KNN — 4.71e7, Decision Tree — 6.43e7).

2. **Impact of Hyperparameter Fine-Tuning**

Fine-tuning substantially improved results. By adjusting the depth of trees, learning rates, and regularization terms, the XGBoost model's total output prediction error dropped to approximately 1.59e9. Similarly, the Random Forest model saw improvements to about 2.03e9 after tuning. For individual buoy output, the best-tuned boosting model reached an MSE as low as about 2.25e7. These improvements illustrate the power of refining the parameter space to uncover model configurations that better capture the complex relationships driving power production in a wave energy farm.

3. **Model Advantages and Tuning Analysis**

Models like Random Forests and XGBoost excelled because they combine multiple weak learners, control model complexity through randomization and regularization, and adapt their structure based on residual errors (in the case of XGBoost). The final hyperparameters reflect the model's attempt to balance complexity and generalization. For XGBoost, strong regularization and shallow trees indicated the data's underlying patterns were not excessively complex, while moderate learning rates and partial subsampling further improved stability. For Random Forest, a large ensemble with controlled feature sampling worked best, suggesting that robust averaging and some randomness are key to handling intricate wave-energy relationships.

4. **Practical Implications**

Accurate predictive models can shape strategic decision-making in wave energy farms. Operators might rely on ML-generated forecasts to adjust buoy placements, altering spacing or orientation to improve power capture efficiency. Likewise, maintenance and operational interventions can be scheduled based on predicted performance dips, ensuring continuous and optimized energy output. In the long run, reliable predictive modeling supports a move toward automated, self-optimizing farms that react dynamically to changing wave climates, reducing the reliance on costly full-scale simulations.

Limitations

Alongside the mean squared error, consider incorporating alternative evaluation metrics that provide additional perspectives. We also would want more research quantifying uncertainty in predictions.

Techniques like prediction intervals or probabilistic models can highlight confidence levels in forecasted power outputs, which can be valuable when making high-stakes operational decisions.
If we had access to temporal and environmental data we could also better understand the role that temperature, wave frequency, tidal ranges, and seasonal variability play on power output. Explaining these may help contextualize the complexity of the prediction task and strengthen the rationale for using ML methods.
While Google Colab provided a practical environment for this research, resource constraints limited the complexity of experiments. Memory restrictions forced dataset sampling, potentially excluding useful variability. Runtime limits capped the depth of hyperparameter searches and precluded extensive ensembling strategies. Additionally, while ML models can approximate the underlying physics, they may lack interpretability and fail to capture all complexities without sufficient training data. Further domain-specific feature engineering and collaboration with hydrodynamic experts could yield improved results.

Future Research Directions

Several promising avenues remain unexplored. Integrating more advanced ML architectures like deep learning models (e.g., convolutional or transformer-based networks) may further improve accuracy, especially if trained on larger datasets enriched with time-series wave conditions. Another critical next step is shifting from mere prediction toward optimization frameworks. Rather than using ML solely to forecast power output, future models could employ reinforcement learning or genetic algorithms to identify configurations that maximize total power generation. Incorporating real-time wave forecasts and higher-fidelity spatial data could also help adapt buoy placements on the fly, increasing total farm output and operational resilience.

Conclusion

In this study, we have shown that machine learning, particularly fine-tuned ensemble methods like XGBoost, can predict wave energy farm outputs with increased accuracy and reduced computational overhead compared to traditional modeling approaches. By bridging the gap between complex hydrodynamic modeling and practical decision-making tools, ML models stand to accelerate the adoption and economic viability of wave energy. As the field matures and data availability grows, these data-driven approaches will become integral to designing, optimizing, and managing large-scale wave energy farms—ultimately contributing to a more diverse and stable renewable energy portfolio.

Acknowledgements

We would like to acknowledge the contributions made by Dr. Lozinski and the A&O SCI C111 instruction team for providing research direction and a web publishing template. I would also like to acknowledge the team of computer scientists at the University of Adelaide who initially gathered and posted this dataset.