# A Machine Learning Approach for Model Selection in Underwater Acoustic Propagation

Finn P. Henman[1], Joshua J. Wakefield[1], Christopher Ashworth[2], Stewart G. Haslinger[2], Jason Ralph[1], and Duncan P. Williams[3]

[1]Department of Electrical Engineering and Electronics, University of Liverpool
[2]Department of Mathematical Sciences, University of Liverpool
[3]Dstl

Finn Henman, University of Liverpool, Liverpool, L69 7ZF, fhenman1@liverpool.ac.uk

***Abstract:*** Underwater acoustic propagation models use different assumptions and approximations to simplify the wave equation. These assumptions make different models more or less suitable for modelling propagation in different ocean environments, such as a ray tracing, parabolic equation, or normal mode model. Ensuring that one's choice of underwater acoustic propagation model is tailored to a given scenario is important to ensure propagation is modelled as accurately as possible.

Current methods for deciding which model to use are often based on simple thresholds for high or low frequency [1]. For example, one may choose to use a ray tracing model for acoustic frequencies above 1000 Hz, and a parabolic equation model below 1000 Hz. A tool that selects the most suitable model based on multiple environmental parameters would be immensely useful for a variety of underwater acoustics applications. This paper presents a novel method to select the most appropriate model for a given scenario by using a random forest classifier to predict the best model based on the input parameters.

A training set of ocean environments was generated using a variety of values for depth, range, bathymetry, and sound speed profiles. These environments were passed through the RAMSurf (parabolic equation) and BellhopCXX (ray tracing) propagation models and results were compared with Kraken (normal mode), which acted as an arbiter. This provides a basis for model selection supported by empirical data, which takes into account a wider selection of environmental parameters, whilst also allowing for edge cases to be modelled more accurately. The results show a high level of accuracy, with the potential to be further expanded to include a wider range of propagation models and alternative selection metrics.

***Keywords:*** *propagation modelling, machine learning, simulation*

## 1. INTRODUCTION

In ocean acoustics, propagation modelling is useful for a wide variety of applications, including the study of marine wildlife, underwater navigation, and seabed surveying [2]. For the selection of an underwater acoustic propagation model, it is important to tailor the chosen model to the environment being modelled by taking into account the underlying assumptions and their strengths and weaknesses.

The ray tracing and beam tracing methods are derived from the Helmholtz equation through high-frequency approximations, where the wavelength of the acoustic wave is much smaller than the dimensions of the environment being modelled [3]. These methods simplify complex wave phenomena into more manageable geometrical constructs. This makes ray tracing most suitable for modelling the propagation of high-frequency sound waves in deep water [4].

The parabolic equation (PE) solution of the Helmholtz equation works under the assumption that the wave only propagates in one direction. The Helmholtz equation is factorized into forward and backward propagating components; the backward components are discarded [5]. These models are well suited for modelling the propagation of low-frequency waves in shallow environments [6].

The aim of this research is to provide a framework for the selection of the most suitable candidate from a choice of PE and ray tracing models for specific environmental and frequency parameter spaces. Specific questions are whether the PE and ray tracing models are capable of good performance beyond the standard parameter spaces for which they are usually implemented, and what does the parameter space in which they both show good performance look like. The propagation models used here for this selection competition have been primarily chosen due to their ubiquity in the propagation modelling literature (see for example [4, 6, 7]):

- BellhopCXX - a C implementation [8] of the ray tracing model Bellhop developed by Porter [4].
- RamSurf - a C implementation [9] of the parabolic equation model RAM developed by Collins [5].

In a previous study considering computational cost and convergence criteria [10], RAM and Bellhop were compared. Here, the focus is on selection between the two models for a variety of environmental parameters, rather than considering only computational efficiency. To facilitate these comparisons, a machine learning (ML) approach was implemented, for which an arbiter was required. In practice, this arbiter can be selected for specific environments under consideration, but for this initial demonstration of the approach, another commonly used model was selected, namely Kraken, a normal modes model developed by Porter [11]. The arbiter model is used to determine the “ground truth” against which the two competing models are compared. Bellhop and RAM were chosen as the initial models of interest because of their common use and the fact that they are generally considered to best suited to contrasting conditions, making the transition region between the suitability of the two of particular interest.

The adoption of an ML approach is ideal considering the number of parameters affecting performance; the alternative of using a simple frequency threshold to decide between two models is straightforward, but this approach will fail to find the most suitable model for edge cases, such as low-frequency for a deep environment. ML is well suited to defining the decision boundary in a multi-dimensional parameter space.

## 2. METHODOLOGY

The training data were generated for a limited selection of generic environments, designed to balance variation and simplification to retain realism whilst avoiding excessive time and computational resources. Every realisation had a flat surface boundary at the air-ocean interface, and a source depth of 25 metres. The environment types were each defined from maximum and minimum values of depth, frequency and surface sound speed, from which each realisation would have a random value chosen from a uniform distribution. Additionally, a random sound speed profile type was chosen. The types of environments are shown in Table 1 along with the range of values for their parameters. This iteration of the modelling framework has been developed to facilitate relatively straightforward expansion, both for the number of environments considered, and their complexity (i.e. defined by the input parameters/labels).

The maximum range of an individual realisation was calculated as a function of frequency, in order to reduce the time spent computing; high-frequency, long-range problems would take a significant amount of time to compute and most of the returned transmission loss plot would be of little use due to signal attenuation. Generally speaking, high-frequency cases are associated with short ranges and low-frequency realisations, with long ranges. This relationship was defined by a linear interpolation of values provided in Table 1 of [10].

| **Environment** | **Max Depth (m)** | **Frequency (kHz)** | **Surface Sound Speed (ms$^{-1}$)** | **Sound Speed Profile types** |
|---|---|---|---|---|
| Shallow | 60 - 200 | 0.2 - 5 | 1480 - 1550 | Constant, Upward Refracting |
| Deep | 1500 - 5500 | 0.05 - 1.5 | 1480 - 1550 | Constant, Munk |
| Coastal | 50 - 60 | 0.5 - 10 | 1480 - 1550 | Constant, Downward Refracting, Upward Refracting |
| Sloped | 252 - 462 | 0.4 - 1.4 | 1480 - 1550 | Constant |

*Table 1: Summary of the different environment archetypes used and their properties, with ranges provided for depth, frequency, surface sound speed and SSP type.*

Four different types of sound speed profiles (SSP) were also defined, and these were prescribed as possible options for each environment type appropriately, e.g. in the Shallow environment, the SSP is typically modelled using an upwards refracting or constant profile. The Sloped environment is characterised by an initial segment of flat seabed at depth 200 metres, followed by a shallow downward incline (fixed to 3°) to a randomly selected depth, followed by another flat stretch out to the maximum range.

To compare the results of the models of interest with the arbiter, a metric was defined. In this case, the modelling output was converted to a single numerical value (mean absolute error), representing the degree to which the results differed from each other. Each propagation model was used to generate a one-dimensional (1-d) array of transmission loss (TL) measurements along the full range of the environment at a fixed receiver depth of 25 metres, with range discretization of one metre. Figure 1 shows an illustrative plot with the three 1-d arrays generated for the same realisation (labelled using values randomly generated from the ranges given in Table 1).

The arrays were pre-processed to remove any outlier values, such as infinities or NaN values which would adversely affect any statistical measure such as the mean. The corresponding points had to also be removed from the arrays generated from the other models to ensure that the same length of range was compared universally. A smoothing filter was then applied to the arrays, again to remove any large outlying values or “jittering” sections. These equalization and smoothing steps were necessary to ensure that the mean absolute error was a meaningful metric for comparing the models. A comparison of a TL array before and after the smoothing filter is shown in Figure 2.

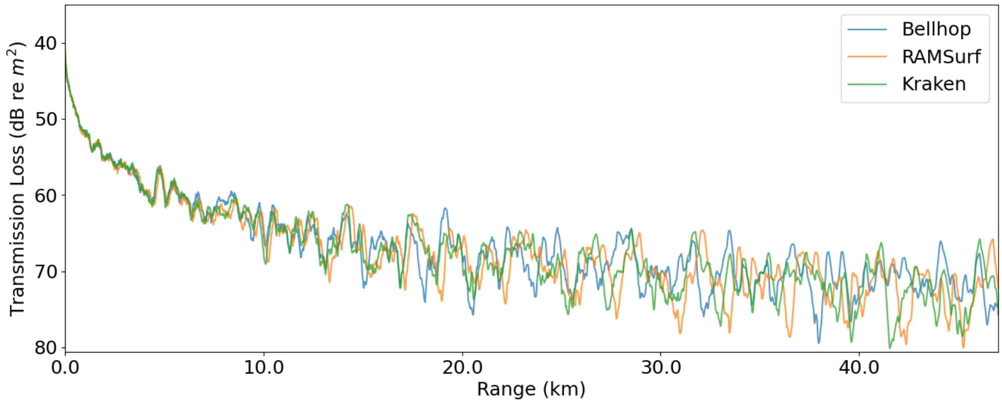


*Figure 1: A comparison of the transmission loss at a fixed depth of 25m between all three propagation models. The Shallow environment parameters used were a maximum depth of 72 m, a frequency of 1160 Hz, a maximum range of 48.5 km, and an upwards refracting sound speed profile with a surface speed of 1508 ms$^{-1}$.*

Following this, the mean absolute error (MAE) is calculated using the equation

$$\frac{1}{N}\sum_{i=1}^{N}\left|X_i - X_i^{\mathrm{Kr}}\right|, \tag{1}$$

where $X_i$ is a TL value for either BellhopCXX or RamSurf at range value $r_i$ for maximum range $N$ metres, and $X_i^{\mathrm{Kr}}$ is the corresponding TL value for Kraken. This MAE metric is used as the similarity score. The similarity scores are compared and the model with the lowest score (closer to the arbiter) is chosen to be the most appropriate selection for the environmental parameters, and the data $y_i$ are labelled as either BellhopCXX or RamSurf, as indicated in Figure 3.

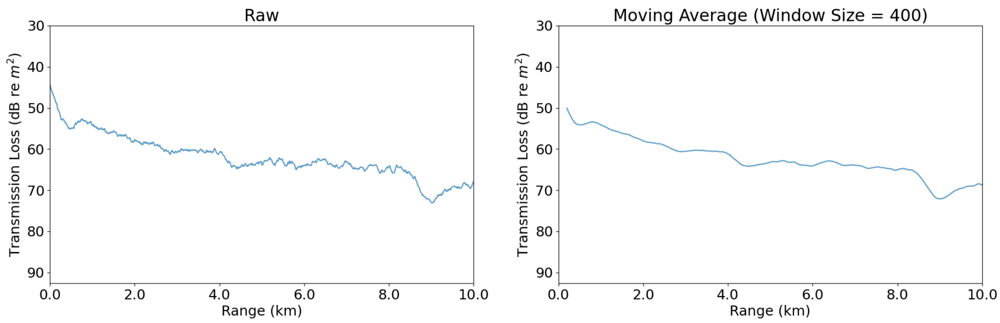


*Figure 2: A comparison of a 1D transmission loss array from BellhopCXX before and after applying the smoothing function. The environment used was Shallow, with a maximum depth of 382 m, a frequency of 722 Hz, a maximum range of 10 km, and a constant sound speed of 1511 ms$^{-1}$.*

Four distinct classification models were employed, with each model trained on 70% of the data and evaluated on the remaining 30%, using a random train-test split. In the results included

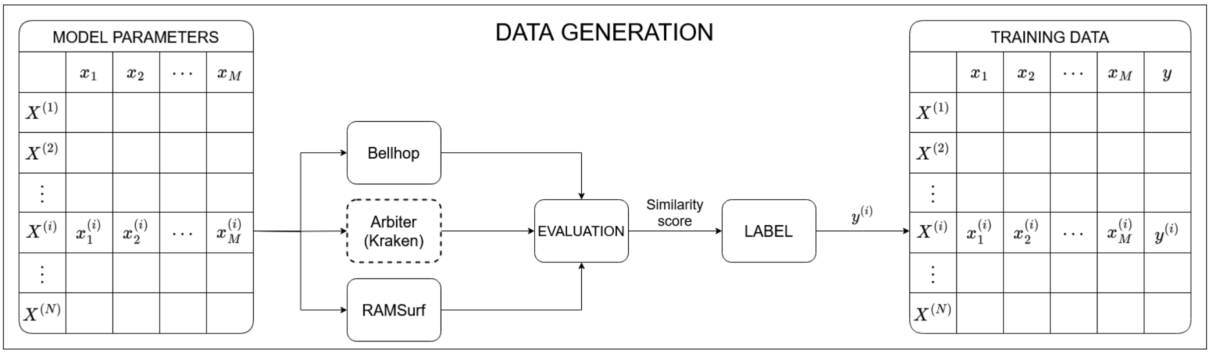


*Figure 3: A flowchart showing how environmental parameters are passed through each model to generate training data.*

in this paper, 2000 individual realisations, across the four environment types, were used to populate the dataset, with a ratio of 4:7:4:5 for the Shallow, Deep, Coastal and Sloped environment types, respectively. During the data generation and testing process, several different ratio combinations were explored, but this ratio proved to be the most balanced and robust. In future investigations, as additional environment types are included, and more complexity is added to the input parameters, it will be interesting to investigate how this ratio changes. It is important to note that the modelling framework is constructed to support changes to the ratio in a simple manner, through an individual input parameter script.

The ML models were implemented using the scikit learn python library. The same training set was given to each model and they were compared against one another to find which one was the most accurate. A neural network was used initially, however it was deemed unsuitable due to the size of the dataset being too small, but this is something that can be explored in future refinements with application to much larger data generation/collection efforts.

The classifier models evaluated are as follows:

- Random Forest Classifier - An ensemble learning method which builds multiple decision trees, the outputs of which are combined using majority voting to improve accuracy and to control overfitting. It works well with both classification and regression tasks and handles large datasets with higher dimensionality effectively.

- Support Vector Machine - A supervised learning algorithm that finds the optimal hyperplane to separate classes in the feature space. It works well for both linear and non-linear classification, and is particularly effective for high-dimensional spaces.

- Logistic Regression - A linear model for binary classification that estimates the probability of a class using the logistic function. Despite the name, it is a classification algorithm and is widely used due to both its simplicity and efficiency.

- K-Nearest Neighbours - A non-parametric, instance-based algorithm that classifies new data points based on the majority class among its $k$ nearest neighbours in the training data. It is simple and effective, although performance can decrease with high-dimensional data.

## 3. RESULTS

Table 2 shows the accuracy of the different classifier algorithms in predicting the most accurate model to use, for both a mixed set of environments and per environment. In general, the random forest classifier was the most effective in predicting the correct model, with an accuracy rating of 84.8%, around 4% better than the other three models investigated. Furthermore, the results show the effect that removing the Sloped environment from the dataset has on the overall accuracy of the results. The Sloped environment proved challenging, with the parameter ranges for depth and frequency being restricted due to convergence issues in the solutions for both Kraken and RamSurf. Initially, it was intended to vary the slope angle up to 10°, but numerical instability in the data generation for both RamSurf and the arbiter prevented this. Instead, a parameter study was conducted to amend the ranges of values of depth and frequency for which the Sloped environment produced reasonable results.

The prediction results for this restricted regime appeared to show some bias towards RamSurf, with the possibility that inaccuracies were arising in both models and so RamSurf was often closer to the "truth" than BellhopCXX. The results in Table 2 also indicate that the uncertainty was much higher for this environment than for the other three types. It is important to explore this environment in more detail in future work to understand how the robustness can be improved to match the results for the Shallow, Deep and Coastal environments.

| Classifier | Shallow | Deep | Coastal | Sloped | All |
|---|---|---|---|---|---|
| Random Forest | 89.1% | 90.6% | 89.0% | 67.9% | 84.8% |
| Support Vector Machine | 89.6% | 91.9% | 76.1% | 68.7% | 80.6% |
| Logistic Regression | 89.5% | 92.3% | 74.5% | 66.6% | 80.6% |
| K-Nearest Neighbours | 89.3% | 90.9% | 76.8% | 65.9% | 80.2% |

*Table 2: The table showing the test accuracy of four different classifiers when trained on datasets made of different environment types*

Figure 4 shows the trends in selection for those realisations labelled as more accurate, using BellhopCXX or RamSurf, versus frequency in Hz. In this dataset, the point at which BellhopCXX generally becomes more accurate than RamSurf appears to be around 1.8 kHz, with BellhopCXX being more accurate above that threshold and less accurate below it. This is not unexpected since ray tracing is essentially a high-frequency approximation. This threshold is considerably higher than the 1kHz mentioned in the abstract, as well as the 500Hz threshold defined for high and low frequency in [1]. However, we can see that in many cases, RamSurf obtains a similarly accurate result at frequencies beyond this threshold, and BellhopCXX also performs relatively well at lower frequencies. It should be noted that there are a higher number of low-frequency environments due to the balancing ratio of environment types of 4:7:4:5, and the frequency limits shown in Table 1.

It is also important to note that the selection metric is entirely based on the accuracy of the results compared with the results of the arbiter model, and does not factor in computation time (although these are recorded by the model). Whilst selecting a model based on accuracy is valid and may often be desirable, this is not always the case. A parabolic equation model may provide more accurate results than a ray tracing model in some cases, but the computation time is often significantly longer (particularly for the Deep environment type considered here). Including computation time relative to the other model(s) as a secondary output and using a weighted combination of the two as the selection criteria would be a potentially useful addition to the

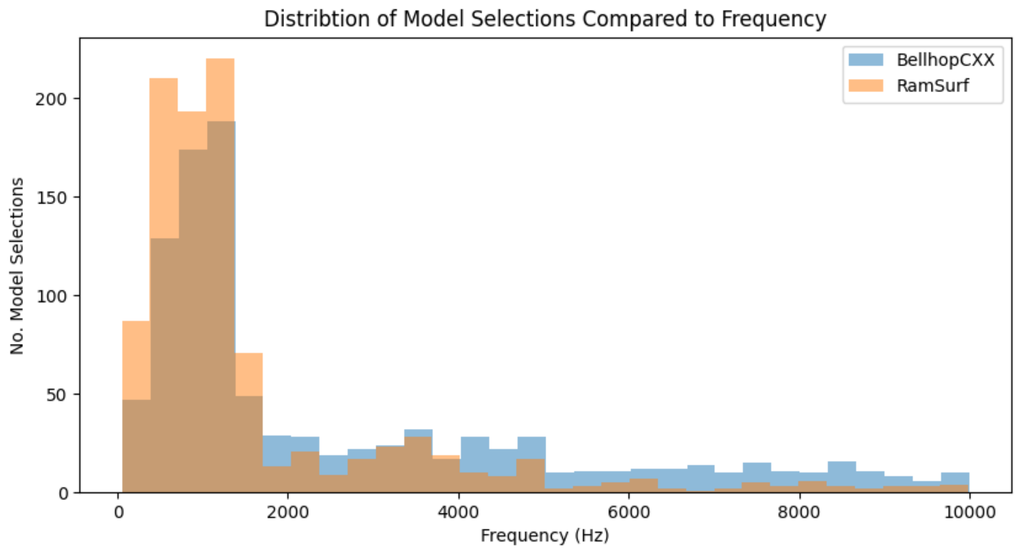


*Figure 4: A histogram showing the distribution of model selections against frequency from a dataset of 2000 environments*

selection algorithm in future refinements and extensions.

## 4. CONCLUSION

The model selection algorithm developed here shows very promising results, generating a selection threshold between ray tracing and parabolic equation propagation models at a frequency threshold close to 1.8 kHz for the four generic environment types implemented. This value was obtained for a dataset consisting of 2000 realisations (with 400 for Shallow, 700 for Deep, 400 for Coastal and 500 for Sloped). Changing the size of the dataset, and varying the ratio of the environment types, will lead to different threshold values so it will be important to design ensemble model efforts for specific implementations to obtain statistically more stable results, with reduced uncertainty.

Considering the environment types separately will also produce different threshold values, emphasizing the value of using this ML-based approach, which can be tailored to specific applications. The use of these algorithms provides data-based guidance on multi-parameter-based criteria which may otherwise be too complex to determine manually or with simpler selection processes. The architecture written provides a basis for potential expansion, such as comparing additional models, defining more complex environments and further input parameters, replacing the arbiter for specific applications (for example, Kraken is a good choice as arbiter for environments which can be modelled as waveguides) and including further parameters and metrics to evaluate model performance.

## 5. ACKNOWLEDGEMENTS

Part of this work was commissioned by Dstl on behalf of the Submarine Delivery Agency as part of the FASTER programme. The authors would like to thank the FASTER team for their support and guidance. The contents include material subject to © Crown copyright (2024), Dstl. This material is licensed under the terms of the Open Government Licence except where otherwise stated. To view this licence, visit `http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3` or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email:
psi@nationalarchives.gov.uk.

## REFERENCES

[1] Jensen, Finn B. : "Numerical Models of Sound Propagation in Real Oceans" in *OCEANS 82*, (Washington, D.C, 1982).

[2] Ainslie, Michael, A. : *Principle of Sonar Performance Modelling*, (Springer-Verlag, Berlin, 2010).

[3] Jensen, Finn B., Kuperman, William A., Porter, Michael B., Schmidt, Henrik : *Computational Ocean Acoustics*, ( Springer, 2011).

[4] Porter, Michael B. : "Beam tracing for two- and three-dimensional problems in ocean acoustics", *The Journal of the Acoustical Society of America* **146**, 2016-2029 **(2019)**.

[5] Collins, Michael D. : "A split-step Padé solution for the parabolic equation method", *The Journal of the Acoustical Society of America* **93**, 1736-1742 **(1993)**.

[6] Kusel, Elizabeth T., Siderius, Martin : "Comparison of Propagation Models for the Characterization of Sound Pressure Fields", *IEEE Journal of Oceanic Engineering* **44**, 598-610 **(2019)**.

[7] Oliveira, Tiago C. A., Lin, Ying-Tsong, Porter, Michael B. : "Underwater Sound Propagation Modeling in a Complex Shallow Water Environment", *Frontiers in Marine Science* **8 (2021)**.

[8] Pisha, Louis A. et al : "Introducing bellhopcxx/bellhopcuda", *Acoustical Society of America Journal* **153**, A218-A218 **(2023)**.

[9] Paille, Serge S., Levaufre, Arnaud : *RAM PARABOLIC EQUATION*, (`https://github.com/quiet-oceans/ramsurf`).

[10] Hartstra, Iris E., Salomons, Erik, Colin, Mathieu E.G.D., Prior, Mark K. : "ENVIRONMENT DEPENDENT TRANSITION FREQUENCIES FOR RAM AND BELLHOP" in *UACE2019 Conference Proceedings*, edited by John S. Papadakis (Hersonisso, 2019).

[11] Porter, Michael B. : *The KRAKEN normal mode program, NRL/MR/5120-92-6920*, (Naval Research Lab., Washington, DC., 1992).