Vessel Noise Correlation Study – Phase 3

ECHO Program study summary

This study was undertaken for the Vancouver Fraser Port Authority-led Enhancing Cetacean Habitat and Observation (ECHO) Program as the third phase of the vessel noise correlations study. The goal of the study was to identify statistical correlations between vessel design and operational characteristics and underwater noise. This third phase aimed to test and evaluate whether the statistical model developed in Phase 1 and 2 could accurately predict vessel underwater noise levels using an independent dataset.

What questions was the study trying to answer?

This third phase of the vessel noise correlations study investigated the following main questions:

- Would the statistical model developed using the ECHO database show an equivalent predictive power on an independent dataset?
- Do predicted Radiated Noise Levels (RNL) or Monopole Source Levels (MSL) show a significant difference or systemic error (distribution of residuals) when compared to the measured RNL and MSL in an independent dataset?

Who conducted the project?

To address these research questions, the Vancouver Fraser Port Authority retained a team led by the Scripps Institution of Oceanography (Scripps), with support provided by both JASCO Applied Sciences (Canada) Ltd. (JASCO) and ERM: Environmental Resources Management (ERM), who developed the original statistical model for the ECHO Program in earlier phases of this project.

What methods were used?

This project used the statistical model developed from the ECHO dataset in earlier phases of the project and the vessel underwater noise dataset collected by Scripps in the Santa Barbara Channel (SBC). The statistical model was developed using the ECHO dataset for both monopole source level (MSL) and radiated noise level (RNL) for six major commercial vessel categories: bulker/general cargo carriers; container ships; large passenger/cruise ships; tankers; tugs; and vehicle carriers.

The Santa Barbara Channel dataset included acoustic data from two sites. The first site, B1, was located approximately 3 km north of the northbound shipping lane at a water depth of 580 m. The second, B2, was located directly under the northbound shipping lane, also in approximately 580 m of water.

- Site B1 data was collected between 2007 and 2018 and included 1654 accepted measurements
- Site B2 data was collected between 2018 and 2021 and included 2468 accepted measurements
- Over both sites, 1242 unique vessels were recorded

The vessel source level measurements from the SBC dataset were matched to the vessel characteristics of each vessel and specific transit. The model inputs include operational parameters: speed through water, actual draft, wind resistance and surface angle of measurement. In addition, the model used vessel design parameters such as: vessel length, vessel design RPM, vessel design main engine power, vessel design speed, and vessel age as provided by Lloyd's List Intelligence.

The range of each of these characteristics varied between the ECHO dataset and the SBC dataset. The SBC dataset's measurements for surface angle, wind resistance and speed through water were outside the range found in the ECHO dataset. Specifically, both SBC sites (B1 and B2) had surface angles well outside of the range found in the dataset used to develop the statistical model. As a result, a fixed surface angle of 30 degrees, equal

to the median value in the ECHO data, was used in the predictive model for the SBC data and as such, only 8 parameters were used to predict MSL and RNL.

To test the model's predictive accuracy, the predicted RNL and MSL values were compared to the actual measured values collected in the Santa Barbara Channel. The residuals, or differences, between predicted and measured RNL and MSL were then compared to see if the model demonstrated similar predictive ability for both the SBC and ECHO datasets. The residuals were also reviewed to look for any obvious trends or systemic differences unique to the SBC dataset not predicted by the model.

What were the key findings?

The main findings of the vessel noise correlation analysis are summarized as follows:

- The statistical model predictions for MSL showed better accuracy than RNL predictions of noise for the SBC dataset. Both RNL and MSL models, however, showed an equivalent ability to predict the general shape of the acoustic measurements as a function of frequency.
- The analysis shows the limitation of RNL as a metric when comparing between sites, as RNL does not account for site-specific sound propagation effects. MSL appears to be a more robust analysis method to account for propagation effects and as a result appears to be a more predictable metric using the statistical model when evaluating independent datasets.
- The statistical model over-estimated noise levels below 100 Hz for most vessel types in the SBC dataset, with the notable exception of container ship MSLs; the largest class. The reason for the over-estimation was unknown but may be related to differences between the ECHO dataset locations and the SBC locations, such as bathymetry, bottom composition, and wind resistance.
- A systemic difference was noted in measured MSL and RNL readings in the SBC dataset that corresponded to site locations (B1 and B2). Calculated source levels at the vessel, as measured at site B1 (farther from the shipping lane) were louder than those at site B2 (beneath the shipping lane). This may be due to several factors including:
  - possible exclusion of quieter or slower vessels from the B1 location if the vessel noise was not sufficiently higher than ambient/background noise over the 3 km distance from the shipping lane to the hydrophone.
  - possible systemic differences between north and southbound vessels, as site B1 tended to only capture northbound traffic whereas site B2 captured both southbound and northbound traffic.
  - potential error introduced by propagation modelling between the site close to the vessel noise source (B2) and that situated farther from the vessel (B1)

Discussion and next steps

Consistent with the findings of the first two phases of the noise correlation study, the statistical model showed the ability to accurately predict overall means and trends in underwater noise emissions, but lacks the detailed information required to precisely predict a particular vessel on transit-by-transit basis.

Use of the statistical model on the SBC dataset also showed the limitations of using a predictive model on a dataset where the parameters exceed the limits of the dataset used to create the model. For the SBC dataset this included surface angle, wind resistance and speed through water. The slower speed through water seen in the SBC dataset may also be partially responsible for over-estimation of noise levels below 100 Hz by the statistical model when compared to measured noise levels.

To better understand the variations in the model between predicted versus measured vessel noise levels, additional analysis would be required. Some potential areas for further study include an analysis to determine the sensitivity of calculated underwater noise to source depth, surface angle and closest point of approach.

Evaluation of ECHO Vessel Noise Correlation Models with a Novel Dataset Collected in the Santa Barbara Channel

Submitted to:
Derek White
Vancouver Fraser Port Authority ECHO Program
Contract: 21-0305

MPL Technical Memorandum 658
January 31, 2022

Kaitlin E. Frasier, Vanessa M. ZoBell
Scripps Machine Listening Lab
Marine Physical Laboratory
Scripps Institution of Oceanography
University of California San Diego
La Jolla, CA 92037

Alexander O. MacGillivray, Joshua N. Dolman
JASCO Applied Sciences (Canada) Ltd
Suite 2305, 4464 Markham St.
Victoria, BC V8Z 7X8

Laurie Ainsworth, Joanna Zhao
ERM Consultants Canada Ltd.
1111 West Hastings St.
Vancouver, BC V6E 2J3 Canada

Executive Summary

Scripps Institution of Oceanography (SIO) assessed the predictive fit of a vessel noise functional regression model developed by JASCO and ERM on an independent database of vessel noise levels measured at recording sites in the Santa Barbara Channel (SBC) in cooperation with the Vancouver Fraser Port Authority led ECHO Program. The functional regression model produced by JASCO and ERM was developed using the ECHO source level database, which comprises vessel transits recorded in Haro Strait, Strait of Georgia, and Boundary Pass at depths of approximately 170 - 250 m. The SIO dataset of vessel transits were recorded at two locations in the SBC, the first of which (site B1) was located approximately 3 km north of the northbound shipping lane, and the more recent site (site B2) was directly in the northbound shipping lane. The differences between the two SBC sites created a bi-modal distribution of surface angles in the dataset. Both of the SBC sites were located at approximately 580 m depth. In addition to the recording depth difference between the databases, the speeds of the vessels were higher in the ECHO database than in the SBC database. On average, the models predicted the monopole source level (MSL) well. Measured MSLs were higher for site B1 transits than site B2 transits, potentially due to data conditioning steps which removed low amplitude site B1 measurements. Other contributing factors may include systematic differences between north and southbound vessels, and incomplete accounting for propagation loss. Model predictions were most comparable to site B1 transits on average. Models over-predicted MSL below 100 Hz for site B2 transits for most of the categories, excluding the container ship category which had the largest sample size. For all categories and site locations, the models predicted MSL well for frequencies above 100 Hz. On a per transit basis, the mean absolute prediction error for MSL was approximately 5 dB. The radiated noise level (RNL) was over-predicted for all ship categories below 100 Hz by approximately 9 dB and above 100 Hz by approximately 4 dB on average, with highest accuracy above 100 Hz for site B1 transits. Overall, the models were able to predict MSL well on the vessel category level, however, they should not be considered to be precise on a transit by transit basis for the SBC dataset.

In order to further understand the variations in the predicted versus observed vessel noise estimates, additional analyses may be required. Specifically, the ECHO and SBC vessel noise levels were recorded in differing depths and geometries. Sensitivity to surface angle, closest point of approach, and propeller depth assumptions should be investigated to identify if the observed variations are due to effects of bottom interaction and Lloyd's mirror. Differences in the ranges of vessel operational and design parameters between the two datasets may highlight opportunities for model improvements on out-of-distribution observations. Additionally, investigating differences in repeat transits of individual vessels may elucidate more information on the variability seen within MSL predictions and observations.

Table of Contents

Introduction

The objective of the current study is to evaluate a vessel noise functional regression model developed by JASCO Applied Sciences and ERM for the Vancouver Fraser Port Authority-led Enhancing Cetacean Habitat and Observation (ECHO) Program to assess the predictive fit on an independent dataset collected by Scripps Institution of Oceanography. The project used data from the Santa Barbara Channel (SBC) collected by the Scripps Whale Acoustics Lab to test the accuracy of the model developed with transits from the ECHO database. Radiated noise levels and monopole source levels were estimated from the SBC dataset and design and operating conditions were obtained for six major commercial vessel categories. The predictor variables included actual operational information from Automated Information System (AIS) data and general vessel characteristics from Lloyd's List Intelligence (LLI).

Dataset Overview

This research project was limited to commercial vessels in the following six categories: bulk carriers and general cargo vessels, container vessels, cruise vessels, tankers, tugs, and vehicle carriers.

Each acoustic measurement in the SBC database relied on Automated Information System (AIS) data obtained from an antenna on Santa Cruz Island maintained by the Santa Barbara Wireless Foundation. Each transit was matched to records from the LLI based on the International Maritime Organization (IMO) number, whenever possible. The IMO number is a 7-digit code that uniquely identifies large cargo vessels (>300 gross tons) and large passenger vessels (>100 gross tons). Data from all sources were merged into a single vessel noise database for subsequent analysis.

Methods

Santa Barbara Channel Acoustic Dataset

High-frequency Acoustic Recording Packages (HARPs) have been maintained at a long-term acoustic monitoring station (Site B1) in the SBC at ~ 580 m depth, 3 km north of the northbound shipping lane from 2007 to 2018 and more recently at a site (Site B2) directly in the northbound shipping lane from 2018 to present (Figure 1). Recordings from both locations were used in this study. The final evaluation dataset consisted of 3,991 transits by 1,242 unique vessels, with 1,111 transits of 429 unique vessels excluded due to missing covariate values.

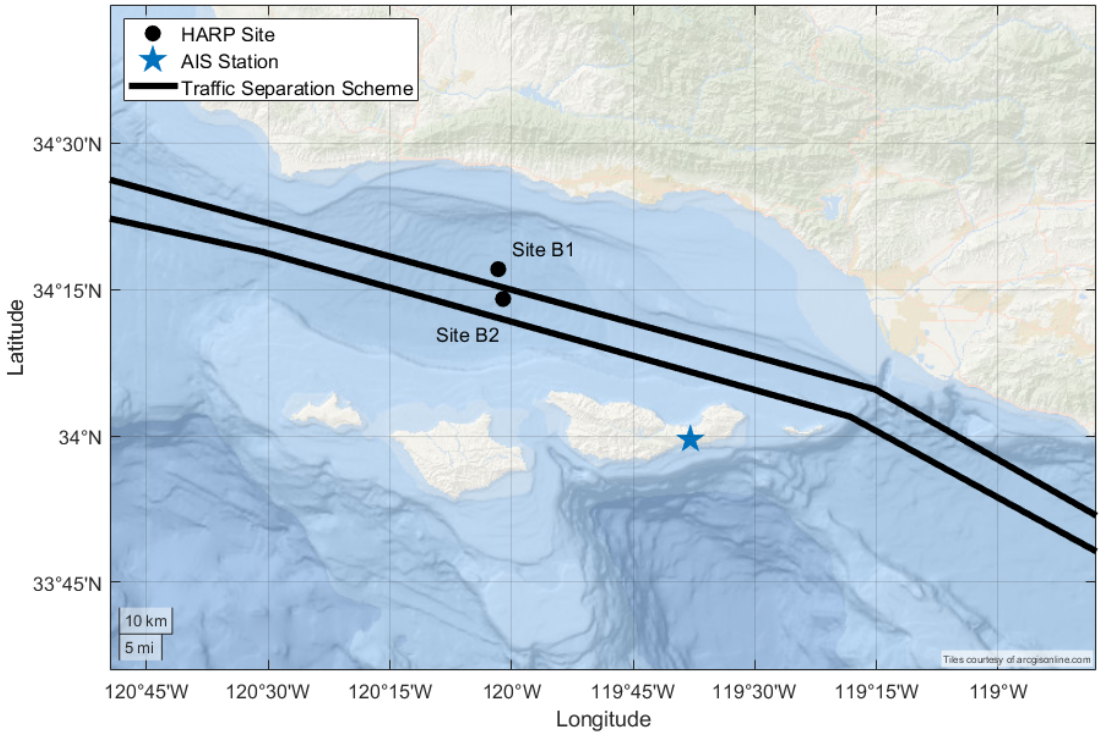

Figure 1. Map of recording locations in relation to the Santa Barbara Channel shipping lanes.

Table 1. Composition of vessel transits by ship category used to assess the ECHO model by category. Values in parentheses indicate numbers prior to removal of transits with missing covariate values.

Vessel Type	All Transits		Unique Vessels
	Site B1	Site B2	
Container	1,103 (1,285)	1,632 (1,888)	458 (560)
Bulker	181 (227)	303 (376)	348 (429)
Tanker	136 (167)	204 (287)	232 (35)
Cruise	4 (11)	21 (43)	15 (35)
Vehicle Carrier	161 (193)	245 (392)	188 (238)
Tug	69 (87)	63 (76)	34 (38)
TOTAL	1,654 (1,970)	2,468 (2,972)	1,242 (1,275)

Calculation of Source and Received Levels

HARP hydrophone electronics were calibrated at Scripps Institution of Oceanography and select full systems were calibrated at the U.S. Navy's Transducer Evaluation Center facility in San Diego, California. Acoustic recordings were collected at a sampling rate of 200 kHz at both sites. To reduce computational requirements, the recordings were decimated by a factor of 20 resulting in a sampling rate of 10 kHz. The data were low-pass filtered with an 8th order Chebyshev Type I IIR filter to prevent aliasing during decimation.

Radiated Noise Levels (RNL) were calculated based on ASA/ANSI (2009) and ISO (2019) specifications. Monopole Source Levels (MSL), correcting for the effect of Lloyd's mirror, were estimated using the approach of Gassmann et al. (2017). Both sound levels, RNL and MSL, were estimated for one-third-octave bands from 10 Hz to 2 kHz. Each vessel transit recording was divided into non-overlapping segments with a duration of 1 s. A 10,000-point (NFFT) Fast Fourier Transform (FFT) was applied to each 1 second segment to provide a frequency bin spacing of 1 Hz. The magnitude of the FFT squared was multiplied by $2/NFFT^2$ to correct for the processing gain of the FFT. Over the duration of the transit, the mean sound pressure level (SPL) was computed over each 5 s segment every 3 s to smooth the time-frequency distribution. The resulting SPLs were reported in decibels (dB) with a reference pressure of 1 μPa^2. To estimate RNL, a spherical spreading propagation loss model (N_{SS}) was calculated with the following equation:

$$N_{SS} = 20\, log_{10}(R/r_0) \qquad \textit{(Equation 1)}$$

where R is the distance from the dipole source to the receiver and r_0 is the reference distance (1 m). The N_{SS} used to compute RNL does not require a source depth (d_S), therefore the d_S is assumed to be the dipole source for all transits. The N_{SS} was applied to the SPL to achieve RNL (Equation 2).

$$RNL = SPL + N_{SS} \qquad \textit{(Equation 2)}$$

A propagation loss model that corrects for the Lloyd's mirror effect (N_{PL}) was applied to estimate MSL to account for image interference at the sea surface and for compliance with ISO (2019, Equation 4). The N_{PL} model ignores sound refraction in the water column and reflections with the seafloor and solely accounts for reflections from the sea surface (Gassmann et al., 2017; Audoly et al., 2017). The propagation loss of a sound source near the surface in deep water considering the Lloyd's mirror effect is given by:

$$N_{PL} = -20\, log_{10}\left(r_0 \left|\frac{e^{-jkr_1}}{r_1} - \frac{e^{-jkr_2}}{r_2}\right|\right) \qquad \textit{(Equation 3)}$$

$$MSL = SPL + N_{PL} \qquad \textit{(Equation 4)}$$

where r_1 is the distance from the source to the receiver, r_2 is the distance from the image source to the receiver, and k is the wave number ($k = 2\pi f/c$) in rad/m. Source depth was taken to be equal to 50% of the actual vessel draft. Harmonic mean sound speeds were calculated from depth, temperature, and salinity data obtained from the California Cooperative Oceanic Fisheries

Investigations (line 81.8, station 46.9) and California Underwater Glider Network using the nine-term equation from Mackenzie (1981) (http://calcofi.org/data.html, Rudnick, 2016).

A modification of the Lloyd's mirror model was established in Gassmann et al. (2017) to remove mismatched interference lobes identified with ship noise measurements in compliance with ANSI/ASA (2009) and ISO (2016). The modification includes using the Lloyd's mirror model from 5 Hz up to the lowest frequency at which the Lloyd's mirror model and the spherical spreading model intersect. At the higher frequencies, the spherical spreading model was used (Gassmann et al., 2017). The intersection frequency was unique for each passage.

Data Conditioning

A background noise measurement was made when each ship was 2 km farther from the hydrophone than the closest point of approach. If the difference between the signal-plus-noise-to-noise ratio (ΔL) was between 3 dB and 10 dB, an adjustment to the measurement was made with equation 4 from ISO (ISO 2016). If ΔL was less than 3 dB, the one-third-octave level was set to NA. No adjustment was made when ΔL was greater than 10 dB.

Each transit was matched with subsequent predictor variables including, length of vessel, main engine design RPM, main engine design power in kW, design speed in knots, vessel age at time of measurement, wind resistance, speed through water, and actual draft. Each transit was paired to records from the LLI, if available, to obtain the predictor variables for design. Vessel age at time of measurement was calculated by subtracting the year built (as obtained from LLI) from the year of the transit. Vessel draft at the time of measurement was obtained from AIS data. Speed through water at the time of measurement was calculated from AIS speed over ground and the current speed and direction obtained from HF Radar Data from the Southern California Coastal Observing System (https://hfradar.msi.ucsb.edu/). Wind resistance was calculated from wind speeds and direction obtained from the National Oceanic and Atmospheric Administration buoy station #46053. Throughout the following analysis, actual draft, speed through water, wind resistance, and surface angle are referred to as "operational covariates", and length of vessel (LOA), main engine design RPM, main engine design power in kW, design speed in knots, and vessel age at time are considered to be "design covariates" following MacGillivray et al. (2020). Actual draft, speed through water, LOA, main engine design RPM and kW and design speed were log-transformed throughout this analysis.

Data Imputation

Minimal imputation was conducted for one category, tugs, because the LLI database only included engine and design speed specifications for one vessel in the set. The values provided for the single represented tug were used to impute the main engine design RPM, main engine design kW and design speed for the remaining tug transits. This reduced covariate missingness for the tug class from 99.4% to 19.0%, however this may result in unrealistic uniformity in predictions for the tug class. Tugs tended to be older, smaller, and poorly-documented. The parameters which were imputed were important predictors for the tug class in MacGillivray et al. (2020), therefore the results for this class in this study should be considered illustrative.

Model Evaluation

The existing model was run for each vessel category using the AIS, buoy and LLI-derived covariate set. A constant surface angle of 30 degrees, equal to the mean surface angle in the ECHO dataset, was used to evaluate the model, following MacGillivray et al. (2020). Model-predicted MSL and RNL one-third-octave spectra for each transit were compared with the data-derived spectra. Model residuals within each vessel category were evaluated as a function of frequency and as a function of each covariate. For the purposes of this analysis, repeat transits by a single same vessel were treated as independent observations. To preserve the relationship to the original units, residuals are reported as mean signed deviation (MSD) when the sign or slope of the residuals is of interest, as

$$MSD_{\hat{\theta}} = \frac{1}{n}\sum_{i=1}^{n}(\hat{\theta}_i - \theta_i) \qquad \textit{(Equation 5)}$$

where $\hat{\theta}_i$ represents the i^{th} model prediction across n frequency bands and θ_i represents the associated observation. Mean absolute error (MAE) is reported when considering model error in the transit-by-transit case as well as overall model uncertainty. MAE was computed as:

$$MAE_{\hat{\theta}} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{\theta}_i - \theta_i\right| \qquad \textit{(Equation 6)}$$

MSD, and MAE values were computed for each transit across all frequency bands. Means and standard deviations of each of these parameters are reported to summarize the overall deviations between the observations and model predictions across transits. To assess potential trends in model residuals, a moving average smoothing filter of the signed deviations was computed across the range of each covariate range using a sliding window length equal to one-fifth of the number of observations.

Results

Exploratory Analysis

Covariates derived from AIS and observational buoys, including draft, speed through water, wind resistance and surface angle had few missing values (Table 2). Missingness was highest amongst design parameters obtained from the LLI, particularly for the cruise ship and tug classes. In total 820 of the original 4,942 transits were excluded due to missing covariates, leaving a total dataset size of 4,122 transits. The majority of transits retained were container ships (66%), followed by bulkers, vehicle carriers, and tankers which each represented approximately 10% of transits.

Missingness in MSL and RNL observations was introduced by the conditioning process, and was strongly related to recording location at frequencies below 100 Hz and above 1 kHz for all vessel categories (Figure 2). Between 20 and 30% of transits recorded at site B1, farthest from the SBC shipping lanes, had missing values below 100 Hz. Missingness for site B1 transits was reduced to approximately 10-20% from 100 to 1000Hz, but increased again at higher frequencies. Transits recorded at site B2, in close proximity to the shipping lanes, had relatively uniform missingness rates of approximately 10% across all frequencies. This between-site difference is likely due to greater attenuation of low frequencies through interaction with the seafloor (i.e., bottom loss) or signals traveling from the shipping lanes to the more distant site B1, as well as

frequency-dependent attenuation of higher frequencies as a function of range. Attenuation at these frequencies reduced amplitudes below a minimum threshold relative to the standardized estimate of background noise. Missingness was identical for MSL and RNL estimates following the application of conditioning criteria.

Table 2. Covariate missingness as a percentage of total transits in each class. Values in parentheses represent missingness prior to data imputation.

	Draft	STW	Wind Resistance	Surface Angle	LOA	Main engine design RPM	Main Engine kW	Design Speed	Vessel Age	Cumulative Missingness
Bulker	0.2%	0%	0.2%	0%	0%	7.1%	9.8%	7.8%	0.3%	19.7%
Container	0%	0%	0.2%	0%	1.4%	7.4%	13.9%	4.9%	1.4%	17.8%
Cruise	0%	0%	0%	0%	5.6%	48.1%	27.8%	29.6%	13.0%	53.7%
Tanker	0%	0	0.2%	0%	0%	11.2%	22.2%	1.3%	0%	25.1%
Tug	0%	0	0%	0%	1.8%	0% (99.4%)	0% (99.4%)	0 % (99.4%)	19.0%	19.0% (99.4%)
Vehicle Carrier	0%	0	0.2%	0%	0%	9.9%	8.9%	8.5%	0%	18.0%

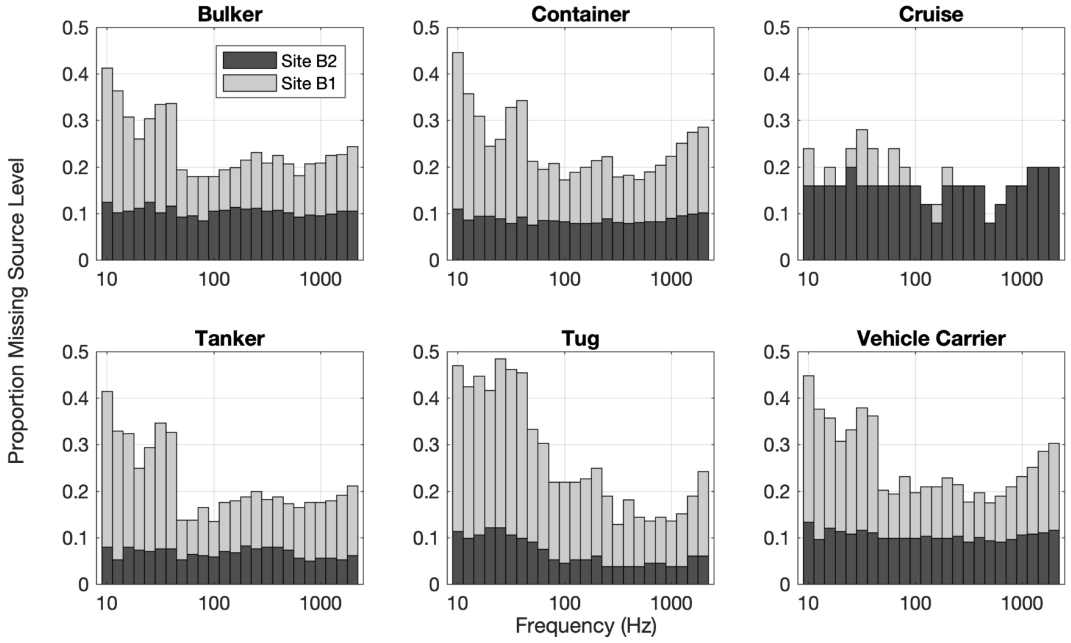

Figure 2. Missingness of MSL and RNL observations as a function of frequency following application of data conditioning criteria. Missingness was higher at low and high frequencies for site B1 (light gray) than site B2 (dark gray) due to attenuation of those frequencies associated with bottom interaction and range respectively. Bars are stacked to show total missingness, with color indicating the contribution of from each site to the total.

Operational covariates including draft, speed through water, wind resistance and surface angle were examined by vessel category (Figure 3). A bimodal distribution was apparent for surface angle, due to the inclusion of transits from the two different monitoring locations, B1 and B2. Cruise ships do not always follow the shipping lanes in the SBC region and therefore surface angles for these vessels were more variable than for other categories. Wind resistance was slightly skewed toward lower values with a long tail associated with occasional high wind conditions. Wind resistance was higher in the B1 observations which consisted to primarily of northbound transits, due to proximity of the northbound shipping lane. The B2 transits were more balanced across the north and southbound lanes. Speed through water was bimodal for vehicle carriers and cargo ships, and highly variable for cruise ships. Drafts were deeper for bulkers and tankers for B2 transits, indicating possible differences in load between north and southbound vessels, while container ship drafts were comparable between the two sites. Tugs stood out as operating at slower speeds and with reduced draft, and experiencing lower wind resistance (presumably related to smaller vessel sizes) relative to other categories.

The distributions of operational covariates differed in some cases from those observed within the original training dataset (Figures 3 and 4). Surface angle was the most notable difference: In the ECHO datasets, mean surface angle was approximately 30 degrees, whereas in the SBC dataset, a bimodal distribution was observed with peaks at approximately 5 and 75 degrees. For this reason, 30 degrees was used as a fixed input to the predictive correlation based model for the SBC dataset. Wind resistance tended to be higher in the SBC dataset, and was rarely negative. This is partially due to a majority of the B1 measurements having been taken from vessels transiting in the northbound shipping lane, however a bias toward primarily positive wind resistance was also observed in the ECHO dataset as any following wind less than vessel transit speed over ground still results in a positive wind resistance factor. Speeds through water were lower for bulkers, container ships, tankers, and vehicle carriers which may reflect partial vessel speed reduction program (VSR) participation (ZoBell et al. 2021). Cruise ship and tug speeds were not significantly different between the studies. Actual drafts were not notably different between the SBC and ECHO datasets.

Design covariate distributions were generally discrete and highly variable (Figure 6). On average, container ships were the largest, fastest and most powerful category according to design specifications, followed by cruise ships. However, operational speed did not always mirror design speed. Variability in operational speed relative to design speed may again reflect cooperation with the VSR program. Most vessels were under 20 years old at the time of passage, with the exception of tugs and cruise ships which tended to be older than other categories. Additionally, outliers existed for container, bulkers, and tankers at approximately 40, 50, 60 years, respectively. Cruise ships had outliers for main engine design RPM and kW at approximately 1,800 RPM and 60,000 kW.

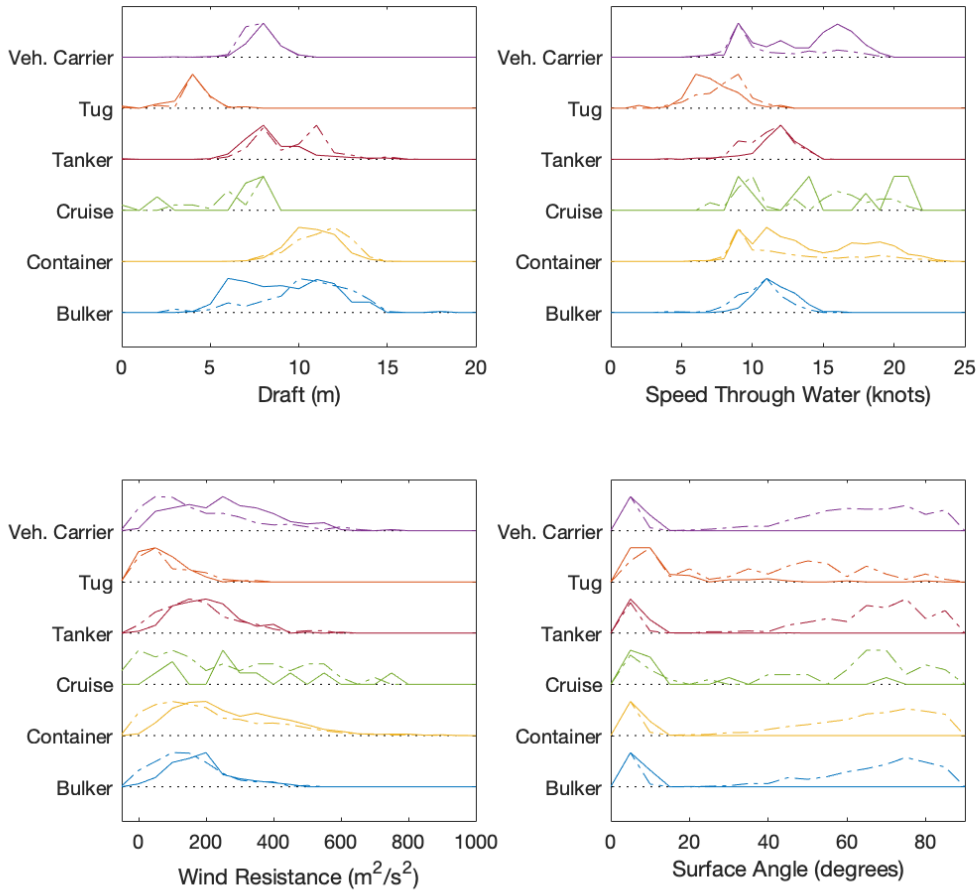

Figure 3. Distributions of operational covariates in SBC dataset by vessel category. Solid line represents B1 transits, dash-dotted line represents B2 transits.

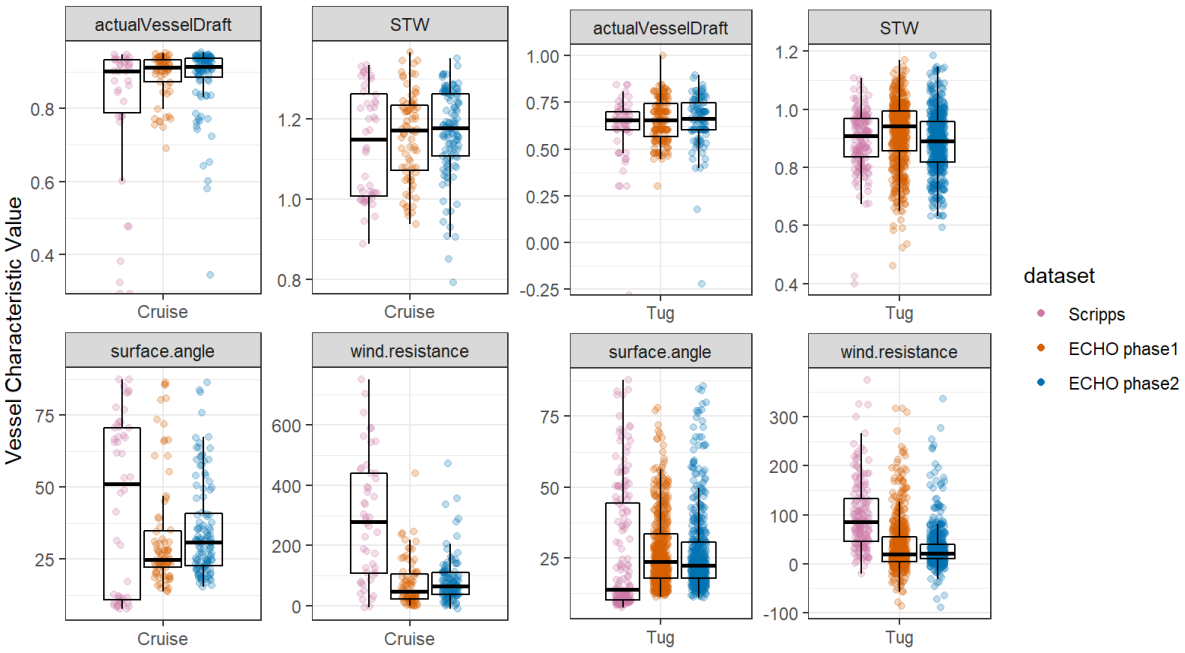

Figure 4. Comparison of the distribution of operational covariates for cruise ships and tugs between the SBC dataset used for model evaluation and the ECHO datasets used for model development. (Note: Values shown for speed through water (STW) and draft have been log-transformed.)

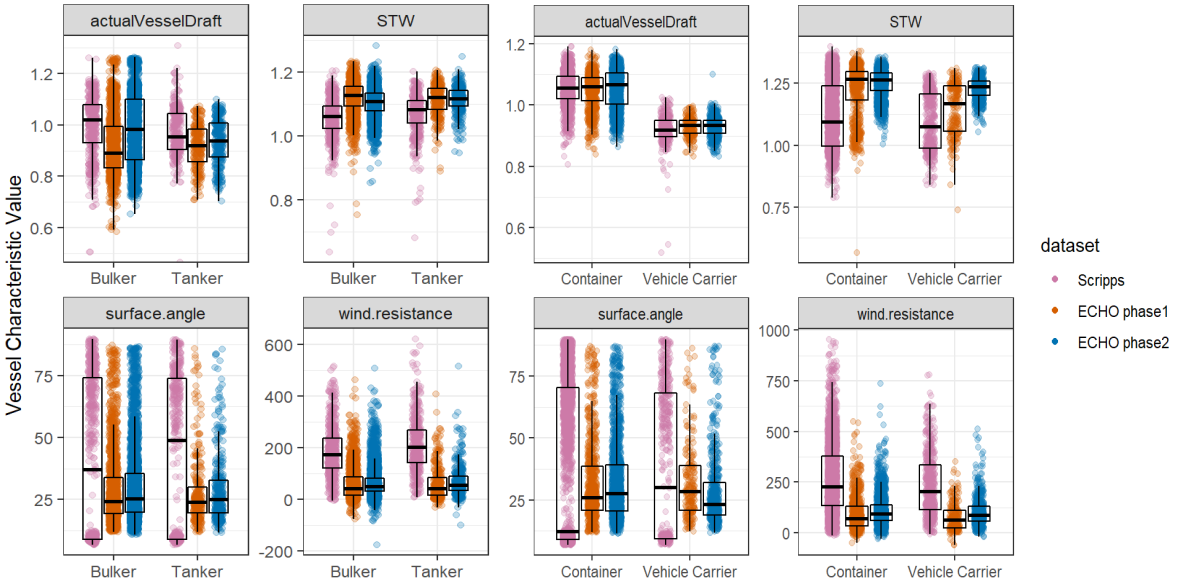

Figure 5. Comparison of the distribution of operational covariates for bulkers, tankers, container ships and vehicle carriers between the Scripps (aka SBC) dataset used for model evaluation and the ECHO datasets used for model development. (Note: Values shown for STW and draft are the log-transformed values rather than the originals.)

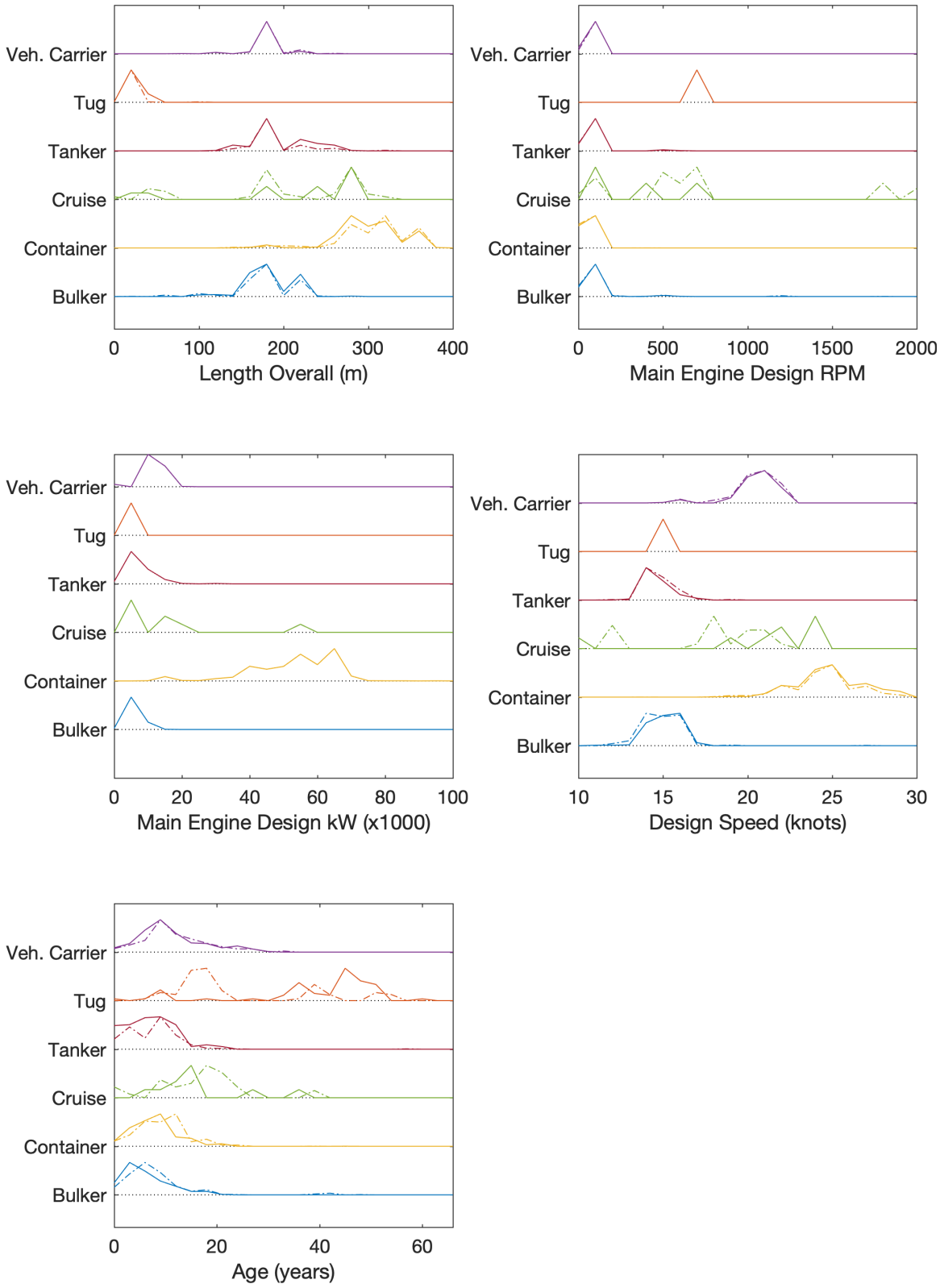

Figure 6. Distributions of design covariates in SBC dataset by vessel category. Solid line represents B1 transits, dash-dotted line represents B2 transits.

Model Evaluation

MSL Model

The MSL model predicts source level at 1 m and is expected to account for propagation loss effects related to source depth (Lloyd's mirror effect) and harmonic mean sound speed. MSL predictions were highest on average for tankers, container ships and bulkers, in general agreement with the observations (Figure 7). Lowest mean MSLs were predicted and observed for cruise ships. Tugs also had lower observed MSLs, however the availability of accurate covariates for this class, as discussed in the data conditioning section above, limited conclusive assessment of the model predictions.

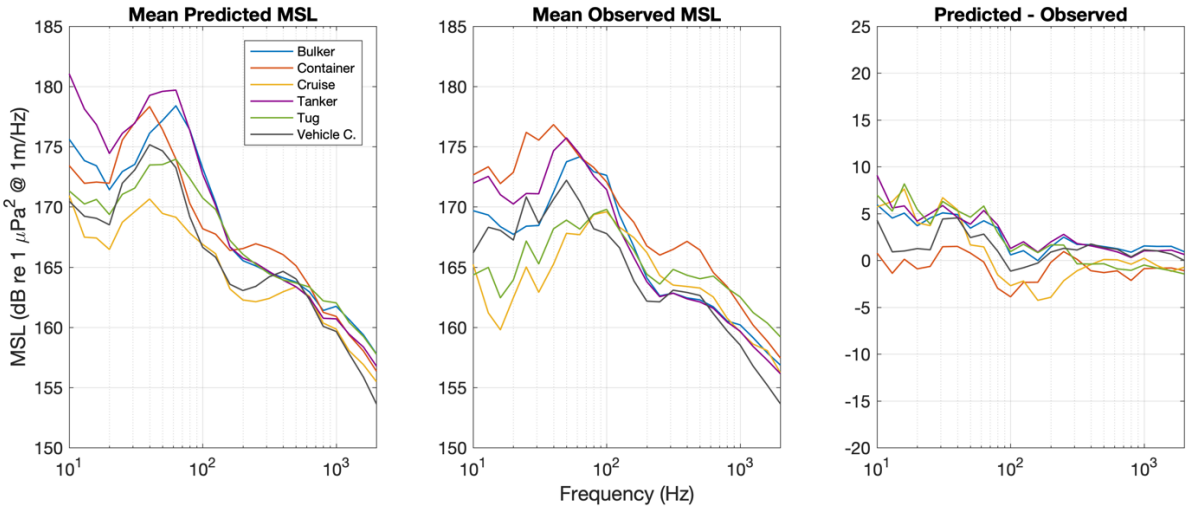

Figure 7. Mean predicted, mean observed and mean differences in MSL by vessel category for the SBC dataset.

The accuracy of model predictions varied by vessel type. Models tended to estimate well on average above 100 Hz. For some vessel classes, including bulkers, tankers, and tugs, MSL tended to be overestimated below 100 Hz (Figure 8 and Figure 9). However, this systematic overestimation was not seen for the largest container vessel category, and was minimal for vehicle carriers. MSL model predictions for container ships, cruise ships, and vehicle carriers showed some bimodality, due to bimodality in ship speeds within these categories. This did not translate into bimodality in the model residuals, suggesting that the predicted differences were observed *in-situ*. Low variability in model predictions for the tug class is directly related to imputation of up to three missing design parameters with identical values across all tug transits.

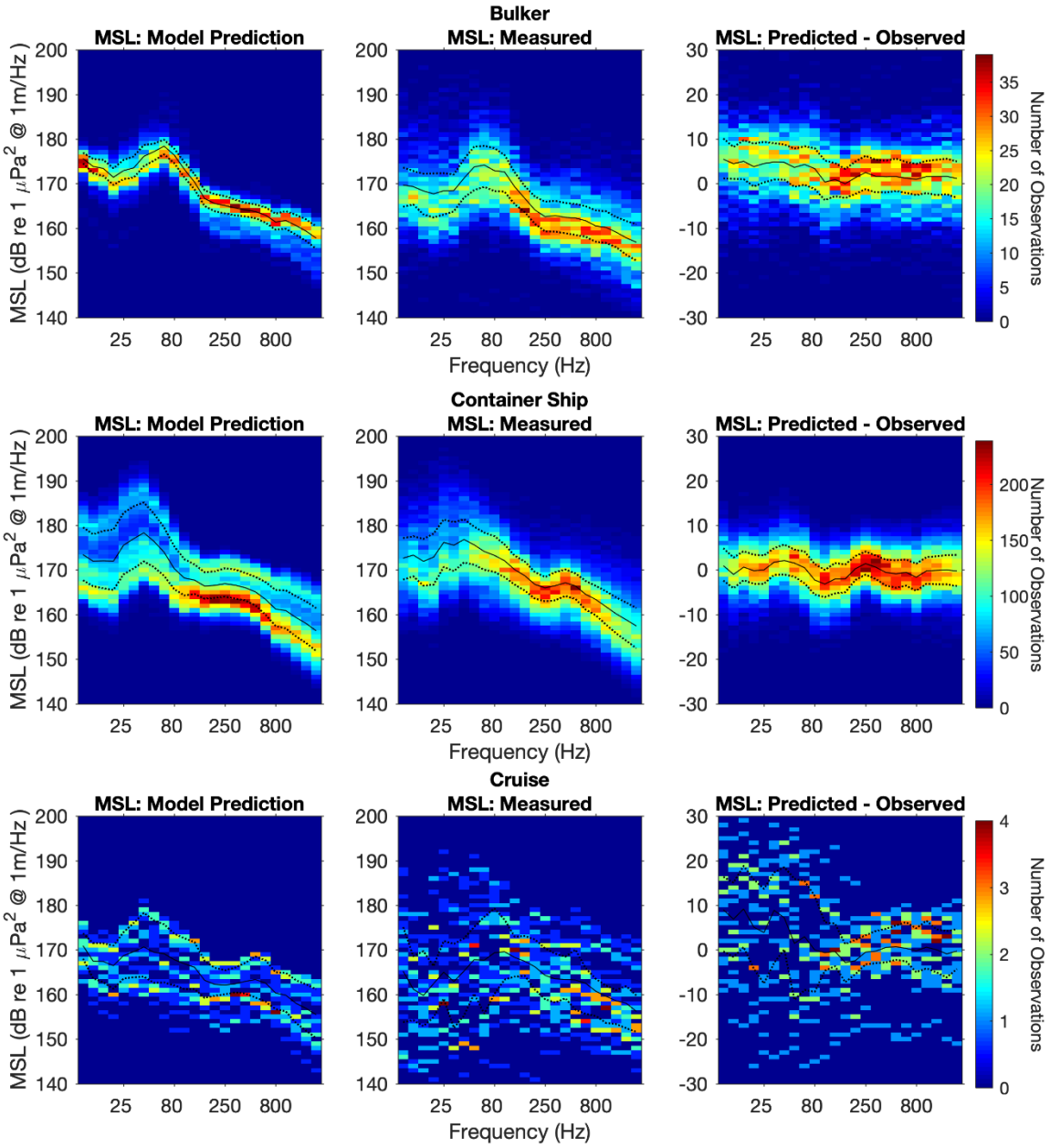

Figure 8. Predicted (left) and observed (center) MSL values for bulkers, container ships and cruise ships displayed as a heatmap with the number of observations in each grid cell represented on a linear color scale from blue (no observations) to red (maximum observations). Panels on the right represent the difference between model predictions and observations (predicted minus observed). Perfect agreement between the model and observations would be represented as a difference of zero. Black line represents the mean, with dotted lines indicating the 25th and 75th percentiles.

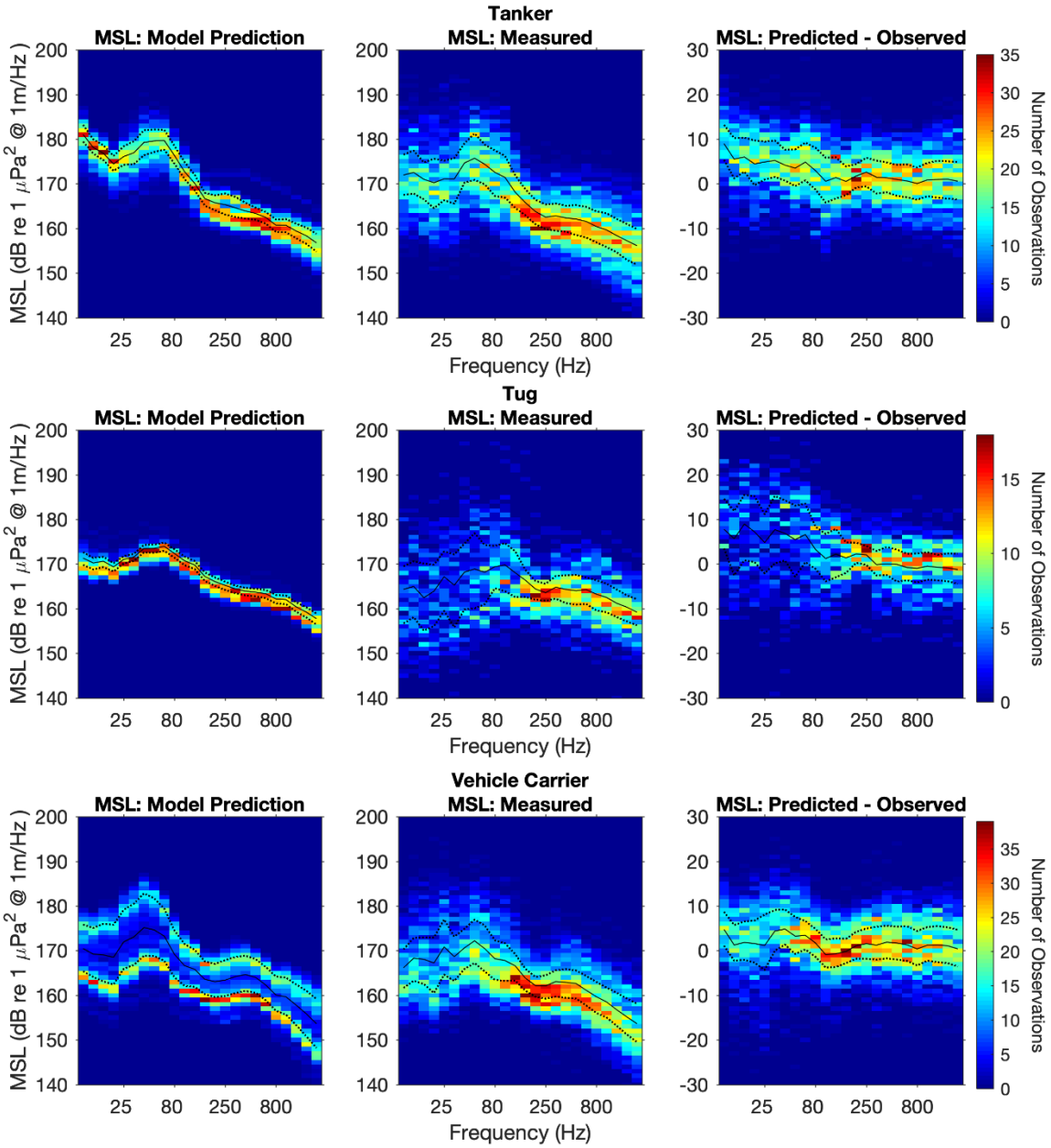

Figure 9. Predicted (left) and observed (center) MSL values for tankers, tugs and vehicle carriers displayed as a heatmap with the number of observations in each grid cell represented on a linear color scale from blue (no observations) to red (maximum observations). Panels on the right represent the difference between model predictions and observations (predicted minus observed). Perfect agreement between the model and observations would be represented as a difference of zero. Black line represents the mean, with dotted lines indicating the 25th and 75th percentiles.

Absolute error in MSL model predictions above 100 Hz was approximately 5 dB across most frequency bins for all vessel categories (Figure 10). This means that any single prediction is expected to be approximately 5 dB different than the observed value on average. This value is similar to the transit-level error in the ECHO dataset. Mean error increased to approximately 5-10 dB below 100 Hz for cruise ships and tugs. Residuals in MSL model predictions are approximately normally-distributed around zero across the frequency band, with broader variability below 100Hz for most classes. Cruise ships and tugs were the two classes with the weakest predictions due to small sample size and data imputation respectively, therefore greater uncertainty was expected.

Few strong trends were visible in the effect of covariates on residuals (Figures 11- 13). Speed through water appeared to be associated with under-prediction at low speeds and over-prediction at high speeds for container ships and vehicle carriers. Large surface angles were associated with slightly increased over-prediction of MSL for bulkers, cruise ships, tankers, and tugs, however the reverse was seen for container ships, with greater error at large surface angles. For container ships and vehicle carriers, MSL tended to be underestimated at low wind resistance conditions and overestimated at high wind resistance, however the opposite was true for bulkers. The remainder of the observed residual trends seem to be related to extrapolating the moving average smoothing filter across a region with sparse data in certain categories (cruise ships and tugs) and gaps in the covariate distributions (e.g. engine RPM), and may not be reliable.

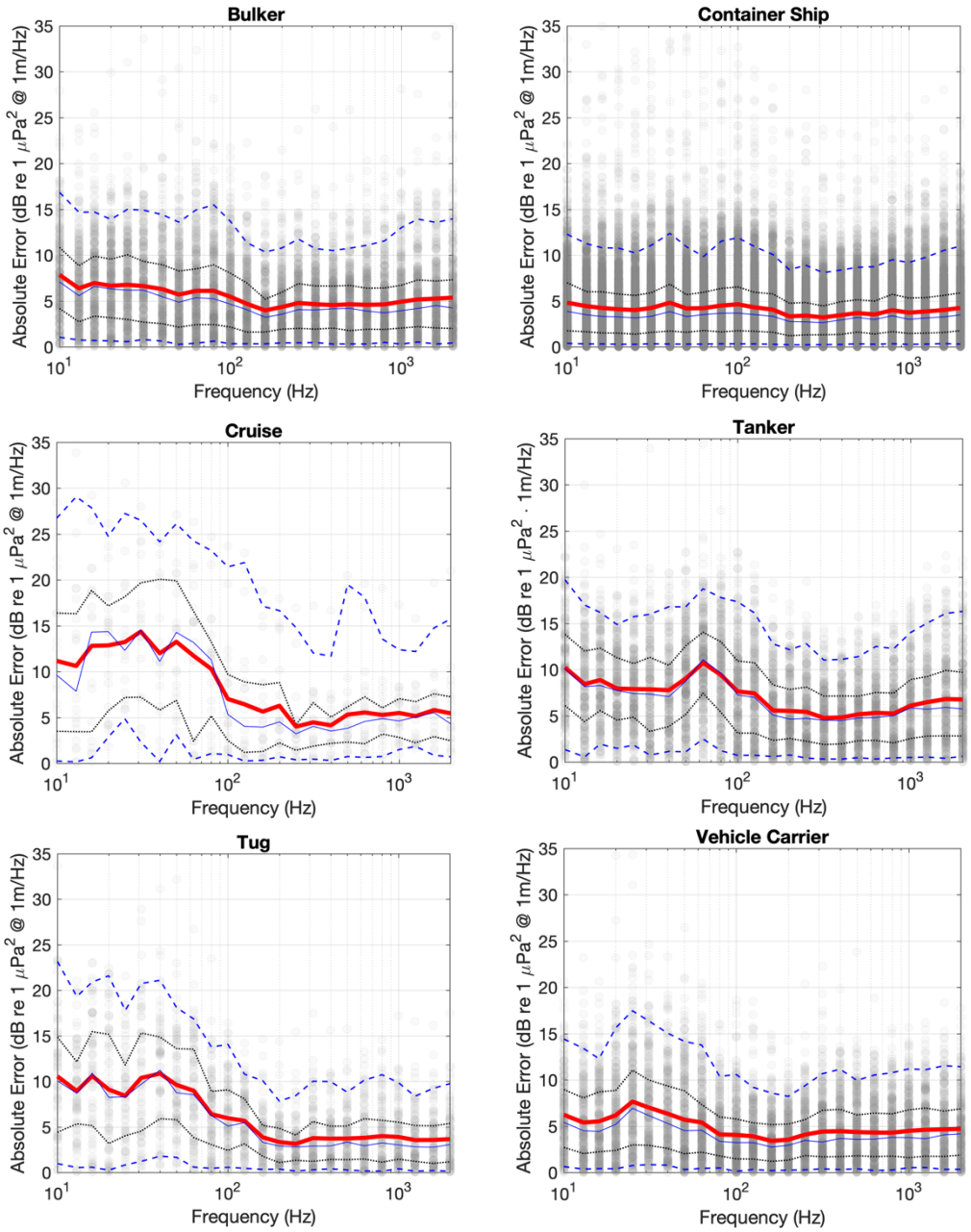

Figure 10. Absolute error as a function of frequency for each vessel category. Each point represents one observed difference between a model prediction and the associated observation. Points are semi-transparent so that the relative density of points can be interpreted, with darker values indicating many overlapping observations and light values indicating sparse observations. Lines denote the mean (solid red), median (solid blue), 25th and 75th percentiles (dotted black), and 5th and 95th percentiles (dashed blue). Offset between the mean and median is an indication of deviance from a symmetric distribution.

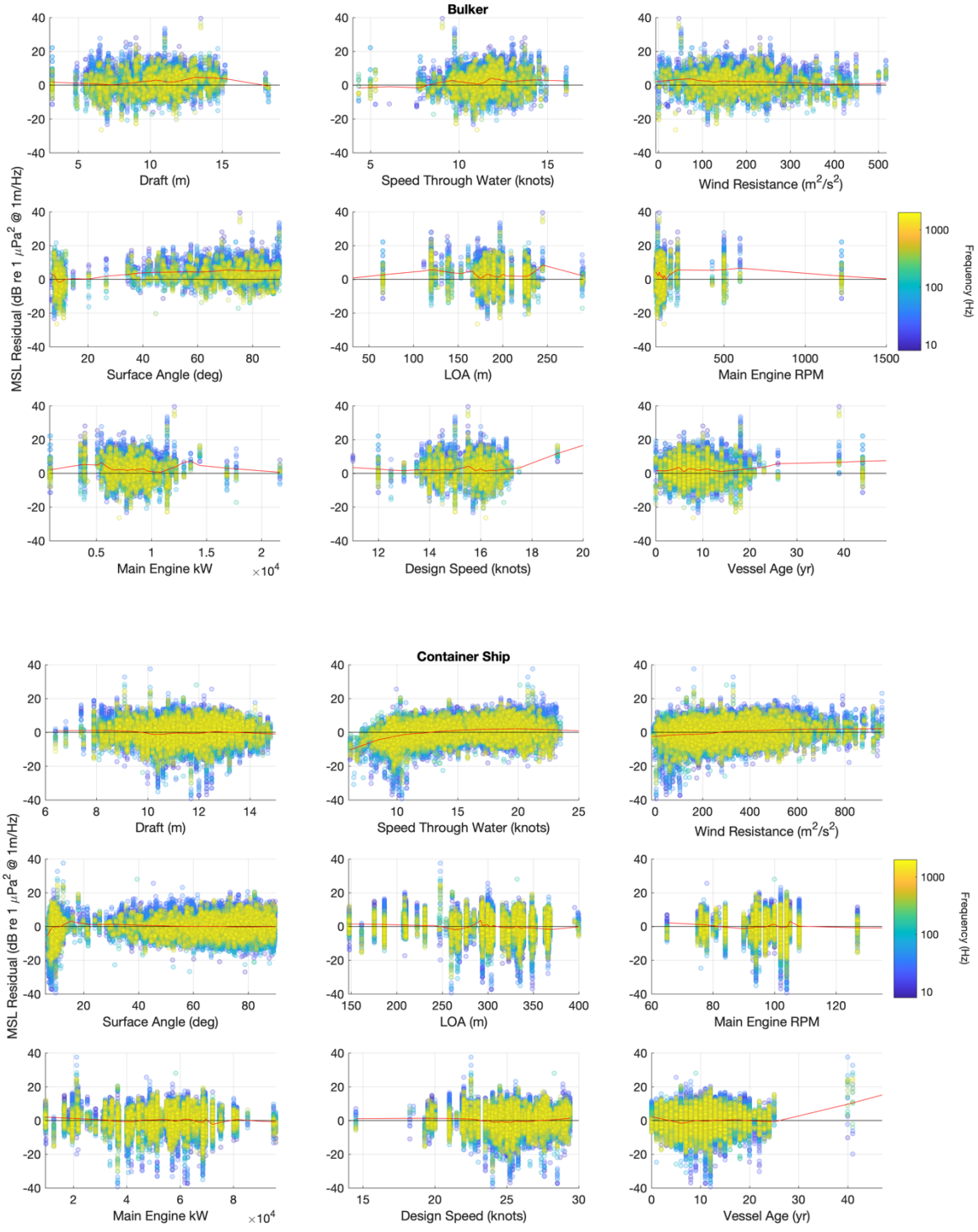

Figure 11. Distribution of MSL model residuals as a function of nine model covariates for bulkers (top) and container ships (bottom). Each point represents a residual value for one transit, within a single frequency band, with color representing frequency. The black line denotes a residual value of zero, and the red line represents a smooth of the mean taken across all frequencies.

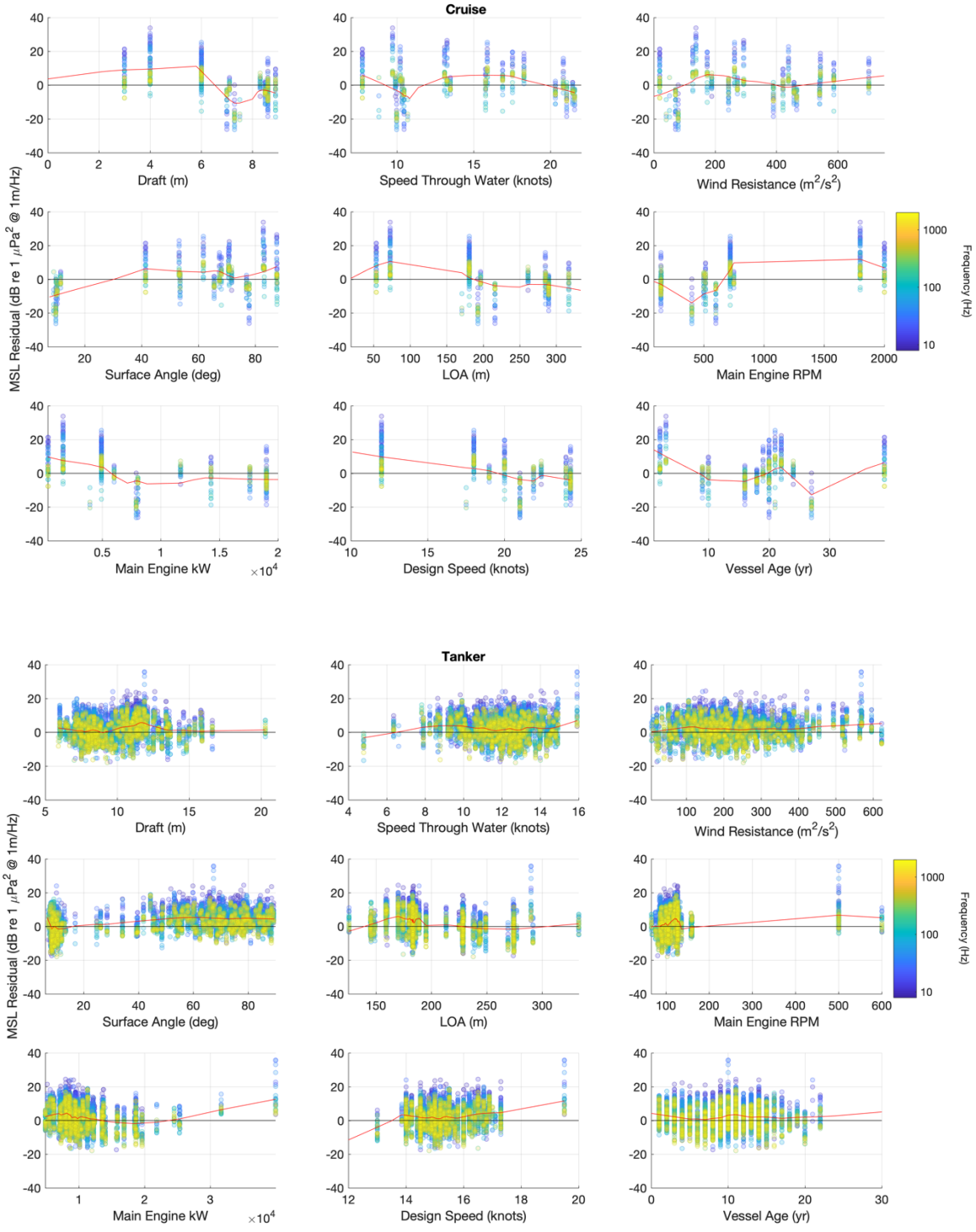

Figure 12. Distribution of MSL model residuals as a function of nine model covariates for cruise ships (top) and tankers (bottom). Each point represents a residual value for one transit, within a single frequency band, with color representing frequency. The black line denotes a residual value of zero, and the red line represents a smooth of the mean taken across all frequencies.

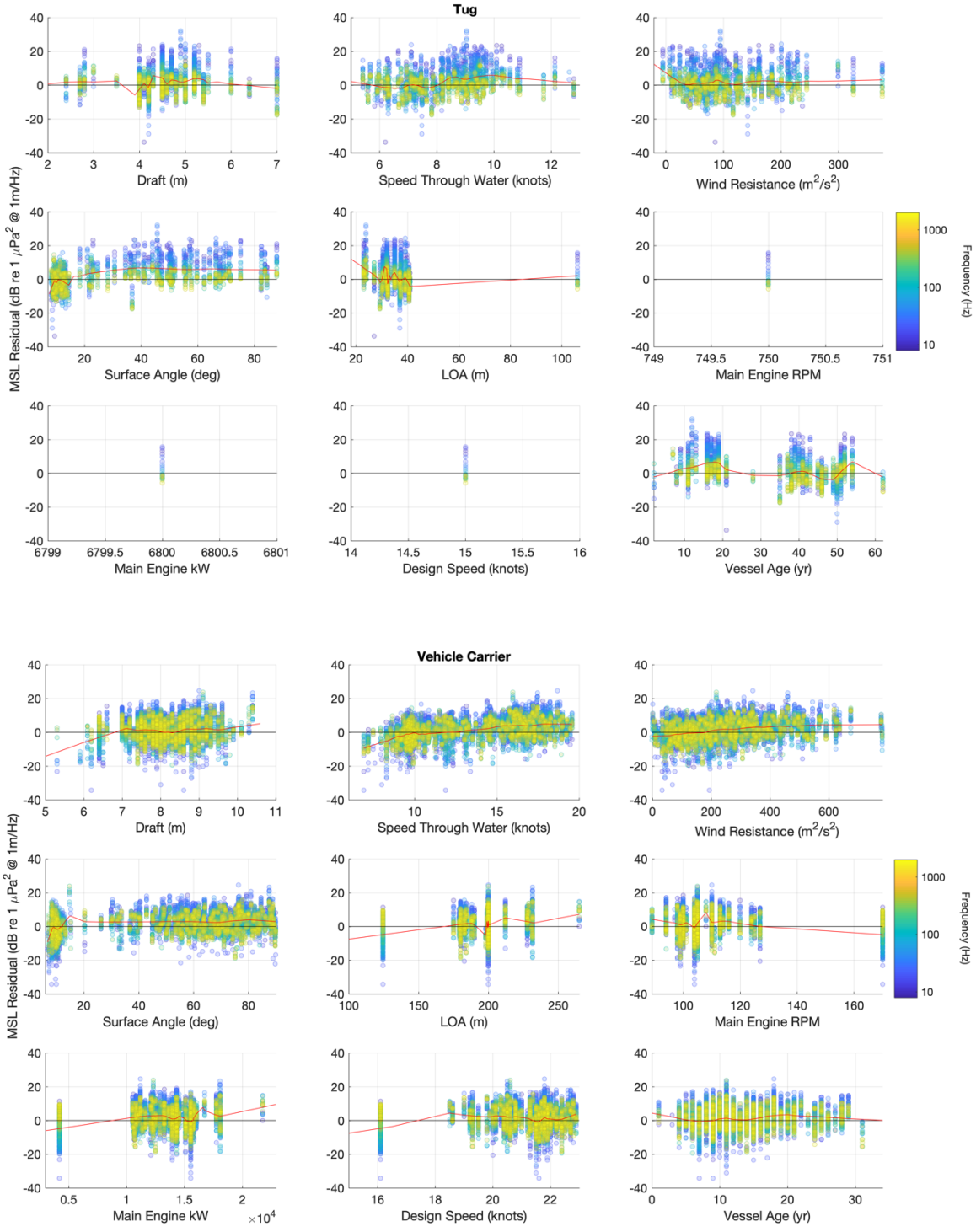

Figure 13. Distribution of MSL residuals as a function of nine model covariates for tugs (top) and vehicle carriers (bottom). Each point represents a residual value for one transit, within a single frequency band, with color representing frequency. The black line denotes a residual value of zero, and the red line represents a smooth of the mean taken across all frequencies.

The effect of surface angle on model predictions for the SBC dataset can be viewed by separating model predictions and observations by site (Figure 14). Model-predicted MSLs for the transits recorded at the B1 and B2 sites were similar, with slightly higher predictions for site B1 transits (1-2 dB higher on average) below 100 Hz. Measured MSLs from site B1 transits were 5 to 10 dB higher on average below 100 Hz than site B2 transits, and 1-5 dB higher above 100 Hz, for the four most common vessel categories. Container ships were an exception, with minimal differences in measured MSL on average above 100Hz. On average, the model's predictions were closest to the MSLs estimated for the site B1 transits.

The difference between measured MSL estimates at the two sites may be related to some combination of preferential exclusion low frequency measurements from low speed as well as lower amplitude transits during the data conditioning step, which may have contributed to the higher mean MSL at this more distant B1 site. Speeds through water were higher on average in the ECHO dataset used to develop the model than the SBC dataset. The observed improvement in low frequency predictions at the more distant site may be partially related to the exclusion of lower amplitude observations below 100 Hz associated with slow transits following the conditioning process. Incomplete accounting for propagation loss with range may also play a role in the observed differences between sites. Other explanations include systematic differences in load, draft, speed through water or wind resistance between northbound and southbound vessels. The site B1 transits include a greater number of northbound vessels with higher wind resistance. Mean differences between model predictions and observations for all four major classes are flatter for B2 transits, indicating a more constant offset across the frequency band. It is interesting to note that the container ship class, which was the largest, showed the least difference between sites and between the model predictions and observations. A deeper investigation of the differences between the vessel-specific models, such as a thorough sensitivity analysis or testing with more complex sound propagation models, could help clarify factors leading to overestimation in some cases.

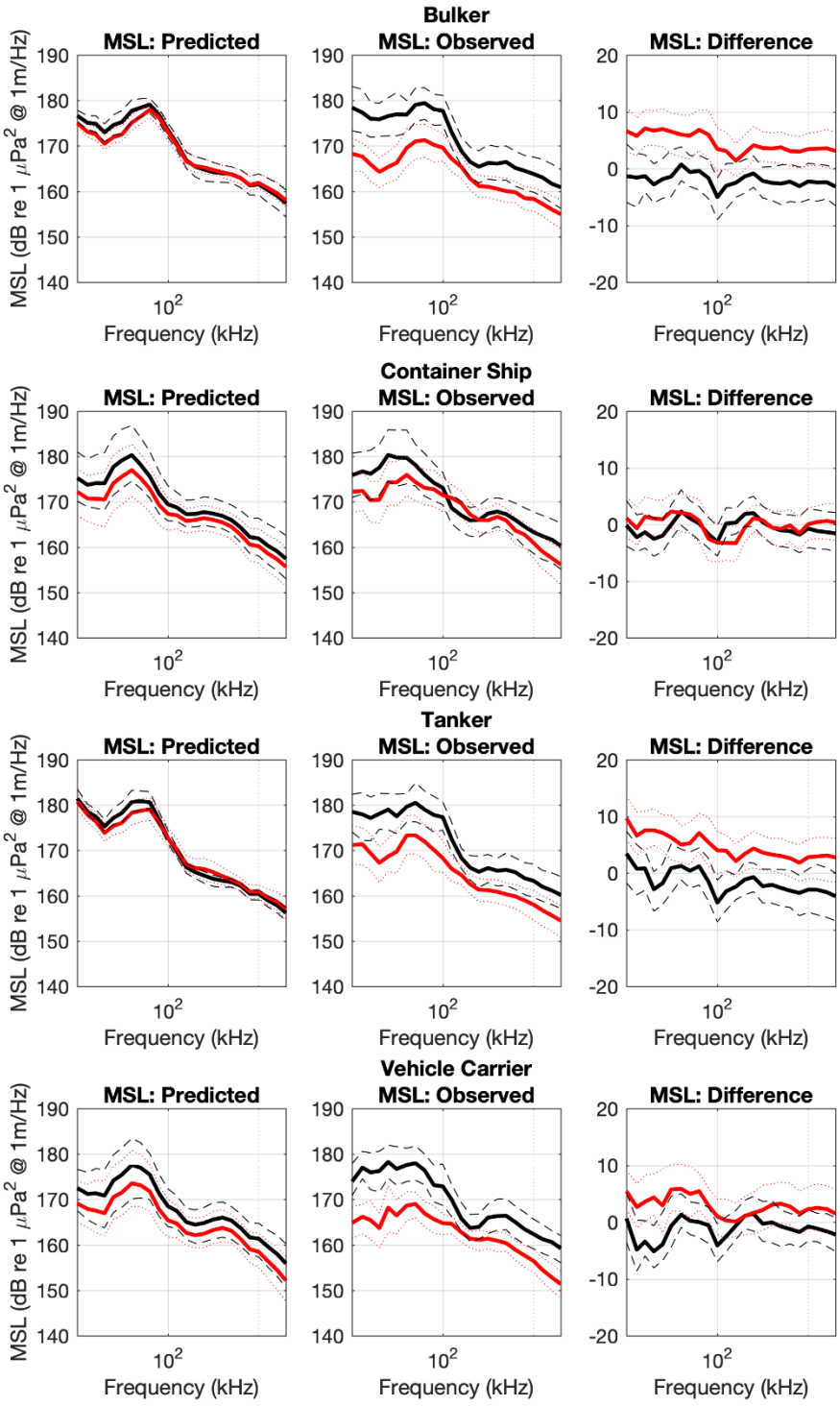

Figure 14. Comparison of model predictions and residuals for the four main vessel categories across the SBC HARP sites, B1 (black) and B2 (red). Means of site B1 transits (solid black line) and B2 transits (solid red line) are shown, with 25th and 75th percentiles as dashed and dotted lines respectively. Differences between site B1 and B2 observations are likely attributable to a combination of data conditioning, incomplete accounting for propagation loss, and systematic differences between north and southbound vessels' operational characteristics.

RNL Model

The RNL model predicts radiated noise as estimated based on a simple spherical spreading model for underwater sound propagation. The observations are not corrected for differences between locations associated with source depth, bathymetry, bottom composition, or water temperature. The Lloyd's mirror effect, which can significantly amplify or reduce lower frequency RNL values through constructive and destructive interference, is not accounted for in the simple propagation model. Although the RNL model could presumably learn to account for environmental effects from data for a particular region, it was not expected to translate as well as the MSL model to a novel recording environment in the Santa Barbara Channel in this study.

In general, the RNL model over-predicted by approximately 10 dB below 100 Hz, and by 2 to 5 dB above 100 Hz for all vessel types (Figure 15). On average, both predicted and observed RNL for bulkers, container ships and tankers were highest, while cruise ships and tugs were lowest. RNL estimates and observations were intermediate for vehicle carriers. It is notable that container ship RNLs were not more accurate than other categories, despite having been considerably more accurate in the MSL case.

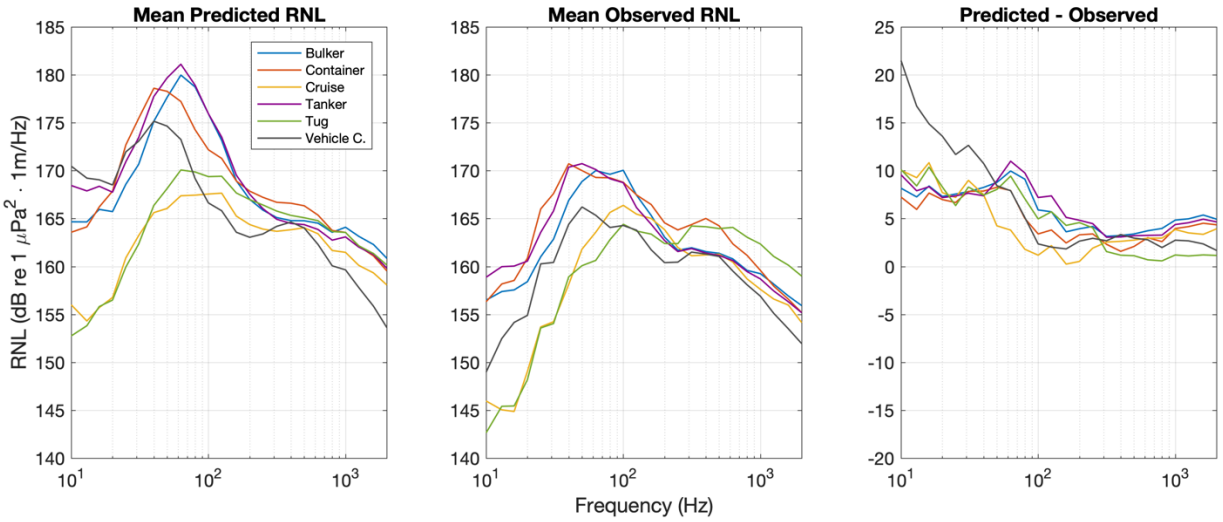

Figure 15. Mean predicted and observed RNL by vessel category for the SBC dataset.

RNL model predictions for container ships, cruise ships and vehicle carriers showed some bimodality, which is attributed to bimodality in observed vessel speeds (Figure 16 and Figure 17). This did not translate into bimodality in the model residuals, suggesting that the model correctly predicted the size of the effect associated with speed differences. Mean absolute errors were higher and more variable below 100 Hz than above (Figure 18). Vehicle carrier RNLs were greatly over-predicted below 100 Hz, and the observations for this category were lower than many of the other vessel categories. This discrepancy may be related to a large number of slow speed (~10 knot) vehicle carrier transits in the SBC dataset, which were well below typical speeds through water represented in model development dataset. Propeller speeds for some of these transits could fall below the point of cavitation inception, leading to large reductions in measured RNL.

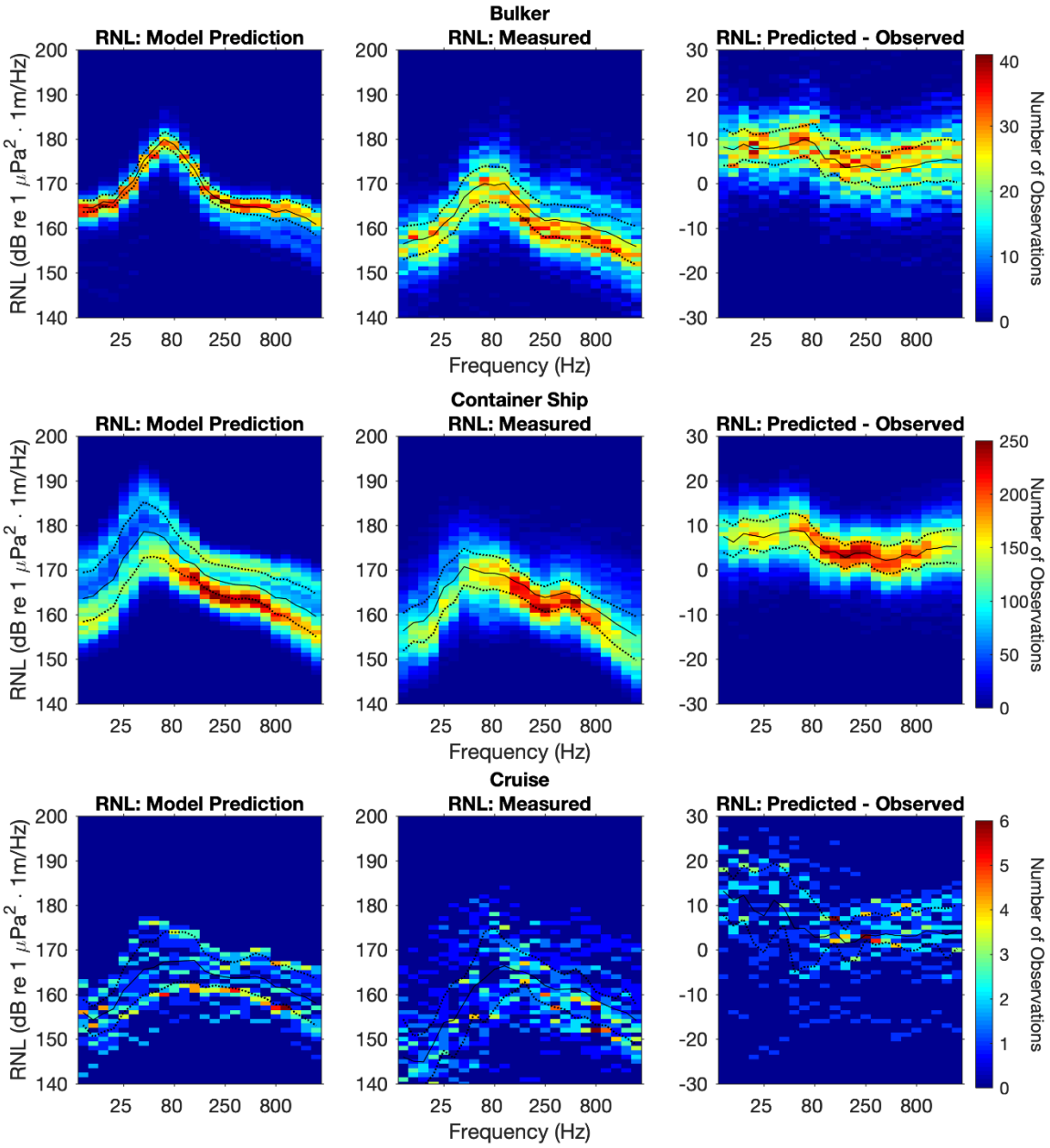

Figure 16. Predicted (left) and observed (center) RNL values for bulkers, container ships and cruise ships displayed as a heatmap with the number of observations in each grid cell represented on a linear color scale from blue (no observations) to red (maximum observations). Panels on the right represent the difference between model predictions and observations (predicted minus observed). Perfect agreement between the model and observations would be represented as a difference of zero.

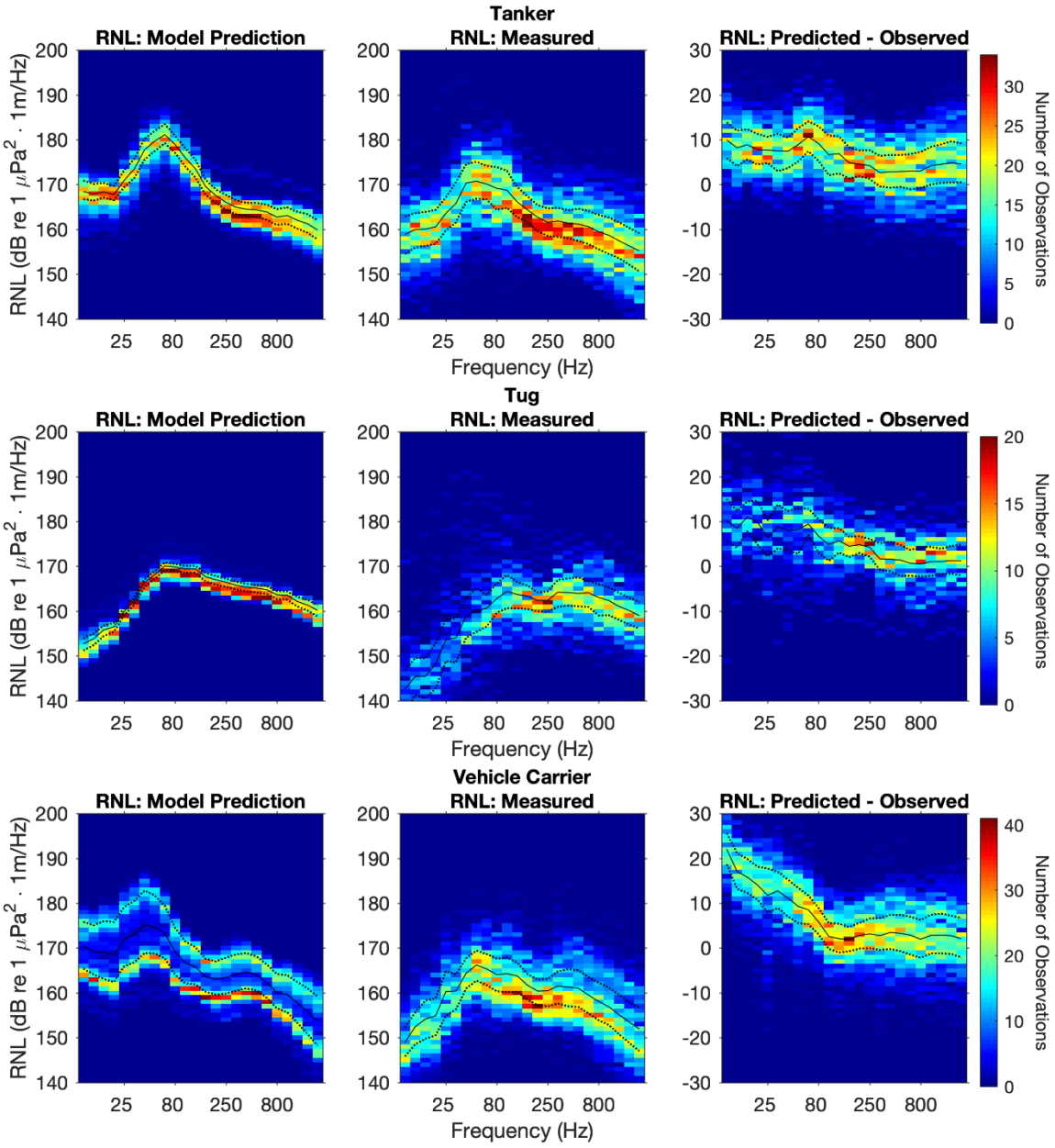

Figure 17. Predicted (left) and observed (center) RNL values for tankers, tugs and vehicle carriers displayed as a heatmap with the number of observations in each grid cell represented on a linear color scale from blue (no observations) to red (maximum observations). Panels on the right represent the difference between model predictions and observations (predicted minus observed). Perfect agreement between the model and observations would be represented as a difference of zero (dotted while line). Black line represents the mean, with dotted lines indicating the 25th and 75th percentiles.

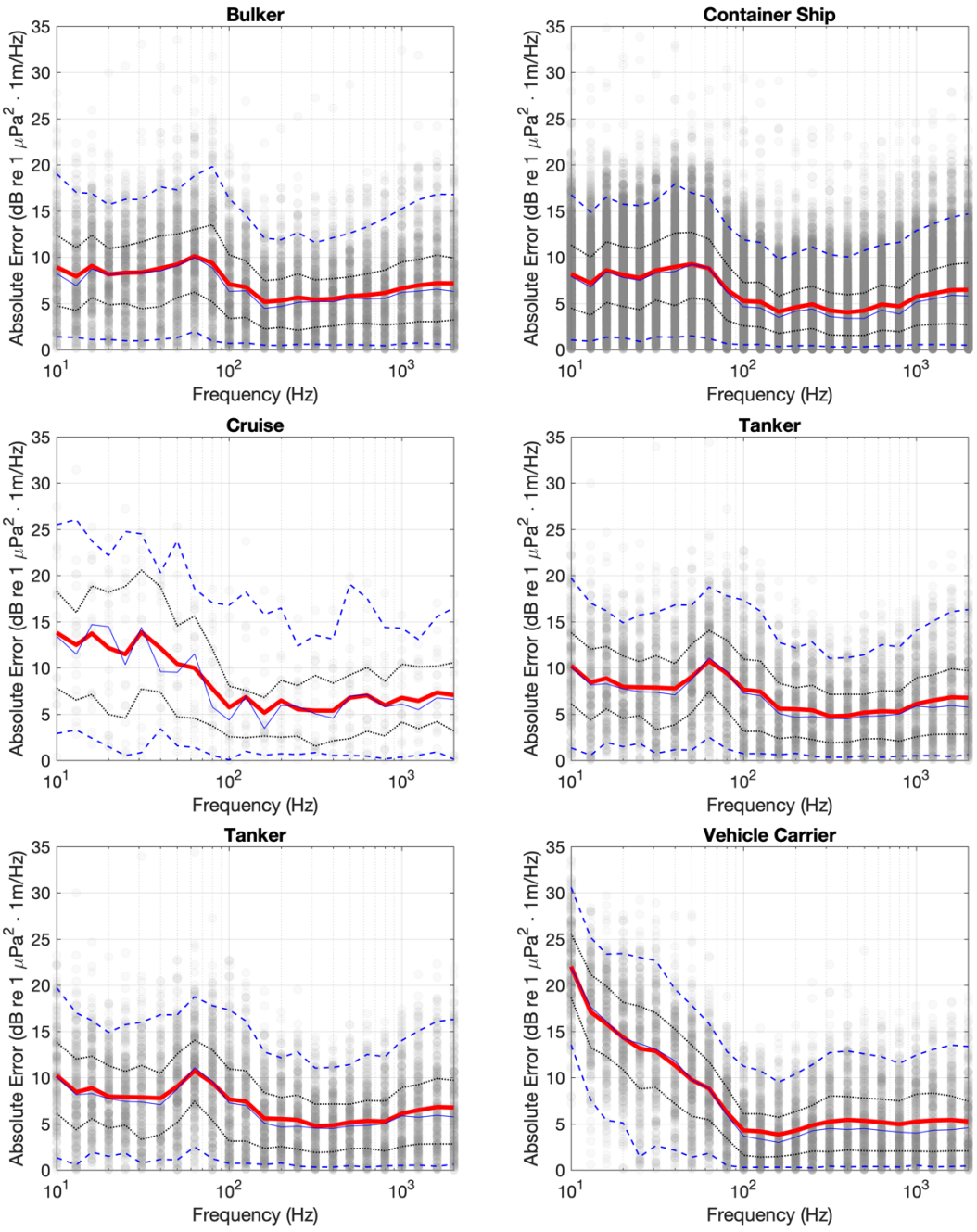

Figure 18. Absolute error as a function of frequency. Each point represents one observed difference between the model prediction and associated observation. Points are semi-transparent so that the relative density of points can be interpreted, with darker values indicating many overlapping observations and light values indicating sparse observations. Lines indicate the mean (solid red), median (solid blue), 25th and 75th percentiles (dotted black), and 5th and 95th percentiles (dashed blue). Offset between the mean and median is an indication of deviance from a symmetric distribution.

Over-prediction of RNLs increased with greater draft for bulkers and vehicle carriers, but decreased for tugs (Figures 19, 20 and 21). Over-prediction occurred at the lowest surface angles for bulkers and tankers, particularly in the low frequencies, however this was not observed for other vessel categories. Over-prediction increased with greater speed for bulkers, container ships, tankers, and vehicle carriers. Lower wind resistance led to over-prediction for bulkers, however the reverse was observed for container ships and vehicle carriers, which overestimated RNL with high wind. For the most part, vessel design characteristics had few consistent effects on the slopes of the residuals. A slight over-prediction appears to occur for young container ships, tankers and vehicle carriers, possibly reflecting improved efficiency in recent designs (ZoBell et al. submitted).

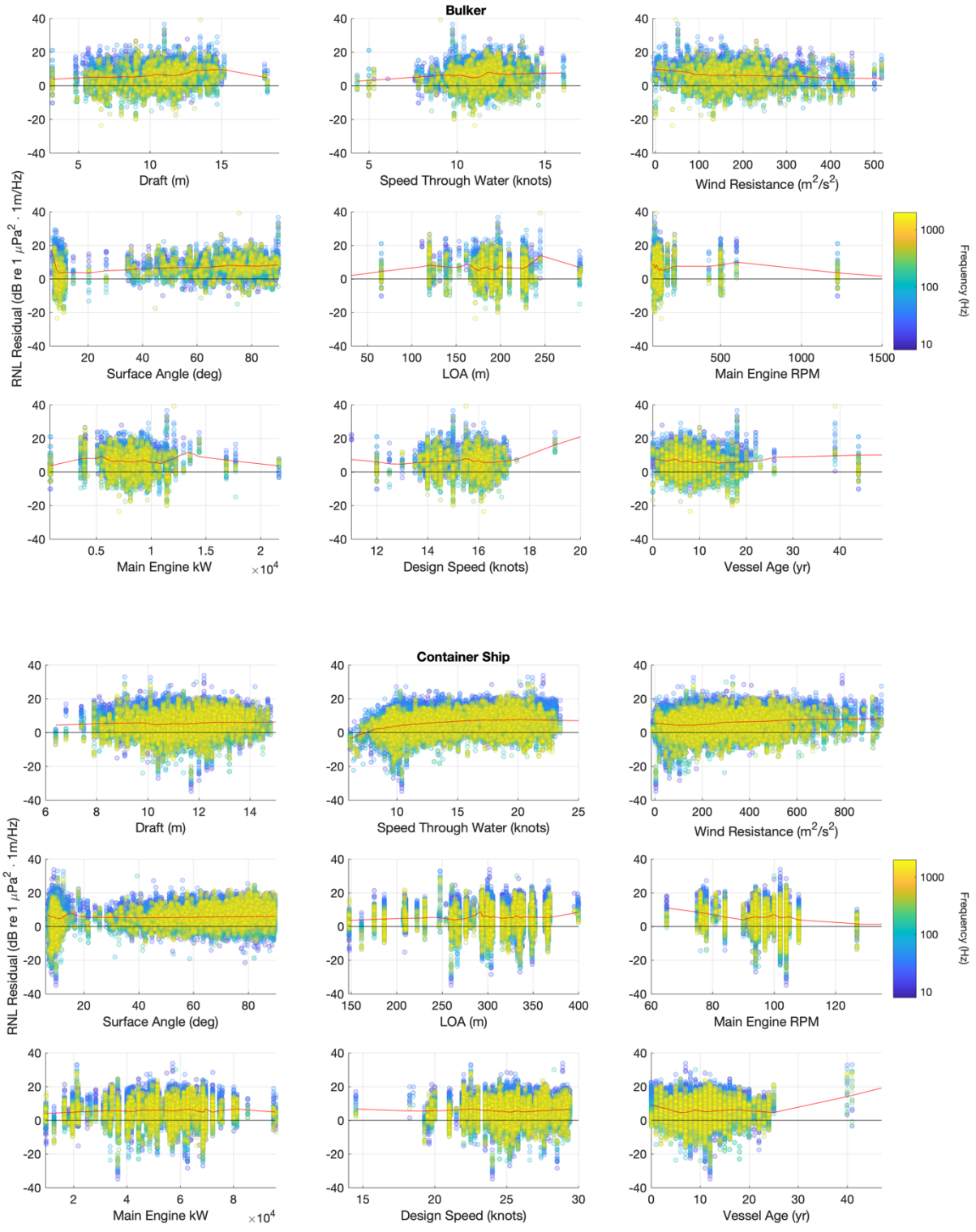

Figure 19. Distribution of MSL residuals as a function of nine model covariates for bulkers (top) and container ships (bottom). Each point represents a residual value for one transit, within a single frequency band, with color representing frequency. The black line denotes a residual value of zero, and the red line represents a smooth of the mean taken across all frequencies.

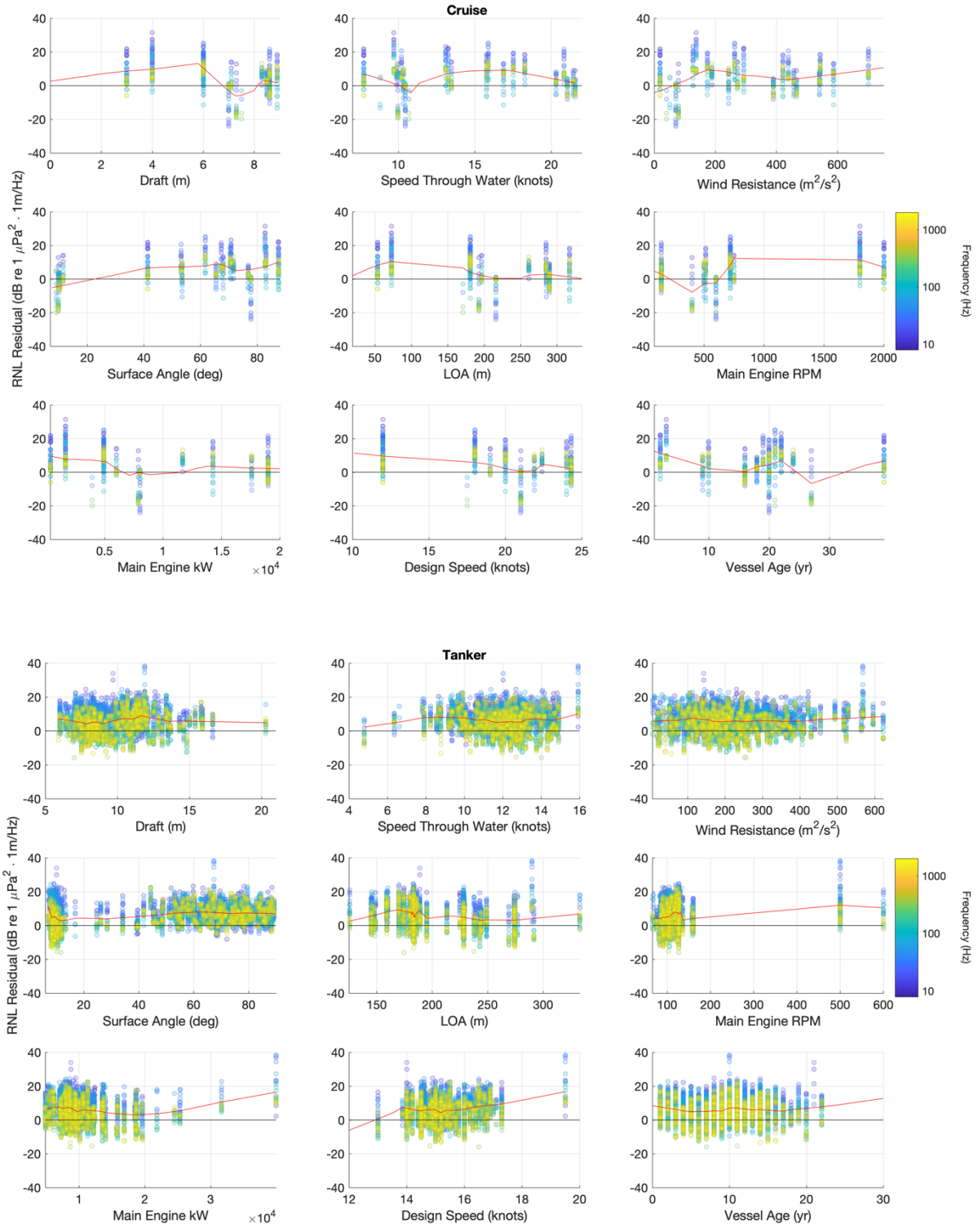

Figure 20. Distribution of MSL residuals as a function of nine model covariates for cruise ships (top) and tankers (bottom). Each point represents a residual value for one transit, within a single frequency band, with color representing frequency. The black line denotes a residual value of zero, and the red line represents a smooth of the mean taken across all frequencies.

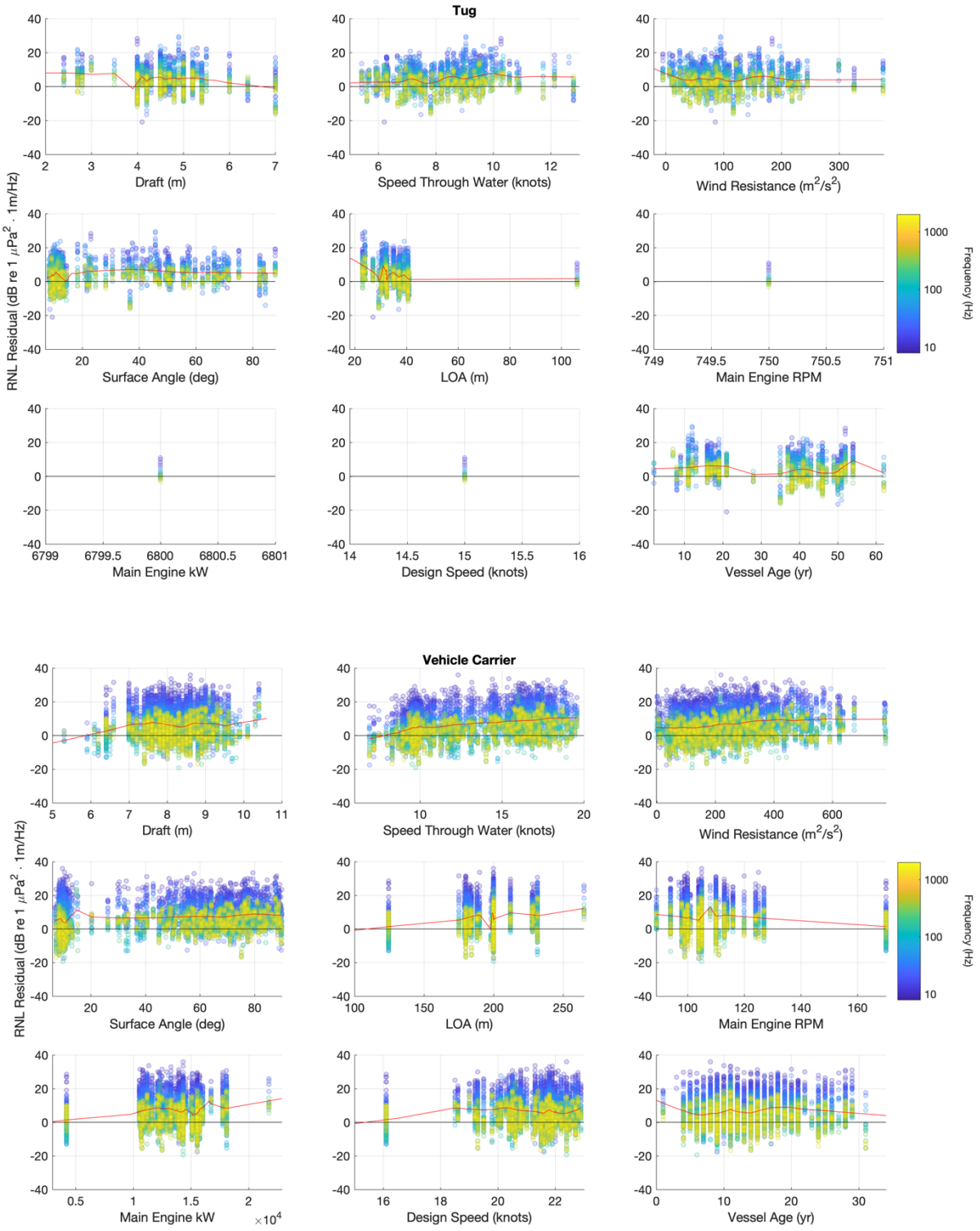

Figure 21. Distribution of MSL residuals as a function of nine model covariates for tugs (top) and vehicle carriers (bottom). Each point represents a residual value for one transit, within a single frequency band, with color representing frequency. The black line denotes a residual value of zero, and the red line represents a smooth of the mean taken across all frequencies.

January 31, 2022

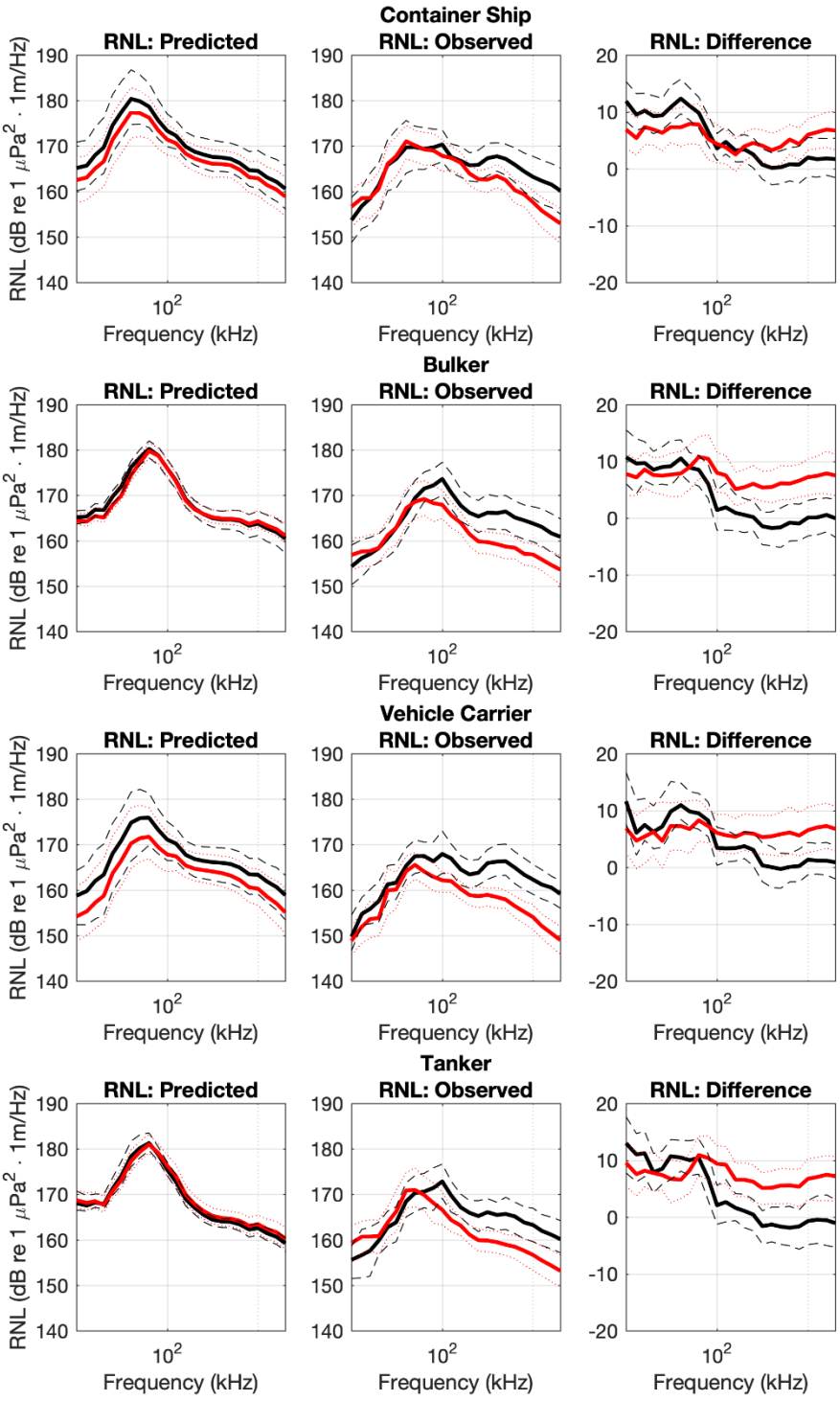

Figure 22. Comparison of RNL model predictions and residuals for the four main vessel categories across the SBC HARP sites, B1 (black) and B2 (red). Means of site B1 transits (solid black line) and B2 transits (solid red line) are shown, with 25th and 75th percentiles as dashed and dotted lines respectively

Average RNL model predictions were similar for B1 and B2 transits, and measured RNL was similar on average below 100 Hz for B1 and B2 transits (Figure 22). Therefore, differences between model predictions and observations below 100 Hz for RNL were not site dependent on average. Above 100 Hz, measured RNL was 5-10 dB higher on average for B2 transits than B1 transits, and model predictions were closer on average to B1 observations.

Overall model accuracy and precision can be considered in multiple ways. Mean signed deviation (MSD, Table 3) indicates the overall accuracy across many predictions. Using this metric, the MSL model predicts the category-level means with an accuracy between 0 to 2 dB on average as estimated by the mean of MSD within each vessel category. Standard deviations of MSD ranged from 4 to 9 dB across categories. The highest MSL model accuracy was achieved for the container ship, cruise ship and vehicle carrier classes, however, the cruise ship class also had the greatest variability. MSD was higher for RNL predictions as illustrated above, with values ranging from 4 to 6.5 dB on average, but variability was similar to that of the MSL predictions. The most accurate RNL models were those for cruise ships and tugs, which were the smallest classes. Frequency bins below 100 Hz contributed more strongly to MSD than those greater than or equal to 100 Hz (Table 4) for both MSL and RNL models.

Mean Absolute Error (MAE), provides insight on model accuracy from a transit-by-transit perspective (Table 5). The mean and standard deviation of MAE in each category is displayed in Table 4. Using MAE, MSL models achieved accuracy between 3.14 and 7.59 dB, with highest accuracy for the container ship, tanker, and vehicle carrier classes. Accuracy was highest above 100 Hz, with a low of 3.28 dB for container ships and tugs, and a high of 5.04 dB for cruise ships. Although the cruise ship category appears accurate across many transits based on MSD, the high variability of those predictions on the transit-by-transit level is reflected in a high MAE value. MAE was higher for the RNL model with values ranging from 7.39 to 10.28 dB. Frequency bins below 100 Hz contributed more strongly to MAE than those greater than or equal to 100 Hz (Table 6) for both MSL and RNL models. MAE was particularly high for RNL in the vehicle carrier class, and further investigation is needed to understand the cause, however it may be related to slow transit speeds.

Table 3. Mean ± standard deviations of mean signed deviation (MSD) computed across all frequency bands.

Vessel Type	MSL MSD (dB re 1 µPa^2 @ 1m)	RNL MSD (dB re 1 µPa^2 • 1m)
Bulker	2.21 ± 5.50	6.39 ± 5.34
Container	-0.29 ± 4.31	5.51 ± 4.36
Cruise	0.73 ± 9.08	4.31 ± 7.60
Tanker	2.16 ± 4.94	6.11 ± 4.47
Tug	1.72 ± 5.61	4.44 ± 4.46
Vehicle Carrier	0.99 ± 5.30	6.53 ± 5.77

Table 4. Mean ± standard deviations of mean signed deviation (MSD) computed separately for frequencies above and below 100 Hz.

Vessel Type	MSL MSD (dB re 1 µPa^2 @ 1m)		RNL MSD (dB re 1 µPa^2 • 1m)	
	< 100Hz	>= 100Hz	< 100Hz	>= 100Hz
Bulker	3.61 ± 6.23	1.37 ± 5.60	9.13 ± 4.99	4.47 ± 5.80
Container	0.06 ± 4.86	-0.46 ± 4.28	8.17 ± 4.66	3.77 ± 4.45
Cruise	3.58 ± 13.24	-1.23 ± 6.83	7.62 ± 9.62	2.39 ± 6.91
Tanker	3.84 ± 5.56	1.06 ± 5.12	8.95 ± 4.22	4.29 ± 5.33
Tug	4.32 ± 8.85	0.22 ± 4.20	8.75 ±5.81	2.42 ± 3.99
Vehicle Carrier	1.71 ± 6.36	0.75 ± 4.69	12.40 ± 5.43	2.56 ± 4.97

Table 5. Mean ± standard deviations of mean absolute error (MAE) computed across all frequency bands.

Vessel Type	MSL MAE (dB re 1 µPa^2 @ 1m)	RNL MAE (dB re 1 µPa^2 • 1m)
Bulker	4.60 ± 3.74	6.79 ± 4.81
Container	3.14 ± 2.96	5.95 ± 3.75
Cruise	7.54 ± 4.85	7.55 ± 4.19
Tanker	4.39 ± 3.12	6.44 ± 4.19
Tug	4.76 ± 3.41	5.12 ± 3.66
Vehicle Carrier	4.10 ± 3.50	7.08 ± 5.08

Table 6. Mean ± standard deviations of per transit mean absolute error (MAE) computed separately for frequencies above and below 100 Hz.

Vessel Type	**MSL MAE (dB re 1 µPa^2 @ 1m)**		**RNL MAE (dB re 1 µPa^2 • 1m)**	
	< 100Hz	**>= 100Hz**	**< 100Hz**	**>= 100Hz**
Bulker	5.62 ± 4.50	4.45 ± 3.65	9.19 ± 4.88	5.82 ± 4.40
Container	3.52 ± 3.36	3.28 ± 2.78	9.43 ± 4.18	4.68 ± 3.48
Cruise	11.56 ± 7.00	5.04 ± 4.65	10.12 ± 6.80	5.93 ± 4.12
Tanker	5.48 ± 3.39	4.22 ± 3.07	9.00 ± 4.12	5.59 ± 3.94
Tug	7.93 ± 5.81	3.28 ± 2.61	9.32 ±4.82	3.71 ± 2.82
Vehicle Carrier	5.02 ± 4.25	3.68 ± 3.68	12.47 ± 5.27	4.33 ± 4.34

Discussion

Both the MSL and RNL ECHO models over-estimated amplitudes below 100 Hz for the majority of vessel types in the SBC dataset, with the notable exception of container ship MSLs, the largest class. The over-estimation may be related to site-specific differences such as bathymetry and bottom composition, systematic differences in design characteristics of vessels between sites, or environmental differences such as wind resistances. It is notable that mean MSL and RNL observations differed between sites B1 and B2, and that model predictions generally showed better agreement with site B1 measurements. A subset of low amplitude measurements was excluded following the data conditioning process under the site B1 conditions, but met threshold criteria in the site B2 conditions. These low amplitude examples, possibly related to slow speed transits, may contribute strongly to the differences between model predictions and observations, and could partially explain the improved agreement with B1 measurements.

It is also possible that interaction with a muddy seafloor in the Santa Barbara channel is attenuating low frequencies more than the seafloor in the original environments in which the ECHO datasets were recorded. If not properly accounted for in propagation loss models this would result in an underestimate of MSL and RNL relative to the model predictions. The methods used to compute MSL and RNL in this and the original study are approximations rather than fully-implemented propagation loss models, and may not account for all of the differences between the recording locations and datasets.

An alternative explanation is that the JASCO model overestimated sound attenuation occurring in the original ECHO dataset, below 100 Hz, and therefore overestimated MSL and RNL in the original dataset. This could happen if the propagation loss model used to compute MSL in the development dataset overestimated attenuation due to bottom interactions in the original shallow water recording environment. Effort to determine why the container ship MSLs were more accurately estimated than other categories might also help identify the underlying cause of overestimates in other classes. The container ship class was the largest, with more high speed transits than other classes, and fewer differences between sites, which may partially explain the improved model performance on this class. However, the RNL model predictions for container ships were comparable with those of other vessel categories, which seems to indicate that the difference does not stem solely from the covariates, which were the same for both MSL and RNL models. A sensitivity analysis could be done to further investigate this. Implementation of a more sophisticated propagation model could help clarify whether systematic differences between the SBC and ECHO datasets are linked to oceanographic differences between recording locations.

Another potential source of error is the assumption that the source depth is equivalent to 50% of the vessel draft. Source depth is a highly influential parameter, and discrepancies between the true and estimated values, or systematic differences between the two datasets (for instance differences in vessel load) may translate into large differences in source level estimates. Verified draft (such as from logbooks), and known propeller diameters for each vessel would be necessary to reduce inaccuracies associated with this assumption (Gassmann et al 2017, ZoBell et al., submitted).

It should be noted that hydrophone calibration becomes increasingly difficult at low frequencies. The sensitivity of the Scripps hydrophone design below 20 Hz is a blend of theoretical estimates and lab measurements. Efforts are underway to improve measurement accuracy through *in-situ* calibrations. A difference in sensor sensitivity, or calibration accuracy could partially explain some of the differences observed at the lowest frequencies (<= 20Hz) in this comparative study.

The models predict well on average above 100 Hz. Across many transits, mean predicted MSL for all vessel types was within +/- 1.7 dB of the mean observed MSL. Mean predicted RNL was higher than observed above 100 Hz for all categories but fell within 0-5 dB of observed means.

Transit-by-transit model predictions were approximately 5 dB different from observed values on average. Error and uncertainty in the MSL models were not strongly frequency dependent for the vessel classes, with the exception of the cruise ship and tug classes. Error was higher below 100 Hz for the RNL model, but uncertainty was relatively constant across the frequency band. Future work could compare MSL predictions for repeat transits of the same vessel.

This analysis clearly demonstrates the limitations of RNL as a metric. The simplified propagation model assumed for RNL calculations does not account for the numerous differences between the physical and operation conditions underlying the ECHO and SBC datasets. Most notably, RNL does not account for the effect of source depth. Signal amplification can occur at low frequencies as a function of draft. Although draft is a predictor in the noise functional regression model, that model does not appear to have been able to account for its complex effects on RNL. It appears that MSL, which includes a more robust preprocessing step to account for propagation effects, is a more predictable metric with fewer site-specific characteristics.

Conclusion

In general, the models were capable of accurately predicting overall means and trends, but may lack the detailed information required to predict precisely on a transit-by-transit basis. The MSL model for container ships was the most accurate on average, followed by cruise ships, vehicle carriers, tankers and bulkers. Slow speed transits in the SBC dataset fell below the observed range of speeds through water measurements in the ECHO dataset, and likely led to over-prediction of MSL and RNL in those cases. Model predictions could likely be improved for these cases by adding slow speed examples in the model development dataset. RNL models were not able to account for the complex acoustic effects of site-specific differences between the ECHO recording station and the SBC sites.

In order to better understand the variations seen in the predicted versus observed vessel noise levels estimates, additional analyses would be required. Sensitivity to assumed source depth, surface angle and closest point of approach should be investigated to identify whether the variations seen are due to effects of bottom interaction and Lloyd's mirror. Further investigation into the effect of assumed source depth on predictions, and the influence of wind resistance, including the bias toward positive resistance in the SBC dataset is recommended. Additionally, investigating differences in repeat transits of individual vessels may elucidate more information on the variability seen within MSL predictions and observations.

References

Gassmann, M., Wiggins, S. M. & Hildebrand, J. A. Deep-water measurements of container ship radiated noise signatures and directionality. J. Acoust. Soc. Am. 142, 1563–1574 (2017).

International Organization for Standardization (ISO). Underwater Acoustics—Quantities and Procedures for Description and Measurement of Underwater Sound from Ships—Part 1: Requirements for Precision Measurements in Deep Water Used for Comparison Purposes (2016).

International Organization for Standardization (ISO). Underwater Acoustics—Quantities and Procedures for Description and Measurement of Underwater Sound from Ships—Part 2: Determination of source levels from deep water measurements (2019).

MacGillivray, A.O., J. Zhao, M.A. Bahtiarian, J.N. Dolman, J.E. Quijano, H. Frouin-Mouy, and L. Ainsworth. 2020. ECHO Vessel Noise Correlations Phase 2 Study: Final Report. Document 02283, Version 1.0. Technical report by JASCO Applied Sciences, ERM Consultants Canada, and Acentech for Vancouver Fraser Port Authority ECHO Program.

Rudnick, D. The California underwater glider network. IDG Scripps Inst. Oceanogr., La Jolla, Calif. 10, S8SPRAY1618 (2016).

ZoBell, V. M., Frasier, K. E., Morten, J. A., Hastings, S. P., Peavey Reeves, L. E., Wiggins, S. M., & Hildebrand, J. A. (2021). Underwater noise mitigation in the Santa Barbara Channel through incentive-based vessel speed reduction. *Scientific reports*, *11*(1), 1-12.

ZoBell, V.M., Frasier, K.E., Gassmann, G., Kindberg L.B., Wiggins S.M, & Hildebrand J.A. (Submitted). *Retrofit-induced Changes in the Underwater Source Levels of Container Ships*. J. Acoust. Soc. Am.