



## Data-driven models for vessel motion prediction and the benefits of physics-based information

Matthew L. Schirrmann, Matthew D. Collette\*, James W. Gose

*University of Michigan, 2600 Draper Drive, Ann Arbor, MI 48109, USA*

### ARTICLE INFO

**Keywords:**

Data-driven models  
Digital twin  
Machine learning  
Motion forecasting  
Neural networks  
Physics-based models

### ABSTRACT

Machine learning approaches, onboard measurements, and widely available wave forecast and hindcast data present an opportunity to develop predictive models for vessel motion forecasting. Detailed vessel motion forecasts would support underway and deployment decisions for safer and more efficient vessel operation. To demonstrate this application, ridge regression and neural network models for heave, pitch, and roll prediction were trained and tested using time-and-place specific, multidirectional wave model parameters as input. Additionally, the performance benefits of providing these predictive models with computationally efficient, physics-based model predictions (PBMPs) of heave, pitch, and roll as additional inputs were examined. Data from approximately 13,500 30-minute windows, measured aboard an operational research vessel, were used to train and test the data-driven models. Data from over 2,500 additional 30-minute windows, measured aboard a sister vessel, were also used to test the versatility of the trained models. The results of this study showed effective reduction of motion amplitude mean-squared error (MSE) values on multiple test datasets relative to the PBMPs alone. The results also showed that inclusion of PBMPs as input to the data-driven models was typically beneficial in terms of MSE reduction, stressing the importance of retaining physics-based information in data-driven models.

### 1. Introduction

Data-driven models and their embedded machine learning (ML) algorithms have proven extremely powerful for learning complex relationships between variables for regression and classification tasks. These ML approaches provide engineers and scientists with opportunities to improve upon standard analyses and practices in a variety of fields, including the marine industry. In the context of marine applications, underway heave, pitch, and roll measurements, often recorded for use by other instruments critical to standard vessel operation, are accessible for many ships. Furthermore, time-and-place specific wave forecast and hindcast data are now widely available, characterizing a vessel's operational environment throughout its life-cycle. In conjunction with speed, heading, and location information, these underway motion measurements and wave model data hold a wealth of knowledge that can be used to train data-driven models for motion forecasting, which is the focus of this work. Reliable motion forecasts, instead of wave forecasts alone, would provide owners and operators with valuable information to support deployment and underway decisions. For example, heave, pitch, and roll motion statistics could be provided as a function of route, vessel heading, or departure

time to aid in decision making. This work demonstrates the application and potential benefits of data-driven models for vessel motion prediction using data from over 16,000 30-minute windows measured aboard two operational sister ships. Schirrmann et al. (2020) presented preliminary results from this study, which only considered data from one of the two vessels, a total of approximately 13,400 30-minute windows. Even given a large pool of data, rather than abandon traditional, computationally efficient seakeeping analyses entirely, this work investigates the benefits of incorporating physics-based information into data-driven models for improved forecasting accuracy.

A number of studies have been performed that employed ML algorithms for vessel motion prediction. Although many of these implementations did not incorporate physics-based information, their successes demonstrate that data-driven models have merit for motion prediction applications. A large portion of these studies used different forms of neural networks (NNs) (e.g., fully-connected, recurrent, convolutional). For example, Bremer (2018) tested linear regression and neural networks for sway, heave, and roll motion prediction in regular wave, beam-seas, model-scale experiments. Using just wave period and height as input variables, the results highlighted the ability of neural networks to accurately predict vessel motions and model highly nonlinear

\* Corresponding author.

E-mail addresses: [mschirma@umich.edu](mailto:mschirma@umich.edu) (M.L. Schirrmann), [mdcoll@umich.edu](mailto:mdcoll@umich.edu) (M.D. Collette), [jgose@umich.edu](mailto:jgose@umich.edu) (J.W. Gose).

phenomena (such as roll), a distinct advantage over linear regression. Using roll angle and roll velocity as input, Xing and McCue (2010) demonstrated the use of neural networks to fit the parameters of two different nonlinear models for roll motion, an example that blends physics-based knowledge with machine learning. Li et al. (2017) demonstrated the use of nonlinear autoregressive exogenous (NARX) networks for predicting pitch and roll angle time series, as well as several other ship motion characteristics. Experiments were performed using online, offline, and hybrid (combined online and offline) learning approaches, and results were discussed for single- and multi-timestep ahead prediction. A discussion of trade-offs between neural networks and other machine learning approaches was also included, which ultimately selected neural networks for vessel motion prediction. Khan et al. (2005) demonstrated the efficacy of NNs for accurate prediction of roll angles up to 10 s in the future for ship system deployment. De Masi et al. (2011) demonstrated radial basis function NNs (RBFNNs) for vessel heave time series predictions using prior heave time series data as input. Similarly, Huang and Zou (2016) compared RBFNNs with two other types of NNs for short-term pitch prediction. Deng et al. (2020) successfully demonstrated the use of long short-term memory (LSTM) networks, a form of recurrent neural network, to predict semi-submersible platform heave, pitch, and roll time-series due to irregular waves from a single direction, where the model inputs were wave elevations. Comparisons were made to motion predictions generated using experimentally determined and physics-based numerical model response amplitude operators (RAOs), which showed that the neural network approach outperformed both RAO-based methods. Although the implementations differ significantly, the present work seeks to build on this idea of data-driven and physics-based model comparisons through injection of RAO-based model outputs into the data-driven models. Unlike many of the implementations above, rather than using previous motion time series as input to predict future motions, the models developed in this study employed multidirectional wave conditions to predict vessel motions statistics.

Numerous NN examples exist in the marine context beyond motion predictions, showing their prominence in the field. For example, Jiang et al. (2020) employed an NN within a submarine maneuvering model, and Najafi et al. (2018) used NNs to predict hydrofoil supported catamaran resistance, effective power, trim, and sinkage based on model-scale experiment data. NNs have also been applied to analyze the operational wave environment using ship motions. For example, Duz et al. (2019) utilized an NN that combined convolution layers with an LSTM network to predict wave parameters using ship motion time series. Kawai et al. (2021) used ship motion and structural response spectra as input to a convolutional NN that predicted wave parameters. NNs have also been applied in the context of oceanography for efficient wave modeling. For example, Wu et al. (2020) trained an NN using the environmental variables and boundary conditions typically used as input for computationally expensive wave models to efficiently generate multi-step wave forecasts. For more examples of NNs in marine applications, please see Gouguolidis (2008).

One common criticism of data-driven models, which this work seeks to combat, is that many implementations abandon the well-understood physics on which traditional methods are based; the models learn from experiences included in the training dataset, but exclude physics-based information that may be useful when a model is applied to novel data. Furthermore, many ML approaches, especially NNs, are often perceived as black boxes whose inner workings are difficult to interpret. These approaches are designed to reduce uncertainty and error between data-driven model predictions and real-world behavior, but the origins of these corrected errors are often untraceable through the model. For example, it can be difficult or impossible to quantify the uncertainty associated with input variables versus the uncertainty due to model form errors, or inadequacies in the underlying prediction model as described by Subramanian and Mahadevan (2019). As such, the combination of physics-based information and machine learning

has garnered significant attention across engineering disciplines in recent years (e.g., Willcox (2019), Raissi et al. (2019), Kapuzoglu and Mahadevan (2020)).

The potential deficiencies of data-driven models mentioned above can be combated through incorporation of physics-based information. For example, Weymouth and Yue (2013) discussed a framework for physics-based learning models and demonstrated their application to several hydrodynamic modeling challenges. These challenges included extrapolation of sparse experimental response amplitude operator (RAO) data in head seas to Froude numbers unseen in the training data. They also demonstrated that coupling numerical models with measured realizations of a system's state presents an opportunity to use more traditional, computationally efficient numerical models rather than time intensive processes (e.g., potential flow codes vs. CFD simulations) to yield similar accuracy. Similarly, a small number of high-fidelity realizations could be used to train these physics-based learning models rather than a large number of computationally expensive runs, which was demonstrated in the context of bow wave behavior prediction. Weymouth (2019) further described this method and demonstrated incorporation of a physical basis into Ridge Regression (RR) models for prediction of roll damping coefficients.

Schirrmann et al. (2019), a predecessor to this work that incorporated physics-based information, employed data-driven models for heave and pitch prediction. The study considered a small dataset (183 three-hour windows) from a single cruise of the *R/V Knorr*, an 85 m research vessel operated by Woods Hole Oceanographic Institution (WHOI). Linear least-squares (LS) and NN regression models were trained using full-scale measurements to predict heave and pitch amplitudes using six input variables: speed, three unidirectional wave model parameters (significant wave height, peak period, and mean direction relative to vessel heading), and predicted heave and pitch amplitudes generated using 2D strip-theory RAOs. These computationally efficient heave and pitch predictions, referred to as physics-based model predictions (PBMPs) moving forward, were primarily included due to the limited amount of training data available for characterization of a relatively large input space. Using only these six input variables, both the LS and NN models significantly reduced the mean-squared error (MSE) of a test dataset relative to the MSE of the PBMPs alone. While the success of these simplistic models showed promise for more intensive data-driven approaches and therefore catalyzed this work, it also prompted a critical question on the importance of physics-based information: how beneficial would PBMPs be given an abundance of training data?

To address this question and demonstrate the application of data-driven models for ship motion prediction in Schirrmann et al. (2020) and this work, linear ridge regression (RR) and (nonlinear) NN models were constructed, trained, and tested both with and without PBMPs included as input variables. The other, consistently included input variables were the vessel's speed and multidirectional wave data parameters, with wave directions adjusted relative to the ship's heading. The performances of the RR and NN models with and without PBMPs were then compared on two test datasets. One of these test datasets was taken from the same vessel used in model training (with preliminary analyses presented in Schirrmann et al. (2020)), and the other was taken from a sister vessel. The application of these trained models to a sister ship was performed to evaluate the versatility of these data-driven models. More explicitly, if data-driven models perform well between similar ships, they may prove useful for fleet management, and multiple ships could be used to accelerate data collection for model training. Although both ships were built to the same specifications, their outfitting differs, and thus, loading conditions may differ. Furthermore, given the difference between home ports (Atlantic Ocean versus Pacific Ocean), operating profiles varied between the two ships, further testing the adaptability of the trained models to unique conditions.

As mentioned previously, reliable data-driven models could be used by vessel owners and operators to support operational guidance and

deployment decisions through motion forecasting. While most of the data-driven model training and testing discussed here was performed using wave hindcast data due to its continued availability, a small test dataset of wave forecast data was also stored for additional demonstration of the most-successful models. This final experiment demonstrated real-world application of these models for vessel motion forecasting at different time horizons.

Compared to the other vessel motion prediction NN studies mentioned previously, the novelty of this work lies in its large scope (two sister vessels and over 16,000 real-world, 30-minute time windows), the use of time-and-place specific multidirectional wave model data, and the direct incorporation of physics-based information and investigation of its benefits.

To summarize, this study's purpose was to evaluate ship motion forecasting using data-driven models through realization of the following goals:

1. Demonstrate the use of linear ridge regression (RR) and nonlinear neural network (NN) approaches for vessel motion prediction using hindcast multidirectional wave model data.
2. Investigate the potential benefits of incorporating physics-based model predictions (PBMPs) as input to these data-driven models.
3. Evaluate model robustness through comparison of predicted responses using forecast and hindcast wave model data.
4. Test the versatility of the data-driven models through application to a sister ship of the original vessel.

Data from approximately 13,500 30-minute windows, resulting from 50 cruises of the *R/V Neil Armstrong*, were considered for training and testing these ML models for motion prediction, which addressed goals 1, 2, and 3. Additionally, another test dataset generated from over 2,500 30-minute windows, resulting from 18 different cruises of the *R/V Sally Ride*, were used to address the final goal and further support goals 1 and 2. Results and preliminary analyses addressing goals 1 and 2 that only used data from the *R/V Neil Armstrong* (approximately 13,400 30-minute windows) were presented in Schirrmann et al. (2020).

## 2. Methodology

This section details the different datasets considered from the sister vessels *R/V Neil Armstrong* and *R/V Sally Ride*, as well as the construction, training, and testing of the linear ridge regression (RR) and nonlinear neural network models (NN) models with and without incorporation of physics-based model predictions (PBMPs) as input. Section 2.1 gives an overview of the two ships, and Section 2.2 details the data collected aboard each vessel and the data processing performed. Sections 2.3 and 2.4 discuss the different datasets considered from the *Neil Armstrong* and *Sally Ride*, respectively. Section 2.5 describes the wave data and their use as input variables to the data-driven models, and Section 2.6 outlines the frequency-domain approach employed to generate PBMPs using these wave data. Section 2.7 discusses the data-driven models' output variables and data normalization, and Sections 2.8, 2.9, and 2.10 detail model construction and training. Finally, Section 2.11 briefly compares the different datasets considered in this work.

Fig. 1 gives an overview of the data processing approach described in this section. Time series data measured aboard the ship included location, time, heading, speed, and motions (heave, pitch, and roll). Data used in model training and testing were processed in 30-minute windows, which were selected such that each window had approximately stationary speed and heading as detailed in Section 2.2.3. As shown in Fig. 1, statistics from each 30-minute window were then used to select time-and-place specific wave data and generate PBMPs. The data types used as model input and the measured motion statistics used in model training and evaluation are also indicated.

**Table 1**

Design characteristics of the sister ships *R/V Neil Armstrong* and *R/V Sally Ride*. All values are molded and based on the hull lines plan created by Guido Perla & Associates and provided by WHOI.

| Design characteristics |             |
|------------------------|-------------|
| Length overall         | 72.54 m     |
| Waterline length       | 70.10 m     |
| Beam                   | 15.24 m     |
| Depth                  | 6.71 m      |
| Draft                  | 4.57 m      |
| Displacement           | 3,043.46 LT |
| LCG (Aft of FP)        | 35.38 m     |
| Block coefficient      | 0.624       |

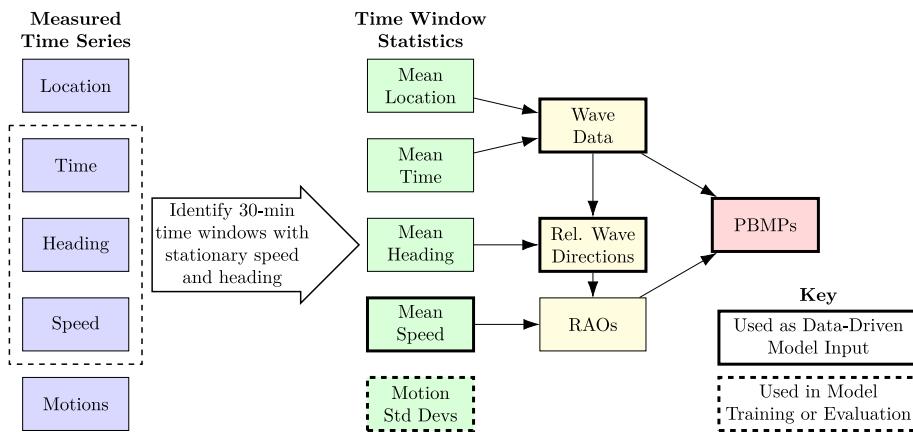
### 2.1. Sister vessels' overview

The *R/V Neil Armstrong* (AGOR-27) was the primary vessel for this study, which is operated by Woods Hole Oceanographic Institution (WHOI) in Woods Hole, Falmouth, Massachusetts. Vessel motions data from numerous cruises of the *Neil Armstrong* and corresponding wave model data were separated into two datasets to train and test the prediction models, respectively. All data from the *Neil Armstrong*'s sister vessel, the *R/V Sally Ride* (AGOR-28), were used for additional testing of these models. Much of the following discussion follows that of Schirrmann et al. (2020), which only considered data from the *Neil Armstrong*. The *Sally Ride* is operated by the Scripps Institution of Oceanography (SIO) at the University of California, San Diego in San Diego, California. Both ships are owned by the Office of Naval Research (ONR) and were constructed at Dakota Creek Industries in Anacortes, Washington (US Navy Office of Information, 2018). The *Neil Armstrong* was delivered in September 2015 and the *Sally Ride* was delivered in July 2016. The two ships are pictured in Fig. 2. The vessels' design speed is 12 knots, and select design characteristics of the vessels are given in Table 1. The values in Table 1 are based on the hull lines plan provided by WHOI, which was created by the vessels' designers at Guido Perla & Associates. This hull lines plan was used for all hydrodynamic modeling performed. A detailed discussion of this hydrodynamic modeling approach is included in Section 2.6. Note that while a theoretical longitudinal center of gravity (LCG) was known from the design specifications, a vertical center of gravity (VCG) was assumed at a value that yielded acceptable transverse stability. The specifics of this VCG selection are revisited in the following discussion. Also consider that vessel loading conditions undoubtedly varied between ships and as a function of time, which would affect both the LCG and VCG; however, these values were not tracked for consideration in this study.

The 50 cruises of the *Neil Armstrong* used to train and test the data-driven models occurred during four calendar years from 2016 through 2019. The additional 18 cruises of the *Sally Ride* used for further model testing occurred during the years 2017 through 2019. Note that these were not the vessels' only voyages during these years, but rather the cruises with available motions data that had appropriate and usable data for model training and testing. As such, some of the considered cruises yielded a small number of usable time windows relative to their duration. The specifications necessary for motions data to be considered appropriate are discussed in Section 2.2.3.

### 2.2. Onboard measurements

Both the *Neil Armstrong* and *Sally Ride* were outfitted with motion measurement instruments that are detailed in the following subsections. The output heave ( $j = 3$ ), pitch ( $j = 5$ ), and roll ( $j = 4$ ) time series recorded by these instruments were processed in 30-minute windows as detailed in Section 2.2.3. The aforementioned values of  $j$  are commonly used motion component indices that are employed in the



**Fig. 1.** An overview of the data processing approach, as well as the data types used as model input variables and for model training and evaluation.



**Fig. 2.** Unmodified photo of WHOI's R/V *Neil Armstrong* (left) from Ken Kostel, WHOI ([https://commons.wikimedia.org/wiki/File:Armstrong\\_fromthefantail\\_406493.jpg](https://commons.wikimedia.org/wiki/File:Armstrong_fromthefantail_406493.jpg)) and U.S. Navy photo of the R/V *Sally Ride* (right) from John F. Williams ([https://commons.wikimedia.org/wiki/File:R-V\\_Sally\\_Ride\\_is\\_currently\\_underway\\_conducting\\_a\\_series\\_of\\_science\\_verification\\_cruises\\_\(31534416962\).jpg](https://commons.wikimedia.org/wiki/File:R-V_Sally_Ride_is_currently_underway_conducting_a_series_of_science_verification_cruises_(31534416962).jpg)).

**Table 2**  
Measured target variables for data-driven model training.

| Index | Variable      | Units | Description              |
|-------|---------------|-------|--------------------------|
| 0     | $\sigma_{3M}$ | m     | Measured heave std. dev. |
| 1     | $\sigma_{5M}$ | deg   | Measured pitch std. dev. |
| 2     | $\sigma_{4M}$ | deg   | Measured roll std. dev.  |

following discussion. These motions time series were also accompanied by status flags, or numerical values to indicate normal operation of the measurement instruments. The standard deviations in each motion degree of freedom (DOF) over the 30-minute windows were the representative motion amplitudes that the data-driven models were designed to predict. For reference, these standard deviations are therefore one half of the vessel's significant amplitude in a given DOF. The three measured target variables are summarized in [Table 2](#).

#### 2.2.1. R/V *Neil Armstrong* (AR) measurements

The data from each cruise of the *Neil Armstrong* were accessed through the WHOI Data Library and Archives at <http://dlacrisedata.whoi.edu/AR/>, where "AR" corresponds to the *Armstrong*. Each cruise was assigned a unique cruise identification tag, or cruise ID (e.g., AR29), which corresponds to a subfolder name within the *Armstrong*'s folder. Additional information about these cruises can be found at <https://www.whoi.edu/what-we-do/explore/cruise-planning/cruise-planning-before-the-cruise/cruise-planning-cruise-synopsis/>.

The heave, pitch, and roll motions of the vessel were recorded at approximately 1 Hz for all 50 cruises using the inertial measurement unit (IMU) of an *Applanix POS MV* system that was located in the transducer room and on the centerline of the vessel. General information about the vessel's equipment and specifications, including a diagram with the *POS MV*'s location, can be accessed at <https://www.whoi.edu/what-we-do/explore/ships/ships-neil-armstrong/>. The speed, heading, and location

(latitude and longitude) of the vessel were also recorded, processed, and provided by WHOI at one-minute intervals.

#### 2.2.2. R/V *Sally Ride* (SR) measurements

Data from the *Sally Ride*'s cruises were accessed through Rolling Deck to Repository (R2R) at <https://www.rvdata.us/search/vessel/Sally%20Ride>. As with the *Neil Armstrong*, each cruise was assigned a cruise ID (e.g., SR1906), which identified each cruise's datasets in R2R. R2R also provides supporting information for each cruise, such as the start/end dates and ports.

The motion reference unit (MRU) of a Kongsberg Seapath 330+ was used to record heave, pitch, and roll measurements at 5 Hz for all cruises considered in this work. General information and layouts for the *Sally Ride* can be accessed at <https://scripps.ucsd.edu/ships/sally-ride>. Similar to the *Neil Armstrong*, speed, heading, and location data were processed and provided via R2R at one-minute intervals.

#### 2.2.3. Measured data processing

One goal of this work and [Schirrmann et al. \(2020\)](#) was to improve upon the data processing approach of the *R/V Knorr* study in [Schirrmann et al. \(2019\)](#), thus increasing the number of time-windows usable for data-driven model training. For the analyses and frequency-domain motion prediction models employed in these studies, it was necessary to assume that the vessel's operational environment and motions were approximately statistically stationary processes during any usable time window. In [Schirrmann et al. \(2019\)](#), vessel motions statistics were calculated for three-hour periods that were centered on wave hindcast output times. After calculating motion statistics for each three-hour window, maximum allowable thresholds on the speed and heading standard deviations of the vessel during the three-hour windows were applied. The goal of this filtering approach was to keep only time intervals in which operating conditions were approximately constant for model training or testing. While this was deemed a reasonably effective

approach, the relatively large time window length and fixed window centroids resulted in many potentially valuable data being discarded. Furthermore, longer time windows are more vulnerable to changes in wave conditions due to variation in time or the ship's location, detracting from the stationary assumption's validity. A shorter window length would be less likely to involve a significant change in speed and heading that would make the window unusable for model training or testing. However, it was also important that the employed window duration was long enough to capture a sufficient number of motion cycles for calculation of heave, pitch, and roll statistics. Although not the primary focus of this work, a study regarding the effect of window size on model performance may be interesting in the future.

The present work reduced the window size to 30-minute intervals and modified the measured data windowing approach to ensure that the assumption of statistical stationarity was followed in good faith while increasing the number of available windows for data-driven model training and testing. The new windowing approach searched for 30-minute time periods in which the speed and heading of the vessel were approximately constant. Beginning with the first 30-minute window for a given cruise, the mean latitude, longitude, speed, and heading of the vessel were calculated. In addition to the requirement that motion data recording was uninterrupted and devoid of irregular status flags, to accept a 30-minute window for use, three requirements needed to be met: the mean latitude and longitude were a minimum of 10 NM from the nearest coast (as determined using the Matplotlib Basemap Toolkit in Python), the maximum absolute speed deviation from the mean within the 30-minute period was less than 1 knot, and the maximum absolute heading deviation from the mean was less than 15°. These speed and heading deviations were calculated using the preprocessed data provided by WHOI and SIO at one-minute intervals, so maximum deviation thresholds were more appropriate than the standard deviation thresholds employed in Schirrmann et al. (2019). If these three requirements were not met for a given 30-minute period, the 30-minute window was shifted one minute in the positive time direction and the three criteria were checked again. When a 30-minute period did meet all three criteria, and was therefore accepted, the search resumed for new windows beginning with the first minute following the accepted time window.

For all accepted 30-minute windows, the standard deviation of the corresponding heave, pitch, and roll time series were calculated, which match the desired target variables from Table 2. Therefore, in each DOF, this windowed time series was comprised of approximately 1,800, 1 Hz measurements for *Neil Armstrong* data, and 9,000, 5 Hz measurements for *Sally Ride* data. Additionally, the mean speed ( $U$ ) for each time window was stored and provided as an input variable to the data-driven models, as discussed in Section 2.5. The vessel's mean heading from each time window was also stored for calculation of wave directions relative to the ship heading.

### 2.3. R/V Neil Armstrong (AR) Train and Test datasets

Fig. 3 shows the 50 *Neil Armstrong* cruises in the Atlantic Ocean. Of the 50 *Neil Armstrong* cruises considered, the train dataset, AR Train, was constructed from the first 42, and the test dataset, AR Test, was constructed from 7 subsequent cruises. The final cruise, AR40, was used as a separate test dataset for comparison of data-driven model performances using wave hindcast and forecast data. This comparison is outlined in Section 2.5. The map of Fig. 3 identifies the dataset that each plotted route belongs to. Although the map reflects all location data available from each cruise, the amount of usable time windows from a given cruise often did not span the full route due to the requirements outlined in Section 2.2.3.

The AR Test and AR Train datasets are summarized in Table 3. In addition to the number of windows in each dataset, the median and maximum significant wave heights,  $H_{s50}$  and  $H_{s100}$ , and the median and maximum mean wave periods,  $T_{m50}$  and  $T_{m100}$ , of each dataset

are included for reference. The source of these wave parameters are discussed in Section 2.5. Table 3 also includes the northernmost and southernmost latitudes,  $\text{Lat}_N$  and  $\text{Lat}_S$ , and the easternmost and westernmost longitudes,  $\text{Lon}_E$  and  $\text{Lon}_W$ , of the usable windows in each dataset. These details are included for the 42 cruises of AR Train and the 7 cruises of AR Test individually in Tables A.10 and B.11 of the Appendix, respectively. Tables A.10 and B.11 also include the month that each cruise began.

### 2.4. R/V Sally Ride (SR) Test dataset

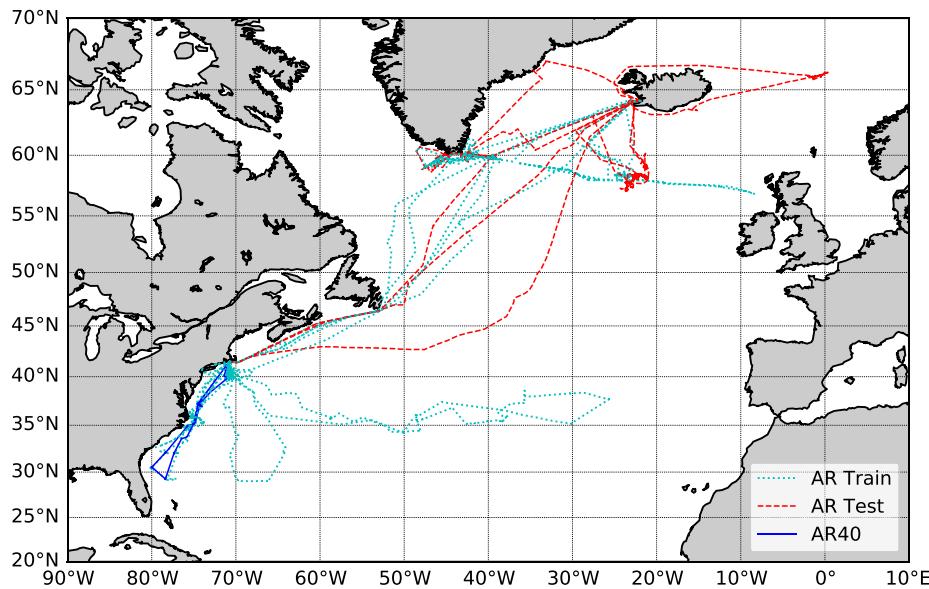
The 18 cruises of the *Sally Ride* used for further model testing are mapped in Fig. 4. As in Fig. 3, Fig. 4 reflects all location data from each cruise rather than the exact time windows with usable data. For instance, less than half of the transpacific voyage shown in Fig. 4 yielded motions data with acceptable status flags, as mentioned in Section 2.2.2. The SR Test dataset is outlined in Table 3, with the 18 individual cruises detailed in Table B.12 of Appendix A.

### 2.5. CMEMS wave data and input variables

The wave data source selected for seakeeping predictions (PBMPs) and data-driven model training came from the European Union's Copernicus Marine Environment Monitoring Service (CMEMS). More specifically, the wave data for all cruises were downloaded through the Global Ocean Waves Analysis and Forecasting Product (GLOBAL\_ANALYSIS\_FORECAST\_WAV\_001\_027). Multidirectional wave data were provided at three-hour time intervals for a majority of the world's oceans in latitude and longitude increments of 5 arcminutes ( $\frac{1}{12}^\circ$ ). Schirrmann et al. (2019) utilized lower resolution, unidirectional wave data and directly used the wave data output at the grid intersection nearest the vessel's mean location. Due to the relatively high spatial resolution of the wave data used in this work and the selected 30-minute window for the onboard measurements in contrast to the three-hour wave hindcast intervals, all wave data parameters were interpolated in three dimensions: latitude, longitude, and time. The specific interpolation approach employed here is described in the following paragraph.

For a given measured data window, interpolated wave data values were calculated at the mean latitude, longitude, and time from a 30-minute measurement period. To do so, eight wave data points, each a corner of the rectangular prism containing these mean value coordinates in 3D (latitude, longitude, and time) space, were used to calculate linearly interpolated wave data parameters. For each time and location, 15 parameter types were extracted from the CMEMS wave hindcast and appropriately interpolated. These 15 parameters included several significant wave height, period, and direction metrics, and are included in Table 4 as indices 1–15. The names of these 15 variables correspond to the names assigned by CMEMS as described in Fernandez and Aouf (2018). Note that select variables include alternate names in parentheses that match more common notation used in other sections of this discussion. Also note that special care was taken to avoid issues with circularity (e.g., transition from 359° to 0°) when calculating means and interpolating directional parameters. More specifically, sine and cosine were employed for these calculations rather than relying solely on wave directions.

As shown in Table 4, in addition to more general parameters for the wave system as a whole, the CMEMS data provided information regarding the multidirectionality of the waves through significant wave height, mean wave period, and mean wave direction parameters for three different wave partitions: wind waves ( $WW$ ), primary swell ( $SW1$ ), and secondary swell ( $SW2$ ). Note the inclusion of multiple period metrics that are each defined in Table 4, where  $m_0$ ,  $m_1$ , and  $m_2$  correspond to the zeroth, first, and second moments of the wave spectrum, respectively. As mentioned in Section 2.2.3, all of the CMEMS direction parameters were adjusted such that they were relative to the vessel's heading, with 180° corresponding to head seas and 0°

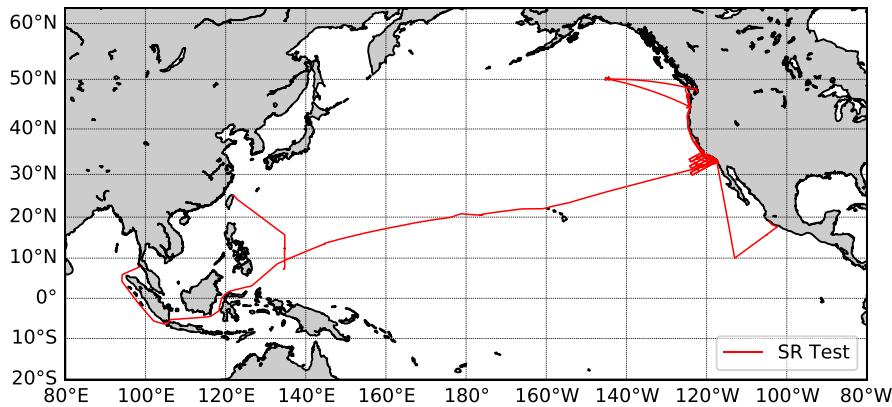


**Fig. 3.** Map of the Atlantic Ocean showing the *Neil Armstrong*'s 42 train and 8 test (with AR40 indicated) cruises. The exact train and test data locations used and omissions of cruise portions due to the windowing approach are not reflected in this map.

**Table 3**

Summary of the datasets considered in this work. Negative longitude values indicate those west of  $0^\circ$  (e.g.,  $-50^\circ$  is equivalent to  $50^\circ$  W in Fig. 3).

| Dataset  | Lat <sub>N</sub> | Lat <sub>S</sub> | Lon <sub>E</sub> | Lon <sub>W</sub> | $H_{s50}$<br>[m] | $H_{s100}$<br>[m] | $T_{m50}$<br>[s] | $T_{m100}$<br>[s] | Windows |
|----------|------------------|------------------|------------------|------------------|------------------|-------------------|------------------|-------------------|---------|
| AR Train | 64.1             | 29.0             | -8.4             | -80.0            | 1.67             | 6.41              | 7.44             | 13.07             | 10,008  |
| AR Test  | 66.9             | 41.4             | 0.3              | -70.3            | 1.86             | 5.84              | 7.86             | 12.02             | 3,384   |
| SR Test  | 50.3             | -6.4             | -102.4           | 94.1             | 1.43             | 3.87              | 9.94             | 14.94             | 2,592   |
| AR40     | 41.3             | 29.2             | -70.9            | -80.0            |                  |                   |                  |                   | 132     |
| 120–96 h |                  |                  |                  |                  | 1.02             | 2.35              | 6.73             | 9.91              |         |
| 96–72 h  |                  |                  |                  |                  | 1.07             | 2.34              | 6.55             | 10.11             |         |
| 72–48 h  |                  |                  |                  |                  | 1.07             | 2.13              | 6.73             | 10.61             |         |
| 48–24 h  |                  |                  |                  |                  | 1.17             | 2.15              | 6.87             | 11.12             |         |
| 24–0 h   |                  |                  |                  |                  | 1.17             | 2.15              | 6.93             | 11.12             |         |
| Hindcast |                  |                  |                  |                  | 1.16             | 2.33              | 6.85             | 11.57             |         |



**Fig. 4.** Map of the Pacific Ocean showing the *Sally Ride*'s 18 test cruises. The exact train and test data locations used and omissions of cruise portions due to the windowing approach are not reflected in this map. Also note the change in latitude/longitude scales relative to Fig. 3.

corresponding to following seas. For this reason, both CMEMS and WHOI/SIO are listed as data sources for these direction parameters in Table 4. Due to the port-starboard symmetry of the vessel, the seakeeping model RAOs described in the next section were identical between

wave systems coming from the port and starboard sides. Therefore, any relative wave directions on  $(180^\circ, 360^\circ)$  were flipped to be on  $[0^\circ, 180^\circ]$  for the purposes of data-driven model training. Note that in applications which require more rigorous handling of phase relationships, such as

**Table 4**

Data-driven model input variables. The primary data source(s) used to generate each variable is(are) included. Although not explicitly listed, note that time and location information from WHOI/SIO's onboard measurements were also required to determine all of the CMEMS wave data variables (indices 1–15).

| Index | Variable                  | Units | Description  | Source(s)             |
|-------|---------------------------|-------|--|-----------------------|
| 0     | $U$                       | m/s   | Vessel speed   | WHOI/SIO              |
| 1     | $VHM0 (H_s)$              | m     | Significant wave height  | CMEMS                 |
| 2     | $VTM10 (T_m)$             | s     | Mean wave period ( $m_0$ and $m_1$ )                           | CMEMS                 |
| 3     | $VTM02$                   | s     | Mean wave period ( $m_0$ and $m_2$ )                           | CMEMS                 |
| 4     | $VTPK$                    | s     | Wave period at spectral peak                                   | CMEMS                 |
| 5     | $VMDR_{Rel}$              | deg   | Mean wave direction (Relative to ship heading)                 | CMEMS/WHOI/SIO        |
| 6     | $VPED_{Rel}$              | deg   | Wave direction at spectral peak (Relative to ship heading)     | CMEMS/WHOI/SIO        |
| 7     | $VHM0_{WW} (H_{s,WW})$    | m     | Wind waves significant wave height                             | CMEMS                 |
| 8     | $VTM01_{WW} (T_{m,WW})$   | s     | Wind waves mean wave period ( $m_0$ and $m_1$ )                | CMEMS                 |
| 9     | $VTMDR_{Rel,WW}$          | deg   | Wind waves mean wave direction (Relative to ship heading)      | CMEMS/WHOI/SIO        |
| 10    | $VHM0_{SW1} (H_{s,SW1})$  | m     | Primary swell significant wave height                          | CMEMS                 |
| 11    | $VTM01_{SW1} (T_{m,SW1})$ | s     | Primary swell mean wave period ( $m_0$ and $m_1$ )             | CMEMS                 |
| 12    | $VTMDR_{Rel,SW1}$         | deg   | Primary swell mean wave direction (Relative to ship heading)   | CMEMS/WHOI/SIO        |
| 13    | $VHM0_{SW2} (H_{s,SW2})$  | m     | Secondary swell significant wave height                        | CMEMS                 |
| 14    | $VTM01_{SW2} (T_{m,SW2})$ | s     | Secondary swell mean wave period ( $m_0$ and $m_1$ )           | CMEMS                 |
| 15    | $VTMDR_{Rel,SW2}$         | deg   | Secondary swell mean wave direction (Relative to ship heading) | CMEMS/WHOI/SIO        |
| 16    | $\sigma_{3P,WW}$          | m     | Predicted heave std. dev. due to wind waves                    | SHIPMO/CMEMS/WHOI/SIO |
| 17    | $\sigma_{SP,WW}$          | deg   | Predicted pitch std. dev. due to wind waves                    | SHIPMO/CMEMS/WHOI/SIO |
| 18    | $\sigma_{4P,WW}$          | deg   | Predicted roll std. dev. due to wind waves                     | SHIPMO/CMEMS/WHOI/SIO |
| 19    | $\sigma_{3P,SW1}$         | m     | Predicted heave std. dev. due to primary swell                 | SHIPMO/CMEMS/WHOI/SIO |
| 20    | $\sigma_{SP,SW1}$         | deg   | Predicted pitch std. dev. due to primary swell                 | SHIPMO/CMEMS/WHOI/SIO |
| 21    | $\sigma_{4P,SW1}$         | deg   | Predicted roll std. dev. due to primary swell                  | SHIPMO/CMEMS/WHOI/SIO |
| 22    | $\sigma_{3P,SW2}$         | m     | Predicted heave std. dev. due to secondary swell               | SHIPMO/CMEMS/WHOI/SIO |
| 23    | $\sigma_{SP,SW2}$         | deg   | Predicted pitch std. dev. due to secondary swell               | SHIPMO/CMEMS/WHOI/SIO |
| 24    | $\sigma_{4P,SW2}$         | deg   | Predicted roll std. dev. due to secondary swell                | SHIPMO/CMEMS/WHOI/SIO |
| 25    | $\sigma_{3P}$             | m     | Predicted heave std. dev. due to all three wave partitions     | SHIPMO/CMEMS/WHOI/SIO |
| 26    | $\sigma_{SP}$             | deg   | Predicted pitch std. dev. due to all three wave partitions     | SHIPMO/CMEMS/WHOI/SIO |
| 27    | $\sigma_{4P}$             | deg   | Predicted roll std. dev. due to all three wave partitions      | SHIPMO/CMEMS/WHOI/SIO |

time-domain analysis, distinction between the port and starboard sides would be necessary.

One drawback of the wave data interpolation approach employed here is that CMEMS' identification of the three wave partitions was not necessarily consistent between the eight corners in latitude, longitude, and time space. For example, the primary and secondary swell partitions could swap between two corners if their significant wave heights were similar, which would yield unreliable interpolated values, especially with regard to the wave partitions' mean directions. While these instances were likely uncommon, it is important to note that this interpolation method was imperfect, and future work may search for data-driven models that circumvent this issue.

In addition to the wave hindcast data used for all work with the AR Train, AR Test, and SR Test datasets, CMEMS provides wave forecast data for up to 120 h (5 days) in the future, updated once daily (Fernandez and Aouf, 2018). Therefore, for any given time of interest, there are five different forecast updates prior. Of these five updates, one update will fall in each of the following time horizon windows, defined as the time between forecast file upload and the time of interest: 120–96 h, 96–72 h, 72–48 h, 48–24 h, and 24–0 h. As mentioned in Section 2.3, the purpose of the final *Neil Armstrong* cruise, AR40, was to compare the performance of the data-driven models using these five forecast wave datasets and the corresponding hindcast wave dataset. Note that CMEMS wave forecast files are only available for a short period surrounding the applicable time windows, after which only the corresponding hindcast files are available. As such, the storage of these forecast files was planned and executed specifically for this single cruise. A summary of AR40, which yielded 132 windows, is included in Table 3. The wave data corresponding to the five forecast time horizon windows and the hindcast data are also summarized in Table 3. The results of this comparison study are discussed in Section 3.4.

## 2.6. Physics-based model predictions (PBMPs)

The following section describes the procedure used to generate sea-keeping predictions for incorporation of a physical-basis into the data-driven models. Note that no spreading was applied because directional

spread information was not provided with the CMEMS multidirectional wave data.

Frequency-domain, 2D strip-theory was employed for heave ( $j = 3$ ), pitch ( $j = 5$ ), and roll ( $j = 4$ ) response prediction using the multidirectional CMEMS hindcast wave data. For a given data window and each of the three wave partitions,  $WP$  (i.e.,  $WW$ ,  $SW1$ , or  $SW2$ ), a Bretschneider (ITTC two-parameter) spectrum,  $S_{WP}^+(\omega_0)$ , was created using Eqs. (1)–(3) and the wave partition's significant wave height,  $H_{s,WP}$ , and mean wave period  $T_{m,WP}$ , as defined in Table 4 (Lloyd, 1989).

$$S_{WP}^+(\omega_0) = \frac{A}{\omega_0^5} \exp\left(\frac{-B}{\omega_0^4}\right) \quad (1)$$

$$A = 172.75 \left( \frac{H_{s,WP}^2}{T_{m,WP}^4} \right) \quad (2)$$

$$B = \frac{691}{T_{m,WP}^4} \quad (3)$$

The creation of an idealized spectrum for each of the three wave partitions was necessary because characteristic parameters were provided by the CMEMS hindcast dataset rather than fully-defined spectra. Using these idealized wave spectra,  $S_{WP}^+(\omega_0)$ , the vessel's one-sided pseudo-response spectrum in the  $j^{th}$  DOF due to an individual wave partition,  $S_{j,WP}^+(\omega_0)$ , was calculated using Eq. (4).

$$S_{j,WP}^+(\omega_0) = S_{WP}^+(\omega_0) |RAO_{j,WP}(\omega_0)|^2 \quad (4)$$

The prefix “pseudo-” indicates that the calculations were performed in wave frequency space,  $\omega_0$ , rather than encounter frequency space,  $\omega_e$ . In Eq. (4),  $RAO_{j,WP}(\omega_0)$  is the 2D strip-theory response amplitude operator (RAO) in the  $j^{th}$  DOF. Note that in addition to wave frequency,  $\omega_0$ , this RAO was a function of the vessel's speed and heading relative to the incoming waves of a given wave partition. SHIPMO.BM, a frequency-domain, 2D strip-theory code developed at the University of Michigan by Beck and Troesch (1990) was used to generate RAOs for this work. Heave, pitch, and roll RAOs were generated for all

combinations of speed and relative heading from 0 to 15 knots in 1 knot increments and 0 to 180° in 5° increments, respectively. The roll calculations of this seakeeping code were based on the method of [Himeno \(1981\)](#). As mentioned previously, the VCG of the vessel was unknown. As such, the VCG was assumed at the waterline because it yielded a marginally stable vessel for the purposes of roll prediction. Given the typical inaccuracy of empirical models for roll prediction in addition to the uninformed VCG assumption, these roll predictions should not be trusted on their own. Given the relatively high RAO resolution in terms of speed and heading, for a given DOF and wave partition, the closest RAO with respect to these two variables was selected for use in Eq. (4) (i.e., interpolation in speed and heading was not applied for the RAOs).

The zeroth response moment due to an individual wave partition in the  $j^{th}$  DOF,  $m_{j0,W P}$ , was calculated through trapezoidal integration of  $S_{WP}^+(\omega_0)$  following Eq. (5).

$$m_{j0,W P} = \int_0^\infty S_{j,W P}^+(\omega_0) d\omega_0 \quad (5)$$

After calculation of  $m_{j0,W P}$  for each of the three wave partitions, the principle of linear superposition was applied to calculate the total zeroth response moment in the  $j^{th}$  DOF,  $m_{j0}$ , as shown in Eq. (6).

$$m_{j0} = m_{j0,WW} + m_{j0,SW1} + m_{j0,SW2} \quad (6)$$

Under the assumptions of this approach, the total zeroth response moment,  $m_{j0}$ , was equivalent to the variance of the response. Therefore, the standard deviation of the response,  $\sigma_{jP}$ , was calculated by taking the square root of  $m_{j0}$  as shown in Eq. (7).

$$\sigma_{jP} = \sqrt{m_{j0}} \quad (7)$$

This response standard deviation is a measure of the vessel's predicted response amplitude in the  $j^{th}$  DOF. The subscript  $P$  was included in  $\sigma_{jP}$  to indicate that this value was the predicted total response amplitude before application of the data-driven models. Note that Eq. (7) was also used to convert response zeroth moments from the individual wave components ( $m_{j0,WW}$ ,  $m_{j0,SW1}$ , and  $m_{j0,SW2}$ ) to standard deviations ( $\sigma_{jP,WW}$ ,  $\sigma_{jP,SW1}$ , and  $\sigma_{jP,SW2}$ ) for use as input variables to the data-driven models. While standard deviations were used as the representative amplitudes input to and output from the data-driven models, for reference, the significant amplitude of the response,  $\zeta_{jP}$ , could be calculated using Eq. (8).

$$\zeta_{jP} = 2\sigma_{jP} \quad (8)$$

As shown by indices 16–27 of [Table 4](#), the representative heave, pitch, and roll amplitudes due to each wave component ( $\sigma_{jP,WW}$ ,  $\sigma_{jP,SW1}$ , and  $\sigma_{jP,SW2}$ ), as well as the resultant amplitudes due to all three wave partitions ( $\sigma_{jP}$ ), were input to the data-driven models as PBMPs when applicable.

## 2.7. Model output variables and data normalization

The three output variables of the data-driven models are listed in [Table 5](#), which correspond to the measured target variables used to train the models given in [Table 2](#). Because machine learning models are often sensitive to significant differences in magnitudes between variables, all of the input and output variables were normalized to take values on [0, 1], approximately, with deviations from this range explained in the following. For all directional input variables listed in [Table 4](#), which were previously adjusted to be on [0°, 180°], this normalization was achieved by dividing all values by 180°. For all significant wave height and period metrics, as well as speed, the maximum value seen in AR Train for a specific variable was used to normalize all of the data for that variable in all datasets. As such, some of the normalized values in the test datasets exceeded 1 if larger magnitudes were present in the test set than the train set; however, using the maximum of all data

(train or test) would compromise the integrity of data-driven model testing.

For the remaining input variables and all output variables, which were all representative amplitude metrics for heave, pitch, or roll, the maximum measured value in AR Train for each DOF was used to normalize all data in that DOF. For example, the maximum measured heave standard deviation ( $\sigma_{3M}$ ) in AR Train, which resulted from all wave partitions, was used to normalize all input PBMPs of heave standard deviations due to wind waves ( $\sigma_{3P,WW}$ ). These heave, pitch, and roll normalization factors are included in [Table 5](#). Note that because the PBMPs could be greater than the measured values, there were many instances in which both train and test normalized values were greater than 1; however, normalization was still important such that the ranges of different variables were more similar than prior to normalization.

## 2.8. Ridge Regression (RR)

The linear ridge regression (RR) models employed in this work were constructed and trained using Python's `scikit-learn` library ([Pedregosa et al., 2011](#)). This RR approach is very similar to the linear LS approach employed in [Schirrmann et al. \(2019\)](#). Given  $N$  train data samples, the input data are organized into an  $N$ -row matrix,  $\mathbf{X}_{Train}$ , and the corresponding measured data are organized into a separate  $N$ -row matrix,  $\mathbf{Y}_{Train}$ . In this work,  $\mathbf{X}_{Train}$  had dimensions  $N \times 17$  when PBMPs were excluded and  $N \times 29$  when PBMPs were included. In each case, respectively, the first 16 or 28 columns corresponded to the input variables given in [Table 4](#), and the one additional column was a vector of ones for multiplication with a bias term. The measured output matrix,  $\mathbf{Y}_{Train}$ , had dimensions  $N \times 3$  with the 3 columns corresponding to the measured heave, pitch, and roll values outlined in [Table 2](#). The minimization problem addressed in RR is posed in Eq. (9), which shows that  $\mathbf{X}_{Train}$  is multiplied by a matrix of weights,  $\mathbf{W}$ . In this work,  $\mathbf{W}$  had dimensions  $17 \times 3$  without PBMPs or  $29 \times 3$  with PBMPs.

$$\mathbf{W}_{RR} = \underset{\mathbf{W}}{\operatorname{argmin}} \|\mathbf{X}_{Train} \mathbf{W} - \mathbf{Y}_{Train}\|_F^2 + \lambda \|\mathbf{W}\|_F^2 \quad (9)$$

In Eq. (9),  $\|\cdot\|_F^2$  denotes the squared Frobenius norm, which is the sum of all squared matrix elements. The key difference between the linear LS approach employed in [Schirrmann et al. \(2019\)](#) and this linear RR approach is the incorporation of an  $L2$  weight regularization parameter,  $\lambda$ , to formally mitigate overfitting of the train data as shown in the second term of Eq. (9). This regularization approach penalizes large weight values in matrix  $\mathbf{W}$ . Large weight values are often undesirable because they can cause a model to rely too heavily on individual input variables. Note that the final row of  $\mathbf{W}$ , which contains three bias terms corresponding to each DOF, is excluded from this penalization. Overfitting can lead to poor generalization, or performance when a model is applied to an unseen test dataset. The tuning of  $\lambda$  using train data is discussed in [Section 2.10](#). Given  $\lambda$ , the solution to this minimization problem is given in Eq. (10).

$$\mathbf{W}_{RR} = [\mathbf{X}_{Train}^T \mathbf{X}_{Train} + \lambda \mathbf{I}]^{-1} \mathbf{X}_{Train}^T \mathbf{Y}_{Train} \quad (10)$$

Using the trained weight matrix,  $\mathbf{W}_{RR}$ , and a matrix of test input data,  $\mathbf{X}_{Test}$ , with dimensions  $M \times 17$  without PBMPs or  $M \times 29$  with PBMPs, where  $M$  is the number of samples in the test set, RR test predictions were calculated as shown in Eq. (11) and compared to the known measured test values for evaluation of the model.

$$\mathbf{Y}_{RR,Test} = \mathbf{X}_{Test} \mathbf{W}_{RR} \quad (11)$$

The RR model's performance on the test dataset was then evaluated as shown in Eq. (12).

$$MSE_{RR,Test} = \frac{1}{3} \|\mathbf{Y}_{RR,Test} - \mathbf{Y}_{Test}\|_F^2 \quad (12)$$

This MSE metric is the mean of the heave, pitch, and roll MSE values, which are also analyzed independently in [Section 3](#). Note that RR can be used to model nonlinear phenomena (e.g., quadratic) via the input

**Table 5**

Data-driven model output variables and their normalization factors, which were the maximum values recorded in the AR Train measurements.

| Index | Variable                         | Units | Norm. factor | Description               |
|-------|----------------------------------|-------|--------------|---------------------------|
| 0     | $\sigma_{3RR}$ or $\sigma_{3NN}$ | m     | 1.822        | Predicted heave std. dev. |
| 1     | $\sigma_{5RR}$ or $\sigma_{5NN}$ | deg   | 2.963        | Predicted pitch std. dev. |
| 2     | $\sigma_{4RR}$ or $\sigma_{4NN}$ | deg   | 6.567        | Predicted roll std. dev.  |

variables (e.g., inclusion of  $U$  and  $U^2$ ) while still considered a linear model with respect to the learned weight parameters in  $\mathbf{W}$ ; however, in this work, the RR approach was linear in both the input variable space and the weight parameters.

### 2.9. Artificial neural network (NN)

NN construction and training was performed using Python's Keras library with the TensorFlow backend (Chollet et al., 2015). The NNs employed in this work were feed-forward networks composed of an input layer with one neuron corresponding to each of the input variables (16 or 28 with or without PBMPs, respectively), multiple hidden layers with varying numbers of neurons, and an output layer with three neurons corresponding to the heave, pitch, and roll output variables given in Table 5. NN training was performed using the MSE loss function, Adam optimizer, and a batch size of 50, which was found to yield similar results to a batch size of 1 with a significant reduction in computation time. The average MSE across heave, pitch, and roll was used as the NN scoring metric during training, analogous to Eq. (12). Note that several single-output NNs (e.g., roll only) were experimented with in the preliminary stages of this study, which yielded no notable performance improvement over these multi-output NNs when applied to a test dataset. For example, in preliminary experiments, the inclusion of heave and pitch standard deviation as additional output variables did not detriment the predictive performance for roll standard deviation.

Because the goal of these networks was regression with input and output variables taking possible values on  $[0, \infty)$ , both the input and output layer neurons used linear activation functions. The hidden layers (HLS) used neurons with the rectified linear unit (ReLU) activation function, which allowed the NN to model nonlinear relationships between variables unlike the linear RR approach. Similar to the RR approach, formal incorporation of  $L2$  weight regularization between layers of the NN with parameter  $\lambda$  was implemented to mitigate overfitting, which penalized large weight values via the same mechanism as RR. Weight regularization was not present in Schirrmann et al. (2019) due to the limited dataset size. The HL architecture (i.e., number of HLS and neurons per HL) and  $\lambda$  were determined via cross-validation as described in the following section. For more detailed background information on NNs, please see relevant textbooks such as Bishop (2006) and Russell and Norvig (2010).

### 2.10. Model training and cross-validation (CV)

For both data-driven approaches, 5-fold cross-validation (CV), described below, was applied to tune model parameters using GridSearchCV of Python's scikit-learn library (Pedregosa et al., 2011). Due to the relatively high computational cost of NN training, the following cross-validation procedures were performed for NNs using computational resources and services provided by Advanced Research Computing at the University of Michigan, Ann Arbor. In this work, the parameters that were formally tuned were  $\lambda$  for both the RR and NN approaches and the HL architecture for the NN only. Due to the previously mentioned computational costs, other NN parameters, such as HL activation functions, batch size, and optimizer, were informally experimented with to arrive at their selected values. For the same reason, note that while significant effort was put toward selecting evenly spread parameters in the search for optimally trained models,

the possibilities for these parameters are infinite. Therefore, selectivity was necessary, especially for the HL architectures of the NN.

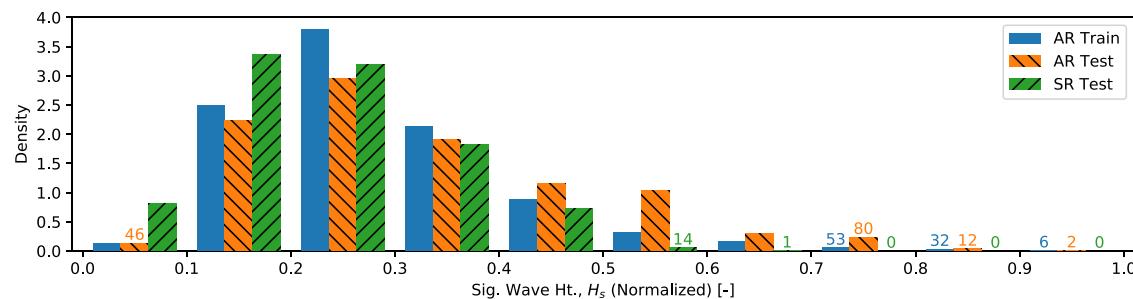
In 5-fold CV, the train dataset is divided into five subsets of approximately equal size. For a given model and all combinations of candidate parameters specified by the user, training is performed using four of the five subsets and then applied to the one remaining, or held out, subset. The MSE between the model's predictions for the held out subset and the corresponding measured values is stored. This operation is then repeated four more times with the same model parameters such that each of the five subsets is held out once. The average of the five MSE losses is calculated and stored as the average CV loss for a given combination of model parameters. This process is then repeated for all combinations of candidate model parameters specified by the user. After all combinations have been tested, the combination of parameters that yielded the lowest average CV loss is selected as the best model from the user-specified candidates. This best model is then reconstructed and retrained using all of the train data (from all five subsets), at which point model training is complete and final model evaluation can begin using the unseen test set(s).

Separation of the train dataset into five subsets for 5-fold CV was performed without shuffling the data; i.e., the first subset was the first 20% of the train data samples chronologically, the second subset was the next 20% of train data samples chronologically, etc. This method of partitioning was selected to avoid separating nearly identical samples into multiple CV subsets, especially given the relatively short 30-minute duration employed here. If nearly identical samples collected in sequence were separated into multiple subsets, overfitting may occur despite the regularization parameter, especially if the sequential samples were outliers. Overfitting would cause the model to have poor generalization, yielding inadequate performance when applied to the test data. On the contrary, it is also possible that random shuffling may improve the model's generalization if input data distributions of chronologically partitioned subsets are drastically different. In general, this is a very complex issue that depends heavily on the available dataset for a given project. As such, given the large number of samples available to train the models in this work, the train data was not shuffled prior to partitioning into CV subsets for all results presented here.

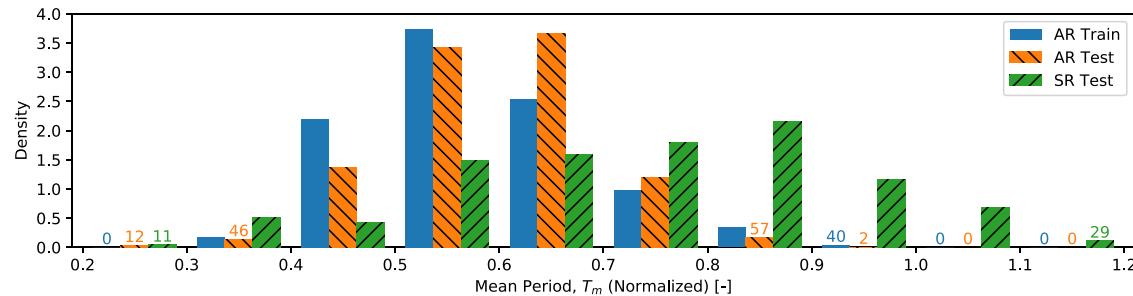
Although not presented, some experiments were performed with random shuffling applied before partitioning the train data into CV subsets. As expected, this random partitioning did lead to lower average CV losses due to increased similarity between the five CV subsets; however, because this similarity was somewhat artificial and did not hold between the train and test sets, with shuffling, both the RR and NN models performed similarly or slightly worse than the results presented in the following section where no shuffling was applied.

### 2.11. Comparison of train and test datasets

In an ideal scenario for data-driven models, the train dataset would contain samples that collectively characterize the input data space occupied by test samples. Assuming reliable measurements used to train and test the models, this similarity between train and test datasets should yield very reliable predictions on the test dataset. Of course, in practice, this is generally not the case. As the number of input variables and their ranges increase, the amount of data needed to adequately characterize the input space increases. Furthermore, these data need to



**Fig. 5.** Histogram of normalized significant wave height comparing the AR Train, AR Test, and SR Test datasets. Although the bars reflect data density, the numbers listed above select bars are the total number of samples for a dataset in a given bin, and are included where these numbers falls below 100.



**Fig. 6.** Histogram of normalized mean wave period ( $T_m$ ) comparing the AR Train, AR Test, and SR Test datasets. Although the bars reflect data density, the numbers listed above select bars are the total number of samples for a dataset in a given bin, and are included where these numbers falls below 100.

contain enough variation to characterize important regions of the input space for future predictions.

A histogram comparing the significant wave height distributions of the AR Train, AR Test, and SR Test datasets is given in Fig. 5. Note that Fig. 5 gives data densities rather than numbers of samples to allow more simple comparison between the AR Train dataset and the smaller test datasets. Also note, however, that the numbers above select bars correspond to numbers of samples, which are present wherever the number of binned samples fell below 100. For context, a lower number of binned samples in the AR Train dataset based on this single input variable increases the likelihood that test samples in this bin lack adequate representation for reliable predictions. As shown in Fig. 5, the AR Train and Test datasets had higher densities at greater significant wave heights than the SR Test dataset, which was also implied by the wave statistics in Table 3. Although normalized significant wave height bins above 0.7 are relatively sparse in terms of AR Train samples, there are not any bins where test data are present without train data.

In contrast, Fig. 6 shows a histogram of normalized mean wave period ( $T_m$ ) that highlights bins below 0.3 and above 1.0 where test samples are present without corresponding train data. The SR Test dataset from the *Sally Ride* had a higher density of samples with longer wave periods than the data from the *Neil Armstrong*, which was also implied in Table 3. This contrast is likely related to the expected wave profile differences between the Atlantic and Pacific Oceans. This shift between datasets adds another level of complexity and intrigue for application of these data-driven models, trained solely using data from the *Neil Armstrong*, to the *Sally Ride*. While the comparisons made between datasets in this section were very limited relative to the number of input variables and possible conditions experienced, it is important to consider the role that dataset similarity (or dissimilarity) plays in the performance of these data-driven models and the potential benefits of including PBMPs to aid in train data-sparse regions of the input space.

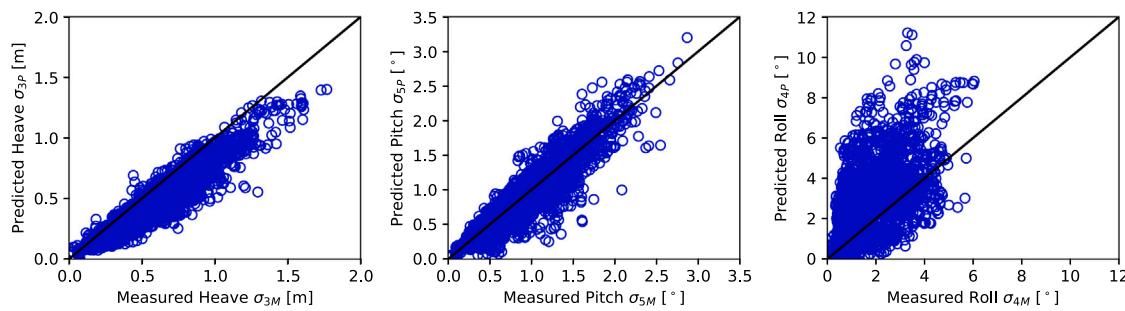
### 3. Results and discussion

The uncorrected PBMPs of heave, pitch, and roll standard deviations for the 3,384 AR Test windows, which were calculated using the 2D strip-theory RAOs, are plotted in Fig. 7. The black lines have a slope of 1.0 and represent perfect agreement between measured and predicted values. Fig. 7 shows reasonably strong agreement of uncorrected heave and pitch PBMPs with their respective measured values, especially in comparison to roll. As expected, these roll PBMPs were very poor, which allowed for the data-driven models to make significant improvements.

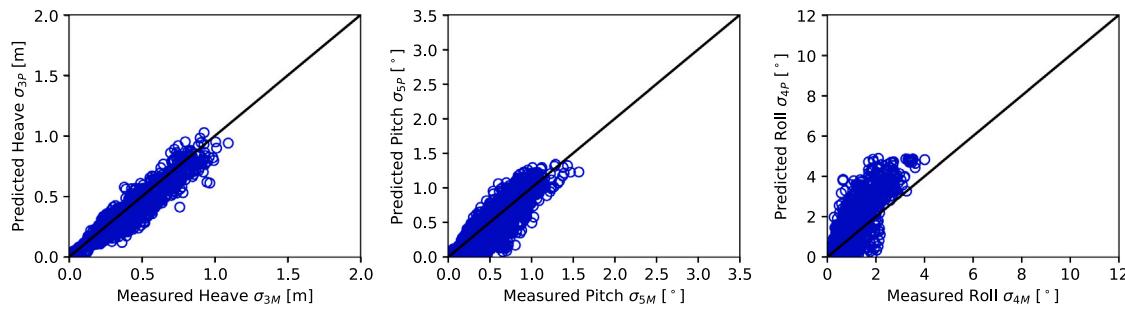
Fig. 8 shows the 2,592 uncorrected PBMPs of the SR Test set. Comparing Figs. 7 and 8, it is evident that the maximum SR Test dataset motions were less than those of AR Test. Furthermore, compared to AR Test, there are fewer roll PBMPs that are drastically different than their respective measured values in the SR Test dataset. These lesser roll magnitudes and discrepancies may be linked to the lower significant wave heights or longer periods seen in the SR Test dataset, as discussed in Section 2.11. It is also possible that the actual and assumed mean loading conditions were more similar for the *Sally Ride* than they were for the *Neil Armstrong*, yielding more accurate PBMPs. Additionally, there could have been differences in the quality of measurements or subsequent filtering between datasets. For instance, it is possible that the measurement instruments and quality assessment for the *Sally Ride* removed more outlying (but not necessarily erroneous) measurements than the same combination for the *Neil Armstrong*. As such, the insights that can be drawn from direct comparison of AR Test and SR Test are limited. The PBMPs shown in Figs. 7 and 8 serve as the baseline for comparisons in the following discussion.

#### 3.1. R/V Neil Armstrong

This section details the results of data-driven model training and testing using the AR Train and AR Test datasets, respectively. Much of the discussion in this section follows that of Schirrmann et al. (2020).



**Fig. 7.** AR Test dataset uncorrected PBMPs of heave, pitch, and roll amplitudes calculated using the 2D strip-theory RAOs versus measured values. The black lines have slope 1.0 and represent perfect agreement between measured and predicted values for reference.



**Fig. 8.** SR Test dataset uncorrected PBMPs of heave, pitch, and roll amplitudes calculated using the 2D strip-theory RAOs versus measured values. The black lines have slope 1.0 and represent perfect agreement between measured and predicted values for reference.

**Table 6**

Results of the 5-fold CV model selection processes for each of the four data-driven models. These selections were performed exclusively using the data in AR Train.

|                           | RR            |            | NN            |              |
|---------------------------|---------------|------------|---------------|--------------|
|                           | Without PBMPs | With PBMPs | Without PBMPs | With PBMPs   |
| Average CV loss           | 0.004438      | 0.003544   | 0.003359      | 0.003273     |
| std. dev. of CV loss      | 0.001524      | 0.001431   | 0.001331      | 0.001463     |
| Reg. parameter, $\lambda$ | 0.155000      | 0.325000   | 0.000003      | 0.000030     |
| HL architecture           | –             | –          | [40, 20, 10]  | [40, 20, 10] |

### 3.1.1. CV selected models and AR Train performance

**Table 6** summarizes the best models determined via 5-fold CV. Recall that the average CV loss, included in the first row of **Table 6**, was the performance metric used to determine the best parameters for each model from the user-specified candidates. The standard deviation of the CV loss from the five different groupings is also included for the best model, which shows that there were notable differences between model performances for each grouping. Intuitively, when random shuffling was applied in preliminary exploration of models, these CV standard deviations were significantly less due to the artificial similarity between subsets; nevertheless, as mentioned previously, there was no notable improvement in performance when applied to the test datasets. The average CV loss and standard deviations of CV loss are based on the normalized values used by the models.

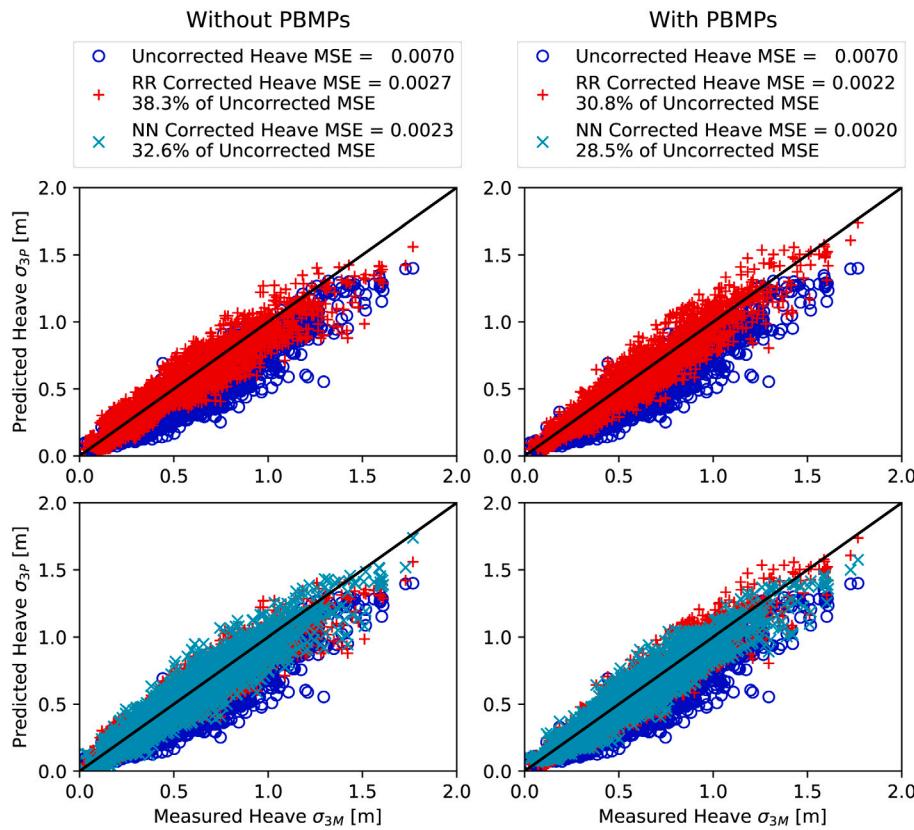
The last two rows of **Table 6** give the parameters that yielded the best models in terms of average CV loss. The HL architecture format [40, 20, 10] means that there were three HMs with 40, 20, and 10 neurons, respectively. While the best HL architectures with and without PBMPs were identical, note that the average CV losses resulting from significantly different HL structures were similar to these best model values; therefore, model performance on train data was not particularly sensitive to the HL architecture selected. For both data-driven model

types, and especially for the NN where computational cost can be significant, it is important to recognize that these best models were selected from a group of user-defined candidates rather than being the absolute best possible performance for a given dataset.

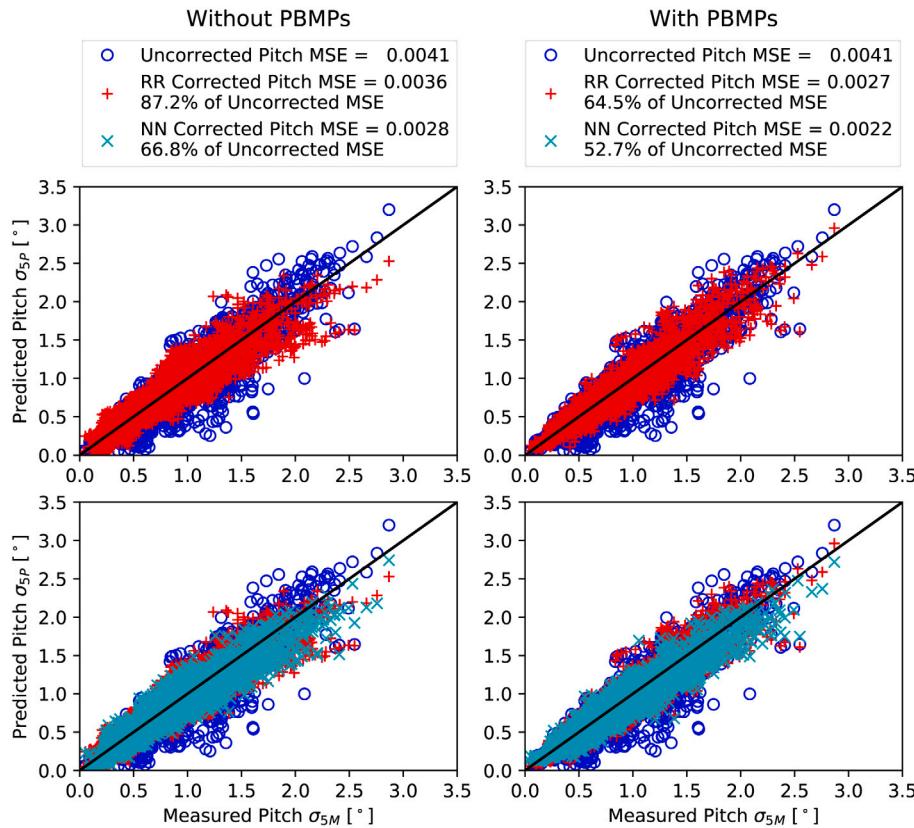
The heave, pitch, roll, and total MSE values of the PBMPs and data-driven approaches (after CV model selection) when applied to the entire AR Train dataset are included in **Table 7**. Recall that total MSE is the mean of the heave, pitch, and roll MSE values and was the metric used to train the models. While values in **Table 7** demonstrate that the models learned from the train data, the insights that can truly be drawn from train data MSE are very limited. The AR Test performance evaluation discussed in Section 3.1.2 is far more significant.

### 3.1.2. AR Test performance

The heave, pitch, roll, and total MSE values of the PBMPs and data-driven approaches when applied to the AR Test dataset are included in **Table 8**. In contrast to train data values of **Table 7**, the test data MSE values in each DOF are the most indicative of a model's real-world performance when applied to an unseen dataset. Therefore, for each DOF in **Table 8**, the rank of each model in terms of test MSE minimization is included as a superscript, with (1) indicating the best model. The performances of these data-driven models on the test data



**Fig. 9.** AR Test dataset uncorrected heave PBMPs and the RR and NN predicted heave amplitudes versus measured values. The black lines have slope 1.0 and represent the line of perfect agreement between measured and predicted values for reference.



**Fig. 10.** AR Test dataset uncorrected pitch PBMPs and the RR and NN predicted pitch amplitudes versus measured values. The black lines have slope 1.0 and represent the line of perfect agreement between measured and predicted values for reference.

**Table 7**

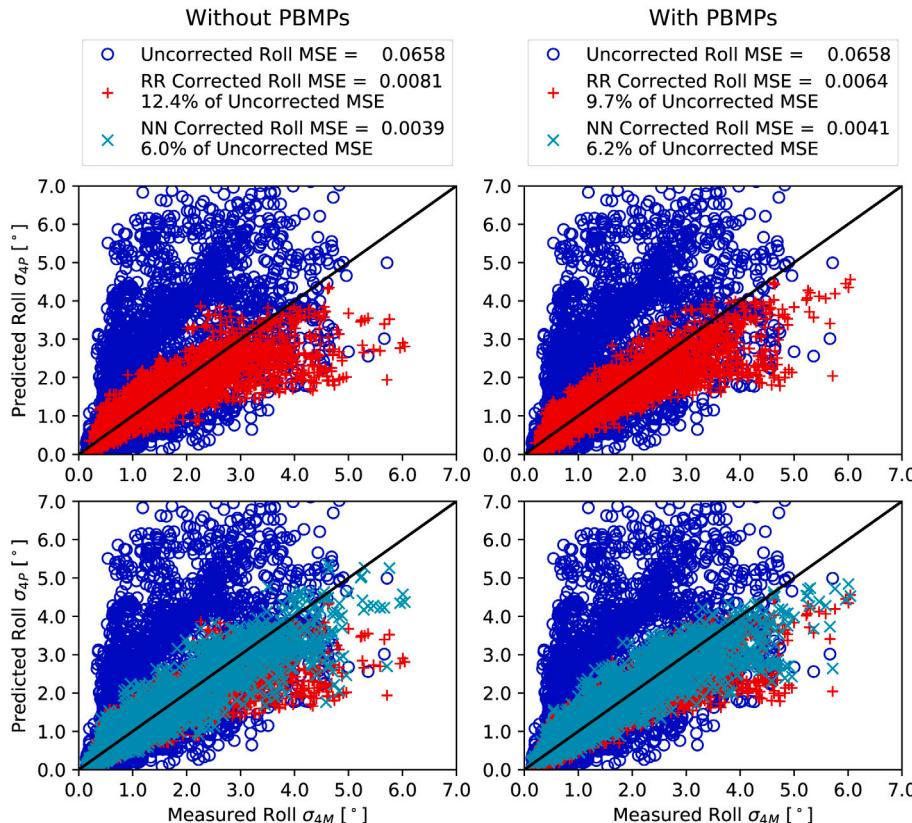
AR Train performance of the data-driven approaches with and without PBMPs, which are also included for reference. After the CV selection process outlined in [Table 6](#) was completed for each model, the model was trained one final time on the entire AR Train dataset. The MSE values in this table are the result of those final training processes.

|       | Uncorrected<br>PBMPs | RR            | NN         |               |            |
|-------|----------------------|---------------|------------|---------------|------------|
|       |                      | Without PBMPs | With PBMPs | Without PBMPs | With PBMPs |
| Heave | 0.005955             | 0.002934      | 0.002319   | 0.001321      | 0.001460   |
| Pitch | 0.004441             | 0.004606      | 0.003255   | 0.001789      | 0.001954   |
| Roll  | 0.075659             | 0.004373      | 0.003571   | 0.001548      | 0.001925   |
| Total | 0.028685             | 0.003971      | 0.003048   | 0.001553      | 0.001780   |

**Table 8**

AR Test performance of the data-driven approaches with and without PBMPs, which are also included for reference. The test data MSE is the best indicator of a model's real-world performance. For this reason, the superscripts are included to indicate the rank of each model in terms of minimizing MSE on the test set for a given DOF with (1) indicating the best model.

|       | Uncorrected<br>PBMPs    | RR                      | NN                      |                         |                         |
|-------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
|       |                         | Without PBMPs           | With PBMPs              | Without PBMPs           | With PBMPs              |
| Heave | 0.007032 <sup>(5)</sup> | 0.002696 <sup>(4)</sup> | 0.002166 <sup>(2)</sup> | 0.002292 <sup>(3)</sup> | 0.002004 <sup>(1)</sup> |
| Pitch | 0.004133 <sup>(5)</sup> | 0.003603 <sup>(4)</sup> | 0.002667 <sup>(2)</sup> | 0.002760 <sup>(3)</sup> | 0.002179 <sup>(1)</sup> |
| Roll  | 0.065751 <sup>(5)</sup> | 0.008134 <sup>(4)</sup> | 0.006372 <sup>(3)</sup> | 0.003944 <sup>(1)</sup> | 0.004108 <sup>(2)</sup> |
| Total | 0.025639 <sup>(5)</sup> | 0.004811 <sup>(4)</sup> | 0.003735 <sup>(3)</sup> | 0.002999 <sup>(2)</sup> | 0.002764 <sup>(1)</sup> |



**Fig. 11.** AR Test dataset uncorrected roll PBMPs and the RR and NN predicted roll amplitudes versus measured values. The black lines have slope 1.0 and represent the line of perfect agreement between measured and predicted values for reference. Note that the roll PBMPs extend above the top of the plot as shown previously in [Fig. 7](#).

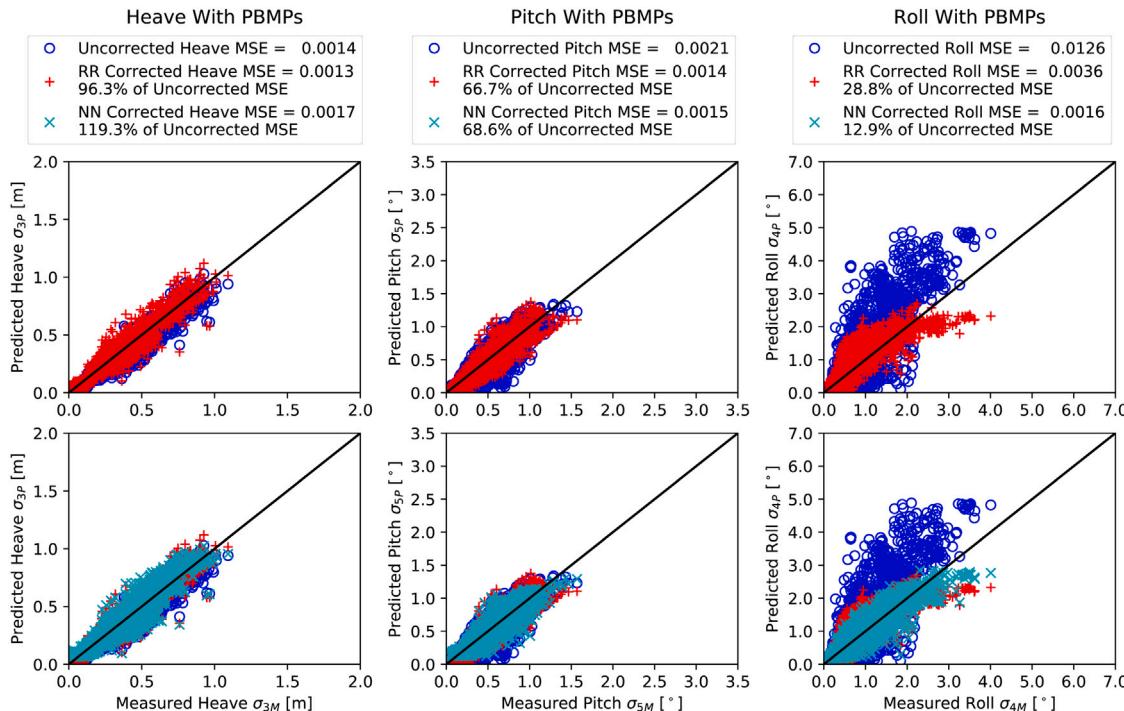
are also demonstrated graphically for heave, pitch, and roll in [Figs. 9, 10](#), and [11](#), respectively. The top row of these figures shows the RR results overlaid on the uncorrected PBMPs (previously shown in [Fig. 7](#)), and the bottom row shows the NN results overlaid on both the RR results and uncorrected PBMPs. As indicated, the left and right columns give the results without and with PBMPs included as input variables, respectively. The legends above each column give the approximate MSE of the normalized results for each approach, which match those presented in [Table 8](#). The RR and NN legend entries also include their

MSE value's relative percentage of the uncorrected PBMPs' MSE. These percentages highlight that for a given DOF and under the same PBMPs assumption (with or without PBMPs), the NN outperformed the RR approach in terms of MSE reduction for AR Test in all cases. This consistently improved performance for the NN is largely attributed to its ability to model nonlinear relationships; however, it is evident that the linear RR approach was still very effective for MSE reduction in all cases. Note that in [Fig. 11](#) for roll, the axes were limited to a value of

**Table 9**

SR Test results of the data-driven approaches with and without PBMPs, which are also included for reference. The test data MSE is the best indicator of a model's real-world performance. For this reason, the superscripts are included to indicate the rank of each model in terms of minimizing MSE on the test set for a given DOF with (1) indicating the best model.

|       | Uncorrected<br>PBMPs    | RR                      |                         | NN                      |                         |
|-------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
|       |                         | Without PBMPs           | With PBMPs              | Without PBMPs           | With PBMPs              |
| Heave | 0.001383 <sup>(2)</sup> | 0.001624 <sup>(3)</sup> | 0.001332 <sup>(1)</sup> | 0.002211 <sup>(5)</sup> | 0.001650 <sup>(4)</sup> |
| Pitch | 0.002148 <sup>(4)</sup> | 0.002439 <sup>(5)</sup> | 0.001432 <sup>(1)</sup> | 0.002077 <sup>(3)</sup> | 0.001473 <sup>(2)</sup> |
| Roll  | 0.012551 <sup>(5)</sup> | 0.004031 <sup>(4)</sup> | 0.003620 <sup>(3)</sup> | 0.001920 <sup>(2)</sup> | 0.001613 <sup>(1)</sup> |
| Total | 0.005361 <sup>(5)</sup> | 0.002698 <sup>(4)</sup> | 0.002128 <sup>(3)</sup> | 0.002069 <sup>(2)</sup> | 0.001579 <sup>(1)</sup> |



**Fig. 12.** SR Test dataset uncorrected PBMPs and the amplitudes predicted by the RR and NN with PBMPs versus measured values for heave, pitch, and roll. The black lines have slope 1.0 and represent the line of perfect agreement between measured and predicted values for reference.

7.0 to highlight the RR and NN results, but the uncorrected PBMPs that were originally presented in Fig. 7 extend to a value of nearly 12.0.

As indicated by the superscripts in Table 8, the NN model with PBMPs yielded the lowest total, heave, and pitch MSE values for the AR Test dataset. Therefore, the introduction of PBMPs as input variables added value and improved the heave and pitch predictions. Furthermore, for both heave and pitch, the second best approach was RR with PBMPs, which means that a linear approach (RR) outperformed a nonlinear approach (NN) when provided with PBMPs. This result was surprising and emphasizes the value that PBMPs can add to a model. This added value is largely attributed to the relatively strong accuracy of the uncorrected PBMPs, which was not the case for roll. As shown in Table 8, for roll, the NN without PBMPs was the best model and slightly outperformed the NN with PBMPs. This result means that no value was added by the PBMPs in the case of the nonlinear NN for roll. In the case of RR, the PBMPs did add value for all three DOFs, which is unsurprising given the limitations of a linear model. However, it should be noted that the improvement in this linear model may have resulted from the added DOFs provided by the PBMPs rather than the values themselves.

### 3.2. R/V Sally Ride test data

The efficacy of the *Neil Armstrong* trained NN and RR models with and without PBMPs when applied to the *Sally Ride* test dataset, SR Test,

are detailed in Table 9. Once again, the SR Test MSE values are ranked, with (1) indicating the best model in each DOF. Additionally, the heave, pitch, and roll uncorrected PBMPs and RR and NN model predictions with PBMPs are plotted versus measured values in Fig. 12.

As shown in Table 9, for roll, the NN with PBMPs yielded the lowest MSE. This value added by roll PBMPs is contrary to the findings for AR Test, which may be due to the better roll PBMP quality for SR Test as discussed with regard to Figs. 7 and 8. For heave and pitch, the top performer was the RR model with PBMPs, which contrasts the superior performance of the NN with PBMPs for the AR Test dataset as given in Table 8. The percentages given in Fig. 12 clarify that for pitch, the performance of the NN with PBMPs was very similar to that of the RR with PBMPs, and both models significantly reduced MSE relative to the uncorrected PBMPs. On the contrary, for heave, Table 9 and Fig. 12 show that the NN with PBMPs actually yielded a higher SR Test MSE than the uncorrected PBMPs. Note that the MSE of these uncorrected heave PBMPs was only slightly greater than MSE of the RR with PBMPs, meaning the best model made minimal improvements upon these PBMPs. While the differences in operating condition distributions of the AR Test and SR Test datasets, highlighted in Section 2.11, did not warrant a direct comparison of MSE values between datasets (Tables 8 and 9), note that the SR Test MSE values of the uncorrected heave, pitch, and roll PBMPs are all less than half of their respective AR Test values. This relationship is also evident visually

when comparing the uncorrected PBMPs given in Figs. 7 and 8. For heave, specifically, the uncorrected PBMP's MSE of SR Test was four times smaller than that of AR Test. As such, the fact that the NN with PBMPs was slightly detrimental to vessel motion predictions may not be a major concern. Nevertheless, it is interesting that the RR with PBMPs outperformed the NN with PBMPs for the SR Test dataset. Additionally, Table 9 shows that for all three DOFs, the RR and NN models with PBMPs outperformed their counterparts without PBMPs, emphasizing the benefits of physics-based information.

### 3.3. Model performance comparisons

While Sections 3.1.2 and 3.2 demonstrated the benefits of including physics-based information for MSE reduction on each test dataset as a whole, there was not a discussion of the models' performances as a function of the measured motions or specific input variables. This section provides these more-detailed model performance analyses.

#### 3.3.1. Motion MSE values vs. Measured motion magnitudes

The true benefits of these data-driven models to a vessel owner or operator may be significantly greater when heave, pitch, or roll response magnitudes are large. Fig. 13 shows heave, pitch, and roll MSE as a function of their respective normalized motion measurements. The three rows correspond to heave, pitch, and roll, and the three columns correspond to the AR Train, AR Test, and SR Test datasets. In each plot, a histogram of the number of samples in 10 different measured motion bins, identified by the right axis, is given. The data overlaid on these histograms are the MSE values calculated from the samples in each bin for a given model or the uncorrected PBMPs.

As discussed in Section 3.1.2, the test dataset MSE values are most indicative of a model's real-world performance. Furthermore, Sections 3.1.2 and 3.2 showed that in general, the RR and NN with PBMPs yielded similar or lower MSE values than their counterparts without PBMPs. As such, for each DOF (row), the y-axis bounds of Fig. 13 were selected such that all test data MSE values (from AR Test and SR Test) were visible for the RR and NN with PBMPs. While the authors felt that it was important to include the uncorrected PBMPs and the RR and NN without PBMPs to highlight bins where these models may have performed adequately or better than those with PBMPs, they are not the focal point of this discussion. Although train MSE is not a strong indicator of a model's real-world performance, the results for the AR Train dataset were included to indicate bins in which the models may have learned better than others. Similar to Figs. 5 and 6, for magnitude bins with less than 100 samples, the number of samples is provided above the bar in Fig. 13. Because train data sparsity can yield poor generalization, causing inadequate performance for a test dataset, the vertical dotted lines in each plot (for all three datasets) indicate the heave, pitch, or roll bin boundary where the number of AR Train samples dropped below 100. Note, however, that the selection of 100 was somewhat arbitrary and primarily motivated by plot legibility. Therefore, the insights provided by these dotted vertical lines are limited because model performance is tied to similarity between the train and test data for a given bin rather than the number of samples.

Fig. 13 shows that typically, greater test MSE values occurred at greater motion magnitudes for all DOFs and predictive models. This increase in MSE was expected for two reasons. First, nonlinearities in vessel response typically grow as magnitudes increase, making accurate prediction more difficult. Second, the train data became more sparse as magnitudes increased, decreasing the likelihood of a test data sample having a similar sample in AR Train. This data imbalance may have also impacted model training. For example, the larger number of samples in lower magnitude bins may have encouraged the model training process to focus on those bins more heavily, as suggested by the lesser AR Train MSE values in lower magnitude bins of Fig. 13. As such, the data imbalance may have detracted from model performance in higher motion magnitude bins. Furthermore, Fig. 13 shows that the differences

in MSE values between models also typically grew as response magnitudes increased. This increase was also anticipated, especially when considering the expected advantages of models with PBMPs in sparse training data bins.

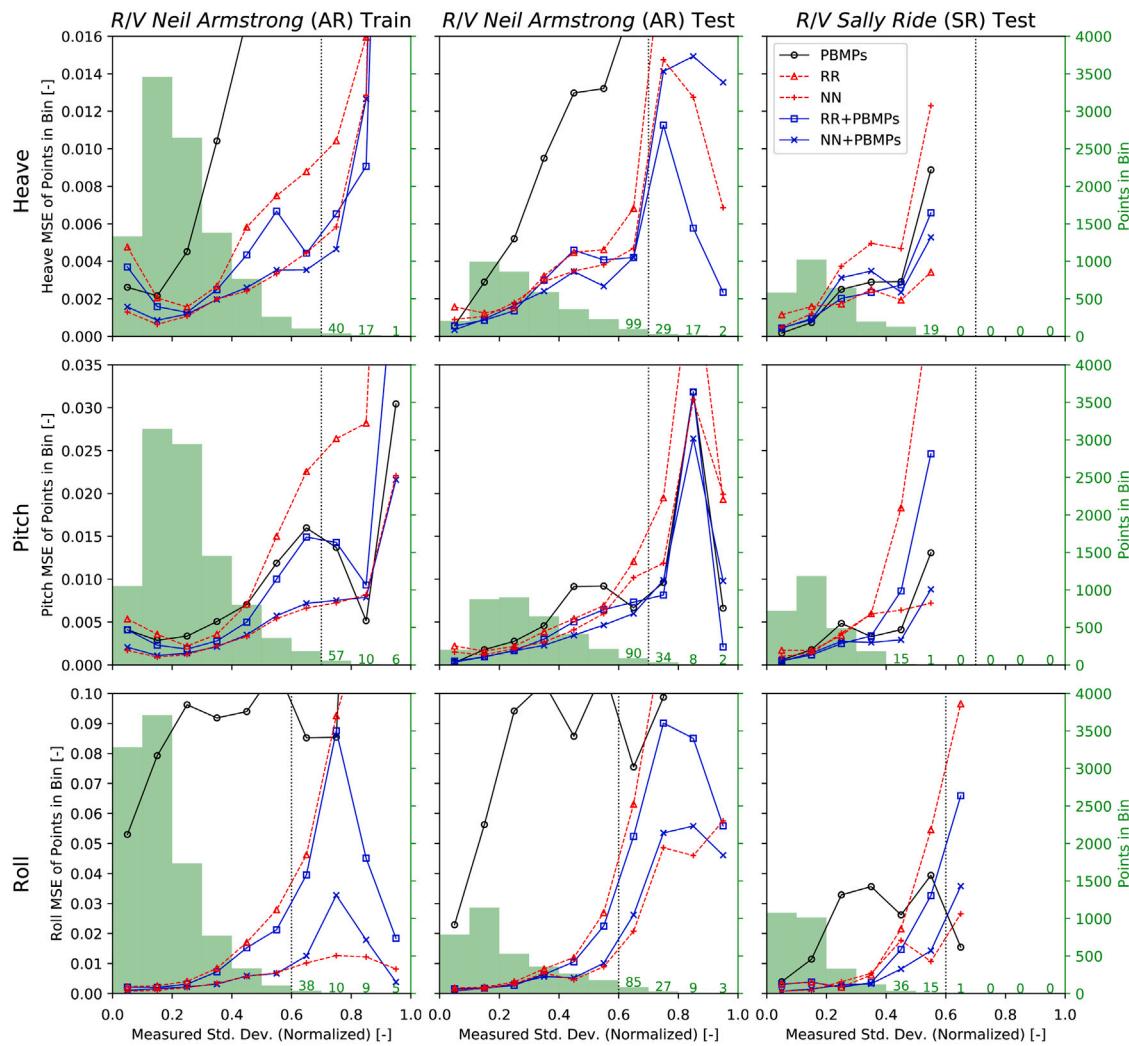
In Fig. 13, for AR Test roll, the NNs with and without PBMPs yielded similar performance in most magnitude bins and outperformed the other models. As mentioned in the discussion of Table 8, this similarity may have resulted from the poor quality of the uncorrected roll PBMPs. Additionally, the advantages of the NNs over the RR models for roll prediction were likely due to the ability of NNs to model nonlinearities. For SR Test, although the results of Table 9 implied that the NN with PBMPs notably outperformed the NN without PBMPs, Fig. 13 suggests that the value added by PBMPs was only evident in a few bins. While the uncorrected roll PBMPs were seemingly better for SR Test than AR Test, the benefits of these PBMPs for roll were relatively insignificant. The authors believe that improved roll PBMPs would improve the predictive capabilities of the NN with PBMPs.

For AR Test heave and pitch in Fig. 13, the top performer for each magnitude bin was either the RR with PBMPs or NN with PBMPs. Furthermore, for some cases, such as AR Test pitch, the RR and NN without PBMPs performed significantly worse than the models with PBMPs and even the uncorrected PBMPs as magnitudes increased and, consequentially, train data were more sparse. This poor performance is an example cause of the mistrust that many engineers have in data-driven models that ignore physics. Note that there were a few select bins in which a model without PBMPs performed better than those with PBMPs (e.g., higher magnitudes for SR Test heave and pitch). Overall, for heave and pitch, the inclusion of PBMPs appears to have been beneficial. However, Fig. 13 does suggest that any existent advantages of the NN with PBMPs over the RR with PBMPs were limited for heave and pitch.

#### 3.3.2. Motion MSE values vs. Mean wave period

As shown in Fig. 13, although some of the greater magnitude heave, pitch, and roll bins in AR Train had few data samples, none of the motion magnitudes in AR Test or SR Test fell outside of the AR Train bounds. As mentioned in Section 2.11, this was also the case for significant wave height, but not mean wave period ( $T_m$ ). To examine the models' performances on test data outside of the AR Train bounds, Fig. 14 follows the same general layout as Fig. 13; however, all data in the plots of Fig. 14 are binned by normalized mean wave period ( $T_m$ ). Therefore, for a given dataset (column) the histograms shown are identical for heave, pitch, and roll. The y-axis bounds were once again selected such that all AR Test and SR Test data samples for the NN and RR with PBMPs were included. Two dotted vertical lines indicating bins with less than 100 AR Train samples are included in the plots of Fig. 14 because for mean wave period, test data extended to values above and below the AR Train range. Although AR Test is included in Fig. 14 for reference, the samples lying outside of the mean wave period range of AR Train were limited. Therefore, SR Test is the focal point of the following discussion. Furthermore, it is difficult to draw conclusions from test data bins with very few samples (i.e., mean wave period bins 0.2–0.3 and 1.1–1.2 for SR Test) given the large number of other input variables. As such, the following discussion will focus on the normalized mean wave period bins 0.9–1.0 and 1.0–1.1. As shown by the histogram annotations, the 0.9–1.0 bin had just 40 AR Train samples, and the 1.0–1.1 bin had 0 AR Train samples.

For SR Test roll in the 0.9–1.0 bin of Fig. 14, the NN with PBMPs performed notably better than any other model. For SR Test roll in the 1.0–1.1 bin, the NN with and without PBMPs were the top performers and yielded nearly identical MSE values. Therefore, for SR Test roll, the PBMPs may have been beneficial in combination with the 40 AR Train samples in the 0.9–1.0 bin, but did not have a notable effect in the 1.0–1.1 bin where 0 AR Train samples were available. For SR Test heave and pitch in the 0.9–1.0 bin, the RR with PBMPs yielded the lowest MSE. For this bin and heave, specifically, the RR with PBMPs was the only



**Fig. 13.** Plots of heave, pitch, and roll MSE as a function of their respective normalized motion measurements. Each plot shows a histogram of the number of samples in 10 different measured motion bins, identified by the right axis. The MSE values, identified by the left axis, calculated from the samples in each bin for a given model or the uncorrected PBMPs are overlaid on these histograms.

model with lower SR Test MSE than the uncorrected PBMPs, and the improvement was slight. For heave in the 1.0–1.1 bin, the uncorrected PBMPs produced notably lower MSE than any of the models. The uncorrected PBMPs were also the best performer for this bin in pitch, with the NN with PBMPs producing similar results. These results suggest that the inclusion of PBMPs was not definitively beneficial for extrapolated input variables when the *Neil Armstrong* trained models were applied to the *Sally Ride*. This conclusion is not particularly surprising given the potential mean loading and operational environment differences between the two vessels. Perhaps a train dataset formed by data from both the *Neil Armstrong* and *Sally Ride* would be beneficial for future response predictions on both ships.

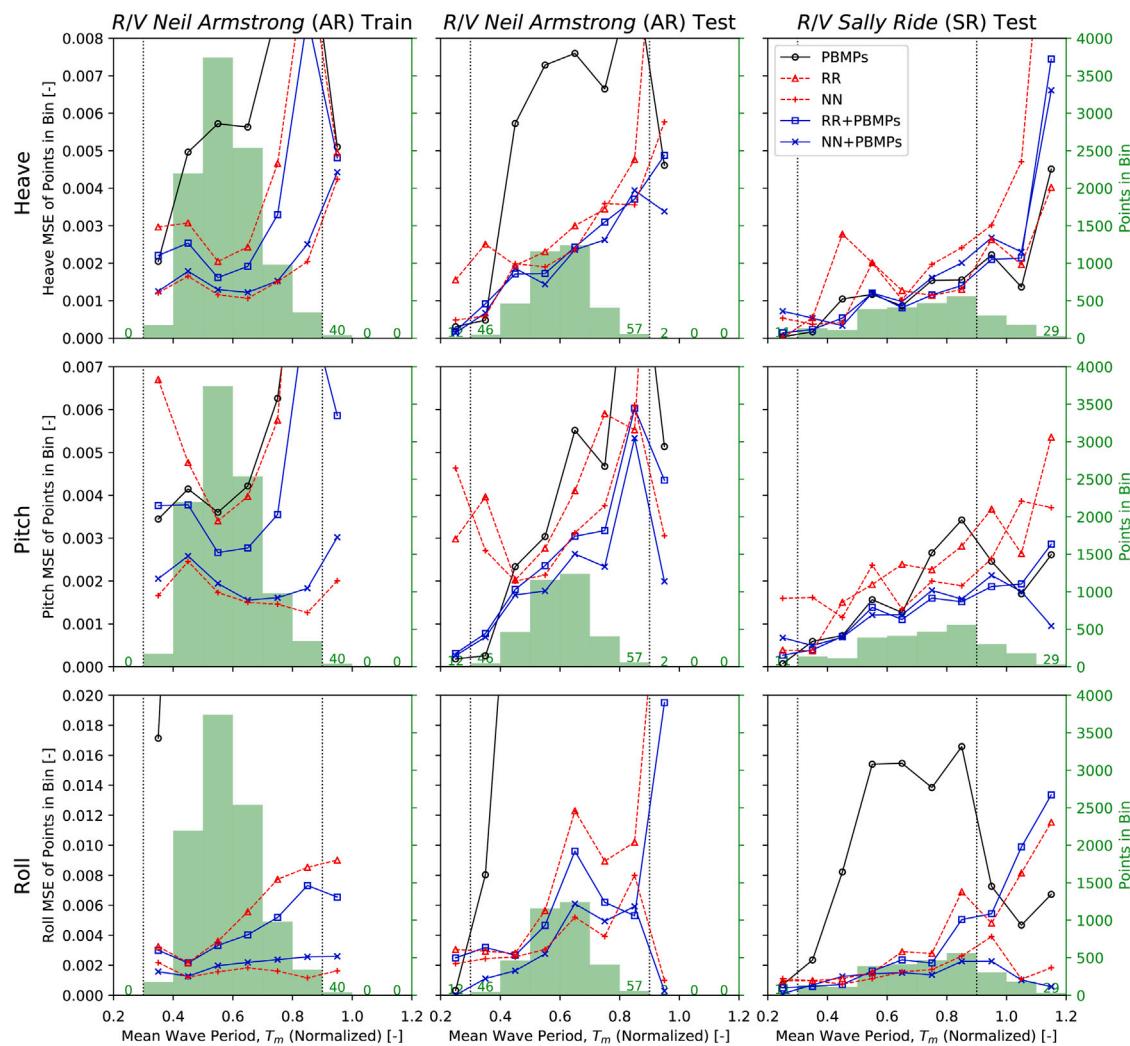
#### 3.4. AR40 demonstration using wave forecast data

As mentioned in Section 2.5, all of the CMEMS wave data used for AR Train, AR Test, and SR Test data were hindcast rather than forecast. In contrast, this section demonstrates the use of wave forecast data in the data-driven models, which emulates this technology's application in real-world operation of a ship for motion forecasting. As described in Section 2.5 and given in Table 3, AR40 yielded 132 30-minute windows for this analysis with five different wave forecast time horizons. Given the general benefits of including PBMPs discussed previously, the RR and NN models without PBMPs were omitted from the following.

**Fig. 15** shows the heave, pitch, and roll MSE values calculated from all 132 samples using wave data from the five different forecast time horizons defined in Section 2.5 and the hindcast wave data. Two plots with different zooms are included for roll due to the drastic improvement of both data-driven models upon the uncorrected PBMPs. For reference, the vertical dashed line in each plot identifies the transition from forecast to hindcast data, and the horizontal lines show the maximum MSE value over the six time horizons for each model.

In Fig. 15, for heave and pitch, the NN with PBMPs yielded the lowest MSE for any given time horizon, followed closely by the RR with PBMPs in most instances. Furthermore, as expected, the use of hindcast data in the models yielded the lowest heave and pitch MSE for each model. With the exception of 72–48 h for pitch, the heave and pitch MSE values for the two data-driven models decreased as the time horizons approached the time of interest (i.e., from left to right). For roll, these trends were not evident. In fact, the hindcast wave data yielded the greatest MSE for both the RR and NN with PBMPs. Note, however, that all the results shown in Fig. 15 are extremely dependent upon the operating conditions and wave data available for this specific cruise; therefore, this discussion serves as a demonstration only, rather than an analysis from which broad conclusions regarding wave forecast reliability could be drawn.

In this scenario, given the vessel's expected location, speed, and heading, an operator could forecast how multidirectional wave conditions will affect his or her vessel up to five days in the future. These



**Fig. 14.** Plots of heave, pitch, and roll MSE as a function of normalized mean wave period ( $T_m$ ). Each plot shows a histogram of the number of samples in 10 different mean wave period bins, identified by the right axis. The MSE values, identified by the left axis, calculated from the samples in each bin for a given model or the uncorrected PBMPs are overlaid on these histograms.

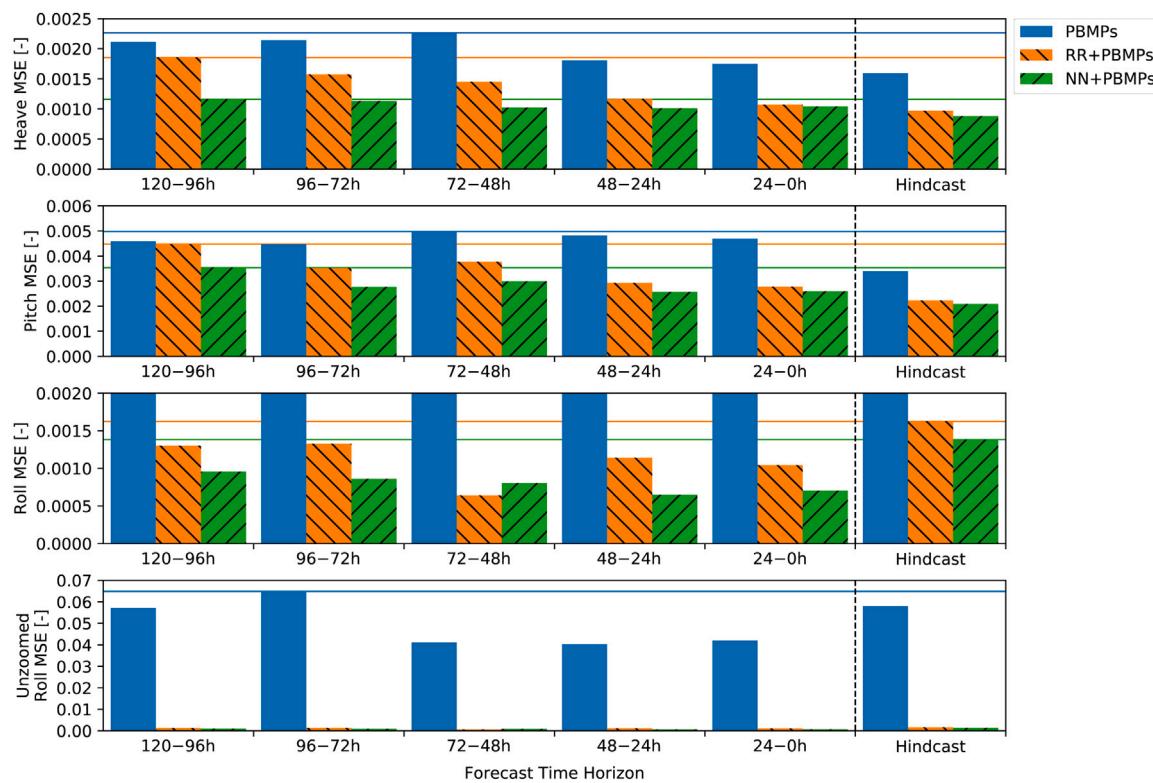
forecast motions would not only be informed by physics, but also by the vessel's past observations. As such, the operator could gain a better understanding of the conditions to come than that achieved using wave parameters alone. The operator could then continue to monitor these forecasts as the time of interest approached, as demonstrated in Fig. 15, and adjust course and/or the vessel's estimated time of arrival if necessary.

#### 4. Conclusions

In this work and the preliminary discussion presented in Schirmann et al. (2020), linear RR and nonlinear NN models were demonstrated for vessel heave, pitch, and roll prediction using multidirectional wave data. The benefits of retaining physics-based information in data-driven models were also addressed through comparison of model performance with and without inclusion of PBMPs as input variables. The models were trained and tested using data from the *R/V Neil Armstrong*. While both the RR and NN approaches effectively reduced the AR Test MSE relative to the uncorrected PBMPs, under the same PBMPs assumption (with or without PBMPs), the NNs outperformed the RRs. This improved performance was attributed to the ability of the NNs to model nonlinear relationships between input variables. However, given the relatively strong performance of the linear RR approach over the underlying PBMPs and that RR is not a black box approach like NN,

RR may be preferable in some instances. For AR Test heave and pitch, which had PBMPs of much higher quality than roll, the inclusion of PBMPs benefited model performance. In fact, the second most effective model behind the NN with PBMPs was the RR with PBMPs, showing that a linear model (RR) with PBMPs could outperform a nonlinear model (NN) without PBMPs. For roll, the low-quality PBMPs did not show any benefit for the NN and limited benefit for the RR approach. In addition to improving the underlying roll PBMPs with an updated GM value (when available) for future studies, given the complexity of roll prediction relative to heave and pitch, a unique structure for roll may be worth investigating. For example, in addition to predicted amplitudes, higher-order moments from the seakeeping code could be used as additional input variables to the data-driven models.

These *Neil Armstrong* trained models were then applied to a test dataset from the *R/V Sally Ride* to test their versatility. In this case, the best approaches for heave and pitch were the RR models with PBMPs, and the best model for roll was the NN with PBMPs. Excluding the NN with PBMPs for heave, all models that included PBMPs reduced the heave, pitch, and roll MSE values of the test dataset relative to the uncorrected PBMPs. The exception for heave may have been due to the relatively high quality of the uncorrected heave PBMPs for the SR Test dataset. Although the relative performance comparisons between models for SR Test did not perfectly match the findings from AR Test, the results did show that in each DOF, the RR or NN with PBMPs



**Fig. 15.** Heave, pitch, and roll MSE calculated from all 132 AR40 samples using the wave data from the five different forecast time horizons and the hindcast wave data.

outperformed their respective counterparts without PBMPs. This result highlights the potential benefits of including physics-based information in data-driven models.

A more in-depth analysis of the motions MSE values as a function of measured motion magnitudes further supported that quality PBMPs, as in the case of heave and pitch, typically improved model performance. Additionally, these comparison demonstrated that in some cases, a lack of PBMPs can cause data-driven models to perform more poorly than the uncorrected PBMPs. This finding reinforces the concerns of some regarding the reliability of data-driven models that abandon physics. An additional analysis of motions MSE values as a function of mean period in train data sparse regions of the input space did not show definitive benefits from including PBMPs when the *Neil Armstrong* trained model was applied to *Sally Ride* data. Nevertheless, training RR and NN models with PBMPs using data from both ships may prove more effective for future response predictions on both ships.

A demonstration of the RR and NN models with PBMPs using forecast, rather than hindcast, wave data was performed for a single cruise of the *Neil Armstrong*. The potential use of these models for real-world operational guidance was highlighted through comparison of model performance ranging from a five-day forecast to a hindcast analysis. Providing vessel operators with these motion predictions may furnish a better understanding of a multidirectional wave forecast's meaning. For example, these predictive models could be used by owners and operators for route and heading selection, leading to safer and more efficient vessel operation. Note that all data-driven models have certain limitations given that they are based on imperfect data from real-world observations; as such, in a real-world implementation, rather than blindly trusting these models, users should cautiously employ them in conjunction with physics-based models and knowledge from their own experiences to make educated decisions.

Using data from over 16,000 30-minute windows measured aboard two operational sister ships, this work demonstrated the efficacy of data-driven models for vessel motion prediction using wave model

data. Furthermore, this work highlighted the importance of including physics-based information, even given an abundance of train data.

#### CRediT authorship contribution statement

**Matthew L. Schirrmann:** Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing – original draft, Visualization. **Matthew D. Collette:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition. **James W. Gose:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgments

The authors wish to thank Dr. Woei-Min Lin, ONR Code 331, for supporting this work under contract N00014-17-1-2982. The authors would also like to thank CEMES for providing the wave data used in this project, as well as WHOI, SIO, the R2R team, and Kongsberg for providing onboard data and technical support. This research was supported in part through computational resources and services provided by Advanced Research Computing at the University of Michigan, Ann Arbor.

#### Appendix A. Table of train cruises

See Table A.10.

**Table A.10**

Summary of the 42 *Neil Armstrong* cruises considered in the AR Train dataset, which are mapped in Fig. 3. Negative longitude values indicate those west of 0° (e.g., -50° is equivalent to 50° W in Fig. 3).

| Start month | Cruise ID | Lat <sub>N</sub> | Lat <sub>S</sub> | Lon <sub>E</sub> | Lon <sub>W</sub> | $H_{s50}$ [m] | $H_{s100}$ [m] | $T_{m50}$ [s] | $T_{m100}$ [s] | Windows |
|-------------|-----------|------------------|------------------|------------------|------------------|---------------|----------------|---------------|----------------|---------|
| 05/2016     | AR3       | 41.3             | 39.8             | -70.5            | -71.3            | 1.78          | 2.91           | 6.65          | 8.40           | 197     |
|             | AR4-A     | 41.3             | 39.9             | -70.8            | -70.9            | 0.97          | 1.50           | 7.35          | 9.79           | 78      |
|             | AR4-B     | 41.3             | 39.8             | -70.3            | -70.9            | 1.04          | 2.70           | 7.26          | 8.19           | 152     |
|             | AR4-C     | 41.3             | 39.8             | -70.5            | -71.0            | 1.13          | 1.41           | 6.02          | 6.98           | 121     |
| 06/2016     | AR5       | 41.3             | 39.2             | -69.5            | -70.9            | 1.47          | 2.35           | 6.04          | 6.65           | 96      |
|             | AR6       | 41.3             | 39.8             | -70.6            | -70.9            | 1.34          | 1.72           | 7.53          | 8.12           | 17      |
| 08/2016     | AR7-02    | 64.1             | 47.7             | -23.2            | -52.5            | 1.63          | 3.33           | 7.36          | 10.22          | 820     |
| 09/2016     | AR7-03    | 46.4             | 41.4             | -53.3            | -70.3            | 1.34          | 2.15           | 6.44          | 8.11           | 149     |
|             | AR8A      | 41.3             | 39.8             | -70.6            | -71.1            | 1.58          | 3.80           | 7.14          | 8.24           | 164     |
| 10/2016     | AR8B      | 41.3             | 39.8             | -70.4            | -71.1            | 1.65          | 3.10           | 8.83          | 9.74           | 151     |
| 11/2016     | AR9-01    | 41.2             | 32.7             | -71.0            | -79.5            | 0.96          | 2.68           | 6.88          | 8.85           | 115     |
| 02/2017     | AR9-03    | 41.3             | 32.9             | -70.9            | -78.8            | 1.11          | 1.83           | 6.05          | 7.56           | 123     |
| 03/2017     | AR10      | 41.3             | 40.2             | -70.4            | -71.3            | 2.18          | 3.24           | 6.32          | 7.37           | 92      |
|             | AR11      | 41.3             | 40.4             | -70.5            | -71.3            | 2.03          | 3.27           | 6.01          | 7.65           | 46      |
|             | AR12      | 41.3             | 40.4             | -70.4            | -71.4            | 1.46          | 3.14           | 6.38          | 8.69           | 88      |
|             | AR13      | 41.3             | 40.3             | -70.4            | -71.2            | 1.62          | 2.34           | 6.82          | 9.73           | 82      |
| 04/2017     | AR14      | 41.3             | 40.5             | -70.5            | -71.0            | 1.70          | 3.35           | 7.08          | 7.75           | 50      |
|             | AR15      | 41.3             | 34.6             | -70.7            | -76.0            | 1.80          | 4.43           | 6.70          | 9.29           | 326     |
| 05/2017     | AR16      | 41.3             | 29.0             | -64.1            | -71.5            | 1.64          | 4.68           | 7.82          | 9.37           | 584     |
|             | AR17-01   | 41.3             | 40.1             | -71.1            | -73.4            | 0.92          | 1.00           | 6.45          | 6.81           | 24      |
|             | AR17-02   | 41.4             | 40.4             | -70.8            | -73.7            | 1.59          | 1.81           | 7.59          | 7.73           | 27      |
|             | AR18A     | 41.3             | 39.9             | -70.8            | -70.9            | 1.27          | 1.62           | 8.28          | 8.94           | 85      |
| 06/2017     | AR18B     | 41.2             | 39.8             | -70.6            | -71.1            | 1.58          | 2.55           | 6.82          | 8.28           | 126     |
|             | AR18C     | 41.3             | 40.1             | -70.6            | -70.9            | 1.03          | 2.18           | 7.18          | 8.50           | 72      |
|             | AR19      | 41.3             | 41.2             | -71.0            | -71.3            | 1.68          | 1.85           | 5.46          | 5.59           | 10      |
| 07/2017     | AR20      | 41.3             | 39.6             | -70.7            | -71.4            | 1.24          | 1.74           | 6.73          | 8.32           | 146     |
|             | AR21      | 60.1             | 41.4             | -39.1            | -70.3            | 1.37          | 3.01           | 7.37          | 10.15          | 1,047   |
| 09/2017     | AR22      | 41.3             | 40.0             | -70.8            | -71.0            | 1.23          | 1.54           | 5.07          | 7.22           | 19      |
|             | AR23-01   | 41.2             | 34.2             | -26.0            | -70.9            | 2.04          | 3.22           | 10.02         | 11.82          | 736     |
| 10/2017     | AR23-02   | 41.1             | 34.3             | -25.8            | -70.8            | 1.66          | 3.84           | 8.24          | 12.85          | 406     |
|             | AR24-A    | 41.2             | 39.9             | -70.8            | -70.9            | 2.24          | 3.97           | 7.05          | 8.93           | 57      |
|             | AR24-B    | 41.4             | 39.8             | -70.6            | -70.9            | 1.94          | 4.26           | 7.82          | 11.57          | 143     |
| 11/2017     | AR24-C    | 41.3             | 39.9             | -70.8            | -71.0            | 1.93          | 3.32           | 7.32          | 10.07          | 153     |
|             | AR25      | 41.4             | 29.1             | -70.8            | -80.0            | 1.50          | 3.05           | 7.14          | 9.54           | 533     |
| 01/2018     | AR26      | 41.3             | 34.5             | -70.9            | -76.1            | 1.73          | 3.96           | 6.81          | 9.51           | 299     |
| 03/2018     | AR28-A    | 41.4             | 39.9             | -70.8            | -70.9            | 1.46          | 4.29           | 6.87          | 12.88          | 219     |
| 04/2018     | AR28-B    | 41.4             | 39.8             | -70.8            | -70.9            | 1.53          | 2.99           | 6.87          | 8.79           | 146     |
|             | AR29      | 41.3             | 39.6             | -70.3            | -71.1            | 1.63          | 4.60           | 7.12          | 9.96           | 110     |
| 05/2018     | AR30-01   | 64.0             | 41.4             | -23.1            | -70.2            | 2.62          | 5.37           | 8.12          | 10.04          | 388     |
|             | AR30-02   | 64.0             | 58.0             | -22.2            | -28.0            | 3.31          | 6.41           | 9.02          | 13.07          | 389     |
| 06/2018     | AR30-03   | 64.1             | 59.7             | -23.1            | -41.3            | 2.00          | 4.19           | 7.75          | 10.38          | 603     |
| 07/2018     | AR30-04   | 64.0             | 56.9             | -8.4             | -36.5            | 1.66          | 3.24           | 7.08          | 9.98           | 819     |
|             |           | 64.1             | 29.0             | -8.4             | -80.0            | 1.67          | 6.41           | 7.44          | 13.07          | 10,008  |

## Appendix B. Tables of test cruises

See Tables B.11 and B.12.

Table B.11

Summary of the 7 *Neil Armstrong* cruises considered in the AR Test dataset, which are mapped in Fig. 3. Negative longitude values indicate those west of  $0^{\circ}$  (e.g.,  $-50^{\circ}$  is equivalent to  $50^{\circ}$  W in Fig. 3).

| Start month | Cruise ID | Lat <sub>N</sub> | Lat <sub>S</sub> | Lon <sub>E</sub> | Lon <sub>W</sub> | H <sub>s50</sub><br>[m] | H <sub>s100</sub><br>[m] | T <sub>m50</sub><br>[s] | T <sub>m100</sub><br>[s] | Windows |
|-------------|-----------|------------------|------------------|------------------|------------------|-------------------------|--------------------------|-------------------------|--------------------------|---------|
| 08/2018     | AR30-05   | 66.8             | 63.1             | 0.3              | -25.0            | 1.38                    | 2.46                     | 7.00                    | 9.76                     | 431     |
|             | AR30-06   | 66.9             | 58.6             | -22.5            | -48.5            | 2.19                    | 5.83                     | 7.68                    | 12.02                    | 620     |
| 10/2018     | AR30-07   | 64.0             | 41.4             | -23.1            | -70.3            | 3.04                    | 5.84                     | 8.85                    | 10.90                    | 365     |
| 04/2019     | AR35-01   | 64.0             | 41.4             | -23.1            | -70.2            | 2.69                    | 5.19                     | 7.97                    | 9.30                     | 467     |
| 05/2019     | AR35-02   | 63.9             | 58.9             | -21.0            | -23.2            | 1.28                    | 4.14                     | 8.13                    | 10.84                    | 259     |
|             | AR35-03   | 64.0             | 57.1             | -21.0            | -29.5            | 2.00                    | 3.89                     | 7.83                    | 10.63                    | 535     |
| 08/2019     | AR35-05   | 64.1             | 41.4             | -23.1            | -70.3            | 1.36                    | 2.89                     | 7.99                    | 10.38                    | 707     |
|             |           | 66.9             | 41.4             | 0.3              | -70.3            | 1.86                    | 5.84                     | 7.86                    | 12.02                    | 3,384   |

**Table B.12**

Summary of the 18 *Sally Ride* cruises considered in the SR Test dataset, which are mapped in Fig. 4. Negative longitude values indicate those west of 0°, but still east of 180°. For example, -170° is equivalent to 170° W, and 170° is equivalent to 170° E in Fig. 4).

| Start month | Cruise ID | Lat <sub>N</sub> | Lat <sub>S</sub> | Lon <sub>E</sub> | Lon <sub>W</sub> | $H_{s50}$<br>[m] | $H_{s100}$<br>[m] | $T_{m50}$<br>[s] | $T_{m100}$<br>[s] | Windows |
|-------------|-----------|------------------|------------------|------------------|------------------|------------------|-------------------|------------------|-------------------|---------|
| 01/2017     | SR1701    | 33.0             | 32.6             | -118.2           | -119.1           | 0.93             | 1.05              | 9.99             | 11.84             | 32      |
| 03/2017     | SR1706    | 32.7             | 32.7             | -118.9           | -119.1           | 1.37             | 1.45              | 11.09            | 11.29             | 2       |
| 03/2017     | SR1708    | 32.9             | 32.9             | -117.5           | -117.5           | 0.99             | 0.99              | 11.66            | 11.66             | 1       |
| 03/2017     | SR1709    | 33.3             | 32.7             | -117.5           | -119.0           | 0.99             | 2.46              | 10.49            | 12.22             | 29      |
| 09/2017     | SR1714    | 35.0             | 32.7             | -117.5           | -120.8           | 1.51             | 2.62              | 7.19             | 10.64             | 45      |
| 10/2017     | SR1716    | 34.9             | 32.7             | -117.5           | -120.9           | 1.04             | 2.81              | 10.90            | 14.94             | 61      |
| 11/2017     | SR1717    | 34.9             | 29.9             | -117.4           | -124.3           | 1.55             | 2.58              | 10.63            | 12.56             | 263     |
| 12/2017     | SR1718    | 34.3             | 32.7             | -117.5           | -120.0           | 0.90             | 2.00              | 10.04            | 12.48             | 46      |
| 03/2018     | SR1804    | 32.7             | 32.7             | -117.5           | -117.5           | 0.99             | 0.99              | 10.28            | 10.28             | 1       |
| 03/2018     | SR1805    | 22.3             | 10.1             | -102.4           | -115.3           | 1.78             | 2.42              | 10.73            | 13.17             | 202     |
| 06/2018     | SR1809    | 44.3             | 32.7             | -117.5           | -124.8           | 1.47             | 2.37              | 8.32             | 12.81             | 95      |
| 07/2018     | SR1811    | 50.3             | 44.7             | -124.5           | -145.2           | 1.38             | 2.26              | 8.21             | 10.91             | 309     |
| 10/2018     | SR1814    | 41.7             | 33.2             | -118.8           | -124.7           | 2.08             | 2.30              | 11.09            | 12.31             | 55      |
| 10/2018     | SR1815    | 35.0             | 29.9             | -117.7           | -124.3           | 1.86             | 2.90              | 10.62            | 14.70             | 246     |
| 03/2019     | SR1904    | 33.8             | 32.7             | -117.6           | -120.3           | 1.49             | 2.05              | 8.50             | 9.31              | 24      |
| 03/2019     | SR1906    | 32.6             | 22.9             | -117.4           | -157.4           | 2.62             | 3.87              | 11.95            | 14.84             | 308     |
| 05/2019     | SR1908    | 13.4             | -6.4             | 144.5            | 94.1             | 0.90             | 1.89              | 7.56             | 14.48             | 684     |
| 10/2019     | SR1914    | 24.2             | 7.3              | 134.8            | 122.9            | 0.99             | 2.43              | 8.76             | 12.98             | 189     |
|             |           | 50.3             | -6.4             | -102.4           | 94.1             | 1.43             | 3.87              | 9.94             | 14.94             | 2,592   |

## References

- Beck, R.F., Troesch, A.W., 1990. Student's Documentation and User's Manual for the Computer Program SHIPMO.BM. University of Michigan Department of Naval Architecture and Marine Engineering.
- Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer, New York.
- Bremer, K.S., 2018. Using neural networks to predict the response of a floating structure. (Master's thesis). Norwegian University of Science and Technology.
- Chollet, F., et al., 2015. Keras. <https://keras.io>.
- De Masi, G., Gaggiotti, F., Bruschi, R., Venturi, M., 2011. Ship motion prediction by radial basis neural networks. In: 2011 IEEE Workshop on Hybrid Intelligent Models and Applications. IEEE, pp. 28–32.
- Deng, Y., Feng, W., Xu, S., Chen, X., Wang, B., 2020. A novel approach for motion predictions of a semi-submersible platform with neural network. J. Mar. Sci. Technol. 26, 883–895.
- Duz, B., Mak, B., Hageman, R., Grasso, N., 2019. Real time estimation of local wave characteristics from ship motions using artificial neural networks. In: Practical Design of Ships and Other Floating Structures, Proceedings of The 14th International Symposium (PRADS 2019). Springer, Yokohama, Japan.
- Fernandez, E., Aouf, L., 2018. Product user manual for global ocean waves analysis and forecasting product global analysis forecast wav\_001\_027. Tech. Rep. 1.0, Copernicus Marine Environment Monitoring Service, pp. 1–12.
- Gougiolidis, G., 2008. The utilization of artificial neural networks in marine applications: An overview. Naval Eng. J. 120 (3), 19–26.
- Himeno, Y., 1981. Prediction of Ship Roll Damping - State of the Art. University of Michigan Department of Naval Architecture and Marine Engineering, (Report).
- Huang, B.-G., Zou, Z.-J., 2016. Short-term prediction of ship pitching motion based on artificial neural networks. In: Proceedings of The ASME 2016 35th International Conference on Ocean, Offshore and Arctic Engineering. Busan, South Korea, pp. 1–5.
- Jiang, L., Signal, S., Jeffries, B., Earley, B., Junghans, K., Hess, D., Faller, W., 2020. A hydrodynamic digital twin concept for underwater vehicles. In: 33rd Symposium on Naval Hydrodynamics (33rd SNH). Osaka, Japan.
- Kapusuzoglu, B., Mahadevan, S., 2020. Physics-informed and hybrid machine learning in additive manufacturing: Application to fused filament fabrication. JOM, J. Minerals Metals Materials Soc. 72 (12), 4695–4705.
- Kawai, T., Kawamura, Y., Okada, T., Mitsuyuki, T., Chen, X., 2021. Sea state estimation using monitoring data by convolutional neural network (CNN). J. Mar. Sci. Technol. 26, 947–962.
- Khan, A., Bil, C., Marion, K.E., 2005. Theory and application of artificial neural networks for the real time prediction of ship motion. In: Knowledge-Based Intelligent Information and Engineering Systems (KES 2005). Springer Berlin Heidelberg, pp. 1064–1069.
- Li, G., Kawan, B., Wang, H., Zhang, H., 2017. Neural-network-based modelling and analysis for time series prediction of ship motion. Ship Technol. Res. 64 (1), 30–39.
- Lloyd, A., 1989. Seakeeping: Ship Behaviour in Rough Weather. Ellis Horwood Ltd., pp. 160–164.
- Najafi, A., Nowruzi, H., Ghassemi, H., 2018. Performance prediction of hydrofoil-supported catamarans using experiment and ANNs. Appl. Ocean Res. 75, 66–84.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.
- Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. 378, 686–707.
- Russell, S., Norvig, P., 2010. Artificial Intelligence: A Modern Approach, third ed. Prentice Hall.
- Schirrmann, M.L., Chen, T., Collette, M.D., Gose, J.W., 2019. Linking seakeeping performance predictions with onboard measurements for surface platform digital twins. In: Practical Design of Ships and Other Floating Structures, Proceedings of The 14th International Symposium (PRADS 2019). Springer, Yokohama, Japan.
- Schirrmann, M.L., Collette, M.D., Gose, J.W., 2020. Improved vessel motion predictions using full-scale measurements and data-driven models. In: 33rd Symposium on Naval Hydrodynamics (33rd SNH). Osaka, Japan.
- Subramanian, A., Mahadevan, S., 2019. Bayesian estimation of discrepancy in dynamics model prediction. Mech. Syst. Signal Process. 123, 351–368.
- US Navy Office of Information, 2018. Auxiliary general purpose oceanographic research vessel - AGOR. <https://www.navy.mil/Resources/Fact-Files/Display-FactFiles/Article/2222992/auxiliary-general-purpose-oceanographic-research-vessel-agor/>.
- Weymouth, G.D., 2019. Roll damping predictions using physics-based machine learning. In: Conference: Conference on Computer Applications and Information Technology in the Maritime Industries (COMPIT). Tullamore, Ireland.
- Weymouth, G.D., Yue, D.K.P., 2013. Physics-based learning models for ship hydrodynamics. J. Ship Res. 57 (1), 1–12.
- Willcox, K.E., 2019. Predictive data science for physical systems: From model reduction to scientific machine learning. invited presentation at SC19, Denver, CO.
- Wu, M., Stefanakos, C., Gao, Z., 2020. Multi-step-ahead forecasting of wave conditions based on a physics-based machine learning (PBML) model for marine operations. J. Mar. Sci. Eng. 8 (12), 1–24.
- Xing, Z., McCue, L., 2010. Modeling ship equations of roll motion using neural networks. Naval Eng. J. 122 (3), 49–60.