

# An integrated framework that combines machine learning and numerical models to improve wave-condition forecasts

Fearghal O’Donncha<sup>a,\*</sup>, Yushan Zhang<sup>b</sup>, Bei Chen<sup>a</sup>, Scott C. James<sup>c</sup>

<sup>a</sup>*IBM Research - Ireland, Dublin*

<sup>b</sup>*University of Notre Dame*

<sup>c</sup>*Baylor University*

---

## Abstract

This study investigates near-shore circulation and wave characteristics applied to a case-study site in Monterey Bay, California. We integrate physics-based models to resolve wave conditions together with a machine-learning algorithm that combines forecasts from multiple, independent models into a single “best-estimate” prediction of the true state. The Simulating WAVes Nearshore (SWAN) physics-based model is used to compute wind-augmented waves. Ensembles are developed based on multiple simulations perturbing data input to the model. A learning-aggregation technique uses historical observations and model forecasts to calculate a weight for each ensemble member. We compare the weighted ensemble predictions with measured data to evaluate performance against present state-of-the-art. Finally, we discuss how this framework that integrates data-driven and physics-based approaches can outperform either technique in isolation.

---

## 1. Introduction

Physics-based numerical models are defined by: (1) the physical formulation, (2) numerical discretization, and (3) input data driving the simulations. Typically, all three involve some degree of uncertainty. Wave modeling

---

\*Corresponding author

*Email address:* feardonn@ie.ibm.com (Fearghal O’Donncha )

5 and predictions result from solution of the spectral-action balance equations,  
6 which are based on an approximation of reality derived from an incomplete  
7 data set (Komen et al., 1996; Mei et al., 1989). Similarly, numerical dis-  
8 cretization approximates the solution to these equations with accuracy de-  
9 pendent upon spatial resolution and time steps. In addition, wave-condition  
10 forecasting involves multicomponent input data such as bathymetry, ocean  
11 characteristics, and meteorological information. Rogers et al. (2005) ob-  
12 served that in global operational nowcast/forecast systems, wind forcing is  
13 a dominant source of error (in global wave models the only time-varying  
14 inputs are typically wind forcing). Here, we combine ensemble-forecasting  
15 and machine-learning techniques to: (1) investigate uncertainty from an ad-  
16 vanced wave-modeling package and (2) generate a forecast that is better than  
17 the best individual model prediction.

18 Ensemble-based machine-learning approaches (Mallet et al., 2009) com-  
19 prise aggregate ensemble predictions based on multiple simulations where  
20 anything from physical parameterizations, numerical discretization, or in-  
21 put data are perturbed. The learning-aggregation technique presented here,  
22 makes use of historical observations and model forecasts to produce a weight  
23 for each model. A linear or convex (i.e., where weights sum to 1) combination  
24 of model forecasts is performed with these weights to generate a best model  
25 forecast. A key component in wave forecasting is representation of extreme  
26 events, an area where traditional machine learning under-performs because  
27 such algorithms typically depend upon predicting the conditional mean of  
28 the data. By combining physical models with machine learning, we propose  
29 to overcome this shortcoming.

30 Ensembles forecasts of wave conditions are typically generated from statis-  
31 tical perturbations of wave-height boundary data, ocean-current input data,  
32 wind forcing (particularly for global models), model physics, discretization,  
33 or parameterization schemes (Chen, 2006). The fundamental objective of  
34 ensemble forecasting or prediction is to investigate the uncertainty inher-  
35 ent in forecasting to provide more information about future states. This  
36 process facilitates transition from single, deterministic forecasting with opti-  
37 mistic assumptions on the fidelity of model inputs, to a multiple, probabilis-  
38 tic forecasting approach that realistically considers the inherent errors and  
39 uncertainty in the model forcing data and fundamental physics. Ensemble  
40 models of large-scale, complex systems are typically created by either com-  
41 bining different models (called a model-intercomparison project MIP) or by  
42 perturbing input conditions and physics of a single model (Falloon et al.,

2014). The MIP approach involves taking a selection of models that differ in their system representation and evaluating forecasts for a range of scenarios. This approach is susceptible to misinterpretation if the models are not independent or if they share approximations or simplifications of certain processes. Further, it does not provide insight into individual model uncertainty or performance (Davie et al., 2013). To avoid these issues (and cognizant of the limited number of operational wave-forecasting systems available), this study considers a single model from which we create ensembles. This limits our focus to one model, but it provides a more systematic analysis of associated uncertainty and sensitivity to forcing data. It also assumes that the wave model is an appropriate simulation platform, an aspect that is assessed in greater detail in section 3.2.

This paper focuses on real-time forecasting of wave conditions at a case-study site in Monterey Bay, California and is structured as follows: The methodology section describes the approach and it includes a description of the model along with the generation of ensemble predictions. This section also details the different aggregation techniques investigated. A short description of the model construction and model set-up are provided including details on inputs and forcings to the model from a suite of real-time operational forecasting platforms. The results section describes the application of the model-aggregation technique to the Bay and the ability of the scheme to generate forecasts is assessed against the measurement data. Finally, conclusions from this research are drawn and the recommendations for future research made.

## 2. Motivation

Circulation and mixing in coastal regions results from the complex, non-linear interactions of waves, ocean currents, and winds. Depending on local conditions, the contribution from each can vary and a comprehensive study requires simultaneous consideration (O’Donncha et al., 2015). The objective of this study is to develop a robust system to forecast wave conditions by combining physics-based models with a machine-learning technique. Our analysis demonstrates that forcing the SWAN model with a single-point measurement of wave conditions provides marginally better performance than when forced by data from WAVEWATCH III (Tolman et al., 2009), despite the greater spatial richness of these forecast data. This illustrates a central point of modeling and operational forecasting; a forecast is not necessarily a

79 prediction on the true state of a system but rather the best estimate based  
80 on available data. The reality is that the WAVEWATCH III data provide  
81 some information on the likely state of the system rather than being viewed  
82 as a specific deterministic forecast. On the other hand, specifying boundary  
83 conditions directly from observation data is not a defensible strategy as one  
84 loses the ability to operate in forecasting mode; i.e., the model can only be  
85 run until the time of the most recent available observation data and in many  
86 respects the model then acts as a spatial interpolation module rather than a  
87 forecasting model.

88 We propose a framework to integrate accurate observational data with  
89 forecast conditions, to improve predictive capabilities. To investigate the  
90 likely true solution state of the system, we consider statistical perturbations  
91 of inputs (lateral wave boundary data) to the SWAN model. This yields a  
92 set of ensemble predictions for the next 48 hours. We propose a non-invasive,  
93 model-aggregating approach that integrates these models based on a set of  
94 learned weights computed by minimizing differences between model outputs  
95 and observations at each time when measured data become available. These  
96 weights are then used to produce a single, deterministic forecast cognizant  
97 of best model performance both historically and at the most recent obser-  
98 vation. The advantages of the proposed framework are: (1) there are no  
99 restrictions on which models may be included in the ensembles and infor-  
100 mation from deterministic physics models, stochastic models, or data-driven  
101 approaches could be readily incorporated and (2) the weights are computed  
102 using model outputs, so no modification to the model is required as would  
103 be necessary for data assimilation (DA), for example. As a comparison,  
104 a number of frameworks exist to integrate DA with existing models. Ex-  
105 amples include the Data Assimilation Research Testbed from the National  
106 Center for Atmospheric Research (NCAR, Anderson et al., 2009), the Java-  
107 based OpenDA (van Velzen et al., 2016), and the Parallel Data Assimilation  
108 Framework (Nerger et al., 2005) developed by the Alfred Wegener Institute.  
109 These provide a variety of different DA libraries and typically interface with  
110 the forecasting model in one of two ways: in-memory coupling by directly  
111 transferring data to the assimilation library (e.g., by calling the model as  
112 a subroutine to the DA framework) or through a “black box” approach in  
113 which all interactions between the models go through input and output files  
114 (Brankart and Melet, 2010). The amount of work required to couple a DA  
115 framework is model dependent and can be reduced to a question of how  
116 difficult it is to pass data in appropriate formats to the DA system (and

117 read back) or whether the model can be readily wrapped in external code  
 118 (i. e., implement the model as a library where the DA framework can readily  
 119 access and update the model state). Browne and Wilson (2015) evaluated  
 120 different strategies to couple DA libraries with complex models and proposed  
 121 an approach using MPI that aimed to reduce changes to source code. This  
 122 paper presents an alternative approach that acts only on model outputs to  
 123 update forecasts based on historical performance against observations.

### 124 **3. Methodology: Physics-based wave forecasting**

125 This effort compares SWAN wave-model predictions in Monterey Bay  
 126 forced by forecasts of wave conditions, ocean currents, and wind speeds to  
 127 data from three National Data Buoy Center (NDBC) buoys. The wave model  
 128 is driven by wave forecasts from NOAA’s NCEP, current forecasts from a  
 129 Regional Ocean Modeling System (ROMS) hydrodynamic model of Monterey  
 130 Bay (Patterson et al., 2012), and wind forecasts from TWC.

SWAN is a third-generation wave model that estimates wave conditions from prescribed wave information along boundary segments, ocean currents, winds, and bottom bathymetry. The spectral action balance equation describes the evolution of the wave-energy density spectrum  $E(\sigma, \theta)$ , over frequencies  $\sigma$  (as observed in a frame of reference moving with the current velocity) and propagation direction  $\theta$  (the direction normal to the wave crest of each spectral component) (The SWAN Team, 2006):

$$\frac{\partial N}{\partial t} + \left( \frac{\partial c_x N}{\partial x} + \frac{\partial c_y N}{\partial y} \right) + \left( \frac{\partial c}{\partial \sigma} N + \frac{\partial c}{\partial \theta} N \right) = \frac{S_{\text{tot}}}{\sigma}, \quad (1)$$

131 where  $N$  is the action density defined as the ratio of energy  $E$  to relative  
 132 frequency  $\sigma$  ( $N = E/\sigma$ ). The first term in the equation represents the local  
 133 rate of change of action density with time while the propagation of action  
 134 in geographical space is represented by the second and third terms. These  
 135 terms incorporate propagation velocities  $c_x$  and  $c_y$  in the  $x$  and  $y$  directions,  
 136 respectively. Depth- and current-induced refractions are represented by the  
 137 fourth term, which describes the propagation velocity,  $c$ , in spectral space  
 138  $(\sigma, \theta)$ . Shifting of the relative frequency due to variations in depth and cur-  
 139 rent are represented with the fifth term. The source term,  $S_{\text{tot}}$ , represents the  
 140 effects of wave generation, nonlinear wave-wave interaction, and dissipation  
 141 (O’Brien and Ragnoli, 2014).

142 The preceding equation can be solved as a hyperbolic equation on the dis-  
 143 cretized spectra and propagated forward in time. However, on large domains  
 144 requiring high spatial resolution, this propagation in time is computationally  
 145 expensive requiring matrix inversion incurring computational costs of up to  
 146  $O(n^3)$  (Toman, 2010), where  $n$  is number of elements. An alternative ap-  
 147 proach is to assume quasi-stationarity in the propagation of boundary effects  
 148 across the model domain. This reduces (1) to an elliptic equation that can  
 149 be resolved directly using an iterative solver.

### 150 3.1. SWAN Model Set-up

151 Figure 1 illustrates the extents of the Monterey Bay modeling domain  
 152 (64 54-km<sup>2</sup> domain discretized across 710 480 computational elements pro-  
 153 viding a horizontal resolution of 0.001° each approximately equal to 90 110 m<sup>2</sup>)  
 154 for the SWAN model, which was originally developed by Chang et al. (2016).  
 155 NOAA NDBC buoys from stations 46042 (white), 46114 (red), and 46240  
 156 (green) provide measurements of wave conditions together with other ocean  
 157 and meteorological data reported every 30 to 60 minutes. Table 1 provides  
 158 further information on the buoy datasets.

159 Primary inputs to the SWAN model are lateral boundary information of  
 160 wave forecasts, along with spatial distribution of ocean-current forecasts, and  
 161 wind forecasts. Lateral boundary information of wave height, direction, and  
 162 period are prescribed on the Southern, Western, and Northern boundaries  
 163 with data extracted from the WAVEWATCH III Eastern North Pacific model  
 164 at 0.25° spacing (turquoise markers denote where WAVEWATCH III data are  
 165 available and red segments where prescribed as lateral boundary conditions).  
 166 Ocean currents are from the Monterey Bay ocean forecasting system (black  
 167 symbols) based on the 3000-m-resolution ROMS model (IOOS, 2017). Wind  
 168 speeds were extracted at 0.25° spacing (turquoise symbols) from TWC appli-  
 169 cation programming interface (API). TWC provides information on a variety

Table 1: NDBC buoys used in this study.

Station name	Longitude	Latitude	Water depth (m)	First data
Buoy 46042	122.452° W	36.791° N	2,098	June 1987
Buoy 46114	122.351° W	36.723° N	1,463	September 2011
Buoy 46240	122.907° W	36.626° N	17.8	January 2009

170 of meteorological conditions, forecasts, alerts, and historical data, which can  
171 be extracted either directly from the TWC API or through the IBM Bluemix  
172 platform. Hourly forecast data out to fifteen days are available along with  
173 historical cleansed (i. e., subject to quality assurance procedures) data for the  
174 past 30 years. Table 2 summarizes details on the datasets used to force the  
175 SWAN model.

### 176 3.2. Verification of the SWAN Model

177 The first step of the study was performance confirmation of the SWAN  
178 model and assessment of the sensitivity to boundary forcing data. Six days of

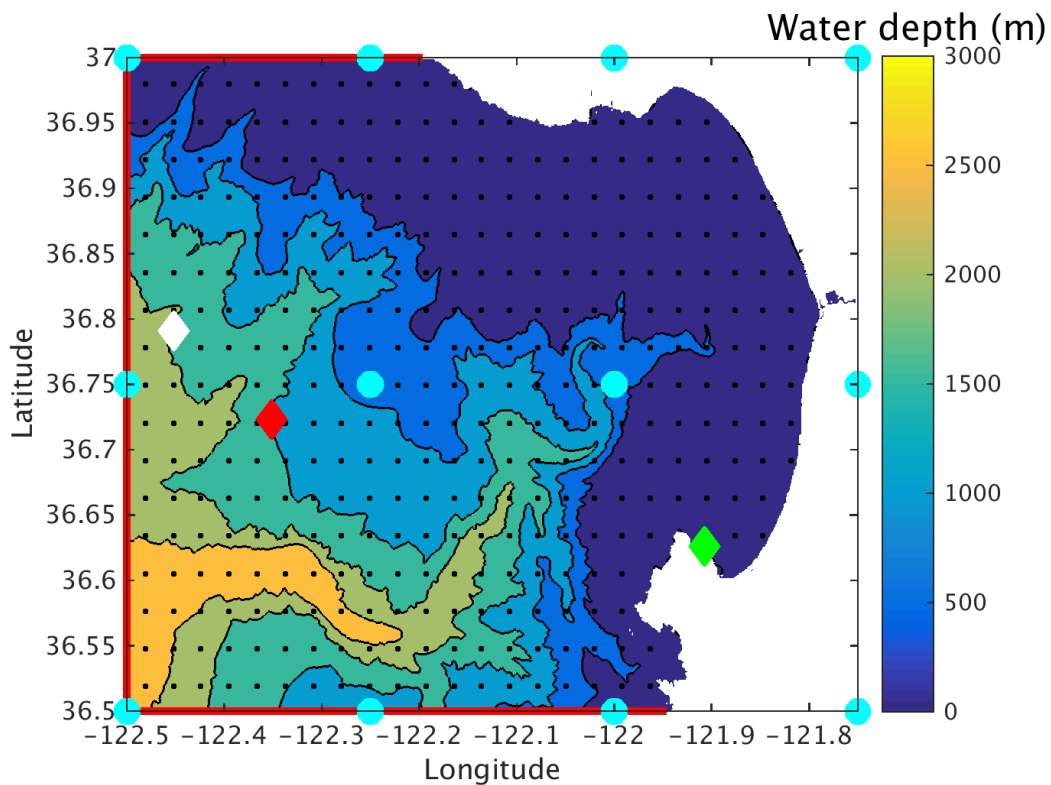


Figure 1: SWAN model domain with color indicating the bathymetric depth. The three buoys used to verify the model are indicated with the symbols where the white diamond is Buoy 46042, the red diamond is Buoy 46114, and the green diamond is Buoy 46240. The red boundaries are where wave-condition data were prescribed. The turquoise symbols indicate where WAVEWATCH III data were available and TWC data were prescribed to the model. Naturally WAVEWATCH III data returned no data over land. The black dots are where ROMS current data were provided.

179 NOAA wave data, ROMS ocean currents, and TWC winds were assembled  
 180 into steady-state SWAN model runs at three-hour intervals. SWAN mod-  
 181 els were run with different lateral wave boundary conditions. Currents were  
 182 supplied by the ROMS 3000-m grid, wind forcings were supplied by TWC,  
 183 and boundary wave conditions were extracted from either WAVEWATCH III  
 184 forecasts or from NOAA Buoy 46042. Figure 2 compares observed wave-  
 185 characteristics data (significant wave height,  $H_s$ , top row; wave period,  $T$ ,  
 186 middle row; wave direction,  $D$ , bottom row) for the three buoys in the Mon-  
 187 terey Bay area (black curves) to the different model forecasts.

188 The black symbols denote data extracted from WAVEWATCH III fore-  
 189 casts at the nearest grid point to each buoy. Given the extent of this rel-  
 190 atively coarse model ( $0.25^\circ$  resolution for the entire Eastern North Pacific),  
 191 it is not surprising that there is significant mismatch. Moreover, the lo-  
 192 cation of NOAA Buoy 46240 (green diamond in Figure 1), which is shel-  
 193 tered from incoming westward waves, results in a significant discrepancy be-  
 194 tween WAVEWATCH III-simulated and buoy-measured wave characteristics  
 195 because the nearest WAVEWATCH III grid point is well to the southwest  
 196 of the buoy ( $122^\circ$  W and  $36.5^\circ$  N). The blue symbols are simulated wave  
 197 characteristics when NOAA wave data from Buoy 46042 were supplied to the  
 198 SWAN model as boundary conditions. Unsurprisingly, blue and black sym-  
 199 bols are quite similar at Buoy 46042 because this buoy served as boundary  
 200 conditions to the SWAN model with results nearly as close at Buoy 46114  
 201 (red diamond in Figure 1), which is about 13 km to the east-southeast. The  
 202 red symbols are simulated wave characteristics at the three buoys when the  
 203 SWAN model was forced by WAVEWATCH III wave conditions along its  
 204 boundaries.

We use both root-mean-squared errors (RMSE) and mean absolute per-

Table 2: Forecast data descriptions including resolutions at which they were applied to the SWAN model.

Model	Data	Resolution	Forecast (days)	Frequency (hr)
WAVEWATCH III	$H_s, T, D$	$0.25^\circ$	7.5	3
TWC	$u, v$ winds	$0.25^\circ$	15	1
ROMS	$u, v$ currents	3000 m	2	3



centage error (MAPE) to quantify model deviation from measured data

$$\text{RMSE} = \sqrt{\frac{\sum_{n=1}^{\mathcal{N}} (X_{\text{model}} - X_{\text{obs}})^2}{\mathcal{N}}}, \quad (2)$$

$$\text{MAPE} = \frac{100}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} |X_{\text{model}} - X_{\text{obs}}|, \quad (3)$$

205 where  $X_{\text{model}}$  and  $X_{\text{obs}}$  represent the model-predicted and observed values,

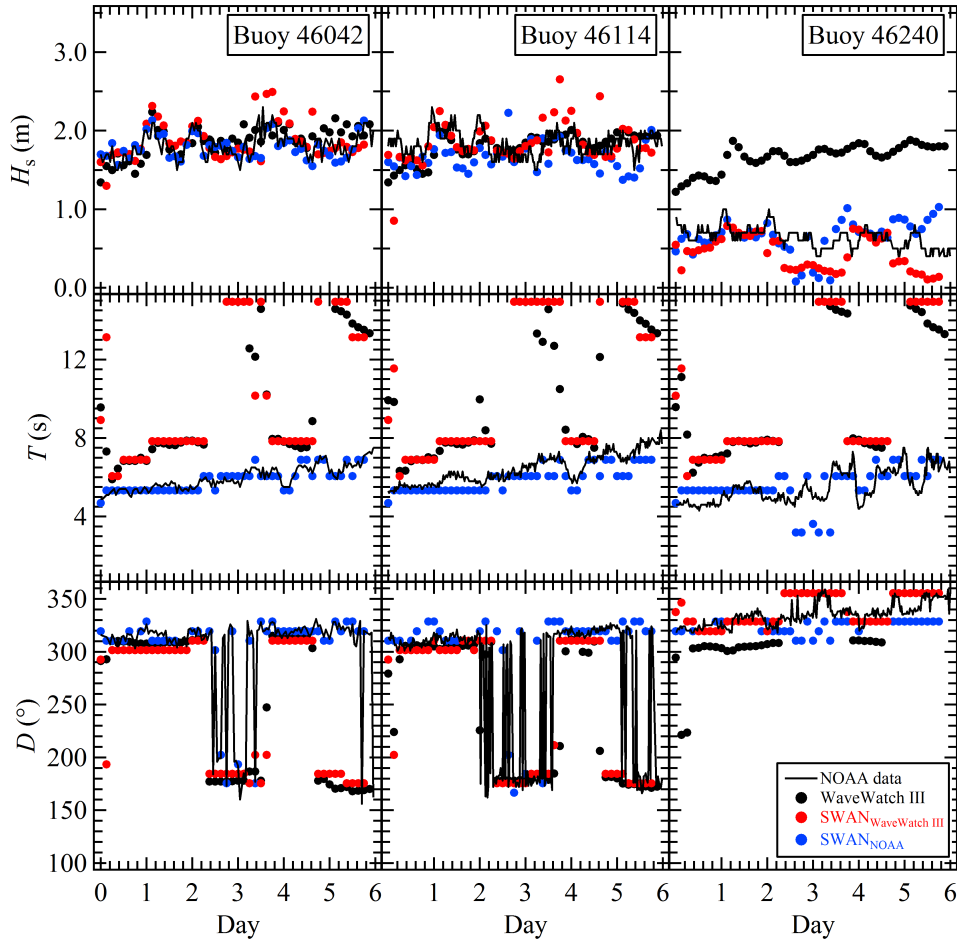


Figure 2: Comparison of simulated and measured wave characteristics at the three NOAA buoys.

206 respectively. As an initial assessment of model performance over the study  
 207 period, RMSEs for significant wave heights at each buoy are listed in Ta-  
 208 ble 3. SWAN provides slightly more accurate results using measured wave  
 209 data from NOAA than when WAVEWATCH III boundary conditions were  
 210 used. This comparison ignores the fact that forcing with NOAA data uses  
 211 relatively accurate observation data, while forcing with WAVEWATCH III  
 212 data incurs the penalty of uncertainty inherent in any operational forecasting  
 213 platform. This is however balanced by the restrictive constraint that forcing  
 214 with measured data does not enable forecasting. Notably, SWAN performs  
 215 better at buoy 46240 when forced by WAVEWATCH III data, a location  
 216 more sensitive to complex directional influences that is likely better resolved  
 217 from the greater spatial coverage of WAVEWATCH III data than from a sin-  
 218 gular buoy. Hence, the remainder of the study focused on ensemble generation  
 219 using forecasts from the NOAA WAVEWATCH III model as forcing data.  
 220 This model configuration produced mean RMSE across the three buoys of  
 221 0.57m (Table 3) corresponding to a MAPE of 22.1%. Bidlot et al. (2002) note  
 222 that 40- to 60-cm RMSEs are typical for modeling studies of wave heights  
 223 demonstrating that the SWAN model and configuration are appropriate as  
 224 a forecasting tool.

225 A notable feature of Table 3 is the relatively high RMSE reported for  
 226 wave direction. This is partly a result of the higher variance and volatility  
 227 in the observed wave directions than simulated. Focusing on the outer buoys  
 228 (46042 and 46114 in Figure 2), wave direction typically alternates between  
 229 175 and 310 degrees (where direction is denoted as direction from which  
 230 waves are coming with units of degrees North, increasing clockwise - with 0°  
 231 being North), denoting waves from South or from West-Northwest. These  
 232 general directional trends are captured well by the model. However, periods

Table 3: RMSEs computed for the different models and model configurations against measured data from NDBC buoys. The first data row compares predictions from the WAVEWATCH III model against the three buoy while the next two rows present RMSEs computed for the SWAN model forced by WAVEWATCH III data and measured data extracted from NOAA Buoy 46042, respectively.

Model	Buoy 46042			Buoy 46114			Buoy 46240		
	$H_s$ (m)	$T$ (s)	$D$ (°)	$H_s$ (m)	$T$ (s)	$D$ (°)	$H_s$ (m)	$T$ (s)	$D$ (°)
WAVEWATCH III	0.67	3.61	80.3	0.62	3.29	74.5	1.62	4.08	103.7
SWAN <sub>WAVEWATCH III</sub>	0.68	3.56	80.0	0.68	3.23	73.7	0.35	4.32	16.5
SWAN <sub>NOAA</sub>	0.50	0.57	62.3	0.41	0.45	73.9	0.42	1.06	12.0

233 with very high volatility in wave direction, as seen at day 3 in Figure 2, are  
234 not replicated by the model, instead predicting relatively stationary incoming  
235 waves from the South. It is likely that the higher spatial resolution (particu-  
236 larly of the coarse-resolution WAVEWATCH III data serving as forcing) do  
237 not encapsulate these small-scale, localized changes in wave characteristics.

## 238 **4. Methodology: Integrating forecasts with machine learning**

### 239 *4.1. Creation of Model Ensembles*

240 A stationary wave model is insensitive to initial conditions; instead an  
241 iterative sweep of wave conditions is conducted until the solution converges  
242 to within some threshold. Hence, we only consider perturbations to forcing  
243 conditions and model physics. Chen et al. (2004) considered the sensitivity  
244 of a global WAVEWATCH III model driven by wind forcing and updated by  
245 information derived from DA. DA provided improved short-term forecasts  
246 (12 to 24 hours) while perturbation of wind forcing had the greatest impact  
247 on forecasts. However, in regional models the sensitivity to wind forcing  
248 is less and the greatest source of uncertainty is the lateral boundary infor-  
249 mation on wave conditions specified from an external model or other data  
250 source. Preliminary studies investigating the sensitivity of the Monterey Bay  
251 SWAN model to perturbed inputs of wind forcing extracted from the NOAA  
252 Global Ensemble Forecast System (GEFS) demonstrated low sensitivity to  
253 perturbation of wind inputs (Zhang et al., 2017). As a result of the limited  
254 spatial scale of the Monterey Bay domain ( $\approx 3,500 \text{ km}^2$ ), perturbing wind  
255 input data based on outputs from NOAA GEFS forecasts yielded changes  
256 in wave height of less than 0.5 cm. Considering the lack of sensitivity, and  
257 to simplify the ensemble generation process, we focused on: (1) perturbing  
258 wave boundary data prescribed on the lateral boundaries and (2) varying  
259 model physics by creating SWAN ensemble members using first-, second-,  
260 and third-generation physics (The SWAN Team, 2006).

261 Creation of perturbed wave boundary condition data considered the de-  
262 terministic forecast from the WAVEWATCH III model and the information  
263 on system dynamics provided by the NOAA buoy nearest to the SWAN  
264 boundary, and consequently most exposed to open-ocean wave conditions  
265 (Buoy 46042, see Figure 1). To quantify the dynamics of the system, we  
266 used the standard deviation of the observation data to generate upper and  
267 lower statistical bounds on the prescribed boundary conditions that encom-  
268 passed all the observation data at this buoy with a 95% confidence interval.

269 The buoy data were first pre-processed to remove outliers by eliminating the  
270 upper and lower 2.5% of observations. Note that we experimented with a  
271 range of percentages between 0.5% and 10% for pre-processing. The esti-  
272 mated standard deviation stabilized at 2.5%, which led to our choice of this  
273 threshold. Next, the standard deviation of observations was computed using  
274 a 48-hour moving window. Upper and lower bounds to the wave boundary  
275 information from the WAVEWATCH III model were computed by adding  
276 1.96 standard deviations to the deterministic WAVEWATCH III forecasts.  
277 Figure 3 presents the deterministic predictions of wave heights extracted  
278 from WAVEWATCH III together with the computed statistical bounds (gray  
279 envelope) while the blue circles denote measured data. The plot illustrates  
280 that the deterministic forecast at times overestimated (e.g., July 17<sup>th</sup>) or  
281 underestimated (e.g., July 29<sup>th</sup>) actual observations. These temporal biases  
282 and general spread of measurements around the model predictions are en-  
283 compassed by the calculated statistical bounds indicating that the generated  
284 spread encompassed most system dynamics.

285 We generated 12 unique ensemble members by perturbing wave boundary  
286 conditions within the statistical bounds estimated as described above. Latin  
287 hypercube sampling (LHS) was used to extract samples from within these  
288 bounds and prescribed as model boundary conditions. LHS is a statistical  
289 method for generating a near-random sample of parameter values from a  
290 distribution (McKay et al., 2000). When sampling a function, its range is  
291 divided into  $N$  equally probable intervals and a random sample is selected  
292 from each interval. This ensures adequate coverage of a distribution where  
293 the tails are important.

294 Additional ensembles were created by manipulating model physics. SWAN  
295 can operate in first-, second-, and third-generation modes. The first- and  
296 second-generation modes are essentially those of Holthuijsen et al. (1988);  
297 first-generation with a constant Phillips “constant” of 0.0081 and second-  
298 generation with a variable Phillips “constant.” Third-generation processes  
299 include wind input, whitecapping, bottom friction, depth-induced wave break-  
300 ing, dissipation due to vegetation, mud, or turbulence, obstacle transmission,  
301 nonlinear wave-wave interactions (quadruplets and triads) and wave-induced  
302 set-up. An overview of these options is given in SWAN User’s Manual (The  
303 SWAN Team, 2006). By perturbing inputs and model physics a total of 15  
304 unique ensemble members were generated (12 with perturbed wave boundary  
305 data using third-generation physics, and individual ensemble elements with  
306 unperturbed inputs and first-, second- and third-generation physics modes).

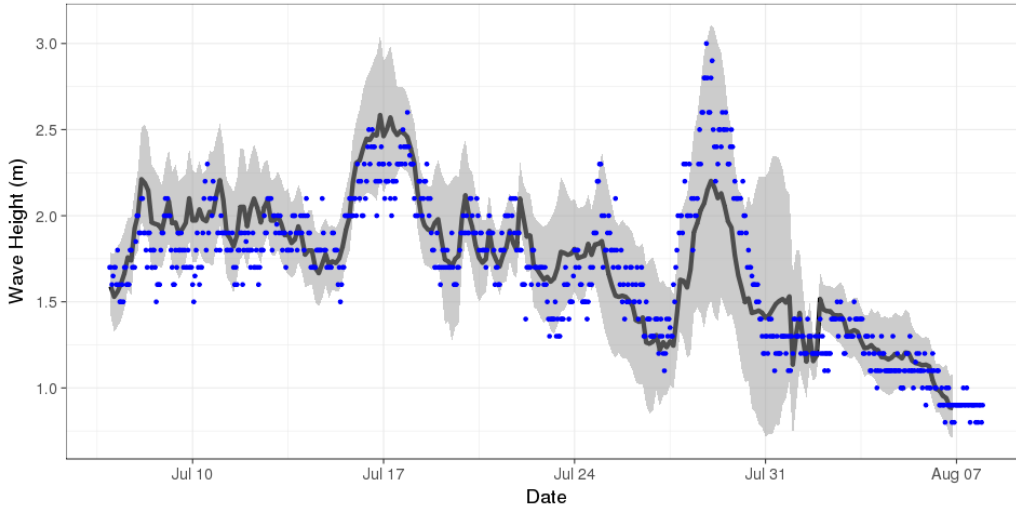


Figure 3: Time series of the mean  $H_s$  computed by WAVEWATCH III at the western model boundary (black curve). The blue circles denote observations from Buoy 46042 (see Figure 1) while the gray band represents the upper and lower bounds on wave height for specifying model lateral boundary conditions.

#### 307 4.2. Model Aggregation

308 The underlying assumption of ensemble modeling is that each model con-  
 309 tains some information pertinent to the true state of the system. The inter-  
 310 play between models is expected to vary in both space and time; i. e., in-  
 311 dividual models of an ensemble perform better at different points in space  
 312 and time depending upon ambient conditions, individual model forcings, and  
 313 other physical interactions. The objective of the aggregation method is to de-  
 314 velop a weight for each member of the ensemble taking into account previous  
 315 predictions and observations.

316 We considered two aggregating approaches to evaluate different sensitivi-  
 317 ties of forecasts to previous model performance. The objective is to generate  
 318 a weight vector,  $\mathbf{u}_t$ , at each time index,  $t$ , that minimizes average mean  
 319 square error (MSE) between predictions and observations by aggregating the  
 320  $N$  ensemble predictions into a single “best-estimate” forecast.

The first technique investigated was a ridge regression (RR) prediction algorithm. The weight vector for each time update was computed as (Mallet

et al., 2009):

$$\mathbf{u}_t = \arg \min_{\mathbf{u} \in \mathbb{R}^n} \left[ \lambda \|\mathbf{u}\|_2^2 + \sum_{t'=1}^{t-1} \sum_{s \in S_{t'}} (\mathbf{u} \cdot \mathbf{x}_{t'}^s - \mathbf{y}_{t'}^s)^2 \right], \quad (4)$$

321 where  $N$  is the number of ensemble members,  $\mathbf{x}_{t'}$  is a vector of dimension  
 322  $N$  containing the prediction from each ensemble member at time  $t'$  (for each  
 323 station  $s$ ),  $\mathbf{y}_{t'}$  represents observations at time  $t'$ ,  $\mathbf{u}_t$  is a vector of weights  
 324 computed for each ensemble member, and  $S_{t'}$  represents the number of ob-  
 325 servation stations for which data are available at each time. Conceptually,  
 326 the objective can be considered as choosing the appropriate weight for each  
 327 ensemble element to minimize the MSE across all observation stations or  
 328 buoys. The training of the weights vector on data progresses on a certain  
 329 subset of data from time  $t'$  to  $t - 1$  from which predictions are then made for  
 330 the next time step  $t$  based on the most recent ensemble predictions.  $\lambda$  is a  
 331 penalty function that keeps the magnitude of  $\mathbf{u}_t$  small and reduces variation  
 332 between consecutive values. Mallet et al. (2009) provided a brief discus-  
 333 sion of this penalization function, which is typically selected in an *ad hoc*  
 334 manner for each study to balance contributions from the most recent model-  
 335 observation datasets and historical data. Scope exists however to use more  
 336 robust cross-validation approaches to guide parameter selection.

The computed weights from time  $t'$  to  $t - 1$  are then used to make a  
 forecast,  $\hat{x}_t^s$ , for each station,  $s$ , at time  $t$  as:

$$\hat{x}_t^s = \mathbf{u}_t \cdot \mathbf{x}_t^s = \sum_{m=1}^N u_{m;t} x_{m;t}^s, \quad (5)$$

337 where  $\mathbf{x}_t^s$  is each member of the ensemble prediction at station  $s$ , and  $\mathbf{u}_t$  is  
 338 the weight applied to each prediction.

339 The second technique implemented here is an Exponentiated Gradient  
 340 (EG) algorithm for linear predictors (Kivinen and Warmuth, 1997). The EG  
 341 algorithm also has a weight vector  $\mathbf{u}_t$  and predicts with  $\hat{x}_t^s = \mathbf{u}_t \cdot \mathbf{x}_t^s$ . The  
 342 update of the model weights vector for each model ensemble member  $x_{m;t'}$  is  
 343 of the form (Kivinen and Warmuth, 1997):

$$u_{m;t} = \frac{\sum_{t'=1}^{t-1} r_{m;t'} u_{m;t'}}{\sum_{j=1}^N \sum_{t'=1}^{t-1} r_{j;t'} u_{j;t'}}, \quad (6)$$

344 for all  $m = 1, \dots, N$  where  $t'$  denotes data at the previous time step and,  $r_{m;t'}$   
 345 is computed as:

$$r_{m;t'} = \exp \left[ \sum_{s \in \mathcal{S}_{t'}} 2\mu(u_{m;t'} \ x_{m;t'}^s \ y_{t'}^s) x_{m;t'} \right], \quad (7)$$

346 where  $\mu$  denotes learning rate.

347 Weights computed with the EG approach were normalized by the sum of  
 348 all weights as expressed in (6). This constrained the weights to a convex com-  
 349 bination where all weights summed to one as opposed to the unconstrained  
 350 weights provided by RR (i. e., where weights could take any values that sat-  
 351 isfy the objective function). A potential advantage of EG-type approaches  
 352 over RR is this constraint on weights, which limits rapid fluctuations over  
 353 time. Convex-combination weight vectors may be more extensible to other  
 354 regions of the model domain away from where observations are available (and  
 355 consequently are included in the weight computations) than unconstrained  
 356 weights (Mallet et al., 2009). These aggregated predictions will always fall  
 357 in the envelope of the ensemble predictions, which avoids unrealistic model  
 358 forecasting.

359 A framework to implement the above aggregation consists of the following  
 360 steps:

- 361 1. Initialize the weight vector  $\mathbf{u}_0 = 0$
- 362 2. Select penalization coefficient  $\lambda > 0$
- 363 3. For each timestep  $t' = 1 \dots t$ 
  - 364 Predict with  $\mathbf{u}_{t'}$
  - 365 Using the prediction and the observations compute updated  $\mathbf{u}_{t'}$
  - 366 that minimizes (4)
  - 367 Using the updated weight,  $\mathbf{u}_{t'}$ , compute forecast,  $\hat{x}_t^s$  for each
  - 368 station,  $s$  at time  $t$  using (5).

369 The Python toolkit SciKit-Learn (Pedregosa et al., 2011) provides a high-  
 370 level programming interface to linear model implementations such as ridge  
 371 regression.

## 372 5. Results and Discussion

373 Analysis of results and performance of the aggregation technique focused  
 374 on one month of simulations from July 7<sup>th</sup>–August 7<sup>th</sup>, 2017. The SWAN

375 model was set up as described in Section 3.1. Forcing data for wave condi-  
 376 tions and wind and current speeds were extracted from the sources described  
 377 in Table 2. The model ensembles comprised 15 individual elements com-  
 378 posed of: a forecast with deterministic, one member with unperturbed wave  
 379 heights, 12 simulations with perturbed wave boundary heights, and two simu-  
 380 lations based on first- and second-generation SWAN model physics (all other  
 381 ensembles used third-generation physics).

382 Model aggregation focused on making on-line forecasting using all avail-  
 383 able past data (real-time) to update weight coefficient applied to each en-  
 384 semble element. Weights were initialized to zero, and the values for each  
 385 model forecast time were computed based on a minimization of the differ-  
 386 ences between forecasts and observations together with historical information  
 387 contained in the first term on the right-hand-side of (4). The fundamental ob-

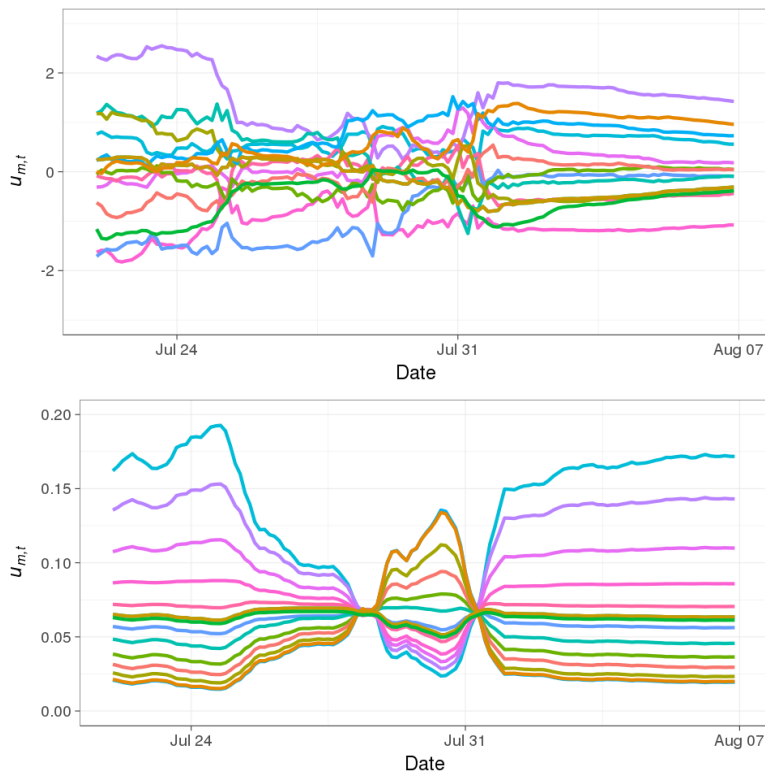


Figure 4: (a) RR and (b) EG weights computed for each of the 15-member model ensemble for each time snapshot (predictions made at three-hour intervals). Each curve represents the weight attached to an individual model and evolution over time.



388 jective is to leverage the observation data available in real-time to improve  
389 short-term forecast capabilities. Figure 4 presents the computed weights  
390 using the RR and EG aggregation method for a one-month period. Most  
391 models contribute to the aggregated forecast with the majority of weights  
392 having strongly non-zero values. Further, the dynamics and variations of  
393 the weighting aggregation over time were apparent with larger magnitude  
394 weights corresponding to periods with largest spread in model forecasts (and  
395 consequently highest uncertainty from the ensemble prediction perspective).  
396 During periods of large model spread, the minimization of model-observation  
397 differences was achieved by applying large weights to models that performed  
398 well and low weights to models that performed poorly.

399 A key consideration for wave forecasting is the temporal dynamics of  
400 the system. The fundamental aspect of weighted model aggregation is that  
401 there is a certain relationship between successive forecasts and observations;  
402 i.e., there is a likelihood that the ensemble element that performs best at  
403 time  $t$  will be the model that performs best at time  $t + 1$ . By constantly  
404 updating weights based on the difference between the latest observations  
405 and forecast, one can ensure that the best-performing model is assigned the  
406 highest weight. This “follow-the-leader” forecasting system works best if  
407 the quantity being modeled is relatively stationary with pronounced historic  
408 influence. Figure 5a presents the spatially averaged  $H_s$  computed at each of  
409 the three buoy locations plotted against spatially averaged observations (over  
410 the three NOAA buoys). The dashed curves represent  $H_s$  computed by each  
411 ensemble element while the solid black curve represents observation data and  
412 the solid red curve represents the model that yielded the best performance  
413 against observed data. The highly dynamic nature of the system is clear  
414 with significant variation in  $H_s$  over short time periods. Further, at various  
415 time periods, the models converged to a similar solution (e.g., August 3<sup>rd</sup>  
416 – 7<sup>th</sup>), while at other times they spanned a broad range (e.g., July 31<sup>st</sup>).  
417 This illustrates the distinct challenges of ensemble aggregation focused on  
418 computing appropriate weights incorporating these highly dynamic events.

419 Figure 5b presents the evolution of MAPE for the corresponding period  
420 for three different cases: (1) where the best-performing ensemble element  
421 was considered (best individual model indicated by the black curve), (2) a  
422 weighted aggregation using the RR method with weights computed for every  
423 three-hour prediction (red curve), and (3) weighted aggregation using the  
424 EG method (blue curve). For the one-month study period, the best indi-  
425 vidual model reported a MAPE of 20.8%, while the weighted aggregation

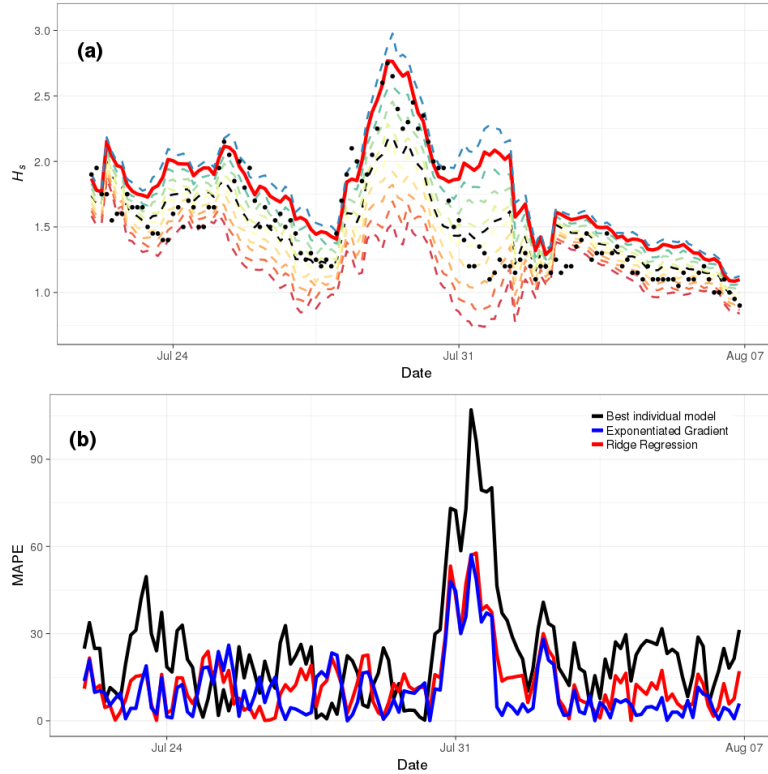


Figure 5: (a) Predicted  $H_s$  against observations spatially averaged across all three buoys. The solid black curve denotes observations, dashed curves represent predictions from each ensemble element, while the solid red curve denotes best-performing individual model. (b) MAPE computed against spatially averaged observations for each of (1) best-individual model (black curve), (2) aggregation using RR (red curve), and (3) aggregation based on EG (blue curve).

426 technique reduced MAPE to 9.76% and 9.26% for the RR and EG forecast-  
 427 ers, respectively. The aggregated prediction significantly outperformed the  
 428 best individual model aside from a short period around July 25<sup>th</sup> and 28<sup>th</sup>  
 429 when the error of the best individual model against observations reduced to  
 430 almost zero.

431 Ocean waves are a volatile natural process subject to rapid variation in  
 432 space and time. This means that the correlation in time is relatively low  
 433 and the “history effect” of prediction is not guaranteed; i. e., often the best-  
 434 performing model at time  $t$  will not be the best performing model three  
 435 hours later. However, these results demonstrated that most of the time

436 the aggregated ensemble prediction provided superior results over simply  
437 selecting the best-performing individual model. The high volatility of wave  
438 conditions means that an aggregating algorithm that adjusts more quickly  
439 to the latest conditions provides better performance overall. Chen et al.  
440 (2004) observed that a wave-forecasting system updated with DA provided  
441 improved forecasts in the near-term (12 to 24 hours). However, due to the  
442 nature of such volatile systems, gains provided by DA diminished for longer  
443 forecasting windows implying that autocorrelation of the system is limited to  
444 12 to 24 hours. Model aggregation is subject to similar trends. The weights  
445 computed based on all past model and observation data inform the best-  
446 performing model at the present time. Depending on the temporal volatility,  
447 this information is valid for a certain time window, beyond which system  
448 dynamics may evolve toward a different set of weights (i. e., ensemble element  
449 scores). The study by Chen et al. (2004) and the results presented here  
450 suggest that the latest information influences future states for short-term  
451 predictions.

452 Scope exists to apply machine-learning techniques to directly forecast  
453 ocean-wave conditions (James et al., 2018). A common shortcoming of data-  
454 driven techniques, however, is that while they may be accurate on average,  
455 they may miss extreme events. Integrating machine-learning approaches with  
456 physics models affords opportunities to leverage available observation data  
457 to update predictions, while also maintaining the ability to resolve extreme  
458 events as encapsulated by the fundamental equations describing the system.  
459 Weights are computed by the machine-learning approach, but the predictions  
460 themselves are made by a physical model. Mallet et al. (2009) discussed the  
461 ability of ensemble-aggregation techniques to predict extreme events applied  
462 to forecasts of ozone concentrations. They demonstrated that for the two  
463 most extreme frequency bins (low and high ozone concentrations), the aggre-  
464 gated predictions always outperformed the best performing individual model.  
465 Similarly, in this study the largest reductions in MAPE were achieved when  
466 wave heights were large and the best-performing model did not adequately  
467 encapsulate the data (i.e., July 30<sup>th</sup> – 31<sup>st</sup> in Figure 5a).

468 Another aspect worth considering is the individual models composing the  
469 ensemble prediction. In this study, a relatively simple experimental design  
470 was adopted to build the ensembles considering the statistical deviation of  
471 observation data to determine perturbations around the mean. These did  
472 not consider in detail any prior knowledge about the model (e.g., biases or  
473 efficiency at forecasting particular events or time periods, etc.). Further,

474 due to the relatively high computational cost of the model, there were limits  
475 to the number of ensemble elements that could be created. Despite the  
476 significant improvement in accuracy provided by the aggregation techniques,  
477 the method could be further complemented by a more extensive design of  
478 the simulation exercise. In particular, the addition of other wave-forecasting  
479 models to the ensemble would be useful to counter any potential biases of  
480 the SWAN model.

## 481 6. Conclusions

482 In this paper, we detailed the creation and generation of ensemble pre-  
483 dictions from statistical perturbation of forcing data. RR and EG aggre-  
484 gating algorithms computed deterministic forecasts from the ensemble that  
485 leveraged past observations and past model performance of each ensemble  
486 member. Results demonstrated that the aggregating forecaster significantly  
487 improved forecasts compared to the traditional civil-engineering approach  
488 based on a calibrated model or the best individual member of the ensemble  
489 to make forecasts. The aggregating algorithm reduced MAPE from 20.8%  
490 (best individual model) to 9.26% using the EG forecaster. Computed weights  
491 demonstrate that the approach leveraged most members of the ensemble to  
492 aggregate the final forecast with most computed weights having magnitudes  
493 notably different from zero.

494 One of the primary advantages of this approach is that it provides a non-  
495 invasive method to leverage data to improve forecasts. As the algorithm  
496 only acts on outputs from the model to compute weighted-sum predictions,  
497 it does not require any amendment or development of the source code as is  
498 necessary with traditional DA approaches. Further, the algorithm can be  
499 readily replaced with alternative local-minima approaches that better reflect  
500 the needs of a particular study (e. g., gradient-descent approaches, etc.).

501 Ensemble-based forecasting is a widely used technique to account for un-  
502 certainty inherent in numerical modeling studies. Leveraging multiple sim-  
503 ulations with perturbed inputs and physics facilitates an expanded explo-  
504 ration of likely future conditions and provides probabilistic information on  
505 forecasts. Many decision processes however, require a single, deterministic  
506 forecast. This is typically done with some form of averaging across all en-  
507 semble members or selection of the best individual model (based on some  
508 metric). The approach presented here provides a comprehensive technique  
509 that leverages information on past model performance and observations to

510 aggregate ensemble elements into a single forecast. The non-invasive frame-  
511 work can be easily integrated into an on-line operational forecasting system.  
512 This can be readily extended to other models and in particular to combin-  
513 ing and aggregating models with different levels of complexity and different  
514 fundamental physics (e. g., combining rule-based models with data-driven  
515 models or deterministic approaches with stochastic). Future work will inves-  
516 tigate aggregating techniques when prior information on model complexity  
517 and physics can be used to provide prior information to the aggregating  
518 technique and parameterization.

### 519 **Acknowledgements**

520 Elements of this research has received funding from the European Union's  
521 Horizon 2020 research and innovation programme under grant agreement No.  
522 773330.

523 **References**

- 524 Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., Avellano,  
525 A., 2009. The data assimilation research testbed: A community facility.  
526 *Bulletin of the American Meteorological Society* 90 (9), 1283–1296.
- 527 Bidlot, J.-R., Holmes, D. J., Wittmann, P. A., Lalbeharry, R., Chen, H. S.,  
528 2002. Intercomparison of the performance of operational ocean wave fore-  
529 casting systems with buoy data. *Weather and Forecasting* 17 (2), 287–310.
- 530 Brankart, J.-M., Melet, A. ., 2010. Multigrid ocean data assimilation: Sesam  
531 with agrif. Tech. rep., LEGI/MEOM.
- 532 Browne, P. A., Wilson, S., 2015. A simple method for integrating a complex  
533 model into an ensemble data assimilation system using mpi. *Environmental*  
534 *Modelling & Software* 68, 122–128.
- 535 Chang, G., Ruehl, K., Jones, C., Roberts, J., Chartrand, C., 2016. Numer-  
536 ical modeling of the effects of wave energy converter characteristics on  
537 nearshore wave conditions. *Renewable Energy* 89, 636–648.
- 538 Chen, H. S., 2006. Ensemble prediction of ocean waves at NCEP. In: Pro-  
539 ceedings of the 28<sup>th</sup> Ocean Engineering Conference.
- 540 Chen, H. S., Behringer, D., Burroughs, L., Tolman, H., 2004. A variational  
541 wave height data assimilation system for ncep operational wave models.  
542 Tech. Proced. Bull., Ser. no. MMAB/2004–2004, 1–13.
- 543 Davie, J., Falloon, P. D., Kahana, R., Dankers, R., Betts, R., Portmann,  
544 F. T., Wisser, D., Clark, D. B., Ito, A., Masaki, Y., et al., 2013. Comparing  
545 projections of future changes in runoff from hydrological and biome models  
546 in ISI-MIP. *Earth System Dynamics* 4(2), 359–374.
- 547 Falloon, P., Challinor, A., Dessai, S., Hoang, L., Johnson, J., Koehler, A.-K.,  
548 2014. Ensembles and uncertainty in climate change impacts. *Frontiers in*  
549 *Environmental Science* 2, 33.
- 550 Holthuijsen, L., Boer, S., et al., 1988. Wave forecasting for moving and sta-  
551 tionary targets.

- 552 IOOS, 2017. Central and Northern California Ocean Observing System, CeN-  
553 COOS.  
554 URL <http://www.cencoos.org/data/models/roms/monterey>
- 555 James, S. C., Zhang, Y., O'Donncha, F., 2018. A machine learning frame-  
556 work to forecast wave conditions. *Coastal Engineering* 137, 1 – 10.  
557 URL <http://www.sciencedirect.com/science/article/pii/S0378383917304969>
- 558 Kivinen, J., Warmuth, M. K., 1997. Exponentiated gradient versus gradient  
559 descent for linear predictors. *Information and Computation* 132 (1), 1–63.
- 560 Komen, G. J., Cavaleri, L., Donelan, M., 1996. *Dynamics and Modelling of*  
561 *Ocean Waves*. Cambridge University Press.
- 562 Mallet, V., Stoltz, G., Mauricette, B., 2009. Ozone ensemble forecast with  
563 machine learning algorithms. *Journal of Geophysical Research: Atmo-*  
564 *spheres* 114 (D5).
- 565 McKay, M. D., Beckman, R. J., Conover, W. J., 2000. A comparison of three  
566 methods for selecting values of input variables in the analysis of output  
567 from a computer code. *Technometrics* 42 (1), 55–61.
- 568 Mei, C. C., Stiassnie, M., Yue, D. K.-P., 1989. *Theory and Applications of*  
569 *Ocean Surface Waves: Part 1: Linear Aspects. Part 2: Nonlinear Aspects.*  
570 *World Scientific.*
- 571 Nerger, L., Hiller, W., Schröter, J., 2005. Pdaf-the parallel data assimilation  
572 framework: experiences with kalman filtering. In: *Use Of High Perform-*  
573 *ance Computing In Meteorology*. World Scientific, pp. 63–83.
- 574 O'Brien, N., Ragnoli, E., 2014. A hybrid grid system for a spectral wave  
575 model. In: *Proceedings of the 9<sup>th</sup> International Conference on Engineering*  
576 *Computational Technology.*
- 577 O'Donncha, F., Hartnett, M., Nash, S., Ren, L., Ragnoli, E., 2015. Char-  
578 acterizing observed circulation patterns within a bay using HF radar and  
579 numerical model simulations. *Journal of Marine Systems* 142, 96–110.
- 580 Patterson, J., Thomas, J., Rosenfeld, L., Newton, J., Hazard, L., Scianna,  
581 J., Kudela, R., Mayorga, E., Cohen, C., Cook, M., et al., 2012. Addressing  
582 ocean and coastal issues at the West Coast scale through regional ocean  
583 observing system collaboration. In: *OCEANS'12. MTS-IEEE*, pp. 1–8.

- 584 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel,  
585 O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011.  
586 SciKit-Learn: Machine learning in Python. *Journal of Machine Learning*  
587 *Research* 12 (10), 2825–2830.  
588 URL <http://scikit-learn.org/stable/>.
- 589 Rogers, W. E., Wittmann, P. A., Wang, D. W., Clancy, R. M., Hsu, Y. L.,  
590 2005. Evaluations of global wave prediction at the fleet numerical meteo-  
591 rology and oceanography center. *Weather and Forecasting* 20 (5), 745–760.
- 592 The SWAN Team, 2006. SWAN scientific and technical documentation. Tech.  
593 Rep. SWAN Cycle III version 40.51, Delft University of Technology.
- 594 Tolman, H. L., et al., 2009. User manual and system documentation of wave-  
595 watch iii tm version 3.14. Technical note, MMAB Contribution 276, 220.
- 596 Toman, H. L., 2010. A note on stationary wave modeling with Wavewatch  
597 III. Tech. rep., NOAA National Centers for Environmental Prediction.
- 598 van Velzen, N., Altaf, M. U., Verlaan, M., 2016. Openda-nemo framework  
599 for ocean data assimilation. *Ocean Dynamics* 66 (5), 691–702.
- 600 Zhang, Y., O’Donncha, F., James, S. C., 2017. An ensemble-based approach  
601 that combines machine learning and numerical models to improve forecasts  
602 of wave conditions. In: OCEANS’17. Anchorage, Alaska, pp. 1–7.