Operational assimilation of spectral wave data from the Sofar Spotter network

Isabel Houghton[1], Christie A Hegermiller[1], Camille Teicheira[1], and Pieter Bart Smit[1]

[1]Sofar Ocean

November 26, 2022

Abstract

Historically, the sparseness of in situ open-ocean wave and weather observations has severely limited the forecast skill of weather over the ocean with major social and economic consequences for coastal communities and maritime industries. Ocean surface waves, specifically, are important for the interaction between atmosphere and ocean, and thus key in modeling weather and climate processes. Here, we investigate the improvements achievable from a large distributed sensor network combined with advances in assimilation strategies. Wave spectra from a global network of over 600 Sofar Spotter buoys are assimilated into an operational global wave forecast via optimal interpolation to update model spectra to best fit observations. We demonstrate end-to-end improvements in forecast skill of significant wave height of 38\%, and up to 45\% for other bulk parameters. This shows distributed observations of the air-sea interface, with advances in assimilation strategies, can reduce uncertainty in forecasts to dramatically improve earth system modeling.

Operational assimilation of spectral wave data from the Sofar Spotter network

Isabel A. Houghton, Christie Hegermiller, Camille Teicheira, Pieter B. Smit

Sofar Ocean Technologies, San Francisco, CA

Key Points:

- A global network of over 600 drifting surface buoys reporting directional wave spectra every hour has been established.
- Assimilation of wave spectra yields quantifiable wave forecast improvements over traditional assimilation using significant wave height.
- Data from a new global ocean sensor and advances in wave data assimilation provide a direct path to improved marine weather forecasts.

Corresponding author: Isabel A. Houghton, `isabel.houghton@sofarocean.com`

Abstract

Historically, the sparseness of in situ open-ocean wave and weather observations has severely limited the forecast skill of weather over the ocean with major social and economic consequences for coastal communities and maritime industries. Ocean surface waves, specifically, are important for the interaction between atmosphere and ocean, and thus key in modeling weather and climate processes. Here, we investigate the improvements achievable from a large distributed sensor network combined with advances in assimilation strategies. Wave spectra from a global network of over 600 Sofar Spotter buoys are assimilated into an operational global wave forecast via optimal interpolation to update model spectra to best fit observations. We demonstrate end-to-end improvements in forecast skill of significant wave height of 38%, and up to 45% for other bulk parameters. This shows distributed observations of the air-sea interface, with advances in assimilation strategies, can reduce uncertainty in forecasts to dramatically improve earth system modeling.

Plain Language Summary

Historically, wave and weather observations are very sparse in the open ocean due to the cost and complexity of instruments and deployments. This lack of real-time weather information results in low-fidelity forecasts. Technological advances have led to the development of the Sofar sensor network, a distributed weather network spanning all the major oceans, consisting of over 600 free-drifting buoys that measure the ocean surface dynamics in great detail (including wave directional spectra). In this work we investigate how such large networks can be successfully used to meaningfully improve forecast accuracy using a new assimilation strategy to ingest the data into operational numerical forecast models. We show substantial improvements in forecast accuracy of the ocean wave field, which has broad implications for earth system modeling and will be directly relevant to coastal communities, marine renewable energy operations, and the efficiency of other maritime industries.

1 Introduction

The ability to observe and accurately predict the dynamics of the ocean interface is critically important for modeling air-sea exchanges, lower-atmosphere dynamics, safety at sea, and mitigation of coastal hazards due to extreme weather events. In general, the skill and accuracy of any weather forecast model fundamentally relies on the availability and successful assimilation of real-time data. In fact, data assimilation (DA) is widely deployed across all disciplines of operational numerical weather prediction and generally contributes as much to the skill of the forecast as the quality of the forecast model itself (Kalnay, 2002; Buizza et al., 2005). With the increase in available data and advances in assimilation strategies, the balance of performance skill will further shift toward data and advances in DA. This work explores how new, globally distributed sensing paradigms combined with advances in assimilation strategies can rapidly accelerate our ability to predict the future state of the air-sea interface.

Despite the importance of the air-sea interface for both ocean and lower-atmosphere dynamics, operational DA in wave models remains uncommon. This is in part due to a lack of suitable data, and in part due to the limitations of existing assimilation strategies that only adjust the total energy of the sea state, but not the distribution of energy. As a result, the benefits of assimilation into wave models is limited (Thomas, 1988; Lionello et al., 1992; Smit et al., 2021). By limiting the assimilation to bulk energy corrections only, traditional wave assimilation cannot address errors across different length scales (e.g. swell or sea components). Consequently, the assimilation improvements usually de-correlate on time scales of typical wind-wave coupling (i.e., under 24 hours) and there is limited value in adding more data to the DA. Fundamental to this, the wave problem is an arbitrary mix of an initial value problem (swell) and boundary value problem (sea), with very different persistence time scales. For example, swell fields exhibit limited interaction with the atmosphere and DA error corrections can persist on the timescale of cross-basin propagation (2-3 weeks). In contrast, shorter waves (sea) are generally strongly coupled to local wind fields, which will dictate persistence of error corrections. To effectively assimilate into a spectral wave model and capture the range of persistence time scales, it is thus critical to correct errors in every component of the spectral distribution.

For such a wave DA strategy to be effective, observations of the wave spectrum are necessary. However, these data have historically been exceedingly sparse – satellite remote sensing is generally limited to bulk parameters (e.g. total energy) and in-situ observations were previously not available in the open ocean. Recently, through advances in mobile technology, satellite communication networks, and improvements in photovoltaic and battery technology, new compact sensor platforms have become available that can deliver scalable, in situ, long-dwell wave spectrum observations. To date, the largest of such wave observing system is the Sofar Spotter network, which is composed of over 600 globally distributed, free-drifting marine weather buoys. (Raghukumar et al., 2019; Voermans et al., 2020; Houghton et al., 2021). This distributed sensor network opens up the opportunity to develop the first operational spectral wave-DA.

Given the historical rarity of buoy spectral information at scale, effective methods to assimilate those data remain uncommon and, to date, have not been widely operationalized. A specific challenge, addressed in the work here, is buoy spectral information is only available as the one-dimensional frequency spectrum and first four directional Fourier components, rather than the two-dimensional frequency-directional (or wavenumber) spectrum that is the model state (Kuik et al., 1988). Thus, an assimilation strategy that relates the observations to the model state is necessary.

Previously, a few studies have explored optimal interpolation based methods for spectra-based assimilation using pitch-and-roll buoy data in a narrow geographic region (Hasselmann et al., 1997; Voorrips et al., 1997). In that method, the analysis was conducted by dividing the spectrum into discrete partitions and updating the model state based on the bulk statistics of each observed partition, substantially reducing the variables describing the wave spectrum. The complexity was then primarily the partitioning of the spectrum and the cross-assignment of partitions between model and observations, which was accomplished with heuristic methods despite possible ambiguities.

Here we present, in tandem, the establishment of a global distributed sensor network and an efficient method for assimilating the observations provided into an operational wave forecast system. This work aims to evaluate the improved forecasting ability made possible by the notable increase in available data, both in terms of geographic coverage and spectral detail. The two step spectra-based DA method outlined here is straightforward to implement and avoids ambiguity with cross-assignment between model and observations. Section 2 describes the buoy network, wave model, and assimilation framework. Results from a month-long reanalysis are presented in Section 3. Finally, impacts and conclusions are described in Section 4.

2 Methods

The DA strategy is built upon the previously established optimal interpolation framework described by Smit et al. (2021), where the initial wave field was updated via sequential optimal interpolation of the observed significant wave height with scaling of the two-dimensional wave action density spectrum to match the analysis wave heights at all grid points. However, this approach has well-documented limitations (Lionello et al., 1992; Portilla-Yandún & Cavaleri, 2016). Specifically, by scaling the spectrum solely by a constant factor derived from the ratio of the analysis wave height to background wave height, model errors in period and direction were left uncorrected. Also, distinct contributions

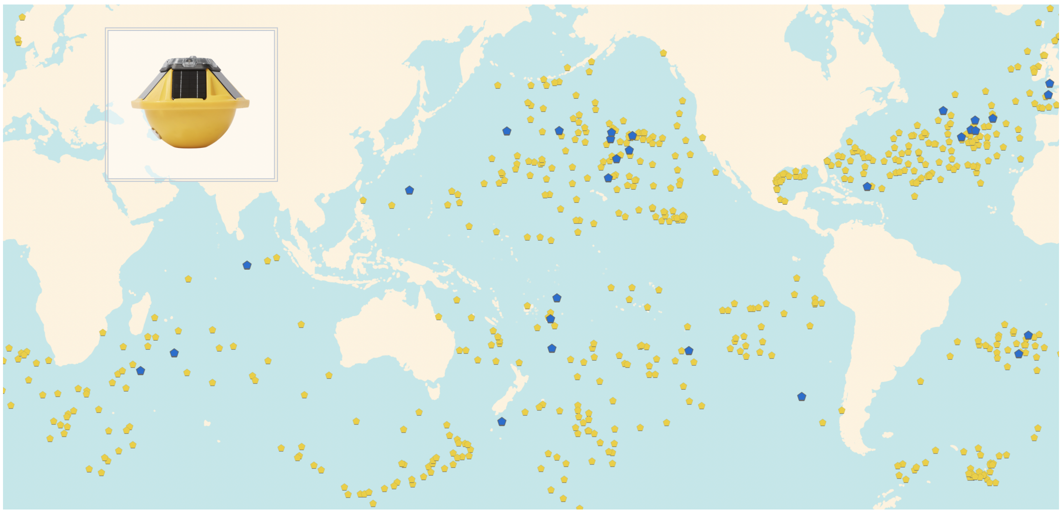

Figure 1. The global Sofar Spotter network (yellow pentagons). Twenty-nine buoys (blue) were randomly selected from the full network to be excluded from the analysis step to provide independent observations to compare with the nowcasts and forecasts. Inset: The 42 cm diameter Spotter buoy represented by pentagonal icons on the map.

to the wave field, e.g. a swell component, could be incorrectly modified despite achieving parity with the bulk significant wave height at that location. While this method was found to produce improvements in both model nowcasts and forecasts, substantially more information is available from the Spotter buoys beyond significant wave height, specifically the variance density spectrum and the four Fourier coefficients. To fully utilize these observational data to update the initial state of the operational forecast model, the optimal interpolation framework is augmented here to update the wave Fourier coefficients on a per-frequency basis and subsequently reconstruct the two-dimensional model spectra.

2.1 Spectral Buoy Data

The wave spectra observations are provided by a global network of free-drifting Spotter buoys developed by Sofar Ocean (Figure 1). The Spotter buoy is lightweight and compact (5.4 kg, 42 cm-diameter approximate sphere) and reports the variance density spectrum and Fourier coefficients along with sea surface temperature, surface drift, barometric pressure, inferred wind and sound level in near-real time (see Raghukumar et al. (2019); Houghton et al. (2021) for further buoy description and validation). The free-drifting buoys provide observations from a network that evolves continuously due to the underlying global currents. Nearly three years of network growth have indicated the sustaining ability to collect long-dwell observations with reliable spatial coverage.

The global Sofar Spotter network surpassed 600 buoys globally in March 2022 and is continuously expanding. The data is stored in a database with a modern API to facilitate operational incorporation of buoy observations at an hourly cadence. As of December 2020, all buoys in the network transmitted spectral data at frequencies from 0.293 Hz to 0.8 Hz, with select buoys transmitting up to 1.25 Hz. The Spotter frequency grid is irregular, with higher resolution bins (0.0098 Hz bins) at frequencies below 0.3 Hz and lower resolution (0.029 Hz bins) at higher frequencies. The frequency dependent variance density spectrum, e^{obs}, and four Fourier coefficients, a_1^{obs}, b_1^{obs}, a_2^{obs}, b_2^{obs} (with $^{\mathrm{obs}}$ denoting observation), at each Spotter location were calculated at thirty minute intervals and reported hourly (i.e. two observations per transmission). Derived wave parameters such as wave height, mean period, and direction are calculated according to standard oceanographic practice ((Kuik et al., 1988)). In order to assimilate the Spotter buoy spectra, the data were interpolated onto the irregular wave model spectral grid (described below) using linear interpolation.

2.2 Wave Forecast Model

The WAVEWATCH3 model (WW3; Tolman et al. (2019)) is implemented over the global ocean at 0.5 degree horizontal resolution and forced by near-surface winds from the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecast System (IFS) High Resolution (HRES) atmospheric and sea ice forecast. The model spectral space is discretized by 36 directions and 36 frequencies. Frequencies are logarithmically distributed with a growth factor of 1.1 from $f_1 = 0.035\,\mathrm{Hz}$ to $f_3 = 0.98\,\mathrm{Hz}$ (see) for full model configuration details). Atmospheric forcing is updated every six hours, at which time a 4- or 10-day operational forecast is initialized from the corresponding analysis for that hour.

The DA uses an hourly analysis cycle. This includes a one-hour WW3 forecast and an instantaneous analysis at the end of each hour to initialize the next forecast. The spectra-based DA method can be summarized as a two step process where (1) the variance density and Fourier coefficients are optimally interpolated for every frequency bin to produce analysis moments and (2) an analysis directional distribution is generated from a cost minimization targeted to match analysis moments and the model background directional distribution. Details of these steps follow.

2.3 Optimal interpolation of Fourier coefficients

We define a reduced background state vector for DA as the variance density and Fourier coefficients at each frequency (e^{bg}, a_1^{bg}, b_1^{bg}, a_2^{bg}, b_2^{bg}, with $^{\mathrm{bg}}$ denoting background). These may be obtained from the full model background state at analysis time through discrete approximations of the Fourier integrals of the directional distribution. Enumerating the N equidistant (resolution $\Delta\theta$) model directions as $\boldsymbol{\theta}^{\mathrm{T}} = [\theta_1, \ldots, \theta_N]$, the discretely sampled directional distribution D_j is defined as $D_j^{\mathrm{bg}}(f;\boldsymbol{x}) = E_j^{\mathrm{bg}}/e^{\mathrm{bg}}$, $E_j^{\mathrm{bg}} = E(f,\theta_j;\boldsymbol{x})$ and

$$e^{\mathrm{bg}}(f;\boldsymbol{x}) = \Delta\theta \sum_{\boldsymbol{\theta}} E^{\mathrm{bg}}(f,\theta;\boldsymbol{x}).$$

The Fourier coefficients, $\boldsymbol{m}$, of the directional distribution then follow as

$$\boldsymbol{m}^{\mathrm{bg}}(f;\boldsymbol{x}) = \begin{bmatrix} (2\pi)^{-1} \\ a_1^{\mathrm{bg}}(f;\boldsymbol{x}) \\ b_1^{\mathrm{bg}}(f;\boldsymbol{x}) \\ a_2^{\mathrm{bg}}(f;\boldsymbol{x}) \\ b_2^{\mathrm{bg}}(f;\boldsymbol{x}) \end{bmatrix} = \Delta\theta \begin{bmatrix} (2\pi)^{-1}\mathbf{1}^{\mathrm{T}} \\ \cos(\boldsymbol{\theta}^{\mathrm{T}}) \\ \sin(\boldsymbol{\theta}^{\mathrm{T}}) \\ \cos(2\boldsymbol{\theta}^{\mathrm{T}}) \\ \sin(2\boldsymbol{\theta}^{\mathrm{T}}) \end{bmatrix} \boldsymbol{D}^{\mathrm{bg}}(f,\boldsymbol{x}) = \mathbf{M}\boldsymbol{D}^{\mathrm{bg}}(f,\boldsymbol{x}) \tag{1}$$

with $\boldsymbol{D}^{\mathrm{T}} = [D_1, \ldots, D_N]$, $\mathbf{1}^{\mathrm{T}} = [1_1, \ldots, 1_N]$ and $\mathbf{M}$ representing the discrete approximation of the Fourier integration. The zeroth coefficient is known a-priori and describes the integration to one of the directional distribution in theta.

The analysis Fourier coefficients $\boldsymbol{m}^{\mathrm{an}}$ ($^{\mathrm{an}}$ denoting analysis) and analysis variance density e^{an} are obtained through optimal interpolation from the analysis equation which – following Smit et al. (2021) – is expressed as

$$\boldsymbol{y}^{\mathrm{an}} = \boldsymbol{y}^{\mathrm{bg}} + \underbrace{\boldsymbol{\rho}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\boldsymbol{\rho}\mathbf{H}^{\mathrm{T}} + \sigma\mathbf{I}\right)^{-1}}_{\mathbf{K}}\left(\mathbf{H}\boldsymbol{y}^{\mathrm{bg}} - \boldsymbol{y}^{\mathrm{obs}}\right) \tag{2}$$

Here $\boldsymbol{y}(f)$ (analysis or background) is the state vector of the model with M grid points for a given frequency, and $\boldsymbol{y}^{\mathrm{obs}} = [y_j^{\mathrm{obs}}(f), \ldots, y_J^{\mathrm{obs}}(f)]^{\mathrm{T}}$ denotes the J observations of the state. Further, $\mathbf{H}$ is a J x M bi-linear interpolation matrix that projects model estimates to observed locations. Lastly, $\mathbf{K}$ is the M x J Kalman Gain matrix that is dependent upon model error correlation, $\boldsymbol{\rho}$, and relative observation errors $\sigma\mathbf{I}$. Here, I is the identity matrix and σ (set to 0.3 here, see Smit et al. (2021)) represents the observational error scaled with a representative model error. Equivalent equations to (2) are used for the Fourier coefficients a_1^{an}, b_1^{an}, etc.

Optimal interpolation requires a-priori specification of the error-covariances (correlations here), which in general are non-trivial to determine. Here, we take ρ to be isotropic, stationary, homogeneous and independent of frequency, and use a parameterized form as in Smit et al. (2021) that de-correlates over a characteristic distance of 300 km. Further, inter-coefficient errors are assumed to be uncorrelated, allowing for independent application of (2) to individual moments.

2.4 Directional Reconstruction

The OI step performs DA in observational space. To return to model space, a subsequent step is needed to reconstruct the two-dimensional directional spectra at each model grid point to serve as the initial condition for the forecast. However, the analysis Fourier coefficients, $\boldsymbol{m}^{\mathrm{an}}$, do not uniquely determine the analysis directional distribution because $\mathbf{M}$ is under-determined and not invertible. To uniquely specify the directional distribution, we assume that the model background distribution estimation, D^{bg}, is in general skillful, and seek a distribution that minimizes the difference with the model background under the constraints that D^{an} reproduces the analysis Fourier coefficients and is positive semi-definite. Considering a single frequency at a single location $\boldsymbol{x}$, the analysis directional distribution is the solution of the quadratic-programming problem,

$$\begin{array}{cc} \min_{\boldsymbol{D}^{\mathrm{an}}} & [\boldsymbol{D}^{\mathrm{an}} - \boldsymbol{D}^{\mathrm{bg}}]^{\mathrm{T}}[\boldsymbol{D}^{\mathrm{an}} - \boldsymbol{D}^{\mathrm{bg}}] \\ \text{subject to} & \mathbf{M}\boldsymbol{D}^{\mathrm{an}} = \boldsymbol{m}^{\mathrm{an}} \\ & \boldsymbol{D}^{\mathrm{an}} \geq 0 \end{array} \tag{3}$$

In practice, the reproduction of the Fourier coefficients is applied as a cost in addition to the difference from the background directional distribution and a least-squares bounded minimization is used. Following Equation 3, an analysis directional distribution is generated for every model grid point. To return to the two-dimensional spectrum, the directional distribution is then multiplied by $e^{\mathrm{an}}(f)$, provided explicitly from the optimal interpolation step.

The optimization approach with constraints, inspired by (Crosby et al., 2017), is chosen over other methods, such as maximum entropy estimation (Lygre & Krogstad, 1986), as it allows for the inclusion of the additional information provided by the model background. This assumes that although the model may be incorrect, it provides a reasonable starting point to generate the analysis distribution and further encourages spatial coherence across the geographic domain despite each grid point being updated independently. Further, this formulation is sufficiently computationally efficient to remain within operational time constraints.

2.5 Reforecast Experiment

The spectra-based data assimilation scheme is evaluated with an approximately 32 day reforecast experiment starting February 20th, 2022 and ending March 24, 2022. Three experiments are run in order to assess the impact of the DA methods: a free-running forced WW3 model forecast, an hourly-cycled DA case assimilating significant wave height observations (henceforth H_s-based), and an hourly-cycled DA case assimilating wave spectra observations as described above (henceforth spectra-based). For each experiment, a

4-day forecast is initialized every 12 hours from the analysis state (or forecast state in the case of the free-running model). Twenty-nine Spotter buoys are excluded from the DA experiments for evaluation. To ensure global coverage of excluded buoys, all buoys are first binned into ten regions by latitude and longitude, and a random selection of 10% in that bin were chosen to be excluded (see Figure 1). In addition to the bulk parameters output hourly over the entire model domain, two-dimensional model spectra are output hourly at the excluded Spotter locations (with buoy drift neglected over forecast timescales).

Forecast skill is evaluated by point-wise comparison of modeled variables to the Spotter observations. Spotter observations are linearly interpolated to the nearest hour and the modeled fields are bilinearly interpolated to the Spotter latitude and longitude.

To assess model skill in different frequency ranges, specifically low frequency swell energy versus high frequency wind sea energy, the observed and modeled variance density spectra ($e(f)$) are partitioned at 0.08789 Hz. Only observations for which the Spotters reported the presence of swell are used to calculate the corresponding root-mean-square error of these partitioned sea states. Following methods from Portilla et al. (2009) for partitioning one-dimensional spectra, an estimate of the ratio between the peak energy of the wave system and the peak energy of a Pierson–Moskowitz spectrum with the same peak frequency, $\gamma^\star$, is calculated as an indicator of swell presence. Observations with a $\gamma^\star$ <0.5 are used to select the observations for assessment of swell forecast skill (see Supplement for further details).

3 Results

3.1 Spectral Updates

The optimal interpolation step updates the frequency-binned moments to balance between the model background and observations. In general, this does not exactly matching either owing to the uncertainty prescribed to both the observations and forecasts in the relative standard deviation of the errors.

Spotter-0890 (Figure 2a), which was excluded from the DA experiments, illustrates the impact of the assimilation of each observation type. For the spectra in Figure 2b, the observed and non-assimilated modeled variance density spectra are different. For the H_s-based assimilation, the distribution is altered with higher energy at the peak frequency and lower energy at the higher frequencies, still different from the Spotter observation. The variance density spectrum for the spectra-based assimilation (blue line), however, closely matches the Spotter observation of the peak frequency as well as the distribution of energy across frequencies, particularly capturing the wind-sea peak at higher frequencies. Further, the H_s-based assimilation does little to improve agreement of the Fourier coefficients with the Spotter observations (Figure 2c-f) while the spectra-based assimilation results in a notable qualitative improvement in agreement. Finally, the two-dimensional spectra from the non-assimilated model (g), spectra-based assimilated model (h), and their difference (i) illustrates the impact of assimilation of spectral information from a network of buoys. Specifically, energy was modified in both direction and frequency space – decreasing the energy and shifting to slightly lower frequencies around 300°, removing a swell field around 100°, and introducing a swell field around 200°.

3.2 Improvements to Bulk Statistics

Direct validation of the analysis two-dimensional spectra, such as shown in Figure 2h, remains challenging because directional wave buoys, like the Spotter, only provide the Fourier coefficients. However, an improved two-dimensional spectrum in the model

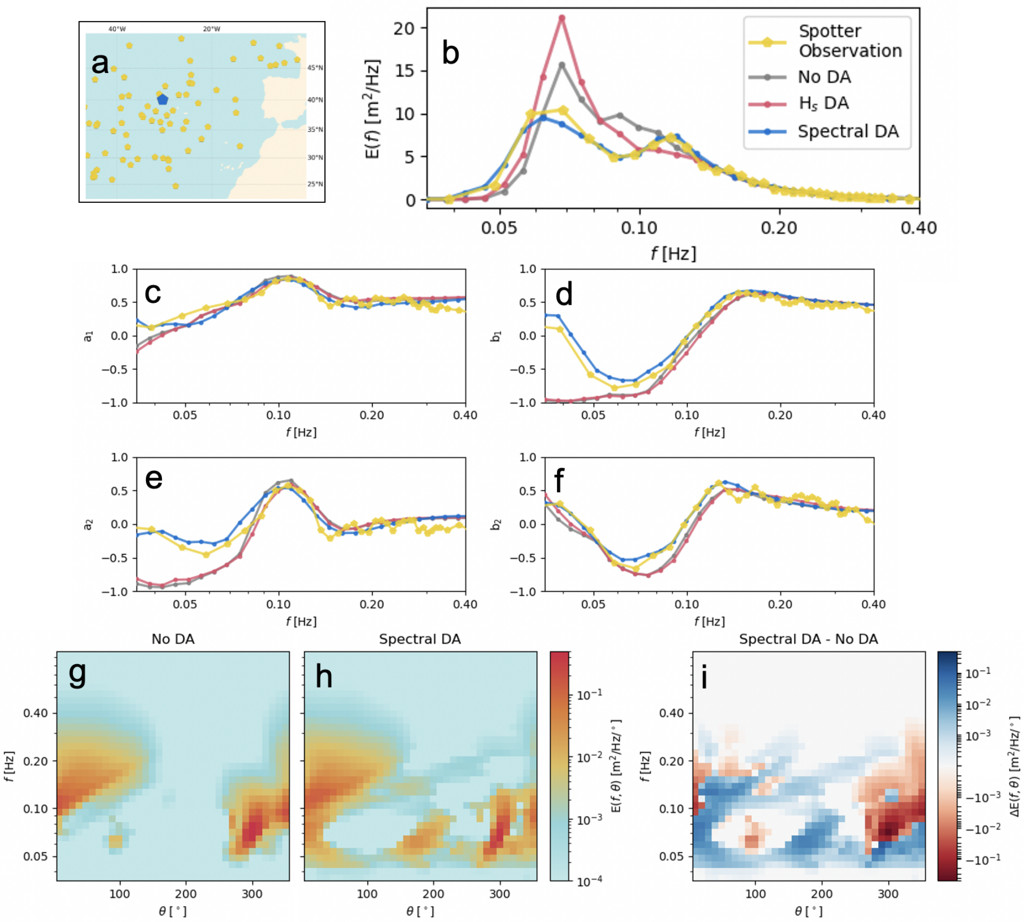

Figure 2. Model states from the different WW3 experiments at an excluded Spotter buoy (SPOT-0980) on Friday, March 4, 2022 12:00 UTC. (a) The location of the Spotter in the North Atlantic. (b) The variance density spectrum and (c-f) Fourier coefficients of the Spotter (yellow), free-running WW3 forecast (grey), wave height-assimilated (pink) and spectra-assimilated (blue). Moments are calculated from the WW3 model spectra. (g-h) The two-dimensional wave spectrum from the free-running WW3 forecast and spectra-assimilated DA case. (i) The difference between the two spectra.

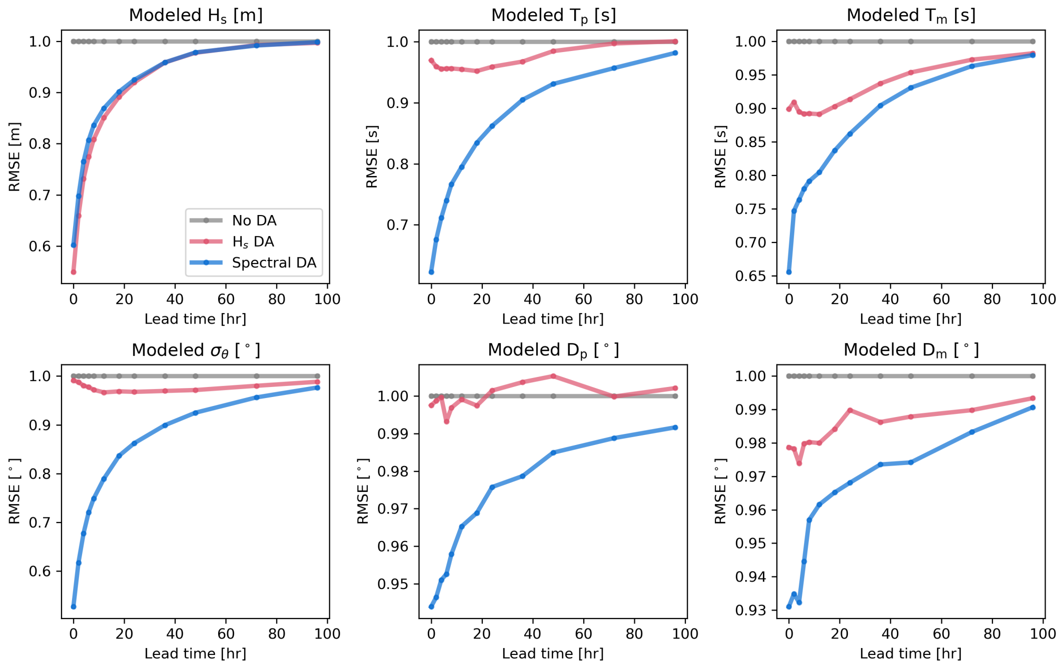

Figure 3. Root-mean-square error (RMSE) of bulk wave parameters in the analyses and forecasts up to four days from all three WW3 experiments. No assimilation (grey), H$_s$-based assimilation (pink), and spectra-based assimilation (blue) were assessed at all Spotter locations. Approximately 25,000 observation-model pairs were used to estimate the RMSE.

will propagate forward in time and space and manifest in the bulk parameters of the downstream wave field.

Substantial improvements are observed in forecast skill when evaluated against Spotter bulk parameter observations (Figure 3). Significant wave height error in the analysis is reduced by approximately 44% by the H$_s$-based assimilation approach and 38% by the spectra-based approach. At 24-hour lead times, the error is reduced by 8.2% and 7.5% for H$_s$-based and spectra-based, respectively. At even longer lead times, the error reductions decay asymptotically to zero, with negligible forecast skill improvement beyond 4 days.

Five other bulk parameters – peak period T_p, mean period T_m, directional width σ_θ, peak direction D_p, and mean direction D_m – consistently exhibited the largest error reductions in the spectra-based DA case, with up to 45% reduction in errors for directional width in the analyses and persistent reductions of 1-2% in 4-day forecasts across bulk parameters.

The full advantage of the spectra-based approach is illustrated in the bulk parameters describing period and direction. The H$_s$-based approach does lead to some improvements in these bulk parameters despite no direct incorporation of this information into the assimilation scheme. Specifically, the H$_s$-based approach scales the energy spectrum equivalently across all frequencies, therefore not initially impacting the peak direction or frequency. However, as different portions of the wave spectrum relax to the forcing field (wind) at different rates (the higher frequencies adjusting the most rapidly), the scaling of the energy spectrum and subsequent relaxation to the background forcing will ultimately modify the shape of the energy spectrum, in turn impacting the period and direction properties of the wave field. This evolution of the spectra results in the interme-

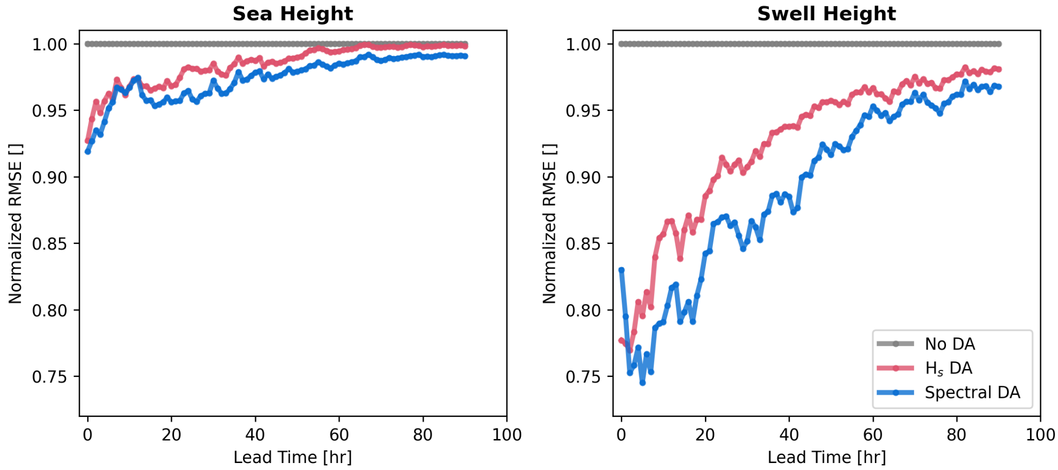

Figure 4. Wind sea significant wave height (left) and swell significant wave height (right) root-mean-square error normalized against the non-assimilated error (grey) for the H$_s$-based assimilation (pink) and spectra-based assimilation (blue) for observations with swell energy present (see supplement for further details).

diate improvements to the frequency and direction properties following the H$_s$-based approach. The spectra-based approach, on the other hand, explicitly updates the spectrum to better match the Fourier coefficients, manifesting in marked improvements to all bulk parameters. In particular, the improvements to the bulk parameters extend to longer lead times, indicating the value of correcting the frequency and direction information to subsequently propagate across the geographic domain. While most modeling efforts are evaluated in terms of significant wave height (likely because this is the primary open ocean data available), other parameters of the wave field are equally important to accurately represent (e.g., large container vessels can be extremely vulnerable to specific frequency waves even at low magnitudes, swell can steer wind stress, and short waves impact air-sea fluxes).

The approximately equivalent performance of the two assimilation strategies (H$_s$-based and spectra-based) when evaluated on just significant wave height is likely a result of the spectra-based approach having additional constraints beyond the significant wave height target. Competing costs in reconstructing the directional distribution would then lead to less direct matching of the bulk parameter of significant wave height, despite better agreement with the spectral shape, with the largest impact at the zero-hour lead time.

While the bulk statistic of total significant wave height is most effectively addressed by H$_s$-based assimilation (Figure 3), when we consider the significant wave height of wind sea (higher frequency) versus swell (lower frequency), a differentiation of the effectiveness of the two assimilation methods becomes clear (Figure 4). Because the wind seas are tightly coupled to the surface winds, any modifications to the initial condition of the high frequency wave field rapidly relax to the wind forcing. However, the propagation of low frequency swell waves is, to the first-order, an initial value problem, and therefore ideally suited to improvement via DA. By updating the wave fields with spectra-based assimilation, the initial state of the swell is better represented and more accurately propagated forward in time. The error of specifically swell-containing events was reduced up to 25% in the analysis, with persistent improvement of approximately 5% out to four days (Figure 4).

4 Discussion and Conclusions

Accurately predicting marine weather is critical to industry, society and the environment – from reducing global shipping emissions and safety risks, to mitigating coastal hazards. Observations and their effective utilization in numerical models play an outsize role in progressing forecasting ability and, for the first time, in situ observations of directional wave spectra are available in the open ocean at a sufficient density for impact at global scales. The operational assimilation scheme described here specifically illustrates the capacity for wave spectral observations to improve forecast accuracy of bulk parameters and spectral characteristics. The incorporation of the wave spectral data in the operational assimilation scheme quantitatively improves the forecast skill of significant wave height up to 38% over the free-running WW3 model, and was further shown to outperform the H_s-based DA in forecasting period and direction, with particular success for swell-dominated fields.

This work focuses on demonstrating the impact of distributed spectral observations on wave forecast skill, but the potential for improvements is not limited to waves alone. All interactions between oceans and the atmosphere are influenced by the ocean surface (Cavaleri et al., 2012), with exchange processes typically strongly dependent on the spectral distribution of energy. Consequently, through coupled data assimilation, a path exists to use spectral observations to improve exchanges between ocean and atmosphere, thus improving earth system modeling more broadly. Overall, this work describes the realization of observational networks to provide the needed data with proven accuracy and reliability for such advances in operational models and lays the groundwork for broad progress in coupled earth systems modeling.

Acknowledgments

Historical data from Spotter buoys, including those used in this study, is freely available for research use by contacting Sofar Ocean Technologies (www.sofarocean.com).

All data used in this analysis will be made openly available.

References

Buizza, R., Houtekamer, P. L., Toth, Z., Pellerin, G., Wei, M., & Zhu, Y. (2005, 5). A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. *Monthly Weather Review*, *133*(5), 1076–1097. Retrieved from `https://journals.ametsoc.org/view/journals/mwre/133/5/mwr2905.1.xml` doi: 10.1175/MWR2905.1

Cavaleri, L., Fox-Kemper, B., & Hemer, M. (2012). Wind waves in the coupled climate system. *Bulletin of the American Meteorological Society*. doi: 10.1175/BAMS-D-11-00170.1

Crosby, S. C., Cornuelle, B. D., O'Reilly, W. C., & Guza, R. T. (2017, 8). Assimilating Global Wave Model Predictions and Deep-Water Wave Observations in Nearshore Swell Predictions. *Journal of Atmospheric and Oceanic Technology*, *34*(8), 1823–1836. Retrieved from `https://journals.ametsoc.org/view/journals/atot/34/8/jtech-d-17-0003.1.xml` doi: 10.1175/JTECH-D-17-0003.1

Hasselmann, S., Lionello, P., & Hasselmann, K. (1997). An optimal interpolation scheme for the assimilation of spectral wave data. *Journal of Geophysical Research C: Oceans*, *102*(C7), 15823–15836. doi: 10.1029/96JC03453

Houghton, I. A., Smit, P. B., Clark, D., Dunning, C., Fisher, A., Nidzieko, N. J., ... Janssen, T. T. (2021, 5). Performance Statistics of a Real-Time Pacific Ocean Weather Sensor Network. *Journal of Atmospheric and Oceanic Technology*, *38*(5), 1047–1058. Retrieved from `https://journals.ametsoc.org/view/journals/atot/aop/JTECH-D-20-0187.1/JTECH-D-20-0187.1.xml` doi: 10.1175/JTECH-D-20-0187.1

Kalnay, E. (2002, 11). Atmospheric Modeling, Data Assimilation and Predictability. *Atmospheric Modeling, Data Assimilation and Predictability*. doi: 10.1017/CBO9780511802270

Kuik, A. J., van Vledder, G. P., & Holthuijsen, L. H. (1988). A Method for the Routine Analysis of Pitch-and-Roll Buoy Wave Data. *Journal of Physical Oceanography*. doi: 10.1175/1520-0485(1988)018⟨1020:amftra⟩2.0.co;2

Lionello, P., Gunther, H., & Janssen, P. A. (1992). Assimilation of altimeter data in a global third-generation wave model. *Journal of Geophysical Research*. doi: 10.1029/92jc01055

Lygre, A., & Krogstad, H. E. (1986). Maximum Entropy Estimation of the Directional Distribution in Ocean Wave Spectra. *Journal of Physical Oceanography*, *16*(12), 2052–2060. Retrieved from `https://journals.ametsoc.org/view/journals/phoc/16/12/1520-0485_1986_016_2052_meeotd_2_0_co_2.xml` doi: 10.1175/1520-0485(1986)016⟨2052:MEEOTD⟩2.0.CO;2

Portilla, J., Ocampo-Torres, F. J., & Monbaliu, J. (2009, 1). Spectral Partitioning and Identification of Wind Sea and Swell. *Journal of Atmospheric and Oceanic Technology*, *26*(1), 107–122. Retrieved from `https://journals.ametsoc.org/view/journals/atot/26/1/2008jtecho609_1.xml` doi: 10.1175/2008JTECHO609.1

Portilla-Yandún, J., & Cavaleri, L. (2016, 1). On the specification of background errors for wave data assimilation systems. *Journal of Geophysical Research: Oceans*, *121*(1), 209–223. doi: 10.1002/2015JC011309

Raghukumar, K., Chang, G., Spada, F., Jones, C., Janssen, T., & Gans, A. (2019). Performance characteristics of "spotter,"' a newly developed real-time wave measurement buoy. *Journal of Atmospheric and Oceanic Technology*, *36*(6), 1127–1141. doi: 10.1175/JTECH-D-18-0151.1

Smit, P. B., Houghton, I. A., Jordanova, K., Portwood, T., Shapiro, E., Clark, D., ... Janssen, T. T. (2021, 3). Assimilation of significant wave height from distributed ocean wave sensors. *Ocean Modelling*, *159*, 101738. doi: 10.1016/J.OCEMOD.2020.101738

Thomas, J. P. (1988, 4). Retrieval of energy spectra from measured data for assimilation into a wave model. *Quarterly Journal of the Royal Meteorological Society*, *114*(481), 781–800. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/full/10.1002/qj.49711448112https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49711448112https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.49711448112 doi: 10.1002/QJ.49711448112

Tolman, H., Abdolali, A., Accensi, M., Alves, J.-H., Ardhuin, F., Babanin, A., . . . Liang, Z. (2019). *User manual and system documentation of WAVEWATCH III (R) version 6.07* (Tech. Rep.). Retrieved from https://www.researchgate.net/publication/336069899_User_manual_and_system_documentation_of_WAVEWATCH_III_R_version_607

Voermans, J. J., Smit, P. B., Janssen, T. T., & Babanin, A. V. (2020). Estimating Wind Speed and Direction Using Wave Spectra. *Journal of Geophysical Research: Oceans*. doi: 10.1029/2019JC015717

Voorrips, A. C., Makin, V. K., & Hasselmann, S. (1997). Assimilation of wave spectra from pitch-and-roll buoys in a North Sea wave model. *Journal of Geophysical Research C: Oceans*, *102*(C3), 5829–5849. doi: 10.1029/96JC03242

1 Supplementary

Following Portilla et al. (2009), an estimate of the ratio between the peak energy of the wave spectrum and the peak energy of a Pierson–Moskowitz (PM) spectrum with the same peak frequency, $\gamma^\star$, was calculated. Specifically,

$$\gamma^\star = \frac{E_{obs}(f_p)}{\alpha g^2 (2\pi)^{-4} f_p^{-5} e^{-5/4} \gamma} \tag{1}$$

where $\gamma = 1$ and $\alpha = \alpha_{PM} = 0.0081$.

In that work, observations with a $\gamma^\star < 1$ were considered swell. Here, a stricter threshold of $\gamma^\star < 0.5$ was used to select the observations at which swell forecast skill was assessed in order to clearly assess wave fields where energy was present in the lower frequency bins. The relationship between $\gamma^\star$ and the distribution of energy across frequencies can be seen in Figure 1, where observations with a peak period proportionally higher than the mean period had a lower $\gamma^\star$.

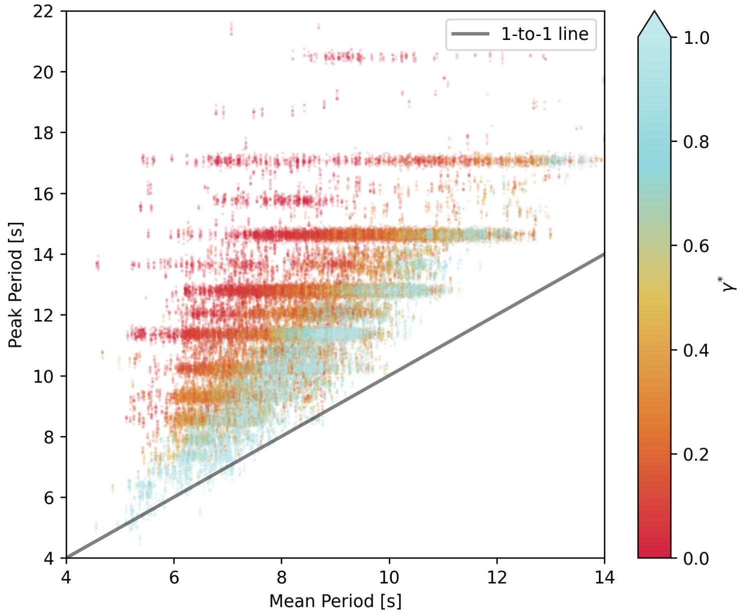

Figure 1. Scatter plot of mean period and peak period from the Spotter observations used in the sea versus swell significant wave height analysis, colored by $\gamma^\star$.

References

Portilla, J., Ocampo-Torres, F. J., & Monbaliu, J. (2009, 1). Spectral Partitioning and Identification of Wind Sea and Swell. *Journal of Atmospheric and Oceanic Technology*, *26*(1), 107–122. Retrieved from `https://journals.ametsoc.org/view/journals/atot/26/1/2008jtecho609_1.xml` doi: 10.1175/2008JTECHO609.1