J. Korean Earth Sci. Soc., v. 45, no. 4, p. 349–362, August 2024
https://doi.org/10.5467/JKESS.2024.45.4.349
ISSN 1225-6692 (printed edition)
ISSN 2287-4518 (electronic edition)

Sensitivity of Data Assimilation Configuration in WAVEWATCH III applying Ensemble Optimal Interpolation

Hye Min Lim[1], Kyeong Ok Kim[2], Hanna Kim[2], Sang Myeong Oh[3], and Young Ho Kim[1,*]

[1]Pukyong National University, Busan 48513, Republic of Korea
[2]Korea Institute of Ocean Science and Technology, Busan 49111, Republic of Korea
[3]National Institute of Meteorological Sciences, KMA, Seogwipo 63568, Republic of Korea

Abstract: We aimed to evaluate the effectiveness of ensemble optimal interpolation (EnOI) in improving the analysis of significant wave height (SWH) within wave models using satellite-derived SWH data. Satellite observations revealed higher SWH in mid-latitude regions (30° to 60° in both hemispheres) due to stronger winds, whereas equatorial and coastal areas exhibited lower wave heights, attributed to calmer winds and land interactions. Root mean square error (RMSE) analysis of the control experiment without data assimilation revealed significant discrepancies in high-latitude areas, underscoring the need for enhanced analysis techniques. Data assimilation experiments demonstrated substantial RMSE reductions, particularly in high-latitude regions, underscoring the effectiveness of the technique in enhancing the quality of analysis fields. Sensitivity experiments with varying ensemble sizes showed modest global improvements in analysis fields with larger ensembles. Sensitivity experiments based on different decorrelation length scales demonstrated significant RMSE improvements at larger scales, particularly in the Southern Ocean and Northwest Pacific. However, some areas exhibited slight RMSE increases, suggesting the need for region-specific tuning of assimilation parameters. Reducing the observation error covariance improved analysis quality in certain regions, including the equator, but generally degraded it in others. Rescaling background error covariance (BEC) resulted in overall improvements in analysis fields, though sensitivity to regional variability persisted. These findings underscore the importance of data assimilation, parameter tuning, and BEC rescaling in enhancing the quality and reliability of wave analysis fields, emphasizing the necessity of region-specific adjustments to optimize assimilation performance. These insights are valuable for understanding ocean dynamics, improving navigation, and supporting coastal management practices.

Keywords: Significant Wave Height, Ensemble Optimal Interpolation, Ensemble Size, Decorrelation Length Scale, Background Error Covariance Rescaling

1. Introduction

Accurate predictions of significant wave height (SWH) play a critical role across various industries and in coastal management practices. SWH, which represents the average height of the highest third of waves in a wave field and serves as a key indicator of wave energy, is essential for understanding and preparing for wave conditions. Advancements in remote sensing and sensor technology have revolutionized the collection of SWH data, enabling the integration of data from satellites and distributed wave sensors into wave models through a process known as data assimilation (Daniel et al., 2011; Hisaki, 2005; Mitsuyasu et al., 1980; Rapizo et al., 2015; Walsh et al., 1989; Gommenginger et al., 2003; Lzaguirre et al., 2011; Queffeulou, 2004). This integration presents a promising opportunity to enhance the accuracy of SWH predictions, thus improving decision making in industries dependent on wave conditions and coastal management strategies.

However, assimilating data from satellites and distributed wave sensors using wave models poses significant challenges. These challenges stem from potential differences in spatial and temporal resolution, as well as errors in measurements that must be

*Corresponding author: yhokim@pknu.ac.kr
Tel: +82-51-629-6571

carefully considered and corrected during the data assimilation process. Despite these obstacles, utilizing satellite and distributed wave sensor data in data assimilation has the potential to significantly boost the precision of SWH forecasts, thereby providing valuable insights into wave behavior and trends.

Wave models such as SWAN (Booij et al., 1999) and WAVEWATCH III (The WAVEWATCH III® Development Group, 2019) (WW3) are widely used for various applications, including weather forecasting, offshore platform operations, marine activities, and disaster prevention. However, these modelling systems have limited performance in prediction due to computational costs and inaccuracies in the input data, including inaccuracies in the wind field used for forcing. It is imperative to refine and calibrate the numerical models by adjusting the parameters and initial conditions. Data assimilation emerges as a crucial technique in this process, as it combines model data with observations to create an analysis field that best represents the state required for accurate numerical model calculations.

Data assimilation methods play a crucial role in enhancing wave prediction accuracy by integrating diverse data sources into wave models. In particular, optimal interpolation (OI), three-dimensional variational assimilation (3D-Var), and four-dimensional variational assimilation (4D-Var) are widely employed methods, with OI standing out due to its operational ease and relatively low cost (Bouttier and Courtier, 2002). Furthermore, ensemble optimal interpolation (EnOI) offers advantages in terms of saving computing time (Evensen, 2003). However, the integration of satellite and distributed wave sensor data presents challenges due to their unique characteristics and potential errors. Despite these challenges, incorporating these data potentially leads a significant advancement in improving wave prediction capabilities.

For ocean forecasting, OI and 3D-Var methods are commonly used due to their computational efficiency (Cummings and Smedstad, 2009). However, these methods have limitations in capturing ocean states because they rely on Gaussian background error covariance (BEC). To address these limitations, there is growing interest in Kalman filter-type assimilation methods and 4D-Var, which consider flow-dependent BEC. The ensemble Kalman filter (EnKF), proposed by Evensen (1994), is a practical data assimilation method that forecasts error statistics using the Monte Carlo method. Despite its advantages, the original version of the EnKF has issues with rank deficiency, which can lead to ensemble collapse when using a small number of members. Various approaches, such as addressing model errors, localization, background error inflation, and adjusting new prior probability density functions, have been suggested to resolve this deficiency.

In situations with limited computational resources, EnOI offers a cost-effective alternative to EnKF (Oke et al., 2007; Kim et al., 2015). EnOI estimates BEC using a stationary ensemble, in contrast to the ensemble model runs of EnKF. To account for flow dependence based on seasonality, Xie and Zhu (2010) introduced seasonally varying quasi-stationary ensembles to the EnOI.

Wittmann and Cummings (2005) utilized the OI method to successfully assimilate altimeter data into a global WW3 model, validating their approach using buoy data. Waters et al. (2013) applied OI to integrate high-frequency radar data into a high-resolution WW3 model in the Celtic Sea, with a particular emphasis on refining the background error covariance matrix to accurately reflect local conditions. Yu et al. (2018) demonstrated that using multiple satellite datasets in a global wave model improved prediction accuracy, while Li and Zhang (2020) employed OI to correct model bias in global wave models using satellite data from Jason-2, Jason-3, and SARAL. Qi and Cao (2015) applied EnOI in the Indian Ocean, focusing on the seasonal impacts of assimilating Jason-2 data. Liu et al. (2023) further advanced this by implementing a local EnOI scheme to assimilate significant wave heights from HY-2B altimeter data in the China Seas. Saulter et al. (2020) used 3DVAR in a regional wave model for the northwest European shelf, exploring the added value of integrating in situ observations alongside

satellite data.

This study aims to evaluate the effectiveness of EnOI in improving the analysis of SWH within wave models using satellite-derived SWH data. By exploring various data assimilation configurations and assessing their sensitivity, this study seeks to contribute to the development of enhanced strategies for SWH analysis, which can be valuable for industries and coastal management reliant on accurate wave forecasts. Building on existing knowledge of data assimilation methods and their applications in wave modeling, this study aspires to enhance the reliability and precision of wave analysis fields.

2. Data and Methods

2.1. WAVEWATCH Ⅲ

In this study, we utilized the WW3 developed by the National Oceanic and Atmospheric Administration (NOAA) and the United States National Centers for Environmental Prediction (NCEP). WW3 is a sophisticated tool designed to predict various characteristics of ocean waves. By employing mathematical equations, the model can simulate wave behavior based on factors such as winds, currents, and bathymetry. As a result, it can generate forecasts for different weather and ocean conditions. WW3 is widely used in marine safety, offshore engineering design, and wave forecasting, offering global-scale forecasts various days in advance, as well as high-resolution regional predictions for specific coastal areas.

The WW3 marks a significant advancement in wave modeling compared to its predecessors, WAVEWATCH I and II. WW3 incorporates various key improvements in its governing equations, model structure, numerical methods, and physical parameterizations (Komen et al., 1994; WAMDI Group, 1988). It is considered a comprehensive framework that incorporates the latest scientific advancements in wind-wave modeling and dynamics.

In this study, we used the WW3 model with reanalyzed wind fields obtained from the Korea Meteorological Administration. These wind fields have a spatial resolution of 1/4° and are updated hourly. To ensure stable conditions for our simulations, we implemented a spin-up period of 1 year. After achieving stability, model simulations were performed to generate the necessary data for 12 months. These results were then used as the historical ensembles for calculating the background error covariance in the EnOI. The data assimilation experiments and historical ensemble run were conducted from May 2020 to April 2021.

2.2. Data Assimilation

In this study, EnOI was employed to assimilate SWH obtained from satellite altimeters. The performance of EnOI was evaluated by assessing its effectiveness under various configurations and comparing it with the results from optimal interpolation. Representing the Kalman gain matrix as K, the optimal wave analysis field (φ^a) is given as follows (Eqs. [1] and [2]):

$$\varphi^a = \varphi^f + \mathrm{K}(d - \mathrm{H}\varphi^f) \tag{1}$$

$$\mathrm{K} = \mathrm{P}^f H^T [\mathrm{H}\mathrm{P}^f \mathrm{H}^T + \mathrm{R}]^{-1} \tag{2}$$

where φ^f represents the forecast state vector. d is the observation vector and H is the observation operator that transforms from the model grid to the observation grid. P^f and R represent the background and observation error covariance matrices, respectively. The superscript T denotes the matrix transpose.

EnOI is based on the work of Evensen (2003) and the analysis approach of Burgers et al. (1998). In this study, we develop an OI scheme in which the analysis is computed in the space spanned by a stationary ensemble of sampled model states over a long integration period. EnOI estimates the BEC matrix by creating a stationary and historical ensemble where statistical errors do not develop over time. EnOI estimates the BEC matrix P^f as follows (Eqs. [3] and [4]):

$$\mathrm{P}^f = \alpha \mathrm{A}'\mathrm{A}'^T/(N_e - 1) \tag{3}$$

$$\mathrm{A}' = [\varphi'_1, \varphi'_2, \ldots, \varphi'_{N_e}] \tag{4}$$

where N_e is the ensemble size, α ($\in$0,1) is introduced

as a scaling factor to reduce the variance (Counillon and Bertino, 2009), A' is a stationary and historical ensemble, and φ'_i is the i-th ensemble member from historical model run. The historical ensemble members were extracted from the historical run data, selecting daily data both before and after the assimilation date, centered around it, to ensure the required number of ensemble members is obtained. In this study, the scaling factor α was set to 0.25 (Oke et al., 2002; Kim et al., 2015). To derive the BEC, ensemble members were chosen from one year of pre-run model integration from May 2020 to May 2021 without data assimilation. For example, the ensemble consisted of N_e members, with $N_e/2$ members representing the days before the target day and the remaining $N_e/2$ members representing the days after the target day.

An ensemble-based data assimilation system can over-estimate correlations between long-distance points that are likely to be independent of each other when a small number of ensemble members are used. To reduce this sampling error, a localization technique (Houtekamer and Mitchell, 2001; Hamill et al., 2001; Oke et al., 2007; Kim et al., 2015) was applied using the Schur product formulation (Gaspari and Cohn, 1999). After applying localization, the Kalman gain matrix was expressed as follows (Eqs. [5] and [6]):

$$\rho^{\circ}\begin{bmatrix} b_{1,1} & .. & b_{i,1} & .. & b_{M,1} \\ .. & .. & .. & .. & .. \\ b_{1,j} & .. & b_{i,j} & .. & b_{M,j} \\ .. & . & . & . & .. \\ b_{1,M} & . & b_{i,M} & . & b_{M,M} \end{bmatrix} = \rho\,,\ [L(x_i,x_j)] \times b_{i,j} \qquad (5)$$

$$\mathrm{K} = \rho^{\circ}\mathrm{A'(HA')}^{T}(\rho^{\circ}\mathrm{HA'(HA')}^{T} + (N_e - 1)\mathrm{R})^{-1} \qquad (6)$$

where ρ°B is calculated as a function of the distance (L) between x_i and x_j. ρ is given as a function of the decorrelation length scale, which has a value of 0-1. As the distance between two points increases, ρ approaches 0, and when the distance exceeds a certain length, ρ becomes 0.

During the data assimilation process, a data assimilation system was constructed by assimilating the SWH observed by satellite with EnOI to produce an analysis field, which was then used as the initial field for the subsequent model run (Fig. 1).

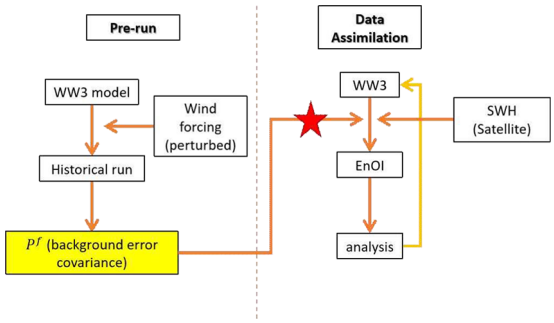

Fig. 1. Schematic diagram of data assimilation system in WAVEWATCH III (WW3). SWH, significant wave height

We assimilated SWH data from satellite altimeters provided by the Copernicus Marine Service. The altimetric data from Jason-3, Sentinel-3A, AltiKa, and CryoSat-2 satellites were utilized for data assimilation, which were provided by the Copernicus Marine Service; the spatial resolution of these data is less than 7 km. Jason-3 and Sentinel-3A orbit the Earth every 10 and 27 days, respectively, whereas AltiKa and CryoSat-2 do not have a constant period.

2.3. BEC rescaling

Generally, in the EnOI (Evensen, 2003), the BEC does not change over time. However, in this study, a data assimilation system was developed to obtain optimal weights by adjusting the BEC to consider the innovation term, which represents the difference between the model and the observation.

The innovation ($d - \mathrm{H}\varphi^f$) at the observed point can be expressed as the sum of the observation error ($\varepsilon^o = \varphi^o - \varphi^t$) and the background error ($\varepsilon^f = \varphi^f - \varphi^t$) where φ^t is the true state (Fig. 4 in Bouttier and Courtier, 2002). As the observation error covariance increases, the value of the BEC decreases for the innovation term. Conversely, when the BEC increased, the observed error covariance decreased. This means that if the BEC is large and the observation error covariance is small, the observation will be assigned greater weight in the data assimilation process, potentially reducing the weight of the background

field. The BEC rescaling helps mitigate the problem of overfitting, in which observations are excessively assimilated.

The relationship between the BEC, observation error covariance, and the rescaling process is shown in the following formula (Bouttier and Courtier, 2002):

$$(d-\mathrm{H}\varphi^{f})^{2} \approx E((\varphi^{o}-\varphi^{t}-\varphi^{f}+\varphi^{t})^{2}) \quad (7)$$

$$E((\varepsilon^{o}-\varepsilon^{f})^{2}) = \sigma_{o}^{2}+\sigma_{f}^{2} \quad (8)$$

Here, σ_{o}^{2} and σ_{f}^{2} are the observation error covariance and BEC at the observation point, respectively. From Eqs. (7) and (8), the square of the innovation is expected to be the sum of the observation error covariance and BEC. The rescaled BEC can be expressed as follows (Eq. [9]):

$$\sigma_{f\_rsl}^{2} \approx (d-\mathrm{H}\varphi^{f})^{2}-\sigma_{o}^{2} \quad (9)$$

However, increasing the BEC directly presents a challenge, as it not only changes the error variance at the inflated grid point but also impacts the error covariance at neighboring grid points (Miyoshi, 2011). Hence, we employed a strategy that involved indirectly adjusting the BEC, by modifying the observation error covariance instead, following Miyoshi (2011). The rescaled observation error covariance is expressed as follows (Eq. [10]):

$$\sigma_{o\_rsl}^{2} = \sigma_{o}^{2} \times \frac{\sigma_{f}^{2}}{\sigma_{o\_rsl}^{2}} \quad (10)$$

Theoretically, the background error covariance and the observation error covariance are independent. However, when estimating these covariances, the background error covariance is rescaled based on the fact that the innovation term at observation points is given by the sum of the background error covariance and the observation error covariance (Bouttier and Courtier, 2002).

This study conducted an experiment to optimize the performance of wave data assimilation by applying the rescaling of BEC to the data assimilation process.

2.4. Experiment Design

Table 1 lists the configurations of sensitivity experiments conducted on the EnOI data assimilation method. These experiments focused on four key parameters: number of ensembles, decorrelation length scale, observation error covariance, and rescaling of BEC. In Liu et al. (2023), the observation error was set to 0.1 by comparing satellite data with buoy data. Following this approach, in the Ex.En3 and Ex.En4 experiments of this study, we set the observation error to 0.1 m. For the Ex.En0, Ex.En1 and Ex.En2, to explore the sensitivity of the model to the observation error, the observation error was set to 0.5 m. Our study evaluates the performance of EnOI under various configurations (labeled Ex.En0-4) by comparing them to a control (CTL) run without data assimilation. This approach allows us to assess the impact of each parameter on the performance of the model and identify the optimal settings for wave prediction in our study area.

2.5. Validation

To compare and validate the results of the WW3 model with data assimilation applied, analyzed satellite SWH data from various sources, including CFOSAT, Cryosat2, HaiYang-2B, Jason-3, SARAL-Altika, Sentinel-3A, and Sentinel-3B, provided by the Copernicus

Table 1. List of experimental cases

Exp. Name	CTL	Ex.En0	Ex.En1	Ex.En2	Ex.En3	Ex.En4
Data Assimilation Method	-	EnOI	EnOI	EnOI	EnOI	EnOI
Number of Ensembles	-	31	61	61	61	61
Decorrelation Length (km)	-	500	500	1000	1000	1000
Observation Error	-	0.5	0.5	0.5	0.1	0.1
Rescaling BEC	-	-	-	-	-	O

EnOI, ensemble optimal interpolation; BEC, background error covariance

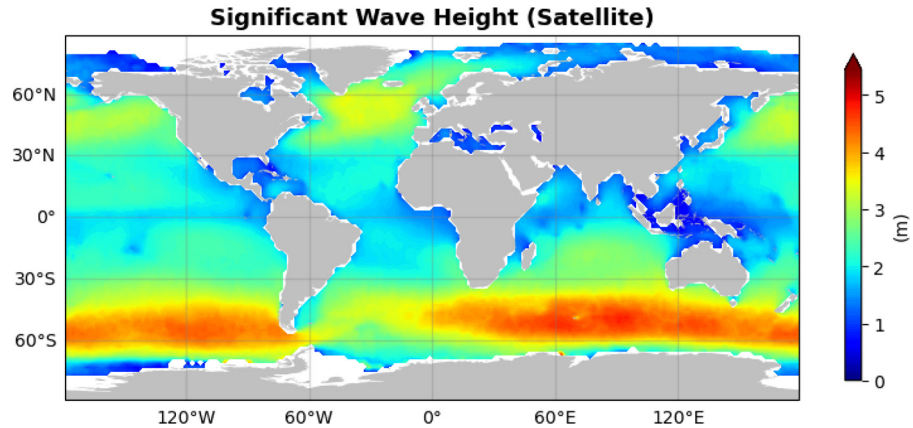

Fig. 2. Mean along-track SWH data from May 2020 to May 2021, compiled from multiple satellite missions: Jason-3, Sentinel-3A, Sentinel-3B, SARAL/AltiKa, CryoSat-2, CFOSAT, and HaiYang-2B.

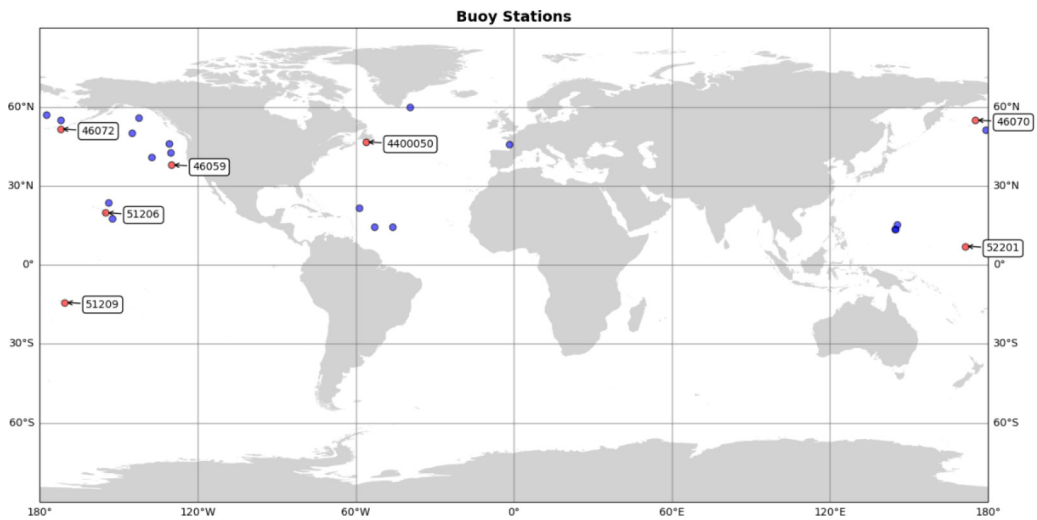

Fig. 3. Locations of the National Data Buoy Center (NDBC) buoy stations used for model comparison and validation. The buoys selected for detailed validation are marked with their identification numbers and represented by red dots.

Marine Service, along with the National Data Buoy Center (NDBC) data, were used. The analyzed SWH data were employed for validation of spatial distribution, whereas the buoy data were utilized for comparing time series at specific locations.

Fig. 2 displays the SWH data obtained from satellite observations, providing a comprehensive view of global wave height characteristics. The data highlight several key patterns. In mid-latitude regions, particularly around 30°-60° north and south, SWHs are notably higher. This is largely due to the stronger winds prevalent in these areas, which generate larger waves. Conversely, in the equatorial regions, the SWHs are considerably lower. Calmer wind conditions in these regions result in smaller waves. Additionally, coastal areas also exhibit lower SWHs, influenced by the interaction of wind patterns, ocean currents, and the presence of landmasses that disrupt wave formation. Understanding the spatial distribution of wave heights is crucial for understanding ocean dynamics, navigation, and coastal management.

The NDBC buoy data were derived from the NOAA (http://www.ndbc.noaa.gov). To compare and validate the results of the low-resolution global wave

model, buoys located in the open ocean were selected rather than those in coastal areas. The selected buoys, shown in Fig. 3, were subsampled to match the model grid.

The validation statistics utilized in this study include the root mean square error (RMSE) and relative improvement (RI), both of which are crucial for evaluating model performance. RMSE measures the average magnitude of the errors between predicted and observed values, providing a clear indication of the accuracy of the model. RI quantifies the improvement in the model compared with a baseline, reflecting the relative enhancement achieved. These metrics are represented by the following Eqs. (11) and (12):

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\Sigma_{i=1}^{N}(x_i-y_i)^2} \tag{11}$$

$$\mathrm{RI}\ (\%)\ =\ \frac{(\mathrm{RMSE}_{\mathrm{ref}}-\mathrm{RMSE}_{\mathrm{e}})}{\mathrm{RMSE}_{\mathrm{e}}}\times 100\,(\%), \tag{12}$$

RI highlights the reduction in RMSE achieved by the model relative to the baseline, thereby emphasizing the effectiveness of the model improvements.

3. Results

Fig. 4a illustrates the RMSE of the CTL run compared with observational data from satellite missions. This CTL run was conducted without implementing any data assimilation techniques, serving as a baseline for assessing the impact of data assimilation in subsequent experiments. In high-latitude areas, particularly around 60°N and 60°S, RMSE values are higher, often exceeding 0.8 m. These elevated RMSE values indicate larger discrepancies between the model outputs and observed data in these regions. This could be due to the more dynamic weather patterns and oceanic conditions at these latitudes. In contrast, the equatorial region, especially around the central Pacific and Atlantic Oceans, shows smaller RMSE values (often less than 0.4 m). This suggests that the model performs better in these regions, with smaller differences between the modeled and observed data. Coastal areas and mid-latitudes show a mixed pattern, with RMSE values generally ranging between 0.4 and 0.6 m, indicating moderate discrepancies. This spatial distribution of RMSE is critical for understanding the performance of the CTL run across different geographic locations, and can guide further model improvements and calibration efforts.

Fig. 4b illustrates the difference in RMSE between the CTL run without data assimilation and the Ex.En0 run with data assimilation. The color scale ranges from –0.30 to 0.30, where blue indicates regions where the RMSE of the CTL is lower than that of the Ex.En0, and red indicates regions where the RMSE of the CTL is higher. The figure highlights significant improvements in RMSE, particularly in the high-latitude regions around 60°N and 60°S. In these areas, the application of data assimilation (Ex.En0) substantially reduced the RMSE, demonstrating a marked enhancement

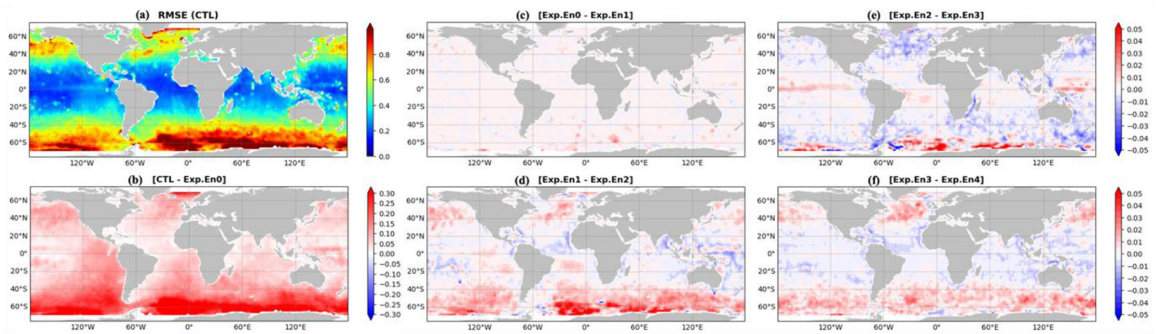

Fig. 4. Temporal mean root mean square error (RMSE) of CTL (a), and difference of temporal mean RMSE (b) between CTR and Ex.En0, (c) between Ex.En0 and Ex.En1, (d) between Ex.En1 and Ex.En2, (e) between Ex.En2 and Ex.En3, and (f) Ex.En3 and Ex.En4. The unit is m.

in model performance. The RMSE reductions were also notable in other areas where high initial errors were present, indicating that data assimilation effectively mitigated the errors in these regions. Overall, this emphasizes the beneficial impact of data assimilation on the quality of the analysis fields, particularly in regions that initially exhibited high RMSE values. The reduced RMSE in high-latitude areas and other regions underscores the importance of incorporating data assimilation techniques for improving the reliability of model outputs.

Fig. 5a shows the RI of the Ex.En0 experiment compared with the CTL run at the buoy stations. In the high-latitude regions of the North Pacific, the RI values were notably high, indicating a significant improvement in the quality of the analyzed fields owing to data assimilation. High RI values were also observed in the Atlantic Ocean, particularly in the waters off the coast of Greenland. Conversely, in the equatorial regions, RI values were generally lower, often below 10%. While there are exceptions, this trend suggests that data assimilation had a less-pronounced impact in these areas. These patterns are consistent with the results shown in Fig. 4. As shown in Fig. 4a, high RMSE values in the CTL run were prevalent in the high-latitude regions. The application of data assimilation in Ex.En0 (Fig. 4b) effectively reduced these errors, leading to the high RI values observed in Fig. 5a. Conversely, the equatorial regions, which initially had lower RMSE values in the CTL experiment, exhibited correspondingly lower RI values, as the scope for improvement was smaller. This consistency between Figs. 4 and 5a underscores the effectiveness of data assimilation in reducing model errors, particularly in regions with initially high discrepancies.

As shown in Fig. 6, a comparison of RMSE values across the seven buoy observation points revealed a significant reduction in RMSE when data assimilation techniques were applied in the Ex.En0 experiment compared with the CTL run. These observation points were widely distributed, with six located across the Pacific Ocean and one situated along the east coast of North America. The RMSE reduction in Ex.En0 was evident at all observation points, emphasizing that the application of data assimilation in Ex.En0 significantly improved the accuracy of the model compared to the CTL run.

3.1. Ensemble Size

Ex.En0 and Ex.En1 represent experiments with different ensemble sizes. The results showed that increasing the ensemble size in Ex.En1 led to slight improvements in the global domain compared to Ex.En0 (Fig. 4c).

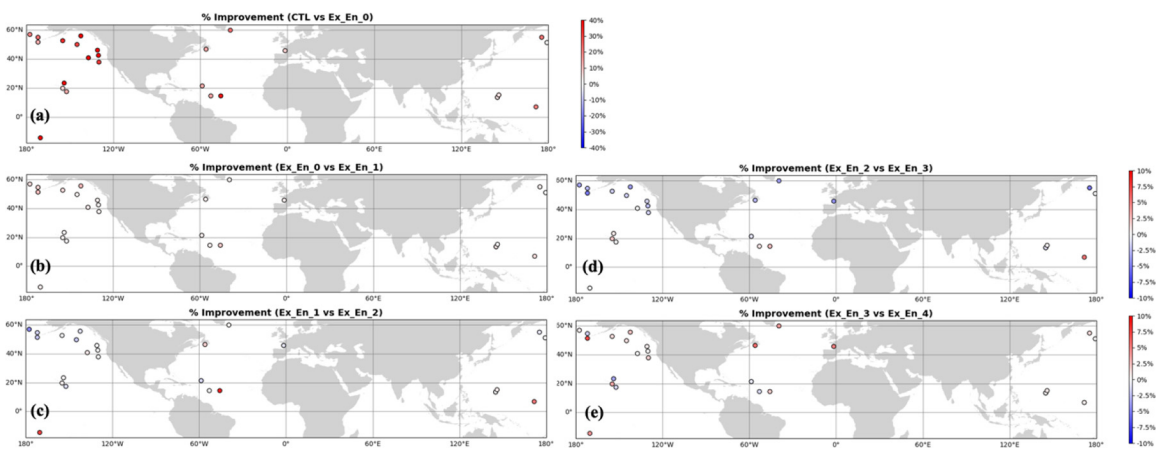

Fig. 5. Relative Improvement (RI) of (a) Ex.En0 against CTR, (b) Ex.En1 against Ex.En0, (c) Ex.En2 against Ex.En1, (d) Ex.En3 against Ex.En2, and (e) Ex.En4 against Ex.En3. Note that the color scale for (a) is -40 to 40%, while for the other figures it is -10 to 10%.

While the degree of improvement is not substantial, it aligns with the findings in the literature regarding the benefits of larger ensemble sizes.

As noted by Houtekamer and Mitchell (1998), increasing the ensemble size generally leads to more accurate representations of the BEC, which in turn can improve the quality of the analysis. This is consistent with our observations, where the difference map (RMSE(Ex.En0)-RMSE(Ex.En1)) indicates areas of improvement in various parts of the global ocean, albeit to a limited extent.

Despite modest global improvements, increasing the ensemble size has shown positive effects. The RI metric showed improvement when validated against buoy data from the western Pacific and Atlantic Oceans (Fig. 5a). While the overall degree of improvement in Ex.En1 was modest, the consistent reduction in RMSE across all seven observation points highlighted the beneficial impact of a larger ensemble size on enhancing the quality of the analysis fields.

3.2. Decorrelation Length Scale

The comparison between the two experiments with different decorrelation length scales is shown in Fig. 4c: Ex.En1 with a 500 km scale and Ex.En2 with a 1000 km scale. The analysis revealed significant improvements in the RMSE when using a larger decorrelation length scale across most global regions. Notably, the improvements were particularly pronounced in high-latitude regions, where wave energy is higher, despite a few exceptions, particularly around the equatorial regions.

The difference map (RMSE(Ex.En1)-RMSE(Ex.En2), Fig. 4d shows predominantly red areas, indicating reduced RMSE in Ex.En2. The figure suggests improvements of up to 0.04 m in RMSE in many regions. Notably, the Southern Ocean displayed the most pronounced improvement, with large dark-red areas indicating RMSE reductions of more than 0.04 m. This substantial improvement in the Southern Ocean aligns with the findings of Sakov et al. (2012), who noted that larger spatial scales can be particularly beneficial in regions with sparse observations and high variability. The Northwest Pacific also shows notable improvement, with RMSE reductions of approximately 0.02-0.03 m. This regional enhancement is particularly relevant for forecasting in the Northwest Pacific and is consistent with the work of Xie et al. (2011), who demonstrated the benefits of optimized spatial scales in regional ocean models. However, it is important to note that some areas, particularly in the equatorial Atlantic and parts of the Indian Ocean, show blue coloration, indicating a slight increase in RMSE (around 0.01-0.02 m) with the larger decorrelation length scale. This suggests that the optimal decorrelation length scale may vary by region, as discussed by Oke et al. (2007) in their study on the impact of length scales on ocean data assimilation.

When comparing the RMSE values across the seven buoy observation points in Fig. 6, it is observed that while most buoys show slight improvements in RMSE with the larger decorrelation length scale, the buoy at location 46070 exhibits a deterioration in performance. This result aligns with the findings shown in Fig. 5d, where mixed outcomes were observed depending on the region. The RI map reveals a complex pattern of performance changes when increasing the decorrelation length scale from 500 to 1000 km (Fig. 5c). The North Atlantic and Eastern Pacific show the most consistent and significant improvements, with many locations seeing 50-100% RI, while the Northwest Pacific also demonstrated notable improvements (25-75% RI). Interestingly, coastal areas, particularly along the east coast of North America, show unexpectedly high RI values (50-100%), suggesting that broader oceanic influences may be better captured with the larger scale. The Southern Ocean presents a more comprehensive view, with high variability ranging from 100% improvement to some degradation, reflecting the region's complex dynamics. The equatorial regions display mixed results, with both positive and negative RI values. This variability across different areas underscores the challenge of finding a globally optimal decorrelation length scale, and suggests that region-specific tuning might be beneficial in data assimilation systems. These findings highlight the importance of

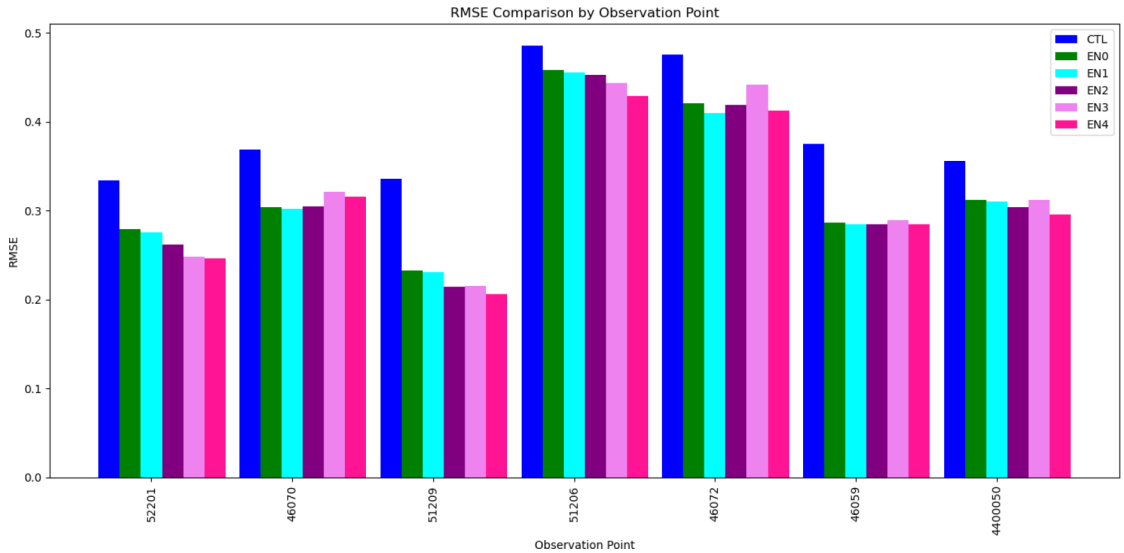

Fig. 6. Comparison of time-averaged RMSE in SWH at ocean buoys in March, 2018. The unit is m.

considering multiple metrics, such as RMSE and RI, when evaluating model performance, as they provide complementary insights into the impacts of parameter changes.

3.3. Observation Error Covariance

In the Ex.En3 experiment, the observation error was reduced from 0.5 to 0.1 m compared to Ex.En2. This change aimed to explore the impact of smaller observation error covariances on model performance. However, as illustrated in Figs. 4e, 5d, and 6, this adjustment led to a general deterioration in performance across most regions, with the exceptions of the equatorial areas and the Southern Ocean.

The RMSE difference between Ex.En2 and Ex.En3 is illustrated in Fig. 4e. The map predominantly displays blue areas, indicating that RMSE has increased in Ex.En3 compared to Ex.En2 across most global regions. This is particularly evident in the Northern Hemisphere, where the performance degradation is more pronounced. The RMSE of Ex.En3 increased by up to 0.02 m in several areas, highlighting the adverse impact of reducing the observation error covariance.

Fig. 5d further corroborates these findings with the RI map showing mixed results. While the equatorial regions and parts of the Southern Ocean showed some positive RI values, most of ocean regions displayed negative RI values, indicating a decline in model performance when the observation error covariance was reduced. Negative RI values were particularly evident in the North Pacific and Atlantic, where Ex.En3 underperformed compared to Ex.En2.

Finally, Fig. 6 provides a buoy-specific comparison of RMSE values, showing that at most buoy locations the RMSE increased in Ex.En3 relative to Ex.En2. This performance deterioration was consistent across the buoys, except in the equatorial and Southern Ocean regions, where some buoys show slight improvements. The buoy data thus reinforce the broader patterns observed in the global maps, highlighting that reducing the observation error covariance did not yield the expected improvements in most regions and instead led to performance degradation.

3.4. BEC Rescaling

The comparison between Ex.En3 without BEC rescaling and Ex.En4 with BEC rescaling applied is shown in Fig. 4f. The RMSE difference map (RMSE(Ex.En3)-RMSE(Ex.En4)) reveals a distinct pattern of improvement at higher latitudes in both the Northern and Southern Hemispheres.

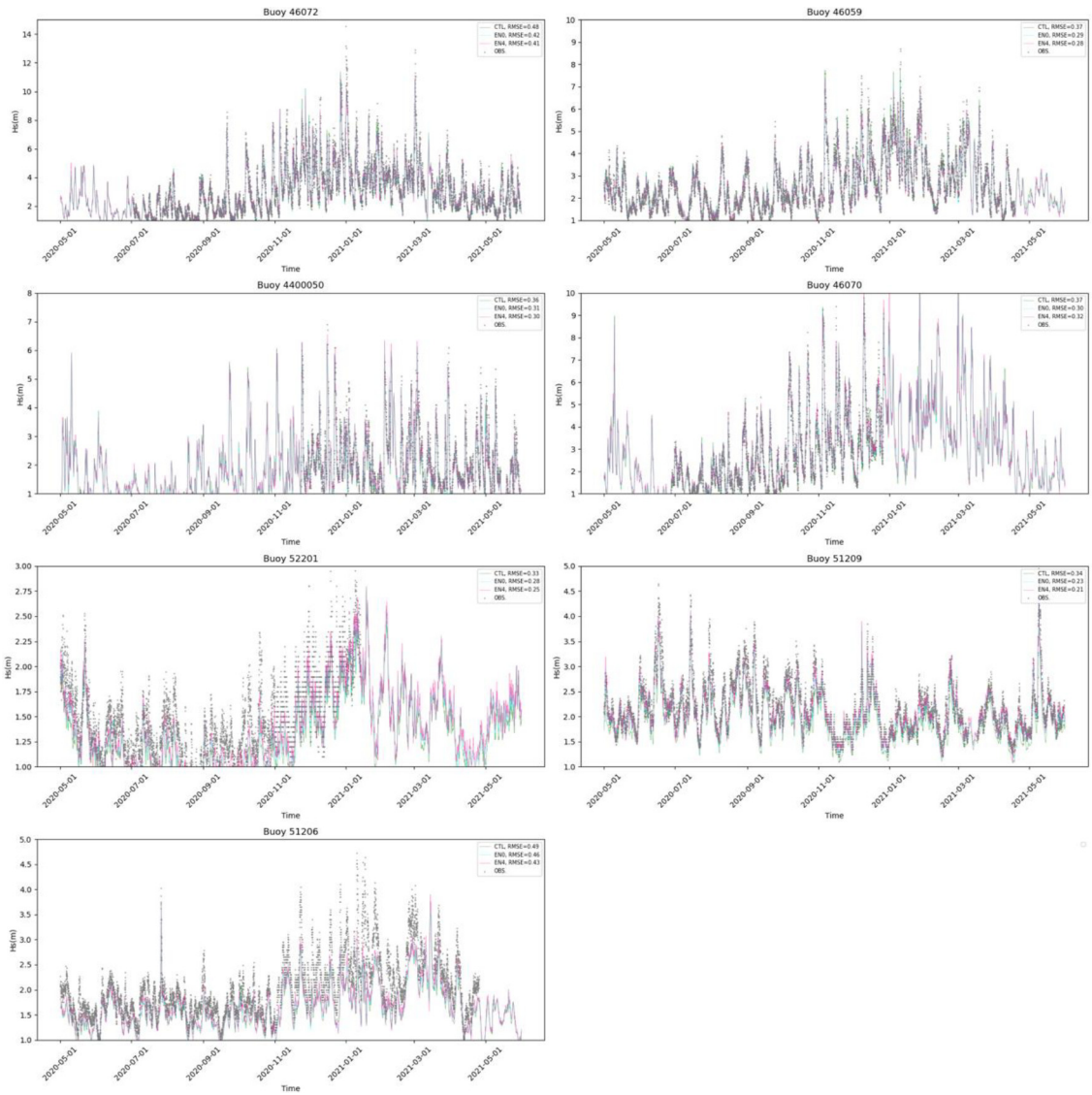

Fig. 7. Comparison of time series plots of SWH simulated from three experimental cases and observed from buoy data from May 2020 to May 2021 at the seven stations (46072, 46059, 4400050, 40670, 52201, 51209, and 51206) provided by the NDBC, which were not assimilated.

In regions poleward of 30° latitude in both hemispheres, there was a consistent reduction in the RMSE, as indicated by the predominant red coloration. This improvement was particularly noticeable in the North Atlantic, North Pacific, and the Southern Ocean, with RMSE reductions ranging from 0.02 to 0.04 m. This pattern suggests that the BEC rescaling technique is especially effective in capturing the error characteristics of these regions, which are often characterized by more dynamic and variable ocean conditions. Conversely, the equatorial and tropical regions (between 30°N and 30°S) show a mixed response, with areas of both slight improvement and degradation. The RMSE differences in these lower latitude areas are generally smaller, ranging from –0.02 to 0.02 m, indicating that the impact of BEC rescaling is less pronounced or consistent in these regions.

The RI analysis (Fig. 5e) corroborated these findings,

demonstrating consistent positive values at higher latitudes and more variable results in the tropics. This aligns with the RMSE difference map and further supports the effectiveness of BEC rescaling in improving model performance, particularly in high-latitude areas.

The effectiveness of BEC rescaling is further demonstrated by the RMSE values for Ex.En4, which show a reduction across all seven buoy observation points compared to Ex.En3 (Fig. 6). The time series of SWH for the selected buoys (Fig. 7) also shows that the Ex.En4 provides a closer analysis to the observed data compared to the Ex.En0 in the North Atlantic and North Pacific regions—specifically buoys 46070, 4400050, 46072, and 51206. Remarkably, Ex.En4 not only improved upon Ex.En3 but also achieved the lowest RMSE values among all five data assimilation experiments, underscoring the significant impact of BEC rescaling on enhancing the quality of the analysis fields.

These results are consistent with the findings of Penny et al. (2013), who noted that BEC rescaling can be particularly beneficial in regions with high variability and complex dynamics, such as those at higher latitudes. Furthermore, the variable impact in tropical regions aligns with the observations of Oke et al. (2007) who discussed the challenges of accurately representing error covariances in areas with different dynamical regimes. The distinct improvement at higher latitudes and the more nuanced impact in tropical regions underscore the importance of carefully considering regional differences when applying data assimilation techniques. As suggested by Sakov et al. (2012), adaptive approaches to BEC estimation and rescaling may be necessary to optimize performance across diverse oceanic environments.

4. Discussion and Conclusions

The results presented in this study emphasize the substantial impact of data assimilation techniques on enhancing the accuracy of SWH analysis fields, as demonstrated across various experiments. The CTL run, which did not utilize data assimilation, consistently exhibited higher RMSE values, particularly in high-latitude regions, indicating significant discrepancies between the model outputs and observed data. This underperformance is evident in Fig. 4a, where the CTL run frequently underestimated the SWH compared with the observational data, especially in dynamic regions such as those around 60°N and 60°S.

The introduction of data assimilation in Ex.En0 resulted in notable improvements, with a marked reduction in RMSE across most regions, including high-latitude areas characterized by significant variability. This enhancement was further supported by the RI metrics, which showed high values in the North Pacific and Atlantic Oceans. These findings underscore the effectiveness of data assimilation in refining model outputs and improving the accuracy of the analysis fields, as reflected in the better alignment between Ex.En0's SWH analysis and the observational data, as shown in Fig. 7.

The application of a larger ensemble size in Ex.En1 resulted in slight yet consistent improvements over Ex.En0, particularly in the western Pacific and Atlantic Oceans. Although the overall improvement was modest, the consistent reduction in RMSE across all seven buoy locations, as illustrated in Fig. 6, highlights the positive impact of increasing the larger ensemble size on the quality of the analysis fields.

Experiments focusing on adjustments to the decorrelation length scale (Ex.En2) demonstrated that increasing the scale from 500 to 1000 km led to significant RMSE reductions, especially in high-latitude regions where wave energy is more intense. However, this adjustment also produced mixed results in equatorial regions, indicating the need for region-specific tuning to optimize performance, as evidenced by both Figs. 4c and 5c. In previous studies applying EnOI, decorrelation length scales ranging from 400 km to 1600 km have been used (Liu et al., 2023; Qi and Cao, 2015). In Liu et al. (2023), it was reported that increasing the decorrelation length scale up to 1600 km resulted in performance improvements. This improvement is attributed to the fact that a larger decorrelation length scale results in stronger correlations within the same

region. However, setting the decorrelation length scale to infinity led to a deterioration in performance, indicating that using an appropriately sized decorrelation length scale can achieve satisfactory performance while avoiding excessive computational costs. In this study, we used decorrelation length scales of 500 and 1000 km, with the experiment using a 1000 km length scale showing better performance. This finding is consistent with the results of Liu et al. (2023).

The exploration of the observation error covariance in Ex.En3, where the observation error was reduced from 0.5 to 0.1 m, revealed a general deterioration in performance across most regions, with improvements observed mainly in the equatorial areas and the Southern Ocean. This outcome, illustrated in Fig. 4e, suggests that reducing the observation error covariance may not always yield better results, and that careful calibration is essential to avoid unintended degradation of the analysis fields.

Finally, the application of BEC rescaling in Ex.En4 proved to be the most effective among the experiments. The rescaling led to a significant reduction in RMSE across all seven buoy observation points, with Ex.En4 achieving the lowest RMSE values of all the data assimilation experiments. These improvements were particularly pronounced in high-latitude regions, where the dynamic and variable ocean conditions benefited the most from BEC rescaling, as shown in Figs. 4f, 5e, and 6.

In conclusion, this study demonstrates that although data assimilation techniques significantly improve the quality of analysis fields, the selection and tuning of parameters such as ensemble size, decorrelation length scale, and observation error covariance are critical for achieving optimal results. The success of BEC rescaling in Ex.En4, in particular, highlights its potential as a powerful tool for enhancing analysis fields in complex and variable oceanic environments, especially in high-latitude regions. While the study demonstrates the potential of data assimilation techniques, the need for careful calibration and adaptive parameter tuning is critical. Future research should focus on addressing these limitations by exploring more adaptive and region-specific approaches, ensuring that the benefits of data assimilation are fully realized across diverse oceanic conditions.

Acknowledgments

This research was funded by the Korea Meteorological Administration Research and Development Program “Development of Marine Meteorology Monitoring and Next-generation Ocean Forecasting System” under Grant (KMA2018-00420).

References

Booij, N., Ris, R.C., and Holthuijsen, L.H., 1999, A third-generation wave model for coastal regions: 1. Model description and validation. Journal of Geophysical Research: Oceans, 104(C4), 7649-7666.

Bouttier, F., and Courtier, P., 2002, Data assimilation concepts and methods. Meteorological Training Course Lecture Series, ECMWF, 59 p.

Burgers, G., van Leeuwen, P.J., and Evensen, G., 1998, Analysis scheme in the ensemble Kalman filter. Monthly Weather Review, 126(6), 1719-1724.

Counillon, F., and Bertino, L., 2009, Ensemble optimal interpolation: multivariate properties in the Gulf of Mexico. Tellus A: Dynamic Meteorology and Oceanography, 61(2), 296-308.

Cummings, J.A., and Smedstad, O.M., 2009, Variational data assimilation for the global ocean. Ocean Dynamics, 59(6), 991-1008.

Daniel, T., Manley, J., and Trenaman, N., 2011, The Wave Glider: Enabling a new approach to persistent ocean observation and research. Ocean Dynamics, 61(10), 1509-1520.

Evensen, G., 1994, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research: Oceans, 99(C5), 10143-10162.

Evensen, G., 2003, The ensemble Kalman filter: Theoretical formulation and practical implementation. Ocean Dynamics, 53(4), 343-367.

Gaspari, G., and Cohn, S.E., 1999, Construction of correlation functions in two and three dimensions. Quarterly Journal of the Royal Meteorological Society, 125(554), 723-757.

Gommenginger, C.P., Strokosz, M.A., Challenor, P.G., and Cotton, P.D., 2003, Measuring ocean wave period with satellite altimeters: A simple empirical model. Geophysical Research Letters, 30(22), 2150.

Hamill, T.M., Whitaker, J.S., and Snyder, C., 2001, Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. Monthly Weather Review, 129(11), 2776-2790.

Hisaki, Y., 2005, Ocean wave directional spectra estimation from an HF ocean radar with a single antenna array: Observation. Journal of Geophysical Research, 110(C11). https://doi.org/10.1029/2005JC002881

Houtekamer, P.L., and Mitchell, H.L., 1998, Data assimilation using an ensemble Kalman filter technique. Monthly Weather Review, 126(3), 796-811.

Houtekamer, P.L., and Mitchell, H.L., 2001, A sequential ensemble Kalman filter for atmospheric data assimilation. Monthly Weather Review, 129(1), 123-137.

Kim, Y.H., Hwang, C., and Choi, B.J., 2015, An assessment of ocean climate reanalysis by the data assimilation system of KIOST from 1947 to 2012. Ocean Modelling, 91, 1-22. doi: 10.1016/j.ocemod.2015.02.006

Komen, G.J., Cavaleri, L., Donelan, M., Hasselmann, S., Hasselmann, K., and Janssen, P.A.E.M., 1994, Dynamics and Modelling of Ocean Waves. Cambridge University Press, 532 p.

Li, J., and Zhang, S., 2020, Mitigation of model bias influences on wave data assimilation with multiple assimilation systems using WaveWatch III v5. 16 and SWAN v41. 20. Geoscientific Model Development, 13(3), 1035-1054.

Liu, J.C., Xu, B., and Wang, J.C., 2023, Ensemble-based assimilation of wave model predictions: contrasting the impact of assimilation in nearshore and offshore forecasting at different distances from assimilated data. Applied Ocean Research, 140, 103726.

Lzaguirre, C., Mendez, F.J., Menendez, M., and Losada, I.J., 2011, Global extreme wave height variability based on satellite data. Geophysical Research Letters, 38(10), 415-421.

Mitsuyasu, H., Tasai, F., Suhara, T., Mizuno, S., Ohkusu, M., Honda, T., and Rikiishi, K., 1980, Observation of the power spectrum of ocean waves using a cloverleaf buoy. Journal of Physical Oceanography, 10(2), 286-296.

Miyoshi, T., 2011, The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. Monthly Weather Review, 139(5), 1519-1535.

Oke, P.R., Allen, J.S., Miller, R.N., and Egbert, G.D., 2002, Assimilation of surface velocity data into a primitive equation coastal ocean model. Journal of Geophysical Research, 107(C9), 5-1-5-25.

Oke, P.R., Sakov, P., and Corney, S.P., 2007, Impacts of localisation in the EnKF and EnOI: Experiments with a small model. Ocean Dynamics, 57(1), 32-45.

Penny, S.G., Behringer, D.W., Carton, J.A., and Kalnay, E., 2013, A hybrid global ocean data assimilation system at NCEP. Monthly Weather Review, 141(8), 2740-2760.

Qi, P., and Cao, L., 2015, The assimilation of Jason-2 significant wave height data in the north Indian ocean using the ensemble optimal interpolation. IEEE Transactions on Geoscience and Remote Sensing, 54 (1), 287-297.

Queffeulou, P., 2004, Long-term validation of wave height measurements from altimeters. Marine Geodesy, 27, 495-510.

Rapizo, H., Babanin, A.V., Schulz, E., Hemer, M.A., and Durrant, T.H., 2015, Observation of wind-waves from a moored buoy in the Southern Ocean. Ocean Dynamics, 65, 1275-1288.

Sakov, P., Counillon, F., Bertino, L., Lisæter, K.A., Oke, P.R., and Korablev, A., 2012, TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic. Ocean Science, 8(4), 633-656.

Saulter, A.N., Bunney, C., King, R.R., and Waters, J., 2020, An application of NEMOVAR for regional wave model data assimilation. Frontiers in Marine Science, 7, 579834.

The WAVEWATCH III® Development Group, 2019, User manual and system documentation of WAVEWATCH III® version 6.07. NOAA/NWS/NCEP/MMAB Technical Note 333, 465 pp.

WAMDI Group, 1988, The WAM model - A third generation ocean wave prediction model. Journal of Physical Oceanography, 18(12), 1775-1810.

Walsh, E.J., Hancock, D.W., and Hines, D.E., 1989, An observation of the Directional Wave Spectrum Evolution from shoreline to fully developed. Journal of Physical Oceanography, 19(5), 670-690.

Waters, J., Wyatt, L.R., Wolf, J., and Hines, A., 2013, Data assimilation of partitioned HF radar wave data into Wavewatch III. Ocean Modelling, 72, 17-31.

Wittmann, P., and Cummings, J., 2005, Assimilation of Altimeter Wave Measurements into Wavewatch III. 12 p.

Xie, J., Counillon, F., Zhu, J., and Bertino, L., 2011, An eddy resolving tidal-driven model of the South China Sea assimilating along-track SLA data using the EnOI. Ocean Science, 7(5), 609-627.

Xie, J., and Zhu, J., 2010, Ensemble optimal interpolation schemes for assimilating Argo profiles into a hybrid coordinate ocean model. Ocean Modelling, 33(3), 283-298.

Yu, H., Li, J., Wu, K., Wang, Z., Yu, H., Zhang, S., Hou, Y., and Kelly R.M., 2018, A global high-resolution ocean wave model improved by assimilating the satellite altimeter significant wave height. International Journal of Applied Earth Observation and Geoinformation, 70, 43-50.

Manuscript received: August 14, 2024
Revised manuscript received: August 26, 2024
Manuscript accepted: August 29, 2024