The assessment of ERA-interim wave data in the China Sea

Hongyuan Shi a,b,*, Jiacheng Sun c, Zaijin You a,b, Qingjie Li c, Delei Li d, Xuefeng Cao e

aSchool of Civil Engineering, Ludong University, Yantai 264025, China, email: shihongyuan1234@163.com (H. Shi)
bInstitute of Ports and Coastal Disaster Mitigation, Ludong University, Yantai 264025, China
cYantai Marine Environmental Monitoring Central Station, State Oceanic Administration, Yantai 264025, China
dChinese Academy of Sciences, Qingdao 266100, China
eNational Marine Environment Monitoring Center, Dalian 116023, China

Received 8 July 2019; Accepted 21 December 2019

ABSTRACT

As we know, wave significant height ($H_s$) and wave period, are critical for maritime commerce, infrastructure design, and hazard mitigation, etc. In this paper, we adopt wave data from eight stations in the data-sparse China Sea to assess the ERA-Interim (ERA-I). We have described the comparison with the short-scale and annual features between the datasets. The annual difference of $H_s$ varies from $-0.43$ to $0.1$ m with a mean value of $-0.24$ m for eight stations. In most of the stations, the ERA-I data of the locations except for the typhoon period has an overall overestimation, and the biases is positive. During the typhoon, due to the underestimation of the typhoon wind field, $H_s$ in the ERA-I data is underestimated, which can be up to $-50\%$. The physics research of typhoon is still poorly at present, causing the discrepancies between the datasets. Hence, we can't adopt the ERA-I $H_s$ for design applications unless the validation has been tested by specific sites.

Keywords: ERA-interim; Reanalysis inter-comparison; Wave height; China Sea

1. Introduction

Wave significant height ($H_s$) and period ($T$) are important for human activities such as maritime commerce, the design of oceanographic engineering, ship design, hazard mitigation, and other things. Until now, there are many ways to obtain these data, such as voluntary observing ships (VOs), synthetic aperture radar (SAR), satellite altimetry, and buoys. As for these, the ship data has the longest history (since the mid-nineteenth century) but the quality can't be guaranteed [1,2]. Due to the shipping routes, it only covers a certain area, and if happening extreme conditions, it may cause data loss [3]. Light vessels and buoys are adopted to observed and measured since the 1950s and 1970s, respectively [4], and they provide comprehensive and diverse measurements and present people an important information source [5–7]. However, they are restricted to shipping routes or coastal sites and they are often not long enough or sparse coverage in the time-series that make research in large area coverage and long-time series impracticable. Satellite altimetry can cover a large field of ocean and the precision is high, so in the field of climate studies, it has adopted as a worthy resource [8–11], but its orbit is periodic for the fixed field and it varies from 10 to 35 d [12]. Hence, the temporal resolution of satellite altimetry is poor and it may lose the extreme events in a high probability [13]. SARs can obtain the frequency-direction spectrum of the sea state, and then get the wave process by the relation between wave elements and wave spectrum [14]. But it has the same shortcoming as satellite altimetry.

Numerical models have been carried out to predict and hindcast the wave conditions since World War II. The physical mechanism and numerical calculation methods in the numerical model are improving all the time. In recent
years, the development of numerical simulation technology has greatly supplemented the measurement of various wave sources, which has the advantages of high spatial and temporal resolution, large coverage area, and long duration [15–17]. They are particularly important for calculating the extreme load design values [18,19] and fatigue loads [20]. What's more important, it can provide useful information for navigation [21].

Generally, significant wave height (SWH) \(H\) and the wave period \(T\) are generally adopted to describe a wave. It has become a common tool to obtain the wave height and wave period in the region without observation by numerical models [13]. Many wave reanalysis datasets are available, and different wave models [such as Wave Watch III, wave model (WAM), and simulating waves nearshore] and different reanalysis wind fields (National Centers for Environmental Prediction, cross calibrated multi-platform, etc.) are used for them. So for the specific purpose, we should know which dataset is more adequate. Many researchers have done this work, for example, the quality of the numerical model of MIKE 21 have been assessed with the measured buoy data along the north Indian Ocean [22]. The result showed that the variation between the MIKE21 and measured \(H\) ranges from 0.68 to 0.91 m and the bias can be up to 0.6 m. And at some sites, the variation can be up to 2 m during the southwest monsoon period. A study assessed the wind speed and \(H\) data of several reanalysis datasets by using altimeter and buoy data. They describe that although the quality of the datasets differs in comparison with the observed data, most of the long-scale features are mainly equal across all the data sets [23].

The ERA-Interim (ERA-I) dataset is a more recent ERA available for all locations of the globe from 1979 to the present. It is widely used at (i) the wave design, (ii) the study on the change of wave characteristics over a long period of time, and (iii) the assessment of wave energy resource [8,24–28]. Thus, it is crucial to know the quality of this data set by compared with observed data from diverse time and diverse sites. Recent study assessed the ERA-I wave data in the shallow water around data-sparse India Ocean by adopting the measured data from six locations [13]. The result shows that ERA-I data overestimates wave height for all regions except the northern part of India's west coast, where the data underestimate. A study made the comparisons between Hindcast of dynamics process of the ocean and coastal areas of Europe (HIPOCAS) and ERA wind and wave reanalysis twice, one is the comparisons between HIPOCAS and ERA-40, another is the comparisons between HIPOCAS and ERA-I [29]. Under calm and moderate weather conditions, the results relatively agree well. The difference in extreme conditions increased greatly, mainly occurring in winter at middle and high latitudes. Around the globe, the correction of reanalysis products has been conducted and several studies have assessed the quality of ERA-I wave datasets in different countries [30]. But the validity of wave height and period in the ERA-I dataset is not well-conducted in the China Ocean due to the data-sparse caused by the confidentiality of observation data and buoy data.

The assessment of datasets will enhance the confidence of datasets application in wave research and marine business. This study assesses the ERA-I datasets by utilizing eight buoy observations, including the performance of characterization, the effectiveness, and consistency. Section 2 describes the details about the reanalysis datasets and the buoys, and the methodology and error metrics for comparison are shown in section 3. Section 4 reports the results, the last section presents the summary and conclusions.

2. Data

2.1. ERA-Interim data

Because reanalysis models can numerically simulate the global ocean conditions, so we can adopt reanalysis models to estimate wave at any nearshore or offshore China locations. These simulations can be carried out over a large space without the disadvantages of irregular and discontinuous data encountered when using synoptic hydrologic observatory data.

ERA-I is a global reanalysis dataset, and it dates back to 1979 that covers the data-rich period. Originally, ERA-I started from 1989, but in the year of 2011, ERA-I extended the data from 1979 to 1988, and the dataset continues updating in real-time. As ERA-I moves on in time, the archive will be updated monthly. The ERA-I project was launched in 2006 for the bridge between the ERA-40 (1957–2002) and ERA5 that is the next generation. The project mainly aimed to improve the quality of ERA-40, for example, the stratospheric circulation and the management of deviations and variations in the observing system. Compared to ERA-40, ERA-I used more observations and adopted the newest assimilation techniques (4D-Var), moreover, it improved the resolution, and adopted better physics and mechanism in the models [24]. The third-generation spectral model (WAM) is utilized for ERA-I to simulate wave. Previous research has updated the source terms of WAM and also have done this work. Now, it improves the ability to simulate the process of wave growth and dissipation [31–33]. What's more, ERA-I can assimilate the measured wave significant height data obtained by polar-orbiting satellites to restrict the predicted spectrum [34]. They utilized the data from satellites to do wave assimilation work since 1991. The resolution of ERA-I wave is approximately \(0.64 \times 0.64^\circ\) and 6 h.

2.2. Buoy data

China has a long coastline of 32,000 km. Most of Chinese largest cities locate along the coastline (almost 70%) and more than 50% of the population woke and live here. Now, more and more people are gathering toward the coast area. Thus, it is of great value to understand the wave characteristics of offshore China.

The buoy data which was measured by the Directional Waverider MKIII and acoustic wave and current are adopted to compare with the ERA-I datasets. The wave measurement was carried out by the State Oceanic Administration (SOA) of China and the Institute of Oceanology, Chinese Academy of Sciences. The instrument of directional wave rider buoy utilizes the heaves to measure the free-surface gravity waves, and the range of application is from \(-20\) to \(+20\) m. The resolution can be up to 1 cm in fluctuation and the measuring range is from 1.6 to 30 s as for wave periods. The reliability of the instrument is high and the cross-sensitivity of sea undulation is less than 3%. The data were recorded at the frequency of 1.28 Hz and the buoys took half an hour...
procession as one record. The default wave characteristics such as SWH are recorded in the file with the format of sdt. In this paper, the records at 6 h intervals from the measured data were used to compare with the ERA-I data.

The locations of buoys are presented in Fig. 1. The locations’ distance from the coast and water depth were taken into account. We have collected eight buoys for the comparison: four of the Yellow Sea (YES), named $Y_1$–$Y_4$, two of the East China Sea (ECS), named $E_1$ and $E_2$, and two of the South China Sea (SCS), named $S_1$ and $S_2$.

The buoy $H_s$ are available hourly from 10 to 20 min-long records. Before using the data for comparison, the data have gone through some quality control. The reanalysis data and measured data are obtained in different ways (the reanalysis data was at the synoptic time, but the buoy measurements were averaged). Thus, the linear interpolation for the reanalysis data is carried out to the buoy location. The measurement data were recorded at Beijing time, which is earlier 8 h than universal time coordinated (UTC). In order to keep time consistency between the reanalysis data and the observation data, all the time of reanalysis data plus 8 h. Table 1 lists the specific information of the eight stations.

3. Methodology used for data comparison

The buoy measurements represent a large spatially disaggregated data and we should adopt a systematic approach for the comparison with the ERA-I. In the background of economic viability and survivability of marine transport and offshore and coastal structures, it is essential to accurately estimate the extreme wave conditions. Underestimating these extremes can have a negative impact on the structure, leading to damage, while overestimating will ensure absolute safety, but it will directly lead to overdesign, causing unnecessarily high capital costs, and financially unattractive for investment [29]. Admittedly, different probabilistic models used can get different assessed extreme values, but the data adopted as the basic source is also very vital [19]. Therefore, this paper analyzed the extreme values at each buoy after the overall approach, finding out typhoon happening around the buoys during the working time.

Thus, sections contain two parts: “the holistic conditions” and “the extreme conditions.” For the whole data, the results are also classified as “all-year comparisons,” and “extreme events comparison” regarding the typhoon wave data. The difference between the two data can be represented by a number of error metrics adopted in this paper. We define $X$ and $Y$ to represent the buoy and ERA-I of $H_s$ over time, respectively. The statistical parameters used for comparison between them were as following:

\[
\text{Bias} = \frac{\sum_{i=1}^{n}(X_i - Y_i)}{n}
\]  
\[
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - Y_i)^2}{n}}
\]  
\[
\text{SI} = \frac{\text{RMSE}}{S}
\]
where the over bar indicates the mean values in time and \( n \) is the number of data.

4. Results and discussions

The comparison result between buoy and ERA-I of eight stations is shown in Fig. 2. The annual mean \( H_s \) in YES is 0.51 m (\( Y_1 \)) and 0.39–0.56 m (\( Y_2 \)), respectively, because the measurement time only lasted 6 months (from January to June) in the site of \( Y_3 \) and \( Y_4 \), so we don’t discuss the annual mean \( H_s \) in these sites. The annual mean \( H_s \) in ECS are 0.23–0.32 m (\( E_1 \)) and 0.76–1.25 m (\( E_2 \)) (Table 2) whereas the mean \( H_s \) in the SCS are 1.2 m and 1.74 m (\( S_2 \)).

In the YES, the variation of \( H_s \) in \( Y_1 \) and \( Y_2 \) are 0.1–2.2 m and 0.1–3.5 m, whereas the variation of \( H_s \) in the ESC is 0.1–1.8 m (\( E_1 \)) and 0.1–6.8 m (\( E_2 \)). In the SCS, \( H_s \) in \( S_1 \) and \( S_2 \) varies from 0.2 to 4.81 m and 0.2 to 9.9 m, respectively. Due to the deeper water of \( Y_3 \) and \( Y_4 \) (see Table 2, the maximum \( H_s \) are greater (4.0 and 4.2 m, respectively) than \( Y_1 \) and \( Y_2 \), even there only 6 month (from January to June) data recorded. The annual bias of wave height in \( Y_1 \) is 0.10 m, and the annual bias of wave height in \( Y_2 \) varies from –0.54 to –0.36 m. The annual bias of wave height in \( Y_3 \) and \( Y_4 \) are –0.08 and –0.097 m, it indicates that the ERA-I \( H_s \) data agrees well with the measured \( H_s \) data in \( Y_3 \) and \( Y_4 \). The data of ERA-I is only slightly bigger than buoy data. The annual bias of wave
Table 2
Comparison result between ERA-I and buoys

<table>
<thead>
<tr>
<th>Station</th>
<th>Annual SWH (m)</th>
<th>Variation of SWH</th>
<th>Bias (m)</th>
<th>RMSE (m)</th>
<th>SI</th>
<th>COR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y_1</td>
<td>0.51</td>
<td>0.1–2.2</td>
<td>0.1</td>
<td>0.24</td>
<td>0.15</td>
<td>0.72</td>
</tr>
<tr>
<td>Y_2</td>
<td>0.39–0.56</td>
<td>0.1–3.5</td>
<td>-0.42 to -0.34</td>
<td>0.57–0.74</td>
<td>1.07</td>
<td>0.59–0.71</td>
</tr>
<tr>
<td>Y_3</td>
<td>/</td>
<td>0.1–4.0</td>
<td>-0.08</td>
<td>0.26</td>
<td>0.25</td>
<td>0.95</td>
</tr>
<tr>
<td>Y_4</td>
<td>/</td>
<td>0.1–4.2</td>
<td>-0.097</td>
<td>0.25</td>
<td>0.23</td>
<td>0.95</td>
</tr>
<tr>
<td>E_1</td>
<td>0.23–0.32</td>
<td>0.1–1.8</td>
<td>-0.42 to -0.35</td>
<td>0.50</td>
<td>0.83</td>
<td>0.65–0.69</td>
</tr>
<tr>
<td>E_2</td>
<td>0.76–1.25</td>
<td>0.1–6.8</td>
<td>-0.44 to -0.11</td>
<td>0.2</td>
<td>0.43–1.69</td>
<td>0.70–0.88</td>
</tr>
<tr>
<td>S_1</td>
<td>1.2</td>
<td>0.2–4.81</td>
<td>0.15</td>
<td>0.89</td>
<td>0.58</td>
<td>0.51</td>
</tr>
<tr>
<td>S_2</td>
<td>1.74</td>
<td>0.2–9.9</td>
<td>0.03</td>
<td>0.38</td>
<td>0.21</td>
<td>0.94</td>
</tr>
</tbody>
</table>

height in E_2 are -0.35 m (2010a), -0.39 m (2011a), -0.40 m (2012a), -0.42 m (2013), -0.34 m (2014), and -0.35 m (2015a), so the data of ERA-I H_s is bigger than the buoy data, and the annual bias of wave height in E_2 are -0.44 m (2010a) and 0.11 m (2011a). As for SCS, the annual bias of wave height in S_2 are 0.70 and 0.03 m, respectively. This indicates that the ERA-I H_s data agree better with the measured H_s in S_2 than in S_1 that maybe caused by the water depth.

Table 2 shows the comparison results, we can see that the comparison statistics also illustrate that the ERA-I H_s data agree well with the buoy-measured H_s data in the deeper water. The values of r for the SWH for Y_1 is 0.72 (2011), and the values of r for the H_s for Y_1 are 0.71 (2014), 0.59 (2015) and 0.65 (2016). There are only half a year data recorded in Y_1 and Y_2, and the correlation coefficient (r) values are 0.95 and 0.95, respectively. In the ECS, the annual correlation coefficient (r) values from 2010 to 2015 for the H_s for E_1 are 0.67, 0.65, 0.67, 0.69, 0.69, and 0.68, respectively. As for E_2, the annual correlation coefficient (r) value is 0.70 (2012) and 0.88 (2011). In the SCS, the correlation coefficient (r) in S_1 and S_2 are 0.51 and 0.94.

The RMSE in Y_1 of 2011a is 0.24 m, and they are 0.57, 0.74 and 0.74 m in Y_1 from the 2014a to 2016a. As for Y_2 and Y_3, the RMSE in the Y_2 are 0.26 and 0.25 m, respectively. In the ECS, the RMSE in the E_1 from 2010 to 2015 are 0.43, 0.78, 0.47, 0.51, 0.41, and 0.42 m, and the RMSE in the E_2 are 0.66 and 0.24 m. In the SCS, the annual RMSE in the S_1 and S_2 are 0.89 and 0.38 m. That also indicates that the data consistency is closely related to water depth: where water depth is large, data consistency agree well.

The SI can show the dispersion of the data, the mean SI from Y_1 to Y_3 in YCS are 0.15, 1.07, 0.25, and 0.23. The SI in E_1 from 2010 to 2015 are 0.81, 0.76, 0.75, 0.83, 0.76, and 0.98. The annual SI in E_2 are 1.69 (2010) and 0.43 (2011), so the data agrees well in 2011 than in 2010 in the station of E_2. The mean SI in S_1 and S_2 are 0.58 and 0.21. The coast of China has been often attacked by tropical cyclones (TCs). There is no other country in the world, where TCs strike with higher frequency and intensity than on the coast of China [35,36]. There are about 10 landfall TCs in a typical year on the coast of China. Recent research has shown that TCs striking East and Southeast Asia have strengthened by 12%–15%, and the percentage of storms that are classified as 4–5 category to the total has doubled or even tripled [37]. Typhoons are likely to occur all year round, mainly in summer and autumn, accounting for more than 80% of the total. In China, typhoons mainly occur in the ECS and the SCS. So the maximum SWH mainly occur in summer or autumn in ESC and SCS. We also discuss the quality of typhoon wave in ERA-I, the result show that the wave height is smaller in ERA-I than in the measured buoys in most stations and they are typically underestimated by 20%–50%, but in the station of E_2, the data of ERA-I is overestimated. We looked up the topographic map and found that E_2 station was surrounded by a small alluvial island. It was located on the west side of the island, and the island played a role of wave protection to the observation point. In the numerical simulation, because the accuracy of the numerical model was rough, the island was treated as water, so the actual measured data was smaller than the numerical simulation data [38–43].

5. Summary and conclusions

In this paper, the performance of ECMWF interim reanalysis H_s data was evaluated by comparing with the data measured by buoys from eight sites along China Sea, four locations off the YES, two locations off East China, and two locations in the SCS were selected. Most of the measured data containing four seasons were adopted for the comparison.

The results showed that ERA-I overestimates the H_s in stations of Y_1, Y_2, Y_3, E_1, and E_2. But underestimate the H_s in the stations of Y_1 and S_1. The difference of H_s is up to 125% for shallow-water locations such as Y_1, Y_2, and E_2. Although the annual mean of bias indicates that ERA-I overestimates H_s under lower wave height (such as typhoon wave), the ERA-I underestimates H_s largely. Due to the geographical location of China, the area of the ECS and SCS are frequently attacked by TCs, and ERA-I underestimates the H_s at these sites distributing there during a typhoon by -30%. Thus, the design applications can't adopt ERA-I to calculate unless the proper validation has been done.

In a slightly more complex basin (such as the Yangtze Estuary), the assessment of the root which causes ERA-I biases and errors with the measured data is difficult by using low-resolution global models. Alluvial island effects and bathymetry may play an important role in this environment. With this in mind, we can expect to find minimal credibility in the ERA-I data, as the partial obstruction of wave energy in the east and the possibility of land-sea interactions play a role.
Acknowledgment

The project was financially supported by National Key R&D Program of China (Grant No.: 2018YFB1501901), the Natural Science Foundation of China (No. 51909114) the Open Research Fund of the Key Laboratory of Ocean Circulation and Waves, Chinese Academy of Sciences (Grant No.: KLOCW1901), the Open Research Fund of State Key Laboratory of Tropical Oceanography, South China Sea Institute of Oceanology, Chinese Academy of Sciences (Project No.: LTO1905), and the Major Research Grant (No.: U1806227) from the Natural Science Foundation of China (NSFC) and the Provincial Natural Science Foundation of Shandong.

References


