

Article

# Intercomparison of Assimilated Coastal Wave Data in the Northwestern Pacific Area

Yukiharu Hisaki 

Department of Physics and Earth Sciences, University of the Ryukyus, Senbaru 1, Nishihara-cho, Okinawa 903-0213, Japan; hisaki@sci.u-ryukyu.ac.jp; Tel.: +81-98-895-8515

Received: 14 July 2020; Accepted: 29 July 2020; Published: 1 August 2020



**Abstract:** The assimilated coastal wave data are useful for wave climate study, coastal engineering, and design for marine disaster protection. However, the assimilated coastal wave data are few. Here, wave analysis data produced by the JMA (Japan Meteorological Agency) and ERA5 wave data were compared with GPS (Global Positioning System) buoy-measured wave data. In addition, the accuracy of ERA5 wave data for various conditions was investigated. The accuracy of JMA analysis wave height was better than that of ERA5 wave height. The ERA5 wave height was underestimated as the wave height increased. The accuracy of the ERA5 wave height was significantly different in fetch-unlimited and fetch-limited conditions. The difference of the skill metrics between fetch-unlimited and fetch-limited conditions was due to the overestimation of the fetch in the ERA5 grid. This result also applied to the wave period.

**Keywords:** wave height; ERA5; coastal areas; Northwestern Pacific; wave period; assimilation

## 1. Introduction

The study of wave climate is critical to the assessment of the utilization of all types of marine vehicles, ships, drilling, and other civilian uses. The wave data on the grid with high spatial resolution are useful for the study of the wave climate. One of the types of wave data with high spatial resolution is the hindcast wave data, which are the output of the wave model forced by the analyzed wind. There are databases of wave hindcast in various areas. Reguero et al. [1] introduced a wave hindcast dataset with a resolution of  $0.5^\circ$  covering the world coastline. Shi et al. [2] established a wave hindcast database with a resolution of 1 km along the Chinese coast. Groll and Weisse [3] produced a wave hindcast dataset in the North Sea covering the period 1949–2014. Haakenstad et al. [4] presented wave hindcasts for the Norwegian Sea, the North Sea, and the Barents Sea.

There are studies of wave climate based on hindcast in the Northwestern Pacific area. Shimura and Mori [5] hindcasted and analyzed the wave climate around Japan. Taniguchi [6] verified the wave hindcast in the Japan Sea and predicted the future wave climate of the Japan Sea. Hu et al. [7] simulated wave parameters associated with typhoons in the Northwestern Pacific and investigated the wave climate by typhoons. Hindcast wave data were used to study the wave climate of various regions, including the Indian Ocean [8], North Sea [9], Hawaiian coast [10], Brazilian coast [11], Swedish coast [12], Black Sea [13,14], and Persian Gulf [15].

Assimilated wave data are more useful for studying the wave climate, because the observed wave data are incorporated into the assimilated wave data. Typical assimilated wave data are ERA5 (Fifth generation of European Centre for Medium Range Weather Forecasting atmospheric reanalyses of the global climate), which was updated from ERA-Interim (European Centre for Medium-Range Weather Forecasts reanalysis data interim version) data. The ERA-Interim and ERA5 reanalysis data are global atmospheric and oceanic reanalysis data produced by the European Centre for Medium Range Weather Forecasting (ECMWF). The ERA5 reanalysis data have been improved compared with

the ERA-Interim data, especially in spatial and temporal resolution [16]. It was shown that the ERA5 wave height was more accurate than the ERA-Interim wave height by comparing with the independent buoy data [17]. However, the spatial resolution of ERA5 wave data is  $0.36^\circ$ , which is not enough for coastal wave data.

The Japan Meteorological Agency (JMA) produces wave analysis data by data assimilation in the marginal seas or coastal seas in the Northwestern Pacific area [18]. The hindcast wave spectra are modified using optimal interpolation with observations of significant wave heights from the radar altimeters of satellites, buoys, coastal wave recorders, and ships. The JMA wave analysis data are used as the initial values of the wave forecast of the JMA. The short-term wave forecast is improved by using assimilated data [18]. The spatial resolution is  $0.05^\circ$ , which is sufficient for coastal wave data in many cases.

Wave analysis data such as JMA and ERA5 data are expected to be used for wave climate research. These data will also provide the basis of the guidelines for designing port structures such as breakwaters, coastal structures such as seawalls, and marine structures such as platforms for offshore oil drilling. In addition, wave analysis data will also be used to select suitable locations for wave power generation [19–21]. Therefore, it is necessary to evaluate the accuracy of JMA and ERA5 wave data.

There were many studies for validating ERA-Interim wave data by comparing measured wave data by buoys in the global ocean [22], Atlantic Ocean [23], Indian Ocean [24], and Arabian Sea [25], which showed good agreement. There were also some studies for validating ERA5 wave data by comparing measured wave data by buoys [26,27] or remotely sensed wave data by altimetry [28].

For example, Bruno et al. [26] compared buoy-measured wave parameters with ERA5 wave parameters. However, the buoy was close to the coast in the shallow area, and it was impossible to compare buoy-measured wave parameters with ERA5 wave parameters directly. Sifnioti et al. [27] also compared buoy-measured wave parameters in the shallow area near the coast with ERA5 wave parameters and hindcast wave data. However, the positions of the ERA5 wave data were different from those of the buoys. No studies have been conducted to intercompare ERA5 wave data, other wave assimilation data, and observed data in the coastal area. The objective of this paper is to evaluate the assimilated wave data near the coast by comparing ERA5 wave data and JMA wave analysis data with buoy-measured wave data. The accuracy of ERA5 wave data for various conditions is also investigated.

Section 2 describes the JMA wave data, ERA5 wave data, and GPS buoy data. The method of analysis is also described in Section 2. Section 3 presents the result of the comparison of wave heights. The comparison of wave periods is also presented in Section 3. Section 4 discusses the interpretation of the results. The conclusions are summarized in Section 5.

## 2. Data and Methods

The JMA wave analysis data are produced from hindcast wave data by the Coastal Wave Model (CWM) and the assimilation. The CWM is based on the MRI (Meteorological Research Institute)-III, which was the third-generation wave model developed by the MRI of JMA [29]. The wave spectra are predicted from the energy balance equation. The parameterization of the source function is described in [30]. The Global Wave Model (GWM) is also operated by the JMA, and the boundary conditions of the CWM are given by the output of the GWM. The wave spectra are corrected based on the significant wave height by using the optimal interpolation (OI) with observations from altimeters, buoys, coastal wave recorders, and ships [30]. The area of the JMA wave analysis data is from  $20^\circ$  N to  $50^\circ$  N and from  $120^\circ$  E to  $150^\circ$  E. The JMA wave analysis data can be downloaded from <http://database.rish.kyoto-u.ac.jp/arch/jmadata/>. The time interval of the JMA wave analysis data is 6 h. The JMA significant wave heights ( $H_j$ ), peak wave periods ( $T_{pj}$ ), peak wave directions, and surface winds  $\mathbf{U}_j = (u_j, v_j)$  are archived in the datasets.

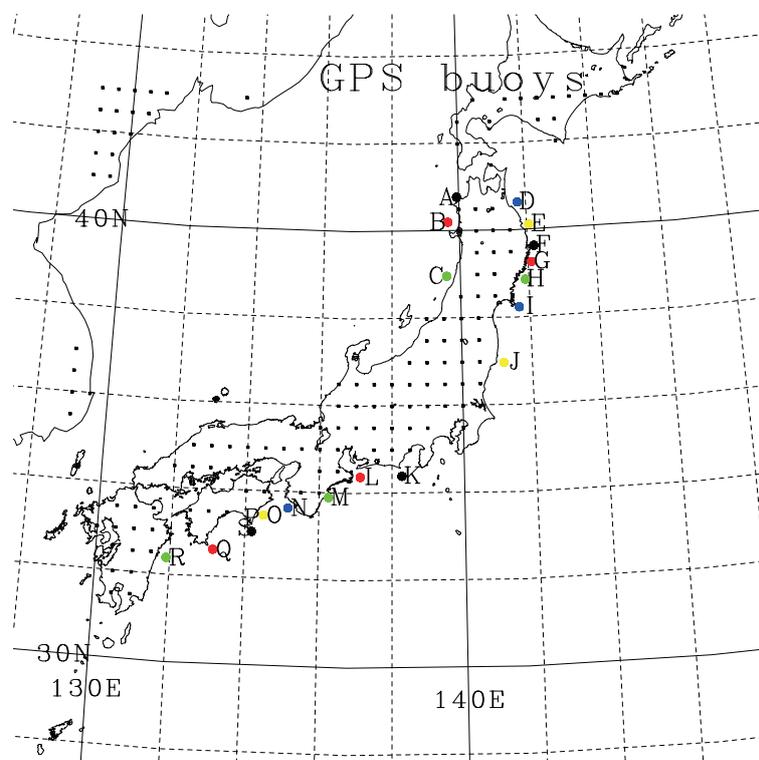
The ERA5 significant wave heights ( $H_e$ ), peak wave periods ( $T_{pe}$ ), and mean wave periods ( $T_{me}$ ) were also used for comparison, where the mean wave period is defined as  $T_{m-1} \equiv M_{-1}M_0^{-1}$ ,

$$M_n = \int_0^\infty f^n F(f) df, \tag{1}$$

where  $f$  is the wave frequency, and  $F(f)$  is the wave frequency spectrum. The native spatial resolution of ERA5 wave data is  $0.36^\circ$ , and the Climate Data Store data (<https://cds.climate.copernicus.eu/>), where it is possible to download ERA5 data, were converted to  $0.5^\circ$  [17]. The ERA5 winds  $U_e$  were also used for the analysis.

The wave parameters observed by the GPS buoys were used for validation. The significant wave heights of the both JMA wave analysis data and ERA5 wave data were estimated from the wave spectrum as  $H_s \equiv 4M_0^{\frac{1}{2}}$  (Equation (1)). On the other hand, the moored GPS-estimated wave height ( $H_g$ ) and period ( $T_g$ ) were estimated by the zero-up-crossing method ( $H_g = H_{\frac{1}{3}}, T_g = T_{\frac{1}{3}}$ ) from 1024 surface elevations [31], where  $H_{\frac{1}{3}}$  and  $T_{\frac{1}{3}}$  are significant wave height and period estimated by the zero-up-crossing method, respectively. The value of  $H_s/H_{\frac{1}{3}}$  ranges from 1.01 to 1.07 [32]. The wave height ( $H_g$ ) and period ( $T_g$ ) were estimated at 20 min intervals. The GPS wave data can be downloaded from <https://nowphas.mlit.go.jp>.

Figure 1 shows the locations of the GPS buoys. There are 18 buoys (A–R) near the coast of Japan. Most of the GPS buoys are located between 10 km and 20 km from the coast. The water depth at most of the GPS buoy locations is over 100 m. Figure 1 also shows the land grid points at  $0.5^\circ$  intervals on the ERA5 grid. The distances from coast lines to GPS buoys are not well resolved in the ERA5 grid. The ERA5 wave data are bilinearly interpolated at the buoy position. If at least one of the four ERA5 grids surrounding the GPS buoy position is a land grid, the buoy wave data were not used for comparison.



**Figure 1.** Locations of the moored GPS buoys (A–R). Black squares: land grid points on ERA5 grids.

The skill metrics for the comparisons are:

$$R_a = R_a(X, Y) = \frac{\langle Y \rangle}{\langle X \rangle} \quad (2)$$

$$R_d = R_d(X, Y) = \langle (X - Y)^2 \rangle^{\frac{1}{2}}, \quad (3)$$

$$r_c = r_c(X, Y) = \frac{\langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle}{\langle (X - \langle X \rangle)^2 \rangle^{\frac{1}{2}} \langle (Y - \langle Y \rangle)^2 \rangle^{\frac{1}{2}}}, \quad (4)$$

$$SI = SI(X, Y) = \frac{1}{\langle X \rangle} \langle [(Y - \langle Y \rangle) - (X - \langle X \rangle)]^2 \rangle^{\frac{1}{2}}, \quad (5)$$

$$C_{rmsd} = C_{rmsd}(X, Y) = \frac{\langle [(Y - \langle Y \rangle) - (X - \langle X \rangle)]^2 \rangle^{\frac{1}{2}}}{(X - \langle X \rangle)^2}, \quad (6)$$

$$S_{sdn} = S_{sdn}(X, Y) = \frac{\langle (Y - \langle Y \rangle)^2 \rangle^{\frac{1}{2}}}{\langle (X - \langle X \rangle)^2 \rangle^{\frac{1}{2}}}, \quad (7)$$

where  $X$  and  $Y$  denote the observed and computed values and  $\langle \dots \rangle$  denotes averaging. The skill metric  $R_d$  is the root mean squared difference (RMSD);  $r_c$  is the correlation coefficient;  $SI$  is the scatter index;  $C_{rmsd}$  is the normalized centered RMSD (CRMSD); and  $S_{sdn}$  is the normalized standard deviation (NSD). The closer  $R_a$  is to 1, the closer  $R_d$  is to 0, and the closer  $r_c$  is to 1, the higher the accuracy of  $Y$ . The  $SI$  is defined as the ratio of the standard deviation of the differences to the mean observed value. The CRMSD and NSD are normalized by the observations, and it is possible to compare across different data groups. The closer  $SI$  is to 0, the closer CRMSD is to 0, and the closer NSD is to 1, the higher the accuracy of  $Y$ .

The period of the intercomparison of GPS buoy wave data, ERA5 wave data, and JMA wave data is from 2014 to 2018. The error of wave forecast by the GWM has been small since 2014 [30]. The period of the validation of ERA5 wave data is from 2012 to 2018.

### 3. Results

#### 3.1. Comparison of Wave Heights

Figure 2 shows an example of significant wave heights by ERA5 and JMA. The wave heights at some of the GPS buoy locations cannot be evaluated from the ERA5 wave heights because of the low spatial resolution. The patterns of wave heights are similar to each other. However, there are some differences between them. For example, the JMA wave heights around the buoys A, B, and C are higher than those of the ERA5 wave heights.

Figure 3 shows the scatter density plots between  $H_g$  and  $H_e$  and between  $H_g$  and  $H_j$ . This figure shows the ratio of the number of wave height data in the 0.2 m bins to the total data. For example, the number of  $(H_g, H_e)$  satisfying  $0.8 \text{ m} \leq H_g < 1 \text{ m}$  and  $0.8 \text{ m} \leq H_e < 1 \text{ m}$  is 3222; the number of comparisons is  $N_c = 58411$ ; and the ratio is  $3222/58411 = 5.5\%$  (Figure 3a). Table 1 summarizes the skill metrics of the comparison of the wave heights for each GPS buoy. The 10 GPS buoys' data were used for the comparison. For the other eight GPS buoys, at least one of the four ERA5 grids that surround them is a land grid.

The skill metrics are  $R_d(H_g, H_e) > R_d(H_g, H_j)$ ,  $r_c(H_g, H_e) < r_c(H_g, H_j)$ , and  $SI(H_g, H_e) > SI(H_g, H_j)$  for most of the buoys, except Buoy K. Thus, as explained in Section 2, the skill metrics for  $H_j$  are better than those for  $H_e$ . The accuracy of the JMA wave heights is better than that of ERA5 wave heights.

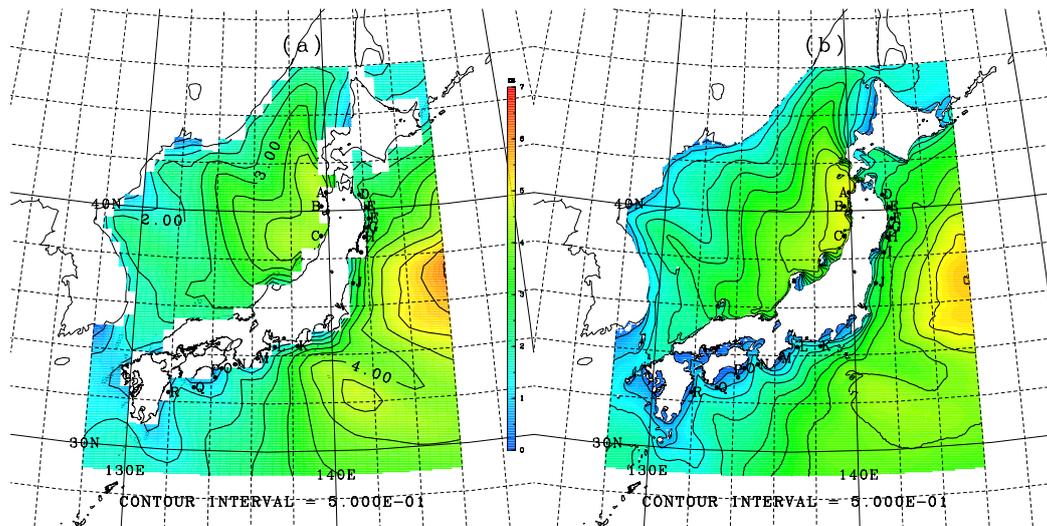


Figure 2. Example of (a) ERA5 wave height and (b) JMA wave height at 18 UT 1 January 2018.

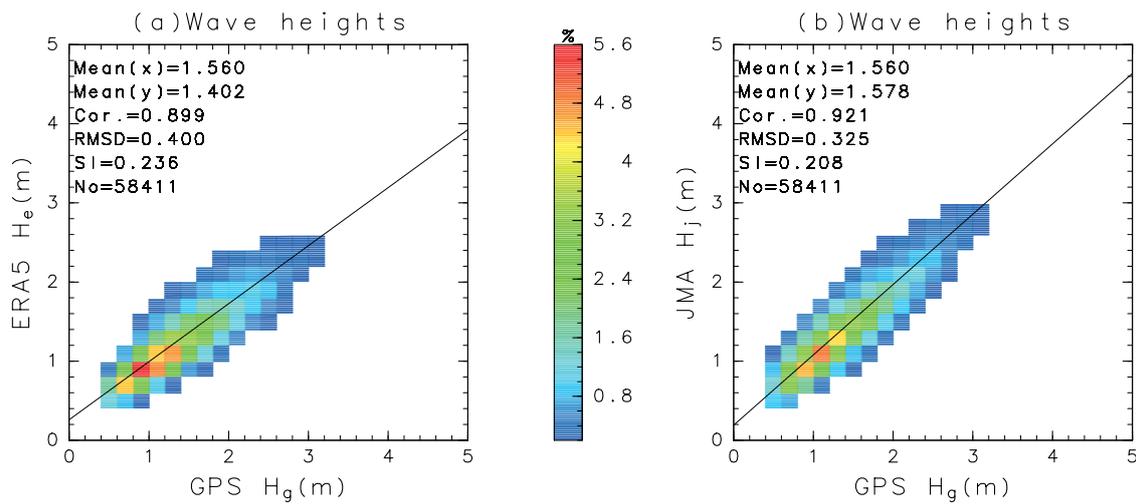


Figure 3. Scatter density plots (a) between  $H_g$  and  $H_e$  and (b) between  $H_g$  and  $H_j$  in the 0.2 m bins. The line indicates linear regression. The period of the comparison is from 2014 to 2018. SI, scatter index.

Table 1. Comparison of  $H_g$  ( $X = H_g$ ) with  $H_e$  and  $H_j$  ( $Y = H_e$  or  $Y = H_j$ ) for various buoys from 2014 to 2018.  $N_c$ : number of comparisons. Other symbols are defined in Equations (2)–(6).

Buoy	Y	$\langle X \rangle$ (m)	$\langle Y \rangle$ (m)	$R_a$	$r_c$	$R_d$ (m)	SI	$C_{rmsd}$	$S_{sdn}$	$N_c$
D	$H_e$	1.473	1.392	0.945	0.884	0.371	0.246	0.470	0.849	7197
D	$H_j$	1.473	1.497	1.016	0.927	0.290	0.196	0.376	0.961	7197
E	$H_e$	1.666	1.381	0.829	0.907	0.455	0.213	0.441	0.779	7179
E	$H_j$	1.666	1.600	0.960	0.946	0.269	0.156	0.324	0.954	7179
F	$H_e$	1.697	1.449	0.854	0.911	0.427	0.205	0.429	0.792	7124
F	$H_j$	1.697	1.669	0.983	0.944	0.271	0.159	0.332	0.966	7124

Table 1. Cont.

Buoy	Y	$\langle X \rangle$ (m)	$\langle Y \rangle$ (m)	$R_a$	$r_c$	$R_d$ (m)	SI	$C_{rmsd}$	$S_{sdn}$	$N_c$
G	$H_e$	1.624	1.425	0.877	0.892	0.433	0.236	0.467	0.773	6842
G	$H_j$	1.624	1.642	1.011	0.939	0.284	0.174	0.345	0.953	6842
J	$H_e$	1.722	1.524	0.885	0.913	0.386	0.192	0.414	0.837	7184
J	$H_j$	1.722	1.657	0.962	0.921	0.320	0.182	0.391	0.964	7184
K	$H_e$	1.739	1.471	0.846	0.910	0.446	0.205	0.418	0.856	3847
K	$H_j$	1.739	1.683	0.968	0.883	0.410	0.233	0.475	0.950	3847
L	$H_e$	1.193	1.162	0.974	0.899	0.305	0.254	0.444	0.828	4344
L	$H_j$	1.193	1.516	1.270	0.915	0.441	0.252	0.440	1.091	4344
O	$H_e$	1.341	1.247	0.930	0.904	0.343	0.246	0.429	0.865	5263
O	$H_j$	1.341	1.306	0.974	0.910	0.335	0.249	0.433	1.038	5263
P	$H_e$	1.408	1.433	1.017	0.918	0.346	0.245	0.400	0.866	5637
P	$H_j$	1.408	1.552	1.102	0.928	0.358	0.233	0.380	1.001	5637
Q	$H_e$	1.608	1.482	0.922	0.892	0.441	0.263	0.459	0.813	3794
Q	$H_j$	1.608	1.633	1.015	0.933	0.333	0.206	0.360	0.955	3794
Total	$H_e$	1.560	1.402	0.899	0.899	0.400	0.236	0.445	0.815	58,411
Total	$H_j$	1.560	1.578	1.012	0.921	0.325	0.208	0.393	0.966	58,411

3.2. Comparison under Various Conditions

We investigated the accuracy of ERA5 and JMA wave heights with increasing wave height.

Figure 4a shows the ratios of mean wave heights ( $R_a$ ) and the CRMSD ( $C_{rmsd}$ ) as a function of  $H_T$  for  $H_g \geq H_T$ . Figure 4b shows the number of comparisons for  $H_g \geq H_T$ , where  $H_T$  is the threshold of wave height. For example, if  $H_g \geq 3$  m ( $H_T = 3$  m),  $R_a(H_g, H_e) = 0.81$ ,  $R_a(H_g, H_j) = 0.96$ ,  $C_{rmsd}(H_g, H_e) = 0.60$ ,  $C_{rmsd}(H_g, H_j) = 0.56$  (Figure 4a), and  $N_c = 3172$  (Figure 4b). As the wave height increases, the ERA5 wave height tends to be underestimated. The CRMSD increases as the wave height increases.

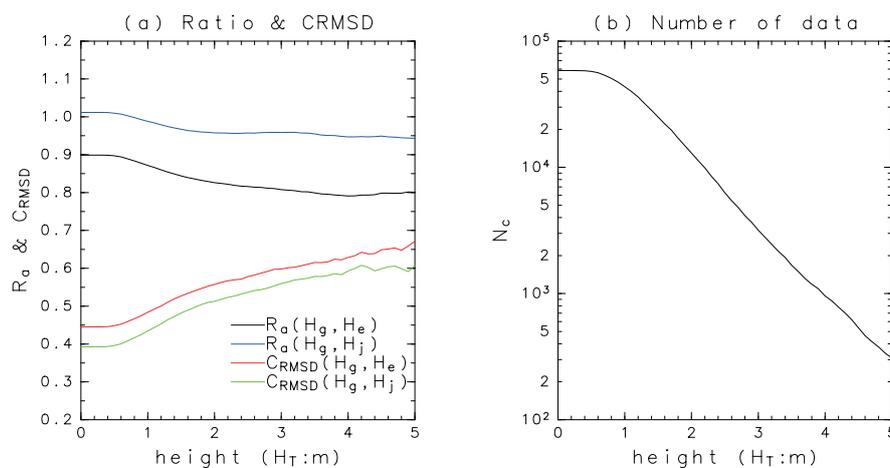


Figure 4. (a) Ratios of mean wave heights ( $R_a$ ) and the centered RMSD (CRMSD) ( $C_{rmsd}$ ) as a function of  $H_T$  and (b) the number of data for comparison, which satisfies  $H_g \geq H_T$ . (a) Black:  $R_a(H_g, H_e)$  (Equation (2)). Blue:  $R_a(H_g, H_j)$ . Red:  $C_{rmsd}(H_g, H_e)$  (Equation (6)). Green:  $C_{rmsd}(H_g, H_j)$ . The period of the comparison is from 2014 to 2018.

Figure 5 shows the Q-Q (quantile-quantile) plots between  $H_g$  and  $H_e$  and between  $H_g$  and  $H_j$ . The ERA5 wave heights are underestimated in higher wave conditions, while the JMA wave heights are not underestimated.

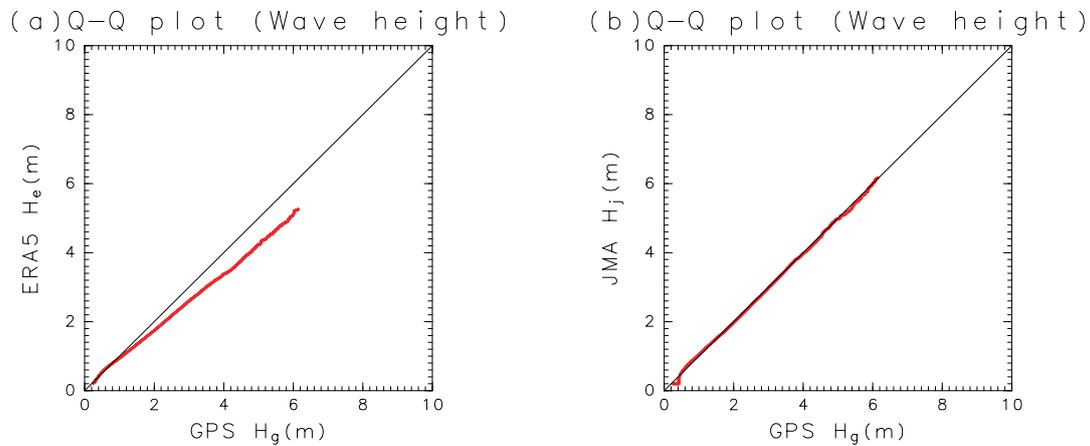


Figure 5. (a) Q-Q plots (from 0.1 to 99.9%) between  $H_g$  and  $H_e$ . Same for (b), but between  $H_g$  and  $H_j$ . The period of the comparison is from 2014 to 2018.

The development of wind waves depends not only on the wind speed, but also on the fetch length. We investigated the accuracy of ERA5 wave heights in the fetch-unlimited condition and the fetch-limited condition. The fetch-unlimited condition and the fetch-limited condition are classified from the JMA winds ( $u_j, v_j$ ) at the buoy locations. The fetch-limited conditions are  $u_j > 0$  for Buoys D, E, F, G, and J and  $v_j < 0$  for Buoys K, L, O, P, and Q (Figure 1). For example, if  $u_j > 0$  at the position of Buoy D, the fetch is limited at that time.

Figure 6 shows the Taylor diagram between  $H_g$  and  $H_e$  for individual buoys in the fetch-unlimited condition and in the fetch-limited condition. The data for validation are from 2012 to 2018. The plots of the fetch-unlimited condition (blue) and of the fetch-limited condition (red) are clustered with each other (Figure 6). The cluster of the fetch-unlimited condition is close to  $(r_c, C_{rmsd}) = (0.94, 0.39)$ , and the cluster of the fetch-limited condition is close to  $(r_c, C_{rmsd}) = (0.9, 0.5)$ . The skill indices in fetch-unlimited conditions are better than those in fetch-limited conditions except the NSD.

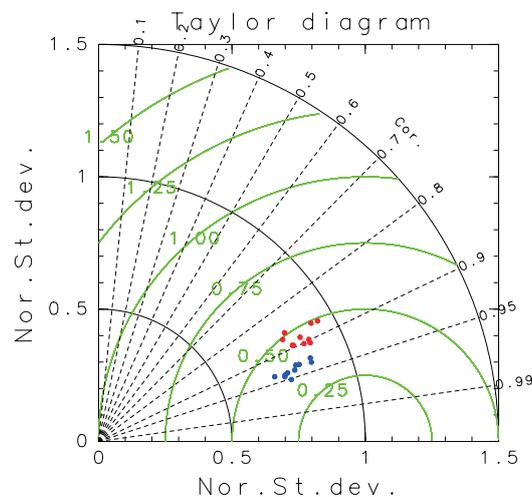


Figure 6. Taylor diagram between  $H_g$  and  $H_e$  for individual buoys in the fetch-unlimited condition (blue) and in the fetch-limited condition (red). The black arcs show normalized standard deviation (NSD) contours, and the green arcs show CRMSD contours. The period of the comparison is from 2012 to 2018.

Table 2 summarizes the skill metrics of ERA5 wave heights in the fetch-limited condition (F) and the fetch-unlimited condition (U) for each GPS buoy. In total,  $R_a(H_g, H_e, U) < R_a(H_g, H_e, F)$ ,  $r_c(H_g, H_e, U) > r_c(H_g, H_e, F)$ ,  $R_d(H_g, H_e, U) > R_d(H_g, H_e, F)$ ,  $C_{rmsd}(H_g, H_e, U) < C_{rmsd}(H_g, H_e, F)$ , and  $SI(H_g, H_e, U) > SI(H_g, H_e, F)$ , where  $R_a(H_g, H_e, U)$  indicates  $R_a$  (Equation (2)) in the fetch-unlimited condition, and  $R_a(H_g, H_e, F)$  indicates  $R_a$  in the fetch-limited condition. The same applies to other skill metrics including U and F.

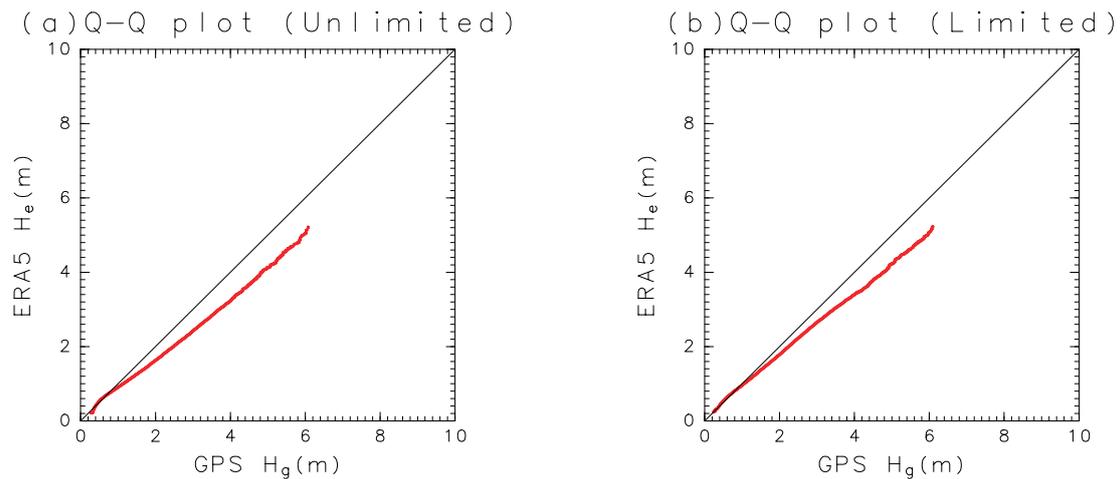
**Table 2.** Comparisons between  $H_g$  and  $H_e$  in the fetch unlimited conditions (U) and the fetch-limited conditions (F) for various buoys from 2012 to 2018. The asterisk (\*) denotes that the probability that  $C_{rmsd}(H_g, H_e, U) < C_{rmsd}(H_g, H_e, F)$  ( $|S_{sdn}(H_g, H_e, U) - 1| > |S_{sdn}(H_g, H_e, F) - 1|$ ) is less than that in the fetch-limited condition is greater than 95%. The asterisks (\*\*) denote the probability is greater than 90%, but less than 95%.

Buoy	Case	$\langle H_g \rangle$ (m)	$\langle H_e \rangle$ (m)	$R_a$	$r_c$	$R_d$ (m)	SI	$C_{rmsd}$	$S_{sdn}$	$N_c$
D	U	1.537	1.309	0.852	0.939	0.399	0.213	0.389 *	0.755	3656
D	F	1.438	1.451	1.009	0.875	0.343	0.238	0.489	0.939 *	5752
E	U	1.668	1.287	0.772	0.938	0.516	0.209	0.418 *	0.704	4380
E	F	1.722	1.482	0.860	0.896	0.434	0.209	0.452	0.816 *	5667
F	U	1.693	1.383	0.817	0.944	0.452	0.194	0.388 *	0.740	4003
F	F	1.740	1.518	0.872	0.895	0.432	0.213	0.454	0.812 *	5985
G	U	1.741	1.424	0.818	0.941	0.466	0.197	0.392 *	0.743	3163
G	F	1.597	1.448	0.907	0.873	0.429	0.252	0.494	0.790 *	6254
J	U	1.803	1.517	0.841	0.933	0.424	0.173	0.382 *	0.806	4251
J	F	1.712	1.546	0.903	0.902	0.387	0.204	0.435	0.854 *	5816
K	U	1.482	1.302	0.878	0.929	0.375	0.222	0.377 **	0.854	2698
K	F	1.948	1.586	0.814	0.898	0.504	0.180	0.440	0.878	4048
L	U	1.193	1.062	0.891	0.951	0.312	0.237	0.363 *	0.760	1858
L	F	1.198	1.212	1.012	0.887	0.311	0.260	0.463	0.852 *	3216
O	U	1.410	1.250	0.886	0.939	0.372	0.238	0.378 *	0.783	2697
O	F	1.331	1.228	0.923	0.872	0.355	0.255	0.491	0.913 *	4284
P	U	1.400	1.428	1.020	0.936	0.335	0.238	0.361 *	0.852	2115
P	F	1.413	1.436	1.016	0.905	0.352	0.249	0.426	0.877	3522
Q	U	1.618	1.402	0.867	0.932	0.423	0.225	0.389 *	0.794	2542
Q	F	1.543	1.468	0.951	0.861	0.452	0.289	0.511	0.809	3706
Total	U	1.593	1.350	0.848	0.931	0.424	0.218	0.398	0.772	31363
Total	F	1.586	1.450	0.914	0.882	0.406	0.241	0.474	0.833	48250

The statistical significance of  $C_{rmsd}(H_g, H_e, U) < C_{rmsd}(H_g, H_e, F)$ , and  $|S_{sdn}(H_g, H_e, U) - 1| > |S_{sdn}(H_g, H_e, F) - 1|$ , which can be compared across different data groups, is explored. The bootstrap method [33,34] is used for the validation. The effective sample size  $N_e$  is evaluated by the method in [33,35], which is smaller than the number of comparisons  $N_c$  ( $1 \leq N_e \leq N_c$ ) from the serial data of  $(H_g, H_e)$ . The  $N_e$  data  $(H_g, H_e)$  were resampled randomly, and they may be sampled in duplicate. The skill metrics were computed from  $N_e$  data  $(H_g, H_e)$ . The resamples and computations of the skill metrics were conducted 1000 times, and the probabilities of  $C_{rmsd}(H_g, H_e, U) < C_{rmsd}(H_g, H_e, F)$ , and  $|S_{sdn}(H_g, H_e, U) - 1| > |S_{sdn}(H_g, H_e, F) - 1|$  were evaluated from 1000 resamples. This method is almost the same as those in [34,36]. The probabilities of  $C_{rmsd}(H_g, H_e, U) < C_{rmsd}(H_g, H_e, F)$  were greater than 95% in nine buoys out of 10 buoys. The result that the CRMSD in the fetch-unlimited case

is smaller than that in the fetch-limited case is statistically significant more than at the 95% confidence level. The probabilities of  $|S_{sdn}(H_g, H_e, U) - 1| > |S_{sdn}(H_g, H_e, F) - 1|$  were greater than 95% in seven buoys out of 10 buoys.

Figure 7 shows the Q-Q plots between  $H_g$  and  $H_e$  in the fetch-limited condition and the fetch-unlimited condition. The ERA5 wave heights are underestimated as the wave height increases in both cases. Although the ratio  $R_a(H_g, H_e)$  is larger in the fetch-limited condition than in the fetch-unlimited condition, the ERA5 wave heights in the fetch-limited condition are also underestimated in high wave conditions.



**Figure 7.** (a) Q-Q plots (from 0.1 to 99.9%) between  $H_g$  and  $H_e$  in the fetch-unlimited condition Same for (b) in the fetch-limited condition. The period of the comparison is from 2012 to 2018.

### 3.3. Comparison of Wave Periods

The comparisons of wave periods are summarized in Table 3. The correlations between peak wave periods ( $T_{pe}$  and  $T_{pj}$ ) and significant wave periods ( $T_g = T_{\frac{1}{3}}$ ) from GPS buoys are  $r_c(T_g, T_{pe}) > r_c(T_g, T_{pj})$ . On the other hand, their SIs are  $SI(T_g, T_{pe}) > SI(T_g, T_{pj})$ , but the difference is small.

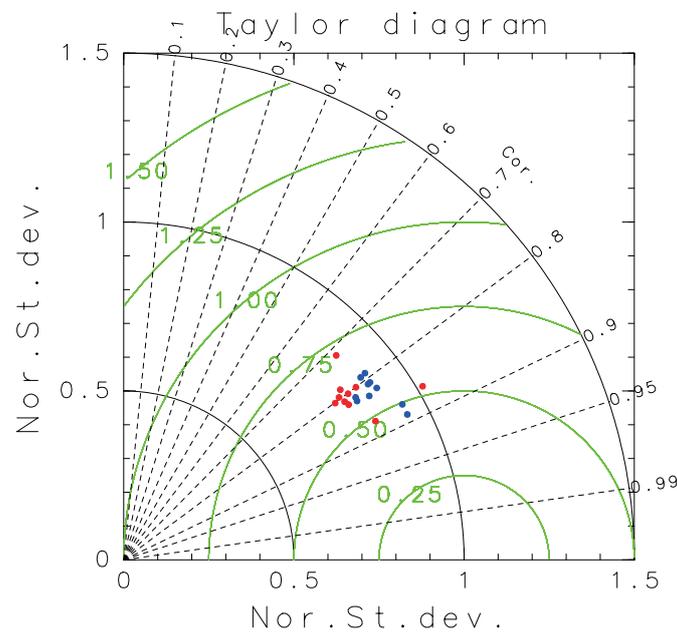
**Table 3.** Comparison of periods ( $X = T_g, Y = T_{pe}$ , or  $Y = T_{pj}$ , or  $Y = T_{me}$ ). T: total. U: fetch-unlimited conditions. F: fetch-limited conditions. Other symbols are defined in Equations (2)–(6). The comparison period with  $T_{pe}, T_{pj}$ , and  $T_{me}$  for the total cases (T) is from 2014 to 2018. The comparison period with  $T_{me}$  for the fetch-unlimited (U) and the fetch-limited (F) cases is from 2012 to 2018.

Y	Case	$\langle X \rangle$ (s)	$\langle Y \rangle$ (s)	$R_a$	$r_c$	$R_d$ (s)	SI	$C_{rmsd}$	$S_{tdn}$	$N_c$
$T_{pe}$	T	7.301	8.910	1.220	0.671	2.376	0.239	1.019	1.370	58,411
$T_{pj}$	T	7.301	8.702	1.192	0.610	2.216	0.235	1.000	1.221	58,411
$T_{me}$	T	7.301	7.320	1.003	0.800	1.038	0.142	0.605	0.872	58,411
$T_{me}$	U	7.503	7.540	1.005	0.801	0.979	0.130	0.606	0.897	32,044
$T_{me}$	F	7.190	7.150	0.994	0.792	1.093	0.152	0.612	0.833	47,561

The mean period  $T_{m-1}$  is the closest to  $T_{\frac{1}{3}}$  in the moment periods ( $T_{m-1}, T_{m1} = M_0/M_1$ , and  $T_{m2} = (M_0/M_2)^{1/2}$ ) [37]. The value of  $R_a(T_g, T_{me})$  (Equation (2)) is close to one. In total, the ERA5 mean wave periods in the fetch-unlimited conditions are closer to  $T_g$  than those in the fetch-limited conditions

Figure 8 shows the Taylor diagram between  $T_g$  and  $T_{me}$  for individual buoys in the fetch-unlimited condition and in the fetch-limited condition from 2012 to 2018. The plots are scattered compared with Figure 6. In particular, the skill metrics in the fetch-limited conditions are more scattered than those in the fetch-unlimited conditions. The CRMSDs ( $C_{rmsd}$ ) in the fetch-unlimited conditions tend to be

smaller than those in the fetch-limited condition. The NSD ( $S_{sdn}$ ) in the fetch-unlimited conditions is closer to one than those in the fetch-limited condition in most of the buoys.



**Figure 8.** Same as Figure 6, but for between  $T_g$  and  $T_{me}$ . The period of the comparison is from 2012 to 2018.

The probabilities of  $C_{rmsd}(T_g, T_{me}, U) < C_{rmsd}(T_g, T_{me}, F)$  and  $|S_{sdn}(T_g, T_{me}, U) - 1| < |S_{sdn}(T_g, T_{me}, F) - 1|$  are also explored by the bootstrap method. The probabilities of  $C_{rmsd}(T_g, T_{me}, U) < C_{rmsd}(T_g, T_{me}, F)$  were more than 90% in four buoys out of 10 buoys. The probabilities of  $|S_{sdn}(T_g, T_{me}, U) - 1| < |S_{sdn}(T_g, T_{me}, F) - 1|$  were more than 95% in seven buoys out of 10 buoys. The accuracy of the ERA5 wave period in the fetch-unlimited condition is better than that in the fetch-limited condition. However, the result is not so robust as wave height, because the skill metrics of wave periods are scattered (Figure 8), while the skill metrics of wave heights are clustered (Figure 6).

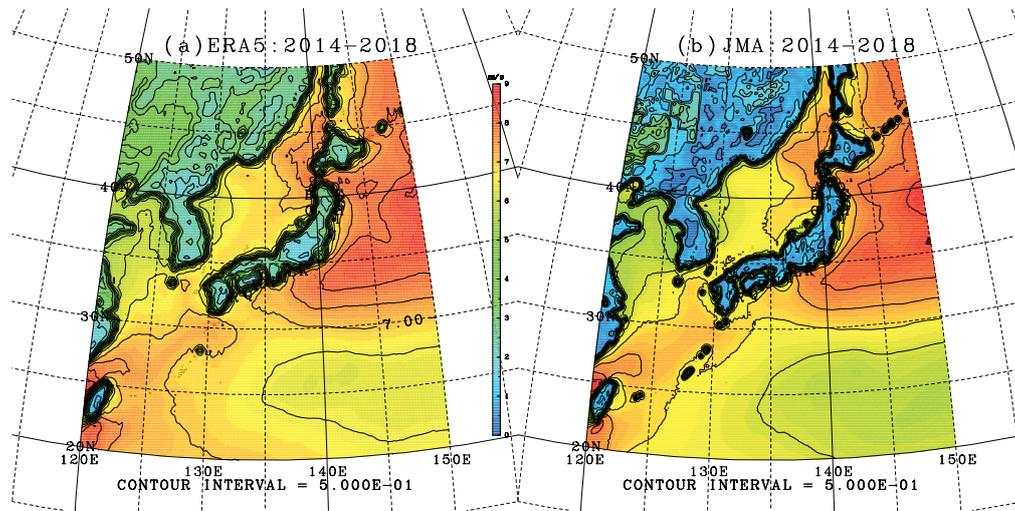
#### 4. Discussion

The intercomparison of ERA5 wave data, JMA wave assimilation data, and GPS-observed wave data in the coastal area was conducted. The period of the intercomparison was from 2014 to 2018. The accuracy of wave forecast by the GWM of JMA was not good in 2013, because the wind field associated with a typhoon was not reconstructed accurately [30]. In fact, the skill metrics from 2012 to 2018 were  $r_c(H_g, H_j) = 0.901$ ,  $R_d(H_g, H_j) = 0.366$  m, and  $SI(H_g, H_j) = 0.230$  for  $N_c = 79613$ . The skill metrics of the JMA analysis wave height were lower than those in Figure 3b and Table 1.

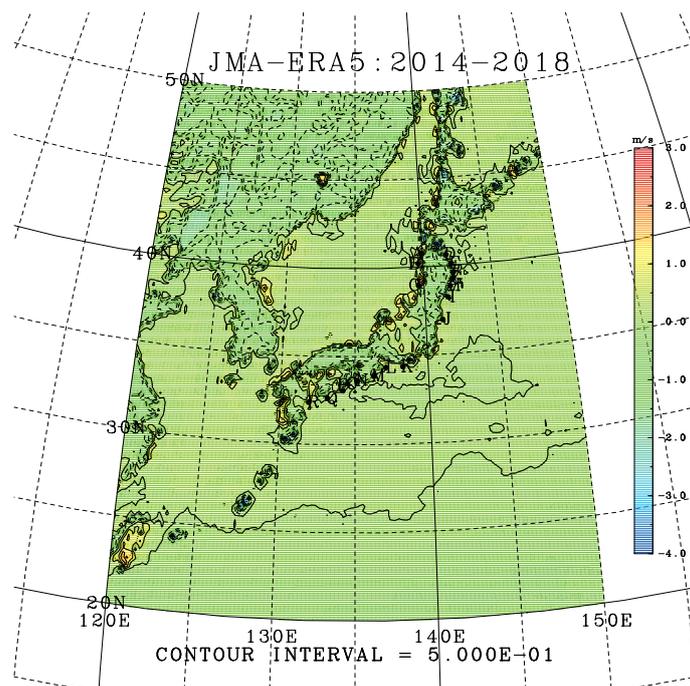
The mean wave heights were  $\langle H_e \rangle < \langle H_j \rangle$  in all of the buoy positions (Table 1).

Figure 9 shows the mean ERA5 wind speeds and mean JMA wind speeds from 2014 to 2018. Figure 9a and Figure 9b are similar to each other. The wind speeds at the GPS buoy positions range from about 5 m/s to 7 m/s.

Figure 10 shows the differences of mean ERA5 wind speeds and mean JMA wind speeds from 2014 to 2018 ( $\langle |U_j| \rangle - \langle |U_e| \rangle$ ). The differences of mean wind speeds ( $\langle |U_j| \rangle - \langle |U_e| \rangle$ ) at the GPS buoy positions range from about  $-0.5$  m/s to  $0.3$  m/s. The mean wind speeds are  $\langle |U_j| \rangle < \langle |U_e| \rangle$  at five buoy positions out of 10 buoy positions. The difference of the local wind speeds cannot explain the result that  $\langle H_e \rangle < \langle H_j \rangle$  in all of the buoy positions.



**Figure 9.** (a) Mean ERA5 wind speeds from 2014 to 2018 ( $\langle |U_e| \rangle$ ). (b) Same as (a), but for JMA wind speed ( $\langle |U_j| \rangle$ ).



**Figure 10.** Differences of mean ERA5 wind speeds and mean JMA wind speeds from 2014 to 2018 ( $\langle |U_j| \rangle - \langle |U_e| \rangle$ ).

GPS buoy-measured wave height data were assimilated into JMA wave data. Therefore, the accuracy of JMA wave height was better than that of ERA5. On the other hand, GPS buoy-measured wave period data were not assimilated into JMA wave data. The accuracy of JMA wave period was not better than that of ERA5.

The spatial resolution of the ERA5 wave model was  $0.36^\circ$ . On the other hand, the JMA model had a spatial resolution of  $0.5^\circ$  outside the CWM region (the CWM region is from  $20^\circ$  N to  $50^\circ$  N and from  $120^\circ$  E to  $150^\circ$  E), which was lower than that of the ERA5 wave model. It is presumed that this is

because the ERA5 wave model had better reproducibility of the swell propagating from outside the CWM region. As a result, the accuracy of JMA wave period was not better than that of ERA5.

The ERA5 wave height was underestimated in higher wave conditions. The underestimation of ERA-Interim wave height in higher wave conditions was shown in [22]. This underestimation is true for ERA5 wave height.

It was found that the accuracy of ERA5 wave height was significantly different between the fetch-limited and fetch-unlimited conditions. In particular, the correlation of wave heights in fetch-limited conditions was lower than that in fetch-unlimited conditions, although the correlation cannot be compared across different data groups.

On the other hand, The ratio  $R_a = \langle H_e \rangle / \langle H_g \rangle$  and NSD ( $S_{sdn}(H_g, H_e)$ ) were closer to one in the fetch-limited conditions than those in the fetch-unlimited conditions. The ERA5 wave heights tended to be underestimated. However, the distance between the buoy position and the coast was overestimated in the ERA5 grid (Figures 1 and 2a). The fetch was overestimated at the buoy locations, and  $R_a$  was larger. The normalized variability of ERA5 wave height in the fetch-limited condition was larger than that in the fetch-unlimited condition. This was also related to the overestimation of fetch in the ERA5 grid. On the other hand, the Q-Q plots in the fetch-limited condition and the fetch-unlimited condition were similar to each other. Even though the fetch was overestimated, ERA5 wave height was underestimated when the wave height was high (Figure 7).

The statistical significance of the difference of the skill metrics between the fetch-limited conditions and the fetch-unlimited conditions was investigated for each buoy's data. The wave height changes seasonally, and the effective sample sizes  $N_e$  were smaller than the number of comparisons  $N_c$ . Therefore, the uncertainties of the skill metrics were large. However, the difference of the skill metrics were statistically significant at more than 90% confidence levels in most of the buoys. The uncertainties of the the skill metrics can be smaller as the data period will be extended.

## 5. Conclusions

The conclusions are summarized as follows:

- The accuracy of JMA analysis wave height is better than that of ERA5 wave height by incorporating the observation data near the coast.
- The accuracy of JMA analysis wave period is not better than that of ERA5 wave period.
- The ERA5 wave height is underestimated as higher wave heights.
- The accuracy of ERA5 wave height in the fetch-limited conditions is significantly lower than that in the fetch-unlimited conditions.
- The accuracy of ERA5 wave period in the fetch-limited conditions is also lower than that in the fetch-unlimited conditions, but this is not so robust as wave height.

From these conclusions, the JMA wave height analysis data can be used as wave climate data around Japan. In addition, if ERA5 wave data were to be used in the various marine development guidelines, ERA5 wave data should be treated separately from the fetch-limited conditions and the fetch-unlimited conditions. Moreover, since ERA5 wave heights tend to be underestimated at higher wave heights, it is necessary to pay attention to this result when using the ERA5 wave height for marine disaster prevention.

**Funding:** This study was financially supported by a Grant-in-Aid for Scientific Research (C-2) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (20K04708).

**Acknowledgments:** The GPS buoy wave data were observed by Maritime Bureau, Ministry of Land, Infrastructure, Transport, and Tourism of Japan and analyzed by the Port and Airport Research Institute. The GFD DENNOU Library (available online at <http://www.gfd-dennou.org/arch/dcl/>) was used for drawing the figures. Comments from the anonymous reviewers were helpful in improving the manuscript.

**Conflicts of Interest:** The author declares no conflict of interest.

## Abbreviations

The following abbreviations are used in this paper:

ERA5	Fifth generation of European Centre for Medium Range Weather Forecasting atmospheric reanalyses of the global climate
JMA	Japan Meteorological Agency
ERA-Interim	European Centre for Medium-Range Weather Forecasts reanalysis data interim version
GPS	Global Positioning System
CWM	Coastal wave model
GWM	Global wave model
OI	Optimal interpolation
MRI	Meteorological Research Institute
RMSD	Root mean squared difference
SI	Scatter index
CRMSD	Normalized centered root mean squared difference
NSD	Normalized standard deviation
Q-Q plot	Quantile-quantile plot

## Notations

$H_e$	ERA5 wave height.
$H_g$	GPS wave height.
$H_j$	JMA wave height.
$T_{me}$	ERA5 mean wave period.
$T_{pe}$	ERA5 peak wave period.
$T_g$	GPS wave period.
$T_{pj}$	JMA peak wave period.
$\mathbf{U}_j = (u_j, v_j)$	JMA wind vector.
$\mathbf{U}_e$	ERA5 wind vector.
$M_{-1}, M_0, M_1, M_2$	Equation (1).
$H_s = 4M_0^{\frac{1}{2}}$	Equation (1).
$H_{\frac{1}{3}}$	significant wave height by the zero-up-crossing method.
$T_{\frac{1}{3}}$	significant wave period by the zero-up-crossing method.
$T_{m1} = M_0/M_1$	
$T_{m-1} = M_{-1}/M_0$	
$T_{m2} = (M_0/M_2)^{1/2}$	
$R_a$	ratio of averages (Equation (2)).
$R_d$	RMSD (Equation (3)).
$r_c$	correlation coefficient (Equation (4)).
$C_{rmsd}$	CRMSD (Equation (6)).
$S_{sdn}$	NSD (Equation (7)).
$H_T$	threshold of wave height.
$N_c$	number of comparisons.
$N_e$	effective sample size.

## References

1. Reguero, B.; Menéndez, M.; Méndez, F.; Mínguez, R.; Losada, I. A Global Ocean Wave (GOW) calibrated reanalysis from 1948 onwards. *Coast. Eng.* **2012**, *65*, 38–55. [[CrossRef](#)]
2. Shi, J.; Zheng, J.; Zhang, C.; Joly, A.; Zhang, W.; Xu, P.; Sui, T.; Chen, T. A 39-year high resolution wave hindcast for the Chinese coast: Model validation and wave climate analysis. *Ocean Eng.* **2019**, *183*, 224–235. [[CrossRef](#)]
3. Groll, N.; Weisse, R. A multi-decadal wind-wave hindcast for the North Sea 1949–2014: coastDat2. *Earth Syst. Sci. Data* **2017**, *9*, 955. [[CrossRef](#)]

4. Haakenstad, H.; Breivik, Ø.; Reistad, M.; Aarnes, O.J. NORA10EI: A revised regional atmosphere-wave hindcast for the North Sea, the Norwegian Sea and the Barents Sea. *Int. J. Climatol.* **2020**. [[CrossRef](#)]
5. Shimura, T.; Mori, N. High-resolution wave climate hindcast around Japan and its spectral representation. *Coast. Eng.* **2019**, *151*, 1–9. [[CrossRef](#)]
6. Taniguchi, K. Variations in winter ocean wave climate in the Japan sea under the global warming condition. *J. Mar. Sci. Eng.* **2019**, *7*, 150. [[CrossRef](#)]
7. Hu, Y.; Shao, W.; Wei, Y.; Zuo, J. Analysis of typhoon-induced waves along typhoon tracks in the Western north Pacific Ocean, 1998–2017. *J. Mar. Sci. Eng.* **2020**, *8*, 521. [[CrossRef](#)]
8. Chowdhury, P.; Behera, M.R. Evaluation of CMIP5 and CORDEX derived wave climate in Indian Ocean. *Clim. Dyn.* **2019**, *52*, 4463–4482. [[CrossRef](#)]
9. Bonaduce, A.; Staneva, J.; Behrens, A.; Bidlot, J.R.; Wilcke, R.A.I. Wave climate change in the North Sea and Baltic Sea. *J. Mar. Sci. Eng.* **2019**, *7*, 166. [[CrossRef](#)]
10. Li, N.; Cheung, K.F.; Stopa, J.E.; Hsiao, F.; Chen, Y.L.; Vega, L.; Cross, P. Thirty-four years of Hawaii wave hindcast from downscaling of climate forecast system reanalysis. *Ocean Model.* **2016**, *100*, 78–95. [[CrossRef](#)]
11. Oliveira, B.A.; Sobral, F.; Fetter, A.; Mendez, F.J. A high-resolution wave hindcast off Santa Catarina (Brazil) for identifying wave climate variability. *Reg. Stud. Mar. Sci.* **2019**, *32*, 100834. [[CrossRef](#)]
12. Waters, R.; Engström, J.; Isberg, J.; Leijon, M. Wave climate off the Swedish west coast. *Renew. Energy* **2009**, *34*, 1600–1606. [[CrossRef](#)]
13. Onea, F.; Rusu, L. A long-term assessment of the Black Sea wave climate. *Sustainability* **2017**, *9*, 1875. [[CrossRef](#)]
14. Saprykina, Y.; Shtremel, M.; Aydoğan, B.; Ayat, B. Variability of the Nearshore Wave Climate in the Eastern Part of the Black Sea. *Pure Appl. Geophys.* **2019**, *176*, 3757–3768. [[CrossRef](#)]
15. Vieira, F.; Cavalcante, G.; Campos, E. Analysis of wave climate and trends in a semi-enclosed basin (Persian Gulf) using a validated SWAN model. *Ocean Eng.* **2020**, *196*, 106821. [[CrossRef](#)]
16. Hoffmann, L.; Günther, G.; Li, D.; Stein, O.; Wu, X.; Griessbach, S.; Heng, Y.; Konopka, P.; Müller, R.; Vogel, B.; et al. From ERA-Interim to ERA5: The considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations. *Atmos. Chem. Phys.* **2019**, *19*. [[CrossRef](#)]
17. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**. [[CrossRef](#)]
18. Kohno, N.; Miura, D.; Yoshita, K. The development of JMA wave data assimilation system. In Proceedings of the 12th International Workshop on Wave Hindcasting and Forecasting, Kona, HI, USA, 30 October–4 November 2011; pp. 1–8. Available online: [http://www.waveworkshop.org/12thWaves/papers/full\\_paper\\_Kohno\\_et\\_al.pdf](http://www.waveworkshop.org/12thWaves/papers/full_paper_Kohno_et_al.pdf) (accessed on 31 July 2020).
19. Pastor, J.; Liu, Y.; Dou, Y. Wave energy resource analysis for use in wave energy conversion. *J. Offshore Mech. Arct. Eng.* **2015**, *137*, 011903-1.
20. Guiberteau, K.L.; Liu, Y.; Lee, J.; Kozman, T. *Potential of Development and Application of Wave Energy Conversion Technology in the Gulf of Mexico*; Energy Systems Laboratory: New Orleans, LA, USA, 2014; pp. 132–141.
21. Pastor, J.; Liu, Y. Wave climate resource analysis based on a revised gamma spectrum for wave energy conversion technology. *Sustainability* **2016**, *8*, 1321. [[CrossRef](#)]
22. Stopa, J.E.; Cheung, K.F. Intercomparison of wind and wave data from the ECMWF Reanalysis Interim and the NCEP Climate Forecast System Reanalysis. *Ocean Model.* **2014**, *75*, 65–83. [[CrossRef](#)]
23. Campos, R.; Soares, C.G. Comparison of HIPOCAS and ERA wind and wave reanalyses in the North Atlantic Ocean. *Ocean Eng.* **2016**, *112*, 320–334. [[CrossRef](#)]
24. Sanil Kumar, V.; Muhammed Naseef, T. Performance of ERA-Interim wave data in the nearshore waters around India. *J. Atmos. Ocean. Technol.* **2015**, *32*, 1257–1269. [[CrossRef](#)]
25. Shanas, P.; Kumar, V.S. Temporal variations in the wind and wave climate at a location in the eastern Arabian Sea based on ERA-Interim reanalysis data. *Nat. Hazards Earth Syst. Sci.* **2014**, *14*, 1371–1381. [[CrossRef](#)]
26. Bruno, M.F.; Molfetta, M.G.; Totaro, V.; Mossa, M. Performance Assessment of ERA5 wave data in a swell dominated region. *J. Mar. Sci. Eng.* **2020**, *8*, 214. [[CrossRef](#)]
27. Sifnioti, D.; Dolphin, T.; Vincent, C. Performance of Hindcast Wave Model Data used in UK Coastal Waters. *J. Coast. Res.* **2020**, *95*, 1284–1290. [[CrossRef](#)]
28. Muhammed Naseef, T.; Sanil Kumar, V. Climatology and trends of the Indian Ocean surface waves based on 39-year long ERA5 reanalysis data. *Int. J. Climatol.* **2020**, *40*, 979–1006. [[CrossRef](#)]

29. Ueno, K.; Kohno, N. The development of the third generation wave model MRI-III for operational use. In Proceedings of the 8th International Workshop on Wave Hindcasting and Forecasting, North Shore, HI, USA, 14–19 November 2004; Available online: <http://www.waveworkshop.org/8thWaves/Papers/G2.pdf> (accessed on 31 July 2020).
30. JMA. Outline of the Operational Numerical Weather Prediction at the Japan Meteorological Agency. 2019. Available online: <https://www.jma.go.jp/jma/jma-eng/jma-center/nwp/outline2019-nwp/index.htm> (accessed on 31 July 2020).
31. Kawaguchi, K.; Inomata, T.; Seki, K.K.; Fujiki, T. *Annual Report on Nationwide Ocean Wave Information Network for Ports and Harbours (NOWPHAS 2012)*; Port and Airport Research Institute Technical Note; Port and Airport Research Institute: Yokosuka, Japan, 2014; Available online: <https://www.pari.go.jp/search-pdf/no1282.pdf> (accessed on 31 July 2020).
32. Goda, Y. *Random Seas and Design of Maritime Structures*; World Scientific Publishing Company: Singapore, 2010; Volume 33. [CrossRef]
33. Emery, W.J.; Thomson, R.E. *Data Analysis Methods in Physical Oceanography*; Pergamon: İzmir, Turkey, 1998.
34. Hisaki, Y. Wave hindcast in the North Pacific area considering the propagation of surface disturbances. *Prog. Oceanogr.* **2018**, *165*, 332–347. [CrossRef]
35. Trenberth, K.E. Some effects of finite sample size and persistence on meteorological statistics. Part II: Potential predictability. *Mon. Weather Rev.* **1984**, *112*, 2369–2379. <2359:SEOFSS>2.0.CO;2. [CrossRef]
36. Hisaki, Y. Sea surface wind correction using HF ocean radar and its impact on coastal wave prediction. *J. Atmos. Ocean. Technol.* **2017**, *34*, 2001–2020. [CrossRef]
37. Yamaguchi, M.; Nonaka, H. Effect of Spectral Form on the Ratio between Surface Elevation-Based Statistics and Spectrum-Based Statistics of Wave Height and Wave Period. *J. Jpn. Soc. Civ. Eng. Ser. B2 (Coast. Eng.)* **2009**, *65*, 136–140. (In Japanese) [CrossRef]



© 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).