A sampling technique to compare climate simulations with sparse satellite observations: performance evaluation of a CMIP5 EC-Earth forced dynamical wave climate ensemble with altimeter observations

Justin E. Stopa[a,b,*], Alvaro Semedo[c], Joanna Staneva[d], Mikhail Dobrynin[e], Arno Behrens[d], Gil Lemos[f]

[a] *Department of Ocean Resources and Engineering, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, Honolulu, HI*
[b] *Univ. Brest, CNRS, IRD, Ifremer, Laboratoire d'Ocanographie Physique et Spatiale (LOPS), IUEM, 29280, Brest, France*
[c] *UNESCO-IHE Institute for Water Education, Deflt, The Netherlands*
[d] *Helmholtz-Zentrum Geesthacht Centre for Materials and Coastal Research, Geesthacht*
[e] *Institute of Oceanography, University of Hamburg, CEN, Germany*
[f] *Instituto Dom Luiz, University of Lisbon, Lisbon, Portugal*

Abstract

Global climate simulations do not capture the exact time history, making it difficult to directly compare them with observations. In this study we simulate the sampling of altimeter observations from a seven-member wind and wave climate ensemble. This allows us to assess the skill of the climate simulations, relative to satellite observations instead of the typical approach which uses reanalysis or hindcast datasets as reference. Out of the sampling methods tested, we find that a systematic sampling technique performs the best. We then apply systematic sampling to wind fields from EC-Earth and wave fields generated using the wave model (WAM) to replicate the changing sampling of the satellite observations. Next we then quantitatively assess the climate simulations and find that the probability density functions (PDFs) computed from the EC-Earth wind speed samples match the shape of the PDFs obtained from the altimeter observations. EC-Earth consistently underestimates the wind speed with respect to the altimeter observations. Contrary to the wind speed underestimation,

*Corresponding author: stopa@hawaii.edu

the wave simulations overestimate wave heights especially in the extra-tropics. The wind speed seasonality in EC-Earth is larger than the seasonality evaluated from altimeter observations while the opposite is true for the wave height seasonality; suggesting the wave physical parameterizations can be improved. We find that the wave height inter-annual variability of the modeled data is considerably less than the inter-annual variability evaluated from the altimeter observations; suggesting long-term climate variability is not well captured. Overall the wave ensemble captures the important features of the global wave climate. The methodology can be adapted to other climate simulations and observational datasets.

Keywords: , wave climate, COWCLIP, wind and wave projections, EC-Earth, altimeter observations, inter-annual variability, climate ensemble

1. Introduction

Global climate models (GCM) are tools to study future changes in climate and can potentially be used to mitigate impacts to humans and infrastructure. The most recent climate projections use ensembles, where the simulations of future conditions are generated using multiple climate models or different initial conditions, rather than a single climate simulation. Ensembles are used to explore and reduce the uncertainties inherent in the simulations that arise from the model's internal variability (Hawkins & Sutton, 2009; Knutti & Sedlacek, 2010; Rauser et al., 2015). Uncertainties in climate modeling inevitably occur due to errors in the physical parameterizations, missing physical parameterizations, or small scale processes not resolved due computational constrains (Stocker et al., 2013). These uncertainties have often been limiting factors in climate studies, particularly on regional scales (Falloon et al., 2014; Payne et al., 2015).

The Intergovernmental Panel on Climate Change (IPCC) recognized ocean waves as a significant driver of hazardous events in the coastal area (Stocker et al., 2013); thus, together with the expected sea level rise, waves will likely play an increasingly important role in dangerous high water levels (Hemer et al., 2013). Despite the important role of waves within the Earth system, there is still no coupled ocean-wave-atmosphere climate model system in operation. Therefore, global wave climate studies rely on the forcing from GCM projections, and are produced as separate simulations. Both statistical and dynamical methods have been used to simulate future wave climate. While statistical methods are less computationally demanding, they require a priori conditions and these are typically based on GCM projections (Perez et al., 2015; Camus et al., 2017). The dynamic approach uses wind speeds and sea-ice coverage from GCMs to drive a wave model and perform wave climate projections. The first global wave climate projections were developed under the auspices of the World Climate Research Program - Joint Technical Commission for Oceanography and Marine Meteorology (WRCP-JCOMM) Coordinated Ocean Wave Climate Projections (COWCLIP) project (Mori et al., 2010; Hemer et al., 2012; Semedo et al., 2013). These studies led to an ensemble of statistical and dynamical global wave climate projections and the ensemble was used to quantify future wave conditions (Hemer et al., 2013). Recent studies used multi- Coupled Model Intercomparison Project Phase 5 (CMIP5) GCM projections to produce dynamical wave climate projections (e.g. Hemer & Trenham, 2016).

Since the wave climate simulations are not time constrained, most studies compare different statistics such as seasonal or long-term averages, between the climate simulations and wave hindcasts (Hemer et al., 2013; Semedo et al., 2013; Hemer & Trenham, 2016). Some examples of wave hindcasts and reanalysis, are the National Center for Environmental Prediction (NCEP) climate forecast system (CFSR) (Chawla et al., 2013) or the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis (ERA-Interim) (Dee et al., 2011). The problem with using reanalysis and hindcast datasets as reference is that there are known errors associated with the driving wind fields (Stopa & Cheung, 2014a), and the physical parameterizations implemented in the wave model (Stopa et al., 2016). In addition, it can be difficult to assess the ability of the wave climate simulations to reproduce extreme waves since hindcasts tend to underestimate the largest sea states (Rascle & Ardhuin, 2013).

The goal of this study is two-fold. Our first objective is to develop a method to compare sparse observational datasets (altimeter observations in our case) to climate simulations from GCMs. Our second objective is to demonstrate the sampling method on an ensemble of CIMP5 EC-Earth wind simulations and associated wave simulations to determine the models' performance. Comparing sparse observations with GCMs is not straightforward because the quantity of the satellite observations changes in time and space as new platforms are activated and others are decommissioned. Therefore, sufficient efforts related to the first objective are taken to adequately sample the climate simulations to capture the statistical properties such as the mean, percentiles, probability density functions, and variance of the altimeter observations. For the second objective, we assess the performance of CIMP5 EC-Earth wind simulations and associated wave simulations in reproducing the wind and wave climate relative to altimeter observations. The EC-Earth and WAM simulation wind and wave ensembles, composed of seven members each, was evaluated with respect to in-situ observations and wave reanalysis datasets (Semedo et al., 2018). This group of simulations was arbitrarily chosen and other CIMP5 simulations are available (e.g. Hemer & Trenham, 2016). Our intent is to demonstrate the method as well as assess the wind speeds and wave heights from the EC-Earth and wave simulations against altimeter observations spatially and for a wide range of sea states.

The study is organized as follows. In section 2, we describe the altimeter observations and the EC-Earth and the wave climate simulations. In section 3, we test several sampling techniques to best capture the variance and sampling of the altimeter observations and in section 4, we compare the wind and wave simulations to the altimeter measurements for a characteristic period of 10 years (1996-2005). Here we emphasize on assessing the model-observation differences spatially and statistically through probability density functions, as well as assessing the extremes, seasonality, and inter-annual variability of the modelled and sampled wave properties. Our discussion and conclusions follow in section 5.

2. Datasets

2.1. EC-Earth ensemble

EC-Earth is a full physics coupled atmosphere-ocean-sea-ice earth system model, developed from the ECMWF Integrated Forecast System (IFS) operational seasonal forecast system (Hazeleger et al., 2011). Note that a wave model is not included in the system. The EC-Earth version 2.2, used here, is based on the ECMWF seasonal forecast system 3 (https://www.ecmwf.int/en/forecasts/documentation-and-support/evolution-ifs/cycles/implementation-seasonal-forecast-system). The atmospheric model in EC-Earth is the same as the ECMWF IFS cycle 31r. The EC-Earth atmospheric model uses a T159 (triangular truncation at wavenumber 159) grid with horizontal spectral resolution of 125 km and 62 vertical levels of a terrain-following mixed sigma-pressure hybrid coordinates, of which about 15 are within the planetary boundary layer. The lowest model level is at 30 m height, and the highest level is at 5 hPa. The ocean model in EC-Earth is the Nucleus for European Modeling of the Ocean (NEMO) (Vancoppenolle et al., 2009). NEMO uses a horizontal resolution of roughly one degree. The EC-Earth performance skills have been evaluated in several studies (e.g. Hazeleger et al., 2011).

The EC-Earth runs were provided by several research groups, as detailed in Table 1. Out of the seven EC-Earth runs used to force WAM, six are part of the regular CMIP5 EC-Earth ensemble. A seventh EC-Earth run (PC20-6), with an increased number of vertical levels, is also used. The EC-Earth runs were initialized between 1850 and 1855, spanning until 2005. Each EC-Earth simulation is independent and no bias corrections were applied.

The seven dynamical wave simulations were produced by forcing the 3rd generation wave model WAM (WAMDI-Group, 1988) with U10 components (East-West and North-South) every 6 hours and daily sea ice concentration from each of the CMIP5 EC-Earth runs. We use WAM cycle 4.5.3, an update of the WAM cycle 4, described in Gunther et al. (1992); Janssen (2008). The source function integration scheme made by Hersbach & Janssen (1999) and the model updates by Bidlot et al. (2007) are incorporated. The WAM simulations were performed on a regular global latitude-longitude grid, covering a latitude range of 78° N to 78° S and using a fixed spatial grid size of 1-degree. The spectral domain is discretized into 25 frequency bins in a geometrical progression with a common ratio of 1.1 that cover the range from 0.04177 to 0.41145 Hz. Wave directions are discretized into 15° bins. The 1-minute world gridded elevations/bathymetry (ETOPO1) data (Amante & Eakins, 2009), defines the water depths. WAM was run for each ensemble member for an (approximate) twentieth century time slice from 1970 to 2005, representing the present or historical climate. The ensemble members will hereafter be mentioned for convenience by PC20 (present climate 20th century) followed by their ensemble number (PC20-i, where i is 1 to 7; see Table 1), and the ensemble as PC20E. For convenience we will use the PC20 for both the EC-Earth and wave climate runs, since we will analyze both the wind speed (U10) and significant wave height (H_s). The WAM cycle 4.5.3 did not implement a sub-grid parametrization which causes the sea states in regions with islands smaller than the computational grid to be overestimated (e.g. Semedo et al., 2013, 2018).

2.2. Multi-platform altimeter dataset

The multi-platform altimeter product, abbreviated as ALT herein, was quality controlled and calibrated by Queffeulou & Croize-Fillon (2017). It was produced as part of the GlobWAVE project and is now extended to form the first version of the Sea State Climate Change initiative database (SeaStateCCI-V0). Here we use a 10 year period to assess the simulations from 1996 through 2005. This period is chosen because there are the largest number of observations during the "historical" EC-Earth simulations 1970-2005. For the period 1996-2005, there are 6 missions: ERS1 (1991-1996), ERS2 (1996-2011), ENVISAT (2002-2012), TOPEX (1993-2005), JASON1 (2002-2013), and GFO (2000-2007). Observations from each platform have been cross-calibrated between platforms and calibrated to moored buoys (Queffeulou & Croize-Fillon, 2017). The multi-mission dataset is expected to be consistent in time with very little deviation between platforms.

Since the 1 Hz altimeter measurements capture the the instantaneous and spatially localized estimate of H_s it is an unfair comparison with the time-space averaged outcome of the spectral wave model (e.g. Chawla et al., 2013; Stopa & Cheung, 2014a). Therefore, we average all observations within the 1-degree bin from the various satellite platforms that fall within a one hour window. The 1-degree bin matches the output of the WAM simulations. This one-hour average represents the satellite observation and is comparable to the time-scales resolved by the phase-averaged spectral wave model at 1-degree resolution (Chawla et al., 2013).

3. Methodology and assessment of sampling techniques

The ensemble uses a single forcing and a single wave model so it is expected that the intra-ensemble variability of PC20E is small. We create a wind speed ensemble and wave height ensemble using equal weighting of the seven simulations from EC-Earth and WAM respectively. In order to compare the performance of the dynamic climate simulations to the altimeter observations it is essential to capture the wind speed and wave height magnitudes and variance of the satellite observations which change as a function of time and space. An example time series is given in Figure 1a taken from a location in the North Atlantic $(20°W, 46°N)$. The location is denoted by the black "X" in Figure 1b. This point was chosen arbitrarily and used as an example to show the effects of satellite sampling. Notice that the number of satellite observations per month changes and there area a larger number of measurements from 2002 to 2005 when there were 4 concurrent missions. We are not concerned with the sparsity of satellite observations; we only consider the altimeter observations to be the reference dataset. The different number of satellite observations in single bins compared to the regularly time-spaced wave simulations impacts the statistics. Therefore, our goal is to properly sample the simulations so that the statistical properties such as the median and various percentiles match the ones from the altimeter observations.

Three fundamental sampling methods are tested: 1) simple random sampling 2) systematic sampling and 3) stratified sampling. Simple random sampling uses an equal weighting to select an event (without replacement) from the larger population. An event in our application is an individual time step. This method minimizes biases but it can be vulnerable to sampling errors especially in the tails of the distribution. Systematic sampling first orders the dataset and then chooses events at regular intervals based on a random starting position. For our application we sort the data by time. A disadvantage of systematic sampling is that it might not capture events that are periodic in nature, such as diurnal cycles. Stratified sampling is the process of first dividing all possible events into mutually exclusive subgroups or strata before sampling. In our case, the strata are months and the events are the individual time steps. For each month or strata, we use simple random sampling or systematic sampling to select the events. We tested a combination of these sampling techniques and summarize them with the following four cases. In each 1-degree bin, let M be the total number of time steps available from the GCM simulation and N be the number of altimeter observations where $M >> N$.

- **Case 1** Simple random sampling - N time steps are randomly chosen from the entire climate simulation time series of length M without replacement.

- **Case 2** Systematic sampling - data are selected every $dx = INT(M/N)$ where INT denotes the greatest integer (the floor function) by randomly choosing an initial index I_0 within the range $[1, dx]$ of the time series.

- **Case 3** Monthly stratified with simple random sampling version 1- here we select a variable number of events from each monthly PC20-i strata. The number of events selected directly corresponds to the number of satellite observations for the given month. For example, in Figure 1a there are 3 observations in January 2001. This means we randomly select 3 events from the model simulations in January 2001.

- **Case 4** Monthly stratified with simple random sampling version 2- here we select the same number of events each month. The number of events for a given location is defined by the average number of satellite observations for the complete time series. Referring to Figure 1a, there is an average of 6 samples per month for (1996-2005); thus, 6 events are selected from each month.

Assuming the complete 6-hour model simulation represents the "true" probability density function f, we can estimate the variance for a given percentile $P \in [0, 1]$ as:

$$\sigma^2 = \frac{1}{\left(f(x_p)\right)^2} \frac{P(1-P)}{N} \tag{1}$$

where x_p represents the given variable H_s or U10 at the given percentile. The total number of satellite observations in 1-degree bins are shown in Figure 1b for the period 1996-2005. The asymptotic variance (Equation 1) was assessed by Brown & Wolfe (1983) for smaller sample sizes (N<160) than our application. Therefore, we expect Equation 1 is an unbiased estimator of the variance since there are at least 350 samples in each 1-degree bin. We repeat the sampling procedures above for a number of trials. The final statistics are created by averaging all of the trials.

Without loss of generality we use the H_s from PC20-1 to assess both the magnitude and variance of the sub-sampled datasets. In Figure 2 we show results of the different sampling methods relative to statistics of the full time series using the time series in Figure 1. For a single trial (Figure 2a,e), all methods capture the percentile within 5% of the expected value, however the variance can have large discrepancies (10-15%). Notice the extremes of the distribution are not well captured especially for percentiles > 90%. If we use ten trials and then average the statistics we can reduce the differences in sampled magnitudes and variance of the full time series as shown in Figure 2b,f. However it is still difficult to capture the variance of the largest events.

Next we assess the ability of the sampling methods to capture the seasonality and use December-January-February (DJF) (Figure 2c,g) and June-July-August

(JJA) (Figure 2d,h) as representative seasons. In DJF (Figure 2c,g) when the waves are large, all cases match the percentiles (<10% difference). All sampling techniques perform similarly and the largest differences are still in the tails of the distribution. For percentiles less than 90-95 all sampling methods capture the magnitude and variance reasonably well. In JJA, when H_s is smaller, we find similar results. Case 3 sampling has the most pronounced deviations from the reference full 6-hourly time series with variances varying ±2% (Figure 2h). Overall, case 2 seems to perform the best because the percentiles are within 1% (Figure 2b) and the variance is well matched for the majority of the percentiles (Figure 2f).

In order to optimize our sampling technique various trials were tested using the 95th percentile (P95) PC20-1 H_s time series in the North Atlantic (not shown). After 25 trials, all of the sampling methods are less than 3% of the P95 observations (not shown). Cases 1 and 2 systematically converged with a lower number of trials than cases 3 and 4. After 10 trials only marginal improvements were observed; therefore, in the remainder of the study, we average 10 trials to represent the statistics from the sampled simulations.

The spatial distribution of the sampling effects for H_s P95 in Figure 3 are analyzed next. We compare both the ratio of H_s P95 ($P95_{sample}/P95_{all}$) and the ratio of H_s P95 variance (Var_{sample}/Var_{all}) for all cases. Cases 1 and 2 perform similarly and the ratios of H_s P95 are close to one, meaning nearly a perfect match. The spatial distribution of the variance of the sampled time series from case 1 and 2 is nearly uniform across the basins (Figure 3b,d). The variance of the time series using case 1 and case 2 sampling is typically 5% larger that the full 6-hourly time series. It is expected that the sampled dataset has a larger variance than the 6-hour time series because the full time series has approximately 10 times more data.

The case 3 ratios ($P95_{sample}/P95_{all}$) of H_s P95 in Figure 3e have a distinct spatial pattern with ±2% deviations from the P95 H_s full time series used as reference. In the North Atlantic, Western Pacific, and Northern Indian Oceans, case 3 sampling underestimates H_s P95; while, in the Indian, North Eastern

Pacific, and South Atlantic Oceans, case 3 sampling overestimates H_s P95. In case 3 sampling, we select the actual number of altimeter observations from the corresponding month of the simulated dataset. This means the number of events selected from the simulated time series in January 1996 corresponds to the actual number of altimeter observations in January 1996. Consequently, case 3 sampling favors months in 2003-2005 since there are more satellites in operation (see Figure 1). This sampling strategy introduces the largest spatial differences with respect to the full time series because the months that have more altimeter observations do not correspond to same months in the GCM forced wave simulations. For all cases the variance of the sampled dataset is 5-10% larger than variance of the full time series (shown in panels b, d, f, and h in Figure 3).

The magnitude of the H_s P95 ratios ($P95_{sample}/P95_{all}$) between the sampled and full time series using Case 4 sampling in Figure 3g is nearly one, similar to cases 1 and 2 (Figure 3a and c). There is a subtle tendency for case 4 to overestimate H_s P95 in the Northwest Pacific, North Atlantic ($> 30°N$), and in the Mediterranean (Figure 3g). The variance ratio of the case 4 sampling in Figure 3h is slightly larger than cases 1 and 2: 6-7% for case 4 sampling compared to 4-5% for cases 1 and 2 (Figure 3b and d).

The performance of each sampling procedure is summarized in Table 2 by comparing various H_s percentile and variance ratios ($\frac{sampled\ time\ series}{full\ time\ series}$) in percentages. We compare the percentiles: 5th, 50th (median), and 95th in different zonal regions: Northern Hemisphere (NH) $> 30°N$, Equatorial Region (EQ) $< 30°N/S$, and Southern Hemisphere (SH) $> 30°S$ and seasons (DJF and JJA). All sampling methods capture the overall variance very well and the errors are typically less than 0.5%. The results in Table 2 reflect similar features seen in Figures 2-3 and are summarized as follows:

- All sampling methods overestimate the small percentiles and the variance is often $> 1\%$.

- The medians are very well matched using any of the sampling methods.

- Case 3 sampling introduces spatial discrepancies.

- Typically all sampled H_s P95 match the full-time series and H_s P95 ratios are $<< 0.1\%$.

- The sampled datasets have larger variance than the full time 6-hourly series.

Table 2 shows the variance can be considerably larger using the sampled time series for particular seasons (DJF and JJA) compared to the full time series, especially for P95. We consistently find case 2 (systematic sampling) performs the best; therefore, this method is implemented to sample the climate simulations to match the satellite observations. We continue to use the average of the 10 trials as a representative sample of the simulated climate wave data since using only one trial can have large differences of magnitude and variability relative to the observations (see Figure 2 a,e).

4. Assessment of the wind and wave climate simulations

In this section we assess the sampled U10 and H_s datasets (PC20-1 to PC20-7) relative to the altimeter observations by analyzing the spatial errors, the probability density functions (PDFs), seasonality, inter-annual variability, and large sea states.

4.1. Spatial features and statistical properties

First we highlight the spatial differences and variability between the ensemble and observations for various statistics such as the median (P50) and upper percentile (P95) wind speeds and wave heights for the entire 10-year period. Figure 4 shows the comparison between PC20E and ALT for both U10 and H_s at P50. The U10 P50 residuals (PC20E-ALT) in Figure 4b show that EC-Earth underestimates U10 by 1-2 ms^{-1} across the majority of the ocean. Near the Equator, there is a strong underestimation of U10 that is persistent for all simulations (PC20-1 to PC20-7). However, we should note that near the

Equator nadir-looking altimeters are not the best source of wind and/or wave data since the calm ocean surface coupled with weak winds can have nearly a specular reflection, thus producing erroneous high wind speeds (Elfouhaily et al., 1998). In addition, there are impacts from the sea state which distort U10 when only the radar cross section (one-parameter approximation) is used to estimate U10 (Gourrion et al., 2002). Consequently, the U10 from altimeters is often higher than the reference buoy wind speeds especially in low wind regions (Young et al., 2017). Near the ice edge in the Southern Ocean, there is a typical underestimation of 1-2 ms^{-1}. Otherwise the ensemble data in the extra-tropics ($30 - 50°N/S$), which are important wave generation regions, agree with altimeter observations ($|U10_{ALT} - U10_{PC20-E}| < 0.5$ ms^{-1}). The intra-ensemble variability (from the seven simulations) in Figure 4c is low (0.2 ms^{-1}) with respect to the U10 differences (typically 1 ms^{-1}) which equates to $< 20\%$ of the variability of the ensemble. Regions in the SH extra-tropics, Eastern-Equatorial Indian Ocean, and the trade wind regions of the NH and SH have the largest intra-ensemble variations.

In Figure 4d,e,f we show the P50 H_s comparison between PC20E and ALT. Overall the H_s P50 of PC20E matches the observations within ± 0.25 m. Across the majority of the ocean, the ensemble overestimates H_s. Otherwise, there are only select regions such as the NW Atlantic, Mediterranean, Gulf of Mexico, and Western Pacific where PC20E underestimates H_s. This might be partially related to discrepancies in U10. Altimeters tend to overestimate wave heights in low sea states (Sepulveda et al., 2015; Kudryavtseva & Soomere, 2017). So, wave intensity in regions like the Mediterranean and Gulf of Mexico, which typically have small wave heights, might be overestimated by the altimeters. Consequently, in low sea states it is difficult to assess the simulations when using the altimeters as reference. In the Pacific trade wind regions, the ensemble systematically overestimates H_s by > 0.5 m with a global maximum difference (P50 H_s PC20E-ALT) coinciding near Micronesia ($130°W, 20°S$). Some of these features are related to unresolved islands smaller than the 1° resolution; similar to features seen in (Semedo et al., 2013). The intra-ensemble variability is low and the H_s deviations are less than 10 cm. So near the noted large differences in the trade wind regions, the intra-ensemble variability is $< 4\%$ of the residuals. The intra-ensemble variability in the Southern Ocean is 3 to 10 cm which is 12 to 25% of the typical 25 cm H_s P50 residual. So the ensemble reduces some of the uncertainty in the SH compared to using only one simulation.

In Figure 5 we show the corresponding plots for the U10 and H_s for the upper percentiles (P95). The U10 P95 residuals in Figure 5b have nearly the same spatial structure as U10 P50 residuals in Figure 4b. The most obvious difference is that the U10 P95 residuals have enhanced underestimation near the western boundaries in the tropics ($5-30°$N/S) relative to the U10 P50 residuals. These regions are affected by tropical cyclones and the ensemble underestimates U10 P95 by 1-2 ms^{-1}. The intra-ensemble variability in Figure 5c is largest in the NH extra-tropics ($30-60°$N). In the NH extra-tropics, there are EC-Earth U10 P95 deviations of 0.2-0.3 ms^{-1}. This is 40-60% of the average value of the residuals (-0.5 ms^{-1}) in Figure 5b. Therefore in these regions (i.e. the gold-colored areas in Figure 5c), the use of the ensembles improves the performance.

The bottom panels of Figure 5 show the corresponding plots for the H_s at P95. The H_s P95 residuals (PC20E-ALT) in Figure 5e are positive in the trade wind regions of the Pacific, meaning that PC20E overestimates H_s P95. The spatial pattern of the H_s P95 residuals in Figure 5e is similar to H_s P50 residual in Figure 4e. In the NH and SH extra-tropics ($30-60°$N/S), the ensemble overestimates H_s at P95 by at least 0.5 m and in some areas the H_s P95 residuals exceed 0.75 m. For example, in the North Pacific near the Aleutians and South Pacific near the ice edge, the H_s P95 is much higher (exceeds 0.75 m) than the observations. EC-Earth underestimates U10 P95 in the Western portion of the Pacific and Atlantic, and likely contributes to a portion of the underestimation of H_s. The underestimation in the Western portion of the Pacific and Atlantic is compounded by the fact that spectral wave models underestimate H_s in rapidly changing "short-fetch" conditions (e.g. Ardhuin et al., 2010). Similar to U10 P95, the intra-ensemble variability helps to reduce the H_s P95 differences mostly in the NH extra-tropics with standard deviations of 0.2 m. This is approximately

20% of the common H_s P95 residual of 1 m.

Next we compare the PDFs and the quantiles to give further insights on the performance of the climate simulations. Figure 6 shows the U10 probability distributions and quantile-quantile (QQ) plots. The shape of the PDFs calculated from the simulations match the ones obtained from observations well, but the PDFs are mis-aligned. In particular, when U10$<(5-10)$ ms^{-1} the probabilities obtained from the simulations are larger than the probabilities obtained from the altimeter observations. When U10$>(5-10)$ ms^{-1} the probabilities obtained from the simulations are lower than those of the altimeters. In these plots each of the seven simulations is analyzed separately and its results plotted with a different color. The PDFs obtained from the simulations are nearly the same and only subtle differences are distinguishable in this representation. The QQ plots show EC-Earth underestimates U10 uniformly across all wind speeds by approximately 0.75 ms^{-1}. Notice that in each region, the QQ plots are similar. Besides the noted spatial differences discussed in the previous section, the simulations perform equally well in the different latitude bands.

We provide the corresponding PDFs for the H_s in Figure 7. The shape of PDFs obtained from the wave simulations is similar to the PDFs obtained from altimeter observations but they are mis-aligned. When $H_s < (2-3)$ m the probabilities obtained from the simulations are smaller than the probabilities obtained from the altimeter observations. When $H_s > (2-3)$ m the probabilities obtained from the simulations are larger than those of the altimeter observations. The wave simulations favor the mid-range ($2 < H_s < 3$ m) more than the altimeter observations, as shown by the reduced width of the PDFs obtained from the wave simulations. We can see some distinction (2% difference) between the PC20-i members near the median. This effect is most evident in the SH. In the SH, there are the largest differences between the PDFs obtained from the wave simulations and the altimeter observations. Here the PDFs obtained from the simulations have higher occurrence of sea states with $2 < H_s < 4$ m compared to the PDFs obtained from the altimeter observations. The QQ plots show that the simulations of the lower percentiles (< median) are often 0.25 m larger than the altimeter observations. For the higher percentiles (such as >P95) the simulations overestimate H_s by 0.25-0.5 m relative to the observations. The QQ H_s results here in Figure 7 are consistent with the buoy comparisons of Semedo et al. (2018) (their Figure 10) and show that H_s is typically overestimated. However, in our analysis the comparisons are global and extend to wave heights of 8 m; thus establishing the validity of the wave simulations to larger sea states.

Notice that the global QQ H_s differences are reflective of the patterns observed in the SH; since in the NH, the PDFs obtained from the wave simulations match those of the observations reasonably well. Near the Equator (EQ: $\in$ 25°N/S) in the lower percentiles (<P50), the wave heights are overestimated. The largest contribution comes from the lower latitudes ($< 15°$) as shown in Figures 4e and 5e. These areas are dominated by swell and the underestimation of the swell dissipation in WAM might be contributing to the discrepancies. In the SH, the differences in the PDFs obtained from the wave simulations and altimeter observations are the largest. The H_s probabilities obtained from the simulations are lower than those of the altimeters for low sea states ($H_s < 2$ m). For sea states with $2 < H_s < 5$ m, the probabilities obtained from the simulations are higher than those of the altimeters. The H_s QQ plots show the thresholds for a given wave height quantile is overestimated in low seas (<P50) and high seas (>P95). We observe the largest variability between the ensemble members in the upper percentiles. In summary, U10 is underestimated by EC-Earth and H_s is overestimated by the wave simulations. This suggests the wave model physical parameterizations are causing the differences and not necessarily the forcing wind.

4.2. Seasonality

To assess the ability of the PC20-E members to capture seasonality we use a metric called the mean annual variability (MAV) (Stopa et al., 2013). It is defined as the average of the annual standard deviation normalized by the annual average

$$MAV = \overline{\left(\frac{\sigma_i}{\overline{x_i}}\right)} \tag{2}$$

where index i refers to the year, σ is the standard deviation, and the overbar denotes average. We compare the MAV between the ensemble average (PC20-E) and ALT in Figure 8. Note that the altimeter patterns of the U10 and H_s are provided as reference in Figure 8a,c. The spatial pattern and magnitudes of the MAV computed from the altimeter observations is similar to that of the MAV computed from the CFSR wave hindcast of Chawla et al. (2013) and presented in (Stopa et al., 2013) (their Figure 6). U10 from EC-Earth has more seasonal variability than the altimeters (see Figure 8b). In particular, PC20E has a larger seasonality in the Southern Ocean extra-tropics. This might partially be influenced by the ice coverage. Near the Equator PC20E has a larger MAV North of the Inter Tropical Convergence Zone (ITCZ) and a smaller MAV South of the ITCZ relative to the observations. However, it is expected the altimeter U_{10} is of poorer quality near the ITCZ due to sea state impacts when the wind is calm and specular reflection is strong (Gourrion et al., 2002). The H_s comparison in Figure 8d shows the ensemble typically underestimates the wave seasonality by as much as 15% but typically 4-8%. It is possible that there are missing physical parameterizations within the wave model or the physical parameterizations are not responding correctly to the wind input; since we observe higher MAV in U10 and lower MAV in the wave field.

The seasonality within the ensemble is further analyzed in Figures 9 and 10. Here we present only results from the P95 since the spatial patterns for other percentiles and the average were nearly the same. In DJF (Figure 9b), the U10 P95 differences between PC20-E and ALT are largest in the SH trade wind regions in the Pacific and in the NW Atlantic. Otherwise the differences are less than 1 ms^{-1}. EC-Earth is overestimating U10 P95 in the SH extra-tropics but usually less than 0.5 ms^{-1}. The intra-ensemble variability in Figure 9c is largest in the NH, SH extra-tropics, and near the EQ in the Indian Ocean. The corresponding H_s P95 residuals of PC20E-ALT in Figure 9e are similar to the U10 P95 residuals with the zonal pattern: PC20E overestimates in $40-60°N/S$ and PC20E underestimates $20-30°N/S$. Near the EQ, the U10 and H_s P95 residuals have opposite signs with an overestimation in H_s and an underestimation of U10.

In JJA (Figure 10b), EC-Earth underestimates U10 P95 relative to ALT across the majority of the global ocean. There are some exceptions where EC-Earth overestimates U10 P95 such as regions in the North Pacific, NW Atlantic, and near Eastern Africa in the NH. The U10 and H_s from the wind and wave ensembles are not capturing the tropical cyclones that are more common this time of the year especially in the Western Pacific (Figure 10b,e). The intra-ensemble variability is largest in the SH for both U10 and H_s P95 (Figure 10c,f). If a single member was used, the differences in the SH might be larger than PC20-E shown in Figure 10b where the average is -1 ms^{-1}. The ensemble produces a better estimation of the seasonality especially in the SH. The spatial pattern of H_s P95 residuals in Figure 10e do not match the U10 P95 residuals except for the region in the Western Pacific. PC20-E overestimates the H_s P95 relative to ALT on average by 0.35 m with some regions in the SH and NH extra-tropics exceeding 0.5 m. In both the NH and SH extra-tropics the intra-ensemble variability is largest and $>$ 0.25 m (Figure 10f). In the wave generation regions of the extra-tropics, PC20E underestimates U10 P95 while PC20E underestimates H_s P95.

Some of the other H_s discrepancies are due to land mask used in the WAM set up, which is different from the one used in ERA-Interim. Additionally the WAM version (v4.5.3) used here is known to dissipate swell improperly in the low latitudes, contributing to an overestimation of the wave heights there (Semedo et al., 2013). These differences such as higher waves in the tropics, around Polynesia, Micronesia, the Maldives, and near the Aleutians Islands might also occur due to unresolved sub-grid scale bathymetry. The intra-ensemble variability in U10 and H_s is largest in the NH and SH extra-tropics.

4.3. Inter-annual variability

Lastly, we assess the ability of the ensemble to capture inter-annual variability (IAV) over this 10-year period (Stopa et al., 2013). The IAV is defined as the standard deviation of the annual averages normalized by the overall average:

$$IAV = \frac{\sigma_{x_i}}{\overline{x}}. \quad (3)$$

We compare the IAV between PC20-E and ALT in Figure 11. The spatial patterns of the altimeter observations in Figure 11a,c qualitatively look similar to the IAV of CFSR presented in Stopa et al. (2013) (their Figure 7). The U10 IAV maxima of the altimeter observations (Figure 11a) are located in the Eastern and Western Equatorial Pacific. The H_s IAV maxima of altimeter observations (Figure 11c) are located in the Southern Ocean near Chile and in the Western Pacific (120°E,15°N). The U10 IAV residuals between PC20E and ALT are largest near the Equator especially in the Pacific which might be related to not properly capturing the El Nino Southern Oscillation (ENSO) which is know to be the dominant mode of inter-annual variability in this region (e.g. Stopa & Cheung, 2014b). Otherwise the EC-Earth U10 ensemble has differences less than 1% across 84% of the global ocean (Figure 11b) suggesting EC-Earth U10 captures a large amount of the U10 IAV. The H_s IAV difference in Figure 11d shows that the IAV for PC20-E is considerably less than the IAV of the altimeter observations. These large differences between PC20-E and ALT mean the IAV of the wave field is not well captured in PC20-E and is typically much smaller than the observations at least over this period of 10 years.

5. Discussion and conclusion

The U10 and H_s climate simulations were sampled such that their statistical properties such as average, variance, and percentiles replicated those of the altimeter observations. We analyzed both the magnitude and the variance of several sampling techniques at various percentiles. We found that systematic sampling (case 2) performed better than the other tested sampling methods.

Properly sampling climate simulations that do not capture the exact time history is particularly important when the reference observations are sparse and/or the number of observations changes in time and space. Our methodology can be adapted to other climate simulation datasets.

We systematically analyzed the skill of the ensemble in reproducing the wind speeds (U10) and the resulting wave heights (H_s) relative to the altimeter observations. EC-Earth underestimates the magnitude of U10 uniformly across all percentiles. The PDFs obtained from the U10 of EC-Earth and those obtained from the altimeter observations are very similiar suggesting that EC-Earth is a satisfactory predictor of wind speeds globally. Even though PC20E underestimates U10, the wave heights are overestimated. This suggests the implementation of WAM is not properly calibrated for the EC-Earth wind field and it is possible to correct this bias by reducing the wind wave growth parameter (β_{max}) in the parametrization of Janssen (1991) as shown by (Stopa, 2018). In the NH, the PDFs computed from U10 and H_s of both the simulations and altimeter observations are similiar; however in the SH, the PDFs are different. Therefore, the performance of the simulations in the NH is better than the SH. The QQ plots also support this point. The global discrepancies in the PDFs and QQ plots strongly reflect the discrepancies of the SH; stressing the importance of future efforts to better simulate the SH. The almost identical match of the PDFs for the 7 simulations limits the possibilities of the ensemble to improve forecasts or hindcasts since the ensemble variance is much less than typical simulation-observation error variances. For example, the H_s standard deviations of wave hindcasts errors at buoys typically ranged from $\pm 40\%$ or 0.3 to 0.8 m as presented by Stopa & Cheung (2014a) (their Table 2). The H_s standard deviation of PC20-E is generally small and less than 0.06 (0.2) m at P50 (P95).

We find PC20-E overestimates H_s in the tropics namely in the Pacific Ocean. This region has an abundance of swell (Semedo et al., 2011) and WAM is most likely underestimating the swell decay. In addition, large H_s discrepancies coincide with island chains in the Pacific and are due to the treatment of sub-grid features not resolved by the model grid resolution. Regions affected by tropical cyclones most notably in the Western Pacific are not well captured by the climate simulations and we observe a severe underestimation of H_s at P95. The use of the ensemble has a minimal effect on improving the predictability in this case. Now that the wave simulations are sampled like the satellite measurements, it is possible to develop a bias correction for the wave simulations and it is topic for future work. Notice that all of the sampling techniques introduce errors of less than 2% for H_s at P95 (see Figure 3) while the H_s model-to-simulation discrepancies at P95 are typically on the order of 10-25% (Figure 5). So we expect that our results are robust and there is minimal impact from the sampling technique applied.

The seasonality is reasonably captured by the U10 and H_s ensembles. We find some differences. For example, EC-Earth overestimates the U10 seasonality while the wave ensemble underestimates the H_s seasonality. This seasonal mismatch was found in other datasets. For example, seasonal residuals between a CFSR wave hindcast and altimeter observationss (both U10 and H_s) were observed in Chawla et al. (2013); Stopa & Cheung (2014a). This suggests the physical parameterizations in spectral wave models like WAVEWATCH and WAM have missing physical processes or the existing parameterizations can be improved to better capture the atmospheric response in both the strong and weak seasons (such as temperature differences or water density differences). We speculate that the current physical parameterizations in spectral wave models have the tendency to underestimate both growth and dissipation which might contribute to a portion of the H_s seasonality residuals. The ensemble improves the prediction of the seasons especially in the Southern Ocean. Otherwise the typical intra-ensemble variability is less than or approximately 10-30% of the PC20E-ALT residuals.

The U10 from EC-Earth captures the important features of the inter-annual variability. On the other hand, the GCM wave simulations have lower inter-annual variability suggesting the time series of wave simulations forced by EC-Earth have a much smoother time series relative to the altimeter observations.

Our comparison of the inter-annual variability is a challenging test for the wave climate simulations. One possible reason why we have such large differences in the IAV between the simulations and satellite observations could be because we use a 10-year period. A longer time series might capture more of the long-term variability. Since the GCM forced wave simulations have difficultly in reproducing the IAV, caution should be taken when analyzing the inter-annual variability of future climate scenarios. Improving the ability of the wave climate simulations to reproduce the inter-annual variability is an opportunity for future efforts.

Previous works use wave reanalysis or wave hindcasts to assess GCM-forced wave simulations. Here we take a novel approach and we use altimeter observations as reference. This is important because the altimeter database is expected to better represent the large sea states and are not subjected to missing or improper wave parameterizations as in models. It also stresses the importance of having an accurate and quality-controlled altimeter database and is currently being re-assessed by the European Space Agency's Sea State Climate Change Initiative. Using the altimeter observations to assess the GCM simulations especially at large sea states (H_s P95) is certainly a benefit of applying the method. In this study, we provide more spatial details of the simulation errors and validate the simulations across a wider range of sea states compared to Semedo et al. (2018) who used reanalysis datasets and in-situ buoys as reference datasets. Future assessments of the historical wave simulations either dynamical or statistical could use a similar methodology and compare to sparse observational datasets like our example of using the altimeter observations as reference. Overall the EC-Earth simulations and associated wave simulations capture the essential features of the climate. Since we understand the discrepancies between the simulations and satellite observations, it is now possible to interpret the wave data for the future simulations which extend until the end of the 22nd century.

6. Acknowledgements

This work has been done under the auspices of the JCOMM COWCLIP (Coordinated Ocean Wave Climate projections) project. Alvaro Semedo has been supported by the project SOLAR (PTDC/GEOMET/7078/2014), financed by Portuguese Foundation for Science and Technology (FCT). Gil Lemos is supported by the Earth Systems Doctoral School, FCT project UID/GEO/50019/2013, University of Lisbon.

References

Amante, C., & Eakins, B. W. (2009). *ETOPO1 1 Arc-minute Global Relief Model: Procedures, Data sources and Analysis*. Technical Report NGDC-24 NOAA Technical Memorandum NESDIS.

Ardhuin, F., Rogers, E., Babanin, A. V., Filipot, J.-F., Magne, R., Roland, A., van der Westhuysen, A., Queffeulou, P., Lefevre, J.-M., Aouf, L., & et al. (2010). Semiempirical dissipation source functions for ocean waves. part i: Definition, calibration, and validation. *Journal of Physical Oceanography*, *40*, 19171941. URL: `http://dx.doi.org/10.1175/2010JPO4324.1`. doi:`10.1175/2010jpo4324.1`.

Bidlot, J., Janssen, P. A. E. M., & Abdalla, S. (2007). *A revised formulation for ocean wave dissipation and its model impact*. Technical Report Memorandum 509 ECMWF.

Brown, M. M., & Wolfe, R. B. (1983). Estimation of the variance of percentile estimates. *Computational Statistics and Data Analysis*, *1*, 167 – 174. doi:`https://doi.org/10.1016/0167-9473(83)90088-9`.

Camus, P., Losada, I. J., Izaguirre, C., Espejo, A., Menéndez, M., & Pérez, J. (2017). Statistical wave climate projections for coastal impact assessments. *Earth's Future*, *5*, 918–933. doi:`10.1002/2017ef000609`.

Chawla, A., Spindler, D. M., & Tolman, H. L. (2013). Validation of a thirty year wave hindcast using the climate forecast system reanalysis winds. *Ocean Modelling*, *70*, 189206. URL: http://dx.doi.org/10.1016/j.ocemod.2012.07.005. doi:10.1016/j.ocemod.2012.07.005.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., & et al. (2011). The era-interim reanalysis: configuration and performance of the data assimilation system. *Q.J.R. Meteorol. Soc.*, *137*, 553597. URL: http://dx.doi.org/10.1002/qj.828. doi:10.1002/qj.828.

Elfouhaily, T., Vandemark, D., Gourrion, J., & Chapron, B. (1998). Estimation of wind stress using dual-frequency TOPEX data. *Journal of Geophysical Research: Oceans*, *103*, 25101–25108. doi:10.1029/98jc00193.

Falloon, P., Challinor, A., Dessai, S., Hoang, L., Johnson, J., & Koehler, A.-K. (2014). Ensembles and uncertainty in climate change impacts. *Frontiers in Environmental Science*, *2*. doi:10.3389/fenvs.2014.00033.

Gourrion, J., Vandemark, D., Bailey, S., Chapron, B., Gommenginger, G. P., Challenor, P. G., & Srokosz, M. A. (2002). A two-parmeter wind speed algorithm for ku-band altimeters. *Journal of Atmospheric Ocean Technology*, *19*, 2030–2048. doi:10.1175/1520-0426(2002)019$<$2030:ATPWSA$>$2.0.CO;2.

Gunther, H., Hasselmann, S., & Janssen, P. (1992). *The WAM model Cycle 4*. Tecnhical Report Report No. 4 Institute of Hydrophysics, GKSS Research Centre.

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.*, *90*, 10951107. doi:10.1175/2009BAMS2607.1.

Hazeleger, W., Wang, X., Severijns, C., Ştefănescu, S., Bintanja, R., Sterl, A., Wyser, K., Semmler, T., Yang, S., Hurk, B., Noije, T., Linden, E., & Wiel,

K. (2011). Ec-earth v2.2: description and validation of a new seamless earth system prediction model. *Climate Dynamics*, *39(2611)*, 1–19. doi:10.1007/s00382-011-1228-5.

Hemer, M. A., Fan, Y., Mori, N., Semedo, A., & Wang, X. L. (2013). Projected changes in wave climate from a multi-model ensemble. *Nature Climate Change*, *3*, 471476. URL: http://dx.doi.org/10.1038/nclimate1791. doi:10.1038/nclimate1791.

Hemer, M. A., & Trenham, C. E. (2016). Evaluation of a cmip5 derived dynamical global wind wave climate model ensemble. *Ocean Modelling*, *103*, 190–203. doi:10.1016/j.ocemod.2015.10.009.

Hemer, M. A., Wang, X. L., Weisse, R., & Swail, V. R. (2012). Advancing wind-waves climate science: The cowclip project. *Bulletin of the American Meteorological Society*, *93(6)*, 791–796. doi:10.1175/BAMS-D-11-00184.1.

Hersbach, H., & Janssen, P. A. E. M. (1999). Improvement of the short-fetch behavior in the wave ocean model (WAM). *Journal of Atmospheric and Oceanic Technology*, *16*, 884–892. doi:10.1175/1520-0426(1999)016$<$0884:iotsfb$>$2.0.co;2.

Janssen, P. A. (2008). Progress in ocean wave forecasting. *Journal of Computational Physics*, *227*, 3572–3594. doi:10.1016/j.jcp.2007.04.029.

Janssen, P. A. E. M. (1991). Quasi-linear theory of wind wave generation applied to wave forecasting. *Journal of Physical Oceanography*, *21*, 1631–1642.

Knutti, R., & Sedlacek, J. (2010). Robustness and uncertainties in the new cmip5 climate model projections. *Nature Climate Change*, *3*, 369373. doi:10.1038/nclimate1716.

Kudryavtseva, N. A., & Soomere, T. (2017). Satellite altimetry reveals spatial patterns of variations in the baltic sea wave climate. *Earth Systems Dynamics*, *8*, 697–706. doi:10.5194/esd-8-697-2017.

Mori, N., Yasuda, T., Mase, H., Tom, T., & Oku, Y. (2010). Projection of extreme wave climate change under the global warming. *Hydrological Research Letters*, *4*, 15–19. doi:`10.3178/hrl.4.15`.

Payne, M. R., Barange, M., Cheung, W. W. L., MacKenzie, B. R., Batchelder, H. P., Cormon, X., Eddy, T. D., Fernandes, J. A., Hollowed, A. B., Jones, M. C., Link, J. S., Neubauer, P., Ortiz, I., Queirós, A. M., & Paula, J. R. (2015). Uncertainties in projecting climate-change impacts in marine ecosystems. *ICES Journal of Marine Science: Journal du Conseil*, *73*, 1272–1282. doi:`10.1093/icesjms/fsv231`.

Perez, J., Menendez, M., Camus, P., Mendez, F. J., & Losada, I. J. (2015). Statistical multi-model climate projections of surface ocean waves in europe. *Ocean Modelling*, *96*, 161–170. doi:`10.1016/j.ocemod.2015.06.001`.

Queffeulou, P., & Croize-Fillon, D. (2017). *Global altimeter SWH data set*. Technical Report 11.1 IFREMER/CERSAT. URL: `ftp://ftp.ifremer.fr/ifremer/cersat/products/swath/altimeters/waves/documentation/altimeter_wave_merge__11.1.pdf`.

Rascle, N., & Ardhuin, F. (2013). A global wave parameter database for geophysical applications. part 2: Model validation with improved source term parameterization. *Ocean Modelling*, *70*, 174188. URL: `http://dx.doi.org/10.1016/j.ocemod.2012.12.001`. doi:`10.1016/j.ocemod.2012.12.001`.

Rauser, F., Alqadi, M., Arowolo, S., Baker, N., Behrens, E., Bedard, J., Dogulu, N., Domingues, L. G., Frassoni, A., Keller, J., Kirkpatrick, S., Langendijk, G., Mohammad, S., Mirsafa, M., Naumann, A. K., Osman, M., Reed, K., Greilinger, M., Schemann, V., Singh, A., Sonntag, S., Tummon, F., Nnamdi, D. V., Villafuerte, M. Q., Walawender, J. P., & Zaroug, M. (2015). Earth system science frontiers - an ecs perspective. *Bulletin of the American Meteorological Society*, *0*. doi:`10.1175/BAMS-D-16-0025.1`.

Semedo, A., Dobrynin, M., Lemos, G., Behrens, A., Staneva, J., de Vries, H., Sterl, A., Bidlot, J.-R., Miranda, P., & Murawski, J. (2018). CMIP5-derived single-forcing, single-model, and single-scenario wind-wave climate ensemble: Configuration and performance evaluation. *Journal of Marine Science and Engineering*, *6*, 90. doi:10.3390/jmse6030090.

Semedo, A., Suelj, K., Rutgersson, A., & Sterl, A. (2011). A global view on the wind sea and swell climate and variability from era-40. *J. Climate*, *24*, 14611479. URL: http://dx.doi.org/10.1175/2010JCLI3718.1. doi:10.1175/2010jcli3718.1.

Semedo, A., Weisse, R., Behrens, A., Sterl, A., Bengtsson, L., & Gunther, H. (2013). Projection of global wave climate change toward the end of the twenty-first century. *Journal of Climate*, *26 (21)*, 8269–8288. doi:10.1175/JCLI-D-12-00658.1.

Sepulveda, H. H., Queffeulou, P., & Ardhuin, F. (2015). Assessment of SARal/altika wave height measurements relative to buoy, jason-2, and cryosat-2 data. *Marine Geodesy*, *38*, 449465. URL: http://dx.doi.org/10.1080/01490419.2014.1000470. doi:10.1080/01490419.2014.1000470.

Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., & Midgley, P. (Eds.) (2013). *IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA. Cambridge University Press.

Stopa, J. E. (2018). Wind forcing calibration and wave hindcast comparison using multiple reanalysis and merged satellite wind datasets. *Ocean Modelling*, *127*, 55 – 69. URL: http://www.sciencedirect.com/science/article/pii/S1463500318301458. doi:https://doi.org/10.1016/j.ocemod.2018.04.008.

Stopa, J. E., Ardhuin, F., Babanin, A., & Zieger, S. (2016). Comparison and validation of physical wave parameterizations in spectral wave models. *Ocean*

Modelling, *103*, 217. URL: http://dx.doi.org/10.1016/j.ocemod.2015.09.003. doi:10.1016/j.ocemod.2015.09.003.

Stopa, J. E., & Cheung, K. F. (2014a). Intercomparison of wind and wave data from the ecmwf reanalysis interim and the ncep climate forecast system reanalysis. *Ocean Modelling*, *75*, 6583. URL: http://dx.doi.org/10.1016/j.ocemod.2013.12.006. doi:10.1016/j.ocemod.2013.12.006.

Stopa, J. E., & Cheung, K. F. (2014b). Periodicity and patterns of ocean wind and wave climate. *Journal of Geophysical Research: Oceans*, *119*, 5563–5584. doi:10.1002/2013jc009729.

Stopa, J. E., Cheung, K. F., Tolman, H. L., & Chawla, A. (2013). Patterns and cycles in the climate forecast system reanalysis wind and wave data. *Ocean Modelling*, *70*, 207220. URL: http://dx.doi.org/10.1016/j.ocemod.2012.10.005. doi:10.1016/j.ocemod.2012.10.005.

Vancoppenolle, M., Fichefet, T., Goosse, H., Bouillon, S., Madec, G., & Maqueda, M. A. M. (2009). Simulating the mass balance and salinity of arctic and antarctic sea ice. 1. model description and validation. *Ocean Modelling*, *27*, 33–53. doi:10.1016/j.ocemod.2008.10.005.

WAMDI-Group (1988). The wam model - a third generation ocean wave prediction model. *Journal of Physical Oceanography*, *12*, 1775–1810. doi:10.1175/1520-0485(1988)018$<$1775:TWMTGO$>$2.0.CO\%3B2.

Young, I. R., Sanina, E., & Babanin, A. V. (2017). Calibration and cross validation of a global wind and wave database of altimeter, radiometer, and scatterometer measurements. *Journal of Atmospheric and Oceanic Technology*, *34*, 1285–1306. doi:10.1175/jtech-d-16-0145.1.

Table 1: Ensemble member details

Ensemble member	CMIP5 experiment	Data Provider
PC20-1	r1i1p1	University of Lisbon
PC20-2	r3i1p1	Danish Meteorological Institute
PC20-3	r1i1p1	Danish Meteorological Institute
PC20-4	r1i1p1	Swedish Meteorological and Hydrological Institute
PC20-5	r2i1p1	Swedish Meteorological and Hydrological Institute
PC20-6	r2i1p1	Danish Meteorological Institute
PC20-7	r3i1p1	Swedish Meteorological and Hydrological Institute

Table 2: H_s statistics for various conditions given as a percentage $\left(\frac{Sample\ time\ series}{Full\ time\ series} - 1\right) \times 100$ for the percentile and variance (given in parenthesis). These values represent the average of ten independent sub-samples.

Selection	**Case**	**All**	**Prc 5%**	**Prc 50%**	**Prc 95%**
Global	C1	(-0.0026)	0.0441(0.8431)	0.0196(-0.4224)	0.0083(5.6070)
	C2	(0.0684)	0.0214(0.6426)	0.0120(-0.2947)	0.0132(4.9086)
	C3	(0.1557)	0.1159(1.6766)	0.2265(-0.3319)	0.0905(5.4032)
	C4	(0.0123)	0.0519(0.7970)	0.0250(-0.4021)	0.0078(5.8263)
NH	C1	(-0.0617)	0.0594(0.6192)	0.0408(-0.3340)	0.0023(5.4733)
	C2	(0.1648)	0.0207(0.5582)	0.0265(-0.1387)	0.0448(4.8654)
	C3	(-0.0698)	0.3133(1.0741)	0.3216(-0.0928)	-0.0084(4.4369)
	C4	(0.2316)	0.1322(1.0685)	0.1989(-0.0783)	0.1415(5.8226)
EQ	C1	(0.0105)	0.0429(0.9913)	0.0203(-0.4609)	0.0070(5.8203)
	C2	(0.0668)	0.0250(0.6745)	0.0122(-0.3039)	0.0067(4.7875)
	C3	(-0.0344)	0.1492(1.7396)	0.2495(-0.5814)	0.0172(4.5424)
	C4	(0.0551)	0.0402(0.9188)	0.0149(-0.4198)	0.0096(6.1059)
SH	C1	(0.0006)	0.0288(0.8190)	0.0129(-0.4390)	0.0095(5.6559)
	C2	(0.0587)	0.0176(0.6951)	0.0093(-0.3403)	0.0115(5.1727)
	C3	(0.6280)	-0.0268(1.8382)	0.1663(-0.0239)	0.2469(6.8550)
	C4	(-0.1088)	0.0178(0.5795)	-0.0534(-0.5374)	-0.0462(5.7805)
DJF	C1	()	0.5127(9.3824)	0.1259(-3.9800)	0.1832(47.9340)
	C2	()	0.2530(7.5153)	0.0857(-2.8300)	0.2699(38.4140)
	C3	()	0.4234(10.2227)	0.1534(-3.6439)	0.3324(49.7517)
	C4	()	0.5016(9.5815)	0.1134(-3.9303)	0.2056(52.3949)
JJA	C1	()	0.4418(9.1540)	0.1212(-22.8537)	0.1260(-34.6021)
	C2	()	0.1770(7.5202)	0.0747(-2.6227)	0.2381(34.9692)
	C3	()	0.5340(9.6684)	0.1755(-4.4505)	-0.0003(45.1828)
	C4	()	0.4230(9.7168)	0.1054(-3.7443)	0.1284(46.9609)

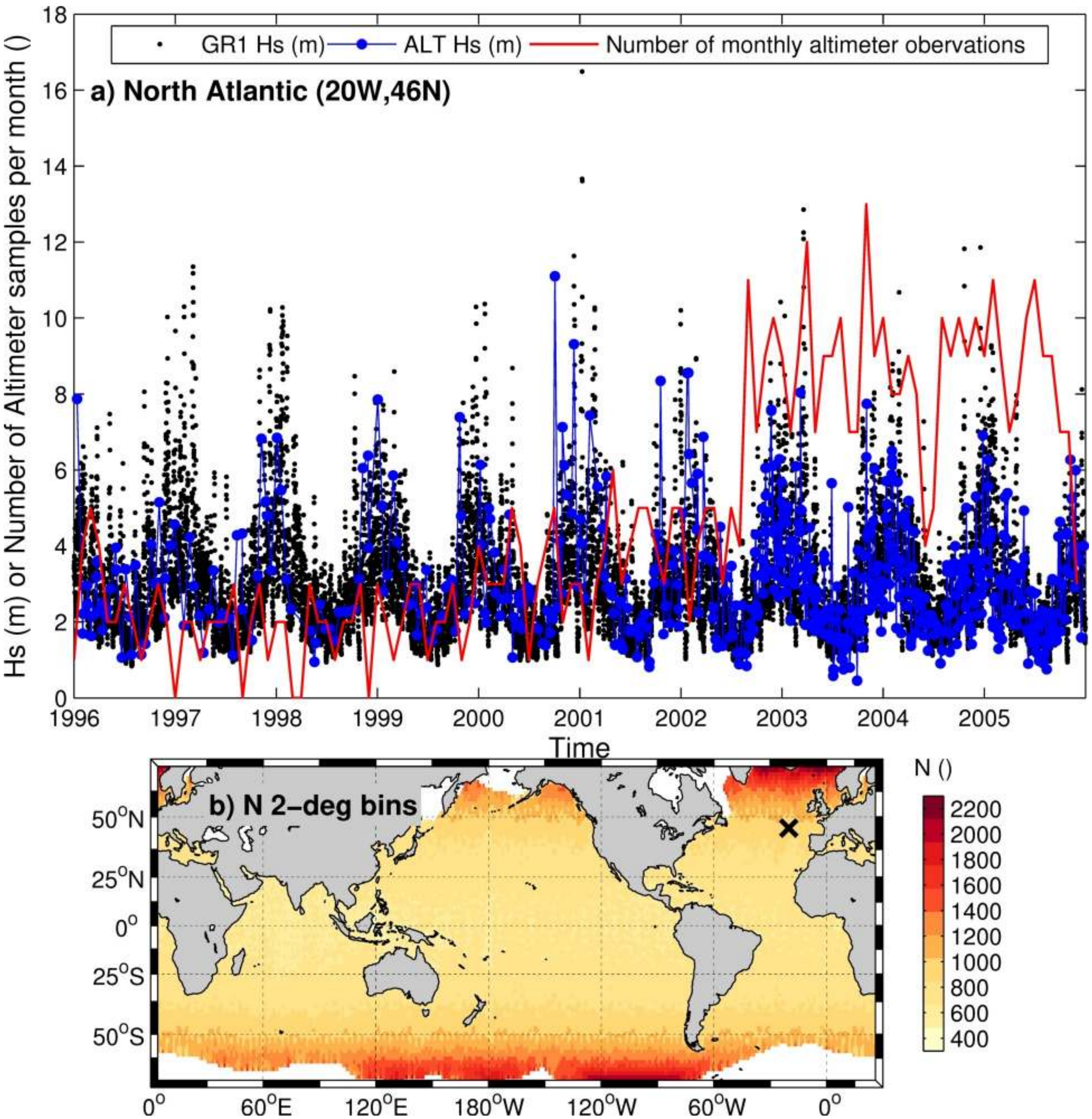

Figure 1: Example H_s time series in the North Atlantic showing the simulated data from ensemble member 1 (black dots), merged altimeters (blue circles), and number of altimeter samples per month (red line) for a 1-degree window (panel a). Panel (b) shows the number of hourly-averaged altimeter observations for the period 1996-2005. The black "X" denotes the location of the example time series shown in panel a.

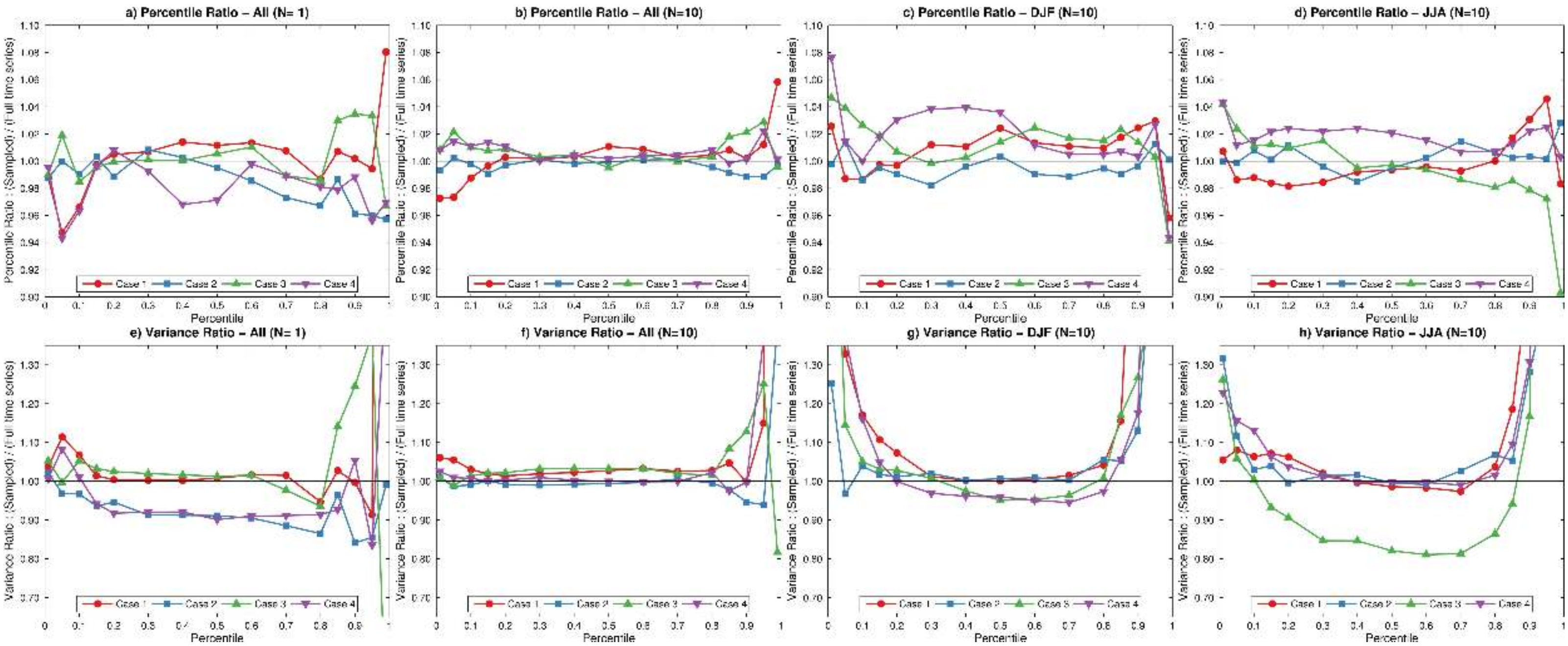

Figure 2: Comparison of the percentiles and variances of the example time series in the North Atlantic for ensemble member 1 for: Case 1 - simple random sampling, Case 2 - systematic sampling, Case 3 - stratified simple random sampling with actual number of altimeters per month, Case 4 - stratified simple random sampling with average number of altimeters per month. The top panels show the ratio of H_s percentiles (sampled/full time series) and the bottom panels show the ratio of H_s variances as a function of percentile. (a,e) represents 1 sample (b,f) represent the average of 10 samples, (c,g) represent the average of 10 samples for the month of January, and (d,h) represent the average of 10 samples for the month of July.

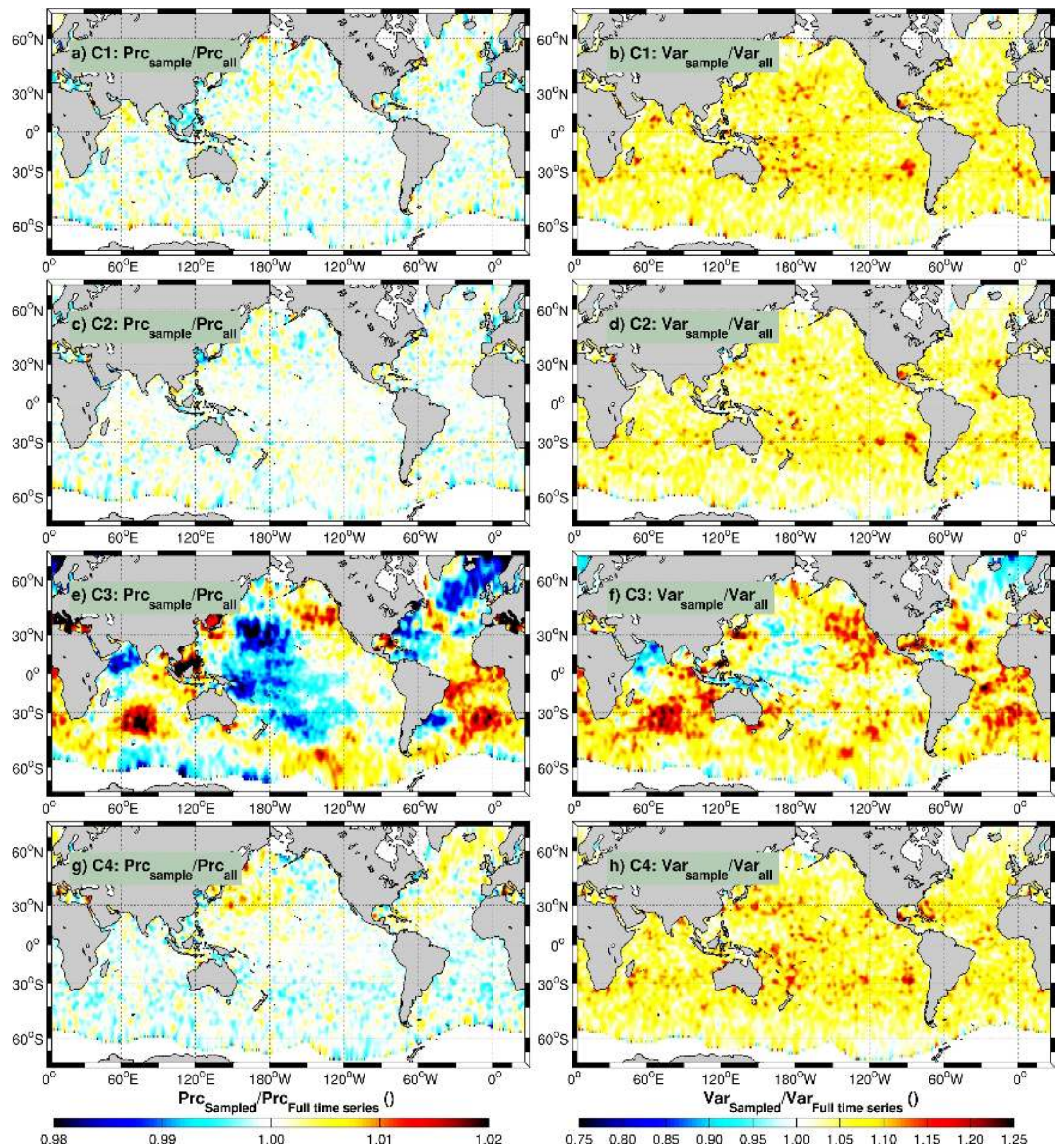

Figure 3: H_s 95% percentile comparison showing the ratios of the subsample time series to the entire time series. The magnitude is given in the left column (a,c,e,g) and variance is given in the right column (b,d,f,h). Each row represents the various sampling strategies averaged using 10 sub-sampled time series. C1, C2, C3, and C4 correspond to the sampling strategies, cases 1 through 4, described in the text.

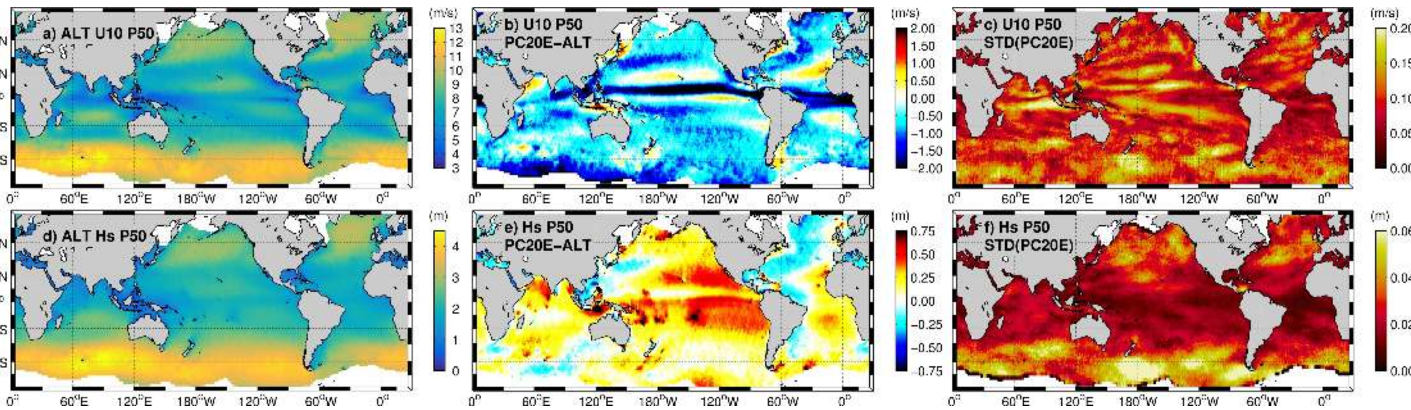

Figure 4: Wind speed (U10) (a,b,c) and wave height (H_s) (d,e,f) comparisons of the median (P50) in units of ms^{-1} and m respectively. a) and d) display the altimeter observations for reference. b) and e) display the difference between PC20E and the altimeters (PC20E-ALT). c) and f) display the standard deviation of the PC20-1 to PC20-7.

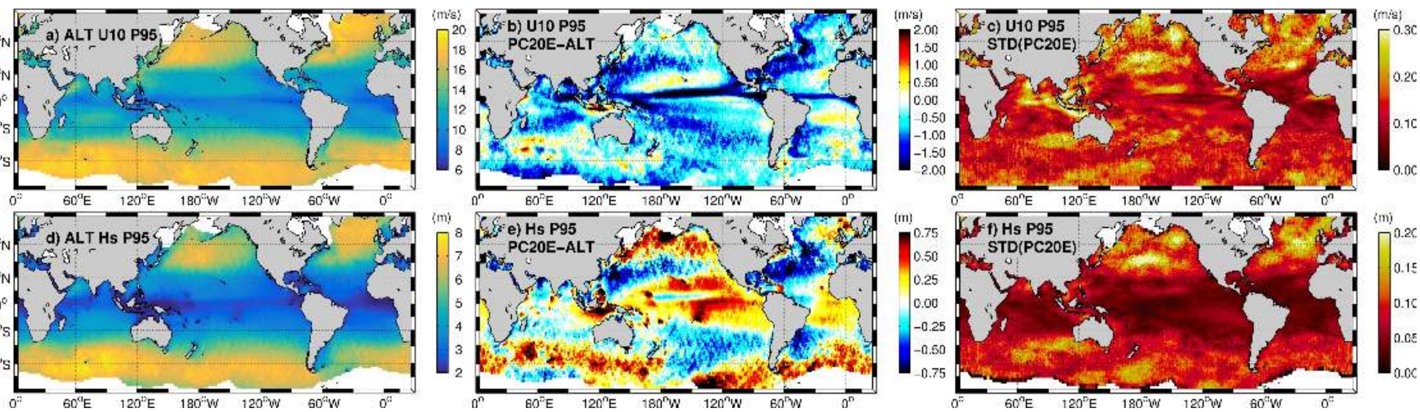

Figure 5: Same as Figure 5 except for the 95th percentile (P95).

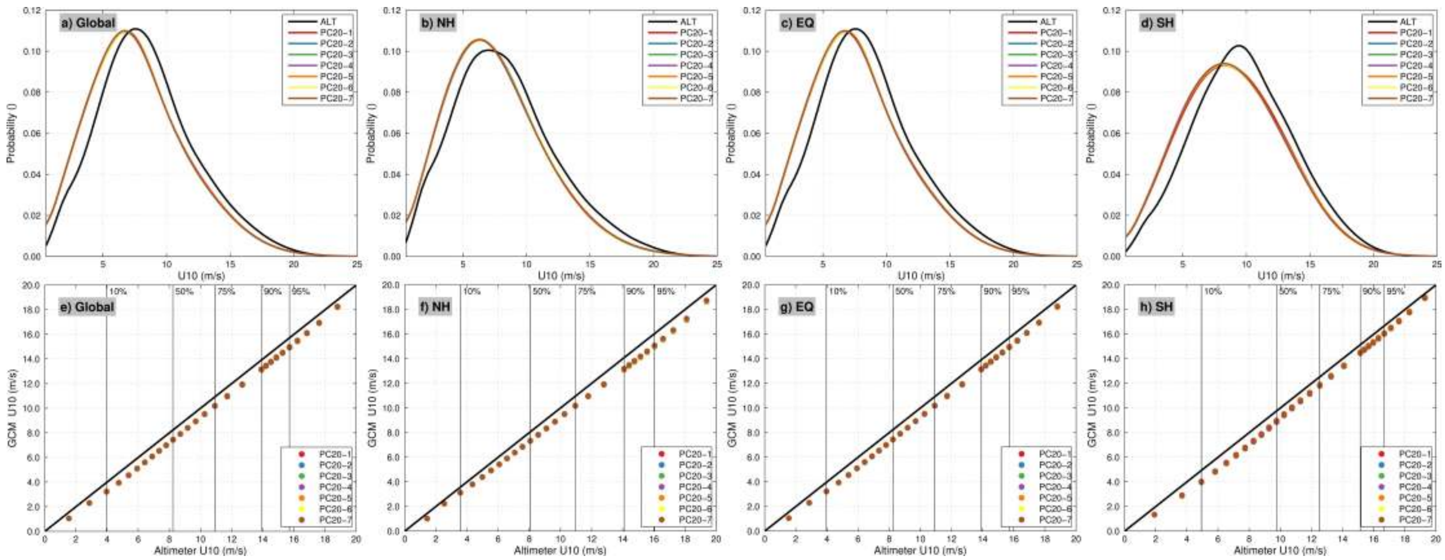

Figure 6: Wind speed probability distribution comparison (a,b,c,d) and quantile-quantile comparison (e,f,g,h) globally (a,e), in the Northern Hemisphere ($> 25°$N) (b,f), near the Equator ($\in 25°$N/S) (c,g), and in the Southern Hemisphere ($> 25°$S) (d,h).

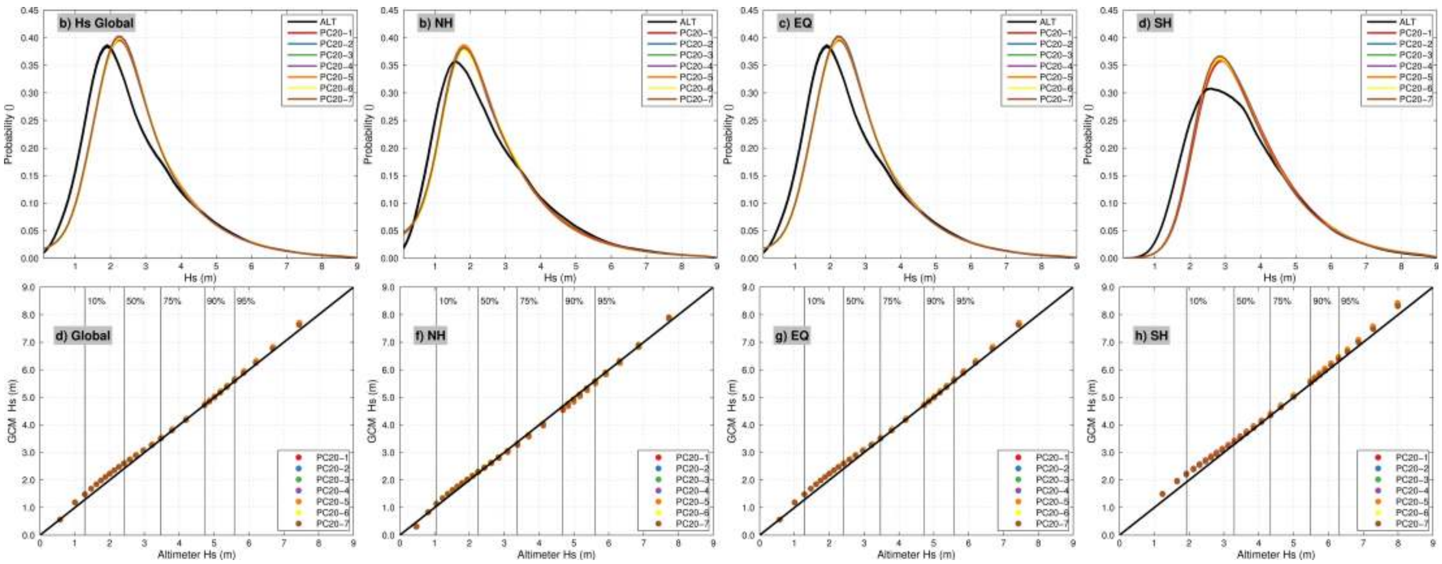

Figure 7: Same as Figure 6 except for H_s.

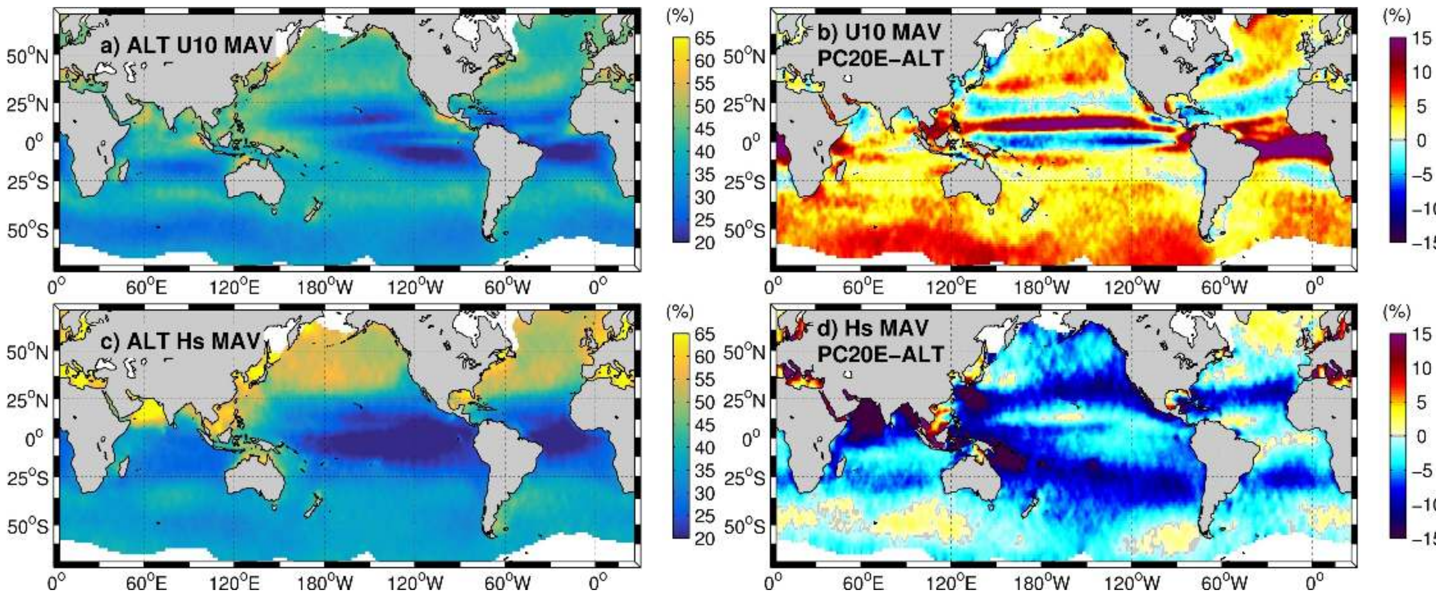

Figure 8: U10 (a,b) and H_s (c,d) of the mean annual variability (MAV) given in a percentage. a) and c) display the altimeter observations for reference. b) and d) display the MAV difference between ensemble and the altimeters (PC20E-ALT).

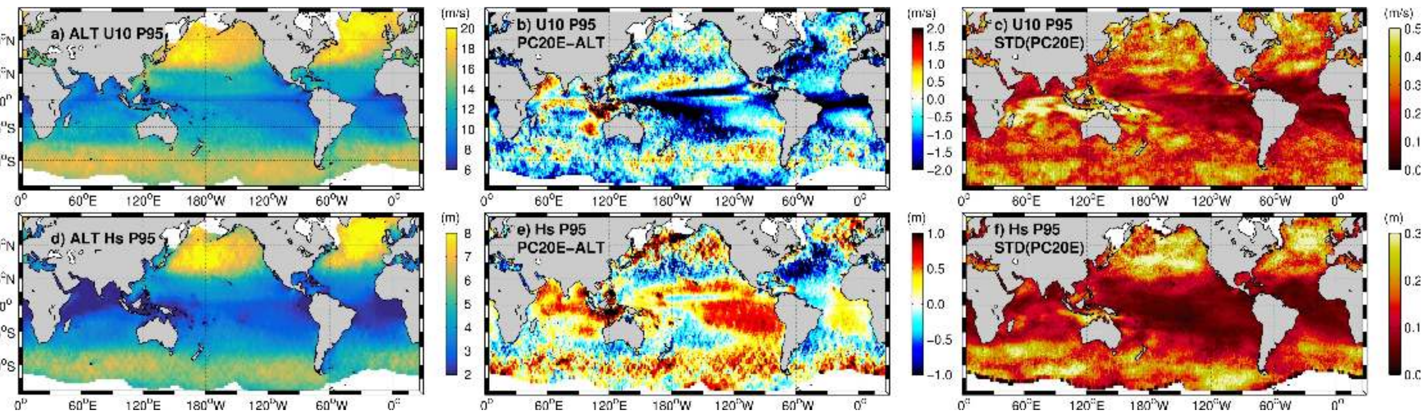

Figure 9: Same as Figure 4 except for the 95th percentile (P95) in the months of December-January-February.

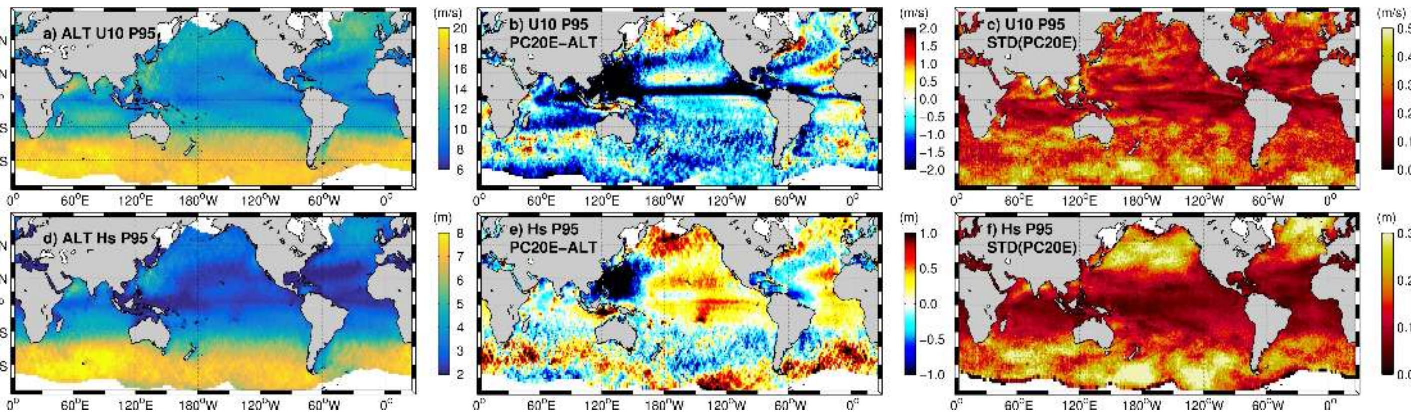

Figure 10: Same as Figure 4 except for the 95th percentile (P95) in the months of June-July-August.

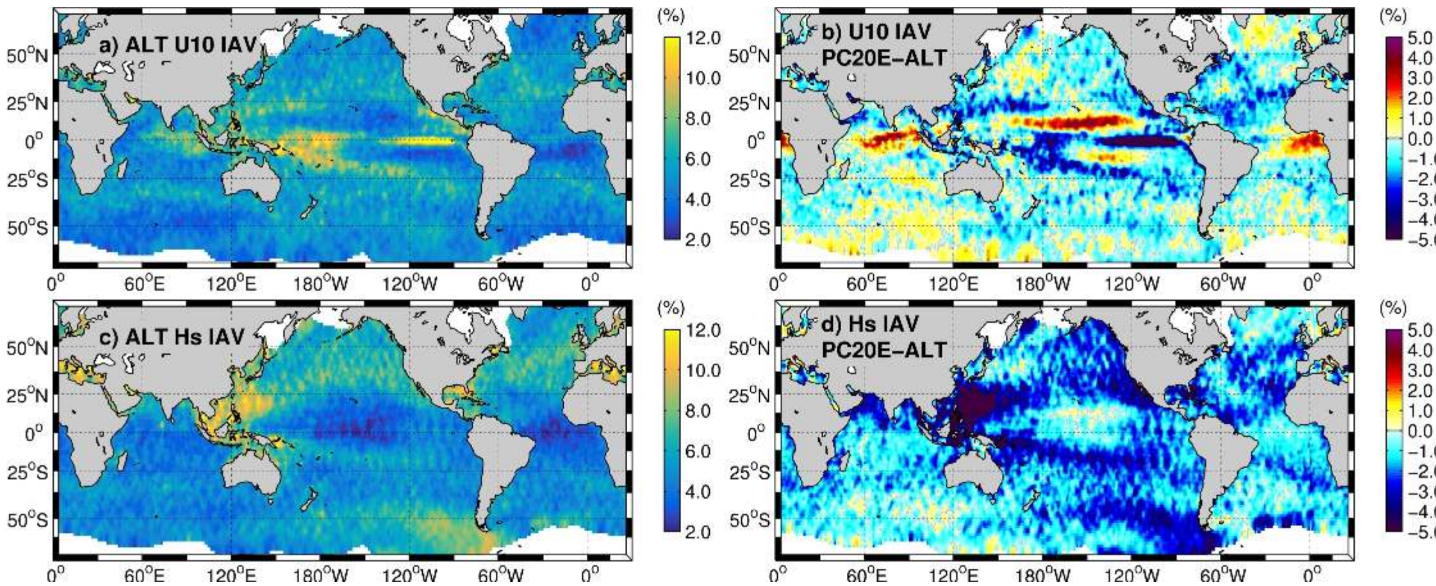

Figure 11: Same as Figure 8 except for the inter-annual variability (IAV).