Ocean Modelling 110 (2017) 52–65

Contents lists available at ScienceDirect

Ocean Modelling

journal homepage: www.elsevier.com/locate/ocemod

Evaluation of bias correction methods for wave modeling output

K. Parker*, D.F. Hill

School of Civil and Construction Engineering, Oregon State University, Corvallis, OR 97330, USA

A R T I C L E I N F O

Article history:
Received 21 June 2016
Revised 12 December 2016
Accepted 25 December 2016
Available online 26 December 2016

Keywords:
Wave modeling
Bias correction
Climate change
Multivariate statistics

A B S T R A C T

Models that seek to predict environmental variables invariably demonstrate bias when compared to observations. Bias correction (BC) techniques are common in the climate and hydrological modeling communities, but have seen fewer applications to the field of wave modeling. In particular there has been no investigation as to which BC methodology performs best for wave modeling. This paper introduces and compares a subset of BC methods with the goal of clarifying a "best practice" methodology for application of BC in studies of wave-related processes. Specific focus is paid to comparing parametric vs. empirical methods as well as univariate vs. bivariate methods. The techniques are tested on global WAVEWATCH III historic and future period datasets with comparison to buoy observations at multiple locations. Both wave height and period are considered in order to investigate BC effects on inter-variable correlation. Results show that all methods perform uniformly in terms of correcting statistical moments for individual variables with the exception of a copula based method underperforming for wave period. When comparing parametric and empirical methods, no difference is found. Between bivariate and univariate methods, results show that bivariate methods greatly improve inter-variable correlations. Of the bivariate methods tested the copula based method is found to be not as effective at correcting correlation while a "shuffling" method is unable to handle changes in correlation from historic to future periods. In summary, this study demonstrates that BC methods are effective when applied to wave model data and that it is essential to employ methods that consider dependence between variables.

1. Introduction

One of the key drivers for the development of wave models has been the need for high resolution data distributed across both large areas and large time windows. These data are integral for navigation, hazard forecasting, recreational purposes, and a broad array of ocean science applications. Due in part to high operational costs, the observational in-situ record is sparse and cannot practically cover all areas of the ocean at all times (Fig. 1). Therefore, models serve to "fill the gaps" and provide a more complete understanding of the wave climate. Models are further essential for any study of future wave conditions where data clearly do not exist. However, data from wave models can and do consistently exhibit bias (defined is this study as a systematic deviation from the corresponding observed "true value") that results from a variety of factors including inherent simplifications and inadequate model physics (parameterizations, assumptions, etc.), numerical solution schemes, resolution, insufficient or imperfect calibration datasets, and incorrect boundary forcing data.

Model bias is not unique to the field of ocean sciences. In particular, the atmospheric and hydrologic science communities have developed a mature body of literature dealing with the subject. This is primarily due to a reliance on general circulation model (GCMs), which are highly prone to bias (Mehran et al., 2014; Mueller and Seneviratne, 2014). Since climate model output is often the input for other models, GCM biases propagate downstream and detrimentally impact other modeling results (Xu, 1999; Christensen et al., 2008). To resolve this, methods for bringing model output back into alignment with observations have been sought. This demand is the foundational driver for the development of bias correction (BC) procedures.

At the conceptual level, BC methods (as explored in this paper) define a transfer function that transforms model data to a new dataset with fewer statistical biases. How this transfer function is defined ranges from a simple shift in the mean value to increasingly complex techniques that can fully correct statistical distributions. For example, the widely used quantile mapping methods (Panofsky et al., 1958; Wood et al., 2004; Déqué, 2007; Piani et al., 2010) attempt to match the CDF (Cumulative Distribution Function) of the model time series to that of a target, typically an observational time series. Recent advances in BC techniques have expanded into the multivariate domain and attempt to incorporate the relationship between variables as well. In cases

* Corresponding author.
E-mail addresses: parkerk@onid.oregonstate.edu, parkerk@oregonstate.edu (K. Parker), david.hill@oregonstate.edu (D.F. Hill).

http://dx.doi.org/10.1016/j.ocemod.2016.12.008
1463-5003/

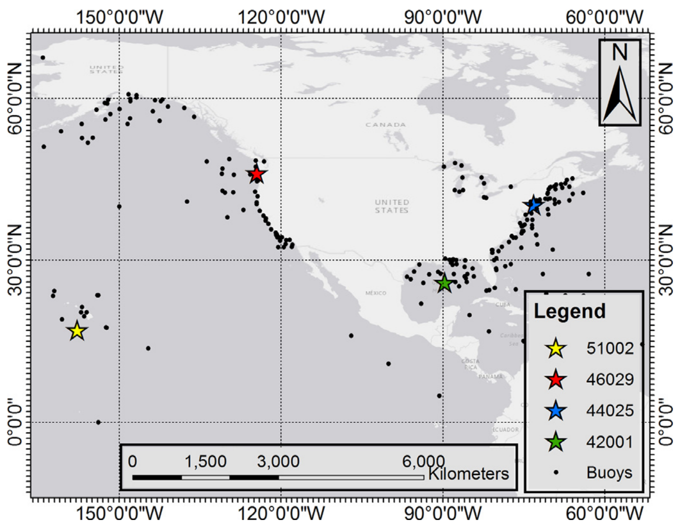

Fig. 1. Site map showing the buoy locations used for bias correction marked as stars and all other buoys (both historic and active) marked with dots. The external tick marks represent the WW3 model 1 degree resolution grid.

of dependent variables, well-intentioned univariate BC can lead to incorrect inter-variable correlations and non-physical results (Chen et al., 2011; Thrasher et al., 2012). This is important in the field of wave modeling as wave parameters (height, direction, period, etc.) are highly correlated (Mathisen and Bitnergregersen, 1990; Ferreira and Soares, 2002; Repko et al., 2004; De Waal and van Gelder, 2006; Corbella and Stretch, 2012). Recent contributions to multivariate BC differ in how they treat inter-variable relationships and include a data binning technique (Piani and Haerter, 2012), a direct bivariate distribution approach based on copulas (Li et al., 2014), and a shuffling technique (Vrac and Friederichs, 2014). For convenience these methods will be called the Binning Method, the Direct Method, and the Shuffling Method respectively. These three BC techniques will be explained in detail below in the methods section.

From a broad perspective, many modeling practices can be considered a form of "bias correction." For example, model tuning or data assimilation are both procedural ways of attempting to bring model output into agreement with observations. In the meteorological community, Model Output Statistics (MOS) are routinely used to remove bias in numerical weather prediction, albiet in a format that may be more easily recognized as statistical downscaling. For brevity, this study does not consider all bias-reducing techniques. Instead, it focuses on statistical bias correction methods that correct statistical distributions of model variables. With this constraint, there are three options when looking at bias correction in a wave modeling problem:

(A) Apply no BC (Leake et al., 2007; Lionello et al., 2008; Grabemann and Weisse 2008; Mori et al., 2010).
(B) Apply BC to the input data (Wang and Swail, 2002; Hemer et al., 2011, Hemer et al., 2012; Durrant, 2013; Wang et al., 2010; Wang et al., 2014).
(C) Apply BC to the output wave fields (Caires and Sterl, 2005, Cavaleri and Sclavo, 2006; Andrade et al., 2007; Tomas et al., 2008; Charles et al., 2012).

Consideration should be given as to which of these methodologies is the most applicable to the particular study since each has associated strengths and weaknesses. Option (**A**) is the ideal case and the most theoretically robust. The gradual improvement of physical models is undeniably the end solution to bias. BC can be thought of a temporary solution to bring currently flawed model predictions into alignment with reality but with associated limitations. Ehret et al. (2012) reviews the broad issues with BC, including the lack of a sound physical basis (Haerter et al., 2011), impossibly restrictive assumptions, a masking of uncertainty, and introduction of physical inconsistencies between other model variables (alteration of the spatial and temporal covariance structure of variable fields (Johnson and Sharma, 2012)). This being said, if model results are sufficiently biased, analysis may be restricted to only being relative (non-absolute). While this is acceptable for a comparison of results (say historic and future simulations), using uncorrected wave model output to force additional models (e.g., coastal sediment transport) will simply propagate the bias.

Option (**B**) has proven effective at improving modeled wave parameters (Caires et al., 2004; Hemer et al., 2011; Durrant et al., 2013) since many wave modeling errors can be traced directly to input wind fields (Cardone et al., 1996; Rogers and Wittmann, 2002; Durrant et al., 2013). Additionally, this option has the advantage of correcting model output over the entire model domain. This said, bias correction of wind fields has significant disadvantages including being computationally expensive (Wang and Swail, 2002) and oftentimes practically difficult. The process is complicated by sparse observational information across the ocean basins, both spatially and temporally, leading to target datasets of limited length, spatial coverage, and accuracy. Furthermore, even with BC of wind fields the wave model output will likely exhibit bias due to wave modeling errors (Rogers et al., 2005).

Option (**C**) is well positioned to ensure that wave model output will be statistically in agreement with wave climate observations. Despite this, there have been relatively few studies of the application of BC methods to wave model output. To discuss a few examples, Andrade et al. (2007) used a variant of quantile mapping that fits a log-normal distribution to significant wave height and changes the parameters to match probability distribution functions (PDFs). Additionally, Caires and Sterl (2005) used non-parametric regression estimators, Cavaleri and Sclavo (2006) used a parametric correction, and Charles et al. (2012) used a univariate quantile mapping method to independently correct wave height, wave period, and wave direction. It should be noted that we are considering a "local" BC problem in the strict sense that model output is corrected only at the observation location. In general, wave model bias can be considered slowly varying (e.g., Fig. 7 of Hemer et al. (2012)) and transfer functions derived at one location can be used to inform the correction at other nearby locations. In this sense, option (**C**) is well suited as an intermediate step in a nested modeling approach. Basin-scale wave model output can be extracted at an observation location, corrected, and then used as open boundary forcing for a local-scale domain. If considering a location with rapidly varying bias structure (e.g. complex nearshore coastal configurations) or looking at corrections across larger regions, Tomas et al. (2008) introduces a spatial-temporal correction based on a nonlinear parametrization of Empirical Orthogonal Functions (EOFs).

The main contribution of this paper is to provide a comparative study of BC methods applied to wave model output (option (**C**) as listed above). This study is the first quantitative comparison of univariate and bivariate methods and is the first to apply bivariate methods to wave model output applications. This paper focuses specifically on the "application" of various BC techniques, leaving more complete expositions of the individual methods to the relevant citations. Section 2 of this paper describes the geographic location and the relevant data used for this study and Section 3 introduces the BC techniques, including a brief theoretical and technical overview as well as the methodology for comparison between them. Comparative results are provided in Section 4 and a discussion of the results, including limitations, is provided in Section 5.

2. Data

The data utilized for inter-comparison of BC techniques were provided by the Australian Commonwealth Scientific and Industrial Research Organization (CSIRO). CSIRO has produced wind-wave climate projection datasets (Hemer et al., 2015) using WAVEWATCH III (v3.14; henceforth WW3; Tolman, 1991; Tolman, 2009) forced by an ensemble of atmosphere-ocean general circulation models (AOGCMs). These datasets cover runs from many different AOGCMs, multiple emissions scenarios (CMIP3 and CMIP5) and several climatological periods (1980–2005; 2026–2045; 2080–2099). The model implementation is detailed in Hemer et al. (2012) and Hemer et al. (2013) and the reader is directed to these primary sources for details regarding the model setup and production methodology for the data used in this study.

The CSIRO datasets are an ideal testbed for this intercomparison of BC methods since: (a) the results were found to demonstrate bias (Hemer et al., 2013), (b) the data are available globally, allowing for the investigation of regional variability, and (c) the models are available for several climatological periods, allowing for testing of nonstationary BC. A single AOGCM, MRI-CGCM3 (Yukimoto et al., 2012), was arbitrarily chosen from the CSIRO ensemble. This choice should not be significant as the statistical testing methodology described in this paper is designed to reveal robust differences between methodologies regardless of the specific dataset. Two timeslices were used for this study, CSIRO's historical (1980–2005) and mid 21st century (2026–2050) simulations with the future run using the RCP 4.5 climate scenario.

It should be noted that Hemer et al. (2013) define bias as the difference between the model being forced by AOGCM winds and the model being forced by reanalysis (CFSR; Saha et al., 2010) winds (either raw or bias corrected). In this paper, bias is instead defined as the difference between buoy observations and model output. This choice was made to include the full range of model biases, including those attributable to the wave model itself. Four National Data Buoy Center (NDBC) locations were chosen as test points (Fig. 1):

- #42001: Mid Gulf of Mexico (25.888 N, 89.658 W)
- #44025: Long Island (40.251 N, 73.164 W)
- #46029: Columbia River Bar (46.159 N, 124.514 W)
- #51002: South West Hawaii (17.056 N, 157.803 W)

These buoys were chosen to sample a wide variety of wave climates and therefore test the spatial variability of the BC effectiveness. Fig. 2 shows the monthly climatology at each of the chosen buoys. Monthly average significant wave height (Hs) and peak wave period (Tp) time series are plotted for the buoy, historic model, and future model. All locations show a distinctive seasonal cycle but with significant spread in terms of magnitude at each location. Of particular interest to this study, each location and variable shows a unique bias structure. This said, wave heights seem to be generally underpredicted while wave period is overpredicted in some cases and underpredicted in others.

It should be noted that in principle, BC methodologies can be extended to the n-dimensional case, incorporating more than the two variables considered here. For wave modeling applications this would mainly be of interest in order to include wave direction within the BC framework. This is excluded from the present paper for two reasons. First, BC methods work best with climatological (typically ∼ 30 years) periods of record and this duration of wave direction timeseries is generally unavailable. Second, the comparative bivariate analysis presented here is relatively complex and lengthy. We therefore defer tri-variate methods to a future study in order to maintain clarity and a reasonable length for this study.

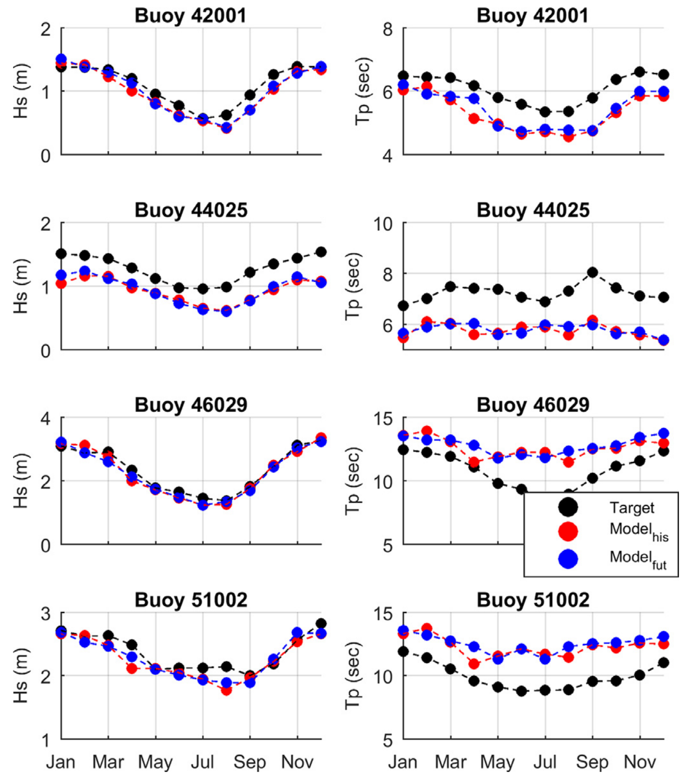

Fig. 2. Monthly means of significant wave height (Hs) and peak wave period (Tp) at each buoy location. Data shown are observational (black), historic model (red) and future model (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Methods

3.1. Bias correction methods

A subset of available BC methods were chosen to illustrate the differences between univariate and multivariate methods, parametric and empirical methods, and to look at the performance of available multivariate methods. In this section we will indicate model data with (M), target buoy data with (T), and bias corrected model data with (B).

3.1.1. Univariate quantile mapping

Several studies have found that distribution based methodologies, in particular quantile mapping, are most effective for BC of climate model data (Themeßl et al., 2011; Teutschbein and Seibert, 2013). Quantile mapping attempts to match the CDF of a modeled time series to that of an observed time series. This is done most commonly using the direct constraint of Déqué (2007) which can be expressed as:

$$B = CDF_T^{-1}[CDF_M(M)] \tag{1}$$

where CDF refers to the cumulative distribution of the time series and the subscripts follow the notation specified above. For example, CDF_T^{-1} is the inverse of the target dataset's CDF. Eq. (1) specifies that each individual model value is replaced with the target value having the same quantile. This will result in a "scaling" of the model time series such that the CDF of the bias-corrected model data exactly matches that of the target. Fig. 3 shows this procedure as applied to a single model variable value (M). Note that in Section 3, all variables will be kept generic and only replaced with wave parameters in the results sections. Also note that the captions for the methodological figures are kept brief, with full explanation in the text. In Fig. 3, the first step of quantile mapping

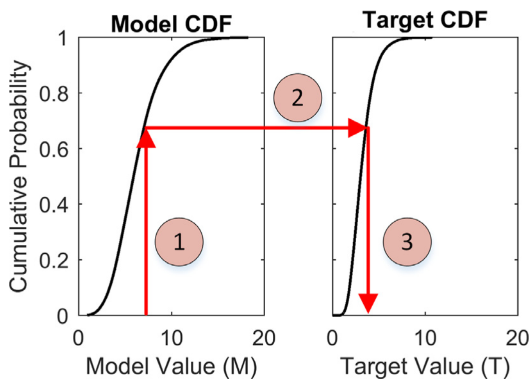

Fig. 3. The Quantile Mapping procedure as applied to a single historic model value.

is to determine the quantile of the model value of interest **(1)**. This quantile is then transferred over to the target dataset **(2)** and a variable value is found by inverting the CDF of the target dataset **(3)**. This value is the new bias corrected model value. This procedure is repeated for each value in the model dataset.

Parametric and empirical versions of quantile mapping differ in how the CDF and its inverse are calculated. The parametric method involves fitting a theoretical (parametric) distribution to the data. Various "long-term" wave parameter distributions have been identified by the literature with the most common being the Log-normal or Weibull for both Hs and Tp (Baur and Staabs, 1998; Ferreira and Soares, 2000; Ferreira and Soares, 2002; De Waal and Van Gelder, 2006; Holthuijsen, 2007). It should be noted that there is no universal best fitting distribution because the statistical properties of waves vary both temporally and spatially, requiring different techniques for each individual case. For this reason multiple commonly used distributions were tried with the best fitting, determined via a goodness of fit test, being ultimately selected. The Rayleigh, Log-normal, Generalized Extreme Value (GEV), Weibull, and Burr distributions were fitted for Hs. For Tp, the same distributions were tried but with the Burr replaced with a Gamma distribution. This fitting and selection process was repeated for each BC operation.

The empirical method determines CDF values directly from the data with no assumption of a distribution. For this study a kernel smoothing function (Bowman and Azzalini, 1997) was used to provide smooth estimates of CDF quantiles and for interpolation when inverting the CDF. Calculated values were checked against the direct empirical estimates to choose a proper bandwidth that avoided oversmoothing.

There exists some disagreement in the literature as to which method (empirical or parametric) is preferable. Li et al. (2010) state that use of an empirical distribution will result in frequent interpolation and extrapolation with unsatisfactory results. This should only be an issue for small datasets and can be circumvented by a kernel smoother with proper bandwidth. Lafon et al. (2013) looked at a comparison of empirical and parametric techniques, with a focus on the effect of using a variable number of quantiles. They cautioned that too many quantiles will result in an "overfitting" to the target dataset, causing a loss of robustness due to an oversensitivity to the calibration period. The climate, if treated as a stochastic process, would be expected to have some random variability around its "true" distribution. A perfect fitting of the model to a training segment of the climate time series being considered (target) could theoretically overfit the correction, causing error between the BC and the true distribution. This issue is subtle and many authors simply fit the data to the empirical CDF at every data point (each model value has is its own quantile). Attempting to resolve this uncertainty is one of the goals of this study.

Eq. (1) is formulated to bias correct model data when observations are available. To bias correct future model data (where observations are not available), the so called equidistant quantile matching method of Li et al. (2010) can be used. This method is described by Eq. (2) and can be seen conceptually in Fig. 4.

$$B_f = M_f + CDF_{T,h}^{-1}\left[CDF_{M,f}\left(M_f\right)\right] - CDF_{M,h}^{-1}\left[CDF_{M,f}\left(M_f\right)\right] \quad (2)$$

In this equation, the subscripts h and f are used to denote historic and future values. As shown in Fig. 4, the first step **(1)** is to find the quantile corresponding to the future model variable value $CDF_{M,\,f}(M_f)$. The variable value of both the target and historic model are evaluated (using the inverse CDF) at this quantile value and the difference is found **(2)**. This difference is the model predicted change from the historic run to the future run. The difference between these two values is then applied to the model future value to acquire the BC value **(3)**. Of note, this method necessarily makes the assumption that the statistical properties of the AOGCM systematic errors are stationary, a result that is not necessarily demonstrated in the context of climate change and is itself a topic of research (Charles et al., 2012; Teutschbein and Seibert, 2013).

3.1.2. The binning method

The Binning method (Piani and Haerter, 2012) is a multivariate method that was originally developed for jointly correcting surface temperature and precipitation meteorological fields. This technique handles correlation by first correcting one of the variables with a standard 1D BC method and then correcting the other variable conditionally based upon bins of the first variable. This is conceptually illustrated in Fig. 5. The first step **(1)**, not shown in the figure, is to use standard quantile mapping on one of the variables (marked with an X subscript). The second step **(2)** is to select a bin (this study uses 10 bins, or deciles) of the first variable (X) from both the target and model datasets. Note that the bins are selected in CDF space so a decile is shown in the figure as a single bar of data ($.2 > CDF_X \geq .1$), indicated with red symbols. The second step is to determine CDFs from these bins of data **(3)**. Finally a normal quantile mapping is used with these decile dependent CDFs **(4)** so as to bias correct the second variable (Y). Step 3 is repeated for each model value within the decile. Then a new bin is selected and the process is repeated.

The Binning method has a free variable in the number of bins to be used. The number must be large enough to properly distinguish correlation but small enough that sufficient data are within each bin to allow an accurate inversion of the CDF. Piani and Haerter [2012] found that surprisingly few bins (<5) were required to accurately correct the correlation. This study's choice of deciles was arbitrary and it is hypothesized, though untested, that a smaller number of bins would have produced similar results.

Note that the Binning method is a framework for dealing with correlation and is not dependent on the univariate BC technique. Any univariate BC could be substituted for quantile mapping and the methodology would be the same. An interesting question is whether or not the choice of the first variable to be corrected is significant. As these methods correct this first variable in a univariate way, and then correct the second conditionally based on this first variable, it is not unreasonable to assume that the first variable will be better matched to the target statistically. Colloquially, the second variable would "sacrifice" its ability to perfectly match the target statistically in order to match the correlation between the variables. For all results presented in this paper Hs was used as the first variable corrected. However, a single buoy (#46029) was used to determine the sensitivity of results to this decision by comparing results to the case when Tp was used as the first variable.

While univariate quantile mapping has an accepted methodology for "future" runs (Eq. (2); Fig. 4), the authors know of no

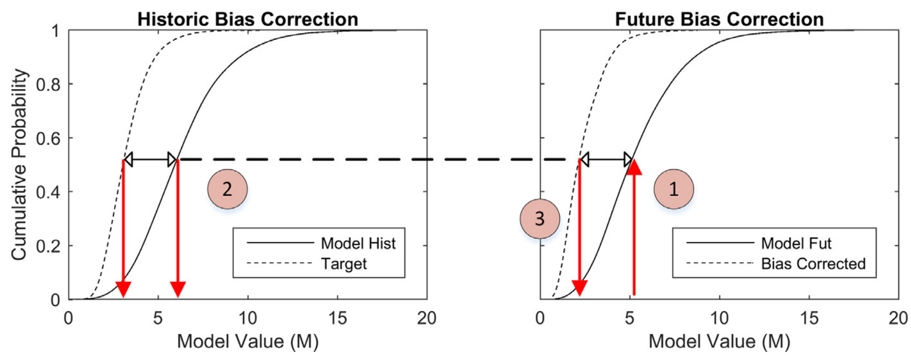

Fig. 4. Equidistant Quantile Matching as applied to a single future model value.

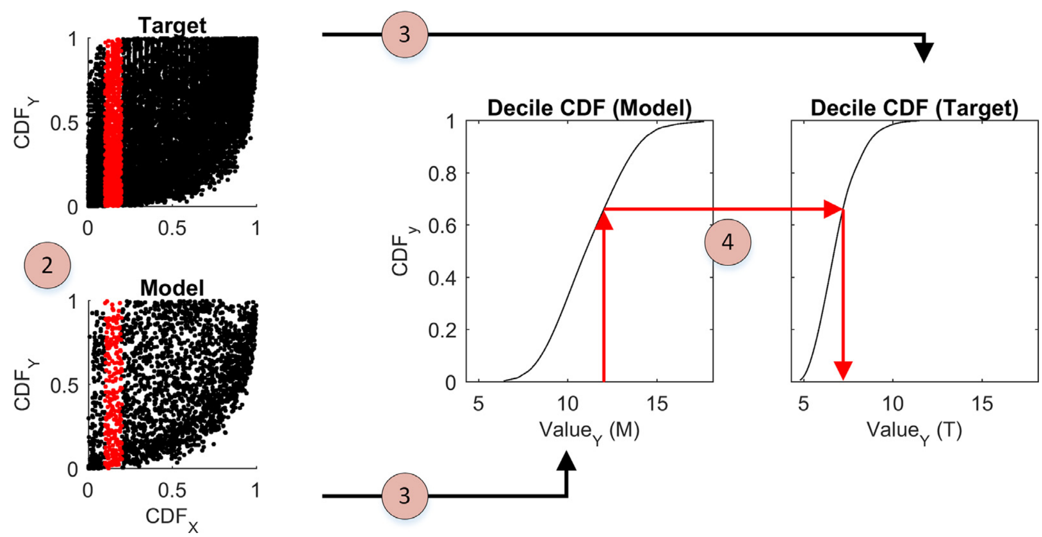

Fig. 5. The Binning Method as applied to the second variable of a bivariate bias correction. Example deciles of bivariate CDF are marked with red dots. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

similar procedure for the multivariate BC methods. As the Binning method is a framework built around univariate BC, an extension based on the principles of equidistant quantile matching was developed for this paper and is detailed in Appendix A.

3.1.3. The direct method

A second technique for bivariate BC was proposed by Li et al. (2014) and is based on a direct defining of a bivariate distribution between the variables of interest using copulas. This allows conditional correcting of one of the variables based upon the bivariate distribution's definition of the relationship between the two variables. The mathematics of the Li et al. (2014) method are significantly more in-depth than those of the previous methods and this paper only presents a procedural overview. Fig. 6 shows the general process for correcting a single model output Hs, Tp pair.

The first step of the methodology **(1)** involves a univariate BC of the first variable, similar to that of Binning method (not shown in Fig. 6). The next step is to fit a copula to both the model and target datasets **(2)**. A variety of Archimedean and Elliptical copulas (Clayton, Frank, Gumbel, Gaussian, and Student T) were considered and the best fitting was chosen for each dataset. Copulas are a way to model dependence between variables, separate from the marginal distributions. Marginal distributions are defined as the PDFs of the individual variables, without reference to the other correlated variable. By looking at CDFs, the marginal distribution is removed since CDFs are both non-dimensional and approximately uniformly distributed. This allows for a more direct look at correlation between variables as the variables have been reduced to a comparable state. A primer on copula mathematics is provided by De Wall and Van Gilder (2006), and De Michele et al. (2007) among others.

Once a copula has been fit to the data, the next step is to determine the bivariate CDF value of the model copula **(3)**. For normal quantile mapping, BC is based on evaluating the model quantiles with target data and this is similar to what will be done here, just in a bivariate capacity. Following this template, the analysis then moves to the fitted target copula. We are trying to find the comparable quantile to that of the model copula but since copulas are bivariate there are multiple locations of identical CDF magnitude, i.e., a so called level curve. This complicates analysis as more than just the bivariate CDF value is required for determining a unique location on the CDF surface. We can escape this problem because of the two bias corrected value pairs, one is already known from the univariate BC (step **(1)**). Using this information it is possible to take a slice of the bivariate CDF at the known BC value **(4)**. Since we know the bivariate CDF value of the model (evaluated at **(3)**), we can take the inverse of this slice to find the CDF value for the second value (based on the target copula) **(5)**. With this quantile we can then invert the target CDF to get the bias corrected model value **(6)**.

Similar to the Binning method, a non-stationary methodology for future runs was developed by the authors and is presented in Appendix A.

3.1.4. The Shuffle method

The "Schaake Shuffle," or Shuffle method for this paper, is a method of multivariate BC that Clark et al. (2004) developed as

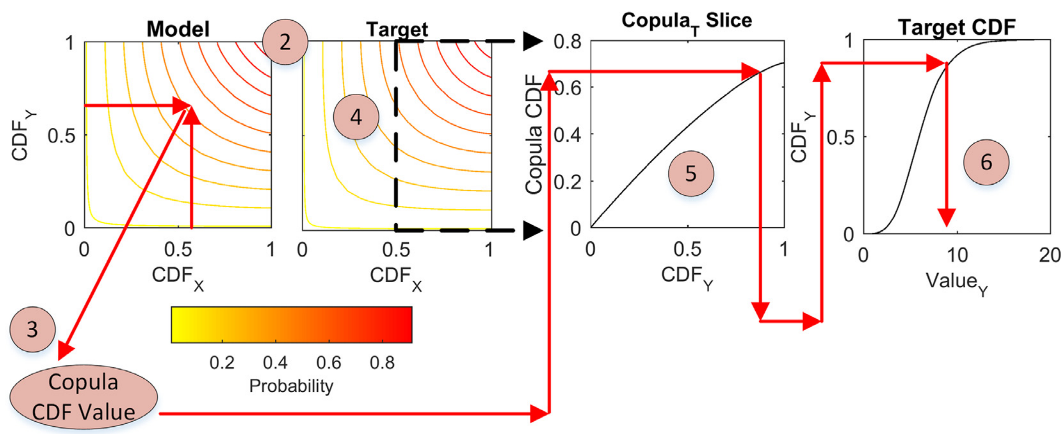

Fig. 6. The Direct Method as applied to the second variable of a bivariate bias correction.

a method for correcting both spatial and temporal covariability in forecast models. Vrac and Friderichs (2014) showed that the method could further be used for multivariate BC. The Shuffle method takes advantage of the fact that the CDF of a time series is not dependent upon the order of the time series. This means it is possible to univariately bias correct the CDF of each variable in a dataset to fix the individual variable's bias and then shuffle each variable's time series to fix the intervariable correlation. Correlation is, simply put, a relationship between two variables and this can be created by properly organizing each variable's timeseries to exhibit the wanted relationship. The target timeseries contains a template for the relationship between variables (as displayed by the way each variable moves in relation to the others) so by rearranging the model timeseries of each variable to mimic that of the target, the method can recreate the wanted correlation. The process is easy to implement and involves first univariately bias correcting each variable to remove the statistical bias. The next step is ranking, in terms of magnitude, both the target and model output (already univariately bias corrected) for each variable. The ranked model output for each variable is then placed into the same order as that of the ranked target. Fig. 7 shows this procedure graphically. The numbers within the circles are the rank of the circles in terms of magnitude for each variable timeseries (represented by the colors brown and blue). For the Target and Model timeseries the colors become darker from right to left representing the original order of the variables. For the shuffling, we use the order of ranks from the target series to rearrange the model to the BC time series. This results in the model having the fifth ranked variable moved to position one from position two for Var_X and to position one from position five for Var_Y. This shuffling continues for all values in the time series.

The Shuffle method is very powerful in that temporal, spatial, and intervariable relationships can be corrected simultaneously. This is only possible because the relationship in time, space, and location (as preserved by the rank) of the target time series is being "stolen" for the model time series. This differentiates the Shuffle method from other BC techniques as it produces a scaled version of the target time series rather than a scaled version of the model output. This further requires that the target time series and model time series have the same sampling rate and size which can be a significant limitation.

3.2. Validation methods

The primary method of assessing the effectiveness of a BC methodology is to look at how well the method is able to change the statistical properties of the model to that of the target. Because our validation methodology involves an ensemble, these statistics were normalized as the relative difference (RD), defined as:

$$RD = \frac{B - T}{T} \tag{3}$$

in which the variables T and B follow the notation from above but represent some metric of performance (mean, standard deviation, 95th quantile, etc.).

A comparison of statistical moments from individual variables ignores the relationship between variables. Since preservation of this relationship is the primary goal of a multivariate BC, we also looked at a comparison of correlation coefficients (a measure of the statistical dependence between two variables). Three common correlation coefficients were examined, (Kendall's τ, Spearman's ρ, and Pearson's r) all of which measure a different form of dependence. Kendall's τ measures the degree of similarity between ranked time series (how the variables move together). Spearman's

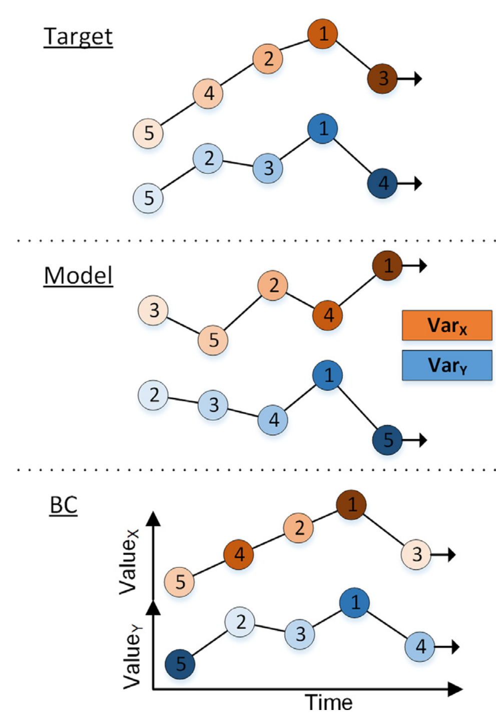

Fig. 7. The Shuffle Method (Schaake Shuffle). Numbers indicate rank in terms of magnitude. Color gradient indicates time with values later in the timeseries being darker. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ρ measures how well the relationship between two variables can be described by a monotonic function. Pearson's r measures the strength and direction of the linear relationship between two variables.

A further metric used as a quantification of BC effectiveness was the shape of the bivariate PDF and CDF. Correlation coefficients would be expected to capture some of the relationships shown by the shape of these distribution functions but not all. Piani and Haerter (2012) noted this when they found that the shape of the bias corrected bivariate CDF was one of the primary means of seeing the difference between a bivariate and univariate correction. Their study noted this qualitatively but did not propose a quantitative metric of this result. We expand upon this by considering the root mean square error (RMSE) between the target's bivariate CDF and the bias corrected model's CDF as a way of quantifying how well correlation is corrected. PDF RMSE was also investigated as an additional perspective into changes of overall distribution shape.

Oftentimes wave modeling studies are more concerned with extreme events than the mean of a wave climate. High return level events are drivers of coastal hazards and can dominate related processes signals (such as geomorphology) (Anderson et al., 2010; Zhang et al., al.,2002). This importance brings an interest in the ability of BC to address incorrect model prediction of high quantile events. For this reason the 95th and 99th quantiles were compared between the target and BC timeseries as a further metric for comparison.

As mentioned earlier, previous studies have brought into question the robustness of BC techniques (Li et al., 2010; Lafon et al., 2013) and in particular empirical methods due to overfitting and the sensitivity of BC to the training period. We decided to address this issue by considering measures of performance using a modified jack-knife (also known as a k-fold) cross-validation technique following the procedure of Lafon et al. (2013). The cross-validation component of this methodology involved breaking up the total target record into a training segment and a validation segment. The metrics of performance are then assessed by developing the BC from the training segment and evaluating its performance with the validation segment. The jack-knife, or k-fold, component of this methodology means that an ensemble of statistics are calculated using different training periods. Specifically, the total target record is broken up into a number of sub-records. For each iteration, one of these sub-records is removed as the validation segment while the others are used as the training segment. For each of these iterations the RD is calculated, resulting in an ensemble of normalized "errors" that will quantify the robustness of the BC technique to the calibration period.

Wave parameters often exhibit a strong seasonal signal (See Fig. 2) that is important to capture in a BC methodology. For this reason it was decided to additionally break the total record up into seasons. As each of these seasons has a characteristically different distribution shape, this can be considered a further test of robustness in how well the BC method performs on different shaped distributions. Each of these seasons was run through the cross-validation procedure. With 4 seasons, each with a 10 bin jack-knife test, the resulting ensemble for each measure of performance was 40 data points. As the RD is normalized, the data were not regrouped by season but treated as a single sample.

4. Results

Although BC was tested on 4 locations, only results from buoy 42001 will be presented graphically in this section. This location was chosen for presentation as being a characteristic case with all other locations exhibiting the same general behavior. It was additionally chosen as the most illustrative with model output being generally the most biased across all metrics. Any cases where individual results vary from the displayed data will be discussed within the text. Furthermore, additional plots for all locations have been provided in the supplementary materials.

4.1. Statistical bias

Fig. 8 presents results of the BC procedure in terms of summary statistics for Hs and Tp. The ensemble, as described above, provides important information about the sample spread, leading to the presentation of the data as box plots. All presented box plots have the centerline of the box at the median value, the edges of the box at the 25th and 75th percentiles, and the whiskers extending to the most extreme data points not considered outliers (as defined using the Tukey boxplot convention (Frigge et al., 1989) where the end of the whiskers are ± 1.5 times the interquartile range from the box edges). Outliers are plotted as black dots. Each methodology has both historic and future results represented by the blue and red colored boxes in each column, respectively. A heavy line is used to connect the means to provide a visual cue as to the change from historic to future runs.

When examining a condensed information plot such as Fig. 8, it useful to consider what an "ideal" result would be. The y-axis is a RD, meaning that the values are both normalized (unitless) and comparative. A value of 0 on this axis (plotted in the figure as a dotted line) for the historic case would mean a perfect match between the target and the bias corrected data. The second aspect to consider is the width of the boxes (and their corresponding whiskers). This is the sample spread and represents the variance in RD, or how well the BC performed for a variety of test cases (the ensemble). An ideal BC technique would have a width approaching zero, meaning that it is robust to both calibration period (k-fold validation) and distribution shape (seasons).

The first column in each panel of Fig. 8 shows the raw model output. The y-axis value in this case indicates the performance of the wave model itself, when compared to the observational data. For example, in looking at RD for the mean values, it is clear that the model under-predicts Hs (negative value for RD) and as well as Tp (also see Fig. 2). The spread in the raw model results is due to the jack-knife procedure, where only a portion of the buoy data is considered the target at a time. The displayed sample spread therefore indicates the natural variability of the target dataset over time.

Since there is no observation (target) dataset for the future, the future result metric is calculated using the historic observational data for normalization. The differences between the historic and the future run results therefore is an indicator of changes in the wave climate. An ideal BC technique would preserve the relationship between the historic and future runs. Conceptually this means that despite biases in the simulated wave data, the future period is assumed to have a useful signal if interpreted relative to the historic period rather than that of the observed (Wood et al., 2002). If the amount of change is the same for the model column and the bias corrected column, then the BC has successfully maintained the relationship between the historic and future runs.

Looking at the Hs column of Fig. 8, all BC techniques perform consistently across the metrics. Furthermore all methods perform very well with the RD of metrics being close to zero with low variance. Noting the change of y-axis scale, it is seen that all techniques perform decreasingly well (both with matching the target and with sample spread) for increasingly higher moments. This is an expected result as higher moments are more variable depending on calibration period. It may be initially counterintuitive that all BC methods perform so uniformly for Hs but recall that all tested multivariate BC methods require one of the variables to be corrected with a univariate BC as a step in the process. For this study this initial variable was chosen to be Hs so all methods are essentially treating this variable the same.

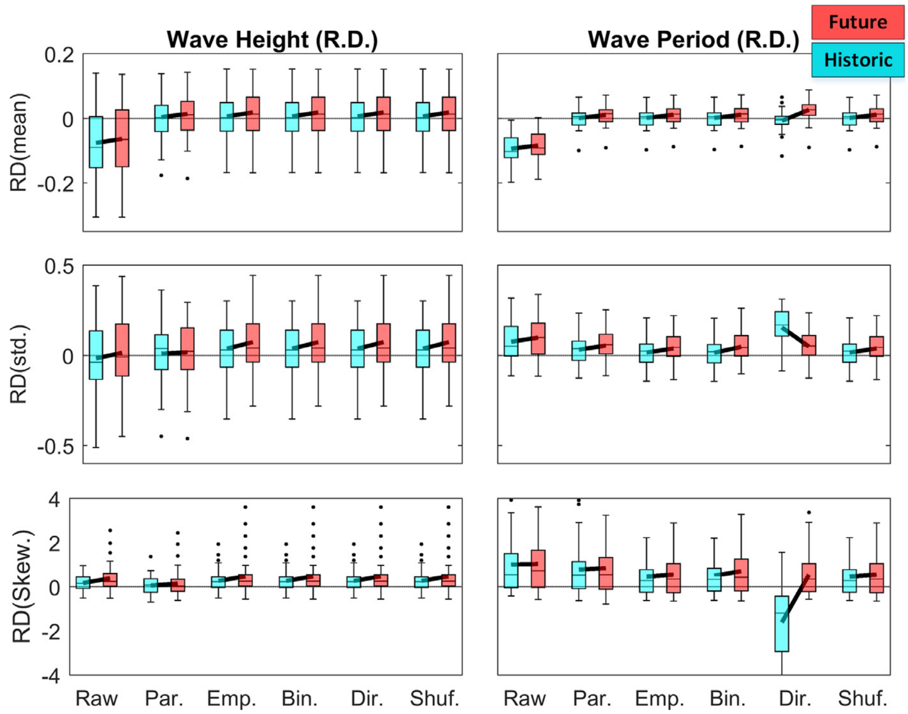

Fig. 8. Bias Correction results for the first three statistical moments: Mean, Standard Deviation, and Skewness. Raw column shows the agreement of raw model output with observations. All data are presented as relative differences (RD) and all boxplots represent an ensemble of 40 data points. Note that some outliers are not visible due to y-axis scaling.

The Tp column of Fig. 8 indicates a similar general level of success in correcting the moments of the Tp distribution. This result is contrary to our hypothesis that the choice of primary BC variable was significant and that the second variable would be less accurately corrected. The only outlier to this generally good performance is the Direct method which did not perform as well for higher statistical moments. It should be noted that the Direct method additionally showed poor performance for mean and standard deviation in some of the other buoys tested (buoy #44025 and #46029).

A sensitivity analysis was performed to explore the significance of Hs being arbitrarily chosen as the primary variable by rerunning a single buoy (#46029) with TP as the primary variable. This test corroborated the above result with no difference found between the means of boxplots run with Hs or Tp as a primary variable.

4.2. Correlation

Fig. 9 shows correlation plotted using the same boxplot conventions as Fig. 8. It is in correlation that multivariate BC techniques would be expected to outperform univariate methods. Fig. 9 confirms this with all univariate procedures having boxplots that have nonzero relative differences, showing that the bias corrected data has the same incorrect correlation as the model itself. This is contrasted with the bivariate methods which all have boxplots shifted towards zero representing an improvement in correlation. Furthermore, this result is shown across all three correlation coefficients hinting at a true correction in correlation, although this will be more accurately accessed by looking at the full bivariate PDF/CDF. In terms of the three bivariate methods, the Binning and Shuffle methods are found to outperform the Direct method. The Shuffle method displays the undesirable characteristic of an inability to preserve the historic-future relationship of correlation (represented by an incorrect bold line slope between historic and future simulation boxes). This is a result of the Shuffle technique methodology which exactly matches both the historic and future periods to the target time series, not allowing for a correct modeling of change in correlation over time.

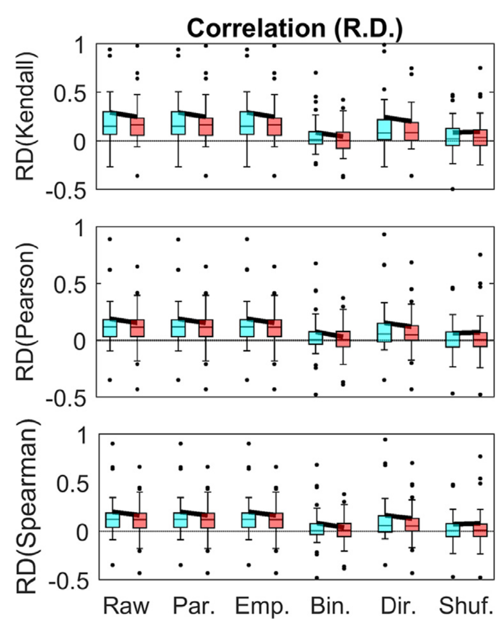

Fig. 9. Bias Correction results for correlation coefficients: Kendall's τ, Spearman's ρ, and Pearson's r. Raw column shows the agreement of raw model output with observations. All data are presented as relative differences and all boxplots represent an ensemble of 40 data points.

4.3. Bivariate PDF/CDF shape and RMSE

While bulk statistics can quickly show the properties of a dataset, analysis of the full distribution provides a more visual

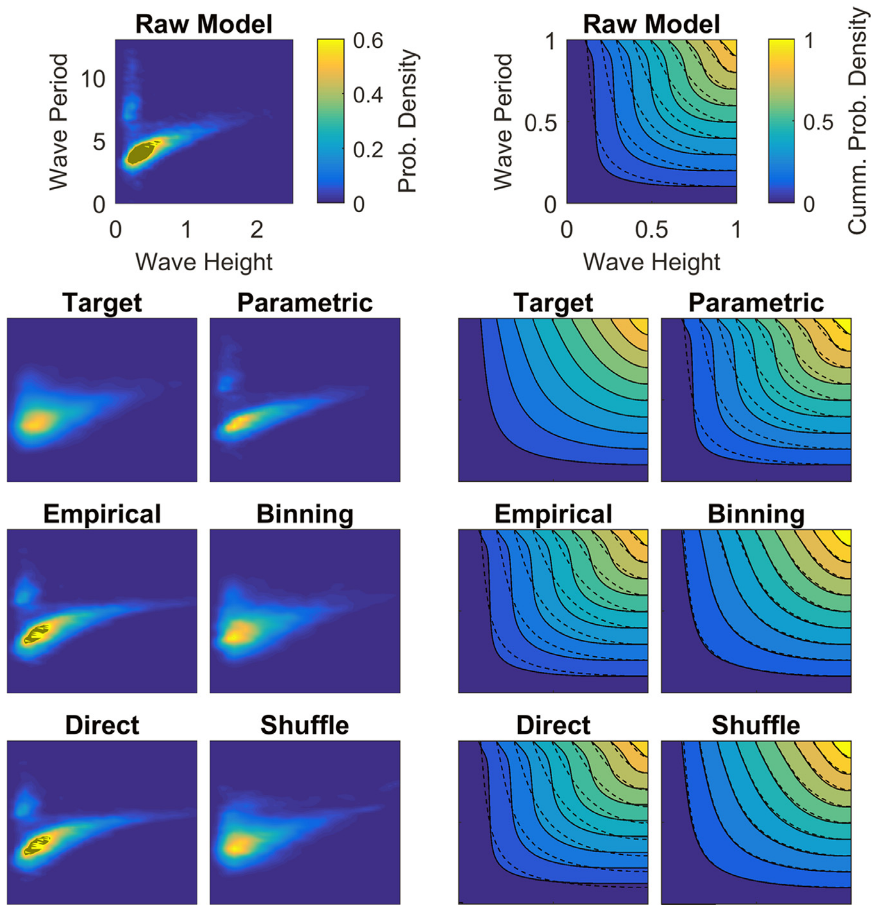

Fig. 10. Comparison of bivariate PDFs (left columns) and CDFs (right columns) for the summer seasons of the historic period. Plotted results are an average of all 10 jack-knife bins. All subplots have the same scale and axes as shown on the Raw Model subplot. The color scale is probability density.

and intuitive understanding of changes. Fig. 10 shows bivariate PDFs and CDFs of all model, target and BC results for the summer seasons of the historic period. The plotted bivariate distributions are an average of the 10 jack-knife runs for the chosen season, each of which has been calculated on a consistent grid for comparison. In terms of the plots types, bivariate PDFs are easier to conceptualize physically (having real units) but are somewhat difficult to compare quantitatively due to noisiness and scale issues. CDF plots are less intuitive but have the advantage of more directly showing the correlation between the variables, a result of removing the marginal distributions.

First considering the bivariate PDFs (displayed as the left columns), Fig. 10 shows that both univariate quantile mapping methods (empirical and parametric) still have the same general shape of the raw model. The distribution has been scaled by the BC (a shrinking or expansion along the x or y axes) but the characteristics of the shape have not been significantly altered. This is contrasted with the bivariate BC methods that not only scale the distribution but also transform the shape into closer alignment with the target. In this case, both the Binning and Shuffle method closely match the target while the Direct method appears to be somewhere in the middle, performing better than the univariate methods but worse than the other bivariate methods. This result is in alignment with what was found by comparing statistical moments and correlation coefficients.

The CDF plots tell a similar story although the removing of the marginal distribution allows us to see that the univariate BC methods are entirely unable to fix correlation. The shape of the model's CDF remains unchanged in each of the univariate methods CDF plots. This is contrasted by the bivariate methods, which for

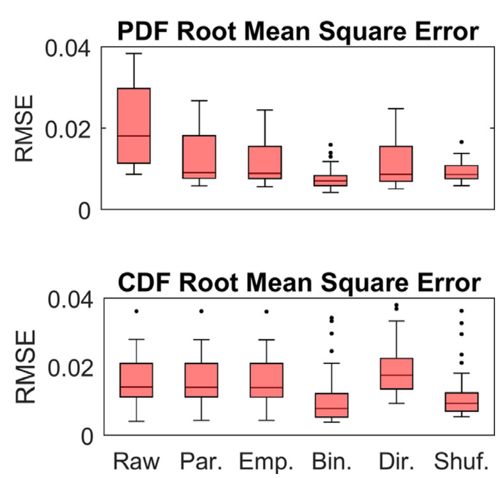

Fig. 11. Root mean square errors for comparison of bivariate PDFs and CDFs.

the Binning and Shuffle methods, match the target to the point that it is difficult to see the dotted line representing the target. The CDF plots show again the direct method is somewhere in between the univariate and bivariate methods.

These plots qualitatively show how the BC process is working but by calculating the RMSE we can attempt to quantitatively compare the methods. Fig. 11 shows this comparison once again as a boxplot with the same format as detailed above. Only historic data are shown since a comparison of the future run to the target would not hold any useful information. The PDF RMSE shows that

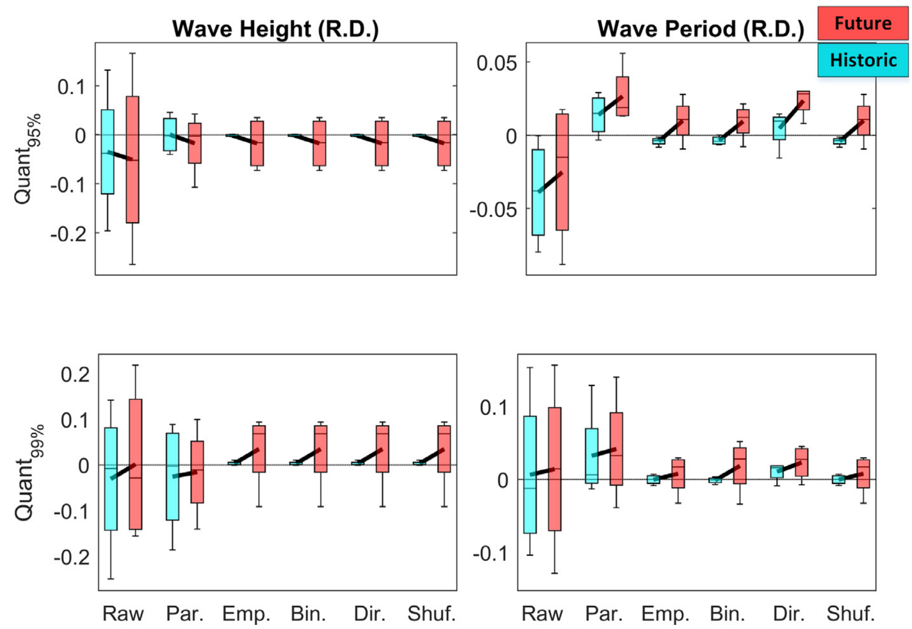

Fig. 12. Bias Correction results for extreme quantiles (95th and 99th). Raw column shows the agreement of raw model output with observations. All data are presented as relative differences (RD) and all boxplots represent an ensemble of 4 data points.

all methods improve the PDF but that the Binning and Shuffle methods perform the best (RMSE closest to zero and the smallest spread). This is because these methods are correcting both the marginal and the shape of the distribution best. Of interest this metric shows the direct method performing quite poorly, at about the same level as the parametric and empirical methods. Upon further investigation this poor performance is season specific and is likely due to the fitted copula not properly describing the correlation structure of the data sufficiently. This said, it should be noted that a direct comparison of bivariate PDFs is a noisy estimate of BC success since values at specific grid cells are very sensitive to scaling.

The CDF RMSE shows the same relationship with the Binning and Shuffle methods outperforming all other methods. Since the marginal distributions are removed, the parametric and empirical methods perform the same as the model. In this case the Direct method is shown to perform worse than the univariate methods and even the raw model. This is evidence that, although the Direct method seems to match correlation coefficients fairly well, it is not properly correcting the full structure of dependence and potentially even making the structure worst in certain cases.

4.4. Extreme quantiles

Fig. 12 shows the results of an upper quantile (95th and 99th percentile) comparison in the same boxplot format as used for Figs. 8 and 9. An important difference is that no jack-knife cross-validation was used to generate the displayed ensemble, only seasonal bins resulting in a total of 4 data points for each method. Extremes are highly dependent on record length and a comparison of longer BC segment to a smaller validation segment would not be valid for this metric. This forced methodology change means that displayed performance is not directly comparable to that seen in Figs. 8 and 9 as the entire record length is used both for bias correction and validation. It would be expected that uncertainty and robustness in this comparison is significantly higher so results should be viewed with caution. Proceeding with this caveat, Fig. 12 shows an improvement of the 95th and 99th percentile for both wave height and wave period. The only outlier to this general improvement is the Parametric method which shows little to no improving for the 99th percentile of wave height and a decrease in accuracy from the model for both the 95th and 99th percentile of wave period.

5. Discussion

This study is intended as an investigation of bias correction applied to wave modeling applications and as a comparative guide to which methods are most applicable in this context. In particular, consideration was given to the relative performances of (1) empirical vs. parametric quantile mapping and (2) univariate vs. bivariate bias correction techniques. The final goal of this paper was, through a rigorous comparison of methods, to identify the best performing method. This section will explore these research questions through statistical testing under the standard framework of the null hypothesis being no difference in the parameter being considered. The reported p-value represents evidence to reject this null hypothesis. For succinctness, this paper will withhold full expositions of the used tests and readers are encouraged to reference a statistics resource as needed.

The first comparison was between parametric and empirical quantile mapping. The literature is somewhat inconclusive in this regard with some studies specifying a loss of robustness due to empirical methods and others a loss of accuracy due to parametric assumptions. This study found little evidence of a gain of accuracy from using the Empirical method nor a gain in robustness from using the Parametric method. This was statistically verified with a paired *t*-test for statistical moments and correlation parameters. For wave height, the test showed an inability to reject the null hypothesis of no difference (p-values $>.03$) for both mean and standard deviation while wave period was unable to reject the null for mean but did show a statistical difference for standard deviation (p-value $<2\times10^{-15}$). That said, the relative difference was very small (95% confidence interval of .0125 to .0187) representing a statistically significant but practically non-relevant difference between methods. This result was further corroborated by the RMSE comparison of bivariate CDF and PDF distributions which showed no noticeable difference between the two methods. Finally, both wave height and period visually showed a less effective correction of extreme events but this result was not found to be statistically significant (p-values >.06). Overall this result of no difference in any metric could either be because the distributions at the buoy locations chosen were well enough fit that the error introduced was negligible or that quantile mapping is insensitive to a well fitted distribution. A goodness of fit test failed at the 95% significance level for the vast majority of cases but this may be because the data record was very long providing strict error bars on the statistical test.

The next comparison was between univariate and multivariate BC techniques. Using a standard one-way analysis of variance (ANOVA) test, it was found that, in terms of statistical moments, there was no evidence of a difference between univariate and bivariate techniques (1 way ANOVA, p-value > .5) for wave height or wave period. This result was somewhat surprising as it was hypothesized that all bivariate methods would perform differently for the second variable due to the additional constraint of matching correlation. When looking at correlation it was found that there was evidence that bivariate techniques improved incorrect model correlations in comparison to univariate methods (1-way ANOVA, p-value ≈ .07). Of note this larger p-value is a result of the model correlation being only moderately incorrect, making the BC correction smaller and therefore less significant statistically. Evidence of the bivariate methods effectiveness was further solidified qualitatively by looking at bivariate PDF shapes and quantitatively by looking at the RMSE of bivariate CDFs. No difference in high percentile corrections were found between univariate and bivariate methods.

As the preservation of the dynamic link between wave parameters (in this case Hs and Tp) is critical for realistic modeling, this result strongly supports the use of bivariate techniques over univariate. While this study looks only at bivariate BC, other correlated parameters (e.g. wave direction) could be included in the BC process. The extension of some copulas (for the Direct method) to multiple dimensions is straightforward and it is hypothesized that the Binning method could also be generalized to an N-dimensional case. As the Shuffle method involves a direct reordering according to the target time series, it is already a multivariate BC technique that would only require keeping track of another variable.

Since bivariate BC was concluded to be superior to univariate, the determination of a “best practice” methodology was limited to a comparison of the three tested multivariate techniques. In terms of statistical moments, strong evidence was found that the Direct method performed less well at correcting the standard deviation of the second variable than the other two bivariate techniques (Multiple Comparison Test, Tukey-Kramer adjustment, p-value $<1 \times 10^{-9}$) as well as higher moments. This result was even more distinct in other buoys (#44025 and #46029), showing a significant incorrect matching of all statistical moments for the second variable. While visually the Direct method appear to not correct correlation as well as other bivariate methods, this result was not found to be statistically significant (Multiple Comparison Test, Tukey-Kramer adjustment, p-value > .5). This result was shared by the correction of extreme quantiles with the Direct method appearing to be less effective but not to the point of being statistically significant (Multiple Comparison Test, Tukey-Kramer adjustment, p-value > .3). A comparison of CDF RMSEs showed the Direct method only slightly outperforming the univariate methods and significantly underperforming in relation to the Shuffle and Binning methods. This result hints at the importance of not just looking at correlation coefficients (which only measure a small subspace of the many ways variables can be related) to measure dependence. Piani and Haerter (2012) found a similar result and concluded that a “full appreciation” of the benefits of a bivariate methodology could not be explained by correlation coefficients alone. We hypothesize that the comparatively poor performance of the Direct method may be a result of the fitted copula not being able to adequately model the correlation between the two variables. It is possible that consideration of a larger number of candidate copulas for fitting could alleviate this problem. The other two methods, the Binning and Shuffle techniques, performed equally well at correcting statistical moments and correlations. This said, the Shuffle technique has the additional requirement of an exact match of sampling rate and record length between the target, model historic and model future time series, a somewhat restrictive condition.

This study also considered the performance of the tested BC methods by examining how well they preserved the non-stationary change from historic to future model runs. It was found that all methods performed well for the first variable (Hs) but that for the second (Tp) the Direct method struggled at reproducing changes for higher moments. In terms of correlation all methods preserved changes well except for the Shuffle technique which is unable to account for changes in correlation. The Shuffle technique also suffers from the philosophical issue that both the historic and future runs have a temporally identical time series (only scaling of variable magnitudes differ).

With current scientific focus on quantifying uncertainty, many wave modeling efforts have evolved into ensemble based frameworks requiring many simulations. This approach brings additional scrutiny to the computational burden of each step in the modeling process. A full computational analysis of the methodologies explored in this paper is beyond the scope of this study but a broad comparative investigation was explored. Looking at the time to correct one season of 30 years of data, Parametric corrections are very fast (order of 10′s of seconds). Empirical methods take longer (order of 10′s of minutes) but still not to the point of being significant in relation to the computational time of model runs. Of Empirical methods, the Direct method is slowest and the Binning method is fastest but with the difference being in the order of minutes. These results are presented cautiously as the magnitude of this comparison is undoubtedly variable dependent on the implementation and system the code is being run on. Overall the computational time of BC was found to be minimal and not a driver of the choice between methodologies or the use of BC in general.

6. Conclusions

In summary, wave forcing is a key component in the accurate modeling of ocean and coastal systems. Oftentimes a direct wave model application will result in biased output due to a variety of factors ranging from errors within the wave model itself to biases within the input meteorological fields (especially with AOGCM derived products). Biases in wave model output can affect the entire system being considered making BC oftentimes requisite for bringing results back into agreement with reality. This said, BC should be used only after noting the various limitations and assumptions ranging from a masking of uncertainty to the potential introduction of physical inconsistencies between model output fields. While BC of model input data is widely accepted, this study investigates the alternative approach of a direct BC to the wave model output itself. In this context, a subset of BC techniques are compared using a statistically robust cross-validation technique in order to investigate which methodologies are most effective.

Key findings of this study were that: (i) the BC of wave model output is an effective tool for bringing wave model output into alignment with observations; (ii) no statistical difference was found between univariate empirical and parametric methods; (iii) bivariate methods were shown to be superior to univariate methods due to improvement in matching correlation; and, (iv) the empirical binning method of Piani and Haerter (2012) (extended in this study to apply to future projection model runs) proved to be the best overall approach across the various metrics considered.

Acknowledgments

The first author conducted this study with support from Oregon Sea Grant (Grant No. NA150AR4170224) as well as the Malouf Scholarship. Further financial support was provided by the ARCS foundation Portland. The data for this study was provided by Mark Hemer at the Commonwealth Scientific and Industrial Research Organization (CSIRO). The utilized data can be found on the CSIRO server at: http://doi.org/10.4225/08/55C991CC3F0E8.

Appendix A

The Binning method (future)

The Binning method as extended to the future case is proposed as procedurally a hybrid of the equidistant quantile mapping method (Fig. 4) and the standard Binning method (Fig. 5). The procedure is illustrated graphically in Fig. 13. The first step is to bias correct the first variable (X) using a standard equidistant quantile mapping. This step is not shown in Fig. 13 but would be identical to the process outlined in Fig. 4. The second step **(2)** is to select a bin of the bias corrected first variable (X) from the target, model historic, and model future datasets. Fig. 13 shows the selected decile bin (CDF_X values ranging from .2–.3) in red. The third step is to determine CDFs from these bins of data **(3)**. With these three bin dependent CDFs it is possible to perform a standard equidistant quantile mapping. This is done by first finding the quantile value corresponding to the model future value for variable (Y) **(4)**. The target value and model historic value are found for this quantile and the difference between these two values is quantified **(5)**. This difference is then applied to the model historic value so as to find the bias corrected model future value **(6)**.

The Direct method (future)

The future extension for the Direct method is detailed in Fig. 14. The overarching goal of the methodology is to find the components necessary for an equidistant quantile mapping: the quantile of the model historic and target datasets (associated by correlation) so that we can apply the difference to the model future quantile of interest. These two quantiles are found using two standard Direct methods applied to their respective datasets. The first step is to bias correct the first variable (X) using a standard equidistant quantile mapping. This step is not shown in Fig. 14. The second step **(2)** is to fit a copula to all three of the model future, model historic, and target datasets. This step is performed only once while all following steps are repeated for each value in the model future dataset. The next step is to find the bivariate CDF value for the model future copula corresponding the CDF value for model future variable X and Y **(3)**. This is done by finding the quantiles corresponding to the model future value pair and evaluating the copula at this point. At this point the two Direct methods diverge with the first branch using the target copula and the second branch using the model historic copula. This divergence corresponds to the goal of finding the quantile of model historic and target datasets. With these two copulas, the next step is to take a slice of the bivariate CDF at the known bias corrected value (variable X). As with the normal direct method (see Fig. 6), the value of CDF_X is already known since it is the bias corrected variable X (step 1). If we slice the copula at this value **(4)** and use our knowledge of the bivariate copula CDF value of the model (evaluated in step **(3)**), we can find

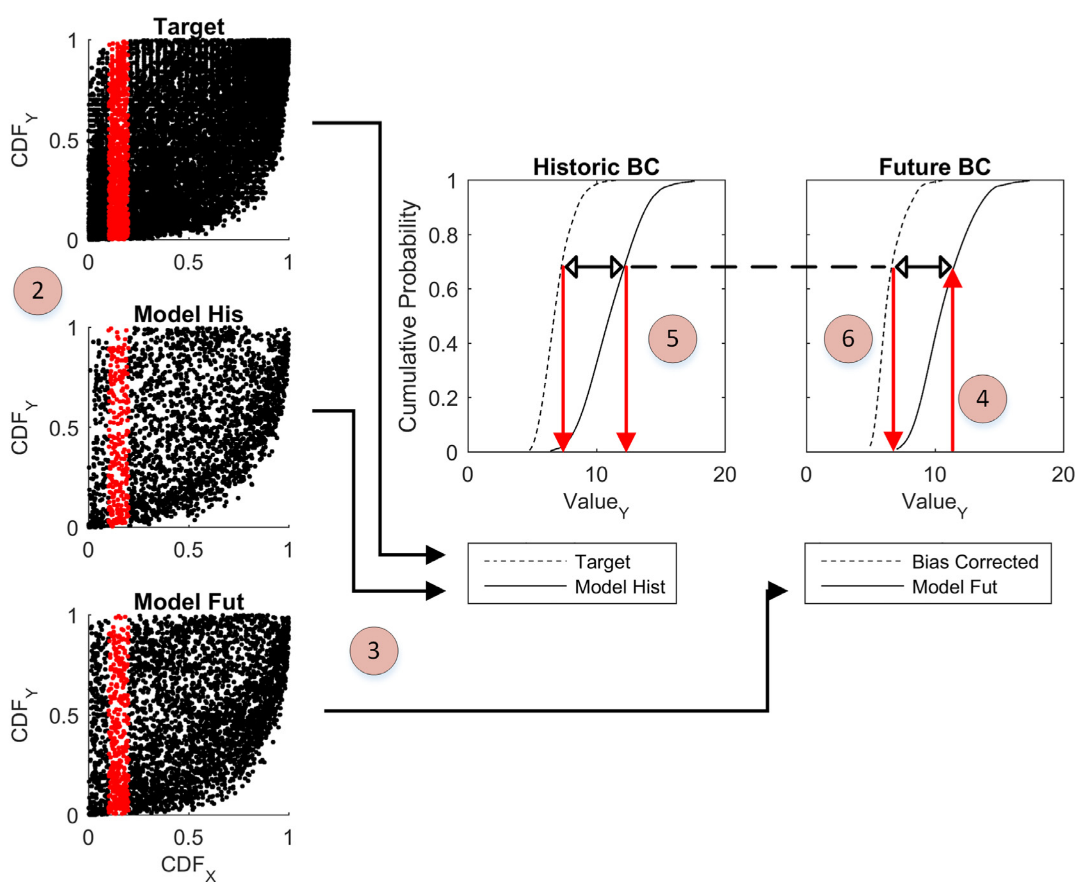

Fig. 13. Binning Method using equidistant quantile mapping.

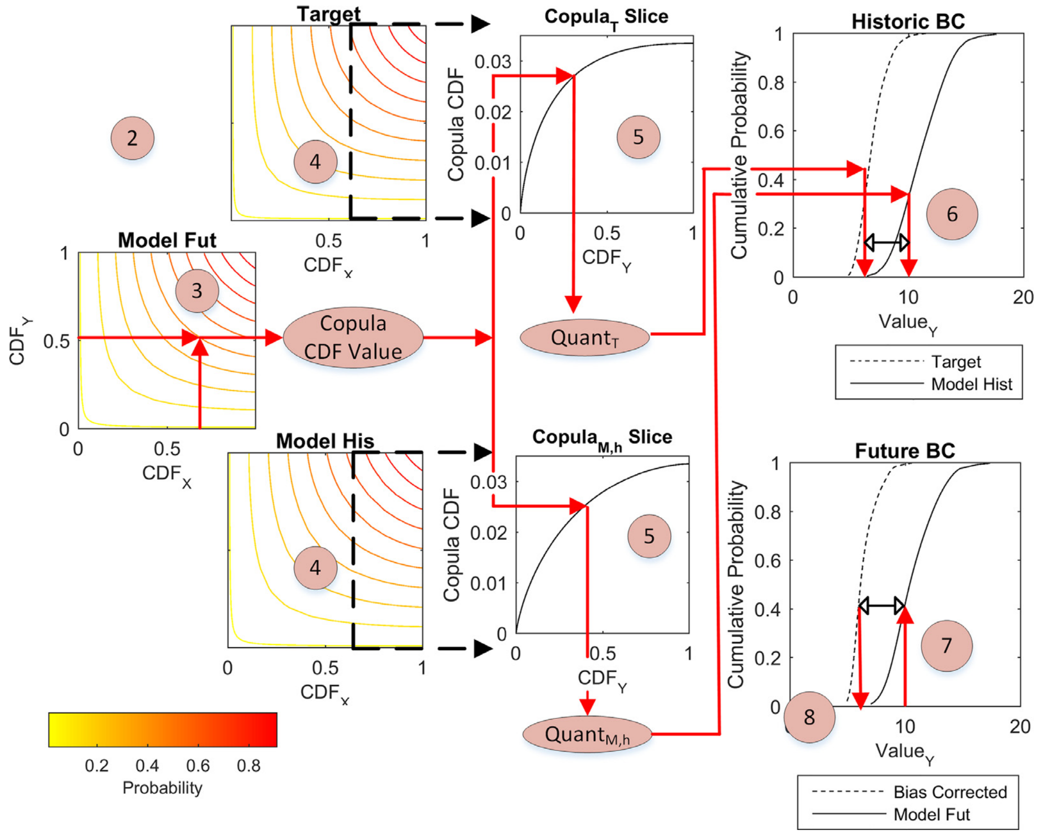

Fig. 14. Direct Method using equidistant quantile mapping.

CDF$_Y$ **(5)**. At this point we found the quantiles of the model historic and target datasets associated with the model future Copula value. With these values, a normal equidistant quantile mapping method is performed by finding the Value$_Y$ corresponding to these quantiles and quantifying the difference between these values **(6)**. Finally, we find the quantile of the model future value **(7)** and apply this difference to find the bias corrected value **(8)**.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ocemod.2016.12.008.

References

Anderson, T.R., Frazer, L.N., Fletcher, C.H., 2010. Transient and persistent shoreline change from a storm. Geophys. Res. Lett. 37. doi:10.1029/2009GL042252, n/a–n/a.

Andrade, C., Pires, H.O., Taborda, R., Freitas, M.D.C., 2007. Projecting future changes in wave climate and coastal response in Portugal by the end of the 21st century. J. Coast. Res. SI 50, 257–263.

Bauer, E., Staabs, C., 1998. Statistical properties of global significant wave heights and their use for validation. J. Geophys. Res. Ocean. 103, 1153–1166. doi:10.1029/97JC02568.

Bowman, A.W., Azzalini, A., 1997. Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations, Vol. 18. OUP Oxford.

Caires, S., Sterl, A., Bidlot, J.-R., Graham, N., Swail, V., 2004. Intercomparison of different wind–wave reanalyses. J. Clim. 17, 1893–1913. doi:10.1175/1520-0442(2004)017⟨1893:IODWR⟩2.0.CO;2.

Caires, S., Sterl, A., 2005. A new nonparametric method to correct model data: application to significant wave height from the ERA-40 Re-Analysis. J. Atmos. Ocean. Technol. 22, 443–459. doi:10.1175/JTECH1707.1.

Cavaleri, L., Sclavo, M., 2006. The calibration of wind and wave model data in the Mediterranean Sea. Coast. Eng. 53, 613–627. doi:10.1016/j.coastaleng.2005.12.006.

Cardone, V.J., Jensen, R.E., Resio, D.T., Swail, V.R., Cox, A.T., Cardone, V.J., Cox, A.T., 1996. Evaluation of contemporary ocean wave models in rare extreme events: the “Halloween Storm” of October 1991 and the “Storm of the Century” of March 1993. J. Atmosp. Oceanic Technol. 13 (1), 198–230 http://doi.org/10.1175/1520-0426(1996)013⟨0198:EOCOWM⟩2.0.CO;2.

Charles, E., Idier, D., Delecluse, P., Déqué, M., Le Cozannet, G., 2012. Climate change impact on waves in the Bay of Biscay, France.. Ocean Dyn 62, 831–848. doi:10.1007/s10236-012-0534-8.

Chen, C., Haerter, J.O., Hagemann, S., Piani, C., 2011. On the contribution of statistical bias correction to the uncertainty in the projected hydrological cycle. Geophys. Res. Lett. 38. doi:10.1029/2011GL049318, n/a–n/a.

Christensen, J.H., Boberg, F., Christensen, O.B., Lucas-Picher, P., 2008. On the need for bias correction of regional climate change projections of temperature and precipitation. Geophys. Res. Lett. 35, L20709. doi:10.1029/2008GL035694.

Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., Wilby, R., 2004. The schaake shuffle: a method for reconstructing space–time variability in forecasted precipitation and temperature fields. J. Hydrometeorol. 5, 243–262. doi:10.1175/1525-7541(2004)005⟨0243:TSSAMF⟩2.0.CO;2.

Corbella, S., Stretch, D.D., 2012. Predicting coastal erosion trends using non-stationary statistics and process-based models. Coast. Eng. 70, 40–49. doi:10.1016/j.coastaleng.2012.06.004.

De Michele, C., Salvadori, G., Passoni, G., Vezzoli, R., 2007. A multivariate model of sea storms using copulas. Coast. Eng. 54, 734–751. doi:10.1016/j.coastaleng.2007.05.007.

De Waal, D.J., van Gelder, P.H.A.J.M., 2006. Modelling of extreme wave heights and periods through copulas. Extremes 8, 345–356. doi:10.1007/s10687-006-0006-y.

Déqué, M., 2007. Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: model results and statistical correction according to observed values. Glob. Planet. Change 57, 16–26. doi:10.1016/j.gloplacha.2006.11.030.

Durrant, T.H., Greenslade, D.J.M., Simmonds, I., 2013. The effect of statistical wind corrections on global wave forecasts. Ocean Model 70, 116–131. doi:10.1016/j.ocemod.2012.10.006.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., Liebert, J., 2012. HESS opinions “Should we apply bias correction to global and regional climate model data? Hydrol. Earth Syst. Sci. 16, 3391–3404. doi:10.5194/hess-16-3391-2012.

Ferreira, J.A., Guedes Soares, C., 2002. Modelling bivariate distributions of significant wave height and mean wave period. Appl. Ocean Res. 24, 31–45. doi:10.1016/S0141-1187(02)00006-8.

Ferreira, J.A., Guedes Soares, C., 2000. Modelling distributions of significant wave height. Coast. Eng. 40, 361–374. doi:10.1016/S0378-3839(00)00018-1.

Frigge, M., Hoaglin, D.C., Iglewicz, B., 1989. Some Implementations of the Boxplot. Am. Stat. 43, 50–54. doi:10.1080/00031305.1989.10475612.

Grabemann, I., Weisse, R., 2008. Climate change impact on extreme wave conditions in the North Sea: an ensemble study. Ocean Dyn 58, 199–212. doi:10.1007/s10236-008-0141-x.

Haerter, J.O., Hagemann, S., Moseley, C., Piani, C., 2011. Climate model bias correction and the role of timescales. Hydrol. Earth Syst. Sci. 15, 1065–1079. doi:10.5194/hess-15-1065-2011.

Hemer, M.A., Fan, Y., Mori, N., Semedo, A., Wang, X.L., 2013. Projected changes in wave climate from a multi-model ensemble. Nat. Clim. Chang. 3, 471–476. doi:10.1038/nclimate1791.

Hemer, M.A., Katzfey, J., Trenham, C.E., 2012. Global dynamical projections of surface ocean wave climate for a future high greenhouse gas emission scenario. Ocean Model. 70, 221–245. doi:10.1016/j.ocemod.2012.09.008.

Hemer, M.A., McInnes, K.L., Ranasinghe, R., 2011. Climate and variability bias adjustment of climate model-derived winds for a southeast Australian dynamical wave model. Ocean Dyn. 62, 87–104. doi:10.1007/s10236-011-0486-4.

Hemer, Mark; Trenham, Claire; Durrant, Tom; Greenslade, Diana, 2015. CAWCR Global wind-wave 21st centry climate projections. V1. CSIRO. Data Collection. http://doi.org/10.4225/08/55C3F0E8.

Holthuijsen, L.H., 2007. Waves in Oceanic and Coastal Waters. Cambridge University Press, Cambridge.

Johnson, F., Sharma, A., 2012. A nesting model for bias correction of variability at multiple time scales in general circulation model precipitation simulations. Water Resour. Res. 48. doi:10.1029/2011WR010464, W01504.

Lafon, T., Dadson, S., Buys, G., Prudhomme, C., 2013. Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods. Int. J. Climatol. 33, 1367–1381. doi:10.1002/joc.3518.

Leake, J., Wolf, J., Lowe, J., Stansby, P., Jacoub, G., Nichols, R., Mokrech, M., Nicholson-Cole, S., Walkden, M., Watkinson, A., Hanson, S., 2007. Predicted wave climate for the UK: towards an integrated model of coastal impacts of climate change (ASCE). Estuary Coast. Model 393–406. doi:10.1061/40990(324)24.

Li, C., Sinha, E., Horton, D.E., Diffenbaugh, N.S., Michalak, A.M., 2014. Joint bias correction of temperature and precipitation in climate model simulations. J. Geophys. Res. Atmos 119 (13), 13–153. doi:10.1002/2014JD022514, 162.

Li, H., Sheffield, J., Wood, E.F., 2010. Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. J. Geophys. Res. 115. doi:10.1029/2009JD012882, D10101.

Lionello, P., Cogo, S., Galati, M.B., Sanna, A., 2008. The Mediterranean surface wave climate inferred from future scenario simulations. Glob. Planet. Change 63, 152–162. doi:10.1016/j.gloplacha.2008.03.004.

Mathisen, J., Bitner-Gregersen, E., 1990. Joint distributions for significant wave height and wave zero-up-crossing period. Appl. Ocean Res. 12, 93–103. doi:10.1016/S0141-1187(05)80033-1.

Mehran, A., AghaKouchak, A., Phillips, T.J., 2014. Evaluation of CMIP5 continental precipitation simulations relative to satellite-based gauge-adjusted observations. J. Geophys. Res. Atmos. 119, 1695–1707. doi:10.1002/2013JD021152.

Mori, N., Yasuda, T., Mase, H., Tom, T., Oku, Y., 2010. Projection of extreme wave climate change under global warming. Hydrol. Res. Lett. 4, 15–19. doi:10.3178/hrl.4.15.

Mueller, B., Seneviratne, S.I., 2014. Systematic land climate and evapotranspiration biases in CMIP5 simulations. Geophys. Res. Lett. 41, 128–134. doi:10.1002/2013GL058055.

Panofsky, H.A., Brier, G.W., Best, W.H., 1958. Some application of statistics to meteorology.

Piani, C., Haerter, J.O., 2012. Two dimensional bias correction of temperature and precipitation copulas in climate models. Geophys. Res. Lett. 39. doi:10.1029/2012GL053839, n/a–n/a.

Piani, C., Haerter, J.O., Coppola, E., 2010. Statistical bias correction for daily precipitation in regional climate models over Europe. Theor. Appl. Climatol. 99, 187–192. doi:10.1007/s00704-009-0134-9.

Repko, A., Van Gelder, P.H.A.J.M., Voortman, H.G., Vrijling, J.K., 2004. Bivariate description of offshore wave conditions with physics-based extreme value statistics. Appl. Ocean Res. 26, 162–170. doi:10.1016/j.apor.2004.10.001.

Rogers, W.E., Wittmann, P.A., 2002. Naval Research Lab Tech, Report. NRL/MR/732-02-8290.

Rogers, W.E., Wittmann, P.A., Wang, D.W.C., Clancy, R.M., Hsu, Y.L., 2005. Evaluations of global wave prediction at the fleet numerical meteorology and oceanography center*. Weather Forecasting 20 (5), 745–760. http://doi.org/10.1175/WAF882.1.

Saha, S., et al., 2010. NCEP climate forecast system reanalysis (CFSR) selected hourly time-series products, January 1979 to December 2010. Research data archive at the National Center for Atmospheric Research. Comput. Inf. Syst. Lab.. http://dx.doi.org/10.5065/D6513W89.

Teutschbein, C., Seibert, J., 2013. Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions? Hydrol. Earth Syst. Sci. 17, 5061–5077. doi:10.5194/hess-17-5061-2013.

Themeßl, M.J., Gobiet, A., Leuprecht, A., 2011. Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. Int. J. Climatol. 31, 1530–1544. doi:10.1002/joc.2168.

Thrasher, B., Maurer, E.P., McKellar, C., Duffy, P.B., 2012. Technical note: bias correcting climate model simulated daily temperature extremes with quantile mapping. Hydrol. Earth Syst. Sci. 16, 3309–3314. doi:10.5194/hess-16-3309-2012.

Tolman, H.L., 1991. A third-generation model for wind waves on slowly varying, unsteady, and inhomogeneous depths and currents. J. Phys. Oceanogr. 21, 782–797. doi:10.1175/1520-0485(1991)021⟨0782:ATGMFW⟩2.0.CO;2.

Tolman, H.L., 2009. NOAA/NWS/NCEP/MMAB Technical Note 276, p. 194.

Tomas, A., Mendez, F.Z., Losada, I.J., 2008. A method for spatial calibration of wave hindcast data bases. Cont. Shelf Res. 28, 391–398. doi:10.1016/j.csr.2007.09.009.

Vrac, M., Friederichs, P., 2014. Multivariate—intervariable, spatial, and temporal—bias correction. J. Clim. 28, 218–237. doi:10.1175/JCLI-14-00059.1.

Wang, X.L., Feng, Y., Swail, V.R., 2014. Changes in global ocean wave heights as projected using multimodel CMIP5 simulations. Geophys. Res. Lett. 41, 1026–1034. doi:10.1002/2013GL058650.

Wang, X.L., Swail, V.R., 2002. Trends of atlantic wave extremes as simulated in a 40-Yr wave hindcast using kinematically reanalyzed wind fields. J. Clim. 15, 1020–1035. doi:10.1175/1520-0442(2002)015⟨1020:TOAWEA⟩2.0.CO;2.

Wang, X.L., Swail, V.R., Cox, A., 2010. Dynamical versus statistical downscaling methods for ocean wave heights. Int. J. Climatol doi:10.1002/joc.1899, n/a–n/a.

Wood, A.W., 2002. Long-range experimental hydrologic forecasting for the eastern United States. J. Geophys. Res. 107, 4429. doi:10.1029/2001JD000659.

Wood, A.W., Leung, L.R., Sridhar, V., Lettenmaier, D.P., 2004. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. Clim. Change 62, 189–216. doi:10.1023/B:CLIM.0000013685.99609.9e.

Xu, C.-y., 1999. From GCMs to river flow: a review of downscaling methods and hydrologic modelling approaches. Prog. Phys. Geogr. 23, 229–249. doi:10.1177/030913339902300204.

Yukimoto, S., Adachi, Y., Hosaka, M., Sakami, T., Yoshimura, H., Hirabara, M., Tanaka, T.Y., Shindo, E., Tsujino, H., Makoto, D., Mmizuta, R., Yabu, S., Obata, A., Nakano, H., Koshiro, T., Ose, T., Kitoh, A., 2012. A new global climate model of the meteorological research institute: MRI-CGCM3; Model description and basic performance. J. Meteorol. Soc. Japan 90A, 23–64. doi:10.2151/jmsj.2012-A02.

Zhang, K., Douglas, B., Leatherman, S., 2002. Do storms cause long term beach erosion along the U.S. East Barrier Coast. J. Geol 110, 493–502. doi:10.1086/340633.