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/

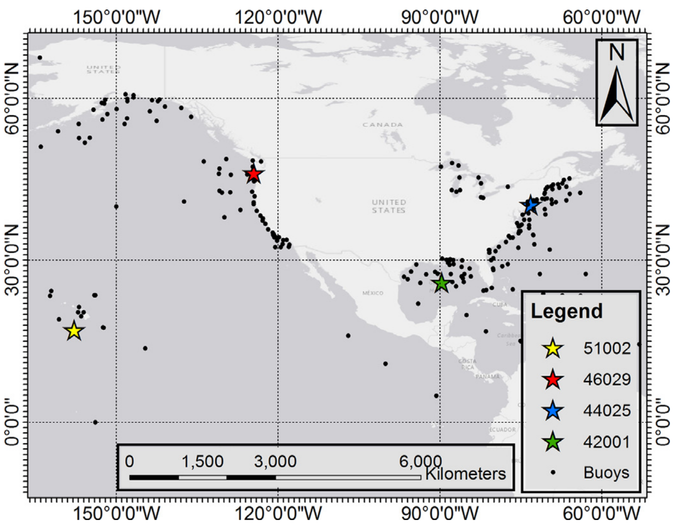

Fig. 1. Site map showing the buoy locations used for bias correction marked as stars and all other buoys (both historic and active) marked with dots. The external tick marks represent the WW3 model 1 degree resolution grid.

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.