Evaluation of a CMIP5 derived dynamical global wind wave climate model ensemble

Mark A. Hemer^{1,*} and Claire E. Trenham^{1,2}.

¹CSIRO Oceans and Atmosphere Flagship, Hobart, Tasmania, Australia.

²Now at: National Computational Infrastructure, Australian National University, Canberra, ACT, Australia

*Corresponding Author: mark.hemer@csiro.au

Abstract

Much effort has gone into evaluating the skill of General Circulation Models (GCMs) for 'standard' climate variables such as surface (air and/or sea) temperature, or precipitation. Whether climate model skill to simulate standard variables translates to the performance of dynamical GCM forced wind-wave simulations is yet to be established.

We assess an ensemble of historical dynamical wave climate simulations whereby surface winds taken from GCMs participating in the Coupled Model Intercomparison Project (CMIP) are used to force a spectral wave model. The GCMs used include 8 CMIP5 models and 2 dynamically downscaled CMIP3 models.

The climatological properties of key integrated wave parameters (significant wave height, maximum wave height, mean wave period and direction) are evaluated, using two independent methods, relative to three historical wave hindcast/reanalysis datasets over 13 areas of the global ocean.

We identify that high performance of GCMs for 'standard' climate variables does not imply high performance for GCM forced wave simulations. We also identify there is little to no benefit in choosing a higher resolution CMIP5 GCM (with resolution of ~1.4degrees) over a lower resolution GCM (~2.8 degree) to improve skill of GCM forced dynamical wave simulations. With the conscious push towards developing projections of waves and storm surges to aid assessments of possible climate driven impacts to coastal communities, we stress the need to evaluate the performance of a

GCM for the marine meteorological climate independently of the performance of the GCM for the 'standard' climate variables.

1. Introduction

The potential coastal and near-shore impacts of climate change driven variations of wind-wave properties have gathered increased attention in recent years (Coelho et al., 2009; Zacharioudaki and Reeve, 2011; Adams et al., 2011; Plecha et al., 2012; Barnard et al., 2014; Sierra and Casas-Prat, 2014). As a consequence, researchers have increasingly taken results from global climate models (GCMs) participating in the World Climate Research Programme's Coupled Model Intercomparison Project phases 3 (CMIP3) and 5 (CMIP5) to infer global or regional wind-wave climate changes, using both statistical (e.g., Wang et al., 2014; Casas-Prat et al., 2014; Laugel et al., 2014) and dynamical approaches (e.g., Dobrynin et al., 2012; Hemer et al., 2013a; Laugel et al., 2014; Fan et al., 2014). The predominant focus of prior studies has been on the uncertainty of the projected 21st Century change signal (e.g., Wang et al., 2014; Hemer et al., 2013b), yet a critical part of evaluating climate projections is testing models' ability to represent the observed climate. To date, the attention given to the ability of dynamical GCM forced wave simulations to represent historical wave fields has been limited, leaving un-answered questions regarding the model uncertainties, deficiencies, and the processes responsible for these. The statistical approaches offer advantages in assessing the uncertainties within a larger ensemble of GCMs, owing to their lower computational demand. The advantage of dynamical studies is that they may better resolve the response of the multi-modal characteristics of the global simulated wave field in a spectral model to varied climate forcings. The dynamical approach has a high computational expense, and consequently, prior studies have considered uncertainties associated with just a select few GCMs. How these forcing GCMs are chosen is a focus of this paper.

The CMIP5 archives include climate scenarios from approximately 60 different GCMs contributed by 28 different international climate modelling groups, and for many GCMs, multiple ensembles are available to investigate intra-model variance. Ideally, wave climate modelling studies would quantify the total inter- and intra-model variance by carrying out wave model simulations with forcing taken from all available GCM simulations. However this is computationally prohibitive for dynamical wave climate studies. Consequently, a subset of GCMs must be chosen. To date, the choice of parent GCM(s) has been largely un-objective. Wave model forcing is either taken from the GCM(s) which is most easily accessed for the study (e.g., Dobrynin et al., 2012, used the EC-EARTH simulations most

easily accessed to this European research group, and Erikson et al., 2015, used four GCMs which were some of the first available models from the CMIP5 Earth System Grid – Centre for Enabling Technologies - ESG-CET, <http://pcmdi9.llnl.gov>, or taken from a regional climate model, where the parent GCM has been predetermined via the downscaling research project (e.g., in global studies such as Mori et al., 2010, Hemer et al., 2013a, Semedo et al., 2013, or regional studies such as Lionello et al., 2008 or Grabemann et al., 2015). A more objective approach to select GCMs for wave climate studies would be to first assess the skill of all GCMs within the CMIP5 ensemble to represent the surface marine wind climatology (being the main forcing variable of wave simulations). A subset of GCMs could then be chosen, either choosing the best performing models, or those which are representative of the range of the complete ensemble, depending on the study objectives. However very few prior studies have assessed the skill of marine winds from the CMIP5 models, and those that have, only a limited set of models have been investigated (e.g., Lee et al., 2013).

Substantially more effort has been put into evaluating the skill of GCMs to represent more commonly considered variables (e.g., Flato et al., 2013). These common variables include surface air temperature – tas, precipitation – pr, and sea-surface temperature – tos, which we will refer to as ‘standard’ climate variables hereafter. This pre-existing body of knowledge of GCM skill provides a potential basis from which an objective choice of GCMs for future wave climate studies could be based. In this study, we aim to determine whether the performance of a GCM (for standard variables, amongst a limited sub-set of models, pre-chosen on the basis of early availability in the ESG-CET) can be used as a proxy for the relative performance of a wave climate simulation forced with that GCM.

Further to these primary objectives, the evaluation of GCM forced wave climate simulations are beneficial for other key research questions. Model evaluation enables researchers to identify the processes which diminish model skill and ultimately improve these for future studies. Furthermore, by providing information on the performance of GCM forced wave climate simulations, the coastal impacts community are better informed when using derived wave projections data in coastal impacts of climate change assessments. For example, users may choose to weight future projection scenarios on the basis of skill of the model in simulating historical conditions. e.g., a model which has poor skill in reproducing a particular historical aspect of the climate system is reasoned to be less likely to produce a realistic representation of that aspect in the future climate, and is given less weight. Model culling being the extreme example of such an approach (e.g., Giorgi and Mearns, 2002). Alternatively, users might adopt a model democracy approach (e.g., Knutti, 2010), where all

simulations are considered equally valid representations of the future regardless of their representation of historical conditions. Our goal is to evaluate the performance of the GCM forced wave models to provide guidance for those using the outputs of these simulations, and to assist other research groups who are choosing CMIP5 models for developing wind-wave or marine-meteorological projections. It is not the intention of this study to guide how the impacts community use this information.

At a practical level, data archived as part of CMIP5 offers a distinct advantage over CMIP3 for wave climate studies. High temporal (3 or 6-hourly) resolution surface wind (defined by directional components u_{as} and v_{as}) and surface pressure (p_{sl}) data are archived from each model for defined time-slices. The data is stored from historical runs (1980-2005), and for 21st Century Representative Concentration Pathway (RCP) scenario experiments (2025-2045 and 2080-2100). CMIP3 data were not archived with sufficient temporal resolution, and thus CMIP3 generation GCM forced dynamical wave climate simulations could not be carried out. This was overcome by either using a high resolution GCM run independently of CMIP to force a wave model (e.g., Semedo et al., 2013), or using a high resolution atmospheric model that had dynamically downscaled the CMIP3 simulations to force a wave model (e.g., Mori et al., 2010, Hemer et al., 2013a). While giving very high resolution forcing to improve representation of reduced scale processes, the different downscaling models introduced another level of model uncertainty which was difficult to quantify (and a focus of study through the CORDEX project; Georgi et al., 2009). In practice, the CMIP5 experiments allow this step to be avoided, but how such direct forcing of the wave model with the CMIP5 GCM outputs affects wave model performance, without gaining the benefits of increased spatial resolution obtained through the dynamical atmospheric downscaling, has not been established.

In this paper, we assess the ability of 8 CMIP5 GCMs to reproduce historical global wave climate when used to directly force a dynamical spectral wave model. These results are compared to wave climate simulations presented by Hemer et al. (2013a), derived using the same wave model but forced with 2 CMIP3 GCM fields downscaled using the cubic-conformal atmospheric model, CCAM (McGregor and Dix, 2008).

The analysis of the model results presented in this paper are directed at the following questions:

- How well do CMIP5 GCM forced dynamical wave simulations represent the past and present climatological annual and seasonal mean and maximum wave climate?

- Are wave climate simulations hampered by using forcing taken directly from coarse resolution CMIP5 GCMs, relative to the CMIP3 wave simulations published by Hemer et al. (2013a), where forcing was derived from atmospheric downscaling simulations?
- Can the ability of a GCM to represent standard climate variables act as a proxy for the accuracy of dynamical wave climate simulations forced with that GCM?

We structure the manuscript as follows. Following this introduction is the methods section (Section 2), including the datasets used and the metrics used to measure model performance. Section 3 presents the results and associated discussion of the study, with Section 4 containing our conclusions.

2. Methodology

2.1 Datasets

2.1.1 Historical Wave Reanalyses

No definitive source of observed wave data is available to describe historical global wave climate. In place of observations, we use three wave model datasets to represent historical wave climate. These include two wave reanalyses - 1) the corrected ERA-40 wave reanalysis (hereafter ERA40C; Sterl and Caires, 2005), and 2) the ERA-Interim reanalysis (Dee et al., 2011); and a wave model hindcast - 3) the NCEP Climate Forecast System Reanalysis (CFSR; Saha et al., 2013) forced hindcast with 1° spatial resolution (G1d_cfsr) described by Hemer et al. (2013a). Despite the G1d_cfsr simulation being a hindcast forced by an atmospheric reanalysis dataset, we refer to these datasets collectively as ‘the reanalyses’ throughout the manuscript to distinguish them from the GCM forced historical simulations. These datasets exhibit differences for many reasons, including the different data sources assimilated (in forcing data, or of wave data for ERA), different spatial resolutions, and different methods of interpolation to account for topographic effects, amongst others. The ECMWF ERA datasets have been widely implemented as benchmark datasets in wave climate studies and are consequently well understood (e.g., Sterl and Caires, 2005; Semedo et al., 2010; Hemer et al., 2013b). The G1d_cfsr dataset is included to provide measure of the performance of the wave model implementation, without the added ambiguity introduced by forcing the wave model with GCM winds. We don’t preference any hindcast over another, each being considered representative of the historical wave climate against which to measure wave climate modelling uncertainty.

The following climate statistics were calculated from each dataset: monthly mean of significant wave height ($H_{s,m}$), mean wave period ($T_{M,m}$) and mean wave direction ($D_{M,m}$), and monthly maximum of significant wave height ($H_{s,x}$). The G1d_cfsr and ERA-Interim datasets span a common time period with the historical GCM simulations from 1980 to 2005, providing a 312 month time-series for each variable, for each dataset. ERA40C data, ends in March 2003. A common time-period thus has a limited span of 1980-2002, providing a 276 month time-series of each variable.

2.1.2 GCM forced wave climate simulations

The WAVEWATCH III wave model (version 3.14, Tolman, 2009) was implemented over a near global domain (latitude 80°S-80°N) at 1° spatial resolution, as detailed by Hemer et al., 2013a. The wave model spectra was defined by a directional resolution of 15° and 25 frequency bands ranging non-linearly from 0.04 to 0.5 Hz. Despite recent advances in available source term packages for WAVEWATCH III, which suggest improvements in simulated global representations of swell (e.g, Ardhuin et al., 2010), WAM-4 parameterisations as adapted by Bidlot et al. (2005), referred to as BAJ in the WAVEWATCH III manual, were implemented to maintain physical consistency with the prior CMIP3 wave climate ensemble members from Hemer et al. (2013a).

In total, we consider 10 GCM forced ‘historical experiment’ wave climate simulations (Table 1). This includes 2 simulations forced with CMIP3 GCMs downscaled using the Cubic Conformal Atmospheric Model (McGregor and Dix, 2008), which were documented in Hemer et al., (2013a), and 8 new simulations forced directly with CMIP5 GCM surface winds. The historical experiment includes both natural and anthropogenic forcings for the years 1850 to 2005, though high temporal resolution surface wind data required to force the wave model is archived only for the period 1980-2005, being the period of focus for this study. For each GCM, only a single ensemble has been used (i.e., no assessment of internal model variability), with our objective being to resolve the uncertainty associated with alternative model formulations from different modelling groups.

Surface wind (uas, vas) at 3-hourly temporal resolution, and sea-ice area fraction (sic) at monthly frequency, from each of the CMIP5 GCMs used, were obtained through the Earth System Grid – Centre for Enabling Technologies (ESG-CET; <http://pcmdi9.llnl.gov/>). These data were then used as forcing for the wave model directly, with data linearly interpolated in time to provide forcing at each time-step. For the HadGEM2-ES and INMCM4 simulations, minor changes to WAVEWATCH III were

required to enable simulation of a 360 and 365-day year, consistent with the respective GCM. Eight CMIP5 GCM forced 26-yr historical wave climate simulations resulted.

An advantage of dynamical wave projection studies is the ability to assess a range of wave characteristics, which can be derived from the modelled spectra. In this paper, we limit our focus to the following climate statistics calculated from the 6-hourly archives from each GCM forced wave simulation: monthly mean of significant wave height ($H_{S,m}$), mean wave period ($T_{M,m}$) and mean wave direction ($D_{M,m}$), and monthly maximum of significant wave height ($H_{S,x}$).

2.2 GCM forced wave climate model evaluation

2.2.1 Performance Metrics

Many metrics are used to evaluate climate models. These typically include measures which summarise features of the climatological mean state and others which summarise temporal variability. To date, no standard methodology for evaluating wave climate models is established and approaches to date have been deficient, concentrating on performance of the model to reproduce annual mean significant wave height, with little regard for other wave parameters (e.g., period and/or directions) (e.g., Wang et al., 2014; Dobrynin et al., 2012) or for temporal or regional variability of these variables.

Beyond the basic bias measures of skill, which we initially present, we have implemented two independent methods to evaluate the performance of the GCM forced wave climate simulations and ensure robust conclusions from our study. Both methods follow those which have been applied to terrestrial climate variables in other published studies. The first method, using the Mielke Measure (or M-score) has been applied by Watterson et al. (2014) amongst others, and the second, which ranks models on the basis of a normalized error matrix, used by Rupp et al. (2013). The two methods are outlined below.

In order to establish model performance across the global ocean, and its regional variability, we define 13 sub-regions of the global ocean (Figure 1), following Alves (2006). The regions are chosen such that the wave climate exhibits similar qualitative characteristics within each area. The metrics, outlined below, were determined for the global domain, and for each of the 13 sub-regions independently.

The focus of this study is on evaluation of the mean and seasonal climatologies of wave properties across the global ocean. We acknowledge the importance of evaluating the skill of our climate simulations to capture the inter-annual and inter-decadal variability of the climate system (Flato et al., 2013). Of particular interest within the GCM forced wave simulations is the skill of the models to represent the wave field response associated with atmospheric climate mode variability (such as the North-Atlantic Oscillation, the El-Nino-Southern Oscillation, or Southern Annular Mode). Our 27-yr time-slice simulations are poorly suited to assessing decadal variability, and the evaluation of inter-annual variability within the simulations is left for a following study.

2.2.1.1 Global Biases of Annual Mean and Seasonality

To support an initial measure of model skill of the GCM forced wave models, we determine the annual mean and seasonality (the difference between the January and July mean values) of $H_{s,m}$, $T_{M,m}$, $D_{M,m}$ and $H_{s,x}$, for each of the ERA40C, ERA-Interim and G1d_cfsr datasets; for each of the historical GCM forced simulations independently; and ensemble means spanning CMIP5 simulations (N=8), CCAM-CMIP3 simulations (N=2) and for the complete GCM forced wave climate ensemble (N=10). We determine the bias, and the absolute bias of each model relative to the reanalyses. To provide a quantitative measure of uncertainty, or inconsistency, of historical observations, we also determine the mean of the absolute pairwise difference between the 3 reanalyses, for both annual mean and seasonality, following Flato et al. (2013).

2.2.1.2 Method 1: Mielke Measure

The first metric used is the non-dimensional *arcsin* Mielke Measure, or *M*-score. This metric has been applied to standard climate variables, tas, pr and psl, by Watterson et al. (2014) to measure skill of the CMIP5 climate models over the continents. The beneficial characteristics of the metric are outlined in Watterson (1996), and it has been used more widely, for example, by Jones et al. (2005), Meehl et al. (2007), and Elguindi et al. (2014). The *M*-score is essentially the mean-square-error (*mse*), non-dimensionalised by the spatial variance of the field. For the model field *X* and observed (or reanalysis) field *Y*, the *M*-score is defined as

$$M = \left(\frac{2}{\pi}\right) \arcsin \left(1 - \frac{mse}{V_x + V_y + (G_x - G_y)^2} \right) * 1000,$$

where V is the spatial variance and G is the spatial mean of a given spatial field (e.g., a temporal [annual or seasonal] mean of [global or sub-domain] significant wave height, or other wave parameter). The *arcsin* transformation was added by Watterson et al. (2014) and means that the deviation from unity asymptotes to the square root of mse , rather than mse itself, which is advantageous for fields where the correlation values tend to be close to 1. All statistics were calculated over the domain, after Y was interpolated onto the X grid. The multiplier gives a skill score that has a maximum possible value (for $mse = 0$) of 1000 “points”, while a zero score indicates no skill (and negative values – signifying anti-correlation - do not occur in our context).

We first determine the three pair-wise similarity values between the three reanalyses (G1d_cfsr, ERA-Interim and ERA40C) over the common period of 1979-2002. The M -score is determined across 14 domains (globally and the 13 sub-domains), for each season (defined by Dec-Feb, Mar-May, Jun-Aug, and Sep-Nov), for four variables ($H_{S,m}$, $H_{S,x}$, $T_{M,m}$ and $D_{M,m}$). To determine the M -score for mean wave direction, the direction is transformed into unit vector zonal and meridional components, and we exploit the property of the mse and variance of a vector being the respective sums from each component.

To measure performance of the GCM forced wave climate fields, The M -score was determined by comparing the climate model forced wave fields to the three reanalyses. This comparison was carried out for the 10 individual climate models, the 8-member CMIP5 model ensemble mean, the 2-member downscaled CMIP3 model ensemble mean, and the total 10-member ensemble mean. The common period between the climate model forced wave fields and the reanalyses is 1980-2005 when assessing against G1d_cfsr and ERA-Interim, but limited to 1980-2002 when measuring against ERA40C. For each reanalysis dataset, the M -score is determined relative to the 10 individual models and 3 ensemble means (CMIP5-EM, CMIP3-EM and the total 10-member EM), comparing spatial fields of seasonal mean (DJF, MAM, JJA, SON) wave fields for four variables ($H_{S,m}$, $H_{S,x}$, $T_{M,m}$ and $D_{M,m}$), for 14 domains (global and 13 sub-domains). i.e., a total of 3 (reanalyses) x 13 (simulations) x 4 (seasons) x 4 (variables) x 14 (domains) = 8736 M -score values.

2.2.1.3 Method 2: Normalised Error Matrix

The second metric is the Normalised Error Matrix (NEM-score; Rupp et al., 2013). This method has been used to assess the performance of CMIP5 models to simulate terrestrial temperature and precipitation over the Pacific Northwest USA. The NEM uses a number of metrics which consider

both properties of the regionally averaged time-series and the spatial variance of the variables within the sub-region. Seven metrics were selected on the basis of being representative of the wave climate system (Table 2). Four of these metrics (spatial correlation, spatial standard deviation, and the time-series variance of annual and five-year running mean) were calculated seasonally. For the four variables of interest ($H_{S,m}$, $H_{S,x}$, $T_{M,m}$ and $D_{M,m}$), these 19 indices lead to a total of 76 metrics for each model (given seasonal values), for each sub-domain.

We define the model error E_{ij} , for a given model i and metric j , as

$$E_{ij} = |x_{\text{reanalysis},j} - x_{ij}|$$

where $x_{\text{reanalysis}}$ and x_i are the reanalysis and simulated metric, respectively. The calculation is repeated for all 3 reanalyses, and values averaged. In the case where x is a correlation coefficient, $x_{\text{reanalysis}}$ is set to 1. A relative error is then defined, to give each metric an equal weight, as

$$E_{ij}^* = \frac{E_{ij} - \min_i(E_{ij})}{\max_i(E_{ij}) - \min_i(E_{ij})}, i = 1, \dots, \text{nmodels}, j = 1, \dots, \text{nmetrics}$$

Where the minimum and maximum operators indicate min/max across all considered models, for a given metric.

The relative error is then summed across all m metrics, where m is 76, to get the total relative error per model.

$$E_{i \text{ tot}}^* = \sum_{j=1}^m E_{ij}^*.$$

The models were ranked by ordering them by their respective total relative error.

The total relative error for each model is determined for each of the 13 sub-regions and for the global domain, to rank performance of models for each sub-basin.

As for the Mielke Measure, we calculate each metric for each of the 13 sub-domains, and globally, for each ensemble member. Furthermore, we also calculate metrics for the 3 ensemble means (2-member CCAM-CMIP3-EM, 8-member CMIP5-EM and 10-member EM), to assess the benefit of the respective ensembles in representing wave climate.

The large number of metrics are able to identify strengths and weaknesses of different elements of each model. However it is somewhat challenging to summarise the performance of models. In order to rank the models, each of the 76 metrics were given an equal weight. We acknowledge that for some applications, some metrics may be more important than others, or that there may be redundancy in some metrics, but equal weighting provided a valuable assessment of overall performance of the model for each region.

3. Results

3.1 Global Biases of Annual Mean and Seasonality

The wave reanalyses show a global distribution of annual mean $H_{s,m}$ with larger wave heights in the extra-tropics and lower heights in the tropics (e.g., Figure 2a). This general pattern is reproduced by all GCM forced wave climate simulations. There are, however, notable differences between CMIP3 and CMIP5 simulations which can be observed in this initial assessment. Hemer et al. (2013a) reported on the increased zonality of the surface atmospheric circulation in the downscaled CMIP3 simulations, driving strong westerly waves in the Southern Ocean, which propagated as swell into the tropics – particularly into the tropical eastern Pacific. The mean bias of annual mean $H_{s,m}$ of the CMIP5 forced simulations (ensemble mean in Figure 2b) relative to the G1d_cfsr hindcast shows a notably different signal. The Southern Ocean wave heights have a strong (~ 0.5 m) negative bias, which we attribute to the reduced intensity of storm systems within the CMIP5 models to drive the large wave heights characteristic of the region. Similar negative biases are observed in the North Pacific and North Atlantic storm-track regions. This reduced generation in the stormy areas has extensive impact on the global wave field, with negative biases reported across the entire global ocean. Bias is smallest, and slightly positive (~ 0.1 m), in the western tropical Pacific and western tropical Indian oceans. In these regions, typical waves can be associated with long fetches, generated by the trade-winds for which high model resolution adds less benefit. Extreme events driven by Cyclone/Typhoon conditions in these regions are poorly resolved by the available model resolution, which impacts heavily on simulation of extremes, but less so on these typical wave conditions. The observed biases (~ 0.5 m, or 10-15% in the Southern Ocean) appear substantive, until compared with reanalysis uncertainty, represented here by the mean of the absolute pairwise difference in annual mean $H_{s,m}$ across the three reanalyses (Figure 2d). In the Southern Ocean (again the region of maximum values), reanalysis inconsistency values of ~ 0.4 m are observed, or more than 50% of the magnitude of the model bias. Similar values are observed in the eastern tropical Pacific

and northern extratropical storm belt. High inconsistency in these regions suggests differences between the reanalyses in the active extra-tropical wave generation zones, which has influence on the strongly swell influenced eastern tropical Pacific region. While not shown here, we also observe the mean of the absolute pairwise difference of the annual mean $H_{s,x}$ values derived from the three reanalyses exceeds the bias between any of the reanalyses and the CMIP5 ensemble mean values. The inconsistencies between reanalyses can likely be attributed to differences in source term physics, the data assimilated, and variable forcing resolution. The magnitude of inconsistency between reanalyses is approximately 80% of the magnitude of the mean absolute bias of the climate model forced simulations. Thus, while climate model forcing results in biases in the global wave fields, a non-trivial portion of this bias may be attributable to wave model uncertainty, and climate model bias being the remaining component.

The strongest seasonality of mean significant wave heights is seen in the extra-tropical storm belts of the North Atlantic and Pacific Oceans, with January wave heights exceeding July wave heights by almost 2m (Figure 3a). Seasonality in the Southern Ocean is somewhat smaller. The Southern Ocean seasonality is strongest in the Indian sector of the Southern Ocean, with July mean wave heights more than 1m greater than January mean values, but the difference in the Pacific sector is smaller (~0.8 m). Through the tropics, mean wave heights remain relatively consistent throughout the year. The CMIP5 forced wave simulations reproduce this seasonality qualitatively, however systematic biases are observed. Across the northern hemisphere, seasonality is underestimated (by about 0.3m) in all ocean sectors. This difference is primarily associated with underestimation of the January wave heights, with July wave heights having little to no bias. In the southern hemisphere, the CMIP5 simulations underestimate seasonality in the Indian Ocean sector by approximately 0.2 m (note that underestimation of seasonality in the southern hemisphere is observed as a positive bias, owing to the negative values of seasonality here) due to July mean wave heights being underestimated. In the Pacific, an underestimation of seasonality is also observed, but is spatially variable and not strong (Figure 3b). This underestimation of stormy season wave heights has the effect of reducing the magnitude of seasonality. This result further reduces our confidence of the ability of the CMIP5 GCMs to generate reliable storm wave heights in our global simulations. It is noteworthy however that in the extra-tropical south Atlantic, seasonality is greater in the GCM forced simulations than the CFSR forced hindcast, indicating there are specificities to each basin. Seasonality of $H_{s,x}$ exhibit similar characteristics as described here for the mean.

The eastwards zonal increase in mean wave period across each ocean basin, with shorter wave periods on the western side of the ocean and longer wave periods on the eastern side, is similarly reflected in all GCM forced wave simulations (Figure 4a). As with wave height statistics, we see that the CMIP5 forced simulations have a larger (negative) bias than the CCAM-CMIP3 simulations of Hemer et al. (2013a). The bias in the CMIP5 runs is relatively uniform at approximately 0.5-0.6s less than G1d_cfsr, although there are regions of lesser bias ($\sim 0.1s$) in the western tropical Pacific and Indian Oceans, and in the eastern tropical Atlantic (Figure 4b). Again, there is notable variance amongst the reanalyses, with the inconsistency (mean of the absolute pairwise difference) among the three datasets being equivalent to approximately one half (0.3s) of the bias of the CMIP5 ensemble mean (Figure 4d). This inconsistency in the observational datasets is greatest in the swell affected eastern tropical Pacific. Differences between the three reanalyses can be expected as a result of differences in wave model source term physics, data assimilated into the wave fields (for ERA products) or associated atmospheric fields (e.g., surface winds - all products), and model resolutions.

As with wave height, mean wave period seasonality is greater in the northern hemisphere. In the extratropical north Pacific and Atlantic oceans, mean wave period is approximately 3s longer in January than July. In the southern hemisphere, this seasonality is less pronounced, with January mean wave periods approximately 1s shorter than July. In the mid-latitudes of the Indian Ocean, west of Australia a region of stronger seasonality is observed (Figure 5a). Seasonality is relatively well represented by the GCM forced simulations, with biases typically less than 10% of the magnitude of seasonality (Figure 5b).

The general distribution of annual mean wave direction is consistent between reanalyses and GCM forced simulations (Figure 6). Systematic biases are observed in the GCM forced simulations. The tendency for the CMIP5 ensemble mean wave directions, as with the CMIP3 results reported in Hemer et al. (2013a), is that the GCM forced wave directions are typically more zonal than hindcast. In the southern-most latitudes, a greater southward directional component (a positive directional bias) is observed, but slightly further north ($\sim$ latitude $40^{\circ}S$), the south-westerly waves become more zonal (westerly), with biases of around 10° . In the mid-latitudes of the Southern Hemisphere, between $10-30^{\circ}S$, the bias is negative, suggesting the south-easterly trade generated waves also tend more zonal (easterly). The magnitude of these biases, between the CMIP5 ensemble mean and the CFSR forced simulation, are not large, and fall within the bounds of the inconsistency (mean of the absolute pairwise difference) in the reanalyses (Figure 6d).

Seasonality of wave direction is most strongly observed in the monsoon affected regions, and also at the sub-tropical ridges as they seasonally shift in position and the associated divergence of flow moves in position. The bias in the directional seasonality between the CMIP5 ensemble mean and the CFSR simulation is of similar magnitude to that observed for the mean, however it is less systematic in its spatial variability (Figure 7).

3.2 Mielke Measure – hindcast intercomparison

The M -scores calculated pairwise between reanalyses shows strongest similarity between ERA-Interim and ERA40C reanalyses. Averaged across four seasons, and four variables, the global M -value is 882 for this pair-wise comparison. This is notably greater than the comparisons between G1d_cfsr and ERA-Interim or ERA40C, which have corresponding M -scores of 782 and 780 respectively (Figure 8).

Strongest similarity is observed for variable $H_{s,m}$, with M -scores (averaged across four seasons) of 935 for the ERA pair-wise comparison, and 869 and 900 for corresponding G1d_cfsr to ERA-Interim and ERA40C comparisons. Lowest M -scores are observed for variable $H_{s,x}$, particularly when comparing G1d_cfsr with either ERA-Interim (global $M = 533$) or ERA40C (global $M=560$), compared to the ERA-Interim to ERA40C value (global $M = 847$). Lower M -scores are observed between G1d_cfsr and the ERA fields in the Southern Ocean (ETSI, ETSP, and ETSA), where $H_{s,x}$ M -score values lie between 479 and 518 (The ERA-Interim – ERA40C pair-wise comparison yields M -score values ranging from 843 – 870 in these subdomains). These results further suggest differences between the reanalyses are a function of forcing characteristics, where the ERA products are more similar than to the NCFSR winds, with differences between products more pronounced in the Southern Ocean.

In other regions, M -scores are more comparable irrespective of the reanalysis pairing (Figure 8). Similarly, M -score values for $T_{M,m}$ and $D_{M,m}$ are similar for all reanalysis pairs, with global four-season averaged values in the range of 881-919 and 791-845 respectively.

While we observe differences between reanalyses, the conclusions which can be drawn from the M -scores of the climate model forced wave fields (presented in the following section) are consistent regardless of which of the three reanalyses the models are compared. Consequently, the remainder

of this study presents the *M*-score of a climate model forced wave field as the average of the 3 values corresponding to the three reanalyses.

3.3 Mielke Measure – climate model forced wave simulations

Averaging *M*-scores across reanalyses, seasons and variables, largest skill scores between GCM forced wave simulations and the reanalyses are observed for the full global domain (mean *M*-score=765, Figure 8). This suggests the global structure of the wave field is relatively well reproduced by the GCM forced simulations. The variance of *M*-scores across all simulations is much smaller for the global domain than for other sub-regions (Figure 8). Over the global domain, the majority of GCM forced wave simulations have an *M*-score comparable with the G1d_cfsr to ERA reanalyses pairwise comparison values. The ensemble mean magnitude of *M* for the global domain (765) indicates the GCM forced wave simulations have a substantial degree of skill when compared to other studies. For comparison, Watterson et al. (2014) report *M*-scores of more widely considered climate model variables (tas, pr and psl) from the CMIP5 model ensemble of between 596 and 755, when compared with ERA-Interim (averaged over the three variables and four seasons). The climate model forced wave simulations thus have similar skill to the climate models from which they are derived.

The skill of the ensemble of GCM forced wave simulations is relatively consistent across the 13 regions, with the ensemble mean *M*-score exceeding 600 in almost all regions - only the Tropical Eastern South Pacific (TESP) region displays lower skill (mean *M*-score of 540). TESP is possibly the most complex region of the global ocean for wave simulation, being characterized by a multi-modal wave field comprised of southern and northern extra-tropical weather system generated swell in addition to locally trade-wind generated seas, and the lower climate wave model skill in this region is a reflection of the models not resolving this complexity.

In almost all regions, an outlier GCM with a notably lower *M*-score is observed. Observing the properties of individual models (Figure 9), CNRM-CM5 is the model which consistently underperforms the other models. Aside from CNRM-CM5, the other climate model forced wave simulations demonstrate comparable skill. Similar relative characteristics are observed in the CNRM-CM5 simulations as in the other simulations, showing higher *M*-values across the TWNP, TWSP, TNAO and global regions, and lower values in ETSI, TENP, TESP and TSAO, for example. However the

CNRM-CM5 *M*-score values are approximately 100-150 unit lower than the other models in almost all regions.

It is of interest that CNRM-CM5 is identified as having the lowest skill of the models used in this study. In prior studies that have ranked the CMIP5 general circulation models on the basis of more widely used variables (e.g., tas, psl and pr; Watterson et al., 2014), CNRM-CM5 is identified as the model with highest skill. Table 3 lists the average global *M*-scores presented by Watterson et al for terrestrial climate variables (tas, psl and pr), along with the average *M*-scores for the climate model wave fields determined in this study. Analysis shows no significant relationship (correlation coefficient, $R = -0.57$, indicating a negative relationship if any) between the two sets of values, including a scenario where CNRM-CM5 is considered an outlier and removed ($R = 0.21$). Watterson et al. (2014) note one factor in the high level of performance of the CNRM-CM5 model for the parameters considered in their study was the model resolution being capable of resolving the topography of the Tibetan plateau, and improving model skill in this region. Resolution of topography is not a first order factor for performance of the marine surface winds for the majority of the global ocean, and the associated benefit observed in this model is lost in our global simulations of surface ocean wind-waves.

A factor considered to be important in the skill of a climate model, particularly when used as forcing for a wave model, is its spatial resolution. Wave models are strongly governed by the representation of the forcing winds, with high resolution required to resolve the intensity of storm systems which generate surface waves. We regress the global *M*-score (and for sub-regions) against the model resolution, represented by length as defined in Watterson et al. (2014), to determine the strength of the assumed relationship. Regardless of whether CNRM-CM5 is included ($R=-0.0247$) or not ($R=-0.3421$), we see no significant correlation between the global *M*-score and model resolution. However, at sub-domain level, we do find significant (at 95% level) correlation between wave model *M*-scores and the resolution of the forcing model in regions ETNA ($R=-0.82$), TNAO ($R=-0.86$) and TSAO ($R=-0.76$), and at a 90% confidence level in regions TENP ($R=-0.62$) and TWSP ($R=-0.65$). If we limit our analysis to mean significant wave height only, we find a significant correlation (90% confidence) between the global *M*-score and model resolution ($R=-0.64$), but looking at a regional level, we see this relationship observed in the Atlantic Ocean, but not elsewhere.

From Table 3, we see that the global *M*-score is greatest for the 10-member ensemble mean. Consistent with the ‘standard’ climate variables that are available directly from the GCMs, the

ensemble mean of the GCM forced wave simulations serves to filter out biases of individual models, retaining only errors which are pervasive (Randall and Wood, 2007). The 10-member ensemble includes models from both CMIP5 models, and down-scaled CMIP3 models. That the global M -score for the 8-member CMIP5 ensemble mean is only 14 less than the full ensemble is suggestive that the 8-member ensemble is near-sufficient to resolve the uncertainty within the models, or perhaps the increased performance of the 10-member ensemble is attributable to the downscaled models being high performing models (which is reflected by the two downscaled models having M -values ranking 3 and 6). The number of models included in our ensemble is consistent with the number which other studies have indicated is sufficient to represent the full ensemble. For example, Pierce et al. (2009), found ensemble skill of average surface temperature over the western US of any 6 GCMs approached that of the full 20+ member ensemble, and Mote et al. (2011) recommended at least 10 models be required within an ensemble to assess the uncertainty associated with the physical response represented by model formulation. That the ensemble mean out-performs any individual model demonstrates the value of coordinated efforts of the wind-wave climate community to increase the ensemble of wave climate projections through the Coordinated Ocean Wave Climate Project (Hemer et al., 2012, 2013b), but points to the need for continued effort to evaluate different wave downscaling methods across the community's efforts.

The 10-member ensemble mean exhibits the highest M -score across all sub-regions, except ETSA and ETNA (Figure 9). In ETSA, we see that two of the CMIP5 models have higher M -score than the ensemble mean (MRI-CGCM3 and INMCM4). In ETNA, the two downscaled CMIP3 models (CCAM-ECHAM5 and CCAM-CSIRO Mk3.0) show substantially more skill than the CMIP5 models, and thus exceed the 10-member mean. Figure 9 shows that no individual model is substantially better than others, however the downscaled CCAM-CMIP3 models tend to have a marginally higher skill than the CMIP5 models, as reflected by the high ranking of global M -scores (3 and 6, Table 3).

Consistent with Hemer et al. (2013a), we find the wave models have greater skill in representing mean conditions than the extremes, with the mean global M -score being much greater for the annual mean wave height (845) than the annual maximum (723; Table 4).

3.4 Relative Error Matrices

A relative error matrix is constructed for each sub-domain, with Figure 10 displaying the relative error matrix for the Global domain. Considerable information can be extracted from these matrices,

but here we limit our attention to using the matrices as an independent check of the findings of the Mielke evaluation measure presented above.

We note that of the 14 matrices (13 sub-domains, and global), simulation 13 - the 10-member ensemble mean, made up of 8 CMIP5 models and 2 CCAM-CMIP3 models – is the best performing simulation in 9 domains, and is the second best performing simulation in remaining domains (Table 5). Of the domains where the ensemble mean is not the best performing simulation, the CMIP5 ensemble mean shows greatest skill in the ETNP, and simulation 8 (MRI-CGCM3) is the best performing simulation in TWSP, ETSP, ETSA and the global domain. MRI-CGCM3 was identified by the Mielke measure analysis above as a high performing model, with it demonstrating greatest skill in ETSA (Figure 9), the highest performance of any individual model in TWSP, and a high performing model in ETSP and Global domains. However, aside from in ETSA, the Mielke measure shows the ensemble means (CMIP5-EM, CCAM-CMIP3-EM, and EM) have greater skill than the MRI-CGCM3 model. While the MRI-CGCM3 model performs very well in many domains, we observe it to be a poorer performing model (ranked 7th or lower) in 4 sub-domains (TNIO, TESP, ETNA and TSAO).

In contrast, the CMIP5-EM (simulation 11) has greatest skill in only one domain, but ranks in the top 3 simulations in all domains except ETSP (rank 6) and GLOB (rank 4). The CCAM-CMIP3 ensemble mean (simulation 12) ranks 3rd for the global domain, but looking across all sub-domains the CCAM-CMIP3-EM simulation does not rank highly (in top 3 in 4 domains, rank 7th or lower in 3 sub-domains).

Consistent with the Mielke score analysis, the relationship between model resolution and model skill is marginal. For the global domain, a Pearson's correlation coefficient of $R=0.498$ is obtained between model skill (sum of relative error across all indices for each model) and model resolution, which has significance at a 90% level. Two other sub-domains have similar relationships (ETNA: $R=0.654$ and TSAO: $R=0.514$), corresponding with regions where a high correlation with model resolution was observed with the Mielke score analysis. In other sub-domains, no significant correlation is observed. In this analysis, the CNRM-CM5 model is again observed to be the poorest performing model, and if removed from analysis we observe an improved relationship between model performance and resolution, with significant relationships (95% confidence level) observed in sub-domains TSIO, ETNA, TSAO and GLOB. i.e., Correlation Coefficient values shift from $0.39 \rightarrow 0.68$, $0.65 \rightarrow 0.70$, $0.51 \rightarrow 0.62$ and $0.49 \rightarrow 0.66$ for these regions respectively when the CNRM-CM5 model is removed as an outlier.

As with the Mielke measure analysis, we find no relationship between the performance of the GCM forced wave simulations, and the skill of the wave models to represent terrestrial climate (as defined by the Watterson M-score rankings).

The relative error matrices also identify the indices for which the models have greater skill. Within each matrix, we observe that the models demonstrate similar skill regardless of which of the four variables are considered ($H_{S,m}$, $H_{S,x}$, $T_{M,m}$ or $D_{M,m}$). We typically observe that the simulations have high skill for many of the indices, regardless of the variable – in particular, relatively high skill is associated with the climatological mean, the seasonal mean amplitude and the spatial correlations and standard deviations. However, it is apparent that poorer skill is associated with the 4 last indices (TVar5.MMM; Table 2), along with the trend (index 3), for all variables. These results suggest the GCM forced wave simulations differ from the wave reanalyses on longer (5+ year) time-scales. Temporal in-homogeneity is widely observed in the wave reanalyses (Hemer, 2010; Chawla et al., 2013; Stopa and Cheung, 2014), so whether the differences can be attributed to problems with GCM forcing or to the broader problems associated with long time-scale wave reanalyses cannot be resolved in this study.

Discussion and Conclusions

The study has evaluated an ensemble of wave climate simulations, where a 1° resolution global configuration of WAVEWATCH III has been forced using surface winds taken directly from the CMIP5 generation of climate models. This follows prior work (Hemer et al., 2013a), where the same model configuration was forced with surface winds taken from two CMIP3 models which were dynamically downscaled using the cubic-conformal atmospheric model (CCAM).

A more rigorous evaluation of the historical model performance has been undertaken using two previously published approaches (the Mielke performance metric and Rupp et al. relative error matrices). We have assessed the seasonal variability of model performance, for four variables ($H_{S,m}$, $H_{S,x}$, $T_{M,m}$ and $D_{M,m}$), against three different reanalyses. From the study, a number of conclusions can be made. Largely consistent results can be concluded from the two methods used.

We use two available wave reanalyses (ERA-Interim and ERA40C) and one reanalysis forced wave hindcast (G1d_cfsr) as reference datasets for our study. Prior studies have evaluated these datasets,

and all display biases relative to available wave (buoy and altimeter) observations. These reanalyses, owing to their coarse resolution, are less able to resolve extreme wave conditions (e.g., Sterl and Caires, 2005, Stopa and Cheung, 2014) and thus the negative biases we report between GCM forced simulations and reanalyses could be expected to be further accentuated relative to observational datasets. The reanalyses however provide a valuable dataset to reference wave simulations carried out at similar spatial resolutions.

The study set out to determine whether the performance of a climate model to represent the commonly assessed variables available in the CMIP data archives (e.g., tas, pr, psl) is representative of the performance of a wave climate simulation forced with that model. We found no relationship between the skill of a climate model to simulate wave fields and the skill of a GCM in representing common climate variables.

The pattern of standard error between GCMs and reanalyses were found to be larger than the mean reanalysis inconsistency. Over the global domain, the majority of GCM forced wave simulations have an M-score comparable with wave reanalysis inconsistency, as represented by the difference between ERA wave reanalyses and the CFSR forced hindcast carried out using the wave model implemented in this study.

Owing to the benefit which is obtained by forcing wave models with higher resolution atmospheric forcing observed in operational and hindcast studies (e.g., Cavaleri, 2009; Bricheno et al., 2013), there is a common perception that forcing a wave model with the higher resolution models which have contributed to the CMIP experiments (with resolution of ~1 degree) might result in improved representation of the simulated wave climate across the global ocean relative to simulations forced with coarser resolution models (of ~2.5-3 degrees). Our study considers only a limited ensemble of models which have contributed to CMIP5, where model resolution ranges from 1.4 degrees (MIROC5 and CNRM-CM5) up to 2.8 degree resolution (BCC-CSM1.1). Amongst this limited ensemble, we find little benefit to using winds from the high resolution CMIP5 GCMs relative to the coarser models. We expect these coarse resolution GCMs which contribute to the CMIP experiments remain too coarse to provide benefit to wind-wave simulations on the basis of resolution. This suggests ongoing refinement of GCM model resolution is required to support wind-wave climate studies.

We set out to determine whether wave climate simulations forced with surface winds taken directly from the CMIP5 GCMs displayed notably poorer skill than the wave climate simulations presented by

Hemer et al. (2013a), using the same wave model, forced with surface winds from a higher resolution atmospheric model which dynamically downscaled CMIP3 GCMs. We find the CMIP5 forced simulations display larger negative biases of wave height (more severe underestimation, particularly of storm wave heights) than the downscaled CMIP3 simulations. We find that on an individual model basis, the dynamically downscaled CMIP3-CCAM-ECHAM5 forced wave simulations is a high performing model (e.g., CMIP3-CCAM-ECHAM5 ranks in the top 5 models the most of any individual model run; Table 5). The CMIP3-CCAM-Mk3.5 model however is a poorer performing model (e.g., ranks in the top 5 simulations – 5th – for one region only – TSAO, Table 5). The added value of dynamically downscaled climate models are well established (e.g, Feser et al., 2011), and the benefits of the dynamical downscaling to the wave simulations can not be tested rigorously without having raw and downscaled forcing available from the same model. However, our study suggests that wave simulations forced directly with CMIP5 model output provide wave climatologies with comparable skill to that presented by Hemer et al. (2013a), with downscaled CMIP3 climate model forcing. We also note that the atmospheric downscaling model removes the observed variance amongst parent GCMs, and introduces its own uncertainty into the simulation.

The skill of the 10-member ensemble mean exceeds the skill of any individual model, when representing the historical wave climate. This is consistent with the well-established climate modelling methodology, where the ensemble is recognized to retain only the pervasive errors, and demonstrates the benefit of ongoing activities of the wave-climate community to increase the wave climate ensemble through the Coordinated Ocean Wave Climate Project (Hemer et al., 2012, 2013b).

By reducing the global ocean into 13 sub-domains, where each sub-domain represents similar wave conditions, we have been able to show that climate model forced wave simulations have variable skill depending on region. The most challenging region is the swell dominated regions of the tropical eastern Pacific. Our sub-domains represent areas of similar wave characteristics, but it is noted that in the swell dominated regions, the waves are likely generated outside the domain.

Consistent with prior wave studies using GCM forcing, we find that the climate model forced simulations have greater skill in representing mean conditions than extremes (e.g., Hemer et al., 2013a, 2013b). The impacts of climatological variability of wind-waves will most likely be experienced in response to the extremes, and it remains an ongoing challenge as to how we can improve skill in representing extreme wave conditions using available climate model forcing.

711 There remain a number of other questions and challenges which are being addressed in ongoing
712 studies of waves in the context of climate science. While prior studies have evaluated marine wind
713 stress climatologies from the CMIP models relative to satellite and atmospheric reanalysis products
714 (Lee et al., 2013), there is little overlap in the CMIP models considered with this study. Consequently,
715 questions remain as to whether the wave model performance we have observed in our study is
716 indicative of the relative skill of the forcing CMIP GCMs to represent surface marine winds. Our study
717 has also identified notable differences between available historical wind-wave reanalyses and
718 hindcasts. Several other wave hindcast and reanalysis datasets not considered in this study (e.g.,
719 ERA40C, ERA-Interim, NCEP CFSR forced hindcasts of NOAA and CAWCR, IFREMER, and JRA-55 wave
720 hindcast) are also available and assessment of the variance and the sources of variance amongst
721 these datasets is warranted.

722
723 This paper can only provide a first order assessment of a limited number of CMIP5 models in terms
724 of their value in forcing a dynamical wave model. Our limited ensemble concentrates only on the
725 inter-model variability amongst the first ensemble member of the few models considered. While
726 multiple ensemble members may be available for some of these models, these have not been
727 considered in this study and we thus provide no quantification of the intra-model variability. It is
728 possible that internal model variability may account for some portion of the observed variance
729 between our simulations, and attributed here to inter-model differences., in prior studies however
730 (e.g., Kharin et al., 2007 evaluating atmospheric extremes, and Wang et al., 2015 looking at
731 statistical wave climate projections), the spread amongst multi-model ensemble members is
732 generally larger than the spread amongst multi-run ensemble members, which has guided this initial
733 assessment for dynamical wave simulations. Following studies should consider multiple ensemble
734 members of individual models in order to distinguish the inter- and intra-model variability of
735 dynamical wave climate simulations.

736
737 Hemer et al. (2013b) speculated the wave downscaling methodology was the largest source of
738 variance in an ensemble of wave climate projections. An evaluation of the skill of the historical
739 simulations from these GCM forced wave models, as carried out in this study, is needed to identify
740 the sources of variance between the wave modelling methodologies. Such analysis is planned with
741 the growing number of contributions being submitted to the second phase of COWCLIP (Hemer et
742 al., 2014), and is much needed as we seek to understand how waves will respond to a future variable
743 climate, given the potential economic and environmental implications of any future changes.

Acknowledgements

This manuscript is a contribution of the Ocean and Climate Dynamics Research Program of the CSIRO Oceans and Atmosphere Flagship to the Coordinated Ocean Wave Climate Project. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 1 of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. We also acknowledge the European Centre for Medium Range Weather Forecasts (ECMWF) for providing ERA wave hindcasts to the research community, and KNMI for access to the corrected ERA40C wave heights. The research was supported via funding from the CSIRO Oceans and Atmosphere Flagship, the Australian Commonwealth Australian Climate Change Science Program, and the Pacific-Australia Climate Change Science and Adaptation Program. We thank our colleague Julian O'Grady and three anonymous reviewers for their helpful comments and discussions on an earlier version of the manuscript.

References

- Adams, P.N., D.L. Inman and J.L. Lovering, 2011. Effects of climate change and wave direction on longshore sediment transport patterns in Southern California. *Climatic Change*, 109. Doi: 10.1007/s10584-011-0317-0.
- Alves, J.H.G.M. Numerical Modeling of Ocean Swell Contributions to the Global Wind-Wave Climate. *Ocean Modelling*. Vol. 11 (1-2), 98-122, 2006
- Ardhuin, F., E. Rogers, A.V. Babanin, J. Filipot, R. Magne, A. Roland, A. van der Westhuysen, P. Queffelecoulou, J. Lefevre, L. Aouf, and F. Collard, 2010: Semiempirical Dissipation Source Functions for Ocean Waves. Part I: Definition, Calibration, and Validation. *J. Phys. Oceanogr.*, 40, 1917–1941. doi: <http://dx.doi.org/10.1175/2010JPO4324.1>

Barnard, PL.; van Ormondt, M; Erikson, LH., Eshleman, J., Hapke, C., Ruggiero, P., Adams, PN, Foxgrover, AC., 2014. Development of the Coastal Storm Modeling System (CoSMoS) for predicting the impact of storms on high-energy, active-margin coasts, *NATURAL HAZARDS*, 74 (2), 1095-1125

Bidlot, J.-R., Janssen, P. and Abdalla, S. 2005. A revised formulation for ocean wave dissipation in CY29R1. ECMWF Technical Memorandum, R60.9/JB/0(1), pp.1–35.

Bricheno, L.M., A. Soret, J. Wolf, O. Jorba and J-M. Baldasano, 2013. Effect of high-resolution Meteorological forcing on near-shore wave and current model performance. *J. Atm. Oc. Tech.*, 30, DOI: 10.1175/JTECH-D-12-00087.1

Casas-Prat, M., X. Wang and JP Sierra, 2014. A physical-based statistical method for modeling ocean wave heights, *Ocean Modelling*, 73, 59-75. 10.1016/j.ocemod.2013.10.008

Cavaleri, L., 2009. Wave modeling—Missing the peaks. *J. Phys.Oceanogr.*, 39,2757–2778.

Chawla, A., D.M. Spindler, H.L. Tolman, 2013. Validation of a thirty year reanalysis using climate forecast system reanalysis winds, *Ocean Modelling*, 70, 189–206

Coelho, C., Silva, R., Veloso-Gomes, F., and Taveira-Pinto, F. 2009. Potential effects of climate change on northwest Portuguese coastal zones. – *ICES Journal of Marine Science*, 66: 1497–1507.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* 137, 553–597. <http://dx.doi.org/10.1002/qj.828>.

Dobrynin, M., J. Murawsky, and S. Yang (2012), Evolution of the global wind wave climate in CMIP5 experiments, *Geophys. Res. Lett.*, 39, L18606, doi:10.1029/2012GL052843.

Elguindi N, Giorgi F, Turuncoglu UU (2014) Assessment of CMIP5 global model simulations over the sub-set of CORDEX domains used in the Phase I CREMA Experiment. Climatic Change. doi:10.1007/S10584-013-0935-9

Erikson, L.H., C.A. Hegermiller, P.L. Barnard, P. Ruggiero and M. van Ormondt (2015) Projected wave conditions in the Eastern North Pacific under the influence of two CMIP5 climate scenarios. Doi:10.1016/j.ocemod.2015.07.004

Fan, Y., S.J. Lin, S.M. Griffies and M.A. Hemer (2014) Simulated global swell and wind sea climate and their responses to anthropogenic climate change at the end of the 21st Century. J. Climate. 27, 3516-3536. Doi: 10.1175/JCLI-D-13-00198.1

Feser, F., B. Rockel, H. von Storch, J. Winterfeldt, and M. Zahn, 2011: Regional Climate Models add Value to Global Model Data: A Review and selected Examples. Bull. Amer. Meteor. Soc., doi: 10.1175/2011BAMS3061.1., 92 (9), pp. 1181-1192

Flato et al., 2013. Evaluation of climate models. In Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Doschung, A. Nauels, Y. Xia, V. Bex, and P.M. Midgley, Eds. Cambridge University Press, 741-882, doi:10.1017/CBO9781107415324.020),

Giorgi, F., and L. Mearns (2002) Calculation of average, uncertainty range and reliability of regional climate changes from AOGCM simulations via the 'reliability ensemble averaging' (REA) method, J. Clim., 15, 1141–1158.

Giorgi, F., C. Jones and G.R. Asrar (2009) Addressing climate information needs at the regional level: the CORDEX framework. WMO Bulletin, 58(3), 175-183.

Grabemann, I., N. Groll, J. Möller, R. Weisse. (2015) Climate change impact on North Sea wave conditions: a consistent analysis of ten projections. *Ocean Dynamics* **65**, 255-267.

Hemer, M.A. (2010) Historical trends in Southern Ocean storminess: long-term variability of extreme wave heights at Cape Sorell, Tasmania. Geophys. Res. Lett., 37, L18601, doi:10.1029/2010GL044595

844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877

Hemer, M.A., X.L. Wang, R. Weisse, and V.R. Swail (2012) Community 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

Hemer, M.A., J. Katzfey and C. Trenham (2013a) Global dynamical projections of surface ocean wave climate for a future high greenhouse gas emission scenario. *Ocean Modelling* (Special Issue on Surface Ocean Waves). DOI:10.1016/j.ocemod.2012.09.008

Hemer, M.A., Y. Fan, N. Mori, A. Semedo and X.L. Wang (2013b) Projected future changes in wind-wave climate in a multi-model ensemble. *Nature Climate Change*, 3(5), 471-476. DOI: 10.1038/nclimate1791

Hemer, M., X. L. Wang, E. Charles, C. Hegermiller and COWCLIP contributors (2014). Report of the 2014 Meeting for the WCRP-JCOMM Coordinated Ocean Wave Climate Project (COWCLIP), Paris, 1-3 October, 2014. JCOMM Technical Report 82.
http://jcomm.info/components/com_oe/oe.php?task=download&id=26840&version=1.0&lang=1&format=1

Jones C, Gregory J, Thorpe R, Cox P, Murphy J, Sexton D, Valdes H (2004) Systematic optimization and climate simulation of FAMOUS, a fast version of HADCM3. *Hadley Centre Technical Note* 60, 33 pp

Kharin, V. V., F. W. Zwiers, X. Zhang, and G. C. Hegerl (2007), Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations, *J. Climate*, **20**, 1419–1444

Knutti, R. (2010), The end of model democracy?, *Clim. Change*, 102, 395–404, doi:10.1007/s10584-010-9800-2.

Laugel, A., M. Menendez, M. Benoit, G. Mattarolo, F. Méndez, 2014. Wave climate projections along the French coastline: Dynamical versus statistical downscaling methods, *Ocean Modelling*, 84, 35-50, <http://dx.doi.org/10.1016/j.ocemod.2014.09.002>

Lee, T., D.E. Waliser, J.-L.F. Li, F.W. Landerer, and M.M. Gierach, 2013. Evaluation of CMIP3 and CMIP5 Wind Stress Climatology Using Satellite Measurements and Atmospheric Reanalysis Products. *J. Climate*, **26**, 5810–5826. doi: <http://dx.doi.org/10.1175/JCLI-D-12-00591.1>

Lionello, P., S. Cogo, M.B. Galati, and A. Sanna, 2008. The Mediterranean surface wave climate inferred from future scenario simulations. *Global and Planetary Change*, 63 (2-3), 152-162.

McGregor, J.L., Dix, M.R., 2008. An updated description of the conformal-cubic atmospheric model. In: Hamilton, K., Ohfuchi, W. (Eds.), *High Resolution Simulation of the Atmosphere and Ocean*. Springer, Berlin, pp. 51–76.

Meehl, G. A., and Coauthors, 2007a: Global climate projections. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 747–845.

Mori, N., Yasuda, T., Mase, H., Tom, T. & Oku, Y. Projection of extreme wave climate change under the global warming. *Hydrol. Res. Lett.* 4, 1519 (2010).

Mote, P.W., L. Brekke, P. Duffy, and E. Maurer, 2011: Guidelines for constructing climate scenarios. EOS, Transactions, Amer. Geophys. Union, 92, doi:10.1029/2011EO310001

Pierce, D. W., T. P. Barnett, B. D. Santer, and P. J. Gleckler, 2009. Selecting global climate models for regional climate change studies. *Proceedings of the National Academy of Sciences*, doi:10.1073/pnas.0900094106

Plecha, S., Lopes, C., Bruneau, N., Ribeiro, N., Silva, P., Fortunato, A., & Dias, J. (2012). INFLUENCE OF THE WAVE REGIME IN COASTAL SEDIMENT BUDGET: PRESENT AND FUTURE SCENARIOS. *Coastal Engineering Proceedings*, 1(33), sediment.85. doi:http://dx.doi.org/10.9753/icce.v33.sediment.85

Randall, D.A., R.A. Wood, S. Bony, R. Colman, T. Fichet, J. Fyfe, V. Kattsov, A. Pitman, J. Shukla, J. Srinivasan, R.J. Stouffer, A. Sumi and K.E. Taylor, 2007: Climate Models and Their Evaluation. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Rupp, D. E., J. T. Abatzoglou, K. C. Hegewisch, and P. W. Mote (2013), Evaluation of CMIP5 20th century climate simulations for the Pacific Northwest USA, *J. Geophys. Res. Atmos.*, 118, 10,884–10,906, doi:10.1002/jgrd.50843.

Saha, Suranjana *et al.*, 2010. The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society* 91, 1015–1057. <http://dx.doi.org/10.1175/2010BAMS3001.1>

Semedo, A., K. Sušelj, A. Rutgersson, A. Sterl, 2010: A Global View on the Wind Sea and Swell Climate and Variability from ERA-40. *Journal of Climate* , 24:1461-1479.

Semedo, A., R. Weisse, A. Behrens, A. Sterl, L. Bengtsson, and H. Günther, 2013: Projection of Global Wave Climate Change toward the End of the Twenty-First Century. *J. Climate*, 26, 8269–8288. doi: <http://dx.doi.org/10.1175/JCLI-D-12-00658.1>

Sierra, J. P.; Casas-Prat, M., 2014. Analysis of potential impacts on coastal areas due to changes in wave conditions, *CLIMATIC CHANGE*, 124 (4), 861-876.

Sterl, A. & Caires, S, 2005. Climatology, variability and extrema of ocean waves - The web-based KNMI/ERA-40 wave atlas. *Int. J. Climatol.* 25, 963997.

Stopa, J.E., K.F. Cheung, 2014. Intercomparison of wind and wave data from the ECMWF reanalysis interim and the NCEP climate forecast system reanalysis, *Ocean Modelling*, 75, 65–83

Tolman, H. L., 2009: User manual and system documentation of WAVEWATCH III version 3.14. NOAA / NWS / NCEP / MMAB Technical Note 276, 194 pp.+ Appendices

Wang, X. L., Y. Feng, and V. R. Swail (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., Y. Feng, and V. R. Swail (2015), Climate change signal and uncertainty in CMIP5-based projections of global ocean surface wave heights, *J. Geophys. Res. Oceans*, 120, 3859–3871, doi:10.1002/2015JC010699.

946 Watterson, I. G. (1996), Non-dimensional measures of climate model performance. *Int. J. Climatol.*,
947 16: 379–391. doi: 10.1002/(SICI)1097-0088(199604)16:4
948
949 Watterson, I. G., J. Bathols, and C. Heady, 2014: What Influences the Skill of Climate Models over the
950 Continents?. *Bull. Amer. Meteor. Soc.*, 95, 689–700. doi: <http://dx.doi.org/10.1175/BAMS-D-12->
951 00136.1
952
953 Zacharioudaki, A. and Reeve, D.E., 2011. Shoreline evolution under climate change wave
954 scenarios. *Climatic Change* 108, 73-105.
955
956
957
958
959
960

List of Tables

Table 1. CMIP Models (Phase 3 and 5) used in the study. Description of the ensemble means assessed in the study are also given.

ID	Full model name	Model	CMIP phase	Atm Resolution Lon x Lat
1	Australian Community Climate and Earth System Simulator 1.0	ACCESS1.0	5	1.88 x 1.25
2	Beijing Climate Centre, Climate System Model, 1-1	BCC-CSM1.1	5	2.8 x 2.8
3	Centre National de Recherches Meteorologiques Coupled Global Climate Model, version 5	CNRM-CM5	5	1.4 x 1.4
4	Geophysical Fluid Dynamics Laboratory Earth System Model 2M	GFDL-ESM2M	5	2.5 x 2.0
5	Hadley Centre Global Environmental Model 2, Earth System	HadGEM2-ES	5	1.88 x 1.25
6	Institute of Numerical Mathematics Coupled Model, version 4.0	INMCM4	5	2.0 x 1.5
7	Model for Interdisciplinary Research on Climate, version 5	MIROC5	5	1.4 x 1.4
8	Meteorological Research Institute Coupled Atmosphere-Ocean General Circulation Model, version 3	MRI-CGCM3	5	1.1 x 1.1
9	ECMWF Hamburg climate model, version 5, dynamically downscaled using CCAM	CMIP3-CCAM-ECHAM5	3	0.5 x 0.5
10	CSIRO Mk3.5 general circulation, version 3.5, dynamically downscaled using CCAM	CMIP3-CCAM-CSIROmk3.5	3	0.5 x 0.5
11	Ensemble mean of CMIP-5 simulations (models 1-8)	CMIP5-EM	-	-
12	Ensemble mean of CMIP3-CCAM simulations (models 9-10)	CMIP3-CCAM-EM	-	-
13	Ensemble mean of all simulations (models 1-10)	EM	-	-

Table 2. List of indices considered in the evaluation analysis using relative error matrices. All indices were determined for each of the four variables ($H_{S,m}$, $H_{S,x}$, $T_{M,m}$ and $D_{M,m}$).

Index	Index name
1.	Climatological mean of annual value [Mean]
2.	Mean seasonal amplitude [SsnAmp]
3.	Linear time trend of annual values [Trend]
4-7.	Spatial Correlation of observed to modeled climatological mean field, by season [SpR.MMM]
8-11.	Spatial Standard Deviation of the climatological mean field, by season [SpSD.MMM]
12-15.	Time-series variance of seasonal mean time-series [TVar1.MMM]
16-19.	Time-series variance of 5-yr running mean of seasonal mean time-series [TVar5.MMM]

Table 3. Comparing model ranks for wave fields relative to terrestrial fields. Length is the variable defined by Watterson et al. (2014), as a measure of the climate model grid resolution. It is defined as the square root of the quotient of the Earth's surface area and the number of points on the model's data grid. The values provided for the CMIP5-EM are the average of the 8 models used in this study. Similarly, the value provided for the overall EM is the average of the 8 CMIP5 models and the two downscaled CMIP3 models. The Terrestrial M-scores are taken from Watterson et al. (2014), (Table 1), as the average of the values using CRU and ERA datasets. The relative rank is given in brackets. The global wave M-score is the M-value, averaged across all four wave variables, across four seasons, across the 3 hindcasts.

Model	Length	Watterson M-score	Wave M-score (global). Rank in brackets.
ACCESS1.0	135	745 (2)	758 (8)
BCC-CSM1.1	250	682 (5)	737 (11)
CNRM-CM5	125	760.5 (1)	548 (13)
GFDL-ESM2M	198	668 (7)	770 (5)
HadGEM2-ES	135	742.5 (3)	755 (9)
INMCM4	154	668 (7)	752 (10)
MIROC5	125	696.5 (4)	727 (12)
MRI-CGSCM3	100	679.5 (6)	764 (7)
CMIP3-CCAM-ECHAM5	62	-	768 (6)
CMIP3-CCAM-CSIROMk3.5	62	-	777 (3)
CMIP5-EM	153	-	785 (2)
CMIP3-CCAM-EM	62	-	773 (4)
EM	135	-	799 (1)

993 Table 4. Ensemble mean M-score, for global domain, by variable.

<i>Variable</i>	<i>M</i>
$H_{S,m}$	845
$H_{S,x}$	723
$T_{M,m}$	814
$D_{M,m}$	830

994

995

996

997 Table 5. Relative ranking of GCMs according to the Relative Error Matrices for the 13 sub-domains,
 998 and the global domain. After determining the total relative error per model, by summing across
 999 metrics listed in Table 2, the models are ranked with the models demonstrating greatest skill having
 1000 rank 1 and those with least skill having rank 13.

<i>Domain</i>	<i>ACCESS1.0</i>	<i>BCC-CSM1.1</i>	<i>CNRM-CM5</i>	<i>GFDL-CM3</i>	<i>HadGEM2-ES</i>	<i>INMCM4</i>	<i>MIROC5</i>	<i>MRI-CGCM3</i>	<i>C3-ECHAM5</i>	<i>C3-Mk3.5</i>	<i>CMIP5-EM</i>	<i>CMIP3-EM</i>	<i>EM</i>
<i>TNIO</i>	5	10	12	9	3	13	8	7	4	11	2	6	1
<i>TSIO</i>	10	12	13	8	9	11	7	5	2	6	4	3	1
<i>ETSI</i>	8	11	13	4	9	10	7	3	5	12	2	6	1
<i>ETNP</i>	5	9	13	7	6	11	4	3	12	8	1	10	2
<i>TWNP</i>	7	6	12	9	10	13	11	3	4	8	2	5	1
<i>TENP</i>	6	7	13	11	10	12	3	4	5	9	2	8	1
<i>TWSP</i>	8	12	13	7	11	9	4	1	5	10	3	6	2
<i>TESP</i>	4	11	13	10	6	9	3	12	5	8	2	7	1
<i>ETSP</i>	10	8	13	5	7	11	12	1	4	9	6	3	2
<i>ETNA</i>	6	13	12	10	11	9	5	8	3	7	2	4	1
<i>TNAO</i>	7	10	13	11	8	12	9	3	5	6	2	4	1
<i>TSAO</i>	8	10	13	12	4	11	9	7	6	5	2	3	1
<i>ETSA</i>	8	7	13	4	5	11	10	1	9	12	3	6	2
<i>GLOB</i>	9	12	13	8	10	11	6	1	5	7	4	3	2

1001

List of Figures

Figure 1. Sub-domain of global ocean. Each sub-region is chosen on basis that the wave climate within each domain is qualitatively similar. The definition of the sub-regions follows that of Alves (2006).

Figure 2. Annual mean $H_{S,m}$ (m) for the period 1980-2005. (a) Multi-model (ensemble) mean (CMIP5-EM) constructed using the 8 CMIP5 models listed in Table 1. (b) Multi-model mean bias, as the difference between CMIP5-EM and the climatology from the G1d_cfsr hindcast. (c) Mean absolute model bias of CMIP5-EM with respect to the G1d_cfsr hindcast. (d) Mean inconsistency between G1d_cfsr, ERA-Interim and the corrected ERA40C wave heights, as the mean of the absolute pairwise difference between the three hindcasts (1980-2002).

Figure 3. Seasonality (January minus July) of $H_{S,m}$ (m) for the period 1980-2005. (a) Multi-model mean (CMIP5-EM) constructed using the 8 CMIP5 models listed in Table 1. (b) Difference between the multi-model CMIP5-EM mean and G1d_cfsr hindcast seasonality.

Figure 4. Annual mean $T_{M,m}$ (s) for the period 1980-2005. (a) Multi-model (ensemble) mean (CMIP5-EM) constructed using the 8 CMIP5 models listed in Table 1. (b) Multi-model mean bias, as the difference between CMIP5-EM and the climatology from the G1d_cfsr hindcast. (c) Mean absolute model bias of CMIP5-EM with respect to the G1d_cfsr hindcast. (d) Mean inconsistency between G1d_cfsr, ERA-Interim and the corrected ERA40C wave heights, as the mean of the absolute pairwise difference of the three hindcasts (1980-2002).

Figure 5. Seasonality (January minus July) of $T_{M,m}$ (s) for the period 1980-2005. (a) Multi-model mean (CMIP5-EM) constructed using the 8 CMIP5 models listed in Table 1. (b) Difference between the multi-model CMIP5-EM mean and G1d_cfsr hindcast seasonality.

Figure 6. Annual mean $D_{M,m}$ (°N) for the period 1980-2005. (a) Multi-model (ensemble) mean (CMIP5-EM) constructed using the 8 CMIP5 models listed in Table 1. (b) Multi-model mean bias, as the difference between CMIP5-EM and the climatology from the G1d_cfsr hindcast. (c) Mean absolute model bias of CMIP5-EM with respect to the G1d_cfsr hindcast. (d) Mean inconsistency between G1d_cfsr, ERA-Interim and the corrected ERA40C wave heights, as the standard deviation of the three datasets (1980-2002).

Figure 7. Seasonality (January minus July) of $D_{M,m}$ (°) for the period 1980-2005. (a) Multi-model mean (CMIP5-EM) constructed using the 8 CMIP5 models listed in Table 1. (b) Difference between the multi-model CMIP5-EM mean and G1d_cfsr hindcast seasonality.

Figure 8. M-score box plot. For each of 13 sub-regions, and globally, the range of M-scores across the 13 model sub-sets (8 CMIP5 models, 2 downscaled CMIP3 models, CMIP5-EM, CMIP3-EM, and whole EM) averaged across four seasons (JFM, AMJ, JAS, OND), four variables (H_s , H_{smax} , T_m and D_m), relative to three hindcast datasets (G1d_cfsr, ERA-Interim and (C)-ERA40C) is shown. The blue boxes represent the interquartile range (25th - 75th percentiles), the red bar shows the mean, the whiskers represent the 95% confidence intervals, and red + symbols as outliers. Also shown are M-scores, averaged across seasons and variables, for comparisons between hindcast datasets. The solid circles / black x's / black triangles representing G1d_cfsr vs ERA-Interim / G1d_cfsr vs ERA40C / ERA-Interim vs ERA40C.

Figure 9. M-scores for the 13 model simulations as outlined in table 1. i.e., 8 CMIP5 models, 2 downscaled CMIP3 models, CMIP5-EM, CCAM-CMIP3-EM, and EM) for each of 13 sub-regions, and globally, averaged across four seasons (JFM, AMJ, JAS, OND), four variables (H_s , H_{smax} , T_m and D_m), relative to three hindcast datasets (G1d_cfsr, ERA-Interim and (C)-ERA40C.

Figure 10. Relative Error Matrix for the global domain. The columns of the matrices correspond with each of the GCM forced wave simulations, with ID numbers given in Table 1. Rows of the matrices correspond with metrics defined in Table 2. The bottom row corresponds to first index for $H_{s,m}$. The rows above correspond to following indices, before being repeated for $H_{s,x}$, $T_{M,m}$ and $D_{M,m}$. The row labelled with the variable corresponds to the annual mean index in each case. The color scale is relative, with white cells indicating zero error (agreement between model and hindcast) to black indicating that model has the maximum relative error of those considered for the given index. Models are ranked from having lowest total relative error in the left most column to having the largest total relative error in the right most column. The model rank varies by region, as indicated in Table 5, where ranks are specified for the global domain (aligned with this figure), and for each sub-domain.

FIGURE 1

FIGURE 2

FIGURE 3

FIGURE 4

FIGURE 5

FIGURE 6

FIGURE 7

FIGURE 8

FIGURE 9

FIGURE 10