Clim Dyn (2014) 43:1271–1283
DOI 10.1007/s00382-013-1939-x

Evaluation of CMIP5 dynamic sea surface height multi-model simulations against satellite observations

Felix W. Landerer · Peter J. Gleckler · Tong Lee

Received: 17 May 2013 / Accepted: 6 September 2013 / Published online: 6 October 2013

Abstract We evaluate the representation of dynamic sea surface height (SSH) fields of 33 global coupled models (GCMs) contributed to the fifth phase of the Coupled Model Intercomparison Project (CMIP5). We use observations from satellite altimetry and basic performance metrics to quantify the ability of the GCMs to replicate observed SSH of the time-mean, seasonal cycle, and inter-annual variability patterns. The time-mean SSH representation has markedly improved from CMIP3 to CMIP5, both in terms of overall reduction in root-mean square differences, and in terms of reduced GCM ensemble spread. Biases of the time-mean SSH field in the Indian and Pacific Ocean equatorial regions are consistent with biases in the zonal surface wind stress fields identified with independent measurements. In the Southern Ocean, the latitude of the maximum meridional gradient of the zonal mean SSH CMIP5 models is shifted equatorward, consistent with the GCMs' spatial biases in the maximum of the zonal mean westerly surface wind stress fields. However, while the Southern Ocean SSH gradients correlate well with the maximum Antarctic circumpolar current transports, there is no significant correlation with the maximum zonal mean wind stress amplitudes, consistent with recent findings that the eddy parameterisations in GCMs dominate over wind stress amplitudes in this region. There is considerable spread across the CMIP5 ensemble for the seasonal and interannual SSH variability patterns. Because of the short observational period, the interannual variability patterns depend on the time-period over which they are derived, while no such dependency is found for the time-mean patterns. The model performance metrics for SSH presented here provide insight into GCM shortcoming due to inadequate model physics or processes. While the diagnostics of CMIP5 GCM performance relative to observations reveal that some models are clearly better than others, model performance is sensitive to the spatio-temporal scales chosen.

Keywords Sea surface height · CMIP5 · GCM skill · Model evaluation · AR5

1 Introduction

Sea level changes reflect the ocean's integral response to a broad spectrum of processes that affect the oceans currents and density structure, as well as the total ocean volume (Milne et al. 2009). Numerous studies have documented that sea level has risen globally throughout the twentieth century at a mean rate of about 1.8 mm/year (Meehl et al. 2007). The rate has increased to about 3.1 mm/year since 1993 (Church and White 2011), and is projected to increase further over the twenty-first century under global warming scenarios (Perrette et al. 2013). Both the observed and projected sea level changes are spatially highly non-uniform due to differential ocean warming, wind changes, the influence of ocean dynamics, and gravitational and solid earth responses (Milne et al. 2009; Yin 2012; Meyssignac et al. 2012; McGregor et al. 2012). The World Climate Research Program's (WCRP's) Coupled Model

Electronic supplementary material The online version of this article (doi:10.1007/s00382-013-1939-x) contains supplementary material, which is available to authorized users.

F. W. Landerer (✉) · T. Lee
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
e-mail: felix.w.landerer@jpl.nasa.gov

P. J. Gleckler
Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, CA, USA

Intercomparison Project (CMIP; Phase 3 and 5) includes an unprecedented set of climate simulations from coupled general circulation models (GCMs) for the recent past as well as for future climate (Meehl et al. 2007). This provides a unique opportunity to assess how well climate models simulate key characteristics of sea surface height (SSH) with high quality satellite data, as has been done for other observables (e.g. Kwok 2011; Li et al. 2012; Lee et al. 2013; Jiang et al. 2012).

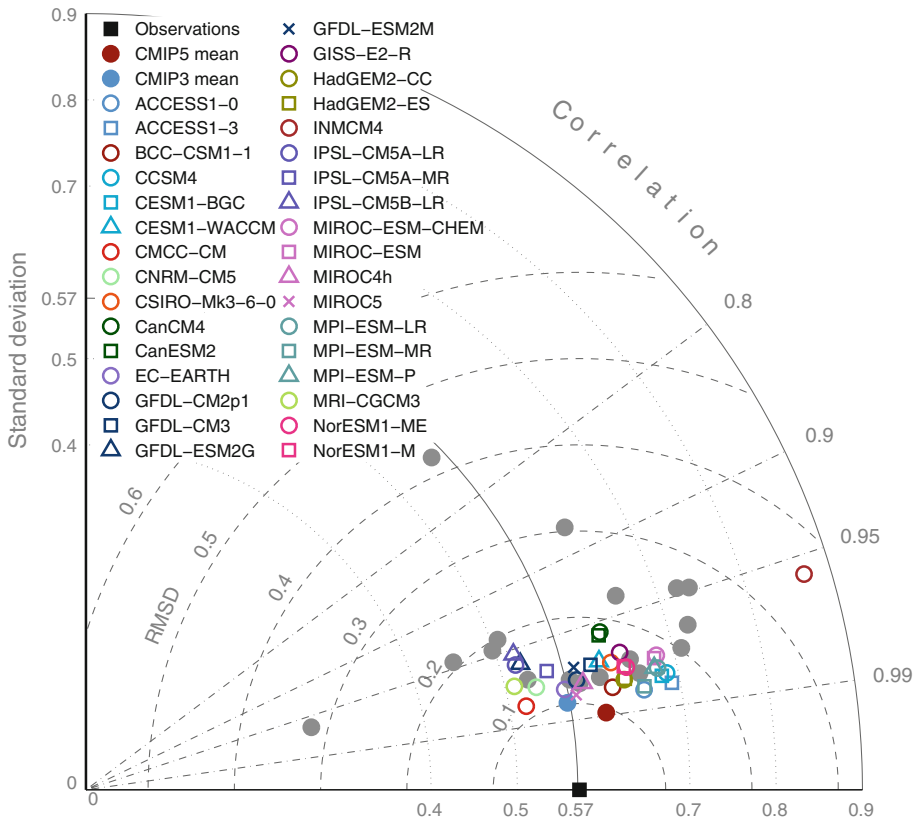
The main goal of our paper is to evaluate how well the spatial and temporal features of SSH patterns simulated by the CMIP5 models compares to available observations. All GCMs show global steric sea level rise from heat uptake under twenty-first century warming scenarios, but the individual magnitudes vary (Yin 2012). In addition, regional sea level changes can deviate by up to $\pm$ 100 % from the global mean changes (Yin 2012; Landerer et al. 2007; Yin et al. 2010a; Pardaens et al. 2010), but these regional patterns show only a few regions of agreement across the various models (Perrette et al. 2013; Yin 2012). The CMIP3 multi-model ensemble (MME) exhibited significant inter-model spread in the magnitude of projected SSH-change patterns, such that over large areas of the global oceans the inter-model differences towards the end of the twenty-first century were larger than the model-mean change (Meehl et al. 2007). Notable exceptions to this were areas of consistent SSH change in the Southern Ocean, and patches in the Arctic, South-East Pacific and West-Indian Ocean (see Fig. 10.32 in Meehl et al. 2007).

In an effort to reduce the ensemble spread (Yin et al. 2010a), used the global mean root-mean square difference (RMSD) between the GCMs' and observations' time-mean SSH fields to exclude 5 of 17 CMIP3 models from the multi-model ensemble, which slightly improved the inter-model agreement for projected SSH changes at the end of the twenty-first century. Here, we expand the model evaluation to annual and interannual time scales as described in Sect. 2 Our results highlight various improvements in the new CMIP5 ensemble over CMIP3, and also aspects of consistent model biases that provide new insights into deficiencies and shortcomings of the underlying model formulations and physics. The paper is organized as follows: in Sect. 2, we describe the methods and data sets and observations; in Sect. 2.2, we summarize methods and analysis schemes; in Sect. 3, we present results of the various skill metrics and comparisons for the CMIP3 and CMIP5 ensembles, and in Sect. 4 we discuss the main conclusions of this paper.

2 Methods and data description

Our analysis focuses on dynamic SSH, which is defined as the local sea surface height deviation from the global mean. Therefore, the global mean of SSH is zero at every time step, and we do not consider global mean sea level changes here. The latter can be steric due to net ocean heat changes, or non-steric due to net ocean mass changes from melting land ice. For ocean warming, global mean sea level can be computed off-line even if the GCMs employ a volume-conserving Boussinesq approximation (Greatbatch 1994; Griffies and Greatbatch 2012). While the dynamic SSH patterns related to surface momentum and buoyancy fluxes are explicitly and adequately simulated (Yin et al. 2010a), the current generation CMIP3 and CMIP5 models do not account for net ocean mass changes from melting glaciers and ice sheets. If associated sea level changes were uniform, the global mean could simply be subtracted from the observations. Hoewever, due to gravitational and loading effects on sea level (Farrell and Clark 1976), net ocean mass changes have a distinct non-uniform SSH pattern, with the largest deviations from a uniform rise occuring in the near field of the mass sources (Tamisiea et al. 2001). For the purposes of this paper, the issue is then whether land-ice fingerprint signals might be present in the 20-year altimetry record, and whether they could lead to a bias in the comparison to the CMIP models.

Several recent papers have explored detection thresholds for land-ice fingerprints, assuming specific levels of ice melt (Kopp et al. 2010; Hay et al. 2012). While these studies differ slightly in their assumptions of melt rates, background SSH variations and observing systems (i.e., relative sea level or sea surface height), a common result is that accelerating ice melt rates should be detectable via their associated sea-level fingeprints over the next decades. However, for the lower ice-melt rates over the last 20 years, the gravitational and loading effects are likely masked by dynamic SSH variability. This is consistent with results from a joint-inversion approach, using both altimetry and time-variable gravity observations, that found altimetry observations to currently have at best only marginal resolution capability of mass-related sea level changes (Rietbroek et al. 2012). Based on contemporary geocentric sea-level fingerprints (Riva, pers. communcation, 2013), we also note that the CMIP-biases discussed below are an order of magnitude larger (and vary in sign) near the ice-sheets than the expected fingerprint amplitudes over the last 20 years. Therefore, we perform our following analysis under the assumption that the non-homogeneous sea-level fingerprints from ocean mass changes do not significantly impact the comparison to CMIP models. Dynamic SSH from the CMIP models can then be directly compared to SSH observations from satellite altimetry as long as the global mean of each data set at every time step is removed using a common land mask (see inset in Fig. 2). In this way we ensure that



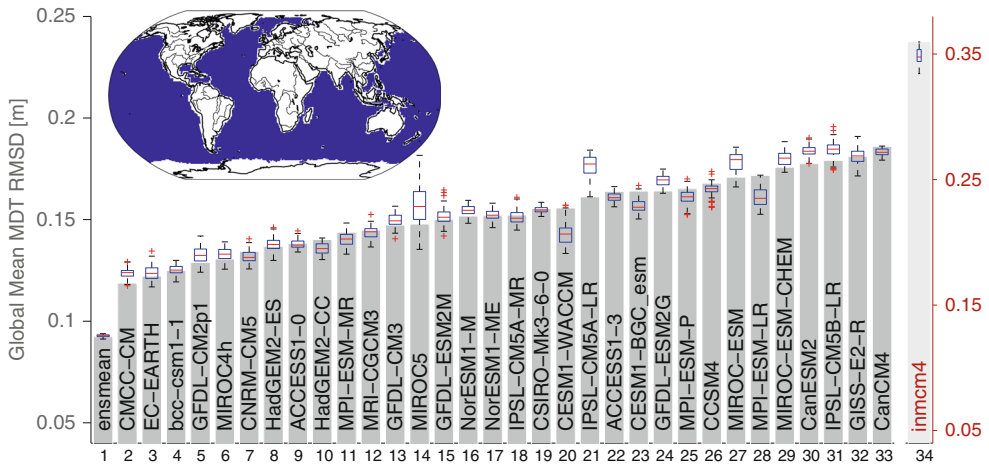


Fig. 1 Taylor diagram of the CMIP3 (*grey dots*) and CMIP5 (*colored symbols*, see legend) mean dynamic topographies (MDTs) compared against the observed MDT from (Maximenko et al. 2009). The CMIP3 and CMIP5 ensemble averages (shown as *stars*) show the best agreement with observations (see map *inset* in Fig. 2 for ocean areas used). Note that CMIP3 RMSD values are consistent with those shown in Fig. 2 of (Yin et al. 2010b)

Fig. 2 Global mean RMSD between 33 CMIP5 models, the CMIP5 model mean, and the observations of the MDT from (Maximenko et al. 2009). The *grey bars* represent the model/observations RMSD value for the time-period 1993–2002, and the whiskers the distribution of RMSDs between the observed MDT and 10-year sliding windows for each CMIP5 GCM for the 'historical' twentieth century runs (*central red mark* is the median, *box edges* are 25th and 75th percentiles). Note the different *vertical scale* for the INMCM4 model on the *right*. CanCM4 is for years 1961–2005 only, HadGEM2-CC for 1959–2005 only, and MIROC4h for 1950–2005 only. The map *inset* shows the ocean area over which the model statistics were computed for all CMIP5 GCMs; marginal seas have been excluded

biases are not due to a global mean offset. As land-ice melt continous to accelerate and contribute to sea level changes in the future, source-specific SSH fingerprint patterns may need to be considered to avoid biases in model-to-observation comparisons.

2.1 Data sets

We use sea surface height above Geoid fields from 33 GCMs from the new CMIP5 multi-model ensemble, and also 18 CMIP3 GCMs to assess improvements from the old

to the new ensemble. We use the first available GCM realizations of the twentieth century experiments (identified as '20c3m' in CMIP3 and 'historical' in CMIP5) if available. Several GCMs in the CMIP5 ensemble differ only in their non-ocean model components (e.g., *NorESM1-ME* and *NorESM1-M*), but we include all GCM variants in our analysis. For CMIP3, the twentieth century experiments start between 1860 and 1900, and end in 2000; in CMIP5, the historical runs start between 1850 and 1960, and end in 2005. As the SSH observations extend into the twenty-first century (see below), we appended the CMIP3 '20c3m' runs with the appropriate model fields from an A1B scenario run (Nakicenovic and Swart 2000) to ensure consistent time overlap, but the choice of scenario is not critical as they are similar over this short extension period. For CMIP5, we use results from the 'historical' simulations only. Tests with a CMIP5 model subset revealed no significant sensitivities to the particular model years (e.g., 1993–2012 vs. 1993–2005 only) for the time-mean and annual climatology comparisons performed here. We exclude some marginal ocean areas or enclosed seas from our analysis where the models exhibit SSH biases that are clearly unrealistic and would unduly distort squared-difference metrics (e.g., MIROC-ESM CMIP5 models have SSH values of +15 m over Hudson Bay, and values of −15 m over the Mediterranean). We suspect that these biases have more to do with model resolution or how the diagnostic SSH calculation is performed rather than a fundamental deficiency in the simulation. An complete overview of CMIP3 and CMIP5 ocean and atmosphere GCMs and their resolutions can be found in Tables 3 and 4 of Jourdain et al. (2013). We note that the spatial resolution of the ocean components in the CMIP3 and 5 archive varies, but even the newer CMIP5 runs generally do not resolve oceanic eddies, which are thus parameterized. For the two CMIP5 GCMs GISS-E2-R and MIROC5, we added the equivalent water thickness of their respective sea ice fields to obtain an effective sea surface height (Griffies et al. 2009).

Insufficient spin-up of the control runs in CMIP-type GCMs often show residual drift that can be removed in the forced runs (Yin 2012; Yin et al. 2010a). In the following analysis we do not correct the historical runs for control run drift because (1) most of the drift maps into the global mean sea level which we remove, and (2) because the drift should have little impact on the mean annual cycle and interannual variations (see discussion below). Significant drift can hypothetically impact the time-mean dynamic SSH topography, but our analysis indicates that this is not the case here (see Sect. 3.1). However, for detection-attribution analyses and projections, a drift correction is crucial for estimating trends and variability. (Yin 2012; Gleckler et al. 2012).

We compare the CMIP SSH fields against observations of the time-mean dynamic topography (MDT) from (Maximenko et al. 2009), which combines observations from satellite altimetry, near-surface drifters, NCEP wind and observations from the Gravity Recovery and Climate Experiment (GRACE) over the time period 1992–2002. For time-variable SSH signals, we use satellite altimetry observations from 1993–2012 provided by AVISO (Ducet et al. 2000), which are now available as part of the 'obs4MIP' project (http://esg-gateway.jpl.nasa.gov). We re-grid all CMIP and observational data sets onto a common 1x1 latitude-longitude grid using bi-linear interpolation. We note that the time-mean model assessment is not dependend on the use of the Maximenko-MDT or the time-mean AVISO data. In fact, the MDT and time-mean AVISO fields have a pattern correlation of $R = 0.998$, and either one yields essentially identical in the following analysis (i.e., in Fig. 1). Since dynamic SSH is significantly influenced by surface momentum fluxes from wind over many regions, we also use the Scatterometer Climatology of Ocean Winds (SCOW) based on QuikSCAT satellite measurements (Risien and Chelton 2008) to assess if common biases between SSH and wind stress exist. This will enable us to evaluate features that we expect to be similar between SSH and wind stress using independent measurements. A detailed evaluation and comparison between CMIP3 and CMIP5 simulated surface wind stress and QuikSCAT observations can be found in (Lee et al. 2013).

2.2 Metrics

The choice of a metric to evaluate model performance is somewhat subjective but we use several basic statistical measures that are routinely used in meteorological and climate analysis, namely: the global mean statistics of root-mean-square differences (centered RMSD), spatio-temporal correlation, and standard-deviation. These basic measures can be examined collectively, e.g., in Taylor-diagrams (Taylor 2001) and are useful first steps for quantifying how well the models agree with observations as well as with each other. We also examine spatial fields of absolute biases between models and observations. While it is reasonable to examine model-to-observation agreement for the time-mean and annual cycle fields, interannual and longer variations are expected to differ as CMIP simulations are forced externally, but free to evolve in terms of internal variability. This, and the relatively short time-span of available global observations, makes it challenging to examine how the CMIP models agree with observations on interannual time scales. Therefore, we limit our CMIP-to-observations comparison on patterns and amplitudes of interannual RMS variability after removing a seasonal

climatology and band-pass filtering between 1 and 10 years, and we assess the impact of choosing different time-periods (see Sect. 3.3 for details).

3 Results

3.1 Time-mean SSH field

We first analyze the representation of the time-mean dynamic SSH fields, which are closely related to the time-mean ocean circulation that governs the large-scale transports of heat, freshwater, and nutrients. A Taylor diagram for the 33 CMIP5 MDT fields averaged over 1992–2002 (Fig. 1) reveals pattern correlations varying between 0.95 and 0.99, and standard deviations slightly larger than what is observed. The RMS differences between observations and CMIP5 models cluster between 12 mm (CMCC-GM) and 19 mm (CanCM4); the INMCM4 model (RMSD = 36mm) appears to be an outlier in this analysis. Similar to previous findings for many other model variables (Gleckler et al. 2008), the CMIP5 ensemble mean SSH-MDT yields the highest correlation and the lowest RMSD at about 9 mm. Based on the MDT Taylor statistics, the CMIP5 simulations have improved markedly over the CMIP3 runs (grey dots in Fig. 1) in several aspects. Firstly, the absolute RMSDs are generally reduced in CMIP5. Secondly, the spread of the global RMSDs among the different CMIP5 models has noticeably decreased (with the exception of INMCM4) from the CMIP3 spread, which featured RMSDs that varied by a factor of up to ≈ 3 across the ensemble (Yin et al. 2010a). Thirdly, the multi-model mean RMSD in CMIP5 is about 25 % smaller than in CMIP3 (note, however, that the CMIP3 ensemble consists of 18 models, whereas our CMIP5 ensemble consists of 33).

Oceanic variability time scales can be longer (several decades and longer) than what the MDT observations used here cover (11 years). Since unforced internal climate variability [e.g., El Niño-Southern Oscillation (ENSO), North Atlantic Oscillation (NAO), etc.] in the GCMs is not constrained to be synchronized to real-world occurrence, we wondered if the RMSDs between observed and CMIP MDTs might be due to long-term variability and changes over time. To examine this, we again computed the global RMSD over a sliding window of all possible 11-year mean SSH fields for each CMIP5 model over the twentieth century (from 1870 to 2005) against the single realization of the observed MDT. While the individual RMS differences of overlapping windows are not independent from each other, this approach demonstrates that differences in a model's RMSD due to an evolving 20 year climatology is relatively small compared to inter-model differences. Alternatively, one could look at all the realizations

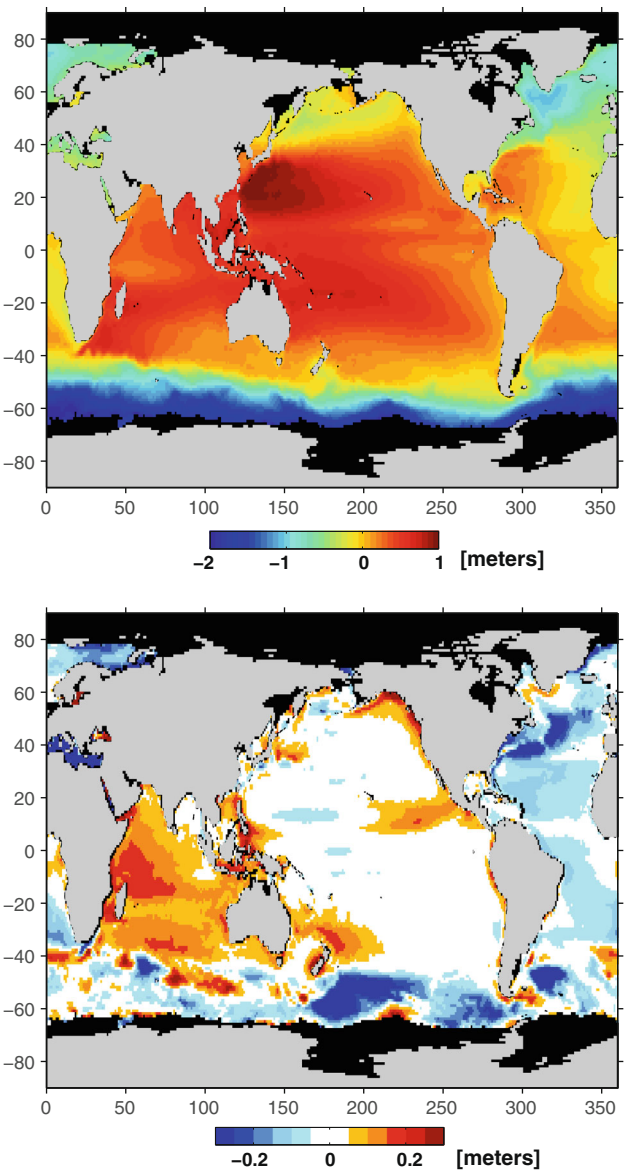


Fig. 3 *Top* mean dynamic topography observations from (Maximenko et al. 2009). *Bottom* bias between the ensemble mean of the 33 CMIP5 models and the observations. The bias was computed over the time-mean SSH for model years 1992–2002. For individual model biases, see Suppl. Material. Units: meters

available, as is commonly done, but the ensemble size available for many models is very small or limited to a single realization. With our approach, the use of a sliding window in a single realization, each model is evaluated in a consistent manner. The results of the sliding window analysis, shown as percentile boxes in Fig. 2, indicate that the RMSDs for each individual model are relatively stable over the twentieth century, with individual model RMSDs varying generally less than 10 %. We interpret these results such that interannual or decadal-scale variability is unlikely the main source of the differences between observed and simulated time-mean SSH fields.

The RMSD variations among the CMIP3 models are generally larger than the twentieth-century RMSD variations for each individual CMIP3 model (not shown), indicating that the biases are systematic—and different– for most of the models. In CMIP5, model-to-model differences are more similar in amplitude to an individual model's twentieth century variations. We also examined if the CMIP5-observations MDT RMSDs are consistently at a minimum (relative to the entire twentieth century) over the contemporaneous time-period (1992–2002, grey bars in Fig. 2). If that were the case, it could indicate that externally forced SSH changes (i.e., from greenhouse gases) have significantly and consistently affected the MDT in the CMIP simulations. The results in Fig. 2 are somewhat inconclusive in this regard: while for some models, the MDT RMSD appears to be at a minimum during the period overlapping with observations, other models have larger RMSDs in the overlapping time period. Not finding a consistent signature in this metric is perhaps not surprising given the global scale of the analysis, as well as the intrinsic variability (see Sect. 3.3). The spatial patterns of the MDT biases vary across the GCMs, but some consistent features emerge (Fig. 3, see also Suppl. Material Fig. 1SM). Most GCMs have prominent biases over the Southern Ocean and the Indian Oceans, and we thus examine these regions in more detail in the following sections. While much of the Pacific Ocean regions away from the equator are relatively well simulated in most GCMs, we also examine the equatorial Pacific biases due to the importance of this region for interannual climate modes.

3.1.1 *Equatorial regions*

Ocean dynamics in the equatorial regions are strongly influenced by surface momentum fluxes (wind stress). We therefore investigated if the CMIP5 MDT biases are consistent with biases of mean zonal surface wind stress in the equatorial regions between 2S and 2N. Because we are looking at individual ocean basin biases, the equatorial zonal mean SSH between 2S and 2N for each ocean basin (Pacific, Atlantic, and Indian Ocean) has been subtracted. In the Indian Ocean, the time-mean strength of Westerlies in the CMIP5 models is generally too weak relative to the QuickSCAT observations (Fig. 4). This in turn implies a too weak Indian Ocean equatorial SSH zonal gradient. Indeed, the mean equatorial upward tilt towards the east is consistently too weak in the CMIP5 models (Fig. 4). Similarly, the CMIP5 mean SSH biases in the tropical Pacific are consistent with the zonal wind stress biases: in the West, the models have a too strong easterly wind stress component that leads to a steeper SSH gradient than observed; in the Central Pacific, the models' zonal wind

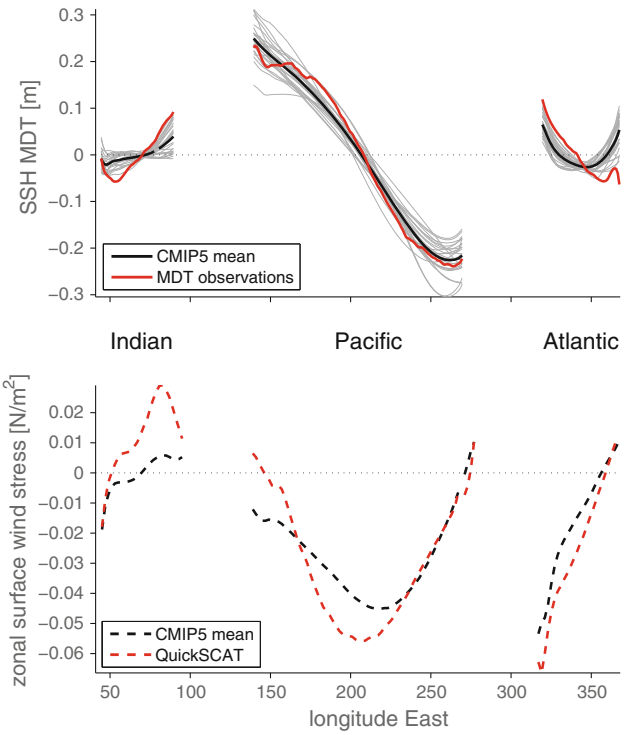


Fig. 4 *Top* absolute MDT bias between individual CMIP5 models (*grey*), the CMIP5 model mean (*black*), and the MDT observations (*red*). The biases were computed between 2S–2N over the time-mean SSH for model years 1992–2002, matching the observations. *Bottom* equatorial zonal wind stress (2S–2N) of the CMIP5 ensemble (*black*) and QuickSCAT observations (*red*); see (Lee et al. 2013) for a detailed analysis of individual CMIP5 GCMs

stress is too weak and leads to a weaker SSH gradient than observed; in the East, models and observations agree well for both zonal wind stress and SSH. Over the Atlantic basin, the easterly wind stress in CMIP5 is generally too weak, which contributes to weaker than observed CMIP5 SSH gradients. Other forcings than wind stress influence SSH variability in this region (McGregor et al. 2012), though it is not clear if this can explain the time-mean bias as well. Qualitatively, very similar results hold for the CMIP3 models (not shown here; see (Lee et al. 2013) for details).

A detailed assessment of surface wind stress in CMIP3 and CMIP5 simulations is discussed in (Lee et al. 2013). In the equatorial regions, the MDT structure is closely related to the time-mean vertical pycnocline structure, and hence any biases in MDT would be mirrored in the pycnocline depth. The too weak SSH gradients in the Indian and Atlantic Oceans correspond to a pycnocline that is too flat, which in turn influences the models' capabilities to properly generate tropical climate modes such as the Indian Ocean Zonal/Dipole Mode. A recent analysis has found that the unrealistic Indian Ocean pycnocline structure (which mostly depends on temperature) in the CMIP

models leads to a thermocline-SST feedback that is too strong, and hence an overestimate of Indian-Ocean dipole amplitude (Cai and Cowan 2013).

3.1.2 Southern Ocean

The time-mean SSH field in the Southern Ocean (primarily a down-sloping meridional gradient towards the South) is closely related to the strength of the Antarctic circumpolar current (ACC). The circumpolar flow is affected by the zonal momentum balance, surface buoyancy fluxes of heat and freshwater, and by the Southern Ocean overturning. The zonal current velocities are related to the meridionally tilted isopycnal surfaces set up by the Ekman transport and overturning circulation in the Southern Ocean (Gill 1982). However, significant poleward eddy-induced transports tend to reduce the meridional isopycnal gradients, which would tend to reduce the ACC transport. Due to their relatively coarse horizontal resolution, the ocean models used in coupled CMIP simulations require a parameterisations for these eddy-induced transports, for example through a quasi-Stokes diffusivity constant based on (Gent and McWilliams 1990) (also Griffies 1998, and references therein).

Here, we evaluate two questions regarding the Southern Ocean mean dynamic topography biases seen in CMIP5 models: (1) Are biases in the latitudinal location of the maximum mean zonal westerly wind stress field consistent with biases in the latitudinal location of the maximum meridional sea surface height gradient $(SSH^{\bar{x}}_{dy})_{max}$, and (2) are there significant correlations between the maximum ACC transports, westerly wind stress maxima, and maximum meridional sea surface height gradients? Several recent papers have documented coherent equatorward biases of the maximum westerly wind stress (and hence the westerly jets that drive the ACC) in CMIP5 models (Swart and Fyfe 2012; Meijers et al. 2012). Consistent with this bias, we find that the maxima of the zonal mean MDT meridional gradients also show a clear tendency for being biased northward compared to observations, although the spread is larger than for the zonal wind stress maxima (Fig. 5). As far as correlations between the maximum amplitudes of $(SSH^{\bar{x}}_{dy})_{max}$, ACC transport and westerly wind stress are concerned, we found no significant relationship between the maxima of westerly wind stress and $(SSH^{\bar{x}}_{dy})_{max}$ (grey dots in Fig. 6). Similar conclusions were presented recently by (Meijers et al. 2012), who showed that the maximum ACC transport is not significantly correlated with the maximum wind stress in the CMIP5-GCM ensemble. Furthermore, (Kuhlbrodt et al. 2012) also found that the ACC transport in CMIP3 models is not significantly correlated to the maximum westerly wind stress

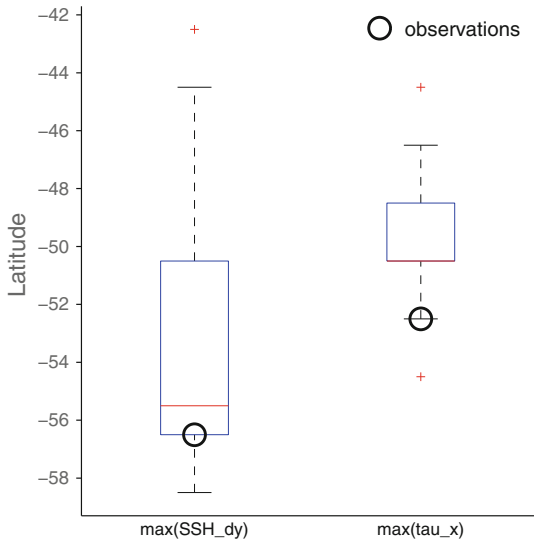


Fig. 5 Comparison of the zonal mean latitudinal position of the maxima of the meridional SSH gradients and westerly surface wind stress maxima in the 33 CMIP5 models (*box-plots*; median is shown in *red*, *box edges* are at the 25th and 75th percentiles), and for the observations (*black dots*). See (Lee et al. 2013) for details on the wind stress analysis

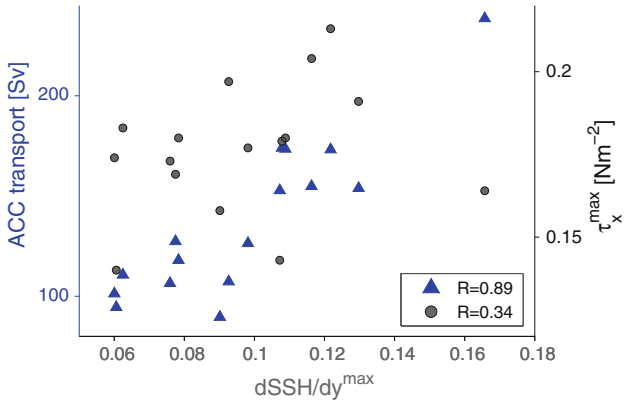


Fig. 6 Scatter *plot* of the maxima of westerly surface wind stress (*gray circles*), and maximum ACC transports through drake passage (*blue triangles*) in CMIP5 models as a function of the zonal mean maxima of the Southern Ocean meridional SSH gradients. Zonal wind stress and ACC values are taken from Table 3 of (Meijers et al. 2012), and plotted against the corresponding SSH values

amplitude, whereas the models' meridional density difference $\Delta\rho_y$ across the ACC was. Our results, using $(SSH^{\bar{x}}_{dy})_{max}$ as a first-order proxy for ACC transport (the correlation is R = 0.89, Fig. 6), are in line with both of these studies. As discussed in detail by (Kuhlbrodt et al. 2012) based on CMIP3 simulations, the quasi-Stokes diffusivity parameterisations κ of eddy-induced transports, used in many CMIP models, can explain the large across-model variance of the simulated ACC maximum transports. Therefore, wind stress is not the dominant factor that determines ACC transports and the corresponding Southern

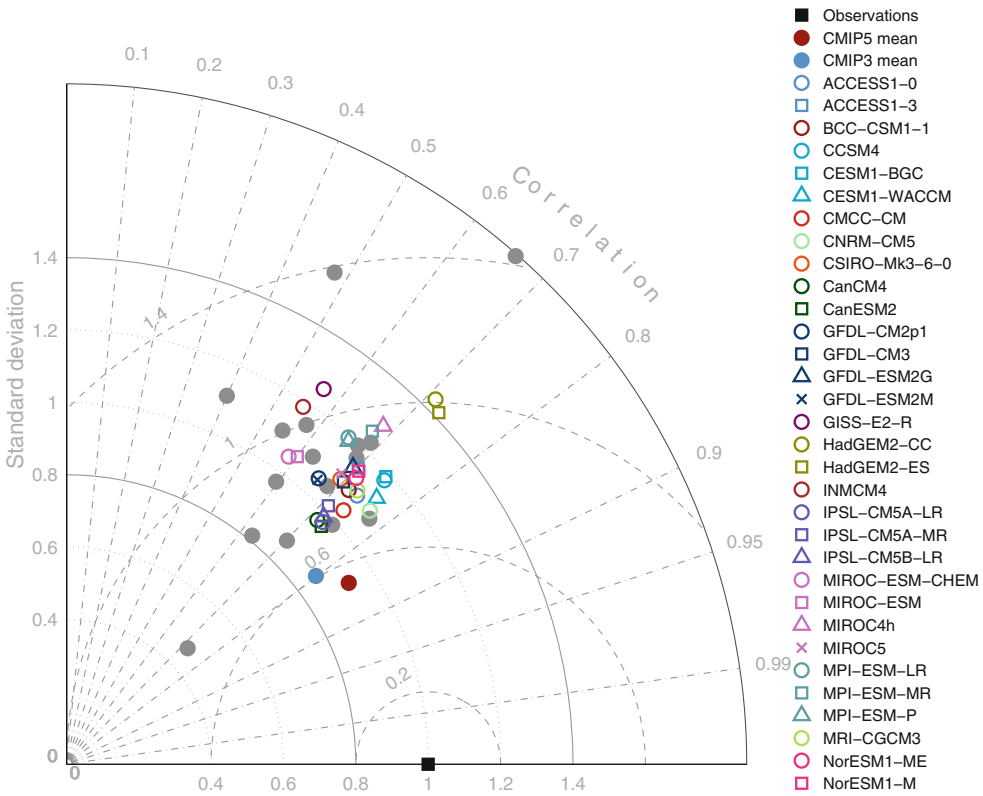


Fig. 7 Taylor diagram summarizing the normalized statistics of the mean monthly climatology SSH fields of the CMIP5 models (see legend), and CMIP3 models (*grey dots*). For each model, 12 monthly maps are used to compute the Taylor statistics (see Methods for details)

Ocean SSH patterns. Many CMIP5 models employ similar parameterisations as used in CMIP3, and hence the dependency of ACC transports and sea surface height in the Southern Ocean would be strongly dependent on the ocean models' values of κ, rather than on zonal wind stress magnitudes.

The detailed processes, in particular from unresolved eddies, that affect the relationship between mean zonal wind stress biases and SSH biases as shown here are not easily discernible from the zonal mean analysis (Figs. 5 and 6). Several studies have emphasized the important role of eddies in the Southern Ocean momentum balance (Boning et al. 2008), such that one may not expect linear relationships between wind stress, ACC transport and SSH changes. Also, surface buoyancy fluxes (heat and freshwater) have a significant influence on ocean dynamics at mid to high latitudes, and therefore any biases in these fields likely contribute to the SSH biases described above (Meijers et al. 2012; Russell et al. 2006; Carman and McClean 2011).

3.2 Monthly climatology SSH variations

In this section, we briefly discuss the representation of the monthly climatology anomalies of SSH in CMIP5 compared to the climatology of the AVSIO altimetric data. The monthly climatologies, relative to the time-mean fields, are derived from the model years 1993 through 2005 for the CMIP5 historical runs, and from the model years 1993 to 2010 for CMIP3 (twentieth century runs extended by the corresponding A1B scenario run); we tested and confirmed that the individual seasonal model statistics are not sensitive to the particular choice of the time periods. Seasonal variations of SSH in the tropical regions are dominated by surface wind field changes, whereas in the higher latitudes buoyancy fluxes and heat content changes tend to be more important. As we did for the time-mean fields, we computed Taylor statistics of each CMIP5 model, but this time for the 12 monthly climatology anomaly SSH fields for each model (Gleckler et al. 2008).

From Fig. 7, it is apparent that the spread of the CMIP5 ensemble is somewhat lower than the spread of the CMIP3 ensemble, though the level of improvement is modest, and the CMIP3 and CMIP5 ensemble mean annual Taylor statistics are almost identical. However, no single model achieves a correlation score above 0.8, in stark contrast to other variables that have a strong annual cycle such as surface temperature (Gleckler et al. 2008). The reasons for these comparatively low correlations may be related to generally smaller ratios of annual versus interannual variability in the oceans when compared to the atmosphere. In addition, SSH is influenced by surface heat fluxes as well as wind stress curl, making the seasonal SSH dynamics more complex than surface temperatures. At least part of the higher than observed seasonal standard variations in many CMIP5 models is contributed by too large a seasonal

SSH amplitude in the equatorial and tropical regions, in particular over the Indian and Pacific Ocean (see Suppl. Material Fig. 2). Compared to observations from the QuikSCAT instrument, many CMIP5 models tend to overestimate the overall magnitude of the seasonal wind stress anomalies, especially in spring and autumn (Lee et al. 2013). Consistent with this, the median of the global mean seasonal SSH amplitudes across the individual CMIP5 GCMs is about 13 % larger than the observed AVISO signal. However, while most CMIP5 models overestimate the observed spatio-temporal standard deviation, the model mean standard deviation is about 10 % lower than observed, indicating that individual model biases tend to average out in the ensemble mean (Fig. 7 and SM-Fig. 2).

3.3 Interannual SSH variations

Comparing variability on interannual to decadal scales between models and observations is not straight forward as the GCMs' internal variability is not constrained to be in phase with observations. In this sense, RMS differences or temporal correlations are not appropriate metrics to assess model performance for longer than seasonal time scales; instead, we focus on comparing the patterns and amplitudes of inter-annual variability, similar to (Gleckler et al. 2008). To extract these interannual patterns, we remove for each model and the observations the mean annual cycle, detrend the SSH data, and apply a band-pass filter with corner frequencies of 1 and 10 years; interannual variability is then the standard variation at each gridpoint of the filtered fields. The choice of 10 years as the maximum period is motivated by the limited length of the observational record. The SSH observations from altimetry cover 18 years, and we use the same record length (i.e., 1988–2005) to generate the interannual variability maps for the CMIP5 models (see SM-Fig. 3). We note that detrending and band-pass filtering the de-seasonalized SSH observations (1993–2010) reduces the global mean inter-annual variability by about 50 %; detrending only reduces the inter-annual variability by 12 %.

In the equatorial-tropical latitude band between 20S and 20N, the spatial pattern of observed interannual SSH variability as defined here is dominated by tropical Pacific variability related to ENSO and variability in the Indian Ocean. In the subtropical to subpolar latitudes, variability peaks appear in the meandering parts of Kuroshio extension and Gulf Stream/North Atlantic Current, and along the ACC boundary (Fig. 8, top). The average interannual variability pattern of the CMIP5 models broadly replicates these features, albeit at lower amplitudes (e.g., in the tropical Pacific Ocean; Fig. 8). The spatial amplitude pattern of the 1-10 year interannual variability varies among the CMIP5 models (see SM-Fig. 3). The main observed

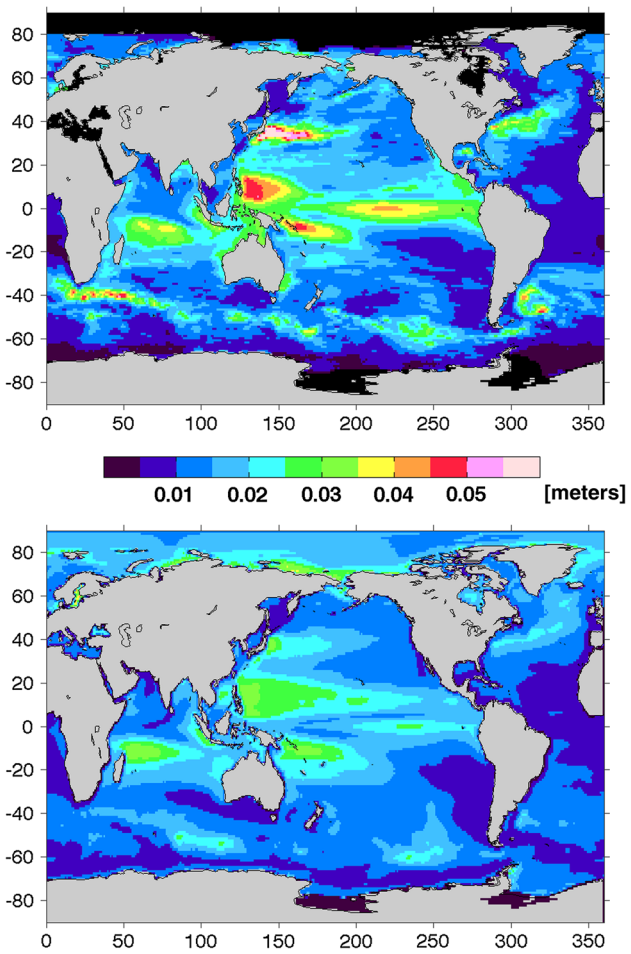
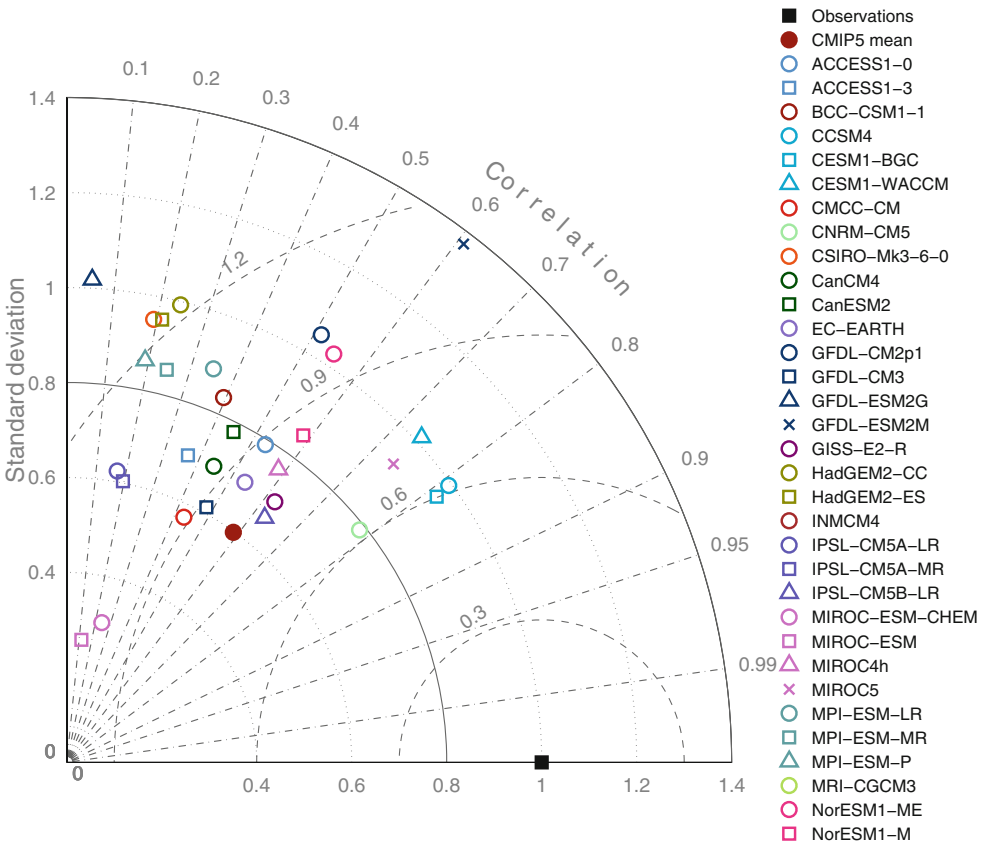


Fig. 8 Spatial patterns of interannual SSH variability of the AVISO observations (*top*) and the mean of the interannual variability patterns of the CMIP5 models (*bottom*). The monthly mean climatology has been subtracted, the data have been detrended, and a band-pass filter (1–10 years) has been applied

features - high variability in the equatorial Central to East Pacific, western Pacific warm pool, tropical Indian Ocean, North Atlantic Gulf Stream, Pacific Kuroshio, and Southern Ocean ACC front - are reproduced in the models, albeit to varying degrees. For example, the NorESM1 models appear to be able to capture the patterns and amplitudes in the Indian and Pacific Oceans quite well, whereas the two MIROC-ESMs underestimate amplitudes in these regions (see SM-Fig. 3). As discussed above, biases in the mean fields can have an effect on a model's capability to generate climate variability modes, in particular in the tropical regions where the pycnocline depth plays an important role in generating the Indian Ocean Dipole (Cai and Cowan 2013).

Moving to a more quantitative comparison between observed and CMIP5-simulated interannual variability, we calculated Taylor-diagrams from the interannual amplitude patterns. Other authors have analyzed interannual variability by using Taylor-diagrams, and pattern correlations



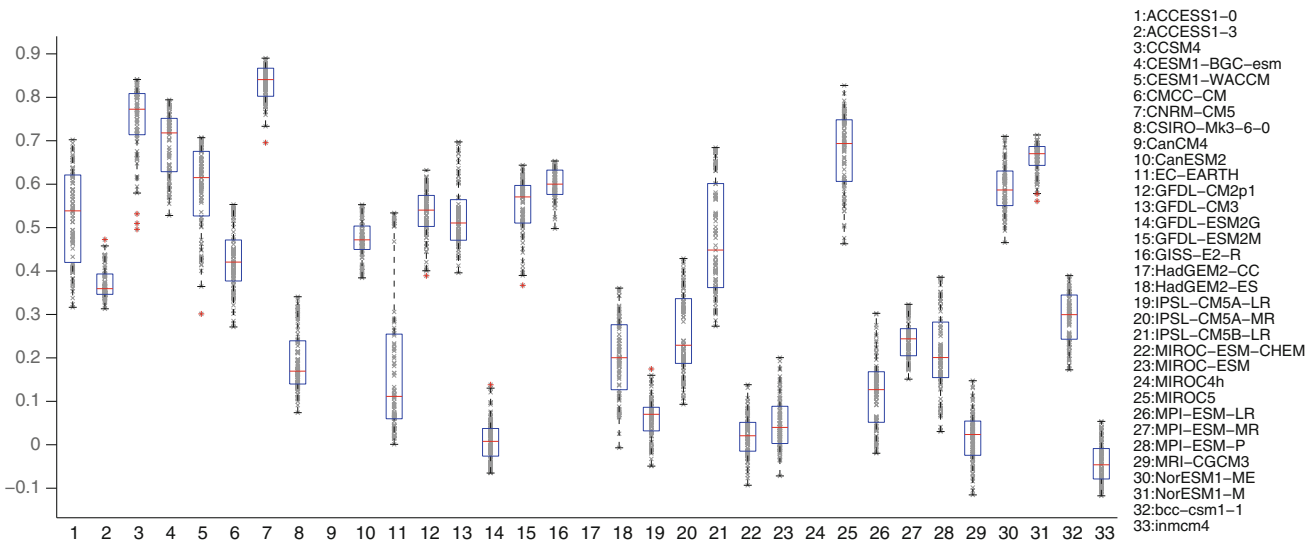


Fig. 9 Taylor diagram summarizing the normalized statistics of interannual SSH variability patterns of the CMIP5 models over the tropical Pacific Ocean area (between 20S and 20N). The monthly mean climatology has been subtracted, the data have been detrended, and a band-pass filter (1–10 years) has been applied. Note that models INMCM4 and MRI-CGCM3 have negative correlations (see also Fig. 10) and are thus not displayed here

Fig. 10 Spread of pattern correlation between interannual SSH variability of the CMIP5 models and observations over the tropical Pacific Ocean area (between 20S and 20N). For each model, the data points are derived by sliding a 18-year window over the historical run, and computing the pattern correlation to the observed interannual variability (the latter remains the same). Data processing and filtering is as in Fig. 9

using the first EOF-moments of SSH-variability or trends (Meyssignac et al. 2012; McGregor et al. 2012). We focus the following analysis of interannual variability on the Pacific between 20S and 20N because interannual variability in this region is comparatively large (Fig. 9; see Suppl. Material for a global analysis). The most striking feature of Fig. 9 is the large ensemble spread, and the reduces performance of the models compared to time-invariable metrics. The highest pattern correlation is about $R = 0.8$, whereas all CMIP5-GCMs reach at least

R = 0.95 for the time-mean SSH fields (Fig. 1). As can already be inferred from the spatial patterns (see SM-Fig. 3), the CMIP5 amplitude variations relative to the observed interannual variability are quite large, with many models underestimating the observed amplitudes in the tropical Pacific, or having different spatial patterns.

Arguably, comparing the interannual variability patterns in this way does not provide detailed insight into the underlying variability dynamics and processes for the individual models. The CMIP5 ensemble as a whole has improved in key aspects of simulating ENSO (Bellenger et al. 2013) over CMIP3; however, our analysis here uses only 18 years of observed SSH data, which is too short to fully capture episodic events such as ENSO, even though the frequency peak of 3-8 years is in principle captured in the band-pass limits. Other large-scale climate modes such as the NAO, PDO, or Atlantic Multi-decadal Oscillation have time scales beyond 18 years, and can thus strongly impact the agreement between simulations and observations. To estimate the potential effects of the limited observational time period, we repeated the analysis of the interannual variability again by sliding 18-year windows over the model years from 1900 through 2005. Figure 10 shows the range of the pattern correlations for each CMIP5 model obtained in this way. It is readily apparent that the variations for individual models can be quite large (e.g., between 0.3 and 0.7 for ACCESS1-0), presumably due to the lack of synchronization between simulated and observed interannual variability and the episodic nature of interannual variations. On the other hand, some models (e.g., CNRM-CN5 and NorESMM1-M) always out-perform other models, which could indicate better skill for the spatio-temporal scales examined here.

4 Conclusions

We have assessed and quantified how well coupled climate models of the CMIP5 climate model ensemble can represent the mean-state, annual climatology, as well as interannual variability pattern of sea surface height variations by comparing them against near-global satellite observations, available since 1993. We find substantial improvements in the time-mean SSH bias in CMIP5 over CMIP3, both in the individual model RMS reduction, as well as in a reduction of ensemble model spread. The GCMs often exhibit biases in the equatorial and Southern Ocean areas, and these biases are broadly consistent with the biases in the models' wind stress fields. The mean-state biases of SSH may also affect how the models are able to correctly simulate climate variability modes that may interact or depend on the mean climate state, which is particularly relevant for Indian Ocean Dipole and ENSO variability.

In the Antarctic circumpolar current region of the Southern Ocean, our results indicate that biases in the latitudinal position of the meridional maximum SSH gradients are consistent with the corresponding latitudinal positioning biases of the zonal wind stress maxima. In both cases, we observe a significant equatorward shift of approximately 4 degrees for $(SSH^{\bar{x}}_{dy})_{max}$, and approximately 3 degrees for τ_x. Consistent with recent findings (Meijers et al. 2012; Kuhlbrodt et al. 2012), we find that the maximum meridional SSH slope across the ACC correlates well with the maximum ACC transport, but not with the maximum zonal surface wind stress. The implication here is that the GCMs' parameterisations schemes of the Southern Ocean eddy-transports are a dominant factor that influence model SSH biases.

Employing a sliding-window analysis over the CMIP5 historical runs, we found no significant dependency on the time-scales over which the time-invariant model-to-observations differences are evaluated. For annual variations, CMIP5 models are clustered closer together than the models of the CMIP3 ensemble. However, CMIP5 models tend to overestimate the annual standard deviation by up to 30 %, and the spatio-temporal correlations to the observations are lower than for atmospheric surface variables (e.g., SST), and have not improved significantly from CMIP3. Over the equatorial regions, the seasonal amplitude CMIP5 model SSH biases are consistent with the seasonal surface wind stress biases (Lee et al. 2013). Comparing simulated and observed variability on interannual time scales remains a challenge because internal, unforced climate variability between models and observations is not synchronized. Focusing on variability time-scales of 1–10 years in the equatorial Pacific between 20S and 20N, we found a large spread of CMIP5 model performance, with correlation values ranging from 0.05 to 0.8. We note, however, that contrary to the time-invariant SSH fields, the interannual metrics used here are quite sensitive to the choice time-period that is used for the comparison.

As several previous studies have found for other climate variables, we also find for sea surface height that the GCMs' performance is rather sensitive to spatial and temporal scales and the choice of metrics (Taylor 2001; Reifen and Toumi 2009; Knutti 2010; Weigel et al. 2010). Model-to-observations differences are complex and frequency-dependent, and the observed seasonal cycle and interannual variability patterns offer more stringent tests of model performance (Santer et al. 2009). The representation of the time-varying fields is more challenging as the accurate representation of the underlying physics is of greater importance, rather than model tuning (Santer et al.

2009). In addition, due to the lack of a clear and consistent relationship between the mean-state SSH skill and the interannual variability representation, caution should be used to derive and apply skill scores from the metrics discussed here to weigh individual models when forming an ensemble average. Additionally, various recent studies (Reifen and Toumi 2009; Knutti 2010; Räisänen and Ylhäisi 2012) have demonstrated that common skill levels of GCMs do not necessarily map into projections that are close to each other. Regardless of the many issues around and approaches for GCM averaging and model weighting, the metrics used here are able to identify model outliers and help to reveal consistent model biases related to inadequate model physics or resolution that need to be improved.

Acknowledgements We acknowledge the GCM modeling groups, the PCMDI, and the WCRP's Working Group on Coupled Modeling for their roles in making available the WCRP CMIP3 and CMIP5 multimodel data sets. Support of these data sets is provided by the Office of Science, US Department of Energy. FWL's and TL's work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA.

References

Bellenger H, Guilyardi E, Leloup J, Lengaigne M, Vialard J (2013) ENSO representation in climate models: from CMIP3 to CMIP5. Clim Dyn 1–20. doi:10.1007/s00382-013-1783-z

Boning CW, Dispert A, Visbeck M, Rintoul SR, Schwarzkopf FU (2008) The response of the Antarctic circumpolar current to recent climate change. Nat Geosci 1(12):864. doi:10.1038/ngeo362

Cai W, Cowan T (2013) Why is the amplitude of the Indian Ocean dipole overly large in CMIP3 and CMIP5 climate models? Geophys Res Lett :n/a–n/a. doi:10.1002/grl.50208

Carman JC, McClean JL (2011) Investigation of IPCC AR4 coupled climate model North Atlantic mode water formation. Ocean Model 40(1):14. doi:10.1016/j.ocemod.2011.07.001 URL http://www.sciencedirect.com/science/article

Church J, White N (2011) Sea-level rise from the late 19th to the early twenty-first century. Surv Geophys 32:585. doi:10.1007/s10712-011-9119-1

Ducet N, Le Traon PY, Reverdin G (2000) Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and-2. J Geophys Res 105(C8):19477. doi:10.1029/2000JC900063

Farrell WE, Clark JA (1976) On postglacial sea level. Geophys J Int 46(3):647. doi:10.1111/j.1365-246X.1976.tb01252.x

Gent PR, McWilliams JC (1990) Isopycnal mixing in ocean circulation models. J Phys Oceanogr 20:150. doi:10.1175/1520-0485(1990)020<0150:IMIOCM>2.0.CO;2

Gill AE (1982) Atmosphere-ocean dynamics. International geophysics series. Academic Press, New York

Gleckler PJ, Taylor KE, Doutriaux C (2008) Performance metrics for climate models. J Geophys Res 113(D6):D06104. doi:10.1029/2007JD008972

Gleckler PJ, Santer BD, Domingues CM, Pierce DW, Barnett TP, Church JA, Taylor KE, AchutaRao KM, Boyer TP, Ishii M, Caldwell PM (2012) Human-induced global ocean warming on multidecadal timescales. Nat Clim Change 2(7):524. doi:10.1038/nclimate1553

Greatbatch RJ (1994) A note on the representation of steric sea level in models that conserve volume rather than mass. J Geophys Res 99(C6):12767

Griffies SM (1998) The Gent-McWilliams skew flux. J Phys Oceanogr 28:831. doi:10.1175/1520-0485(1998)028<0831:TGMSF>2.0.CO;2

Griffies SM, Greatbatch RJ (2012) Physical processes that impact the evolution of global mean sea level in ocean climate models. Ocean Model 51(0):37. doi:10.1016/j.ocemod.2012.04.003 URL http://www.sciencedirect.com/science/article

Griffies S, Adcroft A, Aiki H, Balaji V, Bentson M, Bryan F, Danabasoglu G, Denvil S, Drange H, England M, Gregory J, Hallberg R, Legg S, Martin T, McDougall T, Pirani A, Schmidt G, Stevens D, Taylor K, Tsujino H (2009) Sampling physical ocean fields in wcrp cmip5 simulations: clivar working group on ocean model development (wgomd) committee on cmip5 ocean model output. Tech. rep., CLIVAR / WGOMD URL http://nldr.library.ucar.edu/repository/collections/OSGC-000-000-002-965

Hay CC, Morrow E, Kopp RE, Mitrovica JX (2012) Estimating the sources of global sea level rise with data assimilation techniques. Proc Nat Acad Sci. doi:10.1073/pnas.1117683109 URL http://www.pnas.org/content/early/2012/04/26/1117683109.abstract

Jiang JH, Su H, Zhai C, Perun VS, Del Genio A, Nazarenko LS, Donner LJ, Horowitz L, Seman C, Cole J, Gettelman A, Ringer MA, Rotstayn L, Jeffrey S, Wu T, Brient F, Dufresne JL, Kawai H, Koshiro T, Watanabe M, Lcuyer TS, Volodin EM, Iversen T, Drange H, Mesquita MDS, Read WG, Waters JW, Tian B, Teixeira J, Stephens GL (2012) Evaluation of cloud and water vapor simulations in CMIP5 climate models using NASA "A-Train" satellite observations. J Geophys Res 117(D14):D14105. doi:10.1029/2011JD017237

Jourdain N, Gupta A, Taschetto A, Ummenhofer C, Moise A, Ashok K (2013) The Indo-Australian monsoon and its relationship to ENSO and IOD in reanalysis data and the CMIP3/CMIP5 simulations. Clim Dyns :1–30. doi:10.1007/s00382-013-1676-1

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

Kopp R, Mitrovica J, Griffies S, Yin J, Hay C, Stouffer R (2010) The impact of Greenland melt on local sea levels: a partially coupled analysis of dynamic and static equilibrium effects in idealized water-hosing experiments. Clim Change 103(3-4):619. doi:10.1007/s10584-010-9935-1

Kuhlbrodt T, Smith R, Wang Z, Gregory J (2012) The influence of eddy parameterizations on the transport of the Antarctic circumpolar current in coupled climate models. Ocean Model 52–53(0):1. doi:10.1016/j.ocemod.2012.04.006 URL http://www.sciencedirect.com/science/article

Kwok R (2011) Observational assessment of Arctic Ocean sea ice motion, export, and thickness in CMIP3 climate simulations. J Geophys Res 116:C00D05 doi:10.1029/2011JC007004

Landerer FW, Jungclaus JH, Marotzke J (2007) Regional dynamic and steric sea level change in response to the IPCC-A1B scenario. J Phys Oceanogr 37(2):296. doi:10.1175/JPO3013.1

Lee T, Waliser DE, Li JLF, Landerer FW, Gierach MM (2013) Evaluation of CMIP3 and CMIP5 wind stress climatology using satellite measurements and atmospheric reanalysis products. J Clim. doi:10.1175/JCLI-D-12-00591.1

Li J-LF, Waliser D, Chen W, Guan B, Kubar T, Stephens G, Ma H, Deng M, Donner L, Seman C, Horowitz L (2012) An observationally based evaluation of cloud ice water in CMIP3 and CMIP5 GCMs and contemporary reanalyses using contemporary satellite data. J Geophys Res 117:D16105. doi:10.1029/2012JD017640

Maximenko N, Niiler P, Centurioni L, Rio MH, Melnichenko O, Chambers D, Zlotnicki V, Galperin B (2009) Mean dynamic topography of the ocean derived from satellite and drifting buoy data using three different techniques. J Atmos Oceanic Technol 26(9):1910. doi:10.1175/2009JTECHO672.1

McGregor S, Gupta AS, England MH (2012) Constraining wind stress products with sea surface height observations and implications for Pacific Ocean sea level trend attribution. J Clim 25(23):8164. doi:10.1175/JCLI-D-12-00105.1

Meehl G, Stocker T, Collins W, Friedlingstein P, Gaye A, Gregory J, Kitoh A, Knutti R, Murphy J, Noda A, Raper S, Watterson I, Weaver A, Zhao ZC (2007) In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt K, Tignor M, Miller H (eds) Climate change 2007-the physical science basis: contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, p 19–91

Meijers AJS, Shuckburgh E, Bruneau N, Sallee JB, Bracegirdle TJ, Wang Z (2012) Representation of the Antarctic circumpolar current in the CMIP5 climate models and future changes under warming scenarios. J Geophys Res 117(C12):C12008. doi:10.1029/2012JC008412

Meyssignac B, Salas y Melia D, Becker M, Llovel W, Cazenave A (2012) Tropical Pacific spatial trend patterns in observed sea level: internal variability and/or anthropogenic signature? Clim Past 8(2):787. doi:10.5194/cp-8-787-2012http://www.clim-past.net/8/787/2012/

Milne GA, Gehrels WR, Hughes CW, Tamisiea ME (2009) Identifying the causes of sea-level change. Nature Geosci 2:471. doi:10.1038/ngeo544

Nakicenovic N, Swart R (eds.) (2000) Emissions scenarios. Cambridge University Press, Cambridge

Pardaens A, Gregory J, Lowe J (2010) A model study of factors influencing projected changes in regional sea level over the twenty-first century. Clim Dyn :1–19. doi:10.1007/s00382-009-0738-x

Perrette M, Landerer F, Riva R, Frieler K, Meinshausen M (2013) A scaling approach to project regional sea level rise and its uncertainties. Earth Syst Dyn 4(1):11. doi:10.5194/esd-4-11-2013http://www.earth-syst-dynam.net/4/11/2013/

Räisänen J, Ylhäisi J (2012) Can model weighting improve probabilistic projections of climate change? Clim Dyn 39:1981–1998. doi:10.1007/s00382-011-1217-8

Reifen C, Toumi R (2009) Climate projections: past performance no guarantee of future skill? Geophys Res Lett 36. doi:10.1029/2009GL038082

Rietbroek R, Brunnabend SE, Kusche J, Schroeter J (2012) Resolving sea level contributions by identifying fingerprints in time-variable gravity and altimetry. J Geodyn 59–60(0):72. doi:10.1016/j.jog.2011.06.007 URL http://www.sciencedirect.com/science/article

Risien CM, Chelton DB (2008) A global climatology of surface wind and wind stress fields from eight years of QuikSCAT scatterometer data. J Phys Oceanogr 38(11):2379. doi:10.1175/2008JPO3881.1

Russell JL, Stouffer RJ, Dixon KW (2006) Intercomparison of the Southern Ocean circulations in IPCC coupled model control simulations. J Clim 19(18):4560. doi:10.1175/JCLI3869.1

Santer BD, Taylor KE, Gleckler PJ, Bonfils C, Barnett TP, Pierce DW, Wigley TML, Mears C, Wentz FJ, Brüggemann W, Gillett NP, Klein SA, Solomon S, Stott PA, Wehner MF (2009) Incorporating model quality information in climate change detection and attribution studies. Proc Nat Acad Sci 106(35):14778. doi:10.1073/pnas.0901736106. URL http://www.pnas.org/content/106/35/14778.abstract

Swart NC, Fyfe JC (2012) Observed and simulated changes in the southern hemisphere surface westerly wind-stress. Geophys Res Lett 39(16):n/a. doi:10.1029/2012GL052810

Tamisiea ME, Mitrovica JX, Milne GA, Davis JL (2001) Global geoid and sea level changes due to present-day ice mass fluctuations. J Geophys Res 106:30849. doi:10.1029/2000JB000011

Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res 106(D7):7183. doi:10.1029/2000JD900719

Weigel AP, Knutti R, Liniger MA, Appenzeller C (2010) Risks of model weighting in multimodel climate projections. J Clim 23(15):4175. doi:10.1175/2010JCLI3594.1

Yin J (2012) Century to multi-century sea level rise projections from CMIP5 models. Geophys Res Lett 39(17):n/a. doi:10.1029/2012GL052947

Yin J, Griffies SM, Stouffer RJ (2010a) Spatial variability of sea level rise in twenty-first century projections. J Clim 23(17):4585. doi:10.1175/2010JCLI3533.1

Yin J, Stouffer RJ, Spelman MJ, Griffies SM (2010b) Evaluating the uncertainty induced by the virtual salt flux assumption in climate simulations and future projections. J Clim 23(1):80. doi:10.1175/2009JCLI3084.1