Cold tongue and warm pool ENSO events in CMIP5: mean state and future projections

Andréa S. Taschetto
Climate Change Research Centre and ARC Centre of Excellence for Climate Systems Science, University of New South Wales, Sydney, Australia

Alexander Sen Gupta
Climate Change Research Centre and ARC Centre of Excellence for Climate Systems Science, University of New South Wales, Sydney, Australia

Nicolas C. Jourdain
Climate Change Research Centre and ARC Centre of Excellence for Climate Systems Science, University of New South Wales, Sydney, Australia

Agus Santoso
Climate Change Research Centre and ARC Centre of Excellence for Climate Systems Science, University of New South Wales, Sydney, Australia

Caroline C. Ummenhofer
Department of Physical Oceanography, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA

Matthew H. England
Climate Change Research Centre and ARC Centre of Excellence for Climate Systems Science, University of New South Wales, Sydney, Australia

1 Corresponding author address: Andréa S. Taschetto, Climate Change Research Centre and ARC Centre of Excellence for Climate Systems Science, University of New South Wales, Sydney, NSW, 2052, Australia
E-mail: a.taschetto@unsw.edu.au
The representation of the El Niño-Southern Oscillation (ENSO) under historical forcing and future projections is analyzed in 34 Coupled Model Intercomparison Project Phase 5 (CMIP5) models. Most models realistically simulate the observed location of maximum sea surface temperature (SST) anomalies during ENSO events. However, there exist biases in the westward extent of ENSO-related SST anomalies, driven by unrealistic westward displacement and enhancement of the wind stress field. The CMIP5 models capture the observed asymmetry in the magnitude between the warm and cold events (i.e. El Niños are stronger than La Niñas), and between the two types of El Niños, i.e., Cold Tongue (CT) are stronger than Warm Pool (WP) El Niños, although the strength of the latter is overestimated. However, CMIP5 models fail to reproduce the asymmetry between the two types of La Niñas, with CT events being stronger than WP, which is opposite to observations. Most models simulate the ENSO peak around December as observed, however the seasonal evolution of ENSO has a large spread across the models. In addition, the CMIP5 models can correctly represent the duration of CT El Niños, but show biases in the evolution of the other types of events. The evolution of WP El Niños suggests that the decay of this event occurs through heat content discharge in the models rather than advection of SST via anomalous zonal currents as seems to occur in observations. No robust changes are seen across the models in the location and magnitude of maximum SST anomalies, frequency or temporal evolution of these events in a warmer world.
1. Introduction

The environmental and societal impacts of the El Niño Southern Oscillation (ENSO) set against a gradual warming of the background climate has prompted concerted efforts to improve our understanding of ENSO behavior. Our capacity to predict the onset and duration of ENSO events has benefitted from sustained observing systems (e.g., McPhaden et al. 1998) coupled with developments in ENSO theories (e.g., Jin 1997), as well as ongoing improvements of climate models such as those facilitated by the Climate Model Intercomparison Project (CMIP). In the present study, we assess the fidelity of climate models submitted to CMIP phase 5 (CMIP5) in simulating the interannual SST variability in the tropical Pacific that is largely associated with ENSO, and examine how this variability is projected to change in the future.

Previous studies have shown that both the atmospheric and oceanic signatures of ENSO events are asymmetric in intensity, frequency, duration and spatial pattern, and also in their large-scale atmospheric responses. For example, Hoerling et al. (1997) noted that the nonlinear response in the Northern Hemisphere precipitation and atmospheric circulation to the warm and cold phases of the Southern Oscillation can be attributed to nonlinearities in deep convection to SST. Kang and Kug (2002) and Frauen and Dommengen (2010) showed that the different air-sea feedback interactions during the warm and cold ENSO phases contribute to the skewness in equatorial eastern Pacific SST anomalies (Burgers and Stephenson 1999). Other studies, on the other hand, attribute ENSO asymmetry to non-linear oceanic processes (e.g., Su et al., 2010). For instance, An and Jin (2004) attribute the warm-cold amplitude asymmetry to a strong nonlinear dynamic heating that enhances the warm events, as occurred in the 1982/83 and 1997/98 events, and weakens subsequent cold events.
The asymmetric characteristics of ENSO also manifest in the location of the associated maximum SST anomalies. Canonical El Niño events generally show largest SST anomalies in the eastern Niño-3 region. In contrast, La Niña anomalies tend to peak in the central Pacific, within the Niño-4 region (e.g., Schopf and Burgman 2006; Sun and Yu 2009). Dommenget et al. (2013) suggested that this spatial asymmetry may be partly related to the non-linear wind stress response to SST anomalies associated with opposite phases of ENSO (e.g., Kang and Kug 2002; Frauen and Dommenget 2010).

The transition between ENSO phases occurs via a negative feedback involving ocean wave dynamics (e.g., Battisti and Hirst 1989; Suarez and Schopf 1988; Jin 1997). In a zonally integrated sense, the action of internal waves leads to a build-up of ocean heat content in the equatorial Pacific as El Niño develops, and is subsequently drained off the Equator, leading to a delayed negative feedback and phase reversal to La Niña conditions (Jin 1997). This so-called recharge-oscillator paradigm has been confirmed by observations (Meinen and McPhaden 2000), but can only explain the linear component of ENSO transitions (McGregor et al. 2013). In reality the warm and cold event transition is not regular and ENSO events are also asymmetric in duration (e.g., Larkin and Harrison 2002; McPhaden and Zhang 2009; Okumura and Deser 2010; Ohba and Ueda 2009; Ohba et al. 2010). Warm SST anomalies associated with strong El Niño events tend to decay relatively quickly after their peak in December and are followed by cold SST anomalies in the equatorial Pacific. On the other hand, strong La Niña events generally persist through the following year.

In addition to the nonlinear duration between the Pacific warm and cold events, asymmetric behavior is also observed between strong and weak events of the same ENSO phase. Previous studies have identified inter-El Niño variations, when the
maximum SST anomalies concentrate in the central rather than the eastern Pacific. This central warming pattern appears as the second mode of tropical Pacific SST variability in an Empirical Orthogonal Function (EOF) or Rotated-EOF analysis (Lian and Chen 2012). Some studies have postulated that the first two modes of variability in tropical Pacific SST anomalies represent dynamically independent processes (e.g., Ashok et al. 2007). Others, however, argue that the central Pacific events can be considered as a non-linear manifestation of the canonical ENSO (e.g., Trenberth and Smith 2009; Takahashi et al. 2011; Dommenget et al. 2013; Johnson 2013). Whether or not a separate mode to canonical ENSOs, these Central Pacific events have drawn a lot of attention as they have occurred more frequently in the past few decades (e.g., Ashok et al., 2007; Lee and McPhaden, 2010; Na et al., 2011). This is particularly important as the rainfall teleconnections associated with warm central Pacific SSTs differ from those induced by traditional east Pacific events (e.g., Taschetto et al. 2010).

The mechanisms that give rise to enhanced central Pacific anomalies are still not fully understood. Ashok et al. (2007) proposed that the recent weakening of equatorial easterlies in the central Pacific and enhanced easterlies to the east have decreased the zonal SST gradient and flattened the thermocline, resulting in a climate state more favorable for the evolution of the central Pacific events. Choi et al. (2011, 2012) suggest that decadal changes in climate can play an important role in modulating the occurrence of El Niño with different warming signatures. It has been proposed that the asymmetries between the cold and warm phases of the Southern Oscillation may produce a non-zero residual effect on the time-mean state of the tropical Pacific that in turn modulate ENSO amplitudes (Yeh and Kirtman 2004; Rodgers et al. 2004). Sun and Yu (2009) suggest that the spatial asymmetries between
El Niño and La Niña lead to an ENSO cycle that shifts the tropical Pacific mean climate from a state favorable for strong ENSO activity to a state that sustains weak ENSO activity; a mechanism that has been reproduced by three of 19 CMIP3 models according to Yu and Kim (2011).

Different names have been ascribed to central Pacific ENSO events, despite referring to essentially the same SST structure. We adopt the “Cold Tongue” (CT) and “Warm Pool” (WP) terminologies to refer to ENSO events with maximum SST anomalies located in the eastern and central equatorial Pacific, respectively: CT El Niño, CT La Niña, WP El Niño and WP La Niña. Regardless of whether the CT and WP events are independent modes of variability, or a manifestation of ENSO nonlinearity, observations demonstrate that such spatial asymmetries are part of the interannual variability of the region and that distinct atmospheric teleconnections and associated climate impacts arise when SST peaks in the central or eastern Pacific (e.g., Ashok et al., 2007; Weng et al., 2007; Taschetto and England, 2009; Taschetto et al., 2009, 2010). As such, it is important that climate models can simulate the characteristics of different flavors of ENSO.

Despite significant advances in climate models, simulating realistic ENSO characteristics is still a major challenge (Guilyardi et al. 2009), largely associated with the difficulty in representing feedback processes (e.g., Collins et al. 2010; Kim and Jin 2011; Kim et al. 2013; Bellenger et al. 2013). Leloup et al. (2008) assessed 23 CMIP3 models and conclude that the majority of the models are not able to simulate the location of maximum amplitude of warm and cold events; only half can properly simulate ENSO onset, and none can represent the correct termination phases of either El Niño or La Niña. Nevertheless, climate models have shown improvement on
ENSO-related SST variability since CMIP3 (e.g., Yu and Kim 2010; Kim and Yu 2012; Bellenger et al. 2013).

The relatively short observational record available to date means that understanding the dynamics behind WP ENSO events, which have only become more frequent in recent decades, relies more on the use of climate models (e.g., Dewitte et al. 2012). Sparse observational records and the lack of a dynamical theory also imply uncertainty in determining whether the recent increase in the frequency of WP relative to CT El Niño can be attributed to greenhouse warming (Wittenberg 2009; McPhaden et al. 2011; Newman et al. 2011; Yeh et al. 2011; Kim et al. 2012). Assessing how well the CMIP models capture the observed temporal and spatial characteristics of ENSO may help to clarify the dynamics underlying the two types of ENSO as well as their future projections. Here we use a large pool of available CMIP5 models to provide a comprehensive assessment on their performance in representing the two types of ENSO regarding (1) their spatial characteristics, (2) their associated atmospheric, ocean surface and sub-surface properties, (3) their evolution and seasonality, and (4) their projections in a warmer climate scenario.

2. Models and Methods

a. Observations and Reanalysis data

The SST dataset used here is the Met Office Hadley Centre Sea Ice and Sea Surface Temperature version 1 (HadISST1; Rayner et al. 2003). Wind stress data is from the National Centers for Environmental Prediction (NCEP) - National Center for Atmospheric Research (NCAR) Reanalysis (Kalnay et al. 1996). In this study, we
consider the period from December 1949 to November 2008 for the SST and wind stress fields.

Sub-surface ocean temperature data is from the Simple Ocean Data Assimilation (SODA) Reanalysis (Carton and Giese 2008), covering the period from December 1958 to November 2008. The upper ocean heat content accumulated in the top 300m averaged between 3°S and 3°N is used as a proxy for the equatorial Pacific thermocline depth (Zebiak 1989).

b. CMIP5 Models

We analyze output from 34 climate models taking part in CMIP5 that is used to inform the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5). A summary of the climate models is shown in Table 1.

We examine two scenarios in this study: (1) historical simulations, which are integrations from around 1850 to at least 2005 using realistic natural and anthropogenic forcings, and (2) Representative Concentration Pathways 8.5 (RCP8.5) simulations, which are subject to increasing radiative forcing from the end of the historical simulation to 2100 when the radiative forcing reaches ~8.5Wm⁻². The last 50 years of the 21st century is analyzed of the RCP8.5 simulations. A description of the CMIP5 experiment design can be found in Taylor et al. (2012).

c. Methodology

For all variables, anomalies were calculated by removing the long-term monthly climatology over the entire period analyzed here. Time series of observations and simulations are then linearly detrended. When required, 3-month averages are
calculated for the examination of particular seasons, namely, December-to-February (DJF), March-to-May (MAM), June-to-August (JJA) and September-to-November (SON).

When a mean across all CMIP5 models is considered, the spatial fields are interpolated onto a common 1-degree x 1-degree grid for comparison with observations. Ensemble members for individual models are averaged prior to computing multi-model mean.

Where necessary, the estimate of the confidence levels or spread across CMIP5 models is calculated via the standard deviation among the models. The estimate of significance levels is computed via null-hypothesis using a Student t-test at the 0.05 significance level.

The selection of ENSO years is based on the DJF season, when observed events typically peak. Classifying ENSO events in models can be challenging, as models contain spatial SST biases. However, defining model-specific ENSO classifications to take into account model biases introduces subjective decisions. In addition, allowing multiple ENSO classifications would make model-observation intercomparison more difficult. As such, we adopt one common classification based on the state of equatorial Pacific SST anomalies. Even for observations, there is still no single method to classify two types of El Niño pattern (e.g., Ashok et al., 2007; Yeh et al., 2009).

Here we classify ENSO events by using the normalized DJF-averaged Niño indices as follows. An event is considered a CT ENSO if the Niño3 index is greater than |0.7| standard deviation and the averaged SST anomaly in the Niño3 region has a larger magnitude than the Niño4 SST anomaly. An event is classified as WP ENSO if the Niño4 index is above |0.7| standard deviation and the magnitude of the SST
anomaly in the Niño4 region is larger than in Niño3. The opposite is used for La Niña events. The selected years from the observations are displayed in Table 2. In the CMIP5 models, a threshold of $|1.0|$ standard deviation was used to distinguish ENSO events. In general, the simulated CT and WP events have similar magnitude (shown in section 3b), while in observations, WP El Niños and CT La Niñas are considerably weaker than CT El Niños and WP La Niñas, respectively. Therefore, a slightly lower standard deviation threshold ($|0.7|$) is needed to pick out enough observed events to obtain robust statistics. Similar classifications have been used in previous studies (e.g., Kug et al., 2009; Ham and Kug, 2012). Table 2 summarizes the ENSO classification for the purpose of this paper. ENSO events are selected for individual members of each CMIP5 model when more than one ensemble simulation is available.

Our analysis simplifies ENSO spatial pattern into CT and WP categories. This does not imply a need for either distinct dynamical modes to explain different ENSO flavors, or that ENSO patterns are strictly bimodal. This is just a convenient way to examine differences in simulated ENSO characteristics (compared to the observations) when the primary variability is shifted more to the east or west.

In order to provide quantitative measures of the spatial structure of different ENSO flavors, we calculate the magnitude and location of the maximum SST anomaly along the Equator, as well as the westward extent of the warm or cold anomalies. A model is considered to overestimate or underestimate the magnitude of El Niño or La Niña events if $|SSTA|_{\text{model}}^{\text{max}}$ exceeds $\left(|SSTA|_{\text{observed}}^{\text{max}} + \sigma_{\text{observed}}^{\text{SSTA}}\right)$, where $|SSTA|_{\text{model}}^{\text{max}}$ and $|SSTA|_{\text{observed}}^{\text{max}}$ are the maximum magnitude of the SST anomaly composites along the Equator (meridionally averaged between $5^\circ$S-$5^\circ$N) for the models and observations, respectively, and $\sigma_{\text{observed}}^{\text{SSTA}}$ is the standard deviation of the
maximum SST anomaly over all composite events in observations. Similarly, a model is considered to have a “realistic” representation of the ENSO position if the longitude of the maximum SST anomaly falls within one standard deviation of the mean longitude of the observed composite events. The metric for the westward extent of the SST anomalies is defined as the most westward longitude where SSTA drops to half the maximum SST anomaly. Note that although this metric can be subjective, it is not restrictive in the sense that it accounts for spatial biases in each model in terms of the magnitude of ENSO SST.

3. Results

a. Number of ENSO events

Figure 1 shows the number of each ENSO type based on the historical simulations for each model. In order to facilitate comparison between models and with observations, the number of events are shown per 100 years and the observed events are scaled to account for the different thresholds in ENSO classification. For most of the models, the number of CT events is comparable to the observations for the same length of record. For instance, there is a median of 11 CT El Niño events/100yrs in historical simulations versus 9 events/100yrs in observations. There are 10 WP El Niño/100yrs in simulations and 12 in observations. While there is a clear preference for WP to CT El Niño occurrence in observations, the CMIP5 models simulate a similar number of CT and WP El Niño events. In addition, there are 12 WP La Niña events/100yrs in simulations, the same as in observations. The number of CT La Niñas, however, is overestimated in the models: the median number of simulated CT La Niña
events/100yrs is 8, twice the observed number of events. Despite overestimating the number of CT La Niñas, the observed asymmetry in the number of cold events is represented in most of the models, with more WP than CT La Niña events.

b. Spatial Pattern of ENSO

The multi-model composite of simulated SST anomalies during austral summer (DJF) for both El Niño and La Niña events (Fig. 2, middle column) shows a number of features in common with the observations (Fig. 2, left column). More specifically, the magnitude and location of the maximum SST anomaly as well as the westward extent of the ENSO pattern are quantified in Figure 3 for each model, the multi-model mean and observations. The numbers of models that underestimate, overestimate or “realistically” represent the ENSO types are summarized in Table 3.

Most of the models simulate the magnitude of CT El Niño anomalies in the equatorial Pacific within the observational range. Only four out of 34 models overestimate, while eight underestimate, the magnitude of CT El Niño events. This does not necessarily imply biases in the location or extension of SST anomalies during CT El Niño. In fact, the majority of the models (26 out of 34 model), regardless of the magnitude of maximum SST anomalies, accurately simulate the position of the maximum SST anomaly during CT El Niño events, between 222°E and 248°E (Table 3). However, over one third of the models has an extension bias with the warm pattern extending too far to the west. On the other hand, only two models reveal a CT El Niño pattern with an eastward bias, namely FIO-ESM and HadGEM2-CC.

The bias in the westward extension and magnitude of SST anomalies is more severe for WP El Niño composites. In the observations the core of the warming is very narrow in the zonal direction (Fig. 2b), while the simulations reveal an elongated
pattern with stronger anomalies (Fig. 2j). None of the CMIP5 models can represent
the relatively confined warming in the central Pacific (Figs. 2f, 2b and 3c).
Additionally, 47% of the CMIP5 models overestimate the magnitude of WP El Niño
events and only 44% of the models are able to simulate the maximum SST anomalies
within the observed range (149°W to 167°W).

Despite biases in the magnitude and spatial extent of SST anomalies, all the
models (except CSIRO-Mk3-6-0) reproduce the observed asymmetry between CT and
WP warm events (Fig. 3a): they simulate relatively strong warm events in the east and
relatively weak warm events in the west (Fig. 3a). The exception is the CSIRO-Mk3-
6-0 model that fails to capture the location and magnitude of maximum SST
anomalies during WP El Niño events and simulates weaker SST conditions during CT
instead of WP El Niño. This results in a reversed asymmetry for the strong and weak
El Niños compared to observations.

While CMIP5 models simulate the observed asymmetry in magnitude between
the two types of warm events, this is not the case for the cold Pacific events. In
contrast to the observations, 22 out of 34 models simulate stronger CT events than
WP La Niña (Figs. 2g-h, 3a). In fact, 65% of the models overestimate the magnitude
of CT La Niña across the equatorial Pacific when compared to observations (Fig. 2k).
Despite this, the location and westward extension of CT La Niña events are within the
observational range for most CMIP5 models. Conversely, most of the models (79%)
exhibit WP La Niña events that extend farther west than in the observations. The bias
in intensity and spatial structure of cold events is reflected in the multi-model mean
with minimum SST anomaly of -1.4°C at 122°W for CT La Niña (Fig. 2g,k) and -
1.3°C at 153°W for WP La Niña (Fig. 2h,l) compared with -1.1°C at 126°W (Fig.
2c,k) and -1.4°C at 149°W (Fig. 2d,l) for observed CT La Niña and WP La Niña, respectively.

In general, the models better represent the magnitude of SST anomalies for the stronger CT El Niño and WP La Niña than for the weaker WP El Niño and CT La Niña whose maximum SST anomalies are too large compared to observations (Figs. 2i-l, green and brown lines). However, there are larger inter-model variations for the CT El Niño and WP La Niña composites (Figs. 2i and 2l) compared to the WP El Niño and CT La Niña composites (Figs. 2j and 2k).

The pattern of SST changes during ENSO events is intimately tied to the changes in surface wind stress. As a result we examine here the composite DJF wind stress anomalies for the different types of ENSO (Fig. 4). The maximum westerly (easterly) wind stress anomalies in the central South Pacific during CT El Niño (La Niña) events are reproduced by the models (Figs. 4a,e,i and 4c,g,k, respectively). Notable biases are, however, apparent. For instance, the CT ENSO in the models is associated with weaker than observed zonal wind stress anomalies (Figs. 4a,e, and 4c,g), which are consistent with the weaker subsurface temperature and heat content anomalies (Figs. 5a,c and 6a,c). Stronger bias is exhibited by the WP ENSO, where zonal wind stress anomalies extend and peak in the far western Pacific just to the east of Papua New Guinea with a larger magnitude compared to reanalysis (Figs. 4b,f and 4d,h). The maximum wind stress anomalies are located around 145°E (multi-model mean) in the WP-ENSO compared to 180°E in the reanalysis (Figs. 4j and 4l).

The overestimated zonal wind stress anomalies over the western Pacific would generate excessive upwelling (downwelling) during WP El Niño (WP La Niña). This is consistent with the composites of subsurface temperature anomalies along the Equator shown in Fig. 5b (Fig. 5d), which indicate colder (warmer) subsurface
temperature in the western Pacific at the depth of the thermocline compared to reanalysis (Figs. 6b and 6d). Concurrently, the overly strong westerly (easterly) wind anomalies tend to excessively suppress (enhance) the climatological equatorial upwelling during WP El Niño (WP La Niña), resulting in stronger warming (cooling) of the mixed layer over the west-central Pacific (Figs. 5b and 5d). The dynamical effect of the winds associated with WP El Niño is to shoal the thermocline, which in turn enhances stratification over the central Pacific leading to cold subsurface and warm mixed layer biases (Fig. 6b). The reverse holds for WP La Niña (Fig. 6d).

These wind stress biases allow the warm water to spread from the far west to the eastern equatorial Pacific during WP El Niño, as shown in the composite of subsurface temperature in Figure 5b. In addition, the opposite pattern is seen for WP La Niña events (Fig. 5d), when biases in zonal wind stress anomalies intensify the easterlies too far west in the equatorial Pacific, allowing an unrealistic extension of cold waters in the equatorial Pacific (Fig. 5d). The response in the ocean surface is such that the associated SST anomaly is too intense in the western Pacific during WP La Niña events (Fig. 2l).

c. Temporal Evolution of ENSO

The temporal evolution of SST, wind stress, and heat content anomalies associated with the two types of ENSO are shown in Figures 7 to 9 for the observations, reanalysis, and the CMIP5 multi-model mean. The evolution of SST anomalies during ENSO events for individual models is presented in Figure 10, where SST anomalies are averaged across the equatorial Pacific from 5°S-5°N, 150°E-90°W. In Figure 10, the evolution patterns are ordered according to the models that exhibit the highest to the lowest correlations with the observed SST evolution.
In general the CMIP5 models correctly reproduce the timing of seasonal peaks in SST anomalies, for all ENSO types (Figs. 9 and 10). However, there is a range of behavior in terms of the duration of ENSO events. The transition from warm to cold or neutral SST conditions and vice-versa are simulated in most models, but the timing of these transitions varies considerably across the models. Here we describe the features associated with ENSO evolution separately for each type of event.

COLD TONGUE EL NIÑO

Overall, the multi-model mean evolution of CT El Niño events is best represented among the other types of ENSO. The SST anomalies in the equatorial Pacific become warmer than average in February-March, peak in December, and vanish in October of the following year (Fig. 9a). As previously documented in the literature (e.g. Okumura and Deser 2010), CT El Niño events are generally followed by a La Niña event that starts in the following year but peaks in December, 2 years after the maximum warming (Fig 7a, 9a). CT El Niños are also generally preceded by cold anomalies one year earlier. The evolution of the observed CT El Niño events shown here is consistent with the results by Hu et al. (2012), based on ERSSTv3 data (see their Fig. 3). The models exhibit this observed evolution with a wide range of fidelity.

While some of the models tend to simulate CT El Niño events that are too long (lasting longer than two years; e.g. GFDL-ESM2M, MIROC5 and MPI-ESM-LR), some exhibit a rapid transition to a strong La Niña with a seasonal cycle that is more extreme than observations (CCSM4, CESM1-CAM5, CESM1-FASTCHEM, CESM1-WACCM, FIO-ESM, GFDL-CM3, GFDL-ESM2M and MIROC5; Fig. 10a). Conversely, some models fail to simulate the transition from warm to cold events altogether (ACCESS1-0, IPSL-CM5A-LR, MIROC-ESM, MPI-ESM-MR, HadGEM2-AO and inmcm4).
The inter-model behavior renders a multi-model mean that resembles the observed evolution (Figs. 7e and 7a), except that the simulated CT El Niños are preceded and followed by much weaker than observed La Niñas. The fact that the weak La Niña following a CT El Niño occurs one year earlier than the observed is apparently associated with the more rapid thermocline adjustment in the CMIP5 models (Figs. 8a and 8e), as indicated by the more rapid transition of the basin-wide wind anomalies from westerly to easterly at the peak of El Niño (Fig. 7i), as well as the narrower meridional extent of the zonal winds compared to reanalysis (Figs. 4a and 4e). A narrower meridional extent of ENSO-related zonal wind anomalies would tend to generate faster off-equatorial Rossby waves, thus a more rapid phase transition (Kirtman 1997), which is a bias also seen in CMIP3 models (Capotondi et al. 2006).

WARM POOL EL NIÑO

The multi-model mean for the WP El Niño evolution captures the correct initiation and peak of SST anomalies in the central-western equatorial Pacific (Figs. 7f), with SST anomalies becoming warmer than average around October and maturing in December of the following year (Fig. 9b). However, in most of the models, the SST anomalies associated with the simulated WP El Niño last longer (4 months longer in the multi-model mean, Fig. 9b) than observed events (Figs. 7b, f, j). For example, CanESM2 and HadCM3 simulate overly strong and prolonged WP El Niño events, followed by slightly cold anomalies in the following year (Fig. 10b). CCSM4, CESM1-CAM5, CESM1-FASTCHEM, GFDL-ESM2M, IPSL-CM5B-LR and MIROC5 simulate prolonged warm SST anomalies in the Pacific, followed by strong cold anomalies two years after the peak of WP El Niño events (Fig. 10b). A few models (MIROC-ESM and MPI-ESM-LR) represent WP El Niño with a much longer
duration compared to observations and do not show a transition to SST anomalies of opposite sign either before or after the peak of an event (Fig. 10b).

Overall the CMIP5 models simulate WP El Niño events of similar duration to the simulated CT El Niños. In observations, however, CT El Niños last longer than WP El Niños (Fig. 7a-b; Hu et al. 2012). This can be explained by biases in the wind stress anomaly field in the western Pacific. Stronger than observed anomalous zonal wind stress is seen in the western Pacific approximately two seasons before the peak of the WP El Niño and lasts a couple of months after its mature phase (Figs. 7b and 7f). This results in a shallower than observed thermocline in the west during the austral summer season (Fig. 8b). It is currently thought that the spatial structure of WP El Niño does not favor a discharge process of the equatorial heat content that is efficient enough to trigger a cold event (Kug et al. 2009) unlike the case of CT El Niño events. Instead, the decay of WP El Niño is thought to be driven by zonal advection of mean SST by anomalous zonal currents. However, the cold condition following the warm events and the pronounced eastward propagating negative thermocline anomalies in the multi-model mean (Fig. 8f) suggest that some degree of thermocline processes influences the evolution of WP El Niño in CMIP5 models. This is further supported by the existence of southward shift of westerly anomalies at the peak of the simulated WP El Niños (marked by northerly anomalies across the Equator in Figure 4f). This feature, which is seen in CT but not WP El Niños in observations (Fig. 4a, b), is associated with heat content discharge (McGregor et al. 2012).

COLD TONGUE LA NIÑA

The temporal evolution of CT La Niña events is generally poorly simulated in CMIP5 models: despite correctly simulating the peak of the event, it shows biases in the
temporal start and end of the cold event when comparing to observations (Figs. 9c and 10c). The multi-model mean shows cooler than average SST in the equatorial Pacific starting in April, 7 months later than observations (Figs. 7g,c and 9c). The event peaks correctly around January, however the SST anomalies reach neutral conditions in April of the following year, 5 months later compared to observations (Fig. 9c).

Observed easterly wind stress anomalies in the central Pacific precede SST anomalies by approximately one season and extend far west during austral summer (Fig. 7c), favoring thermocline deepening in the west (Fig. 8c). In the multi-model mean, however, the strengthened easterlies in the central-western Pacific last throughout the calendar year after the mature phase of the CT La Niña (Fig. 7k), resulting in a longer duration of the cold event especially around the dateline (Fig. 7g).

In observations, CT La Niña events are preceded by warm SST conditions in the equatorial Pacific two years before their peak (Figs. 7c, 9c and 10c). CESM1-CAM5 is the only model that correctly simulates the timing and magnitude of this pre-CT La Niña warm event (Fig. 10c). In contrast, the CT La Niña events simulated by most of the CMIP5 models are preceded by a warming in the previous year (Figs. 7g, 8g and 9c), particularly strong in CCSM4, CESM1-FASTCHEM, CESM1-WACCM, FIO-ESM, GFDL-CM3, GFDL-ESM2M, HadCM3, NorESM1-M and NorESM1-ME (Fig. 10c).

WARM POOL LA NIÑA
CMIP5 models also have considerable difficulty in reproducing the evolution of WP La Niña (Fig. 9d). Most of the models realistically simulate the timing of the initiation of WP La Niña events, with negative SST anomalies in the equatorial Pacific starting around March (Fig. 9d). However, most of the CMIP5 models do not simulate the cold SST anomalies that last throughout the following year as in the observations
(Figs. 7d, 8d and 9d), except for CESM-CAM5 (Fig. 10d). Instead, the simulated multi-model mean WP La Niña events terminate six months earlier (Figs. 7h, 8h and 9d). The peak of observed and simulated WP La Niña events occurs in December and is preceded only by weak warm anomalies (Figs. 7d and 7h). Exceptions are GFDL-ESM2M and MIROC5 that simulate overly strong positive SST anomalies two years before the peak of the cold event in the central Pacific (Fig. 10d).

The initiation timing of the zonal wind stress anomaly in the central equatorial Pacific is well captured in the multi-model mean (Fig. 7l) although the maximum amplitude occurs two months earlier than in observations. Similar to WP El Niño events, unrealistically strong wind stress anomalies appear in the western Pacific (Fig. 7h). This biased wind stress in the west leads to an SST pattern that extends westward along the Equator, as previously discussed. Additionally, the simulated wind stress persists throughout the year, resulting in a rapid thermocline adjustment (Fig. 8h) and an early termination of WP La Niña events in the CMIP5 models.

d. Seasonality of ENSO

Figure 11 displays the standard deviation of Niño indices for each model and observations. In general, most of the models show good fidelity in the timing and amplitude of SST variability in the central equatorial Pacific. 27 out of 34 CMIP5 models realistically represent the maximum amplitude of ENSO during November-to-January in the Niño3.4 region (Fig. 11b). Greater disagreement is evident in the seasonality of ENSO in the eastern and western part of the tropical Pacific Ocean. For instance, only 13 out of 34 models capture the correct timing of maximum variability in the Niño3 region. The discrepancies in the timing of the events among the models
are reflected in the notably weaker multi-model mean ENSO seasonality compared to observations.

GFDL-ESM2M, GFDL-CM3 and ACCESS1-3 exhibit maximum variability that is about two months late for the Niño3.4 region compared to observations and the bcc-csm1-1 is three months too early. This phase bias is even more extreme in some models, in particular IPSL-CM5A-MR, IPSL-CM5A-LR and CSIRO-Mk3-6-0, where the maximum variability occurs approximately 6 months after the observed peak of ENSO. In addition to representing ENSO events in the wrong season, the IPSL-CM5A-MR model has very weak seasonality, which is also true for FIO-ESM, MPI-ESM-LR, MPI-ESM-MR, IPSL-CM5B-LR and CMCC-CM, although the ENSO indices in these models tend to peak in the correct season.

As shown in this analysis, the substantial spread in the seasonal peak of warm and cold events compared to observations suggests that ENSO timing is one of the aspects requiring improvement in future CMIP simulations.

4. Future Projections

Here we analyze how the different types of ENSO events may change in the future as projected by 27 CMIP5 simulations that had archived RCP8.5 simulations at the time of writing. Figure 12 shows the multi-model mean difference in the equatorial Pacific SST anomalies between the RCP8.5 and historical simulations for each event. For the CT El Niños, the multi-model mean shows significant cooling in the eastern South Pacific and western Pacific but a warming in the eastern North Pacific (Fig. 12a). The WP El Niños reveal a slight warming in the central-west equatorial Pacific and cooling on both sides of the Equator, suggesting a more confined warming in the
future scenario than the historical simulation (Fig. 12b). The CT La Niña changes exhibit cooling in the west and warming in the east equatorial Pacific (Fig. 12c). The WP La Niña pattern suggests strengthening of the cold events in a warmer scenario (Fig. 12d). However, these future changes in the amplitude of ENSO events are overall small and not consistent across the models. Analysis of the spatial metrics shown in Figure 3 reveals no clear change in the multi-model mean magnitude or location of maximum SST anomaly. The westward extent of ENSO also does not show significant changes in the future projections, except for CT La Niña events that extend 15 degrees west on average, which is consistent with the cooling around the dateline shown in Figure 12c.

The change in the amplitude of SST anomalies is also quantified in Figure 13 via the difference between the standard deviation of Niño indices from the historical simulation to the RCP8.5 scenario. There is little agreement in the projections of Niño indices across the models, indicating that the changes derived for the multi-model mean (Fig. 12) are not statistically significant. Our analysis based on 27 CMIP5 models does not reveal any enhancement of WP-to-CT ENSO intensity from historical to RCP8.5 scenario (Fig. 13d). This contradicts the findings of Kim and Yu (2012) who reported increased WP-to-CT intensity ratio from historical to RCP4.5 scenario using a smaller set of CMIP5 models. Indeed, when individual models are considered, the WP-to-CT ENSO asymmetry in regard to intensity shows significant changes: for instance, there is a robust increase in the Niño4/Niño3 ratio for all 10 members of the CSIRO-Mk3-6-0 model, 5 members of the CanESM2 model and 3 members of the MIROC5 model. Stevenson (2012) found that the SST difference between the Niño4 and Niño3.4 regions from 20th century to RCP4.5 is only statistically significant in four of the eleven CMIP5 models containing more than
three members. This shows the importance of considering large ensembles for future climate projections.

An evaluation of the frequency of ENSO events from historical to RCP8.5 scenarios shows no significant result (Fig. 14). The evolution of ENSO events also exhibits little change in the future. On average, the timing of the initiation, peak and termination of the Pacific events show similar behavior in the RCP8.5 compared to the historical scenario (Figure 9b, red and blue curves).

5. Discussion and Conclusions

Now as in the past there remain substantial problems in the realistic simulation of ENSO in climate models, despite good progress over the past decade. Of particular importance for ENSO teleconnections is the correct simulation of the characteristics of different ENSO flavors, classified here into cold tongue and warm pool El Niños and La Niñas. This study assesses ENSO in 34 CMIP5 models and finds that while most models do simulate events that can be classed as either CT or WP, there is varying fidelity across the models. The CMIP5 models can simulate the location of maximum SST anomalies during ENSO events within the observational bounds, consistent with the findings of Kim and Yu (2012). However, Kim and Yu (2012) find a relatively good representation of WP-ENSO events, with relatively more biases for CT-ENSOs. In contrast, our assessment of 34 CMIP5 models indicates that the intensity of CT El Niño is in general the most realistically represented across the four ENSO types, while the magnitude of WP events is rather poorly simulated (Table 3) based on our metrics. The discrepancy between our study and that of Kim and Yu (2012) is due to the small set of CMIP5 models used in their study (discussed below).
The observed asymmetries in the intensity between warm and cold events (i.e. El Niños stronger than La Niñas) and between warm events (i.e. CT stronger than WP El Niños) are captured in most of the CMIP5 models analyzed here. However, most of the models fail to reproduce the observed asymmetry between the cold events, i.e. simulated CT La Niñas are stronger than WP La Niñas. Notable biases also exist in the extension of the WP events: none of the models can simulate the confinement of observed SST warming within the central Pacific during WP El Niños. The simulated WP El Niños and La Niñas generally extend too far into the western and eastern Pacific.

Most CMIP5 models can simulate an evolution of CT El Niño events that is similar to that observed, with correct time of initiation, duration and peak in December. The simulated CT El Niños are often followed by cold events one year after the peak of the warm event in most of the models, while cold events more commonly occur two years after in observations. The duration of WP El Niños is overestimated for most of the models, a bias related to the simulated wind stress anomalies in the central-to-western equatorial Pacific being too strong and persistent. The evolution of cold events also exhibits biases. In particular, the simulated CT La Niña starts about 2 seasons later than the observed event and ends 5 months too late, even though it correctly peaks around January. Most of the models simulate WP La Niña with the correct initiation and peak, but ending earlier than the observed event, lasting for one year, which is shorter than the observed 2 years.

The seasonality of ENSO shows varying degrees of fidelity depending on the Niño region. Better agreement in the timing of ENSO peak among CMIP5 models is seen in the Niño3.4 region (27 of 34 models peak in the correct season) while a large spread occurs in the Niño3 region (only ~ 1/3 of the models peak in the correct
season). Even in models where the December peak of the Pacific SST variability is correctly simulated, good skill in simulating other aspects of ENSO seasonality is not guaranteed. Several models show too weak a seasonality, suggesting that many ENSO events are also occurring at the wrong time of year. Particularly in the Niño3 and Niño4 regions, ENSO events in many CMIP5 models peak in the wrong seasons. The substantial spread in the seasonal peak of ENSO events in CMIP5 compared to observations suggests that ENSO seasonality is still an aspect that needs to be improved in models.

Most of the biases in the ENSO SST anomalies can be linked to biases in the wind stress anomalies, which are likely in turn related to mean state biases in the SST. For all ENSO flavors the wind stress extends too far westward, particularly during WP events. These biases in the wind stress anomalies generate spurious thermocline anomalies that propagate eastward as upwelling and downwelling Kelvin waves which can in turn influence the evolution of ENSO events. It is noted here that spurious thermocline anomalies over the central Pacific may influence the evolution of warm pool events in the models. Furthermore, the narrower meridional extent of the CT ENSO wind stress anomalies seen in the multi model mean likely contributes to the more rapid and somewhat more regular ENSO phase transition. This systematic bias was also seen in CMIP3 models (Capotondi et al. 2006), which is likely to be associated with the classical ‘cold tongue’ problem, requiring improvements in the representation of the physics of the coupled climate system (e.g., Luo et al. 2005; Guilyardi et al. 2009b). In addition, the representation of the two types of ENSO in climate models seems to be sensitive to the atmospheric response, in particular of convection, to the SST anomaly patterns (Ham and Kug, 2012). As demonstrated by Ham et al. (2012), and indicated by the analysis of Bellenger et al. (2013), convection
parameterization in climate models can strongly affect the seasonal phase locking of ENSO. Recent studies have also attributed biases in ENSO simulations to remote influences, particularly climate over the Indian Ocean basin (see Santoso et al. 2012 and references therein). As suggested by Okumura and Deser (2010), remote forcing from the Indian Ocean can influence the asymmetry in the duration of El Niño and La Niña. Thus, some of the model ENSO biases reported here could relate to how CMIP5 models simulate Indian Ocean climate and its variability.

The different types of ENSO do not show robust changes even when subject to large changes in radiative forcing. This result is consistent with previous studies based on CMIP3 that concluded there is little agreement among the models for projected ENSO changes (e.g., van Oldenborgh et al. 2005; Guilyardi 2006; Collins et al. 2010; Stevenson 2012), and is also consistent with the idea that changes in ENSO amplitude and frequency can be hard to detect given the level of natural variability present in the climate system (e.g., Wittenberg 2009; Aiken et al. 2013).

In contrast to previous studies (e.g., Yeh et al., 2012; Kim and Yu, 2012), the ratio of WP-to-CT variability from the historical run to the RCP8.5 scenario does not show consistent change. The frequency of ENSO events also exhibits no robust variations. The difference between these results and the previous literature is related to the much larger ensemble used in our study. For example, if we sub-select the 12 CMIP5 models used by Yeh et al. (2012), we also find that the WP-to-CT ratio slightly decreases, agreeing with the eastern Pacific trend reported in their study. However this result is not robust when considering the considerable larger ensemble of models used here.

In summary, our study suggests that CMIP5 models can simulate the two types of ENSO with varying degrees of fidelity. The features that models represent
well include: (1) stronger El Niños than La Niñas; (2) stronger CT El Niños than WP
El Niños; (3) the location of maximum SST anomaly for all events; (4) the magnitude
of CT El Niños and WP La Niñas; (5) the ENSO peak around December; (6) the
timing evolution of CT El Niño events. On the other hand, the majority of models (1)
cannot simulate the asymmetry between cold events (stronger CT than WP La Niñas);
(2) overestimate the magnitude of WP El Niño and CT La Niña events; (3)
overestimate wind stress in the western Pacific; (4) simulate SST anomalies extending
too far west in the equatorial Pacific; (5) agree poorly on the ENSO seasonal
evolution; (6) overestimate the termination duration of WP El Niño and CT La Niña
and underestimate for WP La Niña. Finally, there are no robust changes in the future
projections of the magnitude or location of maximum SST anomalies, nor the
frequency of ENSO events. This study motivates further analyses to understand the
disagreement among models and projections, via assessments of climate simulations
remote from the tropical Pacific, as well as understanding the feedback mechanisms
operating in the Pacific region in CMIP5 models.

Acknowledgments

We acknowledge the WCRP’s Working Group on Coupled Modelling, which is
responsible for CMIP, and we thank the climate modeling groups for producing and
making available their model output. The U.S. DoE’s PCMDI provided coordinating
support and led development of software infrastructure in partnership with the Global
Organization for Earth System Science Portals. This work was also supported by the
NCI National Facility at the ANU via the provision of computing resources to
the ARC Centre of Excellence for Climate System Science. We acknowledge all the
Institutions responsible for the observations and reanalysis products for having made
their data available. This project was supported by the Australian Research Council.
References


List of Figures

FIG. 1. Number of ENSO events/100yrs in the historical simulations for each model, multi-model mean and observations. Number of events is averaged for models containing more than one member. Top panel: Warm events, i.e. CT El Niño (CTEN) represented by black circles and WP El Niño (WPEN) by white squares. Bottom panel: Cold events, i.e. CT La Niña (CTLN) as black circles and WP La Niña (WPLN) as white squares. Bars in the multi-model mean indicate the interquartile range.

FIG. 2. Composite of SST anomalies during Dec-Feb season for ENSO events. Units in Celsius. Left column: Observations from HadISST based on the period Dec/1949 to Nov/2008. Middle column: Multi-model mean based on 34 CMIP5 models. Areas within the thin grey line are statistically significant across the observed events and the composited events of CMIP5 models at the 0.05 significance level based on a Student t-test. Right column: Dec-Feb SST anomalies averaged over 5°S-5°N. Brown line represents the multi-model mean, while the green line represents observations. Light brown shade indicates the standard deviation of simulated composites, an estimate of the spread among CMIP5 models.

FIG. 3. (a) Magnitude and (b) location of the maximum SST anomaly for each ENSO type. (c) Westward extension of SST anomaly estimated as the location of half of simulated maximum SST anomaly. NB: The y-axis in (b) and (c) is asymmetric to account for nonlinearity in warm/cold events. Error bars (also represented as dashed lines) in HadISST show the 90% confidence interval for the mean of observed events. Error bars in the multi-model mean (MMM) represents the 90% confidence intervals for the mean of composited events across the CMIP5 models. The mean and
associated error of each type of ENSO event are specified with the same color as in the legend for the MMM and observations. Models that are within the dashed lines are considered to have a “realistic” simulation of the metric.

Fig. 4. Composite of wind stress anomalies (zonal component shaded) during Dec-Feb season for ENSO events. Units in Pa. Left column: NCEP/NCAR Reanalysis. Middle column: Multi-model mean based on 20 CMIP5 models. Areas within the thin grey line are statistically significant across the observed events and the compositing events of CMIP5 models at the 0.05 significance level based on a Student t-test. Right column: Dec-Feb zonal wind stress anomalies averaged over 3°S-3°N. Red line represents the multi-model mean, while blue line represents reanalysis. A 15-degree window running mean was applied to the curves. Light red shade indicates the standard deviation of simulated composites, an estimate of the spread among CMIP5 models.

Fig. 5. Multi-model mean ocean temperature anomalies (shaded) averaged across 3S-3N along the Pacific during Dec-Feb season for ENSO events. The difference between the multi-model mean and observations are contoured in 0.3 Celsius intervals. Red (blue) contours are positive (negative) differences. (a) Cold Tongue El Niño, (b) Warm Pool El Niño, (c) Cold Tongue La Niña, and (d) Warm Pool La Niña. Units in Celsius. Multi-model mean based on 23 CMIP5 models.

Fig. 6. Composite of ocean heat content anomalies averaged across 3S-3N along the Pacific during the Dec-Feb season for ENSO events. The light grey curve is the multi-model mean heat content and the dark grey curve represents the SODA reanalysis. Light grey shade indicates the standard deviation of simulated composites, an estimate of the spread among CMIP5 models. The curves were smoothed with an 11-longitude
point running mean. (a) Cold Tongue El Niño, (b) Warm Pool El Niño, (c) Cold Tongue La Niña, and (d) Warm Pool La Niña. Multi-model mean based on 23 CMIP5 models.

Fig. 7. Hovmöller diagram of the SST (shaded) and wind stress (vectors) anomalies averaged between 5S and 5N across the Pacific Ocean during ENSO events. Left column: Observations from HadISST dataset and reanalysis from NCEP/NCAR. Middle column: Multi-model mean of 20 CMIP5 models. Right column: Evolution of zonal wind stress anomalies (Pa) averaged between 5S-5N, 120E-110W. Red line: Multi-model mean. Blue line: NCEP/NCAR reanalysis. Red and blue lines were smoothed with an 11-month running mean. Light red area represents the standard deviation of the multi-model mean as an estimate of the spread across the models. (a,e,i) Cold Tongue El Niño, (b,f,j) Warm Pool El Niño, (c,g,k) Cold Tongue La Niña, and (d,h,l) Warm Pool La Niña.

Fig. 8. Hovmöller diagram of the ocean heat content (shaded) and zonal wind stress (contours) anomalies averaged between 3°S and 3°N across the equatorial Pacific Ocean during ENSO events. Brown (green) contours are westerly (easterly) anomalies, plotted on 0.003Pa intervals. Multi-model mean of 16 CMIP5 models. An 11-month window running mean was applied to the data. (a,e) Cold Tongue El Niño, (b,f) Warm Pool El Niño, (c,g) Cold Tongue La Niña, and (d,h) Warm Pool La Niña. Units in Celsius and Pa, respectively.

Fig. 9. Evolution of averaged SST anomalies averaged across the equatorial Pacific (5°S-5°N, 150°E-90°W). Black curve: Observations. Blue curve: Multi-model mean of historical simulations. Red curve: Multi-model mean of RCP8.5 scenario. Shading indicates the standard deviation of the multi-model mean for the historical (blue) and
RCP8.5 (red) simulations. Composite for (a) Cold Tongue El Niño events, (b) Warm Pool El Niño, (c) Cold Tongue La Niña, and (d) Warm Pool La Niña.

Fig. 10. Evolution of SST anomalies averaged across the equatorial Pacific (5S-5N, 150E-90W) for individual models. Composite for (a) Cold Tongue El Niño events, (b) Warm Pool El Niño, (c) Cold Tongue La Niña, and (d) Warm Pool La Niña. Models are ordered according to correlations with observations. Units in Celsius.

Fig. 11. Monthly standard deviation of (a) Niño4 (b) Niño3.4 and (b) Niño3 indices for CMIP5 models. For comparison purposes, the monthly standard deviation is divided by the maximum value, which number is indicated in white in the month when it peaks. Models are ordered according to correlations with observations.

Fig. 12. Left column (a-d): Difference in the simulated ENSO SST anomaly composites during Dec-Feb season between the RCP8.5 and historical scenarios. Units in Celsius. Areas within the thin grey line are statistically significant at the 0.05 significance level based on a Student t-test. Right column (e-h): SST anomaly averaged over the equatorial Pacific (5°S-5°N). Red line represents the multi-model mean for RCP8.5, while the blue line represents the historical simulation. Light red (blue) shade indicates the standard deviation of simulated composites for the RCP8.5 (historical), an estimate of the spread among CMIP5 models of each scenario. Based on 27 CMIP5 models.

Fig. 13. Difference in the standard deviation of (a) Niño3, (b) Niño3.4, and (c) Niño4 indices between RCP8.5 and historical simulations for 27 CMIP5 models. (d) Difference in the ratio of the standard deviation between Niño4 and Niño3. Grey dashed line represents the difference in the multi-model mean. Zero appears as the
black dashed line. Vertical bars represent the range of ensemble members when available and circles the respective ensemble mean.

Fig. 14. Difference in the number of events/100yr between RCP8.5 and historical simulations for 27 CMIP5 models. (a) CT El Niño (b) WP El Niño (c) CT La Niña (d) WP La Niña. Grey dashed line represents the difference in the multi-model mean. Zero appears as the black dashed line. Vertical bars represent the range of ensemble members when available and circles the respective ensemble mean.
Table 1. List of the CMIP5 models, with respective institutes, variables and number of ensemble members used in this study.

<table>
<thead>
<tr>
<th>Model</th>
<th>Institute, Country</th>
<th>SST</th>
<th>Wind Stress</th>
<th>Sub-surface Temp</th>
<th>Thermocline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>HIST</td>
<td>RCP8.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACCESS-1.0</td>
<td>CSIRO-BOM, Australia</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ACCESS-1.3</td>
<td>BOM, Australia</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>BCC-CSM1-1</td>
<td>BCC, CMA, China</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>CanESM2</td>
<td>CCCMA, Canada</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>CESM1-CAM5</td>
<td>NSF-DOE-NCAR, USA</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CESM1-FASTCHEM</td>
<td>NSF-DOE-NCAR, USA</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CESM1-WACC3</td>
<td>NSF-DOE-NCAR, USA</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCSM4</td>
<td>NCAR, USA</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>CMCC-CM</td>
<td>CMCC, Italia</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNRM-CM5</td>
<td>CNRM-CERFACS, France</td>
<td>10</td>
<td>5</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>CSIRO-Mk3-6-0</td>
<td>CSIRO-QCCCE, Australia</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>FGOALS-g2</td>
<td>LASG-CESS, China</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGOALS-s2</td>
<td>LASG-IAP, China</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIO-ESM</td>
<td>FIO, SOA, China</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GFDL-CM3</td>
<td>NOAA-GFDL, USA</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GFDL-ESM2G</td>
<td>NOAA-GFDL, USA</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GFDL-ESM2M</td>
<td>NOAA-GFDL, USA</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GISS-E2-H</td>
<td>NASA/GISS, USA</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GISS-E2-R</td>
<td>NASA/GISS, USA</td>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HadCM3</td>
<td>MOHC, UK</td>
<td>9</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>HadGEM2-CC</td>
<td>NIMR-KMA,Korea</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HadGEM2-ES</td>
<td>MOHC, UK</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>INMCM4</td>
<td>INM, Russia</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>IPSL-CM5A-LR</td>
<td>IPSL, France</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>IPSL-CM5B-LR</td>
<td>IPSL, France</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>IPSL-CM5A-MR</td>
<td>IPSL, France</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MIROC5</td>
<td>AORI-NIES-JAMSTEC, Japan</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>MIROC-ESM</td>
<td>AORI-NIES-JAMSTEC, Japan</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>MPI-ESM-LR</td>
<td>MPI-N, Germany</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>MPI-ESM-MR</td>
<td>MPI-N, Germany</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MRI-CGCM3</td>
<td>MRI, Japan</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>NorESM1-M</td>
<td>NCC, Norway</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>NorESM1-ME</td>
<td>NCC, Norway</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Summary of criteria employed for the ENSO classification for observations and CMIP5 models, using the standardized Niño3 and Niño4 indices. For further details see text.

<table>
<thead>
<tr>
<th>Event</th>
<th>Observations</th>
<th>CMIP5 models</th>
<th>Years Selected in Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cold Tongue La Niña</strong></td>
<td>Niño3 &lt; -0.7 and Niño3 &lt; Niño4</td>
<td>Niño3 &lt; -1.0 and Niño3 &lt; Niño4</td>
<td>1950, 1968, 1985, 2006</td>
</tr>
</tbody>
</table>
Table 3. Number of models out of 34 that underestimate (↓), overestimate (↑), and reproduce “realistic” (○) types of ENSO. Largest numbers are in bold. See methodology for explanation.

<table>
<thead>
<tr>
<th></th>
<th>CT El Niño</th>
<th>WP El Niño</th>
<th>CT La Niña</th>
<th>WP La Niña</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Magnitude of maximum SST anomaly</strong></td>
<td>8</td>
<td>22</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>16</td>
<td>9</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>14</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td><strong>Location of maximum SST anomaly</strong></td>
<td>3</td>
<td>26</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>13</td>
<td>0</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>15</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td><strong>Extension of the western 0.5*max magnitude</strong></td>
<td>12</td>
<td>20</td>
<td>2</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
Fig 1. Number of ENSO events/100yrs in the historical simulations for each model, multi-model mean and observations. Number of events is averaged for models containing more than one member. Top panel: Warm events, i.e. CT El Niño (CTEN) represented by black circles and WP El Niño (WPEN) by white squares. Bottom panel: Cold events, i.e. CT La Niña (CTLN) as black circles and WP La Niña (WPLN) as white squares. Bars in the multi-model mean indicate the interquartile range.
Fig 2. Composite of SST anomalies during Dec-Feb season for ENSO events. Units in Celsius. Left column: Observations from HadISST based on the period Dec/1949 to Nov/2008. Middle column: Multi-model mean based on 34 CMIP5 models. Areas within the thin grey line are statistically significant across the observed events and the composited events of CMIP5 models at the 0.05 significance level based on a Student t-test. Right column: Dec-Feb SST anomalies averaged over 5°S-5°N. Brown line represents the multi-model mean, while the green line represents observations. Light brown shade indicates the standard deviation of simulated composites, an estimate of the spread among CMIP5 models.
Fig 3. (a) Magnitude and (b) location of the maximum SST anomaly for each ENSO type. (c) Westward extension of SST anomaly estimated as the location of half of the simulated maximum SST anomaly. NB: The y-axis in (b) and (c) is asymmetric to account for nonlinearity in warm/cold events. Error bars (also represented as dashed lines) in HadISST show the 90% confidence interval for the mean of observed events. Error bars in the multi-model mean (MMM) represent the 90% confidence intervals for the mean of composited events across the CMIP5 models. The mean and associated error of each type of ENSO event are specified with the same color as in the legend for the MMM and observations. Models that are within the dashed lines are considered to have a “realistic” simulation of the metric.
Fig 4. Composite of wind stress anomalies (zonal component shaded) during Dec-Feb season for ENSO events. Units in Pa. Left column: NCEP/NCAR Reanalysis. Middle column: Multi-model mean based on 20 CMIP5 models. Areas within the thin grey line are statistically significant across the observed events and the composited events of CMIP5 models at the 0.05 significance level based on a Student t-test. Right column: Dec-Feb zonal wind stress anomalies averaged over 3°S-3°N. Red line represents the multi-model mean, while blue line represents reanalysis. A 15-degree window running mean was applied to the curves. Light red shade indicates the standard deviation of simulated composites, an estimate of the spread among CMIP5 models.
Fig 5. Multi-model mean ocean temperature anomalies (shaded) averaged across 3°S-3°N along the Pacific during the Dec-Feb season for ENSO events. The difference between the multi-model mean and observations are contoured using 0.3 Celsius intervals. Red (blue) contours are positive (negative) differences. (a) Cold Tongue El Niño, (b) Warm Pool El Niño, (c) Cold Tongue La Niña, and (d) Warm Pool La Niña. Units in Celsius. Multi-model mean based on 23 CMIP5 models.
Fig 6. Composite of ocean heat content anomalies averaged across 3S-3N along the Pacific during the Dec-Feb season for ENSO events. The light grey curve is the multi-model mean heat content and the dark grey curve represents the SODA reanalysis. Light grey shade indicates the standard deviation of simulated composites, an estimate of the spread among CMIP5 models. The curves were smoothed with an 11-longitude point running mean. (a) Cold Tongue El Niño, (b) Warm Pool El Niño, (c) Cold Tongue La Niña, and (d) Warm Pool La Niña. Multi-model mean based on 23 CMIP5 models.
Fig 7. Hovmoeller diagram of the SST (shaded) and wind stress (vectors) anomalies averaged between 5°S and 5°N across the Pacific Ocean during ENSO events. Left column: Observations from HadISST dataset and reanalysis from NCEP/NCAR. Middle column: Multi-model mean of 20 CMIP5 models. Right column: Evolution of zonal wind stress anomalies (Pa) averaged between 5°S-5°N, 120°E-110°W. Red line: Multi-model mean. Blue line: NCEP/NCAR reanalysis. Red and blue lines were smoothed with an 11-month running mean. Light red area represents the standard deviation of the multi-model mean as an estimate of the spread across the models. (a,e,i) Cold Tongue El Niño, (b,f,j) Warm Pool El Niño, (c,g,k) Cold Tongue La Niña, and (d,h,l) Warm Pool La Niña.
Fig 8. Hovmoeller diagram of the ocean heat content (shaded) and zonal wind stress (contours) anomalies averaged between 3°S and 3°N across the equatorial Pacific Ocean during ENSO events. Brown (green) contours are westerly (easterly) anomalies, plotted on 0.003Pa intervals. Multi-model mean of 16 CMIP5 models. An 11-month window running mean was applied to the data. (a,e) Cold Tongue El Niño, (b,f) Warm Pool El Niño, (c,g) Cold Tongue La Niña, and (d,h) Warm Pool La Niña. Units in Celsius and Pa, respectively.
Evolution of SST anomalies averaged across the equatorial Pacific (5°S-5°N, 150°E-90°W) for individual models. Composite for (a) Cold Tongue El Niño events, (b) Warm Pool El Niño, (c) Cold Tongue La Niña, and (d) Warm Pool La Niña. Models are ordered according to correlations with observations. Units in Celsius.
Fig 11. Monthly standard deviation of (a) Niño4 (b) Niño3.4 and (b) Niño3 indices for CMIP5 models. For comparison purposes, the monthly standard deviation is divided by the maximum value, which is indicated in white in the month when it peaks. Models are ordered according to correlations with observations.
Fig 12. Left column (a-d): Difference in the simulated ENSO SST anomaly composites during Dec-Feb season between the RCP8.5 and historical scenarios. Units in Celsius. Areas within the thin grey line are statistically significant at the 0.05 significance level based on a Student t-test. Right column (e-h): SST anomaly averaged over the equatorial Pacific (5°S-5°N). Red line represents the multi-model mean for RCP8.5, while the blue line represents the historical simulation. Light red (blue) shade indicates the standard deviation of simulated composites for the RCP8.5 (historical), an estimate of the spread among CMIP5 models of each scenario. Based on 27 CMIP5 models.
Fig 13. Difference in the standard deviation of (a) Niño3, (b) Niño3.4, and (c) Niño4 indices between RCP8.5 and historical simulations for 27 CMIP5 models. (d) Difference in the ratio of the standard deviation between Niño4 and Niño3. Grey dashed line represents the difference in the multi-model mean. Zero appears as the black dashed line. Vertical bars represent the range of ensemble members when available and circles the respective ensemble mean.
Fig 14. Difference in the number of events/100yr between RCP8.5 and historical simulations for 27 CMIP5 models. (a) CT El Niño (b) WP El Niño (c) CT La Niña (d) WP La Niña. Grey dashed line represents the difference in the multi-model mean. Zero appears as the black dashed line. Vertical bars represent the range of ensemble members when available and circles the respective ensemble mean.