



An evaluation of ENSO dynamics in CMIP simulations in the framework of the recharge oscillator model

Asha Vijayeta¹ · Dietmar Dommenges¹

Received: 8 March 2017 / Accepted: 16 October 2017 / Published online: 18 November 2017
© Springer-Verlag GmbH Germany 2017

Abstract

The CMIP model simulations show wide spread uncertainties in ENSO statistics and dynamics. In this study, we use the concept of the linear recharge oscillator (ReOsc) model to diagnose the ENSO-dynamics in CMIP3 and CMIP5 model simulations. The ReOsc model parameters allow us to quantify SST and thermocline damping, SST coupling to thermocline and vice-versa, sensitivity to wind stress and heat flux forcings and separate atmospheric from oceanic processes. Our results show that the ENSO-dynamics and their diversity within the CMIP ensemble can be well represented with the linear recharge oscillator model diagnostics. We also illustrate that the ENSO dynamics show larger biases relative to observations and spread within the models than simple large-scale statistics such as SST standard deviation would suggest. The CMIP models underestimate the atmospheric positive and negative feedbacks, they have compensating atmospheric and oceanic errors, the thermocline damping is too strong and stochastic noise forcings in models is too weak. The CMIP5 models show only marginal improvements relative to CMIP3. The results suggest that models can still be significantly improved and our analysis gives directions to what needs to be improved.

Keywords El Nino southern oscillation · ENSO · Ocean and atmospheric dynamics · ENSO dynamics · El Nino dynamics · Coupled general circulation models · CGCM · Model evaluation · Recharge oscillator model · Climate feedbacks · CMIP simulations

1 Introduction

El-Nino-Southern Oscillation (ENSO) is the dominant mode of interannual climate variability in the tropical Pacific. ENSO has its origins in the tropical Pacific but it is known to influence the weather all over the world. The dynamics that control ENSO are important for global seasonal climate predictions, but are also important for long-term global climate change.

ENSO being a result of complicated dynamical processes encompasses several atmospheric and oceanic feedbacks. The main linear atmospheric feedbacks are the Bjerknes

feedbacks (Bjerknes 1969) and the net atmospheric heat flux feedback (Zebiak and Cane 1987). The Bjerknes feedbacks are essentially a positive feedback loop leading to ENSO growth and the net atmospheric heat flux feedback is a negative feedback. Several conceptual models have been proposed which condense the dynamics of ENSO into simple theoretical frameworks like the delayed action oscillator (Suarez and Schopf 1988), the recharge oscillator (Jin 1997a, b) and the further simplified recharge oscillator (Burgers et al. 2005). For our analysis, we use the latter and refer to it as the ReOsc model henceforth.

State-of-the-art coupled general circulation models (CGCMs) are capable of simulating the ENSO dynamics albeit with some biases with respect to the observed ENSO characteristics. Latif et al. (2001) carried out an intercomparison of an ensemble of twenty-four CGCMs in terms of performance of the annual mean state, the seasonal cycle and the interannual variability and almost all models (even those employing flux corrections) exhibited problems in simulating the sea surface temperature (SST) climatology. Although our understanding of ENSO has improved over the last decades

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00382-017-3981-6>) contains supplementary material, which is available to authorized users.

✉ Dietmar Dommenges
dietmar.dommenges@monash.edu

¹ School of Earth, Atmosphere and Environment, ARC Centre of Excellence for Climate System Science, Monash University, Melbourne, VIC 3800, Australia

CGCM simulations of the phase 3 and 5 of the Coupled Model Intercomparison Project (CMIP3 and CMIP5) still have difficulties in simulating the climatology of the tropical Pacific, this includes the correct intensity and spatial structure of the East Pacific cold tongue along the equatorial Pacific (Bellenger et al. 2014; Reichler and Kim 2008), mean thermocline depth and slope along the equator and the structure of the equatorial currents (Brown and Fedorov 2008), mean zonal equatorial wind stress (Guilyardi 2006) and meridional extent of the wind variability which is important for ENSO phase change (Zelle et al. 2005).

Because of the scale of ENSO effects on the weather all over the world, changes in dynamics of ENSO in the future is an important scientific question that needs to be answered. Oldenborgh et al. (2005) showed no statistically significant changes in amplitude of ENSO variability in the future. Uncertainties in the variability being large, they estimated very little influence of global warming on ENSO. On the other hand, Collins et al. (2010) report that global warming may change mean climate of the Pacific region, which in turn may modify one or more of the physical processes responsible for determining the characteristics of ENSO. Expected changes would be weakening of tropical easterly trade winds, faster warming of ocean surface near the equator and more slowly farther away; shoaling of the equatorial thermocline along with steeper temperature gradients across the thermocline. Change in non-linear feedbacks like cloud-albedo or thermocline-SST feedback could also lead to different ENSO dynamics. However, confidence in these findings is undermined by the diversity in the model projections. Simulations from different CGCMs result into very different future projections of ENSO (Collins et al. 2010). Here again the uncertainty in the CGCM simulation of the ENSO dynamics is a limiting factor. The relative importance of different processes contributing to the ENSO dynamics is different from model to model and different to those observed.

The aim of this study is to evaluate the CGCM simulation of the CMIP 3 and 5 databases in their performance of ENSO dynamics. We will base our analysis on the ReOsc model, which has been used in many studies to analyze ENSO dynamics in different ways (Burgers et al. 2005; Frauen and Dommenget 2010; Jansen et al. 2009; Levine et al. 2015; Yu et al. 2016). The ReOsc model being a simplified representation of ENSO dynamics, allows us to diagnose the dynamical parameters of the ENSO variability from the statistics of the model simulations or observations. The advantage of this approach in comparison to simple statistical parameters (e.g. SST standard deviation, SST pattern) or heat budget analysis (e.g. the BJ-index from Jin et al. 2006) is that it defines a dynamical framework with only a small numbers of parameters. The dynamical framework allows us to determine the sensitivity of ENSO statistics to different dynamical parameters. Further, the simplicity of the model allows for a large reduction of complexity

in the ENSO dynamics. This approach will help us to get from model evaluations towards direct model developments by linking error in large-scale statistics with dynamical processes that are closer to the processes that are simulated in CGCMs.

This paper is organized as follows. The data sets, model and methods used are described in Sect. 2. The statistical analysis and test of the ReOsc model skill to represent ENSO dynamics is presented in Sect. 3. The main results of this study are presented in Sect. 4. It addresses the CMIP model ENSO dynamics in terms of ReOsc model parameters along with the effect of CMIP model parameter distributions and biases on ENSO dynamics. Section 5 presents a dynamical skill (bias) score of each CMIP model based on the results of the previous sections. The study is concluded with a summary and discussion in Sect. 6.

2 Data, models and methods

For observational data we use the 1950–2014 HadISST1.1 data (Rayner et al. 2003) for SST and the 1982–2002 BMRC 20 °C isotherm depths of (Smith 1995) as an estimate for the thermocline depth. Since the BMRC 20 °C isotherm depths record is much shorter than the SST record we restrict our analysis of SST and thermocline depth co-variability to 1982–2002. The 1979–2014 ERA Interim data set is used for estimates of zonal surface wind stress (Dee et al. 2011) and the 1984–2004 OA Flux for estimates of surface heat fluxes (Yu and Weller 2007).

We analyze model simulations from the CMIP3 and CMIP5 databases (Meehl et al. 2007; Taylor et al. 2012). For the CMIP3 models we use the “20 cm³” simulations and the “historical” scenario for the CMIP5 models for the years 1900–1999. We used all available models that provide all climate variables needed for the following analysis. These are 10 CMIP3 and 29 CMIP5 models; see Table 1.

All analysis is based on monthly mean anomaly time series for thermocline depth averaged over the equatorial Pacific (130°E–80°W, 5°S–5°N) and NINO3 (150°W–90°W, 5°S–5°N) SST index and net heat flux, and central Pacific (160°E–140°W, 6°S–6°N) zonal surface wind stress. Monthly anomalies are computed by subtracting the mean seasonal cycle for each dataset.

The ReOsc model from (Burgers et al. 2005) is given by two tendency equations of the NINO3 region SST anomalies, T , and equatorial Pacific mean thermocline depth anomalies, h :

$$\begin{aligned}\frac{dT(t)}{dt} &= a_{11}T(t) + a_{12}h(t) + \zeta_1 \\ \frac{dh(t)}{dt} &= a_{21}T(t) + a_{22}h(t) + \zeta_2\end{aligned}\quad (1)$$

The model parameters a_{11} and a_{22} represent the damping (or growth rate) of T and h , and the parameters a_{12} and a_{21}

Table 1 CMIP3 and CMIP5 models with the corresponding model numbers

No.	CMIP3 model	No.	CMIP3 model
1	CGCM3.1	6	GISS-ER
2	CGCM3.1.T63	7	IPSL-CM4
3	CNRM-CM3	8	MPI-ECHAM5
4	GFDL-CM2.0	9	MRI-CGCM2.3.2
5	GISS-AOM	10	UKMO-HADCM3
No.	CMIP5 model	No.	CMIP5 model
1	ACCESS1-0	16	GISS-E2-R
2	ACCESS1-3	17	GISS-E2-R-CC
3	CCSM4	18	HadCM3
4	CESM1-BGC	19	HadGEM2-CC
5	CESM1-CAM5	20	HadGEM2-ES
6	CESM1-FASTCHEM	21	IPSL-CM5A-MR
7	CESM1-WACCM	22	IPSL-CM5B-LR
8	CNRM-CM5	23	MPI-ESM-LR
9	CSIRO-Mk3-6-0	24	MPI-ESM-MR
10	CanESM2	25	MRI-ESM1
11	FGOALS-g2	26	NorESM1-M
12	GFDL-ESM2G	27	NorESM1-ME
13	GFDL-ESM2M	28	bcc-csm1-1-m
14	GISS-E2-H	29	CMCC-CM
15	GISS-E2-H-CC		

the coupling between T and h . The two equations are forced by stochastic noise terms ζ_1 and ζ_2 . The parameters of the 2-dimensional model Eq. (1) are estimated for observations and also for each CMIP model simulation by multivariate linear regressing the monthly mean tendencies of T and h against monthly mean T and h , respectively. Following the approach in previous studies (Burgers et al. 2005; Frauen and Dommenget 2012; Jansen et al. 2009). The residual of the linear regression fit can be interpreted, as the random noise forcings with the standard deviation (stdv) of the residuals being the stdv of the noise forcings for the T and h equations (ζ_1 and ζ_2).

The ReOsc model strongly simplifies the ENSO dynamics and each of the 4 parameters and the two noise forcing terms can be a result of many different physical processes in the atmosphere and oceans. To further untangle the complexity the parameters, a_{11} and a_{21} can be split up into an atmospheric, a_{11A} and a_{21A} , and oceanic part, a_{11O} and a_{21O} following the approach of Frauen and Dommenget (2010) and Yu et al. (2016):

$$\begin{aligned} a_{11} &= a_{11O} + a_{11A} \\ a_{21} &= a_{21O} + a_{21A} \end{aligned} \tag{2}$$

The atmospheric damping (or growth rate) of T , a_{11A} , is effectively a coupling to wind stress and net heat flux (Frauen and Dommenget 2010):

$$a_{11A} = a_{12} \lambda C_{\tau T} + \frac{C_{TT}}{\gamma} \tag{3}$$

The coefficient $C_{\tau T}$ is the linear regression of zonal wind stress, τ_x , in the central Pacific box and NINO3 SST. This essentially represents one of the three Bjerknes feedbacks. C_{TT} is a linear regression between net atmospheric heat flux and SST in the NINO3 region. λ is a positive free coupling parameter and γ the ocean mixed layer depth following the approach of Frauen and Dommenget (2010), which is based on the study of Jin et al. (1997a). The atmospheric part of the coupling of h to T , a_{21A} , can be expressed as:

$$a_{21A} = \frac{a_{22}}{2} \lambda C_{\tau T} \tag{4}$$

The oceanic parts of the couplings to T , a_{11O} and a_{21O} , in Eq. (2) can be estimated as the residuals of a_{11} and a_{21} in Eq. (2) when a_{11A} and a_{21A} are estimated from Eqs. (3) and (4). All parameters values as estimated from observations or calculated are listed in Table 2 for an overview. In the analysis part the parameters $C_{\tau T}$, C_{TT} , a_{11O} and a_{21O} will be estimated for the CMIP simulations in the same way as for the observations. To reduce the complexity in the analysis we assumed γ to be the same for all models. The coupling parameter λ is fixed for all analysis, as it was estimated in

Table 2 All model parameters values as estimated from observations or calculated otherwise

a_{11} (T damping/growth rate)	-0.074 [1/month]
a_{12} (T coupling to h)	0.021 [C/m/month]
a_{22} (h damping/growth rate)	-0.022 [1/month]
a_{21} (h coupling to T)	-1.23 [m/C/month]
stdv (ζ_1) (amplitude of T noise)	0.25 [C/month]
stdv (ζ_2) (amplitude of h noise)	2.22 [m/month]
$C_{\tau T}$ (wind-SST feedback)	0.011 [N/m ² /C]
C_{ff} (atmos. heat-SST feedback)	-16.9 [W/m ² /C]
λ (free coupling parameter)	2100 [m ³ /N]
γ (scaled ocean mixed layer depth)	79.0 [C m ² /W/month]

(Frauen and Dommenget 2010) from running CGCM sensitivity studies.

It should be noted here that in model simulations the regions of wind-SST or other interactions may be shifted in location (e.g. east–west shift), but be otherwise similar to observations. This would lead to changes in the estimated parameters. We will not discuss such variations due to regional shifts, but leave this to future studies.

The ReOsc model Eqs. (1) can be integrated with stochastic noise forcing terms ζ_1 and ζ_2 to generate stochastic time series of T and h . We therefore integrated the equations with a time step of 24 h and red noise forcing terms ζ_1 and ζ_2 . The decorrelation time of ζ_1 and ζ_2 is set to 3 days to mimic weather fluctuations that effectively results into white noise for monthly mean data. Monthly mean stdv of the noise forcings is the stdv as observed or estimated from the CMIP simulations in Eq. (1) as described above.

3 Proof of concept

Before we apply the ReOsc model to diagnose the ENSO dynamics in the CMIP model simulations we first like to start with a proof of concept. We therefore do a number of different analyses to illustrate the skill of the ReOsc model in diagnosing the ENSO behavior in different model simulations.

We start the analysis by evaluating the observed time series of T and h , see Fig. 1a. In addition to the standard deviations of T and h , the power spectrum of T and the lag-lead correlation between T and h are two important statistical characteristics describing the ENSO behavior; see Fig. 1c, d. The observed time evolution of h leads that of T by about 5–6 months (peak of cross-correlation), which is indicative of the ENSO recharge and discharge mechanism, which is the fundament of the ReOsc model.

The stochastic integration of the ReOsc model with the observed parameters (see Table 2) shows very similar stochastic behavior. Figure 1b shows a 20 years sample from

the 1000 years long integration. The statistical properties stdv(T), stdv(h), power spectrum of T and the cross correlation between T and h based on the 1000 years integration of the ReOsc model are all similar to observed. The stdv(h), spectral slope and the cross-correlation are, however, slightly overestimated (Table 3). In summary, we conclude that the ReOsc model does replicate the main stochastic characteristic of T and h and their interaction fairly well.

Figure 2 shows the power spectrum and cross correlation between T and h for four CMIP5 models. Here we highlighted four CMIP simulations that have fairly different ENSO behaviors. They differ in the shape of the power spectrum, the overall variance and in the cross correlation between T and h . The stochastic integrations of the ReOsc model with the parameter from the four CMIP model simulations (see Sect. 2 for details) do replicate the differences in these main statistical characteristics fairly well. For instance, the ReOsc integration mimicking the bcc-csm1-1-m simulation has a much stronger peak in the power spectrum than the one for the CMCC-CM, replicating the same difference seen in the original CMIP simulations of bcc-csm1-1-m and CMCC-CM. Similarly, the weak cross correlation between T and h in the CMCC-CM simulation is well captured by the ReOsc integration. These preliminary results suggest that the ReOsc model is capable of capturing the most important characteristics of ENSO behavior in CMIP simulations.

We now compare four main statistical properties [stdv(T), stdv(h), power spectrum slope and cross correlation between T and h] of all CMIP simulations against the ReOsc integration mimicking the CMIP models, see Fig. 3. Here we first of all focus on the proof of concept, but we will in the next sections discuss the results further. The ReOsc model is able to replicate the stdv(T) and stdv(h) with very high correlations for both CMIP3 and 5 (see Fig. 3). It may, however, be argued that this is not totally unexpected, as we have fitted the model parameters to the T and h tendency Eqs. (1) for each model. This essentially enforces a close match in stdv(T) and stdv(h) if the ReOsc model has any value in presenting the main dynamics of the CMIP simulations. However, the fact that we closely reproduce the stdv(T) and stdv(h) does suggest that the ReOsc is capturing some essential elements of the ENSO dynamics in the CMIP simulations. It should be pointed out here that this model does not consider seasonal difference in the parameters (all parameters are constants) neither does it consider non-linearities. Thus, to first order these are not essential to understand the main ENSO statistics presented in this study.

The ReOsc model also has some skill in reproducing the variations in the power spectrum. To capture the difference in the power spectrum we estimated the slope (in log-scale) of the power spectrum for each model and observations for periods from 6 months to 5 years (see Fig. 3c). This marks the range of the power spectrum in which the variance is

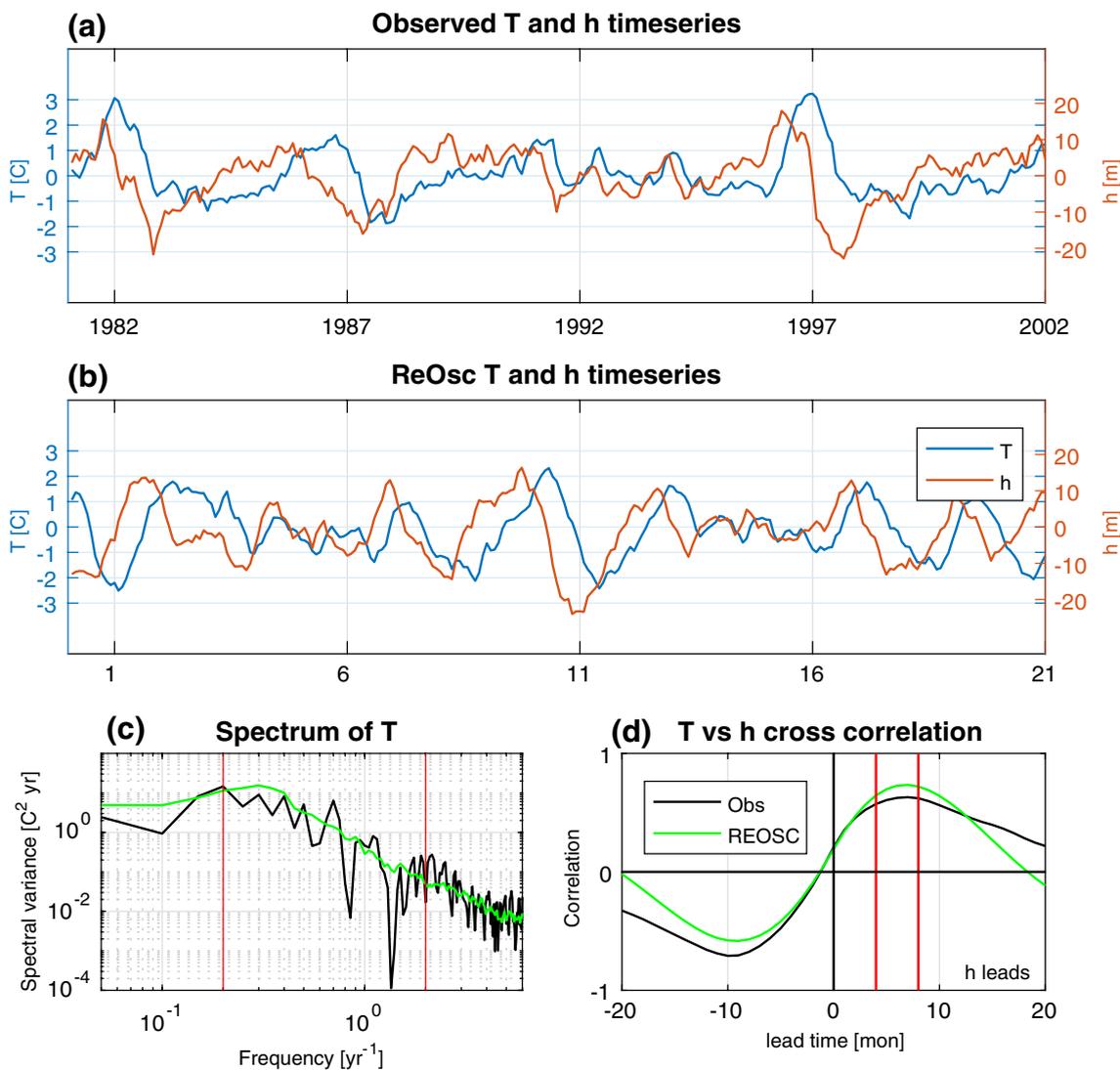


Fig. 1 **a** Time series of observed NINO3 SST anomaly (T) and mean equatorial Pacific thermocline depth (h) anomaly. **b** Time series of ReOsc toy model T and h . **c** Power spectra of T . The red vertical lines mark the 5 years and 0.5 years periods, which mark the period range

used to estimate the spectral slopes in the analysis sections. **d** Cross correlation between T and h . The solid vertical red lines are the 4 and 8 months lead, which mark the lag range used in the analysis sections

Table 3 Comparison of observational data and REOSC toy model estimates

	stdv of T	stdv of h	Spectral slope	Mean correlation (4–8 months lead)
Observation (1982–2002)	1.0	6.8	– 2.5	0.61
Recharge oscillator toy model	0.99	7.8	– 3.1	0.71

strongly increasing with period length (e.g. see Fig. 2c). The ReOsc model estimates are closely following the variations in the CMIP models suggesting that the ReOsc model describes much of the large-scale variations in the power spectra shape.

The mean cross correlation between T and h for 4–8 months lag (h leading T) is also fairly well captured by the ReOsc model (Fig. 3d). However, the ReOsc model tends to overestimate the cross correlation between T and h , suggesting that T and h are more tightly related in the ReOsc model than they are in the CMIP simulations. Given the simplicity of the ReOsc model this is not entirely unexpected.

Another way of testing the ReOsc model is to evaluate the tendencies of T and h . According to the ReOsc model

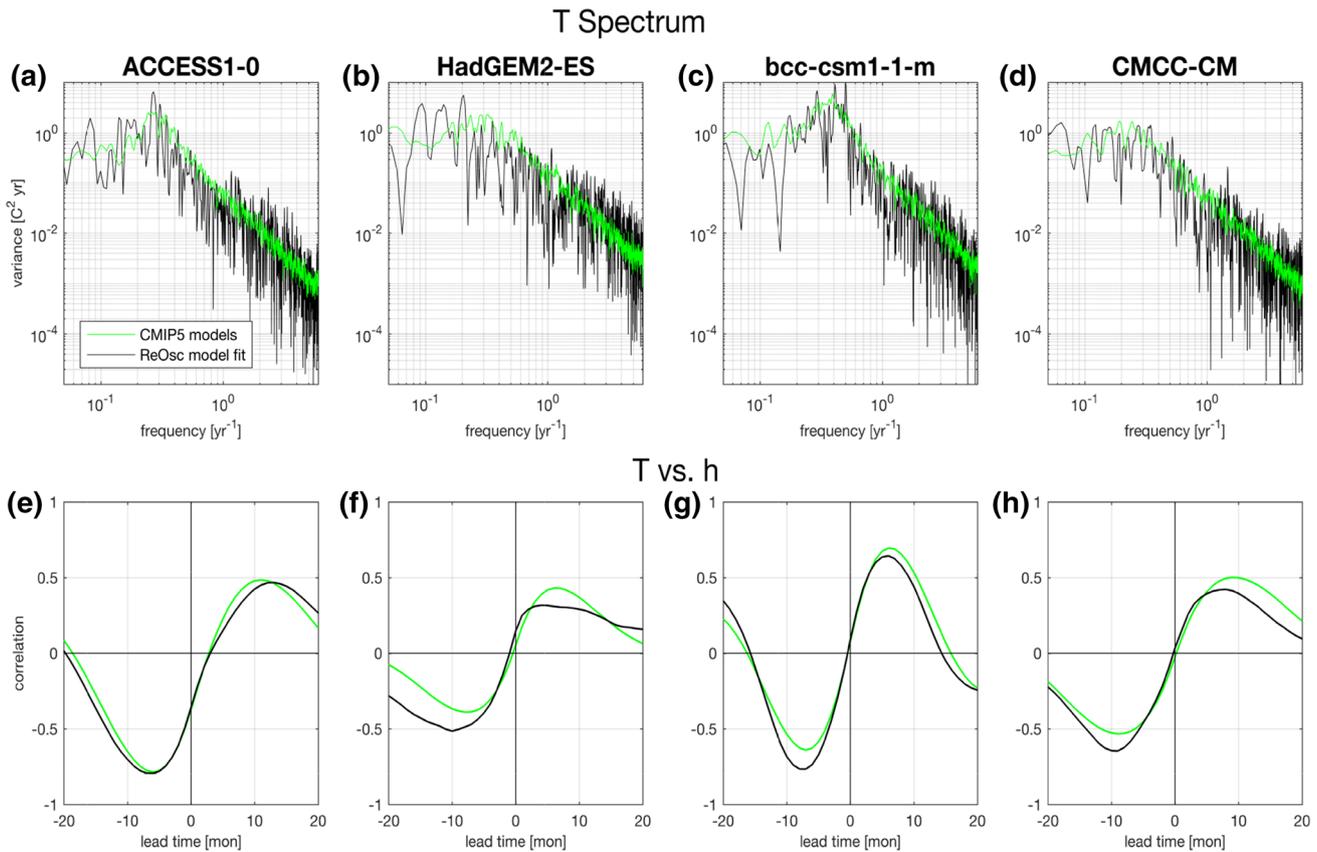


Fig. 2 Power spectra of T (upper) and cross correlations between T and h (lower) for some example CMIP5 models (black) and for the ReOsc model regenerated data (green). Positive lead times in the cross-correlations indicate h leading T

Eq. (1) the tendencies of T and h should be related to T and h themselves [first two terms on the rhs of Eq. (1)] and to the noise terms ζ_1 and ζ_2 . Given the estimated parameters we can evaluate what the correlation between the sum of the two T and h terms in Eq. (1) and the tendencies of T and h for all CMIP simulations, see Fig. 4a, b. The correlations for the tendencies of T in the ReOsc model are distributed somewhere between 0.2 and 0.8 with a mean of about 0.5. The spread in this distribution reflects the different ReOsc model parameters in the CMIP5 simulations. The distribution is similar for h , but the correlation for the h tendencies tend to be larger, indicating that h is more strongly forced by the two T and h terms than by the noise term [last term in Eq. (1)].

We can estimate what distributions of correlations we should expect from the ReOsc model, by doing Monte Carlo integrations of the ReOsc model with the same CMIP parameters. We therefore integrated a 1000 years long time series of the ReOsc model for each CMIP parameter set and computed the correlation values. We did these 100 times for each model. The ReOsc model distributions are very similar to those of the CMIP models, see red lines in Fig. 4 a and b. Spread, mean values and even more detailed variations in the distributions are similar to

those of the CMIP5 simulations, suggesting that the ReOsc model Eq. (1) does give a good approximation of the T and h tendencies and their relations to T and h themselves.

Finally, we test the noise estimates of the ReOsc model tendencies Eq. (1). The assumption of the ReOsc model is that the noise forcings ζ_1 and ζ_2 have a low lag-1 auto correlations. Thus, essentially being white noise. The distribution of the lag-1 auto correlations for ζ_1 and ζ_2 in the ReOsc model integrations are shown in Fig. 4c, d. The distributions have nearly no spread and have a mean of about 0.2, which is close to what is expected for white noise. The estimates of the residual noise for the CMIP5 simulations show a much wider spread and for ζ_2 (for h tendencies) the mean of the distribution is higher (~ 0.4). This indicates that the residual noise forcing estimates in the CMIP model show some deviations from the white noise estimate, in particular for the h tendencies. It illustrates some limitation of the ReOsc model in describing the fully complex CMIP simulations. However, in summary we conclude from this section that the ReOsc model gives a good first order representation of the main ENSO behaviors in the CMIP simulations and that the ReOsc

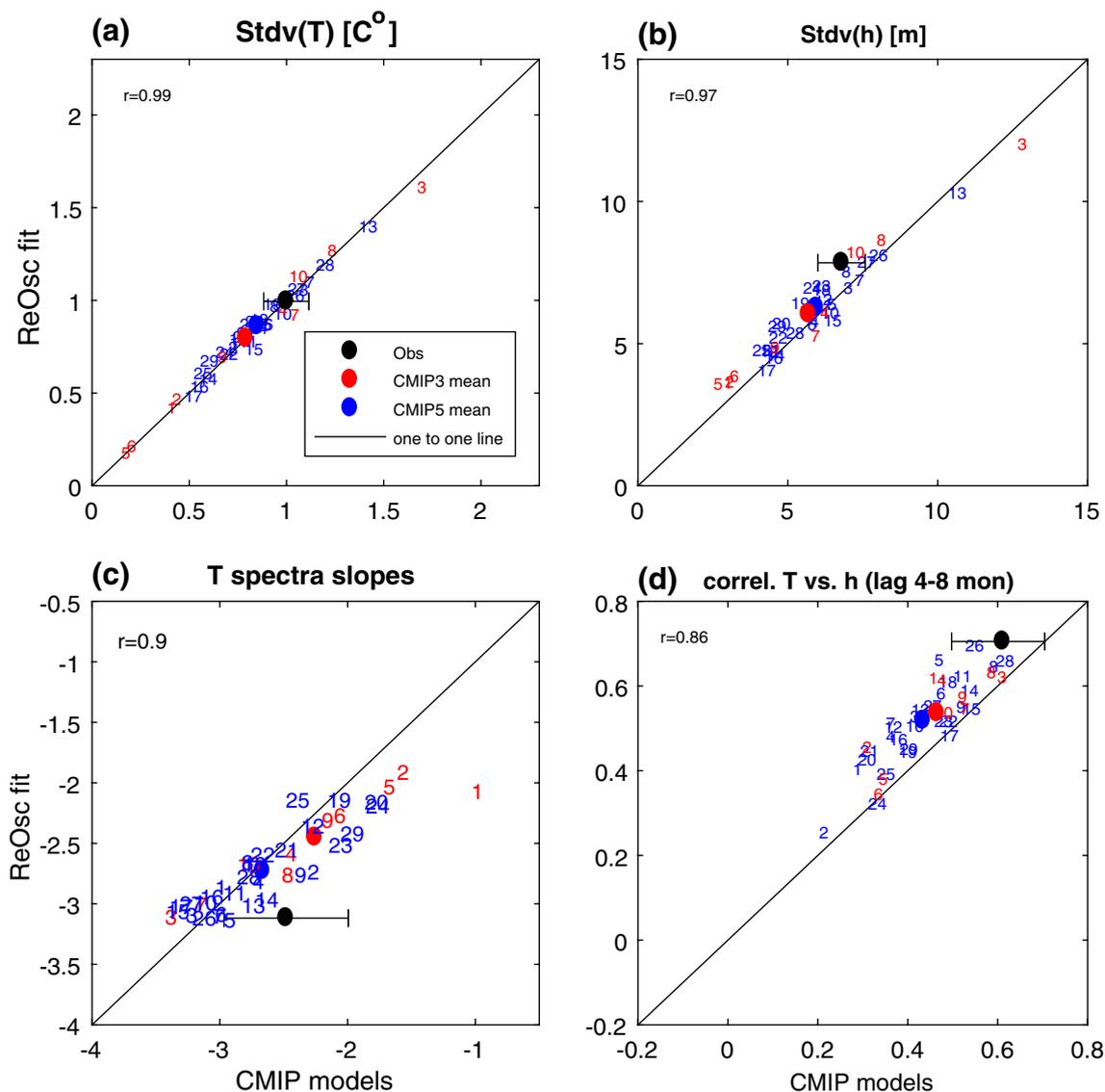


Fig. 3 Scatter plots of observed (black), CMIP3 (red) and CMIP5 (blue) data on *x*-axis and ReOsc toy model regenerated data on *y* axis for **a** standard deviation of *T* (°C); **b** standard deviation of *h* (m); **c** Spectral slope of *T* (log-scale) estimated from 5 to 0.5 years periods ($\log(^{\circ}\text{C}^2)/\log(\text{year}^{-1})$); **d** mean of the cross correlation between *T* and

h for lags 4–8 months. (*h* leading *T*). Observed error bars are the 90% confidence intervals. The *r* value marks the correlation between the *x*-axis vs. the *y*-axis of CMIP data points. Supplemental Table S1 lists all model values shown in this figure

model parameter estimates therefore are a good diagnostic to describe the ENSO behavior in these model simulations.

4 CMIP model ENSO dynamics

After establishing the efficiency of the ReOsc model in replicating the CMIP representation of ENSO, we now focus on the characteristics of the CMIP model simulations ensemble utilizing the ReOsc model parameter diagnostics in combination with other statistical parameters.

We start with a comparison of the CMIP models against observations in Sect. 4.1. This will be followed by an analysis of the atmospheric and oceanic contributions to the diversity in the dynamics of ENSO in Sect. 4.2 and a closer look at the CMIP model ensemble spread in Sect. 4.3. This will be followed by an analysis of the sensitivity of the ENSO dynamics to the different model parameters in Sect. 4.4, which will be utilized to develop a dynamical bias score for the models in Sect. 4.5 and to summarize the CMIP model parameter spread.

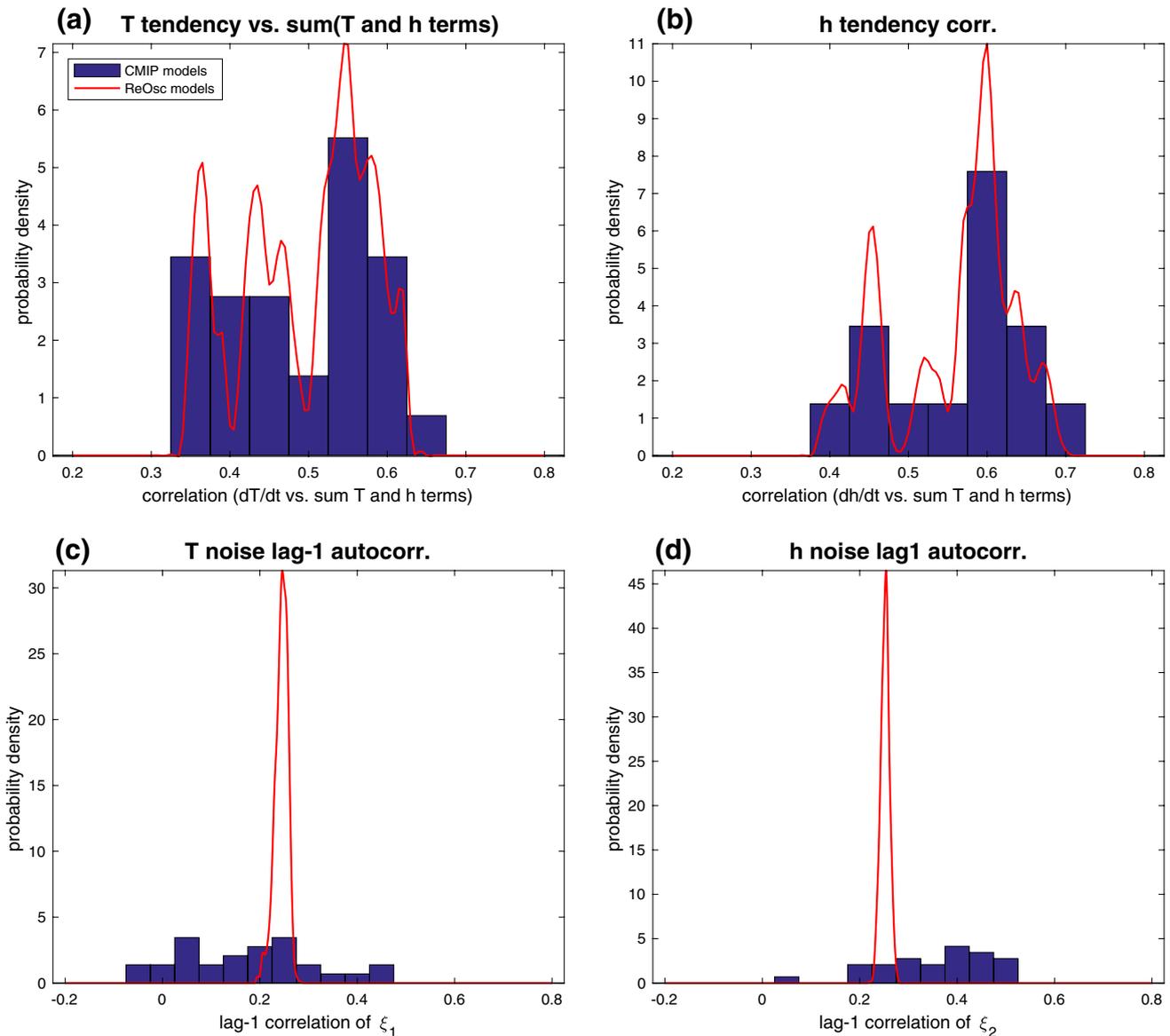


Fig. 4 Statistics of the T and h tendencies for proof of concept. Probability density of: **a** the correlation between dT/dt and the sum of the T and h term of Eq. (1); **b** the correlation between dh/dt and the sum of the T and h term of Eq. (1); **c** the lag-1 auto-correlation of ξ_1 , **d** the

lag-1 auto-correlation of ξ_2 . Blue bars are CMIP model data and the red lines are based on the ReOsc model stochastic integrations using parameter sets from the CMIP models. See text for details

4.1 Comparison towards observations

The main statistical properties [$\text{stdv}(T)$, $\text{stdv}(h)$, spectral slope and lag cross correlation between T and h] of the CMIP3 and CMIP5 ensemble means are fairly close to the observations (Fig. 3). The results of Bellenger et al. (2014) for the observed $\text{stdv}(T)$ agree fairly well with our findings. The $\text{stdv}(T)$, $\text{stdv}(h)$ and the cross correlation between T and h for 4–8 months. lag (h leading T) are all on average slightly underestimated by the CMIP3 and CMIP5 ensembles. However, more remarkable are the quite large spreads

in the CMIP3 and CMIP5 ensembles in $\text{stdv}(T)$ and $\text{stdv}(h)$. They are clearly inconsistent with the observed uncertainties. The spread within the models will be discussed in more detail in Sect. 4.3.

We now focus on the ReOsc model parameters; see Fig. 5. Here we combined the two parameters that influence the T tendency equation in Fig. 5a. Both T damping (a_{11}) and coupling to h (a_{12}) are very close to the uncertainties boundary of the observed values for the CMIP3 and CMIP5 ensemble means. However, the spread of these parameters is again larger than expected from statistical

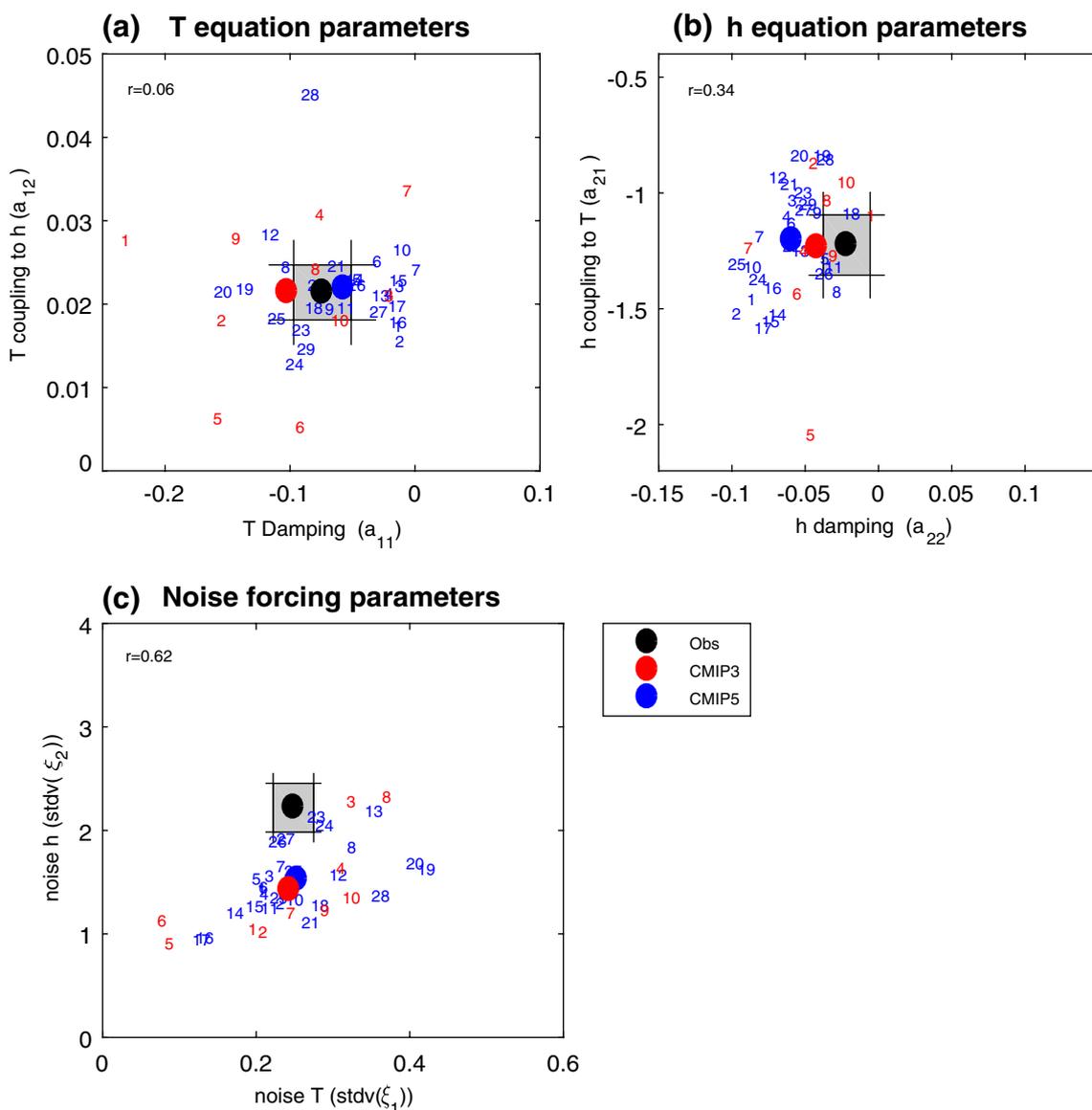


Fig. 5 Equation (1) parameters of the ReOsc model for observed (black with grey shaded area marking the 90% confidence interval), CMIP3 (red) and CMIP5 (blue) models. **a** T damping (a_{11}) vs. T coupling to h (a_{12}); **b** h damping (a_{22}) vs. h coupling to T (a_{21}); **c** noise

forcing for T [$\text{stdv}(\xi_1)$] vs. noise forcing h ($\text{stdv}(\xi_2)$). The r value marks the correlation between the x-axis vs. the y-axis of CMIP data points. See models and corresponding numbers in Table 1. Supplemental Table S2 lists all model values shown in this figure

uncertainties, suggesting that many CMIP3 and CMIP5 simulations have T damping (a_{11}) and coupling to h (a_{12}) significantly different from that observed.

The parameters influencing the h tendency equation are shown in Fig. 5b. The coupling to T is within the observed uncertainties for the CMIP3 and CMIP5 ensemble means suggesting a fairly good agreement of the models with observations in this parameter in average. The h damping (a_{22}) is in most model simulations clearly over estimated (more negative). This has even increased slightly from the CMIP3 to the CMIP5 ensemble (Fig. 5b).

The strength of the noise forcing estimates for T and h ($\text{stdv}(\xi_1)$ and $\text{stdv}(\xi_2)$) are shown in Fig. 5c. The mean T noise forcing is very similar to observed, but the strength of the h noise forcing is underestimated in most CMIP3 and CMIP5 simulations. There is also a very large spread within the model ensembles with some models having stronger noise forcings than observed. In particular, the noise forcing of the T tendencies (ξ_1) are much more widely spread within the models than would be expected from statistical uncertainties. Some models have less than half as much $\text{stdv}(\xi_1)$ than observed. Here it is interesting to note that despite

underestimating the forcing strengths and overestimating the damping of h (a_{22}) the overall variability of T ($\text{stdv}(T)$) is not strongly underestimated, but is in the ensemble mean close to observed.

An alternative way of evaluating the ReOsc model is to look at the relative contribution of the three forcing terms in rhs of Eq. (1). We estimated the relative contribution of each term by building the mean fraction that each of the three terms contribute to the sum of the absolute values of each term of each monthly mean, see Fig. 6. The largest (~50%) contribution to the observed T and h tendencies comes in average from the noise forcings, illustrating that ENSO is a strongly noise driven process. The second largest contribution (~36%) to the observed T tendencies comes from the coupling to h indicating the strong influence of

the thermocline variability onto the SST variability. This coupling is even stronger than the direct effect of T (~17%). Similarly, the tendencies of h are also more strongly forced by the coupling to T (~36%) and less so by h itself (~7%).

The ensemble means of CMIP3 and CMIP5 are very similar to the observed contributions to the T tendencies. However, both ensembles tend to underestimate the relative contribution of the noise forcings to the h tendencies. This is consistent with the above finding that the strength of the h noise forcing ($\text{stdv}(\zeta_2)$) was underestimated by the CMIP models and that the h damping was too strong. The relative contribution of h to the tendencies of T is also slightly underestimated.

Individual models show a fairly wide spread in the relative contributions of the different terms to the T and h

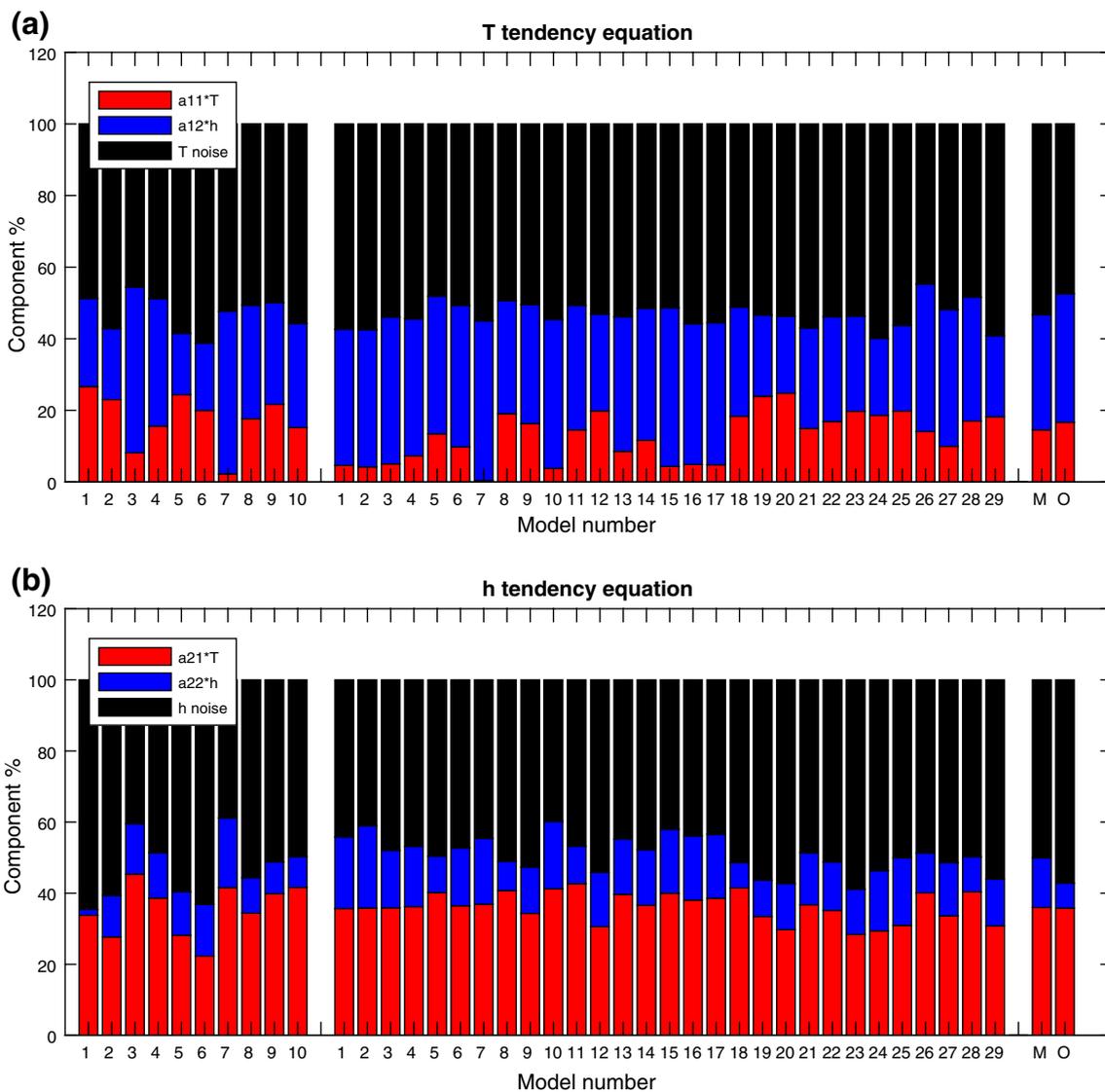


Fig. 6 Mean percentage contributions of the three rhs terms of Eq. (1) of the ReOsc model to **a** the monthly mean dT/dt and **b** the monthly mean dh/dt for observations (last bar), CMIP3 (first 10 bars), CMIP5 models (second rows of bars) and all model mean (second last bar)

tendencies. These can in general be linked to the variations in the model parameters (Fig. 5). CMIP5 model #20, for instance, has a strong contribution of T to the tendencies of T (Fig. 6a), which is consistent with the strong damping (a_{11}) in this model (Fig. 5a). In turn, CMIP5 model #7 with a near zero a_{11} value has almost no contribution of T to the tendencies of T . However, here we have to consider that the relative contributions cannot always be related to the absolute values of the parameters. Low forcing strengths in the CMIP models (Fig. 5c), for instance, do not necessarily imply low relative contributions to the tendencies, because the relative contributions of the forcings also result from the integrations with the other parameters of the model. For instance, CMIP5 model #14 does have a relative contribution of the forcings to the T tendency close to those observed, but at the same time it has much weaker noise forcing ($\text{stdv}(\zeta_1)$; see Fig. 5c) and weaker over all T variability (Fig. 3a). Biases in the other parameters of the model also contribute to the relative contribution of the forcings.

4.2 Contribution from atmospheric and oceanic dynamics

The ReOsc model parameters a_{11} and a_{21} can be split into an atmospheric and a residual oceanic part following the approach of Frauen and Dommenges (2010) (see Sect. 2). We will use this approach to get some further separation between atmospheric and oceanic processes.

We start the discussion with the atmospheric processes contributing to the T damping (a_{11A}), which are the wind-SST Bjerknes feedback ($C_{\tau T}$) and the atmospheric net heat flux (C_{IT}), see Fig. 7a. The observed $C_{\tau T}$ is positive representing a positive feedback and C_{IT} is negative representing a negative feedback for a_{11} (see Eqs. 2, 3). The combined observed atmospheric feedback on T is a positive growth rate a_{11A} (Fig. 7b), which is dominated by the positive Bjerknes feedback ($C_{\tau T}$).

The CMIP models ensemble means and indeed almost all model simulations underestimate both Bjerknes ($C_{\tau T}$) and the atmospheric net heat flux (C_{IT}), feedbacks. $C_{\tau T}$ is in ensemble mean only half as strong as observed and similarly C_{IT} is only half as strong as observed. This is qualitatively similar to the finding of Bellenger et al. (2014) and Lloyd et al. (2009). The combined atmospheric growth rate a_{11A} is in the ensemble mean still positive, but weaker than observed and some models have even negative atmospheric growth rates (damping). Even though the two biases in $C_{\tau T}$ and C_{IT} do compensate each other a little bit, as we have a positive and a negative feedback underestimated, they do not compensate each other completely, because the $C_{\tau T}$ term is much stronger for a_{11A} than the C_{IT} term.

The oceanic contribution to the growth rate of T (a_{11O}) is estimated as the residual of a_{11} after considering the

atmospheric part (see Eq. (2)). The observed oceanic feedback to T is a strong damping (Fig. 7b, c) that counter acts the positive atmospheric growth rates leading to the much weaker total damping of a_{11} . Thus, the weak total T damping (a_{11} in Fig. 5a) is a result of strong positive atmospheric growth rate and strong oceanic damping. Again, nearly all CMIP5 model simulations underestimate the oceanic damping and the ensemble means are significantly smaller than the observed value.

The coupling of h to T (a_{21}) can also be split into an atmospheric and oceanic part (see Eqs. (2) and (4)). Here the observed a_{21} is dominated by the oceanic part (a_{21O}) with a smaller contribution from the atmospheric part (a_{21A} ; Fig. 7c, d). This is similar in the CMIP ensembles, but the models tend to underestimate the oceanic part and overestimate the atmospheric part. The overestimation of a_{21A} results mostly from the too strong h damping (a_{21A} ; see Eq. (4) and Fig. 5b).

4.3 Spread within the model ensembles

In addition to how the models compare to the observations we can consider the spread within the model ensemble. If the models are consistent with each other within the statistical uncertainties of the data, then they should have a spread similar to the statistical uncertainties of the observations (Figs. 3, 5, 7). However, the CMIP3 and CMIP5 model ensembles show much larger spreads in all statistics and parameters shown. In particular, the ENSO statistics $\text{stdv}(T)$, $\text{stdv}(h)$ and the cross-correlation between h and T show large spread. Also, the spread in a_{11} , $\text{stdv}(\zeta_1)$ and $\text{stdv}(\zeta_2)$, $C_{\tau T}$, C_{IT} , a_{11A} and a_{11O} are very large. This highlights large diversity in the model simulations with models not only having very different ENSO statistics, but also having very different process parameters, suggesting that the processes controlling ENSO in the different model simulations can be very different from model to model.

The variation in the parameters within the CMIP model ensemble have many interesting cross-correlations. Some of them are shown in Figs. 5 and 7 and all cross-correlations are listed in Table 4. It is beyond this study to discuss all of these cross-relations in the parameters and most of these will be addressed in future further studies. In the following we like to point out a few interesting relations.

The ReOsc model parameters (Eq. (1); a_{11} , a_{12} , a_{22} , a_{21} , $\text{stdv}(\zeta_1)$ and $\text{stdv}(\zeta_2)$) show very little cross correlations within the CMIP ensemble. Only the strength of the noise forcings ($\text{stdv}(\zeta_1)$ and $\text{stdv}(\zeta_2)$) have a stronger positive correlation, suggesting that models that have stronger noise forcing on T also have strong noise forcing on h (see Fig. 5c). We can also notice a positive correlation between the strength of the noise forcing on T ($\text{stdv}(\zeta_1)$) and the

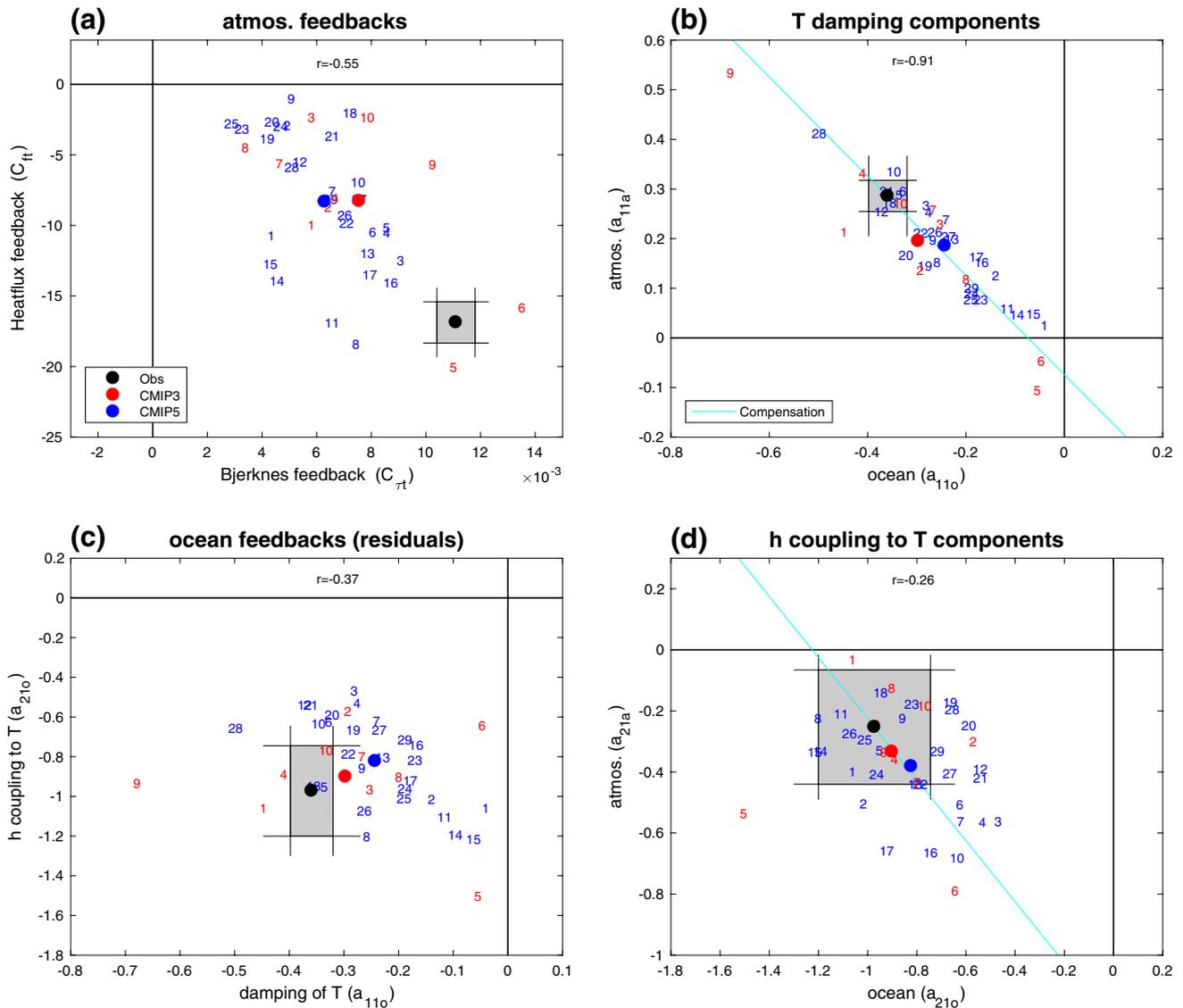


Fig. 7 Atmospheric and oceanic parameter components for observed (black with grey shaded area marking the 90% confidence interval), CMIP3 (red) and CMIP5 (blue) models. **a** Atmospheric Bjerknes feedback ($C_{\tau T}$) ($N/m^2/^\circ C$) vs. atmospheric heat flux feedback ($W/m^2/^\circ C$); **b** oceanic (a_{11o}) and atmospheric (a_{11a}) components of T damping ($1/month$); **c** oceanic feedbacks of T damping (a_{11o}) ($1/month$) vs. h coupling to T (a_{21o}) ($m/^\circ C/month$); **d** oceanic (a_{21o}) and

atmospheric (a_{21a}) components of h coupling to T ($m/^\circ C/month$). The cyan line in **b** and **c** marks the compensation line at which atmospheric and oceanic components add up to the observed total values of a_{11} and a_{21} , respectively. The r value marks the correlation between the x-axis vs. the y-axis of CMIP data points. See models and corresponding numbers in Table 1. Supplemental Table S2 lists all model values shown in this figure

coupling of h to T (a_{21}). A reason for his correlation is unclear to us and would need further future investigations.

If we extend the discussion on all cross-correlation with the atmospheric and oceanic parameters (a_{11a} , a_{11o} , a_{21a} , a_{21o} , $C_{\tau T}$ and C_{TT}), we see many more significant cross-correlations. The most remarkable cross relation here is that the models tend to underestimate all three elements ($C_{\tau T}$, C_{TT} and a_{11o}) contributing to a_{11} yet the total a_{11} is about the same as observed (Fig. 5a). This suggests that the variations of the model simulations relative to the observed

atmospheric and oceanic process have strong compensating effects: model simulations that have too strong atmospheric growth rates also have too strong oceanic damping leading to a total T damping close to observed. This is illustrated in Fig. 7b: the models line up closely to the compensation line, which marks the line on which the sum of a_{11a} and a_{11o} adds up to the observed a_{11} . We are not aware of any physical process that could explain this behavior. We will discuss possible implications of this further in the final summary and discussion section.

Table 4 Cross-correlation in the variations of the parameters in CMIP ensembles

Cross correlations in CMIP model parameters

	a_{11}	a_{12}	a_{22}	a_{21}	$\text{stdv}(\xi_1)$	$\text{stdv}(\xi_2)$	a_{11a}	a_{11o}	a_{21a}	a_{21o}	$C_{\tau T}$	$C_{\tau \Gamma}$
a_{11}		0.06	-0.52	-0.21	-0.18	0.12	0.09	0.33	-0.49	0.14	0.04	-0.07
a_{12}			0.12	0.42	0.47	0.06	0.73	-0.67	0.29	0.23	-0.31	0.22
a_{22}				0.34	0.26	0.08	0.22	-0.42	0.64	-0.11	0.22	-0.03
a_{21}					0.64	0.28	0.48	-0.55	0.44	0.75	-0.31	0.61
$\text{stdv}(\xi_1)$						0.62	0.41	-0.47	0.59	0.26	-0.51	0.61
$\text{stdv}(\xi_2)$							0.04	0.01	0.33	0.05	-0.37	0.36
a_{11a}								-0.91	0.1	0.45	0.1	0.4
a_{11o}									-0.3	-0.37	-0.08	-0.41
a_{21a}										-0.26	-0.57	0.4
a_{21o}											0.09	0.35
$C_{\tau T}$												-0.55

Correlation values with magnitudes < 0.5 are shown in light grey. Values > 0.5 are blue if they are >0.5 in both CMIP3 and CMIP5 ensembles, otherwise they are in dark grey

Some other cross-correlations are quite remarkable too. The atmospheric feedbacks $C_{\tau T}$ and $C_{\tau \Gamma}$, for instance, are negatively correlated. Thus the variations in these two feedbacks seem to be having compensating effects on a_{11a} (see Eq. (3)). Subsequently, variations in a_{11a} are largely unrelated to variations in $C_{\tau T}$ and $C_{\tau \Gamma}$ (Table 4) despite a_{11a} being a direct function of both (Eq. (3)). Other studies have pointed out a similar relationship between $C_{\tau T}$ and $C_{\tau \Gamma}$. Bayr et al. (2017) argue that such a negative relationship can result from a shift in the mean state position of the uplifting branch of the walker circulation.

The atmospheric part of the h coupling to T (a_{21a}) is positively correlated to the h damping (a_{22}) as expected from Eq. (4). However, the Bjerkness feedback ($C_{\tau T}$) is also expected to be positively correlated following Eq. (4), but it has a significant negative cross-correlation. Again, we have no explanation for this unexpected relation.

4.4 Sensitivities of ENSO statistics to the ReOsc model biases and spreads

We now focus on the sensitivity of the ENSO statistics to the ReOsc model parameter biases and spreads discussed above. The advantage of using the conceptual framework of the ReOsc model is that we can evaluate the relative importance of different parameters or feedbacks biases by integrating

the dynamics ReOsc model and analyzing how the ENSO statistic changes.

We estimate the sensitivity of an ENSO statistic, σ_k , to a parameter, p_i , of the ReOsc model by integrating the ReOsc model with all other parameters set to the CMIP ensemble mean parameters and perturbing p_i by a small δp_i . The change in the ENSO statistic, $\delta \sigma_k$, relative to the control integration in which all parameters are set to the CMIP ensemble mean parameters gives us an estimate of the sensitivity λ_{ki} :

$$\lambda_{ki} \approx \frac{\delta \sigma_k}{\delta p_i} \tag{5}$$

All integrations are done for 500 years using the same white noise forcing. We estimate the sensitivities to $C_{\tau T}$, $C_{\tau \Gamma}$, a_{11o} and a_{21o} by incorporating them into the ReOsc model Eq. (1) replacing a_{11} and a_{21} using Eqs. (2–4). By multiplying the sensitivity, estimated with Eq. (5), with a difference in a particular parameter, Δp_i , we can estimate the change in the ENSO statistic $\Delta \sigma_{ki}$:

$$\Delta \sigma_{ki} = \Delta p_i \cdot \lambda_{ki} \tag{6}$$

It needs to be noted here that we made a first order linear approximation in this approach. However, the ENSO statistics in the ReOsc model are in general not a linear function of the parameters, but do show some non-linear behavior (e.g. changing $a_{11} = 0.1$ by $\Delta a_{11} = -0.1$ does not have the exact opposite effect of changing it by

$\Delta a_{11} = +0.1$). It also needs to be considered that the sensitivity of the λ_{ki} will in general depend on the values of all other parameters as well as.

Figure 8a shows the changes in $\text{stdv}(T)$ and $\text{stdv}(h)$ for the biases of the CMIP ensemble mean (of both CMIP3 and 5) relative to the observed values. First, we can note that the $\text{stdv}(T)$ and $\text{stdv}(h)$ behave similar: parameter changes that increase the $\text{stdv}(T)$ also increase $\text{stdv}(h)$.

Suggesting that the strength in variability T and h are strongly linked.

The largest changes in $\text{stdv}(T)$ and $\text{stdv}(h)$ result from the biases in the Bjerknes feedback $C_{\tau T}$, the residual oceanic T damping a_{110} and the atmospheric net heat flux C_{TT} . This reflects that all three have relatively large biases to the observed values (Fig. 7a, b) combined with a relatively large sensitivity of the ENSO statistics in the ReOsc model to

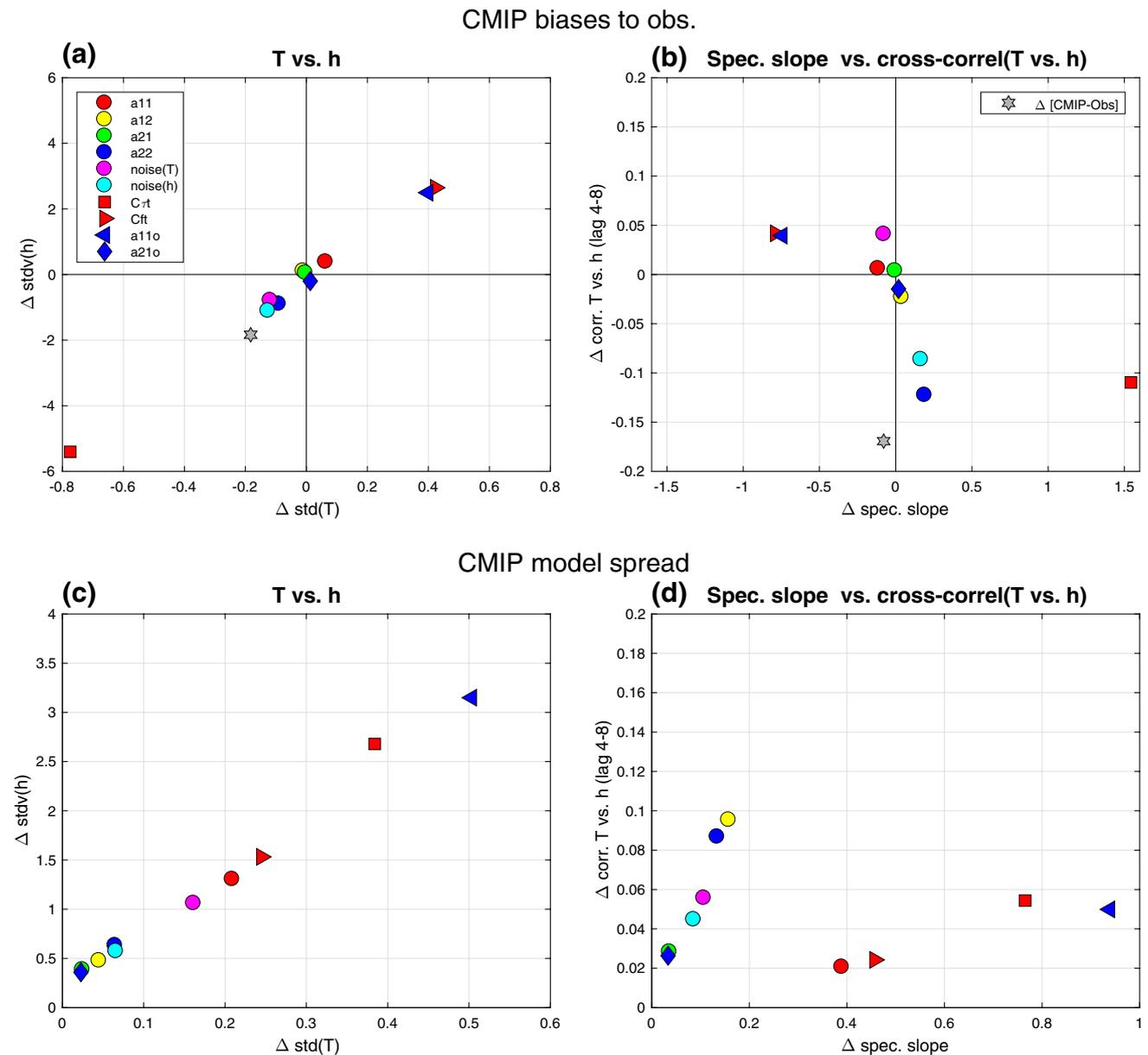


Fig. 8 Sensitivity of statistical properties to different parameters variations for model biases towards observations using Eq. (6) (upper) and for the model ensemble internal spread using the absolute values of Eq. (6) (lower). **a** Changes in $\text{stdv}(T)$ ($^{\circ}\text{C}$) and $\text{stdv}(h)$ (m) due to biases in the CMIP models relative to the observed; **b** changes in the spectral slope ($\log(^{\circ}\text{C}^2)/\log(\text{year}^{-1})$) and the cross correlation

between T and h for lags 4–8 months (h leading T) due to biases in the CMIP models relative to the observed; **c** same as **a** but due to the CMIP model internal spread; **d** same as **b** but due to the CMIP model internal spread. Negative $\Delta \text{std}(T \text{ or } h)$ in **a** and **b** indicate a reduction in the $\text{std}(T \text{ or } h)$ of the ReOsc model. See text for details

these parameters. In turn, the total T damping a_{11} has only a small sensitivity, although it has the same sensitivity (λ_{ki}) as a_{11o} . However, the bias to observed in a_{11} (Fig. 5a) is much smaller than in a_{11o} (Fig. 7b).

The much too weak $C_{\tau T}$ in the CMIP models would result into a much weaker $\text{stdv}(T)$ and $\text{stdv}(h)$, and the too weak a_{11o} would result into a stronger $\text{stdv}(T)$ and $\text{stdv}(h)$. The strongly opposing effects combine (Eqs. (2, 3)) into the total T damping a_{11} , which has very little bias to observations and therefore has very little effect on $\text{stdv}(T)$ and $\text{stdv}(h)$. The biases in the strength of the noise forcings and the h damping a_{22} also have some significant influence on $\text{stdv}(T)$ and $\text{stdv}(h)$, whereas the biases in a_{12} , a_{21} and a_{21o} have very little influence.

The effects of the parameters on the power spectrum slope and the cross-correlation between h and T , (r_{Th}), are shown in Fig. 8b. The spectral slope changes are remarkably similar to the changes in $\text{stdv}(T)$, but with reversed sign: parameters that increase the spectral slope (thus becoming less negative; see Fig. 3c) decrease the $\text{stdv}(T)$ (compare with Fig. 8a). This suggests that models with stronger SST variability also have a stronger increase in SST variance with increasing periods (more negative spectral slope) and thus have a more pronounced interannual variability. There is indeed a negative correlation of -0.6 between the $\text{std}(T)$ and the spectral slopes in the CMIP ensemble (compare Fig. 3a, c).

The cross-correlation r_{Th} is most strongly affect by the biases in the h damping (a_{22}), h forcing strength ($\text{stdv}(\zeta_2)$) and $C_{\tau T}$. Most other parameters have little effect on r_{Th} . There is a weak indication that changes in the parameters that lead to a stronger r_{Th} also lead to more negative spectral slope. This suggests that models with stronger delayed coupling between h and T have a stronger increase in SST variance with increasing periods (more negative spectral slope) and thus have a more pronounced interannual variability.

The CMIP models have fairly large spread within the ensemble in nearly all parameters. We estimate the sensitivity of the statistics to these parameter variations by using absolute values of Eq. (6) (neglecting the signs) with Δp_i being the stdv of the parameters within the CMIP ensemble, see Fig. 8c, d. The largest changes in statistics of $\text{stdv}(T)$ and $\text{stdv}(h)$ again result from the spread in $C_{\tau T}$, a_{11o} and $C_{\Gamma T}$. However, the spread in a_{11} and $\text{stdv}(\zeta_1)$ are now more important than in the bias towards observations. This is due to the fact that a_{11} had little bias towards observations, but have fairly large spread within the CMIP models (see Fig. 5a).

The spectral slope again behaves similarly to the $\text{stdv}(T)$ with the noticeable difference that the T coupling to h (a_{12}) has stronger influence on the spectral slope than it does on the $\text{stdv}(T)$. The T coupling to h (a_{12}) also has the most strongly influence on the cross-correlation r_{Th} . The damping of h (a_{22}) has a similarly strong influence, but most other parameters have weaker effects on r_{Th} .

5 Bias score of CMIP models

We can summarize the CMIP model biases relative to the observed ENSO ReOsc model parameters by combining all parameter biases into a normalized bias score, $S\sigma_{ki}$:

$$S\sigma_{ki} = \frac{1}{|\sigma_k(obs)|} \sqrt{\frac{1}{8} \sum_{n=1}^8 (\Delta p_n^i \cdot \lambda_{kn})^2} \tag{7}$$

with $S\sigma_{ki}$ the bias score for the ENSO statistic σ_k for the model with index i , the model bias in the parameter p_n relative to observed, Δp_n^i , and the observed ENSO statistic, $\sigma_k(obs)$. This bias score is effectively a root mean square error (RMSE) in the eight model parameters (a_{12} , a_{22} , $\text{stdv}(\zeta_1)$, $\text{stdv}(\zeta_2)$, $C_{\tau T}$, $C_{\Gamma T}$, a_{11o} and a_{21o}) scaled by the sensitivity of the ENSO statistics to these parameters (λ_{kn}) in the ReOsc model and normalized by the absolute value of the observed ENSO statistic, $\sigma_k(obs)$. It thus provides a dynamics-based bias score of ENSO. The higher the bias score the more the model dynamics deviate from the observed dynamics, hence the ideal model should have a bias score close to or within the observed uncertainties.

Figure 9 shows the bias scores for all CMIP models for the ENSO statistics $\text{stdv}(T)$, $\text{stdv}(h)$, spectral slope and the mean cross-correlation between h and T at lags 4–8 months (r_{Th}). In addition, we show an estimate of the observed uncertainties by replacing Δp_n^i in Eq. (7) with the 90% uncertainty values of the observed parameters as shown in Figs. 5 and 7.

There are a number of interesting aspects in the bias score results. Starting with the bias scores in $\text{stdv}(T)$ and $\text{stdv}(h)$ (Fig. 9a) we can first of all notice that all models are relatively far away from the observed uncertainty estimate. Much more than they are from any of the individual parameters (see. Figs. 5, 7). This is due to the role of compensating errors. In the definition of the bias score (Eq. (7)) we have implicitly assumed that the biases are independent. However, some of the parameters with the strongest sensitivities have very strongly compensating biases (see $C_{\tau T}$, $C_{\Gamma T}$ and a_{11o} Fig. 7a, b). While these compensating errors lead to relatively small biases in the ENSO statistics, they do not lead to small bias values in Eq. (7). Thus, the main reason why the models are far away from good bias values is the dominance of compensating errors. This is in particular captured by the bias in a_{11o} , which combines $C_{\tau T}$ and $C_{\Gamma T}$, and by the bias in a_{11o} (Fig. 7b). Models that perform well in these have in general small bias scores and vice versa.

Further we can notice that the bias scores in $\text{stdv}(T)$ and $\text{stdv}(h)$ are nearly identical. Models that perform well in $\text{stdv}(T)$ also perform well in $\text{stdv}(h)$. This first of all reflects that the correlation between $\text{stdv}(T)$ and $\text{stdv}(h)$ in the CMIP ensemble is fairly high ($r=0.9$). Thus, the

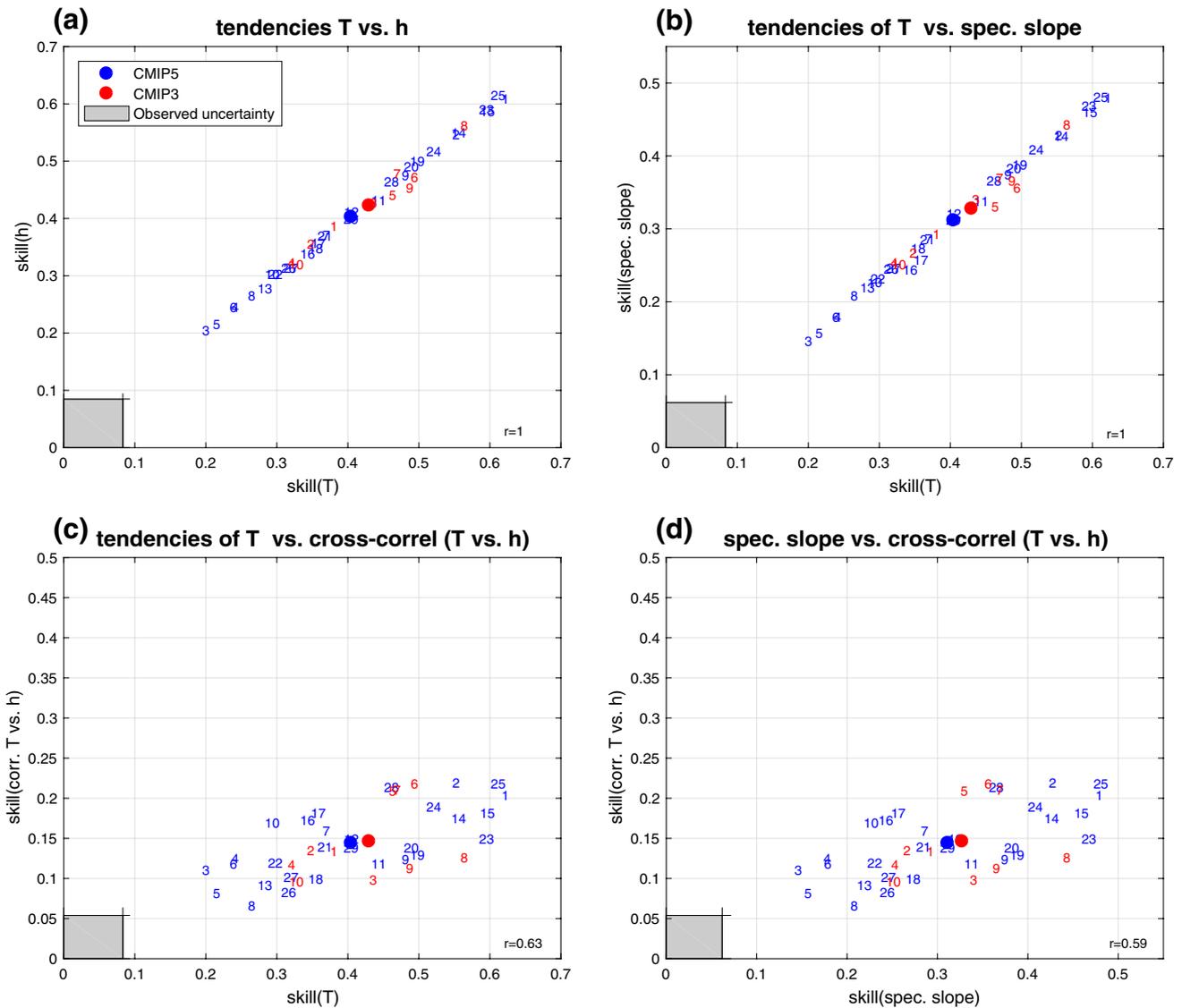


Fig. 9 Dynamical skill scores of CMIP model for different statistical properties. Skill scores for **a** $\text{stdv}(T)$ ($^{\circ}\text{C}$) vs. $\text{stdv}(h)$ (m); **b** $\text{stdv}(T)$ ($^{\circ}\text{C}$) vs. spectral slope ($\log(^{\circ}\text{C}^2)/\log(\text{year}^{-1})$); **c** $\text{stdv}(T)$ ($^{\circ}\text{C}$) vs. cross correlation between T and h for lags 4–8 months (h leading T); **d** spectral slope ($\log(^{\circ}\text{C}^2)/\log(\text{year}^{-1})$) vs. cross correlation

between T and h for lags 4–8 months (h leading T). The r value marks the correlation between the x-axis vs. the y-axis of CMIP data points. See models and corresponding numbers in Table 1. Supplemental Table S3 lists all model values shown in this figure. See text for details

strength of the SST variability in the CMIP model simulations is strongly linked to the strength of the thermocline variability. The tight relation in the bias scores also indicates that the sensitivities, λ_{ki} , in the ReOsc model for $\text{stdv}(T)$ and $\text{stdv}(h)$ are very strongly linked to each other. A parameter change that leads to a change in $\text{stdv}(T)$ also leads to an equivalent change in $\text{stdv}(h)$.

The bias scores of $\text{stdv}(T)$ and $\text{stdv}(h)$ are also relatively wide spread out with a clear separation between the models. Again, this is much stronger than in any of the individual parameters (see Figs. 5, 7). It is indicating that

the dynamical skills of the CMIP model is indeed very different between the models.

The bias score in the spectral slope behaves very similar to the skill score of the $\text{stdv}(T)$ (Fig. 9b). Here it has to be noted that the statistical estimate of the spectral slope of the T time series is independent of the $\text{stdv}(T)$. The tight relation therefore suggests that the spectral slope and the $\text{stdv}(T)$ are dynamically linked. This we already indicated in the discussion of the sensitivities above with the moderate negative correlation between spectral slope and the

Fig. 10 Summary of CMIP model parameter biases and spreads. ▶ Observed values and spread (90% interval) are marked by the black line and grey shaded area. The combined distributions of CMIP3 and CMIP5 models are shown as blue bars with the blue line marking the mean of the distributions. The parameters are sorted by how strongly the CMIP model’s parameter biases influence the stdv(T) as shown in Fig. 8a. The most influential is the upper most

stdv(T) in the CMIP ensembles ($r = -0.57$), but here the dynamical skill suggests an even tighter relationship.

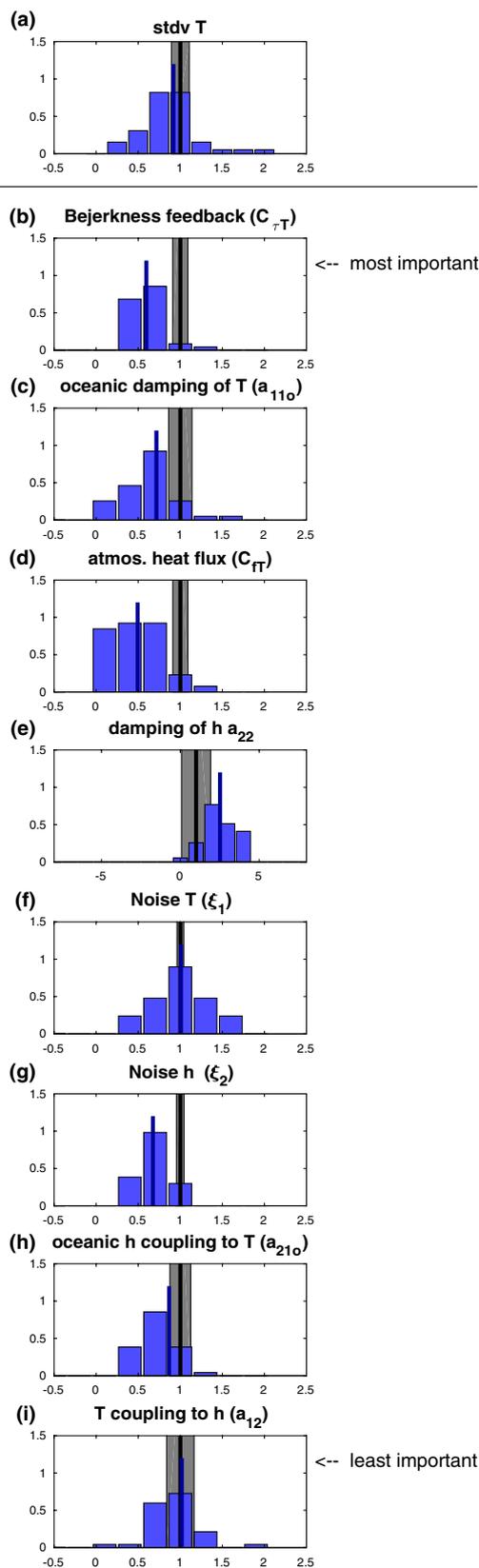
The bias score for cross-correlation r_{Th} is more independent from the other bias scores, but also shows less spread within the models and is also closer to the observed uncertainties. In summary of all skill scores we cannot see much difference in the bias scores of the ensembles means of the CMIP3 and CMIP5 simulation, although there is a weak tendency for the CMIP5 ensemble to have slightly smaller biases than the CMIP3 ensemble.

The bias scores of the models as a whole are surprisingly bad considering that the models match observed ENSO statistic much better than the bias scores would suggest. As mentioned above this is due to the fact that the models have compensating biases. A comparison of the CMIP model spread in stdv(T) with the CMIP model spread in the model parameters illustrates this, see Fig. 10. We can note that the ensemble mean stdv(T) is very close to the observed with the model distribution wider than the observed uncertainty, but still close to the observed. However, the most important parameters (Fig. 10b–e) all have substantial biases towards the observed values and relatively large spread. It is surprising that the models manage to simulate the observed stdv(T) so well given such large biases in the controlling ENSO dynamics.

6 Summary and discussion

In this study, we introduced the linear recharge oscillator model as a diagnostic tool to evaluate the representation of ENSO dynamics in the CMIP model database. We presented a proof of concept analysis that illustrated that ENSO-statistics and their diversity within the CMIP5 ensemble, can be well represented with the linear recharge oscillator model diagnostics. Although simplified to only represent the first order dynamics, presenting only a linear system, we believe it is an efficient tool to replicate ENSO dynamics in CMIP models. It provides a very effective bridge between simple statistical analysis of ENSO variability and the fully complex dynamical ENSO system with all its feedbacks and processes.

Starting with the simple statistics of the variability in T and h we found that the CMIP ensembles in the mean can present those fairly well, but the model ensemble spread is relatively large suggesting that many models are not



consistent with the observed values. Furthermore, as we will point out further below, the good fit in these simple statistics seem to mask bigger problems in the model dynamics of ENSO, as they result from the analysis of the ReOsc model parameters.

The CMIP model simulations present the six parameters of the ReOsc model (Eq. (1)) with different quality. All six parameters show fairly large spread within the CMIP ensemble, with some models being nearly undamped in T tendencies ($a_{11} \approx 0$) and others having more than twice as much damping than the ensemble mean. Other biases of significance are a too strong damping of h and in general too weak stochastic noise forcing on h .

More significant problems in the CMIP ensemble dynamics became apparent when we split up the growth rate (damping) of T (a_{11}) into atmospheric and oceanic feedbacks. Atmospheric feedbacks (Bjerknes wind–SST feedback, $C_{\tau T}$, and atmos. heat flux feedback, $C_{\Gamma T}$) are largely underestimated, which is consistent with previous studies (Lloyd et al. 2009; Bellenger et al. 2014) and also consistent with the idea of Dommenget et al. (2014) that CGCM simulations are often closer to a slab ocean El Nino dynamics (Dommenget 2010) than to the observed ENSO dynamics.

All three elements that make up the growth rate of T (Bjerknes wind–SST feedback, $C_{\tau T}$, atmos. heat flux feedback, $C_{\Gamma T}$, and oceanic damping of T , a_{11o}) are strongly underestimated in the CMIP ensemble, while at the same time the total of T (a_{11}) is essentially unbiased. This is achieved by strongly compensating biases: too weak atmospheric growth rates (a_{11a}) combined with too weak oceanic damping (a_{11o}).

Here it has to be noted that the oceanic damping of T (a_{11o}) was estimated as the residual of a_{11} minus a_{11a} (Eqs. (2 and 3)). A possible explanation for the strong compensation between atmospheric and oceanic feedbacks may therefore be a limitation in the ReOsc model approach used here. The assumption that a_{11a} results from Eq. (3) may be limited. However, it is very likely that oceanic processes do contribute to damping of a_{11} and it therefore seems reasonable to assume that the results presented do hold to some degree. It is beyond this study to solve this unexpected result and future study need to address this issue in more detail.

Assuming that there is indeed a compensating effect of atmospheric feedbacks (a_{11a}) and oceanic feedbacks (a_{11o}) in the CMIP ensemble, then we would need to find a reason of why that is. We are not aware of any physical mechanism that could explain such a relationship, but we cannot exclude such a possibility. For instance, Bayr et al. (2017) and partly Kim et al. (2008) argue that $C_{\tau T}$ and $C_{\Gamma T}$ are dynamically linked by the position of the Walker circulation. Whether some mechanism like this could explain the link between a_{11a} and a_{11o} is unclear. An alternative possibility is that the CGCM models are tuned to produce the observed ENSO

statistics: by doing so model developers may have, by coincident, tuned errors into the atmospheric and oceanic dynamics that compensate each other nearly completely. Thus, CMIP models may produce apparently good ENSO simulations for the wrong reasons.

The combination of errors we find in the CMIP models suggests that the relative importance of ocean dynamics for the simulated ENSO SST variability is underestimated. This results from underestimated oceanic processes such as noise forcing for h , ocean coupling of h to $T(a_{21o})$ and too strong damping of h . This is at large consistent with the findings of Kim et al. (Kim et al. 2014), who also report an underrepresentation of oceanic processes in ENSO. At the same time CMIP models underestimate atmospheric damping, which allows the models to create ENSO variability with realistic amplitudes. Thus, it appears that the relative importance of atmospheric processes in the CMIP models is larger than observed.

The ReOsc model allowed us to estimate the sensitivity of the ENSO statistics to each model parameter. It illustrated that in terms of model biases towards observed the most important parameter errors are in $C_{\tau T}$, $C_{\Gamma T}$, and a_{11o} , and to a lesser extent in the noise forcing and damping of h . In terms of CMIP ensemble spread the same parameters contribute to the ENSO statistics uncertainty, but in addition the growths rate of T (a_{11}) is important too.

Based on these sensitivities we defined a dynamical bias score that is essentially a normalized root mean squared in the model parameters. All CMIP model have fairly bad bias scores relative to what we could expect from observed uncertainties. These bad bias scores mostly result from the compensating errors in the dynamics as describes above. Even if we would exclude the less clear estimate of a_{11o} the bias scores would remain fairly bad (correlation with original bias score of Eq. (7) is still 0.85 if a_{11o} is excluded). Bias scores that are based on just the ENSO statistics (e.g. $\text{stdv}(T)$, etc.) or just the ReOsc model parameters of Eq. (1) (not shown) would be much better. However, these would exclude the most important elements of ENSO dynamics ($C_{\tau T}$, $C_{\Gamma T}$, and a_{11o}) and would therefore not present a complete picture. Bellenger et al. (2014) also found little relation between skill scores based on simple ENSO statistical properties vs. those based on dynamics properties. From the bias scores that we estimated we find a fairly clear ranking in the CMIP model performance with no substantial improvement from CMIP3 to CMIP5. The best models in this skill score are the family of models from NCAR/UCAR (CCSM4, CESM1-BGC, CESM1-CAM5, CESM1-FASTCHEM).

The results we found here have some implications for ENSO studies with CGCMs. The fact that the models produce realistic ENSO statistics with compensating errors in dynamics highlights some concerns in the models skill in predicting future SST evolutions. This is relevant not only

for seasonal to interannual forecasting, but also for long time climate change projections. Even though models in the ensemble mean project a systematic shift towards El Niño like conditions Collins et al. (2010), this projection is undermined by the CGCM having systematic biases in the dynamics. How ENSO will change in the future strongly depends on the right sensitivity of the tropical Pacific region to the different forcings (e.g. heat fluxes and winds) that drive these changes.

The recent global warming hiatus and the mismatch of the CMIP simulations in predicting tropical Pacific climate trends (e.g. in winds or the Walker Circulation) may be some further indication that the model ENSO dynamics do not have the right balance in feedbacks (England et al. 2014; Kociuba and Power 2015; McGregor et al. 2014). The framework that we have introduced here may help to address such biases. Therefore, the good news from this study is that we can expect that future CGCM simulations will improve in ENSO dynamics, if model developers can use approaches like the one we have introduced here to improve the dynamics of the model.

Acknowledgements We are grateful and thank Tobias Bayr, Shayne McGregor, Sarah Perry, Peter van Rensch and Christian Wengel for their suggestions and comments. This study was supported by the ARC project “Beyond the linear dynamics of the El Niño Southern Oscillation”, Australian Research Council (Grant no. DP120101442) and the ARC Centre of Excellence for Climate System Science, Australian Research Council (Grant no. CE110001028). The work presented here has a long history going back to 2007 with a number of people contributing to the analysis over time that we like to acknowledge here: Malte Jansen, Claudia Frauen, Simona Trefalt, Chevillard Jeanne and Payan Timothée and Yanshan Yu.

References

- Bayr T, Latif M, Dommenges D, Wengel C, Harlaß J, Park W (2017) Mean-state dependence of ENSO atmospheric feedbacks in climate models. *Clim Dyn*. <https://doi.org/10.1007/s00382-017-3799-2>
- Bellenger H, Guilyardi E, Leloup J, Lengaigne M, Vialard J (2014) ENSO representation in climate models: from CMIP3 to CMIP5. *Clim Dyn* 42:1999–2018
- Bjerknes J (1969) Atmospheric teleconnections from the equatorial Pacific. *Mon Wea Rev* 97:163–172
- Brown JN, Fedorov AV (2008) Mean energy balance in the tropical Pacific Ocean. *J Mar Res* 66:1–23
- Burgers G, Jin FF, van Oldenborgh GJ, 2005: The simplest ENSO recharge oscillator. *Geophys Res Lett* 32
- Collins M et al (2010) The impact of global warming on the tropical Pacific ocean and El Niño. *Nat Geosci* 3:391–397
- Dee DP et al (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q J R Meteorol Soc* 137:553–597
- Dommenges D (2010) The slab ocean El Niño. *Geophys Res Lett* 37
- Dommenges D, Haase S, Bayr T, Frauen C (2014) Analysis of the Slab Ocean El Niño atmospheric feedbacks in observed and simulated ENSO dynamics. *Clim Dyn* 42:3187–3205
- England MH et al (2014) Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. *Nat Clim Change* 4:222–227
- Frauen C, Dommenges D (2010) El Niño and La Niña amplitude asymmetry caused by atmospheric feedbacks. *Geophys Res Lett* 37
- Frauen C, Dommenges D (2012) Influences of the tropical Indian and Atlantic Oceans on the predictability of ENSO. *Geophys Res Lett* 39
- Guilyardi E (2006) El Niño-mean state-seasonal cycle interactions in a multi-model ensemble. *Clim Dyn* 26:329–348
- Jansen MF, Dommenges D, Keenlyside N (2009) Tropical atmosphere–ocean interactions in a conceptual framework. *J Clim* 22:550–567
- Jin FF (1997a) An equatorial ocean recharge paradigm for ENSO. 1. Conceptual model. *J Atmos Sci* 54:811–829
- Jin FF (1997b) An equatorial ocean recharge paradigm for ENSO. 2. A stripped-down coupled model. *J Atmos Sci* 54:830–847
- Jin FF, Kim ST, Bejarano L (2006) A coupled-stability index for ENSO. *Geophys Res Lett* 33
- Kim D, Kug JS, Kang IS, Jin FF, Wittenberg AT (2008) Tropical Pacific impacts of convective momentum transport in the SNU coupled GCM. *Clim Dyn* 31:213–226
- Kim ST, Cai WJ, Jin FF, Yu JY (2014) ENSO stability in coupled climate models and its association with mean state. *Clim Dyn* 42:3313–3321
- Kociuba G, Power SB (2015) Inability of CMIP5 Models to simulate recent strengthening of the Walker circulation: implications for projections. *J Clim* 28:20–35
- Latif M et al (2001) ENSIP: the El Niño simulation intercomparison project. *Clim Dyn* 18:255–276
- Levine AF, Z. MJ, McPhaden (2015) The annual cycle in ENSO growth rate as a cause of the spring predictability barrier. *Geophys Res Lett* 42:5034–5041
- Lloyd J, Guilyardi E, Weller H, Slingo J (2009) The role of atmosphere feedbacks during ENSO in the CMIP3 models. *Atmos Sci Lett* 10:170–176
- McGregor S, Timmermann A, Stuecker MF, England MH, Merrifield M, Jin FF, Chikamoto Y (2014) Recent Walker circulation strengthening and Pacific cooling amplified by Atlantic warming. *Nat Clim Change* 4:888–892
- Meehl GA et al (2007) The WCRP CMIP3 multimodel dataset—a new era in climate change research. *Bull Am Meteorol Soc* 88:1383–+
- Rayner NA et al (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J Geophys Res Atm* 108
- Reichler T, Kim J (2008) How well do coupled models simulate today’s climate? *Bull Am Meteorol Soc* 89:303
- Smith NR (1995) An improved system for tropical ocean subsurface temperature analyses. *J Atmos Ocean Tech* 12:850–870
- Suarez MJ, Schopf PS (1988) A delayed action oscillator for ENSO. *J Atmos Sci* 45:3283–3287
- Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of Cmp5 and the experiment design. *Bull Am Meteorol Soc* 93:485–498
- van Oldenborgh GJ, Philip SY, Collins M (2005) El Niño in a changing climate: a multi-model study. *Ocean Sci* 1:81–95
- Yu LS, Weller RA (2007) Objectively analyzed air–sea heat fluxes for the global ice-free oceans (1981–2005). *Bull Am Meteorol Soc* 88:527–+
- Yu YS, Dommenges D, Frauen C, Wang G, Wales S (2016) ENSO dynamics and diversity resulting from the recharge oscillator interacting with the slab ocean. *Clim Dyn* 46:1665–1682
- Zebiak SE, Cane MA (1987) A model El-Niño Southern oscillation. *Mon Weather Rev* 115:2262–2278
- Zelle H, van Oldenborgh G, Dijkstra H (2005) El Niño and greenhouse warming: results from ensemble simulations with the NCAR CCSM. *J Clim* 18:4669–4683