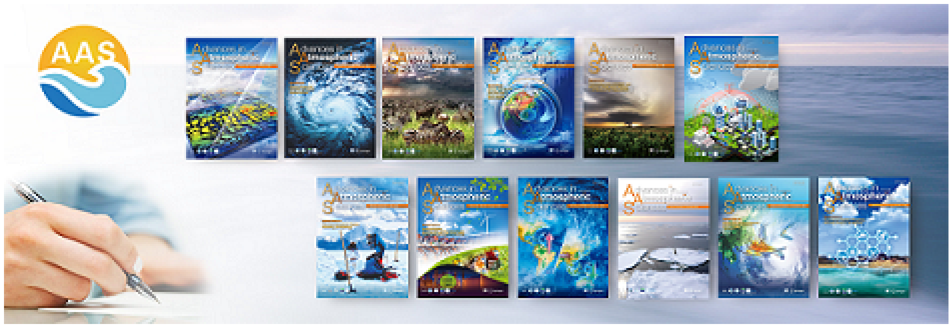

Correcting Climate Model Sea Surface Temperature Simulations with Generative Adversarial Networks: Climatology, Interannual Variability, and Extremes

Journal:	*Advances in Atmospheric Sciences*
Manuscript ID	AAS-2023-0288.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	21-Dec-2023

Complete List of Authors:	wang, Ya; Institute of Atmospheric Physics Chinese Academy of Sciences Huang, Gang; Institute of Atmospheric Physics Chinese Academy of Sciences; Qingdao National Laboratory for Marine Science and Technology; Nanjing University of Information Science & Technology Pan, Baoxiang; Chinese Academy of Sciences LIN, Pengfei; IAP, LASG Boers, Niklas tao, weichen; Institute of Atmospheric Physics Chinese Academy of Sciences Chen, Yutong; Institute of Atmospheric Physics Chinese Academy of Sciences Liu, Bo; Institute of Atmospheric Physics Chinese Academy of Sciences Li, Haijie; Chinese Academy of Sciences

Correcting Climate Model Sea Surface Temperature Simulations with Generative Adversarial Networks: Climatology, Interannual Variability, and Extremes

Ya Wang[1], Gang Huang[1, 2, 3*], Baoxiang Pan[1], Pengfei Lin[1], Niklas Boers[4, 5, 6], Weichen Tao[1], Yutong Chen[1], Bo Liu[7], Haijie Li[1]

[1]State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

[2]Laboratory for Regional Oceanography and Numerical Modeling, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

[3] College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing, China

[4] Earth System Modelling, School of Engineering and Design, Technical University of Munich, Munich, Germany

[5] Potsdam Institute for Climate Impact Research, Potsdam, Germany

[6]Department of Mathematics and Global Systems Institute, University of Exeter, Exeter, UK

[7] CMA Earth System Modeling and Prediction Centre (CEMC), Chinese Academy of Meteorological Sciences, Beijing, China

Corresponding author: Prof. Gang Huang (hg@mail.iap.ac.cn)

Abstract

Climate models are vital for understanding and projecting global climate change and associated impacts. However, these models suffer from biases that limit their accuracy in historical simulations and the trustworthiness of future projections. Addressing these challenges requires addressing internal variability, hindering direct alignment between model simulations and observations and thwarting conventional supervised learning methods. Here, we employ an unsupervised Cycle-consistent Generative Adversarial Network (CycleGAN), to correct daily Sea Surface Temperature (SST) simulations from the Community Earth System Model 2 (CESM2). Our results reveal that the CycleGAN not only corrects climatological biases but also improves the simulation of major dynamic modes including the El Niño-Southern Oscillation (ENSO) and the Indian Ocean Dipole mode, as well as SST extremes. Notably, it substantially corrects climatological SST biases, decreasing the globally averaged Root Mean Square Error (RMSE) by 58%. Intriguingly, the CycleGAN effectively addresses the well-known excessive westward bias in ENSO SST anomalies, a common issue in climate models that traditional methods, like quantile mapping, struggle to rectify. Additionally, it substantially improves the simulation of SST extremes, raising the pattern correlation coefficient (PCC) from 0.56 to 0.88 and lowering the RMSE from 0.5 to 0.32. This enhancement is attributed to better representations of interannual variability and variabilities at intraseasonal and synoptic scales. Our study offers a novel approach to correct global SST simulations, and underscores its effectiveness across different time scales and primary dynamical modes.

Keywords: Generative Adversarial Networks, Model Bias, Deep Learning, El Niño-Southern Oscillation, Marine Heatwaves.

https://doi.org/10.1007/s00376-024-3288-6

Article highlights:

- We employ an unsupervised Cycle-consistent Generative Adversarial Network (CycleGAN), to correct daily Sea Surface Temperature (SST) simulations, addressing the issue that supervised learning methods cannot be directly applied to climate model corrections.
- CycleGAN corrects climatological biases and improves the simulation of primary dynamic modes including the El Niño-Southern Oscillation (ENSO) and Indian Ocean Dipole mode, as well as SST extremes.
- CycleGAN addresses the well-known excessive westward bias into the equatorial Pacific in ENSO SST anomalies, a common bias in climate models.

1 Introduction

Climate Model simulations play an indispensable role in contemporary climate science, serving as fundamental resources for climate change detection, attribution, and projection. Their simulations, especially from the Coupled Model Intercomparison Project (CMIP), underpin the assessments conducted by the Intergovernmental Panel on Climate Change (IPCC), offering comprehensive evaluations and pivotal references that inform global climate science and guide critical decision-making (*Taylor et al.*, 2012; *Eyring et al.*, 2016). However, the utility of CMIP models in both detection and attribution of climate change, as well as climate projections, is significantly impeded by various persistent model biases (*Zheng et al.*, 2012; *Li and Xie*, 2012; 2014; *Liu et al.*, 2013; *Tao et al.*, 2018; *Jiang et al.*, 2019; *Capotondi et al.*, 2020; *Danabasoglu et al.*, 2020; *Jiang et al.*, 2021).

Among these biases, a prominent concern centers on the representation of sea surface temperatures (SSTs) in coupled general circulation models (CGCMs), especially the persistent cold tongue bias found in the equatorial Pacific (*Li and Xie*, 2012; 2014; *Zheng et al.*, 2012). This bias, characterized by a cold bias in the climatological SSTs, has long perplexed the climate community due to its potential repercussions on climatological precipitation patterns and the accurate simulation of the El Niño-Southern Oscillation (ENSO), thereby affecting the fidelity of global climate simulations (*Tao et al.*, 2018; *Jiang et al.*, 2019; *Tang et al.*, 2023). Additionally, substantial biases manifest at interannual scales, with ENSO SST variability exhibiting an excessive westward extension into the equatorial western Pacific, consequently impacting the ENSO decay and regional climate variability. These biases, commonly present in most of the CMIP models, are hardly canceled out even by multi-model means (*Huang and Ying*, 2015; *Jiang et al.*, 2021; *Tang et al.*, 2023), thus impacting historical simulations and future projections.

Efforts to rectify these climate model biases have spawned the development of various methodologies. Traditional methods predominantly center around the local adjustment of specific statistical characteristics. These adjustments typically target parameters like the mean and variance (involving variance adjustment; *Teutschbein and Seibert*, 2012; *Chen et al.*, 2011; *Chen et al.*, 2013; *Li et al.*, 2019), and frequency distributions (via techniques like quantile mapping; QM; *Themessl et al.*, 2011). However, these approaches fall short of providing precise corrections for model errors at the daily timescale and in dynamical climate variabilities. Traditional methods primarily focus on specific statistical features or target adjustments at individual grid points and can by design not correct spatial patterns (*Hess et al.*, 2022; 2023).

The challenge of climate model bias correction is exacerbated by the internal variability of the Earth system, characterized by inherent nonlinearity (*Hawkins and Sutton*, 2011; *Deser et al.*, 2012; *Deser et al.*, 2016; *Hu et al.*, 2018; *Wang et al.*, 2020). In particular, discrepancies arise due to internal variability in model simulations and observations at corresponding times. Consequently, discerning whether disparities between model fields and observational datasets arise from internal variability or genuine model biases poses a substantial challenge; identifying and on that basis correcting the bias in a particular day in model data is inherently difficult. This presents a hurdle when trying to employ traditional statistical methods and supervised learning techniques to correct model simulations.

The advent of deep learning, particularly the development of unsupervised and semi-supervised learning techniques, has introduced novel avenues for addressing climate model biases. Generative Adversarial Networks (GANs), pioneered by *Goodfellow et al.* (2014), leverage a generator-discriminator framework to generate images that are in their characteristics and statistical properties virtually indistinguishable from target images. Subsequently, *Hoffman et al.* (2017), *Yi et al.* (2017), and *Zhu et al.* (2017) expanded upon the GAN methodology, introducing the Cycle-consistent GAN (CycleGAN), which incorporates a cycle-consistent loss to facilitate the bidirectional transformation of data styles. *Pan et al.* (2021) have explored the application of this method to rectify precipitation data in CGCM simulations, yielding promising results in correcting precipitation biases in the United States. Moreover, the extension of this method to global precipitation simulations demonstrated its efficacy in correcting spatial patterns and in ameliorating the Pacific double Intertropical Convergence Zone (ITCZ) bias (*Hess et al.*, 2022; 2023), a prevalent issue across most CGCMs. These applications have predominantly concentrated on evaluating the enhancement of fundamental statistical aspects of precipitation.

The critical inquiry here is whether this method can also bias-correct modelled SST fields. Furthermore, can it correct the dynamical oceanic modes? And can it adapt to variations driven by distinct physical mechanisms at various time scales? These questions presently lack definitive answers.

In this paper, we employ a CycleGAN-based approach to correct global daily SST fields to answer these questions. We find that our method not only addresses biases at climatological averages but also tackles errors associated with ENSO, the Indian Ocean Dipole (IOD), and SST extremes. This paper is structured as follows: Section 2 provides introductions of datasets and methodology employed, Section 3 unveils the primary results, and Section 4 discusses and summarizes our findings.

2 Data and method

2.1 Data

We employ daily surface temperature (ST) from the National Center for Environmental Prediction-Department of Energy Atmospheric Reanalysis (NCEP; *Kanamitsu et al. 2002*) at a resolution of 2.5°× 2.5° for the period 1950 to 2014. Modelled daily ST fields from 1950 to 2014 are obtained from the Community Earth System Model 2 (CESM2; *Danabasoglu et al.*, 2020), a widely used and well-organized model, which can be considered a state-of-the-art climate model. The global gridded monthly SST datasets from Extended Reconstructed Sea Surface Temperature (ERSST.v5; *Huang et al. 2017*), the NOAA 1/4° Daily Optimum Interpolation SST (OISST; *Huang et al.*, 2021), and the Hadley Centre Global Sea Ice and SST (HadISST) are utilized. These three SST datasets and NCEP ST, ranging from 1991~2014, are utilized for model testing. All datasets are interpolated to a resolution of 2.5°× 2.5°.

To extract ENSO-related SST variability, we utilize the unstandardized Niño-3.4 index. This index is defined as the SST anomalies (SSTA) averaged over the region between 5°S~5°N, and 120°W~170°W from December to February (DJF). We performed a regression analysis of this unstandardized Niño-3.4 index on the interannual SSTA, thus revealing the underlying ENSO-driven SST variability. The IOD SST is obtained in a similar way as that in ENSO SST but uses the Dipole Mode Index defined as the anomalous SST gradient between the western equatorial Indian Ocean (50°E-70°E and 10°S-10°N) and the southeastern equatorial Indian Ocean (90°E-110°E and 10°S-0°N) during September-October-November (SON).

2.2 Methods

CycleGAN is a groundbreaking deep learning model designed for unpaired image-to-image translation. It enables the transformation of images from one domain to another even when paired training data is unavailable (Zhu et al. 2017), such as in the case tackled in the present study. Unpaired image-to-image translation involves the mapping of images from one domain (the source domain, denoted by X) to another (the target domain, denoted by Y) without requiring a direct one-to-one correspondence between the images in the two domains. In our case, we apply the principles of unpaired image-to-image translation for the correction of CGCM daily SST data. Specifically, we consider CESM2 SSTs as domain X and NCEP STs as domain Y. The training and validation data comprises CESM2 and NCEP ST data from 1950 to 1990, while the testing data spans from 1991 to 2014. A normalization process is employed to scale the data.

Fig. 1 gives the schematic of the CycleGAN. Within the CESM2, biases exist in the internal variability and external forced responses within model simulations on any given day. However, the internal variability in observations on the same day may not correspond to that in the model. For example, the model might simulate an El Niño year, while observations indicate a La Niña year. Therefore, directly adjusting the model to match the observational state is evidently inaccurate. Our goal is to eliminate biases in internal variability and external forced responses while simultaneously preserving the model's internal variability (e.g., the ENSO phase).

To achieve this objective, a generator (G(x)) and a discriminator are first employed in this model (Fig. 1a). The primary function of the generator is to transform CESM2 data to observational SST, while the discriminator assesses whether the generated outcomes originate from observations. A balance between these two components yields the best results, wherein the discriminator should be incapable of distinguishing between images generated by the generator and observations. This phase aids in correcting the GCM simulations. Subsequently, a secondary generator (F(y)) and discriminator are utilized to retain internal variability. The secondary generator, F(y), is tasked with reconverting the previously generated images back into the original CESM2 data. This process ensures that the generated images by G(x) retain the essential characteristics of CESM2. From a mathematical perspective, the architecture of F(y) ensures a bidirectional mapping between CESM2 and the observational data. In a certain context, this secondary generator facilitates the preservation of internal variability within CESM2, as internal variability is a predominant aspect on a daily scale.

The loss function of this model is as follows:

$$\begin{aligned}\mathcal{L}(G,F,D_X,D_Y) = \mathcal{L}_{\mathrm{GAN}}(G,D_Y,X,Y) + \mathcal{L}_{\mathrm{GAN}}(F,D_X,Y,X)\\ +\lambda_c\mathcal{L}_{\mathrm{cyc}}(G,F) + \lambda_i\mathcal{L}_{\mathrm{id}}(G,F)\end{aligned} \tag{1}$$

The loss includes the adversarial loss of X -> Y ($\mathcal{L}_{\mathrm{GAN}}(G,D_Y,X,Y)$), Y->X ($\mathcal{L}_{\mathrm{GAN}}(F,D_X,Y,X)$), the consistency loss ($\mathcal{L}_{\mathrm{cyc}}(G,F)$), and the identity loss ($\mathcal{L}_{\mathrm{id}}(G,F)$). G denotes a generator that tries to generate images G(x) that are indistinguishable from those in Y, F represents a generator that tries to generate images F(y) similar to those in X. D_Y and D_X denote the discriminators aiming to distinguish between G(x) and y, and F(y) and x, respectively.

$$\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y\sim p_{\text{data}}(y)}[\log D_Y(y)] + \mathbb{E}_{x\sim p_{\text{data}}(x)}\big[\log\big(1 - D_Y\big(G(x)\big)\big] \quad (2)$$

$$\mathcal{L}_{\text{GAN}}(F, D_X, Y, X) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D_x(X)] + \mathbb{E}_{y\sim p_{\text{data}}(y)}\big[\log\big(1 - D_X\big(F(y)\big)\big] \quad (3)$$

Moreover, as by *Zhu et al.* (2017), the consistency loss and identity loss are further added to help retain the internal variability and avoid introducing additional bias, the weights of these terms are controlled by λ_c and λ_i, which are set to 10 and 0.5, respectively.

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x\sim p_{\text{data}}(x)}\big[\| F\big(G(x)\big) - x \|_1\big] + \mathbb{E}_{y\sim p_{\text{data}}(y)}\big[\| G\big(F(y)\big) - y \|_1\big] \quad (4)$$

$$\mathcal{L}_{\text{identity}}(G, F) = \mathbb{E}_{y\sim p_{\text{data}}(y)}\,[\| G(y) - y \|_1] + \mathbb{E}_{x\sim p_{\text{data}}(x)}[\| F(x) - x \|_1] \quad (5)$$

The generative networks consist of three components: down-sampling with two convolutional layers, followed by nine residual blocks comprising a total of 21 convolutional layers, and finally two up-sampling layers (as depicted in Fig. 1). The first residual block, represented in Fig. 1c, contains five convolutional layers, while the subsequent blocks, as shown in Fig. 1d, each consist of two convolutional layers. This deep generative network helps generate finer SST images. The architecture of the discriminator aligns with the model presented by Zhu et al. (2017). Gradient clipping is employed to optimize the training of the model. In training GAN, the Wasserstein Generative Adversarial Network is utilized, chosen for its known stability (*Arjovsky et al., 2017*).

In the following we employ linear regression and composite analysis, and the statistical significance is determined using the two-tailed Student's t-test.

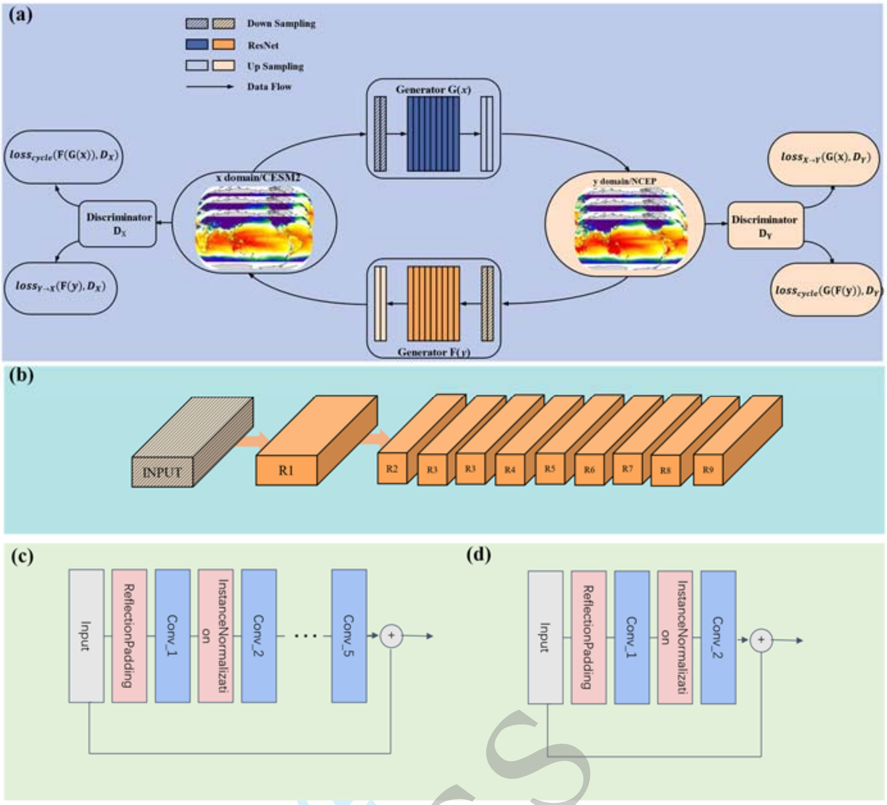

Figure 1 (a) Schematic of the CycleGAN. (b) The details of ResNet, which has 9 residual blocks. (c) The details of the first residual block, and (d) other blocks.

3 Results

Figure 2 presents a qualitative comparison of daily ST on the same date (3th November 2004). CycleGAN essentially preserves most features present in CESM2 results, notably the La Niña-like SSTA (Figs. 2b and 2c), while also exhibiting a greater resemblance to observations (Figs. 2a and 2c). This improvement is particularly notable in regions such as the tropical Atlantic, South Africa, and western South America. This comparison effectively signifies that the data post-processed by CycleGAN, in addition to correcting model outcomes, preserves the internal variability of the model.

In the subsequent analysis, we evaluate the correction from three key perspectives: climatology, interannual variability, and extreme events. To validate the effectiveness of the correction, we performed comparisons between results obtained from CESM2, GAN-corrected SSTs, and four reference datasets (HadISST, ERSST, NCEP, OISST).

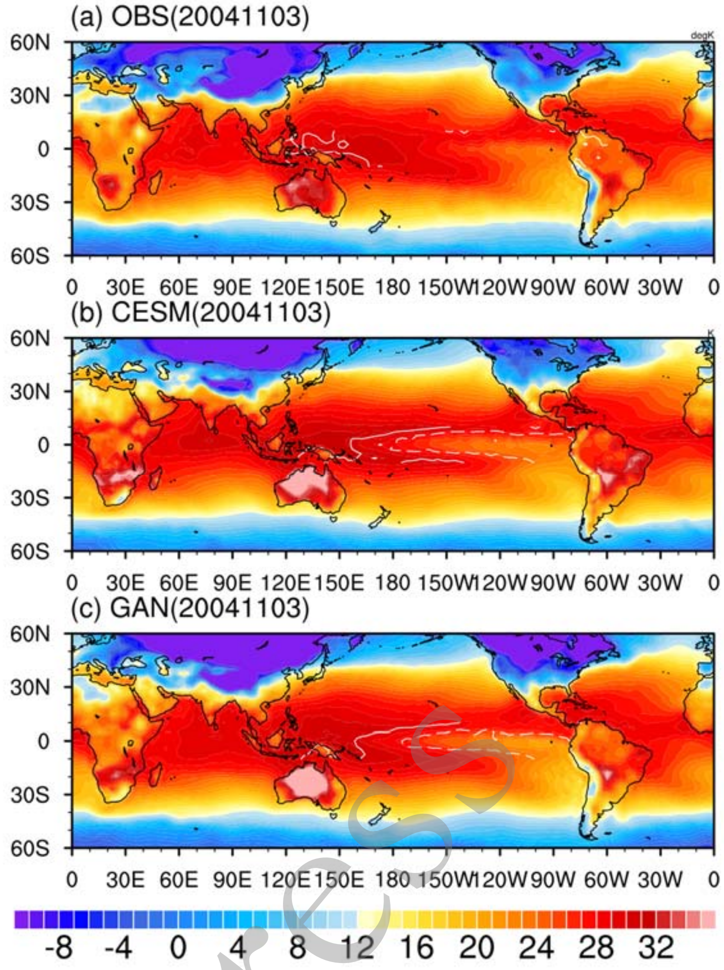

Figure 2 Qualitative comparison of the intermittency in daily ST on the same date (3th November 2004), for the (a) observation, (b) CESM2 model and (c) CycleGAN-based post-processing. White contour lines (with interval of 1K) denote monthly SSTA in November 2004.

3.1 The climatological bias

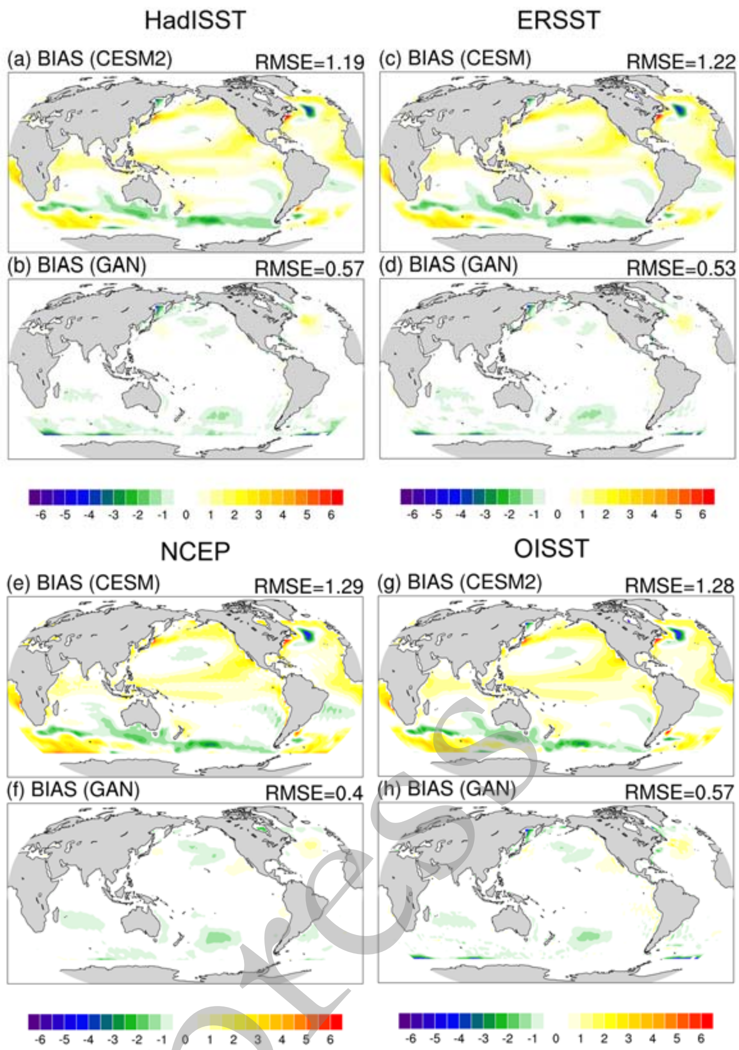

Figure 3 The disparities between climatological annual mean SSTs in (a, c, e, g) CESM2 simulations, (b, d, f, h) GAN-corrected SST, and (a, b) HadISST, (c, d) ERSST, (e, f) NCEP, (g, h) OISST.

Figure 3 illustrates the disparities between CESM2 model data, GAN-corrected outcomes, and observational datasets. Overall, CESM2 displays a substantial warming bias when compared to the four SST datasets. There is a distinct dipole bias in SSTs in the North Atlantic, particularly along the Gulf Stream and North Atlantic Current, where warm and cold biases coexist in proximity. A similar dipole bias is observed in the Southern Ocean at mid to high latitudes. Notably, prominent warm biases are noticeable near eastern boundary upwelling regions, encompassing areas close to California, South America, and Africa (Figs. 3a, 3c, 3e, and 3g). These biases are likely associated with discrepancies in wind stress patterns within these regions and the spatial resolution of the model (*Capotondi et al.*, 2020). Additionally, a significant warm bias is evident in the tropical Pacific. This warm bias is distinct from the well-known cold tongue bias observed in most CMIP5 models and even differs from CESM1, signifying disparities between CESM2 and its predecessors, as also acknowledged in previous CESM2 assessment studies (e.g., *Capotondi et al.*, 2020).

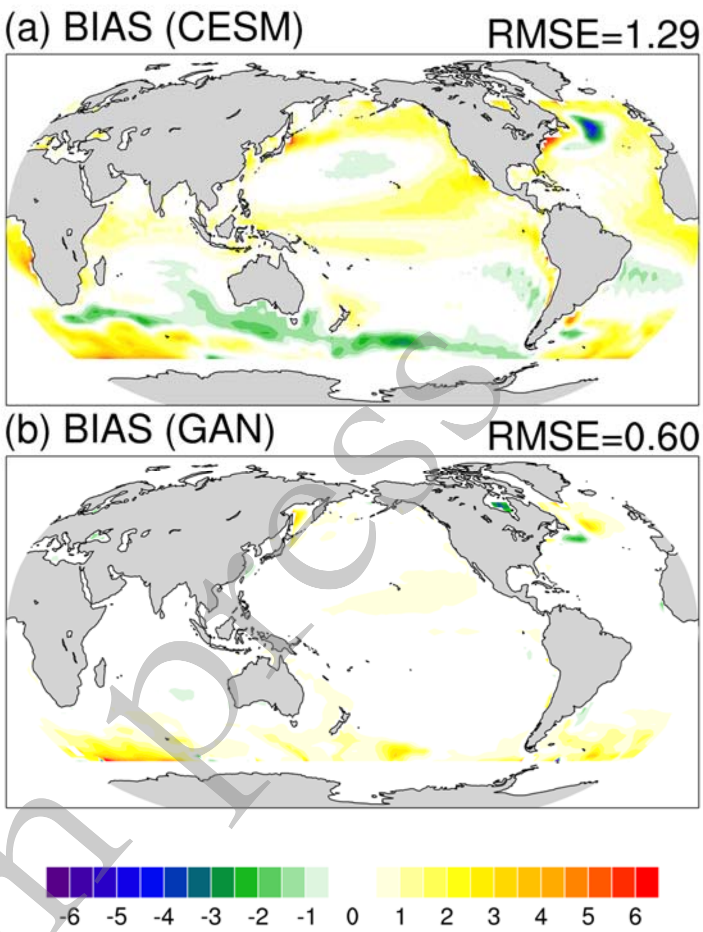

Figure 4 The disparities between climatological annual mean SSTs in (a) CESM2 simulations, (b) QM-corrected SST, and NCEP.

After the correction by the CycleGAN, there is a substantial reduction in the bias in the climatological SSTs (Figs. 2b, 2d, 2f, and 2h). The dipole bias in the North Atlantic and the Southern Hemisphere are notably reduced. Only minor cold biases persist in the vicinity of eastern Africa and the east of New Zealand. Furthermore, warm biases in the tropical Pacific are significantly attenuated. The overall bias diminishes from 1.25°C (1.19°C to 1.28°C) in CESM2 to 0.52°C (0.4°C 0.57°C). Through a comparative analysis involving four sets of SST data, we find that the biases in both magnitude and spatial pattern are highly consistent across the different datasets. This robust agreement underscores the reliability of our evaluation. In addition, CycleGAN exhibits clear advantages over traditional methods. We compare its performance with the widely utilized approach, modified QM (Themeßl et al., 2011; Bai et al., 2016), widely employed for the correction of historical simulations and future projections within CMIP. When applied to SST correction, modified QM yields an error of 0.6, which is inferior to the results obtained by CycleGAN (Fig. 4). It is noteworthy that the performance of QM and CycleGAN is remarkably close in tropical regions, and in the North Pacific, QM even outperforms the latter. However, QM exhibits larger errors in the high-latitude regions of the Southern Ocean.

The errors calculated based on the NCEP data align closely with those from other datasets, further affirming the suitability of this dataset for model training. Given the substantial convergence in results across three SST datasets (HadISST, ERSST, OISST), for the sake of brevity, we will exclusively employ HadISST as the observational reference in the subsequent sections.

3.2 The bias in interannual variability

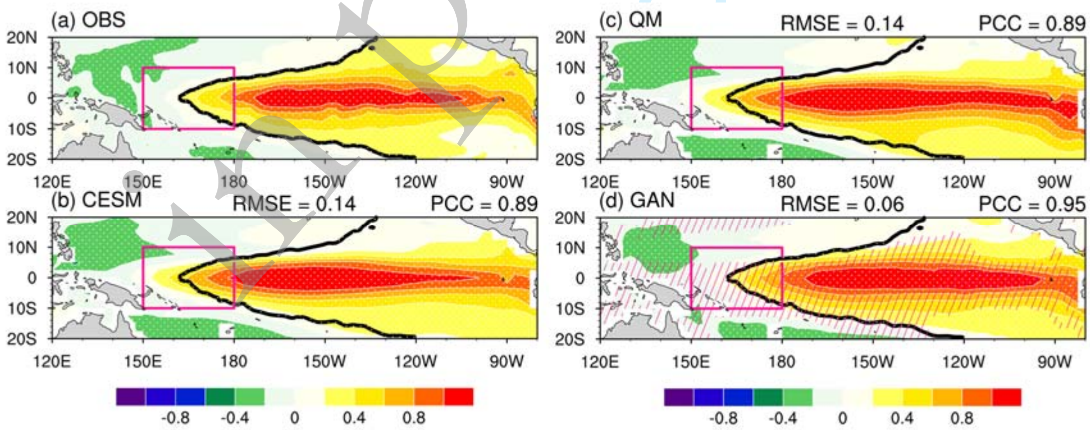

Figure 5 ENSO-related SSTA in (a) observation, (b) CESM2, (c) QM, and (d) CycleGAN. Stippling indicates where the regressions are significant at the 95% confidence level, based on the student's t-test. The hatched areas represent significant differences in El Niño composite SSTs within the CycleGAN-corrected fields compared to CESM2. The black contour line denotes +0.2K in observations.

In addition to addressing biases at the climatological scale, an essential aspect of climate model correction, and perhaps even more crucial, is the capability to simulate interannual dynamical variability such as ENSO and IOD. The accurate representation of these variabilities directly influences the ability to capture global and regional climate variability. Figure 5 displays DJF ENSO SSTA, in observations, CESM2, and CycleGAN-corrected SSTs. Notably, CESM2 exhibits an excessive westward bias in the equatorial Pacific, with a pronounced warm bias in the equatorial western Pacific. This bias is well-known and prevalent in many models, constituting a common bias in most CMIP5 and CMIP6 models (*Tao et al.*, 2015; *Tao et al.*, 2018; *Tao et al.*, 2019; *Jiang et al.*, 2021). It is worth noting that this substantial warm bias in ENSO SSTA can be primarily attributed to the climatological cold tongue bias (*Li and Xie*, 2012; 2014; *Jiang et al.*, 2021). The cold tongue bias manifests as a phenomenon in the model simulation, signifying that the climatological annual mean Sea Surface Temperature (SST) in the central-eastern tropical Pacific is colder than observed. This is characterized by an excessively strong and westward-extending cold tongue in the equatorial Pacific. The presence of the cold tongue bias can influence SSTA in this region through its impact on temperature advection and other related processes. Despite the warm bias in the western Pacific in CESM2, its climatology also exhibits a warm bias. This suggests that the warm bias in the western Pacific of CESM2 may be attributed to other mechanisms.

The modified QM method exhibits limited efficacy in the correction of ENSO SST. Its RMSE and PCC closely align with CESM2, and is ineffective in addressing distributional biases, such as excessive westward bias (Fig. 5c). After correction by the CycleGAN, in the equatorial western Pacific, the warm bias is greatly diminished, with its intensity comparable to observations. Furthermore, the warm bias in the central and eastern Pacific, as observed in CESM2, is also notably reduced. A significance test was performed for composite El Niño SSTs, revealing significant differences between CESM2 and the CycleGAN-corrected results in several regions, including the equatorial western Pacific, the equatorial sides of the central Pacific, and the tropical eastern Pacific (Fig. 5d). This indicates a statistically reliable improvement by the

CycleGAN in addressing warm biases in CESM2. Overall, upon applying the correction, we observe a significant reduction in the excessive westward bias of ENSO SST variability, with the distribution closely resembling observations. The RMSE decreases from 0.14 in CESM2 to 0.06, and the PCC increases from 0.89 to 0.95, corresponding to a 57% reduction and a 6.7% improvement, respectively.

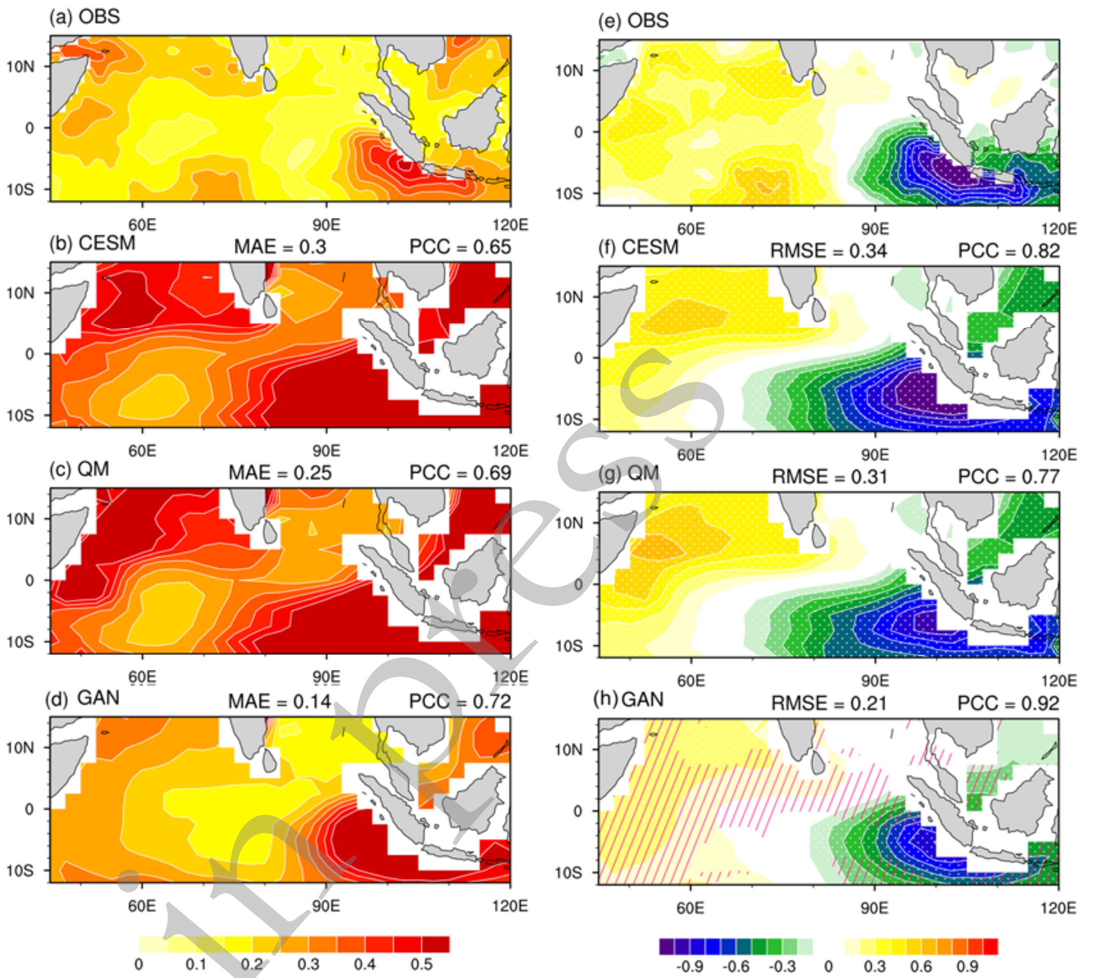

Figure 6 (Left) The standard deviation of SST and (b) the IOD-related SSTA during SON in (a, e) observation, (b, f) CESM2, (c, g) QM, and (d, h) CycleGAN, respectively. Stippling indicates the regressions are significant at the 95% confidence level, based on the student's t-test. The hatched areas represent significant differences in positive IOD composite SSTs within the CycleGAN compared to CESM2.

Figure 6 shows the standard deviation of SSTs in the Indian Ocean and the IOD SSTA. In the CESM2 simulation, there is a significant overestimation of variability in the southeastern and western equatorial Indian Ocean, with the center of variability extending too far westward in the southeastern Indian Ocean when compared to observations. Similar to its limited efficacy in ENSO correction, the modified QM method demonstrates constrained effectiveness in the correction of SST variability. While the CycleGAN markedly reduces the bias in the southeastern and western equatorial Indian Ocean compared to CESM2, and the excessive westward extension is substantially attenuated. The IOD mode exhibits bias like the standard deviation of SSTs (Fig. 6f). In CESM2, the IOD SSTA in the southeastern Indian Ocean is significantly stronger and extends too far westward compared to observations. In observations, negative anomalies roughly extend to 85°E, while in CESM2, the negative anomaly extends to around 70°E. CycleGAN successfully reduces this bias, with the negative anomaly located east of 80°E, significantly closer to observations, the performance is noticeably superior to modified QM, with the latter making only marginal adjustments to the intensity of SSTA. The RMSE decreases substantially compared to CESM2, from 0.34 to 0.21, while the PCC increases from 0.82 to 0.92.

3.3 The bias in SST extremes

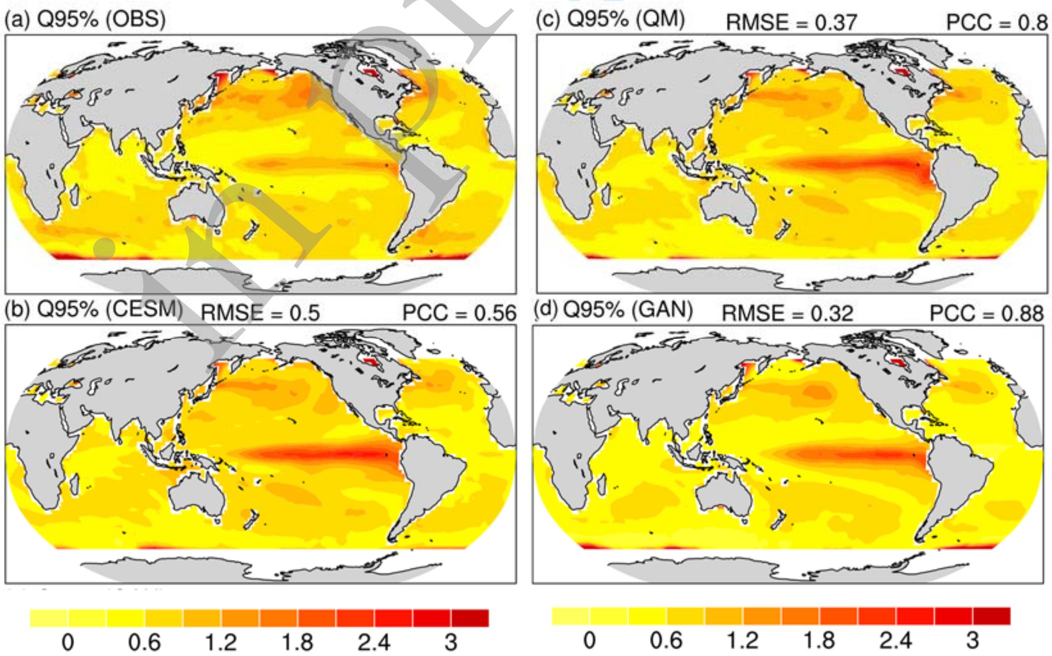

Figure 7 As Fig. 5, but for the 95th percentile of SSTA.

In addition to interannual dynamical variability, a recent focus in oceanic research is marine heatwaves. Frequent marine heatwaves have significant impacts on marine ecosystems, particularly on marine fisheries and coral reefs (*Oliver et al.*, 2017; *Oliver et al.*, 2018; *Holbrook et al.*, 2019; *Oliver et al.*, 2021; *Liu et al.*, 2022; *Liu et al.*, 2023). We primarily use the 95th percentile of daily SSTA to quantify SST extremes and evaluate the performance of CESM2 and the GAN-correction in representing them. To conduct the evaluation, we utilized NCEP data spanning from 1991 to 2014 for testing. Figure 7 illustrates the 95th percentile of NCEP, CESM2, and CycleGAN-corrected SSTA, respectively. In the NCEP data, regions with high values of the 95th percentile are concentrated in the western boundary current extension regions, the central and eastern equatorial Pacific, and the northeast Pacific Ocean, consistent with previous studies (*Chen et al.*, 2014; *Echevin et al.*, 2018; *Oliver et al.*, 2017; *Oliver et al.*, 2021). When compared to NCEP, the spatial distribution of the 95th percentile in CESM2 is generally similar. However, CESM2 significantly overestimates the intensity of extreme SSTA in the central and eastern tropical Pacific while underestimating it in the northwest Pacific. In contrast, CycleGAN-corrected extreme SSTA shows better agreement with observations in terms of intensity and spatial distribution. The PCC improves from 0.56 to 0.88, and the RMSE decreases from 0.5 to 0.32, slightly outperforming those obtained with modified QM.

Particularly noteworthy is the significant reduction of the overestimation bias in the central and eastern tropical Pacific and the enhancement in the underestimation in the northeast Pacific, making CycleGAN's results more consistent with NCEP. Figures 8 and 9 present probability density functions (PDFs) of SSTA in the tropical (-5°S-5°N) and northeast (40°N~60°N; 180~120°W) Pacific, respectively. On average, CESM2 exhibits a wider distribution in the tropics, indicating a stronger variability. CESM2 has a 95th percentile of 1.41 degrees for the tropical region, while NCEP only has 1.18 degrees, resulting in a 19.5% overestimation. This aligns with the earlier result of the overestimation of SSTA in the tropical region, especially in the central and eastern Pacific (Fig. 7). After correction, the distribution of SSTA in the tropics closely matches observations, with 95th percentiles of approximately 1.17, indicating a substantial reduction in the overestimation bias of CESM2 in the tropics. A similar situation is observed in the northeast Pacific. In this area, SSTA distribution in CESM2 is overall narrower than observations, indicating a lower variability. Correspondingly, its 95th percentile is lower than that of NCEP. After correction, the distribution and the 95th percentile of SSTA are much closer to NCEP.

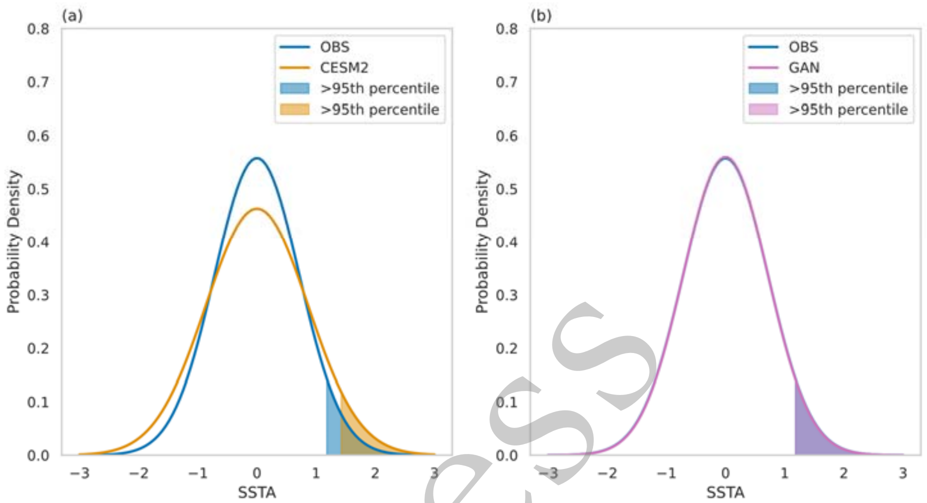

Figure 8 PDFs of average tropical SSTA for (a) NCEP and CESM2, and (b) NCEP and CycleGAN. Shading indicates areas where the SSTA is above the 90th percentile.

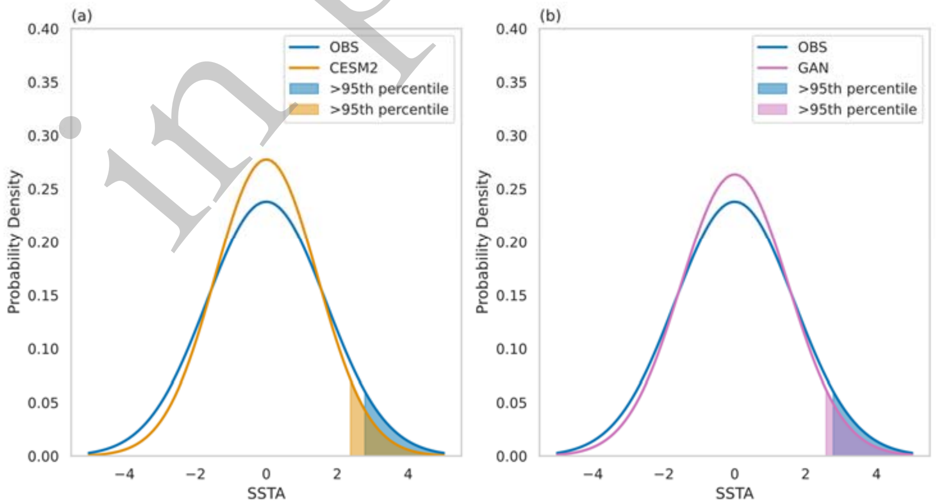

Figure 9 As Fig. 8, but for the Northeast Pacific SSTA.

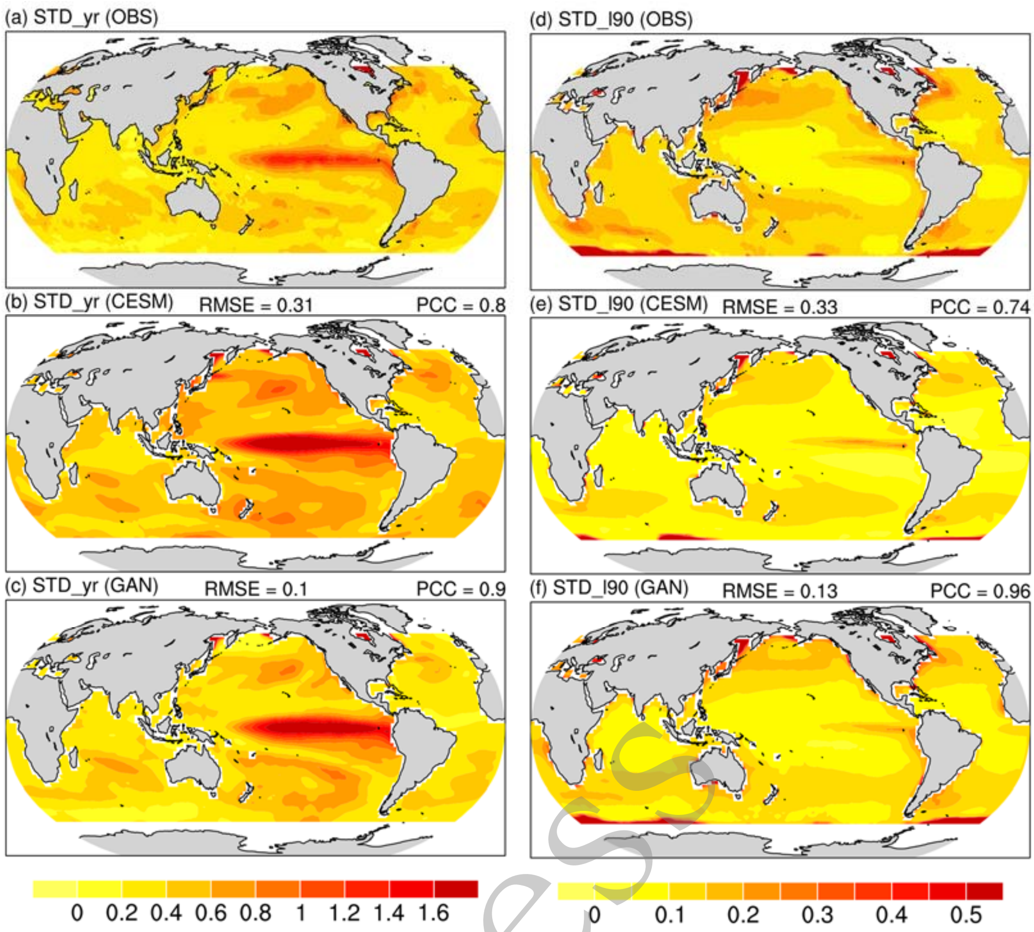

Figure 10 Standard deviations of (left) interannual DJF SSTA and (right) SSTA filtered by a band of 2–90 days in (a, d) NCEP, (b, e) CESM2, and (c, f) CycleGAN, respectively.

Unlike methods such as modified QM that directly adjust data by quantiles to make the local distribution closer to observations, the CycleGAN learns and transforms high-dimensional data features. An important question naturally arises: which key processes does CycleGAN capture that result in the improvement in simulating SST extremes? To address this question, we further performed an analysis of variability at various time scales, involving interannual variability and frequency exceeding 90 day^{-1} (encompass Madden-Julian Oscillation (MJO), Intraseasonal Oscillation (ISO), and synoptic-scale eddies). Figure 10a presents the interannual standard deviation of DJF SSTs based on observations. It primarily exhibits an ENSO-like pattern in the equatorial central and eastern Pacific. Moreover, notable high-variability regions are evident in the Northeast Pacific, North Pacific, mid-latitudes in the South Pacific, and adjacent to western boundary currents. In general, the spatial distribution of SST interannual variability in observations closely resembles the distribution of the 95th percentile of SSTA (shown in Fig. 7), indicating the direct influence of interannual variability, especially ENSO, on extreme SSTs as in previous studies (*Doi et al.*, 2015; *Oliver et al.*, 2021). Additionally, ENSO can induce SST variations in the Northeast Pacific and North Atlantic regions by exciting Rossby wave trains (*Wang et al.*, 2021; 2023; *Wang et al.*, 2022), further influencing extreme events in those areas (*Johnson and Kosaka*, 2016; *Trenberth et al.*, 1998). When compared to observations, CESM2 exhibits an overall overestimation of interannual variability, particularly pronounced in the central-eastern Pacific. This overestimation is likely related to the higher ENSO variability within CESM2 (*Capotondi et al.*, 2020). The consistency of the stronger interannual variability in the tropical central-eastern Pacific and the larger bias of extreme SSTA in Fig. 7 suggest that the overestimation of interannual variability is a key factor contributing to the high bias in extremes in CESM2. In contrast to CESM2, SST fields corrected by the CycleGAN show a substantial reduction in the error in interannual variability, with the RMSE decreasing from 0.31 to 0.1. CycleGAN-corrected results exhibit a closer match to observations in terms of intensity, particularly in the tropical Pacific, South Pacific, Atlantic, and Indian Ocean. While the intensity in the tropical central-eastern Pacific remains higher than observed, notable improvements are observed compared to CESM2, consistent with the enhancement seen in extreme SSTA for this region in Fig. 7. Concerning intraseasonal and synoptic-scale variability, the variability in NCEP is predominantly concentrated in the tropical Pacific and western boundary current regions (Fig. 10d), likely associated with boundary currents and active mesoscale eddy activity (*Oliver et al.*, 2021). CESM2 simulations generally underestimate variability at these scales (Fig. 10e). In contrast, CycleGAN-corrected SSTs exhibit substantial improvements, especially in the Northeast Pacific, where the distribution is in closer agreement with NCEP (Fig. 10f). In summary, the CycleGAN demonstrates significant improvements in SST variability at different time scales, facilitating more accurate simulation of the complex oceanic dynamic processes, particularly regarding interannual variability and variability at intraseasonal and synoptic scales.

4 Conclusion and Discussion

In this study, we employed CycleGAN to correct the daily SSTs from CESM2 historical simulations. We conducted a comprehensive assessment of this model, considering various aspects such as climatology, interannual variability, and extremes. Our findings reveal significant improvements across these evaluation dimensions. At the climatological scale, the CycleGAN substantially reduced bias in annual mean climatological SST. Specifically, there is a 58% reduction in RMSE relative to CESM2, from 1.25 (1.19~1.28) degrees in CESM2 to 0.52 (0.4~0.57) degrees. At the interannual scale, involving two primary tropical modes, ENSO and IOD, we observe significant enhancements in simulating these modes with the CycleGAN. For ENSO SSTA, the RMSE decreases from 0.14 to 0.06, corresponding to a 57% reduction. CycleGAN effectively addresses a common bias of ENSO SST found in many climate models, known as the excessive westward bias in the equatorial Pacific that traditional methods, like quantile mapping, struggle to rectify. In the case of IOD, CESM2 tends to produce excessively strong and westward-extending anomalies in the southeastern Indian Ocean. After the correction, these biases are substantially reduced, resulting in an increased PCC of up to 0.92.

Moreover, we investigate the performance in simulating SST extremes. The CycleGAN corrects the overestimation in extremes in most regions and addresses the underestimation in the Northeast Pacific. The improved performance of the CycleGAN in simulating the distribution of SSTA and extremes can be attributed to its ability to capture different temporal scales of variability, including interannual variability and variability at intraseasonal and synoptic scales, encompassing periods shorter than 90 days.

In summary, our study demonstrates that the CycleGAN offers comprehensive enhancements across various time scales and physical processes. Its utility extends beyond merely correcting specific statistical measures such as first and second-order moments locally; it also enhances the simulation of critical air-sea coupling modes like ENSO and IOD. These findings underscore the significant potential of CycleGAN as a valuable tool for climate model correction and climate projection.

We employ NCEP data for the evaluation of SST extremes. It is crucial to highlight the potential inconsistency among various observational datasets at the daily time scale, particularly in oceanic data. Consequently, it is essential to clarify that the presented results regarding SST extremes should be interpreted as an assessment of the capacity of CycleGAN to enhance the simulation of SST extremes rather than its precision in replicating observations. Achieving the latter would demand a more extensive dataset, particularly observational station data.

Moreover, it is crucial to acknowledge that, beyond the correction of historical simulations, the central consideration is the adaptability of these methods to future projections. Recent studies, exemplified by *Hess et al.* (2022), suggest that CycleGAN, through post-processing or incorporating constraints like global mean values, can reproduce trend signals under global warming. Further research is needed to refine the capability in capturing regional-scale warming responses. Subsequent investigations should also involve the development of correction datasets encompassing multiple models and variables, accompanied by a comprehensive analysis of physical mechanisms and laws in the corrected data. Comparisons with methods integrated with climate models, such as surface flux adjustments, should be explored. Furthermore, the direct coupling of the proposed correction method with climate models represents a critical avenue for future inquiry.

Acknowledgement:

The study is jointly supported by the CTG Colombia holding S.A.S., National Natural Science Foundation of China (41831175, 42141019, 41721004), the Second Tibetan Plateau Scientific Expedition and Research (STEP) program (Grant 2019QZKK0102), the National Natural Science Foundation of China (42175049 and 42175040) and Youth Innovation Promotion Association of CAS (2021072). W.Y. would like to thank the National Key Scientific and Technological Infrastructure project "Earth System Science Numerical Simulator Facility" (EarthLab) and Beijing Paratera Technology Co., Ltd. for computational support and Prof. Xiang Shiming for many useful discussions. NB acknowledges funding from the Volkswagen foundation. The code of CycleGAN could be found in https://github.com/junyanz/CycleGAN.

Reference

Bai K, Chang N B, Yu H, et al (2016), Statistical bias correction for creating coherent total ozone record from OMI and OMPS observations, *Remote Sensing of Environment*, 2016, 182:150-168.DOI:10.1016/j.rse.2016.05.007

Capotondi, A., C. Deser, A. S. Phillips, Y. Okumura, and S. M. Larson (2020), ENSO and Pacific Decadal Variability in the Community Earth System Model Version 2, *Journal of Advances in Modeling Earth Systems*, *12*(12).

Chen, J., F. P. Brissette, and R. Leconte (2011), Uncertainty of downscaling method in quantifying the impact of climate change on hydrology, *J Hydrol*, *401*(3-4), 190-202.

Chen, J., F. P. Brissette, D. Chaumont, and M. Braun (2013), Performance and uncertainty evaluation of empirical downscaling methods in quantifying the climate change impacts on hydrology over two North American river basins, *J Hydrol*, *479*, 200-214.

Chen, K., G. G. Gawarkiewicz, S. J. Lentz, and J. M. Bane (2014), Diagnosing the warming of the Northeastern US Coastal Ocean in 2012: A linkage between the atmospheric jet stream variability and ocean response, *J Geophys Res-Oceans*, *119*(1), 218-227.

Danabasoglu, G., et al. (2020), The Community Earth System Model Version 2 (CESM2), *Journal of Advances in Modeling Earth Systems*, *12*(2).

Deser, C., L. Terray, and A. S. Phillips (2016), Forced and Internal Components of Winter Air Temperature Trends over North America during the Past 50 Years: Mechanisms and Implications, *Journal of Climate*, *29*(6), 160107131720009.

Deser, C., A. Phillips, V. Bourdette, and H. Y. Teng (2012), Uncertainty in climate change projections: the role of internal variability, *Clim. Dyn.*, *38*(3-4), 527-546.

Doi, T., C. X. Yuan, S. K. Behera, and T. Yamagata (2015), Predictability of the California Nino/Nina*, *Journal of Climate*, *28*(18), 7237-7249.

Echevin, V., F. Colas, D. Espinoza-Morriberon, L. Vasquez, T. Anculle, and D. Gutierrez (2018), Forcings and Evolution of the 2017 Coastal El Nino Off Northern Peru and Ecuador, *Front Mar Sci*, *5*.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor (2016), Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geosci Model Dev*, *9*(5), 1937-1958.

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio (2014), Generative Adversarial Nets, paper presented at Neural Information Processing Systems.

Hawkins, E., and R. Sutton (2011), The potential to narrow uncertainty in projections of regional precipitation change, *Clim. Dyn.*, *37*(1), 407-418.

Hess, P., M. Drüke, S. Petri, F. M. Strnad, and N. Boers (2022), Physically constrained generative adversarial networks for improving precipitation fields from Earth system models, *Nat Mach Intell*, *4*(10), 828-+.

Hess, P., S. Lange, C. Schötz, N. Boers (2023), Deep learning for bias-correcting

CMIP6-class Earth system models, *Earth's Future* 11, e2023EF004002 (2023)

Hoffman, J., E. Tzeng, T. Park, J. Y. Zhu, and T. Darrell (2017), CyCADA: Cycle-Consistent Adversarial Domain Adaptation.

Holbrook, N. J., et al. (2019), A global assessment of marine heatwaves and their drivers, *Nat Commun*, *10*.

Hu, K., G. Huang, and S.-P. Xie (2018), Assessing the internal variability in multi-decadal trends of summer surface air temperature over East Asia with a large ensemble of GCM simulations, *Clim. Dyn.*

Huang, B., C. Liu, V. Banzon, E. Freeman, G. Graham, B. Hankins, T. Smith, and H.-M. Zhang (2021), Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) Version 2.1, *Journal of Climate*, *34*(8), 2923-2939.

Huang, P., and J. Ying (2015), A Multimodel Ensemble Pattern Regression Method to Correct the Tropical Pacific SST Change Patterns under Global Warming, *Journal of Climate*, *28*(12), 4706-4723.

Jakob Themeßl, M., Gobiet, A. and Leuprecht, A. (2011), Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. Int. J. Climatol., 31: 1530-1544. https://doi.org/10.1002/joc.2168.

Jiang, W., P. Huang, G. Huang, and J. Ying (2021), Origins of the Excessive Westward Extension of ENSO SST Simulated in CMIP5 and CMIP6 Models, *Journal of Climate*, *34*(8), 2839-2851.

Jiang, W., G. Huang, P. Huang, R. Wu, K. Hu, and W. Chen (2019), Northwest Pacific Anticyclonic Anomalies during Post–El Niño Summers Determined by the Pace of El Niño Decay, *Journal of Climate*, *32*(12), 3487-3503.

Johnson, N. C., and Y. Kosaka (2016), The impact of eastern equatorial Pacific convection on the diversity of boreal winter El Nio teleconnection patterns, *Clim. Dyn.*, *47*(12), 3737-3765.

Li, D. L., J. L. Feng, Z. H. Xu, B. S. Yin, H. Y. Shi, and J. F. Qi (2019), Statistical Bias Correction for Simulated Wind Speeds Over CORDEX-East Asia, *Earth Space Sci*, *6*(2), 200-211.

Li, G., and S. P. Xie (2012), Origins of tropical-wide SST biases in CMIP multi-model ensembles, *Geophysical Research Letters*, *39*.

Li, G., and S. P. Xie (2014), Tropical Biases in CMIP5 Multimodel Ensemble: The Excessive Equatorial Pacific Cold Tongue and Double ITCZ Problems, *Journal of Climate*, *27*(4), 1765-1780.

Liu, L., S.-P. Xie, X.-T. Zheng, T. Li, Y. Du, G. Huang, and W.-D. Yu (2013), Indian Ocean variability in the CMIP5 multi-model ensemble: the zonal dipole mode, *Clim. Dyn.*, *43*(5-6), 1715-1730.

Liu, Z., W. Zhou, and Y. Yuan (2023), 3D DBSCAN detection and parameter sensitivity of the 2022 Yangtze river summertime heatwave and drought, *Atmospheric and Oceanic Science Letters*, *16*(4), 100324.

Liu, Z., W. Zhou, R. Zhang, Y. Zhang, and Y. Wang (2022), Global-Scale Interpretable Drought Reconstruction Utilizing Anomalies of Atmospheric Dynamics, *Journal of Hydrometeorology*, *23*(9), 1507-1524.

Oliver, E. C. J., J. A. Benthuysen, N. L. Bindoff, A. J. Hobday, N. J. Holbrook, C. N.

Mundy, and S. E. Perkins-Kirkpatrick (2017), The unprecedented 2015/16 Tasman Sea marine heatwave, *Nat Commun*, *8*.

Oliver, E. C. J., J. A. Benthuysen, S. Darmaraki, M. G. Donat, A. J. Hobday, N. J. Holbrook, R. W. Schlegel, and A. Sen Gupta (2021), Marine Heatwaves, *Ann Rev Mar Sci*, *13*, 313-342.

Oliver, E. C. J., et al. (2018), Longer and more frequent marine heatwaves over the past century, *Nat Commun*, *9*.

Pan, B. X., G. J. Anderson, A. Goncalves, D. D. Lucas, C. J. W. Bonfils, J. Lee, Y. Tian, and H. Y. Ma (2021), Learning to Correct Climate Projection Biases, *Journal of Advances in Modeling Earth Systems*, *13*(10).

Tang, H., G. Huang, K. Hu, W. Jiang, W. Tao, Y. Wang, and H. Hou (2023), Weak persistence of Northwest Pacific anomalous anticyclone during post-El Niño summers in CMIP5 and CMIP6 models, *Clim. Dyn.*, *61*(7), 3805-3830.

Tao, W., G. Huang, P. Wang, Y. Liu, and D. Dong (2019), Dominant modes of CMIP3/5 models simulating northwest Pacific circulation anomalies during post-ENSO summer and their SST dependence, *Theoretical and Applied Climatology*, *138*(7–8).

Tao, W., G. Huang, K. Hu, X. Qu, G. Wen, and H. Gong (2015), Interdecadal modulation of ENSO teleconnections to the Indian Ocean Basin Mode and their relationship under global warming in CMIP5 models, *International Journal of Climatology*, *35*(3), 391-407.

Tao, W. C., G. Huang, R. G. Wu, K. M. Hu, P. F. Wang, and H. N. Gong (2018), Origins of Biases in CMIP5 Models Simulating Northwest Pacific Summertime Atmospheric Circulation Anomalies during the Decaying Phase of ENSO, *Journal of Climate*, *31*(14), 5707-5729.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An Overview of Cmip5 and the Experiment Design, *Bull. Amer. Meteorol. Soc.*, *93*(4), 485-498.

Teutschbein, C., and J. Seibert (2012), Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, *J Hydrol*, *456*, 12-29.

Themessl, M. J., A. Gobiet, and A. Leuprecht (2011), Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, *International Journal of Climatology*, *31*(10), 1530-1544.

Trenberth, K. E., G. W. Branstator, D. Karoly, A. Kumar, N. C. Lau, and C. Ropelewski (1998), Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures, *J Geophys Res-Oceans*, *103*(C7), 14291-14324.

Wang, Y., G. Huang, and K. Hu (2020), Internal variability in multidecadal trends of surface air temperature over antarctica in austral winter in model simulations, *Clim. Dyn.*, *55*(9-10), 2835-2847.

Wang, Y., K. M. Hu, G. Huang, and W. C. Tao (2021), Asymmetric impacts of El Nino and La Nina on the Pacific-North American teleconnection pattern: the role of subtropical jet stream, *Environ Res Lett*, *16*(11).

Wang, Y., K. M. Hu, G. Huang, and W. C. Tao (2023), The Role of Nonlinear Energy

Advection in Forming Asymmetric Structure of ENSO Teleconnections Over the North Pacific and North America, *Geophysical Research Letters*, *50*(17).

Wang, Y., G. Huang, K. M. Hu, W. C. Tao, H. A. Gong, K. Yang, and H. S. Tang (2022), Understanding the Eastward Shift and Intensification of the ENSO Teleconnection Over South Pacific and Antarctica Under Greenhouse Warming, *Front Earth Sc-Switz*, *10*.

Yi, Z., H. Zhang, P. Tan, and M. Gong (2017), DualGAN: Unsupervised Dual Learning for Image-to-Image Translation, *IEEE Computer Society*.

Zheng, Y. X., J. L. Lin, and T. Shinoda (2012), The equatorial Pacific cold tongue simulated by IPCC AR4 coupled GCMs: Upper ocean heat budget and feedback analysis, *J Geophys Res-Oceans*, *117*.

Zhu, J. Y., T. Park, P. Isola, and A. A. Efros (2017), Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, *IEEE*.