Author/s:
Shao, Y;Bishop, CH;Hobeichi, S;Nishant, N;Abramowitz, G;Sherwood, S

Title:
Time Variability Correction of CMIP6 Climate Change Projections

Date:
2024-02-01

Citation:
Shao, Y., Bishop, C. H., Hobeichi, S., Nishant, N., Abramowitz, G. & Sherwood, S. (2024). Time Variability Correction of CMIP6 Climate Change Projections. Journal of Advances in Modeling Earth Systems, 16 (2), https://doi.org/10.1029/2023MS003640.

Persistent Link:
https://hdl.handle.net/11343/351180

JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2023MS003640

Time Variability Correction of CMIP6 Climate Change Projections

Y. Shao[1,2], C. H. Bishop[1,2], S. Hobeichi[3,4], N. Nishant[3,4], G. Abramowitz[3,4], and S. Sherwood[3,4]

[1]ARC Centre of Excellence for Climate Extremes, The University of Melbourne, Parkville, VIC, Australia, [2]School of Geography, Earth and Atmospheric Sciences, The University of Melbourne, Parkville, VIC, Australia, [3]ARC Centre of Excellence for Climate Extremes, The University of New South Wales, Sydney, NSW, Australia, [4]Climate Change Research Centre, The University of New South Wales, Sydney, NSW, Australia

Abstract Accurate projections of climate change and associated extreme events under differing emission scenarios are linked to realistic representations of the temporal variability of the atmosphere at a variety of time scales, for example, annual, seasonal, synoptic, and daily. Here a new method is employed to explicitly quantify a model's ability to accurately represent covariance at and between differing time scales. From our global-scale analysis, on average, raw Coupled Model Intercomparison Project Phase 6 (CMIP6) models misrepresent temporal variances at differing time scales for maximum temperature (tasmax) by a considerable margin, particularly at 183-, 92- and 46-day time scales. To ameliorate such variability errors, we propose a novel Time Variability Correction (TVC) method that corrects temporal covariances while preserving the essential time-event sequence of the model simulations. We adopt a model-as-truth framework to evaluate the effectiveness of the TVC method under future forcing conditions when applied to daily tasmax simulations from 23 CMIP6 models for 1% of the global grid cells. TVC-corrected temperatures generally show improved matching of temporal variance and lag correlations in the out-of-sample projection period compared to simple mean-corrected projections. By imparting more realistic temporal-correlations to model series, TVC is expected to improve the projections of extreme events associated with persistent heat, such as heatwaves. Applying TVC to future temperature projections using actual observations significantly increases the temperature variance in most middle to high latitude land regions in the Northern Hemisphere, while decreasing it in most low to middle latitude land regions, compared to simple mean-corrected projections.

Plain Language Summary Climate change projections are often used to estimate future weather and climate conditions under differing emission scenarios. Accurate projections are required to usefully inform decision making and long-term planning. In this work, we propose a novel statistical technique to (a) quantify time series variance errors at a range of time scales, such as annual, seasonal, synoptic and so on; (b) correct these errors. By implementing the first step, we find that the time variability of the daily maximum temperature is poorly represented across a range of time scales in raw Coupled Model Intercomparison Project Phase 6 (CMIP6) models. These variability errors are then corrected in the second step. After applying the new technique to the global cases, we demonstrate that it is a powerful tool to improve variance and time-lag correlation attributes of model simulations.

Key Points:

- Time series variability within and across time scales is quantified based on backward looking time averages of differing lengths
- Coupled Model Intercomparison Project Phase 6 models grossly misrepresent temporal variances of maximum temperatures at differing time scales, particularly seasonal time scales
- This new Time Variability Correction method significantly improves the temporal variability and autocorrelation of model time series

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

Y. Shao,
yawen.shao@unimelb.edu.au

Citation:

Shao, Y., Bishop, C. H., Hobeichi, S., Nishant, N., Abramowitz, G., & Sherwood, S. (2024). Time variability correction of CMIP6 climate change projections. *Journal of Advances in Modeling Earth Systems*, *16*, e2023MS003640. https://doi.org/10.1029/2023MS003640

Received 29 JAN 2023
Accepted 25 AUG 2023

1. Introduction

Climate change projections from global climate models are indicative of how future climate may evolve under a spectrum of emission scenarios in the next few decades. The Coupled Model Intercomparison Project Phase 6 (CMIP6) archives the latest generation of climate model ensembles that are generated from climate research groups worldwide (Eyring et al., 2016). Outputs from CMIP6 multi-model ensembles often inform large-scale impact studies (Aadhar & Mishra, 2020; Martel et al., 2022; Xiang et al., 2021) and assessments of weather and climate extremes in a changing climate (Deng et al., 2021; Kim et al., 2020). They also facilitate long-term mitigation and adaptation planning in various Shared Socioeconomic Pathways scenarios (Pörtner et al., 2022).

Over the historical period, CMIP6 models often exhibit means, variances and extreme climate indices that are different to those of observation-based estimates of the climate probability distribution (Almazroui et al., 2020; Deng et al., 2021; Kim et al., 2020; Richter & Tokinaga, 2020; You et al., 2021). To ameliorate unrealistic aspects

of the model climates, various post-processing methods have been developed to minimally adjust model outputs to better align with historical probability distributions while still preserving the models' ability to project the change in climate (Abramowitz & Bishop, 2015; Bishop & Abramowitz, 2013; Maraun et al., 2017; Teutschbein & Seibert, 2012).

One post-processing approach is quantile mapping (QM). QM matches the cumulative probability distribution (CDF) of the model time series to that of the observations (Maraun et al., 2010; Tong et al., 2021). This simply formulated method adjusts all moments while maintaining the rank correlation between models and observations at a certain time scale (e.g., daily, or monthly). Other forms of QM variants are capable of allowing climate distributions to change over time (Li et al., 2010; Yang et al., 2018), or preserving historical trends in the climate projections (Cannon et al., 2015; Grillakis et al., 2017). In many multivariate correction methods, QM models, along with other CDF-based techniques, are frequently performed to correct the marginal distribution of each climate variable before the inter-variable and/or spatial correlations are adjusted (François et al., 2020; Vrac et al., 2022). The adjustment of the multivariate dependence structure essentially alters the temporal sequencing of the model series (Doblas-Reyes et al., 2021), thereby leading to bias-adjusted data that often exhibit weaker temporal dependencies than observations (François et al., 2020). It is noteworthy that when the temporal correlations of the model time series consistently align with those of the observations, it indicates a better representation of the persistence of temperature anomalies over time, and a more accurate simulation of climate extremes. On the other hand, the failure to capture key persistence attributes, such as lag-1 autocorrelation, in the observed series suggests that bias-adjusted results may not adequately simulate extreme events.

Another group of bias adjustment techniques has been developed to correct systematic model errors associated with both distribution and persistence characteristics of the time series across aggregated time scales. Multivariate Recursive Nesting Bias Correction and Multivariate Recursive Quantile-matching Nested Bias Correction aim to correct mean, variance, as well as auto and cross correlation of multiple variables at daily and multiple aggregated time scales, such as monthly, seasonal, annual, and triannual time scales (Johnson & Sharma, 2012; Mehrotra et al., 2018; Mehrotra & Sharma, 2021). In these methods, the lag auto and cross correlations in the standardized model series are matched to the observed correlations using a standard multivariate autoregressive model. However, these methods are all based on the assumption that the variability in the time series is caused by a recursive autoregressive (AR1) process. In the one-dimensional case, the correlation of variations of variables separated by n nominal time steps is simply ρ^n where ρ is the correlation over a single time step. This recursive assumption implies that the AR model does not correct time lags that are significantly longer than the longest lag used. This limitation becomes apparent when considering quasi-periodic fluctuations that exhibit varying fluctuation amplitude over time. For example, temperature fluctuations associated with the passage of a wave packet, compromised of high and low pressure systems in mid-latitudes, are expected to exhibit negative correlations over longer time lags and positive correlations over short ones.

The idea of correcting variances at differing time scales within a time series is related to the Fourier Transform (FFT) (Bracewell, 1986). Recent developments of frequency analysis based methods include Nguyen et al. (2019), who employed the FFT to correct the variability in the frequency domain, and Kusumastuti et al. (2021, 2022), who proposed discrete and continuous wavelet-based bias correction approaches to apply the wavelet transform to rectify systematic errors at all time-frequency domains, including non-dyadic frequency components in the latter study. One common limitation of these approaches is that the choice of the length of training data sets affects the model's ability to represent different frequency components, and there is a trade-off between the loss in frequency resolution and the gain in the reliability of spectral estimates. Moreover, in these methods, the coefficients of the frequency components are estimated using the entire training period, suggesting that the high-frequency components may be determined by the data further before and after.

To introduce our Time Variability Correction (TVC) method, we begin by noting that people often interpret today's temperature in the context of recent weather, such as "it's hot today and it has been a hot summer, but Spring was cold." TVC is based on a backward-looking perspective and represents today's temperature in a location in terms of differing backward-looking time series averages, each associated with a specific time scale. These averages are summed to represent today's temperature, and the sum of time-scale averages varies from one day to the next. Using this framework, for each location, we can compute the covariance between the timescale averages, and reveal how model covariances differ from observed ones across differing time scales. This approach not only diagnoses errors in model time variability, but also allows us to map a model series to a new one with

Table 1
Summary of CMIP6 Models Used in This Work

No	Institution	Model	Variant
1	CSIRO-ARCCSS	ACCESS-CM2	r1i1p1f1
2	CSIRO	ACCESS-ESM1-5	r1i1p1f1
3	AS-RCEC	TaiESM1	r1i1p1f1
4	AWI	AWI-CM-1-1-MR	r1i1p1f1
5	BCC	BCC-CSM2-MR	r1i1p1f1
6	CAS	FGOALS-g3	r1i1p1f1
7	CCCma	CanESM5	r1i1p1f1
8	CNRM-CERFACS	CNRM-CM6-1	r1i1p1f2
9	CNRM-CERFACS	CNRM-ESM2-1	r1i1p1f2
10	EC-Earth-Consortium	EC-Earth3	r1i1p1f1
11	EC-Earth-Consortium	EC-Earth3-Veg	r1i1p1f1
12	EC-Earth-Consortium	EC-Earth3-Veg-LR	r1i1p1f1
13	INM	INM-CM4-8	r1i1p1f1
14	INM	INM-CM5-0	r1i1p1f1
15	IPSL	IPSL-CM6A-LR	r1i1p1f1
16	MIROC	MIROC6	r1i1p1f1
17	MIROC	MIROC-ES2L	r1i1p1f2
18	MPI-M	MPI-ESM1-2-LR	r1i1p1f1
19	MRI	MRI-ESM2-0	r1i1p1f1
20	NCC	NorESM2-LM	r1i1p1f1
21	NCC	NorESM2-MM	r1i1p1f1
22	CNRM-CERFACS	CNRM-CM6-1-HR	r1i1p1f2
23	KIOST	KIOST-ESM	r1i1p1f1

more realistic inter-timescale covariances. We expect that correcting the variances and covariances improves the time-lag correlations of the time series. To provide some insight into this statement, we describe the direct relationship between variances at differing time frequencies and lag correlations in a simple periodic stochastic system in Text S1 in Supporting Information S1. Improved temporal correlations are expected to enhance the prediction of extreme events. We will show that TVC can preserve the time sequence of events in the raw model series.

Using daily maximum temperature projections from 23 CMIP6 models at the global scale, the TVC method is evaluated under a model-as-truth framework, also known as perfect model test. This has been widely applied for validating model weighting techniques (Abramowitz & Bishop, 2015; Abramowitz et al., 2019), statistical downscaling methods (Dixon et al., 2016; Maraun et al., 2015), as well as bias correction models (Vrac et al., 2022). In this setup, among a collection of models, one model is left out, and treated as "observation" that has data available over both historical and projection periods. The remaining models are trained toward the "observation" during the in-sample historical period and tested out-of-sample over the projection period. This process is repeated for each model in turn as "truth" and results are collated. In this regard, one can evaluate the ability of the TVC method to improve future predictions given only current climate forcing conditions.

2. Model and Observational Data

The analysis is conducted on daily near surface 2-m maximum (tasmax) air temperature (i.e., at 2 m above the surface of land, sea, or inland waters) from 23 CMIP6 models (listed in Table 1) that have data available in the historical, SSP1-2.6 (ssp126) and SSP5-8.5 (ssp585) experiments. In this study, only one member from each model's climate projection ensemble is used. This is because: (a) time correlation errors from differing projections produced by the same model are expected to be similar; (b) computational resources are limited, and we want to test our technique on a wide variety of differing climate models. In the analysis, the in-sample historical period is 1950–2014 while the out-of-sample projection period is 2015–2100. The ERA5 reanalysis product (Hersbach et al., 2020) is used as the observational data set in the historical period of 1950–2014.

For consistent comparison, the ERA5 reanalysis and all CMIP6 models are re-gridded to $1° \times 1°$ resolution using the bilinear interpolation method. We obtain ERA5 daily tasmax data by extracting the maximum value of the hourly 2-m near surface temperatures in each day.

3. Methods

3.1. Time Variability Correction (TVC) Method

In this new approach, we quantify the time variability by measuring the temporal covariance across different time scales and correct the time variability of the raw model towards the observed one. In this section, the raw historical model series is denoted as $\mathbf{x}_h$, observed series as $\mathbf{y}$, and raw projected model series as $\mathbf{x}_f$.

3.1.1. Representing the Climate State in Terms of Differing Time Scales

We associate each time scale with a backward-looking time average. The time length of the backward average defines the time scale.

Consider a time series of daily observations $\mathbf{y}^{\mathrm{T}} = (y_1, y_2, \ldots, y_n)$. Our initial objective is to represent each value in the time series in terms of a linear combination of backward-looking time averages of differing lengths. The ith backward looking averaging operator or smoother of the time series at time step q is given as,

$$y_q^i = \sum_{j=0}^{P_i-1} w_j^i y_{q-j} = \left(\mathbf{w}_j^i\right)^{\mathrm{T}} \mathbf{y}_{q-P_i+1:q}, \text{where } \sum_{j=0}^{P_i-1} w_j^i = 1 \text{ and } w_j^i \geq 0 \text{ for all } j \tag{1}$$

where P_i is the number of days in the ith smoother, and w_j^i are chosen as the averaging weights ($w_j^i = \frac{1}{P_i}$) in this study. Note that other forms of weights, time-varying weights for example, are also feasible. We select $k = 9$ smoothers, with $P_i (i = 1, 2, \ldots, k)$ set as (365, 183, 92, 46, 23, 12, 6, 3, 2). The level of pre-defined smoothing is determined from a preliminary analysis where different numbers of smoothers and days of averaging were tested, but this result is not explicitly explored here. Better performance and greater computational efficiency were achieved when 9 smoothers were used.

All levels of smoothers are then employed to separate different time scales of evolution. For example, the time series associated with the yearly-average time scale is given by

$$\mathbf{y}_{P_1:n}^1 = \begin{bmatrix} y_{P_1}^1 \\ y_{P+1_1}^1 \\ \vdots \\ y_n^1 \end{bmatrix} = \begin{bmatrix} w_0^1 & w_1^1 & \cdots & w_{P_1-1}^1 & 0 & \cdots & 0 \\ 0 & w_0^1 & w_1^1 & \cdots & w_{P_1-1}^1 & 0 & \ldots \\ \vdots & & \vdots & & \vdots & \vdots & \\ 0 & \cdots & & w_0^1 & w_1^1 & \cdots & w_{P_1-1}^1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \mathbf{W}_1 \mathbf{y}_{1:n} \tag{2}$$

where $\mathbf{W}_1$ is a $(n - P_1 + 1) \times n$ weight matrix.

The time series $\tilde{\mathbf{y}}_{P_1:n}^1 = \mathbf{y}_{P_1:n} - \mathbf{y}_{P_1:n}^1$ obtained by subtracting the time series of annual averages from the original daily time series is then free of variations associated with annual averages. The time series associated with half-year 183-day averages is given by

$$\mathbf{y}_{P_1+P_2-1:n}^2 = \begin{bmatrix} y_{P_1+P_2-1}^2 \\ y_{P_1+P}^2 \\ \vdots \\ y_n^2 \end{bmatrix} = \begin{bmatrix} w_0^2 & w_1^2 & \cdots & w_{P_2-1}^2 & 0 & \cdots & 0 \\ 0 & w_0^2 & w_1^2 & \cdots & w_{P_2-1}^2 & 0 & \ldots \\ \vdots & & \vdots & & \vdots & \vdots & \\ 0 & \cdots & & w_0^2 & w_1^2 & \cdots & w_{P_2-1}^2 \end{bmatrix} \begin{bmatrix} \tilde{y}_{P_1} \\ \tilde{y}_{P_1+1} \\ \vdots \\ \tilde{y}_n \end{bmatrix} = \mathbf{W}_2 \tilde{\mathbf{y}}_{P_1:n}^1 \tag{3}$$

The time series

$$\tilde{\mathbf{y}}_{P_1+P_2-1:n}^2 = \mathbf{y}_{P_1+P_2-1:n} - \mathbf{y}_{P_1+P_2-1:n}^2 - \mathbf{y}_{P_1+P_2-1:n}^1 \tag{4}$$

is then free of variations associated with both annual and 183-day averages.

Generalizing this procedure, the time series

$$\tilde{\mathbf{y}}_{n_k:n}^k = \mathbf{y}_{n_k:n} - \sum_{i=1}^{k} \mathbf{y}_{n_k:n}^i \text{ where } n_k = 1 - k + \sum_{i=1}^{k} P_i \tag{5}$$

is free from variations associated with all k smoothing periods.

Rearranging Equation 5 shows that if a total of k time scales are used, the n_k:n elements of the original time series are given by

$$\mathbf{y}_{n_k:n} = \tilde{\mathbf{y}}_{n_k:n}^k + \sum_{i=1}^{k} \mathbf{y}_{n_k:n}^i \tag{6}$$

To analyze how these differing time scales covary, we place the time series associated with each of the time scales as columns of a matrix

$$\mathbf{y}_{n_k:n} = \tilde{\mathbf{y}}^9_{n_9:n} + \sum_{i=1}^{9} \mathbf{y}^i_{n_9:n} = \left[\tilde{\mathbf{y}}^9_{n_9:n}, \mathbf{y}^9_{n_9:n}, \mathbf{y}^8_{n_9:n}, \ldots, \mathbf{y}^2_{n_9:n}, \mathbf{y}^1_{n_9:n}\right] \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \mathbf{Y}\mathbf{1}_{10} \tag{7}$$

where $\mathbf{1}_{10}$ is a column vector of 10 ones. Hence, the time series of the true state at any time is equal to the sum $\mathbf{Y}\mathbf{1}_{10}$ of 10 differently filtered time series $\mathbf{Y}$. Note that the column vectors in Equation 7 have $N = n + 1 - n_9 = n + 9 - \sum_{i=1}^{9} P_i$ rows. Each column of $\mathbf{Y}$ describes how the current instantaneous state is represented by one particular time scale of variation.

Similarly, we can derive the counterpart of Equation 7 as,

$$\mathbf{x}_{h,n_k:n} = \tilde{\mathbf{x}}^9_{n_9:n} + \sum_{i=1}^{9} \mathbf{x}^i_{n_9:n} = \left[\tilde{\mathbf{x}}^9_{n_9:n}, \mathbf{x}^9_{n_9:n}, \mathbf{x}^8_{n_9:n}, \ldots, \mathbf{x}^2_{n_9:n}, \mathbf{x}^1_{n_9:n}\right] \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \mathbf{X}_h\mathbf{1}_{10} \tag{8}$$

where $\mathbf{X}_h$ is the matrix of filtered raw model series over the in-sample historical period.

Note that the process of decomposing the time series can also be cast in the form of a wavelet transform (Chui, 1992; Torrence & Compo, 1998). The proof is detailed in Text S2 in Supporting Information S1.

3.1.2. Correcting Time-Scale Dependent Variability

An "ideal" (but impossible to realize) climate projection tool would be a large ensemble of replicate Earths in different weather states (because of chaos) but all experiencing the same changes in Greenhouse Gas (GHG) forcings as our Earth (Bishop & Abramowitz, 2013). That is, these replicate earths would sample the internal climate variability under the same forcing condition. In this regard, the mth member of this ensemble would produce a time series of future observations $\mathbf{y}_f$ which could then be decomposed into the time scale components $\mathbf{Y}$ using the method described in Equations 1–8. A data cube of observed values could then be formed coordinated by three indices: q, the calendar day in the time series; i, the time scale component; and m, the member of the ensemble of replicate Earths from which the observation is taken. With such a data cube, on any specific calendar day q, one could compute the time-varying ensemble mean and covariance of differing time scale components.

However, we do not have access to future observations of replicate Earths. We only have historical observations from our Earth. To crudely estimate these "ideal" means and covariances, one can make quasi-ergodic assumptions that averages through time and covariances around the time-dependent running means will yield similar values. In this study, as an initial experiment, we assume that the mean and the covariance of the time scales are time invariant, and the model warming trend is not corrected.

The time series in each lth column (time scale) of $\mathbf{X}_h$ and $\mathbf{Y}$ have climatological means $\mu_l^{\mathbf{X}_h}$ and $\mu_l^{\mathbf{Y}}$, respectively. We correct the mean of each column of $\mathbf{X}_h$ using

$$\mathbf{x}_h^{c,l} = \left[\mathbf{x}^l_{n_9:n} + \left(\mu_l^{\mathbf{Y}} - \mu_l^{\mathbf{X}_h}\right)\right] \tag{9}$$

to create a mean-corrected matrix of time series of time scale components $\mathbf{X}_h^c$. The next step is to ensure that the elements of $\mathbf{X}_h^c$ covary in time in the same way as the elements of $\mathbf{Y}$. Before variance correction, the two covariances can be estimated from the mean-corrected values as

$$\mathbf{C_Y} = \frac{\mathbf{Y}^{\mathrm{T}}\left(\mathbf{I} - \frac{\mathbf{1}_N\mathbf{1}_N^{\mathrm{T}}}{N}\right)\left(\mathbf{I} - \frac{\mathbf{1}_N\mathbf{1}_N^{\mathrm{T}}}{N}\right)\mathbf{Y}}{N-1} \text{ and } \mathbf{C_X} = \frac{\mathbf{X}_h^{c\,\mathrm{T}}\left(\mathbf{I} - \frac{\mathbf{1}_N\mathbf{1}_N^{\mathrm{T}}}{N}\right)\left(\mathbf{I} - \frac{\mathbf{1}_N\mathbf{1}_N^{\mathrm{T}}}{N}\right)\mathbf{X}_h^c}{N-1} \tag{10}$$

where $\mathbf{1}_N$ denotes a column vector whose N row elements all equal to 1, and $\mathbf{I}$ is an $N \times N$ identity matrix. We can then define a new, covariance-corrected time series $\mathbf{Z}_h$ as

$$\mathbf{Z}_h = \frac{\mathbf{1}_N\mathbf{1}_N^{\mathrm{T}}}{N}\mathbf{X}_h^c + \left(\mathbf{I} - \frac{\mathbf{1}_N\mathbf{1}_N^{\mathrm{T}}}{N}\right)\mathbf{X}_h^c\mathbf{C}_\mathbf{X}^{-1/2}\mathbf{C}_\mathbf{Y}^{1/2} \tag{11}$$

To see that the covariance $\mathbf{C}_{\mathbf{Z}_h}$ of $\mathbf{Z}_h$ is identical to $\mathbf{C}_y$, note that

$$\begin{aligned}\mathbf{C}_{\mathbf{Z}_h} &= \frac{\mathbf{Z}_h{}^{\mathrm{T}}\left(\mathbf{I} - \frac{\mathbf{1}_N\mathbf{1}_N^{\mathrm{T}}}{N}\right)\left(\mathbf{I} - \frac{\mathbf{1}_N\mathbf{1}_N^{\mathrm{T}}}{N}\right)\mathbf{Z}_h}{N-1} = \frac{\mathbf{C}_\mathbf{Y}^{1/2}\mathbf{C}_\mathbf{X}^{-1/2}\mathbf{X}_h^{c\mathrm{T}}\left(\mathbf{I} - \frac{\mathbf{1}_N\mathbf{1}_N^{\mathrm{T}}}{N}\right)\left(\mathbf{I} - \frac{\mathbf{1}_N\mathbf{1}_N^{\mathrm{T}}}{N}\right)\mathbf{X}_h^c\mathbf{C}_\mathbf{X}^{-1/2}\mathbf{C}_\mathbf{Y}^{1/2}}{N-1} \\ &= \mathbf{C}_\mathbf{Y}^{1/2}\mathbf{C}_\mathbf{X}^{-1/2}\mathbf{C}_\mathbf{X}\mathbf{C}_\mathbf{X}^{-1/2}\mathbf{C}_\mathbf{Y}^{1/2} = \mathbf{C}_\mathbf{Y}\end{aligned} \tag{12}$$

In other words, after TVC has been applied to the in-sample period, the rows of $\mathbf{Z}_h$ have the same mean and covariance as $\mathbf{Y}$ and the final time series $\mathbf{z}_h = \mathbf{Z}_h\mathbf{1}_{10}$ is consequently expected to have much more similar time-scale dependent variability to the observed time series $\mathbf{y}$.

Note that over the in-sample period, the parameters $\mu_l^{\mathbf{X}_h}$, $\mu_l^{\mathbf{Y}}$, $\mathbf{C}_\mathbf{X}$, and $\mathbf{C}_\mathbf{Y}$ are estimated using the entire training time series. The nature of the TVC method means that the first $n_k - 1$ days are discarded in the post-processed time series. The final series is approximately 2 years shorter than the raw series.

3.1.3. Out-Of-Sample Application

To post-process future out-of-sample projections, raw projected model series $\mathbf{x}_f^T = \left(x_{f(1)}, x_{f(2)}, \ldots, x_{f(n_f)}\right)$ is first decomposed into differing time scales following the same procedures as shown in Equations 2–7. That is

$$\mathbf{x}_{f,n_k:n_f} = \mathbf{X}_f\mathbf{1}_{10} \tag{13}$$

where $\mathbf{X}_f$ is the matrix of filtered raw model series over the out-of-sample projection period. The column vectors in Equation 13 have $N_f = n_f + 1 - n_9 = n_f + 9 - \sum_{i=1}^{9} P_i$ rows.

Then, the mean of each column in $\mathbf{X}_f$ is corrected as

$$\mathbf{x}_f^{c,l} = \left[\mathbf{x}_{f,n_9:n_f}^l + \left(\mu_l^{\mathbf{Y}} - \mu_l^{\mathbf{X}_h}\right)\right] \tag{14}$$

which creates a mean-corrected matrix $\mathbf{X}_f^c$. The covariance is corrected by defining a new post-processed matrix $\mathbf{Z}_f$ as

$$\mathbf{Z}_f = \frac{\mathbf{1}_{N_f}\mathbf{1}_{N_f}^{\mathrm{T}}}{N_f}\mathbf{X}_f^c + \left(\mathbf{I} - \frac{\mathbf{1}_{N_f}\mathbf{1}_{N_f}^{\mathrm{T}}}{N_f}\right)\mathbf{X}_f^c\mathbf{C}_\mathbf{X}^{-1/2}\mathbf{C}_\mathbf{Y}^{1/2} \tag{15}$$

where the scaling factor of the covariance, $\mathbf{C}_\mathbf{X}^{-1/2}\mathbf{C}_\mathbf{Y}^{1/2}$, is obtained from the corresponding in-sample historical runs. $\mathbf{1}_{N_f}$ denotes a column vector whose N_f row elements are all equal to 1.

Finally, after summing up all time scales, we derive the new TVC post-processed projection series $\mathbf{z}_f = \mathbf{Z}_f\mathbf{1}_{10}$.

3.2. Method Evaluation

3.2.1. In-Sample Testing of a Grid Cell

To explore why and how TVC works, we apply TVC to post-process CMIP6 tasmax simulations at a grid cell (37.5°S 144.5°E) over the in-sample period of 1950–2014. For one run, time series based on daily ERA5 temperatures and raw CMIP6 simulations are decomposed into differing time scales using the filtering approach elaborated in Section 3.1.1. For individual models, we plot back-filtered time series and temporal covariance of all

time scales before applying TVC. The performance of mean-corrected raw and TVC post-processed simulations is compared and summarized.

3.2.2. Model as Truth Experiment

The model-as-truth framework, also known as perfect model test, is implemented to comprehensively evaluate the TVC method when applied to future emission scenarios (Abramowitz & Bishop, 2015; Abramowitz et al., 2019). Specifically, one CMIP6 model is nominated as "observations," in which data of both historical (1950–2014) and projection (2015–2100) period are available. The remaining 22 models are treated as raw simulations and are post-processed against the "observations" separately. For each of the 22 models, TVC is trained over the in-sample historical period, 1950–2014, and applied to the out-of-sample projection period, 2015–2100. This process is repeated for each CMIP6 model as "observations," giving 506 TVC and raw simulations (23 ("observation") × 22 (training model)) for each grid cell, each temperature variable, and each scenario in the following analyses.

In this study, we post-process 1% of the global grid cells, including one-tenth of the longitude and one-tenth of the latitude. In total, there are 648 cells (36 × 18) for evaluation. These 1% of the cells are selected because: (a) these cells cover both land and ocean region; (b) computational resources are limited and the model-as-truth results from 648 cells should be sufficient to draw the conclusion about the robustness of TVC for general cases.

3.2.3. Application to Future Emission Scenario

After conducting the model-as-truth validation where the method is trained on historical model output, we employ TVC for the future application by training on observational data. That is, for each CMIP6 model and each variable at each grid cell, we train the raw model against corresponding ERA5 reanalysis data over the historical period, 1950–2014, and then apply the trained TVC model to the projected emission scenarios over 2015–2100. In this future application, TVC is employed to post-process all the grid cells at the global scale and produce mean- and variance-corrected projections according to each model.

3.2.4. Verification Metric

For each time series, we compute sample variance and persistence attributes. Here, the sample variance

$$\sigma^2 = \frac{\sum_{t=1}^{T} \left(x_t - x_t^{\mathrm{MA}}\right)^2}{T-1} \tag{16}$$

measures the variability relative to the 30-year running mean over the evaluation period, where T is the length of the time series, x_t is the data at time step t, x_t^{MA} is the 30-year moving average centered at time step t. Note for the first (last) 15-year intervals, the data from the initial (final) 30 years are averaged.

Persistence attributes are checked by single time-lag autocorrelation (lag-1 correlation) and 5-day lag autocorrelation of 5-day averages (lag-5 correlation). Lag-1 correlation is the correlation between two sets of the time series that are one day apart, formulated as

$$\rho_1 = \frac{\sum_{t=1}^{T-1} \left(x_t - x_t^{\mathrm{MA}}\right)\left(x_{t+1} - x_{t+1}^{\mathrm{MA}}\right)}{\sqrt{\sum_{t=1}^{T-1} \left(x_t - x_t^{\mathrm{MA}}\right)^2 \sum_{t=1}^{T-1} \left(x_{t+1} - x_{t+1}^{\mathrm{MA}}\right)^2}} \tag{17}$$

The lag-5 correlation is estimated as the correlation between the block series of the 5-day averages that are one 5-day time step apart. The length of the 5-day averages time series is approximately 1/5 of the original one.

Climate indices are often used to characterize weather and climate extreme events. We calculate and examine a temperature-based index, warm spell duration index (WSDI) following the general procedures used by the Expert Team on Climate Change Detection and Indices (ETCCDI) (Zhang et al., 2011). By standard definition, WSDI refers to the annual count of at least 6 consecutive days when the daily maximum temperature is larger than the 90th percentile value. For each calendar day, the 90th percentile is calculated from a centered 5-day window over a 30-year base period. The selection of the base period for evaluation is slightly different from case to case, which

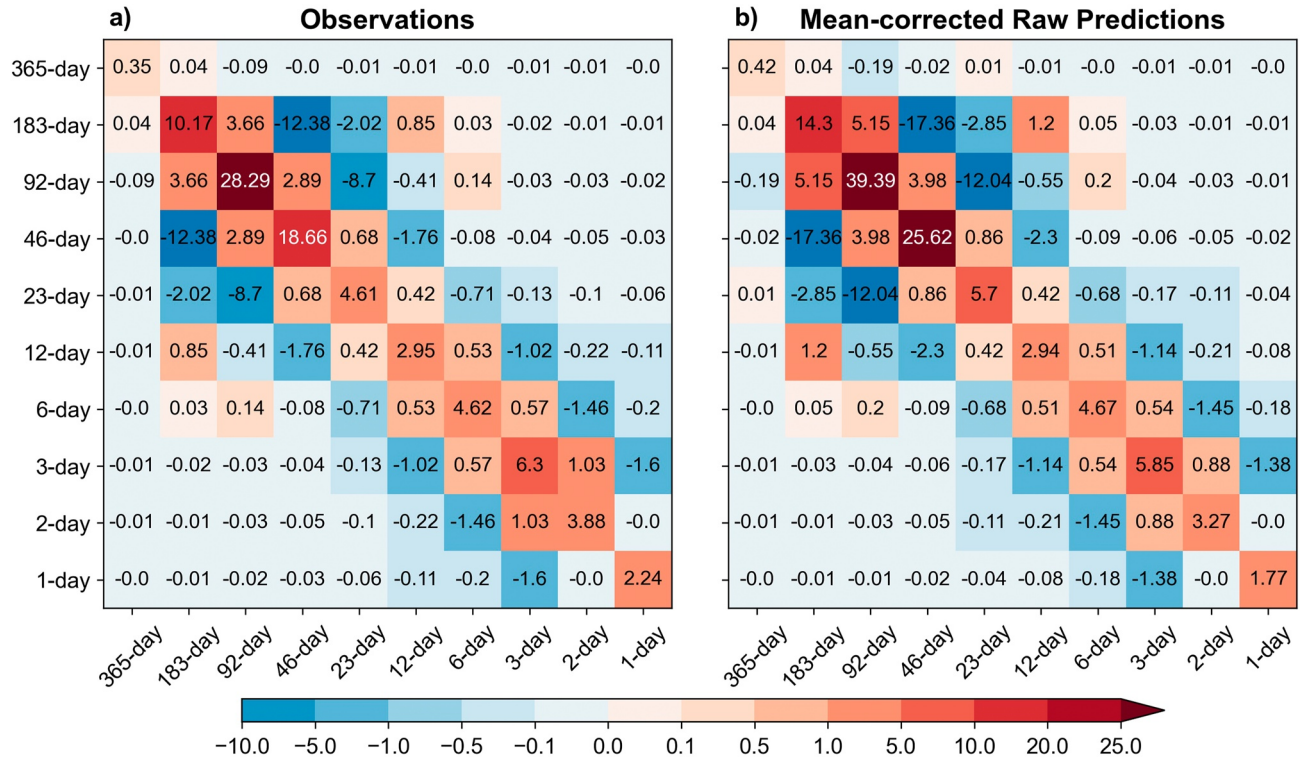

Figure 1. Heatmaps of matrix of covariance across time scales for (a) observations and (b) mean-corrected raw ACCESS-ESM1-5 simulations for tasmax at a grid cell (37.5°S 144.5°E) over the in-sample period, 1950–2014.

will be elaborated in the relevant results section. We compute the averaged WSDI values across the evaluation period for the comparison.

For each model-as-truth, we compare the TVC post-processed and mean-corrected raw projections by calculating the percentage improvements in averaged mean absolute errors (MAE) for variance, lag correlations and climate indices, formulated as

$$\mathrm{MAE} = \frac{1}{22}\sum_{m=1}^{22}|\mathrm{Model}_m - \mathrm{Obs}| \tag{18}$$

$$\%\,\mathrm{improvement} = \frac{\mathrm{MAE}_{\mathrm{RAW}} - \mathrm{MAE}_{\mathrm{TVC}}}{\mathrm{MAE}_{\mathrm{RAW}}} \times 100\% \tag{19}$$

where $\mathrm{MAE}_{\mathrm{RAW}}$ stands for the MAE value for mean-corrected raw temperature projections, while $\mathrm{MAE}_{\mathrm{TVC}}$ is the MAE for TVC post-processed projections.

4. Results

4.1. In-Sample Application of TVC to a Single Grid Cell

In this section, we apply TVC to post-process a single grid cell and examine the performance of mean-corrected raw and TVC post-processed tasmax simulations over the in-sample period, 1950–2014.

Figure 1 shows the covariance matrices of both observations and mean-corrected raw ACCESS-ESM1-5 data. This ACCESS-ESM1-5 model is an arbitrary choice to show the details of modeled results. Diagonal elements refer to the temporal variance of each time scale, while off-diagonal elements are the covariances between differing time scales. Raw ACCESS-ESM1-5 simulations misrepresent the temporal variance of tasmax across all time scales (Figure 1), over-estimating the variance across longer time scales, particularly the 183-, 92-, and 46-day time scales by the respective ratios 1.41, 1.39, and 1.37 (compare diagonal elements between Figures 1a and 1b).

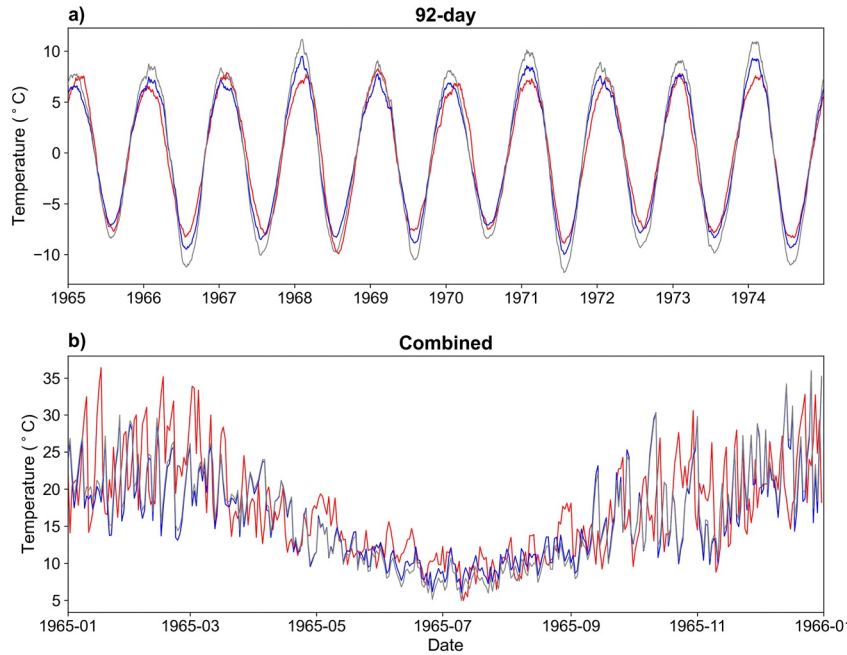

Figure 2. Time series plot of in-sample (a) 92-day time scale and (b) combined series for ERA5 (red line), mean-corrected raw (gray line) and TVC post-processed (blue line) ACCESS-ESM1-5 simulations for tasmax at a grid cell (37.5°S 144.5°E) over selected time periods.

Conversely, most of the remaining time scale variances were underestimated. This finding is also evident in the time series of 92-day time scale plotted in Figure 2, as well as the time series of all time scales presented in Figures S6 and S7 in Supporting Information S1. Apparently, the seasonal cycles shown in the mean-corrected raw 92-day time series (Figure 2a) generally have larger magnitudes than the observations. In Figure 1, the covariances between differing time scales are also not well represented by ACCESS-ESM1-5, notably the covariance between adjacent time scales, 183-day and 92-day time scales for example. In addition, we represent other CMIP6 models in terms of differing time scales and find that the covariance is widely misrepresented in the models (not shown). Similarly, over 80% of the CMIP6 models overestimate the variances at 183-, 92-, and 46-day time scales while over 60% of the models underestimate the variances at shorter than 6-day time scales.

The time series after TVC post-processing is also plotted in Figure 2, Figure S6 and S7 in Supporting Information S1. TVC significantly turns the seasonal cycles much closer to observations. Moreover, as shown in Figure 2b, and the last plot of Figure S7 in Supporting Information S1, to the eye, TVC appears to preserve the time sequencing of the events of the original model time series, because the TVC and mean-corrected raw time series look very similar to each other. This finding indicates that the TVC approach does not seem to introduce spurious unrealistic artifacts into the time series.

Figure 3 compares and summarizes the pooled results of the statistics and climate indices. Referring to Figure 3a, mean-corrected raw CMIP6 simulations largely misrepresent the temporal variance relative to the 30-year running mean, particularly MIROC6 (in red circle) and KIOST-ESM (in pink square). After TVC post-processing, the variance is almost identical to the observations as expected. For time-correlation attributes, it is shown that most mean-corrected raw models were over-correlated through time (Figures 3b and 3c). With TVC post-processing, both lag-1 correlation and lag-5 correlation become much closer to the observed ones. Regarding climate indices (Figure 3d), WSDI is calculated using the base period of 1961–1990 and averaged across the evaluation period of 1952–2014. Overall, the predictions of WSDI are improved through TVC post-processing. We found stronger improvements in 1-day and 5-day lag correlations with relatively weaker improvements in WSDI. One major reason for the weaker improvement seen for WSDI is the fact that the current WSDI definition of the 90th percentile is subject to random fluctuations, and such "statistical stability of metric" issues might have obfuscated the ability of TVC to improve duration indices. Another potential reason is TVC could not directly correct the quantiles of the climate distribution. More discussion of these issues is given in Section 5.

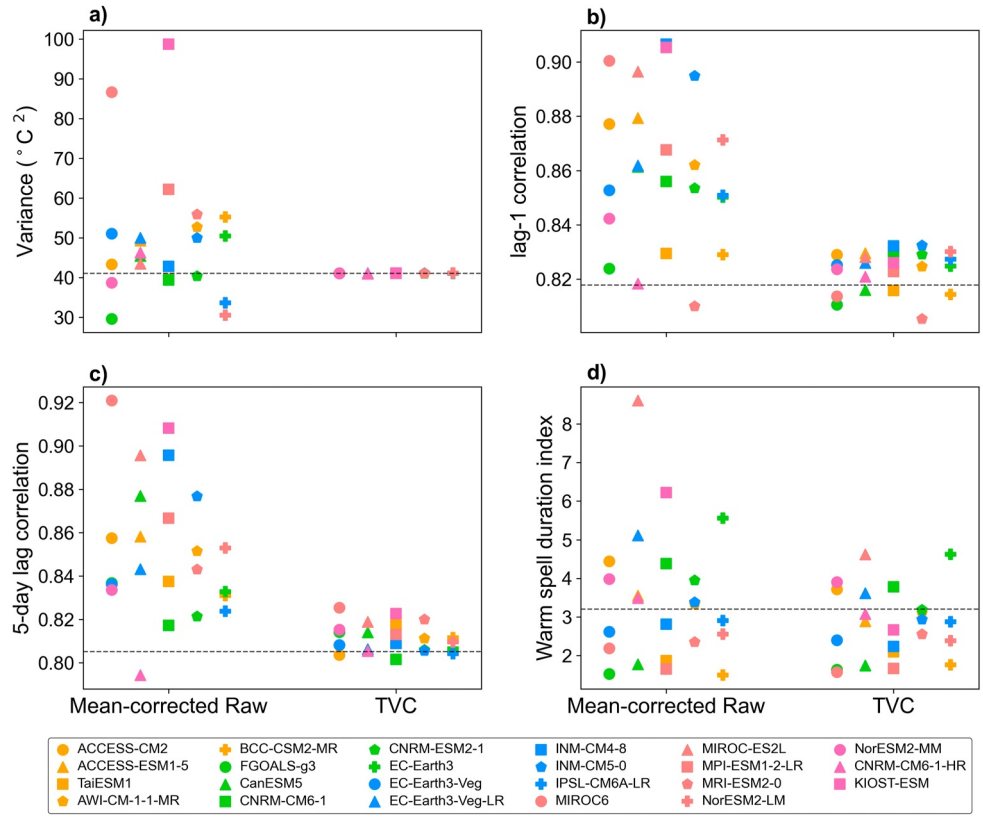

Figure 3. Scatter plots of (a) in-sample variance, (b) lag-1 correlation, (c) lag-5 correlation, and (d) warm spell duration index of mean-corrected raw and TVC post-processed CMIP climate model simulations of tasmax at a grid cell (37.5°S 144.5°E). The observed values are plotted as horizontal gray dashed line.

4.2. Model-As-Truth Experiment

Here, we present the results from the perfect model test in the form of boxplots, representing the spread over all evaluated grid cells.

Figure 4 shows the percentage of improvement in MAE of variance and lag-1 correlation associated with 30-year running mean for all model-as-truth and all scenarios (i.e., in-sample historical, out-of-sample ssp126 and ssp585 periods). We expect to observe greater improvements during the in-sample period than in the out-of-sample periods because this is when TVC is being trained. Since TVC directly corrects the covariance across pre-defined time scales, it is unsurprising to observe that in most cells, the improvement of the variance relative to the 30-year running mean over the historical period is close to 100% (Figure 4a). In both ssp126 and ssp585 scenarios, compared to mean-corrected raw predictions, TVC improves the variance across most grid cells in each model-as-truth. Specifically, approximately half of the grid cells show a percentage improvement larger than 75% in the ssp126 scenario and over 60% in the ssp585 scenario.

In terms of lag-1 correlation (Figure 4b), positive improvements are noticeable for most cases across both in-sample and out-of-sample periods. During historical period, more than half of the cells are with improvements greater than 70% in most model-as-truth setups, except for CNRM-CM6-1-HR. Particularly, when BCC-CSM2-MR and KIOST-ESM models are considered as "observations," percentage improvements are over 80%. Compared to the in-sample period, less improvement is shown for projections in both low and high emission scenarios. Having said that, post-processing by TVC still leads to over 60% and 50% improvements in half of the grid cells in low emission and high emission scenario in all other model-as-truth setups, respectively.

Figure 5 presents the percentage of improvement for lag-5 correlation and WSDI. The general pattern of the lag-5 correlation is similar to that of the lag-1 correlation (Figure 4a). The most substantial improvement is observed for the historical period, while a more modest improvement is seen for the ssp585 scenario. Generally, median

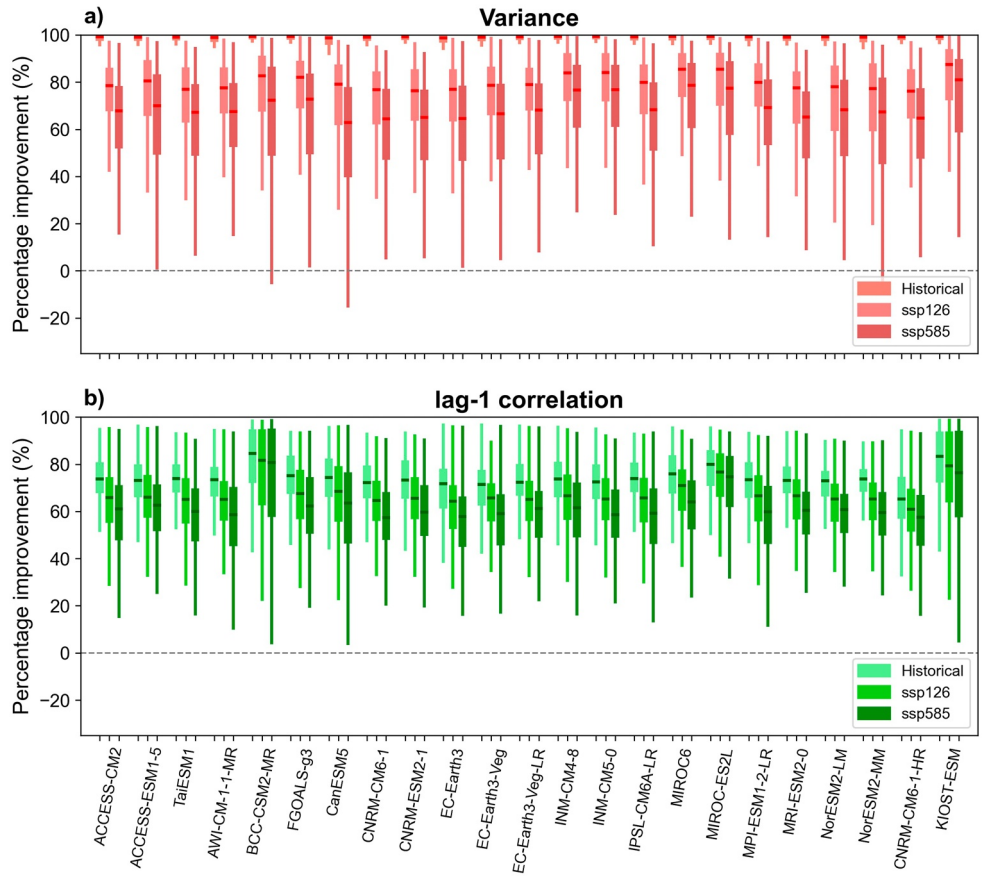

Figure 4. Box plot showing the percentage of improvement in mean absolute errors (MAE) of tasmax (a) variance and (b) lag-1 correlation for each model-as-truth and each experiment. For the calculation of these two metrics, the in-sample historical period is 1950–2014, and the out-of-sample projection period for ssp126 and ssp585 is 2015–2100. The middle horizontal line is the median value, the middle box represents the inter-quartile range between 25th percentile (Q1) and 75th percentile (Q3) of the data, the upper whisker equals to Q3 + 1.5 × (Q3 − Q1), and the lower whisker equals to Q1 − 1.5 × (Q3 − Q1). The outliers beyond the whiskers are not plotted.

percentage improvement of lag-5 correlation is greater than that of lag-1 correlation across all scenarios. Interestingly, across all grid cells, the range of historical improvement in lag-5 correlation is much narrower than that for the lag-1 correlation in each model-as-truth. In contrast, for the ssp585 scenario, the improvement brought by TVC varies a lot across all the cells in many model-as-truth setups. For example, for CanESM5, EC-Earth3, and EC-Earth3-Veg models, the range of improvements spans from negative values to exceeding 90%.

Since remarkable improvements are evident for time correlations, it will be interesting to examine how such improvements contribute to the predictions of threshold-based duration indices. Here, WSDI is measured in a way slightly different from the standard procedure. Specifically, we ensure that errors in the prediction of the climate mean do not affect WSDI scores to a large extent by computing the 90th percentiles based on the climate probability distribution functions of the "true" climate model both in the historical training period and the future testing period. For example, during the in-sample TVC training period, the period of the time series that defines percentiles is from 1961 to 1990, while for the out-of-sample projection period, the period of the time series that defines the critical climate percentiles is from 2061 to 2090. Furthermore, we only calculate and compare the averaged climate indices for observations, mean-corrected raw, and TVC post-processed projections over the 30-year base period.

For WSDI (Figure 5b), median percentage improvements relative to raw projections are greater than 20% in each model-as-truth experiment. On average, the TVC performance does not differ a lot across all scenarios, while more improvements are observed for the in-sample historical period. Apparently, the predictions of WSDI are worsened in around 15%–25% of the grid cells, regardless of the improvement in lag correlations. As mentioned

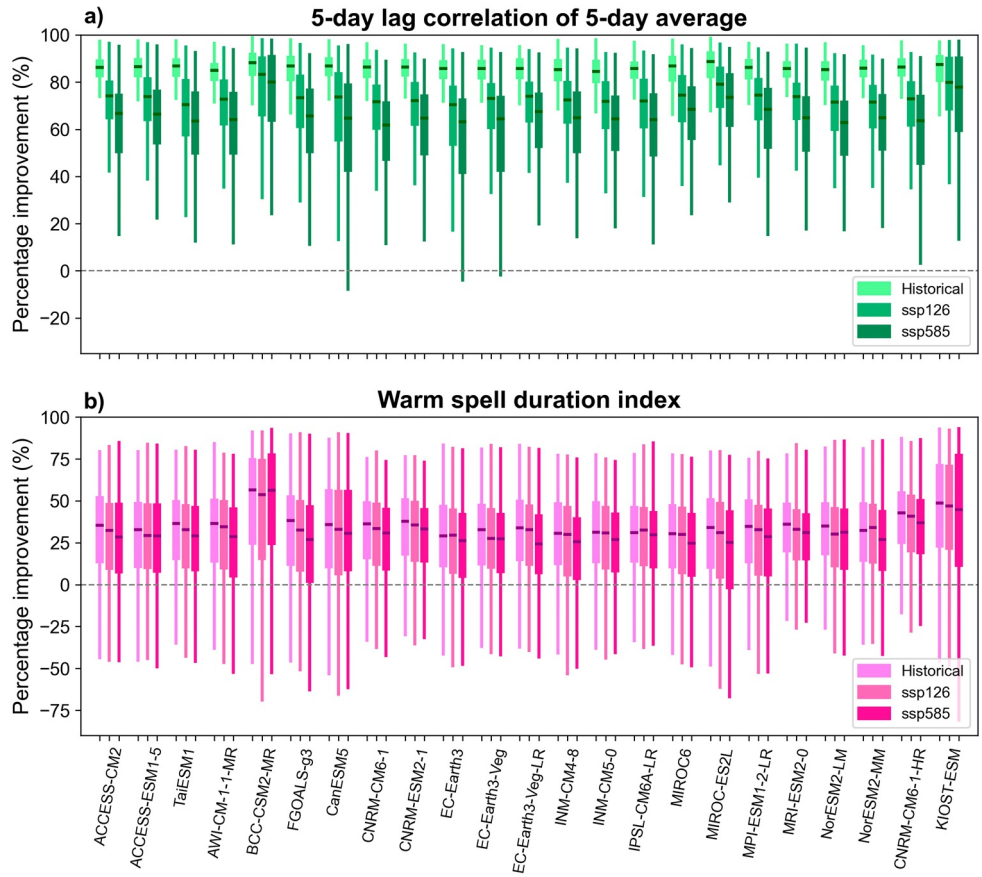

Figure 5. Box plot showing the percentage of improvement in mean absolute errors (MAE) of tasmax (a) lag-5 correlation and (b) warm spell duration index (WSDI) for each model-as-truth and each experiment. For TVC applications, the in-sample historical period is 1950–2014, and the out-of-sample projection period for ssp126 and ssp585 is 2015–2100. The lag-5 correlation is computed over the same period. The WSDI is calculated over the 30-year base period, which is 1961–1990 for the historical period and 2061–2090 for the projection period. The interpretation of the box plot is the same as Figure 4.

in Section 4.1, the calculation of the extreme indices here is subject to large uncertainty, and the results might not be statistically meaningful.

4.3. Impact of TVC on Climate Projections

Having demonstrated the effectiveness of the TVC method in the model-as-truth framework, in this section, we apply TVC to post-process global tasmax projections from individual models. Figure 6 shows the absolute variance difference between TVC and mean-corrected raw data at each time scale over the in-sample historical and out-of-sample ssp585 projection period for all the selected CMIP6 models. The globally weighted difference for each model is calculated by applying a scaling factor of cosine(latitude) to weight each grid cell. Figure 6 indicates that the size of the variance change made by TVC varies with time scale in a qualitatively similar way for all the models. Notably, the biggest variance changes made by TVC are at the 92-day time scale. The general patterns of the scatters for both in-sample historical (Figure 6a) and out-of-sample ssp585 (Figure 6b) periods are very similar, despite slightly different values. This is because we apply the same variance correction factor obtained in the historical training period to the out-of-sample projection period. Over the historical period (last panel in Figure 6a), the variance difference between TVC (equivalent to ERA5 data) and mean-corrected raw historical data for each daily series is larger than $5°C^2$, indicating that on average, raw CMIP6 models misrepresent the temporal variability of ERA5 maximum temperatures at the global spatial scale. In both plots, the largest absolute differences are seen in KIOST-ESM ($>15°C^2$) and MIROC6 ($15°C^2$) models, aligned with the case results shown in Figure 3. The overall variance mismatch can be largely explained by the variance difference at

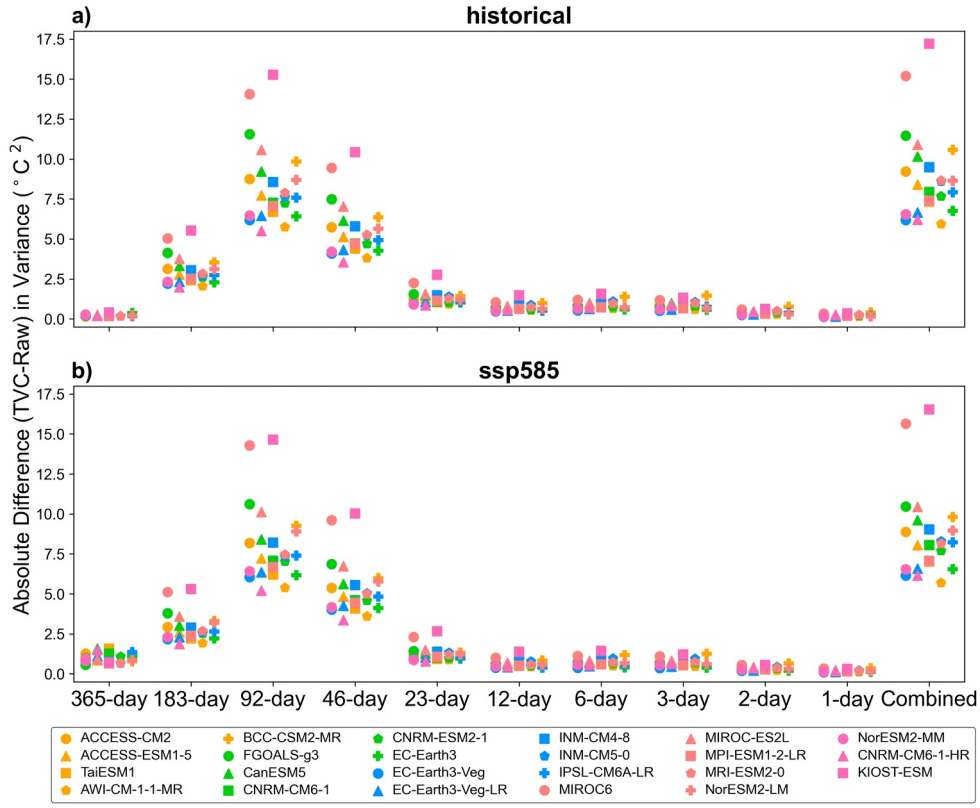

Figure 6. Globally weighted absolute difference between TVC and mean-corrected raw simulations in time-scale-dependent variance at each time scale for each CMIP6 model over the in-sample historical period and out-of-sample ssp585 projection period.

the 183-, 92- and 46-day time scale for both in-sample and out-of-sample periods, particularly the 92-day time scale where the seasonal cycle dominates the variation in the model series.

Figure 7 presents the averaged absolute variance difference between ERA5 and mean-corrected raw simulations for ACCESS-ESM1-5 and all CMIP6 models across each longitude over the in-sample historical period. Figure 7 indicates that the largest changes to variance made by TVC occurred in middle to high latitude regions. In the ACCESS-ESM1-5 model (Figure 7a), the greatest variance discrepancy is shown between 49.5° and the pole in the Northern Hemisphere. Note that very large variance differences ($30°C^2 \sim 50°C^2$) occur in these mid-high latitude regions. Figure 7b presents the variance difference averaged across all 23 CMIP6 models. More variance errors take place in the middle to high latitude (49.5°∼90°) region in the Northern Hemisphere, and high latitude region (−69.5°∼−90°) in the Southern Hemisphere. In general, greater variance differences in these middle to high latitude regions result from the larger amplitudes of the seasonal cycles throughout the year, in contrast to the smaller variations of the temperatures near the equator.

Figure 8 plots the metric difference between TVC post-processed and mean-corrected raw projections (i.e., $M_{TVC} - M_{RAW}$) over the in-sample historical and out-of-sample ssp585 periods. Metrics include temporal variance and lag correlations relative to the 30-year running mean over the evaluation period. Differences are averaged across all 23 CMIP6 models. The results for the ssp126 scenario are similar to the ssp585 scenario, so they are not shown here. After TVC post-processing, general spatial patterns of the metric differences are similar for both in-sample historical and out-of-sample ssp585 projections. For the averaged variance across all CMIP6 models (Figures 8a and 8b), areas with red (blue) color correspond to the regions where the observation informed TVC time series suggests that the models are under-variant (over-variant).

For both evaluation periods, raw models are under-variant in a large portion of the North and North-East Asian continent, and parts of North America. Elsewhere across land regions, on average, CMIP6 models tend to overestimate the variance (in blue). TVC produces lower variances in the high latitude (70°∼90° and −90°∼−70°),

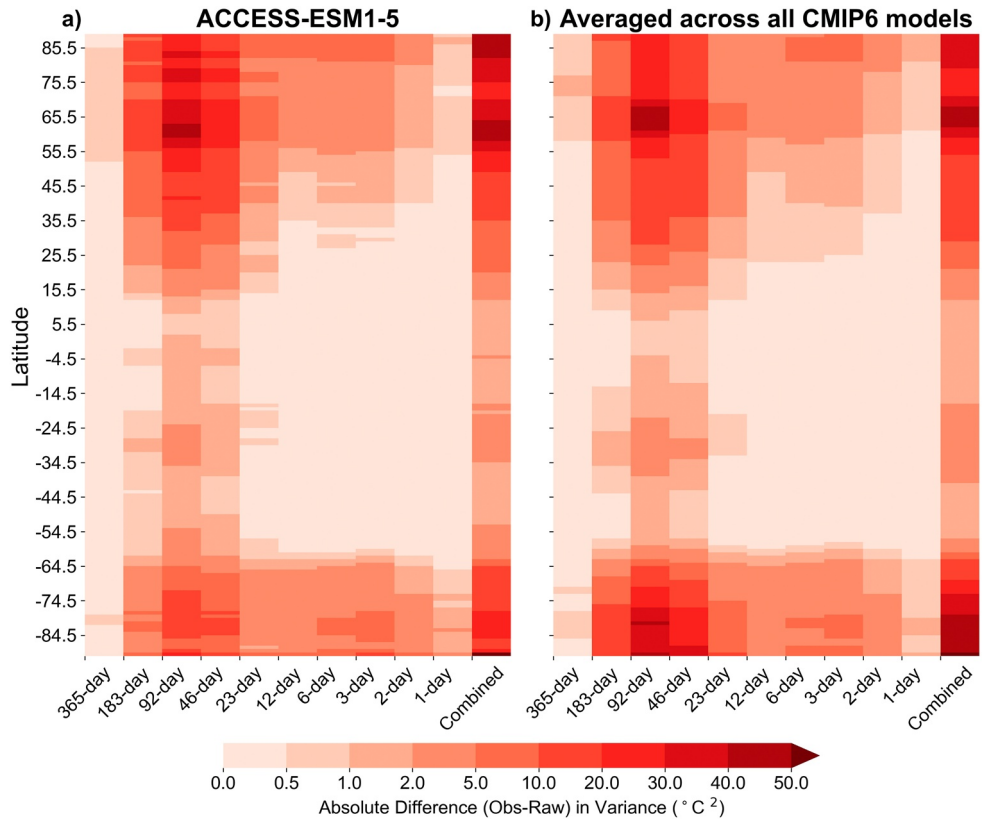

Figure 7. Absolute difference between ERA5 reanalysis and mean-corrected raw simulations in time-scale-dependent variance averaged at each latitude for (a) ACCESS-ESM1-5 only and (b) averaged across all CMIP6 models during the in-sample historical period.

indicating that the averaged variance is likely to be over-estimated in raw CMIP6 models, possibly due to the poor simulation of sea ice cover near the poles. In general, less variance change can be seen across ocean than land regions. Exceptions are parts of the North Atlantic Ocean, the North Pacific Ocean, and Southern Ocean, where raw CMIP6 models are under-variant near the coastlines and also, roughly speaking, underneath the North Pacific and North Atlantic storm tracks.

For a single model, the metric differences for the ACCESS-ESM1-5 model are provided in Figure S8 in Supporting Information S1. For variance (Figures S8a and S8b in Supporting Information S1), the historical variance is considerably underestimated by the model (in red) across middle and high latitude land regions in the Northern Hemisphere, and the variance difference has a larger magnitude than the averaged one across all CMIP6 models. TVC also leads to weak, but still noticeable lower variance in most of African, Australian, South American, and Antarctica continents. Again, less variance change is observed across ocean than land regions.

The spatial patterns for both lag-1 correlations are shown in Figures 8c and 8d and for lag-5 correlation in Figures 8e and 8f. The predominance of brown colors in Figures 8c and 8d and pink colors in Figures 8e and 8f show that in general, correlations of tasmax over 1 day and 5 days are higher in the models than the TVC time series. Figures S8c–S8f in Supporting Information S1 show the corresponding correlation plots for the ACCESS-ESM1-5 model. They are qualitatively similar to those in Figure 8 and hence demonstrate that the time correlation errors of ACCESS-ESM1-5 time series are, unfortunately, very similar to those found in the other CMIP6 models. For both historical and ssp585 future projection periods, there is a noticeable correction by TVC post-processing that is concentrated in ocean areas, particularly in the Southern Hemisphere. Raw projections are often over-correlated in the tropical rainfall zones, and the Southern Hemisphere storm track (in brown and pink). TVC results in much higher temporal correlations in the west Pacific Ocean along the equator. The correlation difference is generally greater than 0.15 for the averaged lag correlations across all CMIP6 models. Similarly, the difference is larger than 0.3 in lag correlations for ACCESS-ESM1-5. One could speculate that the coarseness of

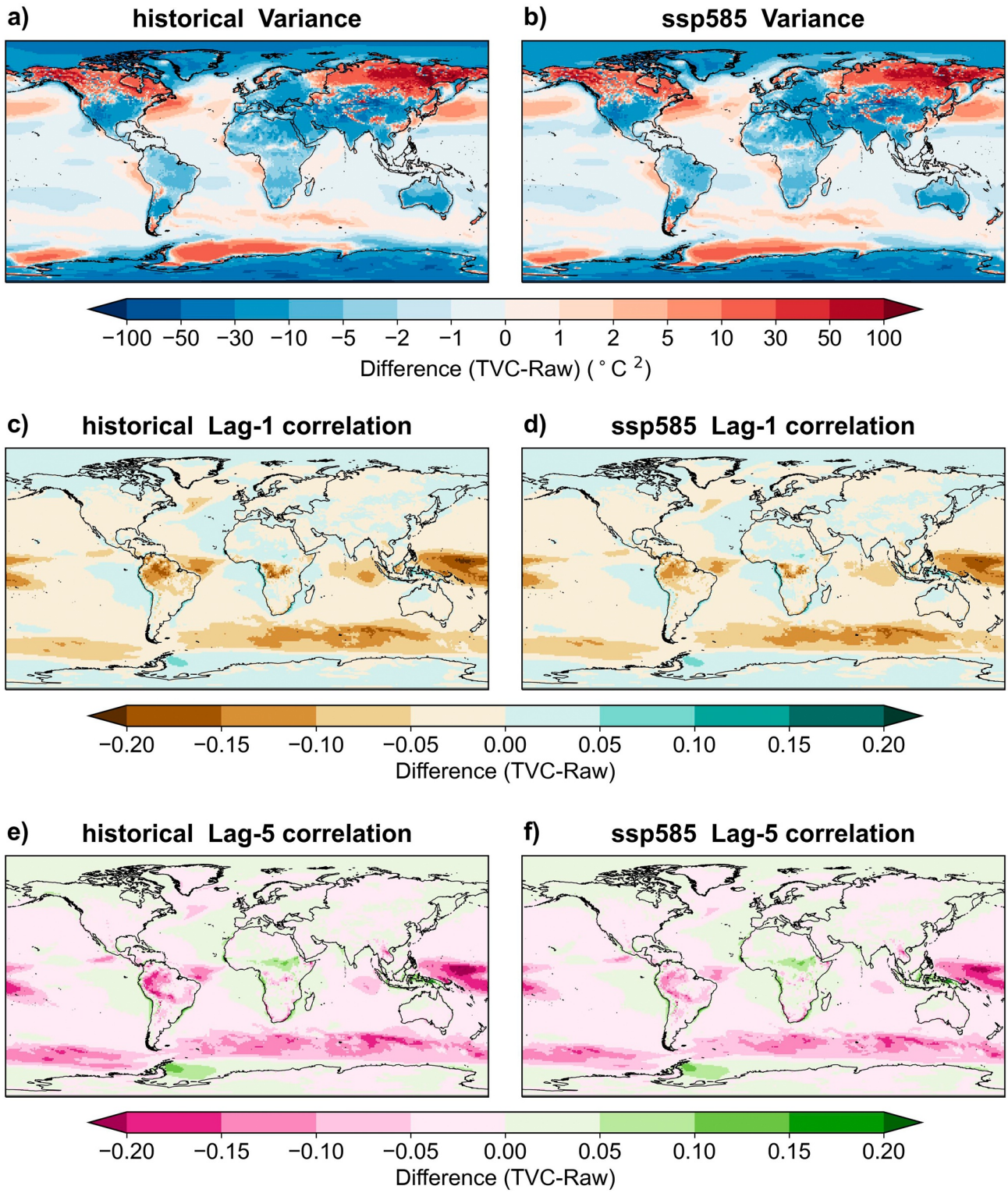

Figure 8. Difference in variance (a, b), lag-1 correlation (c, d), and lag-5 correlation (e, f) between TVC and mean-corrected raw tasmax predictions across the globe for in-sample historical over December 1951–2014 and out-of-sample ssp585 over December 2016 to December 2100. The variance is the averaged value across all CMIP6 models.

CMIP6 model representations of the ocean and the ocean-atmosphere interface has played a role in suppressing higher frequency climate variations leading to over correlated temperature time series.

5. Discussion and Future Work

The new TVC method represents each point in the time series as a sum of differing length backward looking time averages. On any given day, the components associated with the shortest time scales are associated with very recent events. This contrasts strongly with a FFT approach where the coefficient of the Fourier component associated with the shortest time scales is determined by the data long before, and possibly, long after the day of interest. Moreover, our TVC approach is distinguished from conventional FFT in the following ways:

1. TVC corrects covariance between differing time scales rather than rectifying temporal variability in individual frequency domains (Kusumastuti et al., 2021);
2. TVC has the potential to be applied within the context of forecasting because the model state at time t is corrected using data no later than time t;
3. The TVC method gives time series for each time scale, which can be easily visualized. This enables one to check if the time series has more variance at short time scales in, for example, summer than in winter. With a FFT approach, such changes are only observed by systematic interference between Fourier modes of differing frequencies.

We note that TVC is closely related to wavelet transform type approaches, and there is an opportunity to integrate more wavelet basis functions into our initial set of time-averaging functions. Future work will explore this possibility.

When applying the post-processing techniques to the climate model output, caution should be exercised to avoid introducing spurious artifacts into the time series. Inevitably, the degree to which a post-processing technique modifies or interferes with the physics-based projections of the climate model varies among different methods (Abramowitz et al., 2019; Maraun, 2016). One attractive feature of the TVC approach is that it largely preserves the ordering of events and time sequencing of the raw model time series to which it is applied. Furthermore, TVC preserves projections of changes in time variability because we apply the same covariance scaling factor obtained in the historical period to the projection period. For example, if a CMIP6 model predicts a reduction in the seasonal variational of temperature in a future climate relative to current climate, TVC would preserve this reduction in the relative change in seasonal variation in the future climate.

In the global analysis, we diagnose variance errors in mean-corrected raw CMIP6 models at each time scale. On top of the variance analyses done for global averages, further work will comprehensively evaluate how each CMIP6 model represents the temporal variability at differing time scales. Ideally, the problem would be corrected by optimizing the model, but it is also not clear how easy or how to correct such problems in models. Consequently, there is a pressing need for statistical post-processing that would make such model time scale covariances more like the observed covariances.

Within the current TVC framework, we do not account for any treatment for climate change signals. Indeed, a few existing methods attempt to preserve trends (Grillakis et al., 2017; Hempel et al., 2013) or remove trend-biases (Kusumastuti et al., 2021) in the corrected climate projections. There is a long-standing controversy about whether the trend needs to be adjusted in the bias correction (Maraun, 2016). Trend-preserving methods work under the assumption that the model biases are time invariant so that the method corrects the time-invariance of the bias rather than the climate trend. Post-processing methods that embed trend-altering components assume that the trend difference between observed and raw simulations over the historical period applies to the future projections. This assumption is problematic in forcing scenarios with different boundary conditions. Both categories of trend-adjustment methods may not be ideal for trend correction. To account for the climate warming in the model correction, a possible solution is to use the CMIP6 ensemble to better infer the hypothetical time-varying means and covariances of time scales discussed in Section 3.1.2. Further work will explore the possibility of combining TVC with model weighting technique to estimate time-varying ensemble means and better gauge the model uncertainty (Abramowitz & Bishop, 2015; Bishop & Abramowitz, 2013). A better representation of climate warming is expected to further improve the attributes and prediction of the climate indices, particularly in the high emission scenarios.

A common approach for testing statistical post-processing techniques is to set up cross validation procedures over the historical period (Johnson & Sharma, 2012; Mehrotra & Sharma, 2016; Yang et al., 2018). However, such validation method does not test the effectiveness of the post-processing under future forcing scenarios. As found in the past (Abramowitz et al., 2019), some corrections that work well on historical data do not perform well under the different forcing conditions during the later twenty-first century. In this work, we implement the model-as-truth setup, another testing framework to comprehensively evaluate the new TVC method for both in-sample historical and out-of-sample projection periods (Abramowitz & Bishop, 2015; Abramowitz et al., 2019). The parallel computing and support from supercomputing systems make it possible to carry out this perfect model test with multi-model ensembles. The model-as-truth and cross validation setups are essentially complementary. Future endeavors can be made to compare various bias correction methods under both validation frameworks for global and regional climate projections.

We mainly follow the standard procedures used by ETCCDI to calculate a threshold-based climate extreme index, WSDI, as an indicator of temperature-based extremes. However, this index is statistically unstable because the amount of pooled data over the base period (i.e., 5-day running window × 30 years) are not sufficient to obtain the true, representative percentile value for each calendar day. In a sample size of 150 days, the 90th percentile corresponds to the temperature of the fifteenth warmest day in the sample. Obviously, if we were to take 150 samples from a normal distribution for a couple of times and look at the fifteenth highest value from each sampling, the value obtained would fluctuate from one to the next. Such fluctuations may be relatively small compared to the overall shift of the climate mean. Nevertheless, in our case, where the focus is on the temporal correlations of temperatures, such fluctuations could have a significant impact on the results. In Text S3 in Supporting Information S1, we use a simple example to illustrate that while a significant improvement in the lag correlations is observed, it may not result in comparable improvements of the WSDI. One possible explanation could be the inadequate representation of the percentile threshold given a short period of data. Furthermore, the index results are sensitive to the selection of the base period, as well as the changes in the mean of the climate distribution. Future work will examine the uncertainty associated with percentile calculations and seek options to establish a more statistically stable characterization of the probability of persistent extremes in current and future climate. We will also attempt to understand the effect of TVC on predicting extremes.

Apart from 2-m temperatures, TVC is also applicable for other continuous variables, such as sea surface temperature (SST) and wind components. It will be interesting to explore the marine climates by understanding the behavior of TVC post-processed SST projections. For non-continuous variables, such as lower bounded precipitation, more complicated solutions need to be proposed. Currently, TVC is developed on a univariate base. Future research avenues may extend TVC to account for spatiotemporal attributes for multivariable post-processing.

6. Conclusion

In this study, we propose a novel wavelet-based TVC method to quantify time-scale-dependent variance errors and to correct the covariance of predicted variables at differing lengths of time scales. TVC is first applied to maximum temperatures (tasmax) at a grid cell in an in-sample setup. Raw CMIP6 models are found to widely misrepresent temporal covariance across differing time scales, greatly overestimating the variance at 183-, 92-, and 46-day time scales. Post-processing by TVC fixes the covariance mismatch issue, improves the realism of the temporal correlation of model time-series while retaining the qualitative order-of-events given by the model. We also apply the TVC method to a wide range of selected cells on tasmax simulations from 23 CMIP6 models under a model-as-truth experiment. Results show that TVC significantly improves variance, single-day time-lag-correlation and 5-day lag correlation of 5-day averages in all in-sample historical periods. TVC also yields substantial improvements in out-of-sample ssp126 and ssp585 projection periods, with larger improvements seen for the low emission ssp126 scenario.

To improve the temporal variability of current climate projections, we train TVC against ERA5 reanalysis data over the historical period and apply trained model to correct future projections. The variance analysis over the in-sample historical period reveals that globally, raw CMIP6 models misrepresent temporal variability at each time scale, particularly the 183-, 92-, and 46-day time scales that are all associated with seasonality. TVC adjusted time series are found to differ from raw model series most strongly over the middle to high latitude regions. With TVC post-processing, for both in-sample historical and out-of-sample high emission scenarios, corrected projections are more variable over middle and high latitude land regions in the Northern Hemisphere, particularly in

North and North-East Asia, and less variable in other land regions. For 1-day and 5-day time correlations, TVC projections of temperature time series are found to be generally less correlated through time than raw model series, especially in the west Pacific Ocean.

The improvements to climate model variability produced by TVC could be used to more accurately inform socio-economic decisions associated with ameliorating and mitigating the effects of climate change. Another merit of TVC is that it is capable of diagnosing time scale related model errors. Such diagnoses allow model developers to explicitly see where, when and at what time scales the simulations of their model variables are systematically over/under variant and/or over/under time correlated. These model systematic errors are likely to affect model's ability to accurately project the climate warming caused by increased GHG Concentrations.

Ongoing research will attempt to develop a hybrid system with TVC, and a model weighting technique combined to boost the performance of individual models and model ensembles, as well as to better gauge the projection uncertainty. The extension of TVC to be applied for discrete variables, rainfall for example, will also be investigated.

Data Availability Statement

The CMIP6 data (World Climate Research Programme, 2022) used in this work can be downloaded from: https://esgf-node.llnl.gov/search/cmip6/ and the ERA5 reanalysis data (Hersbach et al., 2023) are on: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=form. The codes are all implemented in Python ≥ 3.8 (Python, 2019). The scripts for the TVC method, data analyses, and figure plotting are available at https://zenodo.org/records/10212122 (Shao, 2023).

Acknowledgments
This work is fully funded by the Australian Research Council Centre of Excellence for Climate Extremes (CLEX; CE170100023). We thank National Computational Infrastructure (NCI) for supplying supercomputing systems for computational analyses.

References

Aadhar, S., & Mishra, V. (2020). On the projected decline in droughts over South Asia in CMIP6 multimodel ensemble. *Journal of Geophysical Research: Atmospheres*, *125*(20), e2020JD033587. https://doi.org/10.1029/2020jd033587

Abramowitz, G., & Bishop, C. (2015). Climate model dependence and the ensemble dependence transformation of CMIP projections. *Journal of Climate*, *28*(6), 2332–2348. https://doi.org/10.1175/jcli-d-14-00364.1

Abramowitz, G., Herger, N., Gutmann, E., Hammerling, D., Knutti, R., Leduc, M., et al. (2019). ESD Reviews: Model dependence in multi-model climate ensembles: Weighting, sub-selection and out-of-sample testing. *Earth System Dynamics*, *10*(1), 91–105. https://doi.org/10.5194/esd-10-91-2019

Almazroui, M., Saeed, S., Saeed, F., Islam, M. N., & Ismail, M. (2020). Projections of precipitation and temperature over the South Asian countries in CMIP6. *Earth Systems and Environment*, *4*(2), 297–320. https://doi.org/10.1007/s41748-020-00157-7

Bishop, C. H., & Abramowitz, G. (2013). Climate model dependence and the replicate Earth paradigm. *Climate Dynamics*, *41*(3–4), 885–900. https://doi.org/10.1007/s00382-012-1610-y

Bracewell, R. N. (1986). *The Fourier transform and its applications* (Vol. 31999). McGraw-Hill.

Cannon, A. J., Sobie, S. R., & Murdock, T. Q. (2015). Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *Journal of Climate*, *28*(17), 6938–6959. https://doi.org/10.1175/jcli-d-14-00754.1

Chui, C. K. (1992). *An introduction to wavelets* (Vol. 1). Academic press.

Copernicus Climate Change Service, Climate Data Store. (2023). ERA5 hourly data on single levels from 1940 to present [Dataset]. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*.

Deng, X., Perkins-Kirkpatrick, S. E., Lewis, S. C., & Ritchie, E. A. (2021). Evaluation of extreme temperatures over Australia in the historical simulations of CMIP5 and CMIP6 models. *Earth's Future*, *9*(7), e2020EF001902. https://doi.org/10.1029/2020ef001902

Dixon, K. W., Lanzante, J. R., Nath, M. J., Hayhoe, K., Stoner, A., Radhakrishnan, A., et al. (2016). Evaluating the stationarity assumption in statistically downscaled climate projections: Is past performance an indicator of future results? *Climatic Change*, *135*(3–4), 395–408. https://doi.org/10.1007/s10584-016-1598-0

Doblas-Reyes, F., Sorensson, A., Almazroui, M., Dosio, A., Gutowski, W., Haarsma, R., et al. (2021). Linking global to regional climate change.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

François, B., Vrac, M., Cannon, A. J., Robin, Y., & Allard, D. (2020). Multivariate bias corrections of climate simulations: Which benefits for which losses? *Earth System Dynamics*, *11*(2), 537–562. https://doi.org/10.5194/esd-11-537-2020

Grillakis, M. G., Koutroulis, A. G., Daliakopoulos, I. N., & Tsanis, I. K. (2017). A method to preserve trends in quantile mapping bias correction of climate modeled temperature. *Earth System Dynamics*, *8*(3), 889–900. https://doi.org/10.5194/esd-8-889-2017

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., & Piontek, F. (2013). A trend-preserving bias correction–the ISI-MIP approach. *Earth System Dynamics*, *4*(2), 219–236. https://doi.org/10.5194/esd-4-219-2013

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2023). ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). https://doi.org/10.24381/cds.adbb2d47

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Johnson, F., & Sharma, A. (2012). A nesting model for bias correction of variability at multiple time scales in general circulation model precipitation simulations. *Water Resources Research*, *48*, W01504. https://doi.org/10.1029/2011wr010464

Kim, Y.-H., Min, S.-K., Zhang, X., Sillmann, J., & Sandstad, M. (2020). Evaluation of the CMIP6 multi-model ensemble for climate extreme indices. *Weather and Climate Extremes*, *29*, 100269. https://doi.org/10.1016/j.wace.2020.100269

Kusumastuti, C., Jiang, Z., Mehrotra, R., & Sharma, A. (2021). A signal processing approach to correct systematic bias in trend and variability in climate model simulations. *Geophysical Research Letters*, *48*(13), e2021GL092953. https://doi.org/10.1029/2021gl092953

Kusumastuti, C., Jiang, Z., Mehrotra, R., & Sharma, A. (2022). Correcting systematic bias in climate model simulations in the time-frequency domain. *Geophysical Research Letters*, *49*(19), e2022GL100550. https://doi.org/10.1029/2022gl100550

Li, H., Sheffield, J., & Wood, E. F. (2010). Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *Journal of Geophysical Research*, *115*, D10101. https://doi.org/10.1029/2009jd012882

Maraun, D. (2016). Bias correcting climate change simulations-a critical review. *Current Climate Change Reports*, *2*(4), 211–220. https://doi.org/10.1007/s40641-016-0050-x

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., et al. (2017). Towards process-informed bias correction of climate change simulations. *Nature Climate Change*, *7*(11), 764–773. https://doi.org/10.1038/nclimate3418

Maraun, D., Wetterhall, F., Ireson, A., Chandler, R., Kendon, E., Widmann, M., et al. (2010). Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, *48*, RG3003. https://doi.org/10.1029/2009rg000314

Maraun, D., Widmann, M., Gutiérrez, J. M., Kotlarski, S., Chandler, R. E., Hertig, E., et al. (2015). VALUE: A framework to validate downscaling approaches for climate change studies. *Earth's Future*, *3*(1), 1–14. https://doi.org/10.1002/2014ef000259

Martel, J. L., Brissette, F., Troin, M., Arsenault, R., Chen, J., Su, T., & Lucas-Picher, P. (2022). CMIP5 and CMIP6 model projection comparison for hydrological impacts over North America. *Geophysical Research Letters*, *49*(15), e2022GL098364. https://doi.org/10.1029/2022gl098364

Mehrotra, R., Johnson, F., & Sharma, A. (2018). A software toolkit for correcting systematic biases in climate model simulations. *Environmental Modelling & Software*, *104*, 130–152. https://doi.org/10.1016/j.envsoft.2018.02.010

Mehrotra, R., & Sharma, A. (2016). A multivariate quantile-matching bias correction approach with auto-and cross-dependence across multiple time scales: Implications for downscaling. *Journal of Climate*, *29*(10), 3519–3539. https://doi.org/10.1175/jcli-d-15-0356.1

Mehrotra, R., & Sharma, A. (2021). A robust alternative for correcting systematic biases in multi-variable climate model simulations. *Environmental Modelling & Software*, *139*, 105019. https://doi.org/10.1016/j.envsoft.2021.105019

Nguyen, H., Mehrotra, R., & Sharma, A. (2019). Correcting systematic biases across multiple atmospheric variables in the frequency domain. *Climate Dynamics*, *52*(1), 1283–1298. https://doi.org/10.1007/s00382-018-4191-6

Pörtner, H.-O., Roberts, D. C., Adams, H., Adler, C., Aldunce, P., Ali, E., et al. (2022). Climate change 2022: Impacts, adaptation and vulnerability. IPCC Sixth Assessment Report.

Python (2019). *Python 3.8.0*. Python Software Foundation. Retrieved from https://www.python.org/downloads/release/python-380/

Richter, I., & Tokinaga, H. (2020). An overview of the performance of CMIP6 models in the tropical Atlantic: Mean state, variability, and remote impacts. *Climate Dynamics*, *55*(9), 2579–2601. https://doi.org/10.1007/s00382-020-05409-w

Shao, Y. (2023). Time variability correction. Version v1.0.0. https://zenodo.org/records/10212122

Teutschbein, C., & Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology*, *456*, 12–29. https://doi.org/10.1016/j.jhydrol.2012.05.052

Tong, Y., Gao, X., Han, Z., Xu, Y., Xu, Y., & Giorgi, F. (2021). Bias correction of temperature and precipitation over China for RCM simulations using the QM and QDM methods. *Climate Dynamics*, *57*(5), 1425–1443. https://doi.org/10.1007/s00382-020-05447-4

Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, *79*(1), 61–78. https://doi.org/10.1175/1520-0477(1998)079<0061:apgtwa>2.0.co;2

Vrac, M., Thao, S., & Yiou, P. (2022). Should multivariate bias corrections of climate simulations account for changes of rank correlation over time? *Journal of Geophysical Research: Atmospheres*, *127*(14), e2022JD036562. https://doi.org/10.1029/2022jd036562

World Climate Research Programme. (2022). CMIP6 search interface [Dataset]. ESGF-CoG. Retrieved from https://esgf-node.llnl.gov/search/cmip6/

Xiang, Y., Wang, Y., Chen, Y., & Zhang, Q. (2021). Impact of climate change on the hydrological regime of the Yarkant River Basin, China: An assessment using three SSP scenarios of CMIP6 GCMs. *Remote Sensing*, *14*(1), 115. https://doi.org/10.3390/rs14010115

Yang, X., Wood, E. F., Sheffield, J., Ren, L., Zhang, M., & Wang, Y. (2018). Bias correction of historical and future simulations of precipitation and temperature for China from CMIP5 models. *Journal of Hydrometeorology*, *19*(3), 609–623. https://doi.org/10.1175/jhm-d-17-0180.1

You, Q., Cai, Z., Wu, F., Jiang, Z., Pepin, N., & Shen, S. S. (2021). Temperature dataset of CMIP6 models over China: Evaluation, trend and uncertainty. *Climate Dynamics*, *57*(1), 17–35. https://doi.org/10.1007/s00382-021-05691-2

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., et al. (2011). Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdisciplinary Reviews: Climate Change*, *2*(6), 851–870. https://doi.org/10.1002/wcc.147