

Research papers

Spectrally transformed CMIP6 decadal projections improve interannual rainfall forecasts

Ze Jiang^{a,*}, Dipayan Choudhury^b, Ashish Sharma^{a,*}^a School of Civil and Environmental Engineering, University of New South Wales, Sydney, Australia^b Department of Climate Change, Energy, the Environment and Water, New South Wales, Australia

ARTICLE INFO

This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Ankit Agarwal, Associate Editor

Keywords:

Decadal prediction experiments
Spectral transformation
Hierarchical linear combination
Long-term rainfall forecasts

ABSTRACT

Forecasting rainfall over long lead times (from interannual to decadal scales) is essential for informing water resource management and adaptation planning, especially for drought-prone countries such as Australia. The latest suite of Coupled Model Intercomparison Project Phase 6 (CMIP6) models provides simulations of decadal prediction experiments, although their skill in simulating rainfall remains limited. This study investigates the merit of using spectrally transformed simulations of sea surface temperature anomaly (SSTA) from CMIP6 decadal experiments for interannual rainfall forecasting over Australia. SSTA indices representative of large-scale modes of climate variability are transformed to enhance the prediction of rainfall at specific locations using Wavelet System Prediction (WASP) method. These spectrally transformed SSTA indices are then used as inputs into a hierarchical linear combination (HLC) model to forecast rainfall over interannual timescales. Results show that this hybrid HLC-WASP model significantly improves the predictability of interannual rainfall compared to the existing approach where no spectral transformation is used, up to lead times of five years ahead. This spectral transformation technique has the potential to enhance the quality of inputs for other rainfall forecasting models, particularly over regions with significant teleconnections between climate variability and regional rainfall. Improved rainfall forecasts have significant implications for water resource planning and management, enabling better drought mitigation, efficient water allocation, informed infrastructure planning, and more effective climate adaptation strategies.

1. Introduction

Effective water resource management and adaptation planning depend heavily on reliable long-term rainfall forecasts, particularly under the impending risks of climate change (IPCC, 2021; Meehl et al., 2009; Moemken et al., 2021). For drought-prone regions like Australia, accurate predictions can significantly aid in mitigating the adverse effects of prolonged dry periods and optimizing water usage during wet seasons (Deb et al., 2020; Liang et al., 2017; Roderick et al., 2020; Xu et al., 2020). The Coupled Model Intercomparison Project Phase 6 (CMIP6) provides a suite of advanced climate models that simulate decadal prediction experiments, offering a potential resource for improving long-term rainfall forecasts. The Decadal Climate Prediction Project (DCPP) is a coordinated multi-model investigation into decadal climate prediction, predictability, and variability. This involves the use of retrospective forecasts (commonly referred to as hindcasts) to validate models against historical observations, as well as the ongoing

production of operational decadal climate forecasts (Boer et al., 2016; Eyring et al., 2016). However, the skill of these models in simulating rainfall remains limited (IPCC, 2014; Mehrotra et al., 2014; Smith et al., 2019; Smith et al., 2020).

Most decadal predictability studies have focused on temperature or temperature-related indices at global or regional scales, demonstrating skill in the prediction of temperature variability at interannual timescales (Befort et al., 2020; Chikamoto et al., 2015; Choudhury et al., 2015; Mahmood et al., 2021; Mehta et al., 2013). In contrast, rainfall predictability is significantly lower (IPCC, 2014; Sheen et al., 2017; Smith et al., 2019). This discrepancy between prediction skills of temperature and rainfall is not only limited to decadal predictions but across multiple forecasting timescales and models. As a workaround, studies have often used statistical and dynamical models using sea surface temperature anomaly (SSTA) indices to forecast rainfall over longer lead times, leveraging established relationships between large-scale climate variability and regional rainfall (Meehl et al., 2010). In Australia, the

* Corresponding authors.

E-mail addresses: ze.jiang@unsw.edu.au (Z. Jiang), a.sharma@unsw.edu.au (A. Sharma).

Australian Bureau of Meteorology (BoM) previously issued operational seasonal rainfall forecasts based on approximates of climate indices derived from empirical orthogonal functions (EOFs). These approximates represent large-scale climate variability, such as El Niño-Southern Oscillation (ENSO), Indian Ocean Dipole (IOD), and the El Niño Modoki (EMI) (Schepen and Wang, 2015; Schepen et al., 2012, 2014). Seasonal rainfall forecasts over Australia that employed the EOF-based approximates of these climate indices as predictor variables showed improved predictability, particularly when utilizing the multi-model combination approach (Choudhury et al., 2019; Khan et al., 2015; Schepen et al., 2014).

One challenge in long-term rainfall prediction is the decaying skill of the decadal outputs with increasing lead time due to model drift (Billbao et al., 2021; Boer et al., 2016). This shift of the model simulations away from the observations towards the model equilibrium state occurs as forecasts progress, despite the drift-correction of model simulations (Choudhury et al., 2017; Fučkar et al., 2014; Smith et al., 2013). Challenges in long-term climate prediction using dynamical models, including aspects such as model initialization, boundary conditions, model resolution, structure, and parameterization, remain significant (Jiang et al., 2019). While enhancing dynamical model resolution or refining convection parameterization could potentially improve rainfall forecasting, it is pertinent to note that advances in statistical forecasting methods present an additional avenue for improvement. This is supported by findings in the realm of statistical forecasts (Khan et al., 2015), emphasizing the importance of the integration of decadal simulations with statistical model frameworks. For instance, the hierarchical linear combination (HLC) model with a distribution-based transformation of rainfall was applied to interannual rainfall forecasting using decadal simulations (Choudhury et al., 2019). Rainfall is often a positively skewed variable, and its associated errors frequently deviate from normality, violating the assumptions of many predictive models. Therefore, applying appropriate transformations to rainfall data is crucial for enhancing the reliability and accuracy of forecasting models. There has been a range of variable transformations as a means of improving model specification in hydrological forecasting applications (Helsel and Hirsch, 2002; Hyndman and Athanasopoulos, 2014; McInerney et al., 2017; Wu et al., 2019), including log and reciprocal transformation, Box-Cox transformation (Box and Cox, 1964), and log-sinh transformation (Wang et al., 2012).

Recent advances in statistical forecasting methods, such as transformation that targets different components of the predictor variable spectrum, offer a promising avenue for improvement. Spectral transformation, introduced by Jiang et al. (2020) modifies the spectral representation of predictor variables to align with the associated response, improving the predictor-response relationship and predictability. This technique, initially applied to downscaling, adopts the discrete wavelet transform (DWT) to decompose signal or time series into separate rapidly changing, high-frequency sub-time series (i.e., details) and slowly changing, low-frequency sub-time series (i.e., approximations). Then, the variances of these decompositions of predictor variables are redistributed to approach those of the corresponding response. As a result, the spectrally transformed predictors can better predict the response in terms of various metrics, including correlation, root mean square error (RMSE) as well as other distributional statistics. This was presented in Jiang et al. (2020, 2021a, 2023), who showed substantial improvements in predicting sustained rainfall deficits (i.e., droughts) using spectrally transformed atmospheric predictors or climate indices. The existing issue of DWT requiring future information in the context of forecasting applications was addressed by replacing DWT with maximal overlap DWT (MODWT). The application of the proposed spectral transformation to forecast an ENSO index using SST and wind stress-based predictors resulted in significant improvements, with an average increase of 20% across the investigated lead times, particularly for lead times up to 24 months (Jiang et al., 2021b). The associated Wavelet System Prediction (WASP) software (Jiang et al., 2021c) facilitates this

spectral transformation, extracting critical information from predictors through the spectral domain.

Despite these advancements, applications for rainfall forecasting using decadal predictions accounting for the difference in spectral properties between the target rainfall and their associated predictors have not been reported at the time of writing this study. This study aims to address this gap by integrating the spectral transformation approach, implemented by WASP, with the existing HLC model, utilizing outputs of the CMIP6 decadal prediction experiments. The resulting hybrid HLC-WASP model presented here accounts for the differences in spectral properties between the response and predictors. The HLC-WASP model is evaluated for its ability to predict Australian interannual rainfall using spectrally transformed climate indices (CIs).

In pursuit of this, we have three fundamental objectives:

- 1) Developing an optimal model, including predictor selection and parameter optimization, for interannual Australian rainfall predictions using the CMIP6 decadal simulations.
- 2) Quantifying the merit of integrating spectral information into statistical forecasting systems, specifically for interannual Australian rainfall.
- 3) Evaluating the performance of the multi-model ensemble mean of the spectrally transformed forecasts against individual transformed CMIP6 models.

Given the extra complexity of the hybrid HLC-WASP forecasting framework and the concern that one may inadvertently inflate predictability due to overfitting, the following questions are additionally investigated:

1. Can the inflation in predictability be addressed using a pure validation forecast where one part of the data is used to develop all models and the other part is used for application?
2. Can this possible over-inflation be further verified using measures of predictability over space (spatial correlation) in addition to the regular time series correlation performance metric?
3. Can the HLC model architecture be further simplified, given the added predictability the spectrally transformed CMIP6 predictors offer?

This paper is structured as follows. In section 2, we introduce the methods, providing a detailed description of the WASP and the hybrid HLC-WASP models. This section also documents the observed rainfall data, SSTA indices, and CMIP6 models used in this study. Next, in section 3, we present the results of interannual rainfall predictions using the novel HLC-WASP model. Section 4 discusses the findings, and a summary of conclusions and avenues of further work are presented in section 5.

2. Methods and data

This section presents the observational and model data used in this manuscript, along with a detailed description of spectral transformation and the hybrid forecasting model.

a. SSTA indices and rainfall data

The ten climate indices that have been shown to have a significant relationship with Australian rainfall and are considered as potential predictors (Choudhury et al., 2019; Schepen et al., 2012) include Niño 3, Niño 4, Niño 3.4, the El Niño Modoki index (EMI), the Indian Ocean Dipole Mode Index (DMI), the Indian Ocean East Pole Index (EPI), the Indian Ocean West Pole Index (WPI), the Indonesian Index (II), the Tasman Sea Index (TSI), and the Tropical Trans-basin Variability Index (TBV). The climatic events associated with these indices have been shown to significantly affect Australian rainfall and have been

demonstrated skillful for Australian rainfall prediction (Cai et al., 2011; Fierro and Leslie, 2013; Kirono et al., 2010). Further details of these indices are mentioned in Table S1 of the Supplemental Material. While other indices may add predictability in different parts of Australia, the focus of the present study is to capture the impact that spectral transformation can have.

Observed values of these indices are obtained from the SST outputs of the Hadley Centre Global Sea Ice and Sea Surface Temperature (HadISST) data set (Rayner et al., 2003) over the period of 1961–2022 at monthly resolution. Observed gridded monthly rainfall over Australia, referred to as Australian Gridded Climate Data (AGCD), for the period 1961–2022 was obtained from the Australian Water Availability Project (Jones et al., 2009). The data was re-gridded from its original resolution of $0.05^\circ \times 0.05^\circ$ to $2.5^\circ \times 2.5^\circ$ in this study to correspond with the resolution of the climate datasets and model simulations used. The grid layout across Australia is illustrated in Fig. S1 of the Supplemental Material, and two grids used for demonstration in the following section are highlighted in red. Grids with missing data located in the central and western deserts of Australia (Jones et al., 2009) are given in black colour.

b. Model outputs and drift correction

Predictions of these ten SSTA indices were obtained from outputs of decadal hindcast experiments from the CMIP suite of models (Boer et al., 2016). We used five CMIP6 models for our analysis, including CanESM5, MIROC6, EC-Earth3, NorCPM1, and MPI-ESM1.2-HR. These models were chosen because they provide outputs initialized annually and to enable comparison with the previous study using five CMIP5 models (Choudhury et al., 2019). The issue of sampling biases in CMIP decadal forecasts not initialized annually has been previously reported in Choudhury et al. (2016). SSTA indices were calculated over 53 overlapping decades, which commence from 1961 onwards, separated by one-year intervals, 1961–70, 1962–71, ..., and 2013–22. The baseline period for the climatology from 1950 to 1979 was used to derive these anomalies, which were calculated using the same ensembles of decadal and historical simulations of the CMIP6 model. Further details of the CMIP6 global climate models (GCMs) used in the study are mentioned in Table 1.

The issue of biases and drift has thoroughly been documented to plague climate projections, but it is especially relevant for decadal predictions. Given that decadal runs are initialized based on an observational state that is further away from a model's preferred mean state, it is typical for model projections to gradually approach the model state as the effect of initial condition diminishes with lead time (Boer et al., 2016). Concurrently, several methods to correct this drift have been previously reported in the literature (Fućkar et al., 2014; Kharin et al., 2012; Smith et al., 2013). A summary and comparison of these methods have been discussed in Choudhury et al. (2017), and it suggested that no method performs consistently as the best, but initial condition and trend-based drift correction appear as the best for most cases. As a result, we adopt the initial condition and trend-based drift correction for all the indices and models considered in the study. It should be noted that the recommended way to bias correct decadal predictions for CMIP6 model outputs is to remove the lead-time dependent climatology (Boer et al., 2016).

The corrected indices from all CMIP6 models were then combined to form a multi-model ensemble mean (MME). The MME case here represents the scenario where each drift-corrected predictor variable is averaged across all CMIP6 decadal prediction models and then used directly as inputs for rainfall forecasting. Consequently, in addition to five individual CMIP6 models, rainfall forecasts were driven by the MME to assess whether its reduced variability results in greater skill for the forecasts issued.

The predictability of climate indices derived from decadal predictions is fundamental to the multi-year rainfall forecasting. We

Table 1
CMIP6 GCMs used in the study.

Model	Group	Resolution	Ensemble Size
CanESM5	Canadian Centre for Climate Modelling and Analysis (CCCma), Canada.	atmos: 500 km, land: 500 km, ocean: 100 km	10
MIROC6	Atmospheric and Ocean Research Institute (AORI), National Institute for Environmental Studies (NIES), and Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Japan.	atmos: 250 km, land: 250 km, ocean: 100 km	10
EC-Earth3	A European community Earth-System Model, part of a Europe-wide consortium.	atmos: 100 km land: 100 km ocean: 100 km	10
NorCPM1	The Norwegian Climate Prediction Model (NorCPM), Norway.	atmos: 250 km, land: 250 km, ocean: 100 km	10
MPI-ESM1.2-HR	Max Plank Institute of Meteorology (MPI-M), Germany.	atmos: 100 km land: 100 km ocean: 50 km	5

evaluated the forecast skills of ten climate indices derived from five CMIP6 models through temporal correlation analysis. Consistent with the findings of Choudhury et al. (2015), this assessment highlights the potential of these indices for long-term rainfall forecasting. Further details are presented in Fig. S2 of the Supplemental Material.

c. Wavelet System Prediction (WASP)

WASP relies on wavelet theory and accounts for the spectral difference between the response and associated predictors in any system. First, wavelet transforms are used to characterize the spectrum for a given variable, through decomposition of the signal into a range of low-frequency (representing slow changing process) and high-frequency (representing rapidly changing phenomenon) components. The decomposition into different components enables variance modulation of individual components to proceed, with the reconstruction resulting in an altered frequency spectrum for the variable. The transformation is based on the idea that if the spectral variance structure of the predictor is similar to that of the response, the predictive model using that predictor will exhibit better accuracy than otherwise. A summary of the rationale used is presented below, with readers referred to Jiang et al. (2020) for more details.

Given a time series \mathbf{x} , it can be decomposed into sub-time series representing information at different frequencies in the spectrum using a wavelet transform as given by Eq. (1):

$$\mathbf{x} = \sum_{j=1}^J \mathbf{d}_j + \mathbf{a}_J \quad (1)$$

where \mathbf{d}_j is the wavelet details at the decomposition level j and \mathbf{a}_J is the wavelet approximations at the maximum decomposition level of J . The wavelet details and approximation vectors are orthogonal to each other. Eq. (1) can be cast into a matrix form as $\mathbf{x} = \hat{\mathbf{R}}\mathbf{s}$, where $\hat{\mathbf{R}}$ is the standardized matrix of $\mathbf{R} = [\mathbf{d}_1, \dots, \mathbf{d}_J, \mathbf{a}_J]$ and \mathbf{s} is the standard deviation vector of $\mathbf{s} = [\sigma_{d_1}, \dots, \sigma_{d_J}, \sigma_{a_J}]^T$. Using the variance transformation

method, the transformed predictor variable \mathbf{x}' associated with the target response \mathbf{y} is given as follows:

$$\mathbf{x}' = \widehat{\mathbf{R}}\boldsymbol{\alpha}$$

$$\boldsymbol{\alpha} = \sigma_x \widehat{\mathbf{c}}^T \tag{2}$$

$$\mathbf{c} = \frac{1}{n-1} \mathbf{y}^T \widehat{\mathbf{R}} = [s_{y\hat{a}_1}, \dots, s_{y\hat{a}_j}, s_{y\hat{a}_r}] \tag{3}$$

where n is the number of data pairs (x, y) ; $\widehat{\mathbf{c}}$ is the normalized covariance vector of \mathbf{c} between the variable set $(\mathbf{y}, \widehat{\mathbf{R}})$ and s represents the covariance between response Y and wavelet decompositions; σ_x denotes the standard deviation of \mathbf{x} ; the resulting $\boldsymbol{\alpha}$ is the transformed standard deviation matrix of \mathbf{s} . It is noted that the estimation of \mathbf{c} requires the response variable, and thus, in order to transform the predictor variable in the validation (or pseudo future) periods, the estimated \mathbf{c} was derived from the calibration (known current) periods alone.

The optimal prediction accuracy measured by root mean square error (RMSE), when a simple linear regression model is used for prediction, can be derived as (Jiang et al., 2020),

$$\text{RMSE}_{\min} = \sqrt{\frac{n-1}{n} (\sigma_y^2 - \|\mathbf{c}\|^2)} \tag{4}$$

where σ_y denotes the standard deviation of the response y , and $\|\cdot\|$ denotes the norm of a vector in Euclidean space. In this study, the maximum decomposition level adopts the rule of thumb equation, $J = \frac{\log(\frac{\sigma_x}{\sigma_y})}{\log(2)}$ from (Kaiser, 2010), and the Daubechies 1 (i.e., Haar) wavelet filter with vanishing moment $v=1$ was used. Daubechies wavelets are commonly applied in the field of hydro-climatology (Kusumastuti et al., 2022; Maheswaran and Khosa, 2012), and wavelet filters with different lengths have been tested, and the conclusions reported here remain unchanged.

The above summarizes the basis of variance transformation processes in the WASP software described in Jiang et al. (2021a). There are three alternatives of wavelet transforms provided for different applications in the software. For the purpose of the present study, MODWT was adopted since it is appropriate for use in forecasting applications (Jiang et al., 2021b). For brevity, the variance transformation process is represented as per Eq. (5):

$$\mathbf{x}' = g(\mathbf{x}, \mathbf{y}) \tag{5}$$

where $g(\cdot)$ denotes the variance transformation operation by Eqs. (2) and (3). These transformed predictors (\mathbf{x}') then form the inputs to the hierarchical prediction model as detailed below.

d. The hybrid HLC-WASP forecasting framework

The HLC model (Choudhury et al., 2019) forecasts seasonal rainfall across Australia using a time-varying combination of empirical and dynamical forecasts. The empirical forecasts use lagged climate predictors (observed values at the initial time step), while the dynamical forecasts use concurrent predictors (as simulated by the dynamical climate model). The rationale behind the approach is that the value of an empirical rainfall forecasting scheme will diminish as lead time increases. On the other hand, dynamical forecasts, despite potentially lower performance at short lead times, tend to exhibit enhanced accuracy over extended leads, and their skill declines more slowly as the lead time further increases compared to empirical models (Cali Quaglia et al., 2022; Choudhury et al., 2019). This improvement occurs as the impact of poor initialization in the dynamical forecasts diminishes. As a result, one needs to merge the two forecasts in a way that recognizes the relative contribution each makes at any given lead time. In the HLC model, both empirical and dynamical forecasts predicted the response

(rainfall) using a simplified version of the Bayesian Joint Probability-Bayesian Model Averaging (BJP-BMA) method (Schepen and Wang, 2015; Schepen et al., 2014). The BJP-BMA model, common to both forecasts, represents an additive linear modelling scheme and was used operationally for seasonal rainfall prediction by BoM in Australia prior to the transition to the ACCESS-S (Australian Community Climate and Earth-System Simulator – Seasonal) system. The BJP-BMA has been widely reported in the literature and considered to offer highly robust predictions necessary for issuance to water users across the country. Readers are referred to Schepen and Wang (2015) and Schepen et al. (2014) and papers cited within for additional details about this prediction approach. A summary of this hierarchical forecasting framework is presented here, with readers referred to Choudhury et al. (2019) for additional details.

Consider the target response at lead l to be denoted (y_l). The additive regression framework used in BJP-BMA identifies linear regression relationships with individual predictor variables (denoted x^i , with i representing the i 'th predictor variable used) that are then combined to convert them to an additive regression model. Equivalent models are developed for each lead time of interest. However, there exist two sets of additive linear models, representing the dynamical (d) and the empirical (e) modelling approaches described above. The empirical model is based on a lagged relationship between observed climate indices at lead time 0 (x_0^e) and the response at lead time l (y_l^e), while the dynamical models use modelled climate indices at the concurrent lead time l (x_l^d) to estimate the response at lead time l (y_l^d). These two sets of models can be represented by the following equations:

$$y_l^e = G^e(x_0^e) + \varepsilon_l^e \tag{6}$$

$$y_l^d = G^d(x_l^d) + \varepsilon_l^d \tag{7}$$

where $G^e(\cdot)$ and $G^d(\cdot)$ represent the additive regression model, defined as follows:

$$G^e(x_0^e) = \sum_{i=1}^m [w_i^e (\beta_1^i x_0^{e,i} + \beta_0^{e,i})]$$

$$G^d(x_l^d) = \sum_{i=1}^m [w_i^d (\beta_1^i x_l^{d,i} + \beta_0^{d,i})] \tag{8}$$

where β_1^i and β_0^i are the associated regression coefficients of the associated predictor, x^i , and m is the total number of individual simple linear regression models that form the additive model, corresponding to the number of predictors used in the regression model. w_i is the weight for each individual regression model and estimated from residual error-covariance. In each linear regression model, the residual error term ε is estimated independently for each lead time l . Finally, the best estimate of the response can be obtained by combining the two sets of modeled responses via Eq. (9):

$$\widehat{y}_l = w_l y_l^e + (1 - w_l) y_l^d \tag{9}$$

where $w_l = \frac{S_1 - S_{12}}{S_1 + S_2 - S_{12}}$ represents the weight of the empirical model while $(1 - w_l)$ is the weight associated with the dynamical model, and $\mathbf{S} =$

$\text{cov}(\varepsilon_l^e, \varepsilon_l^d) = \begin{bmatrix} S_1 & S_{12} \\ S_{12} & S_2 \end{bmatrix}$ is the covariance of residual errors from each regression model. The HLC approach inherently mitigates collinearity by selectively identifying only the most relevant predictors. This process, which uses combination weights derived from error-covariance, ensures that the selected predictors contribute to improving the model's accuracy without introducing multicollinearity. The usage of error-covariance to derive combination weights has been proposed and investigated in many previous studies (Bates and Granger, 1969;

Choudhury et al., 2019; Khan et al., 2014; Kim et al., 2022; Rashid et al., 2020).

Unlike the previous study by Choudhury et al. (2019), our hybrid HLC-WASP model incorporates a spectral transformation for each predictor variable that constitutes the HLC model. As introduced in section 2c, WASP transforms the predictor variable to an altered frequency spectrum that improves its prediction accuracy with respect to the response. Consequently, the HLC modelling framework in Equations (6) – (9) now uses transformed predictor variables $g(x_0^{e,1\dots n})$ or $g(x_t^{d,1\dots n})$ instead of just using $x_0^{e,1\dots n}$ and $x_t^{d,1\dots n}$, where $g(\cdot)$ represents the WASP variance transformation given in Eq. (5). As individual predictor variables are used in the additive regression framework adopted in HLC, the WASP transformation amounts to a decomposition of the predictor variable into additive components representing different frequencies using wavelet transformation (Percival and Walden, 2000). Next, these different components are re-weighted so as to result in a frequency spectrum that is analogous to the target response. For instance, if the target response exhibits variability at only one frequency in a hypothetical setting, the decomposed predictor variable will be re-assembled by imparting higher variability to that frequency alone, with other components (representing other frequencies in the spectrum) de-weighted. The inverse wavelet transformation of this re-weighted decomposed variable will mimic the power spectrum of the target response, resulting in an improved predictive relationship. Readers are referred to implementational details on WASP in Jiang et al. (2021a).

To assess the performance of the hybrid HLC-WASP model, first, we carried out the split sample validation approach where the first 26 decades were used for calibration and the remaining 27 decades were used for validation, allowing for an evaluation of the hybrid model in a practical setting. With this validation, the overlap only occurs at the end of calibration and the beginning of the validation, which is regarded as a strict blind validation. We also carried out a double leave-one-out cross-

validation (LOOCV), following the methodology of Choudhury et al. (2019), and results were consistent with the split sample validation and are available upon request. Additionally, since the HLC model consists of both empirical and dynamical components, it is important to assess the performance of each individual component and determine if the model is over-parameterized. Furthermore, the implications of spectral transformation on each component of the HLC need careful assessment. The schematic of the complete framework adopted is presented in Fig. 1.

3. Application of the hybrid HLC-WASP forecasting system

This section presents the application of the HLC-WASP model to interannual rainfall forecasting over Australia. Firstly, the implications of spectral transformation for climate indices derived from CMIP6 decadal predictions are shown. Next, rainfall forecasts using the hybrid HLC-WASP model are evaluated under the split sample validation approach. Finally, the effects of spectral transformation on the empirical, dynamical, and combined HLC models are comprehensively investigated. The model produces monthly rainfall forecasts, which are aggregated at an annual scale for the presentation of the reported metrics in the subsequent sections. While the aggregation process reduces the focus on detailed month-to-month variations, the primary goal for long-term forecasts spanning several years is to estimate total water availability.

a. Spectral transformation of climate indices corresponding to gridded rainfall.

Climate indices are transformed into a new set of indices designed to mimic the frequency spectrum of the target rainfall at each grid point using the WASP. Essentially, this leads to independent sets of transformed SSTA indices per grid point. Hence, the predictor variable used in the HLC is unique to each grid cell, instead of the logic in Choudhury

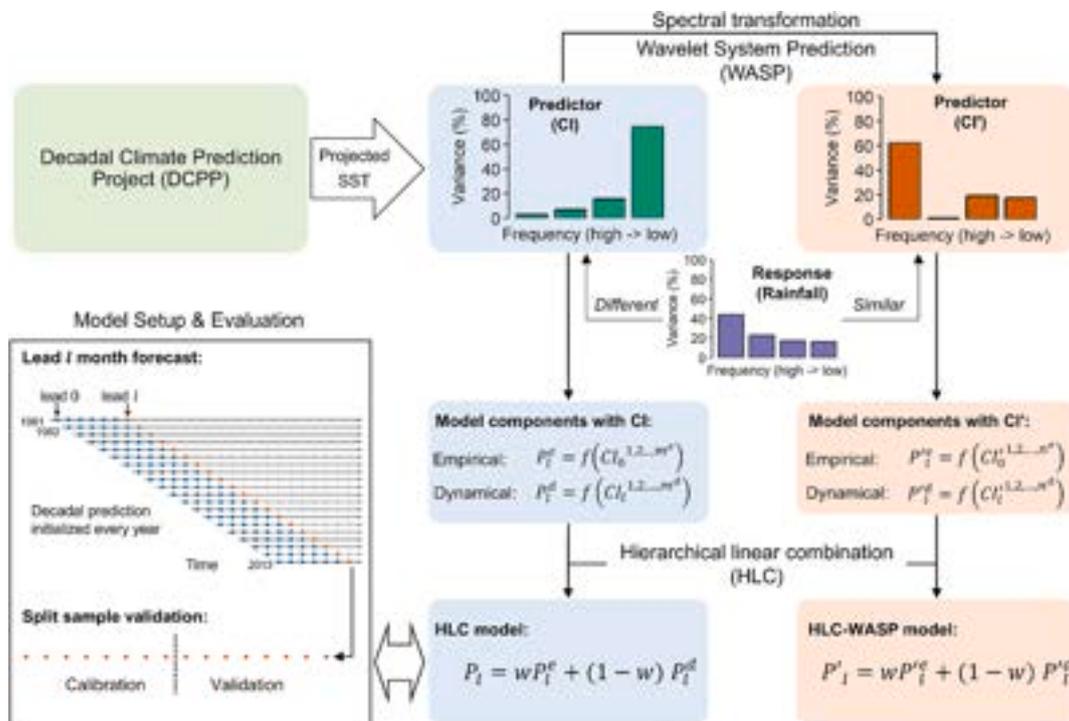


Fig. 1. Schematic representation of the modelling framework for interannual rainfall forecasts using the hybrid HLC-WASP model. The framework consists of three main steps: (1) derivation of a climate index (CI) related to Australian rainfall from the decadal predictions of sea surface temperature (SST); (2) spectral transformation of the CI into a response spectrum-adjusted index (CI') using WASP; and (3) development of a rainfall forecasting model using HLC logic. This model integrates an empirical component, based on lagged CI values at a lead time of 0 months (CI_0), with a dynamical component that uses current decadal predictions (CI_t). The black box illustrates the schematic diagram of the model's setup and evaluation at a given lead time *l*.

et al. (2019) where all predictors stayed the same across the entire Australian region, with only the additive combination weights changing regionally. Fig. 2 presents an example of this transformation for 2 grid points in northern Queensland (Grid 24) and southern Western Australia (Grid 89), refer to Fig. S1 of the Supplemental Material for grid point numbers, with very different climatic conditions. They are selected due to their remarkably different spectrum properties (blue bars in Fig. 2c and d), the rainfall at Grid 24 (Fig. 2a) being higher and having a stronger periodicity at sub-seasonal to seasonal cycle than Grid 89 (Fig. 2b). Fig. 2c and d present the variance distribution of the Niño 3 index before and after spectrally transforming as per the rainfall at grids 24 and 89. The percentage of variance distribution for the original index (teal bars) gradually increases from high frequency to low frequency, and after the variance transformation, the spectral properties of the predictors become closer to mimicking that of the response (local rainfall). Note that wavelet transform decomposes time series into separate rapidly changing sub-time series corresponding to high-frequency levels (i.e., details, d1, d2, and d3) and slowly changing, low-frequency levels

(i.e., approximations, a3) of the spectrum.

It is evident that both the original and the transformed Niño 3 corresponding to rainfall at Grid 24 exhibit greater variability in the lower frequency (a3). However, the medium frequency bands (d1 to d3) are enhanced on transforming in comparison, illustrating the effect of the variance transformation procedure used (orange bars in Fig. 2c). In contrast, when transforming with respect to rainfall at Grid 89, the variance at the high-frequency level (d1) is greatly increased while the variance at the low-frequency level (a3) is reduced (orange bars in Fig. 2d). This allows the transformation to create an independent estimate of the predictor variable that is closest in representation to the target response in its spectrum (purple bars). Fig. 2e and f show the Niño 3 index transformed as per rainfall time series at Grids 24 and 89, respectively. Compared to the original time series, transforming as per rainfall at Grid 24 leads to relatively smaller changes, with only slight alterations in variance distribution in the frequency domain (Fig. 2c and e). However, transforming as per rainfall at Grid 89 leads to much larger changes (Fig. 2f), especially with much higher variability at shorter

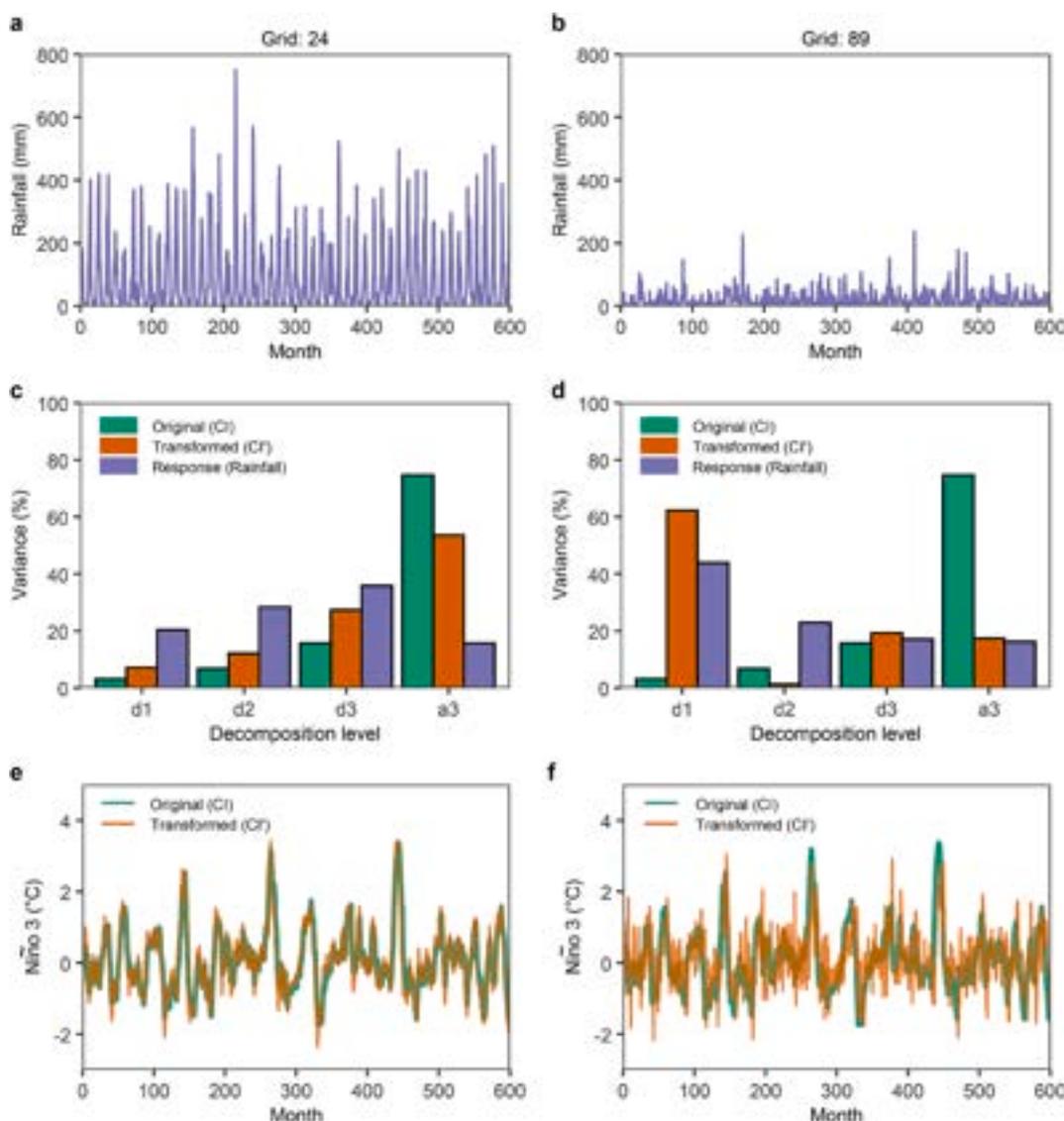


Fig. 2. Illustration of transformed and non-transformed climate indices, taking Niño 3 as an example, corresponding to rainfall at two selected grids using observed data from 1961 to 2010: (a) and (b) represent observed rainfall at northern Queensland (Grid 24) and southern Western Australia (Grid 89), respectively; (c) and (d) show the variance distribution of a predictor, Niño 3, and its target response, rainfall, by percentage in the frequency domain before (teal) and after (orange) transformation as per rainfall at Grids 24 and 89, respectively; (e) and (f) present original and transformed Niño 3 in the time domain linking to the change in the frequency domain in (c) and (d), respectively. It must be noted that the same predictor variable (Niño 3) has been spectrally transformed into locally representative predictors (e and f), which enhances the predictability of the local response variables (a and b).

timescales (Fig. 2d).

We have shown the implication of WASP on individual climate index above, and under the forecasting framework, the implication of WASP on the linear relationship between each climate index and rainfall at different lead times is illustrated in Fig. 3. Taking the lead 12 month of the Grid 89 for example, an improved relationship can be found in all of the empirical models using transformed climate indices except for the WPI, and a new relationship between climate index and rainfall (i.e., change sign of regression coefficient) is established in most of the dynamical model using transformed climate indices except for the DMI, WPI, and TSI. The new relationship has been established due to the negative covariance in c , defined in Equation (3), that identifies the frequency component negatively correlated with the local rainfall of interest. Fig. 4 presents the corresponding c values used to modify the original climate indices as shown in Fig. 3. It is evident that, compared to

DMI, WPI, and TSI, other dynamical models exhibit a higher incidence of negative covariance in concurrent climate indices.

b. Assessing the hybrid HLC-WASP model under the split sample validation

In the hybrid HLC-WASP forecasting framework, combination weights for the individual linear regression models based on their error covariance, given in Equation (9), and the resulting predictability, differ from using the HLC approach. Fig. 5 presents predictor variables with the most significant weight among all the linear regression models in the empirical and dynamical model components of the HLC and HLC-WASP framework. It has been noted that the predictability horizon of climate indices such as EMI and WPI extends up to 24 months in several decadal prediction studies, while TBV shows forecast skill up to 36 months. On the other hand, climate indices like Niño 3, Niño 4, and Niño 3.4

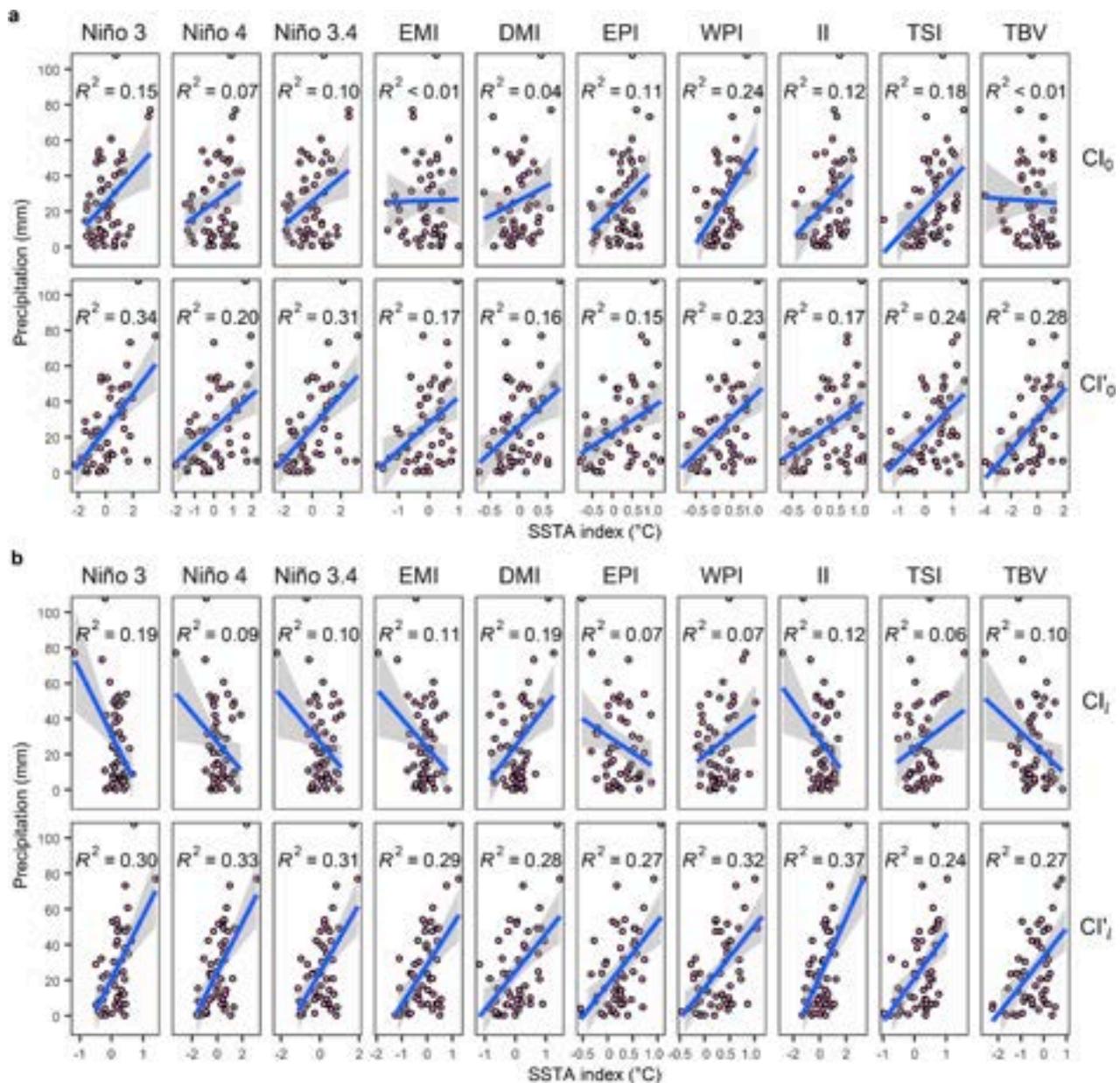


Fig. 3. Illustration of an improved relationship between each climate index and the rainfall at lead 12 months of Grid 89. (a): Climate indices used in empirical model; (b): Climate indices from MIROC6 decadal predictions used in the dynamical model. The first row represents the relationship using non-transformed climate indices (Cl_0 and Cl_1), while the second row shows the improved relationship using transformed climate indices (Cl'_0 and Cl'_1). The blue lines represent linear regression between the SSTA indices and precipitation, with the 95 % confidence interval shown by grey bands and the coefficient of determination (R^2) indicated by black text.

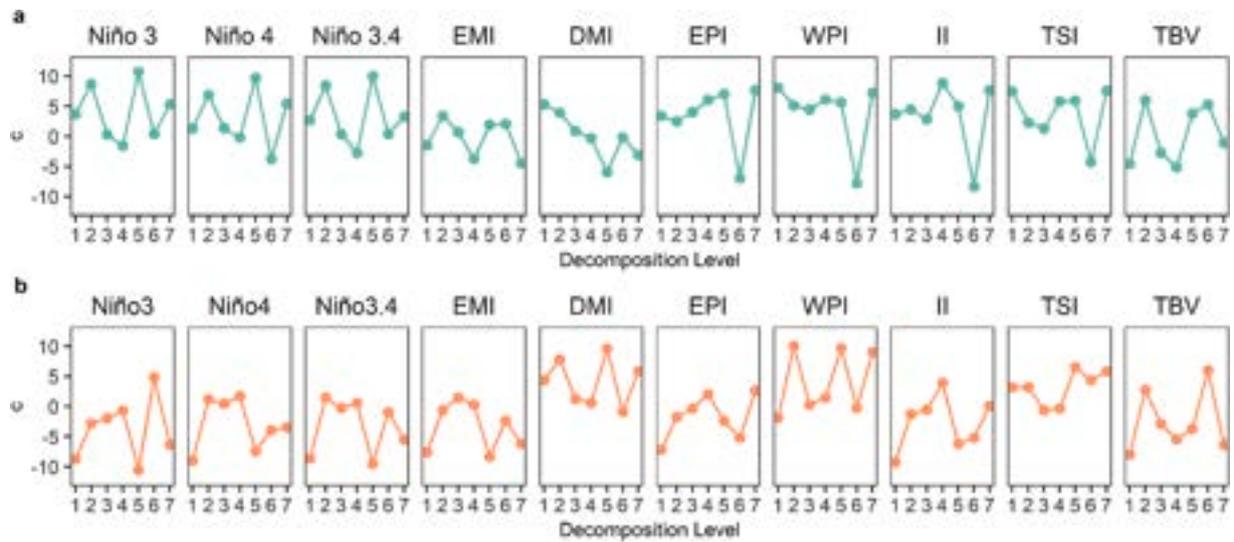


Fig. 4. Transforming covariance c corresponding to Fig. 3. (a): Covariance c used in the empirical model; (b): Covariance c used in the dynamical model.

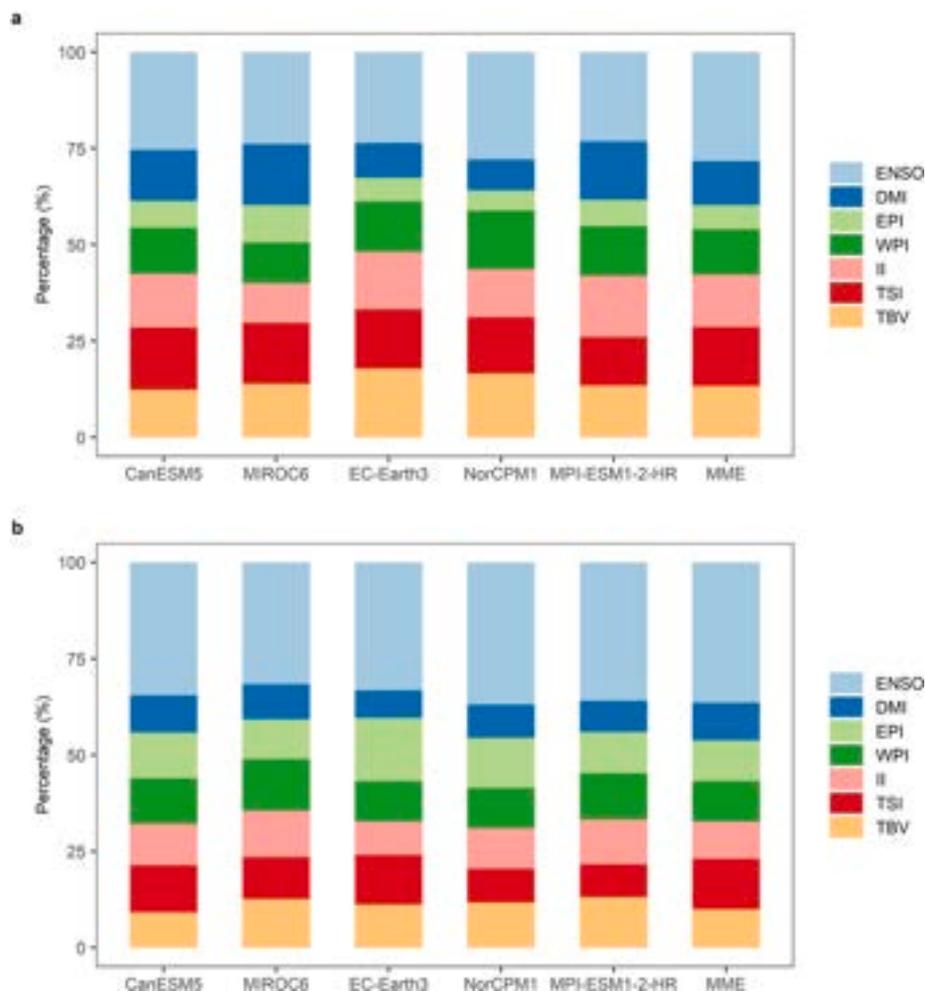


Fig. 5. Percentage of grid cells with the highest weight for each predictor, averaged over lead 1 to 5 years, for (a) the HLC model and (b) the HLC-WASP model. ENSO represents the combined percentage contribution of Niño 3, Niño 4, Niño 3.4, and EMI.

typically exhibit predictability horizons less than 12 months (Chikamoto et al., 2015; Choudhury et al., 2019; Choudhury et al., 2016). Additionally, indices such as DMI, EPI, II, and TSI generally show the poorest predictability among the ten climate indices (Choudhury et al., 2015).

Note that the predictability horizon of an index and model is defined as the lead time in which the model error exceeds 5% of the random model error distribution, indicating the limit of predictability for the CI. The results in Fig. 5 show the averaged percentage of grid cells over lead 1 to

5 years that exhibit the highest weight for each of the predictor variables considered. For both HLC and HLC-WASP models, ENSO indices (including Niño 3, Niño 4, Niño 3.4, and EMI), TSI, and TBV account for a larger percentage across all the grid cells considered. However, there is a larger proportion of ENSO indices in the HLC-WASP model, while the HLC model has a higher percentage of WPI and TSI. It suggests that WASP is capable of identifying the physically meaningful predictors, such as ENSO indices for Australian rainfall. The percentage of grid cells varying with lead time can be found in Fig. S3 of the Supplemental Material. Investigating the difference across all five individual CMIP6 models, although CIs are substantially different from each other, their relative importance to rainfall forecasting is similar. After averaging the CIs based on all five CMIP6 models, the MME can still capture the underlying signal of climate variability, and CIs with a long predictability horizon are more frequently selected at long-lead times, while ENSO indices with a short predictability horizon are often selected at short-lead times.

It is worth reiterating that the HLC-WASP model consists of two components, representing an empirical and a dynamical model. The empirical model uses lagged relationships established between observed CI at lead time 0 and the response at lead l , and uses the lagged observed SSTA values to predict the response. The dynamical model uses concurrent SSTA simulations (not observations) to forecast the response at lead l . Fig. 6 shows that, across the CMIP6 climate models investigated, there is generally a decreasing trend in the dynamical model weights in the HLC-WASP framework, while increasing weights are found for the empirical model. On the other hand, empirical forecasts have a relatively lower weight at short lead times, but as lead time increases, their weight gradually increases as the skill of empirical forecasts gets better than the dynamical model. This is different from the previous study

(Choudhury et al., 2019) as now the dynamical model component considers all ten SSTA indices. In addition, compared to the weights of the HLC model (Fig. 6a), there is no major difference in trend between weights derived from the HLC and HLC-WASP models, but the HLC-WASP model has substantially higher dynamical weights than that of the HLC model, which can be attributed to the fact that the transformation performed by WASP on the concurrent SSTA simulations improves the performance of the dynamical component. The MME, which averages the drift-corrected predictor variables from all models, shows similar variations in weight between the empirical and dynamical components in both the HLC and HLC-WASP models. The HLC-WASP model places a higher weight on the dynamical component compared to the HLC model, with lead-time averaged weights of 0.55 and 0.64, respectively. In contrast, the averaged weights from the empirical model are 0.45 and 0.36 for the HLC and HLC-WASP models, respectively.

With estimated weights of empirical and dynamical components, the forecasting skills of each CMIP6 climate model averaged over Australia using both HLC and HLC-WASP models are shown in Fig. 7. In addition to the two proposed models, a baseline model is represented by a solid teal line, which utilizes lagged rainfall at lead time 0 as the predictor for rainfall forecasting. This model is considered as a reference to assess the forecasting skills of HLC and HLC-WASP models. In terms of temporal correlation, our analysis shows that except for the HLC model at lead 1 year, the forecasting skills of both proposed models surpass that of the baseline model, confirming the authenticity and effectiveness of the forecast skills demonstrated by the HLC and HLC-WASP models. For RMSE, the baseline model exhibits consistently lower values compared to the HLC model because this persistence-based model essentially predicts the mean rainfall, which inherently minimizes variance but does not capture rainfall variability. The solid line presents the MME

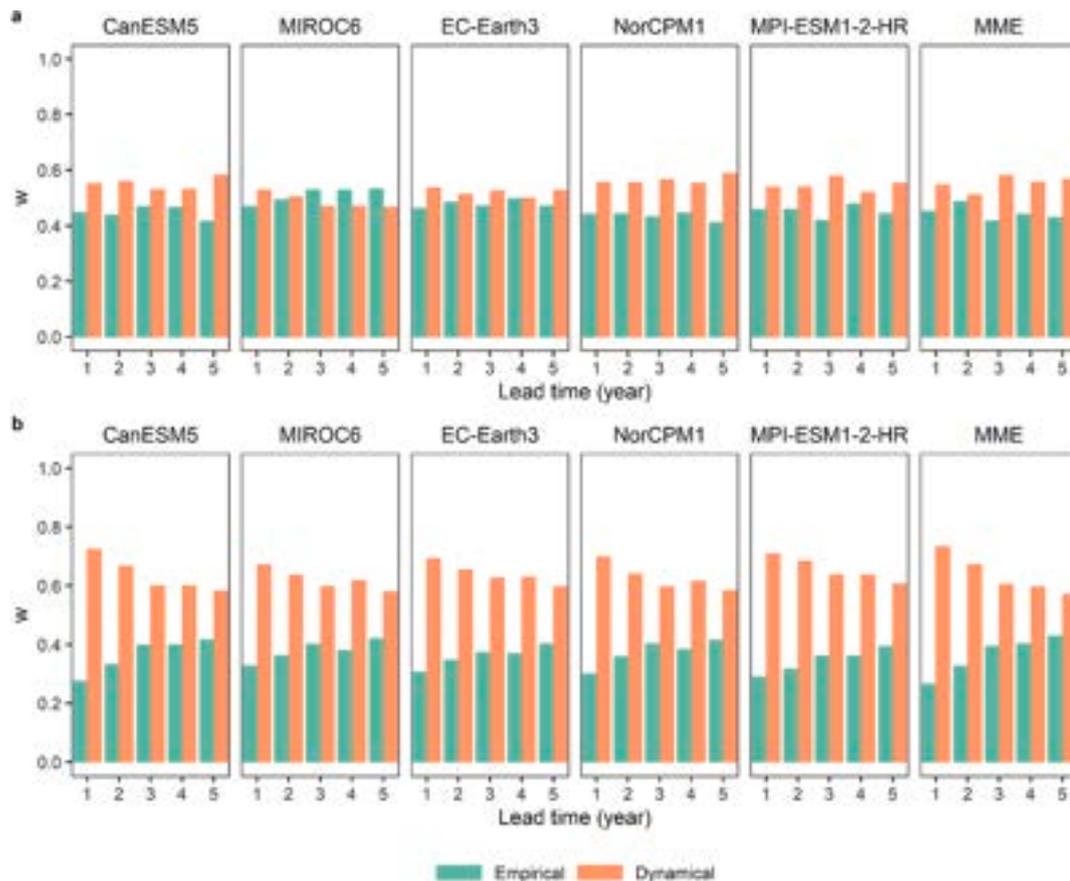


Fig. 6. Empirical and dynamical model component weights averaged over Australia from the individual CMIP6 models and their MME using (a) the HLC model and (b) the HLC-WASP model.

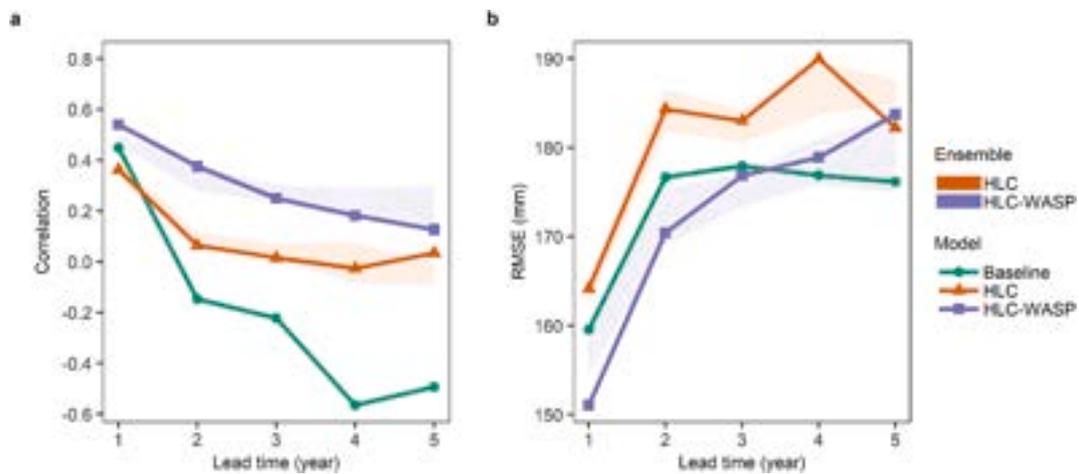


Fig. 7. Forecast skills of rainfall from both HLC and HLC-WASP frameworks, including (a) Correlation and (b) RMSE of the predicted rainfall averaged across all grids over Australia. Results are shown for the baseline model (teal), and the HLC (orange) and HLC-WASP (purple) models using decadal outputs from the multi-model ensemble mean (MME). The shaded area is the uncertainty band from the ensemble of CMIP6 models.

while the shaded area shows the uncertainty band from the ensemble of CMIP6 models. Please note the MME case here represents the scenario where drift-corrected predictor variables are averaged and then either used directly as inputs (in HLC) or spectrally transformed and used as inputs (in HLC-WASP). The corresponding results from the individual CMIP6 models are given in Fig. S4 of the Supplemental Material. As shown in Fig. 7, the HLC-WASP model has a consistently higher correlation range from over 0.5 to around 0.1, and a lower RMSE from around 150 to 180 mm, compared to the HLC model. Both the prediction models show similar highest skills at the earliest leads that gradually deteriorate with lead time. This is expected because the predictability skill in simulated SSTA indices from decadal predictions decreases and the linear relationship between observations at lead time 0 and forecast at lead time l decays gradually.

This improvement in skill between HLC-WASP and HLC is mostly noticed over the lead times of 2–3 years. This is because transforming

decadal projections at longer lead times notably increases their value for rainfall prediction compared to using untransformed values. Beyond lead time 3-year forecast skills in terms of both correlation and RMSE for both models decays and smaller improvements are observed. Compared to the individual CMIP6 models, the performance of the MME is mostly within the variability of the ensemble of CMIP6 models, but there is an improvement in skill towards the lead times of 1–2 years. This is probably because the MME can improve the reliability of SSTA predictions and thus the derived CIs when individual CMIP6 model has higher skills at short lead times. In contrast, at long lead times, the reduced variability by averaging SSTA indices from all CMIP models results in worse skill. We also examined using projections from the previous generation (CMIP5) of decadal experiments (similar to HLC results in Choudhury et al., 2019), and the resulting prediction skills were consistent across the two sets of inputs.

To further assess the performance of the HLC-WASP model from the

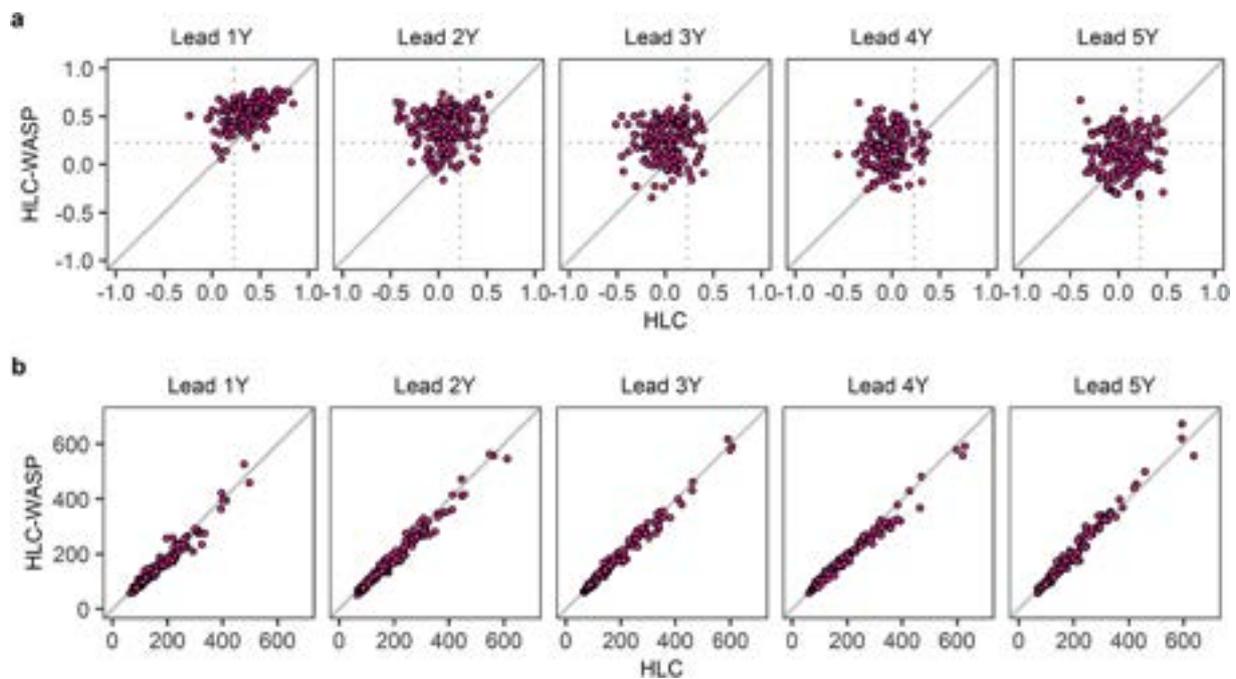


Fig. 8. Scatter plots of (a) Correlation, where dashed lines represent the critical value for correlation significantly different from zero (0.228, corresponding to the 51 degrees of freedom, as we have 53 decades in total); (b) RMSE of predicted rainfall (mm) at each grid point, comparing HLC-WASP and HLC models for different lead times using inputs of SSTA indices from the MME.

MME, scatter plots are used to compare the correlation and RMSE between the HLC-WASP and the HLC model across all five lead times. In Fig. 8, the correlation and RMSE of the predicted rainfall using HLC and HLC-WASP models from each grid and lead time are compared. The performance of the HLC-WASP model is better compared to the HLC model across various lead times. As lead time increases, the HLC-WASP model results in a higher percentage of improved correlations as compared to the HLC model with 84.7%, 85.3%, 78.3%, 73.9%, and 58.0% for lead 1 to 5 years, respectively (Fig. 8a). In terms of RMSE, the associated comparison between HLC-WASP and HLC model is shown in Fig. 8b. Among the various lead years, while the improvements across different grids are small, the HLC-WASP model using transformed predictors has a lower RMSE across most grids (additional assessment in Fig. 9b). The improvement in lead 1–2 years is the most evident, with 81.5% and 82.8% of grids demonstrating reduced RMSE, respectively. The range of RMSE is large, from around 100 mm to over 600 mm, when considering the entire Australia, which covers different climatic regions. The above discusses the results from the MME while the corresponding results from the individual CMIP6 models are given in Fig. S5 of the Supplemental Material. In summary, the benefits of using transformed indices are consistent across both individual CMIP6 models and the MME, particularly for lead times up to 2 years, where the dynamical model component has the highest forecast skills and contributes the most weight in the combined model.

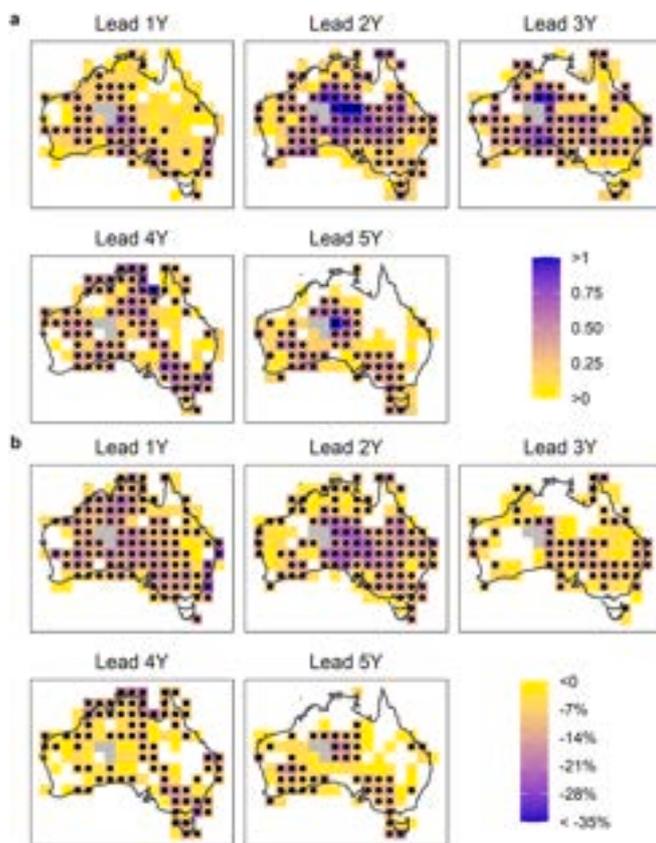


Fig. 9. Difference between HLC-WASP and HLC models over Australia using SSTA indices from the MME: (a) Difference in correlation ($r_{\text{HLC-WASP}} - r_{\text{HLC}}$); (b) Relative change in RMSE by percentage ($\frac{RMSE_{\text{HLC-WASP}} - RMSE_{\text{HLC}}}{RMSE_{\text{HLC}}} \times 100\%$). Compared to the HLC model, the yellow colour region shows minor improvements; blue colour regions present large improvements, while regions with white colour represent the area with no improvement. Grey grids in the central and western deserts of Australia are grids with missing data. The black dot marks the significant value at a 90 % significance level with a difference in correlation greater than 0.214 and a relative change in RMSE less than -5.42 %.

Fig. 9 assesses the performance of the HLC-WASP model against the HLC model from a spatial perspective to better understand where improvements are most prominent. At short lead years, improvements are large, while the degree of improvement decreases with lead time. Compared to the HLC model, the yellow colour region shows minor improvements; blue colour regions present large improvements, while regions with white colour represent the area with no improvement. Grey grids in the central and western deserts of Australia are grids with missing data, as seen in Fig. S1 of the Supplemental Material. The regions with larger improvements are located in southeast Australia. Compared to RMSE, the enhancements in forecasting accuracy are similar for a 2-year lead time when assessing both correlation and RMSE. However, for lead times of 3–5 years, the improvements are more pronounced in correlation metrics, whereas at a 1-year lead time, the improvements are notably more evident in RMSE. As shown in Table 2, the percentage of grids that show improvements in RMSE (with significant improvements at 90% significance in parentheses) are 81.5% (66.2%) and 82.8% (61.8%) for lead 1 and 2 year, with the equivalent numbers for correlation being 84.7% (36.3%) and 85.4% (58.6%) for lead 1 and 2 years, respectively. Note that only a small percentage of the significant improvements in correlation are observed at lead 1 year because both models had reasonably good skills at this lead time. At lead 3, 4, and 5 years, these improvements become 65.6% (37.6%), 71.9% (47.8%), and 53.5% (28.7%) using RMSE, and 78.3% (50.3%), 73.9% (51.6%), and 58.0% (32.5%) using correlation. At the 90% significance level, improvements are considered statistically significant if the difference in correlation ($r_{\text{HLC-WASP}} - r_{\text{HLC}}$) is greater than 0.214, and the relative change in RMSE ($\frac{RMSE_{\text{HLC-WASP}} - RMSE_{\text{HLC}}}{RMSE_{\text{HLC}}} \times 100\%$) is less than -5.42%. In short, the HLC-WASP model provides significant benefits compared to the HLC model, as over 50% of the grids show improved correlation and RMSE across all lead times. This improvement is highlighted in Fig. 7, which presents the results from the MME. Additionally, Fig. S6 in the Supplemental Material presents results from the individual CMIP6 models, highlighting the differences between the HLC-WASP and HLC models.

Furthermore, to evaluate the extent of performance enhancement across Australian grids, a one-sided paired *t*-test is conducted at a 95% significance level. This test assesses whether HLC-WASP outperforms HLC from the MME in terms of improved metrics – specifically, higher correlation and lower RMSE. The results help quantify the proportion of grids demonstrating significant advancements in forecast accuracy. As shown in Table 2, the improvement in correlation of the hybrid HLC-WASP model compared to the HLC model is significant across all lead years, while the RMSE of the HLC-WASP model is significantly lower than the HLC model across all lead times. The paired *t*-test *p*-values for correlations in the MME are less than 0.05, indicating significant differences between the HLC and HLC-WASP models. Similarly, the *p*-values for RMSE are also less than 0.05 across all lead times, except for the lead 5 year. The associated results across all the climate models investigated are provided in Table S2 of the Supplemental Material. In addition, to account for spatial dependence between grid cells, we applied a modified *t*-test (Dutilleul et al., 1993). The results confirm that, except for correlations at long lead times, the conclusion remains consistent: the HLC-WASP model demonstrates statistically significant superiority over the HLC model. Further details can be found in Table S3 of the Supplemental Material.

c. Evaluating individual components of the HLC-WASP model

Here, we aim to further evaluate the hybrid HLC-WASP model by separately investigating its empirical and dynamical components. Our observation that the dynamical component contributed the most to the model raises the question of whether the combined model is over-parameterized. To address this, we included forecasts using the empirical and dynamical models individually, in addition to the combined

Table 2

Percentage of grids that show improvements in correlation and RMSE, and paired *t*-test results for HLC-WASP exhibiting improved performance (with respect to correlation and RMSE) than HLC from the MME.

Lead time	Percentage of improvements in correlation (%)		Percentage of improvements in RMSE (%)		p-value of paired <i>t</i> -test	
	Improved	Significantly improved	Improved	Significantly improved	Correlation	RMSE
Lead 1Y	84.7 %	36.3 %	81.5 %	66.2 %	1.77E-25	4.58E-15
Lead 2Y	85.4 %	58.6 %	82.8 %	61.8 %	4.75E-29	7.10E-20
Lead 3Y	78.3 %	50.3 %	65.6 %	37.6 %	2.37E-19	9.05E-07
Lead 4Y	73.9 %	51.6 %	72.0 %	47.8 %	2.03E-19	7.11E-13
Lead 5Y	58.0 %	32.5 %	53.5 %	28.7 %	1.77E-04	8.80E-01

model, for comparison.

Fig. 10a presents the results of three different modelling alternatives, namely empirical, dynamical, and combined models, and compares the performance of the HLC and the hybrid HLC-WASP frameworks. Notably, the empirical model solely utilizes observed CIs at lead 0 to forecast rainfall at lead *l*, resulting in no shaded area representing the outcomes from different GCMs. The results demonstrate substantial improvements in both RMSE and correlation across lead times, while the

empirical model exhibits an equivalent forecasting skill between the HLC and HLC-WASP methods. This suggests that WASP can perform better when transforming concurrent CIs used in dynamical models but has limited influence when applied to predictors of CIs at lead 0 used in empirical models. Moreover, the MME displays higher skills at lead 1–2 years compared to individual CMIP6 models.

Furthermore, Fig. 10b presents the outcomes of the split sample validation, the dynamical model has similar or better skill than the

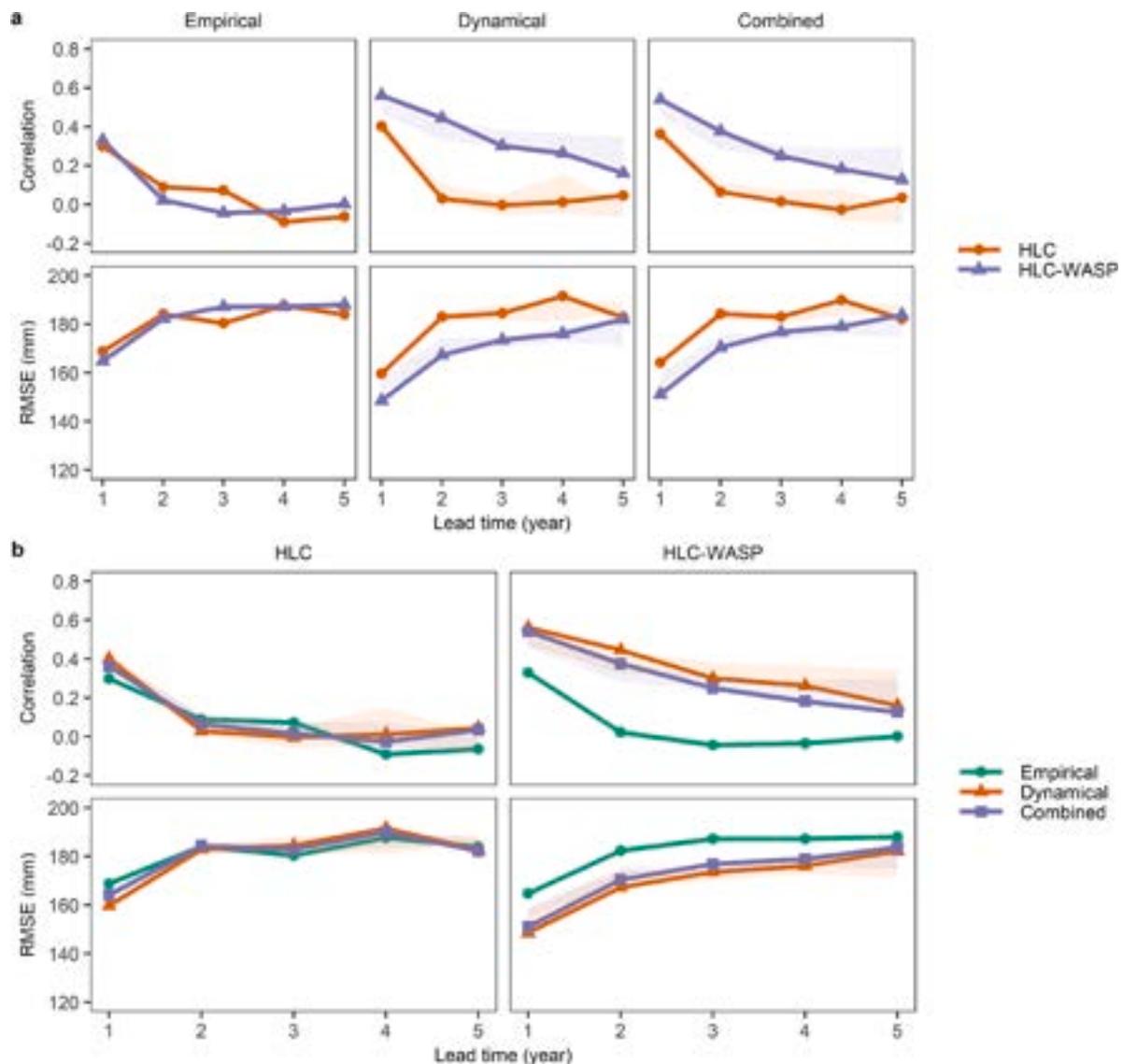


Fig. 10. Forecast skills of empirical, dynamical, and combined models using both HLC and HLC-WASP frameworks under the split sample validation. Subplot (a) compares the performance of the HLC and the hybrid HLC-WASP forecasting framework using three different modelling alternatives. Subplot (b) compares the forecast skills of three models using the HLC and HLC-WASP frameworks. The shaded area is the uncertainty band from the ensemble of CMIP6 models, and the solid lines within the shaded area represent the results from their MME.

combined model, and the empirical model exhibits the worst skill under both HLC and HLC-WASP frameworks. The results indicate that the dynamical model component contributes more to the forecast skill, with higher weights across all lead times, as shown in Fig. 6. In contrast, the calibrated results show that the combined model is the best among the three alternatives. In conclusion, we confirm that the combined model is over-parameterized, and using the dynamical component alone can serve as the optimal model for rainfall forecasts. Note again that these results represent performance in validation, and the corresponding calibration results are available in Fig. S7 of the Supplemental Material.

So far, we have evaluated the performance of the hybrid HLC-WASP model and found that it outperforms other alternatives for interannual rainfall forecasting up to five years. Integrating spectral information into the HLC modelling system proved to be beneficial. We also found that the combined model may not necessarily be the best option and that using the dynamical model alone can be sufficient. Given our findings, we focus on the best-performing model (HLC-WASP with the dynamical component) to investigate Australian rainfall anomalies recorded in recent years, influenced in part by ENSO (Doi et al., 2020; Freund et al., 2021; Ma et al., 2023). We use MIROC6 as it is the dynamical model, as, at the time of writing, recent decadal hindcasts up to 2021 were not available from the other modelling groups that form the earlier set of results.

Fig. 11 demonstrates the performance of the hybrid HLC-WASP model from MIROC6 in predicting rainfall over Australia. The results show that the model not only has a higher temporal correlation (Fig. 10) but also a higher spatial correlation when compared to observed rainfall, as displayed in Fig. 11a. Furthermore, Fig. 11b shows the spatial correlation of predicted rainfall against year averaged across all grids from the hybrid HLC-WASP model. The results suggest that the forecasts for lead 1 year have the best spatial correlation, particularly during El Niño

and La Niña events when Australian rainfall exhibits high variability, as indicated by the shadow red and blue colours, respectively. However, it should be noted that during ENSO events, there may be a lower spatial correlation at longer lead times due to the diminishing skill of SSTA forecasts in predicting such rainfall anomalies.

Additionally, we specifically examined the recent years' rainfall anomalies at lead 1 year during the out-of-sample period spanning from 2018 to 2021, which are depicted by the grey shadow areas in Fig. 11. The forecasted rainfall during this period has a spatial correlation ranging from 0.3 to 0.6, covering a range of extreme dry to wet years. Fig. 12 clearly depicts the substantial rainfall anomalies observed during this period compared to the past decade's climatology. The dynamical model under the HLC-WASP framework was able to capture these patterns over the years.

4. Discussion

In the present section, we address the concerns raised earlier about the possible over-inflation of predictive accuracy due to overfitting. Specifically, we revisit three questions we had raised earlier and provide our answers based on our findings. The implications of our results and what they suggest about the use of spectral transformation and the hybrid HLC-WASP model for interannual rainfall forecasting in Australia is discussed as well. We reflect below on answers to the three questions we had raised earlier in the paper:

1. Can the above question be addressed using a pure validation forecast where one part of the data is used to develop all models and the other part is used for application? The pure validation results in Fig. 10 confirm the superiority of the HLC-WASP over the non-spectrally

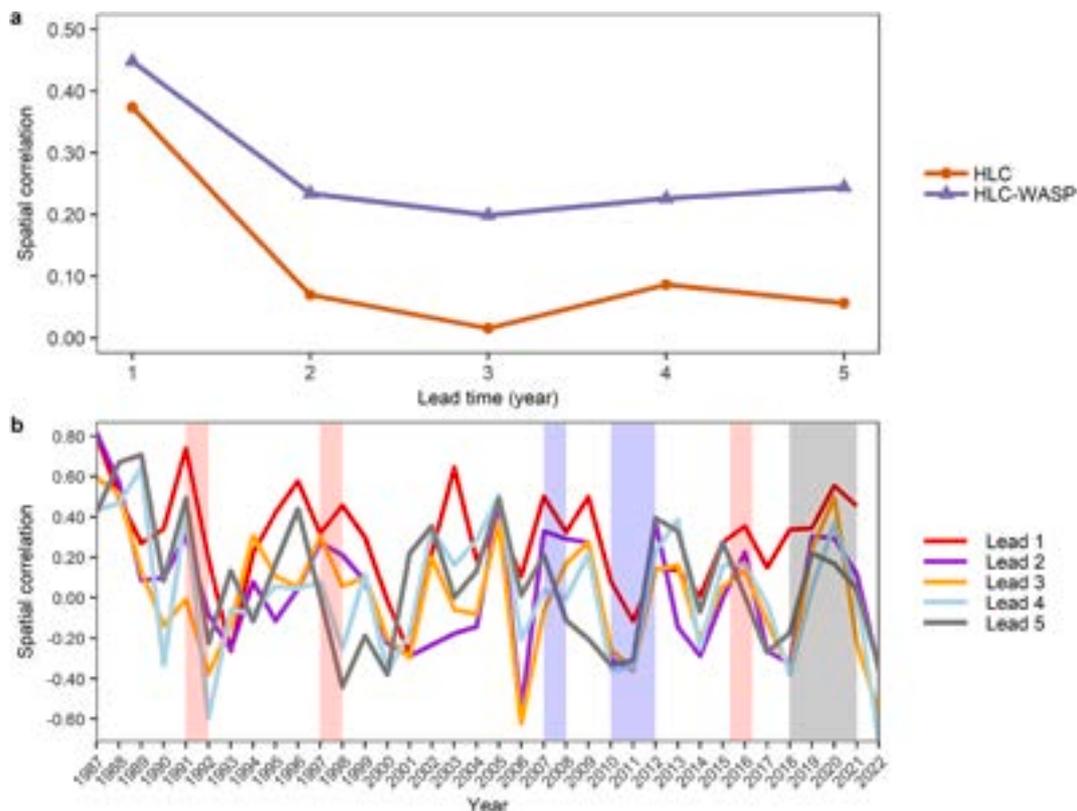


Fig. 11. Forecast skills over space (spatial correlation) under the split sample validation: (a) Spatial correlation of predicted rainfall (mm) against lead time averaged across all grids, comparing HLC-WASP to HLC using inputs of SSTA indices. (b) Spatial correlation of predicted rainfall against year averaged across all grids, using spectrally transformed SSTA indices (HLC-WASP model). The red and blue shadow areas indicate the Australian rainfall during El Niño and La Niña events, respectively. The grey shadow areas represent recent years (2018–2021) with strong rainfall anomalies.

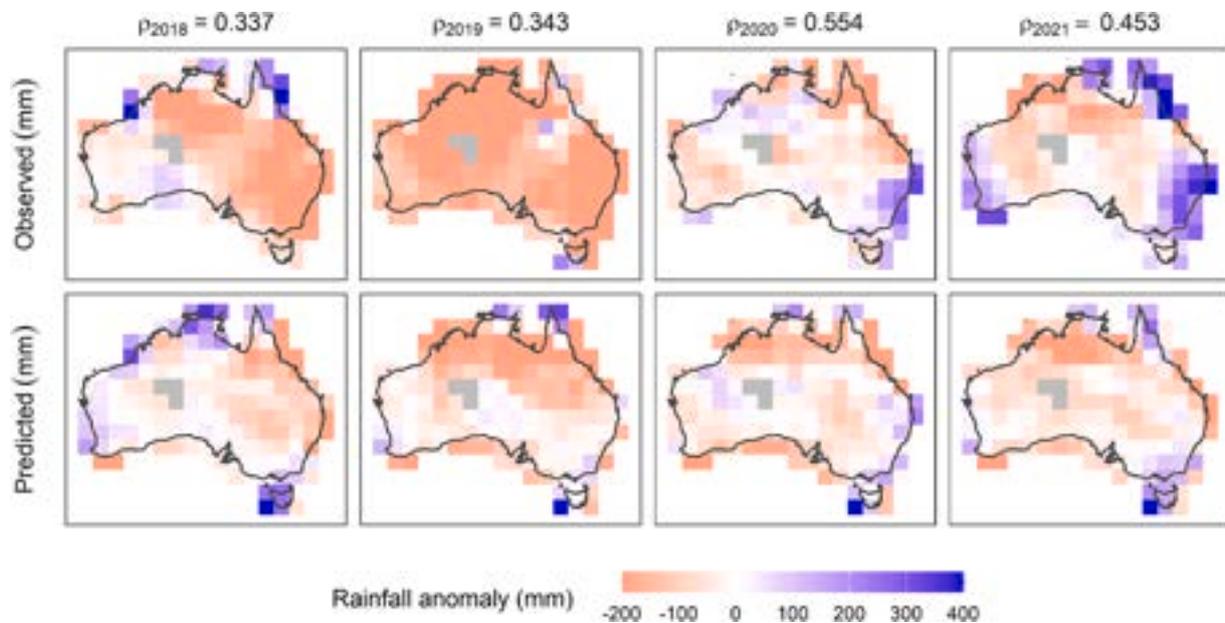


Fig. 12. Observed and predicted rainfall anomaly (mm) at lead 1 year from the split sample validation across all grids over Australia using spectrally transformed climate indices. The spatial correlation between observed and predicted rainfall anomalies is shown on the top of subplots. Compared to the past decade's climatology, the blue colour regions show positive rainfall anomalies (wetter); red colour regions present negative rainfall anomalies (drier), while regions with white colour represent the area with no significant changes relative to the past decade. Grey grids in the central and western deserts of Australia are grids with missing data.

transformed HLC, as the relative improvements illustrated remain consistent, even though actual correlations differ.

2. Can this possible over-inflation be further verified using measures of predictability over space (spatial correlation) in addition to the regular time series correlation performance metric? The use of spatial correlation in Fig. 11 again confirms the superiority of the HLC-WASP over the HLC case.
3. Can the HLC model structure be further simplified, given the added predictability the spectrally transformed CMIP6 predictors offer? The results in Fig. 10 confirm that the use of the spectrally transformed predictors is sufficient to exclude the empirical model component of the HLC, a notable outcome that highlights the need for sensible transformations of direct climate model simulations before use in a prediction context

Our findings suggest that the combination of spectral transformation and the HLC model can lead to improved interannual rainfall forecasting in Australia. The use of spatial correlation and the deletion of the empirical model component of the HLC further confirm the superiority of the HLC-WASP forecasting framework. To sum up, the enhanced multi-year forecast skills can be attributed to three main factors: the intrinsic predictability of certain climate indices, the combination of dynamical forecasts using HLC logic, and the enhancement achieved through the integration of the WASP method in the modelling framework.

In addition, this study primarily focuses on deterministic forecasts; hence, deterministic measures such as correlation and RMSE are used for evaluation. We incorporate uncertainty by employing different initializations to generate an ensemble of simulations. This accounts for uncertainties from model parameters, emission scenarios, and internal climate variability. The use of multiple ensemble members from individual CMIP6 models and a selection of five CMIP6 models allows for an explicit representation of forecast uncertainty. As shown in Figs. 7 and 10, the uncertainty in rainfall forecasts escalates with increasing lead time, consistent with expectations.

5. Conclusions

Climate predictions at interannual-to-decadal timescales provide information that has substantial socio-economic benefits, especially for sectors relying on water availability. While various studies have shown promise in predicting at this timescale using the latest suite of decadal prediction experiments from CMIP6, the prediction skill of rainfall remains poor compared to other variables. As an alternative, novel statistical methods based on the decadal prediction outputs have the potential to aid rainfall prediction at this critical timescale. Here, we presented a hybrid HLC-WASP model for rainfall forecasting in Australia using climate indices derived from decadal prediction experiments. Climate indices representative of large-scale modes of climate variability are spectrally transformed corresponding to the gridded rainfall across Australia and used as predictors for rainfall prediction at interannual timescales. Through the split sample validation, the best model for the interannual Australian rainfall forecasting is the hybrid HLC-WASP model with the dynamical component alone. Using outputs from five CMIP6 GCMs, we showed that spectrally transformed SSTA indices significantly improve the predictability of rainfall over long-lead times, up to a maximum of five years ahead.

The multi-model ensemble mean emphasizes the uncertainty in climate predictions. The HLC-WASP, driven by the MME inputs, shows the best performance at lead times of 1–2 years compared to individual climate models. However, its reduced variability at longer lead times does not result in greater skill for the forecasts issued. The largest improvements were observed in northern and southeast Australia, and the HLC-WASP performed significantly better than its predecessor, the HLC model (Choudhury et al., 2019), across all lead years and climate models. Additionally, these improvements become clearer with increasing lead times, suggesting its potential for aiding risk management and adaptation assessments.

Alongside its potential for long-term rainfall forecasting, the HLC-WASP method has broader applications, including drought downscaling, streamflow prediction, and paleoclimate reconstruction. The integration of spectral transformation with statistical forecasting systems offers a promising approach for enhancing the predictability of hydro-climatic variables, with significant implications for the fields of

hydrology and civil engineering. By improving the accuracy of long-term rainfall forecasts, this work supports more effective water resource management, infrastructure planning, and adaptation strategies in the face of climate variability and change.

CRedit authorship contribution statement

Ze Jiang: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis. **Dipayan Choudhury:** Writing – review & editing. **Ashish Sharma:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the Australian Research Council (ARC) Discovery project DP200101326 and ARC Industry Fellowship IE240100041, in partnership with WaterNSW, Australia. We thank the climate modelling groups for producing and making available their model outputs used in our study. This research was undertaken with the assistance of resources from the National Computational Infrastructure (NCI), which is supported by the Australian Government. We are also thankful to two anonymous referees and the editor for their constructive comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jhydrol.2025.132870>.

Data availability

I have shared the link to access data used in the study.

References

- Bates, J.M., Granger, C.W., 1969. The combination of forecasts. *J. Oper. Res. Soc.* 20 (4), 451–468.
- Befort, D.J., O'Reilly, C.H., Weisheimer, A., 2020. Constraining Projections Using Decadal Predictions. *Geophys. Res. Lett.* 47 (18), e2020GL087900. <https://doi.org/10.1029/2020GL087900>.
- Bilbao, R., Wild, S., Ortega, P., Acosta-Navarro, J., Arsouze, T., Bretonnière, P.A., Vegas-Regidor, J., 2021. Assessment of a full-field initialized decadal climate prediction system with the CMIP6 version of EC-Earth. *Earth Syst. Dynam.* 12 (1), 173–196. <https://doi.org/10.5194/esd-12-173-2021>.
- Boer, G.J., Smith, D.M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Eade, R., 2016. The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geosci. Model Dev.* 9 (10), 3751–3777. <https://doi.org/10.5194/gmd-9-3751-2016>.
- Box, G.E.P., Cox, D.R., 1964. An Analysis of Transformations. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 26 (2), 211–252. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>.
- Cai, W., Van Rensch, P., Cowan, T., Hendon, H.H., 2011. Teleconnection pathways of ENSO and the IOD and the mechanisms for impacts on Australian rainfall. *J. Clim.* 24 (15), 3910–3923. <https://doi.org/10.1175/2011jcli4129.1>.
- Calli Quaglia, F., Terzago, S., von Hardenberg, J., 2022. Temperature and precipitation seasonal forecasts over the Mediterranean region: added value compared to simple forecasting methods. *Clim. Dyn.* 58 (7), 2167–2191. <https://doi.org/10.1007/s00382-021-05895-6>.
- Chikamoto, Y., Timmermann, A., Luo, J.-J., Mochizuki, T., Kimoto, M., Watanabe, M., Jin, F.-F., 2015. Skilful multi-year predictions of tropical trans-basin climate variability. *Nat. Commun.* 6 (1), 6869. <https://doi.org/10.1038/ncomms7869>.
- Choudhury, D., Mehrotra, R., Sharma, A., Sen Gupta, A., Sivakumar, B., 2019. Effectiveness of CMIP5 Decadal Experiments for Interannual Rainfall Prediction Over Australia. *Water Resour. Res.* 55 (8), 7400–7418. <https://doi.org/10.1029/2018WR024462>.
- Choudhury, D., Sen Gupta, A., Sharma, A., Mehrotra, R., Sivakumar, B., 2017. An Assessment of Drift Correction Alternatives for CMIP5 Decadal Predictions. *J. Geophys. Res. Atmos.* 122 (19), 10282–210296. <https://doi.org/10.1002/2017JD026900>.
- Choudhury, D., Sharma, A., Sen Gupta, A., Mehrotra, R., Sivakumar, B., 2016. Sampling biases in CMIP5 decadal forecasts. *J. Geophys. Res. Atmos.* 121 (7), 3435–3445. <https://doi.org/10.1002/2016JD024804>.
- Choudhury, D., Sharma, A., Sivakumar, B., Sen Gupta, A., Mehrotra, R., 2015. On the predictability of SSTA indices from CMIP5 decadal experiments. *Environ. Res. Lett.* 10 (7), 074013. <https://doi.org/10.1088/1748-9326/10/7/074013>.
- Deb, P., Moradkhani, H., Abbaszadeh, P., Kiem, A.S., Engström, J., Keellings, D., Sharma, A., 2020. Causes of the Widespread 2019–2020 Australian Bushfire Season. *Earth's Future* 8 (11), e2020EF001671. <https://doi.org/10.1029/2020EF001671>.
- Doi, T., Behera, S.K., Yamagata, T., 2020. Predictability of the Super IOD Event in 2019 and Its Link With El Niño Modoki. *Geophys. Res. Lett.* 47 (7), e2019GL086713. <https://doi.org/10.1029/2019GL086713>.
- Dutilleul, P., Clifford, P., Richardson, S., Hemon, D., 1993. Modifying the t test for assessing the correlation between two spatial processes. *Biometrics* 49 (1), 305–314. <https://doi.org/10.2307/2532625>.
- 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. *Geosci. Model Dev.* 9 (5), 1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>.
- Fierro, A.O., Leslie, L.M., 2013. Links between central west Western Australian rainfall variability and large-scale climate drivers. *J. Clim.* 26 (7), 2222–2246.
- Freund, M.B., Marshall, A.G., Wheeler, M.C., Brown, J.N., 2021. Central Pacific El Niño as a Precursor to Summer Drought-Breaking Rainfall Over Southeastern Australia. *Geophys. Res. Lett.* 48 (7), e2020GL091131. <https://doi.org/10.1029/2020GL091131>.
- Fuckar, N.S., Volpi, D., Guemas, V., Doblas-Reyes, F.J., 2014. A posteriori adjustment of near-term climate predictions: Accounting for the drift dependence on the initial conditions. *Geophys. Res. Lett.* 41 (14), 5200–5207. <https://doi.org/10.1002/2014GL060815>.
- Helsel, D.R., Hirsch, R.M., 2002. *Statistical methods in water resources* (04-A3). Retrieved from Reston, VA: <http://pubs.er.usgs.gov/publication/twri04A3>.
- Hyndman, R.J., Athanasopoulos, G., 2014. *Forecasting: principles and practice: Heathmont. Vic. OTexts*.
- Intergovernmental Panel on Climate Change, 2014. *Near-term Climate Change: Projections and Predictability*. In C. Intergovernmental Panel on Climate (Ed.), *Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 953–1028). Cambridge: Cambridge University Press.
- IPCC, 2021. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (ISBN 978-1-107-66182-0). Retrieved from Cambridge, United Kingdom and New York, NY, USA: <https://www.ipcc.ch/report/ar6/wg1/>.
- Jiang, Z., Johnson, F., Sharma, A., 2023. Do Derived Drought Indices Better Characterize Future Drought Change? *Earth's Future* 11 (7), e2022EF003350. <https://doi.org/10.1029/2022EF003350>.
- Jiang, Z., Rashid, M.M., Johnson, F., Sharma, A., 2021a. A wavelet-based tool to modulate variance in predictors: An application to predicting drought anomalies. *Environ. Model. Softw.* 135, 104907. <https://doi.org/10.1016/j.envsoft.2020.104907>.
- Jiang, Z., Rashid, M. M., Johnson, F., & Sharma, A. (2021). WASP: Wavelet System Prediction. The Comprehensive R Archive Network. Retrieved from <https://cran.r-project.org/web/packages/WASP/index.html>.
- Jiang, Z., Sharma, A., Johnson, F., 2019. Assessing the sensitivity of hydro-climatological change detection methods to model uncertainty and bias. *Adv. Water Resour.* 134, 103430. <https://doi.org/10.1016/j.advwatres.2019.103430>.
- Jiang, Z., Sharma, A., Johnson, F., 2020. Refining Predictor Spectral Representation Using Wavelet Theory for Improved Natural System Modeling. *Water Resour. Res.* 56 (3). <https://doi.org/10.1029/2019WR026962>.
- Jiang, Z., Sharma, A., Johnson, F., 2021b. Variable transformations in the spectral domain – Implications for hydrologic forecasting. *J. Hydrol.* 603, 126816. <https://doi.org/10.1016/j.jhydrol.2021.126816>.
- Jones, D.A., Wang, W., Fawcett, R., 2009. High-quality spatial climate data-sets for Australia. *Aust. Meteorol. Oceanogr. J.* 58 (4), 233.
- Kaiser, G., 2010. *A friendly guide to wavelets*. Springer Science & Business Media, Berlin/Heidelberg, Germany.
- Khan, M.Z.K., Mehrotra, R., Sharma, A., Sankarasubramanian, A., 2014. Global Sea Surface Temperature Forecasts Using an Improved Multimodel Approach. *J. Clim.* 27 (10), 3505–3515. <https://doi.org/10.1175/JCLI-D-13-00486.1>.
- Khan, M.Z.K., Sharma, A., Mehrotra, R., Schepen, A., Wang, Q.J., 2015. Does improved SSTA prediction ensure better seasonal rainfall forecasts? *Water Resour. Res.* 51 (5), 3370–3383. <https://doi.org/10.1002/2014WR015997>.
- Kharin, V.V., Boer, G.J., Merryfield, W.J., Scinocca, J.F., Lee, W.-S., 2012. Statistical adjustment of decadal predictions in a changing climate. *Geophys. Res. Lett.* 39 (19). <https://doi.org/10.1029/2012gl052647>.
- Kim, S., Sharma, A., Liu, Y.Y., Young, S.I., 2022. Rethinking Satellite Data Merging: From Averaging to SNR Optimization. *IEEE Trans. Geosci. Remote Sens.* 60. <https://doi.org/10.1109/Tgrs.2021.3107028>.
- Kirono, D.G., Chiew, F.H., Kent, D.M., 2010. Identification of best predictors for forecasting seasonal rainfall and runoff in Australia. *Hydrological Processes: an International Journal* 24 (10), 1237–1247.

- 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>.
- Liang, X.-Z., Wu, Y., Chambers, R.G., Schmoltdt, D.L., Gao, W., Liu, C., Kennedy, J.A., 2017. Determining climate effects on US total agricultural productivity. *Proc. Natl. Acad. Sci.* 114 (12), E2285–E2292. <https://doi.org/10.1073/pnas.1615922114>.
- Ma, Y., Sun, J., Dong, T., Yu, W., Dong, W., 2023. More profound impact of CP ENSO on Australian spring rainfall in recent decades. *Clim. Dyn.* 60 (9), 3065–3079. <https://doi.org/10.1007/s00382-022-06485-w>.
- Maheswaran, R., Khosa, R., 2012. Comparative study of different wavelets for hydrologic forecasting. *Comput. Geosci.* 46, 284–295. <https://doi.org/10.1016/j.cageo.2011.12.015>.
- Mahmood, R., Donat, M.G., Ortega, P., Doblas-Reyes, F.J., Ruprich-Robert, Y., 2021. Constraining Decadal Variability Yields Skillful Projections of Near-Term Climate Change. *Geophys. Res. Lett.* 48 (24), e2021GL094915. <https://doi.org/10.1029/2021GL094915>.
- McInerney, D., Thyer, M., Kavetski, D., Lerat, J., Kuczera, G., 2017. Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. *Water Resour. Res.* 53 (3), 2199–2239. <https://doi.org/10.1002/2016wr019168>.
- Meehl, G.A., Goddard, L., Murphy, J., Stouffer, R.J., Boer, G., Danabasoglu, G., Hawkins, E., 2009. Decadal prediction: can it be skillful? *Bull. Am. Meteorol. Soc.* 90 (10), 1467–1486.
- Meehl, G.A., Hu, A., Tebaldi, C., 2010. Decadal Prediction in the Pacific Region. *J. Clim.* 23 (11), 2959–2973. <https://doi.org/10.1175/2010JCLI3296.1>.
- Mehrotra, R., Sharma, A., Bari, M., Tuteja, N., Amirthanathan, G., 2014. An assessment of CMIP5 multi-model decadal hindcasts over Australia from a hydrological viewpoint. *J. Hydrol.* 519, 2932–2951. <https://doi.org/10.1016/j.jhydrol.2014.07.053>.
- Mehta, V.M., Wang, H., Mendoza, K., 2013. Decadal predictability of tropical basin average and global average sea surface temperatures in CMIP5 experiments with the HadCM3, GFDL-CM2.1, NCAR-CCSM4, and MIROC5 global Earth System Models. *Geophys. Res. Lett.* 40 (11), 2807–2812. <https://doi.org/10.1002/grl.50236>.
- Moemken, J., Feldmann, H., Pinto, J.G., Buldmann, B., Laube, N., Kadow, C., Marotzke, J., 2021. The regional Miklip decadal prediction system for Europe: Hindcast skill for extremes and user-oriented variables. *Int. J. Climatol.* 41 (S1), E1944–E1958. <https://doi.org/10.1002/joc.6824>.
- Percival, D.B., Walden, A.T., 2000. *Wavelet methods for time series analysis*. Cambridge University Press, Cambridge.
- Rashid, M.M., Sharma, A., Johnson, F., 2020. Multi-model drought predictions using temporally aggregated climate indicators. *J. Hydrol.* 581, 124419. <https://doi.org/10.1016/j.jhydrol.2019.124419>.
- Rayner, N.A., Parker, D.E., Horton, E.B., Folland, C.K., Alexander, L.V., Rowell, D.P., Kaplan, A., 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res. Atmos.* 108 (D14). <https://doi.org/10.1029/2002jd002670>.
- Roderick, T.P., Wasko, C., Sharma, A., 2020. An Improved Covariate for Projecting Future Rainfall Extremes? *Water Resour. Res.* 56 (8), e2019WR026924. <https://doi.org/10.1029/2019WR026924>.
- Schepen, A., Wang, Q.J., 2015. Model averaging methods to merge operational statistical and dynamic seasonal streamflow forecasts in Australia. *Water Resour. Res.* 51 (3), 1797–1812. <https://doi.org/10.1002/2014WR016163>.
- Schepen, A., Wang, Q.J., Robertson, D.E., 2012. Evidence for Using Lagged Climate Indices to Forecast Australian Seasonal Rainfall. *J. Clim.* 25 (4), 1230–1246. <https://doi.org/10.1175/JCLI-D-11-00156.1>.
- Schepen, A., Wang, Q.J., Robertson, D.E., 2014. Seasonal Forecasts of Australian Rainfall through Calibration and Bridging of Coupled GCM Outputs. *Mon. Weather Rev.* 142 (5), 1758–1770. <https://doi.org/10.1175/MWR-D-13-00248.1>.
- Sheen, K.L., Smith, D.M., Dunstone, N.J., Eade, R., Rowell, D.P., Vellinga, M., 2017. Skillful prediction of Sahel summer rainfall on inter-annual and multi-year timescales. *Nature Communications* 8 (1), 14966. <https://doi.org/10.1038/ncomms14966>.
- Smith, D.M., Eade, R., Pohlmann, H., 2013. A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. *Clim. Dyn.* 41 (11), 3325–3338. <https://doi.org/10.1007/s00382-013-1683-2>.
- Smith, D.M., Eade, R., Scaife, A.A., Caron, L.P., Danabasoglu, G., DelSole, T.M., Yang, X., 2019. Robust skill of decadal climate predictions. *npj Clim. Atmos. Sci.* 2 (1), 13. <https://doi.org/10.1038/s41612-019-0071-y>.
- Smith, D.M., Scaife, A.A., Eade, R., Athanasiadis, P., Bellucci, A., Bethke, I., Zhang, L., 2020. North Atlantic climate far more predictable than models imply. *Nature* 583 (7818), 796–800. <https://doi.org/10.1038/s41586-020-2525-0>.
- Wang, Q.J., Shrestha, D.L., Robertson, D.E., Pokhrel, P., 2012. A log-sinh transformation for data normalization and variance stabilization. *Water Resour. Res.* 48 (5). <https://doi.org/10.1029/2011wr010973>.
- Wu, X., Marshall, L., Sharma, A., 2019. The influence of data transformations in simulating Total Suspended Solids using Bayesian inference. *Environ. Model. Softw.* 121, 104493. <https://doi.org/10.1016/j.envsoft.2019.104493>.
- Xu, L., Zhang, C., Chen, N., Moradkhani, H., Chu, P.-S., Zhang, X., 2020. Potential precipitation predictability decreases under future warming. *Geophys. Res. Lett.* 47 (22), e2020GL090798. <https://doi.org/10.1029/2020GL090798>.