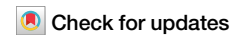




# Tropical monsoon rainfall can be predicted with lead times up to 10 months



Guanghao Ran <sup>1</sup>, Jun Meng <sup>2</sup> & Jingfang Fan <sup>3,4</sup>

Tropical monsoons play a critical role in shaping regional and global climate systems, with profound ecological and socio-economic impacts. However, their long-term prediction remains challenging due to the complex interplay of regional dynamics, global climate drivers, large-scale teleconnections, and inherent non-stationarities in the climate system. Here, we introduce a unified network-based framework for predicting monsoon precipitation across diverse tropical regions. By leveraging global 2-meter air temperature fields, this approach captures large-scale climate teleconnections, such as the El Niño–Southern Oscillation and Rossby waves, enabling accurate forecasts for four key monsoon systems: the South American, East Asian, West African, and Indian monsoons. Our framework achieves remarkable forecasting accuracy with lead times of 4–10 months, outperforming traditional systems such as Seasonal Forecast System 5 and Climate Forecast System version 2. Beyond its predictive capabilities, the framework offers flexibility for application to other regions and climate phenomena, advancing our understanding of global climate dynamics. These findings have far-reaching implications for disaster preparedness, resource management, and sustainable development.

Tropical monsoon regions, concentrated within 30° latitude of the equator, are marked by abundant precipitation and exceptional ecological diversity. Monsoon precipitation is vital for maintaining regional ecosystems<sup>1</sup>, supporting agriculture<sup>2,3</sup>, and water resource management<sup>4</sup>. Beyond their local significance, tropical monsoons influence global atmospheric circulation, water cycling<sup>5</sup>, and energy balance<sup>6</sup>.

However, accurately forecasting monsoon precipitation over long timescales remains a formidable challenge due to the complex interplay of initial conditions, external forcings, and internal variability<sup>7</sup>. For instance, small errors in initial conditions can amplify over time, making accurate long-term forecasts difficult<sup>8</sup>. Additionally, uncertainties in future external forcings, such as greenhouse gas emissions and volcanic activity, further complicate predictions<sup>9,10</sup>. The strong natural variability of monsoons adds another layer of difficulty, posing significant challenges for seasonal and decadal-scale forecasting<sup>11</sup>.

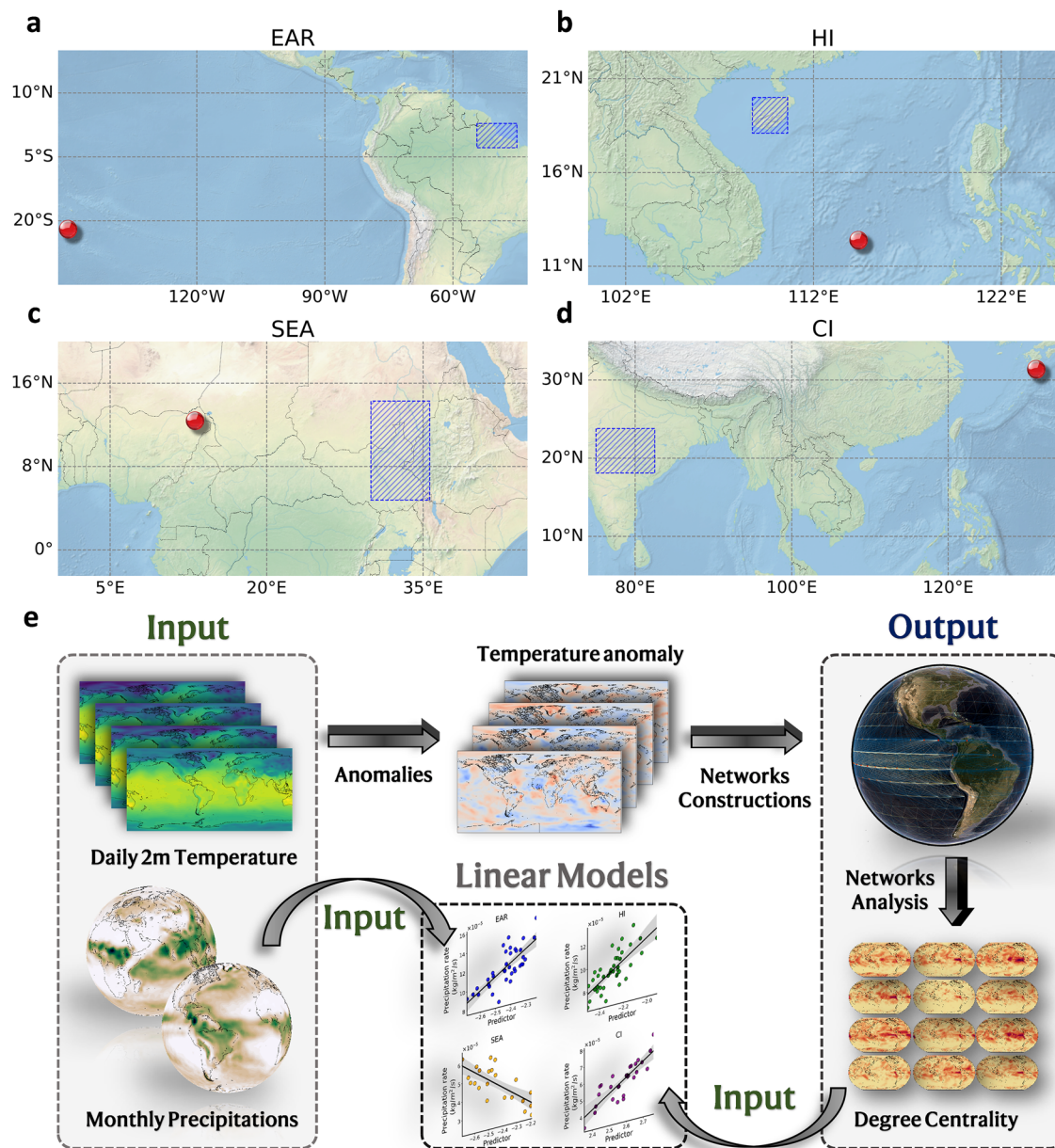
Despite progress, climate models struggle to capture the nonlinear processes and complex interactions that define the climate system, limiting their long-term forecasting accuracy<sup>12</sup>. Furthermore, these models require extensive observational data and sophisticated parameterizations, complicating their use in regions with limited historical records<sup>13,14</sup>. Recently, AI-

based methods have enabled the development of machine learning (ML) weather prediction systems, such as FourCastNet<sup>15</sup>, GraphCast<sup>16</sup>, Pangu-Weather<sup>17</sup>, NowcastNet<sup>18</sup>, and GenCast<sup>19</sup>. However, these systems excel at short- to medium-term (up to 15 days) predictions and are currently ill-suited for addressing the complexities of long-term and seasonal-scale precipitation forecasting.

To address these challenges, network theory offers a robust approach for analyzing the dynamic and structural properties of complex systems<sup>20–23</sup>. Its application in climate science<sup>24–27</sup> has revealed hidden connections and teleconnections within the global climate system, providing a novel basis for prediction<sup>28–30</sup>. By leveraging network-derived metrics, researchers have successfully predicted phenomena like El Niño events<sup>31,32</sup> and monsoonal rainfall<sup>26,33</sup>, demonstrating the approach's feasibility and reliability.

This study aims to address the longstanding challenge of long-lead seasonal monsoon precipitation prediction by proposing a climate network-based framework that leverages global teleconnections. Specifically, we (1) develop a network-based prediction system grounded in spatiotemporal climate connectivity; (2) apply this unified framework to forecast monsoon precipitation across four major tropical monsoon systems—the South American monsoon (Eastern Amazon Region (EAR)),

<sup>1</sup>Department of Physics, Hong Kong Baptist University, Hong Kong, China. <sup>2</sup>National Key Laboratory of Earth System Numerical Modeling and Application, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China. <sup>3</sup>School of Systems Science/Institute of Non-equilibrium Systems, Beijing Normal University, Beijing, China. <sup>4</sup>Potsdam Institute for Climate Impact Research, Potsdam, Germany. e-mail: [jun.meng.phy@gmail.com](mailto:jun.meng.phy@gmail.com); [jingfang@bnu.edu.cn](mailto:jingfang@bnu.edu.cn)



**Fig. 1 | Network-based prediction framework.** The red dots on the map indicate the locations of the identified predictors, while the blue boxes delineate the targeted forecast regions: **a** EAR, **b** HI, **c** SEA, and **d** CI. **e** presents the proposed prediction

framework, highlighting the integration of network-based metrics and statistical modeling for monsoon precipitation forecasting. The geographic visualization was generated using Gephi combined with Google Earth Engine (Image ©2023 Google).

the East Asian monsoon (Hainan Island (HI)), the West African monsoon (Sahel region of East Africa (SEA)), and the Indian monsoon (Central India (CI)); and (3) assess its forecasting skill relative to established dynamical models, namely European Center for Medium-Range Weather Forecasts Seasonal Forecast System 5 (SEAS5)<sup>34</sup> and Climate Forecast System version 2 (CFSv2)<sup>35</sup>. These four regions represent diverse climatic regimes, enabling robust evaluation of the framework’s generalizability. Notably, the SEA focus area, encompassing South Sudan within the Sahel belt, exemplifies a region where predictive urgency intersects with ecological and socioeconomic vulnerability: Around 80% of South Sudan’s population depends on rain-fed agriculture<sup>36</sup>, while its ecologically sensitive landscapes-including Boma National Park, a critical refuge for African wildlife-face escalating climate risks. This dual pressure underscores the need for reliable monsoon forecasts to support food security, biodiversity conservation, and climate-resilient development. Therefore, by integrating network theory with climate predictability science, we aim to provide both improved

forecasting capabilities and enhanced insight into the underlying drivers of monsoon variability.

## Results

### Mining network predictors

The rainy season, characterized by significant increases in precipitation, varies widely in timing, duration, and intensity across different regions. For this study, we focused on four regions within the tropical monsoon belt, illustrated as blue boxes in Fig. 1a–d: EAR, HI, SEA, and CI (See Table 1 for the specific latitudinal/longitudinal ranges of these regions). These regions represent diverse monsoonal systems driven by distinct atmospheric and oceanic processes, making them ideal test cases for our predictive framework (as depicted in Fig. 1e). To standardize the analysis, we defined the rainy season as May–October in the Northern Hemisphere and November to next year’s April in the Southern Hemisphere, based on historical precipitation patterns.

**Table 1 | Monsoon regions and Predictors' information for monsoon precipitation forecasting**

Monsoon regions	Geographical locations	Rainy seasons	Predictor's locations	Predictors
EAR	(0.9524°S–0.9524°N) (46.8750°W–52.5000°W)	Nov.–Apr.	(21.9044°S, 150.0000°W)	$k_{4279}^{out-}$
HI	(18.0950°N–19.9997°N) (108.7500°E–110.6250°E)	May–Oct.	(12.3808°N, 114.3750°E)	$k_{2500}^{out-}$
SEA	(6.6666°N–12.3808°N) (31.8750°E–33.7500°E)	May–Oct.	(12.3808°N, 13.1250°E)	$k_{2446}^{out-}$
CI	(19.9997°N–21.9044°N) (76.8750°E–80.6250°E)	May–Oct.	(31.4281°N, 131.2500°E)	$k_{1549}^{out+}$

The subscripts in the Predictors column represent the index of the node among the 6242 nodes.

In our endeavor to identify reliable predictors for rainy season precipitation, we constructed a series of climate networks (CNs) spanning January 1950–December 2022. The dataset used for this analysis comprises daily mean 2 m air temperature data from NCEP-Reanalysis I, covering 6242 global grid points. Each grid point was treated as a node within the network, and its in-degree and out-degree were computed based on correlations with other nodes (See Methods). In our climate network framework, we quantify regional climate interactions through two fundamental metrics: in-degree and out-degree. For a given year  $y$ , the in-degree of node  $i$ , denoted as  $k_i^{in+}(y)$  (positive) or  $k_i^{in-}(y)$  (negative), measures the cumulative influence the region receives from global climate patterns. Physically, this metric represents the number of strong climate teleconnections (e.g., ENSO or Rossby wave activity) affecting the region, serving as an indicator of its sensitivity to external climate drivers. A high in-degree suggests the region is strongly coupled to large-scale climate variability, making it particularly responsive to global forcing mechanisms.

Conversely, the out-degree  $k_i^{out+}(y)$  or  $k_i^{out-}(y)$  quantifies the region's influence on other parts of the climate system, where the sign indicates whether the relationship is reinforcing (positive) or opposing (negative). Regions exhibiting high out-degree values function as important climate signal transmitters- their temperature variations can significantly impact remote areas through atmospheric teleconnections. For instance, a node with elevated  $k_i^{out+}(y)$  might amplify similar temperature anomalies downstream, while high  $k_i^{out-}(y)$  could induce opposite thermal responses. In our predictor selection, we particularly focus on high out-degree nodes located within known teleconnection zones, as their robust network connectivity often translates into strong predictive relationships with monsoon precipitation variability across our target regions.

Our primary objective was to identify specific regions within the tropical monsoon belt where rainy season precipitation could be reliably predicted, alongside their corresponding network predictors. To identify robust climate predictors, we systematically examined network nodes whose in-degree or out-degree metrics exhibited strong correlations with monsoon precipitation across the four target regions (EAR, HI, SEA, and CI). A multi-stage validation procedure was employed to ensure the temporal stability and physical plausibility of these relationships under evolving climate conditions. First, we evaluated correlation strength and statistical significance ( $p < 0.05$ ) across two distinct historical periods, 1950–1989 and 1990–2022, retaining only those predictors that remained significant in both intervals and exhibited stronger associations during the recent, warming-dominated period.

To further assess robustness, we performed sensitivity analyses across varying time windows and emphasized predictors that consistently demonstrated stable performance. The final set of predictors (highlighted in red dots in Fig. 1 a–d) were selected based on three complementary criteria: (1) highest correlation magnitude during the recent 33-year period (1990–2022); (2) sustained statistical significance across the full analysis window; and (3) physical alignment with established large-scale climate teleconnections, such as ENSO variability and Rossby wave propagation pathways. This selection strategy, supported by the global node screening presented in Supplementary Fig. S1, ensures that our predictors are both statistically sound and mechanistically interpretable within the broader context of climate system dynamics.

We designate the corresponding in-degree or out-degree of these nodes as predictors (as shown in Table. 1) for rainy season precipitation in these four regions. The calculation of rainy season precipitation for these regions is based on their respective geographical locations (See Methods). An illustrative depiction of the linear relationship between network predictors and rainy season precipitation is presented in Fig. 2, offering a clear visualization of their correlation.

### Forecasting the monsoon precipitation

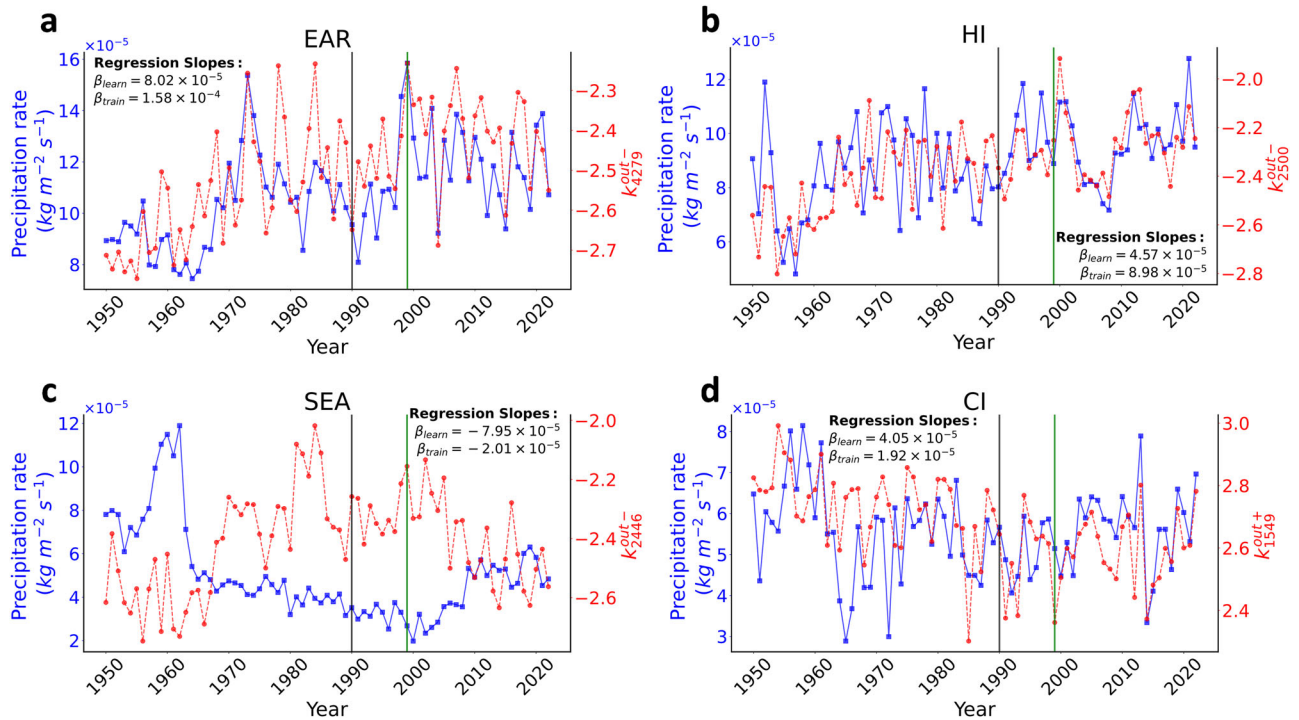
We identified a strong linear relationship between specific predictors and precipitation rates in each of the targeted regions (illustrated in Fig. 2). While the SEA region exhibits a negative correlation between predicted and observed precipitation, making visual pattern matching less intuitive, its predictive power remains statistically robust. Leveraging this relationship, we developed reliable forecasting algorithms using linear regression models. However, given the non-stationary nature of the climate system, where dynamics change over time, we assumed that the linear relationships would also evolve dynamically. To address this, we partitioned the most recent 33 years into two phases and fine-tuned our model by determining the optimal window length,  $L_{opt}$ , during the training phase (1990–1999). Rather than relying on all historical data since 1950, the model employs a moving window of data with a fixed length  $L_{opt}$ . Subsequently, the forecasting phase (2000–2022) is used to perform actual predictions and evaluate the algorithm's performance. This adaptive approach allows the model to adjust to recent trends and variability, mitigating the influence of outdated or fluctuating data.

To assess the performance of our forecasting framework, we employed three key metrics: root mean square error (RMSE)<sup>37</sup>, mean absolute percentage error (MAPE)<sup>38</sup>, and Pearson correlation coefficient (PCC)<sup>39</sup>. RMSE measures the average magnitude of prediction errors, highlighting significant deviations. MAPE expresses errors as a percentage, enabling easy comparison across datasets. PCC evaluates the linear correlation between predicted and observed values, indicating trend consistency. We systematically optimized these metrics to determine the best parameter set for our prediction framework by minimizing RMSE and MAPE while maximizing PCC. Specifically, we defined an evaluation criterion  $\lambda$  and selected the value of  $L$  that maximized  $\lambda$  as the optimal window length:

$$L_{opt} = \arg \max_L \lambda \triangleq \arg \max_L \left( \frac{PCC_{norm}}{RMSE_{norm} \cdot MAPE_{norm}} \right), \quad (1)$$

where  $PCC_{norm}$ ,  $RMSE_{norm}$  and  $MAPE_{norm}$  represent the normalized PCC, RMSE, and MAPE respectively. Our investigation revealed that employing a fixed  $L$  of 40 years, corresponding to the 40-year period preceding the predicted year  $y$ , resulted in poor performance in terms of RMSE and MAPE. This finding suggests the presence of significant systematic errors (see Supplementary Fig. S2 for details).

Consequently, we determined the optimal learning set lengths  $L_{opt}$  for four monsoon regions as follows: 37 years for EAR, 40 years for HI, 11 years for SEA, and 33 years for CI (see Supplementary Fig. S3 for details). Using these optimal learning set lengths, we forecasted rainy season total precipitation for the respective regions from 2000 to 2022, yielding favorable results. The forecasts achieved: RMSE values below  $1.2 \times 10^{-5} \text{ kgm}^{-2}\text{s}^{-1}$ , MAPE values below 16%, and PCC values higher than 0.6 (refer to Table 2).



**Fig. 2 | A visualization of the linear relationship between network predictors and rainy season precipitation.** **a** EAR, **b** HI, **c** SEA, and **d** CI. The blue curves (left-hand y-axis) represent the time series of observed seasonal total precipitation rates in tropical monsoon regions, while the red curves (right-hand y-axis) represent their corresponding network predictors, which are derived from the

red dots in Fig.1 for the corresponding regions. The linear regression slopes are displayed in each panel. The black and green vertical lines indicate the years 1990 and 1999, respectively, dividing the entire time span into three phases: the learning phase (1950–1989), the training phase (1990–1999), and the forecasting phase (2000–2022).

**Table 2 | A comparison of the forecast performance among SEAS5, CFSv2, and the network-based prediction method**

Regions	PCC			MAPE (%)			RMSE ( $\times 10^{-5}$ kg/m <sup>2</sup> /s)		
	SEAS5	CFSv2	Networks	SEAS5	CFSv2	Networks	SEAS5	CFSv2	Networks
EAR	<b>0.65</b>	0.67	0.63 ( <b>0.77</b> )	10.48	35.70	<b>7.57 (6.72)</b>	1.55	4.45	<b>1.13 (0.93)</b>
HI	0.47	0.29	<b>0.74 (0.82)</b>	15.81	43.91	<b>8.24 (7.96)</b>	1.77	4.43	<b>0.95 (0.81)</b>
SEA	0.36	0.44	<b>0.81 (0.77)</b>	58.10	27.27	<b>15.31 (17.20)</b>	2.28	1.11	<b>0.76 (0.82)</b>
CI	0.36	0.53	<b>0.76 (0.84)</b>	47.63	14.89	<b>12.65 (14.27)</b>	3.02	1.01	<b>0.80 (0.89)</b>

SEAS5 (2000–2022), CFSv2 (2000–2016). Bracketed values: network-based method (2000–2016). Bold highlights the optimal values (maximized PCC, minimized RMSE/MAPE).

This prediction model demonstrates a significant advantage over alternative forecasting methods, particularly in terms of lead time. For the SEA, CI, and HI regions, our model enables precipitation forecasts with a lead time of up to 4 months. For the EAR region, the forecast lead time extends to 10 months. This advantage arises from our methodology: the climate network is constructed annually using 2 m air temperature data from January to December of the preceding year, enabling predictor values to be generated in January and forecasts to be issued well before the onset of each region’s monsoon season. Such extended forecasting capabilities are crucial for societal, economic, and environmental applications, offering improved responses to climate-related disasters, enhanced resource management, and additional temporal windows for informed decision-making and planning.

**Comparative study: seasonal forecast systems vs. network-based prediction method**

We conducted a comparative analysis to evaluate the predictive performance of our proposed network-based prediction method against two established seasonal climate prediction systems: the European Center for Medium-Range Weather Forecasts (ECMWF) Seasonal Forecast System 5

(SEAS5) developed by the ECMWF and the Climate Forecast System version 2 (CFSv2) developed by the National Oceanic and Atmospheric Administration (NOAA). Both SEAS5 and CFSv2 are known to provide a reasonable fit for the observed spatial distribution of precipitation, as shown in Supplementary Fig. S4. It is noteworthy that the hindcast data for SEAS5 is only available up to 2022, while the hindcast data for CFSv2 has gaps during 1990–1992 and 2017–2018. Therefore, only the period from 2000 to 2016 was used to calculate the hindcast performance of SEAS5 and CFSv2. Numerous studies have evaluated the performance of SEAS5 and CFSv2 in tropical monsoon regions<sup>40–42</sup>. For instance, Ferreira et al.<sup>40</sup> demonstrated that SEAS5 effectively captures extreme precipitation anomalies in northern South America and northeastern Brazil, despite some spatial heterogeneity. Chevuturi et al.<sup>41</sup> reported a PCC of approximately 0.33 between SEAS5 hindcast data and observed data for the Indian monsoon. Similarly, Lang et al.<sup>42</sup> noted that CFSv2’s forecast skill scores for rainy season precipitation in 17 regions across China, including the Pearl River Basin (HI), generally remained below 0.3.

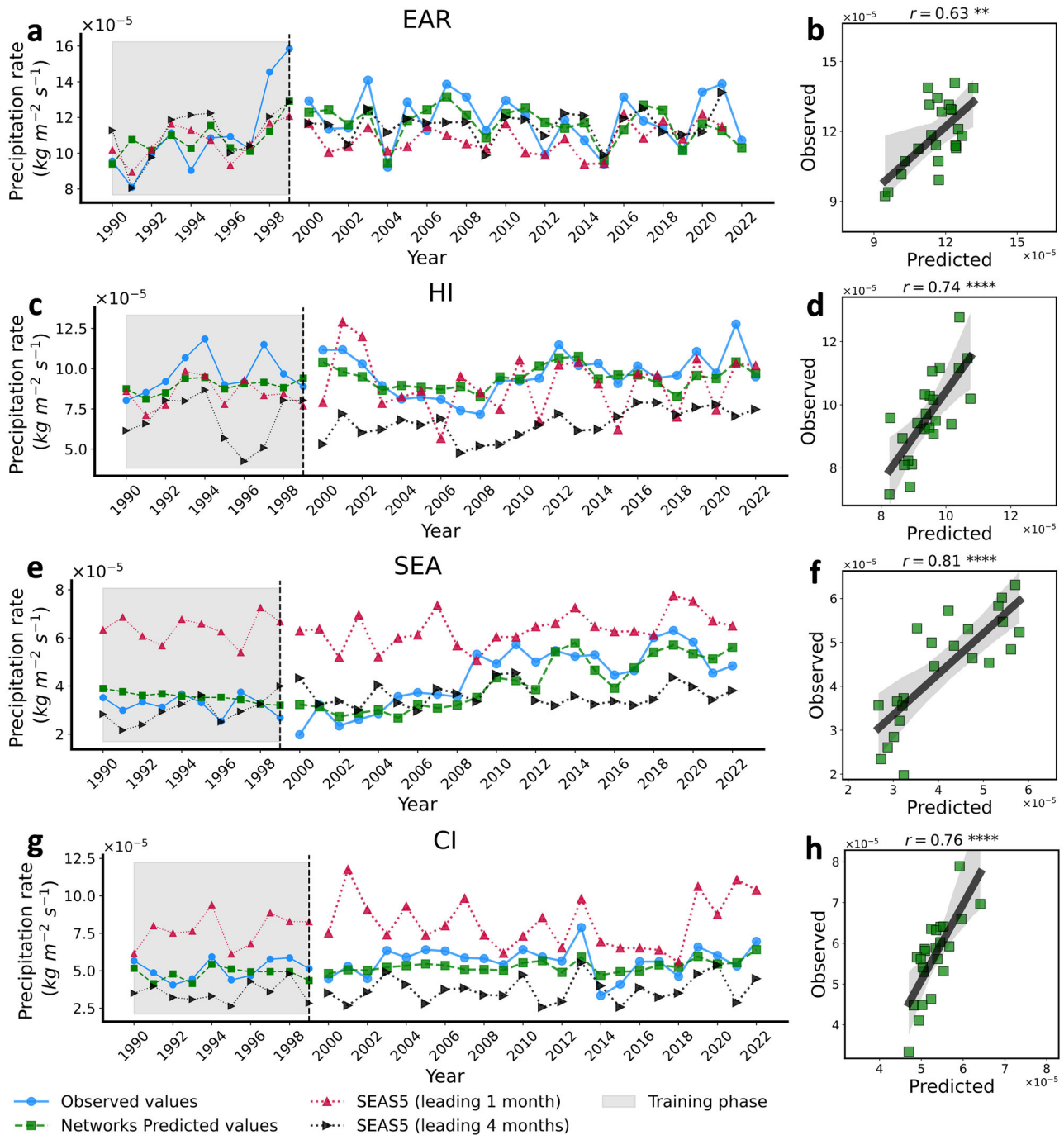
To compare the performance of SEAS5, CFSv2, and our network-based method, we evaluated their predictive capabilities for four tropical monsoon regions—EAR, HI, SEA, and CI—using hindcast data. SEAS5 and

CFSv2 hindcast data covered the periods 1990–2022 and 1993–2016, respectively, with lead times of 1 and 4 months. SEAS5 utilizes 25 ensemble members<sup>34</sup>, while CFSv2 employs 24 ensembles<sup>35</sup>. We compared these systems' predictive performance against our network-based method using key metrics (e.g., RMSE, MAPE, and PCC). Table 2 summarizes their performance for lead times of 1 month (refer to Table S1 for lead-4 month results).

SEAS5 and CFSv2 exhibited their best performance in the EAR region, achieving a maximum PCC of 0.67 at a lead time of 1 month. However, our network-based method matched their PCC (0.63) at a

much longer lead time of 10 months, while also delivering significantly lower MAPE and RMSE values (Table 2). In contrast, SEAS5 and CFSv2 showed poorer performance in the other three regions, with PCC values ranging from 0.26 to 0.47 and higher errors compared to our method, as illustrated in Fig. 3 and Supplementary Fig. S5. These results are consistent with findings from prior studies<sup>40–42</sup>, further underscoring the advantages of our forecasting method.

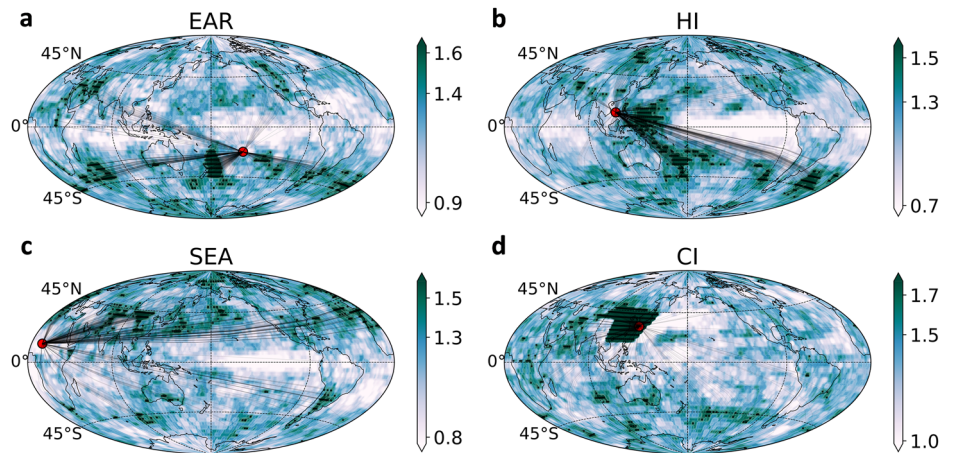
Beyond its superior predictive performance, the network-based approach provides valuable insights into the physical mechanisms driving climate variability, addressing the opacity of AI models that often act as



**Fig. 3 | Prediction results.** a, c, e, and g display time series of predicted values (green dashed lines) and observed values (blue solid lines). The SEAS5 hindcast values with lead times of 1 month and 4 months are shown as pink and black dotted lines, respectively. The gray shaded area indicates the training phase (1990–1999), while the white area represents the forecasting phase (2000–2022). b, d, f, and h display the

scatter plot of predicted values versus observed values during the forecasting phase (2000–2022) for corresponding regions. The black solid line represents the linear regression line, with the gray shaded area indicating its 99% confidence interval. The Pearson correlation coefficient (*r*) is provided for each relationship. Significance levels are denoted by markers: \*\*\*\* ( $p < 10^{-4}$ ) and \*\* ( $p < 10^{-2}$ ).

**Fig. 4 | Spatial distribution of the annually averaged (1950–2022) absolute link strengths (unitless) that comprise the predictor indices. a** EAR, **b** HI, **c** SEA, and **d** CI. Red dots represent the locations of the predictors, while black points indicate locations of the predictors which show the top 5% strongest link, and the thickness of black lines represents the strength of the link.



“black boxes.” By analyzing the annual averaged absolute weights of network links connected to predictors, we identified regions with significant influences on predictor dynamics. These insights also helped relate network indices’ predictability to underlying climate mechanisms. Key regions influencing predictors for the EAR and SEA (depicted in Fig. 4a, c) align with the Rossby wave train, consistent with the findings of Gelbrecht et al.<sup>43</sup>, which highlight the pivotal role of Rossby waves in modulating South American precipitation patterns. The propagation of Rossby waves across the atmosphere establishes teleconnections that influence precipitation variability in both the Eastern Amazon Rainforest and the Sahel region. For the HI region, influential predictors are primarily located in the eastern Pacific (Fig. 4b), a region strongly governed by the ENSO. ENSO’s global impacts on atmospheric circulation and weather patterns, including its modulation of East Asian monsoons, have been extensively documented<sup>44–46</sup>. The connection between ENSO phases and HI predictors underscores the critical role of large-scale ocean-atmosphere interactions in shaping rainfall patterns over East Asia. In the CI region, predictors cluster predominantly in East Asia, illustrating a teleconnection with the East Asian monsoon (Fig. 4d). This is consistent with prior studies, such as Kripalani et al.<sup>47</sup>, which demonstrated an out-of-phase relationship between monsoon precipitation in southern Japan and India. This teleconnection reflects the interplay between subtropical and tropical circulation systems, contributing to the variability of Indian monsoon rainfall.

These findings validate the network-based method as a powerful tool for understanding and forecasting monsoonal precipitation. By identifying and quantifying the influence of key teleconnections, this approach not only enhances predictive accuracy but also provides critical insights into the underlying physical mechanisms driving climate variability.

## Discussion

Traditional global climate models (GCMs), such as SEAS5 and CFSv2, simulate atmospheric processes by numerically solving fundamental physical equations. While effective for capturing large-scale patterns, these models struggle with regional precipitation variability—particularly over land in monsoon-affected areas. For instance, SEAS5’s performance over regions like SEA and CI is constrained by its sensitivity to cumulus parameterizations and the dominance of sea surface temperature (SST)-driven signals over tropical oceans<sup>12,34</sup>. The homogenization effect in gridded GCM outputs often dilutes land-based precipitation signals, which are governed by localized land-atmosphere feedbacks and convective processes.

In contrast, our network-based framework achieves dual advantages: (1) targeted extraction of dynamically relevant signals from observational data, and (2) implicit encoding of complex interactions, such as ENSO- and Rossby wave-driven teleconnections—within the network topology. By annually updating the network using global temperature fields, the framework reduces dimensionality while preserving key spatiotemporal features governing monsoon variability. This strategy enables early-warning signal

detection and robustness against structural uncertainties inherent in GCMs. Taking EAR as an illustrative example, the superior predictive performance in this region results primarily from the inherently stable interannual variability of its climatic predictors. EAR’s regional precipitation is predominantly influenced by well-defined ENSO-driven teleconnections and Rossby wave activities. These processes generate coherent, large-scale atmospheric signals, which our network is particularly adept at capturing. Conversely, regions like CI or SEA exhibit more complex land-ocean-topography interactions, introducing significant localized noise and thus decreasing signal stability and predictive accuracy. EAR’s more robust ocean-atmosphere coupling yields a higher signal-to-noise ratio, significantly enhancing the network’s predictive reliability and accuracy. Consequently, our method outperformed SEAS5 and CFSv2, achieving superior predictive accuracy (higher PCCs, lower MAPEs and RMSEs) with lead times up to 10 months.

The framework’s interpretability further addresses a critical limitation of purely AI-driven models. By identifying influential predictor regions (e.g., ENSO-associated zones) and their climatic linkages, it clarifies physical mechanisms, such as Rossby wave propagation—that underpin monsoon variability. This transparency bridges the gap between AI’s predictive power and the “black box” critique. Importantly, the framework complements rather than replaces GCMs, it bridges planetary-scale predictability with region-specific decision-making by leveraging observational constraints and teleconnection structures, thereby translating global model outputs into actionable regional insights.

Despite its strengths, several limitations require attention. First, relying on single-node predictors risks overfitting; ensemble-mean predictors (tested preliminarily) reduced overfitting but slightly lowered accuracy during extreme events. Second, while predictor regions align with known teleconnections (e.g., ENSO), causal relationships remain hypothetical, necessitating causality-based validation. Third, persistent climate shifts may demand periodic re-evaluation of network structures, despite our adaptive training strategy. Lastly, the linear framework may overlook nonlinear dynamics, suggesting opportunities to integrate ML for hybrid modeling<sup>48,49</sup>.

Emerging approaches combining CNs with ML, such as embedding network topology into AI architectures, could enhance nonlinear pattern recognition while retaining interpretability. Such hybrids may advance monsoon teleconnection analysis and refine predictions, synergizing physical insights with computational power. Future work should also explore incorporating additional variables (e.g., soil moisture, atmospheric pressure) to better resolve land-atmosphere feedbacks and extend the framework to extratropical systems or drought prediction.

By addressing GCM limitations and offering actionable insights into monsoon mechanisms, our framework advances climate prediction for disaster preparedness, agriculture, and water management. Its interpretability, adaptability, and extended lead times establish it as a critical tool for

climate resilience. Future integration with AI/ML and causality-driven methods promises to further bridge the gap between global modeling and regional decision-making, fostering a more sustainable society.

## Methods

### Data pre-process

For each node  $i$  (i.e., longitude–latitude grid point), we calculate daily 2 m air temperature standardized anomalies  $T_i(d)$  (actual temperature value minus the climatological average and divided by the climatological standard deviation) for each calendar day. Specifically, given an actual record  $\tilde{T}^y(d)$ , where  $y$  is the year (from 1948 to 2021), and  $d$  stands for the day (from 1 to 365), the anomalies record is defined as

$$T^y(d) = \frac{\tilde{T}^y(d) - \text{mean}(\tilde{T}^y(d))}{\text{std}(\tilde{T}^y(d))}, \quad (2)$$

where

$$\text{mean}(\tilde{T}^y(d)) = \begin{cases} \frac{1}{y-1948} \sum_{k=1949}^{y-1} \tilde{T}_i^k(d) & y \geq 1953 \\ \frac{1}{4} \sum_{k=1949}^{1952} \tilde{T}_i^k(d) & y < 1953 \end{cases}, \quad (3)$$

Similarly,  $\text{std}(\tilde{T}^y(d))$  is calculated in the same segments. Leap days were excluded to maintain consistency in the calendar year duration and simplify the analysis.

To analyze regional precipitation patterns, we selected multiple uniformly distributed latitude-longitude grid points (referred to as nodes) to represent four monsoon regions in the tropical monsoon belt: EAR, HI, SEA, and CI. The number of nodes in each region is denoted as:  $N_{\text{EAR}}$ ,  $N_{\text{HI}}$ ,  $N_{\text{SEA}}$ , and  $N_{\text{CI}}$ , respectively. The spatially averaged total precipitation rates, denoted as  $R$ , were calculated for each region during their respective rainy seasons from 1950 to 2022. Taking HI as an example, where the rainy season spans from May to October, the total rainy season precipitation is computed as:

$$R_{\text{HI}}(y) = \frac{1}{N_{\text{HI}}} \sum_{i=1}^{N_{\text{HI}}} \sum_{m=\text{May}}^{\text{Oct}} \mathcal{R}_i^{m,y}, \quad y = 1950, 1951, \dots, 2022, \quad (4)$$

where  $y$  indicates the year when the rainy season starts, and  $\mathcal{R}_i^{m,y}$  is the monthly averaged precipitation rate for node  $i$  in month  $m$ , year  $y$ .

### Climate network construction

In this study, we construct a series of CNs for each calendar year  $y$  from 1949 to 2021. Each CN is built based on statistical relationships between temperature anomaly time series from different spatial locations (nodes), where each node corresponds to a specific latitude–longitude grid point on the Earth’s surface. The strength and direction of links between node pairs are determined by the measure of similarity of their respective temperature time series. To account for Earth’s spherical geometry, we selected 6242 latitude-longitude grid points as CN nodes, ensuring an approximately homogeneous global coverage. Specifically, we define the base longitude resolution near the Equator (at latitude  $\phi \approx 0.9524^\circ$ ) as  $\Delta = 1.875^\circ$ , which corresponds to  $n_0 = 192$  nodes per latitude circle. Then, for a given latitude  $\phi_i$  (in degrees), the number of nodes is adjusted to  $n_i = \lfloor 192/b_i \rfloor$ , where  $b_i = \text{round}(1/\cos \phi_i)$  dynamically adjusts the longitudinal sampling to maintain near-constant surface density. The total number of nodes is  $N = \sum_{i \in \mathcal{I}} n_i \approx \sum_{k=0}^{46} 2 \cdot \lfloor 192 \cdot \cos \phi_{2k} \rfloor = 6242$ , where  $\mathcal{I} = \{0, 2, \dots, 192\}$  indexes latitudes sampled every two grid points, and the factor of two accounts for symmetric contributions from the Northern and Southern Hemispheres. This cosine-weighted thinning ensures quasi-uniform spacing across the sphere, countering the original grid’s  $\rho \propto 1/\cos \phi$  density divergence.

**Links.** To construct the CN for a given year  $y$ , we compute the time-lagged Pearson correlation between all possible pairs of nodes  $i$  and  $j$  following<sup>25</sup>. This yields a cross-correlation function as follows:

$$C_{ij}^y(-\tau) = \frac{\langle T_i^y(t)T_j^y(t-\tau) \rangle - \langle T_i^y(t) \rangle \langle T_j^y(t-\tau) \rangle}{\sqrt{\langle (T_i^y(t) - \langle T_i^y(t) \rangle)^2 \rangle} \sqrt{\langle (T_j^y(t-\tau) - \langle T_j^y(t-\tau) \rangle)^2 \rangle}}, \quad (5)$$

and

$$C_{ij}^y(\tau) = \frac{\langle T_i^y(t-\tau)T_j^y(t) \rangle - \langle T_i^y(t-\tau) \rangle \langle T_j^y(t) \rangle}{\sqrt{\langle (T_i^y(t-\tau) - \langle T_i^y(t-\tau) \rangle)^2 \rangle} \sqrt{\langle (T_j^y(t) - \langle T_j^y(t) \rangle)^2 \rangle}}, \quad (6)$$

where  $\tau \in [0, \tau_{\text{max}}]$  is the time lag, with  $\tau_{\text{max}} = 200$  days ensuring a reliable estimate of the background noise level, and  $\langle \cdot \rangle$  denote the average over the past 365 days, i.e.,

$$\langle f(t) \rangle = \frac{1}{365} \sum_{k=1}^{365} f(t-k), \quad (7)$$

where  $t$  stands for the end of time series (i.e., 31 Dec. of the year  $y$ ).

Next, we identify two key correlation values:

- $\max C_{ij}^y$ : the maximum (most positive) correlation and corresponding lag  $\tau^+$ ;
- $\min C_{ij}^y$ : the minimum (most negative) correlation and corresponding lag  $\tau^-$ .

These correlations extrema define the directionality of the network:

- if  $\tau^+ > 0$ , the positive link is directed from node  $i$  to node  $j$ , indicating that changes in  $i$  precede similar changes in  $j$ ;
- if  $\tau^- > 0$ , a negative link is directed from  $i$  to  $j$ , indicating that changes in  $i$  precede opposite changes in  $j$ .

This process yields two directed unweighted networks: one for positive and one for negative correlations. We then assign weights to each link based on the z-scores of the correlation values (standardized relative to the null distribution), resulting in two directed weighted CNs per year as follows:

$$W_{ij}^+(y) = \frac{C_{ij}^y(\tau^+) - \text{mean}(C_{ij}^y)}{\text{std}(C_{ij}^y)}, \quad (8)$$

and

$$W_{ij}^-(y) = \frac{C_{ij}^y(\tau^-) - \text{mean}(C_{ij}^y)}{\text{std}(C_{ij}^y)}, \quad (9)$$

where mean and std are the mean and standard deviation of the cross-correlation function, respectively.

Subsequently, we establish two distinct networks for each calendar year to differentiate between positive and negative link strengths. This dual-network approach allows us to capture both positive and negative interactions between nodes. Each network is characterized by an anti-symmetric adjacency matrix denoted as  $\mathbf{A}(y)$ , where each entry  $a_{ij}(y)$  signifies the presence of a link from node  $i$  to node  $j$  with a value of 1, while a link from node  $j$  to node  $i$  is represented by  $-1$  for the year  $y$ .

**Node degree.** To explore the dynamic aspects of the global climate system related to precipitation within tropical monsoon regions, we analyze the temporal evolution of key CNs properties. Our focus is on the degree centrality of nodes, a fundamental parameter in network theory that quantifies the total number of connected links<sup>21</sup>. Within the positive

and negative CNs of a given year  $y$ , each node has two distinct degree measures: in-degree and out-degree. The in-degree represents the average strength of incoming links, while the out-degree signifies the average strength of outgoing links. Mathematically, these degree measures are defined as:

$$k_i^{in\pm}(y) = \frac{\sum_{j=1, j \neq i}^N W_{ij}^{\pm}(y) I_{a_{ij}(y)=-1}}{\sum_{j=1, j \neq i}^N I_{a_{ij}(y)=-1}}, \quad (10)$$

and

$$k_j^{out\pm}(y) = \frac{\sum_{i=1, i \neq j}^N W_{ij}^{\pm}(y) I_{a_{ij}(y)=1}}{\sum_{i=1, i \neq j}^N I_{a_{ij}(y)=1}}, \quad (11)$$

where  $N$  equals to the total number of nodes (i.e., 6242) and  $I$  is the indicator function.

### Linear Models Construction

Within our proposed framework, we employ a linear model with an updating learning set to forecast precipitation for a given year  $y$ . In this approach, we utilize data from the  $L$  preceding years to make the prediction, where  $L$  ranges from 5 to 40 years. Specifically, for predicting the precipitation rate in year  $y$ , the learning set comprises the  $L$  years immediately preceding that target year as follows:

$$\mathbf{Y}^y \triangleq \{R(\eta) : \eta = y - 1, y - 2, \dots, y - L + 1, y - L\}. \quad (12)$$

Moreover, we have a vector of *Predictor* corresponding to the precipitation rate vector  $\mathbf{Y}^y$ , which is denoted as:

$$\mathbf{X}^{y-1} \triangleq \{k(\eta) : \eta = y - 2, y - 3, \dots, y - L, y - L - 1\}. \quad (13)$$

Next, to solve the linear model,

$$\mathbf{Y}^y = \beta^y \cdot \mathbf{X}^{y-1} + \alpha^y. \quad (14)$$

we employ the ordinary least squares regression technique, which provides the optimal linear unbiased estimator of the model parameters. Thus, the predicted value of year  $y$  can be calculated as:

$$\tilde{R}(y) = \beta^y \cdot k(y - 1) + \alpha^y. \quad (15)$$

In constructing the climate network, we use 2 m air temperature anomaly time series from January 1st to December 31st of each year. Therefore, the full set of network-based predictors for a given year  $y$ , can be computed in early January of year  $y + 1$ . This allows us to generate precipitation forecasts several months ahead of the upcoming monsoon seasons. Specifically, For HI, CI, and SEA regions, where the monsoon rainy seasons begin in May, predictions can be made in January, yielding a lead time of ~4 months. For the EAR region, where the monsoon typically begins in November, this approach provides a lead time of up to 10 months.

### Evaluation metrics

In addition to the conventional use of PCC, we also evaluate the forecasting skill associated with each specific  $L$  using metrics, such as RMSE and MAPE. RMSE is widely employed in climate and environmental research<sup>37</sup>. For the predicted values  $\tilde{R}(y)$  and the observed values  $R(y)$  of our proposed model,

the RMSE is calculated as

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{y=y_1}^{y_n} (R(y) - \tilde{R}(y))^2}. \quad (16)$$

On the other hand, MAPE provides a measure of the relative error and is defined as:

$$\text{MAPE} = \frac{1}{n} \sum_{y=y_1}^{y_n} \left| \frac{R(y) - \tilde{R}(y)}{R(y)} \right|. \quad (17)$$

MAPE is particularly useful in tasks where sensitivity to relative variations is more important than sensitivity to absolute variations. Both RMSE and MAPE have a numerical range of  $[0, +\infty)$ , with smaller values indicating better predictive performance.

### Data availability

The datasets used in this study were obtained from authoritative and publicly accessible sources as detailed below: (1) 2 m air temperature data (daily mean, 1948–2021, UTC 00:00) were obtained from the NCEP/NCAR Reanalysis I dataset. The data are provided on a Gaussian grid with a spatial resolution of approximately  $1.875^\circ \times 1.875^\circ$ . This dataset was selected due to (a) its well-established correlation with SST fields, facilitating the representation of large-scale climate interactions, and (b) its regular updates, supporting long-term climate analysis. (Access: <https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html>). (2) Monthly mean precipitation rate data (1950–2022) were also sourced from NCEP/NCAR Reanalysis I, covering the same spatial resolution. (Access: [https://downloads.psl.noaa.gov/Datasets/ncep.reanalysis/Monthlies/surface\\_gauss/](https://downloads.psl.noaa.gov/Datasets/ncep.reanalysis/Monthlies/surface_gauss/)). (3) Hindcast precipitation data from two operational seasonal forecasting systems were used for validation: the ECMWF and the National Centers for Environmental Prediction. These institutes support two seasonal forecast systems, SEAS5 and CFSv2, respectively. (a) The SEAS5 (system id = 5) system provides continuous data coverage from 1981 to 2022, with subsequent data being generated by the updated SEAS5 system (system id = 51). (b) The CFSv2 system offers data from 1993 to present, with missing records for 2017 and 2018. (Data accessible at: <https://cds.climate.copernicus.eu/datasets/seasonal-monthly-single-levels?tab=download>). (4) ERA5 monthly total precipitation data (1990–2022) were used for supplementary validation and were sourced from the Copernicus Climate Change Service. (Access: <https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels-monthly-means?tab=download>).

### Code availability

The Python codes used for the analysis are available on GitHub (<https://github.com/fanjingfang/Monsoon>).

Received: 6 February 2025; Accepted: 15 May 2025;

Published online: 28 May 2025

### References

1. Fu, C. & Wen, G. Variation of ecosystems over East Asia in association with seasonal, interannual and decadal monsoon climate variability. *Clim. Change* **43**, 477–494 (1999).
2. Wang, B., Gadgil, S. & Rupa Kumar, K. The asian monsoon-agriculture and economy. *Asian Monsoon* 651–683 (Springer, 2006).
3. Gadgil, S. & Gadgil, S. The indian monsoon, GDP and agriculture. *Econ. Polit. Weekly* **41**, 4887–4895 (2006).
4. Balek, J. *Hydrology and Water Resources in Tropical Regions* (Elsevier, 1983).
5. Worden, J., Noone, D. & Bowman, K. Importance of rain evaporation and continental convection in the tropical water cycle. *Nature* **445**, 528–532 (2007).

6. Trenberth, K. E., Stepaniak, D. P. & Caron, J. M. The global monsoon as seen through the divergent atmospheric circulation. *J. Clim.* **13**, 3969–3993 (2000).
7. Zhou, T., Hsu, H.-H. & Matsumoto, J. Summer monsoons in East Asia, Indochina and the western North Pacific. In *Proc. Global Monsoon System: Research and Forecast*, 43–72 (World Scientific, 2011).
8. Slingo, J. & Palmer, T. Uncertainty in weather and climate prediction. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **369**, 4751–4767 (2011).
9. Collins, M. et al. *Long-term Climate Change: Projections, Commitments and Irreversibility* (IPCC, 2013).
10. Wang, B. et al. Rethinking indian monsoon rainfall prediction in the context of recent global warming. *Nat. Commun.* **6**, 7154 (2015).
11. Kajtar, J.B., Santoso, A., England, M.H. & Cai, W. Tropical climate variability: interactions across the Pacific, Indian, and Atlantic oceans. *Clim. Dyn.* **48**, 2173–2190 (2017).
12. Sperber, K. et al. The Asian summer monsoon: an intercomparison of CMIP5 vs. CMIP3 simulations of the late 20th century. *Clim. Dyn.* **41**, 2711–2744 (2013).
13. Fan, K., Liu, Y. & Chen, H. Improving the prediction of the east asian summer monsoon: new approaches. *Weather Forecast.* **27**, 1017–1030 (2012).
14. Shi, P. et al. Significant land contributions to interannual predictability of east asian summer monsoon rainfall. *Earth's Future* **9**, e2020EF001762 (2021).
15. Pathak, J. et al. Fourcastnet: a global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214* (2022).
16. Lam, R. et al. Learning skillful medium-range global weather forecasting. *Science* **382**, 1416–1421 (2023).
17. Bi, K. et al. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* **619**, 533–538 (2023).
18. Zhang, Y. et al. Skillful nowcasting of extreme precipitation with nowcastnet. *Nature* **619**, 526–532 (2023).
19. Price, I. et al. Gencast: diffusion-based ensemble forecasting for medium-range weather. *arXiv preprint arXiv:2312.15796* (2023).
20. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
21. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47 (2002).
22. Newman, M. *Networks* (Oxford University Press, 2018).
23. Cohen, R. & Havlin, S. *Complex Networks: Structure, Robustness and Function* (Cambridge University Press, 2010).
24. Tsonis, A. A. & Roebber, P. J. The architecture of the climate network. *Phys. A Stat. Mech. Appl.* **333**, 497–504 (2004).
25. Yamasaki, K., Gozolchiani, A. & Havlin, S. Climate networks around the globe are significantly affected by El Niño. *Phys. Rev. Lett.* **100**, 228501 (2008).
26. Boers, N. et al. Prediction of extreme floods in the eastern central Andes based on a complex networks approach. *Nat. Commun.* **5**, 5199 (2014).
27. Boers, N. et al. Complex networks reveal global pattern of extreme-rainfall teleconnections. *Nature* **566**, 373–377 (2019).
28. Feng, Q. Y. & Dijkstra, H. Are North Atlantic multidecadal SST anomalies westward propagating? *Geophys. Res. Lett.* **41**, 541–546 (2014).
29. Dijkstra, H. A., Hernández-García, E., Masoller, C. & Barreiro, M. *Networks in Climate* (Cambridge University Press, 2019).
30. Ludescher, J. et al. Network-based forecasting of climate phenomena. *Proc. Natl. Acad. Sci.* **118**, e1922872118 (2021).
31. Ludescher, J. et al. Improved El Niño forecasting by cooperativity detection. *Proc. Natl. Acad. Sci.* **110**, 11742–11745 (2013).
32. Meng, J. et al. Complexity-based approach for El Niño magnitude forecasting before the spring predictability barrier. *Proc. Natl. Acad. Sci.* **117**, 177–183 (2020).
33. Fan, J. et al. Network-based approach and climate change benefits for forecasting the amount of indian monsoon rainfall. *J. Clim.* **35**, 1009–1020 (2022).
34. Johnson, S. J. et al. Seas5: the new ECMWF seasonal forecast system. *Geosci. Model Dev.* **12**, 1087–1117 (2019).
35. Saha, S. et al. The NCEP climate forecast system version 2. *J. Clim.* **27**, 2185–2208 (2014).
36. Abdalla, A. A. & Abdel Nour, H. The agricultural potential of Sudan. *Exec. Intell. Rev.* **28**, 37–45 (2001).
37. Willmott, C. J. & Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **30**, 79–82 (2005).
38. De Myttenaere, A., Golden, B., Le Grand, B. & Rossi, F. Mean absolute percentage error for regression models. *Neurocomputing* **192**, 38–48 (2016).
39. Cohen, I. et al. Pearson correlation coefficient. *Noise Reduction in Speech Processing 1–4* (Springer, 2009).
40. Ferreira, G. W., Reboita, M. S. & Drumond, A. Evaluation of ECMWF-SEAS5 seasonal temperature and precipitation predictions over South America. *Climate* **10**, 128 (2022).
41. Chevuturi, A. et al. Forecast skill of the indian monsoon and its onset in the ECMWF seasonal forecasting system 5. *Clim. Dyn.* **56**, 2941–2957 (2021).
42. Lang, Y. et al. Evaluating skill of seasonal precipitation and temperature predictions of NCEP CFSv2 forecasts over 17 hydroclimatic regions in China. *J. Hydrometeorol.* **15**, 1546–1559 (2014).
43. Gelbrecht, M., Boers, N. & Kurths, J. Phase coherence between precipitation in South America and Rossby waves. *Sci. Adv.* **4**, eaau3191 (2018).
44. Changnon, S. A. *El Niño 1997-1998: the Climate Event of the Century* (Oxford University Press, 2000).
45. Ronghui, H. & Yifang, W. The influence of ENSO on the summer climate change in China and its mechanism. *Adv. Atmos. Sci.* **6**, 21–32 (1989).
46. Yuan, Y. et al. The 2016 summer floods in China and associated physical mechanisms: a comparison with 1998. *J. Meteorol. Res.* **31**, 261–277 (2017).
47. Kripalani, R. H. & Kulkarni, A. Monsoon rainfall variations and teleconnections over south and east Asia. *Int. J. Climatol. A J. R. Meteorol. Soc.* **21**, 603–616 (2001).
48. Nooteboom, P. D., Feng, Q. Y., López, C., Hernández-García, E. & Dijkstra, H. A. Using network theory and machine learning to predict El Niño. *Earth Syst. Dyn.* **9**, 969–983 (2018).
49. Petersik, P. J. & Dijkstra, H. A. Probabilistic forecasting of El Niño using neural network model. *Geophys. Res. Lett.* **47**, e2019GL086423 (2020).

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 42450183, 12275020, 12135003, 12205025, and 42461144209) and the Ministry of Science and Technology of China (2023YFE0109000). J.F. is supported by the Fundamental Research Funds for the Central Universities.

## Author contributions

J.M. and J.F. designed the research. G.R. performed the analysis. G.R., J.M., and J.F. prepared the manuscript, G.R., J.M., and J.F. discussed results, and contributed to writing the manuscript. J.M. and J.F. led the writing of the manuscript.

## Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s43247-025-02391-1>.

**Correspondence** and requests for materials should be addressed to Jun Meng or Jingfang Fan.

**Peer review information** *Communications Earth & Environment* thanks and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary handling editor: A.L. [A peer review file is available].

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025