Geophysical Research Letters

RESEARCH LETTER
10.1029/2024GL111076

Advancing Subseasonal Surface Air Temperature and Heat Wave Prediction Skill in China by Incorporating Scale Interaction in a Deep Learning Model

Jiehong Xie[1,2], Pang-Chi Hsu[1], Yamin Hu[3], Hualong Zhang[4], and Mengxi Ye[5]

[1]Key Laboratory of Meteorological Disaster of Ministry of Education, Joint International Research Laboratory of Climate and Environment Change, Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science & Technology, Nanjing, China, [2]School of Atmospheric Sciences, Sun Yat-sen University and Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, China, [3]Guangdong Climate Center, China Meteorological Administration, Guangzhou, China, [4]Guangdong Meteorological Observatory, China Meteorological Administration, Guangzhou, China, [5]Guangzhou Sub-branch of Guangdong Ecological and Environmental Monitoring Center, Guangzhou, China

Abstract Accurate subseasonal predictions of high surface air temperature (SAT) and heat wave events 10–30 days in advance are crucial for mitigating the risks of extreme weather; however, they pose a challenge for current operational models. In this study, we trained a convolutional neural network (CNN)-based deep learning model to exploit the modulations in China's SAT by precursor signals across different timescales to improve predictions of future SAT and heat wave events. This CNN model demonstrated superior capability in capturing the evolution of SAT anomalies and the occurrence of heat wave events with forecast lead times beyond 20 days, compared with that of the operational models of the China Meteorological Administration and European Centre for Medium-Range Weather Forecasts. Explainability analysis highlighted that subseasonal SAT predictability in China is driven primarily by large-scale intraseasonal perturbations from both lower- and higher-latitude regions of Eurasia and interannual variability.

Plain Language Summary Heat waves, marked by prolonged periods of extremely high surface air temperature (SAT), are among the most devastating natural disasters. Improving the predictive accuracy and extending the forecast lead times for heat wave events are vital for disaster prevention and mitigation. However, current operational models typically exhibit limited skill in predicting heat waves at subseasonal scales (2–6 weeks ahead). This is particularly true in relation to the densely populated regions of southern East China, for which the operational prediction systems of the China Meteorological Administration and European Centre for Medium-Range Weather Forecasts are useful only up to approximately 2 weeks ahead. To address this limitation, we developed a deep learning model using a convolutional neural network approach that can reliably predict SAT anomalies and heat wave events in China with lead times of 10–30 days. Our analysis indicated that successful enhancement of the prediction skill of this model hinges on effectively leveraging the modulating effects of intraseasonal and interannual signals on China's SAT. Rather than focusing solely on specific timescale components, our findings suggest that considering interactions across multiple timescales could enhance subseasonal predictability.

Key Points:

- Summertime surface air temperature (SAT) and heat waves in China are influenced by variabilities across different timescales
- A deep learning model skillfully predicted SAT and heat waves for periods extending beyond 20 days by incorporating scale interactions
- Explainability analysis linked model skill to modulation of local SAT by large-scale intraseasonal and interannual signals

Supporting Information:

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

Correspondence to:

P.-C. Hsu,
pangchi@nuist.edu.cn

Citation:

Xie, J., Hsu, P.-C., Hu, Y., Zhang, H., & Ye, M. (2024). Advancing subseasonal surface air temperature and heat wave prediction skill in China by incorporating scale interaction in a deep learning model. *Geophysical Research Letters*, *51*, e2024GL111076. https://doi.org/10.1029/2024GL111076

Received 2 JUL 2024
Accepted 13 OCT 2024

1. Introduction

Extremely high surface air temperatures (SATs) or heat wave events can cause considerable harm to human health, ecosystems, and social services. According to the 6th Assessment Report of the Intergovernmental Panel on Climate Change (Masson-Delmotte et al., 2021), climate change has intensified heat extremes globally. Enhancing prediction skill for extreme weather events and extending forecast lead times beyond the conventional 10-day weather timescale are crucial for effective risk reduction preparedness. Given the difficulties current operational models face in forecasting extreme events accurately on a subseasonal timescale (2–6 weeks) (Vitart & Robertson, 2018; Xie et al., 2020), the World Meteorological Organization launched the Subseasonal-to-Seasonal (S2S) prediction project to improve predictability of such high-impact events (Vitart et al., 2017).

The equatorial Madden–Julian Oscillation (Madden & Julian, 1971, 1972) and extratropical intraseasonal variability (Hannachi et al., 2017) represent the dominant modes of subseasonal variability in the atmosphere. These

modes critically affect SAT variations in China by inducing anomalous circulations and thermodynamic conditions (Chen et al., 2018; Hsu et al., 2017), recognized as key sources of subseasonal predictability for heat waves across various regions of China (Qi & Yang, 2019; Xie et al., 2020; Zhu et al., 2023). Previous studies effectively used these large-scale intraseasonal signals as predictors in empirical models to enhance subseasonal forecasts in China (Wu et al., 2022; Zhu & Li, 2018). Additionally, perturbations related to the Madden–Julian Oscillation and extratropical intraseasonal variability are influenced by background conditions associated with the El Niño–Southern Oscillation and other interannual modes of sea surface temperature (Liu et al., 2016; Pourasghar et al., 2019). Incorporating El Niño–Southern Oscillation signals in statistical models can thus enhance subseasonal prediction capabilities, particularly in relation to North American SAT (Johnson et al., 2014) and tropical cyclone activity in the western North Pacific (Qian et al., 2024).

Although the sources of subseasonal predictability in extreme weather events are well recognized, effectively leveraging this information to enhance forecasting skill presents another challenge. With increase in data availability and advances in computing power, the use of machine learning algorithms, especially deep learning, has gained increasing attention in relation to weather forecasting (Bi et al., 2023; Lam et al., 2023; Zhang et al., 2023) and long-term climate prediction (Ham et al., 2019; Ling et al., 2022; Tang & Duan, 2021). However, their application to subseasonal predictions remains in the nascent stage. Recent studies employed deep learning models, such as Convolutional Neural Networks (CNNs) (Lecun et al., 2015), to correct biases in the Madden–Julian Oscillation (Kim et al., 2021), rainfall and SAT predictions (Lyu et al., 2023; Zhou & Liu, 2024; Zhou & Zhao, 2023), and to forecast regional rainfall and SAT conditions using preceding signals (Weirich-Benet et al., 2023; Xie et al., 2023). Compared to dynamical and statistical-dynamical hybrid prediction approaches, statistical prediction generally requires fewer computing resources and allows for easier adjustments to model structures, facilitating the investigation of predictability sources. Therefore, this study aims to develop a skillful statistical model for SAT and heat wave predictions in China by employing deep learning-based methods. Furthermore, we seek to enhance the understanding of their subseasonal predictability, with a particular focus on the role of various timescale information, through a series of forecast experiments and explainability analysis.

2. Data, Methods and Model Constructions

2.1. Data

Observational daily maximum SAT data from across China, with 0.5° × 0.5° resolution, were obtained from the CN05 data set (Xu et al., 2009). Large-scale fields used as potential predictors included air temperature at 850 hPa (T850); zonal winds at 200, 500, and 850 hPa (U200, U500, and U850, respectively); meridional winds at 200, 500, and 850 hPa (V200, V500, and V850, respectively); geopotential height fields at 200, 500, and 850 hPa (H200, H500, and H850, respectively); relative humidity at 700 hPa (RH700); precipitable water (PW); sea level pressure; surface soil moisture (SSM; top 0–10 cm); and outgoing longwave radiation (OLR). These variables were sourced from the National Centers for Environmental Prediction/Department of Energy (NCEP/DOE) reanalysis (Kistler et al., 2001), except the SSM and OLR data, which were obtained from the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis (Kalnay et al., 1996) and National Oceanic and Atmospheric Administration (NOAA) polar-orbiting satellites (Liebmann & Smith, 1996), respectively. All large-scale fields have horizontal resolution of 2.5° × 2.5°, except that of SSM, which uses a T62 Gaussian grid format (192 × 94). Before model training, the SSM field was regridded to horizontal resolution of 2.5° × 2.5°. The data obtained covered the period 1979–2021, and the study focused on predictions during boreal summer (May–October).

The forecast skill of the deep learning model was compared with reforecast products from the operational models of the China Meteorological Administration (CMA) and the European Centre for Medium-Range Weather Forecasts (ECMWF), which participated in the S2S prediction project (Vitart et al., 2017). For details regarding the S2S models and their reforecast data, see Text S1 in Supporting Information S1.

2.2. Analysis Methods

A heat wave event was identified in both observations and predictions when the gridded SAT anomaly exceeded the 90th percentile for at least three consecutive days. Although an absolute threshold of 35°C is used by operational centers to define high-temperature days, both statistical and dynamical models show systematic biases in SAT magnitudes (Figure S1 in Supporting Information S1). To ensure more useful predictive information and

enable an efficient comparison of model performance, a relative threshold is therefore applied for the identification of extreme events. The forecast skills for SAT evolution and heat wave occurrence were assessed using several metrics. The temporal correlation coefficient (TCC) and pattern correlation coefficient (PCC) can evaluate the similarities in the temporal evolution and spatial patterns, respectively, between the predicted SAT anomalies and observations. The root mean square error (RMSE) quantifies the magnitude of bias. The extremal dependence index (EDI; Ferro & Stephenson, 2011), defined in Equation 1, can evaluate the accuracy of heat wave occurrence predictions:

$$\mathrm{EDI} = \frac{\log F - \log H}{\log F + \log H}, \tag{1}$$

where H represents the hit rate, defined as the ratio of correctly forecast observed heat waves to the total observed heat waves, and F is the false alarm rate, calculated as the ratio of false alarms to the total number of observed non-events. The EDI scale ranges from -1 to 1, with 1 indicating perfect prediction skill, while negative values represent incorrect predictions. EDI values greater than 0 indicate skillfulness; a higher value signifies better model performance in predicting extreme events.

We utilized the Gradient-weighted Class Activation Mapping (Grad-CAM) method (Selvaraju et al., 2020) to interpret why CNNs yield useful predictions. Grad-CAM, designed specifically for CNN-based models, is typically applied to the final convolutional layer within a specific convolutional block. It computes the gradients of the target output with respect to the specified layer, averages them across each output channel, and then multiplies the average gradients by the layer activations. By analyzing the resulting attribution map, it is possible to identify the major contributors to forecast accuracy.

2.3. Construction of the Deep Learning Model

In developing the deep learning model to focus on the subseasonal timescale, for which day-to-day weather is largely unpredictable, we defined the predictand as SAT anomalies by removing high-frequency signals (<10 days). Predicted heat wave occurrences were identified when predicted SAT anomalies met certain criteria (Section 2.2). For real-time applications, we employed a non-band-pass filtering method (Hsu et al., 2015) to extract intraseasonal (10–90-day) components and the low-frequency background state (LFBS; monthly anomaly). The filtering process for intraseasonal variability involves eliminating the climatological annual cycle through subtracting the climatological daily mean from the raw data. Then, the average values of the preceding 45 days are subtracted from the daily anomaly to eliminate signals exceeding 90 days. Finally, a 5-day mean over the preceding 5 days is applied to suppress high-frequency disturbances.

To reduce computing resources, we applied empirical orthogonal function (EOF) analysis to the summer SAT data throughout China (Figure S2 in Supporting Information S1). The principal components (PCs) of the first six leading modes, accounting for 73% of the total variance, were selected as outputs (i.e., labels) for model training. Using the predicted PCs and observed EOF modes, the SAT patterns were reconstructed, and heat wave occurrences were then defined as the final prediction products.

Potential predictors were selected based on previous studies identifying precursory intraseasonal signals linked to heat wave occurrences in China (Hsu et al., 2017; Qi & Yang, 2019; Zhu et al., 2023). To validate their influence on SAT anomalies in China, lead–lag correlations between the potential predictors and the individual PCs, used as predictands, were analyzed. Figures S3 and S4 in Supporting Information S1 represent the correlation maps of large-scale fields with SAT variability corresponding to the first two EOF modes. While these large-scale signals primarily originate from upstream Eurasia and propagate southeastward and southward toward China, they display distinct spatial structures and follow different paths, penetrating various regions across the country. We summarized the results and found that the subseasonal perturbations propagating from three main pathways are critical in modulating summer SAT variability over China: (a) southeastward-propagating signals originating from northwestern Eurasia, (b) southeastward-propagating signals originating from Northeast Asia, and (c) northward-propagating signals initiated from the tropical Western Pacific and Indian Ocean. The key preceding LFBS signals affecting the SAT modes were also found important over Eurasia and the tropical oceans. Consequently, the domain for these predictors was established over the statistically significant areas, spanning Eurasia and the tropical regions (2.5°S–75.0°N, 40°–167.5°E).

Given the advantages of CNN in spatial signal recognition, efficient parallel training, relatively low computational cost, and interpretability (Delaunay & Christensen, 2022; Shin et al., 2024; Xie et al., 2023; Yang et al., 2023), we employed it to develop SAT prediction models that incorporate the source of predictability provided by multiscale variability. In approaches based on deep learning, a notable challenge often arises when the size of the training data set is relatively small, potentially leading to overfitting. To address this issue, we implemented a pre-training model strategy (Yosinski et al., 2014) to improve the generalization performance of the model while efficiently utilizing the signals from optimal predictors. We first trained individual models separately based on candidate predictors. Note that data from all seasons rather than solely from summer were used for model training. This approach produced models that outperformed those trained on summer data. Then, we selected the best four intraseasonal predictors and three LFBS predictors (Table S1 in Supporting Information S1) according to the performance of the individual model predictions, constructed in the first step, at the forecast lead time of 20 days during the validation period. The number of predictors used (four for the intraseasonal predictors and three for the LFBS predictors) was determined based on a set of sensitivity experiments, in which the current number of predictors showed the most stable and skillful results (figures not shown). Finally, we trained a multivariable CNN with the signals from the seven selected predictors for each PC of the SAT modes. The predictions by these models were assessed against observations during the independent forecast period. Thus, we divided the historical data into three phases: 1979–2009 as the training period, 2010–2015 as the validation period for selecting optimal predictors, and 2016–2021 as the independent forecast period (totaling 1,104 time points during the studied summers).

Figure 1 displays the architectural design of the CNN-based model. The individual predictor model (Figure 1a) includes two input layers, a convolutional block with three 2D convolutional (Conv2d) layers and two max-pooling (MP) layers, a fully connected layer, and an output layer. The first input layer's feature map (orange boxes) has dimensions $N \times 32 \times 52$, representing values over the domain 2.5°S–75.0°N (32 grid points) and 40°–167.5°E (52 grid points), derived from the preceding 30 days at 5-day intervals for large-scale subseasonal anomalies ($N = 6$). For the LFBS predictors, the feature maps include the LFBS (i.e., monthly anomaly) components from the previous 90 days at 30-day intervals ($N = 3$) in the same region. The second input layer (red box) encodes initial month information to understand the annual cycle of subseasonal variability (Liu et al., 2020). The output layer (black box) predicts future PC values at forecast lead times of 10, 15, 20, 25, or 30 days.

The multivariate CNN (Figure 1b) has seven input layers to encompass the predictable information provided by the four intraseasonal predictors and the three LFBS predictors. For example, when forecasting PC1 at a 20-day lead, the predictors include subseasonal perturbations of PW, T850, U850, and OLR, together with the LFBS components of T850, PW, and H500 (Table S1 in Supporting Information S1). The pre-trained convolutional blocks (Figure 1a) from these predictors were utilized for training. After convolutional block processing, the features from the multivariate predictors are concatenated across the channel dimensions and undergo two convolution operations. This approach enables leveraging of the joint predictive information from the optimal predictors, considering the interactions between intraseasonal and LFBS variability. Details on model hyperparameters can be found in Text S2 in Supporting Information S1.

3. Evaluations of Predicted SAT and Heat Waves by CNN-Based Model

The TCC skills for SAT prediction, measuring the models' capability to predict the timing of hot and cold periods, performed by the deep learning model and the operational models of the CMA and ECMWF are compared in Figure 2. The deep learning model shows lower skill at lead times of 2–3 pentads (Figures 2a and 2b) than the dynamical models (Figures 2f, 2g, 2k, and 2l), but outperforms the CMA and ECMWF predictions for forecast lead times extending beyond 20 days (Figures 2c–2e, 2h–2j, and 2m–2o). Even for a long forecast lead of six pentads, the deep learning model still offers useful predictions across most of China (Figure 2e), whereas the dynamical models excel mainly over western and southwestern China (Figures 2j and 2o).

The prediction fidelity of the spatial distribution of SAT anomalies over China was evaluated using PCC skill (Figure S5 in Supporting Information S1). We computed the PCC at each forecast time point and calculated the temporal mean (Figure S5a in Supporting Information S1). Similar to the TCC skill, the CNN-based predictions cannot compete with the dynamical models within the lead time of three pentads. However, they show superior prediction capability 4–6 pentads in advance compared with that of the CMA and ECMWF models. The results are further confirmed by estimating the percentage of time points with statistically significant PCC values (Figure

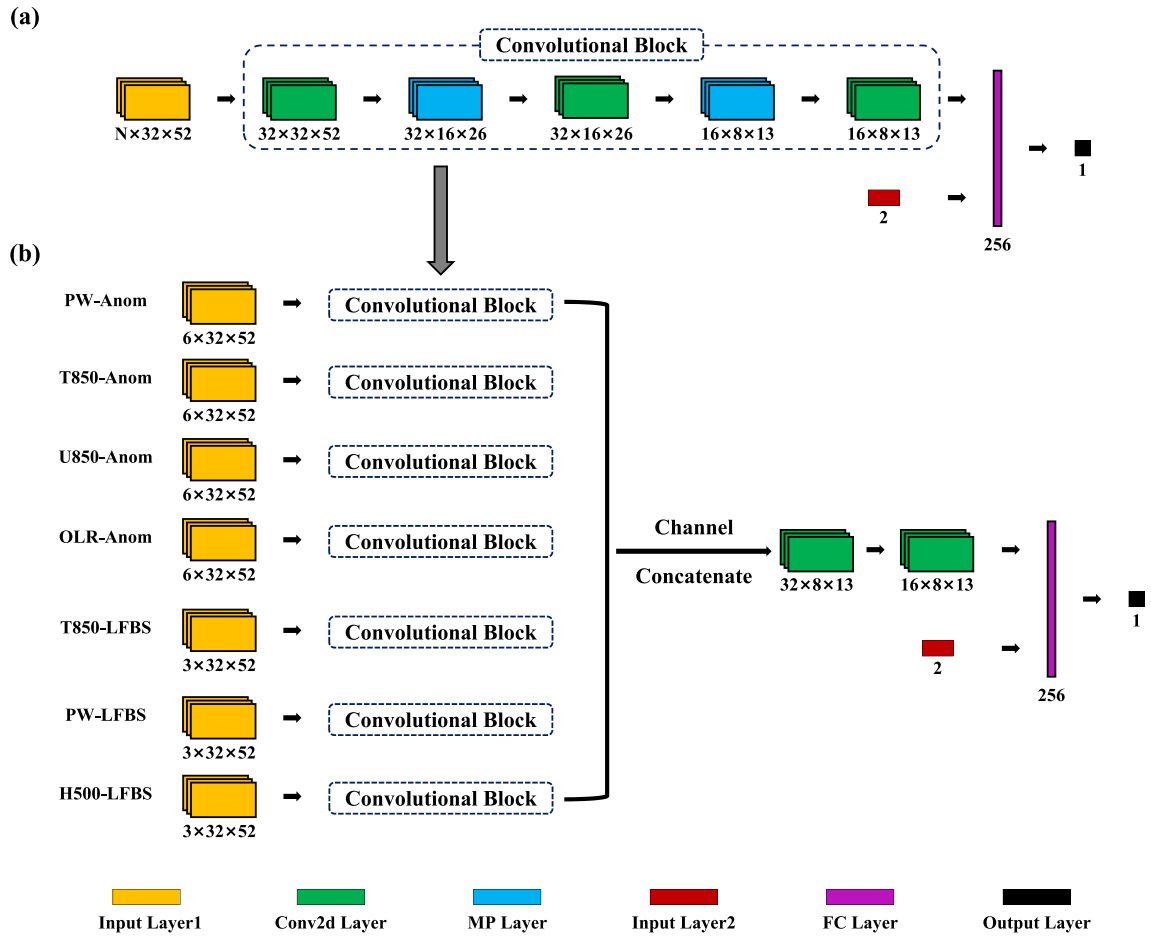

Figure 1. Architecture of the (a) pre-training Convolutional Neural Network (CNN) and (b) multivariate CNN. Different colors represent various layers, and their dimensions are indicated in the bottom corners of each part. Input layers highlighted in orange comprise predictors from different preceding time steps, whereas input layers indicated in dark red provide encoded information related to the initial prediction month. A convolutional block, consisting of three Conv2d layers and two MP layers, is designed to extract latent features. The fully connected layer is employed to link the extracted features from the input to the output layer (colored black), which represents the predictand at a future time step (i.e., 10, 15, 20, 25, or 30 days).

S5b in Supporting Information S1). Regarding prediction amplitude bias, the deep learning model exhibits smaller RMSE values than the dynamical models when the forecast lead times are beyond 20 days (Figure S6 in Supporting Information S1), with domain-averaged RMSEs over China consistently smaller than one standard deviation for forecast leads from 10 to 30 days. Generally, regions with higher TCCs exhibit small prediction errors (Figure 2 and Figure S4 in Supporting Information S1).

Figure 3 further compares the EDI among the CNN-based and dynamical predictions. Both the deep learning and the dynamical models can predict partial heat waves (EDI > 0) across most of China at lead times of 2–6 pentads. In the short-term extended-range forecast (within three pentads), the ECMWF model demonstrates the highest skill (Figures 3k and 3l), with the CMA model also performing well at the two-pentad forecast lead (Figure 3f). However, in longer-term predictions (beyond 20 days), the deep learning model outperforms the dynamical models (lower three panels in Figure 3), particularly over the densely populated area of East China.

Based on the Grad-CAM method (Section 2.2), Figure 4 illustrates the attribution maps when the CNN-based model accurately forecasts hot periods (PC1 exceeding the 90th percentile threshold) 20 days in advance. Figures 4a–4g display results from the final Conv2d layers of each convolutional block, highlighting the critical features from each predictor over Northwest Eurasia and the tropical region that are important for accurate predictions. Their detailed evolutions are confirmed by Figures S7 and S8 in Supporting Information S1, where signals from both the tropics and the extratropics are observed. This suggests that the deep learning model can identify the origin and propagation of precursor intraseasonal signals related to hot periods. It also demonstrates

10.1029/2024GL111076

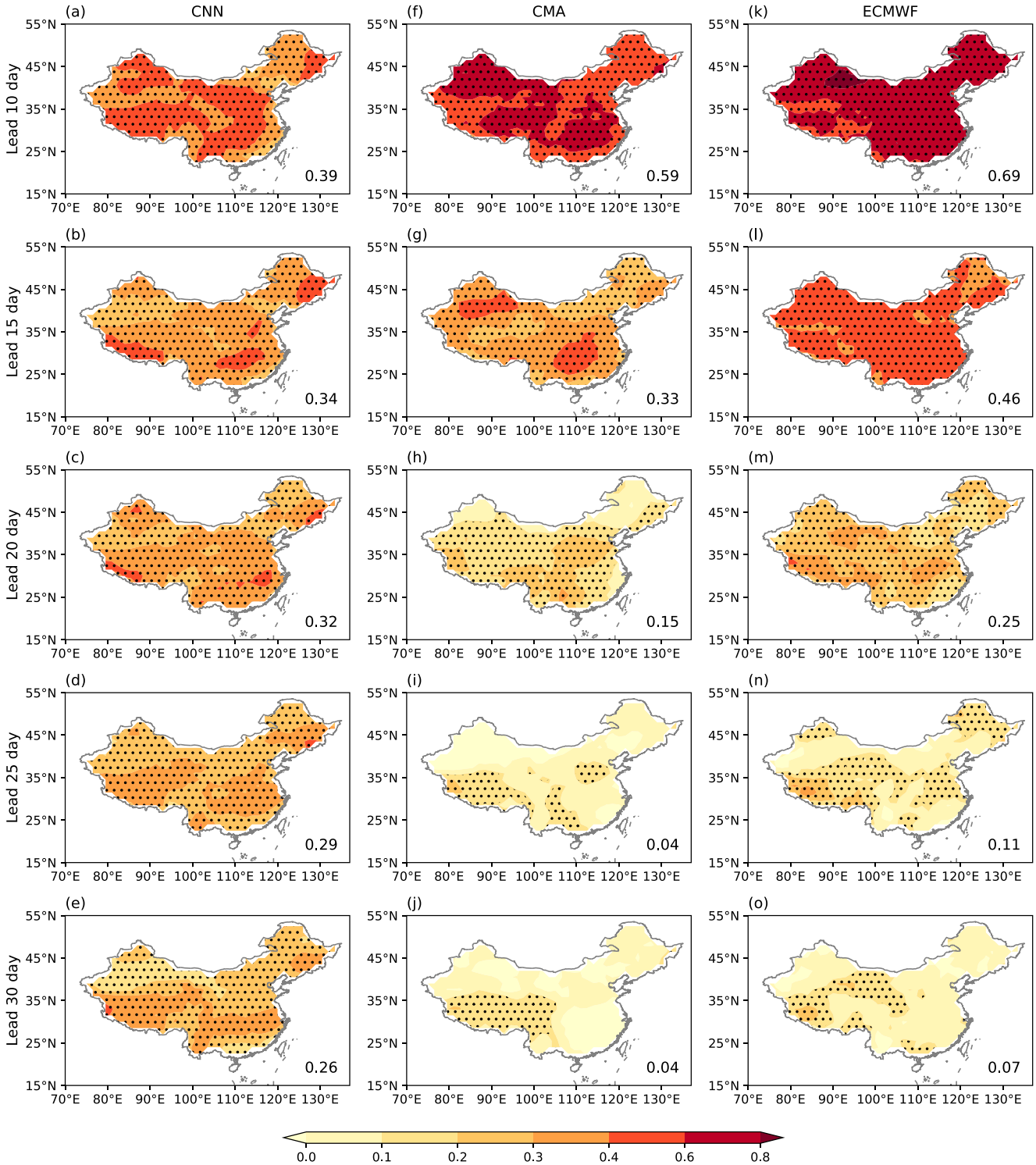

Figure 2. Distributions of the temporal correlation coefficient (TCC) skill of the (a)–(e) Convolutional Neural Network, (f)–(j) China Meteorological Administration, and (k)–(o) European Centre for Medium-Range Weather Forecasts models at forecast lead times of 10–30 days (2–6 pentads), evaluated based on results during the independent prediction period. Stippling indicates regions with useful prediction, with a significant TCC at the 99% confidence level. The area-mean TCC value over China is displayed in the lower-right corner of each panel.

10.1029/2024GL111076

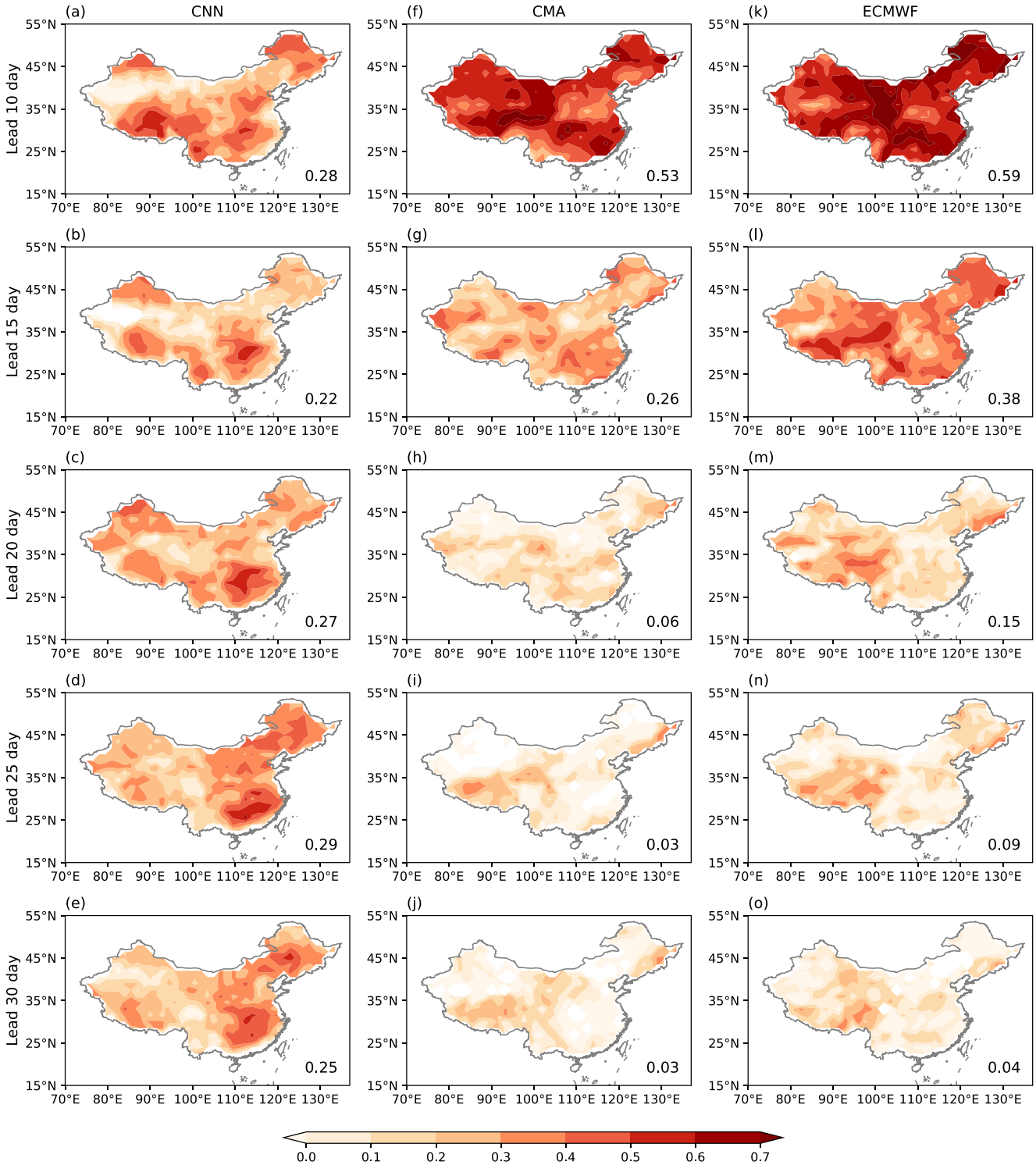

Figure 3. Heat wave prediction skill. As in Figure 2 but for the EDI skill of predicted heat waves.

the model's capability to understand how interannual variability modulates local SAT variations. The attribution map from the final Conv2d layer (Figure 4h) demonstrates how all predictors interact to produce predictions in the multivariable CNN model (Figure 1b). Input features over Northwest Eurasia, Northeast Asia, the tropical Indian Ocean, and local regions in China all contribute to the 20-day lead forecasts (Figure 4h).

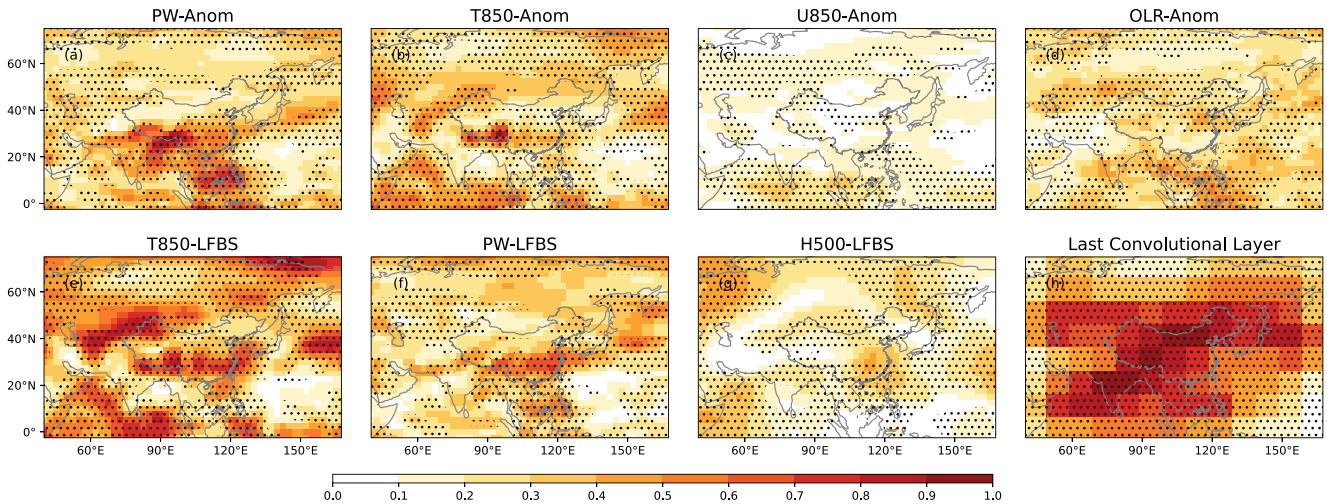

Figure 4. Attribution maps corresponding to accurate forecasts of hot periods of the first EOF mode at a lead time of 20 days during the independent forecast period. (a)–(g) Attribution maps derived from the final convolutional layer in each convolutional block, associated with the intraseasonal predictors of precipitable water (PW), T850, U850, and outgoing longwave radiation (OLR) and the LFBS predictors of T850, PW, and H500, respectively. (h) Similar to (a)–(g) but from the final convolutional layer in the model. All values are normalized to the range 0–1. Stippling indicates regions with significant differences at the 99% confidence level.

4. Conclusion and Discussion

To bridge the gap between short-term weather forecasts and long-term climate predictions among current operational systems, we developed a CNN-based model that considers multiscale and multivariate interaction processes. This deep learning model demonstrated its capability to predict SAT anomalies and heat waves over China at forecast lead times of 10–30 days. Based on TCC, PCC, RMSE, and EDI metrics, its performance surpassed that of two dynamical models from the S2S prediction project (i.e., the CMA and ECMWF models) in predicting SAT and heat waves beyond a 20-day forecast lead. Explainability analysis revealed that the evolution of large-scale intraseasonal perturbations over the tropics and the mid–higher-latitude regions of Eurasia, and the modulation of interannual variability on subseasonal disturbances, are notable sources of predictability for SAT and heat waves across China.

To provide insights for future enhancements of subseasonal prediction models, we discussed potentially effective strategies based on our encouraging results. When the model employed the same CNN structures as shown in Figure 1, but used only intraseasonal predictors (Figure S9a in Supporting Information S1) and only LFBS predictors (Figure S9b in Supporting Information S1), the forecast skill decreased by 5%–15% and by 25%–60%, respectively, compared with that of the model that utilizes both types of information (green and brown bars vs. blue bars in Figure S10 in Supporting Information S1). Including the spatiotemporal evolution features of predictors also enhanced the forecasting capability. When only the most recent preceding intraseasonal information was used ($N = 1$ for intraseasonal predictors; Figure S9c in Supporting Information S1), the prediction skills in terms of SAT and heat waves dropped by 5%–30% compared with that of the models utilizing the preceding 30-day information ($N = 6$) (purple vs. green bars in Figure S10 in Supporting Information S1). Regarding the model training approach, we emphasize the importance of additional convolutional operations to integrate the collective information from various predictors (Figure 1b). Without this integration, ensemble predictions derived from averaging individual predictor outputs exhibit lower skill (red vs. green bars in Figure S10 in Supporting Information S1). Moreover, omitting pre-trained weights also leads to reduced prediction skill (orange vs. blue bars in Figure S10 in Supporting Information S1), highlighting the advantages of pre-training to enhancement of the generalization capability of the model.

To verify the advantages of accounting for nonlinear connections between large-scale predictors and the predictand, we compared the prediction skill of the previously proposed statistical model based on the spatial-temporal projection method (STPM) via singular value decomposition (Hsu et al., 2015). As demonstrated in Zhu and Li (2018), the STPM shows skillful SAT prediction in China at the extended range. However, its ability to forecast SAT variability and heat wave occurrence appears to be lower than that of the CNN-based model

beyond a forecast lead time of 5 days (Figure S11 in Supporting Information S1). This suggests that the superior performance of the CNN model is partly attributable to its ability to capture the nonlinear processes influencing SAT and heat wave activity.

Although the current CNN model demonstrates encouraging results, there is still considerable potential for enhancing its predictive accuracy. As deep learning methods continue to evolve rapidly, various modules offer distinct advantages. Integrating the strengths of different deep learning techniques into a hybrid model could provide a valuable approach for further enhancing predictive performance. Besides, applying deep learning for bias correction is another promising direction (Jin et al., 2022). Our evaluations show that both statistical CNN models and dynamical S2S models tend to misestimate SAT amplitude. Implementing bias correction methods or developing dynamic-statistical hybrid models with deep learning could further improve prediction skill. Finally, optimizing the selection of training domains and potential predictors is also a crucial area for future research. In this study, we primarily utilized data from the Eurasia, tropical Pacific and tropical Indian Ocean regions, potentially overlooking the influence of Atlantic and southern Indian Ocean sea surface temperature anomalies on extreme events in China (Dong, Yang, Cao, Wang, & Cheng, 2023, Dong, Yang, Cao, Wang, & Yang, 2023). In our prediction experiments, we expanded the training domain (CNN-Global) and included sea surface temperature as a predictor, but the skill slightly decreased (Text S3 and Figure S10 in Supporting Information S1), likely due to overfitting. Incorporating more data from reanalysis and model predictions may help achieve more stable and accurate results.

Data Availability Statement

The NCEP/DOE and NCEP/NCAR reanalysis were acquired from the NOAA NCEP (https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html and https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html), and the OLR data are available from NOAA (https://psl.noaa.gov/data/gridded/data.olrcdr.interp.html). The reforecast data from two dynamical models (CMA and ECMWF) were downloaded from the server of the S2S database (http://apps.ecmwf.int/datasets/data/s2s).

Acknowledgments
We sincerely appreciate the insightful comments from the anonymous reviewers, which have significantly enhanced the quality of this study. This study is supported by the National Natural Science Foundation of China (42225502 & 42075173), Special Fund of CMA Youth Innovation Team Project (CMA2024QN01), Special Fund of CMA for Innovation and Development (CXFZ2023J027) and Special Fund for Forecasters of CMA (CMAYBY2020-094). We appreciate the operational centers for providing their model outputs through the S2S database. We also acknowledge the High Performance Computing Center of Nanjing University of Information Science and Technology for their support of this work.

References

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, *619*(7970), 533–538. https://doi.org/10.1038/s41586-023-06185-3

Chen, R., Wen, Z., & Lu, R. (2018). Large-scale circulation anomalies and intraseasonal oscillations associated with long-lived extreme heat events in South China. *Journal of Climate*, *31*(1), 213–232. https://doi.org/10.1175/JCLI-D-17-0232.1

Delaunay, A., & Christensen, H. M. (2022). Interpretable deep learning for probabilistic MJO prediction. *Geophysical Research Letters*, *49*(16), e2022GL098566. https://doi.org/10.1029/2022GL098566

Dong, Z., Yang, R., Cao, J., Wang, L., & Cheng, J. (2023). A strong high-temperature event in late-spring 2023 in Yunnan province, Southwest China: Characteristics and possible causes. *Atmospheric Research*, *295*, 107017. https://doi.org/10.1016/j.atmosres.2023.107017

Dong, Z., Yang, R., Cao, J., Wang, L., & Yang, G. (2023). Sea surface temperature anomalies in different ocean basins affecting the interannual variations of summer precipitation in low-latitude highlands of Southeast Asia. *Climate Dynamics*, *61*(11–12), 5517–5531. https://doi.org/10.1007/s00382-023-06868-7

Ferro, C. A. T., & Stephenson, D. B. (2011). Extremal dependence indices: Improved verification measures for deterministic forecasts of rare binary events. *Weather and Forecasting*, *26*(5), 699–713. https://doi.org/10.1175/waf-d-10-05030.1

Ham, Y.-G., Kim, J.-H., & Luo, J.-J. (2019). Deep learning for multi-year ENSO forecasts. *Nature*, *573*(7775), 568–572. https://doi.org/10.1038/s41586-019-1559-7

Hannachi, A., Straus, D. M., Franzke, C. L. E., Corti, S., & Woollings, T. (2017). Low-frequency nonlinearity and regime behavior in the Northern Hemisphere extratropical atmosphere. *Reviews of Geophysics*, *55*(1), 199–234. https://doi.org/10.1002/2015RG000509

Hsu, P.-C., Lee, J.-Y., Ha, K.-J., & Tsou, C.-H. (2017). Influences of boreal summer intraseasonal oscillation on heat waves in monsoon Asia. *Journal of Climate*, *30*(18), 7191–7211. https://doi.org/10.1175/JCLI-D-16-0505.1

Hsu, P.-C., Li, T., You, L., Gao, J., & Ren, H.-L. (2015). A spatial–temporal projection model for 10–30 day rainfall forecast in South China. *Climate Dynamics*, *44*(5–6), 1227–1244. https://doi.org/10.1007/s00382-014-2215-4

Jin, W., Zhang, W., Hu, J., Weng, B., Huang, T., & Chen, J. (2022). Using the residual network module to correct the sub-seasonal high temperature forecast. *Frontiers in Earth Science*, *9*, 760766. https://doi.org/10.3389/feart.2021.760766

Johnson, N. C., Collins, D. C., Feldstein, S. B., L'Heureux, M. L., & Riddle, E. E. (2014). Skillful wintertime North American temperature forecasts out to 4 weeks based on the state of ENSO and the MJO. *Weather and Forecasting*, *29*(1), 23–38. https://doi.org/10.1175/WAF-D-13-00102.1

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437–472. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Kim, H., Ham, Y. G., Joo, Y. S., & Son, S. W. (2021). Deep learning for bias correction of MJO prediction. *Nature Communications*, *12*(1), 3087. https://doi.org/10.1038/s41467-021-23406-3

Kistler, R., Collins, W., Saha, S., White, G., Woollen, J., Kalnay, E., et al. (2001). The NCEP–NCAR 50-Year Reanalysis: Monthly means CD-ROM and documentation. *Bulletin of the American Meteorological Society*, *82*(2), 247–268. https://doi.org/10.1175/1520-0477(2001)082,0247:TNNYRM.2.3.CO;2

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., et al. (2023). Learning skillful medium-range global weather forecasting. *Science*, *382*(6677), 1416–1421. https://doi.org/10.1126/science.adi2336
Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539
Liebmann, B., & Smith, C. A. (1996). Description of a complete (interpolated) outgoing longwave radiation dataset. *Bulletin of the American Meteorological Society*, *77*(6), 1275–1277.
Ling, F., Luo, J.-J., Li, Y., Tang, T., Bai, L., Ouyang, W., & Yamagata, T. (2022). Multi-task machine learning improves multi-seasonal prediction of the Indian Ocean Dipole. *Nature Communications*, *13*(1), 7681. https://doi.org/10.1038/s41467-022-35412-0
Liu, F., Li, T., Wang, H., Deng, L., & Zhang, Y. (2016). Modulation of boreal summer intraseasonal oscillations over the western North Pacific by ENSO. *Journal of Climate*, *29*(20), 7189–7201. https://doi.org/10.1175/JCLI-D-15-0831.1
Liu, F., Ouyang, Y., Wang, B., Yang, J., Ling, J., & Hsu, P.-C. (2020). Seasonal evolution of the intraseasonal variability of China summer precipitation. *Climate Dynamics*, *54*(11–12), 4641–4655. https://doi.org/10.1007/s00382-020-05251-0
Lyu, Y., Zhu, S., Zhi, X., Ji, Y., Fan, Y., & Dong, F. (2023). Improving subseasonal-to-seasonal prediction of summer extreme precipitation over Southern China based on a deep learning method. *Geophysical Research Letters*, *50*(24), e2023GL106245. https://doi.org/10.1029/2023GL106245
Madden, R. A., & Julian, P. R. (1971). Detection of a 40–50 day oscillation in the zonal wind in the tropical pacific. *Journal of the Atmospheric Sciences*, *28*(5), 702–708. https://doi.org/10.1017/CBO9781107415324.004
Madden, R. A., & Julian, P. R. (1972). Description of Global-Scale circulation cells in the tropics with a 40–50 day period. *Journal of the Atmospheric Sciences*, *29*(6), 1109–1123. https://doi.org/10.1175/1520-0469(1972)029<1109:dogscc>2.0.co;2
Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., et al. (2021). IPCC, 2021: Summary for policymakers. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
Pourasghar, F., Oliver, E. C. J., & Holbrook, N. J. (2019). Modulation of wet-season rainfall over Iran by the Madden–Julian Oscillation, Indian Ocean Dipole and El Niño–Southern Oscillation. *International Journal of Climatology*, *39*(10), 4029–4040. https://doi.org/10.1002/joc.6057
Qi, X., & Yang, J. (2019). Extended-range prediction of a heat wave event over the Yangtze River Valley: Role of intraseasonal signals. *Atmospheric and Oceanic Science Letters*, *12*(6), 451–457. https://doi.org/10.1080/16742834.2019.1669408
Qian, Y., Hsu, P.-C., Murakami, H., Gao, J., Wang, H., & Duan, M. (2024). Influences of ENSO and intraseasonal oscillations on distinct tropical cyclone clusters over the western North Pacific. *Climate Dynamics*, *62*(3), 1861–1885. https://doi.org/10.1007/s00382-023-07000-5
Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, *128*(2), 336–359. https://doi.org/10.1007/s11263-019-01228-7
Shin, N.-Y., Kim, D., Kang, D., Kim, H., & Kug, J.-S. (2024). Deep learning reveals moisture as the primary predictability source of MJO. *Npj Climate and Atmospheric Science*, *7*(1), 11. https://doi.org/10.1038/s41612-023-00561-6
Tang, Y., & Duan, A. (2021). Using deep learning to predict the East Asian summer monsoon. *Environmental Research Letters*, *16*(12), 124006. https://doi.org/10.1088/1748-9326/ac34bc
Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., et al. (2017). The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, *98*(1), 163–173. https://doi.org/10.1175/BAMS-D-16-0017.1
Vitart, F., & Robertson, A. W. (2018). The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *Npj Climate and Atmospheric Science*, *1*(1), 3. https://doi.org/10.1038/s41612-018-0013-0
Weirich-Benet, E., Pyrina, M., Jiménez-Esteve, B., Fraenkel, E., Cohen, J., & Domeisen, D. I. V. (2023). Subseasonal prediction of central European summer heatwaves with linear and random forest machine learning models. *Artificial Intelligence for the Earth Systems*, *2*(2), e220038. https://doi.org/10.1175/AIES-D-22-0038.1
Wu, J., Ren, H.-L., Zhang, P., Wang, Y., Liu, Y., Zhao, C., & Li, Q. (2022). The dynamical-statistical subseasonal prediction of precipitation over China based on the BCC new-generation coupled model. *Climate Dynamics*, *59*(3–4), 1213–1232. https://doi.org/10.1007/s00382-022-06187-3
Xie, J., Hsu, P.-C., Hu, Y., Ye, M., & Yu, J. (2023). Skillful extended-range forecast of rainfall and extreme events in East China based on deep learning. *Weather and Forecasting*, *38*(3), 467–486. https://doi.org/10.1175/WAF-D-22-0132.1
Xie, J., Yu, J., Chen, H., & Hsu, P.-C. (2020). Sources of subseasonal prediction skill for heatwaves over the Yangtze River basin revealed from three S2S models. *Advances in Atmospheric Sciences*, *37*(12), 1435–1450. https://doi.org/10.1007/s00376-020-0144-1
Xu, Y., Gao, X. J., Shen, Y., Xu, C. H., Shi, Y., & Giorgi, F. (2009). A daily temperature dataset over China and its application in validating a RCM simulation. *Advances in Atmospheric Sciences*, *26*(4), 763–772. https://doi.org/10.1007/s00376-009-9029-z
Yang, J., Zhu, T., Gao, M., Lin, H., Wang, B., & Bao, Q. (2018). Late-July barrier for subseasonal forecast of summer daily maximum temperature over Yangtze River basin. *Geophysical Research Letters*, *45*(22). https://doi.org/10.1029/2018GL080963
Yang, Y.-M., Kim, J.-H., Park, J.-H., Ham, Y.-G., An, S.-I., Lee, J.-Y., & Wang, B. (2023). Exploring dominant processes for multi-month predictability of western Pacific precipitation using deep learning. *Npj Climate and Atmospheric Science*, *6*(1), 157. https://doi.org/10.1038/s41612-023-00478-0
Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems*, *27*, 3320–3328.
Zhang, Y., Long, M., Chen, K., Xing, L., Jin, R., Jordan, M. I., & Wang, J. (2023). Skilful nowcasting of extreme precipitation with NowcastNet. *Nature*, *619*(7970), 526–532. https://doi.org/10.1038/s41586-023-06184-4
Zhou, J., & Liu, F. (2024). Improving subseasonal forecasting of East Asian monsoon precipitation with deep learning. *Atmospheric and Oceanic Science Letters*, 100520. https://doi.org/10.1016/j.aosl.2024.100520
Zhou, Y., & Zhao, Q. (2023). Taking advantage of quasi-periodic signals for S2S operational forecast from a perspective of deep learning. *Scientific Reports*, *13*(1), 4108. https://doi.org/10.1038/s41598-023-31394-1
Zhu, T., Yang, J., Wang, B., & Bao, Q. (2023). Boreal summer extratropical intraseasonal waves over the Eurasian continent and real-time monitoring metrics. *Journal of Climate*, *36*(12), 3971–3991. https://doi.org/10.1175/JCLI-D-22-0788.1
Zhu, Z., & Li, T. (2018). Extended-range forecasting of Chinese summer surface air temperature and heat waves. *Climate Dynamics*, *50*(5–6), 2007–2021. https://doi.org/10.1007/s00382-017-3733-7