

# Geophysical Research Letters<sup>®</sup>

## RESEARCH LETTER

10.1029/2025GL116339

### Key Points:

- The sea surface temperature (SST) cooling tendency over the north tropical Atlantic contributes to the summer hot days in western North America (WNA)
- Integrating SST tendency as critical input features refined the convolutional neural network (CNN) prediction skills of the hot days in WNA
- The improved prediction skill of CNN relies on the successfully capturing physically meaningful predictor of Atlantic SST tendency

### Supporting Information:

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

### Correspondence to:

Z. Zhu,  
zww@nust.edu.cn

### Citation:

Tan, H., Zhu, Z., Ling, F., & Wang, B. (2025). Refine extreme hot day predictions with the sea surface temperature tendency. *Geophysical Research Letters*, 52, e2025GL116339. <https://doi.org/10.1029/2025GL116339>

Received 3 APR 2025

Accepted 4 SEP 2025

### Author Contributions:

**Conceptualization:** Hui Tan, Zhiwei Zhu, Bin Wang  
**Data curation:** Hui Tan  
**Formal analysis:** Hui Tan, Zhiwei Zhu  
**Funding acquisition:** Zhiwei Zhu  
**Investigation:** Hui Tan, Zhiwei Zhu  
**Project administration:** Zhiwei Zhu  
**Supervision:** Zhiwei Zhu  
**Validation:** Zhiwei Zhu  
**Writing – original draft:** Hui Tan, Zhiwei Zhu  
**Writing – review & editing:** Zhiwei Zhu, Fenghua Ling, Bin Wang

© 2025. The Author(s).

This is an open access article under the terms of the [Creative Commons Attribution License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

## Refine Extreme Hot Day Predictions With the Sea Surface Temperature Tendency



Hui Tan<sup>1,2</sup> , Zhiwei Zhu<sup>1,2</sup> , Fenghua Ling<sup>3</sup>, and Bin Wang<sup>4</sup>

<sup>1</sup>State Key Laboratory of Climate System Prediction and Risk Management/Key Laboratory of Meteorological Disaster, Ministry of Education/Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing, China, <sup>2</sup>School of Atmospheric Sciences, Nanjing University of Information Science and Technology, Nanjing, China, <sup>3</sup>Shanghai AI Laboratory, Shanghai, China, <sup>4</sup>Department of Atmospheric Sciences, University of Hawaii at Manoa, Honolulu, HI, USA

**Abstract** The extreme high temperature in western North America (WNA) exerts profound impacts on industrial and agricultural production, and trigger catastrophic wildfires. Exploring the underlying mechanisms influencing extreme hot days over WNA (WEHDs) and improving the seasonal prediction are of great scientific and social significance. This study reveals that two independent precursor signals, the persistent negative sea surface temperature (SST) anomalies in tropical eastern Pacific and the cooling tendency in tropical North Atlantic SST during springtime exhibit significant influence on WEHDs. A physics-based empirical model constructed using these two predictors exhibits robust independent prediction skills. Guided by the underlying physical mechanisms, we integrate SST tendency fields as critical input features into convolutional neural network (CNN) to further enhance the prediction accuracy. The physically informed CNN achieves significantly improved performance and successfully predicts the extreme WEHD events of 2021. The results emphasize the pivotal role of physical cognition in advancing deep learning-based climate prediction.

**Plain Language Summary** Western North America stands as one of the most vulnerable regions globally to extreme high temperature, rendering the accurate seasonal prediction for such events a matter of critical practical significance. While deep learning has demonstrated considerable success in weather forecast and climate prediction, its application to extreme events prediction remains relatively underexplored. Furthermore, the integration of physical understanding of extreme events with deep learning to improve prediction skills warrants deeper investigation. This study identifies that, aside from the traditional persistent sea surface temperature (SST) precursor, the tendency of the SST independently influences summer extreme hot days in western North America (WNA). By incorporating SST tendency fields as additional input fields in convolutional neural network achieves superior independent skill in prediction of the seasonal mean WNA extreme hot days. The results offer a novel perspective and a methodological guidance for operational agencies and the climate-hydrology research community in leveraging deep learning for seasonal climate prediction.

## 1. Introduction

North America has emerged as one of the regions severely impacted by extreme heatwaves (Grotjahn et al., 2016). For example, the unprecedented western North America (WNA) heatwave during June-July 2021 has caused nearly 1,400 deaths (Lin et al., 2022; X. Zhang et al., 2023), widespread infrastructure damage, and economic losses amounting to worth several billion USD (Lucarini et al., 2023). Under global warming, this region is projected to experience more frequent, prolonged, and intense extreme heat events (Bartusek et al., 2022; Thompson et al., 2022). Concurrent with the increasing frequency of extreme heatwaves, North America has witnessed a marked escalation in the frequency and intensity of wildfires, affecting millions of acres of land and leading to extensive property destruction, ecological degradation, deteriorated air quality, and severe haze pollution (White et al., 2023). Therefore, understanding the physical processes controlling the variability of summer extreme hot days (EHDs) in North America and improving its seasonal prediction are of profound scientific significance and practical implications.

At the seasonal timescale, a tropospheric high-pressure (anticyclonic) anomaly leads to the formation of a heat dome, which triggers heatwaves through clear-sky radiative forcing (Pfahl & Wernli, 2012), subsidence-induced adiabatic heating (Nabizadeh et al., 2021) or low-level anomalously warm horizontal advection (Miralles et al., 2014; Woollings et al., 2018). The formation of a high-pressure anomaly is physically linked to large-scale

teleconnection patterns and remote underlying boundary forcing. For example, anomalous convection over the tropical western Pacific and Maritime Continent region (Jong et al., 2020; Luo & Lau, 2020), the Indian Ocean (Song et al., 2024; B. Wang et al., 2022), along with East Asian summer monsoon (Lopez et al., 2019; Zhu & Li, 2016, 2018), can excite cross-Pacific Rossby wave trains in different shapes, ultimately generating a high-pressure anomaly over North America. The sea surface temperature (SST) anomalies (SSTA) in the tropical Atlantic can induce meridional circulation anomalies that produce anomalous subsidence over North America, reducing local cloud cover and enhancing downward shortwave radiation, thereby favoring heatwaves (Lopez et al., 2022; Ruprich-Robert et al., 2018). The reduced Arctic sea ice can induce a Rossby wave train propagating southward, leading to the high-pressure (anticyclonic) anomaly over North America (Neal et al., 2022; H. Wang et al., 2022). Additionally, the occurrence of EHDs over North America are also modulated by circum-global teleconnection (Luo & Lau, 2020), Arctic Oscillation (Loikith & Broccoli, 2014), Atlantic Multidecadal Oscillation (Ruprich-Robert et al., 2018), Pacific Decadal Oscillation (Hulley et al., 2020), and anthropogenic aerosol forcing (Bercos-Hickey et al., 2022).

The physics-based empirical models (PEM) leveraging these mechanisms have demonstrated moderate success in seasonal prediction (B. Wang et al., 2013, 2015). For instance, Wei et al. (2023) effectively predicted autumn temperature in central North America by incorporating spring western Pacific SST and preceding winter ENSO indices. However, PEM exhibits certain limitations. When the climate system undergoes significant interdecadal changes, previously identified predictors may no longer impact the predictand, leading to substantial degradation in prediction skills (Li et al., 2023; B. Wang et al., 2015). Furthermore, as PEM relies on linear combinations of predictors, the prediction skills significantly diminishes when nonlinear processes start to play roles (Camps-Valls et al., 2025; Y. Wang et al., 2023).

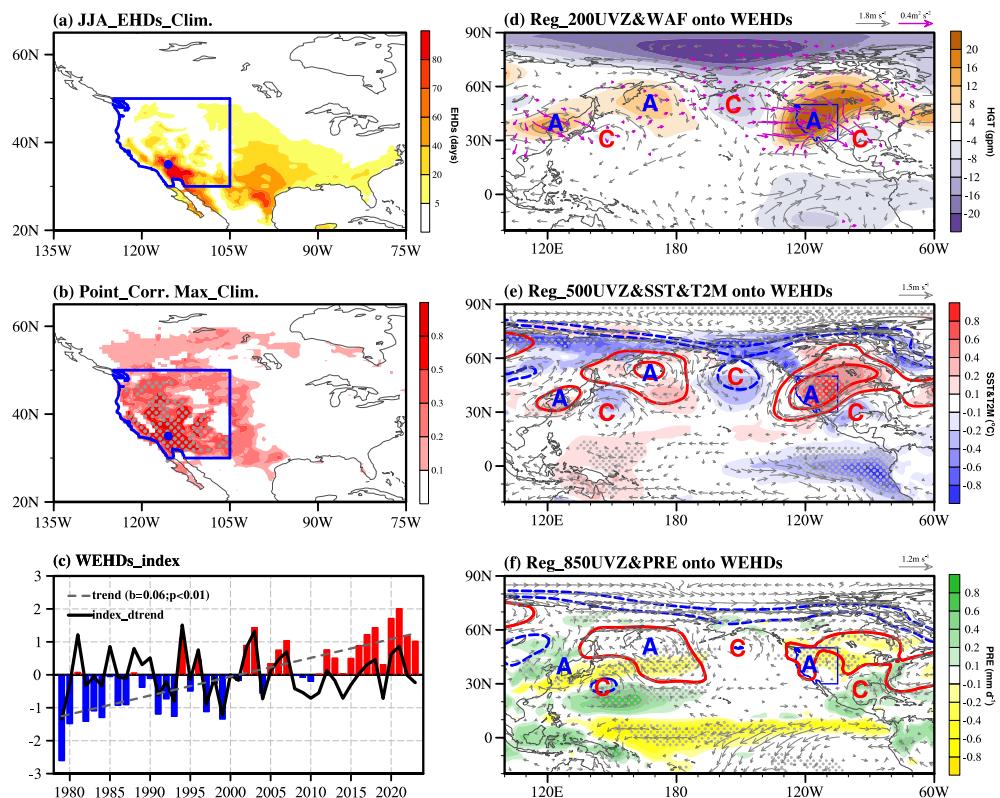
In recent years, deep learning approaches present a promising alternative, with convolutional neural network (CNN) demonstrating particular success in capturing spatial-temporal patterns for climate prediction (Ham et al., 2019; Hwang et al., 2019). Unlike PEMs, these deep learning models autonomously extract nonlinear predictive signals from raw input fields without requiring explicit physical assumptions. A notable example is the application of deep learning models in predicting the Indian Ocean Dipole (IOD) index, where the model autonomously focuses on distinct regional information during different IOD phases, consequently achieving good prediction skills (Ling et al., 2022). However, the relatively limited observational data restricts the generalization ability of deep learning in the ever-changing climate system (Oh & Ham, 2024; Shen et al., 2024), and their “black box” nature obscures physical interpretability—a critical barrier for improvements in prediction skills and scientific insight (M. Chen et al., 2024).

Consequently, the integration of physical understanding with deep learning approaches to genuinely improve climate prediction emerges as a critical frontier in scientific research. Therefore, this study addresses the following pivotal questions: What are the physical processes affecting summer EHDs in WNA? What prediction skills can be achieved by a PEM and a classical CNN architecture? How to improve the prediction skills of CNN based on physical cognition?

## 2. Data and Methods

### 2.1. Data

The data applied in this study includes: The daily maximum temperature data from the Climate Prediction Center (CPC) global temperature data set with a horizontal resolution of  $0.5^\circ \times 0.5^\circ$ ; Monthly SST data from the National Oceanic and Atmospheric Administration’s (NOAA) Extended Reconstructed SST data set version 5 (ERSST.v5) with a horizontal resolution of  $2^\circ \times 2^\circ$  (Huang et al., 2017a); Monthly precipitation data from NOAA’s Precipitation Reconstruction (PREC) with a horizontal resolution of  $2.5^\circ \times 2.5^\circ$  (M. Chen et al., 2002); Monthly sea-ice concentration data from the Hadley Center Sea Ice and SST data set (HadISST) with a horizontal resolution of  $1^\circ \times 1^\circ$  (Rayner et al., 2003); Monthly soil moisture data from ERA5 reanalysis data sets with a horizontal resolution of  $1^\circ \times 1^\circ$  (Hersbach et al., 2020); monthly atmospheric variables, including zonal and meridional winds, geopotential height, 2-m air temperature (T2M), and sea level pressure (SLP) from the National Centers for Environmental Prediction/National Center for Atmospheric Research Reanalysis-1 (NCEP/NCAR) with a horizontal resolution of  $2.5^\circ \times 2.5^\circ$  (Kalnay et al., 1996); The historical simulations from Coupled Model Intercomparison Project Phase 6 (CMIP6) (Eyring et al., 2016) (see Table S1 in Supporting Information S1 for detail information). The studying period is 1979–2023.



**Figure 1.** The definition, time series of the extreme hot days over western North America (WEHDs) index and the simultaneous associated anomalous fields. (a) The distribution of climatological mean summer EHDs (shading, units: days month<sup>-1</sup>) over North America. (b) The correlation coefficient between the year-to-years variations of EHDs in the maximum climatological EHDs point (34.5°N, 116°W) and in each grid. (c) The time series of original and detrend WEHDs. The dashed gray line is the trend of the WEHDs. (d) The regressed 200 hPa geopotential height (shading; units: gpm) and wind (black vectors; units: m s<sup>-1</sup>), and the associated wave activity flux (purple vectors; units: m<sup>2</sup> s<sup>-2</sup>) onto the standardized detrend WEHDs. Panels (e) and (f) are same as (d) but for the sea surface temperature and 2-m air temperature (shading; units: °C), 500 hPa geopotential height (contours; units: gpm) and wind (black vectors; units: m s<sup>-1</sup>), and the precipitation (shading; units: mm d<sup>-1</sup>), 850 hPa geopotential height (contours; units: gpm) and wind (black vectors; units: m s<sup>-1</sup>). Values that are significant at 90% confidence level by a two-tailed Student's *t*-test are dotted.

## 2.2. Definition of EHDs

In this study, a EHD is defined as daily maximum temperature exceeding 35°C. Clearly, the considerable climatological EHDs and year-by-year standard deviation of EHDs are primarily observed from June to August (Figure S1 in Supporting Information S1). Therefore, we will focus on the summer season (JJA) in the following study. While maximum centers of summer EHDs appear in both western and southern North America (Figure 1a), the correlation map between the maximum EHDs point and the EHDs in each grid shows that the coherent variation of EHDs are mainly confined to WNA (Figure 1b). Therefore, a local index for EHDs over WNA (WEHDs) is defined as the area-mean EHDs over 30°–50°N and 125°–105°W.

## 2.3. Wave Activity Flux

The phase-independent wave activity flux (WAF) is calculated based on the following formula (Takaya & Nakamura, 2001):

$$W = \frac{1}{2\overline{|U|}} \left[ \frac{\overline{u}(\psi_x'^2 - \psi' \psi'_{xx}) + \overline{v}(\psi_x' \psi_y' - \psi' \psi'_{xy})}{\overline{u}(\psi_x' \psi_y' - \psi' \psi'_{xy}) + \overline{v}(\psi_y'^2 - \psi' \psi'_{yy})} \right],$$

where an overbar and a prime represent the climatological mean and anomaly, respectively;  $\psi$  and  $U = (u, v)$  represent the stream function and the horizontal wind, respectively; and  $W$  denotes the two-dimensional Rossby WAF.

#### 2.4. Physics-Based Empirical Model

The PEM is established based on understanding of the physical linkage between the predictors and the predictand (B. Wang et al., 2013, 2015) using the data sets from 1979 to 2013. It is established using the physically meaningful and independent predictors, and an independent forecast is made for the remaining 10 years from 2014 to 2023. The detailed description is shown in Text S1 in Supporting Information S1.

Four verification metrics are used to evaluate the prediction skills: The temporal correlation coefficient (TCC), the root mean squared error (RMSE), the mean square skill score (MSSS) and the same sign rate (SSR). The detailed calculation of these metrics is shown in Text S2 in Supporting Information S1.

#### 2.5. CNN Model and Explainability

Following the same methodology of Ham et al. (2019), we construct a CNN model to predict the WEHDs index. The SST, T2M, and SLP anomalies from January to May over the region north of the 30°S are used as the input layer. The CNN model has three convolutional layers and three max-pooling layers, and the last pooling layer is linked to a fully-connected layer and the final output. We also apply the transfer learning, the CNN model is first trained using the CMIP6 data, and then the trained weights are used as initial weights to train the final CNN model with the reanalysis (Ham et al., 2019). A dropout layer with a rate of 0.2 is added on the fully-connected layer not only during training, but also during the testing phase (Oh & Ham, 2024; Srivastava et al., 2014). Additionally, we employ weight decay in the Adam optimizer as a L2 regularization technique to further mitigate overfitting (Kingma & Ba, 2014). To ensure the robustness of the model predictions, we trained the CNN model with 12 different seeds, representing different ensemble members. Notably, to ensure an equitable comparison with the PEM, training and independent prediction periods are strictly maintained for the CNN model (detailed data division is shown in Table S2 in Supporting Information S1).

Gradient-weighted Class Activation Mapping (Grad-CAM) is used to visualize the contribution of input signals for output in CNN model (Selvaraju et al., 2020). The key regions that contribute to the prediction can be identified through heat maps. More detailed descriptions of model hyperparameter setting and Grad-CAM calculation are shown in Text S3 in Supporting Information S1.

#### 2.6. Model Experiments

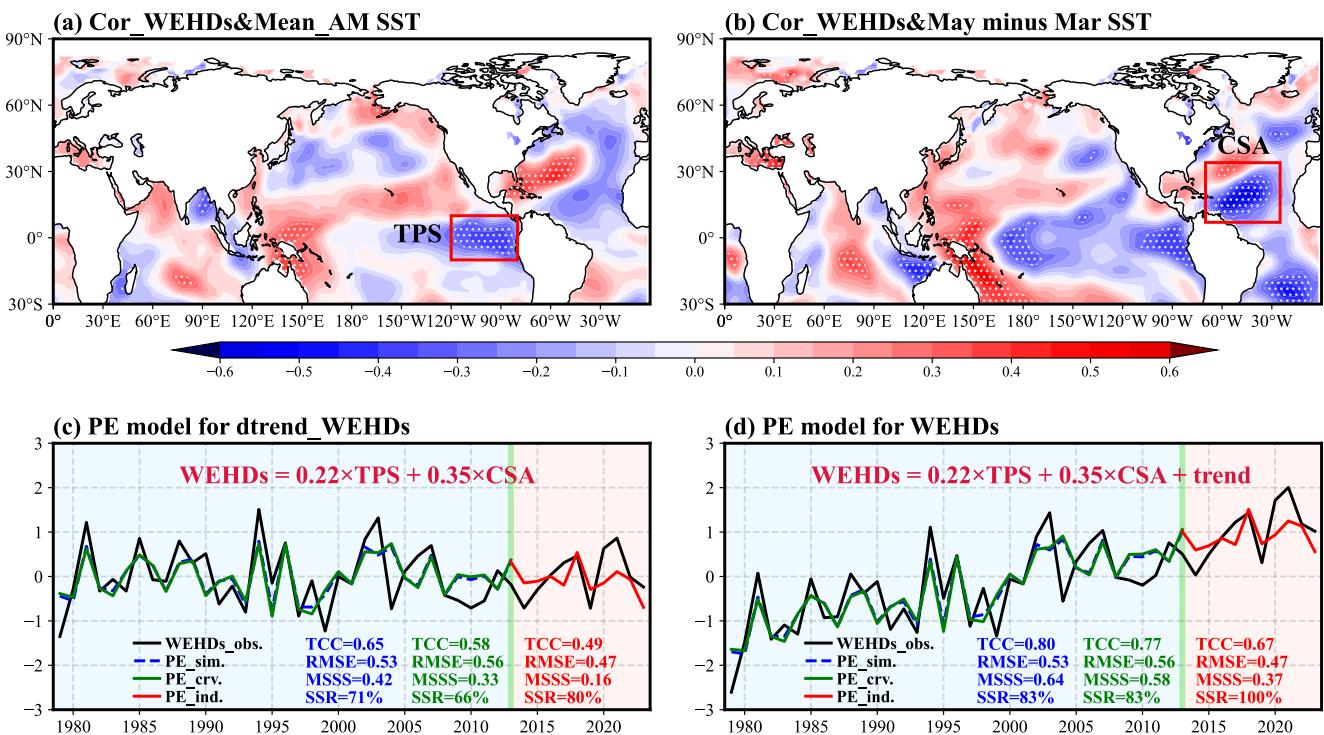
To justify the physical connections between the selected predictors and WEHDs in PEM, a linear baroclinic model (LBM) (Watanabe & Kimoto, 2000) is employed. This model is run at a horizontal resolution of triangular truncation of T21 with 20 sigma levels in the vertical. The 1981–2010 climatological JJA mean of the NCEP/NCAR reanalysis is used as a realistic mean state in the model. LBM has been employed to investigate the atmospheric circulation responses to diabatic heating or vorticity forcing. The integration of the model span 30 days, and the averaged output in the last 10 days represents the equilibrium state of atmospheric responses.

### 3. Results

#### 3.1. Simultaneous Anomalous Fields Associated With WEHDs

The WEHDs index exhibits a significant increasing trend of 0.06 days year<sup>-1</sup> (Figure 1c). The detrended WEHDs index demonstrates pronounced interannual variability. Notably, the WEHDs in 2021 remains an extreme outlier in both the original and detrended indices, indicating that the record-breaking EHDs was influenced not only by the global warming but also by internal climate variability.

To reveal the physical mechanisms driving the variations of WEHDs, the simultaneous circulation and SST fields are regressed onto the detrended WEHDs index. Over the WNA, a quasi-barotropic anomalous high-pressure system, acting as a heat dome, is the dominant circulation pattern. This feature enhances local temperatures and increasing the frequency of EHDs. Tracing the origins of this anomalous high-pressure system, the associated upper-level WAF shows a cross-Pacific Rossby wave train (Figure 1d), consistent with the negative phase of the



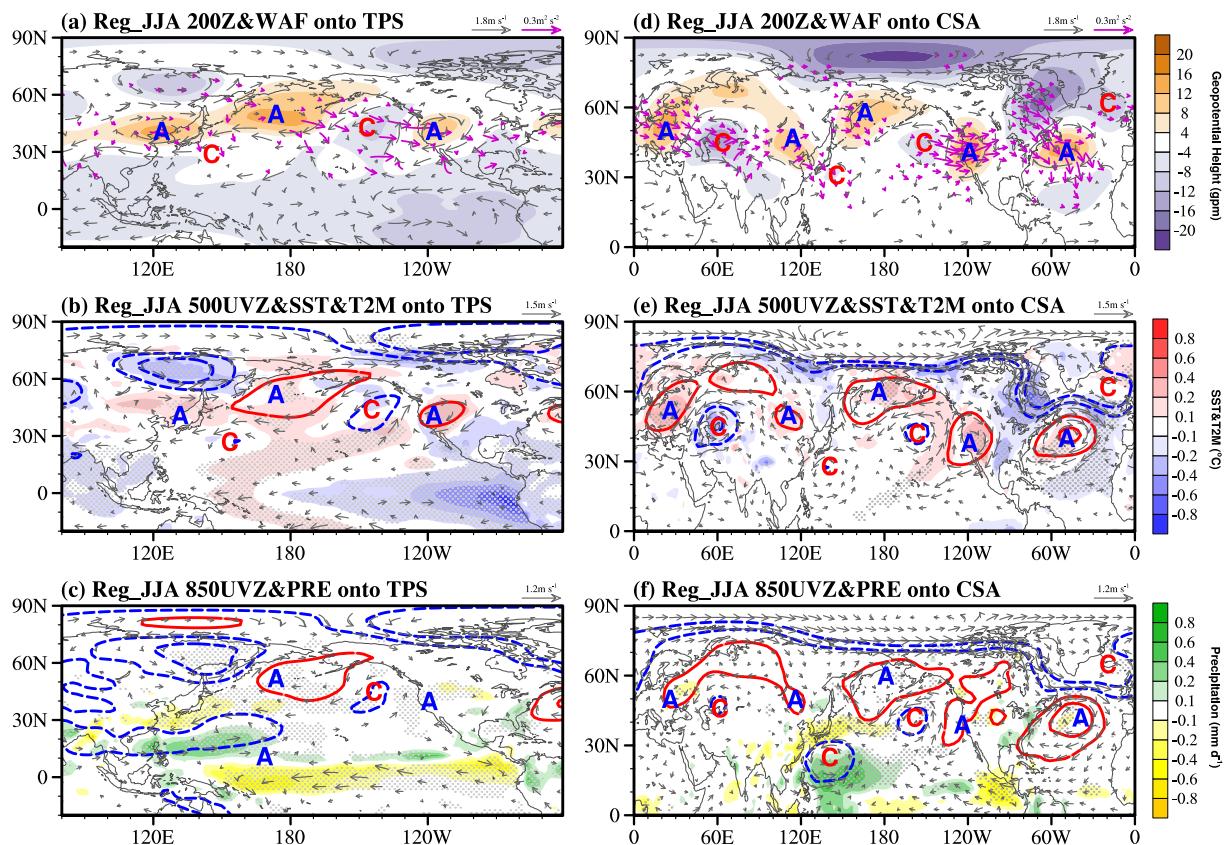
**Figure 2.** The selected predictors and establishment of P-E models. Correlation between detrend extreme hot days over western North America (WEHDs) and (a) April-May mean sea surface temperature (SST), (b) May minus March SST. Values that are significant at 90% confidence level by a two-tailed Student's *t*-test are dotted. The red boxes denote the domains used for the definition of each predictor: TPS ( $10^{\circ}\text{S}$ – $10^{\circ}\text{N}$ ,  $120^{\circ}\text{W}$ – $80^{\circ}\text{W}$ ), CSA ( $7^{\circ}\text{N}$ – $34^{\circ}\text{N}$ ,  $70^{\circ}\text{W}$ – $25^{\circ}\text{W}$ ). (c) Time series of observed (black line), simulated (blue line), cross-validated reforecast (green line), and independently predicted (red line) detrend WEHDs. The temporal correlation coefficient, root mean squared error, mean square skill score and same sign rate during training period (blue for simulated, and green for cross-validated reforecast) and independent prediction period (red) are showed. Panel (d) as in (c), but for original WEHDs.

Asia-North America teleconnection pattern (Zhu & Li, 2016, 2018). Simultaneously, negative SSTA appear over the tropical eastern Pacific (TPS), while positive SSTA emerge over the western Pacific (Figure 1e). The SSTA pattern corresponds to the enhanced precipitation over the western North Pacific and suppressed precipitation over the subtropical East Asia, presenting a pattern opposite to that of the enhanced East Asian Meiyu front (Figure 1f). Then, a critical question arises: What precursor signals could potentially induce these simultaneous circulation anomalies?

### 3.2. Selection of the Predictors and Their Physical Processes

Based on the stepwise regression and the physical cognition validated by the numerical simulations, two precursors are selected as predictors for enhanced WEHDs from 44 potential precursors (Figures S2 and S3 in Supporting Information S1), including the persistent negative SSTA in the TPS during April-May (Figure 2a) and the cooling tendency of SST over the tropical North Atlantic (CSA) from March to May (Figure 2b). While other precursors are excluded either because of their dependency with the selected predictors or their unexplainable linkage with WEHDs, the two selected predictors are independent with each other and represent different physical processes toward the WEHDs (Table S3 in Supporting Information S1).

The persistent negative SSTA in the TPS during April and May for the first predictor TPS persist into the following summer (Figure 3b), which leads to negative diabatic heating anomalies in the central and western equatorial Pacific (Figure 3c). As a Gill-type response to the negative diabatic heating anomalies, an upper level cyclonic anomaly appears to the southeast of Japan (Figure 3a). The easterlies over the northern flank of the cyclonic anomaly may further perturb the East Asian subtropical westerly jet and leading to a downstream quasi-barotropic Rossby wave train (Figure 3a) with an anticyclonic anomaly over the WNA and enhanced WEHDs (Qian et al., 2022; W. Zhang et al., 2020).



**Figure 3.** The physical processes of each predictor impacts the extreme hot days over western North America (WEHDs). Regressions of the JJA mean (a) geopotential height (shading; units: gpm) and wave activity flux (vectors; units:  $m^2 s^{-2}$ ) at 200 hPa, (b) sea surface temperature and 2-m air temperature (shading; units:  $^{\circ}C$ ), geopotential height (contours; units: gpm) and wind (vectors; units:  $m s^{-1}$ ) at 500 hPa, (c) precipitation (shading; units:  $mm d^{-1}$ ), geopotential height (contours; units: gpm) and wind (vectors; units:  $m s^{-1}$ ) at 850 hPa onto the predictor TPS. (d–f) Similar with (a–c) but onto the predictor CSA. Values that are significant at 90% confidence level by a two-tailed Student's *t*-test are dotted.

The aforementioned physical mechanism can be well reproduced by the LBM. Whether the numerical experiment is conducted with the TPS-related negative diabatic heating (Figure S4 in Supporting Information S1), or with the positive vorticity forcing associated with the cyclonic anomaly over the subtropical western Pacific (Figure S5 in Supporting Information S1), the steady atmospheric responses in two simulations both show a downstream cross-Pacific Rossby wave train and an anticyclonic anomaly over WNA, consistent with the observation.

The cooling tendency of SST over the tropical North Atlantic from March to May for the second predictor CSA is related to a quasi-barotropic anticyclonic anomaly over the mid-latitude Atlantic (Figures 3d–3f). The northerly wind anomalies along the eastern flank of the anomalous anticyclone tend to enhance climatological northerly winds in the tropical North Atlantic, thereby facilitating the local SST cooling via the wind-evaporation-SST feedback mechanism (Figures 3e and 3f). This is also confirmed by the heat budget analysis on the mixed layer ocean temperature (Jin et al., 2003), suggesting that the net surface heat flux plays a dominant role in driving the negative tendency of the mixed layer temperature in the tropical North Atlantic (Figure S6 in Supporting Information S1). Therefore, CSA reflects the maintenance and intensification of the mid-latitude Atlantic anticyclonic anomaly coupling with SSTA. In the following summer, the negative SSTA over the tropical North Atlantic induces easterly wind anomalies to its west, resulting in the anticyclonic wind shear over the WNA. In addition, the Atlantic anticyclonic anomaly can act as a Rossby wave source to excite a downstream quasi-barotropic Rossby wave train, leading to a cyclone anomaly over the western Pacific. The easterlies along the northern flank of the anomalous cyclone (Figure 3f) further perturb a downstream cross-Pacific Rossby wave train, with an anticyclone anomaly over the WNA (Figures 3d–3f).

This physical mechanism can be also successfully reproduced by a two-step process in LBM. First, when the LBM is forced by prescribed negative vorticity over the mid-latitude Atlantic, which represents the CSA's related

anticyclonic anomaly (Figures S7a–S7c in Supporting Information S1). The steady atmospheric response shows a quasi-barotropic Rossby wave train from the Atlantic across Eurasian continent to the western Pacific, with an cyclonic wind shear over the western Pacific. However, because the LBM cannot simulate localized air-sea coupling phenomena, an additional LBM experiment is forced by prescribed positive vorticity induced by the cyclonic anomaly over the western Pacific (Figures S7d–S7f in Supporting Information S1). The positive vorticity forcing perturbs the westerly jet and induce a downstream quasi-barotropic Rossby wave train with an anticyclonic anomaly over the WNA, similar with the observation.

Based on these two physically meaningful predictors, the PEM is conducted using regressions in the training period of 1979–2013 with the simulation equation of WEHDs =  $0.22 \times \text{TPS} + 0.35 \times \text{CSA}$ . In the training period, the PEM obtained a TCC of 0.65 ( $p < 0.01$ ), RMSE of 0.53, MSSS of 0.42 and SSR of 71%, while they are 0.49, 0.47, 0.16, and 80% in independent prediction period (Figure 2c). Besides, incorporating the trend component of the WEHDs further improves the prediction skill during both the training and independent prediction periods (Figure 2d).

However, while the PEM effectively captures the primary physical mechanisms influencing WEHDs, it struggles to capture the extreme value of the 2021 EHDs over WNA. Such extreme records are often driven by intricate and nonlinear feedbacks that are not fully encapsulated within the linear regression (LR) framework of the PEM (Bartusek et al., 2022). Therefore, to better account for these nonlinear dynamics and improve the prediction of extreme events, we explore how to improve the prediction framework using CNN.

### 3.3. Prediction of WEHDs With Convolutional Neural Network

We construct a CNN model similar to Ham et al. (2019) to predict the WEHDs index. The SST, T2M, and SLP anomalies from January to May over the region north of the 30°S are used as the input layer. Note that both the input and output data retain their original trends, which enables the CNN to autonomously identify and incorporate global warming signals in its predictions.

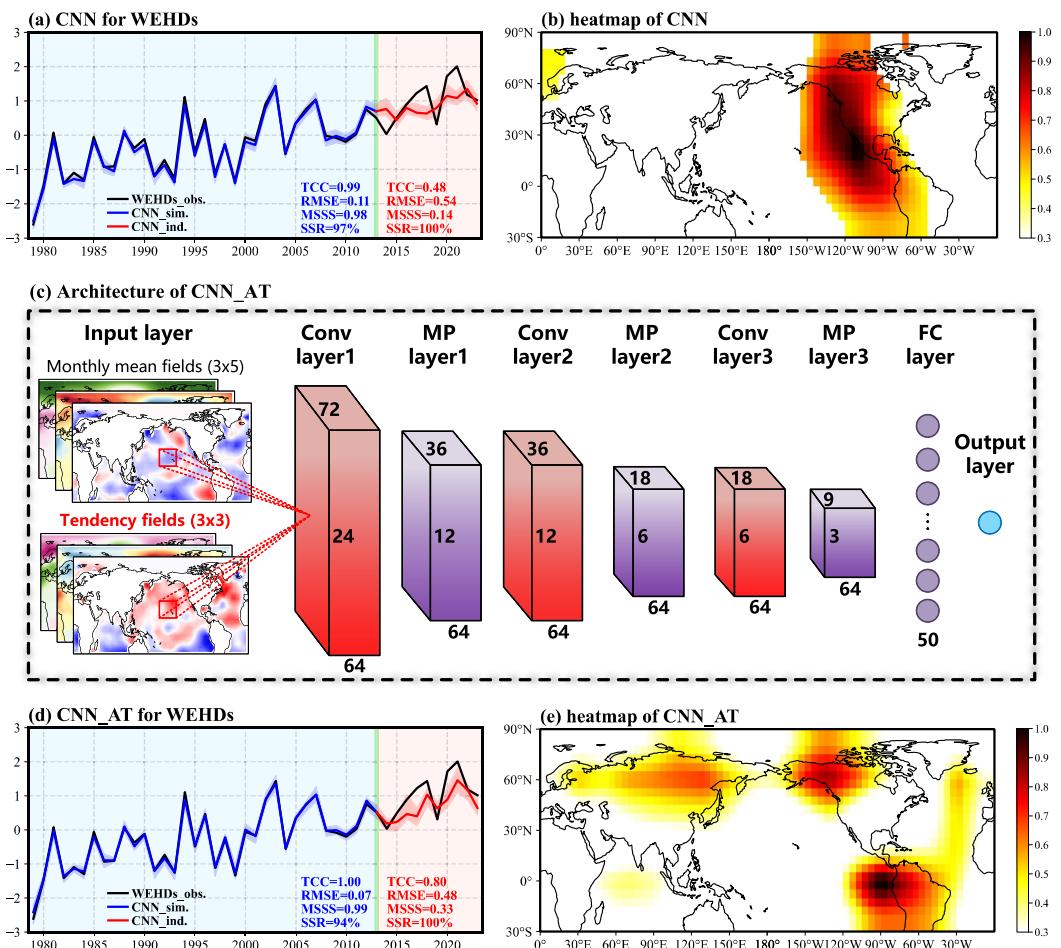
While the CNN model demonstrates nearly perfect prediction skill during the training period (Figure 4a), its independent prediction performance with TCC of 0.48, RMSE of 0.54 and MSSS of 0.14 falls short of the PEM. Meanwhile, the CNN exhibits limited prediction performance for the values in 2014 and 2017–2019, and crucially, fails to capture the extreme record in 2021.

To understand why the prediction skill of CNN model in independent period is limited, we produce the composite heatmap for the training period (Figure 4b). The heatmap quantifies the contributions of the variables at each grid to the predictand, and a larger value denotes a greater contribution to the WEHDs. It shows that large values mainly appear over the eastern Pacific and along the western North American coast, indicating that the CNN model assigns greater importance to the variables in these regions. This may imply that the model may have recognized the physically meaningful predictor TPS. However, the absence of large values over the Atlantic suggests that the CNN model may fails to capture the other physically meaningful predictor CSA.

Constrained by limited model parameters and observational data, the current CNN model struggles to identify critical signals and achieve robust prediction skills. Then, the understanding of the physical processes provides important insights for feature engineering within existing frameworks. Because the CNN missed the ocean surface tendency predictor CSA, we directly integrate SST tendency fields as critical input features into the CNN, guided by the underlying physical mechanisms of WEHDs.

### 3.4. Improve the CNN Prediction From Physical Cognition

Based on the understanding of the physical processes influencing WEHDs, both the monthly mean fields and tendency fields are used as input fields to train the CNN model (Figure 4c). This refined CNN model that adds SST tendency fields (CNN\_AT) shows significant improvement in independent prediction skill (Figure 4d, and Table S4 in Supporting Information S1) in both TCC (from 0.48 to 0.80) and MSSS (from 0.14 to 0.33). Additionally, the CNN\_AT model exhibits encouraging performance for the extreme WEHDs record in both 2021 and 2018. The difference in performance is significant and consistent with different random seeds (Figures S8 and S9 in Supporting Information S1). Among the 12 CNN ensemble members, several members can perfectly predict these extreme events (Figures S9i and 9l in Supporting Information S1).



**Figure 4.** The convolutional neural network (CNN) prediction, explainability and architecture. (a) Time series of observed (black line), CNN simulated (solid blue line) and independently forecasted (solid red line) extreme hot days over western North America (WEHDs). The shaded areas indicate the CNN ensemble spread. The temporal correlation coefficient, root mean squared error, mean square skill score, and same sign rate during training period (blue) and independent prediction period (red) are showed. (b) Grad-CAM map during training period. Only the values statistically significant at 10% level are displayed. (c) Architecture of the CNN model used for predicting WEHDs. The monthly fields are three variables (sea surface temperature (SST), 2-m air temperature, sea level pressure (SLP)) from January to May. The tendency fields include May minus April, May minus March, April-May minus December-January SST/2-m air temperature and SLP. The total input channel of input layer is 24. Panels (d)–(e) as in (a)–(b), but for CNN\_AT model which use both monthly mean and tendency fields as input fields.

The heatmap of the newly developed CNN\_AT shows large values in both the eastern Pacific, Atlantic and the Eurasian Continent (Figure 4e), suggesting that the model may capture both the physically meaningful persistent signal in tropical Pacific (TPS), and tendency signal in Atlantic (CSA) and its downstream Rossby wave trains over Eurasian Continent, thereby contributing to the substantial improvement in prediction skill of the CNN\_AT compared to the original CNN.

One may doubt that the improved performance and the appearance of Atlantic signals in heatmap are purely because of more input data, not the physical process understanding. To clarify this issue, an ablation experiment is conducted with using monthly fields and noise fields as input fields. It shows a decrease of prediction skill and no large values appear over Atlantic and Eurasian Continent in heatmap (Figure S10 in Supporting Information S1). The results confirm the physical information of the SST tendency fields improves the prediction skill of the CNN model.

#### 4. Conclusions

The present study reveals two independent precursor signals influencing WEHDs, each representing different physical processes. The first predictor TPS represents the persistent negative SSTA in the TPS. It leads to the negative diabatic heating anomalies and further stimulates a cyclonic anomaly over the western Pacific. The cyclonic anomaly perturbs the East Asian subtropical westerly jet, leading to a downstream quasi-barotropic Rossby wave train with an anticyclonic anomaly over the WNA. The second predictor CSA reflects the cooling tendency of the SSTA over the tropical North Atlantic, which is related to an anomalous anticyclone over the Atlantic. It acts as a Rossby wave source, exciting a quasi-barotropic Rossby wave train that leads to a high-level anomalous cyclone over the western Pacific, which further perturbs the East Asian jet stream and a downstream cross-Pacific Rossby wave train toward WNA.

A CNN model similar to Ham et al. (2019) is constructed, which utilizes monthly mean fields from January to May as input fields to predict WEHDs. However, it shows limited prediction skill. Then, based on the understanding of the physical processes impacting WEHDs, we further incorporate the tendency fields as additional input fields. The newly developed CNN\_AT model shows significantly improved prediction skills and successfully predicts the 2021 extreme WEHDs. While both CNN and CNN\_AT exhibit nearly perfect prediction skills during training, the CNN dropped its skills during independent period. Conversely, CNN\_AT has significantly better skills in the independent prediction period, which overcomes the overfitting problem as shown in CNN. This comparison highlights that incorporating physically meaningful SST tendency signals effectively alleviates overfitting and enhances independent prediction accuracy. These findings highlight the importance of integrating physical understanding into machine learning-based seasonal prediction frameworks by feature engineering to enrich the physical information of the input data.

To further confirm the importance of the SST tendency predictors not only applicable to CNN, we also use the LR and two widely used tree-based models, Random Forest (RF, Breiman, 2001) and Extreme Gradient Boosting (XGBoost, T. Chen & Guestrin, 2016) to predict WEHDs. For the linear regression model, although the incorporation of tendency signals did not result in a significant improvement in the prediction skills, it still exhibited improved capability in capturing the interannual variability of WEHDs (Figures S11a and S11b in Supporting Information S1), suggesting the importance of the Atlantic Ocean tendency predictor. For the tree-based models, we use Bayesian Optimization (Shahriari et al., 2015) to find optimal hyperparameters for each ML method. It shows that adding the Atlantic Ocean tendency predictor could significantly improve the prediction skills in the independent prediction period. For RF, the TCC skill improves from 0.32 to 0.61 (Figures S11c and S11d in Supporting Information S1). The XGBoost exhibited much greater improvement with the TCC skill increasing from 0.27 to 0.72 (Figures S11e and S11f in Supporting Information S1). In sum, the tendency signals are universally important for other types of machine learning models because they include the key physical processes for the formation of WEHDs.

It is noted that the heatmap of the CNN\_AT also shows large values over Alaska and the Indian Ocean (Figure 4e). This raises critical questions: why does the CNN\_AT focus on these regions? Are they related to the Arctic signals or Eurasian teleconnection? Do these regions reflect genuine physical processes or merely serve as statistical artifacts for model fitting? These questions merit further investigation to elucidate the implication of the heatmap in these identified regions.

In this study, a straightforward three-layer CNN model was utilized for the prediction of EHDs. However, the potential improvement in predictive performance and ability to capture physically meaningful signals through more complex deep learning architectures, such as Residual Neural Networks and Long Short-Term Memory Networks, warrants further investigation.

#### Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

#### Data Availability Statement

The CPC data are from <https://psl.noaa.gov/data/gridded/data.cpc.globaltemp.html>. ERSST v5 data and ERA5 reanalysis are respectively available at Huang et al. (2017a, 2017b), Hersbach et al. (2020, 2023). NOAA's

Precipitation Reconstruction (PREC) data is from M. Chen et al. (2002) at <https://psl.noaa.gov/data/gridded/data.prec.html>. HadISST data is from Rayner et al. (2003) obtained from <https://www.metoffice.gov.uk/hadobs/hadisst/data/download.html>. NCEP/NCAR Reanalysis is from Kalnay et al. (1996) at <https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html>. The CMIP6 models used in this study are listed in Table S1 in Supporting Information S1 and are available at Eyring et al. (2016) (<https://esgf-node.ipsl.upmc.fr/projects/cmip6-ipsl/>).

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (42175033 & 42375045). We acknowledge the High-Performance Computing Center at Nanjing University of Information Science & Technology for their computing support.

## References

- Bartusek, S., Kornhuber, K., & Ting, M. (2022). 2021 North American heatwave amplified by climate change-driven nonlinear interactions. *Nature Climate Change*, 12(12), 1143–1150. <https://doi.org/10.1038/s41558-022-01520-4>
- Bercos-Hickey, E., O'Brien, T. A., Wehner, M. F., Zhang, L., Patricola, C. M., Huang, H., & Risser, M. D. (2022). Anthropogenic contributions to the 2021 Pacific Northwest heatwave. *Geophysical Research Letters*, 49(23), e2022GL099396. <https://doi.org/10.1029/2022GL099396>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Camps-Valls, G., Fernández-Torres, M. Á., Cohrs, K. H., Höhl, A., Castelletti, A., Pacal, A., et al. (2025). Artificial intelligence for modeling and understanding extreme weather and climate events. *Nature Communications*, 16(1), 1–14. <https://doi.org/10.1038/s41467-025-56573-8>
- Chen, M., Qian, Z., Boers, N., Creutzig, F., Camps-Valls, G., Hubacek, K., et al. (2024). Collaboration between artificial intelligence and Earth science communities for mutual benefit. *Nature Geoscience*, 17(10), 949–952. <https://doi.org/10.1038/s41561-024-01550-x>
- Chen, M., Xie, P., Janowiak, J. E., & Arkin, P. A. (2002). Global land precipitation: A 50-yr monthly analysis based on gauge observations. *Journal of Hydrometeorology*, 3(3), 249–266. [https://doi.org/10.1175/1525-7541\(2002\)003<0249:GLPAYM>2.0.CO;2](https://doi.org/10.1175/1525-7541(2002)003<0249:GLPAYM>2.0.CO;2)
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785–794). <https://doi.org/10.1145/2939672.2939785>
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model inter-comparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5), 1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>
- Grotjahn, R., Black, R., Leung, R., Wehner, M. F., Barlow, M., Bosilovich, M., et al. (2016). North American extreme temperature events and related large scale meteorological patterns: A review of statistical methods, dynamics, modeling, and trends. *Climate Dynamics*, 46(3), 1151–1184. <https://doi.org/10.1007/s00382-015-2638-6>
- 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>
- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2023). ERA5 monthly averaged data on single levels from 1940 to present [Dataset]. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. <https://doi.org/10.24381/cds.f17050d7>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>
- Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., et al. (2017a). Extended reconstructed sea surface temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *Journal of Climate*, 30(20), 8179–8205. <https://doi.org/10.1175/JCLI-D-16-0836.1>
- Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., et al. (2017b). NOAA Extended reconstructed Sea Surface temperature (ERSST), version 5 [Dataset]. *NOAA National Centers for Environmental Information*. <https://doi.org/10.7289/V5T72FNM>
- Hulley, G. C., Dousset, B., & Kahn, B. H. (2020). Rising trends in heatwave metrics across southern California. *Earth's Future*, 8(7), e2020EF001480. <https://doi.org/10.1029/2020EF001480>
- Hwang, J., Orenstein, P., Cohen, J., Pfeiffer, K., & Mackey, L. (2019). Improving subseasonal forecasting in the western US with machine learning. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 2325–2335). <https://doi.org/10.1145/3292500.3330674>
- Jin, F. F., An, S. I., Timmermann, A., & Zhao, J. (2003). Strong El Niño events and nonlinear dynamical heating. *Geophysical Research Letters*, 30(3), 20-1. <https://doi.org/10.1029/2002GL016356>
- Jong, B. T., Ting, M., Seager, R., & Anderson, W. B. (2020). ENSO teleconnections and impacts on US summertime temperature during a multidecadal La Niña life cycle. *Journal of Climate*, 33(14), 6009–6024. <https://doi.org/10.1175/JCLI-D-19-0701.1>
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR reanalysis 40-year project. *Bulletin of the American Meteorological Society*, 77(3), 437–471. [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2)
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. <https://doi.org/10.48550/arXiv.1412.6980>
- Li, J., Zheng, C., Yang, Y., Lu, R., & Zhu, Z. (2023). Predictability of spatial distribution of pre-summer extreme precipitation days over southern China revealed by the physical-based empirical model. *Climate Dynamics*, 61(5), 2299–2316. <https://doi.org/10.1007/s00382-023-06681-2>
- Lin, H., Mo, R., & Vitart, F. (2022). The 2021 western North American heatwave and its subseasonal predictions. *Geophysical Research Letters*, 49(6), e2021GL097036. <https://doi.org/10.1029/2021GL097036>
- 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>
- Loikith, P. C., & Broccoli, A. J. (2014). The influence of recurrent modes of climate variability on the occurrence of winter and summer extreme temperatures over North America. *Journal of Climate*, 27(4), 1600–1618. <https://doi.org/10.1175/JCLI-D-13-00068.1>
- Lopez, H., Kim, D., West, R., & Kirtman, B. (2022). Modulation of North American heat waves by the tropical Atlantic warm Pool. *Journal of Geophysical ResearchAtmosphere*, 127(21), e2022JD037705. <https://doi.org/10.1029/2022JD037705>
- Lopez, H., Lee, S. K., Dong, S., Goni, G., Kirtman, B., Atlas, R., & Kumar, A. (2019). East Asian monsoon as a modulator of US Great Plains heat waves. *Journal of Geophysical ResearchAtmosphere*, 124(12), 6342–6358. <https://doi.org/10.1029/2018JD030151>
- Lucarini, V., Melinda Galfi, V., Riboldi, J., & Messori, G. (2023). Typicality of the 2021 western North America summer heatwave. *Environmental Research Letters*, 18(1), 015004. <https://doi.org/10.1088/1748-9326/acab77>
- Luo, M., & Lau, N. C. (2020). Summer heat extremes in northern continents linked to developing ENSO events. *Environmental Research Letters*, 15(7), 074042. <https://doi.org/10.1088/1748-9326/ab7d07>
- Miralles, D. G., Teuling, A. J., Van Heerwaarden, C. C., & Vilà-Guerau de Arellano, J. (2014). Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. *Nature Geoscience*, 7(5), 345–349. <https://doi.org/10.1038/ngeo2141>

- Nabizadeh, E., Lubis, S. W., & Hassanzadeh, P. (2021). The 3D structure of northern hemisphere blocking events: Climatology, role of moisture, and response to climate change. *Journal of Climate*, 34(24), 9837–9860. <https://doi.org/10.1175/JCLI-D-21-0141.1>
- Neal, E., Huang, C. S., & Nakamura, N. (2022). The 2021 Pacific Northwest heat wave and associated blocking: Meteorology and the role of an upstream cyclone as a diabatic source of wave activity. *Geophysical Research Letters*, 49(8), e2021GL097699. <https://doi.org/10.1029/2021GL097699>
- Oh, S. H., & Ham, Y. G. (2024). Few shot learning for Korean winter temperature forecasts. *npj Climate and Atmospheric Science*, 7(1), 279. <https://doi.org/10.1038/s41612-024-00813-z>
- Pfahl, S., & Wernli, H. (2012). Quantifying the relevance of atmospheric blocking for co-located temperature extremes in the Northern Hemisphere on (sub-) daily time scales. *Geophysical Research Letters*, 39(12), L12807. <https://doi.org/10.1029/2012GL052261>
- Qian, Y., Hsu, P. C., Yuan, J., Zhu, Z., Wang, H., & Duan, M. (2022). Effects of subseasonal variation in the East Asian monsoon system on the summertime heat wave in western North America in 2021. *Geophysical Research Letters*, 49(8), e2021GL097659. <https://doi.org/10.1029/2021GL097659>
- Rayner, N. A. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., et al. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, 108(D14), 4407. <https://doi.org/10.1029/2002JD002670>
- Ruprich-Robert, Y., Delworth, T., Msadek, R., Castruccio, F., Yeager, S., & Danabasoglu, G. (2018). Impacts of the Atlantic multidecadal variability on North American summer climate and heat waves. *Journal of Climate*, 31(9), 3679–3700. <https://doi.org/10.1175/JCLI-D-17-0270.1>
- 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>
- Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & De Freitas, N. (2015). Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1), 148–175. <https://doi.org/10.1109/JPROC.2015.2494215>
- Shen, Z., Sun, Q., Lu, X., Ling, F., Li, Y., Wu, J., et al. (2024). Current Progress in subseasonal-to-Decadal prediction based on machine learning. *Applied Computing and Geosciences*, 24, 100201. <https://doi.org/10.1016/j.acags.2024.100201>
- Song, Y., Lu, M., & Wang, B. (2024). The Indo-Pacific rim at risk: How Rossby waves contribute to extreme precipitation clustering. *Geophysical Research Letters*, 51(6), e2023GL107690. <https://doi.org/10.1029/2023GL107690>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1), 1929–1958. <https://doi.org/10.5555/2627435.2670313>
- Takaya, K., & Nakamura, H. (2001). A formulation of a phase-independent wave-activity flux for stationary and migratory quasigeostrophic eddies on a zonally varying basic flow. *Journal of the Atmospheric Sciences*, 58(6), 608–627. [https://doi.org/10.1175/1520-0469\(2001\)058<0608:AFOAPI>2.0.CO;2](https://doi.org/10.1175/1520-0469(2001)058<0608:AFOAPI>2.0.CO;2)
- Thompson, V., Kennedy-Asser, A. T., Vosper, E., Lo, Y. E., Huntingford, C., Andrews, O., et al. (2022). The 2021 western North America heat wave among the most extreme events ever recorded globally. *Science Advances*, 8(18), eabm6860. <https://doi.org/10.1126/sciadv.abm6860>
- Wang, B., Lee, M. Y., Xie, Z., Lu, M., & Pan, M. (2022). A new Asian/North American teleconnection linking clustered extreme precipitation from Indian to Canada. *npj Climate and Atmospheric Science*, 5(1), 90. <https://doi.org/10.1038/s41612-022-00318-7>
- Wang, B., Xiang, B., & Lee, J. Y. (2013). Subtropical high predictability establishes a promising way for monsoon and tropical storm predictions. *Proceedings of the National Academy of Sciences of the United States of America*, 110(8), 2718–2722. <https://doi.org/10.1073/pnas.1214621110>
- Wang, B., Xiang, B., Li, J., Webster, P. J., Rajeevan, M. N., Liu, J., & Ha, K. J. (2015). Rethinking Indian monsoon rainfall prediction in the context of recent global warming. *Nature Communications*, 6(1), 7154. <https://doi.org/10.1038/ncomms8154>
- Wang, H., Gao, Y., Wang, Y., & Sheng, L. (2022). Arctic sea ice modulation of summertime heatwaves over western North America in recent decades. *Environmental Research Letters*, 17(7), 074015. <https://doi.org/10.1088/1748-9326/ac765a>
- Wang, Y., Hu, K., Huang, G., & Tao, W. (2023). The role of nonlinear energy advection in forming asymmetric structure of ENSO teleconnections over the North Pacific and North America. *Geophysical Research Letters*, 50(17), e2023GL105277. <https://doi.org/10.1029/2023GL105277>
- Watanabe, M., & Kimoto, M. (2000). Atmosphere-ocean thermal coupling in the north Atlantic: A positive feedback. *Quarterly Journal of the Royal Meteorological Society*, 126(570), 3343–3369. <https://doi.org/10.1002/qj.49712657017>
- Wei, T., Sun, C., & Li, J. (2023). Seasonal predictability of North American surface temperature arises from SST propagation over the western Pacific. *Journal of Climate*, 36(18), 6273–6286. <https://doi.org/10.1175/JCLI-D-22-0783.1>
- White, R. H., Anderson, S., Booth, J. F., Braich, G., Draeger, C., Fei, C., et al. (2023). The unprecedented Pacific Northwest heatwave of June 2021. *Nature Communications*, 14(1), 727. <https://doi.org/10.1038/s41467-023-36289-3>
- Woollings, T., Barriopedro, D., Methven, J., Son, S. W., Martius, O., Harvey, B., et al. (2018). Blocking and its response to climate change. *Current Climate Change Reports*, 4(3), 287–300. <https://doi.org/10.1007/s40641-018-0108-z>
- Zhang, W., Villarini, G., & Vecchi, G. A. (2020). The East Asian subtropical jet stream and Atlantic tropical cyclones. *Geophysical Research Letters*, 47(15), e2020GL088851. <https://doi.org/10.1029/2020GL088851>
- Zhang, X., Zhou, T., Zhang, W., Ren, L., Jiang, J., Hu, S., et al. (2023). Increased impact of heat domes on 2021-like heat extremes in North America under global warming. *Nature Communications*, 14(1), 1690. <https://doi.org/10.1038/s41467-023-37309-y>
- Zhu, Z., & Li, T. (2016). A new paradigm for continental US summer rainfall variability: Asia–North America teleconnection. *Journal of Climate*, 29(20), 7313–7327. <https://doi.org/10.1175/JCLI-D-16-0137.1>
- Zhu, Z., & Li, T. (2018). Amplified contiguous United States summer rainfall variability induced by East Asian monsoon interdecadal change. *Climate Dynamics*, 50(9), 3523–3536. <https://doi.org/10.1007/s00382-017-3821-8>