Geophysical Research Letters

RESEARCH LETTER
10.1029/2023GL103405

Regional Heatwave Prediction Using Graph Neural Network and Weather Station Data

Peiyuan Li[1], Yin Yu[2], Daning Huang[2], Zhi-Hua Wang[3], and Ashish Sharma[1,4,5]

[1]Discovery Partners Institute, University of Illinois System, Chicago, IL, USA, [2]Department of Aerospace Engineering, Pennsylvania State University, University Park, PA, USA, [3]School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ, USA, [4]Department of Atmospheric Sciences, University of Illinois at Urbana-Champaign, Champaign, IL, USA, [5]Environmental Science Division, Argonne National Laboratory, Lemont, IL, USA

Key Points:

- Trained by the daily weather data from 91 ground stations over the CONUS, the Graph Neural Network (GNN) model can predict regional heatwaves with 90% accuracy
- With the interpretable structure of GNN, we analyze the interplays among meteorological features and the connections between climate zones
- From a deep-learning perspective, the spatiotemporal patterns of climate dynamics are reproduced with a finer granularity

Supporting Information:

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

Correspondence to:

P. Li,
peiyuanl@illinois.edu

Citation:

Li, P., Yu, Y., Huang, D., Wang, Z.-H., & Sharma, A. (2023). Regional heatwave prediction using Graph Neural Network and weather station data. *Geophysical Research Letters*, *50*, e2023GL103405. https://doi.org/10.1029/2023GL103405

Received 21 FEB 2023
Accepted 18 MAR 2023

Author Contributions:

Conceptualization: Peiyuan Li, Zhi-Hua Wang
Formal analysis: Peiyuan Li
Funding acquisition: Ashish Sharma
Investigation: Peiyuan Li
Methodology: Peiyuan Li, Yin Yu, Daning Huang
Software: Yin Yu, Daning Huang
Supervision: Daning Huang, Ashish Sharma
Validation: Peiyuan Li
Visualization: Peiyuan Li, Zhi-Hua Wang
Writing – original draft: Peiyuan Li, Yin Yu, Daning Huang

Abstract Heatwaves lead to catastrophic consequences on public health and the economy. Accurate and timely predictions of regional heatwaves can improve climate preparedness and foster decision-making to alleviate the burdens due to climate change. In this paper, we propose a heatwave prediction algorithm based on a novel deep learning model, that is, Graph Neural Network (GNN). This new GNN framework can provide real time warnings of the sudden occurrence of regional heatwaves with high accuracy at lower costs of computation and data collection. In addition, its interpretable structure unravels the spatiotemporal patterns of regional heatwaves and helps to enrich our understanding of the general climate dynamics and the causal influences between locations. The proposed GNN framework can be applied for the detection and prediction of other extreme or compound climate events, which calls for future studies.

Plain Language Summary Understanding the occurrence and propagation of regional heatwaves is of vital importance to mitigate the consequence of heat extremes. A low-cost, accurate, and timely prediction algorithm for regional heatwaves is desirable. In this study, we use the measurement data set collected from the ground weather stations, together with an advanced deep learning algorithm, to predict the occurrence of regional heatwaves in the lower 48 states of the U.S. The prediction model is trained by the daily weather observations from 91 weather stations and achieves over 90% accuracy in validation. In addition, we summarize the spatiotemporal patterns of the climate dynamics learned by the deep learning model. Our results align with the previous findings while providing a much finer resolution. The proposed modeling framework can be applied to predict other types of extreme events, such as extreme precipitation, drought, and compounded events. The analysis of the model structure will also enhance our understanding of the causal inference between climate regions in the U.S. from a brand-new deep learning perspective.

1. Introduction

Heatwaves (HWs) lead to catastrophic consequences such as the mortality and morbidity of human (Xu et al., 2016), animal (Vitali et al., 2015), and crops (Brás et al., 2021), and severely deteriorate socioeconomic development. Furthermore, global climate change tends to increase the intensity, frequency, and duration of regional HWs (Clarke et al., 2022; Perkins-Kirkpatrick & Lewis, 2020), thus exacerbates the overall vulnerability of the natural environment and human society. In the past decades, studies have been focused on the evaluation of the adverse impacts from HWs and its prediction (Campbell et al., 2018; Ford et al., 2018; Ragone et al., 2018), aiming to derive the corresponding actionable mitigation and adaptation strategies.

Generally, a heat wave can be defined as a prolonged period of excessive hot weather, which usually happens during summer months with a notable and sudden increase of temperature compared to the historical average (Perkins, 2015). It is recognized that the main drivers for HWs are the large-scale atmospheric circulation such as land-atmosphere exchanges and feedbacks on soil moisture and energy (Miralles et al., 2014), the external forcings such as land use change and anthropogenic emission of aerosols (Findell et al., 2017; IPCC, 2021; Zipper et al., 2019), and synchronization of these dynamic contributors (Z.-H. Wang et al., 2021). Physical climate models (e.g., Saha et al., 2014) and the data-driven statistical models (e.g., McKinnon et al., 2016; Zhu & Li, 2018) are prevailing tools for predicting HW events across spatial scales. The climate models solve prognostic equations to quantify energy and matter interact in different parts of the ocean, atmosphere, and land at certain

Writing – review & editing: Peiyuan Li, Yin Yu, Daning Huang, Zhi-Hua Wang, Ashish Sharma

spatiotemporal resolutions (usually from ∼100 km in Global Climate Models to a few kilometers in mesoscale weather models). While statistical models aim to detect potential correlations between temperature and other meteorological variables, often driven by measurement or reanalysis datasets downscaled to urban scales. The final forecasts are usually the ensemble from multiple sub-models.

Recent years have witnessed the emergence of machine learning (ML)-based algorithms to predict atmospheric phenomena including temperature and HWs. Such models use training data from Global Climate Models (Jacques-Dumas et al., 2022), climate reanalysis (Asadollah et al., 2022), and their combination (Jose et al., 2022). ML algorithms based on observations are rather limited primarily due to the spatial discontinuity of weather stations. The Graph Neural Network (GNN), a recent variant of deep learning models, has emerged as a powerful tool for the modeling of unstructured data defined on graphs or networks. In the context of timeseries forecasting, the GNN learns the temporal evolution on the node of interest (NoI) while considering the dynamics between connected nodes (Kipf & Welling, 2017). This graph-theoretic approach has been implemented on multiple climate applications, for example, global weather forecasts (Keisler, 2022; Lam et al., 2022), El Niño (Cachay et al., 2020), air quality (Ejurothu et al., 2022; S. Wang et al., 2020), runoff (Xiang & Demir, 2021), and frost (Lira et al., 2022), to name a few. Yet, for problems of larger scales, the current studies have been limited to data defined on regular Cartesian grids from numerical models rather than observations; for these scenarios, the conventional convolutional neural networks will suffice. One apparent advantage of GNN, however, is to process and exploit irregular graph structure via a generalized convolution operation (Defferrard et al., 2016).

In this study, we propose a novel algorithm based on GNN to predict the occurrence of regional summer heatwaves. The model exclusively utilizes the daily weather summaries of basic meteorological variables from selected ground stations in North America to forecast HW events. Moreover, the analysis on the model structure identifies key meteorological variables that lead to HW and their spatiotemporal patterns across the climate regions in the Contiguous United States (CONUS). The investigation will enhance our understanding of the climate dynamics and temperature extremes and shed new lights on the HW propagation mechanisms from a deep learning perspective. The proposed framework will also reduce the computational burdens on immediate HW warning, and facilitate fast decision-making to mitigate the impacts of extreme heat.

2. Data

2.1. Historical Weather Record

Global Surface Summary of the Day (GSOD) is derived from the Integrated Surface Hourly (ISH) data set and provides daily summaries of near-surface meteorological variables including temperatures, pressures, wind speed, and precipitation. The earliest observation in GSOD can be traced back to 1929, while the data availability and quality are subject to the working conditions of the individual weather stations. In this study, we selected 91 stations over CONUS (Figure 1a) with the consecutive daily recording in the 15 years from 1 January 2006 to 31 December 2020 (5,479 days). This 15-year observation data set captures the long-term norms as well as the seasonal and interannual variability of the climate. Missing data can introduce large uncertainties to data-driven algorithms (Emmanuel et al., 2021; Hariri et al., 2019). In the initial selection, we deliberately avoid the stations with measurement gaps longer than 60 days. This criterion, therefore, limits the number of stations used in this study.

2.2. Identification of Heatwaves

The HW events are identified from the timeseries of the measured maximum temperature (T_{max}) for each station following the criteria from Perkins-Kirkpatrick and Lewis (2020), that is, when T_{max} is above 90th percentile of T_{max} for each calendar day and last at least three consecutive days during warm months (May to September). The percentile threshold is calculated based on a 15-day moving window of daily T_{max} from the temperature record from 2006 to 2020 (15 years) at each location. We then define a new variable, HW flag (β), with dichotomous values. We pre-label the days that HW occurs with $\beta = 1$ and $\beta = 0$ otherwise. Note that there are many other definitions of HW with various criteria on intensity, duration, and frequency. These variations may lead to differences when evaluating their impact on the environment and human health (Xu et al., 2016). But for the development of prediction algorithms, a set of generic criteria will be sufficient and more appropriate when applied to the continental scale. Figure 1b shows the statistics of HW events identified by the criteria over 91 stations during

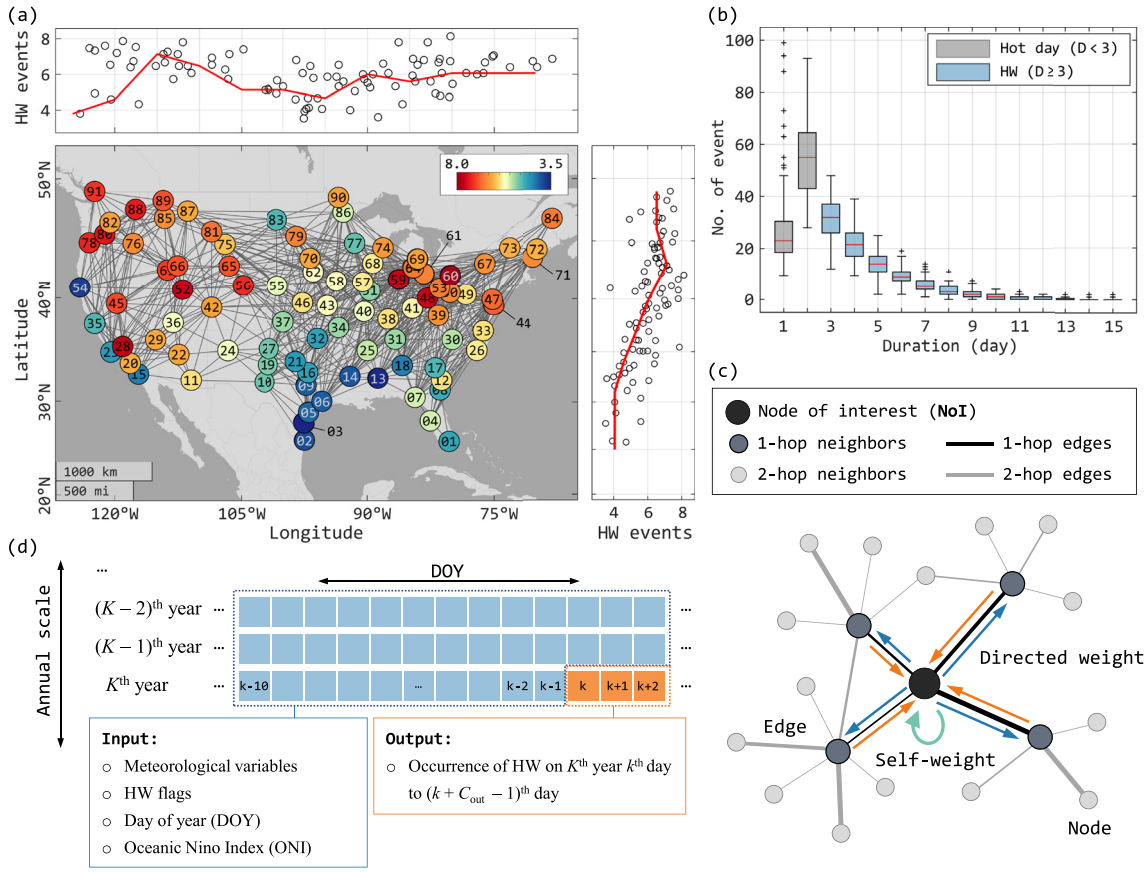

Figure 1. (a) Geographic locations of the selected 91 weather stations; Filled color in (a) shows the annual average number of heatwave (HW) events at each node. The sub-panels above and to the right of (a) show the number of HW events with longitude and latitude, respectively, with red solid lines showing the moving average. (b) Statistics of HW frequency and duration over all stations from 2006 to 2020 (15 years). (c) A diagram illustrating network structure centered on the node of interest; The value and direction of the attention weights on edges in (c) are represented by the thickness of the lines and arrows. (d) A diagram illustrating the model input and output in temporal scale when $C_{in} = 10$, $C_{out} = 3$, $N_{hist} = 2$.

the 15-year study period. On average, each station experiences 28.6 HW days and 6.0 HW events each year. These two numbers are comparable to the HW statistics in the 2010s (24.1 HW days and 6.0 HW events) from the US Environmental Protection Agency (EPA).

2.3. Graph Formulation

A graph or network represents data through a set of nodes, and a set of edges that defines the pairwise relations of the corresponding nodes. Every node that is connected to a NoI by an edge is known as a neighbor of the NoI. These connections are represented by an adjacency matrix that is used in the model. In this study, we formulate the network by treating weather stations as nodes and generate edges based on the correlation coefficients (CC) between two nodes calculated from the daily maximum temperature and precipitation timeseries over 15 years. We set the cut-off thresholds to be the 0.75 quantiles of the CCs calculated from all nodes. This treatment ensures every node has at least one and an average of 10 neighbor nodes. The initial formation identifies the edges (gray lines in Figure 1a) that allow for information exchange between the nodes. Then weights are assigned to the edges to quantify the strength of the connection between two nodes. In many cases, the edge weights are pre-defined and fixed at the formulation of a graph. In our proposed model, we use a set of attention-based edge weights (AWs) that are learned adaptively from the historical weather data during the training of the model (detailed in Section 3.2), aiming to reveal the interactions between climate regions and their impact on the evolution of weather (detailed in Section 4.4).

We select the features (model inputs) that are closely related to the formation and propagation of regional HWs at local and large scales. The features can be classified into three groups. The first group contains the basic

meteorological observations and information from local weather stations, including the HW flag (β), the daily maximum and minimum air temperature ($T_{\max}$ and $T_{\min}$), daily average temperature (T_{avg}), dew point temperature (T_{dew}), air pressure (P_{a}), precipitation (P), and daily average wind speed (U). The second group has variables that relate to large-scale climate variabilities. We use the long-term temperature normals (T_{thres}) at each station, daily sea-level air pressure (P_{sea}), and monthly Oceanic Niño Index (ONI). The last group contains temporal and geographic information, that is, the day of the year (DOY), and the location of the weather station (Lon and Lat). At time instance k and station i, these features define a vector $x_i^{(k)} \in \mathbb{R}^{14}$, where

$$x_i^{(k)} = \left[\beta, T_{\max}, T_{\min}, T_{\text{dew}}, T_{\text{avg}}, U, P, P_{\text{a}}, T_{\text{thres}}, P_{\text{sea}}, \text{ONI}, \text{DOY}, \text{Lon}, \text{Lat}\right] \tag{1}$$

Features on the NoI and its linked neighbors are used for model prediction (Section 3.2). We select the first 13 years of data (2006–2018) for model training and the last 2 years (2019 and 2020) for model validation.

On the temporal scale, we also inform the model with a short history of all features from NoI and its neighbors. The length of input weather history (in days) is defined as C_{in}, while the prediction period is C_{out}. The model also allows the input from previous years with the same DOY, defined as the maximum number of historical years (N_{hist}), to inform the model with the recurrence of HW. Therefore the input data at one node amounts to a sequence of $x_i^{(k)}$ of N_d days, where

$$N_d = C_{\text{in}}(N_{\text{hist}} + 1) + C_{\text{out}} N_{\text{hist}}. \tag{2}$$

Figure 1d shows the model input and output when $C_{\text{in}} = 10$, $C_{\text{out}} = 3$, and $N_{\text{hist}} = 2$. However, due to the contingency of the extreme events, data from the historical years will not affect the model performance in our preliminary tests, hence we use $N_{\text{hist}} = 0$ in the analysis afterward (i.e., $N_d = C_{\text{in}}$). Note that for predictions of the other weather events, information from historical years might be necessary. In the model ablation analysis, we vary the values of C_{in}, and C_{out}, and to study their impacts on model performance (Section 4.2).

3. Methods

3.1. Model Architecture

The GNN model employs an encoder-processor-decoder architecture (Text S1 in Supporting Information S1). Under this architecture, forecasting HW occurrence is a binary classification problem, which can be stated mathematically as: Given a sequence of input features, predict the occurrence of HW at the future C_{out} days,

$$\left[\beta^1, \beta^2, \ldots, \beta^{C_{\text{out}}}\right] = F(X_i, \mathcal{G}; \boldsymbol{\Theta}) \tag{3}$$

where F is a GNN model; $\mathcal{G}$ is graph topology (the network); $\boldsymbol{\Theta}$ is the set of GNN parameters; $\beta \in \mathbb{R}^N$ is the binary HW flag, a positive value of which indicates an occurrence of HW event in the future 1 to C_{out} days.

The model performance will be evaluated by accuracy, recall (or true positive rate, TPR), and precision that are defined from the confusion matrix (Text S2 in Supporting Information S1). Accuracy represents the overall rate of correct predictions regardless of HW days or non-HW days. Recall indicates the success rate of HW prediction within the known HW events. Conversely, precision represents the correct rate in the predicted HW by the model. It is noteworthy that the HW days, as well as the other extreme events, are rare, meaning only a small portion of the data (<8%) are labeled with HW events. The imbalanced HW label will cause a model bias toward the underprediction of HW when the conventional cross-entropy loss (Good, 1952) is used for model training. To avoid this issue, we implement soft F1-score (Text S3 in Supporting Information S1), which combines recall and precision into a single metric to limit model bias while keeping high precision. Specifically, the model parameters ($\boldsymbol{\Theta}$) are trained by maximizing the F1-score between the predicted and true labels.

3.2. Attention-Based Graph Neural Network

In the purposed model, each node represents a weather station. The sequence of node features X_i inform the model of local weather information of previous C days, as well as the global properties (e.g., ONI). Using the message passing (MP) mechanism (Bruna et al., 2013), the output of each NoI (HW flag) is determined by the input features of itself and its neighbors, through the composition of a sequence of GNN modules. One MP

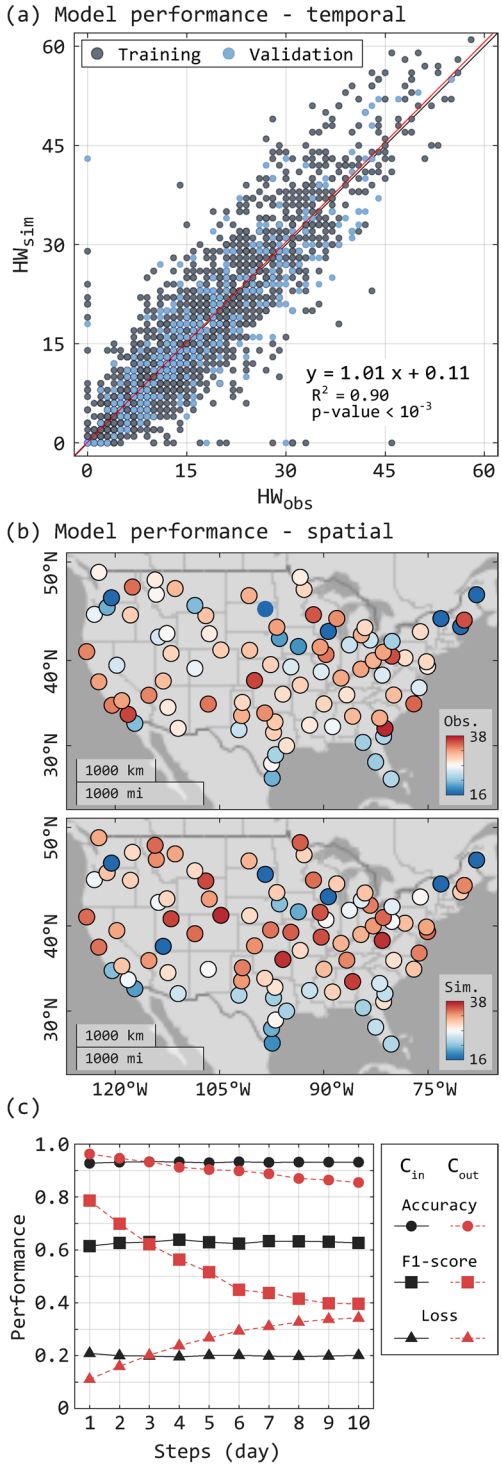

Figure 2. Model performance on (a) temporal scale; (b) spatial scale, and (c) variations when changing C_{in} and C_{out}. Scatters in (a) shows the observed and predicted number of nodes that experience heatwave (HW) at individual timesteps over 15 years. Color map in (b) shows the average number of annual HW days at individual nodes over 15 years. When changing C_{in}, C_{out} is fixed at 3. When changing C_{out}, C_{in} is fixed at 10.

step corresponds to the information exchange between 1-hop neighbors (i.e., the nodes that are directly connected). With multiple layers of GNN modules (more MP steps), the output of an NoI is also influenced by more distant neighbors (Figure 1c). The selection of MP steps is determined by the node density on a graph. With 91 nodes in this study, we limit long-distance information exchange by using only two MP steps.

Specifically, we choose the Graph Attention Layer (GAL) as the GNN module (see Text S4 in Supporting Information S1 for details). The GAL is a nonlinear form of graph convolution that allows the adjacency matrix to dynamically adjust according to the node features using the attention mechanism (Veličković et al., 2017), thus introducing stronger expressiveness when compared to the conventional graph convolution operation. The AW matrix is not symmetric, therefore it is possible that a node is strongly influenced by its neighbor, quantified by a large weight, but not vice versa. Specifically, there are three types of AW: weight with a direction pointing toward NoI (orange arrows in Figure 1c), leaving from NoI (blue arrows), and self-weight (green arrow). They reflect the influence *from* the neighbors, *to* the neighbors, and from the history of NoI itself, respectively. The first two are conventionally calculated as the weighted in-degree (WID) and weighted out-degrees (WOD) of the nodes (Barrat et al., 2004; Opsahl et al., 2010), while the self-weight is exclusively computed from the GAL. Note that the GAL also normalizes AW to ensure the summation of the WID of NoI is equal to 1. Due to these special treatments, the exact values of the weighted degrees computed from GAL will be different from the studies with other approaches, though their spatial patterns should be similar. The values of AW are initialized randomly and then trained together with other network parameters. We repeat the model training five times with the same configuration to ensure the convergence of AWs.

4. Results and Discussions

4.1. Model Performance on HW Prediction

The overall accuracy of the model on the validation data set (2019–2020) is 94.1%, with an average recall of 58.5% and a precision of 62.5%. As mentioned in Section 3.1, a model with a high recall will identify many true HW events but is inclined to false alarms. A model with a high precision, on the other hand, may miss many true HW events, but the ones it catches are very likely to be correct. The precision-recall trade-off can be judiciously tweaked by adjusting the composition of F1-score. The current model uses a balanced configuration (i.e., equal weights of recall and precision), thus capturing the general trend of HW occurrence on the temporal scale without an apparent bias (Figure 2a). In contrast, a high recall rate (>92%) can be achieved with model tuning at the cost of overprediction (Figure S1 in Supporting Information S1). Based on the risk preference in decision-making and the vulnerability to HWs, the model can be adjusted practically. Vulnerable places may prefer a sensitive forecast to increase preparedness; while the places that suffer less from HW will prefer precise prediction and avoid false alarms.

The configuration in this study is designed for immediate warning and decision-making regarding regional HW events. For such large-scale problems, it matters more to predict *when* it occurs rather than *where*. The model can replicate the general spatial patterns of HW occurrence (Figure 2b). Its performance is only tolerable when focusing on specific nodes. Due to the limited number of nodes on the graph, the model has a higher chance to

miss the occurrence of localized HW events. Furthermore, the volume of spatial information, represented by the network density, is not comparable to the length of the timeseries (daily data over 15 years) on the temporal scale. This discrepancy mainly contributes to the performance difference on temporal and spatial scales. It is noteworthy that the model performance will benefit from a denser graph and data with a finer temporal resolution, but it will also require a more complicated model to process the dynamics correspondingly. This trade-off needs to be carefully considered according to the availability of the input data and the intended use of the model.

4.2. Ablation Analysis

We also conduct an ablation analysis on the GNN structure by varying the lengths of model input and output (Figure 2c). The model is not sensitive to the input history when $C_{in} > 2$, and performs reasonably well for shorter C_{in} (2–4), indicating that it only needs the immediate weather history to make reliable predictions. The optimal C_{in} and its relation to performance are largely affected by the other hyperparameters in GNN structure, for example, the number of hidden layers. The current setup limits the information flow within two edges from the NoI, thus can effectively utilize short weather history. To detect HW events at larger spatial scales in conjunction with a denser network, a deeper model with more MP steps and longer C_{in} may be needed.

The model performance degrades quickly with the increase of C_{out} (Figure 2c). The model is accurate with short lead-times, which is sufficient for immediate warning and fast decision-making purposes. It seems difficult for the model to capture the long-term evolution of the system from the daily weather summaries, which usually fluctuate significantly during extreme events. The prediction lead-time can potentially be improved with hourly weather data at a cost of a more sophisticated model architecture, a denser graph network, and extra effort in data collection.

4.3. Features Sensitivity

The GNN model predicts the HW based on the non-linear relations it learns from the historical data. We conduct the sensitivity test over the 14 input features using the nominal model (i.e., model with best overall performance, $C_{in} = 4$, $C_{out} = 3$, $N_{hist} = 0$). The relative sensitivity is quantified by the normalized sensitivity score (SS). Figure 3a shows the temporally averaged SS values. Positive SS indicates positive relation between the chance of HW and the input feature, or vice versa. The SS values reflect the physical dynamics that the model perceived from the training data. For example, air pressure (P_a) has positive SS at most of the stations, which matches the physics-based knowledge that the occurrence of HW is associated with high-pressure system. Similar situations apply to features that indicate hot weather in general, such as T_{avg}, T_{max}, and T_{thres}. Feature sensitivity also has a spatial variation. We find the antecedent HW event β has large SS values, especially in the southwest US and southern Great Plains (Figure 3b), corresponding to a higher chance of prolonged HW in these regions (First Street Foundation, 2022). In contrast, SS is lower in the eastern US, meaning the HW episodes there may not last long or occur consecutively.

Intriguingly, SS of precipitation (P) varies drastically across the US (Figure 3c), though its median SS is close to zero. We find a notable negative-to-positive pattern from the west to the east. It is not surprising to see a strong correlation between precipitation and HW, which are usually reported as sequential hazardous events (Zhang et al., 2021) in the US (Zhang & Villarini, 2020) and other countries (Chen et al., 2021; Sedlmeier et al., 2018). But the SS values from the GNN model show different sequences of occurrence. In the east US, it is likely to have a HW after a rain (Cloutier-Bisbee et al., 2019), but in the west, antecedent rainfall tends to relieve the heat stress and not like to be followed immediately by HWs (Raghavendra & Milrad, 2021). The SS values also vary from month to month as the GNN model captures seasonal dynamics from the training data. Precipitation will retract HW in May but contribute to HW in August (Figure S2 in Supporting Information S1). Nevertheless, the spatial patterns in each month still follow the increasing trend from west to east US.

4.4. Interpretation of Climate Dynamics From GNN

The evolution of regional weather systems is largely affected by the complex interactions between climate regions. In the GNN framework, the strength of these interconnections is quantified by the number of edges and the values of directed attention weights (self-weight, WID, and WOD). It is noteworthy that nodes in the Midwest

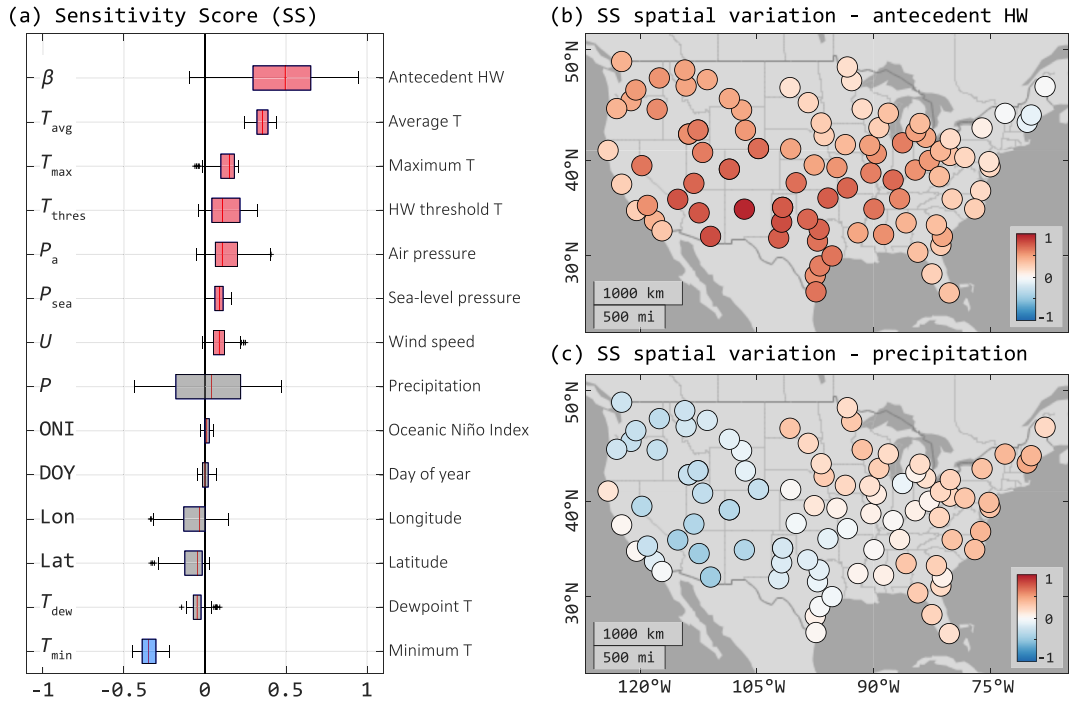

Figure 3. (a) Normalized sensitivity score (SS) of each feature over 91 stations; and spatial variations of relative sensitivity for (b) antecedent heatwave (HW) event (β) and (c) precipitation (P). Features in (a) are ranked by descending order based on the median value of SS. The boxplots show the variation of SS among 91 nodes with the upper and lower whiskers showing 1.5 times of interquartile range (IQR) above 0.75 quantiles and below 0.25 quantile, respectively. The boxes are colored with red, gray, and blue for the median SS above, around, and below zero, respectively.

region have more edges (Figure 4a) as well as larger WID and WOD (Figures 4c and 4d), as compared to the other zones. This indicates that the Midwest is a convergent region of atmospheric transport where the GNN needs more information from the network to predict the HW occurrence. Correspondingly, the self-weights, which reflect the dependence on its own weather history, are low in the Midwest and Southeast US (Figure 4b). In contrast, the self-weights in the west US are relatively high. The phenomenon implies that the HW is likely to originate in the western US and propagates inland. This pathway aligns with the direction of jet streams, which are the main driver of the high-pressure system in the west and the subsequent HW events (Bartusek et al., 2022).

At continental scale, the GNN model can successfully replicate the spatial patterns of the causal influence among the climate regions but with a finer granularity. Figure 4d shows the comparison between the weighted degrees from GNN model to the average casual effect (ACE) and susceptibility (ACS) of thermal environments found in Yang et al. (2022) based on long-term observation. ACE and ACS are obtained using the causal inference algorithm based on the Convergent Cross Mapping (CCM) method (Sugihara et al., 2012). The CCM is method is particular fitful to detect causality in moderately-coupled nonlinear dynamic system, such as Earth's climate System. More specifically, in Yang et al. (2022), ACE and ACS were calculated as the average causality strength among the 9 climate regions in U.S., which can be physically interpreted as measures of the causal strength of a given climate region in modulating (ACE) or being affected by (ACS) other regions. In this study, nodes in Ohio Valley (OV), which are described as "the causal gateway and climate mediator" (Konapala & Mishra, 2017; Yang et al., 2022), exhibit the largest weighted in- and out-degrees (Figures 4c and 4d). In our GNN framework, this phenomenon indicates intense information exchange between these well-connected nodes. On the one hand, if the perturbation from the surrounding (primarily from the west) is not strong enough, its influence will dissipate quickly via the dense connections. Once the impact exceeds a certain tipping point, these connected nodes are likely to be affected as an integrated region. This also explains why OV is both the hub of causal effect and susceptibility from prior studies.

Nodes in the coastal regions (east/west coast and Gulf of Mexico) are generally more isolated from the network and have higher self-weights. Coastal climate is strongly regulated by the sea breezes caused by the differential

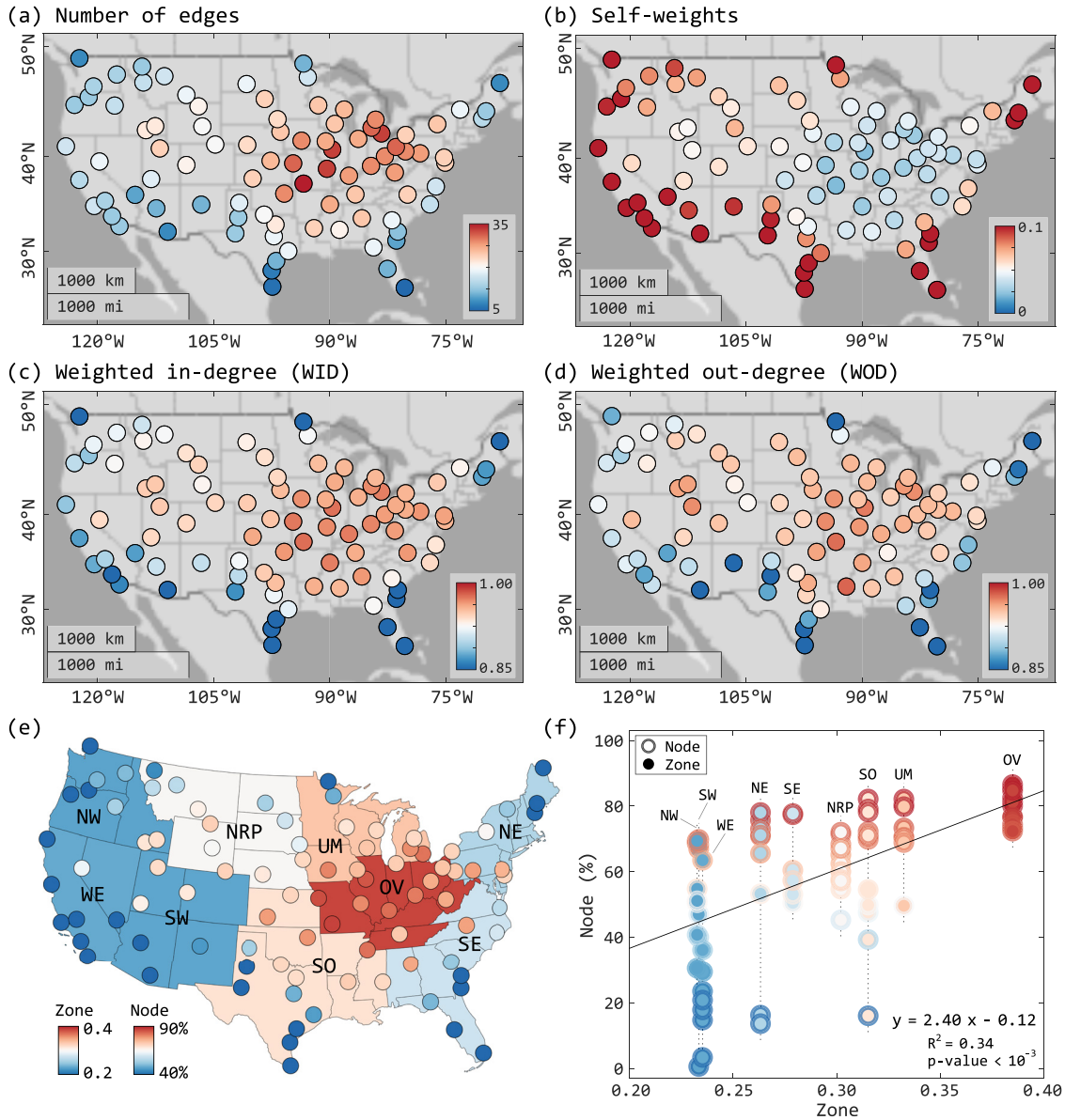

Figure 4. Spatial distribution of (a) number of edges; (b) self-weights; (c) weighted in-degree (WID); and (d) weighted out-degree. (e) Shows a comparison of spatial distribution between WID from Graph Neural Network model to the causal influence (ACS) of thermal environments among climate regions in Yang et al. (2022). (f) Shows the linear regression. Climate regions: NW, Northwest; WE, West; SW, Southwest; NRP, Northern Rockies and Plains; SO, South; UM, Upper Midwest; OV, Ohio Valley; SE, Southeast; NE, Northeast.

heating of the land and the ocean. Therefore, the values of directed weights are distinctive compared to the inland nodes (Figures 4b–4d). Similarly, a few inland nodes (node 11, 22, 86, 90 in Figure 1a) also have high self-weights. These are influenced by specific regional characteristics. For example, summer weather in Southwest US (nodes 11 and 22) is influenced by seasonal monsoon from the Gulf of California; while the weather in the north US (nodes 86 and 90) is more closely connected to the system from the far north (Lee & Sheridan, 2018). Therefore, to capture the dynamics of these nodes, one needs information outside of the current graph. Nonetheless, previous complex network analyses rarely report this limitation. It becomes manifest with increased granularity, network density, and spatial coverage. This shortcoming can be resolved by expanding the graph to the offshore ocean, which may have relatively stable meteorological (i.e., temperature) conditions as a system boundary; this treatment is a prevailing practice in current physics-based weather predictions. It is noteworthy that AWs from GNN change temporally, corresponding to the climate variation between late spring (May) to early fall (September).

Some specific seasonal events, such as North American Monsoon, have certain prevailing directions, which could vary from month to month. The characteristics of these variations from spatial and temporal scales, therefore, are worth further investigation.

5. Concluding Remarks

In this study, we propose a GNN model to predict the occurrence of regional HW based on the daily meteorological data collected from ground weather stations. This framework can be implemented as an immediate warning algorithm for fast decision-making purposes. From its interpretable structure, we further investigate the feature sensitivities. Despite the simplicities of the input data and model structure, the GNN framework is capable of capturing the climate dynamics at the synoptic scale with a very good accuracy. The results of sensitivity analysis align with the evidence from the observations and the physical knowledge reported previously. Moreover, we analyze the directed attention weights and reconstruct the spatiotemporal characteristics of the climate dynamics perceived by GNN. The results highly match the findings from the network analysis and agree with the climate statistics from long-term observations. In addition, the results from GNN have finer granularity.

The application of the GNN framework is promising as a useful analysis and predictive tool for climate studies, but its full potential is yet to be excavated. First, future implementations are expected to incorporate hourly weather data from in-situ observation sites or numerical models. A denser network will also be helpful for the model to capture the temperature evolution at finer spatial scales. In this case, a model with a more sophisticated structure will be needed to tackle the higher data volume on each node and over the denser graph. Nevertheless, it is still critical to seek a balance between the model complexity, input data requirements, and forecast precision under the restrictions of computational cost. The optimum model design is also worth in-depth investigation (Li et al., 2022). In addition, the proposed model structure can be implemented in the detection and prediction of other extreme events such as cold waves, precipitation, drought, and compound events. The GNN framework is not only proven efficient for forecast purposes at a relatively low cost, but also has good potential for disentangling the complex mechanistic processes among the climate and weather extremes from a unique deep-learning perspective.

Conflict of Interest

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

Data Availability Statement

All the processed data, workflow, and Py-Torch code have been deposited into an open data repository, Zenodo, and can be accessed via link https://doi.org/10.5281/zenodo.7504753 with Creative Commons Attribution 4.0 International licensing. Global Surface Summary of the Day (GSOD) data set can be accessed via https://www.ncei.noaa.gov/access/search/data-search/global-summary-of-the-day.

Acknowledgments
This research is supported by the Walder Foundation, NSF awards #139316 and #2230772. This work was also supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research, under contract DE-AC02-06CH11357. The GNN model is implemented using PyTorch Geometric (PyG) (Fey & Lenssen, 2019), an open-source machine learning framework with Graph Network architectures built upon PyTorch (Paszke et al., 2019). We would like to acknowledge the National Center of Environmental Information (NCEI) for providing the data used in this study. We also thank Ms. Xueli Yang for sharing the research data reported in Yang et al. (2022).

References

Asadollah, S. B. H. S., Khan, N., Sharafati, A., Shahid, S., Chung, E.-S., & Wang, X.-J. (2022). Prediction of heat waves using meteorological variables in diverse regions of Iran with advanced machine learning models. *Stochastic Environmental Research and Risk Assessment*, *36*(7), 1959–1974. https://doi.org/10.1007/s00477-021-02103-z

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

Barrat, A., Barthélemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(11), 3747–3752. https://doi.org/10.1073/pnas.0400087101

Brás, T. A., Seixas, J., Carvalhais, N., & Jägermeyr, J. (2021). Severity of drought and heatwave crop losses tripled over the last five decades in Europe. *Environmental Research Letters*, *16*(6), 065012. https://doi.org/10.1088/1748-9326/abf004

Bruna, J., Zaremba, W., Szlam, A., & LeCun, Y. (2013). Spectral networks and locally connected networks on graphs. *arXiv*. https://doi.org/10.48550/ARXIV.1312.6203

Cachay, S. R., Erickson, E., Bucker, A. F. C., Pokropek, E., Potosnak, W., Osei, S., & Lütjens, B. (2020). Graph neural networks for improved El Niño forecasting. *arXiv*. https://doi.org/10.48550/ARXIV.2012.01598

Campbell, S., Remenyi, T. A., White, C. J., & Johnston, F. H. (2018). Heatwave and health impact research: A global review. *Health & Place*, *53*, 210–218. https://doi.org/10.1016/j.healthplace.2018.08.017

Chen, Y., Liao, Z., Shi, Y., Tian, Y., & Zhai, P. (2021). Detectable increases in sequential flood-heatwave events across China during 1961–2018. *Geophysical Research Letters*, *48*(6), e2021GL092549. https://doi.org/10.1029/2021GL092549

Clarke, B., Otto, F., Stuart-Smith, R., & Harrington, L. (2022). Extreme weather impacts of climate change: An attribution perspective. *Environmental Research: Climate*, *1*(1), 012001. https://doi.org/10.1088/2752-5295/ac6e7d

Cloutier-Bisbee, S. R., Raghavendra, A., & Milrad, S. M. (2019). Heat waves in Florida: Climatology, trends, and related precipitation events. *Journal of Applied Meteorology and Climatology*, *58*(3), 447–466. https://doi.org/10.1175/JAMC-D-18-0165.1

Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional Neural Networks on graphs with fast localized spectral filtering. *arXiv*. https://doi.org/10.48550/ARXIV.1606.09375

Ejurothu, P. S. S., Mandal, S., & Thakur, M. (2022). Forecasting PM2.5 concentration in India using a cluster based hybrid Graph Neural Network approach. *Asia-Pacific Journal of Atmospheric Sciences*. https://doi.org/10.1007/s13143-022-00291-4

Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., & Tabona, O. (2021). A survey on missing data in machine learning. *Journal of Big Data*, *8*(1), 140. https://doi.org/10.1186/s40537-021-00516-9

Fey, M., & Lenssen, J. E. (2019). Fast graph representation learning with PyTorch Geometric. In *ICLR workshop on representation learning on graphs and manifolds*.

Findell, K. L., Berg, A., Gentine, P., Krasting, J. P., Lintner, B. R., Malyshev, S., et al. (2017). The impact of anthropogenic land use and land cover change on regional climate extremes. *Nature Communications*, *8*(1), 989. https://doi.org/10.1038/s41467-017-01038-w

First Street Foundation. (2022). First Street Foundation's 6th national risk assessment: Hazardous heat [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.6980285

Ford, T. W., Dirmeyer, P. A., & Benson, D. O. (2018). Evaluation of heat wave forecasts seamlessly across subseasonal timescales. *npj Climate and Atmospheric Science*, *1*(1), 20. https://doi.org/10.1038/s41612-018-0027-7

Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society: Series B*, *14*(1), 107–114. https://doi.org/10.1007/978-1-4612-0919-5_23

Hariri, R. H., Fredericks, E. M., & Bowers, K. M. (2019). Uncertainty in big data analytics: Survey, opportunities, and challenges. *Journal of Big Data*, *6*(1), 44. https://doi.org/10.1186/s40537-019-0206-3

IPCC. (2021). Framing, Context, and Methods. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, et al. (Eds.). *The physical science basis in the sixth assessment report of the intergovernmental panel on climate change*. Cambridge University Press. https://doi.org/10.1017/9781009157896.003

Jacques-Dumas, V., Ragone, F., Borgnat, P., Abry, P., & Bouchet, F. (2022). Deep learning-based extreme heatwave forecast. *Frontiers in Climate*, *4*, 21. https://doi.org/10.3389/fclim.2022.789641

Jose, D. M., Vincent, A. M., & Dwarakish, G. S. (2022). Improving multiple model ensemble predictions of daily precipitation and temperature through machine learning techniques. *Scientific Reports*, *12*(1), 4678. https://doi.org/10.1038/s41598-022-08786-w

Keisler, R. (2022). Forecasting global weather with graph neural networks. *arXiv*. https://doi.org/10.48550/ARXIV.2202.07575

Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *5th international conference on learning representations, ICLR 2017—Conference track proceedings* (pp. 1–14).

Konapala, G., & Mishra, A. (2017). Review of complex networks application in hydroclimatic extremes with an implementation to characterize spatio-temporal drought propagation in continental USA. *Journal of Hydrology*, *555*, 600–620. https://doi.org/10.1016/j.jhydrol.2017.10.033

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Pritzel, A., et al. (2022). GraphCast: Learning skillful medium-range global weather forecasting. *arXiv*. https://doi.org/10.48550/ARXIV.2212.12794

Lee, C. C., & Sheridan, S. C. (2018). Trends in weather type frequencies across North America. *npj Climate and Atmospheric Science*, *1*(1), 41. https://doi.org/10.1038/s41612-018-0051-7

Li, P., Xu, T., Wei, S., & Wang, Z.-H. (2022). Multi-objective optimization of urban environmental system design using machine learning. *Computers, Environment and Urban Systems*, *94*, 101796. https://doi.org/10.1016/j.compenvurbsys.2022.101796

Lira, H., Martí, L., & Sanchez-Pi, N. (2022). A graph neural network with spatio-temporal attention for multi-sources time series data: An application to frost forecast. *Sensors*, *22*(4), 1486. https://doi.org/10.3390/s22041486

McKinnon, K. A., Rhines, A., Tingley, M. P., & Huybers, P. (2016). Long-lead predictions of eastern United States hot days from Pacific sea surface temperatures. *Nature Geoscience*, *9*(5), 389–394. https://doi.org/10.1038/ngeo2687

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

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, *32*(3), 245–251. https://doi.org/10.1016/j.socnet.2010.03.006

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). PyTorch: An imperative style, high-performance deep learning library (Software). In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 32, pp. 8024–8035). Curran Associates, Inc.

Perkins, S. E. (2015). A review on the scientific understanding of heatwaves—Their measurement, driving mechanisms, and changes at the global scale. *Atmospheric Research*, *164–165*, 242–267. https://doi.org/10.1016/j.atmosres.2015.05.014

Perkins-Kirkpatrick, S. E., & Lewis, S. C. (2020). Increasing trends in regional heatwaves. *Nature Communications*, *11*(1), 3357. https://doi.org/10.1038/s41467-020-16970-7

Raghavendra, A., & Milrad, S. M. (2021). On the relationship between heat waves and extreme precipitation in a warming climate. In *Extreme events and climate change* (pp. 183–203). John Wiley Sons, Ltd. https://doi.org/10.1002/9781119413738.ch12

Ragone, F., Wouters, J., & Bouchet, F. (2018). Computation of extreme heat waves in climate models using a large deviation algorithm. *Proceedings of the National Academy of Sciences*, *115*(1), 24–29. https://doi.org/10.1073/pnas.1712645115

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., et al. (2014). The NCEP climate forecast system (version 2) [Dataset]. Journal of Climate, *27*(6), 2185–2208. https://doi.org/10.1175/JCLI-D-12-00823.1

Sedlmeier, K., Feldmann, H., & Schädler, G. (2018). Compound summer temperature and precipitation extremes over central Europe. *Theoretical and Applied Climatology*, *131*(3), 1493–1501. https://doi.org/10.1007/s00704-017-2061-5

Sugihara, G., May, R., Ye, H., hao Hsieh, C., Deyle, E., Fogarty, M., & Munch, S. (2012). Detecting causality in complex ecosystems. *Science*, *338*(6106), 496–500. https://doi.org/10.1126/science.1227079

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2017). Graph attention networks. *arXiv*. https://doi.org/10.48550/ARXIV.1710.10903

Vitali, A., Felici, A., Esposito, S., Bernabucci, U., Bertocchi, L., Maresca, C., et al. (2015). The effect of heat waves on dairy cow mortality. *Journal of Dairy Science*, *98*(7), 4572–4579. https://doi.org/10.3168/jds.2015-9331

Wang, S., Li, Y., Zhang, J., Meng, Q., Meng, L., & Gao, F. (2020). PM2.5-GNN: A domain knowledge enhanced graph neural network for PM2.5 forecasting. In *Proceedings of the 28th international conference on advances in geographic information systems* (pp. 163–166). Association for Computing Machinery. https://doi.org/10.1145/3397536.3422208

Wang, Z.-H., Wang, C., & Yang, X. (2021). Dynamic synchronization of extreme heat in complex climate networks in the contiguous United States. *Urban Climate*, *38*, 100909. https://doi.org/10.1016/j.uclim.2021.100909

Xiang, Z., & Demir, I. (2021). High-resolution rainfall-runoff modeling using graph neural network. *arXiv*. https://doi.org/10.48550/ARXIV.2110.10833

Xu, Z., FitzGerald, G., Guo, Y., Jalaludin, B., & Tong, S. (2016). Impact of heatwave on mortality under different heatwave definitions: A systematic review and meta-analysis. *Environment International*, *89–90*, 193–203. https://doi.org/10.1016/j.envint.2016.02.007

Yang, X., Wang, Z.-H., Wang, C., & Lai, Y.-C. (2022). Detecting the causal influence of thermal environments among climate regions in the United States. *Journal of Environmental Management*, *322*, 116001. https://doi.org/10.1016/j.jenvman.2022.116001

Zhang, W., Luo, M., Gao, S., Chen, W., Hari, V., & Khouakhi, A. (2021). Compound hydrometeorological extremes: Drivers, mechanisms and methods. *Frontiers of Earth Science*, *9*, 673495. https://doi.org/10.3389/feart.2021.673495

Zhang, W., & Villarini, G. (2020). Deadly compound heat stress-flooding hazard across the central United States. *Geophysical Research Letters*, *47*(15), e2020GL089185. https://doi.org/10.1029/2020GL089185

Zhu, Z., & Li, T. (2018). Extended-range forecasting of Chinese summer surface air temperature and heat waves. *Climate Dynamics*, *50*(5), 2007–2021. https://doi.org/10.1007/s00382-017-3733-7

Zipper, S. C., Keune, J., & Kollet, S. J. (2019). Land use change impacts on European heat and drought: Remote land-atmosphere feedbacks mitigated locally by shallow groundwater. *Environmental Research Letters*, *14*(4), 044012. https://doi.org/10.1088/1748-9326/ab0db3