Project Report : CS 7643
Climate Cascade: Predicting Compound Extreme Events Using Graph-Based Deep Learning with Causal Inference

Abhishek Khanal
Georgia Tech
akhanal9@gatech.edu

Astha Rai
Georgia Tech
arai303@gatech.edu

Vasu Ganesh
Georgia Tech
vganesh32@gatech.edu

Benjamin Koob
Georgia Tech
bkoob3@gatech.edu

Abstract

As extreme weather events become more frequent due to climate change, predicting when and where these catastrophic events occur becomes increasingly important for aftereffect preparedness. In addition, these weather events may combine to form compound weather events, which can have even more significant impacts on environments and infrastructure. We therefore propose utilizing causal inference techniques in conjunction with a graph neural network (GNN) to identify and quantify causal climate variables that lead to the formation of compound weather events. Our method shows a strong predictive performance (macro-AUROC of 0.9499 across eight compound hazards) while retaining interpretability via a per-variable causal effect estimation. For that, we leverage DoWhy, a causal inference python library to estimate average causal effects of climate drivers—such as temperature, snow depth, and precipitation—on the likelihood of next-week compound events. The results reveal physically consistent causal pathways (e.g., heat driving hot–dry sequences, precipitation persistence driving multi-storm events), validating that the GNN internalizes meaningful mechanisms rather than spurious correlations. This integration of GNNs with causal inference provides a scalable, interpretable framework for regional compound climate hazard prediction under climate variability.

1. Introduction/Background/Motivation

The increasing frequency of severe extreme weather events, such as heat waves, droughts, floods, and wildfires, poses major challenges to societies around the world. Beyond discrete event hazards, compound extreme events (where multiple stressors occur simultaneously or sequentially) can amplify impacts on infrastructure, agriculture, and public health [29]. The complexity of these events, which span space and time across multiple meteorological variables, underscores the need for advanced prediction and risk assessment tools.

Deep learning and machine learning, in general, offer interesting tools for solving the problem of extreme weather prediction [2]. While convolutional neural networks and recurrent neural networks show promising results in forecasting such events [8], graph neural networks [28, 24] present a unique perspective as they can allow for spatiotemporal representations of climate data, a method which has been documented in previous work utilizing complex network theory [16, 12]. Indeed, there are numerous studies which have displayed the strength of various GNN-based methods for weather prediction [13, 4, 10, 1]. Thus, it stands to reason that they could be useful for prediction of compound weather events as well.

Another useful technique is causal inference [14, 25], which in the context of weather prediction can help garner an understanding of the confounding climate variables: those which affect other independent variables in addition to being a contributing factor in an extreme weather event. Previous works have combined causal inference with GNNs [6] and even applied such models to climate-related problems such as wildfire detection [27].

In order to explore the ways in which climate variables interact to drive compound extreme events, models such as Single-Model Initial-condition Large Ensemble [15] and General Circulation Models [11] are used. GCMs use numerical methods to model different components (such as atmospheric models and ocean models) in tandem to determine climate variability, while SMILE models use numerical methods alongside GCMs. Simulations of these models are run multiple times with different initial conditions to observe how the climate information influences long-term trends in weather. However, GCMs can struggle with short-term events and specific specialized terrains due to how they represent the spatial aspect, which in turn limits the usefulness of the resulting data about climate patterns. SMILEs, which use GCMs, therefore suffer from the same limitations. Furthermore, if initial conditions are not chosen carefully for SMILEs, they can not cover the true range of possibilities for climate patterns. Reexamining the use of causal GNNs in the context of compound events may thus prove effective.

Our objectives are as follows:

- Develop a graph-based deep learning framework to model spatio-temporal dependencies among climate variables associated with compound extremes such as heat-drought or drought-flood sequences.

- Use GNN-based representations and predictions as inputs to a causal inference framework that quantifies the effect of individual climate drivers using back-door adjustment.

If successful, our methods will yield new insight into the strength of GNN-based methods in predicting compound events. This will not only help us better understand how climate variables interact to drive compound events, but also understand the impact and importance of these interactions. This knowledge is crucial for developing climate adaptation strategies and mitigating the socioeconomic impact of compound events, which affect populations all over the world.

1.1. Data

For this project, the dataset was created by combining daily summary data collected from Climate Data Online [17], an NOAA archive of NCDC weather data. These daily summaries are observed at each individual station and include measurements such as minimum & maximum temperature, precipitation, cloudiness, wind speed, and weather type labels. 30 years worth of data (stretching from 1995-01-01 to 2025-01-01) were gathered for 35 well-maintained weather stations, each in a different city around the Great Lakes Basin of the United States.The Great Lakes region was chosen because it presents a uniquely compelling testbed for compound extreme event prediction due to its complex hydro-climatic interactions and diverse seasonal weather patterns. The region experiences strong seasonal contrasts, with abrupt transitions from heatwaves in summer to snowstorms and ice cover in winter. Additionally, the dense distribution of weather stations around the lakes makes the Great Lakes area both data-rich and impact-sensitive—ideal for both modeling and actionable insights.

Initially, the idea was to supplement CDO data with the ERA5 Dataset [7], a reanalysis dataset which provides an extensive tracking of global atmospheric data and can be constrained to specific latitudes and longitudes. However, this posed a problem in that ERA5 uses the GRIB format, which is a bit more complicated to read and parse when compared to just obtaining CSV files. Therefore, for the sake of time, only CDO data was used for this project.

2. Approach

Our approach utilized the PyTorch [20] deep learning framework with some additional functionality from scikit-learn [21, 5], scipy [22], and pandas [19, 23]. The code for the project can be viewed on GitHub.

Specifically, we focused on predicting compound climate events using multi-variable climate data from 35 weather stations around the Great Lakes Basin of the United States over 30 years between 1995 to 2025. In order to accomplish this, we fed climate data to a spatio-temporal GNN with the aim of producing a model which would predict whether or not a compound weather event might occur in the future given current weather measurements. Additionally, we employed causal inference to quantify the impact of individual climate variables on the predicted likelihood of compound extreme events. We believed our methodology would be successful because organizing city-specific climate data in a graph structure was an intuitive approach to showing the relationships between the different locations. This structure could also be well represented by the GNN and would not require an overly complex structure or design, meaning it should be easier to tune and train. While our GNN-based approach is related to much of the work discussed above, our method differs due to the focus on compound weather events as explained earlier.

2.1. Preprocessing

Dimensionality reduction and imputation were performed to preprocess the raw daily summary data. First, feature columns with a large amount of missing data were dropped as they would likely hinder performance of the network. These included features such as multiday precipitation totals, temperature at time of observation, and labels for weather type in nearby vicinities (there were separate features for weather types recorded at the stations proper). Columns related to 1-minute and 2-minute winds were also dropped as it was decided that 5-second wind data would be sufficient information in the context of extreme weather events.Next, each station's dataset was reindexed to the 30-year period between 1995 and 2025 to fill in any missing rows for dates where the station had not made a daily observation. Missing values for numerical features were filled in with simple mean imputation. Finally, PCA was performed to obtain only the features with 95% of explained variance.

Furthermore, some exploratory data analysis was performed on the raw data set after PCA to better understand the data and understand the relationship between the variables with the highest variance. As seen in Figure 3, most of the variables are weakly correlated, as the heat map is uniform in color. There is some variation in the values, but they are all generally low, suggesting that linear relationships between the features are unlikely. Figure 1 is the resulting plot for the PCA with k-means clustering. Although there are a few outliers, the blue and turquoise clusters that are continuous likely account for the majority of the variance. Given what we see of the feature contributions in Figure 2, the blue cluster which is higher for PC2 likely corresponds to stormy weather. The blue cluster eases into the turquoise cluster, which reflects lighter rainfall and milder weather conditions. The brown cluster is very high along PC2, indicating extreme storm conditions. Figure 2 illustrates the contributions of the original features to the two principal components. PC1 is most influenced by mist, drizzle, rain, snow, and dust, indicating more persistent, continuous weather conditions. However, for PC2, the biggest contributors are fog, thunder, sleet, hail, and high winds, which seem to represent storm-like weather conditions.

2.2. Compound Event Labeling

After obtaining and preprocessing CDO data, each daily measurement was labeled with a classification for an extreme weather event based on the values of the numerical features. For example, a day would be labeled

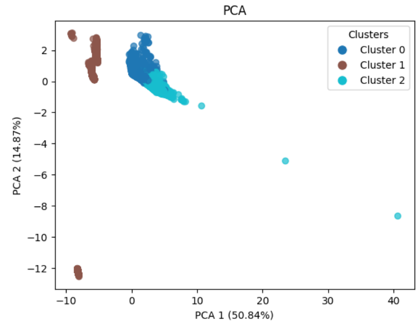

Figure 1. PCA

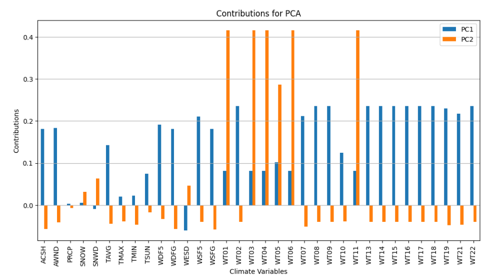

Figure 2. PCA Contributions

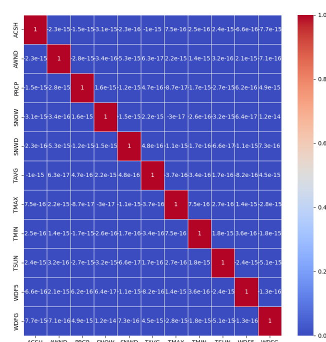

Figure 3. Correlation Matrix

as a *blizzard* if the measurement for snowfall was in the 95th percentile and if the measurement for average wind speed was over 17 m/s.

Following this single-instance labeling, we focused on establishing criteria for compound events by researching specific climate event definitions [3, 18, 29]. From there, we determined which compound events were feasible and which to target for predictions. Furthermore, we opted to aggregate the data over seven-day periods prior to compound event labeling. This ensured that the model's predictions would be performed on a weekly basis, thus allowing us to focus on trends and patterns.

To operationalize this, we defined compound events as the co-occurrence or sequential occurrence of two or more single hazards within the same 7-day window, such as a heatwave followed by extreme rainfall. Each compound label (e.g., `COMPOUND_HEAT_DRY`) was derived using threshold-based logic on rolling weekly statistics, such as temperature exceeding the 90th percentile combined with precipitation below the 10th percentile. Finally, we applied post-label quality control to ensure balance and class frequency, excluding compound types that were too rare for a meaningful prediction or too ambiguous in definition. The resulting label set included eight compound events that captured both temporally concurrent and sequential multivariate extremes.

2.3. Graph Structure and Neural Network Architecture

The GNN's structure reflected the spatiotemporal nature and variable relationships of the data. With inspiration from the work performed in [4], weather stations were represented as nodes with the climate variables embedded within each node. The nodes will contain multiple attributes: information about the station and its climate patterns. The station information consists of the station name, the date, longitude and latitude. Climate information consists of cloudiness, precipitation, snowfall, snow depth, average/maximum/minimum temperature, daily total sunshine time, average daily wind speed, directions of fastest 5-second wind and peak wind gust. Edges were the spatial distance between stations based on their latitudinal and longitudinal coordinates. Initially, we planned to handle the temporal aspect by stacking the graphs, where each section is a spatial graph between stations. However, we struggled to implement this properly. Figure 4 shows our initially proposed architecture for the GNN. After further research, we found it was better to focus on the weeks leading up as it had better predictive power, so we aggregated the data by weeks. The climate variables would then be passed along the edges for inference.

The GNN was composed of a series of SAGEConv [9] layers and ReLU activations with dropout, followed by a single linear layer for binary classification of whether or not a compound weather event had occurred. These output predictions were then fed into a Binary Cross-Entropy (BCE) loss function. This loss function was used because of the binary classification nature of the problem; additionally, given that compound climate events do not tend to occur with a particularly high frequency, BCE loss could help temper the imbalance of such events. Finally, AdamW was chosen for the optimizer for its better regularization of weight decay which helps prevent overfitting and improve generalization. Through optimization, the GNN would learn weights for the graph convolutions as well as the final fully-connected layer.

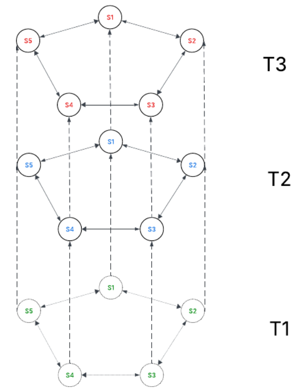

Figure 4. Sample Graph Structure

The GNN's hyperparameters included the number of SAGEConv layers, the hidden dimension size for the final linear layer, dropout rate, learning rate, batch size, weight decay and the number of epochs. Specific tuning for these hyperparameters will be discussed below with the experimentation procedures.

3. Experiments and Results

Our objective was to learn a transductive node-level predictor to predict next-week compound severe weather events across the US Great Lakes region. Success is measured by macro-AUROC across the eight compound labels and binary cross entropy loss for optimization and evaluating convergence via train-validation loss gap.

3.1. Data Split & Protocol

For training the model, 2 weather stations were dropped due to $> 10\%$ nulls, 26 were used as training data, and the remaining 7 were held out-of-network and used for validation. For each Monday-start week, we build a PyG graph with 8-nearest neighbors. The graph is mapped below. All validation metrics are computed only on the validation nodes of each graph; training nodes from those weeks are ignored at evaluation time.

3.2. Hyper-Parameter Search

We proceeded through three search phases, starting with the default GraphSAGE configuration shipped in PyG, no dropout, and Adam with a fixed 8e-4 LR. Val-

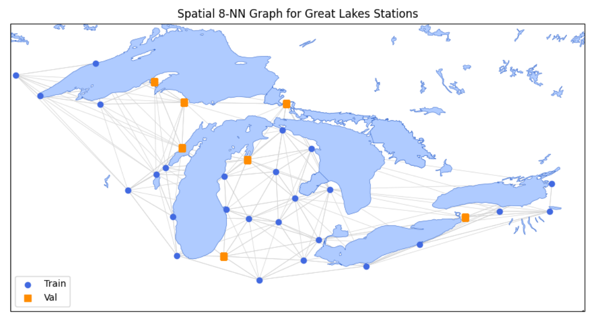

Figure 5. Mapping of actual graph structure

Hyperparameter	Value
Number of SAGEConv layers	2
Linear layer hidden dimension size	96
Dropout	0.05
Learning rate	0.001
Batch size	128
Weight decay	1e-4
Number of epochs	500

Table 1. Final GNN hyperparameters.

idation AUROC was only 0.79. For each phase we ran 20–30 Optuna trials, using 30-epoch budgets; the objective was lowest validation BCE.

Phase 1: architecture sweep We varied the number of SAGEConv layers $L \in \{1, 2, 3\}$ and hidden widths $h \in \{32, 64, 96, 128\}$. Two layers gave the best bias–variance trade-off; three layers over-smoothed node embeddings (-0.015 AUROC). The sweet-spot width was $h = 96$, which we kept for the remainder.

Phase 2: regularization & optimizer We scanned dropout $p \in [0, 0.3]$ and weight decay $\lambda \in [10^{-6}, 10^{-3}]$. A mild dropout of 0.05 and $\lambda = 10^{-4}$ maximized AUROC (+0.01 over no dropout). We also replaced vanilla Adam with **Adam W** because it decouples weight-decay from the adaptive moments, improving stability.

Phase 3: learning-rate scheduling We compared (i) fixed LR, (ii) cosine decay, and (iii) cosine warm restarts (WR). Cosine–WR with $T_0 = 50$ epochs and $\eta_{\min} = 10^{-5}$ out-performed the others by 0.006 AUROC while removing the need for fine LR annealing.

The final settings—summarized in Table 1—were then re-trained for up to 500 epochs with an early-stopping patience of 80. They yielded the best checkpoint, reported below.

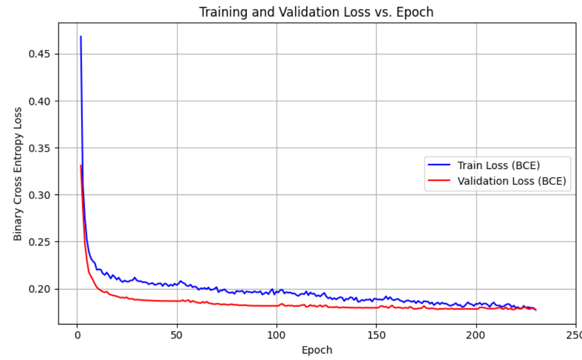

Figure 6. Loss Curve

3.3. Discussion

- A simple 2-layer GCN + inverse-distance weights is sufficient once high-quality extreme-event features are provided.
- Cosine warm-restarts yield a stable plateau around epoch ~180; longer training mainly performs fine-scale weight averaging.
- The extremely tight train–val loss gap confirms that station masking (node-level split) is an effective regularizer compared with the earlier graph-level split.

The results of the best run are very promising. Figure 6 shows the BCE loss curves, with a final validation loss of 0.1775. This is a strong loss and compared to the training loss of 0.1778, shows very strong convergence without overfitting. Figure 7 shows the AUC curve, with 0.9499 AUROC. The AUROC score indicates very high accuracy in predicting compound climate events. Per [26], high-quality deep CNN-based extreme weather event prediction models can achieve up to 89% - 99% performance, placing our working firmly within the state-of-the-art.

3.4. Causal analysis of driver influences

After training the GNN, we freeze all network weights and compute the *node-level* probabilities $\hat{p}_{c,t+1}$ for every compound event c at week t. Following Pearl's back-door criterion for causal inference, we estimate for every *pair* (compound c, driver d) the **Average Causal Effect** (ACE)

$$\mathrm{ACE}_{c,d} = \mathbb{E}\Big[\hat{p}_{c,t+1} \,\Big|\, \mathrm{do}\big(d = d_0 + 1\sigma\big)\Big] - \mathbb{E}\Big[\hat{p}_{c,t+1} \,\Big|\, \mathrm{do}\big(d = d_0\big)\Big]. \quad (1)$$

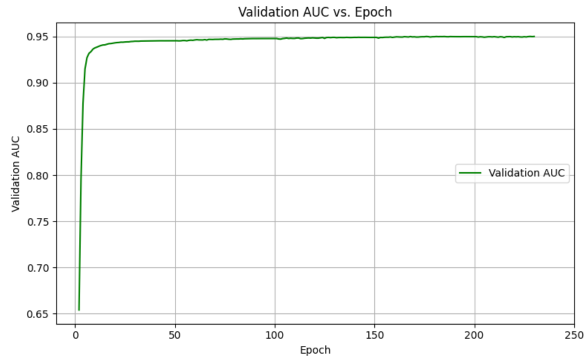

Figure 7. AUC Curve

where d_0 denotes the empirical mean of driver d (all drivers are z-standardized beforehand) and $do(T = t)$ refers to an intervention that sets T to a specific value t, breaking any causal dependencies that normally determine T. The remaining drivers form the confounder set **Z**. We apply DoWhy library's linear back-door estimator with robust HC3 standard errors; 95 % confidence intervals (reported in Table 2) are obtained from 1 000 non-parametric bootstrap resamples of weeks. The results in Table 2 can be interpreted as following:

Table 2. Average causal effect (ACE) of a $+1\sigma$ change in each driver on next–week compound hazards. Values are probability-point changes; $\pm$ gives the 95 % bootstrap CI.

Compound $(t+1)$	**ACE (pp)**			
	ΔTMAX	**ΔSNWD**	**ΔSNOW**	**ΔPRCP**
Heat + dry (HEAT_DRY)	**+3.62** $\pm$0.44	+0.98 $\pm$0.31	+0.10 $\pm$0.09	$-$0.47 $\pm$0.22
Thaw $\rightarrow$ freeze	**+2.19** $\pm$0.37	+0.79 $\pm$0.28	+0.24 $\pm$0.11	$-$0.09 $\pm$0.06
Rain on snow	**+2.10** $\pm$0.35	+0.66 $\pm$0.24	$-$0.12 $\pm$0.10	+0.22 $\pm$0.08
Back-to-back heavy rain	+0.79 $\pm$0.29	$-$0.18 $\pm$0.12	$-$1.00 $\pm$0.19	**+2.56** $\pm$0.41

Hot–dry sequences: A one-sigma warm anomaly increases the probability of a **HEAT_DRY** week ahead by 3.6%, raising the risk from a 6 % baseline to roughly 9.6 %. Snow-related drivers and rainfall exert only minor, partly compensatory, effects ($\leq$ 0.5%).

Thaw–freeze chains Warmer weeks also heighten the odds of a subsequent **thaw–freeze** event by 2.2%, supporting the mechanism of daytime melt followed by nocturnal refreezing. Rainfall is essentially neutral in this pathway.

Rain-on-snow events Both deeper snow cover (+0.7%) and higher air temperature (+2.1%) favour **rain-on-snow**; the direct contribution of extra rainfall is modest (+0.2%), indicating that meltwater dominates over rainfall volume in triggering the hazard.

Successive heavy-rain weeks For **back-to-back heavy rain** the controlling driver is PRCP itself (+2.6%). Temperature and snow variables play a secondary or suppressive role, implying that moisture persistence—rather than prior heat or snowpack—governs the cascade of consecutive wet extremes.

Discussion on causal results: Physical consistency across all four events supports the causal frame: heat drives hot-dry and freeze-thaw, moisture persistence drives storm-storm sequences, and snow depth modulates rain-on-snow. The effects sizes appear small in absolute terms, but imply relative risk changes of 40 to 70%. Therefore, causal analysis confirms that the latent representation of the GNN has captured meaningful driver–hazard links rather than artifacts of spatial correlation.

4. Future Work and Conclusions

Results from our experiments provide the foundations for future research. In this paper, we combined a Graph Neural Network (GNN) with causal inference to not only forecast compound hazards one week ahead, but also to quantify the influence of individual climate drivers such as temperature, snow depth, and precipitation. The resulting model demonstrated strong predictive performance (macro-AUROC = 0.9499) while maintaining interpretability via average causal effect (ACE) estimates derived using the back-door adjustment criterion. By identifying which drivers are most causally linked to different compound events, we provide physically consistent insights that complement the model's raw accuracy. These findings underscore the promise of combining deep learning with causal frameworks to enhance both prediction and understanding of climate extremes.

While we compute ACE values averaged across all locations, a critical next step is to estimate *Conditional Average Treatment Effects (CATEs)* at the station level. This would enable us to discover spatially heterogeneous drivers, such as specific locations being disproportionately vulnerable to heat-induced thaw cycles. We plan to integrate causal forests or doubly robust learners to achieve this. In doing so, we aim to build a fully interpretable and generalizable compound hazard prediction tool for climate risk assessment.

5. Work Division

Table 3 lists the contributions of each team member.

Student Name	Contributed Aspects	Details
Vasu Ganesh	Data Creation and Preprocessing	Gathered data for the project and preprocessed the data for use with neural network frameworks.
Astha Rai	GNN Graph Structure	Set up an extensible GNN graph structure by researching appropriate nodes and edges, to accurately represents key variables and events. Performed EDA on the raw data for insight into relationships in data.
Benjamin Koob	GCN Writing and Visualization	Preprocessed the data by merging stations and reducing features with PCA. Wrote all training scripts and trained the model. Mapped the graph.
Abhishek Khanal	Causal Analysis and Preprocessing	Preprocessed the data and wrote code to create compound label after PCA, took GNN weight and wrote to perform the causal experiments and analysis. Wrote scripts to create visualization related to causal analysis.

Table 3. Contributions of team members.

References

[1] Mihir Agarwal, Progyan Das, and Udit Bhatia. Spatially regularized graph attention autoencoder framework for detecting rainfall extremes, 11 2024. 1

[2] Sina Ardabili, Amir Mosavi, Majid Dehghani, and Annamária R. Várkonyi-Kóczy. Deep learning and machine learning in hydrological processes climate change and earth systems a systematic review. In Annamária R. Várkonyi-Kóczy, editor, *Engineering for Sustainable Future*, pages 52–62, Cham, 2020. Springer International Publishing. 1

[3] Emanuele Bevacqua, Carlo De Michele, Colin Manning, Anaïs Couasnon, Andreia F. S. Ribeiro, Alexandre M. Ramos, Edoardo Vignotto, Ana Bastos, Suzana Blesić, Fabrizio Durante, John Hillier, Sérgio C. Oliveira, Joaquim G. Pinto, Elisa Ragno, Pauline Rivoire, Kate Saunders, Karin van der Wiel, Wenyan Wu, Tianyi Zhang, and Jakob Zscheischler. Guidelines for studying diverse types of compound weather and climate events. *Earth's Future*, 9(11):e2021EF002340, 2021. e2021EF002340 2021EF002340. 3

[4] Harish Chandra Bhandari, Yagya Raj Pandeya, Kanhaiya Jha, and Sudan Jha. Recent advances in electrical engineering: exploring graph neural networks for weather prediction in data-scarce environments. *Environmental Research Communications*, 6(10):105010, oct 2024. 1, 4

[5] Lars Buitinck, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae, Peter Prettenhofer, Alexandre Gramfort, Jaques Grobler, Robert Layton, Jake VanderPlas, Arnaud Joly, Brian Holt, and Gaël Varoquaux. API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pages 108–122, 2013. 2

[6] Xiaojing Du, Feiyu Yang, Wentao Gao, and Xiongren Chen. Causal gnns: A gnn-driven instrumental variable approach for causal inference in networks, 2024. 1

[7] ECMWF. Era5, 2025. 2

[8] Wei Fang, Qiongying Xue, Liang Shen, and Victor S. Sheng. Survey on the application of deep learning in extreme weather prediction. *Atmosphere*, 12(6), 2021. 1

[9] William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs, 2018. 4

[10] Arnold Kazadi, James Doss-Gollin, Antonia Sebastian, and Arlei Silva. Floodgnn-gru: a spatio-temporal graph neural network for flood prediction. *Environmental Data Science*, 3:e21, 2024. 1

[11] Dmitrii Kochkov, Janni Yuval, Ian Langmore, Peter Norgaard, Jamie Smith, Griffin Mooers, Milan Klöwer, James Lottes, Stephan Rasp, Peter Düben, Sam Hatfield, Peter Battaglia, Alvaro Sanchez-Gonzalez, Matthew Willson, Michael P. Brenner, and Stephan Hoyer. Neural general circulation models for weather and climate. *Nature*, 632(8027):1060–1066, Aug 2024. 2

[12] Goutam Konapala, Somnath Mondal, and Ashok Mishra. Quantifying spatial drought propagation potential in north america using complex network theory. *Water Resources Research*, 58(3):e2021WR030914, 2022. e2021WR030914 2021WR030914. 1

[13] Peiyuan Li, Yin Yu, Daning Huang, Zhi-Hua Wang, and Ashish Sharma. Regional heatwave prediction using graph neural network and weather station data. *Geophysical Research Letters*, 50(7):e2023GL103405, 2023. e2023GL103405 2023GL103405. 1

[14] Christos Louizos, Uri Shalit, Joris Mooij, David Sontag, Richard Zemel, and Max Welling. Causal effect inference with deep latent-variable models, 2017. 1

[15] N. Maher, S. Milinski, and R. Ludwig. Large ensemble climate model simulations: introduction, overview, and future prospects for utilising multiple types of large ensemble. *Earth System Dynamics*, 12(2):401–418, 2021. 2

[16] Somnath Mondal and Ashok Kumar Mishra. Complex networks reveal heatwave patterns and propagations over the usa. *Geophysical Research Letters*, 48(2):e2020GL090411, 2021. e2020GL090411 2020GL090411. 1

[17] NOAA. Climate data online daily summaries, 2025. 2

[18] Intergovernmental Panel on Climate Change (IPCC). *Summary for Policymakers*, page 3–32. Cambridge University Press, 2023. 3

[19] The pandas development team. pandas-dev/pandas: Pandas, Feb. 2020. 2

[20] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Z. Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. *CoRR*, abs/1912.01703, 2019. 2

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. 2

[22] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. 2

[23] Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61, 2010. 2

[24] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S. Yu. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1):4–24, Jan. 2021. 1

[25] Zikun Ye, Zhiqi Zhang, Dennis J. Zhang, Heng Zhang, and Renyu Zhang. Deep learning based causal inference for large-scale combinatorial experiments: Theory and empirical evidence. In *Proceedings of the 24th ACM Conference on Economics and Computation*, EC '23, page 1160, New York, NY, USA, 2023. Association for Computing Machinery. 1

[26] Prabhat Joaquin Correa Amir Khosrowshahi David Lavers Kenneth Kunkel Michael Wehner William Collins Yunjie Liu, Evan Racah. Application of deep convolutional neural networks for detecting extreme weather in climate datasets. 2016. 5

[27] Shan Zhao, Ioannis Prapas, Ilektra Karasante, Zhitong Xiong, Ioannis Papoutsis, Gustau Camps-Valls, and Xiao Xiang Zhu. Causal graph neural networks for wildfire danger prediction, 2024. 2

[28] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications, 2021. 1

[29] Jakob Zscheischler, Seth Westra, Bart J. J. M. van den Hurk, Sonia I. Seneviratne, Philip J. Ward, Andy Pitman, Amir AghaKouchak, David N. Bresch, Michael Leonard, Thomas Wahl, and Xuebin Zhang. Future climate risk from compound events. *Nature Climate Change*, 8(6):469–477, Jun 2018. 1, 3