Geophysical Research Letters

RESEARCH LETTER
10.1029/2021GL093787

Using Machine Learning to Analyze Physical Causes of Climate Change: A Case Study of U.S. Midwest Extreme Precipitation

Frances V. Davenport[1] **and Noah S. Diffenbaugh**[1,2]

[1]Department of Earth System Science, Stanford University, Stanford, CA, USA, [2]Woods Institute for the Environment, Stanford University, Stanford, CA, USA

Key Points:

- We use a neural network to predict extreme precipitation from daily sea level pressure and 500-hPa geopotential height fields
- Increasing Midwest extreme precipitation is linked to increasing moisture transport and precipitation during specific circulation patterns
- Our method is generalizable to studying how climate change affects the physical causes of various types of extreme events in other regions

Supporting Information:

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

Correspondence to:

F. V. Davenport,
fvdav@stanford.edu

Citation:

Davenport, F. V., & Diffenbaugh, N. S. (2021). Using machine learning to analyze physical causes of climate change: A case study of U.S. Midwest extreme precipitation. *Geophysical Research Letters*, *48*, e2021GL093787. https://doi.org/10.1029/2021GL093787

Received 8 APR 2021
Accepted 9 JUL 2021

Abstract While global warming has generally increased the occurrence of extreme precipitation, the physical mechanisms by which climate change alters regional and local precipitation extremes remain uncertain, with debate about the role of changes in the atmospheric circulation. We use a convolutional neural network (CNN) to analyze large-scale circulation patterns associated with U.S. Midwest extreme precipitation. The CNN correctly identifies 91% of observed precipitation extremes based on daily sea level pressure and 500-hPa geopotential height anomalies. There is evidence of increasing frequency of extreme precipitation circulation patterns (EPCPs) over the past two decades, although frequency changes are insignificant over the past four decades. Additionally, we find that moisture transport and precipitation intensity during EPCPs have increased. Our approach, which uses deep learning visualization to understand how the CNN predicts EPCPs, advances machine learning as a tool for providing insight into physical causes of changing extremes, potentially reducing uncertainty in future projections.

Plain Language Summary Extreme precipitation and flooding cause widespread impacts on human society. While global warming has increased the occurrence of these damaging events, there is still uncertainty about how climate change will affect precipitation and flooding, making it difficult to adequately prepare for future hazards. We use machine learning to understand why extreme precipitation is becoming more common in the U.S. Midwest by analyzing the atmospheric circulation patterns during extreme precipitation events. Our results show that there is heavier precipitation when extreme precipitation patterns occur, but the patterns themselves have not changed significantly in frequency over the past four decades. Our method could be used to better understand changes in extreme events in the Midwest and in other regions of the world.

1. Introduction

Severe precipitation and flooding are widespread hazards impacting >70 million people globally each year (CRED, 2018). Climate change has increased the frequency and intensity of extreme precipitation (Diffenbaugh et al., 2017; Min et al., 2011; Papalexiou & Montanari, 2019), which increases the costs associated with these hazards (Davenport et al., 2021). To adapt to future precipitation and flooding extremes, it is critical to understand how these hazards are changing.

Increasing precipitation intensity due to higher atmospheric moisture is an expected response to global warming (Allen & Ingram, 2002; Trenberth, 1999). Climate change could also cause dynamic changes such as altering the location and speed of storm tracks (O'Gorman, 2010; Shaw et al., 2016; Yin, 2005), or changing the occurrence of the atmospheric environments associated with precipitation extremes (Barlow et al., 2019; Diffenbaugh et al., 2013; Prein & Mearns, 2021; Trapp et al., 2007). However, due to the challenges of simulating precipitation processes in general circulation models (GCMs; Kharin et al., 2013; Prein & Pendergrass, 2019), understanding of how global warming affects regional and local precipitation extremes remains incomplete.

One approach is to analyze the large-scale atmospheric conditions during which extremes occur. Such studies have commonly used unsupervised methods, such as k-means clustering and self-organizing maps (Agel et al., 2018; Cassano et al., 2007; Esteban et al., 2005; Fragoso & Tildes Gomes, 2008; Gibson et al., 2017; Horton et al., 2015; Loikith et al., 2017; Merino et al., 2016; Singh et al., 2014, 2016; Wilson et al., 1992;

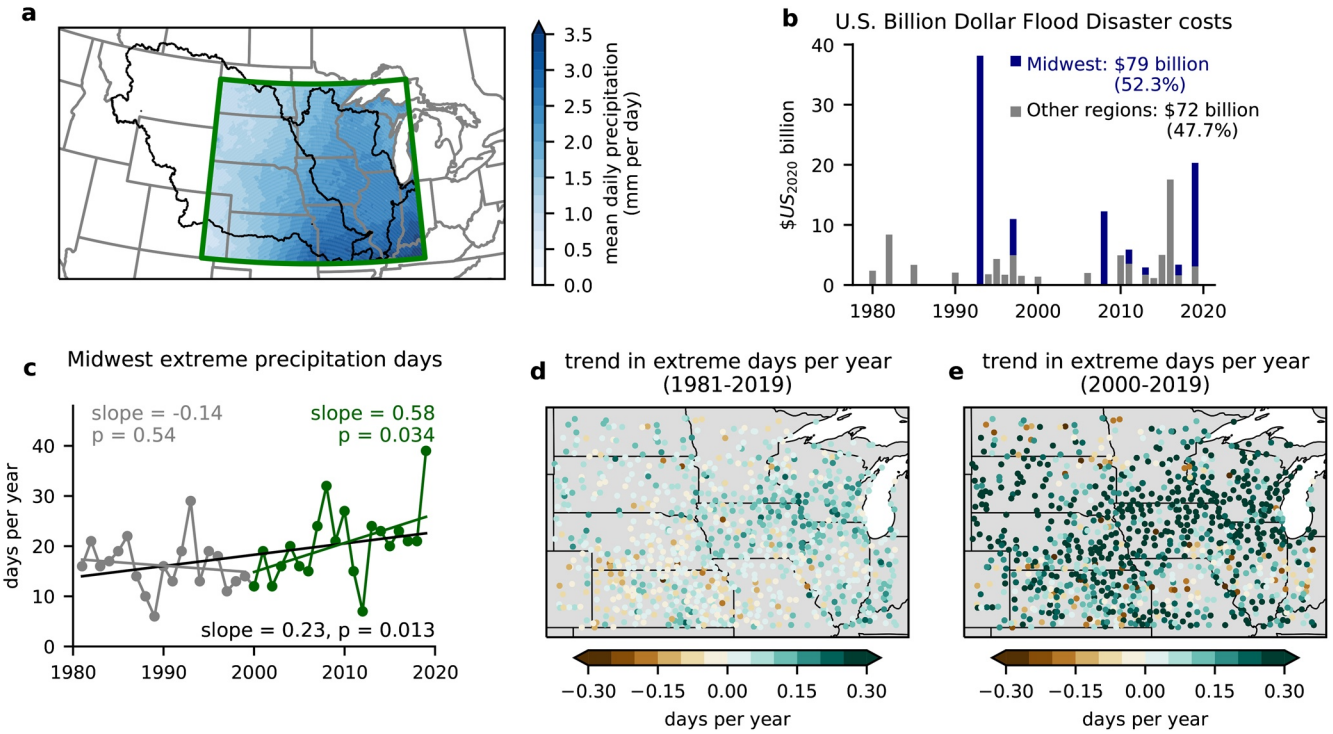

Figure 1. (a) Mean daily precipitation within the study region (green box). (b) Annual damages from U.S. Billion Dollar Disaster floods. Damages from floods affecting the Midwest are shown in blue. (c) Linear trend in the number of Midwest extreme precipitation days per year for the overall (black), early (gray), and late (green) periods. Points show individual years. (d) Linear trend in number of extreme precipitation days per year at Global Historical Climatology Network (GHCN) stations. (e) Same as (d), for the late period (2000–2019).

Zhang & Villarini, 2019). In addition to generating insight into mechanisms by which dynamic and thermodynamic changes may affect extreme events, this approach provides an opportunity to link observed extremes with changes in the large-scale circulation simulated by GCMs (Barlow et al., 2019; Diffenbaugh et al., 2017; Grotjahn et al., 2016).

Here, we demonstrate the use of a convolutional neural network (CNN) to identify large-scale atmospheric circulation patterns associated with extreme precipitation. Because days with extreme precipitation represent a small subset of overall days, the circulation patterns associated with those events may be missed by unsupervised methods that learn the most common patterns in the data. Therefore, we use a supervised approach in order to focus specifically on circulation patterns related to extreme precipitation.

CNNs are neural networks designed for analyzing gridded data, such as images or spatial data (Goodfellow et al., 2016). CNNs are named for their convolutional layers, which use sliding filters (or "kernels") to learn spatial features in the input image. Deep neural networks, including CNNs, often achieve superior performance compared to traditional machine learning or statistical models due to their ability to learn complex, non-linear patterns (LeCun et al., 2015). Their use has grown rapidly, including in climate science and meteorology (Barnes et al., 2020; Chattopadhyay et al., 2020; Ham et al., 2019; Lagerquist et al., 2019). Additionally, recent examples of deep learning visualization within the geosciences have shown that neural networks can provide insight into physical processes (Ebert-Uphoff & Hilburn, 2020; Gagne et al., 2019; McGovern et al., 2019; Toms et al., 2020).

We focus our analysis on the U.S. Midwest. Floods affecting the Midwest account for more than half of damages from U.S. Billion Dollar Flood disasters (Figure 1b; NOAA, 2021). While multiple physical and socio-economic factors influence flood damage, U.S. damages are highly sensitive to extreme precipitation (Davenport et al., 2021), and extreme precipitation days have become more frequent in the region (Figures 1c–1e). To understand these increases, we train a CNN to learn large-scale "extreme precipitation circulation patterns" (EPCPs). We interpret the EPCPs using layerwise relevance propagation (LRP), a deep

learning visualization tool. We use the trained CNN to identify EPCP and non-EPCP days, and investigate the hypotheses that changes in precipitation are due to (a) changes in the frequency of EPCPs and (b) changes in water vapor transport and precipitation intensity when EPCPs occur.

2. Data and Methods

2.1. Data

We use PRISM 4 km daily precipitation (PRISM Climate Group) to calculate extreme precipitation days over the U.S. Midwest from 1981 to 2019. For each day, we calculate the average precipitation over a rectangular region covering the Upper Mississippi Watershed and the eastern portion of the Missouri Watershed (37°N to 48°N, 104°W to 86°W; green box in Figure 1a). We use this regional daily precipitation timeseries to calculate the 95th percentile (p95) of daily precipitation, and define extreme precipitation days as those exceeding the p95 threshold (e.g., Zhang & Villarini (2019); Sillmann et al. (2013)). Because Midwest extreme precipitation can occur throughout the year (Figure S1), we analyze EPCPs across all seasons.

In addition to the PRISM gridded precipitation, we also evaluate the performance of the CNN and analyze long-term changes in precipitation using station data from the Global Historical Climatology Network (GHCN). We include all stations in the Midwest with data available for the full period of analysis (1981–2019), excluding any stations with >10% missing data (leaving a total of 923 stations). At individual stations, we define extreme precipitation days as those exceeding the 95th percentile of daily precipitation for that station. We also define regional extreme days as those where ≥20% of stations exceeded the 95th percentile (this occurs on approximately 5% of days).

We use daily mean sea level pressure (SLP) and 500-hPa geopotential height (GPH) anomalies calculated from the NCEP/NCAR-R1 reanalysis (Kalnay et al., 1996), which provides global coverage at 2.5° x 2.5° horizontal resolution. For the atmospheric variables, we analyze a larger spatial domain that covers the continental U.S. and surrounding oceans (20°N to 55°N and 140°W to 55°W; Figure 2). To remove uniform thermal dilation caused by recent tropospheric warming (Horton et al., 2015), we first subtract the area-weighted average 500-hPa GPH trend over the atmospheric domain, preserving spatially non-uniform changes in 500-hPa GPH that could impact extreme precipitation (Cattiaux et al., 2013; Swain et al., 2016). We then calculate daily standardized anomalies (z-scores) by subtracting the grid-cell calendar-day mean and dividing by the grid-cell calendar-day standard deviation.

We use zonal wind (u), meridional wind (v), and specific humidity (q) fields from the NCEP/NCAR-R1 reanalysis to evaluate differences in moisture flux on days with EPCP and non-EPCP patterns.

2.2. Identification of Extreme Precipitation Circulation Patterns (EPCPs) Using a Convolutional Neural Network

2.2.1. Neural Network Model

To identify Midwest EPCPs, we train a CNN to predict extreme precipitation days using daily SLP and 500-hPa GPH anomalies (Figure 2a). (We also test a model using only 500-hPa GPH, but the model with both input variables has higher classification performance; Figure S2). The CNN is a sequential classifier, meaning input data for each day are passed through multiple layers to generate an output classification of EPCP (Class 1) or non-EPCP (Class 0). For each day, the input to the CNN is a three-dimensional matrix with dimensions 15 × 35 × 2 (i.e., latitude × longitude × 2 input variables). The output labels for the CNN training are generated from the precipitation data. Extreme precipitation days are assigned to Class 1, and non-extreme precipitation days are assigned to Class 0. (We also train the model using a moving p95 threshold to assign extreme and non-extreme days, which produces similar results; Figure S3).

More details on the model architecture, training, hyperparameter selection, and validation results are provided in the supporting information.

After training, the model weights are frozen. The trained model is then used to classify each day as EPCP (Class 1) or non-EPCP (Class 0) based on the 500-hPa GPH and SLP anomaly patterns (without considering precipitation). In other words, EPCP days include all days with a model-predicted EPCP probability >0.5,

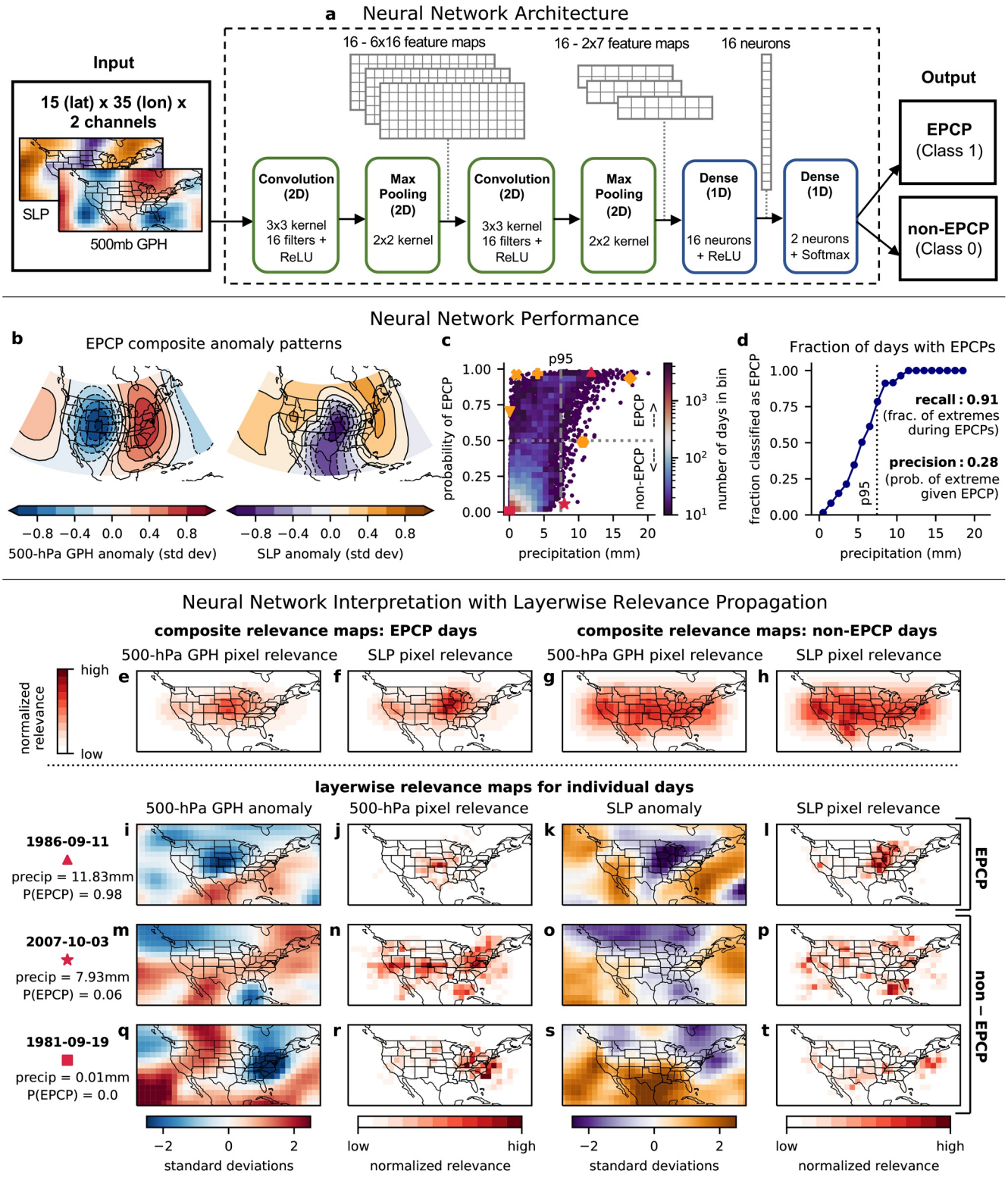
a Neural Network Architecture
Input
15 (lat) x 35 (lon) x 2 channels
SLP
500mb GPH
16 - 6x16 feature maps
16 - 2x7 feature maps
16 neurons
Convolution (2D) 3x3 kernel 16 filters + ReLU
Max Pooling (2D) 2x2 kernel
Convolution (2D) 3x3 kernel 16 filters + ReLU
Max Pooling (2D) 2x2 kernel
Dense (1D) 16 neurons + ReLU
Dense (1D) 2 neurons + Softmax
Output
EPCP (Class 1)
non-EPCP (Class 0)
Neural Network Performance
b EPCP composite anomaly patterns
−0.8 −0.4 0.0 0.4 0.8
500-hPa GPH anomaly (std dev)
−0.8 −0.4 0.0 0.4 0.8
SLP anomaly (std dev)
c
p95
probability of EPCP
precipitation (mm)
EPCP
non-EPCP
number of days in bin
d Fraction of days with EPCPs
fraction classified as EPCP
recall : 0.91
(frac. of extremes during EPCPs)
precision : 0.28
(prob. of extreme given EPCP)
p95
precipitation (mm)
Neural Network Interpretation with Layerwise Relevance Propagation
composite relevance maps: EPCP days
composite relevance maps: non-EPCP days
500-hPa GPH pixel relevance
SLP pixel relevance
500-hPa GPH pixel relevance
SLP pixel relevance
normalized relevance
high
low
e
f
g
h
layerwise relevance maps for individual days
500-hPa GPH anomaly
500-hPa pixel relevance
SLP anomaly
SLP pixel relevance
1986-09-11
precip = 11.83mm
P(EPCP) = 0.98
2007-10-03
precip = 7.93mm
P(EPCP) = 0.06
1981-09-19
precip = 0.01mm
P(EPCP) = 0.0
i
j
k
l
m
n
o
p
q
r
s
t
EPCP
non − EPCP
−2 0 2
standard deviations
low high
normalized relevance
−2 0 2
standard deviations
low high
normalized relevance

Figure 2.

even if precipitation on that day did not exceed the p95 threshold. (Conversely, non-EPCP days include all days with a predicted EPCP probability <0.5). The results shown are calculated across all days in the period, except when noted as training or validation results.

2.2.2. Neural Network Interpretation

We use LRP to interpret the network's classification of individual days. Starting from the final classification layer and working back toward the input layer, LRP traces the most important, or "relevant," information that was passed between layers. Following previous applications in the geosciences (Hilburn et al., 2021; Toms et al., 2020) and recommendations from Montavon et al. (2018), we use the alpha-beta rule with $\alpha = 1$ and $\beta = 0$ to track information passed between layers. With this formulation, LRP tracks information that positively contributes to the predicted class. This rule is known to produce relevance heatmaps which are usually straightforward to interpret. We also tested the alpha-beta rule with $\alpha = 2$ and $\beta = 1$, which tracks both positive and negative information, but the resulting heatmaps were noisy and less interpretable, which is a known challenge of some LRP rules (Montavon et al., 2018).

For a given day, LRP produces a heatmap with the same dimensions as the input, where each pixel's value corresponds to the importance of that input pixel for the final classification. Thus, a primary advantage of LRP is the ability to visualize features important for the prediction in the original spatial domain.

We implement LRP using the iNNvestigate package in Python (Alber et al., 2019).

2.3. Evaluating Changes in Extreme Precipitation Days

After each day is classified, we first count the number of EPCP days and average precipitation intensity across EPCP days for each calendar year. We then calculate linear trends in EPCP frequency and precipitation intensity across the annual timeseries for the early (1981–1999), late (2000–2019), and full (1981–2019) periods. (The sensitivity to the cut-off year between the early and late periods is shown in Table S1). We also compare the distribution of daily precipitation for EPCP and non-EPCP days during the early and late periods (Figures 3f–3i).

We calculate the vertically integrated daily moisture flux (MF_x and MF_y) in the zonal and meridional directions using the equations:

$$MF_x = -\frac{1}{g}\int_{1{,}000\ \mathrm{hPa}}^{300\ \mathrm{hPa}} qu\,dp \tag{1}$$

$$MF_y = -\frac{1}{g}\int_{1{,}000\ \mathrm{hPa}}^{300\ \mathrm{hPa}} qv\,dp \tag{2}$$

For each grid cell, the total moisture flux is calculated as:

$$\text{Total moisture flux} = \sqrt{MF_x^2 + MF_y^2} \tag{3}$$

We calculate composite maps of total moisture flux on EPCP and non-EPCP days. We calculate changes in moisture flux for each class as the difference in average moisture flux for the late period (2000–2019) compared to the early period (1981–1999), with the statistical significance of the difference calculated at each grid cell using the Mann-Whitney U-test.

Figure 2. (a) Schematic diagram of neural network architecture, input data, and output prediction. (b) Composite anomaly maps for extreme precipitation circulation patterns (EPCPs). (c) Probability of EPCP predicted by convolutional neural network (CNN) versus daily precipitation. Heatmap covers areas with >10 days/bin (points show individual days). Red shapes correspond to the three days in (i)–(t). Orange shapes correspond to examples in Figure S5. Vertical dashed line shows 95th percentile of daily precipitation. (d) Fraction of days classified as EPCPs for different precipitation amounts. (e)–(h) composite relevance maps for each class (EPCP and non-EPCP), calculated using layerwise relevance propagation (LRP; see Methods). (i) 500-hPa geopotential height (GPH) anomaly map for the >p95 day from the validation set with the highest EPCP probability (1986-09-11). (j) Relevance of each pixel in (i) for the final prediction. (k) SLP anomaly map on 1986-09-11. (l) Relevance of each pixel in (k) for the final prediction. (m–t) same as (i–l) for two more days from the validation set: >p95 day with lowest EPCP probability (m–p) and <p95 day with lowest EPCP probability (q–t).

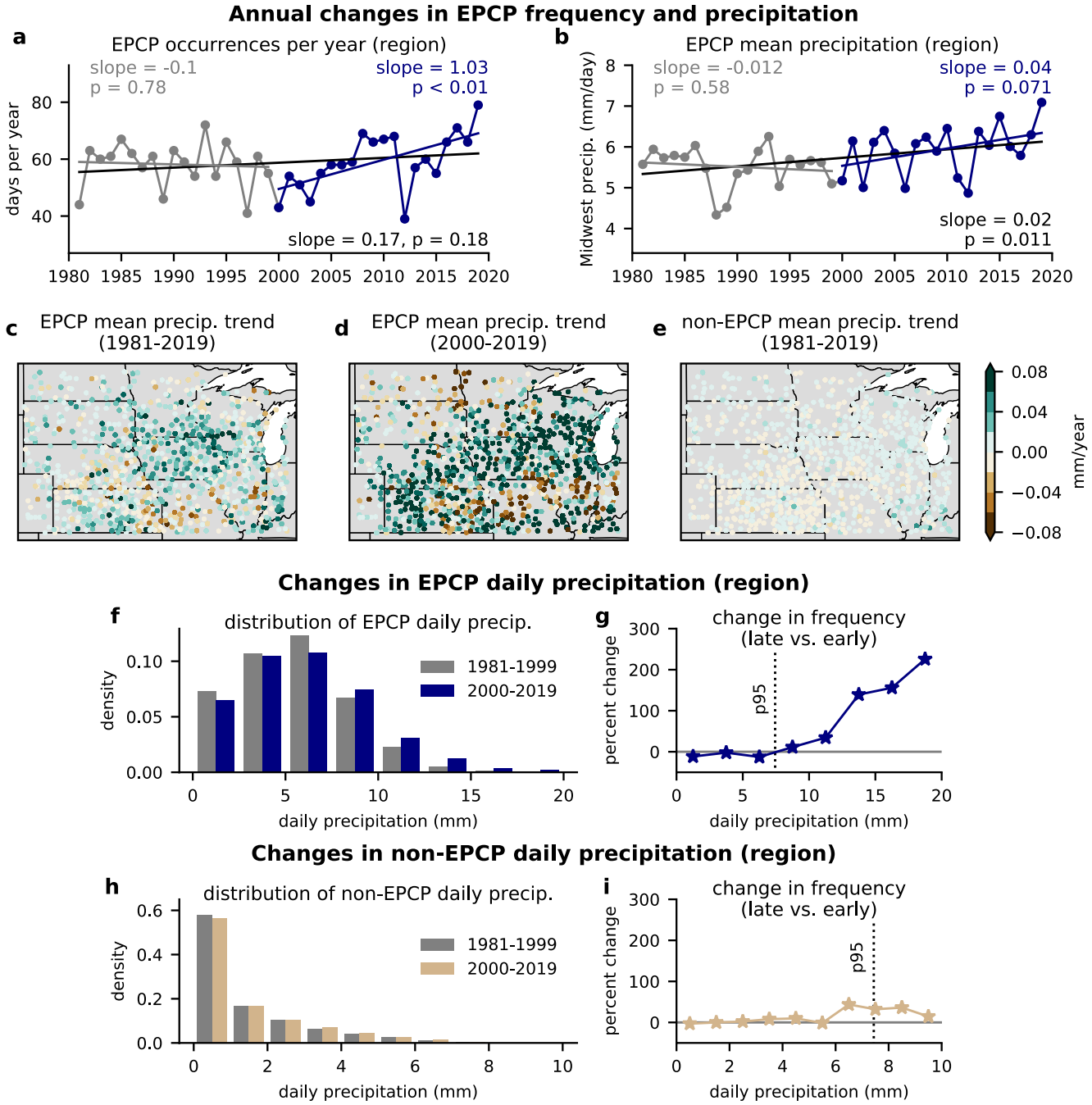

Figure 3. (a) Linear trend in annual extreme precipitation circulation pattern (EPCP) occurrences for the overall (black), early (gray), and late (blue) periods. Points show individual years. (b) Same as (a) for mean EPCP precipitation intensity. (c) Linear trend in EPCP mean precipitation (1981–2019) at Global Historical Climatology Network (GHCN) stations. (d) Same as (c), for the late period (2000–2019). (e) Same as (c) for non-EPCP days. (f) Distribution of EPCP daily precipitation for the early and late periods. (g) Change in relative frequency for each bin shown in (e). (h–i) Same as (f–g) for non-EPCP days (note different *x*-axis scale).

3. Results

3.1. Network Performance and Interpretation

The CNN correctly classifies 91% of Midwest extreme precipitation days as EPCPs, with an overall accuracy of 88% across both classes. (The CNN shows similar performance when evaluated using precipitation station data; Figure S4). Composite anomaly maps across EPCPs show a 500-hPa trough over the western

U.S., high pressure at 500-hPa along the U.S. East Coast, and an SLP low over the Midwest and extending south (Figure 2b). Of all days classified as EPCP patterns, 28% resulted in extreme precipitation. In contrast, only 0.5% of days classified as non-EPCP patterns resulted in extreme precipitation. The CNN shows predictive skill across all seasons, although there are fewer low-precipitation EPCP days in summer compared to other seasons (Figure S5). The CNN also correctly predicts extreme precipitation days at individual stations, although the predictive performance decreases for stations near the edge of the precipitation domain (Figure S4).

Although the network was trained using a fixed extreme precipitation threshold, the network shows skill in identifying days that are slightly below this threshold (Figures 2c, 2d and S6). On average, EPCP days that did not have > p95 precipitation still have higher precipitation than non-EPCP days (Figure 2c). Additionally, 100% of days that exceed higher precipitation thresholds (e.g., >12 mm) are correctly identified as EPCPs (Figure 2d).

We show composite LRP maps for each class (Figures 2e–2h), along with LRP results for selected days from the validation set (Figures 2i–2t and S3). The composite LRP maps on EPCP days show the highest relevance for pixels over the Midwest, indicating that the CNN focuses on circulation features in that area, despite having no information about the geographic location of the binary extreme precipitation timeseries. In contrast, there is no particular region that stands out as having the highest relevance on non-EPCP days, nor any clear spatial coherence to the composite LRP pattern (Figures 2g and 2h).

Examining LRP maps for individual days shows that the CNN identifies strong anomalous features in the input variables. For example, on the extreme precipitation day with highest EPCP probability (Figures 2i–2l), we find that the CNN identifies an anomalous 500-hPa trough over the Midwest (Figure 2j) and a negative SLP anomaly slightly eastward of the elevated trough (Figure 2l). In contrast, on the non-extreme day with the lowest EPCP probability (1981-09-19), the CNN identifies a negative 500-hPa GPH anomaly over the East Coast (Figures 2q–2t). On the extreme precipitation day with the lowest EPCP probably (2007-10-03; Figures 2m–2p), the areas of the highest relevance for the non-EPCP prediction include positive 500-hPa GPH anomalies over much of the U.S. Taken together, the relevance maps indicate that the network identifies a variety of features in the input data, but is particularly sensitive to the presence – and location – of strong, negative 500-hPa GPH or SLP anomalies.

3.2. Changes in Pattern Frequency, Precipitation, and Moisture Flux

To identify underlying causes of increasing Midwest extreme precipitation days (Figure 1), we calculate trends in the frequency of EPCPs, and trends in precipitation intensity on days when EPCPs occur. Over the recent period (2000–2019), the frequency of EPCPs increased at a rate of 1.03 days year^{-1} ($p < 0.01$; Figure 3a). However, this increase followed a period of equivocal – or even slightly decreasing – trend in EPCP frequency, and there is no significant trend in frequency for the 1981–2019 period overall (Figures 3a and S3). (We find similar results using different breakpoints between the early and late period; Table S1). Given that the wettest days are associated with high predicted EPCP probabilities (Figure 2c), we also assess changes in the frequency of days with EPCP probabilities >0.75 or >0.9. There is some evidence of an increasing frequency of days with EPCP probability >0.75, but no significant change in the frequency of days with EPCP probability >0.9 (Figures S7 and S8).

In contrast, precipitation intensity has increased significantly when EPCPs occur (Figure 3b), with mean precipitation on EPCP days increasing at a rate of 0.02 mm day^{-1} year^{-1} for the overall period ($p = 0.011$), and 0.04 mm day^{-1} year^{-1} over the last 20 years ($p = 0.071$). (Using a different breakpoint for the early and late periods yield similar trend estimates; Table S1). We also find similar increases in precipitation on days with EPCP probability >0.75 or >0.9 (Figures S7 and S8). Further, many individual stations have experienced increases in EPCP precipitation of a similar magnitude as the regional trend (Figures 3c and 3d), indicating that this result is robust across spatial scales and precipitation datasets.

The proportion of EPCPs that resulted in extreme precipitation also increased in the late period compared to the early period, while the proportion of EPCPs with low precipitation decreased (Figures 3f and 3g).

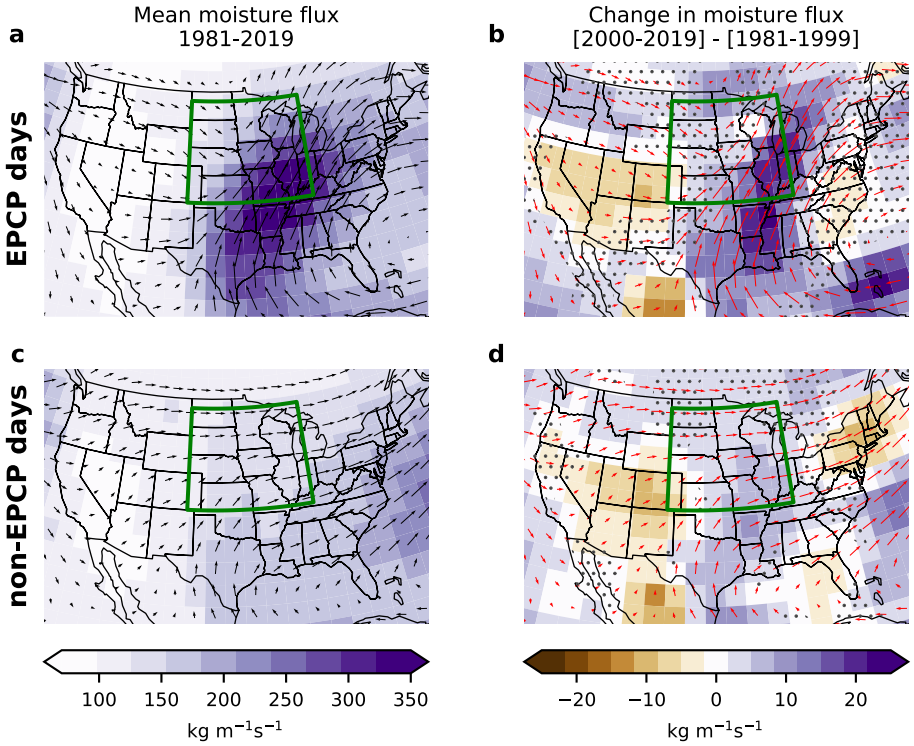

Figure 4. (a) Mean moisture flux on extreme precipitation circulation pattern (EPCP) days (precipitation region shown in green box). (b) Change in absolute moisture flux on EPCP days between the late and early periods. Red arrows show the magnitude and direction of mean moisture flux during the late period. Stippling indicates areas without statistically significant changes in moisture flux ($p > 0.05$), using a two-tailed Mann-Whitney U-test. (c–d) Same as (a–b) for non-EPCP days.

In contrast, changes in the distribution of daily precipitation were small for non-EPCP days (Figures 3h and 3i), providing further evidence that the observed regional increase in the frequency of extreme precipitation was primarily a result of increases in precipitation intensity during EPCPs.

The increases in precipitation intensity during EPCPs have co-occurred with preferential increases in moisture flux into the Midwest region. On average, EPCP days have high moisture flux from the Gulf of Mexico into the region (Figure 4a). In addition, there were statistically significant ($p < 0.05$) increases in moisture flux into the Midwest region on EPCP days in the late period compared to the early period (Figure 4b). In contrast, non-EPCP days show lower – and more zonal – moisture flux over the Midwest (Figure 4c), and smaller changes in moisture flux have occurred for non-EPCPs, especially over the Midwest (Figure 4d). Outside of the Midwest region, changes have been more similar across the two classes, including slight decreases in moisture flux over the Southwest. Together, these results show that there have been distinct changes in moisture flux into the Midwest on EPCP versus non-EPCP days, with the largest increases in moisture flux occurring during EPCPs.

4. Discussion and Conclusions

We use a CNN to identify days with EPCPs, and assess causes of observed increases in extreme Midwest precipitation. While there is some indication that EPCP frequency may have increased over the past two decades (Table S1), there have been no significant changes in EPCP frequency over the full 1981–2019 period (Figure 3). However, precipitation intensity has increased significantly when EPCPs occur, increasing the proportion of EPCP days exceeding the p95 threshold (Figure 3). The increase in EPCP precipitation intensity has co-occurred with preferential increases in moisture flux to the Midwest during EPCPs (Figure 4).

These results suggest an earlier emergence of thermodynamically driven changes in extreme precipitation compared to dynamically driven changes.

The CNN shows high accuracy in distinguishing EPCPs and non-EPCPs. On average, the CNN identifies EPCPs based on negative 500-hPa GPH anomalies in the western portion of the Midwest domain combined with negative SLP anomalies over the eastern part of the Midwest domain (Figure 2). This SLP anomaly pattern is physically consistent with the pattern of high moisture flux into the Midwest region from the Gulf of Mexico, as reflected in Figure 4. Thus, despite not having prior information about the geographic provenance of the extreme precipitation timeseries, the CNN uncovers physically meaningful features in this region.

However, some extremes are missed by the algorithm, such as October 3, 2007 (Figures 2m–2p). This day occurred during a period of unusual weather, with record heat and dry conditions for much of the region, and heavy precipitation concentrated in northern areas (Midwest Regional Climate Center, 2007). This example highlights that extreme precipitation is often controlled by localized processes not reflected in the large-scale SLP and GPH fields or the regional-mean daily precipitation. Additional input variables that reflect smaller-scale meteorological processes (e.g., vertical velocity, wind shear, and convective available potential energy; Flora et al., 2021) could improve the CNN accuracy for forecasting-related applications. The CNN output could also be redefined to focus on a smaller region or to capture days with sub-regional heterogeneity.

Although the CNN shows predictive accuracy across all seasons (Figure S5), there are seasonal differences in the distribution of precipitation during EPCPs, likely reflecting seasonality in the dominant causes of precipitation. Extreme precipitation in the Midwest results from distinct causes, including fronts, mesoscale convective systems, and tropical and extra-tropical cyclones (Kunkel et al., 2012). While the CNN used here groups all EPCPs into one category, extending our LRP analysis of "features" learned by the CNN may reveal distinct types of EPCPs, and would complement recent approaches to identify specific extreme weather "objects" (e.g., fronts or tropical cyclones) using deep learning (Biard & Kunkel, 2019; Prabhat et al., 2021; Racah et al., 2017). Additionally, incorporating temporal information, such as through a three-dimensional CNN, could provide insight into different causes of extreme precipitation.

Finally, because the CNN is trained on variables that are available globally, it could be readily retrained to analyze EPCPs in other regions. This approach could be useful in regions where limitations of climate models in simulating precipitation processes lead to high uncertainty in future changes in extreme precipitation. Because the CNN uses relatively coarse input data (2.5° horizontal resolution), the CNN could be used to evaluate GCM-simulated changes in EPCPs, allowing for analysis of longer historical trends, and separation of forced climate change from internal variability. Combined with deep learning visualization tools, the CNN could also provide a method to evaluate GCM biases in simulating EPCPs. This method of evaluation would represent a new pathway to reduce uncertainty in future projections of extreme precipitation phenomena that are closely linked with large-scale circulation patterns.

Although our initial case study is confined to a single type of event in a single region, the results demonstrate that machine learning can provide critical insight into the physical processes underlying changes in climate extremes. Because it is generalizable to a range of extreme events and regions, it represents a promising tool for both scientific understanding and the planning and adaptation required to reduce vulnerability to current and future climate change.

Data Availability Statement

PRISM daily precipitation is available from the PRISM Climate Group (http://www.prism.oregonstate.edu). The GHCN data is available from the National Oceanic and Atmospheric Administration (NOAA; https://www.ncdc.noaa.gov/ghcnd-data-access). The NCEP/NCAR-R1 reanalysis is also available from NOAA (https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html). Code supporting the analysis is provided in a Zenodo archive (https://doi.org/10.5281/zenodo.5104473).

Acknowledgments
The authors thank two anonymous reviewers for insightful and constructive comments. The authors thank the PRISM Climate Group and NOAA for data access. The Center for Computational Earth & Environmental Sciences and the Stanford Research Computing Center at Stanford University provided computational resources. Funding was provided by Stanford University.

References

Agel, L., Barlow, M., Feldstein, S. B., & Gutowski, W. J. (2018). Identification of large-scale meteorological patterns associated with extreme precipitation in the US northeast. *Climate Dynamics*, *50*(5–6), 1819–1839. https://doi.org/10.1007/s00382-017-3724-8

Alber, M., Lapuschkin, S., Seegerer, P., Hägele, M., Schütt, K. T., Montavon, G., et al. (2019). INNvestigate neural networks. *Journal of Machine Learning Research*, *20*, 1–8.

Allen, M. R., & Ingram, W. J. (2002). Constraints on future changes in climate and the hydrologic cycle. *Nature*, *419*(September), 224–232. https://doi.org/10.1038/nature01092

Barlow, M., Gutowski, W. J., Gyakum, J. R., Katz, R. W., Lim, Y. K., Schumacher, R. S., et al. (2019). *North American extreme precipitation events and related large-scale meteorological patterns: A review of statistical methods, dynamics, modeling, and trends. Climate Dynamics* (Vol. 53, pp. 6835–6875). Springer Berlin Heidelberg. https://doi.org/10.1007/s00382-019-04958-z

Barnes, E. A., Toms, B., Hurrell, J. W., Ebert-Uphoff, I., Anderson, C., & Anderson, D. (2020). Indicator patterns of forced change learned by an artificial neural network. *Journal of Advances in Modeling Earth Systems*, *12*(9). https://doi.org/10.1029/2020MS002195

Biard, J. C., & Kunkel, K. E. (2019). Automated detection of weather fronts using a deep learning neural network. *Advances in Statistical Climatology, Meteorology and Oceanography*, *5*(2), 147–160. https://doi.org/10.5194/ascmo-5-147-2019

Cassano, J. J., Uotila, P., Lynch, A. H., & Cassano, E. N. (2007). Predicted changes in synoptic forcing of net precipitation in large Arctic river basins during the 21st century. *Journal of Geophysical Research*, *112*(G4). https://doi.org/10.1029/2006JG000332

Cattiaux, J., Douville, H., & Peings, Y. (2013). European temperatures in CMIP5: Origins of present-day biases and future uncertainties. *Climate Dynamics*, *41*(11–12), 2889–2907. https://doi.org/10.1007/s00382-013-1731-y

Chattopadhyay, A., Hassanzadeh, P., & Pasha, S. (2020). Predicting clustered weather patterns: A test case for applications of convolutional neural networks to spatio-temporal climate data. *Scientific Reports*, *10*(1), 1–13. https://doi.org/10.1038/s41598-020-57897-9

CRED. (2018). *Natural disasters 2018*. Retrieved from https://emdat.be/sites/default/files/adsr_2018.pdf

Davenport, F. V., Burke, M., & Diffenbaugh, N. S. (2021). Contribution of historical precipitation change to US flood damages. *Proceedings of the National Academy of Sciences of the United States of America*, *118*, 1–7. https://doi.org/10.1073/pnas.2017524118

Diffenbaugh, N. S., Scherer, M., & Trapp, R. J. (2013). Robust increases in severe thunderstorm environments in response to greenhouse forcing. *Proceedings of the National Academy of Sciences*, *110*(41), 16361–16366. https://doi.org/10.1073/pnas.1307758110

Diffenbaugh, N. S., Singh, D., Mankin, J. S., Horton, D. E., Swain, D. L., Touma, D., et al. (2017). Quantifying the influence of global warming on unprecedented extreme climate events. *Proceedings of the National Academy of Sciences*, *114*(19), 4881–4886. https://doi.org/10.1073/pnas.1618082114

Ebert-Uphoff, I., & Hilburn, K. (2020). Evaluation, tuning, and interpretation of neural networks for working with images in meteorological applications. *Bulletin of the American Meteorological Society*, *101*(12), E2149–E2170. https://doi.org/10.1175/BAMS-D-20-0097.1

Esteban, P., Jones, P. D., Martín-Vide, J., & Mases, M. (2005). Atmospheric circulation patterns related to heavy snowfall days in Andorra, Pyrenees. *International Journal of Climatology*, *25*(3), 319–329. https://doi.org/10.1002/joc.1103

Flora, M. L., Potvin, C. K., Skinner, P. S., Handler, S., & McGovern, A. (2021). Using machine learning to generate storm-scale probabilistic guidance of severe weather hazards in the warn-on-forecast system. *Monthly Weather Review*, *149*(5), 1535–1557. https://doi.org/10.1175/MWR-D-20-0194.1

Fragoso, M., & Tildes Gomes, P. (2008). Classification of daily abundant rainfall patterns and associated large-scale atmospheric circulation types in Southern Portugal. *International Journal of Climatology*, *28*(4), 537–544. https://doi.org/10.1002/joc.1564

Gagne, D. J., II, Haupt, S. E., Nychka, D. W., & Thompson, G. (2019). Interpretable deep learning for spatial analysis of severe hailstorms. *Monthly Weather Review*, *147*(8), 2827–2845. https://doi.org/10.1175/MWR-D-18-0316.1

Gibson, P. B., Perkins-Kirkpatrick, S. E., Uotila, P., Pepler, A. S., & Alexander, L. V. (2017). On the use of self-organizing maps for studying climate extremes. *Journal of Geophysical Research: Atmospheres*, *122*(7), 3891–3903. https://doi.org/10.1002/2016JD026256

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. Retrieved from http://www.deeplearningbook.org

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* (Vol. 46). Springer Berlin Heidelberg. 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

Hilburn, K. A., Ebert-Uphoff, I., & Miller, S. D. (2021). Development and interpretation of a neural-network-based synthetic radar reflectivity estimator using GOES-R satellite observations. *Journal of Applied Meteorology and Climatology*, *60*(1), 3–21. https://doi.org/10.1175/JAMC-D-20-0084.1

Horton, D. E., Johnson, N. C., Singh, D., Swain, D. L., Rajaratnam, B., & Diffenbaugh, N. S. (2015). Contribution of changes in atmospheric circulation patterns to extreme temperature trends. *Nature*, *522*(7557), 465–469. https://doi.org/10.1038/nature14550

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437–471. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Kharin, V. V., Zwiers, F. W., Zhang, X., & Wehner, M. (2013). Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, *119*(2), 345–357. https://doi.org/10.1007/s10584-013-0705-8

Kunkel, K. E., Easterling, D. R., Kristovich, D. A. R., Gleason, B., Stoecker, L., & Smith, R. (2012). Meteorological causes of the secular variations in observed extreme precipitation events for the conterminous United States. *Journal of Hydrometeorology*, *13*(3), 1131–1141. https://doi.org/10.1175/JHM-D-11-0108.1

Lagerquist, R., McGovern, A., & Gagne, D. J., II (2019). Deep learning for spatially explicit prediction of synoptic-scale fronts. *Weather and Forecasting*, *34*(4), 1137–1160. https://doi.org/10.1175/WAF-D-18-0183.1

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Loikith, P. C., Lintner, B. R., & Sweeney, A. (2017). Characterizing large-scale meteorological patterns and associated temperature and precipitation extremes over the northwestern United States using self-organizing maps. *Journal of Climate*, *30*(8), 2829–2847. https://doi.org/10.1175/JCLI-D-16-0670.1

McGovern, A., Lagerquist, R., John Gagne, D., Jergensen, G. E., Elmore, K. L., Homeyer, C. R., & Smith, T. (2019). Making the black box more transparent: Uderstanding the physical implications of machine learning. *Bulletin of the American Meteorological Society*, *100*(11), 2175–2199. https://doi.org/10.1175/BAMS-D-18-0195.1

Merino, A., Fernández-Vaquero, M., López, L., Fernández-González, S., Hermida, L., Sánchez, J. L., et al. (2016). Large-scale patterns of daily precipitation extremes on the Iberian Peninsula. *International Journal of Climatology*, *36*(11), 3873–3891. https://doi.org/10.1002/joc.4601

Midwest Regional Climate Center. (2007). *Midwest weekly highlights – October 1-10, 2007*. Retrieved from https://mrcc.illinois.edu/cliwatch/0710/071010.htm

Min, S. K., Zhang, X., Zwiers, F. W., & Hegerl, G. C. (2011). Human contribution to more-intense precipitation extremes. *Nature*, *470*(7334), 378–381. https://doi.org/10.1038/nature09763

Montavon, G., Samek, W., & Müller, K.-R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, *73*, 1–15. https://doi.org/10.1016/j.dsp.2017.10.011

NOAA. (2021). *National Centers for Environmental Information (NCEI) U.S. Billion-Dollar Weather and Climate disasters*. https://doi.org/10.25921/stkw-7w73

O'Gorman, P. A. (2010). Understanding the varied response of the extratropical storm tracks to climate change. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(45), 19176–19180. https://doi.org/10.1073/pnas.1011547107

Papalexiou, S. M., & Montanari, A. (2019). Global and regional increase of precipitation extremes under global warming. *Water Resources Research*, 4901–4914. https://doi.org/10.1029/2018WR024067

Prabhat, M., Kashinath, K., Mudigonda, M., Kim, S., Kapp-Schwoerer, L., Graubner, A., et al. (2021). ClimateNet: An expert-labeled open dataset and deep learning architecture for enabling high-precision analyses of extreme weather. *Geoscientific Model Development*, *14*(1), 107–124. https://doi.org/10.5194/gmd-14-107-2021

Prein, A. F., & Mearns, L. O. (2021). U.S. extreme precipitation weather types increased in frequency during the 20th century. *Journal of Geophysical Research: Atmospheres*, *126*. https://doi.org/10.1029/2020JD034287

Prein, A. F., & Pendergrass, A. G. (2019). Can we constrain uncertainty in hydrologic cycle projections? *Geophysical Research Letters*, *46*(7), 3911–3916. https://doi.org/10.1029/2018GL081529

PRISM Climate Group. *Daily precipitation product 1981-2019*. Retrieved from http://prism.oregonstate.edu

Racah, E., Beckham, C., Maharaj, T., Kahou, S. E., Prabhat, & Pal, C. (2017). ExtremeWeather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. In *Advances in Neural Information Processing Systems, 2017-Decem(Nips)* (pp. 3403–3414).

Shaw, T. A., Baldwin, M., Barnes, E. A., Caballero, R., Garfinkel, C. I., Hwang, Y. T., et al. (2016). Storm track processes and the opposing influences of climate change. *Nature Geoscience*, *9*(9), 656–664. https://doi.org/10.1038/ngeo2783

Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W., & Bronaugh, D. (2013). Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *Journal of Geophysical Research Atmospheres*, *118*(4), 1716–1733. https://doi.org/10.1002/jgrd.50203

Singh, D., Horton, D. E., Tsiang, M., Haugen, M., Ashfaq, M., Mei, R., et al. (2014). Severe precipitation in Northern India in June 2013: Causes, historical context, and changes in probability. *Bulletin of the American Meteorological Society*, *95*(9), 558–561.

Singh, D., Swain, D. L., Mankin, J. S., Horton, D. E., Thomas, L. N., Rajaratnam, B., & Diffenbaugh, N. S. (2016). Recent amplification of the North American winter temperature dipole. *Journal of Geophysical Research: Atmospheres*, *121*(17), 9911–9928. https://doi.org/10.1002/2016JD025116

Swain, D. L., Horton, D. E., Singh, D., & Diffenbaugh, N. S. (2016). Trends in atmospheric patterns conducive to seasonal precipitation and temperature extremes in California. *Science Advances*, *2*(4), e1501344. https://doi.org/10.1126/sciadv.1501344

Toms, B. A., Barnes, E. A., & Ebert-Uphoff, I. (2020). Physically interpretable neural networks for the geosciences: Applications to earth system variability. *Journal of Advances in Modeling Earth Systems*, *12*(9), 1–20. https://doi.org/10.1029/2019MS002002

Trapp, R. J., Diffenbaugh, N. S., Brooks, H. E., Baldwin, M. E., Robinson, E. D., & Pal, J. S. (2007). *Changes in severe thunderstorm environment frequency during the 21st century caused by anthropogenically enhanced global radiative forcing*.

Trenberth, K. E. (1999). Conceptual framework for changes of extremes of the hydrological cycle with climate change. *Climatic Change*, *42*, 327–339. https://doi.org/10.1007/978-94-015-9265-9_18

Wilson, L. L., Lettenmaier, D. P., & Skyllingstad, E. (1992). A hierarchical stochastic model of large-scale atmospheric circulation patterns and multiple station daily precipitation. *Journal of Geophysical Research*, *97*(D3), 2791. https://doi.org/10.1029/91JD02155

Yin, J. H. (2005). A consistent poleward shift of the storm tracks in simulations of 21st century climate. *Geophysical Research Letters*, *32*(18). https://doi.org/10.1029/2005GL023684

Zhang, W., & Villarini, G. (2019). On the weather types that shape the precipitation patterns across the U.S. Midwest. *Climate Dynamics*, *53*, 4217–4232. https://doi.org/10.1007/s00382-019-04783-4

References From the Supporting Information

Araujo, A., Norris, W., & Sim, J. (2019). Computing receptive fields of convolutional neural networks. *Distill*, *4*(11). https://doi.org/10.23915/distill.00021

Chollet, F. (2015). *Keras*. Retrieved from https://keras.io