Potential Predictability of Regional Precipitation and Discharge Extremes Using Synoptic-Scale Climate Information via Machine Learning: An Evaluation for the Eastern Continental United States®

JAMES KNIGHTON

Department of Biological and Environmental Engineering, Cornell University, Ithaca, New York

GEOFF PLEISS

Department of Computer Science, Cornell University, Ithaca, New York

ELIZABETH CARTER

Department of Biological and Environmental Engineering, Cornell University, Ithaca, New York

STEVEN LYON

The Nature Conservancy, Southern New Jersey Office, Delmort, New Jersey

M. TODD WALTER AND SCOTT STEINSCHNEIDER

Department of Biological and Environmental Engineering, Cornell University, Ithaca, New York

(Manuscript received 13 September 2018, in final form 4 March 2019)

ABSTRACT

Current generation general circulation models (GCMs) simulate synoptic-scale climate state variables such as geopotential heights, specific humidity, and integrated vapor transport (IVT) more reliably than mesoscale precipitation. Statistical downscaling methods that condition precipitation on GCM-based, synoptic-scale climate features have shown promise in the reproduction of local precipitation. However, current approaches to climate-state-informed downscaling impose some limitations on the skill of precipitation reproduction, including hard clustering of climate modes into a discrete set of states, utilization of numerical clustering methodologies poorly suited to nonnormal data, and a tendency to focus on relationships to a limited set of large-scale climate modes. This study presents a methodology based on emerging machine learning techniques to develop global approximators of regional precipitation and discharge extremes given a suite of synoptic-scale climate state variables. Archetypal analysis is first used to define regional modes of winter and summer extreme precipitation and discharge across the eastern contiguous United States. A 2D convolution neural network (NN) is then used to predict the co-occurrence of the archetypes using 300- and 700-hPa geopotential heights, 300- and 700-hPa specific humidity, and IVT. Results suggest that 300-hPa geopotential height, 700-hPa specific humidity, and IVT yield the most reliable predictions, although with some important differences by season and region. Finally, we demonstrate that the trained activations of NN convolutional layers can be used to infer the causal pathways between synoptic-scale climate features and regional extremes.

1. Introduction

Extreme precipitation and associated hydrological extremes have major social and economic impacts across the United States and globally. A review of U.S. insurance claims exceeding one billion dollars from 1980 to 2011 shows that the most severe economic losses are dominated by large-scale storm events that induce riverine flooding (NCDC 2018). It is generally accepted

DOI: 10.1175/JHM-D-18-0196.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JHM-D-18-0196.s1.

Corresponding author: James Knighton, jok8@cornell.edu

that global climate change will result in thermodynamic intensification of the hydrologic cycle, which will increase the frequency and magnitude of precipitation extremes (Trenberth 2011). Somewhat more uncertain though is the impact that climate change will have on extremes related to shifting atmospheric dynamics, that is, how the atmospheric circulation and storm tracks will change under warming and the implications for the frequency of extreme storms in different regions (e.g., Shaw et al. 2016).

Precipitation in the Northern Hemisphere midlatitudes is strongly influenced by large-scale atmospheric circulation (Woollings et al. 2010; Farnham et al. 2017; Nakamura and Huang 2018; Conticello et al. 2018). During North American winters, for example, shifts in the polar jet stream on daily to weekly time scales steer Pacific (Farnham et al. 2017) and Atlantic (Nakamura and Huang 2018) storm tracks, significantly altering surface precipitation and hydroclimatic risk across the United States (Lukens et al. 2018). Agel et al. (2018) and Roller et al. (2016) demonstrate robust correlations between regional jet stream features and seasonal precipitation extremes across the northeastern United States, while similar teleconnections have been demonstrated for daily tropical moisture export (TME)derived rainfall across the eastern United States (Steinschneider and Lall 2015) and Europe (Conticello et al. 2018).

General circulation models (GCMs) and later generation Earth system models (ESMs) are valuable tools for projecting global changes in atmospheric dynamics that control extreme precipitation under future greenhouse gas emissions scenarios. However, the fidelity of localscale weather in GCMs is heterogeneous across seasons (e.g., Knighton et al. 2017a), space (e.g., Wójcik 2015; Bock et al. 2018), and climate variables (e.g., surface temperatures, geopotential heights, precipitation; Strobach and Bel 2017) owing to differences in the complexity of the physics driving each process, data availability, and numerical limitations imposed by the horizontal spatial grid. These errors are often particularly large for precipitation (Stephens et al. 2010), which, in addition to low GCM resolution, complicates the direct use of climate model projections for hydrologic impact studies (Xu 1999). Recently, these challenges have led some to argue that more robust projections of shifting extreme events can be achieved by characterizing the linkages between large-scale, dynamical features of atmospheric circulation and regional hydroclimate extremes directly (Merz et al. 2014). These statistical downscaling approaches offer flexibility in how information is extracted from GCMs, leveraging the fact that synoptic-scale state variables are often more reliably simulated by GCMs than local-scale weather (e.g., Murawski et al. 2016, 2018; Farnham et al. 2018). Importantly though, these methods rely on careful consideration of the most significant relationships between synoptic-scale climate features and precipitation extremes, and are best suited for regional extreme events rather than localized events such as those driven by spatially confined areas of strong convection.

Studies linking extreme precipitation to features of atmospheric circulation have often focused on relationships to popular and well understood climate oscillations (e.g., El Niño-Southern Oscillation, North Atlantic Oscillation, Pacific decadal oscillation; e.g., Steinschneider and Lall 2016; Lee et al. 2018), which are usually defined by the first few EOFs of large-scale geopotential height or SST fields. While a focus on major modes of climate variability can be useful in simplifying the analysis, they may not be linked to extreme events in a particular region, limiting the utility of this approach. Precipitation extremes are likely a complex function of atmospheric conditions across multiple scales (Agel et al. 2019; Conticello et al. 2018; Farnham et al. 2018), and so new insights and predictive skill may be possible by relaxing the reliance on a priori beliefs about which physical teleconnections are most meaningful. Accordingly, many recent studies have begun to examine a larger suite of synoptic-scale "climate states" identified for smaller regions. They have also explored the use of alternative clustering algorithms to identify these states, including archetypal analysis (AA; e.g., Steinschneider and Lall 2015; Hannachi and Trendafilov 2017), K-means clustering (e.g., Roller et al. 2016), simulated annealing and diversified randomization (SANDRA) classification (e.g., Murawski et al. 2016), and self-organizing maps (e.g., Conticello et al. 2018; Agel et al. 2018). Similar clustering is also often conducted on local precipitation or wind speeds to identify weather types, and connections between climate states and weather types are then evaluated based on coincident frequency of the previously identified clusters or through the development of conditional pdfs of precipitation extremes. For instance, Roller et al. (2016) used K-means clustering to cluster 850-hPa wind fields across the northeast United States into five climate states. They demonstrate that extreme winter precipitation in the Northeast is most frequently associated with synoptic-scale states indicative of strong extratropical cyclones. Agel et al. (2019) furthered this work and incorporated both synoptic-scale and subregionalscale states to characterize localized climate features in the northeast United States. Conticello et al. (2018) similarly clustered near-surface geopotential heights

and integrated vapor transport (IVT) through selforganizing maps to derive conditional pdfs of precipitation extremes across a collection of rain gauges in Italy.

While these approaches have advanced the identification of linkages between large-scale climate patterns and regional extremes, some limitations remain. First, clustering algorithms, while useful to reduce the dimension of the problem, may conceal important linkages between synoptic-scale climate fields and regional extremes. This is particularly true if clustering is conducted on both large-scale climate and extreme precipitation, since the identified clusters may highlight independent modes of variability in each field at the expense of joint patterns of variation.

In addition, the clustering techniques used may obscure important features of climate states and extreme precipitation. For instance, it is common to impose subjective spatial constraints on the domain under examination, and some of the methodologies (e.g., EOFs, K-means clustering) are limited in their ability to represent patterns in nonnormal and skewed data, both of which can lead to spurious patterns in the identified clusters. Further, many of the methods discussed above also tend to assume that each day in a historical record can be meaningfully assigned to one weather type or climate state from a discrete set of patterns. While this approach provides an unambiguous classification, hard clustering possibly obfuscates useful information on the distance of each record from its prototypical mean, and possibly ignores simultaneous physical expression of multiple climate patterns.

The present study is motivated by the need to better define the limits of predictability of regional hydroclimate extremes using synoptic-scale climate features, while minimizing the methodological limitations described above. We aim to understand the degree to which large-scale circulation features that are credibly simulated by GCMs can be used to understand changes in regional extreme event risk due to shifts in atmospheric dynamics under climate change. The results of this work can help set expectations regarding the limits of inference on future extremes through a focus on large-scale circulation dynamics, as has been recently suggested (Merz et al. 2014).

We employ a methodology centered on emergent machine learning techniques, a broad family of algorithms that can relax the need to overly constrain the problem statement a priori. The approach consists of two primary steps: 1) the use of AA to characterize dominant modes of extreme regional precipitation and discharge by season, and 2) the use of a 2D convolutional neural network (NN) to predict seasonal archetypes using gridded, synoptic-scale climate information. These methodological advances are forwarded as a way to address the following questions:

- How well can regional hydroclimatic extremes be characterized using only features of large-scale circulation?
- Which features of large-scale circulation are most relevant for predicting regional extreme events?
- How does this vary by season and region?

We evaluate these questions through a case study of observed precipitation and simulated discharge extremes across the eastern contiguous United States (CONUS). This region is selected for the wide availability of hydrometeorological observations, the propensity for extreme precipitation and riverine flooding hazards (e.g., Lukens et al. 2018; Buchanan et al. 2018), and strong influence of synoptic-scale circulation on regional extreme events (e.g., Nakamura et al. 2013; Steinschneider and Lall 2015, 2016; Farnham et al. 2017, 2018; Roller et al. 2016; Agel et al. 2018). The remainder of this paper proceeds as follows. Section 2 describes the use of AA and an NN to determine the potential predictability of regional extremes based on synoptic-scale climate information. Details of our application to the eastern CONUS, including the data, land surface model, and evaluation methodology, are described in section 3. Results are presented in section 4, and the paper concludes with a discussion of broader impacts and potential future research needs that follow from this work.

2. Methodology

a. Archetypal analysis to identify space-time organization of regional hydroclimate extremes

The AA technique, first applied in the context of facial pattern recognition, is a method that defines modes of variability in multivariate databased on extremal points in the dataset (Cutler and Breiman 1994). When applied to hydrometeorological data, AA uncovers "archetypal patterns" or "idealized examples" of space-time organization in regional precipitation. As demonstrated in Steinschneider and Lall (2015), AA is designed to objectively define coherent patterns of regional extreme events in an underlying data field. These patterns are based on observations along the convex hull (i.e., outer edge) of the multivariate dataset. Each archetype represents a spatial pattern in the data, similar to loading patterns in EOF analysis. Individual observations of the data at each time step are then represented as convex combinations of the archetypes (i.e., linear combinations with nonnegative weights that sum

to unity), similar to how principal components are a linear combination of EOFs and the original data. In this way, AA is analogous to EOF analysis, but presents an alternate and novel approach for identifying the spatiotemporal structure that avoids the limitations of EOF analysis when applied to extremes (Steinschneider and Lall 2015; Hannachi and Trendafilov 2017).

Consider an $n \times k$ matrix **X** with *n* observations across *k* locations, with $\mathbf{x}_i^{\mathrm{T}} = [x_{i,1}, \ldots, x_{i,k}]$ a single multivariate observation. AA defines *p* characteristic archetypes, or spatial patterns, across the *k* locations, with each of the $j = 1, \ldots, p$ archetypes represented by a different weighting vector $\mathbf{w}_j^{\mathrm{T}} = [w_{1,j}, \ldots, w_{k,j}]$. By construction, each archetype is defined as a convex combination (i.e., a mixture) of the original observations:

$$\mathbf{w}_{j} = \sum_{i=1}^{n} \boldsymbol{\beta}_{i,j} \mathbf{x}_{i},$$

where

$$\sum_{i=1}^{n} \boldsymbol{\beta}_{i,j} = 1 \quad \text{and} \quad \boldsymbol{\beta}_{i,j} \ge 0$$

The archetypes can be organized into a $p \times k$ matrix **W**, with the convex weights grouped into an $n \times p$ matrix **B**, such that **W** = **B**^T**X**. Cutler and Breiman (1994) showed that the archetypes **W** lie on the convex hull of **X**, suggesting that the resulting archetypes will represent hydrometeorological extremes in different regions of the domain (as well as a null archetype, associated with zero values across the domain). The observations at each time step can then be reconstructed as a separate convex combination of the archetypes:

$$\hat{\mathbf{x}}_i = \sum_{j=1}^p T_{i,j} \mathbf{w}_j$$

where

$$\sum_{j=1}^{p} T_{i,j} = 1 \quad \text{and} \quad T_{i,j} \ge 0$$

The $n \times p$ matrix **T** contains the convex weights such that $\hat{\mathbf{X}} = \mathbf{TW}$. Matrix **T** describes the temporal expression of the archetypes, where $T_{\cdot,j}$ is a vector of values between 0 and 1 that can be interpreted as a time series of the relative expression of the *j*th archetype across time $i = 1, \ldots, n$. Extreme precipitation events are defined as days on which the value of $T_{i,j}$ approaches 1 for any archetype *j*, on any day *i*, associated with a spatial loading of precipitation. For a given *p*, archetypes are identified through numerical approaches that iteratively perturb the matrices **T** and **B** to search for the minimization

of the sum of square residuals (RSS), $\mathbf{X} - \mathbf{TB}^{T}\mathbf{X}$ (see Steinschneider and Lall 2015).

To determine a suitable number of archetypes p to analyze, the optimal value of RSS is calculated for different values of p. The number of archetypes is then selected by determining when declines in RSS decelerate as additional archetypes are added (i.e., similar to a scree test for EOF analysis). Appropriate selection of p guarantees that large expressions of **T** are representative of precipitation extremes.

b. 2D convolutional neural networks to identify teleconnections

We seek to determine the potential predictability of regional extreme events, as represented by the archetypes above, given varied climate state variables that are reliably simulated by GCMs. We use a NN instead of other approaches (e.g., multivariate regression) because NNs are well suited for the identification of meaningful relationships in high-dimensional space and require few a priori assumptions about the structure of the underlying data and relationships between the inputs and outputs. Neural networks can approximate any function to an arbitrary degree of precision (Goodfellow et al. 2016), and are therefore able to model complex relationships between input and output variables that might not be captured by simpler models. As more data become available for training, NNs will approach the Bayes optimal classifier (Duin and Tax 2005), even if this comes at the expense of interpretability of the underlying relationships (Montavon et al. 2018). Of course a true upper limit on predictability using NNs cannot be guaranteed, as this method (like all others) still has underlying assumptions in its architecture that may limit prediction. However, NNs are sufficiently flexible in design that they provide a better approximation of the potential predictability of extremes using high-dimensional synoptic-scale climate information, as compared to simpler methods (Rawat and Wang 2017).

We develop a deeply connected 2D convolutional NN (Table 1, Fig. 1) within Keras (Chollet 2015), similar to that described in Krizhevsky et al. (2012). This network architecture utilizes the spatial structure of a 2D image input layer [i.e., a horizontal ($x \times y$) grid of synoptic-scale climate at time *i*; layer 1 in Table 1] to produce an output of *p* classifications (i.e., simultaneous expression of each archetype $T_{i,.}$ at time *i*; layer 10 in Table 1). A series of hidden (i.e., intermediate) layers connect the horizontal grid to the output of *p* classifications. These layers are composed of a series of neurons *h*, where each neuron is defined by a function of connected input neurons from the previous layer, and a corresponding

Layer	Description	Dimensions of features/outputs	n parameters	Activation
1	Input	$x \times y \times 1$	_	_
2	Convolutional 2D (4×4 kernel, stride 1)	$77 \times 21 \times 1$	15	ReLU
3	Max pooling $(2 \times 2 \text{ pool size})$	$35 \times 10 \times 1$	0	None
4	Batch normalization	$35 \times 10 \times 1$	12	None
5	Convolutional 2D (4×4 kernel, stride 1)	$32 \times 7 \times 3$	147	ReLU
6	Max pooling $(2 \times 2 \text{ pool size})$	$16 \times 3 \times 3$	0	None
7	Batch normalization	$16 \times 3 \times 3$	12	None
8	Flatten	144×1	0	None
9	Dense (40% dropout)	64 imes 1	20 880	ReLU
10	Dense	$p \times 1$	580	Softmax

TABLE 1. Deep convolutional neural network architecture for prediction of archetype expression given a daily climate input field. Intermediate layer dimensions presented are established for the data used in this study (inputs x = 78, y = 22, and output p = 5). The column "*n* parameters" indicates the number of calibrated parameters within each layer.

series of weights u. Each layer type performs a similar operation, but differs in the number of input neurons connected, and how those input neurons are used to predict values of neurons in the next layer. Convolutional 2D and max pooling layers retain important information on the spatial patterns of inputs, using a sliding window to pool information across a matrix of input neurons that captures spatial information in the horizontal grid of synoptic-scale climate information. A convolutional 2D layer calculates the dot product of connected input neurons across all layer weights, while max pooling layers retain only the maximum value neuron within each pool. Flatten layers do not modify the neuron values, but reshape the neurons from a spatial grid to a vector, while dense layers connect each input neuron in that vector to each output neuron in the next layer. In dense layers, dropout is used to randomly remove a proportion of

neuron connections for network training in an attempt to limit network overfitting. In addition, L2 regularization with a weight of 0.01 was applied to each convolutional and densely connected layer. Rectified linear unit (ReLU) activation is employed so that negative neuron values are set to 0. Finally, Softmax activation on the output of the densely connected layer forces the algorithm to assign probabilities between (0, 1) to each output node, which respects archetype structure (i.e., $\sum_{j=1}^{p} T_{ji} = 1$ for each observation i). By simultaneously predicting the p archetype values for each day, the NN can help explore the feasibility of predicting more complex combinations of the archetypal patterns, rather than a hard binning of each historical day into one prototypical pattern. This is akin to a fuzzy classification problem (Huth et al. 2008). The network is evaluated with the Adam optimizer using mean square error (MSE) as the



FIG. 1. Schematic of computational workflow in a deeply connected convolutional NN showing a convolutional 2D layer, a max pooling layer, a flatten layer, and a densely connected layer. The sliding window in the convolutional 2D layer maps subgrids of $a \times b$ cells in the original $x \times y$ grid of synoptic climate information to a neuron in the next layer, while maintaining the gridded structure of information. Similarly, the sliding window in the max pooling layer maps subgrids of $c \times d$ neurons to a neuron in the next layer, which also maintains the gridded structure. Eventually, the grid of neurons is flattened into an e-vector, which is used to predict archetype expressions through a Softmax activation.

objective function, and is trained with a batch size of 40 samples, across 40 epochs.

3. Application to the eastern CONUS

NN predictions are evaluated for daily precipitation and discharge archetypes across the eastern CONUS (east of 90°W) for cold and warm seasons and different synoptic-scale climate fields. More detail on the development and evaluation of the precipitation and discharge archetypes, their prediction using 2D convolutional NNs, and the interpretability of the NN output is provided below.

a. Discharge and precipitation archetypes

Cold season (October–March) and warm season (April–September) precipitation archetypes are developed from CPC Unified Gauge Network daily gridded precipitation collected between 1948 and 2017 at a 0.25° horizontal resolution (NOAA 2018). This precipitation product is derived from an interpolation of a dense network of weather stations (Xie et al. 2007; Chen et al. 2008) and therefore avoids some of the problematic issues with GCM- and reanalysis-derived precipitation (e.g., Wuebbles et al. 2014; Janssen et al. 2016; Knighton et al. 2017a). For each season, we develop p = 1, ..., 20 archetypes and select the number of archetypes p based on inflection points in a scree plot.

Cold season (November-April) and warm season (May-October) stream discharge archetypes are developed from simulated daily discharge at a 0.25° resolution using a land surface model (LSM) to translate available daily meteorological data into daily streamflow. The seasonal ranges are shifted forward by one month compared to the precipitation archetypes to avoid the influence of spring snowmelt on warm season discharge archetypes. Simulated discharge is developed using a semiphysically based lumped LSM (Archibald et al. 2014) developed for the United States east of 90°W that computes surface runoff via the curve number approach modified to simulate variable source area (VSA) hydrology (Schneiderman et al. 2007; Easton et al. 2008). The model simulates snow accumulation and melt with the physically based energy budget model of Walter et al. (2005). For conciseness, additional details on the LSM model, including additional data inputs, fitting procedures, and validation, are presented in supplementary material (Fig. S1 in the online supplemental material). Scree tests are also used to select the number of discharge archetypes for each season.

To ensure that the identified archetypes correlate with extreme events, we compare the time expression $T_{.j}$ of each archetype to catalogs of historical weather and

socioeconomic loss data. For cold season precipitation archetypes, we utilize the U.S. National Centers for Environmental Information (NCEI) Storm Events Database between 1982 and 2017 (NCEI 2018). This record is filtered to include only weather events east of 90°W and from the following reported precipitation mechanisms: blizzard, flash flood, flood, heavy rain, heavy snow, hurricane, lake effect snow, tropical storm, tropical depression, thunderstorm, winter storm, and winter weather. We cluster each record in T_{ij} into days with and without an extreme weather event as determined by the Storm Events Database, and compare the conditional cumulative distribution function (cdf) of T_{ij} across these two groups. The degree of separation between the cdfs highlights the association between increases in a given archetype and observed extremes. The analysis is repeated for daily discharge archetypes using the NCDC record of weather extremes resulting in billion dollar losses (NCDC 2018). This loss record is filtered to only include losses associated with riverine flooding resulting from atmospheric mechanisms.

b. Predictions of archetypes via 2D convolutional NN

Of the four sets of archetypes above, we use three in a NN prediction model: cold season precipitation, warm season precipitation, and warm season discharge. We do not attempt to predict cold season discharge with synoptic-scale fields because snow accumulation and melt processes are not captured in the climate data alone. Five climate fields are considered as predictors for the precipitation and discharge archetypes: daily 300- and 700-hPa geopotential heights, 300- and 700-hPa specific humidity, and IVT. All data above are taken from the NCEP-NCAR Reanalysis 1 project (Compo et al. 2011) over the region 12.5°-65°N, 227.5°-35°W and are available at a $2.5^{\circ} \times 2.5^{\circ}$ resolution between 1948 and 2017. IVT is calculated as in Zhu and Newell (1998) by integrating the product of specific humidity q and both zonal u and meridional v wind speeds between 1000 and 300 hPa:

$$IVT = \sqrt{\left(\int_{1000}^{300} \frac{uq}{g} dp\right)^2 + \left(\int_{1000}^{300} \frac{vq}{g} dp\right)^2},$$

where g is the standard acceleration of gravity (9.8 m s^{-2}) . We use Reanalysis I for all fields above to support the inclusion of more extreme events in the analysis, prior to the start of more resolved reanalysis products in 1979. In addition, the selection of the rather coarse (2.5°) reanalysis dataset was intentionally chosen to align with the average horizontal resolution of GCMs used in the CMIP5 ensemble of long-term climate projections (Taylor et al. 2012). We note that higher-resolution input datasets could yield stronger identification of atmospheric



FIG. 2. Winter and summer archetypes of daily precipitation across the eastern United States.

teleconnections, although this would increase the ratio of grid cells to observations and could lead to overfitting.

Prediction skill is evaluated for one input climate variable at a time to understand the relative importance of each for different precipitation and discharge archetypes. Accordingly, the NN inputs are a batch of twodimensional 78 \times 22 \times 1 inputs for each day of the record, that is, one 2.5° synoptic-scale climate state variable ($x \times y$ in Table 1). All input climate state variables are locally centered and scaled prior to use in the NN. The NN to predict all archetypes is trained only for days when the null archetype (i.e., the archetype associated with zero precipitation or discharge) is expressed less than 0.5. This approach emphasizes extreme and heavy (high) precipitation (discharge) events when training the NN. The resulting datasets are 2062, 2734, and 1071 days for cold season precipitation, warm season precipitation, and warm season discharge, respectively. Due to the limited samples available, we do not use distinct training, validation, and testing datasets. Instead, we utilize an approach described in Goodfellow et al. (2016) in which the model is fit to 80% of the available data and validated against the remaining 20%. This analysis is repeated for five iterations, each time selecting a new 20% of the samples as validation. This approach allows all samples to be used in validation to more accurately assess model prediction skill, although some overfitting to the validation dataset, and deflation of MSE scores, is still possible (Russell and Norvig 2016).

Finally, the results of the NN prediction for one of the winter precipitation archetypes is interpreted through an evaluation of the activations associated with the first 2D convolutional layer trained on the IVT data, similar to methodologies described by Zeiler and Fergus (2014) and Yosinski et al. (2015). The relative simplicity of the NN applied in this problem allows for some interpretation of the first layer activations. By reviewing the activations we hope to understand what climate information is preserved and passed on to subsequent layers, and is therefore predictive of teleconnections. This enables some inference on physical mechanisms linking the climate fields to the archetypes, although it is important to recognize that additional layers within the network obscure the final linkages made by the model. Interpretation of NN activations for a second winter precipitation archetype and 300-hPa geopotential height inputs are provided in the supplementary material.

4. Results

a. Archetypes of daily precipitation and stream discharge

1) DAILY PRECIPITATION

Scree tests suggest that p = 5 archetypes are sufficient to describe a substantial proportion of variability within the dataset, and additional explained variance increases only marginally with more archetypes. Figure 2 shows

JOURNAL OF HYDROMETEOROLOGY



FIG. 3. Winter and summer archetypes of daily discharge across the eastern United States.

the five archetypes for cold and warm season precipitation across the eastern CONUS. The null archetype (first column, Fig. 2) represents the case of no precipitation over the domain and is the most common. The remaining archetypes represent spatially coherent modes of precipitation across the eastern United States and are very consistent across the two seasons. Most nonnull archetypes are elliptical with a characteristic elongation from southwest to northeast, which is consistent with the westward and poleward direction of storm tracks over the CONUS and associated moisture flux in the warm conveyer belt of extratropical cyclones. This is particularly the case in the cold season for all regions (Great Plains, P-W2; Northeast, P-W3; Southeast, P-W4; and Gulf region, P-W5). In contrast, some archetypes for warm season precipitation archetypes are less elliptical (Great Plains, P-S2), suggesting a different delivery mechanism. The warm season archetypes are generally consistent with the June-August archetypes identified by Steinschneider and Lall (2015), with precipitation over the Great Plains region (P-S2) and the Southeast (P-S4) being the most common, and less frequent, intense precipitation identified in the Northeast (P-S3). In contrast to our conclusion, Agel et al. (2015) suggest that daily extreme summertime precipitation over the Northeast is not spatially cohesive, although Agel et al. (2015) considered a relatively low density of point measurements of precipitation and defined the Northeast to be bounded below by New York. Our

archetypal analysis of daily precipitation suggests that extreme weather events affecting the Northeast (Fig. 2, P-W3 and P-S3) may commonly center on New York and extend south to 35°N. Further, the small number of archetypes selected here likely ignores spatially heterogeneous extremes linked to localized convection, and rather focuses on larger-scale mechanisms of moisture delivery to the region.

2) DAILY DISCHARGE

Similar to precipitation, p = 5 archetypes were selected based on scree tests for the discharge archetypes on both seasons. In part, the similarity in the number of selected archetypes is driven by the link between the two processes, that is, large-scale heavy precipitation drives regional floods. However, the number of archetypes to select was less clear from the scree tests for discharge compared to precipitation. We select five discharge archetypes for each season as a reasonable choice based on the scree test and for consistency with the precipitation archetypes. Figure 3 shows these cold and warm season discharge archetypes.

As with precipitation, the first archetype of each season represents the null case of little significant streamflow (i.e., base flow conditions) across the eastern United States. In the cold season, some discharge archetypes spatially align with the precipitation archetypes, for instance (D-W3, P-W5) and (D-W4 and P-W2). Other winter discharge archetypes (D-W2 and D-W5)



FIG. 4. CDFs of precipitation archetype expressions on days with no reported NCEI storm event (blue) and a reported storm event (red).

deviate from winter precipitation archetypes, particularly in the northern latitudes where they exhibit stark latitudinal gradients linked to snowpack accumulation and melt. The spatial coherence of discharge archetypes is larger during the cold season, particularly for archetypes D-W3 and D-W4. During this season the land surface is generally saturated with reduced evapotranspiration and latent heat losses. As a result, precipitation as rainfall is more directly translated into stream discharge. Warm season discharge archetypes are more spatially constrained than cold season discharge archetypes, as well as warm season precipitation archetypes (Fig. 2). This result is mostly likely due to canopy interception and land surface infiltration (Knighton et al. 2017b), allowing only precipitation inputs in excess of the initial abstraction to contribute to streamflow, as surface runoff or groundwater respectively. Summer discharge archetypes (at least the five assessed here) are more indicative of Gulf and Atlantic Coast landfalling weather systems with high discharge directly along the coasts. If seven discharge archetypes are utilized, the algorithm highlights a Great Plains discharge archetype (see Fig. S2).

3) COMPARISON TO WEATHER EVENT CATALOGS

The daily precipitation archetype time series \mathbf{T} are grouped into two datasets: days on which at least one

precipitation event was reported in the NCEI Storm Events Database, and days with no reported weather events. Figure 4 shows the conditional cdfs of the time series expressions across all precipitation archetypes (i.e., columns of **T**). Days with no NCEI reported storm events have a higher probability that the null archetype is expressed (Fig. 4, column 1). Conversely, the nonnull winter and summer precipitation archetypes become more likely during days with reported weather events.

The separation between cdfs is generally greater across all cold season archetypes than warm season archetypes. This result is possibly attributable to cold season precipitation extremes predominantly occurring as regional events which are more likely to be captured by AA, versus more localized mechanisms that dominate summer extremes. This comparison generally supports the precipitation archetypes as a meaningful clustering of weather events. Despite this result, it is worth mentioning that the NCEI Storm Events Database is likely an incomplete record of historical events, since the magnitude of variations in annual storm frequency tends to correlate with state population density, suggesting issues with how events are observed and reported. A similar comparison of discharge archetypes to a catalog of billion dollar loss estimates is further discussed in the supplementary material (Fig. S3).

b. Convolutional neural network

1) NN PREDICTIONS OF PRECIPITATION ARCHETYPES

The NN was trained on each synoptic climate field separately to predict the expression of all five warm and cold season precipitation archetypes on each day. The pdf of each target archetype is highly nonnormal, with a positive skew. The mode of each archetype is 0, yielding an NN preference to set predicted archetypes to low values to minimize MSE, presenting an inherent challenge for network training. However, despite these challenges, the resulting network demonstrates skill in the reproduction across all archetypes for both warm and cold season precipitation (Figs. 5, 6). We only show predictions for nonnull archetypes to focus the discussion on the ability of the NN to predict extreme precipitation events.

In the cold season, Fig. 5 shows that both upper-(300 hPa) and lower-level (700 hPa) geopotential height provide similar information on extreme precipitation with respect to MSE, but upper-level pressures show a less biased prediction. This result is not surprising, as perturbations to the jet stream tend to guide winter storm tracks (Nakamura and Huang 2018). For several cold season archetypes, the specific humidity field at the 700-hPa level outperforms the 300-hPa specific humidity field, which is expected as specific humidity at 700 hPa is more representative of the mass of available atmospheric moisture for precipitation. IVT represents the most consistent predictor of cold season precipitation archetypes, outperforming the other predictors when considering all archetypes.

In the warm season, the relative predictability afforded by lower- (700 hPa) and upper-level (300 hPa) geopotential heights is reversed as compared to the cold season. This reflects the fact that jet stream dynamics are less dominant over North America in the warm season, and low level circulation patterns (e.g., the North Atlantic subtropical high) play a larger role in controlling moisture transport over the continent (Steinschneider and Lall 2015). Similar to the cold season though, low-level specific humidity and particularly IVT provide strong predictors of warm season archetypes. Overall, our results are consistent with the conclusions of Conticello et al. (2018) that IVT can serve as a strong, synoptic-scale predictor of precipitation extremes, and suggest this is possible in both seasons. Our results are also consistent with the conclusion of Murawski et al. (2016) that near-surface

atmospheric humidity provides a better predictor of local precipitation than lower-level pressure states.

Importantly, the predictive performance of the NN is not uniform across archetypes. Better out-ofsample predictions are observed for certain archetypes, such as P-W3, P-S2, and P-S3 and, to a lesser extent, P-W2, given 300-hPa geopotential heights, 700-hPa specific humidity, and IVT. For the remaining archetypes, minimization of MSE occurs at the expense of increased prediction bias, particularly for T > 0.5, which serves as a threshold for storm identification (see Fig. 4). The NN struggles to predict extremes of the remaining archetypes, possibly for two reasons. First, winter archetypes P-W4 and P-W5 do not lie on the predominant Pacific or Atlantic Northern Hemisphere winter storm tracks (Lukens et al. 2018), and therefore these modes of precipitation are possibly not as influenced by regional atmospheric flow as are P-W2 and P-W3. Similarly, summer precipitation is generally less influenced by regional atmospheric circulation than occurs in winter. These differences highlight the fact that synoptic-scale information has different utility for characterizing the occurrence of extreme events depending on the season and region. Second, even in the 70-yr record used here, there are only a limited number of extreme events for each region, and the available data is less than the many thousands of samples that NNs typically require for model training. It is possible that a larger dataset could uncover stronger relationships between archetypes P-W4, P-W5, P-S4, and PW-5 and the synoptic-scale climate products.

2) NN PREDICTIONS OF SUMMER DISCHARGE ARCHETYPES

Figure 7 shows NN predictions of nonnull warm season discharge archetypes. Overall, no synoptic-scale field provided strong predictions of land surface discharge patterns across any of the archetypes under cross validation. These results could suggest that the processes involved in large runoff events are too complex to be predicted with synoptic-scale climate information alone, and additional information (e.g., antecedent conditions) needs to be considered to understand how dynamical climate change could impact discharge extremes. Alternatively, the issue could be one of data limitations. In contrast with winter and summer precipitation, the discharge dataset was limited to 1062 records (when the null archetype was < 0.5) as a result of drier summer soils allowing for larger infiltration rates, tempering runoff extremes (Ivancic and Shaw 2015; Knighton et al. 2017b). In this way, the land surface acts as a high-pass filter on precipitation,



FIG. 5. Deep convolutional network prediction skill for the validation set of cold season precipitation archetypes. MSE scores are presented for the validation set, where an asterisk (*) indicates the best performing climate state variable per archetype. The 1:1 line is presented in gray. The scatterplot presents only validation points from the most recent 20% of samples.

allowing only higher discharge events to contribute to the expression of nonnull archetypes. While this filtering effect could simplify the problem in that only days of high archetype expression are provided for NN training, it is possible that this case is too data limited to properly evaluate teleconnections given the overly parameterized nature of NNs. If we remove the null archetype from the prediction set (i.e., reduce the dimensionality of the problem), there is a substantial improvement in NN prediction skill for summer discharge archetypes (see Fig. S4). This occurs even though the model formulation is biased, since Softmax activation requires predictions to sum to unity, which will not occur in the observed archetype expressions if the null archetype is removed. This analysis demonstrates that direct downscaling from synoptic-scale climate features to regional discharge may be possible, but likely will rely on a



FIG. 6. Deep convolutional network prediction skill for the validation set of warm season precipitation archetypes. MSE scores are presented for the validation set, where an asterisk (*) indicates the best performing climate state variable per archetype. The 1:1 line is presented in gray. The scatterplot presents only validation points from the most recent 20% of samples.

modified network architecture and a larger training dataset.

3) INTERPRETATION OF CONVOLUTIONAL NNS: ARCHETYPE P-W2

Here we review characteristic events and NN predictions for P-W2, and present a similar analysis for P-W3 in the supplementary material (Fig. S5). Figure 8 shows four characteristic events where archetype P-W2 was very active (T > 0.95). All four events show similar spatial patterns of intense precipitation across the Great Plains and Midwest regions of the eastern United States. The 8 December 1978, 2 March 1997, and 19 March 2008 precipitation events (and subsequent flooding) occurred via nearly identical atmospheric mechanisms. In each case, rainfall was induced by a stationary front stalled in the Northern Kentucky region, which steadily drew moisture from the Gulf of Mexico poleward, resulting



FIG. 7. Deep convolutional network prediction skill for the validation set of summer discharge archetypes. MSE scores are presented for the validation set, where * indicates the best performing climate state variable per archetype. The 1:1 line is presented in gray. The scatterplot presents only validation points from the most recent 20% of samples.

in a line of intense thunderstorms that persisted for greater than 24 h (NOAA 2008; Sullavan et al. 1979; Jackson and Vivian 1997).

Across each event, the geopotential height field exhibits a trough over the western United States or eastern Pacific and a high pressure ridge along the U.S. Atlantic coast extending from Florida to Maine, consistent with the presence of meridional moisture flux from the Gulf of Mexico. Strong and Liptak (2012) performed EOF analysis of upper-level wind speeds to uncover correlations with Midwest winter precipitation. They determined that an extratropical cyclone structure of the wind field that drew and propagated moisture from the Gulf through the East Coast was correlated with days of winter precipitation. They also showed that, conversely, an equatorward movement of the jet stream tended to suppress winter precipitation. Nakamura et al. (2013) reached similar conclusions through a diagnostic



FIG. 8. Four characteristics events of P-W2: total daily rainfall, 300-hPa geopotential heights (black contours) and anomalies (color), IVT, and trained first convolutional layer activations.

review of 20 extreme flooding events within the Ohio basin. They emphasized the importance of the stationary high pressure ridge along the eastern U.S. coast that guides moisture poleward toward the Ohio basin, as opposed to the seasonal mean westerly moisture transport. These findings are in line with the observed IVT for each event, showing a southwest–northeast flux of moisture across the Great Plains region.

The first convolution layer activations from the NN trained on IVT are evaluated to understand how the NN screens out noninformative climate features. Similar results for 300-hPa geopotential heights and P-W2 and P-W3 are presented in the supplemental material (Figs. S6 and S7, respectively). Across all four events, the first convolutional 2D layer activations on IVT show strong spatial consistency across the Great Plains and eastern United States, mirroring the characteristic shape of winter storm tracks (Lukens et al. 2018). The trained layer assigns low values to most IVT fluxes that substantially deviate from the southwest to northeast elongated shape of moisture fluxes common to P-W2 events. Notably though, the first convolutional layer activations preserve moisture tracks over the Pacific during 23 February 1975 and 2 March 1997 events that have similar shapes to moisture fluxes resulting in Great Plains precipitation events. This is because the NN architecture used here seeks to identify characteristic shapes in the input data that are related to the archetypes. Network weighting of similar spatial patterns in

different locations of the input layer could result in false archetype prediction, although the complexity of the method makes it difficult to determine if this issue is ameliorated in deeper layers in the network.

5. Discussion

a. General circulation model precipitation downscaling

Kennel et al. (2016) suggest that we currently lack appropriate demonstrations of how a physical understanding of climate dynamics can inform practices to reduce systemic risk and aid decision making. In the context of long-term planning, this gap is sustained in part because of the difficulty in understanding the relationships between the general circulation and regional to local precipitation and hydrologic extremes, and utilizing those relationships in the design of engineered projects with long planning horizons. Statistical downscaling techniques, which often attempt to both bias-correct GCM estimates and to relate grid-scale physically based processes (e.g., atmospheric moisture transport) to parameterized subgrid processes (e.g., local precipitation), carry issues of skill (e.g., Castellano and DeGaetano 2017) and climate transferability (e.g., Dixon et al. 2016). This uncertainty often leads to low reliability in predictions of extreme precipitation (e.g., Stephens et al. 2010; Bock et al. 2018) and land surface responses (Hirabayashi et al. 2013).

However, GCMs provide relatively reliable estimates of synoptic-scale climate variables (surface air temperatures, geopotential height fields, IVT; e.g., Johnson and Sharma 2009; Perez et al. 2014) and the structure of some synoptic-scale weather systems (Zappa et al. 2013), though potentially with some regional heterogeneity (e.g., Sachindra et al. 2013). Murawski et al. (2016, 2018) and Farnham et al. (2018) demonstrate the viability of relating synoptic-scale climate patterns to local weather as an alternative methodology for downscaling. These approaches avoid some of the issues of climate transferability (e.g., Dixon et al. 2016) in the explicit consideration of future changes to the frequency of synoptic-scale climate states. Our results support this general approach, and suggest that 300-hPa geopotential heights, 700-hPa specific humidity, and IVT all show some utility for predicting extreme precipitation patterns across the Eastern CONUS, at least in some regions and seasons.

The use of emerging machine learning techniques like archetypal analysis and a 2D convolutional NN represents a novel approach for identifying spatiotemporal clusters and predictability of precipitation and discharge extremes. This methodology offers the possibility of several improvements over previous methods. First, previous methods have predominantly utilized hard clustering of synoptic-scale weather types to uncover relationships to regional precipitation (Steinschneider and Lall 2015, 2016; Murawski et al. 2016; Roller et al. 2016; Zhao et al. 2017; Agel et al. 2018; Conticello et al. 2018; Farnham et al. 2017). We demonstrate that AA circumvents the need to classify each day as exclusively one weather type. Previous studies have also focused on teleconnections to well-understood modes of global climate variability (e.g., ENSO, NAO, PDO; e.g., Lee et al. 2018), but such approaches are perhaps too limiting, as demonstrated by Agel et al. (2019) and Conticello et al. (2018). NNs can avoid the need to subjectively define climate modes a priori, or overly constrain the bounding window of climate information considered.

b. Limitations and opportunities for further research

Though our research approach is novel and introduces new opportunities for downscaling GCM projections to regional extremes, we adopted several assumptions that require additional discussion.

First, we only use daily data, which is a common approach but neglects processes occurring over shorter time scales relevant to the evolution of extreme events. Flooding within the United States predominantly affects smaller catchments (e.g., Marjerison et al. 2016) that are sensitive to rainfall characteristics occurring at subdaily scales (Knighton and Walter 2016). The atmospheric

patterns that we relate to regional precipitation and streamflow in this work may have less effect on subdaily extremes, which are driven primarily by localized weather events (Barbero et al. 2017). It is worth expanding the scope of this analysis to consider 6-hourly to hourly data to determine if climate-weather relationships can be found at time scales that are most relevant to flooding hazards. This would also increase the available data for model training.

In addition, the relationship between archetypes and extreme events was verified on two storm event catalogs, but these databases are potentially an incomplete record of past events, and so a more robust approach for benchmarking AA should be conducted. Further, increasing the number of archetypes could more accurately represent localized patterns of discharge and precipitation extremes, and potentially more meaningful climate teleconnections.

We applied a relatively simplistic hydrologic model that partitioned rainfall into runoff and infiltration via an imposed representation of VSA hydrology across the eastern United States. While this is not a poor assumption for this region (Buchanan et al. 2018), a more physically based land surface model could potentially improve the representation of high-intensity precipitation-runoff events.

Finally, the strength of the relationships between spatial patterns of precipitation and discharge across the Eastern CONUS and daily large-scale climate features was at times limited. This indicates that expectations should be tempered when relating extremes to synoptic-scale fields, at least for some regions and seasons. However, some of this error may be attributable to limitations of reanalysis products in capturing regional geopotential heights and near-surface humidity that co-occur with precipitation extremes (Sachindra et al. 2013). Biases in the hydrologic budgets of NCEP-NCAR Reanalysis 1, particularly in the southeast CONUS could lead to limitations in reproduction of precipitation archetypes (Mo and Higgins 1996; Trenberth and Guillemot 1998). More generally, data availability and network architecture could also limit the usefulness of the NN approach. Methods to extend the historical record, like the use of longer but sparser gauging networks, could be paired with longer reanalysis datasets (NOAA-CIRES Twentieth Century Reanalysis; Compo et al. 2011) to develop more robust training datasets. More complex network architectures should also be explored that can better distinguish between shapes in the synoptic-scale fields that are most relevant to extremes in a particular region of interest. Future research could also expand the temporal window of NN inputs to consider a combination of current day and past weather conditions in predicting precipitation, or multiple combinations of inputs (e.g., 300-hPa geopotential heights and 700-hPa specific humidity). Both of these changes would increase the dimensionality of the problem and would possibly require alternate computational strategies to avoid model overfitting.

6. Conclusions

This study contributes a novel approach to link the occurrence of regional precipitation and discharge extremes to synoptic-scale climate fields that has the potential to improve downscaling of climate model information. We developed seasonal archetypes for daily precipitation and discharge across the eastern United States, which were verified against catalogs of historical storms and then predicted using a 2D convolutional neural network and five daily climate reanalysis products (300-hPa geopotential height, 700-hPa geopotential height, 300-hPa specific humidity, 700-hPa specific humidity, and IVT). The first convolutional layer activations of the NN were also examined to understand how NNs identify teleconnections between synoptic-scale climate features and regional extreme events. Results suggested that IVT (and, to a lesser extent, 300- and 700-hPa geopotential height) yielded the most reliable predictions of regional archetypes. In particular, cold season precipitation across the Great Plains and Midwest region were reliably predicted by preserving information on moisture fluxes across the eastern United States. These results support the use of IVT as a key synoptic-scale variable that should be considered in future downscaling applications. However, prediction skill varied significantly across archetypes of precipitation and discharge for different seasons and regions. Therefore, there may be a limit to the utility of synoptic-scale information for downscaling applications targeted to extreme events, depending on the hydrologic variable of interest and the predominant circulation regimes for the region and time of year under study. This study demonstrates that emergent machine learning techniques like AA and 2D convolutional NNs can help assess whether such a downscaling strategy is viable for regional extremes, and the degree of predictability that can be expected given synoptic-scale climate information.

Acknowledgments. This work was supported by Cornell University and The Nature Conservancy. The Nature Conservancy gratefully acknowledges the William Penn Foundation's generous support of the development of this work (Grant 141-16). The opinions expressed in this report are those of the author(s) and do not necessarily reflect the views of the William Penn Foundation.

REFERENCES

- Agel, L., M. Barlow, J. H. Qian, F. Colby, E. Douglas, and T. Eichler, 2015: Climatology of daily precipitation and extreme precipitation events in the northeast United States. *J. Hydrometeor.*, **16**, 2537–2557, https://doi.org/10.1175/ JHM-D-14-0147.1.
- —, —, S. B. Feldstein, and W. J. Gutowski, 2018: Identification of large-scale meteorological patterns associated with extreme precipitation in the US northeast. *Climate Dyn.*, **50**, 1819–1839, https://doi.org/10.1007/s00382-017-3724-8.
- —, —, F. Colby, H. Binder, J. L. Catto, A. Hoell, and J. Cohen, 2019: Dynamical analysis of extreme precipitation in the US northeast based on large-scale meteorological patterns. *Climate Dyn.*, **52**, 1739–1760, https://doi.org/ 10.1007/s00382-018-4223-2.
- Archibald, J. A., B. P. Buchanan, D. R. Fuka, C. B. Georgakakos, S. W. Lyon, and M. T. Walter, 2014: A simple, regionally parameterized model for predicting nonpoint source areas in the northeastern US. J. Hydrol. Reg. Stud., 1, 74–91, https:// doi.org/10.1016/j.ejrh.2014.06.003.
- Barbero, R., H. J. Fowler, G. Lenderink, and S. Blenkinsop, 2017: Is the intensification of precipitation extremes with global warming better detected at hourly than daily resolutions? *Geophys. Res. Lett.*, 44, 974–983, https://doi.org/10.1002/ 2016GL071917.
- Bock, A. R., L. E. Hay, G. J. McCabe, S. L. Markstrom, and R. D. Atkinson, 2018: Do downscaled general circulation models reliably simulate historical climatic conditions? *Earth Interact.*, 22, https://doi.org/10.1175/EI-D-17-0018.1.
- Buchanan, B., and Coauthors, 2018: Estimating dominant runoff modes across the conterminous United States. *Hydrol. Processes*, **32**, 3881–3890, https://doi.org/10.1002/hyp.13296.
- Castellano, C. M., and A. T. DeGaetano, 2017: Downscaling extreme precipitation from CMIP5 simulations using historical analogs. J. Appl. Meteor. Climatol., 56, 2421–2439, https:// doi.org/10.1175/JAMC-D-16-0250.1.
- Chen, M., W. Shi, P. Xie, V. B. Silva, V. E. Kousky, R. W. Higgins, and J. E. Janowiak, 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. J. Geophys. Res., 113, D04110, https://doi.org/10.1029/2007JD009132.
- Chollet, F., 2015: Keras. https://keras.io.
- Compo, G. P., and Coauthors, 2011: The Twentieth Century Reanalysis Project. *Quart. J. Roy. Meteor. Soc.*, **137**, 1–28, https:// doi.org/10.1002/qj.776.
- Conticello, F., F. Cioffi, B. Merz, and U. Lall, 2018: An event synchronization method to link heavy rainfall events and large-scale atmospheric circulation features. *Int. J. Climatol.*, 38, 1421–1437, https://doi.org/10.1002/joc.5255.
- Cutler, A., and L. Breiman, 1994: Archetypal analysis. *Technometrics*, **36**, 338–347, https://doi.org/10.1080/00401706.1994.10485840.
- Dixon, K. W., J. R. Lanzante, M. J. Nath, K. Hayhoe, A. Stoner, A. Radhakrishnan, V. Balaji, and C. F. Gaitán, 2016: Evaluating the stationarity assumption in statistically downscaled climate projections: Is past performance an indicator of future results? *Climatic Change*, **135**, 395–408, https://doi.org/ 10.1007/s10584-016-1598-0.
- Duin, R. P., and D. M. J. Tax, 2005: Statistical pattern recognition. Handbook of Pattern Recognition and Computer Vision,

C. H. Chen, L. F. Pau, and P. S. P. Wang, Eds., World Scientific, 3–24, https://doi.org/10.1142/1802.

- Easton, Z. M., D. R. Fuka, M. T. Walter, D. M. Cowan, E. M. Schneiderman, and T. S. Steenhuis, 2008: Re-conceptualizing the soil and water assessment tool (SWAT) model to predict runoff from variable source areas. *J. Hydrol.*, **348**, 279–291, https://doi.org/10.1016/j.jhydrol.2007.10.008.
- Farnham, D. J., S. Steinschneider, and U. Lall, 2017: Zonal wind indices to reconstruct CONUS winter precipitation. *Geophys. Res. Lett.*, 44, 12 236–12 243, https://doi.org/10.1002/ 2017GL075959.
- —, J. Doss-Gollin, and U. Lall, 2018: Regional extreme precipitation events: robust inference from credibly simulated GCM variables. *Water Resour. Res.*, **54**, 3809–3824, https:// doi.org/10.1002/2017WR021318.
- Goodfellow, I., Y. Bengio, A. Courville, and Y. Bengio, 2016: Deep Learning. MIT Press, 775 pp.
- Hannachi, A., and N. Trendafilov, 2017: Archetypal analysis: Mining weather and climate extremes. J. Climate, 30, 6927– 6944, https://doi.org/10.1175/JCLI-D-16-0798.1.
- Hirabayashi, Y., R. Mahendran, S. Koirala, L. Konoshima, D. Yamazaki, S. Watanabe, H. Kim, and S. Kanae, 2013: Global flood risk under climate change. *Nat. Climate Change*, 3, 816–821, https://doi.org/10.1038/nclimate1911.
- Huth, R., C. Beck, A. Philipp, M. Demuzere, Z. Ustrnul, M. Cahynová, J. Kyselý, and O. E. Tveito 2008: Classifications of atmospheric circulation patterns. *Ann. N. Y. Acad. Sci.*, **1146**, 105–152, https://doi.org/10.1196/annals.1446.019.
- Ivancic, T. J., and S. B. Shaw, 2015: Examining why trends in very heavy precipitation should not be mistaken for trends in very high river discharge. *Climatic Change*, **133**, 681–693, https:// doi.org/10.1007/s10584-015-1476-1.
- Jackson, K. S. and S. A. Vivian, 1997: Flood of March 1997 in Southern Ohio. USGS Water-Resources Investigations Rep. 97–4149, 21 pp., https://pubs.usgs.gov/wri/1997/4149/ report.pdf.
- Janssen, E., R. L. Sriver, D. J. Wuebbles, and K. E. Kunkel, 2016: Seasonal and regional variations in extreme precipitation event frequency using CMIP5. *Geophys. Res. Lett.*, 43, 5385– 5393, https://doi.org/10.1002/2016GL069151.
- Johnson, F., and A. Sharma, 2009: Measurement of GCM skill in predicting variables relevant for hydroclimatological assessments. J. Climate, 22, 4373–4382, https://doi.org/10.1175/ 2009JCLI2681.1.
- Kennel, C. F., S. Briggs, and D. G. Victor, 2016: Making climate science more relevant. *Science*, **354**, 421–422, https://doi.org/ 10.1126/science.aag3248.
- Knighton, J. O., and M. T. Walter, 2016: Critical rainfall statistics for predicting watershed flood responses: Rethinking the design storm concept. *Hydrol. Processes*, **30**, 3788–3803, https:// doi.org/10.1002/hyp.10888.
- Knighton, J., S. Steinschneider, and M. T. Walter, 2017a: A vulnerability-based, bottom-up assessment of future riverine flood risk using a modified peaks-over-threshold approach and a physically based hydrologic model. *Water Resour. Res.*, 53, 10043–10064, https://doi.org/10.1002/2017WR021036.
 - —, A. DeGaetano, and M. T. Walter, 2017b: Hydrologic state influence on riverine flood discharge for a small temperate watershed (Fall Creek, United States): Negative feedbacks on the effects of climate change. J. Hydrometeor., 18, 431–449, https://doi.org/10.1175/JHM-D-16-0164.1.
- Krizhevsky, A., I. Sutskever, and G. E. Hinton, 2012: Imagenet classification with deep convolutional neural networks.

Advances in Neural Information Processing Systems 25, F. Pereira et al., Eds., Neural Information Processing Systems Foundation, 1097–1105.

- Lee, D., P. Ward, and P. Block, 2018: Attribution of large-scale climate patterns to seasonal peak-flow and prospects for prediction globally. *Water Resour. Res.*, 54, 916–938, https:// doi.org/10.1002/2017WR021205.
- Lukens, K. E., E. H. Berbery, and K. I. Hodges, 2018: The imprint of strong-storm tracks on winter weather in North America. J. Climate, 31, 2057–2074, https://doi.org/10.1175/ JCLI-D-17-0420.1.
- Marjerison, R. D., M. T. Walter, P. J. Sullivan, and S. J. Colucci, 2016: Does population affect the location of flash flood reports? J. Appl. Meteor. Climatol., 55, 1953–1963, https:// doi.org/10.1175/JAMC-D-15-0329.1.
- Merz, B., and Coauthors, 2014: Floods and climate: Emerging perspectives for flood risk assessment and management. *Nat. Hazards Earth Syst. Sci.*, 14, 1921, https://doi.org/10.5194/ nhess-14-1921-2014.
- Mo, K. C., and R. W. Higgins, 1996: Large-scale atmospheric moisture transport as evaluated in the NCEP/NCAR and the NASA/ DAO reanalyses. J. Climate, 9, 1531–1545, https://doi.org/ 10.1175/1520-0442(1996)009<1531:LSAMTA>2.0.CO;2.
- Montavon, G., W. Samek, and K. R. Müller, 2018: Methods for interpreting and understanding deep neural networks. *Digital Signal Process.*, 73, 1–15, https://doi.org/10.1016/ j.dsp.2017.10.011.
- Murawski, A., G. Bürger, S. Vorogushyn, and B. Merz, 2016: Can local climate variability be explained by weather patterns? A multi-station evaluation for the Rhine basin. *Hydrol. Earth Syst. Sci.*, **20**, 4283, https://doi.org/10.5194/ hess-20-4283-2016.
- —, S. Vorogushyn, G. Bürger, L. Gerlitz, and B. Merz, 2018: Do changing weather types explain observed climatic trends in the Rhine Basin? An analysis of within-and between-type changes. J. Geophys. Res. Atmos., 123, 1562–1584, https:// doi.org/10.1002/2017JD026654.
- Nakamura, J., U. Lall, Y. Kushnir, A. W. Robertson, and R. Seager, 2013: Dynamical structure of extreme floods in the U.S. Midwest and the United Kingdom. J. Hydrometeor., 14, 485–504, https://doi.org/10.1175/JHM-D-12-059.1.
- Nakamura, N., and C. S. Y. Huang, 2018: Atmospheric blocking as a traffic jam in the jet stream. *Science*, **361**, 42–47, https:// doi.org/10.1126/science.aat0721.
- NCDC, 2018: Billion dollar weather and climate disasters. NOAA/ NCDC, https://www.ncdc.noaa.gov/billions/.
- NCEI, 2018: Storm events database. NOAA/NCEI, https:// www.ncdc.noaa.gov/stormevents/.
- NOAA, 2008: Flood event March 2008. NOAA/NWS, 9 pp., https://www.weather.gov/media/lsx/Events/03_18_2008.pdf.
- —, 2018: CPC Unified gauge-based analysis of daily precipitation over CONUS. NOAA/ESRL/PSD, https://www.esrl.noaa.gov/ psd/data/gridded/data.unified.daily.conus.html.
- Perez, J., M. Menendez, F. J. Mendez, and I. J. Losada, 2014: Evaluating the performance of CMIP3 and CMIP5 global climate models over the north-east Atlantic region. *Climate Dyn.*, 43, 2663–2680, https://doi.org/10.1007/s00382-014-2078-8.
- Rawat, W., and Z. Wang, 2017: Deep convolutional neural networks for image classification: A comprehensive review. *Neural Comput.*, **29**, 2352–2449, https://doi.org/10.1162/ neco_a_00990.
- Roller, C. D., J. H. Qian, L. Agel, M. Barlow, and V. Moron, 2016: Winter weather regimes in the northeast United States.

J. Climate, **29**, 2963–2980, https://doi.org/10.1175/JCLI-D-15-0274.1.

- Russell, S. J., and P. Norvig, 2016: *Artificial Intelligence: A Modern Approach*. Pearson Education Limited, 1152 pp.
- Sachindra, D. A., F. Huang, A. Barton, and B. J. C. Perera, 2013: Least square support vector and multi-linear regression for statistically downscaling general circulation model outputs to catchment streamflows. *Int. J. Climatol.*, 33, 1087–1106, https:// doi.org/10.1002/joc.3493.
- Schneiderman, E. M., T. S. Steenhuis, D. J. Thongs, Z. M. Easton, M. S. Zion, A. L. Neal, G. F. Mendoza, and M. T. Walter, 2007: Incorporating variable source area hydrology into a curvenumber-based watershed model. *Hydrol. Processes*, 21, 3420– 3430, https://doi.org/10.1002/hyp.6556.
- Shaw, T. A., and Coauthors, 2016: Storm track processes and the opposing influences of climate change. *Nat. Geosci.*, 9, 656– 664, https://doi.org/10.1038/ngeo2783.
- Steinschneider, S., and U. Lall, 2015: Daily precipitation and tropical moisture exports across the eastern United States: An application of archetypal analysis to identify spatiotemporal structure. J. Climate, 28, 8585–8602, https://doi.org/10.1175/ JCLI-D-15-0340.1.
- —, and —, 2016: Spatiotemporal structure of precipitation related to tropical moisture exports over the eastern United States and its relation to climate teleconnections. *J. Hydrometeor.*, **17**, 897–913, https://doi.org/10.1175/JHM-D-15-0120.
- Stephens, G. L., and Coauthors, 2010: Dreary state of precipitation in global models. J. Geophys. Res., 115, D24211, https:// doi.org/10.1029/2010JD014532.
- Strobach, E., and G. Bel, 2017: The contribution of internal and model variabilities to the uncertainty in CMIP5 decadal climate predictions. *Climate Dyn.*, **49**, 3221–3235, https://doi.org/ 10.1007/s00382-016-3507-7.
- Strong, C., and J. Liptak, 2012: Propagating atmospheric patterns associated with Midwest winter precipitation. J. Hydrometeor., 13, 1371–1382, https://doi.org/10.1175/JHM-D-11-0111.1.
- Sullavan, J. N., F. Quinones, and R. F. Flint, 1979: Floods of December 1978 in Kentucky. USGS Open-File Rep. 79–977, 53 pp., https://pubs.usgs.gov/of/1979/ofr79977/pdf/ofr_79-977.pdf.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, 93, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1.
- Trenberth, K. E., 2011: Changes in precipitation with climate change. *Climate Res.*, 47, 123–138, https://doi.org/10.3354/ cr00953.
 - —, and C. J. Guillemot, 1998: Evaluation of the atmospheric moisture and hydrological cycle in the NCEP/NCAR

reanalyses. *Climate Dyn.*, **14**, 213–231, https://doi.org/10.1007/s003820050219.

- Walter, M. T., E. S. Brooks, D. K. McCool, L. G. King, M. Molnau, and J. Boll, 2005: Process-based snowmelt modeling: Does it require more input data than temperatureindex modeling? J. Hydrol., 300, 65–75, https://doi.org/ 10.1016/j.jhydrol.2004.05.002.
- Wójcik, R., 2015: Reliability of CMIP5 GCM simulations in reproducing atmospheric circulation over Europe and the North Atlantic: A statistical downscaling perspective. *Int. J. Climatol.*, 35, 714–732, https://doi.org/10.1002/joc.4015.
- Woollings, T., A. Hannachi, and B. Hoskins, 2010: Variability of the North Atlantic eddy-driven jet stream. *Quart. J. Roy. Meteor. Soc.*, **136**, 856–868, https://doi.org/10.1002/qj.625.
- Wuebbles, D., and Coauthors, 2014: CMIP5 climate model analyses: Climate extremes in the United States. Bull. Amer. Meteor. Soc., 95, 571–583, https://doi.org/10.1175/ BAMS-D-12-00172.1.
- Xie, P., M. Chen, S. Yang, A. Yatagai, T. Hayasaka, Y. Fukushima, and C. Liu, 2007: A gauge-based analysis of daily precipitation over East Asia. J. Hydrometeor., 8, 607–626, https://doi.org/ 10.1175/JHM583.1.
- Xu, C. Y., 1999: From GCMs to river flow: A review of downscaling methods and hydrologic modelling approaches. *Prog. Phys. Geogr.*, 23, 229–249, https://doi.org/10.1177/ 030913339902300204.
- Yosinski, J., J. Clune, A. Nguyen, T. Fuchs, and H. Lipson, 2015: Understanding neural networks through deep visualization. ArXiv, https://arxiv.org/abs/1506.06579.
- Zappa, G., L. C. Shaffrey, K. I. Hodges, P. G. Sansom, and D. B. Stephenson, 2013: A multimodel assessment of future projections of North Atlantic and European extratropical cyclones in the CMIP5 climate models. J. Climate, 26, 5846–5862, https://doi.org/10.1175/JCLI-D-12-00573.1.
- Zeiler, M. D., and R. Fergus, 2014: Visualizing and understanding convolutional networks. *Computer Vision – ECCV 2014*, D. Fleet et al., Eds., Lecture Notes in Computer Science, Vol. 8689, Springer, 818–833, https://doi.org/10.1007/978-3-319-10590-1_53.
- Zhao, S., Y. Deng, and R. X. Black, 2017: A dynamical and statistical characterization of U.S. extreme precipitation events and their associated large-scale meteorological patterns. *J. Climate*, **30**, 1307–1326, https://doi.org/10.1175/JCLI-D-15-0910.1.
- Zhu, Y., and R. E. Newell, 1998: A proposed algorithm for moisture fluxes from atmospheric rivers. *Mon. Wea. Rev.*, **126**, 725–735, https://doi.org/10.1175/1520-0493(1998)126<0725: APAFMF>2.0.CO;2.