



# Modeling the contributions of global air temperature, synoptic-scale phenomena and soil moisture to near-surface static energy variability using artificial neural networks

Sara C. Pryor<sup>1</sup>, Ryan C. Sullivan<sup>1</sup>, and Justin T. Schoof<sup>2</sup>

<sup>1</sup>Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, NY 14853, USA

<sup>2</sup>Department of Geography and Environmental Resource, Southern Illinois University, Carbondale, IL 62901, USA

Correspondence to: Sara C. Pryor (sp2279@cornell.edu)

Received: 22 April 2017 – Discussion started: 13 July 2017

Revised: 1 November 2017 – Accepted: 4 November 2017 – Published: 6 December 2017

**Abstract.** The static energy content of the atmosphere is increasing on a global scale, but exhibits important subglobal and subregional scales of variability and is a useful parameter for integrating the net effect of changes in the partitioning of energy at the surface and for improving understanding of the causes of so-called “warming holes” (i.e., locations with decreasing daily maximum air temperatures ( $T$ ) or increasing trends of lower magnitude than the global mean). Further, measures of the static energy content (herein the equivalent potential temperature,  $\theta_e$ ) are more strongly linked to excess human mortality and morbidity than air temperature alone, and have great relevance in understanding causes of past heat-related excess mortality and making projections of possible future events that are likely to be associated with negative human health and economic consequences. New non-linear statistical models for summertime daily maximum and minimum  $\theta_e$  are developed and used to advance understanding of drivers of historical change and variability over the eastern USA. The predictor variables are an index of the daily global mean temperature, daily indices of the synoptic-scale meteorology derived from  $T$  and specific humidity ( $Q$ ) at 850 and 500 hPa geopotential heights ( $Z$ ), and spatiotemporally averaged soil moisture (SM). SM is particularly important in determining the magnitude of  $\theta_e$  over regions that have previously been identified as exhibiting warming holes, confirming the key importance of SM in dictating the partitioning of net radiation into sensible and latent heat and dictating trends in near-surface  $T$  and  $\theta_e$ . Consistent with our a priori expectations, models built using artificial neural networks (ANNs) out-perform linear models that do not permit interaction of the predictor variables (global  $T$ , synoptic-scale mete-

orological conditions and SM). This is particularly marked in regions with high variability in minimum and maximum  $\theta_e$ , where more complex models built using ANN with multiple hidden layers are better able to capture the day-to-day variability in  $\theta_e$  and the occurrence of extreme maximum  $\theta_e$ . Over the entire domain, the ANN with three hidden layers exhibits high accuracy in predicting maximum  $\theta_e > 347$  K. The median hit rate for maximum  $\theta_e > 347$  K is  $> 0.60$ , while the median false alarm rate is  $\approx 0.08$ .

## 1 Motivation and objectives

Extreme heat is associated with significant societal and environmental impacts, and a number of prior studies have elaborated on the drivers of extreme air temperatures ( $T$ ) and made projections of extreme  $T$  and the associated human health impacts and socioeconomic consequences (Sanderson and Ford, 2017; de’Donato et al., 2015; O’Neill and Ebi, 2009; Garcia-Herrera et al., 2010). Previous studies have sought to quantify the predictability of extreme  $T$  as a function of lead-time and variables describing teleconnections to remote sea surface temperature anomalies (McKinnon et al., 2016) and/or local soil moisture (Brabson et al., 2005). Physiological stress is maximized under the co-occurrence of elevated  $T$  and specific humidity ( $q$ ) (Zhang et al., 2014). Thus to understand spatiotemporal variability in heat-related mortality and/or morbidity there is a need to consider integrative variables derived from both  $T$  and  $q$ , such as equivalent potential temperature ( $\theta_e$ ) computed herein using the following

approximation:

$$\theta_e = T \left( \frac{1000}{P} \right)^{R_d/C_{pd}} + \frac{L(T)}{C_{pd}} q, \quad (1)$$

where  $T$  is air temperature (K),  $P$  is atmospheric pressure (hPa),  $R_d$  is specific gas constant for air ( $\text{J kg}^{-1} \text{K}^{-1}$ ),  $C_{pd}$  is specific heat for dry air ( $\text{J kg}^{-1} \text{K}^{-1}$ ),  $L(T)$  is latent heat of vaporization ( $f(T)$ ) ( $\text{J kg}^{-1}$ ), and  $q$  is specific humidity of water vapor ( $\text{kg kg}^{-1}$ ) (Bolton, 1980).

Potential temperature is the temperature an unsaturated air parcel would have if brought adiabatically to a standard pressure. Thus, potential temperature is conserved for an unsaturated air parcel if it remains unsaturated as it rises and sinks. Equivalent potential temperature is conserved under vertical motion even if there is phase change of water vapor contained within the air. Hence, the use of metrics such as  $\theta_e$  permits more accurate depictions of near-surface energy budgets and surface heating trends for use in climate change detection studies (Davey et al., 2006). Equivalent potential temperature rather than equivalent temperature is used here to allow comparison of values from other reanalyses (or other model output) that use a different discretization of terrain elevation.

When considering the local surface energy balance (Eq. 2), near-surface  $T$  is responsive only to changes in the sensible heat flux from or to the surface, while  $\theta_e$  responds to changes in both the sensible and latent heat flux terms:

$$S(1 - \alpha) + R_{li} - R_{lo} = H + LE + G, \quad (2)$$

where  $S$  is incoming solar radiation,  $\alpha$  is albedo,  $R_{li}$  is counter-radiated longwave radiation from the atmosphere,  $R_{lo}$  is outgoing longwave radiation from the surface,  $H$  is sensible heat flux,  $LE$  is latent heat flux, and  $G$  is ground heat flux.

Accordingly, the total static energy of the near-surface atmosphere (and thus  $\theta_e$ ) has significantly increased since the 1970s due to the increase in net radiation (left-hand side of Eq. 2) (Peterson et al., 2011) and in response to increased  $T$  and the associated response in  $q$  (Ribera et al., 2004; Willett et al., 2007). However, climate modes (such as El Niño–Southern Oscillation, ENSO, the North Atlantic Oscillation, NAO, and Pacific–North American mode, PNA) cause inter-annual variability in global (Huang et al., 2017) and regional (Llamedo et al., 2017)  $T$  and  $q$  anomalies, and there is substantial regional and subregional variability in absolute magnitude of static energy and temporal trends therein (Fall et al., 2010; Pryor and Schoof, 2016; Pryor et al., 2016).

The following is a précis of the scales and processes we seek to explore and include in our new statistical downscaling models for summertime daily maximum and minimum equivalent potential temperature ( $\theta_e$ ).

- *Global-scale forcing due to enhanced greenhouse gas concentrations and internal climate variability (e.g., ENSO; Mann et al., 1998).* As  $T$  increases, the atmospheric water vapor content responds in proportion to

the saturation vapor pressure (Willett et al., 2007; Allen and Ingram, 2002). Thus, both components ( $T$  and  $q$ ) of static energy ( $\theta_e$ ) are enhanced in years and seasons with high global mean air temperatures. Previous research has indicated that variability in equivalent temperature ( $T_e$ , i.e., the temperature computed from Eq. (1) but excluding the correction for bringing the air parcel adiabatically to a reference pressure of 1000 hPa) in the North Atlantic is strongly linked to NAO (Ribera et al., 2004), and the probability of heat waves across the US is linked to hemispheric waves (Teng et al., 2013) and thus the PNA (Trenberth, 1990).

- *Synoptic-scale forcing (Grotjahn et al., 2016).* Much of the eastern USA broke records for heat indices during the summer of 2012 in part due to persistent anticyclonic conditions (Peterson et al., 2013), and many heat watch warnings employ a synoptic-typing-based methodology (Sheridan and Kalkstein, 2004). Further, variability and temporal trends in summertime moist static energy and dew point temperatures ( $T_d$ ) in the Midwestern USA appear to be linked to enhanced horizontal atmospheric moisture advection due in part to northward expansion of the southeast summertime anticyclone (Kalkstein et al., 1998; Rogers et al., 2007; Pryor and Schoof, 2016; Ford and Schoof, 2017).
- *Local–regional changes in water availability and energy partitioning at the surface.* These changes are due to factors such as land cover change and soil moisture modification as a result of irrigation of cropland (Davey et al., 2006; Pryor et al., 2016).

Observed tendencies in  $T$ ,  $q$  and  $\theta_e$  are naturally a product of a combination of these drivers (Horton et al., 2016). The expression of internal climate modes (e.g., ENSO, PNA and NAO) influences the frequency and intensity of different synoptic-scale phenomena (Sheridan, 2003; Weaver, 2013) and has been found to be partly responsible for an increase in the number of oppressively hot days in many urban areas across the USA over recent decades (Sheridan et al., 2009). Further, there are important feedbacks between the land–atmosphere coupling, the synoptic-scale circulation patterns and boundary-layer structure (Lee et al., 2016). Thus, for example, extremely high  $T_d$  (and by association elevated  $q$ ) in the Midwestern USA is associated with (a) development and propagation of low pressure from the High Plains to the upper Great Lakes, (b) healthy crops and sufficient surface soil moisture, and (c) restricted vertical mixing (Bentley and Stallins, 2008).

The objectives of this research are as follows:

1. to use nonlinear (machine learning) models applied to a three-tiered suite of predictors: (a) an index of the daily global mean temperature, (b) indices of the conditions on a synoptic scale based on principal components analysis (PCA) of upper-air variables, and (c) soil

moisture estimates, to investigate spatiotemporal variations in  $\theta_e$  and enhance understanding of the causes of variability and change in  $\theta_e$  over the eastern USA.

2. to examine whether the performance of the artificial neural networks (ANNs) used to predict daily maximum and minimum  $\theta_e$  is enhanced by inclusion of additional hidden layers in the ANNs.

## 2 Data and methods

### 2.1 Study region

The eastern USA (Fig. 1a) is the focus of this research because of the following reasons.

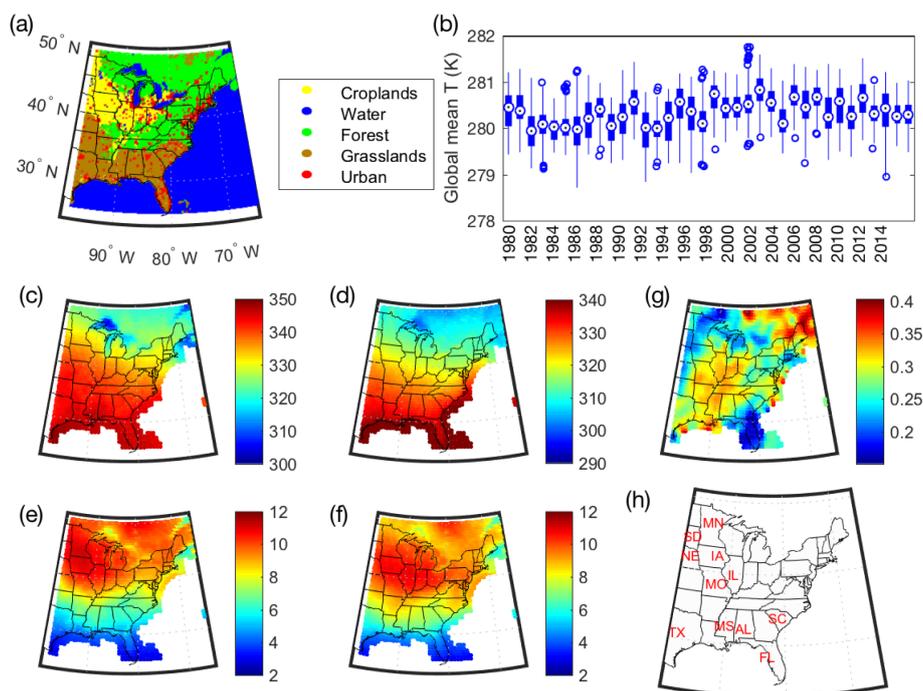
- It is home to over 200 million people (based on the 2013 census).
- It is characterized by high summertime  $T_e$  (and  $\theta_e$ ), and in situ data have indicated trends in  $T_e$  exceed those in  $T$  alone (Schoof et al., 2015). Further, the region is largely congruent with an area of coherence for extreme  $T$  events in prior research (McKinnon et al., 2016).
- It encompasses major urban areas that have experienced a number of past extreme heat events associated with substantial excess mortality and morbidity (see summary in Vanos et al., 2015 and also Anderson and Bell, 2011). Further, a range of reanalyses exhibit a consistent signal of increasing frequency of both dry (i.e., high  $T$  but low  $q$ ) and humid (i.e., combined high  $T$  and  $q$ ) heat wave days over the study region during the period 1981–2015 (Schoof et al., 2017).
- It exhibits strong spatial gradients in terms of the nature of land cover and rate of change of both land management and soil moisture (SM) (Fig. 1a and g; Pryor et al., 2016; Pryor and Schoof, 2016; Ellenburg et al., 2016). For example, over the period 1950–2000 the region as a whole experienced rapid population growth (though this was not spatially uniform), expansion of area classified as exurban, an overall reduction of land in agriculture (though again this was highly heterogeneous in space) and an increase in the intensity of water management (including expansion of irrigation) (Brown et al., 2005).
- Parts of it (focused on the southern Great Plains, i.e., the southern and westernmost portion of the current study area) were identified in the global land–atmosphere coupling experiments (GLACEs) as exhibiting atypically strong atmosphere–surface coupling in some global climate models. Further, soil moisture makes a large contribution to subseasonal forecast skill for air temperatures and precipitation in this region (Koster et al., 2006, 2011; Dirmeyer and Halder, 2017; Guo et al., 2006).

- It also incorporates two areas of reduced daily maximum  $T$  during multiple consecutive years extending over part or all of the 20th century (i.e., warming holes). One is located along the border of Iowa (IA), Nebraska (NE) and South Dakota (SD), and one is centered on Mississippi (MS) and Alabama (AL) (locations shown in Fig. 1h). The lack of warming in both regions has been attributed to changing land-surface characteristics and enhanced soil moisture availability (Kalnay and Cai, 2003; Pan et al., 2009; Ellenburg et al., 2016). In the case of MS and AL, up to 60% of the variance of summer temperatures has been ascribed to soil moisture (and thus an increase in the LE at the surface at the expense of  $H$ , Eq. 2) and cloud cover (reducing the net radiation) (Ellenburg et al., 2016). However, additional factors may account for these warming holes including large-scale climate modes such as the Interdecadal Pacific Oscillation and the associated sea surface temperature anomalies in the tropical Pacific (Meehl et al., 2015), as well as aerosol radiative forcing (Leibensperger et al., 2012; Yu et al., 2014).

### 2.2 Data set

All variables used herein derive from the MERRA-2 reanalysis data set (Molod et al., 2015; Reichle et al., 2017a). This minimizes uncertainty in models linking our predictors to near-surface static energy resulting from the use of different data sets to derive the predictor suite and/or the response variables (daily minimum and maximum  $\theta_e$ ). The MERRA-2 reanalysis product assimilates an unprecedented array of remote sensing and in situ data streams, but does not assimilate in situ observations of near-surface  $T$  or  $q$ . MERRA-2 output is available at a resolution of  $0.625^\circ \times \sim 0.5^\circ$  (longitude by latitude). We use MERRA-2 output for all summer days (June, July and August: JJA) during 1980–2015 to compute the following variables.

1. *A globally averaged daily mean air temperature computed using hourly 2 m  $T$  from all MERRA-2 grid cells.* This predictor (Fig. 1b) is intended to represent the long-term tendency in global mean temperatures and interannual or interdecadal variability caused by internal climate modes (Huang et al., 2017).
2. *Indices of synoptic-scale meteorology.* Air temperature ( $T_{850}$ ) and specific humidity ( $q_{850}$ ) at 850 hPa along with 500 hPa geopotential heights ( $Z_{500}$ ) in the domain ( $25.5\text{--}50^\circ$  N,  $97.5\text{--}65^\circ$  W) for 20:00 UTC (i.e., 16:00 Eastern Daylight Time, 15:00 Central Daylight Time) are converted to  $z$  scores (that indicates how many standard deviations an individual value is from the mean) and used in a rotated PCA to generate daily principal component scores (PCs) that represent the proximity of each day to the major modes of synoptic-scale variabil-

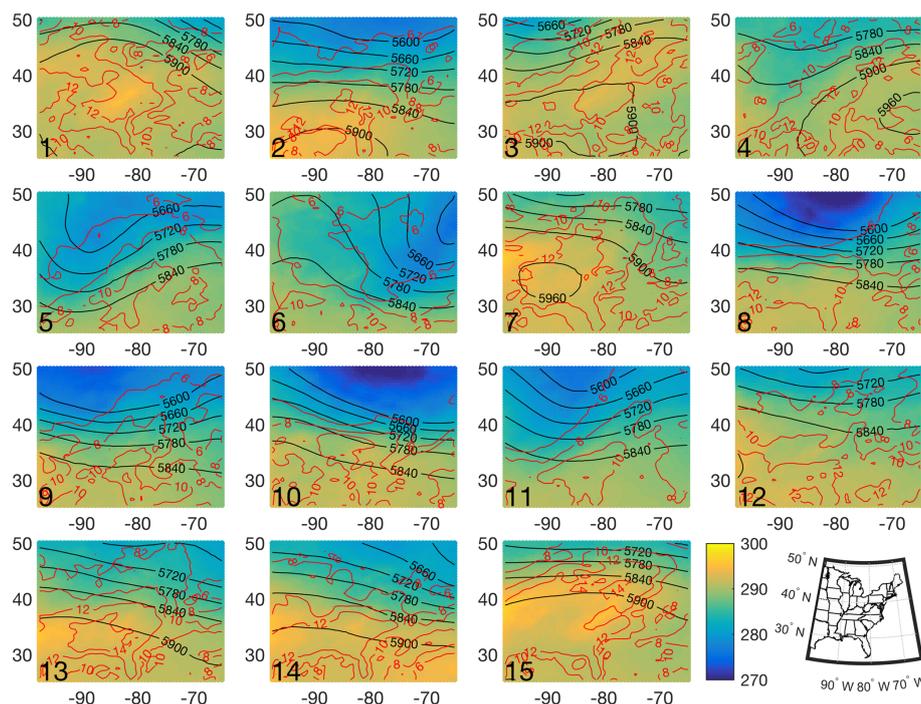


**Figure 1.** (a) The study domain used herein ( $25.5^{\circ}$ – $50^{\circ}$  N,  $97.5^{\circ}$ – $65^{\circ}$  W) and five consolidated land use and land cover (LULC) classes as represented at a resolution of  $0.05^{\circ} \times 0.05^{\circ}$  from the MODIS land cover data set for 2014 (MCD12C). (b) A box plot of daily global mean JJA temperature values (K) (1980–2015) as computed from the MERRA-2 output. (c) Mean summertime (JJA, 1980–2015) maximum  $\theta_e$  and (d) minimum  $\theta_e$  (K) as computed from hourly  $T$  and  $q$  at 2 m and surface pressure ( $P$ ) in each MERRA-2 grid cell. Standard deviation of daily (e) maximum  $\theta_e$  and (f) minimum  $\theta_e$  (K) (JJA, 1980–2015). (g) Mean soil moisture estimates (1980–2015) from MERRA-2 used as a predictor in the ANN and computed as follows: the daily mean value for total profile soil moisture content (PRMC) for each grid cell is averaged over a spatial area of  $3 \times 3$  grid cells centered on the grid cell of interest, and used to compute a 90-day running JJA mean soil moisture ( $\text{m}^3 \text{m}^{-3}$ ). Panel (h) shows the location of states that are explicitly referred to herein.

ity. These predictors (i.e., the PCs) are intended to represent variability in the synoptic-scale circulation (e.g., presence of anticyclonic conditions likely to be associated with subsidence and thus retarded vertical mixing) and also large-scale advection of static energy. A total of 15 components are retained based on a scree plot analysis (Cattell, 1966) and are subject to Varimax rotation (Richman, 1986). Daily PC scores for all 15 PCs are used as predictors in the statistical models, to allow each day to exhibit partial membership of multiple synoptic types. Spatial fields of the three variables sampled once per day are used in the PCA due to the high temporal autocorrelation present in these variables, and 20:00 UTC is selected to approximately coincide with the timing of the afternoon peak in surface  $T$  over the eastern USA. Figure 2 shows centroids of synoptic modes of variability as defined using the PCA (so-called key days) as represented by the spatial patterns of  $T_{850}$ ,  $q_{850}$  and  $Z_{500}$  computed as the mean conditions on the 7 days that exhibit highest PC scores on each PC. As shown, many of the synoptic types thus identified are readily interpretable as representing a diversity of zonal vs. meridional circulation (cf. type 9, 5 and 6), and

some are characterized by conditions known to be associated with strong southerly low-level advection of high  $T$  and  $q$  into the region (e.g., types 1, 3, and 15) (Pryor and Schoof, 2016; Weaver, 2013). Further, most types exhibit a high degree of similarity with other synoptic-scale classifications derived for the region (e.g., type 7 is very similar to one that is associated with summertime precipitation over the southeastern USA, Diem, 2006).

3. *An index of soil moisture.* This predictor is included to represent the availability of moisture at the atmosphere–surface boundary. Due to its spatial heterogeneity, following previous research (Ford and Schoof, 2016), we use a time-integrated and spatially integrated metric of SM. Specifically we use a 90-day running mean estimate of antecedent SM (i.e., the value for 1 June is an average of values from approx. 1 March to 1 June) over  $3 \times 3$  grid cells centered on the grid cell in question (i.e., integrated over an area of approximately  $30\,000 \text{ km}^2$ , see Fig. 1g). MERRA-2 SM has been evaluated relative to in situ measurements of surface and root-zone SM and exhibits an unbiased root mean square error (RMSE) of  $0.05 \text{ m}^3 \text{m}^{-3}$  and a variance explanation



**Figure 2.** Summary of the conditions under each of the PCA-derived synoptic modes of variability. The panels show the “mode centroid” (i.e., mean of the 7 days that exhibit highest PC scores for the type). The color depicts  $T_{850}$  (K), the solid black lines show  $Z_{500}$  (gpm: geopotential meters) and the red lines depict  $q_{850}$  ( $\text{g kg}^{-1}$ ). The PC number is shown in the lower left of each panel and the panels are arranged so that the first mode of variability is shown in the upper left panel, the fifth PC is shown in the second row in the left-most panel and so forth. To aid legibility, the panels in this figure denote the study region as a rectangle, with state outlines excluded, thus also shown in the lower right of this figure is a map depicting the state outlines for the same domain as shown in the other panels but using a plate carrée projection (i.e., the equirectangular projection) used in Figs. 1, 4–7.

( $R^2$ ) value of the average root-zone SM anomaly of 0.56 (Reichle et al., 2017a). Further, since the MERRA-2 system applies bias correction to the precipitation estimates used in the land surface model (Reichle et al., 2017b), this may result in enhanced accuracy of SM estimates. However, it should be noted that there are relatively few direct measurements of SM and thus the evaluation of MERRA-2 is focused on agricultural locations.

4. An estimate of (i) daily maximum  $\theta_e$  and (ii) daily minimum  $\theta_e$  in each grid cell (see Fig. 1c and d). The daily minimum and maximum values are used as the predictands in the downscaling and are derived using Eq. (1) applied to hourly  $T$  at 2 m,  $q$  at 2 m and surface pressure ( $P$ ). The domain used to compute the gridded fields of the predictands (daily minimum and maximum  $\theta_e$ ) is truncated by one grid cell on each boundary of the domain used to generate the predictors to accommodate the spatial averaging used to generate the SM predictor. The range of grid-cell  $\theta_e$  estimates are consistent with those derived from station observations within the study region (Pryor and Schoof, 2016), and in accordance with a priori expectations both daily maximum  $\theta_e$

and minimum  $\theta_e$  exhibit primarily latitudinal variability (Fig. 1c and d). Mean maximum  $\theta_e$  decreases from values of  $\sim 350$  K ( $\sim 80$  °C) in the south of the domain to approximately 320 K ( $\sim 50$  °C) in the north. MO and IA exhibit anomalously high mean maximum  $\theta_e$  compared with grid cells at the same latitudes (Fig. 1c) reflecting, in part, the advection of air with high  $T$  and  $q$  from the south by the Great Plains low level jet (GPLLJ) (Weaver, 2013; Pryor and Schoof, 2016; Schoof et al., 2015) and the presence of abundant SM (Fig. 1g) (Al Bitar et al., 2017).

There are also important spatial patterns of the day-to-day variability of daily maximum  $\theta_e$  and minimum  $\theta_e$  that provide key context for considering the performance of different transfer functions (i.e., statistical models linking indices of the global mean  $T$ , synoptic-scale meteorology and soil moisture (the predictors) to the response variables; maximum  $\theta_e$  and minimum  $\theta_e$ ). The variance (and standard deviation) of daily maximum  $\theta_e$  values (i.e., a measure of the dispersion of individual days around the mean values of minimum and maximum  $\theta_e$  shown in Fig. 1c and d) is largest over IA and southern Minnesota (MN) (Fig. 1e), while the variance of minimum  $\theta_e$  is greatest over Illinois (IL) (Fig. 1f).

In both cases the day-to-day variability as measured by the standard deviation is  $\sim 10$  K. The standard deviation of daily maximum  $\theta_e$  and minimum  $\theta_e$  decreases with decreasing latitude and both variables exhibit the lowest variability over the southern portions of the study domain (e.g., Florida (FL) has a standard deviation  $< 5$  K) (Fig. 1e and f).

Thus the time series of predictors (1) and (2) (global mean  $T$  and the 15 PC scores) are common to models built for all grid cells, but predictor (3) (SM) and the response variables (predictands, daily maximum  $\theta_e$  and minimum  $\theta_e$ ) are grid-cell specific.

### 2.3 Methods

Artificial neural network architectures are potentially highly useful in developing statistical models for response variables such as  $\theta_e$  because ANNs do not require any assumptions about the form of the relationship between individual predictors and between predictors and predictands (minimum and maximum  $\theta_e$ ) and can treat complex and nonlinear term interactions. ANNs are data-driven self-adaptive multi-layer perceptrons that model relationships between input variables and dependent output variables. Term interactions are described using mathematical functions encoded within hidden layers and weights that connect all nodes within the network layers including the input (predictors) and output (predictand) layers to perform the nonlinear mapping between the input and output variables (Gardner and Dorling, 1998). The number of hidden layers within the ANN determines the degree of nonlinearity that can be modeled. Hence, if the data are linearly separable, no hidden layers are required. Our a priori expectation is that the different predictors of daily maximum and minimum  $\theta_e$  will interact in complex, nonlinear ways. Thus, we apply ANN to develop models relating the global mean  $T$ , PC scores of the synoptic-scale meteorology and antecedent SM to daily maximum or minimum  $\theta_e$  in each grid cell. Because we seek to examine spatial variability in model performance, we build and test the ANNs at the grid-cell level and then examine the resulting spatial coherence of model skill. A range of different learning algorithms can be employed in ANN. Herein the neural networks are constructed within MATLAB using the Levenberg–Marquardt back-propagation algorithm (in which the sum of the squares of the deviations between the observations and model predictions is minimized) (Papageorgiou and Poczeta, 2017). Although there is no single “best-practice” regarding the number of hidden layers to use with ANN, there is evidence that a single hidden layer is sufficient for the large majority of problems (Toth et al., 2000). To test the dependence of model skill on the number of hidden layers, three independent models are constructed for each MERRA-2 grid cell using the following:

1. no hidden layers, i.e., a linear regression model with no interaction between the predictors;

2. ANN with a single hidden layer;
3. ANN with three hidden layers.

To examine the importance of SM in determining the downscaling model skill, a fourth ANN model (with three hidden layers) is also built for each grid cell and each of the two predictands (daily maximum and minimum  $\theta_e$ ) that excludes SM from the input variables. Lastly, it is challenging to determine which measures of SM are most appropriate to use within statistical downscaling models. Therefore, in addition to developing models using the MERRA-2 variable “PRMC”, which is the “total profile soil moisture content” ( $\text{m}^3 \text{m}^{-3}$ ; that is summed across all six soil layers and represents the total water potentially available for evapotranspiration to the atmosphere), a fifth ANN model (with three hidden layers) is also built that uses the variable “GWETTOP” that describes the SM content in the upper 5 cm of the soil (unitless) (Reichle et al., 2017a) and thus best represents the SM that is readily available for evaporation into the overlying atmosphere.

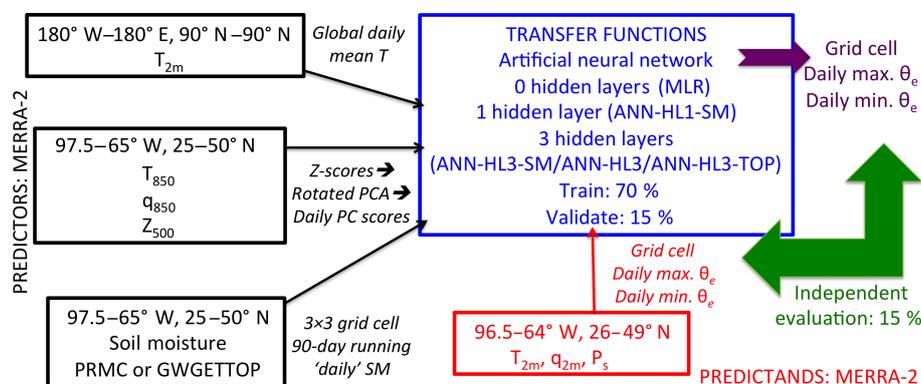
Table 1 summarizes the model abbreviations used herein. A schematic of the downscaling model architecture and data flows is given in Fig. 3. For each model (and thus each grid cell) 70 % of the data set is randomly selected to be used for training of the models, 15 % is used for internal validation and 15 % is withheld and used as an independent sample for model testing. We use two primary metrics of model performance: RMSE and Pearson correlation coefficient ( $r$ ) between observed and predicted daily minimum and maximum  $\theta_e$  in each grid cell on each day in the test data set (i.e., independent observations), summarized at both the grid-cell level and also averaged over all 1962 MERRA-2 grid cells that have some land areas within them (i.e., a domain average). The correlation coefficient is thus used as a relative measure of model performance, while RMSE provides an absolute measure of degree of agreement between the model “predictions” and the observed values (i.e., it is the typical value of the prediction error). Given the importance of extreme heat to human health, we further examine the ability of the models to capture the occurrence of very high  $\theta_e$ . In this analysis we set a threshold of 347 K ( $73^\circ\text{C}$ ) to indicate extreme maximum  $\theta_e$  (based on information provided in (Buzan et al., 2015) for the eastern USA) and a threshold of 337 K for minimum  $\theta_e$ . A contingency table approach is used to evaluate the accuracy of the model predictions of extreme  $\theta_e$  using the hit rate (HR) (Wilks, 2011):

$$\text{HR} = \frac{\text{No. of hits}}{\text{No. of hits} + \text{No. of misses}}, \quad (3)$$

where “No. of hits” is the number of days in the independent data set in which the observations and predictions both indicate exceedance of the threshold, and “No. of misses” is the number of days when the observations indicate an exceedance of the threshold but the model prediction does not.

**Table 1.** Domain-averaged model performance statistics (root mean square error (RMSE) and the mean Pearson correlation coefficient  $r$ ) vs. independent test data for daily maximum equivalent potential temperature (maximum  $\theta_e$ ) and daily minimum equivalent potential temperature (minimum  $\theta_e$ ) during JJA, 1980–2015. The results are shown for the five model formulations expressed in terms of the number of hidden layers in the ANN (where 0 hidden layers indicates results for a multiple linear regression model with no term interactions), the presence or absence of soil moisture as a predictor, and whether the SM is total columnar (PRMC) or surface only (WGETTOP). The total number of grid cells considered herein is 1960.

Abbreviation used herein	Model architecture	Soil moisture predictor	Maximum $\theta_e$			Minimum $\theta_e$		
			$\langle$ RMSE $\rangle$ (K)	$\langle r \rangle$	No. of grid cells with $r > 0.8$ , RMSE < 5 K, and both criteria met	$\langle$ RMSE $\rangle$ (K)	$\langle r \rangle$	No. of grid cells with $r > 0.8$ , RMSE < 5 K, and both criteria met
MLR	Multiple linear regression	PRMC	4.05	0.849	1575, 1727, 1370	4.39	0.836	1537, 1665, 1311
ANN-HL1-SM	ANN with one hidden layer	PRMC	4.03	0.850	1602, 1754, 1394	4.34	0.839	1572, 1750, 1370
ANN-HL3-SM	ANN with three hidden layers	PRMC	3.97	0.855	1621, 1798, 1482	4.29	0.844	1588, 1782, 1450
ANN-HL3	ANN with three hidden layers, but excluding SM	None	4.10	0.846	1549, 1675, 1300	4.33	0.843	1580, 1765, 1427
ANN-HL3-TOP	ANN with three hidden layers	WGETTOP	3.97	0.856	1623, 1802, 1476	4.29	0.844	1604, 1782, 1432



**Figure 3.** Schematic of the ANN model architecture (model naming convention), data flow, predictors and predictands.

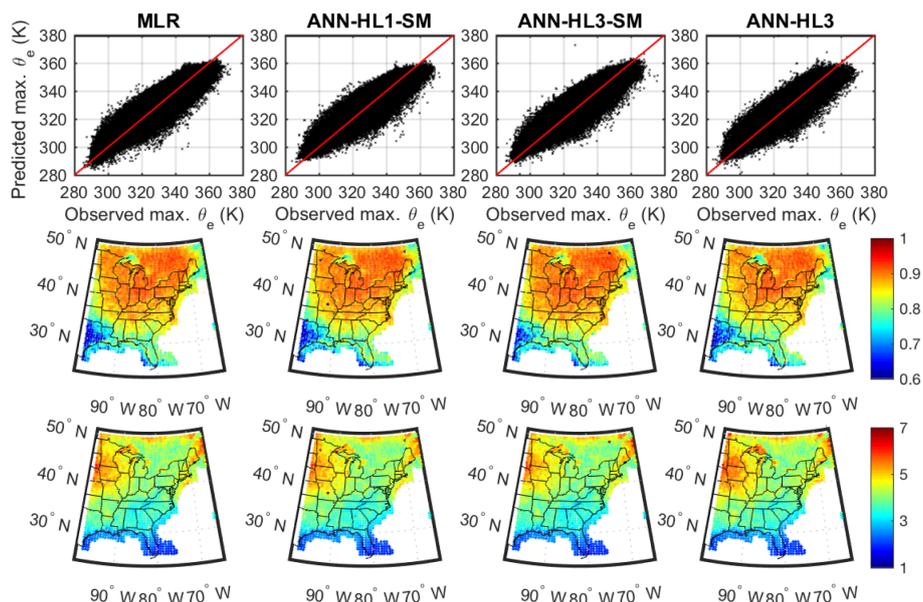
The false alarm rate (FAR) of each individual grid cell is as follows:

$$\text{FAR} = \frac{\text{No. of false alarms}}{\text{No. of false alarms} + \text{No. of correct nonevents}}, \quad (4)$$

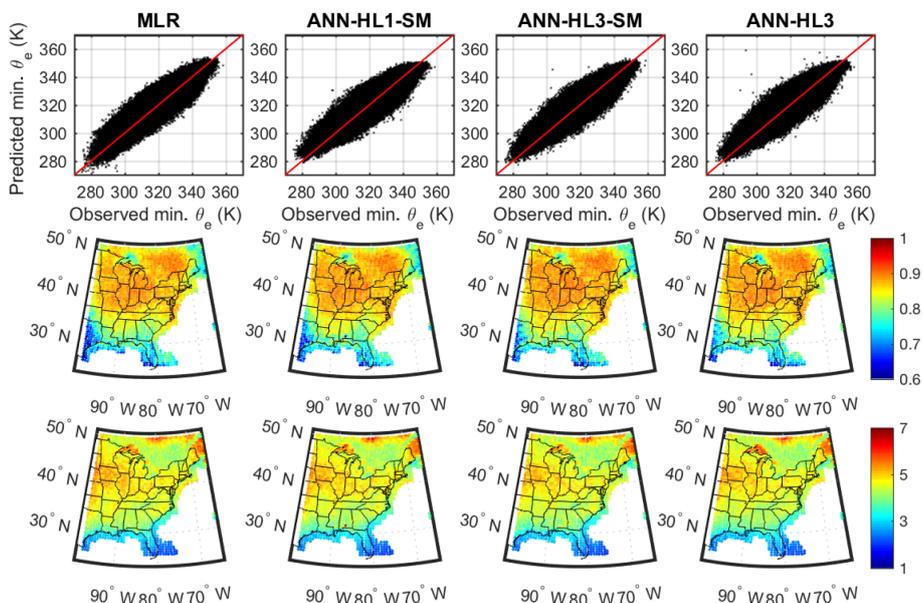
where “No. of false alarms” is the number of days in the independent data set where the observations did not indicate exceedance of the threshold but the prediction was for an exceedance, and “No. of correct nonevents” is the number of days when the observations and predictions both indicate the threshold is not exceeded.

### 3 Results

For both daily maximum  $\theta_e$  and daily minimum  $\theta_e$ , the worst model performance statistics (highest RMSE as a fraction of the temporal variability  $\theta_e$ , and lowest  $r$ ) are associated with the linear models that do not include parameter interactions (i.e., MLR) (Table 1 and Figs. 4 and 5). Nevertheless, output from all model architectures for minimum  $\theta_e$  and maximum  $\theta_e$  exhibit high correlation coefficients ( $r > 0.8$ ) with independent data over most of the study domain. Correlation coefficients exceed 0.8 for 84 % of grid cells for ANN-HL3-SM (ANN with three hidden layers) applied to maxi-



**Figure 4.** Précis of the downscaling model performance for daily maximum  $\theta_e$  (max.  $\theta_e$ ). The upper panels show scatter plots of observed and predicted maximum  $\theta_e$  for all grid cells and all days within the test (independent) data, along with a red 1 : 1 line. The columns show results for four model configurations: multiple linear regression (MLR, left), ANN with one hidden layer (ANN-HL1-SM, second column), ANN with three hidden layers (ANN-HL3-SM, third column) and ANN with three hidden layers but excluding information regarding the soil moisture (ANN-HL3, right column). The second row shows the Pearson correlation ( $r$ ) of predicted and observed maximum  $\theta_e$  values in the independent data for the four different models, while the third row shows the mean root mean square error (RMSE) (K) computed for independent data from each grid cell for the four models. Results for ANN-HL3-TOP are virtually identical to those for ANN-HL3-SM and thus are not shown.



**Figure 5.** Précis of the downscaling model performance for daily minimum  $\theta_e$  (min.  $\theta_e$ ). The upper panels show scatter plots of observed and predicted minimum  $\theta_e$  for all grid cells and all days within the test (independent) data, along with a red 1 : 1 line. The columns show results for four model configurations: multiple linear regression (MLR, left), ANN with one hidden layer (ANN-HL1-SM, second column), ANN with three hidden layers (ANN-HL3-SM, third column) and ANN with three hidden layers but excluding information regarding the soil moisture (ANN-HL3, right column). The second row shows the Pearson correlation ( $r$ ) of predicted and observed minimum  $\theta_e$  values in the independent data for the four different models, while the third row shows the mean root mean square error (RMSE) (K) computed for independent data from each grid cell for the four models. Results for ANN-HL3-TOP are virtually identical to those for ANN-HL3-SM and thus are not shown.

imum  $\theta_e$  and 81 % for minimum  $\theta_e$ . Further, 92 % of grid cells for ANN-HL3-SM exhibit RMSE < 5 K for maximum  $\theta_e$  and 91 % exhibit RMSE < 5 K for minimum  $\theta_e$  (Table 1, Figs. 4 and 5). The ANN-HL3-SM model also exhibits the highest number of grid cells that have both a RMSE < 5 K and a  $r > 0.8$  for both maximum  $\theta_e$  and minimum  $\theta_e$  (Table 1). Grid cells with the highest RMSE for minimum  $\theta_e$  and maximum  $\theta_e$  also generally have the highest variance (i.e., largest day-to-day variability). For example, grid cells in IA exhibit the highest variance and highest RMSE for maximum  $\theta_e$  (Figs. 1e and 4), and grid cells within IL are generally characterized by large RMSE and variance of minimum  $\theta_e$  (Figs. 1f and 5). Thus, while noting the RMSE (i.e., typical prediction error) is largest over IA for maximum  $\theta_e$  ( $\sim 5$  K), it is less than half the standard deviation computed from the day-to-day variability in maximum  $\theta_e$  ( $\sim 10$  K).

Generally all models exhibit slightly worse performance across both measures ( $r$  and RMSE) for minimum  $\theta_e$  than maximum  $\theta_e$  at the grid-cell level and integrated over all land grid cells (Table 1 and Figs. 4 and 5). The reduced model skill for minimum  $\theta_e$  may reflect use of output at 20:00 UTC values of the predictors used in the synoptic classification due to our particular focus on daytime maximum  $\theta_e$ .

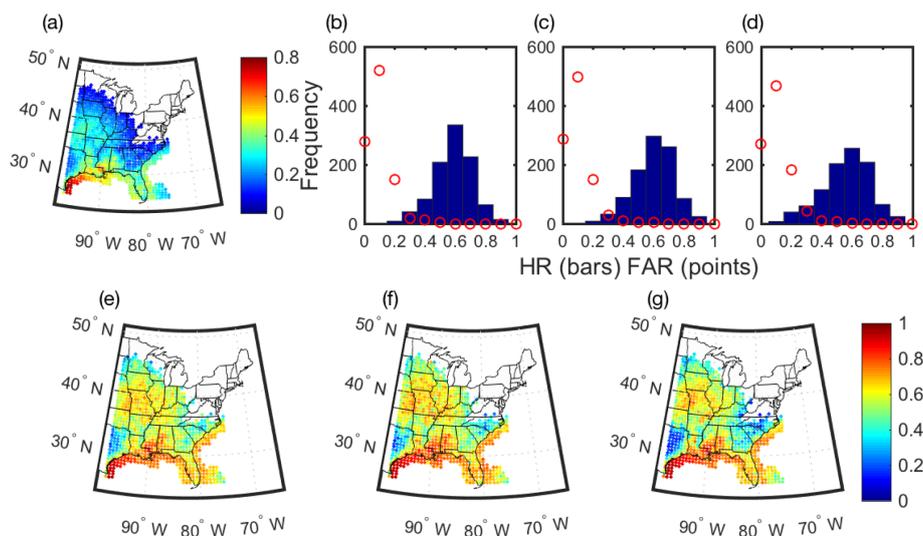
Although performance differences between the five model architectures for daily maximum  $\theta_e$  and daily minimum  $\theta_e$  are comparatively modest when averaged over the entire domain (Table 1), there are important regional variations in the performance of the different model functional forms. Over two-thirds of all grid cells (1332 of 1962) exhibit lower RMSE in the ANN model with three hidden layers and including SM (i.e., ANN-HL3-SM) than in any of the other models (e.g., MLR). The enhancement of model performance as measured by a decrease in RMSE for the more complex model of maximum  $\theta_e$  and minimum  $\theta_e$  is particularly marked in the west-central region of the domain (over parts of Missouri (MO) and Iowa (IA), close to or within one of the “warming holes”) (Figs. 4 and 5). This is a region where a substantial fraction of  $T$  variance is explained by thermal and moisture advection by the GPLLJ (Weaver, 2013), where modeled land–atmosphere coupling is particularly intense (Koster et al., 2006, 2011) and where there are strong longitudinal gradients of SM (Fig. 1g). The lowest correlations between predicted and observed minimum and maximum  $\theta_e$  values occur over east Texas (TX) for all model formulations although the RMSE of model predictions are not particularly high in this area (Figs. 4 and 5). The low RMSE may reflect the small day-to-day variability in minimum and maximum  $\theta_e$  over this region (Fig. 1e and f), possibly due to the proximity to the ocean, while the low  $r$  may indicate that the synoptic types derived herein are not able to represent mesoscale features such as dry lines that play a key role in dictating day-to-day variability in  $\theta_e$  over this subregion. It is also worthy of note that this area was excluded from the eastern US in terms of the area of coherence for extreme  $T$  over the eastern USA (McKinnon et al., 2016)

and that eastern TX is on the southwestern boundary of the study domain. These two factors may indicate that the synoptic types derived herein do not fully represent the range of meteorological conditions associated with  $\theta_e$  variability in the lower central Great Plains.

Interpreting weights from complex ANNs is very challenging in the context of predictor relevance and the values (and sign) of the weights vary in space and with the complexity of the model architecture (number of hidden layers). However, it is worthy of note that the weights on the index of global temperatures (and thus expression of internal climate modes) in ANN-HL3-SM are uniformly close to zero across the grid cells in the interior of the continent, but are of large magnitude in land grid cells close to the Gulf of Mexico (i.e., the south of the domain) and around the Great Lakes. Given this index is strongly influenced by global sea surface temperatures (SSTs), the implication is that this predictor contains important information about the SST and thus potentially evaporation from the Great Lakes and the Gulf of Mexico, leading to higher  $q$ .

Consistent with prior research that has indicated the importance of atmosphere–surface interactions (Cai et al., 2014) and specifically soil moisture (Pryor et al., 2016; Seneviratne et al., 2010) in surface energy partitioning and thus near-surface  $T$  and  $q$  regimes and static energy, exclusion of SM from the ANN with three hidden layers (i.e., ANN-HL3) decreases model performance relative to ANN-HL3-SM and increases the RMSE for maximum  $\theta_e$  in 70 % of grid cells. The regions for which this impact is most strongly manifest are close to or within the “warming holes” described above and/or are located downstream of regions of significant county-level irrigation and anthropogenic enhancement of SM (Pryor et al., 2016; DeAngelis et al., 2010) (Fig. 1g), indicating the potential for anthropogenic enhancement of SM to strongly influence static energy and human heat stress in these regions. For example, RMSE for maximum  $\theta_e$  is increased in models excluding SM in all grid cells within MO, and all but one grid cell each in IA and IL (Fig. 4). This finding is also replicated in the second region of weak or negative air temperature trends described above and centered on MS and AL (Ellenburg et al., 2016). The RMSE is lower in ANN-HL3-SM than ANN-HL3 over all but one grid cell in these two states. Thus this analysis strongly supports prior assertions that SM plays a key role in dictating the surface energy balance and in the suppression of daily maximum  $T$ , while increasing maximum  $\theta_e$ .

The statistical (downscaling) models show similar dynamic range to independent observations, although there is some evidence that the models underestimate the total variance in maximum  $\theta_e$  leading to underestimation of extreme maximum  $\theta_e$ , as is evident from the flattening of the scatter plots for very high values of daily maximum  $\theta_e$  (see the upper row of panels in Fig. 4). To examine this further we conduct an analysis of the HR and FAR for maximum  $\theta_e$  in excess of 347 K. This threshold is exceeded by daily max-



**Figure 6.** (a) Map of the fractional frequency of occurrence of maximum  $\theta_e > 347$  K in each grid cell based on the independent (test) data set. (b) Histogram of hit rate (HR) and false alarm rate (FAR) for all grid cells and (e) spatial map for correct identification of maximum  $\theta_e > 347$  K for all grid cells where this threshold is exceeded by  $> 8\%$  of days in the independent data based on output from the multiple linear regression model (MLR). (c, f) As in panels (b) and (e) but for output from the ANN model with three hidden layers and including SM (ANN-HL3-SM). (d, g) As in panels (b) and (e) but for output from the ANN model with 3 hidden layers excluding SM (ANN-HL3).

imum  $\theta_e$  derived from the MERRA-2 reanalysis on an average of  $\sim 15\%$  of summer days when all eastern US grid cells are considered, but naturally exhibits a higher frequency of exceedance (of up to  $75\%$  of days) along the southeastern portion of the TX gulf coast and is observed on nearly  $50\%$  of days over coastal portions of the Gulf coast states and FL (Fig. 6a). Conversely, it is seldom or never observed within grid cells in the north of the domain (Fig. 6a). To ensure a sufficiently robust sample size on which to compute the HR for extreme maximum  $\theta_e$  we consider only grid cells where more than 40 days in the independent data sample (i.e.,  $8\%$ ) exceed this threshold. The mean HR values for the linear model (MLR), the ANN with three hidden layers (ANN-HL3-SM) and the ANN with three hidden layers but excluding SM (ANN-HL3) computed over all these grid cells are 59, 60 and 56 %, respectively (Fig. 6b–d), indicating that over the entire study domain the role of SM and predictor interactions in explaining the occurrence of extreme maximum  $\theta_e$  is modest. All model forms perform least well in terms of predicting the occurrence of maximum  $\theta_e > 347$  K over eastern TX and South Carolina (SC) (Fig. 6e–g). However, the model excluding SM exhibits a particularly poor performance (i.e., low HR) in these regions. The causes of the poor model performance in eastern TX and SC are currently not fully understood, although it is worthy of note that data from MERRA-2 grid cells in SC exhibit a relatively low overall frequency of exceedance of this threshold and are also characterized by comparatively low 99th percentile  $\theta_e$  in an analysis of heat indices derived from the Community Land Model v4.5 (Buzan et al., 2015). Grid cells along the Gulf coast and over the states of MO, IA and IL exhibit high HR

for prediction of extreme maximum  $\theta_e$  and substantial improvement in HR is noted in IA, IL and MO (Fig. 6e–g) in the ANN-HL3-SM relative to the other model forms. This is consistent with strong spatial gradients in SM (Fig. 1e), findings of the GLACE projects of strong atmosphere–surface coupling (Koster et al., 2006, 2011), and analyses for stations in IL that also show a strong dependence of high  $T_e$  on soil moisture (Ford and Schoof, 2016). To contextualize the HR presented above it is important to note that they are associated with comparatively low false alarm rates (FAR). Indeed, FAR for the occurrence of minimum  $\theta_e > 342$  K or maximum  $\theta_e > 347$  K are very modest for all model formulations (Figs. 6 and 7). For example, over  $94\%$  of grid cells indicate FAR for maximum  $\theta_e > 347$  K that are below 0.25 for the ANN-HL3-SM models. Thus, the relatively high HR reported herein are not being artificially inflated by unrealistically high predictions of the occurrence of extreme  $\theta_e$ . The inclusion of SM as a predictor enhances HR in regions previously identified as exhibiting high variance in extreme  $\theta_e$  without a concomitant increase in FAR (Figs. 6 and 7). It should be acknowledged that even the ANN with three hidden layers and soil moisture (ANN-HL3-SM) exhibits a modal grid-cell HR of 0.6–0.7, and thus misses a substantial fraction of extreme  $\theta_e$ . Nevertheless, these HR and FAR are indicative of positive relative operating characteristics (ROC) (i.e., plots of the true positive rate greatly exceeds false positive rates) (Wilks, 2011). Further, HR and FAR computed for maximum  $\theta_e$  and minimum  $\theta_e$  are comparable to (or better than) seasonal re-forecasts of summertime  $T$  at 2 m over the land areas of southern Europe developed using the European Centre for Medium-Range Weather

Forecasts (ECMWF) seasonal ensemble forecasting system (Weisheimer et al., 2011).

In contrast to the results for prediction of extreme maximum  $\theta_e$  the model architecture has virtually no impact on HR for minimum  $\theta_e > 337$  K, and neither does the inclusion of SM in the model. In all cases the domain-averaged HR = 59 % and no region exhibits consistent improved or degraded performance for ANN-HL3-SM or ANN-HL3 over MLR (Fig. 7). This finding is consistent with the overall results for models of minimum  $\theta_e$  that exhibit only modest decreases in model performance (increased RMSE and lower  $r$ ) when SM is excluded from the predictor suite (Table 1 and Fig. 5). Consistent with the interpretation of the surface energy balance (Eq. 2) this re-emphasizes that SM more directly impacts near-surface  $T$  and  $q$  during the daytime under conditions of positive net radiation.

Differences in model performance between ANNs conditioned on total SM and using wetness only in the top soil layer (upper 5 cm) are very small when averaged across the domain (Table 1) and indeed for virtually all grid cells. Only 26 grid cells exhibited a  $\Delta|\text{RMSE}| > 0.5$  K for models using PRMC vs. those using GWETTOP (out of a total of 1962), while 155 exhibited an increase in RMSE  $> 0.5$  K when SM was excluded from the model. Thus, although the weights within the ANNs differ for use of the two SM parameters, the overall model skill is unchanged by use of the two SM estimates, possibly due to the spatial and temporal averaging applied herein, or uncertainty in reanalysis-derived SM variables.

#### 4 Summary and conclusions

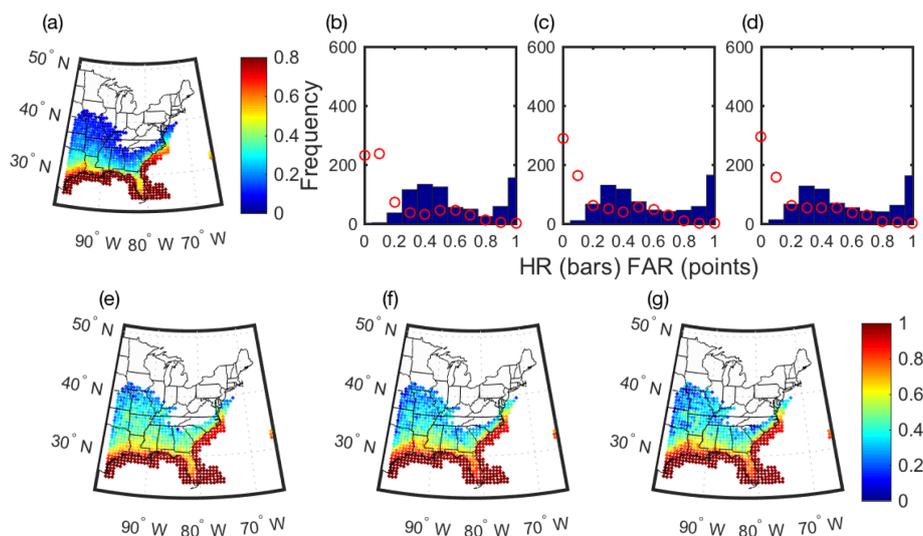
Very few statistical downscaling analyses focus on integrative variables such as  $\theta_e$  that explicitly incorporate co-variability of  $T$  and  $q$ , but such variables have direct applications to climate change impact analyses (such as analysis of heat waves; Buzan et al., 2015). Further, this is an application of climate downscaling where statistical approaches may be particularly useful given evidence that even when nested within observed lateral boundary conditions regional climate models (RCMs) have difficulty in capturing the joint probability distributions of  $T$  and  $q$  and thus in accurately representing either the probability distribution of static energy or the spatiotemporal variability therein (Pryor and Schoof, 2016). Analyses of  $\theta_e$  are also essential to advancing fundamental understanding of changes in the total static energy content of the lower atmosphere, and may reveal important information of relevance to both model performance analyses and attribution studies of global change.

The goal of this work is to develop a hierarchy of statistical models with increasing complexity and use them to determine the degree to which increased complexity enhances the skill of model predictions of  $\theta_e$  and to attribute variability in minimum and maximum  $\theta_e$  over eastern North America.

Prior to discussing the results from application of this analysis framework to output from the MERRA-2 reanalysis, it is worthy of note that previous research on regional heat wave characteristics over the contiguous US using a suite of reanalyses indicated some important differences in the magnitude of derived equivalent temperature ( $T_e$ ) between the reanalysis products (Schoof et al., 2017) as well as in the strength of land–atmosphere coupling between the reanalysis products (Ferguson et al., 2012). Thus, there would be value in applying this framework to additional observationally constrained data sets to evaluate (1) the degree to which the findings of a key role of SM to determining the model skill for daily maximum  $\theta_e$  in specific subregions are generalizable and spatially consistent between reanalyses, and further if the predictability of  $\theta_e$  exhibits sensitivity to the spatiotemporal averaging used in deriving the SM predictors; (2) if use of a reanalysis product (or forecast model) that does not employ bias correction of precipitation amounts would substantially alter the ANN model structure; and (3) if the partial truncation of the upper percentiles of daily maximum  $\theta_e$  in the model predictions is also a generalizable finding when our model framework is applied to different data sets.

Consistent with our a priori expectations, models built using ANN out-perform those that do not permit interaction of the predictor variables. Domain-averaged RMSE for minimum and maximum  $\theta_e$  is smallest in the more complex models (i.e., for ANN-HL3-SM, RMSE  $< 4$  and  $< 4.3$  K, respectively, cf. mean maximum  $\theta_e \approx 333$  K and mean minimum  $\theta_e \approx 321$  K). Particularly in regions with high variability in minimum and maximum  $\theta_e$ , the more complex models with multiple hidden layers are better able to capture the day-to-day variability in  $\theta_e$ . Correlation coefficients exceed 0.8 for 84 % of grid cells for ANN-HL3-SM applied to maximum  $\theta_e$  and 81 % for minimum  $\theta_e$ . Further, 92 % of grid cells for ANN-HL3-SM exhibit a RMSE  $< 5$  K for maximum  $\theta_e$  and 91 % for minimum  $\theta_e$ .

The primary purposes of this research are to enhance understanding of the causes of variability and change in  $\theta_e$  over the eastern USA and to propose a new downscaling approach to allow projections of daily minimum and maximum  $\theta_e$  using variables commonly available from reanalyses and global and regional climate models. However, although prognostic thermal physiological models are required to make accurate assessments of human heat stress, the ANN models developed here may also have utility in assessments of possible climate change impacts on human health. Further, these analyses also may have applications to short-term forecasting of human-health-relevant heat events (McKinnon et al., 2016; Weisheimer et al., 2011), since the methodological framework developed herein could be applied to observed antecedent SM, and modeled forecasts of the global mean  $T$  and conditions on a synoptic scale over the eastern USA. Many of the heat watch-warning systems implemented across the United States currently employ a synoptic typing methodology (Sheridan and Kalkstein, 2004), but the perfor-



**Figure 7.** (a) Map of the fractional frequency of occurrence of minimum  $\theta_e > 337$  K in each grid cell based on the independent (test) data set. (b) Histogram of hit rate (HR) and false alarm rate (FAR) for all grid cells and (e) spatial map for correct identification of minimum  $\theta_e > 337$  K for all grid cells where this threshold is exceeded by  $> 8\%$  of days in the independent data based on output from the multiple linear regression model (MLR). (c, f) As in panels (b) and (e) but for output from the ANN model with three hidden layers and including SM (ANN-HL3-SM). (d, g) As in panels (b) and (e) but for output from the ANN model with 3 hidden layers excluding SM (ANN-HL3).

mance of such systems may be aided by implementation of other variables and/or analysis methodologies such as those used herein. The ANN-HL3-SM models developed herein exhibit relatively high skill in predicting the occurrence of extreme minimum and maximum  $\theta_e$ , and indeed out-perform the simpler models. The ANN with three hidden layers that includes SM as a predictor (i.e., ANN-HL3-SM) exhibits a domain-averaged median hit rate for maximum  $\theta_e > 347$  K that is  $> 0.60$ , while the median FAR is  $\approx 0.08$ . Results from the ANN models further indicate that maximum  $\theta_e$  and the occurrence of extreme maximum  $\theta_e$  appear to be considerably more sensitive to SM than minimum  $\theta_e$ , which in turn appears to exhibit a stronger dependence on the precise prevailing synoptic-scale conditions based on the ANN weights.

Our results imply there are large spatial gradients in the importance of the predictors we used herein. For example, in the northeastern portions of our study region inclusion of SM as a predictor has considerably lower impact on model skill for either maximum  $\theta_e$  or minimum  $\theta_e$  (Figs. 4–7). Global  $T$  substantially contributes to model skill near the Gulf coast and close to the Great Lakes but is less important over the remainder to the eastern USA, while SM exhibits the greatest importance in subregions previously noted as exhibiting warming holes. Our framework has greater skill for maximum  $\theta_e$  than minimum  $\theta_e$ . It is possible that inclusion of additional predictors could lead to enhanced model skill, particularly for extreme high values of maximum  $\theta_e$  or minimum  $\theta_e$  that are of greatest importance to human health, and/or that our methodology could be evolved to allow derivation of persistence indices (e.g., the occurrence of consecutive nights with high minimum  $\theta_e$ ).

We can not conclusively discount contributions from other phenomena (e.g., aerosol forcing, cloud cover) to the occurrence of warming holes (areas with declining or no-trends in  $T$ ) (Meehl et al., 2015), and these features may be a complex response to multiple drivers. However, results presented herein are consistent with past work that has indicated the importance of soil moisture in determining partitioning of the surface energy budget, and thus the spatiotemporal patterns of  $\theta_e$  over the central and eastern USA (Koster et al., 2006, 2011; Pryor and Schoof, 2016; Pryor et al., 2016; Ford and Schoof, 2016, 2017; McKinnon et al., 2016). Indeed, SM is particularly important in determining the surface energy partitioning and the magnitude of  $\theta_e$  over regions that have previously been identified as exhibiting warming holes, and for all grid cells the RMSE for models including SM as a predictor is smaller than the temporal variability of  $\theta_e$  as measured using the standard deviation of the daily  $\theta_e$  values. Specifically, only a model including SM is able to predict the occurrence of extreme (and highly health-relevant) values of  $\theta_e$  over the western portion of Midwestern states such as IA, MO, IL and also in MS and AL. This research thus implies that SM has played and may continue to play a key role in dictating the presence and intensity of warming holes that have been previously noted in analyses of near-surface air temperature data (from both in situ measurements and re-analysis products).

*Data availability.* The MERRA-2 data used herein are available from <http://disc.sci.gsfc.nasa.gov/daac-bin/FTPSubset2.pl> (NASA,

2016; see Bosilovich et al., 2015 for a detailed description of the file structures).

*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* The authors gratefully acknowledge funding from the US National Science Foundation under grants 1339629, 1339655 and 1502400, the US Department of Energy (DE-SC0016438) and the US National Aeronautics and Space Administration (14-EARTH14F-0207). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the view of the National Science Foundation, National Aeronautics and Space Administration or Department of Energy.

Edited by: Qiang Fu

Reviewed by: three anonymous referees

## References

- Al Bitar, A., Mialon, A., Kerr, Y. H., Cabot, F., Richaume, P., Jacquette, E., Quesney, A., Mahmoodi, A., Tarot, S., Parrens, M., Al-Yaari, A., Pellarin, T., Rodriguez-Fernandez, N., and Wigneron, J.-P.: The global SMOS Level 3 daily soil moisture and brightness temperature maps, *Earth Syst. Sci. Data*, 9, 293–315, <https://doi.org/10.5194/essd-9-293-2017>, 2017.
- Allen, M. and Ingram, W.: Constraints on future changes in climate and the hydrologic cycle, *Nature*, 419, 224–232, 2002.
- Anderson, G. B. and Bell, M. L.: Heat waves in the United States: mortality risk during heat waves and effect modification by heat wave characteristics in 43 US communities, *Environ. Health Persp.*, 119, 210–218, 2011.
- Bentley, M. L. and Stallins, J. A.: Synoptic evolution of Midwestern US extreme dew point events, *Int. J. Climatol.*, 28, 1213–1225, 2008.
- Bolton, D.: The computation of equivalent potential temperature, *Mon. Weather Rev.*, 108, 1046–1053, 1980.
- Bosilovich, M. G., Lucchesi, R., and Suarez, M.: MERRA-2: File Specification, GMAO Office Note No. 9 (Version 1.0), 73 pp., available at: [http://gmao.gsfc.nasa.gov/pubs/office\\_notes](http://gmao.gsfc.nasa.gov/pubs/office_notes) (last access: 14 July 2016), 2015.
- Brabson, B., Lister, D., Jones, P., and Palutikof, J.: Soil moisture and predicted spells of extreme temperatures in Britain, *J. Geophys. Res.-Atmos.*, 110, D05104, <https://doi.org/10.1029/2004JD005156>, 2005.
- Brown, D. G., Johnson, K. M., Loveland, T. R., and Theobald, D. M.: Rural land-use trends in the conterminous United States, 1950–2000, *Ecol. Appl.*, 15, 1851–1863, <https://doi.org/10.1890/03-5220>, 2005.
- Buzan, J. R., Oleson, K., and Huber, M.: Implementation and comparison of a suite of heat stress metrics within the Community Land Model version 4.5, *Geosci. Model Dev.*, 8, 151–170, <https://doi.org/10.5194/gmd-8-151-2015>, 2015.
- Cai, X., Yang, Z. L., Xia, Y., Huang, M., Wei, H., Leung, L. R., and Ek, M. B.: Assessment of simulated water balance from Noah, NoahMP, CLM, and VIC over CONUS using the NLDAS test bed, *J. Geophys. Res.-Atmos.*, 119, 13751–13770, 2014.
- Cattell, R. B.: The scree test for the number of factors, *Multivar. Behav. Res.*, 1, 245–276, 1966.
- Davey, C. A., Pielke Sr., R. A., and Gallo, K. P.: Differences between near-surface equivalent temperature and temperature trends for the Eastern United States – equivalent temperature as an alternative measure of heat content, *Global Planet. Change*, 54, 19–32, <https://doi.org/10.1016/j.gloplacha.2005.11.002>, 2006.
- DeAngelis, A., Dominguez, F., Fan, Y., Robock, A., Kustu, M. D., and Robinson, D.: Evidence of enhanced precipitation due to irrigation over the Great Plains of the United States, *J. Geophys. Res.-Atmos.*, 115, D15115, <https://doi.org/10.1029/2010JD013892>, 2010.
- de’Donato, F. K., Leone, M., Scortichini, M., De Sario, M., Katsouyanni, K., Lanki, T., Basagana, X., Ballester, F., Astrom, C., Paldy, A., Pascal, M., Gasparrini, A., Menne, B., and Michelozzi, P.: Changes in the effect of heat on mortality in the last 20 years in nine European cities. Results from the PHASE project, *Int. J. Environ. Res. Public Health*, 12, 15567–15583, <https://doi.org/10.3390/ijerph121215006>, 2015.
- Diem, J. E.: Synoptic-scale controls of summer precipitation in the Southeastern United States, *J. Climate*, 19, 613–621, <https://doi.org/10.1175/jcli3645.1>, 2006.
- Dirmeyer, P. A. and Halder, S.: Application of the Land–Atmosphere Coupling Paradigm to the Operational Coupled Forecast System, Version 2 (CFSv2), *J. Hydrometeorol.*, 18, 85–108, <https://doi.org/10.1175/jhm-d-16-0064.1>, 2017.
- Ellenburg, W. L., McNider, R. T., Cruise, J. F., and Christy, J. R.: Towards an understanding of the twentieth-century cooling trend in the southeastern United States: biogeophysical impacts of land-use change, *Earth Interact.*, 20, 20–018, <https://doi.org/10.1175/ei-d-15-0038.1>, 2016.
- Fall, S., Diffenbaugh, N. S., Niyogi, D., Pielke Sr., R. A., and Rochon, G.: Temperature and equivalent temperature over the United States (1979–2005), *Int. J. Climatol.*, 30, 2045–2054, <https://doi.org/10.1002/joc.2094>, 2010.
- Ferguson, C. R., Wood, E. F., and Vinukollu, R. K.: A global inter-comparison of modeled and observed land–atmosphere coupling, *J. Hydrometeorol.*, 13, 749–784, <https://doi.org/10.1175/jhm-d-11-0119.1>, 2012.
- Ford, T. W. and Schoof, J. T.: Oppressive heat events in Illinois related to antecedent wet soils, *J. Hydrometeorol.*, 17, 2713–2726, 2016.
- Ford, T. W. and Schoof, J. T.: Characterizing extreme and oppressive heat waves in Illinois, *J. Geophys. Res.-Atmos.*, 122, 682–698, 2017.
- Garcia-Herrera, R., Diaz, J., Trigo, R. M., Luterbacher, J., and Fischer, E. M.: A review of the European summer heat wave of 2003, *Crit. Rev. Env. Sci. Tec.*, 40, 267–306, 2010.
- Gardner, M. W. and Dorling, S.: Artificial neural networks (the multilayer perceptron) – a review of applications in the atmospheric sciences, *Atmos. Environ.*, 32, 2627–2636, 1998.
- Grotjahn, R., Black, R., Leung, R., Wehner, M. F., Barlow, M., Bosilovich, M., Gershunov, A., Gutowski Jr., W. J., Gyakum, J. R., and Katz, R. W.: North American extreme temperature events and related large scale meteorological patterns:

- a review of statistical methods, dynamics, modeling, and trends, *Clim. Dynam.*, 46, 1151–1184, 2016.
- Guo, Z. C., Dirmeyer, P. A., Koster, R. D., Bonan, G., Chan, E., Cox, P., Gordon, C. T., Kanae, S., Kowalczyk, E., Lawrence, D., Liu, P., Lu, C. H., Malyshev, S., McAvaney, B., McGregor, J. L., Mitchell, K., Mocko, D., Oki, T., Oleson, K. W., Pitman, A., Sud, Y. C., Taylor, C. M., Verseghy, D., Vasic, R., Xue, Y. K., and Yamada, T.: GLACE: The Global Land–Atmosphere Coupling Experiment. Part II: Analysis, *J. Hydrometeorol.*, 7, 611–625, <https://doi.org/10.1175/jhm511.1>, 2006.
- Horton, R. M., Mankin, J. S., Lesk, C., Coffel, E., and Raymond, C.: A review of recent advances in research on extreme heat events, *Curr. Clim. Change Rep.*, 24, 242–259, 2016.
- Huang, J. P., Xie, Y. K., Guan, X. D., Li, D. D., and Ji, F.: The dynamics of the warming hiatus over the Northern Hemisphere, *Clim. Dynam.*, 48, 429–446, <https://doi.org/10.1007/s00382-016-3085-8>, 2017.
- Kalkstein, L. S., Sheridan, S. C., and Graybeal, D. Y.: A determination of character and frequency changes in air masses using a spatial synoptic classification, *Int. J. Climatol.*, 18, 1223–1236, 1998.
- Kalnay, E. and Cai, M.: Impact of urbanization and land-use change on climate, *Nature*, 423, 528–531, 2003.
- Koster, R. D., Sud, Y., Guo, Z., Dirmeyer, P. A., Bonan, G., Oleson, K. W., Chan, E., Verseghy, D., Cox, P., and Davies, H.: GLACE: The Global Land–Atmosphere Coupling Experiment. Part I: Overview, *J. Hydrometeorol.*, 7, 590–610, 2006.
- Koster, R. D., Mahanama, S., Yamada, T., Balsamo, G., Berg, A., Boisserie, M., Dirmeyer, P., Doblas-Reyes, F., Drewitt, G., and Gordon, C.: The second phase of the global land–atmosphere coupling experiment: soil moisture contributions to subseasonal forecast skill, *J. Hydrometeorol.*, 12, 805–822, 2011.
- Lee, E., Bieda, R., Basara, H., and Shanmugasundaram, J.: Land surface and atmospheric conditions associated with heat waves over the Chickasaw Nation in the South Central United States, *J. Geophys. Res.-Atmos.*, 121, 6284–6298, 2016.
- Leibensperger, E. M., Mickley, L. J., Jacob, D. J., Chen, W.-T., Seinfeld, J. H., Nenes, A., Adams, P. J., Streets, D. G., Kumar, N., and Rind, D.: Climatic effects of 1950–2050 changes in US anthropogenic aerosols – Part 2: Climate response, *Atmos. Chem. Phys.*, 12, 3349–3362, <https://doi.org/10.5194/acp-12-3349-2012>, 2012.
- Llamedo, P., Hierro, R., de la Torre, A., and Alexander, P.: ENSO-related moisture and temperature anomalies over South America derived from GPS radio occultation profiles, *Int. J. Climatol.*, 37, 268–275, <https://doi.org/10.1002/joc.4702>, 2017.
- Mann, M. E., Bradley, R. S., and Hughes, M. K.: Global-scale temperature patterns and climate forcing over the past six centuries, *Nature*, 392, 779–787, 1998.
- McKinnon, K. A., Rhines, A., Tingley, M. P., and Huybers, P.: Long-lead predictions of eastern United States hot days from Pacific sea surface temperatures, *Nat. Geosci.*, 9, 389–394, <https://doi.org/10.1038/ngeo2687>, 2016.
- Meehl, G. A., Arblaster, J. M., and Chung, C. T. Y.: Disappearance of the southeast US “warming hole” with the late 1990s transition of the Interdecadal Pacific Oscillation, *Geophys. Res. Lett.*, 42, 5564–5570, <https://doi.org/10.1002/2015gl064586>, 2015.
- Molod, A., Takacs, L., Suarez, M., and Bacmeister, J.: Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2, *Geosci. Model Dev.*, 8, 1339–1356, <https://doi.org/10.5194/gmd-8-1339-2015>, 2015.
- NASA (The US National Aeronautics and Space Administration): The Modern-Era Retrospective Analysis for Research and applications, Version 2 (MERRA-2), available at: <http://disc.sci.gsfc.nasa.gov/daac-bin/FTPSubset2.pl>, last access: 14 July 2016.
- O’Neill, M. S. and Ebi, K. L.: Temperature extremes and health: impacts of climate variability and change in the United States, *J. Occup. Environ. Med.*, 51, 13–25, <https://doi.org/10.1097/JOM.0b013e318173e122>, 2009.
- Pan, Z. T., Segal, M., Li, X., and Zib, B.: Global climate change impact on the Midwestern USA – a summer cooling trend, in: *Understanding Climate Change: Climate variability, predictability and change in the Midwestern United States*, edited by: Pryor, S. C., Indiana University Press, Bloomington, IN, USA, 29–41, 2009.
- Papageorgiou, E. I. and Poczeta, K.: A two-stage model for time series prediction based on fuzzy cognitive maps and neural networks, *Neurocomputing*, 232, 113–121, <https://doi.org/10.1016/j.neucom.2016.10.072>, 2017.
- Peterson, T. C., Willett, K. M., and Thorne, P. W.: Observed changes in surface atmospheric energy over land, *Geophys. Res. Lett.*, 38, L16707, <https://doi.org/10.1029/2011GL048442>, 2011.
- Peterson, T. C., Alexander, L. V., Allen, M. R., Anel, J. A., Barriopedro, D., Black, M. T., Carey-Smith, T., Castillo, R., Cattiaux, J., Chen, X. L., Chen, X. Y., Chevallier, M., Christidis, N., Ciavarella, A., de Vries, H., Dean, S. M., Deans, K., Diffenbaugh, N. S., Doblas-Reyes, F., Donat, M. G., Dong, B., Eilerts, G., Funk, C., Galu, G., Garcia-Herrera, R., Germe, A., Gill, S., Gimeno, L., Guemas, V., Herring, S. C., Hoell, A., Horling, M. P., Huntingford, C., Husak, G., Imada, Y., Ishii, M., Karoly, D. J., Kimoto, M., King, A. D., Knutson, T. R., Lewis, S. C., Lin, R. P., Lyon, B., Massey, N., Mazza, E., Michaelsen, J., Mollard, J., Mori, M., Mote, P. W., Nieto, R., Otto, F. E. L., Park, J., Perkins, S. E., Rosier, S., Rowland, J., Rupp, D. E., Salas y Melia, D., Scherer, M., Shiogama, H., Shukla, S., Song, F. F., Sparrow, S., Scott, P. A., Sutton, R., Sweet, W., Tett, S. F. B., Trigo, R. M., van Oldenborgh, G. J., van Westrhenen, R., Verdin, J., Watanabe, M., Wittenberg, A. T., Woollings, T., Yiou, P., Zeng, F. R., Zervas, C., Zhang, R., and Zhou, T. J.: Explaining extreme events of 2012 from a climatic perspective, *B. Am. Meteorol. Soc.*, 94, S1–S74, <https://doi.org/10.1175/bams-d-13-00085.1>, 2013.
- Pryor, S. C. and Schoof, J. T.: Evaluation of nearsurface temperature, humidity, and equivalent temperature from regional climate models applied in type II downscaling, *J. Geophys. Res.-Atmos.*, 121, 3326–3338, 2016.
- Pryor, S. C., Sullivan, R. C., and Wright, T.: Quantifying the roles of changing albedo, emissivity, and energy partitioning in the impact of irrigation on atmospheric heat content, *J. Appl. Meteorol. Clim.*, 55, 1699–1706, 2016.
- Reichle, R. H., Draper, C. S., Liu, Q., Griotto, M., Mahanama, S. P. P., Hoster, R. D., and Lannoy, G. J. M.: Assessment of MERRA-2 land surface hydrology estimates, *J. Climate*, 30, 2937–2960, 2017a.
- Reichle, R. H., Liu, Q., Koster, R. D., Draper, C. S., Mahanama, S. P., and Partyka, G. S.: Land surface precipitation in MERRA-2, *J. Climate*, 30, 1643–1664, 2017b.

- Ribera, P., Gallego, D., Gimeno, L., Perez-Campos, J., García-Herrera, R., Hernández, E., de La Torre, L., Nieto, R., and Calvo, N.: The use of equivalent temperature to analyse climate variability, *Stud. Geophys. Geod.*, 48, 459–468, 2004.
- Richman, M. B.: Rotation of principal components, *J. Climatol.*, 6, 293–335, 1986.
- Rogers, J. C., Wang, S. H., and Coleman, J. S. M.: Evaluation of a long-term (1882–2005) equivalent temperature time series, *J. Climate*, 20, 4476–4485, 2007.
- Sanderson, M. G. and Ford, G. P.: Projections of severe heat waves in the United Kingdom, *Clim. Res.*, 71, 63–73, <https://doi.org/10.3354/cr01428>, 2017.
- Schoof, J. T., Heern, Z. A., Therrell, M. D., and Remo, J. W. F.: Assessing trends in lower tropospheric heat content in the central United States using equivalent temperature, *Int. J. Climatol.*, 35, 2828–2836, 2015.
- Schoof, J. T., Ford, T., and Pryor, S. C.: Recent changes in United States heat wave characteristics derived from multiple reanalyses, *J. Appl. Meteorol. Clim.*, 56, 2621–2636, 2017.
- Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A. J.: Investigating soil moisture–climate interactions in a changing climate: A review, *Earth-Sci. Rev.*, 99, 125–161, 2010.
- Sheridan, S. C.: North American weather-type frequency and teleconnection indices, *Int. J. Climatol.*, 23, 27–45, 2003.
- Sheridan, S. C. and Kalkstein, L. S.: Progress in heat watch–warning system technology, *B. Am. Meteorol. Soc.*, 85, 1931–1941, 2004.
- Sheridan, S. C., Kalkstein, A. J., and Kalkstein, L. S.: Trends in heat-related mortality in the United States, 1975–2004, *Nat. Hazards*, 50, 145–160, 2009.
- Teng, H., Branstator, G., Wang, H., Meehl, G. A., and Washington, W. M.: Probability of US heat waves affected by a subseasonal planetary wave pattern, *Nat. Geosci.*, 6, 1056–1061, 2013.
- Toth, E., Brath, A., and Montanari, A.: Comparison of short-term rainfall prediction models for real-time flood forecasting, *J. Hydrol.*, 239, 132–147, 2000.
- Trenberth, K. E.: Recent observed interdecadal climate changes in the Northern Hemisphere, *B. Am. Meteorol. Soc.*, 71, 988–993, 1990.
- Vanos, J., Kalkstein, L., and Sanford, T.: Detecting synoptic warming trends across the US Midwest and implications to human health and heat-related mortality, *Int. J. Climatol.*, 35, 85–96, 2015.
- Weaver, S. J.: Factors associated with decadal variability in Great Plains summertime surface temperatures, *J. Climate*, 26, 343–350, <https://doi.org/10.1175/jcli-d-11-00713.1>, 2013.
- Weisheimer, A., Doblas-Reyes, F. J., Jung, T., and Palmer, T.: On the predictability of the extreme summer 2003 over Europe, *Geophys. Res. Lett.*, 38, L05704, <https://doi.org/10.1029/2010GL046455>, 2011.
- Wilks, D. S.: *Statistical Methods in the Atmospheric Sciences*, International Geophysics Series, Academic Press, Oxford, UK, 2011.
- Willett, K. M., Gillett, N. P., Jones, P. D., and Thorne, P. W.: Attribution of observed surface humidity changes to human influence, *Nature*, 449, 710–712, 2007.
- Yu, S., Alapaty, K., Mathur, R., Pleim, J., Zhang, Y., Nolte, C., Eder, B., Foley, K., and Nagashima, T.: Attribution of the United States “warming hole”: Aerosol indirect effect and precipitable water vapor, *Sci. Rep. UK*, 4, 6929, <https://doi.org/10.1038/srep06929>, 2014.
- Zhang, K., Li, Y., and Schwartz, J. D.: What weather variables are important in predicting heat-related mortality? A new application of statistical learning methods, *Environ. Res.*, 132, 350–359, 2014.