

Response to:

Interactive comment on “Modeling the contributions of global air temperature, synoptic-scale phenomena and soil moisture to near-surface static energy variability using artificial neural networks” by Sara C. Pryor et al.

Anonymous Referee #2 Received and published: 30 July 2017

Below we list the comments of the reviewer and our responses (in italics below each point). At the end of this document we also provide a tracked changes version of the manuscript showing the changes we have made in full.

...<Preamble deleted for brevity> ...

I am providing below a number of minor issues that should be resolved prior to publication.

Abstract: “... measure of static energy... are more strongly linked to excess human mortality and morbidity than air temperature alone”- while this statement provides sound motivation for the study, the two are only weakly connected in the article’s discussion and not at all connected in the results. Connecting this idea back in for the Results and Conclusion would be desirable

*Response: Done. We have extensively modified section 4 to include the requested linkages back to human health.*

Ln 16 model[s]

*Response: Done*

Ln17 the “drivers”, should be explicitly named in the abstract: global T2m, synoptic T Q at 850 and 500 hPa geopotential, and SM all taken from MERRA2.

*Response: Done. Added text: The predictor variables include an index of near-surface daily global mean temperature, daily indices of the synoptic scale meteorology derived from T and specific humidity (Q) at 850 hPa and 500 hPa geopotential heights (Z), and spatiotemporally averaged soil moisture (SM).*

Ln 25 “Over the [eastern U.S.], the ANN...alarm rate [is] ~0.08.”

*Response: Done*

Main Body Pg1Ln30 “elaborated [on] the drivers...health [and socioeconomic impacts](Sanderson...”

*Response: Done*

Pg2Ln1 “associated with annual [increased] welfare losses of ...”- clarify that \$57 billion is annual increase due to warming, not annual total.

*Response: We do not refer to welfare losses of \$57 billion, so no clarification needed.*

Pg2Ln3 suggest deleting “Many”

*Response: Done*

Eq1. There should be references to Bolton, 1980; Bryan, 2008; and/or Davies-Jones, 2009. Theta-e should be defined in words here, as well. e.g., “potential temperature plus the temperature increase that would be caused by latent heat of saturation of water contained in air”. Suggest starting from potential temperature definition first, then moving to equivalent potential temperature.

*Response: Reference added along with following text; ‘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. Use of metrics such as  $\theta_e$  permits more accurate depictions of near-surface energy budgets and surface heating trends (Davey et al., 2006)..’*

Pg2ln14-15 units for each variable need to be added here.

*Response: Done*

Pg2ln16 should read [g/g], not g/kg in Q

*Response: Can not find what the reviewer was referring to we did not state units of g/kg, but addressed by the above.*

Pg2ln18 isn’t T responsive to advection-driven forcing, as well?

*Response: Yes. We have reworded this to clarify that here we are referring to the local surface energy balance prior to describing non-local processes in the following paragraphs. The rewrite reads; ‘When considering the local surface energy balance (Eq. 2), near-surface T is responsive only to changes in the sensible heat flux from/to the surface, while  $\theta_e$  responds to changes in both the sensible and latent heat flux terms.’*

Pg2ln22 “[surface broadband] albedo”. What is the meaning of “counter-radiated”? Is it “surface downward”?

*Response: Yes, “counter-radiated longwave radiation from the atmosphere” is ‘surface downward’ but I believe ‘counter-radiated longwave radiation from the atmosphere’ is in common usage in atmospheric science and thus should be clear to the readership of ACP.*

Pg2ln27 “although [heat extremes] such as ... (Garcia-Herrera et al., 2010)...

(Vanos et al., 2015)”. Other references are required here to strengthen assertion being made. Vanos et al., 2015 covers Midwest only. References list needs to be expanded to cover SE, NE U.S. cases, as well.

*Response: The following paper that we also cite does cover cities drawn from the entire contiguous USA: 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 U.S. communities, *Environmental Health Perspectives*, 119, 210-218, 2011.*

Pg2ln31 is Peterson et al., 2011 reference for 2m static energy? Please clarify level to which “lower atmosphere” refers.

*Response: This is an excellent point. Peterson et al. 2011 refer to ‘surface atmospheric energy ‘ (in the title of their article) but do not specify the height they draw T, Q and wind speeds from in order to compute the components of the atmospheric energy budget. I suspect (based on NWS protocols in the USA) that T and Q are LIKELY to be drawn from 2-m, but wind speeds are more likely to be taken at 10-m. However, the reviewer is correct, the lack of specificity on the part of those authors led me to state “lower atmosphere” in the absence of more specific information. I have changed this to near-surface to be consistent with the authors to whose work we are referring.*

Pg3ln2 Davey et al., 2006 reference expected after Eq. 1

*Response: Done*

Pg3ln3 statistical models are ill-suited for pursuits of physical, process-level understanding

*Response: Respectfully, this is only partly true. Statistical relationships MAY not be reflective of an underlying physical process. HOWEVER, the fundamental basis of all big-data analytics and deep-learning (statistical) data analysis methods is that one can extract high-level, complex abstractions as data representations through a hierarchical learning process. This implies there are repeatable associations that are manifesting causal mechanisms. Nevertheless, we have removed the word ‘mechanistic’ from ‘enhance mechanistic understanding’.*

Pg3ln6 suggest using the more common equivalent term of “summary”

*Response: I can not find what the reviewer was referring to ... Pg3 line 6 read; ‘the summer of 2012 in part due to persistent anticyclonic conditions (Peterson et al., 2013). Further, variability’*

Pg3ln34 “extreme [high] Td”

*Response: Done*

Pg3ln36 “High Plains to [] the upper Great Lakes”. Restricted vertical mixing = subsidence?

*Response: This is an interesting point. The authors of the article to which we are referring actually DID NOT evaluate subsidence but do report ‘restricted vertical mixing’. Thus we have maintained their statement (restricted vertical mixing) and not inferred information about an increase in subsidence.*

Pg4lns2-4 scale length for (b) and (c) need also to be specified as it is for (a) (i.e., global).

*Response: Done*

Pg4ln11 technically, Canada should be masked out of Figure 1a. Domain lat-lon extents should be provided here, as on pg5ln16

*Response: Although our primary focus is on the eastern USA and perhaps we should mask out Canada, we prefer to include grid cells within Canada since there have been*

*heat-wave incidents in Canada (see (Smoyer-Tomic et al., 2003)). I have added the domain extent to the caption to Figure 1.*

Pg4ln12 “trends in  $T_e$ ” please specify over what period these trends were computed.  
*Response: Done (its 1981-2015).*

Pg4ln15 could you provide an estimate of the affected population residing in the eastern U.S. domain?

*Response: the eastern US (as defined in our study) has a population of > 200 million (we have added this to the text; ‘It is home to over 200 million people (based on the 2013 census).’*

Pg4ln19 sentence beginning “Therefore,..” is confusing and should be reworded.

*Response: Unfortunately there is no sentence on page 4 line 19 that begins ‘therefore’.*

Pg4ln21 it is confusing to bundle land management and SM rates of change in the same sentence while they occur on very different time scales. Clarify relative rate of change and period of change.

*Response: We believe the reviewer is referring to ‘It exhibits strong spatial gradients in terms of the nature of land cover and rate of change of both land management and soil moisture.’ They are quite correct, these are complex processes. We have added the following: ‘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).’*

Pg4ln23 specify the “Parts of the region” that are being referenced. Is this the southern Great Plains? There is no reason to be non-specific here. The GLACE hotspot and two regions of reduced  $T_{max}$  should be demarcated on Fig 1a, which could be enlarged.

*Response: At least to our knowledge the two references we cite not provide precise lat/long coordinates for the regions of strong coupling (though they are shown on maps). We note the reviewer did state (later in their review); ‘The central U.S. hotspot of GLACE-1 was squarely contained within 27-41N; 106-97W’... In order to address this concern we have elaborated in the text so that it now reads: “Parts of the region (focused on the southern Great Plains, i.e. the southern and western-most portion of the current study area) were identified in the global land–atmosphere coupling experiments (GLACE) as exhibiting atypically strong atmosphere–surface coupling in some global climate models. Further, soil moisture makes a large contribution to sub-seasonal forecast skill for air temperatures and precipitation in this region (Koster et al., 2011;Koster et al., 2006).”*

Pg4ln26 Guo et al., 2006 and Dirmeyer and Halder (2017) should be added to reference list.

*Response: Done*

Pg4ln27 “The study region”. The preceding three bullets referred to the study region as “it”. Why the change here? Also regarding: “maximum T during parts or all of the twentieth century”, it is unclear whether “parts” in this context alludes to seasons, years, consecutive years, etc. This should be clarified.

*Response: Each bullet now started with it. We have clarified; ‘It also incorporates two areas of reduced daily maximum T during multiple consecutive years extending over part or all of the twentieth century.’*

Pg4ln29 “lack of warming” The warming hole label was included in the abstract and this is probably the appropriate place to introduce the term in the article.

*Response: Done*

Pg4ln31 suggest new sentence begins with “... Ellenburg et al., 2016). In the case of Mississippi ... (AL), [up to ]...”

*Response: Done*

Pg4ln32 specify whether “summer temperature” is JJA T, JJA Tmax, JJA Tmin, etc. This section should be more carefully worded and details added to improve clarity of meaning.

*Response: Where hourly data are used it is specified (e.g. for T, Q and P).*

Pg5ln7 consider adding a note that MERRA-2 uses bias-corrected P to drive the land surface model, which lends strong confidence to the SM estimates. The skill of MERRA-2 lower-atmospheric fields used in the synoptic airflow classification is yet to be well established.

*Response: Yes, excellent suggestion. Done. In terms of the evaluation of MERRA-2 – this is clearly NOT our product but NASA scientists have been extensively evaluating the reanalysis system and output. What we can assert is that the output from our PCA analysis of the synoptic scale meteorology generates reasonable and interpretable types. This is why we wrote (and write); ‘As shown, many of the synoptic types thus identified are readily interpretable as representing a diversity of zonal versus meridional circulation (cf. type 9 and 5 and 6), and some are characterized by conditions known to be associated with strong low-level advection of 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)).’ Naturally, this does not constitute a comprehensive evaluation!*

Pg5ln16 suggest adding 12 EDT /2 CDT after 20 UTC.

*Response: We have added a note regarding time zones. Thus the text now read; ‘for 20:00 UTC (i.e. 16:00 Eastern Daylight Time, 15:00 Central Daylight Time)’. Please note: Eastern Daylight Time (EDT) is UTC MINUS 4 hr, not 8 hours as the reviewer wrote (maybe this is a typographic error).*

Pg5ln30 “strong [southerly] low-level advection of [high] T and Q into the region”?

*Response: Done*

Pg5ln36 “[Due to its spatial heterogeneity,] soil moisture is...”

*Response: Done*

Pg6ln3 SM plays much less of a role in the radiation-limited temperate NE forests.

*Response: Yes.*

Pg6ln6 I would not agree that MERRA-2 SM has been “extensively” evaluated- it is a very new product and only one (biased) reference is provided

*Response: We have removed ‘extensively’.*

Pg6ln8 it should be specified that validation occurs only over non-forested sites, where in-situ SM is available.

*Response: Very good point – well made. We have added a caveat that reads; ‘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.’*

Pg6ln18 was the time difference (2PM EDT, 1PM CDT) accounted for in the comparison? If so, how?

*Response: No.*

Pg6ln19 is the GPLLJ the only source of moisture for MO and IA? Sentence is currently misleading.

*Response: Fair point we have add ‘in part’ to clarify.*

Pg6ln21 “and the presence of abundant SM”. Intent of statement is unclear. SM is abundant in the eastern states, as well.

*Response: This is a very interesting point. There are relatively few homogeneous SM data sets. IF (and I would emphasize, if) one treats remote sensing SM estimates as equally valid over the entire study area then for our study region SM is most abundant in the region I was referring to (lower GP) and in the NE, but is less abundant in the eastern states. See figure below:*

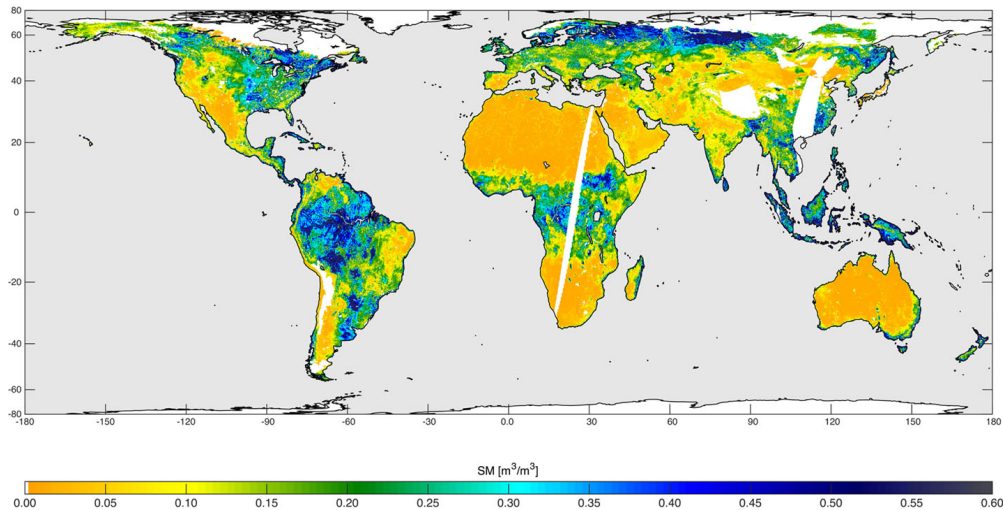


Image from: <https://www.jpl.nasa.gov/spaceimages/details.php?id=PIA19337>. Caption: High-resolution global soil moisture map from SMAP's combined radar and radiometer instruments, acquired between May 4 and May 11, 2015 during SMAP's commissioning phase. The map has a resolution of 5.6 miles (9 kilometers).

We have added a reference to (Al Bitar et al., 2017) in support of the assertion regarding abundant SM so we are not solely relying on the MERRA-2 product (and our figure 1).

Pg6ln21 suggest list item (5) should be added, beginning from “There are also Important ...”

Response: Done

Pg6ln25-30 suggest moving these stats out of the data section and into Results.

Response: We decided against doing so.

Pg6ln30 “over [the] southern ...”

Response: Done

Pg7ln8 please include commentary on how these predictors were selected. What was the logic or metrics employed in the selection process?

Response: Done (but we do so where the predictors are first introduced – i.e. in the bulleted list in section 2.2 not where the reviewer suggested).

Pg7ln24 “readily available to [soil evaporation, as opposed to integrated soil moisture profile that constitutes the water availability to evapotranspiration]”.

Response: We have modified the sentence to read; ‘Therefore, in addition to developing models using the MERRA-2 variable ‘PRMC’, which is the ‘Total profile soil moisture content’ in  $m^3m^{-3}$  (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 3 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., 2017), and thus best represents the SM that is readily available for evaporation into the overlying atmosphere.'*

Pg7ln25 suggest "Table 1 summarizes the acronyms used herein for each of the five models considered. A schematic of the model architecture and data flows is provided in Fig. 3.

*Response: Done*

Pg7ln27 suggest including statement that the 70-15-15 sample subsets were identical for all model architectures.

*Response: Done*

Pg8ln17 this statement appears at odds with Table 1, which lists ANN-HL3 as having the worst model statistics.

*Response: I believe this is a misunderstanding ANN-HL3 excludes SM and indeed has highest RMSE.*

Pg9ln2 is there any statistical significance to this difference?

*Response: The reviewer is raising an interesting point – can one undertake an appropriate non-parametric test, correcting (of course) for multiplicity. We did not, instead choosing to draw out spatial consistency as a basis for inferring 'skill'.*

Pg9ln6 "where [modeled] land-atmosphere ... and where strong [longitudinal] gradients of SM"

*Response: Done*

Ph9ln13 "such [as] dry lines ..."

*Response: Done*

Pg9ln24 I question "evaporation from the Great Lakes". Is Great Lakes evaporation a strong predictor of theta-e in the eastern U.S.? Evaporation over the Great Lakes is much more substantial in the winter months and drives tremendous lake-effect snow bands, but the affected region is isolated.

*Response: Interesting point. As a resident of upstate New York I can confirm the lake-effect snows are a strong function of ice cover (and hence evaporation) from the Great Lakes! However, observationally-derived estimates of evaporation from the Great Lakes reported in (Notaro et al., 2015) are certainly 'non-trivial' (i.e. approx. an average of 45 mm in August, relative to over-lake precipitation receipt of approx. 85 mm for the period 1980-1999..). But it is an open question as to the impact on downstream theta-e, so we have added 'potentially'.*

Pg9ln30 how is water management signal being linked to the present analysis? MERRA2 has no accounting of water management.

*Response: Right, its not.*



Pg9ln32 suggest DeAngelis et al. 2010 in addition to Pryor et al., 2016 reference for warming hole

*Response: Done*

Pg10ln5 HL3-SM seems to perform the best; 3 hidden layers as opposed to 1.

*Response: Correct*

Pg10ln7 "...when all [eastern U.S.] grid cells are considered..."

*Response: Done*

Pg10ln12 please clarify that "test period" here is *synonymous with "independent sample"*, or 15% of all JJA days

*Response: Reworded to independent data sample.*

Pg10lns5-36. Why can't these stats on HR and FAR be added to an expanded Table 1 for each model?

*Response: This is a style consideration. Table 1 is designed to summarize the overall performance of the models (and is already quite complex), the HR/FAR naturally reflect the performance for extremes.*

Pg10ln22 statement on CLM (Buzan et al., 2015) does not appear directly relevant to MERRA-specific results being discussed here. It is out of place, given non-MERRA results are not specifically called out elsewhere in the paper. I suggest removing this statement.

*Response: It is POTENTIALLY interesting at least to some readers to note that; '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 99<sup>th</sup> percentile  $\theta_e$  in an analysis of heat indices derived from the Community Land Model v4.5 (Buzan et al., 2015).'*

Pg10ln26 To include MO, IA and IL is a stretch. The central U.S. hotspot of GLACE-1 was squarely contained within 27-41N; 106-97W.

*Response: Right, the other references we cite do focus more on the eastern states. We have reworded to; 'central and eastern USA'*

Pg10ln32 please quantify "greatly" Is this statistically significant?

*Response: Fair point 'greatly' is subjective. I have deleted this word.*

Pg11ln1 "HR and FAR are comparable to (or better than) seasonal re-forecasts of summertime T at 2-m..." Please clarify were the HR and FAR of T or theta-e-max,min compared against the HR and FAR of T2m from ECMWF (ERA-I)?

*Response: Yes. I have reworded to; 'Further, HR and FAR computed for max- $\theta_e$  and min- $\theta_e$  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).'*

Pg11In15 the comparison between HL3-SM and HL3-TOP needs to be included and thoroughly discussed, esp. for theta-e-max.

*Response: There is a discussion which reads; 'Differences in model performance between ANN 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  $\Delta|RMSE| > 0.5 K$  for models using PRMC versus 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.'*

Pg11In32 "from the [MERRA-2]..."

*Response: Done*

Pg11In33 suggest "important differences in the magnitude of derived equivalent temperature ( $T_e$ ), as well as in strength of land-atmosphere coupling between the reanalysis products (e.g., Ferguson et al., 2012; Schoof et al., 2017)]"

*Response: Done*

Pg12In9 sentence beginning "Correlation coefficients exceed..." needs to be reworded and probably split into two or more sentences.

*Response: Done*

Pg12In11 sentence beginning "This is true for the simulation..." is confusing. Does this imply that the prior sentence holds exactly for max theta e?

*Response: Reworded for clarity to ; 'ANN-HL3-SM models also exhibit highest skill for simulation of extreme min- and max- $\theta_e$ .'*

Pg12In15 the statement that "min theta-e exhibits a stronger dependence on the precise prevailing synoptic scale conditions" is unsupported by the analysis and accompanying discussion. Where has the link between min theta-e predictability and PC number been established and probed?

*Response: Very good point. We have re-written this sentence to read; 'Results from the ANN models further indicate that max- $\theta_e$  and the occurrence of extreme max- $\theta_e$  appear to be considerably more sensitive to SM than min- $\theta_e$  which in turn appears to exhibit a stronger dependence on the precise prevailing synoptic scale conditions based on the ANN weights.' To reflect the assessment I made of the weights.*

Pg12In28 what are the alternative hypotheses in literature for the 'warming hole'? references here or previously upon the introduction of the term would be useful.

*Response: Done*

Pg12ln30 I disagree that the statistical modeling exercise documented herein has “enhance[d] mechanistic understanding of the causes of variability and change in  $\theta_e$ ”. This claim should be deleted.

*Response: Deleted mechanistic.*

Pg12ln35. The authors should comment on lessons learned and insight gleaned. For example, the weighting for the global T was stated to be *negligible*. Then, is their recommendation to forgo global T requirements in future ANN pursuits? Similarly, is the total integrated column soil moisture necessary or merely the surface layer? Does the fact that MERRA2 uses bias-corrected P make any allusions of similar success in full forecast models (with biased, model P) unfair and misleading? In addition to those predictors tested here, which other predictors would be meaningful to explore/consider?

*Response: Well what we actually state is; ‘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.’ So no we would not suggest foregoing global T.*

*We have elaborated on ‘next steps’ and the implications of our work by adding text to the conclusions. Section 4 now read;*

*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 spatio-temporal 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 min- and max- $\theta_e$  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 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 sub-regions 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. (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 min- and max- $\theta_e$  is smallest in the more complex models (i.e. for ANN-HL3-SM, RMSE < 4 K and < 4.3 K, respectively, c.f. mean max- $\theta_e \approx 333$  K and mean min- $\theta_e \approx 321$  K).*

*Particularly in regions with high variability in min- and max- $\theta_e$  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 max- $\theta_e$  and 81% for min- $\theta_e$ . Further, 92% of grid cells for ANN-HL3-SM exhibit a RMSE < 5 K for max- $\theta_e$  and 91% for min- $\theta_e$ .*

*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 at the 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 performance of such systems may be aided by implementation of other variables/analysis methodologies such as those used herein. The ANN-HL3-SM models developed herein exhibit relative high skill in predicting the occurrence of extreme min- and max- $\theta_e$ , and indeed out-perform the simpler models. The ANN with 3 hidden layers and that includes SM as a predictor (i.e. ANN-HL3-SM) exhibits a domain averaged median hit rate for max- $\theta_e > 347$  K is > 0.60, while the median FAR is  $\approx 0.08$ . Results from the ANN models further indicate that max- $\theta_e$  and the occurrence of extreme max- $\theta_e$  appear to be considerably more sensitive to SM than min- $\theta_e$  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 max- $\theta_e$  or min- $\theta_e$  (Figure 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 greatest importance in sub-regions previously noted as exhibiting 'warming holes'. Our framework has greater skill for max- $\theta_e$  than min- $\theta_e$ . It is possible that inclusion of additional predictors could lead to enhanced model skill particularly for*

*extreme high values of  $\max-\theta_e$  or  $\min-\theta_e$  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 (SM) 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., 2011;Koster et al., 2006;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 reanalysis products).*

Table 1. Caption should specify over JJA and 1980-2014; “coefficient [(r)]”, “the presence [or] absence”; the total number of grids (1962) should be specified OR the stats should be presented as percentages to be consistent with the text. Column with  $r > 0.8$  and RMSE  $< 5k$ : what about the number of grids that satisfy BOTH criteria?

*Response: Done and done (new data added to Table 1, and a note added to the text that reads; ‘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  $\max-\theta_e$  and  $\min-\theta_e$ . (Table 1).’).*

Fig 1. For (c-g) it is unclear from the caption whether 1980-2014 is still the averaging period; “standard deviation of daily [JJA] (e)...”; Mean [JJA daily] soil moisture...”; explained computation would be more appropriate to include in the main text of article. Is SM also averaged over 1980-2014? ; “total profile [0-1m? profile total thickness]”; the variable names and units should be displayed on each subpanel. Is the 90-day running JJA just June1-Aug31 average or on June 1 the ~March1-June1 average and so-on? Please clarify.

*Response: Clarifications added. In the text we have added the following details about 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 3x3 grid cells centered on the grid cell in question (i.e. integrated over an area of approximately 30,000 km<sup>2</sup>, see Figure 1g).’ This is already a relatively compact figure so have maintained the units in the caption.*

Fig3. The 5 models from Table 1 should be alluded to here by their acronyms in that table; the domain of actions on the right hand side should be specified as for the

predictors. From the flowchart it appears that z-scores are fed into the ANN. From the text I understood that daily z-scores are used to compute daily PC's which are passed to the ANN. This Figure could be eliminated by adding a data table, which demarcates the predictors and predictands, and describing the rest in text. Currently the flowchart is a bit unclear.

*Response: I've remade the figure to define the acronyms and clarify the data flows.*

Fig 4. ANN-HL3-TOP is missing and needs to be included. The labels and units for rows 2 and 3 should be added to the far right side.

*Response: ANN-HL3-TOP is qualitatively so similar to ANN-HL3-SM that in order to enhance the legibility of the figure I excluded it. I have added a note to the caption to reflect this. This is already a relatively compact figure so have maintained the units in the caption.*

References (Bolton 1980; Bryan 2008; Davies-Jones 2009; DeAngelis et al. 2010; Dirmeyer and Halder 2017; Ferguson et al. 2012; Guo et al. 2006)

...<References supplied by the reviewer deleted for brevity> ...

References cited in this response

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 System Science Data*, 9, 293-315, 10.5194/essd-9-293-2017, 2017.

Brown, D. G., Johnson, K. M., Loveland, T. R., and Theobald, D. M.: Rural land-use trends in the conterminous United States, 1950-2000, *Ecological Applications*, 15, 1851-1863, 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, *Geoscientific Model Development*, 8, 151-170, 10.5194/gmd-8-151-2015, 2015.

Davey, C. A., Pielke, R. A., Sr., 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 and Planetary Change*, 54, 19-32, 10.1016/j.gloplacha.2005.11.002, 2006.

Diem, J. E.: Synoptic-scale controls of summer precipitation in the Southeastern United States, *Journal of Climate*, 19, 613-621, 10.1175/jcli3645.1, 2006.

Ferguson, C. R., Wood, E. F., and Vinukollu, R. K.: A Global Intercomparison of Modeled and Observed Land-Atmosphere Coupling, *Journal of Hydrometeorology*, 13, 749-784, 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, *Journal of Hydrometeorology*, 17, 2713-2726, 2016.

Ford, T. W., and Schoof, J. T.: Characterizing extreme and oppressive heat waves in Illinois, *Journal of Geophysical Research: Atmospheres*, 122, 682-698, 2017.

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, *Journal of Hydrometeorology*, 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, *Journal of Hydrometeorology*, 12, 805-822, 2011.

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, *Nature Geoscience*, 9, 389-394, 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, *Geophysical Research Letters*, 42, 5564-5570, 10.1002/2015gl064586, 2015.

Notaro, M., Bennington, V., and Lofgren, B.: Dynamical Downscaling-Based Projections of Great Lakes Water Levels, *Journal of Climate*, 28, 9721-9745, 10.1175/jcli-d-14-00847.1, 2015.

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., Hoerling, 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, *Bulletin of the American Meteorological Society*, 94, S1-S74, 10.1175/bams-d-13-00085.1, 2013.

Pryor, S. C., and Schoof, J. T.: Evaluation of near - surface temperature, humidity, and equivalent temperature from regional climate models applied in type II downscaling, *Journal of Geophysical Research: Atmospheres*, 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, *Journal of Applied Meteorology and Climatology*, 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, *Journal of Climate*, 30, 2937-2960, 2017.

Schoof, J. T., Ford, T., and Pryor, S. C.: Recent changes in United States heat wave characteristics derived from multiple reanalyses, *Journal of Applied Meteorology and Climatology*, In review, 2017.

Sheridan, S. C., and Kalkstein, L. S.: Progress in heat watch–warning system technology, *Bulletin of the American Meteorological Society*, 85, 1931-1941, 2004.

Smoyer-Tomic, K. E., Kuhn, R., and Hudson, A.: Heat wave hazards: an overview of heat wave impacts in Canada, *Natural hazards*, 28, 465-486, 2003.

Weaver, S. J.: Factors associated with decadal variability in Great Plains summertime surface temperatures, *Journal of Climate*, 26, 343-350, 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, *Geophysical Research Letters*, 38, doi: 10.1029/2010GL046455, 2011.



# 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>

5 <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: S.C. Pryor ([sp2279@cornell.edu](mailto:sp2279@cornell.edu))

**Abstract.** The static energy content of the atmosphere is increasing at the global scale, but exhibits important sub-global and sub-regional 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 hPa 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 (ANN) out-perform linear models that do not permit interaction of the predictor variables (global  $T$ , synoptic-scale meteorological conditions and  $SM$ ). This is particularly marked in regions with high variability in min- and max- $\theta_e$ , 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 max- $\theta_e$ . Over the entire domain the ANN with 3 hidden layers exhibits high accuracy in predicting max- $\theta_e > 347$  K. The median hit rate for max- $\theta_e > 347$  K is  $> 0.60$ , while the median false alarm rate ~~is~~  $\approx 0.08$ .

## 30 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

Prof. Sara C. Pryor 10/30/2017 11:17 AM  
**Deleted:** A n

Prof. Sara C. Pryor 10/30/2017 11:17 AM  
**Deleted:** is

Prof. Sara C. Pryor 10/30/2017 11:27 AM  
**Formatted:** Font:Italic

Prof. Sara C. Pryor 10/30/2017 11:27 AM  
**Formatted:** Font:Italic

Prof. Sara C. Pryor 10/30/2017 11:28 AM  
**Deleted:** )

Prof. Sara C. Pryor 10/30/2017 11:29 AM  
**Formatted:** Font:Italic

Prof. Sara C. Pryor 10/30/2017 11:26 AM  
**Formatted:** Font:Italic

Prof. Sara C. Pryor 11/1/2017 9:11 AM  
**Deleted:** It is shown that soil moisture (

human health impacts and socioeconomic consequences (Sanderson and Ford, 2017; deDonato 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 ( $f(T)$ ) ( $\text{J Kg}^{-1}$ ),  $q$  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, 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 uses 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/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 sub-regional 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

Prof. Sara C. Pryor 10/30/2017 11:33 AM  
Deleted: Many p

Prof. Sara C. Pryor 10/30/2017 11:50 AM  
Formatted: Superscript

Prof. Sara C. Pryor 10/30/2017 11:38 AM  
Formatted: Widow/Orphan control

Prof. Sara C. Pryor 10/31/2017 11:30 AM  
Formatted: Font:(Default) Times New Roman, 10 pt, Not Bold, Font color: Auto,

Prof. Sara C. Pryor 10/31/2017 11:30 AM  
Formatted: Font:

Prof. Sara C. Pryor 10/31/2017 11:30 AM  
Formatted: Check spelling and grammar

Prof. Sara C. Pryor 10/31/2017 11:30 AM  
Formatted: Font color: Auto

Prof. Sara C. Pryor 11/1/2017 9:15 AM  
Deleted: While

Prof. Sara C. Pryor 11/1/2017 9:16 AM  
Deleted: within the surface energy balance

Prof. Sara C. Pryor 10/31/2017 11:30 AM  
Formatted: Font:(Default) Times

Prof. Sara C. Pryor 11/1/2017 9:15 AM  
Deleted: T

Prof. Sara C. Pryor 10/30/2017 12:21 PM  
Deleted: global lower

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 (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-warning 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, 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, extreme 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 through 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:

- 1) To use non-linear (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 at the 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 (Figure 1a) is the focus of this research because:

- It is home to over 200 million people (based on the 2013 census).

Prof. Sara C. Pryor 11/1/2017 11:24 AM  
Deleted: .

Prof. Sara C. Pryor 10/30/2017 12:35 PM  
Deleted: mechanistic

- 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  $SM$  (Figure 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 western-most portion of the current study area) were identified in the global land–atmosphere coupling experiments (GLACE) as exhibiting atypically strong atmosphere–surface coupling in some global climate models. Further, soil moisture makes a large contribution to sub-seasonal forecast skill for air temperatures and precipitation in this region (Koster et al., 2011; Koster et al., 2006; 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 twentieth century (i.e. warming holes). One is located along the border of Iowa (IA)–Nebraska (NE)–South Dakota (SD) and one is centered on Mississippi (MS) and Alabama (AL) (locations shown in Figure 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 upto 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), and aerosol radiative forcing (Leibensperger et al., 2012; Yu et al., 2014).

Prof. Sara C. Pryor 10/30/2017 1:40 PM

Deleted:

Prof. Sara C. Pryor 10/30/2017 2:14 PM

Deleted: , and

Prof. Sara C. Pryor 10/30/2017 1:47 PM

Deleted: soil moisture

Prof. Sara C. Pryor 10/30/2017 1:47 PM

Formatted: Font:Italic

Prof. Sara C. Pryor 10/30/2017 2:14 PM

Deleted: the region

Prof. Sara C. Pryor 10/30/2017 2:06 PM

Deleted: and

Prof. Sara C. Pryor 10/30/2017 2:15 PM

Deleted: during parts

Prof. Sara C. Pryor 10/30/2017 2:16 PM

Deleted:

Prof. Sara C. Pryor 10/30/2017 2:17 PM

Deleted: , and i

Prof. Sara C. Pryor 11/1/2017 9:20 AM

Deleted: regions of slow or absent warming

## 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 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°× ~0.5° (longitude-by-latitude). We use MERRA-2 output for all summer days (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 (Figure 1b) is intended to represent the long-term tendency in global mean temperatures and interannual/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 (25.5-50°N, 97.5-65°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 variability. 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. Fifteen 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 coincide with approximately 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 seven 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 versus meridional circulation (cf. type 9 and 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 ( $SM$ ). 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 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 km<sup>2</sup>, see Figure 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 m<sup>3</sup>m<sup>-3</sup> and a variance explanation ( $R^2$ ) 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

Prof. Sara C. Pryor 10/30/2017 2:30 PM

Deleted: F

Prof. Sara C. Pryor 10/30/2017 6:26 PM

Deleted: in

Prof. Sara C. Pryor 10/30/2017 4:49 PM

Deleted: extensively

Prof. Sara C. Pryor 10/30/2017 12:45 PM

Formatted: Superscript

(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.

Prof. Sara C. Pryor 10/30/2017 4:50 PM  
Formatted: Font:Italic

4) An estimate of (i) daily maximum  $\theta_e$  and (ii) daily minimum  $\theta_e$  in each grid cell (see Figure 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 accord with *a priori* expectations both daily max- $\theta_e$  and min- $\theta_e$  exhibit primarily latitudinal variability (Figure 1c and d). Mean max- $\theta_e$  decreases from values of  $\sim 350$  K ( $\sim 80^\circ\text{C}$ ) in the south of the domain to approximately 320 K ( $\sim 50^\circ\text{C}$ ) in the north. MO and IA exhibit anomalously high mean max- $\theta_e$  compared with grid cells at the same latitudes (Figure 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 (Figure 1g)(Al Bitar et al., 2017).

Prof. Sara C. Pryor 10/31/2017 12:35 PM  
Deleted: along with

5) There are also important spatial patterns of the day-to-day variability of daily max- $\theta_e$  and min- $\theta_e$  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; max- $\theta_e$  and min- $\theta_e$ ). The variance (and standard deviation) of daily max- $\theta_e$  values (i.e. a measure of the dispersion of individual days around the mean values of min- and max- $\theta_e$  shown in Figures 1c and 1d) is largest over IA/southern Minnesota (MN) (Figure 1e), while the variance of min- $\theta_e$  is greatest over Illinois (IL)(Figure 1f). In both cases the day-to-day variability as measured by the standard deviation is  $\sim 10$  K. The standard deviation of daily max- $\theta_e$  and min- $\theta_e$  decreases with decreasing latitude and both variables exhibit lowest variability over the southern portions of the study domain (e.g. Florida (FL) has a standard deviation  $< 5$  K) (Figure 1e,f).

Prof. Sara C. Pryor 10/30/2017 4:52 PM  
Deleted: ,

Prof. Sara C. Pryor 10/30/2017 5:10 PM  
Formatted: Font color: Auto, Pattern: Clear

Prof. Sara C. Pryor 10/31/2017 12:01 PM  
Formatted: Font:Italic

Prof. Sara C. Pryor 10/31/2017 12:01 PM  
Deleted:

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 max- $\theta_e$  and min- $\theta_e$ ) are grid-cell specific.

### 2.3 Methods

Artificial neural network (ANN) architectures are potentially highly useful in developing statistical models for response variables such as  $\theta_e$  because ANN do not require any assumptions about the form of the relationship between individual predictors, between predictors and predictands (min- and max- $\theta_e$ ), and can treat complex and non-linear 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 non-linear mapping between the input and output variables (Gardner and Dorling, 1998). The number of hidden layers within the ANN determines the degree of non-linearity 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, non-linear 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:

- 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 3 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’ in  $m^3 m^{-3}$  (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 3 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 Figure 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

Prof. Sara C. Pryor 10/30/2017 5:20 PM  
Deleted: and

Prof. Sara C. Pryor 10/30/2017 5:22 PM  
Deleted: , and a

models to capture the occurrence of very high  $\theta_e$ . In this analysis we set a threshold of 347 K (73°C) to indicate extreme max- $\theta_e$  (based on information provided in (Buzan et al., 2015) for the eastern USA) and a threshold of 337 K for min- $\theta_e$ . 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):

$$HR = \frac{\#hits}{\#hits + \#misses} \quad (3)$$

Where  $\#hits$  is number of days in the independent data set where the observations and predictions both indicate exceedance of the threshold, and  $\#misses$  is the number of days when the observations indicate an exceedance of the threshold but the model prediction does not.

And the false alarm rate ( $FAR$ ) of each individual grid cell:

$$FAR = \frac{\#false\ alarms}{\#false\ alarms + \#correct\ non-events} \quad (4)$$

Where  $\#false\ alarms$  is 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  $\#correct\ non-events$  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 Figures 4 and 5). Nevertheless, output from all model architectures for min- $\theta_e$  and max- $\theta_e$  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 applied to max- $\theta_e$  and 81% for min- $\theta_e$ . Further, 92% of grid cells for ANN-HL3-SM exhibit RMSE < 5 K for max- $\theta_e$  and 91% exhibit RMSE < 5 K for min- $\theta_e$  (Table 1, Figure 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 max- $\theta_e$  and min- $\theta_e$  (Table 1). Grid cells with highest RMSE for min- $\theta_e$  and max- $\theta_e$  also generally have highest variance (i.e. largest day-to-day variability). For example, grid cells in IA exhibit highest variance and highest RMSE for max- $\theta_e$  (Figure 1e and Figure 4), and grid-cells within IL are generally characterized by large RMSE and variance of min- $\theta_e$  (Figure 1f and Figure 5). Thus, while noting the RMSE (i.e. typical prediction error) is largest over IA for max- $\theta_e$  (~ 5 K), it is less than half the standard deviation computed from the day-to-day variability in max- $\theta_e$  (~ 10 K).

Generally all models exhibit slightly worse performance across both measures ( $r$  and RMSE) for min- $\theta_e$  than max- $\theta_e$  at the grid-cell level and integrated over all land grid cells (Table 1 and Figures 4 and 5). The reduced model skill for min- $\theta_e$  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 max- $\theta_e$ .

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

Prof. Sara C. Pryor 10/31/2017 11:50 AM  
 Deleted: and RMSE < 5 K (  
 Prof. Sara C. Pryor 10/31/2017 11:50 AM  
 Deleted: a  
 Prof. Sara C. Pryor 10/31/2017 11:50 AM  
 Deleted: RMSE below this threshold  
 Prof. Sara C. Pryor 10/31/2017 11:50 AM  
 Deleted: )



lower RMSE in the ANN model with 3 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 max- $\theta_e$  and min- $\theta_e$  is particularly marked in the west-central of the domain (over parts of Missouri (MO) and Iowa (IA), close to or within one of the ‘warming holes’) (Figure 4 and 5). This is a region where a substantial fraction of  $T$  variance is explained by thermal and moisture advection by the GPLJ (Weaver, 2013), where modeled land-atmosphere coupling is particularly intense (Koster et al., 2011;Koster et al., 2006) and where there are strong longitudinal gradients of  $SM$  (Figure 1g). Lowest correlations between predicted and observed min- and max- $\theta_e$  values occur over east Texas (TX) for all model formulations although the RMSE of model predictions are not particularly high in this area (Figure 4 and 5). The low RMSE may reflect the small day-to-day variability in min- and max- $\theta_e$  over this region (Figure 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 sub-region. 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 ANN 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 (SST), 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 3 hidden layers (i.e. ANN-HL3) decreases model performance relative to ANN-HL3-SM and increases the RMSE for max- $\theta_e$  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) (Figure 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 max- $\theta_e$  is increased in models excluding  $SM$  in all grid cells within MO, and all but one grid cell each in IA and IL (Figure 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

Prof. Sara C. Pryor 10/30/2017 5:28 PM  
Deleted: horizontal

Prof. Sara C. Pryor 10/30/2017 5:36 PM  
Deleted: evaporation  
Prof. Sara C. Pryor 11/1/2017 9:31 AM  
Formatted: Font:Italic

Prof. Sara C. Pryor 11/1/2017 11:39 AM  
Formatted: Font:Italic

maximum  $T$ , while increasing  $\max-\theta_e$ .

The statistical (downscaling) models show similar dynamic range to independent observations, although there is some evidence that the models underestimate the total variance in  $\max-\theta_e$  leading to underestimation of extreme  $\max-\theta_e$  as is evident from the flattening of the scatterplots for very high values of daily maximum  $\theta_e$  (see the upper

row of panels in Figure 4). To examine this further we conduct an analysis of the  $HR$  and  $FAR$  for  $\max-\theta_e$  in excess of 347 K. This threshold is exceeded by daily maximum  $\theta_e$  derived from the MERRA-2 reanalysis on an average of

~ 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 (Figure 6a). Conversely, it is seldom or never

observed within grid cells in the north of the domain (Figure 6a). To ensure a sufficiently robust sample size on which to compute the  $HR$  for extreme  $\max-\theta_e$  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 3-hidden layers (ANN-HL3-SM), and the ANN with 3-hidden layers but excluding SM (ANN-HL3) computed over all these grid cells is 59, 60 and 56%, respectively (Figure 6b-d), indicating that over the entire study domain the role of  $SM$  and predictor interactions in explaining the occurrence of extreme  $\max-\theta_e$  is modest. All model forms

perform least well in terms of predicting the occurrence of  $\max-\theta_e > 347$  K over eastern TX and South Carolina (SC) (Figure 6e-g). However, the model excluding  $SM$  exhibits 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 99<sup>th</sup> 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  $\max-\theta_e$  and substantial improvement in  $HR$  is

noted in IA, IL and MO (Figure 6e-g) in the ANN-HL3-SM relative to the other model forms. This is consistent with strong spatial gradients in  $SM$  (Figure 1e), findings of the GLACE projects of strong atmosphere-surface coupling (Koster et al., 2011; Koster et al., 2006), 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  $\min-\theta_e > 342$  K or  $\max-\theta_e > 347$  K are very modest for all model formulations (Figure 6 and 7). For example, over 94% of

grid cells indicate  $FAR$  for  $\max-\theta_e > 347$  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  $FAR$  (Figure 6 and 7). It should be acknowledged that even the ANN with 3 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  $\max-\theta_e$  and  $\min-\theta_e$  are comparable to (or better than) seasonal re-

Prof. Sara C. Pryor 10/30/2017 5:51 PM

Deleted: test period

Prof. Sara C. Pryor 11/1/2017 11:40 AM

Formatted: Font:Italic

Prof. Sara C. Pryor 10/30/2017 5:56 PM

Deleted: greatly

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 max- $\theta_e$  the model architecture has virtually no impact on  $HR$  for min- $\theta_e > 337$  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 (Figure 7). This finding is consistent with the overall results for models of min- $\theta_e$  that exhibit only modest decreases in model performance (increased RMSE and lower  $r$ ) when  $SM$  is excluded from the predictor suite (Table 1 and Figure 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 ANN 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  $\Delta|RMSE| > 0.5$  K for models using PRMC versus 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 spatio-temporal 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 min- and max- $\theta_e$  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 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

Prof. Sara C. Pryor 11/1/2017 11:42 AM  
Deleted: 342

Prof. Sara C. Pryor 10/30/2017 6:01 PM  
Formatted: Widow/Orphan control

Prof. Sara C. Pryor 10/30/2017 5:59 PM  
Deleted: EM

Prof. Sara C. Pryor 10/31/2017 12:04 PM  
Deleted: t

sub-regions 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. (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 min- and max- $\theta_e$  is smallest in the more complex models (i.e. for ANN-HL3-SM, RMSE < 4 K and < 4.3 K, respectively, c.f. mean max- $\theta_e \approx 333$  K and mean min- $\theta_e \approx 321$  K). Particularly in regions with high variability in min- and max- $\theta_e$  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 max- $\theta_e$  and 81% for min- $\theta_e$ . Further, 92% of grid cells for ANN-HL3-SM exhibit a RMSE < 5 K for max- $\theta_e$  and 91% for min- $\theta_e$ .

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 at the 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 performance of such systems may be aided by implementation of other variables/analysis methodologies such as those used herein. The ANN-HL3-SM models developed herein exhibit relative high skill in predicting the occurrence of extreme min- and max- $\theta_e$ , and indeed out-perform the simpler models. The ANN with 3 hidden layers and that includes SM as a predictor (i.e. ANN-HL3-SM) exhibits a domain averaged median hit rate for max- $\theta_e > 347$  K is > 0.60, while the median FAR is  $\approx 0.08$ . Results from the ANN models further indicate that max- $\theta_e$  and the occurrence of extreme max- $\theta_e$  appear to be considerably more sensitive to SM than min- $\theta_e$  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 max- $\theta_e$  or min- $\theta_e$  (Figure 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 greatest importance in sub-regions previously noted as exhibiting 'warming holes'. Our framework has greater skill for max- $\theta_e$  than min- $\theta_e$ . It is possible that inclusion of additional predictors could lead to enhanced model skill particularly for extreme high values of max- $\theta_e$  or min- $\theta_e$  that are of greatest importance to human health, and/or that

Prof. Sara C. Pryor 10/31/2017 12:32 PM  
**Formatted:** Font:Italic

Prof. Sara C. Pryor 10/31/2017 12:06 PM  
**Deleted:** and spatially consistent between reanalyses

Prof. Sara C. Pryor 10/31/2017 12:06 PM  
**Deleted:** and equally to determine i

Prof. Sara C. Pryor 10/30/2017 6:01 PM  
**Formatted:** Font:

Prof. Sara C. Pryor 10/31/2017 11:56 AM  
**Deleted:** an ANN with 3 hidden layers

Prof. Sara C. Pryor 10/30/2017 6:03 PM  
**Deleted:** and

Prof. Sara C. Pryor 10/30/2017 6:04 PM  
**Deleted:** This is also true for the simulation of

Prof. Sara C. Pryor 10/30/2017 6:04 PM  
**Deleted:** .

Prof. Sara C. Pryor 10/30/2017 6:08 PM  
**Deleted:** exhibits a

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 ( $SM$ ) 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., 2011;Koster et al., 2006;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 reanalysis products).

## 5 Data availability

The MERRA-2 data used herein are available from <http://disc.sci.gsfc.nasa.gov/daac-bin/FTPSubset2.pl> (see (Bosilovich et al., 2015) for a detailed description of the file structures).

## 6 Acknowledgments

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.

## 7 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 System Science Data, 9, 293-315, 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.

Prof. Sara C. Pryor 10/31/2017 12:21 PM

Formatted: Widow/Orphan control

Prof. Sara C. Pryor 10/31/2017 12:17 PM

Deleted: or that

Prof. Sara C. Pryor 10/31/2017 12:17 PM

Deleted: are a

Prof. Sara C. Pryor 10/31/2017 12:30 PM

Deleted:

Prof. Sara C. Pryor 11/1/2017 10:41 AM

Deleted: The primary purposes of this research are to enhance mechanistic 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, 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 at the synoptic scale over the eastern USA. -

- 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 U.S. communities, *Environmental Health Perspectives*, 119, 210-218, 2011.
- Bentley, M. L., and Stallins, J. A.: Synoptic evolution of Midwestern US extreme dew point events, *International Journal of Climatology*, 28, 1213-1225, 2008.
- Bolton, D.: The computation of equivalent potential temperature, *Monthly Weather Review*, 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 from [http://gmao.gsfc.nasa.gov/pubs/office\\_notes/](http://gmao.gsfc.nasa.gov/pubs/office_notes/), 2015.
- Brabson, B., Lister, D., Jones, P., and Palutikof, J.: Soil moisture and predicted spells of extreme temperatures in Britain, *Journal of Geophysical Research: Atmospheres*, 110, doi: 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, *Ecological Applications*, 15, 1851-1863, 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, *Geoscientific Model Development*, 8, 151-170, 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, Noah-MP, CLM, and VIC over CONUS using the NLDAS test bed, *Journal of Geophysical Research: Atmospheres*, 119, 13,751-713,770, 2014.
- Cattell, R. B.: The scree test for the number of factors, *Multivariate Behavioral Research*, 1, 245-276, 1966.
- Davey, C. A., Pielke, R. A., Sr., 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 and Planetary Change*, 54, 19-32, 10.1016/j.gloplacha.2005.11.002, 2006.
- 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., Gasparini, 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, 10.3390/ijerph121215006, 2015.
- 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, doi:15110.11029/12010JD01389, 10.1029/2010jd013892, 2010.
- Diem, J. E.: Synoptic-scale controls of summer precipitation in the Southeastern United States, *Journal of Climate*, 19, 613-621, 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), *Journal of Hydrometeorology*, 18, 85-108, 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 Interactions*, 20, Paper 20-018, doi: 010.1175/ei-d-1115-0038.1171, 2016.

- Fall, S., Diffenbaugh, N. S., Niyogi, D., Pielke, R. A., Sr., and Rochon, G.: Temperature and equivalent temperature over the United States (1979-2005), *International Journal of Climatology*, 30, 2045-2054, 10.1002/joc.2094, 2010.
- Ferguson, C. R., Wood, E. F., and Vinukollu, R. K.: A Global Intercomparison of Modeled and Observed Land-Atmosphere Coupling, *Journal of Hydrometeorology*, 13, 749-784, 10.1175/jhm-d-11-0119.1, 2012.
- 5 Ford, T. W., and Schoof, J. T.: Oppressive Heat Events in Illinois Related to Antecedent Wet Soils, *Journal of Hydrometeorology*, 17, 2713-2726, 2016.
- Ford, T. W., and Schoof, J. T.: Characterizing extreme and oppressive heat waves in Illinois, *Journal of Geophysical Research: Atmospheres*, 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. Environ. Sci. Technol.*, 40, 267-306, 2010.
- 10 Gardner, M. W., and Dorling, S.: Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences, *Atmospheric environment*, 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, *Climate Dynamics*, 46, 1151-1184, 2016.
- 15 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, *Journal of Hydrometeorology*, 7, 611-625, 10.1175/jhm511.1, 2006.
- 20 Horton, R. M., Mankin, J. S., Lesk, C., Coffel, E., and Raymond, C.: A review of recent advances in research on extreme heat events, *Current Climate Change Reports*, 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, *Climate Dynamics*, 48, 429-446, 10.1007/s00382-016-3085-8, 2017.
- 25 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, *International Journal of Climatology*, 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, *Journal of*
- 30 *Hydrometeorology*, 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, *Journal of Hydrometeorology*, 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, *Journal of Geophysical Research: Atmospheres*, 121, 6284-6298, 2016.
- 35

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, *Atmospheric Chemistry and Physics*, 12, 3349-3362, 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, *International Journal of Climatology*, 37, 268-275, 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, *Nature Geoscience*, 9, 389-394, 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, *Geophysical Research Letters*, 42, 5564-5570, 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, *Geoscientific Model Development*, 8, 1339-1356, 2015.

O'Neill, M. S., and Ebi, K. L.: Temperature Extremes and Health: Impacts of Climate Variability and Change in the United States, *Journal of Occupational and Environmental Medicine*, 51, 13-25, 10.1097/JOM.0b013e318173et22, 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, 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, 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, *Geophysical Research Letters*, 38, doi: 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., Hoerling, 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, *Bulletin of the American Meteorological Society*, 94, S1-S74, 10.1175/bams-d-13-00085.1, 2013.



- Pryor, S. C., and Schoof, J. T.: Evaluation of near-surface temperature, humidity, and equivalent temperature from regional climate models applied in type II downscaling, *Journal of Geophysical Research: Atmospheres*, 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, *Journal of Applied Meteorology and Climatology*, 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, *Journal of 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, *Journal of 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, *Studia Geophysica et Geodaetica*, 48, 459-468, 2004.
- Richman, M. B.: Rotation of principal components, *Journal of Climatology*, 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, *Journal of Climate*, 20, 4476-4485, 2007.
- Sanderson, M. G., and Ford, G. P.: Projections of severe heat waves in the United Kingdom, *Climate Research*, 71, 63-73, 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, *International Journal of Climatology*, 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, *Journal of Applied Meteorology and Climatology*, 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-Science Reviews*, 99, 125-161, 2010.
- Sheridan, S. C.: North American weather-type frequency and teleconnection indices, *International Journal of Climatology*, 23, 27-45, 2003.
- Sheridan, S. C., and Kalkstein, L. S.: Progress in heat watch-warning system technology, *Bulletin of the American Meteorological Society*, 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, *Natural 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, *Nature Geoscience*, 6, 1056-1061, 2013.
- Toth, E., Brath, A., and Montanari, A.: Comparison of short-term rainfall prediction models for real-time flood forecasting, *Journal of Hydrology*, 239, 132-147, 2000.

- Trenberth, K. E.: Recent observed interdecadal climate changes in the Northern Hemisphere, *Bulletin of the American Meteorological Society*, 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, *International Journal of Climatology*, 35, 85-96, 2015.
- 5 Weaver, S. J.: Factors associated with decadal variability in Great Plains summertime surface temperatures, *Journal of Climate*, 26, 343-350, 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, *Geophysical Research Letters*, 38, doi: 10.1029/2010GL046455, 2011.
- 10 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, *Scientific Reports*, 4, Article # 6929, doi:6910.1038/srep06929, 2014.
- 15 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, *Environmental Research*, 132, 350-359, 2014.

5

Table 1. Domain averaged model performance statistics (root mean square error (RMSE) and the mean Pearson correlation coefficient ( $r$ ) versus independent test data for daily maximum equivalent potential temperature ( $\max-\theta_e$ ) and daily minimum equivalent potential temperature ( $\min-\theta_e$ ) 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	$\max-\theta_e$			$\min-\theta_e$		
			<RMSE> (K)	<r>	# grid cells with $r > 0.8$ , RMSE < 5K, and both criteria met	<RMSE> (K)	<r>	# grid cells with $r > 0.8$ , RMSE < 5K, and both criteria met
MLR	Multiple linear regression	PRMC	4.05	0.849	1575, 1727, <u>1370</u>	4.39	0.836	1537, 1665, <u>1311</u>
ANN-HL1-SM	ANN with 1 hidden layer	PRMC	4.03	0.850	1602, 1754, <u>1394</u>	4.34	0.839	1572, 1750, <u>1370</u>
ANN-HL3-SM	ANN with 3 hidden layers	PRMC	3.97	0.855	1621, 1798, <u>1482</u>	4.29	0.844	1588, 1782, <u>1450</u>
ANN-HL3	ANN with 3 hidden layers, but excluding SM	None	4.10	0.846	1549, 1675, <u>1300</u>	4.33	0.843	1580, 1765, <u>1427</u>
ANN-HL3-TOP	ANN with 3 hidden layers	WGETTOP	3.97	0.856	1623, 1802, <u>1476</u>	4.29	0.844	1604, 1782, <u>1432</u>

Prof. Sara C. Pryor 10/30/2017 6:19 PM  
Deleted: r

Prof. Sara C. Pryor 11/1/2017 10:49 AM  
Deleted: &

Prof. Sara C. Pryor 11/1/2017 10:50 AM  
Deleted: &

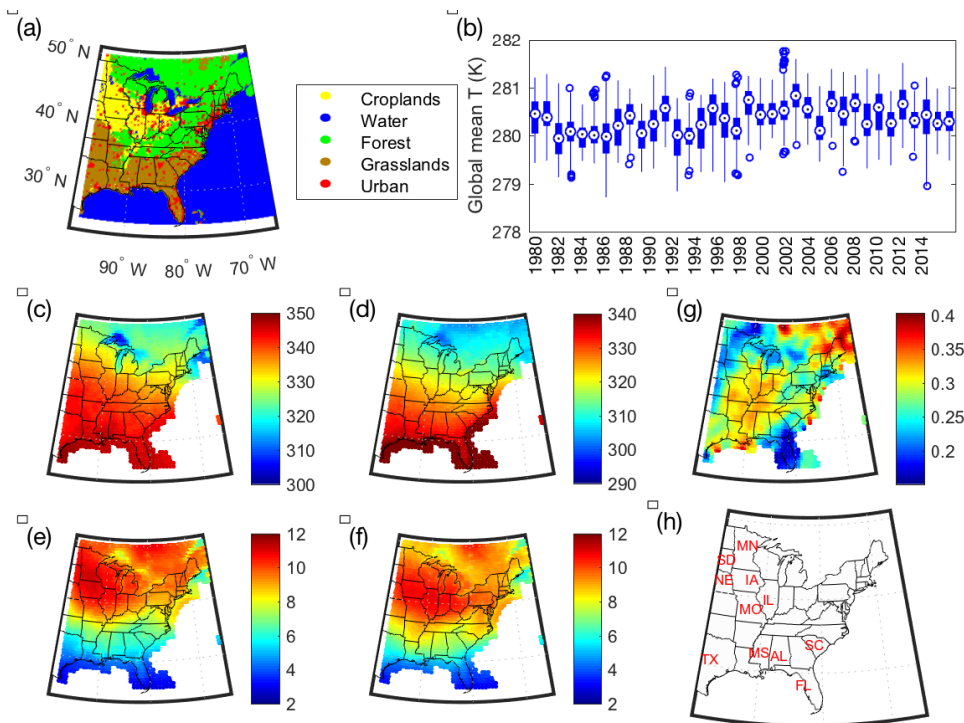


Figure 1. (a) The study domain used herein (25.5-50°N, 97.5-65°W) and five consolidated land use land-cover (LULC) classes as represented at a resolution of 0.05°x0.05° from the MODIS land cover data set for 2014 (MCD12C). (b) A boxplot of daily global mean JJA temperature values (K) (1980-2015) as computed 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 3x3 grid cells centered on the grid cell of interest, and used to compute a 90-day running JJA mean soil moisture ( $m^3m^{-3}$ ). Panel (h) shows the location of states that are explicitly referred to herein.

Prof. Sara C. Pryor 11/1/2017 11:29 AM  
Deleted: 4

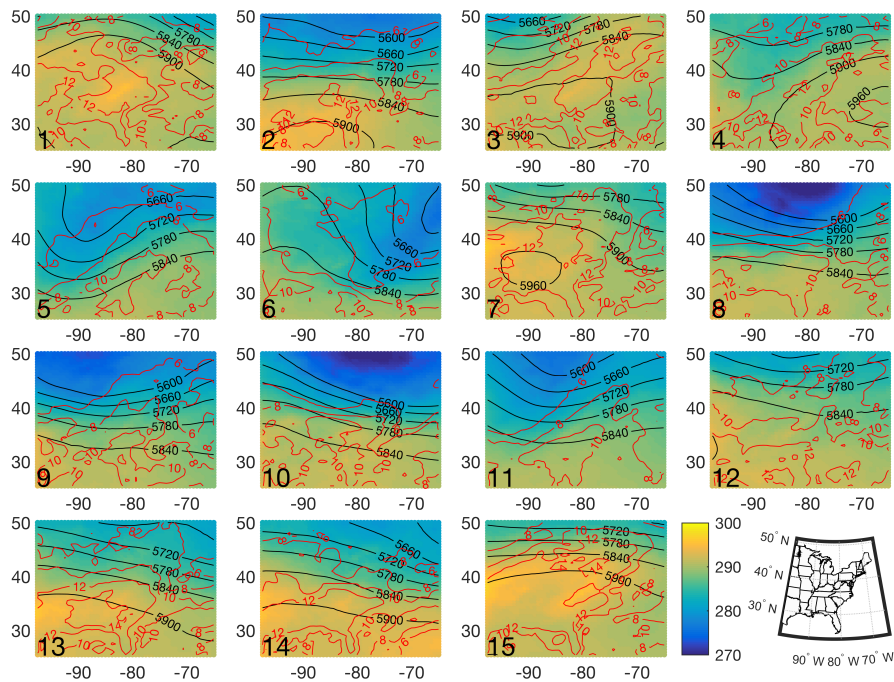
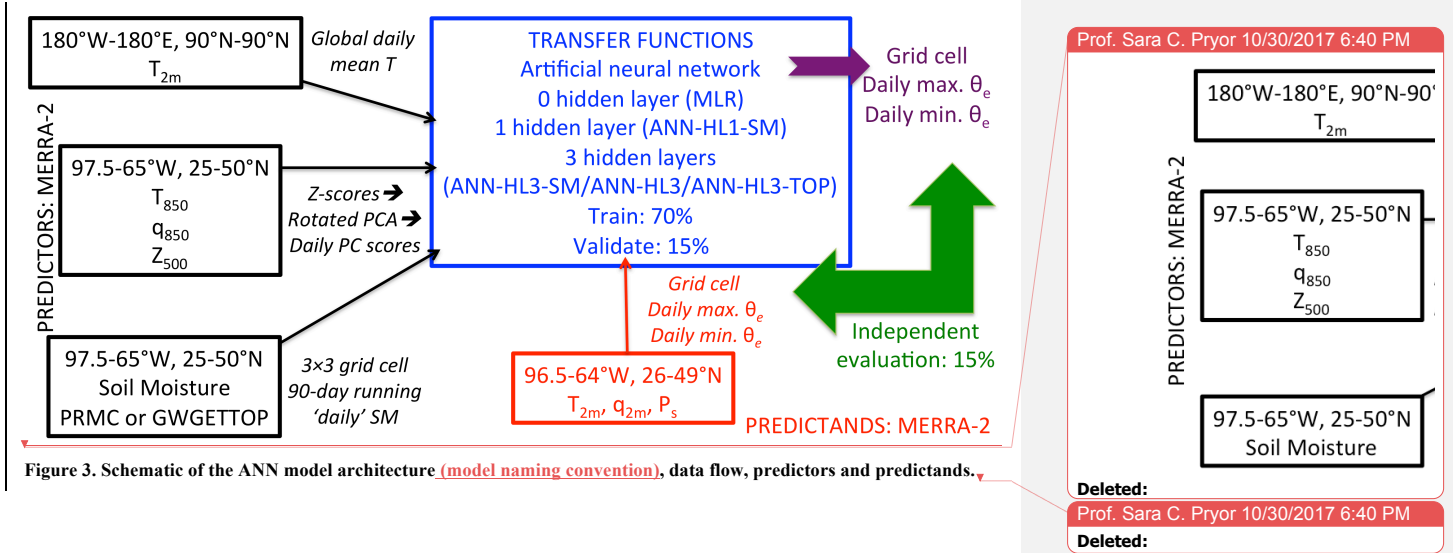


Figure 2. Summary of the conditions under each of the PCA-derived synoptic modes of variability. The panels show the “mode centroid” (i.e. mean of the 7 days that exhibit highest PC scores for the type). The color depicts  $T_{850}$  (K), the solid black lines show  $Z_{500}$  (gpm) and the red lines depict  $q_{850}$  ( $\text{gKg}^{-1}$ ). 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 Carree projection (i.e. the equi-rectangular projection) used in Figures 1, 4, 5, 6 and 7.

5

10



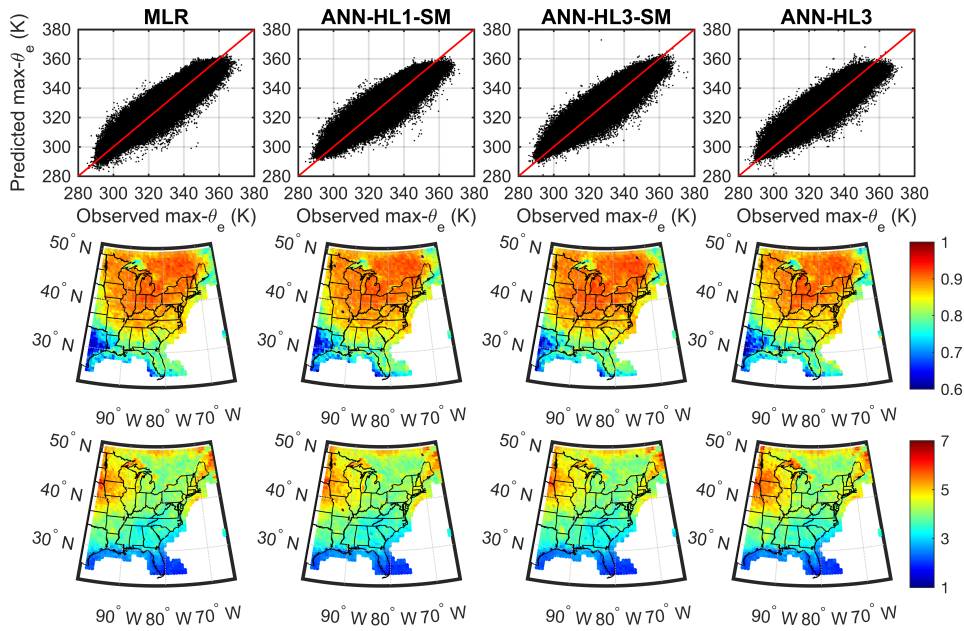


Figure 4. Précis of the downscaling model performance for daily maximum- $\theta_e$  ( $\text{max-}\theta_e$ ). The upper panels show scatterplots of observed and predicted  $\text{max-}\theta_e$  for all grid cells and all days within the test (independent) data, along with a red 1:1 line. The columns show results for four model configurations: multiple linear regression (MLR, left), ANN with 1-hidden layer (ANN-HL1-SM, second column), ANN with 3-hidden layers (ANN-HL3-SM third column) and ANN with 3-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  $\text{max-}\theta_e$  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.

5

10

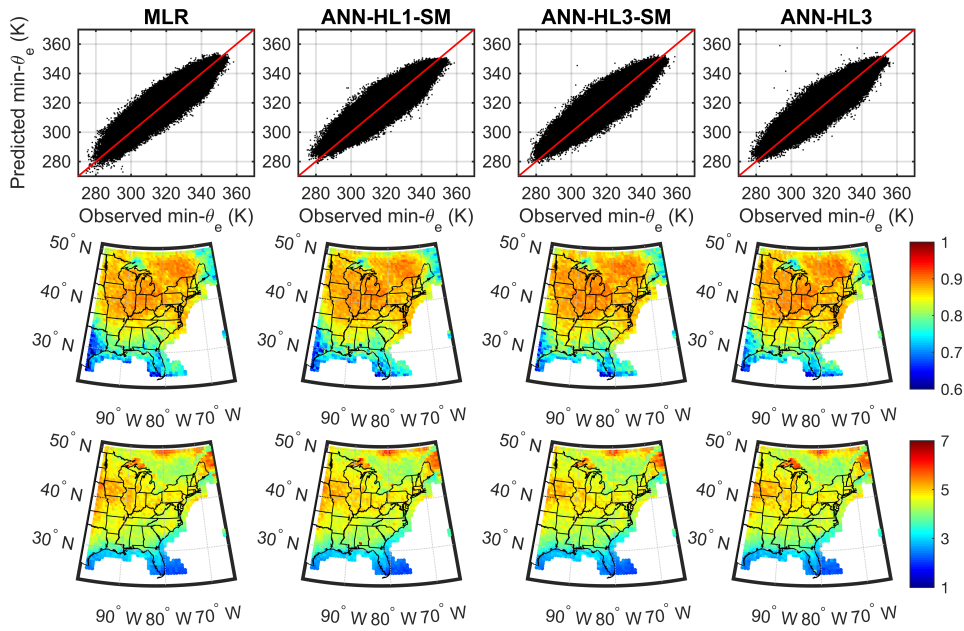
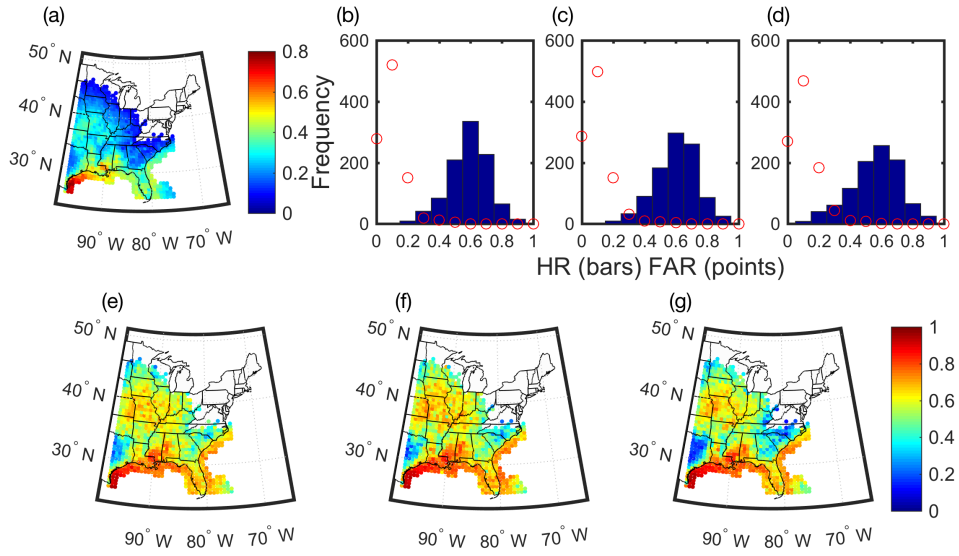


Figure 5. Précis of the downscaling model performance for daily minimum- $\theta_e$  ( $\text{min-}\theta_e$ ). The upper panels show scatterplots of observed and predicted  $\text{min-}\theta_e$  for all grid cells and all days within the test (independent) data, along with a red 1:1 line. The columns show results for four model configurations: multiple linear regression (MLR, left), ANN with 1-hidden layer (ANN-HL1-SM, second column), ANN with 3-hidden layers (ANN-HL3-SM third column) and ANN with 3-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  $\text{min-}\theta_e$  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.

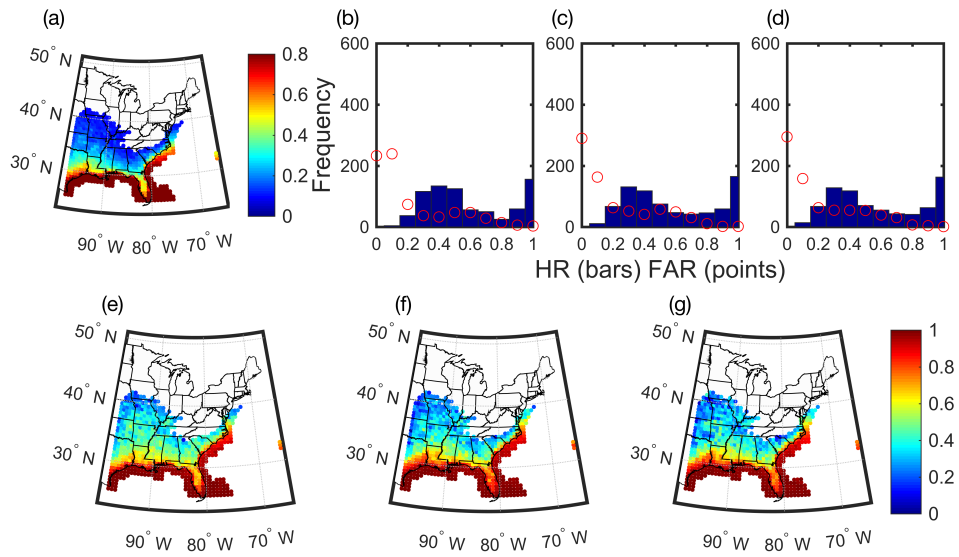
5

10





5 **Figure 6.** (a) Map of the fractional frequency of occurrence of  $\max\text{-}\theta_e > 347$  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  $\max\text{-}\theta_e > 347$  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) and (f) As in panels (b) and (e) but for output from the ANN model with 3 hidden layers and including  $SM$  (ANN-HL3-SM). (d) and (g) As in panels (b) and (e) but for output from the ANN model with 3 hidden layers excluding  $SM$  (ANN-HL3).



5 **Figure 7.** (a) Map of the fractional frequency of occurrence of  $\min-\theta_e > 337$  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  $\min-\theta_e > 337$  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) and (f) As in panels (b) and (e) but for output from the ANN model with 3 hidden layers and including  $SM$  (ANN-HL3-SM). (d) and (g) As in panels (b) and (e) but for output from the ANN model with 3 hidden layers excluding  $SM$  (ANN-HL3).