




# The critical role of humidity in modeling summer electricity demand across the United States

Debora Maia-Silva <sup>1✉</sup>, Rohini Kumar <sup>2✉</sup> & Roshanak Nateghi <sup>1,3</sup>

Cooling demand is projected to increase under climate change. However, most of the existing projections are based on rising air temperatures alone, ignoring that rising temperatures are associated with increased humidity; a lethal combination that could significantly increase morbidity and mortality rates during extreme heat events. We bridge this gap by identifying the key measures of heat stress, considering both air temperature and near-surface humidity, in characterizing the climate sensitivity of electricity demand at a national scale. Here we show that in many of the high energy consuming states, such as California and Texas, projections based on air temperature alone underestimates cooling demand by as much as 10–15% under both present and future climate scenarios. Our results establish that air temperature is a necessary but not sufficient variable for adequately characterizing the climate sensitivity of cooling load, and that near-surface humidity plays an equally important role.

<sup>1</sup>Environmental and Ecological Engineering, Purdue University, West Lafayette, IN 47906, USA. <sup>2</sup>Department Computational Hydrosystems, Helmholtz Centre for Environmental Research-UFZ, Leipzig 04318, Germany. <sup>3</sup>School of Industrial Engineering, Purdue University, West Lafayette, IN 47906, USA. ✉email: [dmaiasil@purdue.edu](mailto:dmaiasil@purdue.edu); [rohini.kumar@ufz.de](mailto:rohini.kumar@ufz.de)

Accurate predictions of demand are a key challenge in electricity adequacy planning. Climate, technological, and socioeconomic factors are commonly used in predictive models of electricity demand<sup>1,2</sup> to ensure reliable planning and operation in the electricity sector by adequately balancing supply and demand. However, more frequent and intense climate extremes such as sustained heat waves<sup>3,4</sup> cause unanticipated changes in load<sup>5</sup>, challenging the reliability of electricity demand predictions. This poses a significant risk to the resilient operation of power systems<sup>6</sup>. Specifically, unanticipated higher demand for space conditioning and refrigeration during heat waves has led to rolling outages with serious socio-economic and public health consequences<sup>7–12</sup>. In light of the expected increase in frequency and intensity of climate extremes<sup>3</sup>, climate-induced outages are an increasing risk to the resilient operation of the electric power infrastructure<sup>13–17</sup>.

Quantifying the climate-sensitive portion of the electricity demand, referred to as the climate-demand nexus, hinges on effective characterizations of the key measures of heat stress. Heat stress measures refer to a combination of temperature and humidity that capture human discomfort levels and its ability to efficiently dissipate heat to avoid life-threatening conditions<sup>9,10,18,19</sup>. Heat stress is generally associated with high morbidity and mortality rates during heat waves. Despite the rich literature in electricity demand prediction<sup>2,20–23</sup>, little prior work has focused on exploring the climate sensitivity of demand with respect to different measures of heat beyond air temperature. While climate science research establishes air temperature as an incomplete measure of the surface heat content<sup>9,10,18,24,25</sup>, the majority of the existing research on climate-demand nexus use air temperature—or features derived from air temperature, such as cooling and heating degree days—as key predictors<sup>2,26–28</sup>.

We address this gap by leveraging advanced data analytics to characterize the climate sensitivity of residential electricity demand as a function of a range of heat stress measures beyond air temperature, including dew point temperature, discomfort index (DI), heat index (HI), humidex, wet bulb temperature, and wet bulb globe temperature (see Methods). These measures are specifically chosen because they have been established in the literature as most effective in capturing heat stress<sup>29</sup>. However, no consensus has yet been reached on which measures are most predictive of the climate-demand nexus. We isolate the climate sensitivity of residential energy demand (see Methods), which has been established as the most climate-sensitive sector<sup>15,16,30–34</sup>. Thus, our results primarily reflect changes in demand due to climate variability. We focus on summer months—that is, May, June, July, and August—as the majority of the US states peak during summer, and climate extremes such as heat waves occur in the summer<sup>35</sup>.

The central thesis in this paper is that air temperature alone underestimates the climate-sensitive portion of energy demand for cooling. To test this hypothesis, we first comprehensively assess the role of air temperature in capturing the climate sensitivity of summer electricity load at a US national scale; then, we identify the measures of heat stress that are most predictive of the climate-demand nexus and finally we quantify the likely underestimation of models based on air temperature alone under current and future climate scenarios.

We use monthly aggregated, state-level electricity consumption data across the contiguous United States, extracted from the Energy Information Administration (EIA) public reports over the years of 1990–2016 as well as population data from the US Census Bureau, and harmonized climate data from NCEP North American Regional Reanalysis (NARR)<sup>36</sup>. The electricity consumption data is carefully adjusted to remove trends associated with demographic and technological

changes to isolate the climate effects on energy demand<sup>37</sup> (see Methods).

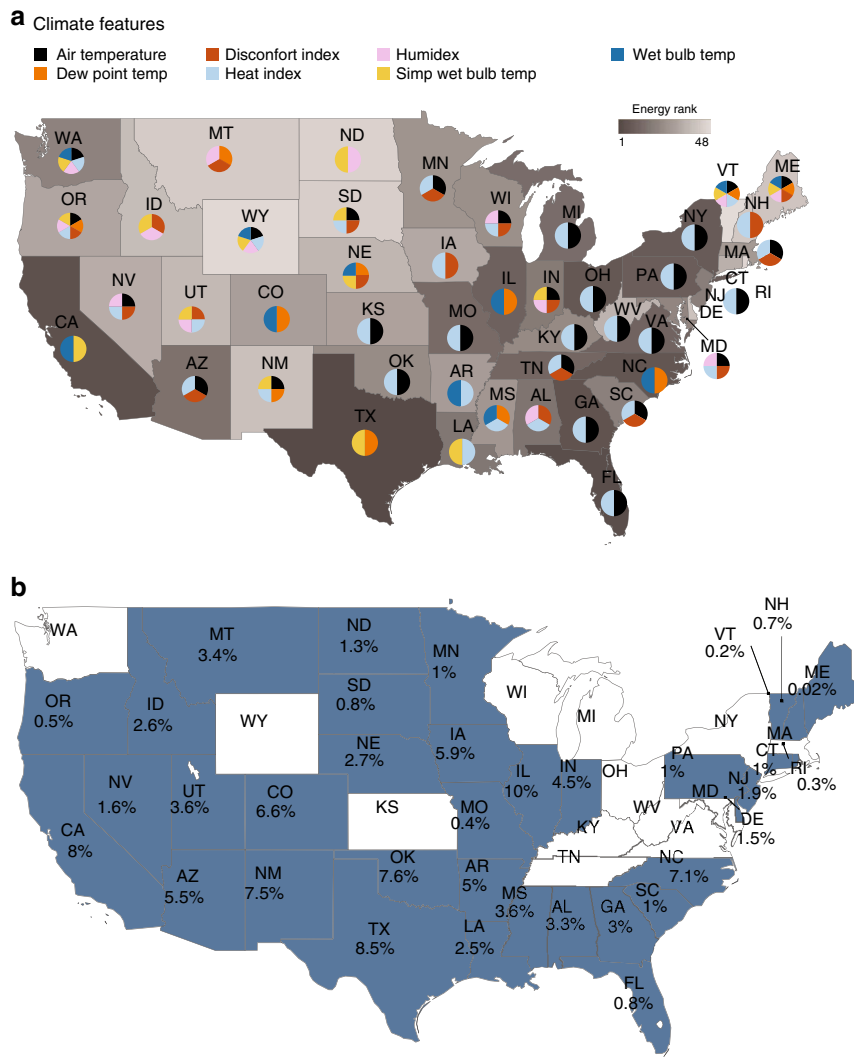
To quantify the effectiveness of air temperature alone in explaining the climate sensitivity of residential sector energy load, we develop two sets of models at the state level: air-temperature-only models, which only use surface air temperature as a predictor, and selected-features models, where data-driven variable selection is performed to identify the key measures of heat stress considered in this study. Comparing the results of the air-temperature-only and selected-features models, we determine the influence of each measure of heat stress on the overall climate sensitivity of demand. The predictive models are developed using a state-of-the-art, stochastic, non-parametric Bayesian ensemble-of-trees algorithm<sup>38</sup> (see Methods), which has been shown to outperform other climate-demand nexus models in terms of predictive accuracy<sup>1,14,15,31,39</sup>.

We conclude that the air-temperature-only models underestimate residential energy consumption, specially for future climate scenarios. With that, we show that near-surface humidity has an equally important role in electricity prediction as air temperature.

## Results

**Climate features identification.** Our results reveal substantial improvements in prediction accuracy in 35 states for the cooling demand as a function of both temperature and humidity over the air-temperature-only model. Fig. 1a illustrates the key predictors of the climate-sensitive portion of electricity demand across the 48 US states as identified by the selected-features model. A darker shade represents states with higher energy consumption and the pie charts depict the selected climate variables for each state. We find significant variability in the identified key measures of heat stress across the United States, illustrating the importance of conducting region-specific analysis in characterizing the climate-energy-demand nexus. The most energy-intensive states such as California and Texas, as well as many of the summer-peaking southern states (Supplementary Fig. 1 and Supplementary Note 1), benefit from the inclusion of measures of heat stress beyond air temperature—such as dew point temperature, wet bulb temperature, wet bulb globe temperature, and HI (see Methods)—in estimating the climate sensitivity of demand. This underscores the importance of accounting for near-surface humidity combined with temperature to better model human comfort levels and response to heat, which translates to residential cooling demand.

The highlighted states in Fig. 1b represent the extent in which the selected-features models outperformed air-temperature-only models in terms of predictive accuracy. Comparing the air-temperature-only models and selected-features models reveal that 35 states benefit from including measures of heat stress beyond air temperature, based on the comparison of root mean square error (RMSE) improvement in prediction accuracy (Fig. 1b). In addition to comparing model performances for the average consumption levels (i.e., 50th quantile predictions, Fig. 1b), we also compare the 90th quantile predictions (Supplementary Figs. 2 and 3). In this way, the effects of extreme heat events can be deduced from the higher quantiles of the aggregated electricity demand data. In other words, since heat waves typically coincide with the energy peaking summer months, the results for the 90th quantile RMSE improvement (Supplementary Fig. 3) reflect months with more intense and frequent heat events. We observe significant model improvements, in terms of out-of-sample RMSE, for a number of high-energy-consuming states of the order of 8% (7.7%) in California, 8.5% (21.1%) in Texas, 8.6% (26.1%) in Illinois, and 7.1% (9%) in North Carolina, for the 50th



**Fig. 1 Key climate predictors of cooling demand. a** States are shaded according to total energy consumption intensity (scale bar goes from 1 to 48, with darker tones representing higher intensity states in terms of energy consumption). The pie charts illustrate the state-level key measures of heat stress for predicting the climate-sensitive portion of electricity demand. The majority of states have up to two most important features. It is important to point out that climate sensitivity is not an exclusively geographical characteristic, and considering the population density weight method for the variables calculation (see Methods), it is possible to expect neighboring states to not necessarily have the same selected climate variables. Non-summer-peaking states with many selected features (e.g., Washington (WA) and Oregon (OR)) reveal relative climate insensitivity in comparison (Supplementary Note 1). Note that the top energy-consuming states of California (CA) and Texas (TX) do not include air temperature as the key predictors of the climate-sensitive portion of demand. **b** Highlighted states show percentage improvement in predictive accuracy when comparing the selected-features models to the air-temperature-only models (comparison based on out-of-sample RMSE). Source data are provided as a Source Data file.

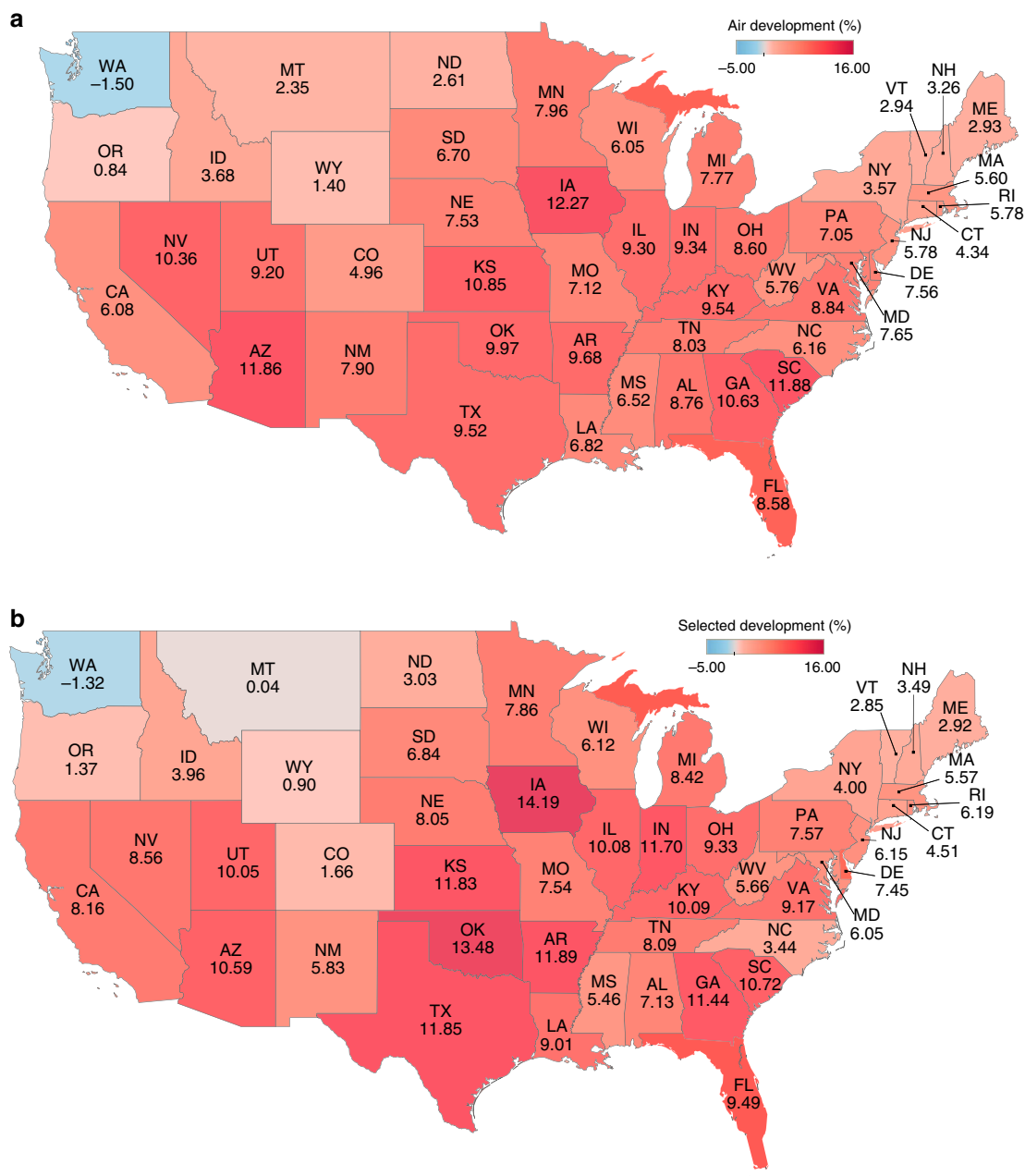
quantile predictions (and 90th quantiles predictions—Supplementary Fig. 3). The observed improvements are substantial. For instance, in a state such as Texas, the selected-features model’s average improvement of 8.5% in a given summer month such as August 2016 is equivalent to 1,498,968 MWh of energy consumption, which could sustain the residents of Austin-TX for more than 4 months<sup>40–43</sup>. In the same year and month, the 8% median increase in the state of California could sustain almost 1.5 million of Californians households<sup>44</sup>.

**Sensitivity analysis under future climate scenarios.** To quantify the extent to which the air-temperature-only models underestimate the climate sensitivity of demand under different climate scenarios, we use projected climate data extracted from five widely used Global Circulation Models (GCMs).

The GCM-extracted climate data are used to train state-level air-temperature-only and selected-features models to predict for

the reference period of 1991–2010, and the future period of 2031–2050 (Methods). Since the objective here is to evaluate the changes in demand pattern as a result of future climate change (i.e., to quantify the relative increase over the periods, as opposed to absolute values), reference period data from the GCM is used instead of the historical data to remove bias induced by the use of data from different sources in the sensitivity analysis. In other words, input data generated from the same source (i.e., the GCMs) over the two time periods of reference (1991–2010) and future (2031–2050) are used in the developed predictive framework to track the relative increase in the climate-sensitive portion of demand as estimated by the air-temperature-only versus selected-features models.

The 1991–2010 reference period was chosen to match the period of the observed data as closely as possible and, at the same time, the same length as the future period (i.e., 20 years). While the common practice in climate science research is making



**Fig. 2 Projected increase in cooling demand.** Increase in the climate-sensitive portion of the demand under climate change (scale bars go from – 5% to 16%) based on: **a** the air-temperature-only models and **b** the selected-features models. Note the significantly higher demand increases for the selected-features models (32 states showed higher values for the selected-features models), supporting the initial hypothesis that the air-temperature-based projections underestimate the climate sensitivity of demand. Source data are provided as a Source Data file.

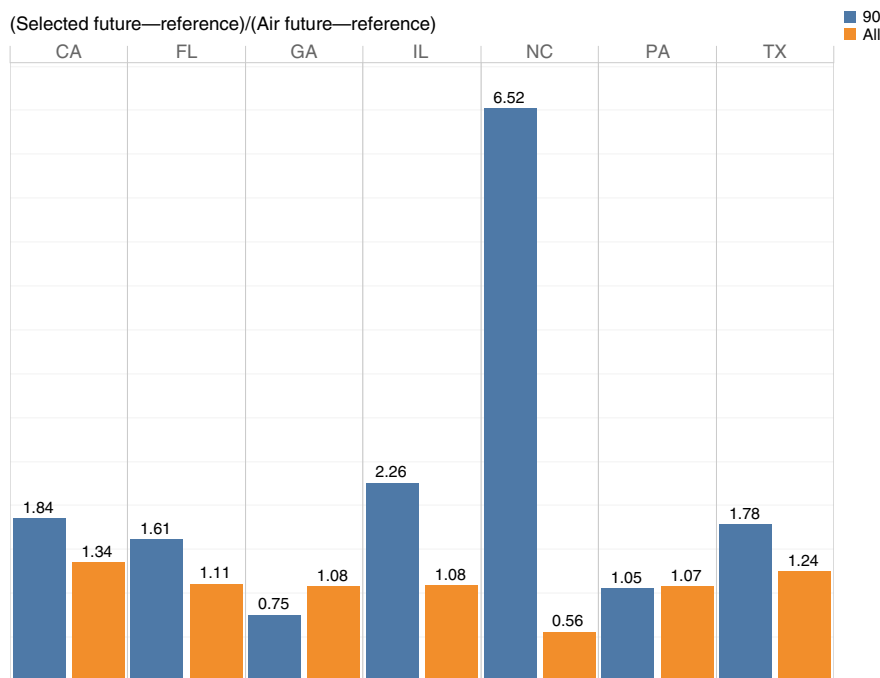
projections until 2100, that timeline is not appropriate in the energy sector, given the life span of the existing energy infrastructure<sup>45–47</sup>. The future period of 2031–2050 was chosen since energy infrastructure planning has the 2050 year as a comparable reference target in most energy projection and planning reports<sup>48,49</sup>.

The selected-features models project a significantly higher increase in the climate-sensitive portion of demand over time compared to the air-temperature-only models, when comparing the relative increase of each model (Fig. 2). The projections based on the selected-features models for the high-energy-consuming states of Texas, California, and Florida show a relative demand increase of about 12%, 8% and 10%, respectively, over the reference period (Fig. 2).

Figure 3 depicts the relative increase of projected demand by the selected-features model over the air-temperature-only model for the average consumption levels and 90th quantile predictions for the top seven energy-consuming states that presented an improvement in prediction accuracy in the historical period. Figure 3 enables the comparison between the ratios of the projections based on the selected features over the air-temperature models for 90th percentile as well as the average values, as shown in the equation below. Complete results for all states are shown in Supplementary Fig. 4.

$$\frac{\text{Selected features future}_{2031-2050} - \text{Selected features reference}_{1991-2010}}{\text{Air-temperature-only future}_{2031-2050} - \text{Air-temperature-only reference}_{1991-2010}}$$

(1)



**Fig. 3 Relative increase in average and extreme projections.** Seven out of the top ten energy-consuming states in the United States that showed improvement with selected-features model. The relative increase of projected demand by the selected-features model over the air-temperature-only model for the average consumption levels and 90th quantile predictions for the top energy-consuming states. Source data are provided as a Source Data file.

This comparison reveals that the underestimation in energy demand attributable to the air-temperature-only models is substantially more significant for the projections of higher demand values that are typically associated with periods of intense heat stress months—where the resilient operation of the grid is most challenged and unexpected demand spikes lead to supply inadequacy and thereby increase the risks of morbidity and mortality.

**Discussion**

The existing projections of climate-induced demand increase based on rising air temperatures alone ignore the fact that rising temperatures are associated with increased humidity—a lethal combination that increase morbidity and mortality in the absence of adequate cooling capacity during extreme heat events. We show that air temperature is a necessary but not sufficient measure to characterize residential space cooling demand during summer times. Humidity levels are also critical in capturing what truly reflects heat sensation for the human body<sup>9</sup>. Ignoring the role of humidity leads to underestimating the climate sensitivity of demand, challenging the resilient operation of power systems—especially under future warming scenarios where summertime energy production will be further constrained<sup>50</sup>. Inadequate supply to meet rising demand will have significant socio-economic costs due to adverse health effects—particularly among the most vulnerable population—and warrants rapid and costly investments in energy infrastructure expansion as well as adaptation measures<sup>26</sup>.

We propose a data-driven framework to quantify the extent of underestimation of the climate-sensitive portion of cooling load, attributable to the use of air temperature alone as a measure of heat stress. Our results reveal a significant increase in predictive accuracy (of the order of 8% for high-energy-consuming states such as California and Texas) in characterizing the climate-sensitive portion of demand through a more holistic inclusion of

measures of heat stress. The results based on projecting demand under future climate scenarios are consistent, in that models based on air temperature alone show a systematic underestimation of the climate-sensitive portion of demand. The underestimation is particularly pronounced at the upper tail of the demand distribution, suggesting that the projected increase would be almost twice as large when considering measures of heat stress beyond air temperature in high-energy states such as California, Florida, and Texas.

It is important to note that our results pertain to only the climate-sensitive portion of the residential cooling demand. To project changes in total residential cooling demand under future climate scenarios, additional information regarding economic and population growth as well as technology and demographic changes need to be considered. What we wish to highlight here is that, under similar cooling technology, the existing projections based on air temperature alone substantially underestimate the anticipated demand increase under future warming scenarios, which could lead to inadequate and ineffective investments in capacity expansion and demand response programs as well as adaptation measures. Furthermore, while this study is focused on the residential energy demand, the methodology can be extended to other climate-sensitive sectors<sup>51</sup>.

**Methods**

**Observed electricity consumption data.** Monthly electricity consumption data was extracted from the EIA<sup>52</sup> over the years of 1990–2016 at a state level, with a total of 68 months per state (4 months per year for a total of 17 years). The source file includes electricity consumption data in Megawatt hours (converted to kWh), separated by sector: residential, commercial, industrial, transportation and others. The electricity consumption of the residential sector was the focus of the analysis presented in this paper. The residential electricity consumption data was detrended to isolate the climate effects on energy demand from techno-demographic changes such that increases in energy were not attributable to non-climatic factors such as technology shifts and population increase<sup>16,37</sup>. The residential energy consumption data in each state was first normalized using state-level population data—from the US Census Bureau<sup>41</sup>—to obtain a per-capita electricity demand, and then

detrended based on the following steps:

$$E(y) = \frac{\sum_{m=1}^{12} \sum_{y=1}^{n_{\text{years}}} E(m, y)}{n_{\text{years}}}, \quad (2)$$

where the total years,  $n_{\text{years}}$ , range from 1990 to 2016,  $m$  denotes the month, and  $y$  the year.

Then, an adjustment factor was calculated for each year as the sum of the per-capita consumption per month divided by the yearly consumption  $E(y)$ . This process was repeated for all states as follows:

$$F_{\text{adj}} = E(y)^{-1} \sum_{m=1}^{12} E(m, y). \quad (3)$$

Finally, the adjusted energy consumption was calculated by dividing the monthly consumption by the adjustment factor:

$$E(m, y)_{\text{adj}} = E(m, y) / F_{\text{adj}}. \quad (4)$$

The detrended monthly electricity consumption data (described above) is referred to as observed electricity consumption data. This is a widely used method for trend-adjusting monthly aggregated regional data. For a comparative assessment of different de-trending methodologies please refer to Bessec and Fouquau<sup>53</sup>.

**Observed climate data.** The observed climate data was acquired at a sub-daily time scale for the period of 1990–2016 from the NCEP NARR<sup>36,54,55</sup>, aggregated at a monthly level to match the temporal scale of electricity consumption data, and weighted by population so as to give a greater weight to areas with higher population density when aggregating to the state level. The input climate data include air temperature, dew point temperature, DI, HI, humidex, wet bulb temperature, and wet bulb globe temperature. All climate variables use a combination of temperature, humidity, and pressure.

Dew point refers to the temperature in which air is saturated with water vapor, calculated in this paper using the equation below:

$$t_d \approx t - \left( \frac{100 - \text{RH}}{5} \right), \quad (5)$$

where  $t_d$  is the dew point temperature and  $t$  is air temperature in Celsius, and RH stands for relative humidity in percent (same RH as in the other equations in this section)<sup>56</sup>.

Wet bulb temperature is the lowest temperature to which air can be cooled by water evaporation at a constant pressure. In this paper, it follows the equation:

$$T_w = T_w - \frac{f(T_w; \pi) - (C/T_E)^\lambda}{f'(T_w; \pi)}, \quad (6)$$

where  $\pi = (p/p_0)^{1/\lambda}$  is the Exner function, used for scaling,  $\lambda$  is the inverse of the Poisson constant for dry air (3.504),  $p$  pressure (mb),  $p_0$  the reference pressure (1000 mb);  $C$  is the freezing temperature in Kelvin, and  $T_E$  the equivalent temperature, which is the moist potential temperature scaled by  $\lambda$ . For reviewing the complete derivation of the equation, refer to the Appendix A of refs. 29,57.

DI was developed in the 1950s to calibrate air conditioners and further adapted by the Israeli Defense Force as a main measure of heat stress:

$$\text{DI} = 0.5T_w + 0.5T_C, \quad (7)$$

where  $T_w$  stands for wet bulb temperature, RH is percentage relative humidity, and  $T_C$  is the temperature in Celsius. DI is unitless<sup>29,58,59</sup>.

Wet bulb globe temperature (sWBGT) is the heat stress measure in direct sunlight:

$$\text{sWBGT} = 0.56T_C + \frac{0.393e_{\text{RH}}}{100} + 3.94, \quad (8)$$

where  $e_{\text{RH}} = (\text{RH}/100)e_{\text{spai}}$ ; wet bulb temperature used here is unitless and widely used.

$e_{\text{RH}}$  represents the vapor pressure in pascals and is calculated from RH in percentage and saturated vapor pressure<sup>29,60</sup>.

Humidex was developed for the Meteorological Service of Canada. It is a unitless<sup>29,61</sup> index, aiming to explain what the temperature feels like for the human body:

$$\text{Humidex} = T_C + \frac{5}{9} \left( \frac{e_{\text{RH}}}{100} - 10 \right). \quad (9)$$

HI is also called apparent temperature, describing what the temperature feels like to the human body when relative humidity is combined with air temperature:

$$\begin{aligned} \text{HI} = & -42.379 + 2.04901523T_F + 10.14333127\text{RH} - 0.22475541T_F\text{RH} \\ & - 6.83783 \times 10^{-3}T_F^2 - 5.481717 \times 10^{-2}\text{RH}^2 + 1.22874 \times 10^{-3}T_F^2\text{RH} \\ & + 8.5282 \times 10^{-4}T_F\text{RH}^2 - 1.99 \times 10^{-6}T_F^2\text{RH}, \end{aligned} \quad (10)$$

where ( $T_F$ ) denotes the air temperature, and HI are measured in degrees Fahrenheit<sup>29,62</sup>.

**Projected climate data.** Five different GCMs were used in this paper, namely, Geophysical Fluid Dynamics Laboratory Earth Systems Model (GFDL-ESM2M), Hadley Global Environment Model 2-Earth System (HadGEM2-ES), IPSL Earth System Model for the 5th IPCC report (IPSL-CM5A-LR)<sup>63</sup>, Atmospheric Chemistry Coupled version of MIROC-ESM, a Earth System model (MIROC-ESM-CHEM), and the Norwegian Earth System Model (NorESM1-M). The data are considered under the emission scenario that has an end-of-century radiative forcing equal to  $8.5 \text{ W m}^{-2}$ , a Representative Concentration Pathway that is characterized by high greenhouse emission levels (RCP8.5)<sup>15,64</sup>. This global data is approximated by a  $0.5^\circ$  spatial resolution ( $\sim 50 \text{ km}$ )<sup>65</sup> and is processed from 1950 to 2099. The same variables were calculated for the projected data, namely, dew point temperature, DI, HI, humidex, wet bulb temperature, and wet bulb globe temperature. Two periods were extracted from the projected data, namely, the reference period of 1991–2010, and the future period of 2031–2050.

There were two distinct stages in model development and analysis, namely, predictive model development using the observed data and sensitivity analysis using the projected data (Supplementary Fig. 5). In the predictive model development stage, electricity consumption data, climate data, and population data were aggregated and normalized at the state level, as explained earlier, yielding monthly observed data for 48 states. Using these data sets, the development of air-temperature-only models and selected-features models involved three steps: for the selected-features model, variable selection was performed to identify the key predictors of cooling demand (details in the next section: BART (Bayesian Additive Regression Trees) algorithm and modeling process); then, the model hyper-parameters were tuned (through cross-validation), and finally a ten-fold cross-validation was performed to assess the predictive accuracy of the final models. For the air-temperature-only models, the same process was followed with the exception of the variable selection step, as there was only one independent variable. After developing these statistical models, the results were processed with the collection of in-sample (training set, 90%) and out-of-sample (test set, 10%)  $R^2$  and RMSE values. The comparative assessments between the models were conducted based on the out-of-sample values of errors (RMSE results shown on Supplementary Figs. 2 and 3 and  $R^2$  results shown on Supplementary Figs. 6 and 7).

In the sensitivity analysis stage, projected climate data from the five GCMs over the reference period (1991–2010) and future period (2031–2050) were used as inputs to the previously developed predictive models. This generated four sets of data, namely, air temperature, reference: estimates from the air-temperature-only model using the reference period data; air temperature, future: estimates from the air-temperature-only model using the future period; selected features, reference: estimates from the selected-features model for the reference period; and finally, the selected features, future, estimates from the selected-features model for the future period. This allowed for calculating the delta change over time between the same models (i.e., calculating the difference between selected-features models over the periods of 2031–2050 and selected 1991–2010). Similar analysis was also conducted on the 90th quantile of data, as electricity inadequacy during extreme heat episodes is typically associated with the tail of the demand distribution.

**BART algorithm and modeling process.** We leveraged a non-parametric Bayesian ensemble-of-trees algorithm to characterize the climate sensitivity of electricity consumption, since the algorithm was shown to outperform other climate-demand nexus models in terms of predictive accuracy<sup>1,14,15,31,38,39</sup>. The independent variables in the development of the BART models were heat stress measures, and the response variable was state-level electricity consumption. Models were trained and tested using a 10-fold cross-validation technique<sup>66</sup>.

The non-parametric Bayesian sum-of-trees model can be represented using the equation below<sup>38,67</sup>:

$$Y = \sum_{j=1}^m g(x; T_j, M_j) + \varepsilon, \varepsilon \sim N(0, \sigma^2). \quad (11)$$

Here,  $Y$  denotes the response variable,  $g$  denotes the regression tree function,  $T_j$  denotes individual tree structures,  $M_j$  denotes a set of parameters (e.g., mean values) associated with the tree nodes. BART is the sum of  $m$  total trees. The decision rules associated with each tree involve binary splitting of the predictor space. So for a given  $T$  and  $M$ ,  $g(x; T, M)$  represents a tree with parameters  $M$  assigning the binary splits for the terminal nodes of  $T$ .  $\varepsilon$  is the stochastic irreducible error, assumed to be normally distributed with a constant variance<sup>38</sup>.

The BART algorithm uses a prior over the sum-of-trees parameters to regularize the trees, that is, to weaken the effect of individual trees with the goal for better prediction when adding all decision trees. A prior is imposed on all parameters,  $(T_j, M_j)$  [ $\forall j = 1, \dots, m$ ] and  $\sigma$ . This regularization is made to keep individual trees effects from being unduly influential over the entire sum of trees. Without a regularization, large tree components would overwhelm the sum of trees and would diminish the benefits of leveraging additive tree structures. The prior regularization can be mathematically represented as:

$$p(M_j | T_j) = \prod_i p(\mu_{ij} | T_j), \quad \text{where } \mu_{ij} \in M_j, \quad (12)$$

where tree components  $T_j$  and  $M_j$  are assumed to be independent of each other and of  $\sigma$ .

As mentioned earlier, to assess the role of air temperature in capturing the climate sensitivity of electricity consumption compared to other measures of heat stress, two models were developed for each state: the air-temperature-only model, and the selected-features model. Air-temperature-only models were constructed with a single input variable of surface temperature and electric demand as the response variable. To develop the selected features a data-driven variable selection algorithm was harnessed. Variable selection: specifically, the variable selection algorithm involved identifying the variables most frequently used as decision/splitting rules in the sum-of-tree model<sup>38</sup>. The calculation was then averaged over 100 model iterations to achieve stable estimates throughout the sum-of-trees distribution. This is a reasonably robust approach for variable selection using the Bayesian ensemble-of-trees algorithm<sup>38,67</sup>.

**Pseudo-code.** Feature selection is performed for 48 states before running BART and out-of-sample RMSE is averaged over 20 models (cross validation performed at line 3). Code was performed in parallel cores.

#### Algorithm 1. BART Machine Model Fit

```

1: for  $i \in \{1 : 48\}$  do
2:   for  $k \in \{1 : 20\}$  do
3:      $BartMachine_k, RMSE_k, R_k^2$ 
4:   end for
5:    $RMSE_i = \text{mean}(RMSE_k) \sim k \in 1 : 20$ 
6:    $R_i^2 = \text{mean}(R_k^2) \sim k \in 1 : 20$ 
7: end for

```

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

#### Data availability

All data sets used in this study are publicly available from the referenced sources.

#### Code availability

Code is available at the corresponding author's Github along with pre-processed data [[https://github.com/DebsM/humidity\\_paper\\_codes](https://github.com/DebsM/humidity_paper_codes)].

Received: 24 October 2019; Accepted: 4 March 2020;

Published online: 03 April 2020

#### References

- Mukherjee, S., Nateghi, R. & Hastak, M. A multi-hazard approach to assess severe weather-induced major power outage risks in the U.S. *Reliab. Eng. Syst. Saf.* **175**, 283–305 (2018).
- Ruijven, B. J. v., Cian, E. D. & Wing, I. S. Amplification of future energy demand growth due to climate change. *Nat. Commun.* **10**, 2762 (2019).
- Coumou, D. & Rahmstorf, S. A decade of weather extremes. *Nat. Clim. Change* **2**, 491–496 (2012).
- Meehl, G. A. & Tebaldi, C. More intense, more frequent, and longer lasting heat waves in the 21st century. *Science* **305**, 994–997 (2004).
- Alipour, P., Mukherjee, S. & Nateghi, R. Assessing climate sensitivity of peak electricity load for resilient power systems planning and operation: a study applied to the Texas region. *Energy* **185**, 1143–1153 (2019).
- US Department of Energy. *US Energy Sector Vulnerabilities to Climate Change and Extreme Weather* <https://apps.dtic.mil/dtic/tr/fulltext/u2/a583709.pdf> (2013).
- Semenza, J. C. et al. Heat-related deaths during the July 1995 heat wave in Chicago. *N. Engl. J. Med.* **335**, 84–90 (1996).
- Electric Choice. *9 of the Worst Power Outages in United States History* <https://www.electricchoice.com/blog/worst-power-outages-in-united-states-history/> (2019).
- Mora, C. et al. Global risk of deadly heat. *Nat. Clim. Change* **7**, 501–506 (2017).
- Im, E.-S., Pal, J. S. & Eltahir, E. A. B. Deadly heat waves projected in the densely populated agricultural regions of South Asia. *Sci. Adv.* **3**, e1603322 (2017).
- US Energy Information Administration. *Texas Heat Wave, August 2011: Nature and Effects of An Electricity Supply* <https://www.eia.gov/todayinenergy/detail.php?id=3010> (2011).
- Creswell, J. & Gebeloff, R. Traders profit as power grid is overworked. *The New York Times*. <https://www.nytimes.com/2014/08/15/business/energy-environment/traders-profit-as-power-grid-is-overworked.html> (2018).
- Sailor, D. J., Baniassadi, A., O'Lenick, C. R. & Wilhelm, O. V. The growing threat of heat disasters. *Environ. Res. Lett.* **14**, 054006 (2019).
- Mukhopadhyay, S. & Nateghi, R. Estimating climate—demand nexus to support longterm adequacy planning in the energy sector. *Climatic Change* **29**, 1–5 (2018).
- Nateghi, R. & Mukherjee, S. A multi-paradigm framework to assess the impacts of climate change on end-use energy demand. *PLoS ONE* **12**, e0188033 (2017).
- Mukherjee, S., Vineeth, C. R. & Nateghi, R. Evaluating regional climate electricity demand nexus: a composite Bayesian predictive framework. *Appl. Energy* **235**, 1561–1582 (2019).
- De Cian, E. & Sue Wing, I. Global energy consumption in a warming climate. *Environ. Resour. Econ.* **72**, 365–410 (2019).
- Sherwood, S. C., Huber, M. & Emanuel, K. A. An adaptability limit to climate change due to heat stress. *Proc. Natl Acad. Sci. USA* **107**, 9552–9555 (2010).
- Coffel, E. D., Horton, R. M. & Sherbinin, A. D. Temperature and humidity based projections of a rapid rise in global heat stress exposure during the 21st century. *Environ. Res. Lett.* **13**, 014001 (2017).
- Kwak, Y., Seo, D., Jang, C. & Huh, J.-H. Feasibility study on a novel methodology for short-term real-time energy demand prediction using weather forecasting data. *Energy Buildings* **57**, 250–260 (2013).
- PJM. *Load Forecast Development Process* <https://www.pjm.com/planning/resource-adequacy-planning/load-forecast-dev-process.aspx> (2016).
- Apadula, F., Bassini, A., Elli, A. & Scapin, S. Relationships between meteorological variables and monthly electricity demand. *Appl. Energy* **98**, 346–356 (2012).
- Mirasgedis, S. et al. Models for mid-term electricity demand forecasting incorporating weather influences. *Energy* **31**, 208–227 (2006).
- Pielke, R. A., Davey, C. & Morgan, J. Assessing global warming with surface heat content. *Eos, Trans. Am. Geophys. Union* **85**, 210–211 (2004).
- Xie, J., Chen, Y., Hong, T. & Laing, T. D. Relative humidity for load forecasting models. *IEEE Trans. Smart Grid* **9**, 191–198 (2018).
- Auffhammer, M., Baylis, P. & Hausman, C. H. Climate change is projected to have severe impacts on the frequency and intensity of peak electricity demand across the United States. *Proc. Natl Acad. Sci. USA* **114**, 1886–1891 (2017).
- Jaglom, W. S. et al. Assessment of projected temperature impacts from climate change on the U.S. electric power sector using the Integrated Planning Model. *Energy Policy* **73**, 524–539 (2014).
- Reyna, J. L. & Chester, M. V. Energy efficiency to reduce residential electricity and natural gas use under climate change. *Nat. Commun.* **8**, 1–12 (2017).
- Buzan, J. R., Oleson, K. & Huber, M. Implementation and comparison of a suite of heat stress metrics within the Community Land Model version 4.5. *Geosci. Model Dev. (Kattlenburg-Lindau)* **8**, 151 (2015).
- Obringer, R., Kumar, R. & Nateghi, R. Analyzing the climate sensitivity of the coupled water-electricity demand nexus in the Midwestern United States. *Appl. Energy* **252**, 113466 (2019).
- Mukherjee, S. & Nateghi, R. A data-driven approach to assessing supply inadequacy risks due to climate-induced shifts in electricity demand. *Risk Anal.* **39**, 673–694 (2018).
- Raymond, L. et al. Projected climate change impacts on Indiana's energy demand and supply. *Clim. Change* <https://doi.org/10.1007/s10584-018-2299-7> (2019).
- Deschenès, O. & Greenstone, M. Climate Change, mortality, and adaptation: evidence from annual fluctuations in weather in the US. *Am. Econ. J. Appl. Econ.* **3**, 152–185 (2011).
- Obringer, R., Kumar, R. & Nateghi, R. Analyzing the climate sensitivity of the coupled water-electricity demand nexus in the Midwestern United States. *Appl. Energy* **252**, 113466 (2019).
- Hadley, S. W., Erickson, D. J., Hernandez, J. L., Broniak, C. T. & Blasing, T. J. Responses of energy use to climate change: a climate modeling study. *Geophys. Res. Lett.* **33**, <https://doi.org/10.1029/2006GL026652> (2006).
- Mesinger, F. et al. North American Regional Reanalysis. *Bull. Am. Meteorol. Soc.* **87**, 343–360 (2006).
- Sailor, D. J. & Muñoz, J. R. Sensitivity of electricity and natural gas consumption to climate in the U.S.A.—methodology and results for eight states. *Energy* **22**, 987–998 (1997).
- Chipman, H. A., George, E. I. & McCulloch, R. E. BART: Bayesian additive regression trees. *Ann. Appl. Stat.* **4**, 266–298 (2010).
- Mukherjee, S. & Nateghi, R. Climate sensitivity of end-use electricity consumption in the built environment: an application to the state of Florida, United States. *Energy* **128**, 688–700 (2017).
- Energy Information Administration. *Residential Energy Consumption Survey (RECS)* <https://www.eia.gov/consumption/residential/index.php> (2015).
- Population and Housing Unit Estimates Datasets. <https://www.census.gov/programs-surveys/popest/data/data-sets.html> (2019).
- Energy Department. Residential Average Monthly kWh and Bills: Residential Average Monthly kWh and Bills, 3KB (Energy Department, 2016).
- U.S. Census Bureau. QuickFacts: Austin city, Texas. <https://www.census.gov/quickfacts/fact/table/austincitytexas/PST045218> (2019).

44. San Francisco, CA Electricity Rates. <https://www.electricitylocal.com/states/california/san-francisco/> (2019).
45. Allen, M. R. & Ingram, W. J. Constraints on future changes in climate and the hydrologic cycle. *Nature* **419**, 228–232 (2002).
46. Levesque, A. et al. How much energy will buildings consume in 2100? A global perspective within a scenario framework. *Energy* **148**, 514–527 (2018).
47. American Society of Civil Engineer (ASCE). *ASCE 2017 Infrastructure Report Card* <https://www.infrastructurereportcard.org/cat-item/energy/> (2017).
48. Energy Information Administration. *EIA—Annual Energy Outlook 2019* <https://www.eia.gov/outlooks/aeo/> (2019).
49. International Renewable Energy Agency. *Global Energy Transformation: A Roadmap to 2050 (2019 edition)* (International Renewable Energy Agency, 2019).
50. Bartos, M. D. & Chester, M. V. Impacts of climate change on electric power supply in the Western United States. *Nat. Clim. Change* **5**, 748–752 (2015).
51. Mu, J. E., Antle, J. M. & Abatzoglou, J. T. Representative agricultural pathways, climate change, and agricultural land uses: an application to the Pacific Northwest of the USA. *Mitig. Adapt. Strateg. Glob. Change* **24**, 819–837 (2019).
52. Electric power sales, revenue, and energy efficiency Form EIA-861 detailed data files. <https://www.eia.gov/electricity/data/eia861/> (2019).
53. Bessec, M. & Fouquau, J. The non-linear link between electricity consumption and temperature in Europe: a threshold panel approach. *Energy Econ.* **30**, 2705–2721 (2008).
54. NCEP North American Regional Reanalysis (NARR). <https://www.esrl.noaa.gov/psd/data/gridded/data.narr.html> (2019).
55. Center for International Earth Science Information Network. <http://www.ciesin.org/data.html> (2019).
56. Lawrence, M. G. The relationship between relative humidity and the dewpoint temperature in moist air: a simple conversion and applications. *Bull. Am. Meteorol. Soc.* **86**, 225–233 (2005).
57. Davies-Jones, R. An efficient and accurate method for computing the wet-bulb temperature along pseudoadiabats. *Mon. Weather Rev. (Washington)* **136**, 2764–2785 (2785).
58. Thom, E. C. The discomfort index. *Weatherwise* **12**, 57–61 (1959).
59. Epstein, Y. & Moran, D. S. Thermal comfort and the heat stress indices. *Indus. Health* **44**, 388–398 (2006).
60. Department of the Army and Air Force. *Heat Stress Control and Heat Casualty Management* (Headquarters, Department of the Army and Air Force, Washington, 2003).
61. Masterton, J. M. & Richardson, F. A. *Humidex; A Method of Quantifying Human Discomfort Due to Excessive Heat and Humidity* (Environment Canada, Atmospheric Environment, Downsview, 1979).
62. Rothfusz, L. P. *The Heat Index Equation*. National Weather Service Technical Attachment (SR 90-23) (National Weather Service, 1990).
63. Fifth Assessment Report—IPCC. <https://www.ipcc.ch/assessment-report/ar5/> (2014).
64. Warszawski, L. et al. The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP): Project framework. *Proc. Natl Acad. Sci. USA* **111**, 3228–3232 (2014).
65. Hempel, S., Frieler, K., Warszawski, L., Schewe, J. & Piontek, F. A trend-preserving bias correction—the ISI-MIP approach. *Earth System Dyn.* **4**, 219–236 (2013).
66. Jerome, F., Trevor, H. & Robert, T. *The Elements of Statistical Learning* ISBN: 978-1-282-82726-4 (2019) (Springer, Dordrecht, 2009).
67. Kapelner, A. & Bleich, J. bartMachine: machine learning with Bayesian additive regression trees. *J. Stat. Software* **70**, 1–40 (2013).

## Acknowledgements

We would like to acknowledge the Purdue Climate Change Research Center (PCCRC) and the Purdue Center for Environment (C4E) for their support. R.N. would also like to acknowledge support from NSF grants CRISP-1832688 and HDBE-1826161. We also acknowledge the efforts of different organization for making the data available for this work, which include the NCEP Reanalysis data, which is kindly provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website at <https://www.esrl.noaa.gov/psd/>; the World Climate Research Program's Working Group on Coupled Modeling and the ISI-MIP community for the CMIP5 simulations.

## Author contributions

D.M.-S. and R.K. implemented data collection and processing. D.M.-S. conceived the method and analysis, and led the manuscript write up. R.N. and R.K. assisted in the result analysis and provided overall guidance for the study. All authors reviewed and edited the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-15393-8>.

Correspondence and requests for materials should be addressed to D.M.-S. or R.K.

Peer review information *Nature Communications* thanks Colin Raymond and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020