Analyzing the climate sensitivity of the coupled water-electricity demand nexus in the Midwestern United States

R Obringer[1,*], R Kumar[2] and R Nateghi[1,3]

[1] Environmental and Ecological Engineering, Purdue University, 500 Central Drive, West Lafayette, Indiana, 47907, USA

[2] Department Computational Hydrosystems, Helmholtz Centre for Environmental Research - UFZ, Permoserstr.15, 04318 Leipzig, Germany

[3] School of Industrial Engineering, Purdue University, 315 N. Grant Street, West Lafayette, Indiana, 47907, USA

[*] Corresponding Author. Email: robringe@purdue.edu

Abstract

Accounting for the nexus between water and electricity demand is critical for ensuring efficiency and conservation measures are successful in lowering the net water and electricity use in a city. Considering the nexus is also critical for accurately estimating the price elasticity of demand and designing effective demand response programs. The importance of the water-electricity demand nexus is rapidly increasing as cities are stressed by factors such as global climatic and socioeconomic changes as well as unprecedented rates of urbanization and growth. Despite the extensive recent research efforts on electricity and water demand modeling, significant knowledge gaps remain that are primarily rooted in (i) the use of univariate approaches that cannot adequately account for the nexus and (ii) the lack of a comprehensive assessment of the role of climate drivers on the demand nexus. To address these gaps, we propose a multivariate (i.e., multi-response), algorithmic framework for assessing the climate-sensitivity of the coupled water-electricity demand nexus. To illustrate the applicability of the proposed framework, we selected six Midwestern cities as our test cases. The results indicated that climate variability alone could account for 23%–71% of variability in the water-electricity demand nexus with the seasonally adjusted dataset, and 47%–87% of the variability on the non-adjusted dataset. The results also revealed that water use was more climate-sensitive than electricity use. Additionally, we demonstrated the importance of the variability in the global climate drivers such as the El Niño/Southern Oscillation cycle, which explained the majority of the covariance in the water-electricity nexus. Our modeling results suggest that stronger El Niños lead to an overall decrease in the climate-sensitive portion of the water and electricity use in the selected cities. Since the El Niño cycle is a well-documented phenomenon, it would be possible for utility managers to make broad predictions about the following season's water and electricity demand and plan accordingly. Moreover, as the climate continues to warm, it is likely that El Niños will become stronger, potentially leading to a reduction in the climate-sensitive portion of the water-electricity demand nexus in the Midwest.

Keywords: Water-Energy Nexus, Multivariate Tree Boosting, Climate Sensitivity, Multidimensional Modeling

1. Introduction

The water-electricity nexus is a concept dating back to the late 1980's, however applying the concept to urban areas began around 2010's [1]. Since the release of these studies and reports, there have been many initiatives surrounding the water-electricity nexus calling for researchers to evaluate the nexus and its impacts at various spatiotemporal scales and for numerous applications. The idea behind studying the nexus, as opposed to studying water and/or electricity in isolation, is that the two systems are interrelated and studying them separately will likely lead to (i) attenuated effects in efficiency and conservation programs to reduce residential energy and water consumption, (ii) overestimating price elasticity of demand, and (iii) designing ineffective demand response programs. On the other hand, considering their co-benefits in conservation measures has demonstrated potential to achieve savings at no net cost in some regions [2].

There are a variety of ways to study the water-electricity nexus, including water for electricity analyses and electricity for water analyses. To understand water for electricity, researchers frequently evaluate the water that is used during electricity generation [3]. An estimated 90% of the electricity in the US comes from thermoelectric power plants, which require water for cooling [4]. The amount of water withdrawn by these plants accounted for 40% of the water withdrawals in the US during 2005 [5], making these plants a crucial aspect to studying water availability in the US, especially during heatwaves and droughts. Higher temperatures and drought conditions have been shown to increase electricity demand, which ultimately leads to increased water withdrawals by electricity generators [4]. Electricity for water analyses, on the other hand, focus on quantifying the electricity it takes to treat and distribute water [3]. It was estimated that in 2012, water utilities in the United States consumed 38,100 GWh of electricity [6], which will likely increase as utilities continue to expand to keep up with urban growth. Given that water-related electricity use is expected to increase in states that are already water stressed, such as Florida, Texas, and Arizona [7], analyses that focus on the water-electricity nexus are becoming increasingly important.

Climate change will likely exacerbate the stress on urban water and electricity utilities, which are already facing unprecedented growth in many parts of the world, including the United States [8]. Water and electricity utilities depend on each other to maintain their respective services, but under the more variable conditions brought on by climate change, including higher temperatures and increased frequency and intensity of drought events [9, 10], utilities may begin to face challenges related to their supply. For example, as mentioned earlier, higher temperatures will increase the demand for cooling in thermoelectric power plants, which will lead to more water withdrawals by the power plants [4]. The higher temperatures and increased frequency of droughts will also put pressure on water resources and the utilities that own them to provide water for public supply and any other major users, including thermoelectric power plants. This pressure could result in temporary reductions in electricity production, such as those that have occurred in a few European countries in the past few years [11]. In this sense, the electricity sector puts pressure on the water sector by requiring a large amount of water supply, and the water sector puts pressure on the electricity sector when there are shortages. This will be compounded by climate change, ultimately putting additional pressure on both sectors.

In addition to the supply-based (inter)dependencies discussed above, there are many aspects of water and electricity use that are interconnected. For example, watering landscapes, washing clothes, taking hot showers, and using a dishwater all require both water and electricity. These dependencies are critical for both electric and water utilities trying to reduce peak load to lower the likelihood of supply inadequacies and service disruption risks, and reduce operations and maintenance cost [12].

In comparison to the studies of water-electricity *supply* nexus, research on the water-electricity *demand* nexus is more nascent [13]. The majority of the work on the demand-side has primarily focused on human behavior and specific tasks (e.g., heating water, using a dishwasher, or landscaping) [14, 15]. These studies provide a wealth of information on people's behaviors and the coupling between the urban water and electricity systems, but there is very little work on the subject that takes climate variability and change into account. The handful of studies that do consider climate, employ only simple and limited measures (e.g., change in precipitation or temperature) to determine the impact [16, 17, 18]. The climate measures impacting the water-electricity nexus likely go beyond simple measures such as precipitation and temperature that have yet to be explored. In particular, the El Niño/Southern Oscillation cycle, which has been shown to impact the water-energy-food nexus [19], has not been included in urban water-electricity demand nexus studies. Moreover, the majority of the existing studies have not harnessed a multivariate approach to simultaneously estimate the water and electricity demand as a function of exogenous factors such as climate variability and change.

The purpose of this study is to bridge these gaps by proposing a multivariate paradigm to harness the dependencies in the urban water and electricity demand data and allow for simultaneously estimating the climate-sensitive portion of the water demand, electricity demand, and their nexus. Given the focus on the climate-sensitive portion of the demand nexus, only climatic variables were used as predictors in the study. The central goal of this paper is to comprehensively assess the climate sensitivity of the urban water-electricity demand nexus, which has largely been overlooked in previous studies. Our proposed framework is designed to handle multiple interdependent response variables. Since the coupled water-electricity nexus model takes the correlation between the response variables into account, it was hypothesized that this multivariate modeling framework would predict the water and electricity use better than similar univariate models. To test this hypothesis, we applied our framework to six large-range cities in the Midwestern United States and evaluated the impacts of climate variability on the demand nexus. We also hypothesized that both local climatic variables, such as precipitation and temperature, and large climatic drivers, such as the El Niño/Southern Oscillation index, would be important predictors of end-use demand for water and electricity.

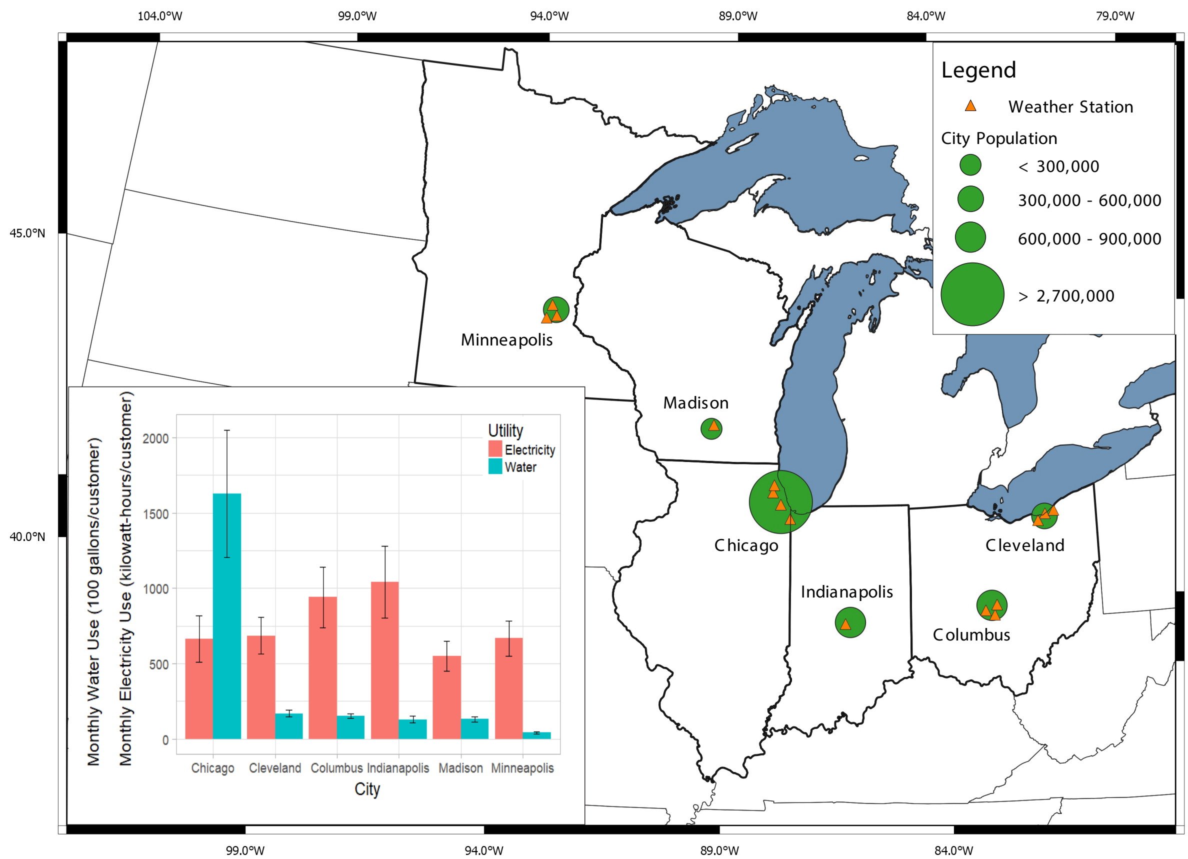

Figure 1: A map of the cities chosen for this study. From left to right: Minneapolis (MN), Madison (WI), Chicago (IL), Indianapolis (IN), Columbus (OH), and Cleveland (OH). The locations of the weather stations used to collect the meteorological data are also included. The inset plot shows the mean $\pm$ one standard deviation for both water and electricity use for each city. The water use is in 100 gallons/customer (or metered account) and the electricity use is in kilowatt-hours/customer.

2. Data and Methods

To demonstrate the applicability of the proposed approach, we used the Midwest region in the United States as a case study. In this section, we will first describe the study sites and the input data used for the analyses presented in this paper, and will then delve into the proposed methodology for assessing the coupled water-electricity nexus in the case study areas.

2.1. Site Description

In this study, we focused on the northern and eastern parts of the Midwest, including Ohio, Indiana, Illinois, Wisconsin, and Minnesota. Within this study area, depicted in Figure 1, six cities of varying population sizes were selected: Chicago (IL), Columbus (OH), Indianapolis (IN), Minneapolis (MN), Cleveland (OH), and Madison (WI). These cities were selected in order to capture a variety of different sizes, while still focusing on some of the most populous cities in the region. Moreover, each city, though they have different demand patterns, will likely experience similar impacts of climate change due to their geographical proximity. In particular, it is likely that the Midwest region as a whole will have higher temperatures and more precipitation as CO_2 levels continue to rise [20], which will in turn affect the urban water-electricity demand nexus.

2.2. Data Description and Preprocessing

The data for this study was obtained from four main sources—the US Energy Information Administration (EIA), National Centers for Environmental Information (NCEI), National Oceanic and Atmospheric Administration (NOAA), and local water utilities. Specifically, monthly residential electricity use was obtained from the EIA [21], meteorological and climate data from the NCEI [22] and NOAA [23], and residential water use was obtained through records requests to local water utilities. The meteorological data was collected from several meteorological towers stationed around each city and aggregated to get an average monthly value for each city between 2007 and 2016. Specifically, there were four active towers in Chicago, Columbus, and Minneapolis, three in Cleveland, and one in Indianapolis and Madison (see Figure 1). Meteorological variables used in the analysis included temperature (dry bulb and dew point), relative humidity, wind speed, and precipitation. The El Niño/Southern Oscillation strength index was also included in the analysis, as a large-scale climatic driver that has been shown to impact the climate of the Midwest [24].

In this study, there were two response variables: residential electricity use and residential water use, both normalized by the number of customers reported by the utility. Often water and electricity are provided by separate utilities, with potentially different service areas, this normalization allowed us to compare these two variables regardless of the differences in service area. Additionally, the response data was adjusted for seasonality to ensure that our results were demonstrating the effect of climate on the water-electricity demand nexus, independent of the natural seasonality present in the usage patterns. In the seasonality adjustment, the time series were decomposed and the seasonality components were subtracted from the original time series [25] (see Supplemental Methods for more information). There were also eight meteorological and climatic predictors (see Table 1), that were included in the initial model run. There was a focus on variables that are easily measured by meteorological stations due to the availability of such data, as well as the results of previous studies, which showed the importance of meteorological variables on water and electricity demand [26, 27, 28?]. Both average and maximum values of meteorological variables were included to establish which statistic (i.e., maximum or mean) would better capture the intensity of the signals in the water and electricity demand data. Similarly, it has been shown that the El Niño/Southern Oscillation plays an important role in affecting hydroclimatic processes across the US [24, 29, 30], making it an important variable to include in the analysis of the climate impact on residential water and electricity use.

2.3. Methodology

The interconnectivity between water and electricity use has been well documented throughout the literature [1, 6, 7], with a few studies focusing on the impacts of climate [14, 16]. However, this is the first time,

Table 1: The input variables used for developing the coupled water-electricity demand nexus model. Each variable was collected at the city-scale from January 2007 through December 2016 and aggregated to ensure a consistent monthly time scale.

Variable Type	Variable Name	Units	Source
Response (2007-2016)	Monthly Water Use (normalized)	gal.	Local Utilities
	Monthly Electricity Use (normalized)	MWh	EIA-861M [21]
Predictor (2007-2016)	Average Maximum Dry Bulb Temperature	°F	NCEI [22]
	Average Dew Point Temperature	°F	NCEI [22]
	Average Relative Humidity	%	NCEI [22]
	Average Maximum Relative Humidity	%	NCEI [22]
	Average Wind Speed	mph	NCEI [22]
	Average Maximum Wind Speed	mph	NCEI [22]
	Accumulated Precipitation	in	NCEI [22]
	El Niño/Southern Oscillation index	–	NOAA [23]

to our knowledge, that the impact of climate on the water-electricity nexus has been evaluated through a *multivariate* framework based on statistical learning theory. The advantages of this framework include (i) assessing the role of a wider range of climatic variables on the water-electricity demand nexus than previous studies, and (ii) leveraging a robust, non-parametric technique to assess the climate-sensitivity of both water and electricity use simultaneously, while taking their complex and non-linear interactions into account. Moreover, the required inputs to the modeling framework are readily available, such that utility managers, researchers, or other interested parties can easily apply the model to their city or cities of interest.

There are four main steps in the modeling process: (1) data collection, preprocessing and aggregation, (2) model training and testing, (3) statistical inferencing, and (4) comparative analysis with a univariate model. A schematic of this process can be seen in Figure 2. The first step was to collect the data, normalize the response variables and implement seasonality adjustments (as described in Section 2.2), and to aggregate the meteorological data spatially across weather stations and temporally from daily to monthly values. The initial model training and testing was performed—within a 5-fold cross validation loop—with all the predictor variables (see Table 1). Cross validation was used for both model hyperparameter tuning as well as model performance assessment. The initial model runs were then followed by a variable selection step to establish the key predictors (see section 2.3.3 for more details). Finally, the statistical inferencing was

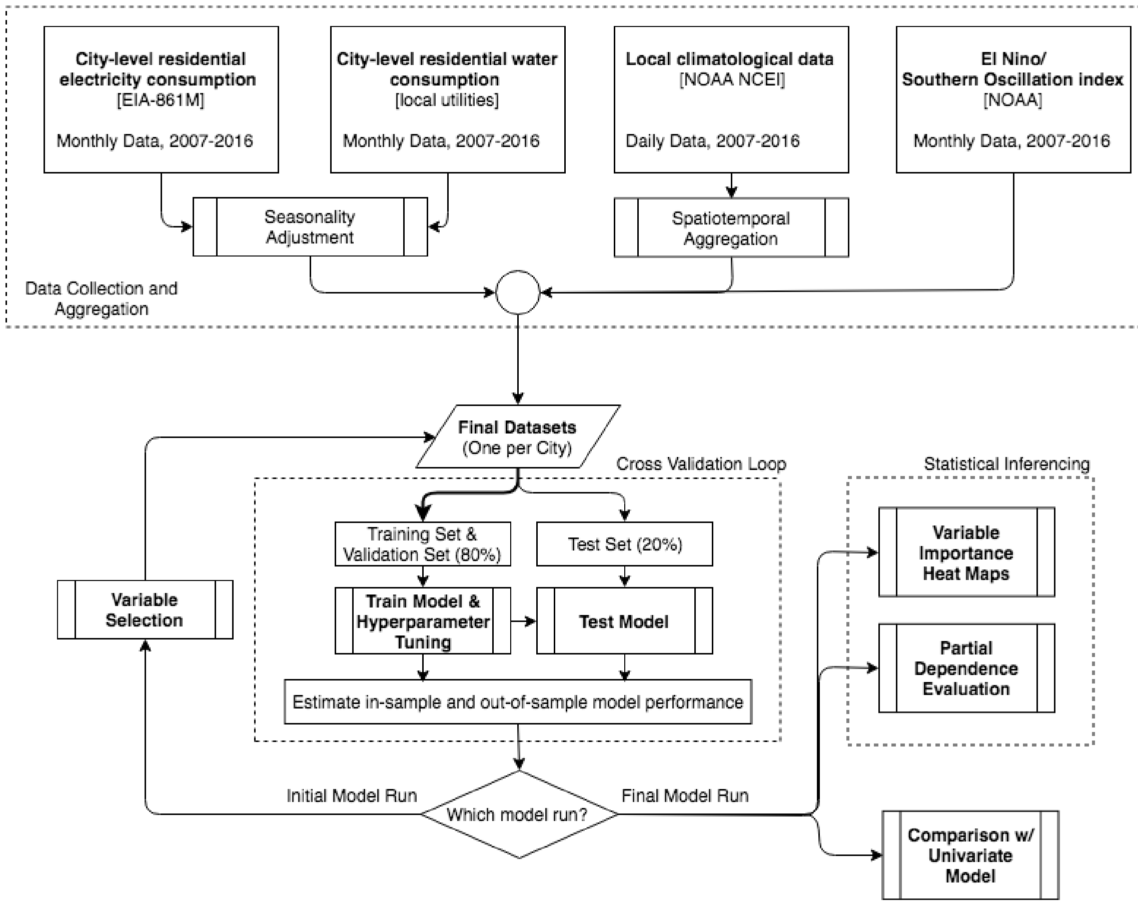

Figure 2: Schematic of the modeling process.

performed using the results from the final best model that included the reduced input variable set, based on the variable selection step (see section 2.3.4 for more details). Each of these steps will be described in further detail in the following sections.

2.3.1. Supervised Learning Theory

The algorithm used throughout this study fall into a larger category of statistical learning theory known as ‘supervised learning’. Supervised learning algorithms are built to predict target variable(s) of interest (i.e., the response variable(s)), given a number of predictor variables. Supervised learning can be mathematically described as:

$$Y = f(X) + \epsilon \tag{1}$$

where Y is the response variable(s) of interest, X is the series of predictor variables used to predict the response, and ϵ is the irreducible error ($\epsilon \sim N(0, \sigma^2)$) [31]. In supervised learning, the aim is to predict the response variable(s) such that the the expected error is minimized as shown below [31].

$$\min \frac{1}{N} \sum_{i}^{N} \Delta[\hat{f}(X_i), f(X_i)] \tag{2}$$

Here $\hat{f}(X_i)$ and $f(X_i)$ represent the estimated and true functions, respectively, and Δ represents some measure of distance (e.g. the Euclidean or Manhattan distance).

Among the wide library of supervised learning algorithms, tree-based methods are one of the most popular non-parametric learning techniques [31]. Tree-based models offer competitive predictive accuracy compared to most of the state-of-the art statistical machine learning algorithms [32], and lend themselves more easily to interpretation and inferencing compared to other "black box" algorithms, such as deep learning and support vector machines [31]. In this paper, we used a multivariate extension of an ensemble-of-trees approach, as described below.

2.3.2. Algorithm Description

We used an advanced supervised learning technique—based on an ensemble-of-trees approach—that leverages the covariance structure of multiple response variables to better estimate the complex interactions between the target variables. Specifically, the predictive model of the coupled residential water and electricity demand was developed based on a multivariate extension of the gradient boosted regression trees algorithm [33].

Gradient boosted regression trees is an ensemble-of trees method that takes advantage of the boosting meta-algorithm to increase the predictive accuracy [33]. The boosting meta-algorithm works by sequentially fitting models (in this case decision trees), where in each iteration more weight is given to the better classifiers and the misclassified points in order to reduce the overall loss function and enhance the predictive accuracy. Boosting is represented mathematically in the equation below.

$$G(x) = \sum_{m}^{M} \alpha_m C_m(x) \tag{3}$$

Here $G(x)$ is the final ensemble model, M is the total number of iterations to be completed, α_m is the weight of each prediction, and C_m is the tree models fitted to the input variable x at iteration m.

In this paper, we leverage multivariate tree boosting which extends gradient boosted regression trees to a multivariate (i.e., multi-response) case. Thus, the multivariate extension of the algorithm enables the simultaneous prediction of multiple response variables [34]. Specifically, this algorithm iteratively builds trees by minimizing the squared error loss for each response variable and maximizing the covariance discrepancy in the multivariate response. In other words, at each iteration, a prediction is made for each response variable, such that the loss function is minimized and the covariance discrepancy between the current and previous predictions is maximized. This allows each subsequent prediction to be incrementally more accurate than the previous, while ensuring the predictors that account for the most covariance in the nexus of the response variables are selected. The steps of the algorithm are summarized below:

Algorithm 1 Multivariate Ensemble Tree Boosting Algorithm D [34]

1: **for** m *in* $1, \ldots, M$ steps (regression trees) **do**
2: **for** r in $1, \ldots, R$ quantitative response variables (e.g., water and electricity demand) **do**
3: train tree $m^{(r)}$ to residuals, and estimate the covariance discrepancy $D_{m,r}$
4: **end for**
5: Select the response $y^{(r)}$ corresponding to the regression tree that yielded the maximum $D_{m,r}$
6: Update residuals by subtracting the predictions of the tree fitted to $y^{(r)}$, multiplied by step-size.
7: **end for**

This algorithm has been tested in a few multivariate predictive applications, ranging from psychological well-being [34] to multi-dimensional infrastructure resilience assessment [28], and we hypothesized would be a good candidate for energy-water nexus modeling.

2.3.3. Variable Selection

Per Occam's razor, we aimed to establish the simplest model (containing a subset of input variables) that best captured the data dependencies and covariance. In other words, we conducted variable selection to reduce model complexity via retaining only the most important or influential predictors in the final model. In our framework, variable selection was based on establishing the relative influence of each variable, via measuring the sum of squared errors obtained on any split of a given predictor, summed over all trees in the the prediction model [31]. The calculated sums of squared errors provide a basis for ranking the predictor variables. Thus, the relative influence is related to the amount of reduction in total error that can be attributed to a given predictor—the higher the reduction in error, the more influential (and important) the variable is in the model. For multi-dimensional response variables, the univariate relative influence is first measured for each independent variable and for each response. Summing the importance over all response variables renders a 'global' measure of influence for the independent variables across all target variables.

In this study, the variables were selected for the final model if they had a relative influence greater than 5% in at least 4 of the 6 cities. Using this threshold, we retained the following five predictors in the final model: average maximum dry bulb temperature, average dew point temperature, average relative humidity, average wind speed, and the El Niño/Southern Oscillation index. These variables were used in the final model run and subsequent analyses/inferencing.

2.3.4. Statistical Inferencing and Analyses

The statistical inferencing for the multi-dimensional water-electricity nexus model—developed using the multivariate tree boosting algorithm described in section 2.3.2—was conducted using the following methods: (1) evaluating the model performance (i.e., model goodness-of-fit and predictive accuracy), (2) assessing the covariance explained by each predictor on individual response variables and identifying the clusters of input variables that jointly influence one or both response variables, (3) visualizing the partial dependence between the important predictors and the response variables, and (4) comparing the multivariate model performance to a similar univariate model.

- *Model Performance*

To evaluate model fit and predictive accuracy, the algorithm was run—within the 5-fold cross validation loop—for each city simultaneously (Figure 2), resulting in one prediction per city per response variable. The performance of the model was assessed using two statistical measures: the out-of-sample root-mean-squared error (RMSE) and the out-of-sample coefficient of determination (R^2). RMSE provides a measure of error that heavily penalizes large deviations, making it ideal for prediction applications. The out-of-sample R^2 value demonstrates the fit of the model predictions made by the test dataset.

- *Heat Maps of the Covariance Structure*

The leveraged algorithm can help identify the pairs of the predictor variables that explain the variance in individual response variables and/or the covariance between multiple response variables. The hierarchical clustering technique can then be used to group the predictors that explain covariance in similar pairs of response variables, and the pairs of responses that are dependent on similar subsets of predictors; the results can then be illustrated as a heat map [34].

- *Partial Dependence*

A crucial aspect of statistical inferencing is determining the nature of the statistical relationship between the most important predictors and the response variables. For non-parametric models, partial dependency analyses are conducted to characterize the association between the inputs and the response variable(s). The partial dependence can be calculated using the following equation [31]:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \hat{f}(x, x_C^{(i)}) \tag{4}$$

Where x is the predictor of interest and $x_C^{(i)}$ represents the other predictor variables that are not of interest. The estimated partial dependence, $\hat{f}(x)$, is the average value of the response variable, when only the predictor variable of interest is considered.

- *Model Comparison*

Finally, the results from the multivariate model were compared to results from a similar univariate model. Specifically, we used gradient tree boosting [33] to predict the water and electricity use as isolated variables. Gradient tree boosting is the basis for multivariate tree boosting [34], thus the main difference between the multivariate and univariate algorithms is the consideration of response variable dependencies. The purpose of this final analysis was to demonstrate the value of the multivariate framework, as this is the first time this coupled methodology has been applied to predicting the climate-sensitive portion of the water-electricity nexus.

3. Results

Following the modeling process outlined above (see Figure 2), we estimated the climate-sensitive portion of the interdependent water and electricity use for each city in the study area. In this section, we will first describe the model performance, then discuss the results from the various statistical inferencing techniques, including the covariance explained evaluations and the partial dependence visualizations, before describing the comparison between the multivariate and univariate model performance.

3.1. Model Performance

To develop a predictive model of interdependent urban water and electricity demand, we leveraged the multivariate tree boosting algorithm described in Section 2.3.2. In the initial training of model, several independent variables that could potentially affect water and/or electricity demand were included (see Table 1). The final model included a reduced variable set based on the relative influence each predictor had over the predictive accuracy. The variables in the final model included maximum dry bulb temperature, average dew point temperature, average relative humidity, average wind speed, and the El Niño/Southern Oscillation index. The selected variables were similar to previous studies on the sensitivity of water and electricity demand [26, 27, 35, 36].

3.1.1. Treatment of Seasonality

As part of the data preprocessing, the response variables were adjusted for seasonality. It has been shown that seasonality aids in the predictive accuracy, but in such a way that is misrepresentative of the actual system [25]. Here we present the results from the model performance using both the original dataset and the seasonally adjusted dataset. Without the seasonality adjustment, the model performance was better (see Table 2a and Figure 3a), however, the seasonality present in the data was likely masking the true signal from the model, thus the improved performance compared to the seasonally adjusted dataset (see Table 2b and Figure 3b).

3.1.2. Measures of Model Performance

The performance of the final model was assessed based on the out-of-sample estimates of the coefficient of determination (R^2) and the root-mean-squared error (RMSE). These measures of error were calculated using the test set. The results from the seasonally adjusted dataset are summarized in Table 2b, which demonstrates that climate variables alone can account for a significant fraction of variability in the electricity and water demand data (ranging from 43%-73% in the in-sample performance, and 30%-71% in the out-of sample performance, for the seasonally adjusted dataset). Thus, while the previous literature primarily focused on explaining the variance in the demand as a function of socioeconomic and technological factors as well as cultural norms, in this study, we focused on isolating the effects of climate variability and demonstrated the significant role of climate in explaining the covariance of the water-electricity demand nexus.

Table 2: The model performance for each city for the final model run. Table 2a shows the results using the original dataset (i.e., the dataset with seasonality intact) and Table 2b shows the results using the seasonally adjusted dataset. The in-sample measures were calculated using the same data used to train the model, while the out-of-sample measures were calculated using the test dataset, which was not included in the model training (see Figure 2).

Table 2a: Model performance for each city using the original dataset (i.e., the dataset with seasonality).

City	Water Use				Electricity Use			
	in-sample R^2	in-sample RMSE	out-of-sample R^2	out-of-sample RMSE	in-sample R^2	in-sample RMSE	out-of-sample R^2	out-of-sample RMSE
Chicago	0.71	0.333	0.47	0.731	0.85	0.235	0.76	0.499
Columbus	0.82	0.285	0.78	0.619	0.89	0.218	0.84	0.496
Indianapolis	0.89	0.222	0.83	0.491	0.94	0.160	0.87	0.385
Minneapolis	0.88	0.221	0.81	0.468	0.91	0.197	0.83	0.431
Cleveland	0.51	0.452	0.31	0.876	0.81	0.306	0.77	0.566
Madison	0.79	0.290	0.71	0.623	0.85	0.226	0.77	0.450

Table 2b: Model performance for each city using the seasonally adjusted dataset.

City	Water Use				Electricity Use			
	in-sample R^2	in-sample RMSE	out-of-sample R^2	out-of-sample RMSE	in-sample R^2	in-sample RMSE	out-of-sample R^2	out-of-sample RMSE
Chicago	0.69	0.344	0.51	0.720	0.53	0.457	0.39	0.932
Columbus	0.63	0.416	0.62	0.894	0.49	0.500	0.31	0.975
Indianapolis	0.73	0.327	0.71	0.739	0.53	0.455	0.41	0.934
Minneapolis	0.69	0.333	0.55	0.761	0.50	0.467	0.42	1.113
Cleveland	0.44	0.490	0.23	0.910	0.46	0.509	0.34	0.943
Madison	0.54	0.444	0.34	0.925	0.43	0.512	0.30	1.003

The results summarized in Table 2 indicate that a significant fraction of variability in the water-electricity demand nexus can be explained by the input climate variables. This is futher demonstrated in Figure 3, which shows the predicted values plotted against the actual values for both the original dataset (Figure 3a) and the seasonally adjusted demand data (Figure 3b). The results are illustrative of the fact that climate variability is an important driver of water and electricity use in Midwestern cities.

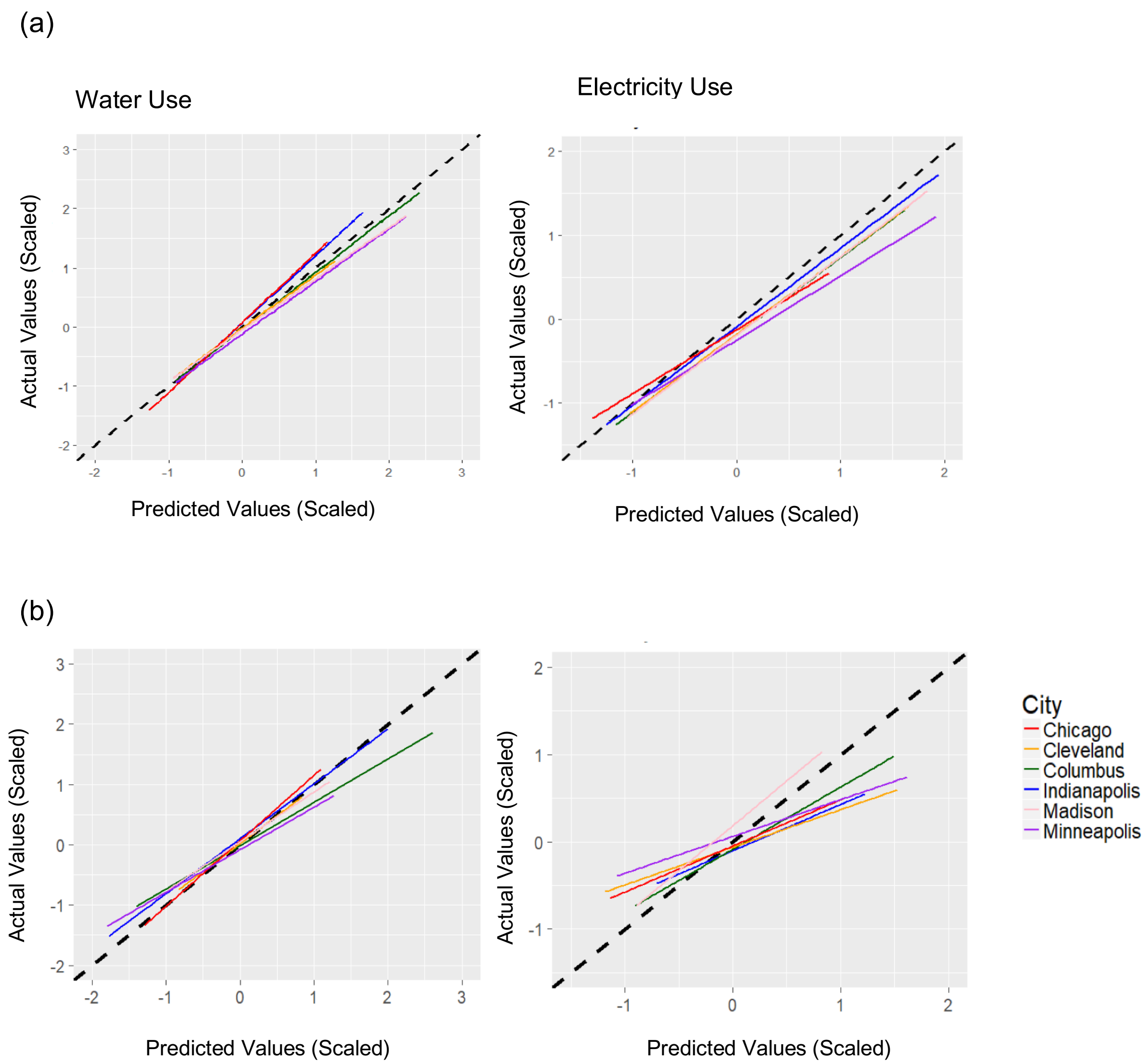

Figure 3: Out-of-sample model performance for (a) the original dataset (i.e., the dataset with seasonality) and (b) the seasonally adjusted dataset with the multivariate model. The response variables, water and electricity use, have been scaled to account for different units of measurement. The lines are best fit lines plotted through the predicted versus actual points, with a 45° dashed line for reference.

3.2. Statistical Inferences from the Multivariate Model

One of the advantages of our proposed multivariate approach is the ability to determine the covariance explained by the predictors for each individual response variable and the nexus between response variables. This feature allows us to see what variables have the most impact on the water-electricity nexus and if those variables differ from those most greatly impacting water or electricity use alone.

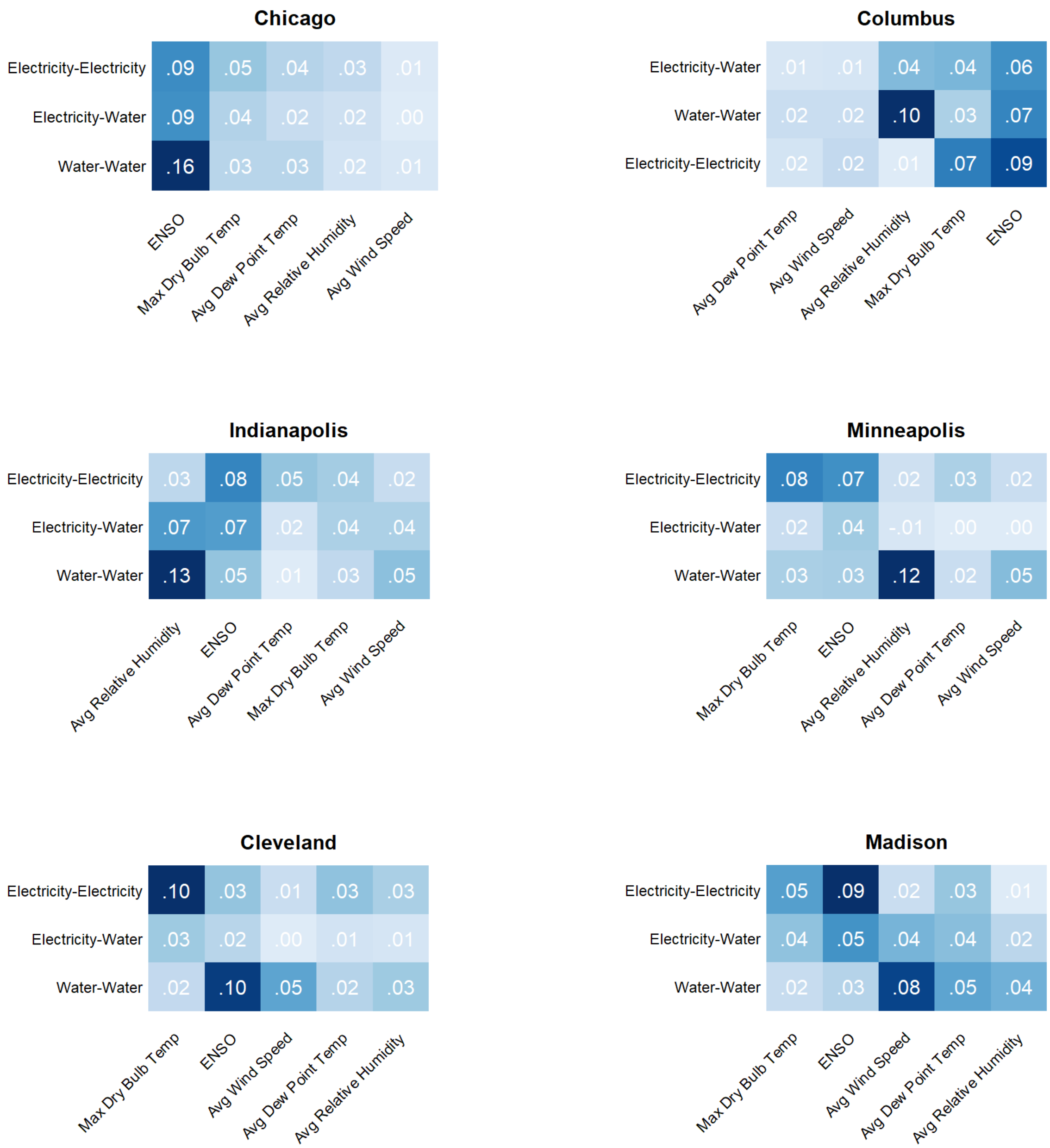

Figure 4: Clustered heat maps showing the covariance explained by each predictor variable in each city, after the seasonality was removed from the dataset. The darker blues represent higher values of covariance explained, while the lighter blues represent less. The variables have been grouped using hierarchical clustering, a method used to group similar objects together. In this figure, predictors clustered together explain the covariance in similar outcome pairs, therefore, the position of the variables on the axes is different for each city due to each city has a different clustering outcome.

Figure 4 shows the clustered heat maps of the covariance explained for each city. These heat maps are clustered via hierarchical clustering, which indicates which predictors are affecting the response variables in similar ways, as well as which response variables pairs are being influenced by similar subsets of predictors. Overall, assessing the covariance explained allows us to investigate the similarities and differences between the cities, as well as any differences between the isolated water use, isolated electricity use, and the water-electricity use nexus. The results from the heat maps demonstrate that although the model itself is generalizable across the different cities, as indicated by the model performance (see Table 2), the covariance explained by the variables will differ from city to city. For example, in the land-locked cities of Columbus, Indianapolis, and Minneapolis, average relative humidity explains the most covariance in water use. This is different than the coastal cities of Chicago and Cleveland, where the ENSO index explains much of the water use and relative humidity has less of an impact.

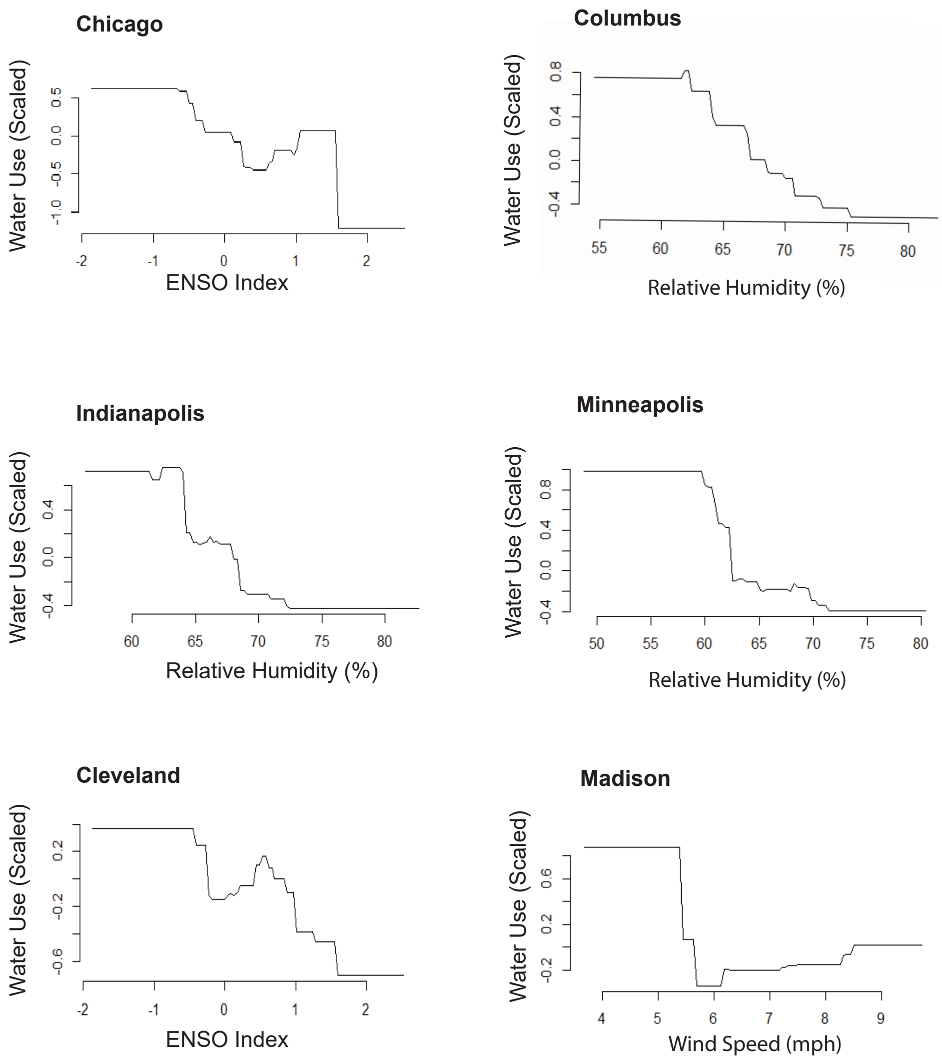

Figure 5: Partial dependence plots between the most important predictor variable and water use in each city. Note that the water use has been scaled, so there are no units.

The covariance explained, however, does not give any indication to the direction of the relationship between the predictors and the response variables—just the magnitude of it. Thus, it is necessary to perform other analyses to determine if higher relative humidity will lead to higher or lower water use in Indianapolis, for example. To answer this question, we evaluated the partial dependence of the predictors on the individual response variables. A selection of these partial dependence plots are shown in Figure 5 (additional partial dependence plots can be seen in Supplemental Figure S2).

These plots show the relationship between the most important variables and water use in each city. In particular we can see that in the cities of Columbus, Indianapolis, and Minneapolis, as relative humidity increases, the water use decreases. A similar pattern appears in Chicago and Cleveland—as the El Niño gets stronger, the water use decreases. This suggests that utility managers trying to reduce water use in Columbus or Indianapolis should focus on the days with intermediate relative humidity, as that is when people are using the most water. Likewise, a manager in Chicago or Cleveland should focus their demand reduction efforts during the cold phase of the El Niño cycle (i.e., La Niña).

3.3. Univariate Model Comparison

One of the goals of this work was to demonstrate the power of including both water and electricity use in the model as interdependent response variables. This was done through a model performance comparison of the multivariate tree boosting model and a univariate version: gradient tree boosting. The results from the univariate model run are shown in Table 3.

Table 3: The in-sample and out-of-sample model performance (R^2 and RMSE) of the univariate model, gradient tree boosting, for each city after the seasonality was removed from the data.

City	Water Use				Electricity Use			
	in-sample R^2	in-sample RMSE	out-of-sample R^2	out-of-sample RMSE	in-sample R^2	in-sample RMSE	out-of-sample R^2	out-of-sample RMSE
Chicago	0.60	0.437	0.50	0.747	0.36	0.600	0.32	0.981
Columbus	0.55	0.500	0.53	0.860	0.36	0.601	0.26	0.987
Indianapolis	0.62	0.429	0.64	0.732	0.39	0.577	0.29	0.938
Minneapolis	0.56	0.451	0.55	0.756	0.32	0.599	0.32	1.003
Cleveland	0.34	0.614	0.36	0.860	0.30	0.630	0.28	0.991
Madison	0.41	0.548	0.37	0.883	0.28	0.663	0.28	1.036

Both approaches revealed that a significant fraction of the variability in the water and electricity use could be accounted for by climate variables alone. Additionally, the relative performance of the various cities matched between the univariate and multivariate models. For example, in both approaches, Indianapolis's water use was found to be most climate-sensitive, while Cleveland's revealed the least amount of climate sensitivity (based on their estimated coefficients of determination). Overall, however, the multivariate model was better at capturing the climate sensitivity of two demands than the univariate model, with the exception of Cleveland's and Madison's water use.

The main difference between the univariate and multivariate models was the inclusion of response variable interdependencies within the multivariate model. This is indicative that, in most cases, the consideration of the interconnectivity between water and electricity use improves the final prediction of both water and electricity use. Of the cities tested as a part of this analysis, the climate sensitivity of water use in Cleveland and Madison—smallest cities included in this study—were better accounted for by the univariate model, which suggests a loose coupling between the climate-sensitive portion of the water and electricity use in those cities than the other cities studied. Additional research is necessary to determine the reason behind this reduced coupling between the climate-sensitive portion of the water and electricity demand.

4. Discussion

This study focused on analyzing the water-electricity demand nexus based solely on climate variables. This allowed us to isolate the effect of climate on residential water and electricity use—a factor that is often not included in demand analyses. Our results show that water use is more climate-sensitive in most of the cities included. This suggests that water use is more dependent on the climate than electricity use, which is an interesting finding, given the documented increase in electricity with increasing temperatures in the Midwest [28].

Given that the model performance for the electricity sector was more impacted by the seasonality adjustment than the water sector, we conclude that in the Midwest, the long-term climatic conditions are more likely to drive changes in water use, while the short-term weather patterns are more likely to act as a driver for electricity use. That is not to say that climate is the only driver of changing water use, but rather it is a potentially important driver that has often been left out of many demand analyses. In this sense, water demand studies, which often focus on population, socioeconomic, and/or cultural factors, ought to also include climatic factors in their analyses. This will become especially important as we try to predict water demand under climate change.

One of the main findings of this study was the importance of the El Niño cycle on the residential water and electricity demand in the region of interest. The ENSO index was consistently among the predictors that explained the most covariance in the response variables. Given that the El Niño cycle is a well-documented climate phenomenon that can be predicted relatively easily, it is an ideal variable for making more general or broad predictions. For example, a common ENSO-based prediction is the type of winter that a given region will have (e.g., a strong El Niño usually leads to warmer, drier winters in the Midwest [24]). Our modeling framework allows us to make a simple, first order forecast for the demand nexus based on large scale climate predictor. In other words, our results suggest that a strong El Niño is more likely to lead to lower water and electricity use . This knowledge would allow utility managers to prepare for the upcoming season based on the predicted El Niño strength that is determined on a monthly basis. The importance of the ENSO index also has implications for climate change. It is likely that El Niños will become stronger as sea surface temperature continues to increase [37], and our results suggest that if this holds true, water and electricity use in the Midwestern cities studied, will decrease as a result of the change in climate, should everything else in the cities remain constant. This assumption—that the population, socioeconomic breakdown, culture, etc. of a city will remain constant—is, of course, highly unlikely; however, the results demonstrate the importance of including climate variables in the overall analysis of water and electricity demand.

Finally, one of the goals of this study was to compare the results from the multivariate model, which considers the coupling between water and electricity demand, and a univariate model that is based on the same algorithm. Our results demonstrate that the multivariate framework is able to better capture the climate-sensitivity of water and electricity use in most cases. Since both models were based on the same algorithm, the only difference between them being the inclusion of multiple interconnected response variables, our results suggest that system coupling are an important consideration for the prediction of water and electricity demand. Ultimately, our results indicate that there needs to be an increased effort to (i) consider the increasing role of climate drivers on demand and (ii) harness a multivariate framework to better account for the interdependent response variables in demand analyses.

5. Conclusions

The purpose of this study was to build a multi-response predictive model of the portion of the urban residential water-electricity demand nexus that was sensitive to climate, using the multivariate tree boosting algorithm. In this study, there were two response variables: water use and electricity use, and five main predictors. The model was tested on six Midwestern cities of variable size, demonstrating the generalizability of the model to the region of interest. The results of the study indicated that a significant fraction of the water-electricity demand nexus can be explained by climate variability alone. Urban water and electricity demand are impacted by a number of factors, including population density, socioeconomic status, and cultural values, in addition to the climate. However, the role of climate has been understudied in comparison to other important drivers of urban water and electricity demand. For this reason, the goal in this study was to isolate the effects of climate and demonstrate the value of their inclusion in future analyses. Our results indicated that water and electricity use are sensitive to climate variables, and will likely be affected by future climate change. The impact of the El Niño cycle was especially important in each city, as the variable consistently explained much of the covariance in the water-electricity nexus and in the individual response variables.

Our proposed framework can be used by utility managers that are interested in tailoring conservation interventions to the times at which they will be most effective. For example, focusing on conservation during the cold cycle of the El Niño (i.e., La Niña) will likely be more effective and result in greater reductions than the same campaign during a strong El Niño. Finally, we compared the model performance to a similar univariate algorithm, known as gradient tree boosting. Our results demonstrated that in the majority of cities studied, the multivariate (i.e., multi-response) algorithm outperforms the univariate version. Since the main difference between the algorithms is the inclusion of multiple interdependent response variables, we recommend that future studies, especially in the Midwest, focus on modeling the water-electricity nexus, even if they are only interested in one of the response variables. Although the focus of this study was to isolate and analyze the effect of climate variables on the water-electricity nexus, the framework could easily be expanded to included other important factors, such as socioeconomic status, housing characteristics, or population density. Moreover, while we focused on modeling the water-electricity demand nexus, our proposed framework could be easily extended to include other critical urban services (e.g., food).

Acknowledgement

The authors would like to acknowledge the Purdue Climate Change Research Center (PCCRC) and the Purdue Center for Environment (C4E) for their support. Roshanak Nateghi would also like to acknowledge support from NSF grants #1826161 and #1728209.

References

[1] J. P. Newell, B. Goldstein, A. Foster, A 40-Year Review of the Food-Energy-Water Nexus Literature with a Focus on the Urban, Environmental Research Letters.

[2] M. D. Bartos, M. V. Chester, The conservation nexus: Valuing interdependent water and energy savings in Arizona, Environmental Science and Technology 48 (4) (2014) 2139–2149. doi:10.1021/es4033343.

[3] S. Derrible, Urban infrastructure is not a tree: Integrating and decentralizing urban infrastructure systems, Environment and Planning B: Urban Analytics and City Science 44 (3) (2017) 553–569. doi:10.1177/0265813516647063.

[4] B. R. Scanlon, I. Duncan, R. C. Reedy, Drought and the water-energy nexus in Texas, Environmental Research Letters 8 (4). doi:10.1088/1748-9326/8/4/045033.

[5] J. F. Kenny, N. L. Barber, S. S. Hutson, K. S. Linsey, J. K. Lovelace, M. a. Maupin, Estimated Use of Water in the United States in 2005 Circular 1344, no. Circular 1344, 2009. doi:http://dx.doi.org/10.3133/cir1405.

[6] C. M. Chini, A. S. Stillwell, The State of U.S. Urban Water: Data and the Energy-Water Nexus, Water Resources Research 54 (3) (2018) 1796–1811. doi:10.1002/2017WR022265.

[7] K. T. Sanders, M. E. Webber, Evaluating the energy consumed for water use in the United States, Environmental Research Letters 7 (3). doi:10.1088/1748-9326/7/3/034034.

[8] The World Bank, Cities and Climate Change: An Urgent Agenda, Tech. rep., The World Bank (2010).

[9] A. Dai, Drought under global warming: A review, Wiley Interdisciplinary Reviews: Climate Change 2 (1) (2011) 45–65. doi:10.1002/wcc.81.

[10] L. Damberg, A. AghaKouchak, Global trends and patterns of drought from space, Theoretical and Applied Climatology 117 (3-4) (2014) 441–448. doi:10.1007/s00704-013-1019-5.

[11] M. T. H. van Vliet, S. Vögele, D. Rübbelke, Water constraints on European power supply under climate change: impacts on electricity prices, Environmental Research Letters 8 (3) (2013) 035010. doi:10.1088/1748-9326/8/3/035010.

[12] S. F. Hoque, Water Conservation: The Answer to Water Shortage in the 21st Century, in: Water Conservation in Urban Households: Role of Prices, Policies and Tec1hnologies, IWA Publishing, 2014, pp. 1–24.

[13] A. Escriva-Bou, J. R. Lund, M. Pulido-Velazquez, R. Hui, J. Medellín-Azuara, Developing a water-energy-GHG emissions modeling framework: Insights from an application to California's water system, Environmental Modelling and Software 109 (February) (2018) 54–65. doi:10.1016/j.envsoft.2018.07.011.

[14] A. K. Plappally, J. H. Lienhard V, Energy requirements for water production, treatment, end use, reclamation, and disposal, Renewable and Sustainable Energy Reviews 16 (7) (2012) 4818–4848. doi:10.1016/j.rser.2012.05.022.

[15] D. M. Ruddell, P. G. Dixon, The energy-water nexus: are there tradeoffs between residential energy and water consumption in arid cities?, International journal of biometeorology 58 (7) (2014) 1421–1431. doi:10.1007/s00484-013-0743-y.

[16] G. Venkatesh, A. Chan, H. Brattebø, Understanding the water-energy-carbon nexus in urban water utilities: Comparison of four city case studies and the relevant influencing factors, Energy 75 (2014) 153–166. doi:10.1016/j.energy.2014.06.111.

[17] K. L. Lam, S. J. Kenway, P. A. Lant, Energy use for water provision in cities, Journal of Cleaner Production 143 (2017) 699–709. doi:10.1016/j.jclepro.2016.12.056.

[18] N. Mostafavi, F. Gándara, S. Hoque, Predicting water consumption from energy data: Modeling the residential energy and water nexus in the integrated urban metabolism analysis tool (IUMAT), Energy and Buildings 158 (2018) 1683–1693. doi:10.1016/j.enbuild.2017.12.005.

[19] D. Conway, E. Archer van Garderen, D. Deryng, S. Dorling, T. Krueger, W. Landman, B. Lankford, K. Lebek, T. Osborn, C. Ringler, J. Thurlow, T. Zhu, C. Dalin, Climate and southern Africa's water–energy–food nexus, Nature Climate Change 5 (9) (2015) 837–846. doi:10.1038/nclimate2735.

[20] Intergovernmental Panel On Climate Change, Climate Change 2013: The Physical Science Basis, Tech. rep. (2013).

[21] US Energy Information Administration, Form EIA-861M Sales and Revenue Data (2019).

[22] NOAA National Centers for Environmental Information, Local Climatological Data (LCD) (2010).

[23] K. Wolter, M. S. Timlin, Measuring the strength of ENSO events: How does 1997/98 rank?, Weather 53 (9) (1998) 315–324.

[24] D. Kluck, M. Woloszyn, El Niño Impacts and Outlook: Midwest Region, Tech. Rep. February, National Oceanic and Atmospheric Administration (2016).

[25] R. J. Hyndman, G. Athanasopoulos, Time Series Decomposition, in: Forecasting Principles and Practice, 2nd Edition, 2018.

[26] R. C. Balling, P. Gober, N. Jones, Sensitivity of residential water consumption to variations in climate: An intraurban analysis of Phoenix, Arizona, Water Resources Research 44 (10) (2008) 1–11. `doi:10.1029/2007WR006722`.

[27] L. House-Peters, B. Pratt, H. Chang, Effects of urban spatial structure, sociodemographics, and climate on residential water consumption in Hillsboro, Oregon, Journal of the American Water Resources Association (JAWRA) 46 (3) (2010) 461–472. `doi:10.1111`.

[28] R. Nateghi, S. Mukherjee, A multi-paradigm framework to assess the impacts of climate change on end-use energy demand, PLoS ONE 12 (11). `doi:10.1371/journal.pone.0188033`.

[29] S. A. Changnon, S. D. Hilberg, K. E. Kunkel, El Niño 1997-1998 in the Midwest, Tech. rep. (2000).

[30] R. Obringer, R. Natgehi, Predicting Urban Reservoir Levels Using Statistical Learning Techniques, Scientific Reports 8 (5164).

[31] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, 2nd Edition, Springer-Verlag New York, 2009. `arXiv:arXiv:1011.1669v3`.

[32] R. Caruana, A. Niculescu-Mizil, An Empirical Comparison of Supervised Learning Algorithms, Proceedings of the 23rd International Conference on Machine Learning`doi:10.1145/1143844.1143865`.

[33] J. H. Friedman, Greedy function approximation: A gradient boosting machine, Annals of Statistics 29 (5) (2001) 1189–1232. `arXiv:arXiv:1011.1669v3`, `doi:DOI10.1214/aos/1013203451`.

[34] P. J. Miller, G. H. Lubke, D. B. McArtor, C. S. Bergeman, Finding structure in data using multivariate tree boosting, Psychological Methods 21 (4) (2016) 583–602. `doi:10.1037/met0000087`.

[35] L. Raymond, D. Gotham, W. McClain, S. Mukherjee, R. Nateghi, P. Preckel, P. Schubert, S. Singh, E. Wachs, Climate Change Impacts on Indiana's Energy Demand and Supply, Climatic Change, Climatic Change.

[36] S. Mukherjee, R. Nateghi, A place-based, sectoral, data-driven approach to assess inadequacy risks in the electric power system under climate variability., Jouranl of Risk Analysis 39 (3) (2019) 673–694.

[37] S.-W. Yeh, J.-S. Kug, B. Dewitte, M.-H. Kwon, B. P. Kirtman, F.-F. Jin, El Niño in a changing climate, Nature 461 (2009) 511.