Impact of Climate Change on Residential Electricity in Western Cities in China

Cuihui Xia (✉ xiacuihui@itpcas.ac.cn)

Institute of Tibetan Plateau Research Chinese Academy of Sciences <https://orcid.org/0000-0002-5648-5014>

Tandong Yao

Institute of Tibetan Plateau Research Chinese Academy of Sciences

Haowen Hou

Tencent

Pengling Wang

China Meteorological Administration

Research Article

Keywords: residential electricity consumption (REC), Lanzhou, Lanzhou and Lhasa, climate adaptation

Posted Date: November 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-971158/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Extreme weather induced by climate change has triggered large-scale power outages worldwide. More insight into the climate impact (especially precipitation) on residential electricity consumption (REC) is needed. With a particular focus on precipitation, our study aimed to quantify the climate impact on REC and project associated changes under the representative concentration pathways (RCPs) 2.6, 4.5, and 8.5 climate change scenarios in Lanzhou and Lhasa, two western cities in China. The climate impact on REC in both cities is driven by heating-related demand, with Lanzhou being more sensitive to climate than Lhasa during the warm season. There is a stronger precipitation impact during the cold season in both cities, since precipitation (especially snowfall) boosts electricity consumption during the cold season. As the two cities become warmer and wetter, precipitation will offset the impact of warming on REC, with Lanzhou being more affected. Based on Lanzhou projections, the offsetting effect will be more pronounced in the cold season across all scenarios, and will be particularly strong under RCP 2.6. For the remainder of the year, the effects of increased precipitation and warming will have competing importance under the RCP 4.5 scenario, whereas temperature effects will generally dominate the climate impact under the RCP 8.5 scenario. Our results provide new references for future climate–energy studies in western China and can help inform policy-making on regional climate adaptation. We also propose an algorithm model that is readily compatible with potential early-warning projects to protect against extreme weather-induced power outages.

1 Introduction

Extreme weather events induced by climate change threaten energy supplies worldwide (Mideksa & Kallbekken, 2010). Understanding how residential electricity consumption (REC) responds to different climatic conditions as well as socioeconomic conditions is key to mitigating these threats, as it will assist in evaluating the social cost of carbon, i.e., a key indicator for policies focused on curbing climate change (Li, Pizer, & Wu, 2019; Nordhaus, 2017), and developing targeted adaptive measures (Auffhammer, Baylis, & Hausman, 2017; Auffhammer & Mansur, 2014; Davis & Gertler, 2015; Deschênes & Greenstone, 2011).

Previous studies suggest that the climate impact on REC is mostly due to heating and cooling demands (Auffhammer et al., 2017; Auffhammer & Mansur, 2014; Deschênes & Greenstone, 2011; Waite et al., 2017). Climate change not only increases the summer temperatures, it can also affect precipitation and induce more extreme weather throughout the year (Cohen et al., 2014). The 2021 Texas power crisis in the United States (Busby et al., 2021) and the 2020 Hunan power shortage in China (Zheng, 2020) both occurred in below-freezing winter temperatures, which justifies the study of the role of precipitation when estimating heating-related REC.

Compared to temperature, precipitation has been largely generalized in econometric models used in previous studies (Auffhammer et al., 2017; Davis & Gertler, 2015; Du, Yu, & Wei, 2020; Li et al., 2019; Zhang et al., 2020). Cooling degree days (CDD) and heating degree days (HDD) are used to group temperature data (Du et al., 2020; Zhang et al., 2020). In contrast, there is no equivalent for precipitation data. Another

method of sorting temperature data is to use distribution-based temperature bins (Davis & Gertler, 2015; Deschênes & Greenstone, 2011; Li et al., 2019). Distribution bins for precipitation have been used; however, there are fewer data points for precipitation than for temperature for a given study period (Davis & Gertler, 2015; Deschênes & Greenstone, 2011). An alternative is to treat precipitation as a covariate or control variable (Auffhammer et al., 2017; Zhang et al., 2020), which effectively means paying less attention to precipitation than to temperature. However, precipitation does affect REC (Auffhammer & Aroonruengsawat, 2011; Deschênes & Greenstone, 2011), and precipitation changes were shown to increase the predictive uncertainty of REC in an eastern province in China (Zhang et al., 2020).

Our study aimed to quantify the impact of climate on REC and project its future trends under climate change in two western cities in China, with a particular focus on the role of precipitation. We hypothesized that precipitation impacts heating-related REC in the cold season, likely in the form of snowfall. If our hypothesis is correct, climate impact on the REC in the cold season may not simply decline with warming; it may also increase if there is more precipitation in the cold season. To test this hypothesis, we built an algorithm model using seven machine learning algorithms and predicted future climate impact on REC under three climate futures projected with different greenhouse gas concentration pathways (RCP). They are RCP 2.6, i.e., whereby greenhouse gas emissions are expected to start to decrease by 2020 and reach zero by 2100; RCP 4.5, whereby the emission is assumed to peak around 2040 and then decline; and RCP 8.5, where the emissions continue to rise throughout the 21st century.

We selected two western cities in China, Lanzhou and Lhasa, as our research objects. The two cities were chosen to represent different climate pattern in western China. Lanzhou is located in the northwest and Lhasa is in the southwest. We also selected these two cities as they represent the different heating infrastructures in northern and southern China, as defined by the Qinling–Huaihe Line. Lanzhou is equipped with a coal-fueled central heating system as in other northern cities, while Lhasa relies mostly on household electric heating as in the rest of southern China.

By investigating climate-REC relationships in these two cities and project the impact of climate change, our results provided new references for future climate–energy studies in western China and can help inform policy-making on regional climate adaptation. In the process, we also developed an algorithm model that is readily compatible with potential early-warning projects to protect against extreme weather-induced power outages.

2 Materials And Methods

2.1 Research Design

Previous studies (Auffhammer & Aroonruengsawat, 2011; Auffhammer & Mansur, 2014; Davis & Gertler, 2015; Deschênes & Greenstone, 2011; Fan, Zhang, & Wang, 2017; Frederiks, Stenner, & Hobman, 2015; Li et al., 2019) suggest that factors impacting REC include temperature, precipitation, income, population, and urbanization. Since Lanzhou and Lhasa are both cities, we considered information on urbanization incorporated in the population data. Thus, we built our model using temperature, precipitation, income,

and population as predictor variables, and REC as the response variable. Unlike previous studies using data modeling, we adopted an algorithm modeling approach. With respect to the predictive accuracy as well as interpretability, we compared the results from our model with those from the previously used parameter-based models to verify and demonstrate the strength of our new modeling approach. We then tapped into the capacity of our model to holistically estimate climate impact by projecting the impact of climate change on REC under RCP2.6, RCP4.5, and RCP8.5 scenarios.

2.2 Data

We based our estimates on 2011–2019 Lanzhou and 2014–2019 Lhasa panel data. Each dataset included monthly REC, monthly mean temperature, monthly precipitation, annual salary, and annual population. The study period was mainly defined by the availability of monthly REC data for the two cities. We obtained Lanzhou REC data from the Gansu State Grid Company (GSGC) and Lhasa REC data from the Tibet State Grid Company (TSGC). Monthly mean temperature and monthly precipitation data were calculated using 2011–2019 Lanzhou and 2014–2019 Lhasa daily temperature and precipitation data from the China Meteorological Administration. Yearly salary and population data were provided by the National Bureau of Statistics of China.

For the prediction of future climate impact on REC in the two cities, we used future climate projection of China based on the regcm4.6 model under the RCP2.6, RCP4.5, and RCP8.5 scenarios (Lei & Xiaoduo, 2020). This regional climate model has yielded more accurate simulation results of present-day mean climatology over Northwest China than the Met Office Hadley Centre Earth System (HadGEM2-ES) global climate model (Pan, Zhang, & Huang, 2020). We extract climate projections for the two cities from the nearest available points in the data; i.e., at the coordinates of 36°N, 103.5° E for Lanzhou and 29.5°N, 91° E for Lhasa. This climate projection data is more reliable for Lanzhou, as it is located in Northwest China where the regcm4.6 model performs well. Lhasa, on the other hand, is located not only in Southwest China, but also in the Tibetan Plateau. While the Tibetan Plateau is found to be getting warmer and wetter (Yao et al., 2019), climate projections for the region are greatly complicated by the uplift of the Tibetan Plateau and there is not reliable near-term climate change projection yet (Hu & Zhou, 2021). Since Lanzhou is also projected to get warmer and have increased precipitation (Lei & Xiaoduo, 2020), we were able to obtain a qualitative reference for Lhasa from results in Lanzhou.

2.3 Methods

2.3.1 Algorithm modeling vs. data modeling

We adopted algorithm modeling rather than data modeling as used in previous studies. Modeling can be considered as a process associating predictor variables with the response variable; the main difference between data modeling and algorithm modeling is the way in which the association is built (Leo Breiman, 2001). Data modeling assumes that real-world data conform to a stochastic regression function that incorporates predictor variables, random noise, and parameters, and attempts to estimate the response variable using functions that are then validated using goodness-of-fit tests (Leo Breiman, 2001).

Algorithm modeling, on the other hand, assumes that the association is unknown and attempts to identify an algorithm that can best predict the response variables using predictor variables, and the results are then validated using predictive accuracy (Leo Breiman, 2001).

Data models attempt to fit data gathered from real-world observations of various systems with unknown mechanisms into a parametric model selected by a statistician. This inevitably compromises accuracy as well as insight (Leo Breiman, 2001). Despite efforts to minimize parametric assumptions (Deschênes & Greenstone, 2011), the econometric models used in previous studies remain as parameter-based data models with similar limitations. Conversely, algorithm modeling is completely data-driven since it makes only one assumption, i.e., that the used data are obtained from an unknown multivariate distribution (Leo Breiman, 2001). The more accurate a model, the more reliable is the information we can derive regarding the underlying data mechanism. Since algorithmic models have generally been proven to make better predictions (Leo Breiman, 2001), we adopted algorithm modeling in our study.

Since algorithm modeling treats the association between predictor and response variables as unknown, there was no means of ensuring that the algorithm worked well with our data prior to testing. We derived a pool of seven commonly used candidate algorithms (Brownlee, 2016).

Three are basic algorithms: k-nearest neighbors (KNN), support vector regression (SVR), and classification and regression tree (CART). KNN was designed with the assumption that similar things are in close proximity to each other. It makes estimations by measuring the distances between an instance and all other instances in the dataset. It selects the specified number of instances (K) closest to the one concerned, then votes for the most frequent label (in the case of classification) or averages the labels (in the case of regression) (Altman, 1992). SVR works by constructing a hyperplane or hyperplanes in a high-dimensional space for either classification or regression. The hyperplane with the longest distance from all the nearest training data points of any class (also known as the functional margin) is considered a good separation since the larger the margin, the lower the generalization error (Cortes & Vapnik, 1995). CART is also known as a decision tree, since it is a tree-like algorithm that is built through repeatedly splitting into two child nodes. It splits on a single input variable that improves the Gini index, a performance measure that calculates the probability of a specific input variable being classified incorrectly when selected randomly (L Breiman, Friedman, Olshen, & Stone, 2017). The structure of CART makes it (along with tree-based algorithms) innately immune to correlation among predictor variables (J. Friedman, Hastie, & Tibshirani, 2001).

The other four are ensemble algorithms. Ensemble is a machine learning method that combines several base algorithms into one in order to decrease variance and bias or to improve predictions. We used Adaptive Boosting (AB), Random Forest (RF), Extra Trees (ET), and Gradient Boosting Machine (GBM), i.e., all CART ensembles with different ensemble methods. RF and ET train their individual trees independently and average predictions across trees, an ensemble method known as bagging. RF and ET are different in how their individual trees split as well as in how they sample data. RF splits where the performance measure is best, whereas ET splits randomly. RF subsamples the data sample with

replacement, or bootstrapping, while ET uses the original data sample (Geurts, Ernst, & Wehenkel, 2006; Ho, 1995). AB and GBM build one tree at a time, where each new tree corrects errors made by the previous tree, an ensemble method known as boosting. AB and GBM differ in how they identify and correct errors of previous trees. AB identifies errors using high-weight data points as each new tree up-weights observations that were misclassified in the previous tree, whereas GBM identifies errors using gradient of loss function computed in the previous trees, which reflects how well the previous trees performed. AB corrects errors by assigning its trees different weights in the final prediction based on their performance. In contrast, GBM weights each tree equally, but restricts their predictive capacity through the learning rate, which represents how quickly an error is corrected from each tree to the next, for greater accuracy (J. H. Friedman, 2001; Kégl, 2013; Mason, Baxter, Bartlett, & Frean, 1999).

We determined the baseline performance of the seven algorithms by running our data with each algorithm using the default hyperparameters of the algorithms in a Python 3.7 environment employing scikit-learn implementation (Pedregosa et al., 2011). We used the scikit-learn utility of Pipeline (Pedregosa et al., 2011) to automate workflow and avoid data leakage. In the workflow, we standardized the data to test all candidate algorithms. We used 80% of the data for algorithm testing and modeling and reserved 20% for validation. The training dataset and the reserved validation dataset were split randomly. We used 10-fold cross validation to estimate algorithmic performance as measured by the mean squared error (MSE).

The stochastic nature of machine learning algorithms means that model results for each run can be different. To ensure we selected the best algorithm from the candidate pool, we ran the seven candidate algorithms 100 times each on our data and averaged their performance in terms of MSE and averaged ranking out of the 100 runs.

For reference, we also used a simple multivariate parameter-based ordinary least squares (OSL) model on the data in parallel with our algorithm model as follows:

$$"REC_i = \beta_0 Temp_i + \beta_1 Preci_i + \beta_2 Salary_i + \beta_3 Pop_i + \omega_i",$$

where i is the sample month, REC_i is residential electricity consumption for month i , $Temp_i$ is mean temperature of month i , $Preci_i$ and is the total precipitation of month i . $Salary_i$ and Pop_i are salary and population levels for month i , respectively, both averaged from their yearly total. The coefficients of interest are β_0 , β_1 , β_2 , and β_3 , representing the impact of temperature, precipitation, income and population, respectively. ω_i represents the bias. We did not regroup the temperature or the precipitation data in this parallel model in order to ensure that no more assumptions were made for this model than for the algorithmic model. We tested this model with the machine learning algorithms for baseline performance.

2.3.2 Comparing predictive accuracy

Figure 1 shows the baseline performance of the candidate algorithms using the Lanzhou and Lhasa data, respectively, with the blue bars indicating the mean MSE and the pink bars showing the averaged ranking of the algorithm out of the 100 runs. It was evident that the ensemble algorithm GBM was the best fit for studying 2011–2019 Lanzhou data and 2014–2019 Lhasa data, as it showed the lowest average MSEs across the board with high consistency.

To validate the model results in case of over fitting, we ran predictor variables from the validation datasets with the prepared models to make predictions. The validation results are shown in Figure 2. The good validation results suggested that the model was not overfitted.

2.3.3 Comparing interpretability

A robust interpretation of models can only be achieved with models with high predicative accuracy. Figures 1 and 2 showed that that GBM yielded highest predicative accuracy for Lanzhou 2011–2019 data and Lhasa 2014–2019 data. To compare its interpretation with the traditional parameter-based models, we presented the results of OSL linear models across different datasets in Table 1 for reference.

Table 1 Results of ordinary least squares (OSL) linear regression across Lanzhou and Lhasa year-round datasets. REC - residential electricity consumption.

Data	Response variable	R ²	Coeff.	Predictor variables	p
Lanzhou 2011–2019 (year-round)	REC	0.901	-1459.1703	Temperature	0.000
			-136.8127	Precipitation	0.441
			3154.4050	Income	0.000
			-277.0476	Population	0.023
Lhasa 2014–2019 (year-round)	REC	0.851	-438.3596	Temperature	0.000
			48.2718	Precipitation	0.374
			62.8833	Income	0.863
			1915.3652	Population	0.012

Table 1 shows that the OSL models can explain 90.1% of the variation in the Lanzhou 2011–2019 year-round data and 85.1% in the Lhasa 2014–2019 data. Hence, it was meaningful to compare the OSL interpretation with the GBM model interpretations for Lanzhou since both were good models for the Lanzhou datasets, with GBM yielding better predictions (Fig. 1).

Despite their reputation for being “black boxes” that are difficult to interpret (Azodi, Tang, & Shiu, 2020), algorithmic models have been proven to be interpretable in ways that not only show the individual impact

of predictor variables, but also the joint impact of multiple predictor variables (Azodi et al., 2020; J. H. Friedman, 2001; J. H. Friedman & Meulman, 2003; Goldstein, Kapelner, Bleich, & Pitkin, 2015; Mehdiyev & Fettke, 2020; Molnar, 2020; Strobl, Boulesteix, Zeileis, & Hothorn, 2007; Zhao & Hastie, 2019). We used partial dependence plots (PDPs) and individual conditional expectation (ICE) plots to interpret the data mechanisms underlying the algorithmic models (Mehdiyev & Fettke, 2020; Molnar, 2020; Zhao & Hastie, 2019).

We used the concept of feature for discussions on PDPs and ICE plots. The concepts of “feature” and “predictor variable” are often used interchangeably, especially in machine learning models. However, we made a distinction between the two in our study for clarity. We use the term “predictor variables” to refer to the individual variables we use for modeling (i.e., temperature, precipitation, income, and population). A “feature” can either be an individual predictor variable such as temperature or two predictor variables combined into one, e.g., climate, which comprises both temperature and precipitation.

The PDPs depict the overall dependence of model prediction on a given input feature by marginalizing over the values of all other input features; ICE plots disaggregate this average by showing the estimated functional relationship of each instance (Goldstein et al., 2015; Mehdiyev & Fettke, 2020; Molnar, 2020; Zhao & Hastie, 2019). While PDPs are useful for depicting the overall marginal effect of a given feature, they can also obscure heterogeneous relationships caused by interactions (J. H. Friedman & Meulman, 2003; Goldstein et al., 2015; Mehdiyev & Fettke, 2020; Molnar, 2020; Zhao & Hastie, 2019). For this reason, we compared ICE plots and PDPs to determine if heterogeneous relationships existed in the estimation.

We used PDP utility offered by scikit-learn (Pedregosa et al., 2011) for ICE and PDP visualization. To compare with the OSL interpretation based on coefficients for predictor variables, we first interpreted the role of predictor-variables for the GBM models as shown in Figure 3. Each individual plot shows both the ICE plots and PDPs of a given feature; in this case, a predictor variable for the corresponding model, with y-axes showing partial dependence, or the marginal effect of a given feature on the REC estimation, and x-axes showing the value of the corresponding feature. Note that marginal effect is not the predicted REC value; rather, it is the way in which REC changes with the feature. It is evident that all ICE plot curves generally follow the same pattern as that shown by the PDP line. This means the PDPs of the models provide a reliable interpretation of the relationships between the features and REC.

Interpretations by GBM and OSL on individual predictor variables are generally consistent in temperature–REC relationships, with the innately non-linear GBM being more informative than OSL models. GBM suggested that precipitation plays a role in influencing REC, whereas OSL appeared to differ since it failed to come to a conclusion on the role of precipitation, as shown by the high p-value and opposite signs of precipitation coefficients in the two cities in Table 1. At this stage, we cannot be sure which interpretation is correct due to the multicollinearity between year-round temperature and precipitation data, as shown in Figures 4a and 4c.

Multicollinearity may be one of the reasons why precipitation has not been fully considered in previous studies. Multicollinearity can reduce the precision of estimated coefficients in linear regressions and it is

common practice to remove or lessen the weight of one predictor (in this case, precipitation). Correlations between temperature and precipitation are also a challenge for GBM interpretation, since the use of PDPs and ICE plots is based on the assumption that features are not correlated (Mehdiyev & Fettke, 2020; Molnar, 2020; Zhao & Hastie, 2019).

To test our hypothesis regarding the role of precipitation during the cold season, we created a data subset consisting of data points from November, December, January, and February, i.e., the cold-season only. Multicollinearity between temperature and precipitation in this subset was no longer a problem for either city as shown in Figures 4b and 4c. For this reason, we used this subset (rather than the year-round dataset) to analyze the individual roles of temperature and precipitation.

Quantification of year-round climate impact, however, needs to be conducted without interference from multicollinearity. Interpretations of an algorithm model can manage the multicollinearity between temperature and precipitation by combining them into one feature (Molnar, 2020). Unlike OSL models that can only rely on coefficients of individual predictor variables for information, PDPs of algorithmic models can be applied to a predictor variable as well as to a feature comprising two combined predictor variables (J. H. Friedman & Meulman, 2003; Goldstein et al., 2015; Molnar, 2020). Since the climate feature was not correlated with the other features of the model in this study, our interpretation of climate impact was not affected by the influence of multicollinearity.

It is noteworthy that the PDP for the climate feature was not a simple combination of the individual PDPs for temperature and precipitation, as shown in Figure 3. Rather, the PDP for this newly created climate feature captured changes in temperature and precipitation as well as their interactions, and showed the marginal effect of all of these as a whole since partial dependence was calculated by marginalizing over all the other inputs of the estimation, i.e., income and population (Friedman and Meulman 2003; Molnar 2020).

When quantifying the marginal effect of the climate feature, we first averaged the monthly mean temperature and monthly precipitation by calendar month to represent the corresponding mean climate patterns for the cities during their respective study periods. We then obtained possible variation ranges of the climate patterns using standard deviations of monthly mean temperature and monthly precipitation for each calendar month in the two cities. Finally, by averaging the marginal effects of all climate conditions within the variation range for each month for the two cities, we obtained the average marginal effects of climate for each calendar month in the two cities.

We tapped further into this potential of the algorithm model for holistically quantifying climate impact on REC to predict the impact of climate change on REC of the two cities. This was achieved by replacing climate observations with climate projections under RCP2.6, RCP4.5, and RCP8.5 scenarios in the PDP-based quantification. As our study focused on the climate impact, we effectively controlled the socioeconomic conditions at the status-quo level of the study periods. We made projections for the mid-century decade from 2036–2045, since this period is more likely to accommodate the socioeconomic presumption than more distant periods, but is sufficiently distant for meaningful climate projections.

3 Results

3.1 Individual roles of temperature and precipitation

The GBM and OSL interpretations described in Section 2.3.3 revealed a negative effect of temperature on REC in both cities, which is consistent with the heating “half” of the U-shaped temperature–REC function in previous studies (Auffhammer & Mansur, 2014; Li et al., 2019). Combined with local knowledge, we are confident that the climate impact on Lanzhou and Lhasa REC is mainly driven by consumption for heating.

However, fewer references were available from previous studies on the year-round role of precipitation, and neither the OSL coefficient nor GBM PDPs were conclusive in this respect. To test our hypothesis regarding the role of precipitation in the cold season, we ran the OSL model using the cold-season data subset and obtained the results shown in Table 2.

Table 2 Results of ordinary least squares (OSL) linear regression across Lanzhou and Lhasa cold-season subsets. REC - residential electricity consumption.

Data	Response variable	R ²	Coeff	Predictor variables	<i>p</i>
Lanzhou 2011–2019 (cold season)	REC	0.959	-1307.5208	Temperature	0.009
			4985.5676	Precipitation	0.001
			3911.2492	Income	0.000
			-299.2837	Population	0.062
Lhasa 2014–2019 (cold season)	REC	0.884	-941.8544	Temperature	0.001
			-1628.8780	Precipitation	0.198
			-355.0618	Income	0.632
			3219.3960	Population	0.042

The OSL model based on the Lanzhou cold-season data subset yielded an R² of 95.9%, and considered precipitation to be a statistically significant predictor variable with a p-value lower than that of temperature. The positive coefficient of precipitation suggested that more precipitation in the cold season will drive up REC. The cold-season subset was too small to model using GBM; however, we were able to obtain confirmation based on its year-round interpretation. In Figure 4a, GBM shows that the marginal effects of precipitation on REC are higher when monthly precipitation falls below 40 mm in Lanzhou and below 20 mm in Lhasa. Climate patterns of the two cities shown in Figures 5c and 5d indicate that such low monthly precipitation only occurs in the cold season, suggesting that the impact of precipitation is stronger during the cold season.

The OSL model for the Lhasa cold season yielded an R^2 of 88.4%, and suggested a negative precipitation–REC relationship. There was no statistically significant precipitation–REC relationship in this case, though the p-value for precipitation was significantly lower for the Lhasa cold-season subset than for the Lhasa year-round data. Despite this, we consider that the result can be used as a reference, since it would be inappropriate to conclude that there is no association simply because the p-value is > 0.05 (Amrhein, Greenland, & McShane, 2019). The GBM interpretation of the precipitation impact in Lhasa shown in Figure 4b also confirmed this theory, as it showed a small peak in a low precipitation range in cold season, suggesting the existence of a precipitation impact during the cold season.

We identified an explanation of the absence of a positive OSL coefficient for precipitation in Lhasa cold season by comparing the Lhasa climate pattern with that of Lanzhou, as shown in Figure 5b. The monthly mean temperature rarely drops far below 0 °C in Lhasa, even in the cold season. Additionally, monthly precipitation during the cold season in Lhasa is considerably lower than that during the cold season in Lanzhou. In other words, compared with Lanzhou, snowfall is not as frequent or as heavy in Lhasa; hence, the less observable impact of precipitation during the cold season in Lhasa.

We conclude that the impact of precipitation is stronger in both cities during the cold season, as a result of snowfall boosting electricity consumption used for heating.

3.2 Impact of climate

As discussed in section 2.3.3, quantifying marginal effects of year-round temperature and precipitation separately would not be accurate due to the presence of multicollinearity. It would be unreasonable to omit either temperature or precipitation from our estimation just because of their numerical correlation, since our study already proved that both play a role as independent climate variables in influencing REC. For this reason, we quantified their joint impact (i.e., the climate impact) by interpreting the climate feature enabled by PDP.

Figures 5a and 5c show the marginal effects of climate on REC, and the climate impact on REC in Lanzhou and Lhasa, respectively. We used real-world climate patterns in Figures 5b and 5d to assist in constraining the way in which we used PDPs so that conclusions were not made based on extremely unlikely combinations of temperature and precipitation. The calculated average marginal effect showing how REC changes with climate conditions is marked in purple in Figures 5b and 5d.

Note that the z-axes of the PDPs in Figures 5a and 5c show the marginal effect of climate on REC rather than the actual predicted REC values; this describes the means by and degree to which REC changes with climate, i.e., the climate impact. The same is true for the purple y-axes in Figures 5b and 5d. A climate impact of zero showed that the climate feature did not have any effect on the estimation. Positive climate impact values show that the corresponding climate conditions will drive up the estimated REC and that the negative values will reduce it. For ease of reference, we describe changes in the climate impact by referring to them as positive and negative climate impacts; i.e., when we state that the climate impact is enhanced, it means it has increased in absolute value.

Interpretation of the GBM model using a holistic climate feature in Figure 5b shows a V-shaped relationship between climate impact on REC and calendar months for Lanzhou, with July showing the strongest negative impact from climate and November, December, January, and February showing the greatest positive climate impact. The climatic impact change during cold-season is more consistent with the change in precipitation, which consolidated our previous finding that precipitation has stronger impact on Lanzhou REC during cold season, likely in the form of snowfall given the mean temperature of the cold season months are slightly above or below 0°C.

For Lhasa, the relationship is U-shaped (Fig. 5d), with June, July, August, and September showing similarly strong negative climate impacts and December and January showing the greatest positive climate impacts. The climatic impact change during cold-season is more consistent with the change in temperature than in precipitation. This can be explained by less snowfall in Lhasa. Precipitation is low in Lhasa during cold season where there is only one month with mean temperature below 0°C.

Given the climate patterns, this indicates that climatic impact on REC is mainly driven by heating demand in the two cities, and that Lanzhou REC is more sensitive to climate variations in warm months than Lhasa.

We also noticed that there is not a consistent temperature threshold for climatic impact on REC in the two cities, as opposed to the assumption of HDD and CDD that there is a generally consistent base temperature at which there is no energy demand for cooling or heating. Climatic impact on REC is around zero when monthly mean temperature is around 12°C for Lanzhou and around 8°C for Lhasa. This further proves our point that temperature alone does not determine the climatic impact on REC. We believe precipitation as well as socioeconomic factors may also play a role in determining the climate threshold where there is no significant energy demand for cooling or heating.

3.3 Projected change in climate impact on REC

For the meaningful comparison of climate impact on REC under status-quo and projected RCP2.6, RCP4.5, and RCP8.5 scenarios, insights are needed regarding bias of the simulated climate patterns; the smaller the simulated bias, the more reliable projected climate impact.

We compared climate patterns simulated by the Regcm4.6 model under different climate scenarios for Lanzhou and Lhasa with the observed climate patterns during 2011–2019 and 2014–2019, respectively. We considered the data that was used to train our model as status-quo. Figures 6a and 6b show the comparison of status-quo and simulated temperature patterns, and Figures 6c and 6d show the precipitation patterns.

With respect to temperature, for Lanzhou, the simulated temperatures were lower than observed temperatures under all three climate scenarios, although they were in an accepted range with the exception of February and March. The simulated temperature in February is < 0 across all scenarios while the actual observation is > 0. This may create confusion given our hypothesis regarding precipitation

taking the form of snowfall. The underestimation of the simulated temperature in March was too large to be taken as reference. For Lhasa, simulated temperatures were significantly lower than the observed temperatures under all three climate scenarios throughout the year, with cold-season months all showing negative simulated values in contrast to positive readings in observations.

For precipitation, the Regcm4.6 model under all scenarios overestimated precipitation for the cold-season months in both Lanzhou and Lhasa, with Lanzhou showing much smaller (and thus more acceptable) bias across the year than Lhasa. Thus, we consider the climate impact projection to be more reliable for Lanzhou than for Lhasa.

With this consideration, we compared the climate impact on REC under the status-quo climate pattern with that under the RCP2.6, RCP4.5, and RCP8.5 climate scenarios for Lanzhou and Lhasa as shown in Figures 6e and 6f, respectively. The absence of projections under some scenarios was the result of the projected climate conditions exceeding the boundary of the PDPs defined by the status-quo observation data.

For Lanzhou, we excluded February and March from our discussion due to unacceptable bias. In terms of the coldseason, climate impact on REC is projected to increase in November, and decrease only slightly in December and January for 2036–2045. The climate impact in November will increase more under RCP 2.6 and RCP 4.5 than under RCP 8.5, and decrease less in December under RCP 2.6 than under RCP 4.5 or RCP 8.5; this suggests the involvement of precipitation. If temperature is the only climatic impact factor, then warming would result in a smaller positive climate impact across all scenarios. The fact that the positive climate impact on REC is projected to increase or drop less in cold season for scenarios exhibiting slower warming (i.e., RCP 2.6), suggests an offsetting role of precipitation. In other words, although warming reduces REC in cold-season, increased precipitation increases REC during the cold season in Lanzhou. This means precipitation has an offsetting effect to temperature on cold-season REC in Lanzhou.

For the remainder of the year, only July, August, and September showed enhanced negative climate impact. This also suggested an offsetting role of precipitation since warming alone would only enhance the negative climate impact. Since July and August (i.e., the two warmest months in Lanzhou) will still experience a negative climate impact, we do not believe that the drop in negative impact in other months will be driven by cooling-related demands. Rather, a more logical explanation would be that more precipitation may have a cooling effect which will offset the warming impact on REC. This offsetting impact of precipitation may be adequately strong to change the climate impact. For example, in October, under RCP 2.6, the climate impact changed from negative to positive, suggesting that the climate impact changed from reducing the REC estimation to increasing it. This can be confirmed by the fact that the negative climate impact was generally more enhanced under RCP 8.5 than under RCP 2.6. The varied ranking of climate impact under RCP4.5 among the three scenarios suggests that under this scenario, the effects of precipitation and temperature have competing importance in shaping the overall climate impact.

We did not thoroughly interpret the quantification of the climate impact on REC in Lhasa as the simulated climate was considered to be excessively biased to elucidate the projected climate impact on REC. However, as Lhasa, like Lanzhou, is becoming warmer and wetter, we also inferred precipitation to offset the warming impact on REC. However, as Lanzhou shows stronger precipitation impact in the cold season and is more sensitive to climate in warm months, we expect Lanzhou to be affected more than Lhasa by this offsetting effect.

4 Discussion And Conclusions

Our study proves the strength of algorithm modeling since it shows improved predictive accuracy and interpretability for studying climate impact on REC than linear models. Compared with previous econometric models used for climate-REC studies, our method allows highly correlated temperature and precipitation to be interpreted as one climate feature to quantify the climate impact. This means that our method has overcome the challenges of multicollinearity, which enables our models to treat temperature and precipitation equally, and has the potential to be used in all climate impact studies.

A data-driven model can only be as good as the data. Our model failed to make predictions for several scenarios that exceeded the bounds of the PDPs of climate feature. Our study was limited as we controlled the non-climate inputs at the status-quo level to predict climatic impact on REC; however, socio-economic conditions will definitely alter REC and, thus, the marginal effects of climate impact. Future studies should have a socio-economic focus to enable more accurate predictions. Our projection for Lhasa is also limited by insufficient climate projection studies in the region. More accurate climate projections are needed to improve the understanding of potential future climate impact on REC.

Our results show that the climate impact on REC in both Lanzhou and Lhasa is driven by heating-related demand, with Lanzhou being more sensitive to climate than Lhasa during the warm season. The model results support our hypothesis. Both Lanzhou and Lhasa show a stronger precipitation impact during the cold season, since precipitation—especially in the form of snowfall—boosts electricity consumption during the cold season. The impact is stronger in Lanzhou, which experiences greater precipitation and more months when temperatures below 0 °C occur during the cold season. As the two cities become warmer and wetter, precipitation will offset the impact of warming on REC across the year, with Lanzhou being more affected. Based on Lanzhou projections, the offsetting effect is more pronounced in cold season across all scenarios, and is particularly strong under RCP 2.6. For the remainder of the year, the effects of increased precipitation and warming will have competing importance under the RCP 4.5 scenario, while the effects of temperature will generally dominate the climate impact under the RCP 8.5 scenario.

Our study highlights the need for climate-related energy studies to address climate holistically rather than to focus solely on the temperature, especially in regions with cold winters as well as in areas vulnerable to extreme weather such as blizzards. It provides previously unavailable references for future climate–energy studies in western China and broader alpine regions where temperature is rapidly increasing just

as the water cycle is becoming ever more unstable. Our study also assists in aiding regional policy-making on climate change adaptation. The developed model is also readily compatible with potential early-warning projects to protect against power outages induced by extreme weather.

Declarations

Data availability

The data that support the findings of this study are available from data sources as stated in section 2.2. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the authors with the permission of data sources.

References

1. Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3), 175-185. doi:<https://doi.org/10.2307/2685209>
2. Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*. doi:<https://doi.org/10.2307/2685209>
3. Auffhammer, M., & Aroonruengsawat, A. (2011). Simulating the impacts of climate change, prices and population on California's residential electricity consumption. *Climatic Change*, 109(1), 191-210. doi:10.1007/s10584-011-0299-y
4. Auffhammer, M., Baylis, P., & Hausman, C. H. (2017). Climate change is projected to have severe impacts on the frequency and intensity of peak electricity demand across the United States. *Proceedings of the National Academy of Sciences*, 114(8), 1886-1891. doi: <https://doi.org/10.1073/pnas.1613193114>
5. Auffhammer, M., & Mansur, E. T. (2014). Measuring climatic impacts on energy consumption: A review of the empirical literature. *Energy Economics*, 46, 522-530. doi:<https://doi.org/10.1016/j.eneco.2014.04.017>
6. Azodi, C. B., Tang, J., & Shiu, S.-H. (2020). Opening the Black Box: Interpretable machine learning for geneticists. *Trends in genetics*, 36(6), 442-455. doi:<https://doi.org/10.1016/j.tig.2020.03.005>
7. Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3), 199-231. doi:<https://doi.org/10.1214/ss/1009213726>
8. Breiman, L., Friedman, J., Olshen, R., & Stone, C. (2017). *Classification Regression Trees*. Routledge.
9. Brownlee, J. (2016). *Machine learning mastery with python (Vol. 527): Machine Learning Mastery Pty Ltd*.
10. Busby, J. W., Baker, K., Bazilian, M. D., Gilbert, A. Q., Grubert, E., Rai, V., . . . Webber, M. E. (2021). Cascading risks: Understanding the 2021 winter blackout in Texas. *Energy Research & Social Science*, 77, 102106. doi:<https://doi.org/10.1016/j.erss.2021.102106>

11. Cohen, J., Screen, J. A., Furtado, J. C., Barlow, M., Whittleston, D., Coumou, D., . . . Overland, J. (2014). Recent Arctic amplification and extreme mid-latitude weather. *Nature geoscience*, 7(9), 627-637. doi:<https://doi.org/10.1038/ngeo2234>
12. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297. doi:<https://doi.org/10.1007/BF00994018>
13. Davis, L. W., & Gertler, P. J. (2015). Contribution of air conditioning adoption to future energy use under global warming. *Proceedings of the National Academy of Sciences*, 112(19), 5962-5967. doi:<https://doi.org/10.1073/pnas.1423558112>
14. Deschênes, O., & Greenstone, M. (2011). Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US. *American Economic Journal: Applied Economics*, 3(4), 152-185. doi:<https://doi.org/10.1257/app.3.4.152>
15. Du, K., Yu, Y., & Wei, C. (2020). Climatic impact on China's residential electricity consumption: Does the income level matter? *China Economic Review*, 63, 101520. doi:<https://doi.org/10.1016/j.chieco.2020.101520>
16. Fan, J.-L., Zhang, Y.-J., & Wang, B. (2017). The impact of urbanization on residential energy consumption in China: An aggregated and disaggregated analysis. *Renewable Sustainable Energy Reviews*, 75, 220-233. doi:<https://doi.org/10.1016/j.rser.2016.10.066>
17. Frederiks, E. R., Stenner, K., & Hobman, E. V. (2015). The socio-demographic and psychological predictors of residential energy consumption: A comprehensive review. *Energies*, 8(1), 573-609. doi:<https://doi.org/10.3390/en8010573>
18. Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1): Springer Series in Statistics, New York.
19. Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
20. Friedman, J. H., & Meulman, J. J. (2003). Multiple additive regression trees with application in epidemiology. *Statistics in medicine*, 22(9), 1365-1381. doi:<https://doi.org/10.1002/sim.1501>
21. Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, 63(1), 3-42. doi:<https://doi.org/10.1007/s10994-006-6226-1>
22. Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *journal of Computational Graphical Statistics*, 24(1), 44-65. doi:<https://doi.org/10.1080/10618600.2014.907095>
23. Ho, T. K. (1995). Random decision forests. Paper presented at the Proceedings of 3rd international conference on document analysis and recognition.
24. Hu, S., & Zhou, T. (2021). Skillful prediction of summer rainfall in the Tibetan Plateau on multiyear time scales. *Science Advances*, 7(24), eabf9395.
25. Kégl, B. (2013). The return of AdaBoost. MH: Multi-class Hamming trees. arXiv preprint arXiv.
26. Lei, Z., & Xiaoduo, P. (2020). Future climate projection of China based on regcm4.6 (2007-2099).

27. Li, Y., Pizer, W. A., & Wu, L. (2019). Climate change and residential electricity consumption in the Yangtze River Delta, China. *Proceedings of the National Academy of Sciences*, 116(2), 472-477. doi: <https://doi.org/10.1073/pnas.1804667115>
28. Mason, L., Baxter, J., Bartlett, P., & Frean, M. (1999). Boosting algorithms as gradient descent in function space. Paper presented at the Proc. NIPS.
29. Mehdiyev, N., & Fettke, P. (2020). Prescriptive process analytics with deep learning and explainable artificial intelligence.
30. Mideksa, T. K., & Kallbekken, S. (2010). The impact of climate change on the electricity market: A review. *Energy Policy*, 38(7), 3579-3585.
31. Molnar, C. (2020). Interpretable machine learning: Lulu. com.
32. Nordhaus, W. D. (2017). Revisiting the social cost of carbon. *Proceedings of the National Academy of Sciences*, 114(7), 1518-1523. doi:<https://doi.org/10.1073/pnas.1609244114>
33. Pan, X., Zhang, L., & Huang, C. (2020). Future climate projection in Northwest China with RegCM4. 6. *Earth Space Science Reviews*, 7(2), e2019EA000819.
34. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
35. Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics*, 8(1), 1-21. doi:<https://doi.org/10.1186/1471-2105-8-25>
36. Waite, M., Cohen, E., Torbey, H., Piccirilli, M., Tian, Y., & Modi, V. (2017). Global trends in urban electricity demands for cooling and heating. *Energy*, 127, 786-802. doi:<https://doi.org/10.1016/j.energy.2017.03.095>
37. Yao, T., Xue, Y., Chen, D., Chen, F., Thompson, L., Cui, P., . . . Mosbrugger, V. (2019). Recent Third Pole's rapid warming accompanies cryospheric melt and water cycle intensification and interactions between monsoon and environment: Multidisciplinary approach with observations, modeling, and analysis. *Bulletin of the American Meteorological society*, 100(3), 423-444. doi:<https://doi.org/10.1175/BAMS-D-17-0057.1>
38. Zhang, M., Zhang, K., Hu, W., Zhu, B., Wang, P., & Wei, Y.-M. (2020). Exploring the climatic impacts on residential electricity consumption in Jiangsu, China. *Energy Policy*, 140, 111398. doi:<https://doi.org/10.1016/j.enpol.2020.111398>
39. Zhao, Q., & Hastie, T. (2019). Causal interpretations of black-box models. *Journal of Business & Economic Statistics*, 39(1), 272-281. doi:<https://doi.org/10.1080/07350015.2019.1624293>
40. Zheng, X. (2020). Power utilities asked to ensure proper supply. *China Daily*.

Figures

Figure 1

Results of algorithm selection for estimating Lanzhou and Lhasa residential electricity consumption (REC). The blue bars indicate negative mean mean square error (MSE) and the pink bars show the average ranking for each algorithm based on 100 runs; the shorter the bar, the better the baseline performance of the corresponding algorithm. GBM - gradient boosting machine, ET - extra trees, KNN - k-nearest neighbors, AB - adaptive boosting, RF - random forest, CART - classification and regression tree, SVR - support vector regression, LR-simple linear regression model, i.e.,ordinary least squares (OSL) model

Figure 2

Validation of the Lanzhou and Lhasa models. The blue line marks the reserved residential electricity consumption (REC) validation data. The red line shows REC estimated by the corresponding model with inputs of temperature, precipitation, income and population. (a) and (b) show validation results for models trained with datasets for Lanzhou 2011–2019 and Lhasa 2014–2019, respectively

Figure 3

Individual conditional expectation (ICE) plots and partial dependence plots (PDPs) for the models. (a) and (b) Results of models estimating Lanzhou 2011–2019, and Lhasa 2014–2019 REC, respectively. Subplots 1, 2, 3, and 4 show the partial dependence of temperature, precipitation, income, and population, respectively

Figure 4

Correlation heatmaps of the Lanzhou 2011–2019 datasets and Lhasa 2014–2019 datasets. (a) and (b) Correlations among data variables in the Lanzhou year-round and cold season datasets, respectively; (c) and (d) correlations for the Lhasa year-round and cold season datasets, respectively. REC - residential electricity consumption

Figure 5

Partial dependence plot (PDP)-based interpretation of gradient boosting machine (GBM) models for Lanzhou 2011–2019 year-round dataset and Lhasa 2014–2019 year-round dataset. (a) and (c) Marginal effects of all climate conditions considered by the GBM model on residential electricity consumption (REC) in Lanzhou and Lhasa, respectively, with x-axes showing temperature, y-axes showing precipitation, and z-axes the the marginal effect of the climate feature on the model outcome. (b) and (d) Average climate impact based on climate patterns in the two cities, with x-axes showing month, purple y-axes showing the marginal effect of the climate feature, or climate impact, blue y-axes showing temperature and green y-axes showing precipitation

Figure 6

Projected climate change and change in climate impact on residential electricity consumption (REC) in Lanzhou and Lhasa. (a) and (b) Comparisons of observed and simulated monthly mean temperature for status-quo periods of Lanzhou and Lhasa respectively. (c) and (d) Comparison of observed and simulated monthly mean precipitation for status-quo periods of Lanzhou and Lhasa respectively. (e) and (f) Comparison between the status-quo climate impact on REC with that under projected climate

scenarios during the period from 2036–2045 in Lanzhou and Lhasa, respectively. Data of status-quo is marked with green, RCP 2.6 in blue, RCP 4.5 in orange, and RCP 8.5 in red