



Article Effect of Climate on Residential Electricity Consumption: A Data-Driven Approach

Cuihui Xia ^{1,2,*}, Tandong Yao ^{1,3}, Weicai Wang ¹ and Wentao Hu ²

- ¹ Big Science Program Center, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China; tdyao@itpcas.ac.cn (T.Y.); weicaiwang@itpcas.ac.cn (W.W.)
- ² University of Chinese Academy of Sciences, Beijing 100101, China; hwt11@itpcas.ac.cn
- ³ Key Laboratory of Tibetan Environment Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China
- * Correspondence: xiacuihui@itpcas.ac.cn; Tel.: +86-137-1862-2997

Abstract: Quantifying the climatic effect on residential electricity consumption (REC) can provide valuable insights for improving climate–energy damage functions. Our study quantifies the effect of climate on the REC in Tibet using machine learning algorithm models and model-agnostic interpretation tools of feature importance scores and partial dependence plots. Results show that the climate contributes about 16.46% to total Tibet REC while socioeconomic factors contribute about 83.55%. Precipitation (particularly snowfall) boosts electricity consumption during the cold season. The effect of the climate is stronger in urban Tibet (~25.06%) than rural Tibet (~14.79%), particularly in September when electricity-aided heating is considered optional, as higher incomes amplified the REC response to the climate. With urbanization and income growth, the climate is expected to contribute more to Tibet REC. Hence, precipitation should be incorporated in climate–REC functions for the social cost of carbon (SCC) estimation, particularly for regions vulnerable to snowfall and blizzards. Herein, we developed a model-agnostic method that can quantify the total effect of the climate while differentiating between contributions from temperature and precipitation, which can be used to facilitate interdisciplinary and cross-section analysis in earth system science. Moreover, this data-driven model can be adapted to warn against extreme weather induced power outages.

Keywords: climate; residential electricity consumption; rural and urban difference; machine learning; heating; multicollinearity; Tibet

1. Introduction

The recent Intergovernmental Panel on Climate Change (IPCC) report [1] highlights the need for stringent and consistent policies on the reduction of greenhouse gas emissions. The social cost of carbon (SCC), that is, the economic cost of an additional ton of carbon dioxide emissions or its equivalent, is used to assess climate change-related policies and their implementation [2]. The current SCC estimates are greatly challenged owing to the limited knowledge regarding damage functions, which define how changes in climate variables impact human life [3,4]. A key knowledge gap in building damage functions lies in the intrinsic economic perspective of SCC estimation, which largely discounts the physical mechanism of the earth system [4]. An earth-system perspective that integrates the physical sciences and social sciences would facilitate the understanding of the damage functions [5].

The climate–electricity damage function is of particular importance for both curtailing, and adapting to, climate change. In fact, the power sector accounts for 41% of carbon emissions globally; thus, replacing fossil fuels with green electricity, and generating electricity using renewable energy sources, can effectively reduce carbon emissions [6]. The climate impacts electricity markets on both the demand and supply ends [7]. On the supply end, efforts have been made to inform green electricity generation [8] with measures boosting



Citation: Xia, C.; Yao, T.; Wang, W.; Hu, W. Effect of Climate on Residential Electricity Consumption: A Data-Driven Approach. *Energies* 2022, *15*, 3355. https://doi.org/ 10.3390/en15093355

Academic Editor: Piotr Kosowski

Received: 20 April 2022 Accepted: 2 May 2022 Published: 5 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). incentives [9]. On the demand side, the end-use electricity analysis is of particular importance to climate and energy management [10]. Among the types of energy consumption, residential electricity is particularly significant as it affects, and is affected by, the climate. Residential energy consumption (REC) is a major contributor to carbon emissions globally [6], with space heating and air conditioning representing the largest contributors [11]. Given that heating and cooling demand is related to climate variables [12–15], quantifying the effect of climate on residential electricity consumption (REC) can help inform climate change-related policymaking as well as climate–energy research methodologies.

Previous literature estimates climate–REC functions primarily with econometric models designed for economic analysis [13,14,16,17], which risk dwarfing the complexity of physical processes in the earth system with model presumptions. "Neutral" models that treat the physical and social aspects equally are needed to reach common understanding across different disciplines under the earth-system perspective. More specifically, datadriven models that make few subjective presumptions would be ideal. Since data-driven techniques such as clustering and neural networks have already been adopted by recent energy [18] and physical science studies [19,20], respectively, data-driven methods can also be used to examine climate–energy interactions for better interdisciplinary insights.

Previous literature on the climate-REC function has largely overlooked precipitation [21], meanwhile, the effect of temperature on the REC has been extensively analyzed [11,13,15,22,23]. Most of these studies have focused on the cooling demand in temperate, subtropical, or tropical regions in which the potential effect of the climate is likely related to higher summer temperatures and higher air conditioning adoption rates [12,14,16,24–26]. The general conclusion of studies that have focused on the heating demand, was that global warming reduces the heating-related energy demand [27,28]. Moreover, precipitation, in the few cases where it is been considered, has consistently been considered to a lesser extent than temperature [12,29]. However, climate is the abstraction of temperature, precipitation, and their seasonal variations [21]. In fact, climate change reflects changes in the earth system, and has a more significant impact than direct warming [30]. For example, global warming intensifies the water cycle and increases the amount of precipitation [31]. Furthermore, precipitation was implicated in multiple recent power crises, including the 2021 power crisis in Texas, United States [32], the 2020 power shortage in Hunan province, China [33], and the most recent power outage in Northeastern US in 2022 [34], all of which occurred under heavy snow conditions. Thus, heating-oriented studies are needed to examine the effect of precipitation on the REC.

In addition to overlooking precipitation as an essential climate variable, previous climate–REC damage functions have not sufficiently focused on rural–urban disparities, which is highly relevant for policy implementation. Nevertheless, the REC in urban areas has been reported as higher than in rural areas [35]. Common drivers of both rural and urban REC have been identified as climate, income, population, and urbanization, with income being the most influential [36]. However, the influence of rural–urban disparities on the effect elicited by climate on the REC has not yet been clarified.

To address this gap in knowledge, we selected Tibet as a subject area as its alpine climate is ideal for heating-oriented analysis, and its demographic structure featuring more rural than urban population provides a distinctive socioeconomic context compared to other studies that have been conducted in more developed and urbanized areas. As a natural laboratory for research on multi-sphere interactions and the human–nature relationship [37], the Tibetan Plateau is experiencing fast climate and environmental changes, and is expected to become warmer and wetter with an increasingly unsteady water cycle [38], making it an ideal region for characterizing the role of precipitation on REC.

The aim of our study was to quantify the effect of climate on the REC in rural and urban Tibet, focusing on the effect of precipitation on the heating-related REC and the effect of rural–urban disparity during the cold season, or months, in which the temperature is below 0 °C. We hypothesized that precipitation in the form of snowfall boosts the heating-related REC in Tibet and that the effect of the climate on the REC is more conspicuous in the

more affluent urban regions of Tibet. To test this hypothesis, we constructed data-driven models using eight algorithms.

We argue that precipitation, as a key indicator of climate in earth system science, should be incorporated in climate–REC functions for SCC estimation and treated equally as temperature. Our results, based on investigating and comparing climate–REC functions of total, rural, and urban data for Tibet, provide new insights regarding the role of precipitation in affecting REC, as well as the rural–urban difference in climatic impact on REC. Collectively, we have developed a new model-agnostic method that can quantify the total effect of the climate, i.e., the combined effect of temperature, precipitation, and their seasonal variations, which can be used to improve damage functions for different sectors, thus enhancing the accuracy of SCC estimation to better inform climate-related policymaking and policy assessment.

2. Materials and Methods

2.1. Research Design

Previous studies indicate that temperature, precipitation, income, population, and urbanization are major factors influencing the REC [13,14,16,25,26,39–41]. Accordingly, we chose temperature, precipitation, income, and population as explanatory variables, and REC as the response variable to model. Urbanization was reflected through rural and urban comparison.

In contrast to previous studies based on the distributional approach and data modeling, we adopted a randomized [42] or data-driven approach based on algorithm modeling [43]. We built models that can estimate the total, rural, and urban REC in Tibet by testing the datasets with eight algorithms and select the one with best baseline performance to tune. We compared the results of these models using a new model-agnostic interpretation method to reveal the individual effects of temperature and precipitation on the REC and how the urban–rural disparity influences the effect of climate on the REC in Tibet (Figure 1).

More specifically, we first built a pool of candidate algorithms including seven nonparametric machine learning algorithms and one parametric algorithm. The non-parametric machine learning algorithms are used to build randomized data-driven models while the parametric algorithm is used to build a distributional model. Their difference will be discussed further in Section 2.3.1. The parametric distributional algorithm is included as a parallel for comparison with non-parametric data-driven algorithms. Detailed descriptions of the algorithms will be provided in Section 2.3.2. We tested the datasets with each algorithm in the candidate pool to select the algorithm with best baseline performance to build models. The modeling process will be described in detail in Section 2.3.2. If the best-performing algorithm is randomized, we will interpret the model results with the tools of feature importance score (FIS) and partial dependence plot (PDP), both of which will be explained in detail in Section 2.3.3. If the best-performing algorithm is not randomized, we will interpret the model results both in the traditional way with coefficients and innovatively with PDPs. We will compare performance and interpretability of the data-driven models and the distributional model to show the strength of the data-driven approach. Interpretations of the models trained with datasets at different scales will be compared quantitatively using FIS and qualitatively using PDP for insights into climatic effect on REC.

2.2. Data

We used total, rural, and urban panel data from 2014 to 2017 comprising REC, temperature, precipitation, income, and population information for Tibet. We used the monthly REC for total, rural, and urban Tibet from the Tibet State Grid Tibet Electric Power Company. The monthly mean temperature and monthly precipitation were extracted based on the administrative boundary of Tibet from the temperature [44] and precipitation data [45] for China, which has a spatial resolution of 1 km \times 1 km and a temporal resolution of month and year, from the National Tibetan Plateau/Third Pole Environment Data Center. For the socioeconomic indicators of income and population, we used the yearly total, rural and urban per capita disposable income, and annual total, rural, and yearly urban population for Tibet from the National Bureau of Statistics of China.



Quantification of climatic and socioeconomic impacts

Figure 1. Research design.

To examine the climate–REC relationships at different scales, we further divided the Tibet data into warm- and cold-season datasets. We divided the cold and warm seasons using the monthly temperature, with months with a mean temperature below and above zero representing the cold and warm season, respectively.

2.3. Methods

2.3.1. Distributional vs. Randomized Approach

Modeling involves associating explanatory variables with response variables [43]. Computer scientists developed two approaches for this purpose [43]: distributional and randomized. In the distributional approach, a certain distribution of the inputs is assumed by visualizing the raw data, and the most efficient algorithm based on this assumption is

identified. In the randomized approach, no assumption regarding the input distribution is made and the data shapes the algorithm based on stochastic moves in the computation [42]. Modeling practices under these two approaches are known as data modeling and algorithm modeling, respectively [43]. Data modeling [43] features a pure algorithm, the computational analog of a mathematical function [42], which always produces the same results with the same inputs without mutation. In contrast, algorithm modeling [43] features a randomized algorithm [42], which effectively randomizes the inputs and produces results based on probability; that is, the same inputs do not always return the same result. Most machine learning algorithms are randomized by default.

The randomized approach has unique advantages in that it can define the complex relationship between climate and human life. First, the randomized approach provides data to decide which algorithm fits best, thereby avoiding potential distortions created by the subjective judgement on how the function should be shaped to best fit the data. Second, the data-driven nature of this approach enables the models to renew themselves with new data inputs automatically and in real-time, which accommodates uncertainties caused by climate change. Finally, as we discuss in Sections 2.3.2 and 2.3.3, model-agnostic interpretation tools developed under the earth-system perspective can enable interpretations to be readily compared quantitatively and qualitatively across different algorithm models, which can effectively facilitate interdisciplinary and cross-section analysis.

2.3.2. Modeling the Climate–REC Functions

Most climate–REC studies have been based on econometric models built using the distribution approach, which assume that real-world REC, climate, and socioeconomic data conform to a mathematical function comprising parameters and random errors. Such a "too-good-to-be-true" assumption can lead to unintended side effects, such as overlooking the key driver of precipitation, as discussed later. Meanwhile, for algorithm modeling with minimal presumptions, it is difficult to determine whether the algorithm works well with the data before testing. We created a pool of eight mainstream candidate algorithms [46]. They are k-nearest neighbors (KNN), support vector regression (SVR), and classification and regression tree (CART), adaptive boosting (AB), random forest (RF), extra trees (ET), and the gradient boosting machine (GBM) and multivariate parameter-based ordinary least squares (OLS).

The KNN algorithm, based on distances between instances in the dataset, assumes that similar elements are close to each other. It selects a specified number of instances (K) that are closest to the one concerned, and then chooses the most frequent label for classification or averages the labels for regression [47]. In the SVR algorithm, one or more hyperplanes are constructed to separate data. Good separation is achieved by the hyperplane that has the longest distance from all data points nearby, as it minimizes the generalization error [48]. The CART algorithm is commonly referred to as the decision tree since it is constructed by splitting nodes into two child nodes, repeatedly. The split occurs on the input variable that minimizes the Gini index, a performance measure that measures how likely a randomly selected instance is wrongly classified [49]. The structure of CART makes it innately immune to correlation among input variables [50].

The algorithms of AB, RF, ET, and GBM are ensembles of CART. An ensemble combines several base algorithms (in this case, CART) into one, so as to achieve better performance. RF and ET are bagging ensembles, where individual trees are independently trained, and predictions are averaged across trees. The RF and ET differ regarding the splitting of individual trees and data sampling. Splits occurs where the performance measure is the best for RF but randomly for ET. Additionally, the RF sub-samples the data by replacement or bootstrapping, whereas ET uses the original samples [51,52]. AB and GBM are boosting ensembles that grow one tree at a time, with each new tree correcting errors of the previous tree. However, the two differ regarding the identification and correction of previous errors. AB finds errors with high-weight data points by up-weighting previously misclassified observations in each new tree. GBM, on the other hand, finds errors with the gradient

loss function derived from previous trees. AB makes corrections by assigning weights to the trees according to their performances. GBM, however, weights each tree equally but restricts their predictive capacities using the learning rate, which controls the speed of correction from one tree to the next [53–55].

In addition to the above-mentioned algorithms, which are all non-parametric, we also included simple linear regression as a representative pure algorithm in our candidate pool—a simple OLS model shown in Equation (1):

$$REC_i = \beta_0 Temp_i + \beta_1 Preci_i + \beta_2 Income_i + \beta_3 Pop_i + \omega_i$$
(1)

where *i* denotes the time frame of the sample instance, with REC_i denoting REC, $Temp_i$ the mean temperature, $Preci_i$ the mean precipitation, $Income_i$ the income level, and Pop_i the population during the period *i*. The coefficients β_0 , β_1 , β_2 , β_3 indicate effects of temperature, precipitation, income, and population, respectively, and ω_i represents the error. Unlike the nonparametric algorithms, we have to presume the shape of the climate-REC function for Tibet. Previous studies found U-shaped temperature-REC relationships [13], i.e., REC increases both as cold temperature falls and warm temperature increases. However, since Mexico, with minimal heating infrastructure, does not show an REC increase with temperature falling on the lower end [25], we presume for alpine Tibet, with minimal cooling infrastructure, that the REC increase with increased temperature at the higher end will also be absent. With the presumption that REC increases as cold temperature falls in Tibet, we did not make the OLS scale logarithmic to avoid over-representing the impact of humans compared to other algorithms, thus providing a fairer data-driven comparison.

To accommodate the stochastic nature of non-parametric algorithms, we ran each candidate algorithm 100 times using the total, rural, and urban data with their default hyperparameters in a Python 3.7 environment with the utility of scikit-learn [56]. Data are standardized to evaluate all candidate algorithms. In the workflow, 80% of the data are used for algorithm testing and modeling, while 20% are reserved for validation. We split the training and reserved validation datasets randomly. We used 10-fold cross validation and the mean squared error (MSE) to estimate the performance of the algorithms. We averaged the performances of the 100 runs based on the MSE, as shown in Figure 2 (cf., results in Table S1). Linear regression yields the best fit for rural data, whereas GBM is preferable for the urban and total data of Tibet. We tuned the hyperparameters of the algorithms with corresponding data using GridSearchCV in the scikit-learn implementation [56] and constructed the models. Default hyperparameters were applied for the linear regression when building the rural model.

That different algorithms were used for rural and urban data for the same location highlights the advantage of algorithm modeling over data modeling. Since algorithm modeling does not make assumptions, it provides insights regarding the rural–urban disparity. The same algorithm was used for the total and urban data, suggesting that the urban mechanism plays a dominant role in Tibet's total REC.

2.3.3. Model Interpretation

Models built based on the distribution approach are straightforward. The signs and absolute values of the coefficients indicate how the corresponding variable affects the result and by how much. Although models built based on the randomized approach are known for being "black boxes" that are hard to interpret [57,58], they have been proven to be interpretable using tools such as FIS [57,59,60] and PDP [59,61–63].



Figure 2. Algorithm baseline performance using the rural (**a**), urban (**b**), and total (**c**) Tibet datasets. Negative MSEs of different algorithms for different data are shown in pink and their average ranking for the 100 runs is shown in blue. The shorter the combination of the two bars, the better the performance of a given 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 (OLS) model.

The FIS addresses the "how much" problem as it reflects how useful each feature—the equivalent of the explanatory variable in machine learning—is at predicting the response variable (REC in our case) [59]. We used the FIS to interpret the GBM-based urban and total models. GBM is a tree-based ensemble and the FIS of a GBM is calculated by averaging the FISs of the features across all individual trees [60]. More specifically, the FIS of a decision tree is calculated as Equation (2) [64] in the scikit-learn [56]:

$$fi_{i} = \frac{\sum_{j: \text{ node } j \text{ splits on feature } i ni_{j}}{\sum_{k \in \text{all nodes } ni_{j}}},$$
(2)

where f_{i_i} denotes the FIS of feature *i*, n_{i_j} the importance of node *j* where feature *i* is used to split the tree; n_{i_j} is calculated using Equation (3) [64]:

$$ni_j = w_j C_j - w_{\text{left}(j)} C_{\text{left}(j)} - w_{\text{right}(j)} C_{\text{right}(j)}$$
(3)

where w_j denotes the weighted samples reaching node *j*, C_j the Gini index of node *j*, *left*(*j*) the child node from the left split on node *j*, and *right*(*j*) the child node from the right split on node *j*.

The scikit-learn enables FIS comparison across models by normalizing each feature importance value to percentages by dividing it by the sum of all feature importance values, as shown in Equation (4) [56,64]:

$$normfi_i = \frac{fi_i}{\sum_{j \in \text{all features }} fi_j} \tag{4}$$

For tree-based ensembles, FIS for each feature is calculated by averaging FISs of all trees involved, as shown in Equation (5) [56,64]:

$$TreeEnsemblef_{i_i} = \frac{\sum_{j \in all trees} normf_{i_j}}{T}$$
(5)

where T denotes the number of trees in the ensemble.

The higher the FIS is, the more important is the corresponding feature for the model. We used FIS to determine how important temperature, precipitation, income, and population are for REC estimation and compared across models.

Regarding the "how", the contribution of a feature to the results of an algorithm model cannot be generalized as the signs of the coefficients because randomized algorithms are multidimensional. However, several methods can reveal the data mechanisms behind, such as PDPs [59,63]. The PDP function is a model-agnostic interpretation tool that is defined in Equation (6) [56]:

$$pd_{X_{S}}(x_{S}) = \int_{-\infty}^{def} \mathbb{E}_{X_{C}}[f(x_{S}, X_{C})] = \int_{-\infty}^{\infty} f(x_{S}, x_{C})p(x_{C})dx_{C},$$
(6)

where X_S represents the feature to be analyzed, X_C represents all other input features, and f is the model function. The PDP function is used to marginalize the model output over the distribution of X_C such that the function reveals the correlation between X_S and the model outcome [50,59,61].

However, the prerequisite of the PDP function is that X_S and X_C are not correlated [59,61–63]. Table 1 shows that the temperature and precipitation as well as the income and population in the Tibet data are correlated, thus the PDP cannot be applied for individual features.

Pair of Variables Season Type Spearman's Correlation Coefficients Highly Correlated (>0.5) 0.89 Year-round Yes _ Temperature-0.5 Cold season _ No Precipitation 0.81 Yes Warm season Total 1 Yes Income-Population -0.95Rural Yes Urban 1 Yes -

Table 1. Results of Correlation Tests for the Tibet Datasets.

2.3.4. Addressing the Challenge of Multicollinearity

Multicollinearity represents a challenge for both data and algorithm models and may account, in part, for precipitation being overlooked in previous literature. Since the precision of the estimated coefficients can be comprised by multicollinearity, it is common practice to remove or lessen the weight of one of the highly correlated variables (in this case, precipitation) [65]. However, in our linear rural model, we did not remove a basic climate indicator, such as precipitation. Instead, we addressed the multicollinearity by dividing the data into subsets in which temperature and precipitation were not strongly correlated and by using a tree-based rural model for verification. More specifically, we divided the

rural data into cold and warm seasons based on whether the monthly temperature was below or above 0 °C, respectively, yielding a cold-season dataset in which the temperature and precipitation were not strongly correlated (Table 1). In addition, we used the best-performing tree-based algorithm for the rural Tibet dataset (RF) to construct a tree-based rural model.

The urban and total models are tree-based GBM ensembles that are innately immune to multicollinearity, as discussed in Section 2.3.2. Based on the calculation of the FIS for GBM, as discussed in Section 2.3.3, the FIS is also immune to multicollinearity. Therefore, we used the FIS to interpret how important each input feature is for the REC estimation with a special interest in the effect of precipitation. We divided the total data into cold and warm seasons for the modeling depending on whether the monthly temperature was below or above 0 $^{\circ}$ C to test our hypothesis regarding the boosting effect of snowfall.

The FIS did not reveal how the features affect the REC; therefore, PDP interpretation was used. However, tree-based models were also challenged by multicollinearity regarding the PDP interpretation. The strong correlation between the temperature and precipitation prevented the quantification of their individual impacts using PDP. Thus, we quantified their joint effect—combined effect of temperature and precipitation or climate effect—by inserting tuples into X_S and X_C in Equation (1). More specifically, we divided the input features in two groups; namely, a climate group *Cl* comprising temperature and precipitation, to accommodate the assumption for PDP that X_S and X_C in Equation (6) are not correlated:

$$Cl = \{c_i | c_i = (Temp_i, Preci_i)\}$$
(7)

$$Sc = \{sc_i | sc_i = (Income_i, Pop_i)\}$$
(8)

$$pd_{Cl}(c_i) \stackrel{def}{=} \mathbb{E}_{X_C}[f(c_i, Sc)] = \int f(c_i, sc_i)p(sc_i)dsc_i$$
(9)

where *Cl* and *Sc* groups do not correlate, thus facilitating quantification of the total contribution of the climate inputs and that of the socioeconomic inputs using PDP functions. As PDP is a model-agnostic tool, it can also be used for linear models. We thus quantified the climate and socioeconomic effects on the REC by applying the PDP function to the total, rural, and urban models of the Tibet REC. We used the PDP tool in scikit-learn [56] for the PDP computation and visualization.

The existence of climate patterns indicates that the combination of temperature and precipitation is not random. We thus obtained the mean climate pattern for Tibet with calendar-month mean temperature $TMEAN_m$ (Equation (10)) and calendar-month mean precipitation $PMEAN_m$ (Equation (11)). We then determined the potential variations of the mean climate pattern by calculating standard deviations of the monthly mean temperature $TMEAN_m$ and monthly mean precipitation $PMEAN_m$ for the calendar months to obtain the climate pattern for Tibet C_m (Equation (12)):

$$TMEAN_m = \frac{1}{n_{year}} \sum_{i=1}^n T_{im}$$
(10)

$$PMEAN_m = \frac{1}{n_{year}} \sum_{i=1}^n P_{im}$$
(11)

$$C_m = [(TMEAN_m \pm STD(T_m)), (PMEAN_m \pm STD(P_m))]$$
(12)

Finally, we averaged the marginal effects of all climate instances within the variation range for each month to obtain the average effect of climate for each calendar month in Tibet as follows:

$$R_m = C_m \cap Cl' \tag{13}$$

$$pd_{cl'}(r_m) \approx \frac{1}{n_{R_m}} \sum_{j=1}^n f\left(r_m, sc_i^{(j)}\right) \tag{14}$$

where R_m is the intersection between the climate pattern of Tibet C_m and all the climate combinations simulated during the PDP calculation Cl'. n_{R_m} denotes the number of instances r_m in R_m and $pd_{Cl'}(r_m)$ denotes the averaged climatic impact on REC in month m. However, this method cannot be applied to the socioeconomic inputs as a socioeconomic pattern governed by physical laws is not available for reference.

3. Results

3.1. Effect of Precipitation

Robust interpretation can be achieved only with models with a high predictive accuracy. As Figure 2 indicated, non-parametric machine learning algorithms show better baseline performances than linear regression for urban and total Tibet datasets, but not for the rural Tibet dataset. We thus first examined the linear regression-based rural models, using R² to measure its performance. Based on the results of the rural model in Table 2, 80.8% of the rural data can be explained with the model, which can be considered as a good performance. However, the performance of the model based on regrouped data notably decreased to 74.5% for cold-season data and significantly increased to 92.7% for warm-season data. Hence, the OLS explains warm-season data better than cold-season data.

Table 2. Results of the rural model for Tibet.

Data	Response Variable	R ²	Coefficient	Predictor Variables	р
Tibet Rural 2014–2017 (year-round)	REC	80.8%	-38.7626	Temperature	0.698
			-71.6497	Precipitation	0.462
			600.9989	Income	0.000
			-207.8950	Population	0.175
Tibet Rural 2014–2017 (cold season)	REC	74.5%	121.7221	Temperature	0.479
			-404.0049	Precipitation	0.285
			689.3523	Income	0.001
			1.0650	Population	0.996
Tibet Rural 2014–2017 (warm season)	REC	92.7%	-80.0322	Temperature	0.757
			26.1458	Precipitation	0.809
			462.7637	Income	0.002
			-507.2098	Population	0.007

The coefficients obtained for the temperature and precipitation in the year-round and warm-season linear rural models may not be accurate owing to the strong correlation between temperature and precipitation. However, the cold-season linear model, with a weaker correlation between temperature and precipitation, performed poorer compared with the year-round and warm season. This shows the limitation of the distributional approach in differentiating the effects of correlated variables. Therefore, we analyzed the coefficients only for inference. The coefficients suggest that precipitation plays a more significant role than temperature throughout the year, particularly in the cold season, and precipitation and temperature may have opposite effects on REC in the cold season. The opposite effects suggest the climate-REC relation in the cold-season in rural Tibet is probably nonlinear.

The *p*-value of the income, which is statistically significant across all scales, indicates that the income is the most decisive factor, which is consistent with the finding of previous studies [16,66]. However, neither temperature nor precipitation is found to be statistically significant by the linear model, which contradicts previous studies [13,16]. This indicates that the distributional model, even with a better baseline performance, cannot provide satisfactory interpretations without transforming the datasets to suit additional human assumptions as previous studies did [13,14].

Given the multicollinearity of year-round and warm-season data, as well as the weak linear model performance for the cold season, we verified this interpretation by building an alternative rural model using the best-performing tree-based algorithm RF for the rural dataset. As discussed in Section 2.3, this tree-based rural model is innately immune to multicollinearity and its FIS results can differentiate contributions from correlated variables such as temperature and precipitation.

The performance of urban and total models based on the ensemble algorithm was validated using the MSE of the data set before modeling. We also validated the linear and tree-based rural models for comparison (Figure 3). The validation results were good for all four models, suggesting that the models were not over-fitted. After tuning, the RF-based rural model outperformed the linear rural model, although the linear model had a higher baseline performance. We thus used both models as valid rural models in our discussion.



Figure 3. Model validation results. Plots (**a**–**d**) show the model validation results of the tree-based rural model, linear rural model, urban model, and total model for Tibet. Plots (**e**,**f**) show the results of the model trained without precipitation and without temperature, respectively. The red curves represent the REC estimates made based on the model and the blue curves show the validation REC data. MSE, mean squared error.

We extracted the FISs from the tree-based year-round rural model as well as the urban and total year-round models (Figure 4). The FIS results indicate that climate contributes about 16.46% and socioeconomic factors contribute about 83.55% to total Tibet REC estimation. The tree-based rural year-round model validated our inferences based on the linear rural model that the income with highest FIS (FIS ~44.52%) is the most important for Tibet's rural REC estimation, while precipitation with a slightly higher FIS (~7.43%) than temperature (FIS ~7.36%) is the more important climatic factor throughout the year. For the urban Tibet REC, the population (FIS ~43.55%) was the most important factor, and precipitation was the more important climatic factor (FIS ~13.72%) than temperature (FIS ~11.34%). For the total Tibet REC, precipitation (FIS ~8.98%) remained the more important climatic factor than temperature (FIS ~7.48%); population (FIS ~44.02%) was identified as the most important factor overall. The total data cancelled out the effect of urbanization on the REC as it generalized the rural-urban disparity. Given that the FISs of different models were comparable, precipitation was more influential in urban Tibet (FIS ~13.72%) than in rural Tibet (FIS ~7.43%).

To be sure of the importance of precipitation, we also estimated Tibet's total REC based on biased datasets, which included only the temperature or the precipitation as climate input (Figure 3e,f, respectively). The validation is considerably better with precipitation (Figure 3f) than without (Figure 3e). The model trained without temperature (Figure 3f), however, performs almost as well as the model trained with both the temperature and precipitation (Figure 3d). This confirms the FIS results that precipitation is the more important climatic factor for estimating Tibet REC.

Considering that the correlation between the temperature and precipitation was weak in the cold season (Table 1), we consider precipitation became the stronger driver of the REC at temperatures below 0 °C. We infer that the climate pattern in the Tibetan Plateau region may be the key for explaining the significant effect of precipitation. Precipitation in months with below-zero temperatures likely occurred as snowfall, suggesting that the effect of precipitation was most likely based on snowfall. The steep and high topography of the Tibetan Plateau led to colder weather; thus, precipitation more likely occurs as snow. Snowy weather can increase the electricity consumption by both generating a high energy demand for heating and keeping residents indoors. To test this theory, we modeled cold-season and warm-season total, rural and urban data, and assessed the validation and FISs.

We considered the performance of cold-season and warm-season total, rural and urban models (cf., results in Figure S1) to be acceptable for FIS interpretation. All the models suggest that precipitation is the more important climatic factor in the cold season (Figure 4), whereas the temperature is more important for estimation in the warm season (Figure 4). This supports our hypothesis that precipitation impacts REC through snowfall events during the cold season.

It is noteworthy that the rural cold-season model registered the highest FIS across the board with income contributing ~57.68% to its estimation, suggesting rural Tibet REC during the cold season is highly sensitive to income. While income is the more important socioeconomic factor than population across all the rural models (Figure 4), population is found to be more important (FIS~43.55%, FIS~49.82%) than income (FIS~31.39%, FIS~28.82%) for urban year-round and cold-season models (Figure 4).

3.2. Climate and Socioeconomic Effects

We used PDPs to quantify the climatic and socioeconomic effects on the REC for the total, rural, and urban data of Tibet (Figure 5). In contrast to the FISs, PDPs are model-specific and thus cannot be compared across models quantitatively. We compared the quantification of the climatic effects using graphs. Both the linear rural model (Figure 5a) and tree-based rural model (Figure 5c) were interpreted. Based on the PDP definition, the PDP of linear models will also be linear, which explains the smoother PDP of the linear rural model compared with the rugged PDP of the nonlinear tree-based rural model. Although the tree-based rural model provides more detailed information, it is consistent with the

linear rural model. Both suggest that the effect of the climate on the REC is heating-driven as the effect of the climate increases with decreasing temperature. The same is true for the urban (Figure 5e) and total (Figure 5g) models. Interpreting precipitation, as discussed in Section 3.1, cannot be isolated from the temperature. We thus used the climate pattern to interpret the effect of the precipitation, which will be discussed in Section 3.3.



Figure 4. Feature importance score (FIS) results of tree-based models estimating Tibet REC at different scales. Plots (**a**,**b**) show the FIS results of climate and socio-economic inputs, respectively. The FIS was standardized and can be compared across models.



Figure 5. Climatic and socioeconomic effects on Tibet's REC. All plots are partial dependence plots (PDPs), with the color panel indicating how much the respective feature contributes to the REC estimation of the respective model. Plots (**a**,**c**,**e**,**g**) are PDPs of climate features and plots (**b**,**d**,**f**,**h**) are PDPs of socioeconomic features. Plots (**a**,**b**), (**c**,**d**), (**e**,**f**), and (**g**,**h**) present the results of the linear rural model, tree-based rural model, urban model, and total model, respectively.

Socioeconomic factors differently affected Tibet's rural REC (Figure 5b,d) and Tibet's urban and total REC (Figure 5f,h). Income growth increased the REC in all models. The main difference can be observed for the population. Rural population growth decreased the REC, whereas urban population growth increased the REC. One of the defining features of urbanization is that the rural population decreased whereas the urban population increased. Based on placing the different roles of the rural and urban population in urbanization, the rural, urban, and total models yielded the same conclusion that urbanization increased the REC. The combination of temperature and precipitation is governed by physical laws—the climate pattern—whereas the combination of income and population does not follow a specific pattern. It is, therefore, difficult to differentiate their precise contributions. Accordingly, we conclude that the effect of the climate on Tibet's rural, urban, and total REC is driven by heating-related demands. Income and urbanization boost the REC in rural, urban, and total Tibet.

3.3. Different Climatic Effects Due to Urban–Rural Disparity

We further interpreted the PDPs of the climate features based on the climate pattern using the method expressed in Equation (14), which resulted in Figure 6. The general shapes of the curves of the climate feature PDPs of all models were consistent. The effect of the climate was stronger during the cold season than during the warm season. This confirms our previous conclusion that the effect of climate on the REC in Tibet is driven by the heating-related demand. However, the curves also exhibited differences.



Figure 6. Effect of the climate on the REC in rural, urban, and total Tibet. All plots show the marginal effect of the climate feature in the respective model (purple curve). The temperature and precipitation averaged by the calendar month are represented by blue and green curves, respectively. Plots (**a**–**d**) show the results of the linear rural model, tree-based rural model, urban model, and total model, respectively.

Similar to the PDPs, the curve of the linear rural model was smoother (Figure 6a) than the tree-based rural model (Figure 6b). The curve representing the rural climatic effect on the REC was V-shaped, whereas that of the tree-based rural model was U-shaped, except for the significant increase in April. Given that the tree-based rural model outperformed the linear rural model after tuning, the April peak as well as other smaller fluctuations were likely smoothed out by the linearity in the linear rural model. This highlights the advantage of the machine learning algorithm in capturing smaller changes in the data mechanism.

The shapes of the curves for the urban and total climate features were similar, with the climatic effect being the strongest during the winter months (November, December, and January), as well as peaks in April and September. This indicates that the data mechanism underlying the urban model is dominant in Tibet.

Both the tree-based rural model and urban model exhibited the April peak when the mean temperature was below zero, and the precipitation was the highest for the below-freezing months. Combined with our earlier conclusion regarding the role of precipitation, this peak was caused by the effect of snowfall. The April peaks suggest that the precipitation (snowfall) in the cold season boosts the REC in Tibet.

However, the September peak visible in the urban model was absent in the rural model. September is characterized by a mean temperature above zero and the highest precipitation after the summer months of June, July, and August. Because the temperature drops below zero in October, the potential of snowfall or freezing rain during the latter part of September is high. Our earlier conclusion that the effect of precipitation is stronger in urban Tibet than in rural Tibet may explain the absence of the September peak in the curve of the rural model.

We further investigated the factors that contribute to the stronger effect of precipitation in urban Tibet. Considering that the climatic inputs of the rural and urban models are the same, the different magnitudes of the effect of the climate in September on the rural and urban REC can only be caused by the contribution of different socioeconomic factors. The difference between the rural–urban population, as discussed in Section 3.2, is related to the urbanization effect, based on which the REC in both communities increases. Therefore, the only factor that might lead to this difference in the effect of the climate is the income. Thus, the difference in the sensitivity to electricity costs due to the rural–urban income disparity has contributed to the differing effect induced by the climate on the REC in September. It is likely that electricity-sensitive rural residents respond to an ever-colder September by wearing more clothes, whereas their urban peers who are less sensitive to electricity costs likely employed electricity-aided heating for comfort.

Therefore, we conclude that the snowfall increases the REC in Tibet. The rural–urban income disparity leads to the stronger effect induced by climate on urban Tibet than on rural Tibet, particularly in September when electricity-aided heating is considered optional.

4. Discussion

Our results provide insights into the contribution of precipitation and socioeconomic disparity to the effect of climate on the REC in alpine regions, as well as elsewhere, during cold seasons. Our results highlight that precipitation plays a major role in shaping climatic impact, which has been largely overlooked in previous studies. Our study highlights that climate-related energy studies should treat climate as a whole rather than focusing on the temperature alone, particularly for studies concerning cold winters and extreme weathers such as blizzards.

This study is useful for policymaking regarding climate change adaptation, for example, based on the climate projection that Tibet is becoming warmer and wetter, we suggest that the climate-related REC in Tibet may not just fall with warming. If more precipitation in the region leads to more snowfall events, then climate-related REC will increase considerably in the region. With urbanization and income growth, climate is expected to contribute more to Tibet REC since income growth can lead to a higher adoption rate of AC and a lower sensitivity of energy cost. As the region is experiencing fast climate change as well as economic growth, monitoring systems should be built for the continuous observation of climate-REC dynamics to inform adaptation efforts of grid planning and power distribution in the region.

This study can also benefit early warning projects aimed at protecting against power outages induced by extreme weather, since the model developed in this study is purely data-driven and can be easily replicated using high-frequency meteorological and REC data for short-term REC prediction. The proven predictive power of machine learning models can enable REC prediction based on meteorological forecast, thus helping warn against extreme weather-induced power surges and prepare the power transmission system in advance.

Moreover, this study emphasizes the advantage of the nonparametric machine learning algorithms based on their better predictive accuracy and interpretability necessary for building data-driven interdisciplinary models. While the distributional models with readily available coefficients are easier to interpret, their interpretations are severely challenged by multicollinearity. Compared with previous studies using a distribution approach, our method facilitates interpretation of the importance of temperature and precipitation separately, as well as their combined effect. By addressing the multicollinearity, our model-agnostic interpretation method has the potential to be used to improve damage functions of different sectors and better inform climate-related policymaking and policy assessment. Our method can be easily replicated for other interdisciplinary and cross-section analysis in earth system science, as human presumptions have been minimized throughout the modeling process.

A limitation of our study is the qualitative analysis of the contribution made by socioeconomic factors. Thus, additional investigation into the correlations among income, population, and urbanization are needed to more accurately determine the role of income and population.

Besides, the quality of data-driven models depends on the data. For REC data, monthly is the highest frequency available in Tibet for now. Our study is thus kept from investigating linkages between snowfall events and corresponding real-time REC due to data resolution. However, the model validation results of our study have proven the strength of machine learning algorithms with limited data. We would expect more insights when our study is replicated using future penal data of a higher frequency.

Also, while the FIS tool has enabled individual contributions of temperature and precipitation to be compared quantitatively across models, the PDP-enabled climatic impact quantification remains model-specific and can only be compared qualitatively across models. We would expect future advances in interpretable machine learning to help shed light on this.

5. Conclusions

Our study found that the climate contributes about 16.46% and socioeconomic factors contribute about 83.55% to total Tibet REC. Climatic effects on REC are stronger in urban Tibet (FIS ~25.06%) than rural Tibet (FIS ~14.79%). For total, rural, and urban Tibet's REC estimation, precipitation is more important (FIS ~8.98%, FIS ~7.43%, FIS ~13.72%) than temperature (FIS ~7.48%, FIS ~7.36%, FIS ~11.34%). The effect of precipitation is largely based on snowfall during the cold season when the monthly mean temperature is below 0 °C. Rural Tibet REC is more sensitive to income, especially during the cold season (FIS ~57.68%). Urban Tibet REC is generally more responsive to population (FIS ~43.55%) except for during the warm season when income becomes more important (FIS ~52.72%). The rural–urban income disparity has resulted in a stronger climate-based effect on urban Tibet than on rural Tibet, particularly in September when electricity-aided heating is considered optional. The results of our study can help improve the climate–energy damage function for SCC estimation and inform climate change adaptation efforts in the region. Our method can be easily replicated with meteorological forecast data to warn against extreme weather

induced power outages. With few presumptions, the method is also readily available for other interdisciplinary and cross-section studies.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/en15093355/s1, Table S1. Results for the Tibet 2014–2017 data at different scales; Figure S1. Model validation results.

Author Contributions: Conceptualization, C.X. and T.Y.; methodology, C.X. and W.H.; validation, C.X., W.H.; formal analysis, C.X.; resources, T.Y. and W.W.; data curation, C.X. and W.W.; writing—original draft preparation, C.X.; writing—review and editing, T.Y., W.W. and W.H.; supervision, T.Y.; project administration, T.Y. and W.W.; funding acquisition, T.Y. and W.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Second Tibetan Plateau Scientific Expedition and Research (STEP) project, grant number 2019QZKK0208; the Strategic Priority Research Program of the Chinese Academy of Sciences, grant number XDA20100300 and the National Natural Science Foundation of China, grant number 41771088.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that supports the findings of this study are available from data sources as stated in Section 2.2. Restrictions apply to the availability of the REC data, which were used under license for this study. Data are available from the authors with the permission of data sources.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. IPCC. Contribution of Working Group I to the Sixth Assessment Report. In *Climate Change 2021: The Physical Science Basis;* IPCC: Geneva, Switzerland, 2021.
- 2. Nordhaus, W.D. Revisiting the social cost of carbon. Proc. Natl. Acad. Sci. USA 2017, 114, 1518–1523. [CrossRef]
- 3. Stern, N.; Stiglitz, J.E. *The Social Cost of Carbon, Risk, Distribution, Market Failures: An Alternative Approach*; National Bureau of Economic Research: Cambridge, MA, USA, 2021.
- 4. Aldy, J.E.; Kotchen, M.J.; Stavins, R.N.; Stock, J.H. Keep climate policy focused on the social cost of carbon. *Science* 2021, 373, 850–852. [CrossRef]
- 5. Steffen, W.; Richardson, K.; Rockström, J.; Schellnhuber, H.J.; Dube, O.P.; Dutreuil, S.; Lenton, T.M.; Lubchenco, J. The emergence and evolution of Earth System Science. *Nat. Rev. Earth Environ.* **2020**, *1*, 54–63. [CrossRef]
- CICC Global Institute. Carbon Neutrality Economics: Macro and Industry Trends under New Constraints; China CITIC Press: Beijing, China, 2021.
- Mideksa, T.K.; Kallbekken, S. The impact of climate change on the electricity market: A review. *Energy Policy* 2010, 38, 3579–3585. [CrossRef]
- Del Río, P.; Silvosa, A.C.; Gómez, G.I. Policies and design elements for the repowering of wind farms: A qualitative analysis of different options. *Energy Policy* 2011, 39, 1897–1908. [CrossRef]
- 9. Himpler, S.; Madlener, R. Repowering of Wind Turbines: Economics and Optimal Timing; FCN: Aachen, Germany, 2011.
- De la Puente-Gil, A.; González-Martínez, A.; Borge-Diez, D.; Martínez-Cabero, M.-A.; de Simón-Martín, M. True power consumption labeling and mapping of the health system of the Castilla y León region in Spain by clustering techniques. *Energy Procedia* 2019, 157, 1164–1181. [CrossRef]
- 11. Li, D.H.; Yang, L.; Lam, J.C. Impact of climate change on energy use in the built environment in different climate zones–A review. *Energy* **2012**, *42*, 103–112. [CrossRef]
- 12. Auffhammer, M.; Baylis, P.; Hausman, C.H. Climate change is projected to have severe impacts on the frequency and intensity of peak electricity demand across the United States. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 1886–1891. [CrossRef]
- 13. Auffhammer, M.; Mansur, E.T. Measuring climatic impacts on energy consumption: A review of the empirical literature. *Energy Econ.* **2014**, *46*, 522–530. [CrossRef]
- 14. Deschênes, O.; Greenstone, M. Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US. *Am. Econ. J. Appl. Econ.* **2011**, *3*, 152–185. [CrossRef]
- 15. Waite, M.; Cohen, E.; Torbey, H.; Piccirilli, M.; Tian, Y.; Modi, V. Global trends in urban electricity demands for cooling and heating. *Energy* 2017, *127*, 786–802. [CrossRef]
- 16. Li, Y.; Pizer, W.A.; Wu, L. Climate change and residential electricity consumption in the Yangtze River Delta, China. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 472–477. [CrossRef]
- 17. Auffhammer, M. Climate Adaptive Response Estimation: Short and Long Run Impacts of Climate Change on Residential Electricity and Natural Gas Consumption Using Big Data; National Bureau of Economic Research: Cambridge, MA, USA, 2018.

- De la Puente-Gil, Á.; González-Martínez, A.; Borge-Diez, D.; Blanes-Peiró, J.J.; de Simón-Martín, M. Electrical Consumption Profile Clusterization: Spanish Castilla y León Regional Health Services Building Stock as a Case Study. *Environments* 2018, 5, 133. [CrossRef]
- 19. Rezaei, K.; Pradhan, B.; Vadiati, M.; Nadiri, A.A. Suspended sediment load prediction using artificial intelligence techniques: Comparison between four state-of-the-art artificial neural network techniques. *Arab. J. Geosci.* **2021**, *14*, 215. [CrossRef]
- Eskandari, E.; Mohammadzadeh, H.; Nassery, H.; Vadiati, M.; Zadeh, A.M.; Kisi, O. Delineation of isotopic and hydrochemical evolution of karstic aquifers with different cluster-based (HCA, KM, FCM and GKM) methods. *J. Hydrol.* 2022, 609, 127706. [CrossRef]
- 21. Strahler, A.H. Introducing Physical Geography; Wiley: New York, NY, USA, 2011.
- 22. Wenz, L.; Levermann, A.; Auffhammer, M. North–south polarization of European electricity consumption under future warming. *Proc. Natl. Acad. Sci. USA* 2017, 114, E7910–E7918. [CrossRef]
- 23. Eskeland, G.S.; Mideksa, T.K. Electricity demand in a changing climate. *Mitig. Adapt. Strateg. Glob. Chang.* **2010**, *15*, 877–897. [CrossRef]
- 24. Auffhammer, M. Cooling China: The weather dependence of air conditioner adoption. Front. Econ. China 2014, 9, 70-84.
- 25. Davis, L.W.; Gertler, P.J. Contribution of air conditioning adoption to future energy use under global warming. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 5962–5967. [CrossRef]
- 26. Auffhammer, M.; Aroonruengsawat, A. Simulating the impacts of climate change, prices and population on California's residential electricity consumption. *Clim. Chang.* 2011, 109, 191–210. [CrossRef]
- 27. Papakostas, K.; Mavromatis, T.; Kyriakis, N. Impact of the ambient temperature rise on the energy consumption for heating and cooling in residential buildings of Greece. *Renew. Energy* **2010**, *35*, 1376–1379. [CrossRef]
- Jylhä, K.; Jokisalo, J.; Ruosteenoja, K.; Pilli-Sihvola, K.; Kalamees, T.; Seitola, T.; Mäkelä, H.M.; Hyvönen, R.; Laapas, M.; Drebs, A. Energy demand for the heating and cooling of residential houses in Finland in a changing climate. *Energy Build.* 2015, 99, 104–116. [CrossRef]
- 29. Zhang, M.; Zhang, K.; Hu, W.; Zhu, B.; Wang, P.; Wei, Y.-M. Exploring the climatic impacts on residential electricity consumption in Jiangsu, China. *Energy Policy* **2020**, *140*, 111398. [CrossRef]
- 30. Cohen, J.; Screen, J.A.; Furtado, J.; Barlow, M.; Whittleston, D.; Coumou, D.; A Francis, J.; Dethloff, K.; Entekhabi, D.; E Overland, J.; et al. Recent Arctic amplification and extreme mid-latitude weather. *Nat. Geosci.* **2014**, *7*, 627–637. [CrossRef]
- 31. Zhang, W.; Furtado, K.; Wu, P.; Zhou, T.; Chadwick, R.; Marzin, C.; Rostron, J.; Sexton, D. Increasing precipitation variability on daily-to-multiyear time scales in a warmer world. *Sci. Adv.* **2021**, *7*, eabf8021. [CrossRef]
- 32. Busby, J.W.; Baker, K.; Bazilian, M.D.; Gilbert, A.Q.; Grubert, E.; Rai, V.; Rhodes, J.D.; Shidore, S.; Smith, C.A.; Webber, M.E. Cascading risks: Understanding the 2021 winter blackout in Texas. *Energy Res. Soc. Sci.* **2021**, *77*, 102106. [CrossRef]
- 33. Zheng, X. Power Utilities Asked to Ensure Proper Supply. Available online: https://global.chinadaily.com.cn/a/202012/19/WS5 fdd5583a31024ad0ba9cc79.html (accessed on 15 April 2022).
- Aljazeera. Huge US Winter Storm Leaves More than 330,000 without Power. Available online: https://www.aljazeera.com/ news/2022/2/4/huge-us-winter-storm-leaves-more-than-330000-without-power (accessed on 15 April 2022).
- 35. Chen, G.; Zhu, Y.; Wiedmann, T.; Yao, L.; Xu, L.; Wang, Y. Urban-rural disparities of household energy requirements and influence factors in China: Classification tree models. *Appl. Energy* **2019**, 250, 1321–1335. [CrossRef]
- 36. Nie, H.-G.; Kemp, R.; Xu, J.-H.; Vasseur, V.; Fan, Y. Drivers of urban and rural residential energy consumption in China from the perspectives of climate and economic effects. *J. Clean. Prod.* **2018**, 172, 2954–2963. [CrossRef]
- 37. Yao, T. Tackling on environmental changes in Tibetan Plateau with focus on water, ecosystem and adaptation. *Sci. Bull.* **2019**, *64*, 417. [CrossRef]
- 38. Yao, T.; Xue, Y.; Chen, D.; Chen, F.; Thompson, L.; Cui, P.; Koike, T.; Lau, W.K.-M.; Lettenmaier, D.; Mosbrugger, V.; et al. 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. *Bull. Am. Meteorol. Soc.* 2019, 100, 423–444. [CrossRef]
- 39. Frederiks, E.R.; Stenner, K.; Hobman, E.V. The socio-demographic and psychological predictors of residential energy consumption: A comprehensive review. *Energies* **2015**, *8*, 573–609. [CrossRef]
- 40. Fan, J.-L.; Zhang, Y.-J.; Wang, B. The impact of urbanization on residential energy consumption in China: An aggregated and disaggregated analysis. *Renew. Sustain. Energy Rev.* 2017, 75, 220–233. [CrossRef]
- 41. Zheng, S.; Huang, G.; Zhou, X.; Zhu, X. Climate-change impacts on electricity demands at a metropolitan scale: A case study of Guangzhou, China. *Appl. Energy* **2020**, *261*, 114295. [CrossRef]
- 42. Yao, A.C.-C. Probabilistic computations: Toward a unified measure of complexity. In Proceedings of the 18th Annual Symposium on Foundations of Computer Science (sfcs 1977), Providence, RI, USA, 31 October–2 November 1977; pp. 222–227.
- 43. Breiman, L. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Stat. Sci.* **2001**, *16*, 199–231. [CrossRef]
- 44. Shouzhang, P. 1-km Monthly Mean Temperature Dataset for China (1901–2017); National Tibetan Plateau/Third Pole Environment Data Center: Beijing, China, 2019. [CrossRef]
- 45. Shouzhang, P. 1-km Monthly Precipitation Dataset for China (1901–2017); National Tibetan Plateau/Third Pole Environment Data Center: Beijing, China, 2020. [CrossRef]

- 46. Brownlee, J. Machine Learning Mastery with Python; Machine Learning Mastery. 2016, Volume 527, pp. 100–120. Available online: https://books.google.com.hk/books?hl=en&lr=&id=BgmqDwAAQBAJ&oi=fnd&pg=PP1&dq=Machine+Learning+Mastery+ with+Python&ots=frmYXAqM4V&sig=JKLaQWQuS8QRJODNyQhS_dWmYi0&redir_esc=y&hl=zh-CN&sourceid=cndr#v= onepage&q=Machine%20Learning%20Mastery%20with%20Python&f=false (accessed on 15 April 2022).
- 47. Altman, N.S. An introduction to kernel and nearest-neighbor nonparametric regression. Am. Stat. 1992, 46, 175–185.
- 48. Cortes, C.; Vapnik, V. Support-vector networks. Mach. Learn. 1995, 20, 273–297. [CrossRef]
- 49. Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. Classification Regression Trees; Routledge: England, UK, 2017.
- 50. Friedman, J.; Hastie, T.; Tibshirani, R. The Elements of Statistical Learning; Springer: New York, NY, USA, 2001; Volume 1.
- Ho, T.K. Random decision forests. In Proceedings of the 3rd international Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995; pp. 278–282.
- 52. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]
- 53. Kégl, B. The return of AdaBoost. MH: Multi-class Hamming trees. arXiv 2013, arXiv:1312.6086. [CrossRef]
- 54. Mason, L.; Baxter, J.; Bartlett, P.; Frean, M. Boosting algorithms as gradient descent. *Adv. Neural Inf. Processing Syst.* **1999**, 12, 512–518.
- 55. Friedman, J.H. Greedy function approximation: A gradient boosting machine. Ann. Stat. 2001, 29, 1189–1232. [CrossRef]
- 56. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- 57. Azodi, C.B.; Tang, J.; Shiu, S.-H. Opening the Black Box: Interpretable machine learning for geneticists. *Trends Genet.* **2020**, *36*, 442–455. [CrossRef]
- Friedman, J.H.; Meulman, J.J. Multiple additive regression trees with application in epidemiology. *Stat. Med.* 2003, 22, 1365–1381. [CrossRef]
- 59. Molnar, C. Interpretable Machine Learning; Ruboss Technology Corporation: Victoria, BC, Canada, 2020.
- 60. Strobl, C.; Boulesteix, A.-L.; Zeileis, A.; Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinform.* 2007, *8*, 25. [CrossRef]
- 61. Goldstein, A.; Kapelner, A.; Bleich, J.; Pitkin, E. Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *J. Comput. Graph. Stat.* **2015**, *24*, 44–65. [CrossRef]
- 62. Zhao, Q.; Hastie, T. Causal interpretations of black-box models. J. Bus. Econ. Stat. 2019, 39, 272–281. [CrossRef]
- Mehdiyev, N.; Fettke, P. Prescriptive process analytics with deep learning and explainable artificial intelligence. In Proceedings of the 28th European Conference on Information Systems (ECIS), Online AIS Conference, 15–17 June 2020.
- Ronaghan, S. The Mathematics of Decision Trees, Random Forest and Feature Importance in Scikit-learn and Spark. Available online: https://towardsdatascience.com/the-mathematics-of-decision-trees-random-forest-and-feature-importance-in-scikitlearn-and-spark-f2861df67e3 (accessed on 15 April 2022).
- Chong, I.-G.; Jun, C.-H. Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab.* Syst. 2005, 78, 103–112. [CrossRef]
- Du, K.; Yu, Y.; Wei, C. Climatic impact on China's residential electricity consumption: Does the income level matter? *China Econ. Rev.* 2020, *63*, 101520. [CrossRef]