

## Article

# Building Energy Models at Different Time Scales Based on Multi-Output Machine Learning

Guangchen Li <sup>1</sup>, Wei Tian <sup>1,2,\*</sup> , Hu Zhang <sup>1</sup> and Bo Chen <sup>3</sup><sup>1</sup> College of Mechanical Engineering, Tianjin University of Science and Technology, Tianjin 300222, China<sup>2</sup> Tianjin International Joint Research and Development Center of Low-Carbon Green Process Equipment, Tianjin 300222, China<sup>3</sup> College of Economics and Management, Tianjin University of Science and Technology, Tianjin 300457, China

\* Correspondence: weitian@tust.edu.cn; Tel.: +86-(022)-6060-0705

**Abstract:** Machine learning techniques are widely applied in the field of building energy analysis to provide accurate energy models. The majority of previous studies, however, apply single-output machine learning algorithms to predict building energy use. Single-output models are unable to concurrently predict different time scales or various types of energy use. Therefore, this paper investigates the performance of multi-output energy models at three time scales (daily, monthly, and annual) using the Bayesian adaptive spline surface (BASS) and deep neural network (DNN) algorithms. The results indicate that the multi-output models based on the BASS approach combined with the principal component analysis can simultaneously predict accurate energy use at three time scales. The energy predictions also have the same or similar correlation structure as the energy data from the engineering-based EnergyPlus models. Moreover, the results from the multi-time scale BASS models have consistent accumulative features, which means energy use at a larger time scale equals the summation of energy use at a smaller time scale. The multi-output models at various time scales for building energy prediction developed in this research can be used in uncertainty analysis, sensitivity analysis, and calibration of building energy models.

**Keywords:** building energy prediction; machine learning; multi-output model; time scales



**Citation:** Li, G.; Tian, W.; Zhang, H.; Chen, B. Building Energy Models at Different Time Scales Based on Multi-Output Machine Learning. *Buildings* **2022**, *12*, 2109. <https://doi.org/10.3390/buildings12122109>

Academic Editors: Graziano Salvalai, Shi-Jie Cao, Antonio Caggiano, Wei Yu and Christopher Yu-Hang Chao

Received: 6 November 2022

Accepted: 28 November 2022

Published: 1 December 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/>).

## 1. Introduction

In 2020, the construction and operation of buildings already account for 36% of global energy use and 37% of global CO<sub>2</sub> emissions [1]. The energy use of the construction industry has a significant impact on the environment and the economy. Hence, it is important to reduce the energy use of buildings reasonably and efficiently [2,3]. The key to building energy conservation is to accurately compute building energy use, so as to formulate a reasonable energy conservation plan according to the characteristics of energy use [4,5].

Machine learning methods have a wide range of applications in the field of building energy prediction [6,7]. Most researchers use single-output models to predict building energy use in which the different energy outputs are predicted by the separately individual models. According to time scales, the research in daily, monthly, and annual time periods would be discussed in the following three paragraphs, respectively [8].

Firstly, the application of a single-output machine learning model to predict the daily energy of buildings is introduced. Liu et al. [9] used the support vector machine (SVM) method to predict the daily cooling energy use. The average relative error of the model was 5.03%, indicating that the method had excellent prediction accuracy, and the authors proposed an energy use diagnosis method based on energy use prediction. Alobaidi et al. [10] proposed an ensemble artificial neural network (ANN) method with a resampling technique to predict daily building energy use, which has high generalization ability and robustness. Hong et al. [11] applied seven machine learning methods including light gradient boosting machine (lightGBM) and random forest (RF) to predict daily electricity

consumption. The results show that the coefficient of variation of the root mean square error (CV(RMSE)) of the seven models is less than 10%, in which the lightGBM model has the best performance with the minimum CV(RMSE) of 4.1%. Jetcheva et al. [12] used the neural network-based ensemble model method for daily power consumption prediction, and compared it with the seasonal autoregressive integrated moving average (SARIMA) model; the results indicate that the neural network model had better accuracy. Ferrantelli et al. [13] defined an analytical bottom-up model for predicting any daily consumption given a benchmark daily profile by comparing regression curves obtained with a frequentist inference to Bayesian inference.

Secondly, the application of a single-output machine learning model in predicting the monthly energy use of a building is described. Wang et al. [14] proposed a method by combining a network model with a long-short-term memory learning model to predict monthly building energy use. Compared with traditional ANN and SVM, the performance of this model was better; the mean absolute error and root mean square error were 6.66% and 0.36 kWh/m, respectively. Tran et al. [15] proposed an ensemble model named the evolutionary neural machine inference model (ENMIM), which consisted of least squares support vector regression (LSSVR) and the radial basis function neural network (RBFNN). The model was used to predict the monthly cooling/heating loads and achieved excellent results. Tian et al. [16] adopted a full linear model to predict monthly power and heating energy use respectively, and obtained excellent results, providing a reliable statistical energy model for model calibration. Zhu et al. [17] used five machine learning models including multivariate adaptive regression splines (MARS) to predict the monthly power consumption of buildings, which provided a basis for estimating unknown parameters based on the approximate Bayesian calibration. Koschwitz et al. [18] proposed a novel recurrent neural network method to predict the monthly thermal loads of buildings, and the results show that the method is more accurate than support vector machines. Jahani et al. [19] used the Genetic algorithm-based numerical moment matching (GA-NMM) method to predict the monthly electricity consumption of buildings. Lin et al. [20] used RF, SVM, and ANN to predict the monthly electricity consumption of buildings.

Thirdly, the application of a single-output machine learning model to predict the annual energy use of buildings is presented. Olu-Ajayi et al. [21] used ANN, gradient boosting (GB), deep neural network (DNN), RF, stacking, K-nearest neighbor (KNN), SVM, decision tree (DT), and linear regression (LR) methods to predict the annual energy use of buildings, and the DNN model has the highest prediction accuracy. Tian et al. [22] used 10 machine learning methods to predict the annual energy use of buildings in London, and conducted variable importance analysis of the learning model and local spatial analysis of model differences. Tian et al. [23] used five machine learning methods to accurately predict the annual energy use of buildings for Dempster-Shafer theory uncertainty analysis and global sensitivity analysis. Building annual energy use forecasts can also use a combination of machine learning methods and statistical methods to further explore the linear or nonlinear relationship between annual energy use and certain factors, including building geometry, occupant, and climate-related factors [24]. A novel multiple linear regression model can predict building energy use based on changes in climate parameters and the intensity of the urban heat island effect, which leads to the conclusion that the urban heat island effect can reduce building energy use in heating-dominated cities [25]. In summary, energy use at three different time scales (day, month, and year) can be accurately predicted using different single-output machine learning methods.

There are very limited studies to explore the multi-output models of building energy and most previous studies use single-output machine learning models. Luo et al. [26] used heating, cooling, lighting load, and building integrated photovoltaic (BIPV) electrical power as the output for prediction, and the multi-output model selected SVM and ANN. The results showed that the ANN model obtained higher accuracy and the SVM model took less computation time. Liu et al. [27] proposed a multi-input multi-output (MIMO) strategy based on recurrent neural networks for hourly building energy use prediction. The

absolute percentage error of each time step was analyzed to explore the building's energy-saving performance, thus establishing an energy quantification system for the building. Li et al. [28] developed a MIMO strategy by returning data information in a single-time step, which avoids accumulated errors compared to recursive strategies and direct strategies.

These previous studies provide useful information on the characteristics of multi-output models in building energy analysis. However, there is still a lack of sufficient studies to explore the predictive performance of multi-output machine learning models in building energy analysis, especially focused on output correlation. Moreover, the accumulative features are not explored yet in creating multi-output models at various time scales (such as daily, monthly, and annual).

In order to address these research gaps, this paper is focused on the construction and performance analysis of multi-output machine learning models for building energy assessment. Two multi-output machine learning methods are Bayesian adaptive spline surface (BASS) and deep neural network (DNN). The performance of multi-output models is evaluated from three aspects: computational time, model accuracy, and output correlation. The main new contributions of this study are:

- (1) This study compares the predictive performance of single-output and multi-output learning models in building energy analysis. This would provide guidelines on how to choose the single-output and multi-output models in creating machine learning models for building energy assessment.
- (2) This study explores the performance of two multi-output models (BASS and DNN) in which the main difference of the two models is whether to maintain output correlation. This would provide the guidelines on how to choose the learning models with or without considering output correlation.
- (3) The additive or accumulative features are investigated in creating various time scale models for building energy analysis. This would provide insight on the methods of obtaining building energy use from a smaller time scale to a larger time scale.

The remaining part of this paper is organized as follows. Section 2 introduces the research methods, including data acquisition methods, modeling techniques, and performance evaluation metrics. In Section 3, the final tuning hyper-parameters from the machine learning models are first presented. Then, the results and discussion of the multi-output machine learning model for predicting cooling and heating energy are presented in turn.

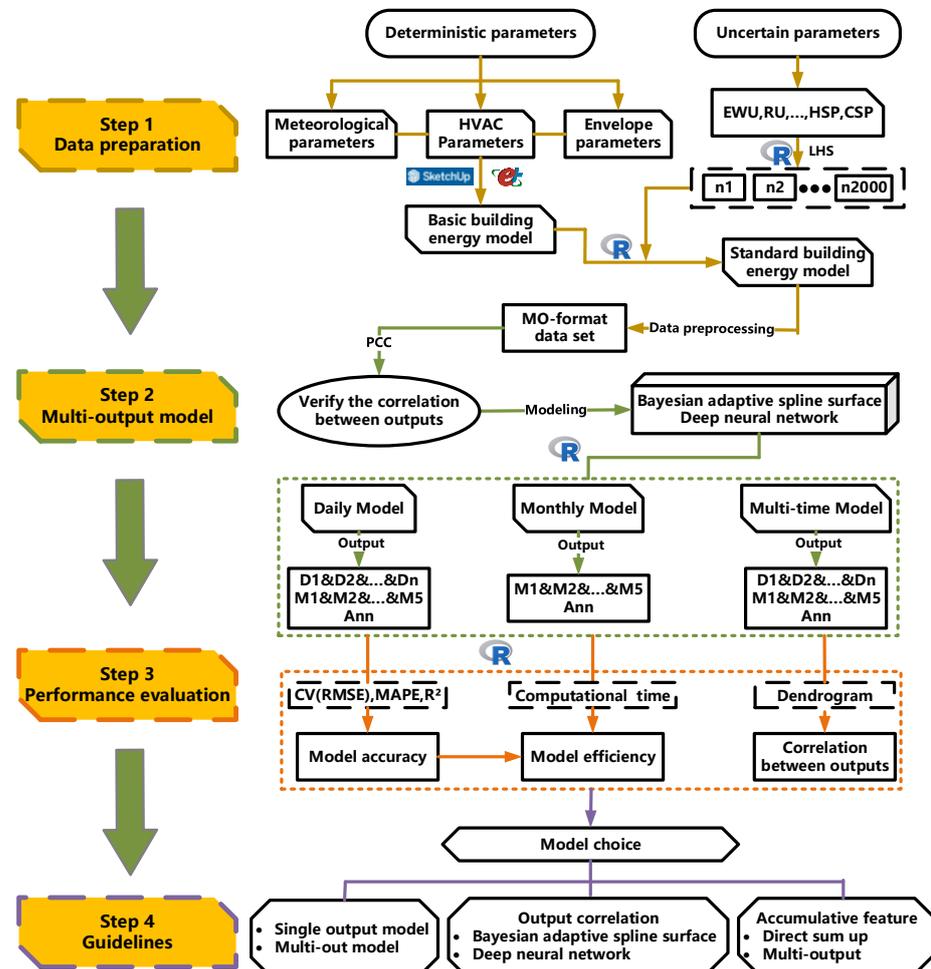
## 2. Method

The research framework based on multi-output machine learning models at different time scales is shown in Figure 1. The analysis procedure can be divided into four steps: data preparation, multi-output models, model performance evaluation, and guidelines. The first step is to prepare the data required to establish the multi-output models as described in Section 2.1. The second step presents the construction of the two multi-output models—Bayesian adaptive spline surfaces (BASS) and deep neural networks (DNN). The third step is model performance evaluation. The fourth step is to provide guidelines for applying the multi-output models in building energy analysis. All the above steps are implemented in R language.

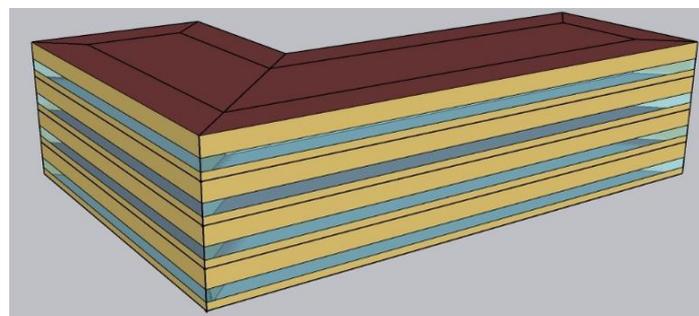
### 2.1. Data Preparation

This section describes the process of acquiring the data required to build machine learning models. Firstly, an L-shaped four-story office building as shown in Figure 2 is created as the basic building energy model. The choice of this building type is due to its simplicity and generalizability compared to other building types (residential or commercial buildings). Each floor of this model has six peripheral zones and two core zones. The building model has a total of 32 thermal zones with a floor area of 1600 m<sup>2</sup>. The fan coil unit is set to provide ventilation, heating, and cooling for the building. The air-cooled chiller and gas boiler provide cold water and hot water, respectively [29]. The internal heat gain schedule of the building is derived from the China national standards of buildings [30].

The building used in this research is located in Tianjin, China. Therefore, the meteorological data of Tianjin in the Chinese standard weather data (CSWD) is used for the calculation of the building energy model.



**Figure 1.** A flow chart of multi-output models of building energy at daily, monthly, and annual scales based on machine learning.



**Figure 2.** A four-story office building.

A basic building energy model is constructed based on deterministic parameters, and then uncertain parameters using the Latin Hypercube sampling (LHS) method are imported into the basic model to obtain a complete standard building energy model. The ranges of uncertain parameters are shown in Table 1, which are derived from previous studies [16,17] and Chinese national standards [30,31]. The building envelope parameters include the U-values of external walls, windows, and roofs. Internal building heat gains

include light power density and equipment power density. Heating and cooling set-point temperatures are also considered in this study. The construction of the modeling data set is divided into two steps. The first step is to perform Latin hypercube sampling within the parameter ranges to obtain 2000 sets of parameter combinations and then generate 2000 idf (EnergyPlus Input Files) format files. The 2000 sets of energy use data at different time scales are obtained by running the EnergyPlus (V22.1.0) program, in which 1000 sets of data are used for model training, 500 sets are used for model validation, and 500 sets are used for model testing. The second step is data preprocessing to create the daily, monthly, and annual energy models. The cooling time periods for this office building is from May 1st to September 30th, and the heating time period is from November 1st to March 31st. The heating and cooling energy on weekends and holidays in this office building is zero and thus excluded from this analysis. As a result, the numbers of daily and monthly cooling energy data are 105 and 5, respectively. The numbers of daily and monthly heating energy data are 102 and 5, respectively.

**Table 1.** Variation of building input parameters.

No.	Parameters	Short Names	Range	Unit
1	Exterior wall U-value	EWU	0.1–0.25	W/(m <sup>2</sup> K)
2	Roof U-value	RU	0.15–0.3	W/(m <sup>2</sup> K)
3	Window U-value	WU	1–2.4	W/(m <sup>2</sup> K)
4	Solar heat gain coefficient	SHGC	0.2–0.48	-
5	Infiltration rate	INF	0.5–0.8	ACH
6	Lighting power density	LPD	5–10	W/m <sup>2</sup>
7	Equipment power density	EPD	9–15	W/m <sup>2</sup>
8	Occupancy density	OPD	9–14	m <sup>2</sup> /person
9	Heating set-point	HSP	20–22	°C
10	Cooling set-point	CSP	24–26	°C

## 2.2. Multi-Output Models

This section describes the construction process of multi-output machine learning models. The Pearson correlation coefficient (PCC) is used to assess the correlation between the output data before the model is established. The computation of the Pearson correlation coefficient is as follows:

$$r = \frac{\sum_{i=1}^n (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^n (a_i - \bar{a})^2} \sqrt{\sum_{i=1}^n (b_i - \bar{b})^2}} \quad (1)$$

where  $r$  represents the Pearson correlation coefficient,  $a_i$  represents the  $i$ th sample  $a$ ,  $b_i$  represents the  $i$ th sample  $b$ ,  $\bar{a}$  represents the average of sample  $a$ , and  $\bar{b}$  represents the average of sample  $b$ . The PCC value is between  $-1$  and  $1$ . The PCC is close to  $1$  to indicate that there is a strong positive correlation. In contrast, a strong negative correlation can be suggested by the PCC value close to  $-1$ .

Two types of machine learning methods—BASS and DNN—are chosen based on these capability of multi-output models.

The BASS method is similar to Bayesian multivariate adaptive regression splines used to learn a set of basis functions from historical data, which are tensor products of polynomial splines. The BASS model can adaptively select the number of basis functions and adjust the number of variables and nodes in each basis. While most hyper-parameters of the BASS model are determined automatically, the priors in the BASS model can be modified to reduce overfitting. The main adjusted parameters include the number of principal component analysis  $n.pc$  and the number of iterations  $nmcmc$ . The principal component analysis (PCA) method is used to calculate the principal components of multiple outputs and then fit a regression model. The R-BASS package (V1.2.2) is used to implement

the multi-output machine learning technique in this research. The advantages of using this BASS package include scalable capabilities to handle large numbers of observations and predictions. In addition, not only regression models but also classification models can be created using the BASS package [32].

The DNN method is one of the current mainstream methods for analyzing building energy, including an input layer, multiple hidden layers, and an output layer. The deep feedforward neural network is selected in this research, and neuron signals are transmitted forwardly between layers without feedback signal transmission. In other words, each layer can only receive the signal of the previous layer, and then pass it to the next layer. The signal transfer between layers depends on the connection of nonlinear functions. By determining the weight and bias terms of the function, the signal of the previous layer is effectively transmitted to the next layer. It is worth noting that there is no signal transmission between neurons in the same layer of a feedforward neural network [33]. The main hyper-parameters of the feedforward neural network include the number of hidden layers, the number of hidden layer neurons, the loss function, and the number of iterations. The DNN parameter adjustment method generally adopts empirical adjustment. The keras and tensorflow packages in the R environment are used to implement the establishment of DNN models.

### 2.3. Performance Evaluation

The predictive performance of learning models is evaluated using three indicators: CV(RMSE) (coefficient of variation of the root mean square error), MAPE (mean absolute percentage error), and  $R^2$  (coefficient of determination). The calculation formulas for these three indicators are as follows:

$$CV(RMSE) = \frac{\sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}}{\frac{1}{n} \sum_{i=1}^n y_i} \quad (2)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (4)$$

where  $n$  represents the sample size,  $i$  represents the  $i^{th}$  sample,  $y_i$  represents the  $i^{th}$  true value, and  $\hat{y}_i$  represents the  $i^{th}$  predicted value. The CV(RMSE), MAPE, and  $R^2$  values are all dimensionless numbers, which are not affected by the order of magnitude of the data and can more effectively express the accuracy of the model. CV(RMSE) and MAPE reflect the error of the model—the smaller, the better. The value of  $R^2$  ranges from 0 to 1, which reflects the relative error of the model relative to the direct average value. The closer to 1, the better.

The computational time of creating machine learning models is also an important indicator to assess the performance of learning algorithms. The computing time includes the modeling time and the prediction time when using the machine learning techniques in building energy analysis. The computing time is counted using the `system.time` function in the R language environment. Simulations were performed on a workstation with an Intel Xeon CPU (E5-2650 2.3 GHz) and 64 GB RAM.

For the multi-output learning models, it is necessary to investigate whether the machine learning model maintains the correlation structure between the outputs of the energy data from the training set. In this research, the cluster dendrogram method is used for model output correlation analysis [34], which can hierarchically cluster multidimensional data to aggregate highly similar data into one category. Through this dendrogram, the clusters and the number of objects belonging to each cluster can be determined [35]. The dendrogram is plotted by computing the distances between clusters according to the Lance-Williams dissimilarity update formula by obtaining the dissimilarities prior to forming

the new cluster. The complete linkage method, one of hierarchical classification methods that uses the distance between the most distant elements from each cluster, is used for hierarchical clustering [36]. The cluster analysis in this research is implemented using the R package tidyverse and hclust functions.

### 3. Results and Discussion

#### 3.1. Results of Model Hyperparameter Tuning

Table 2 shows the optimal hyper-parameters for machine learning models. The SO-BASS and MO-BASS denote the single-output models and multiple-output models, respectively, based on the BASS technique. The same rules are also applicable to the SO-DNN and MO-DNN models. In the BASS models, the degree represents the degree of interaction between the input parameters. The degree of interaction between the input parameters of daily cooling energy is lower than that of daily heating energy. For the MO-BASS model, the n.pc represents the number of principal components used. The daily cooling energy use can be fitted with fewer principal components than the daily heating energy while still producing good results. Therefore, it is easier to predict daily cooling energy use. The nmcmc represents the number of iterations, and all BASS models can achieve excellent accuracy after 10,000 iterations. Note that the hyper-parameters for cooling and heating energy are the same for month-to-month and multi-time scale energy models, indicating the same difficulty in predicting cooling and heating energy use.

The final hyper-parameters from the DNN models are also listed in Table 2. For the SO-DNN model, there is only one neuron in the output layer. The three layers of the SO-DNN hidden layer are selected in the daily and monthly energy use prediction. The activation function of monthly energy use is different from that of daily energy use. For the MO-DNN model, the output layer neurons should be adjusted according to the number of outputs. For example, if the building has 105 cooling days throughout the year, then the MO-DNN must have 105 output layer neurons. The number of hidden layers selected by the MO-DNN model is four layers for daily energy use and three layers for monthly energy use.

**Table 2.** The optimal tuning parameters from machine learning models.

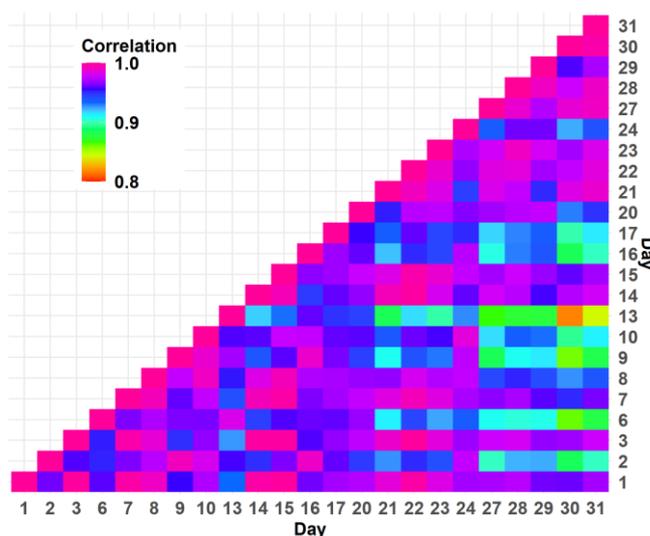
Model	Hyperparameters	Daily		Monthly		Multi-Time	
		Cooling	Heating	Cooling	Heating	Cooling	Heating
SO-BASS	degree nmcmc	3	4		4		-
				10,000			
MO-BASS	n.pc degree nmcmc	7 3	10 4		5 4		7 3
				10,000			
SO-DNN	activation number of hidden layers output layer neurons	tanh,relu,liner			tanh,elu,relu,liner		-
				3			
				1			
MO-DNN	activation number of hidden layers output layer neurons		4		tanh,elu,relu,liner		
					3		4
		105	102		5	111	108

#### 3.2. Results of Multi-Output Cooling Energy Models

Section 3.2.1 presents the results of daily cooling energy models and Section 3.2.2 describes the results of monthly cooling energy models. Section 3.2.3 discusses the results of multi-time scale models that simultaneously predict daily, monthly, and annual cooling energy. Section 3.2.4 presents the results from the models in which monthly and annual energy can be directly or accumulatively obtained.

### 3.2.1. Daily Cooling Energy Models

The correlation coefficient of daily cooling energy in the work days from the training set of size 1000 is shown in Figure 3. The correlation coefficients are all above 0.8, and most of them are above 0.9, as is evident from the color labels in Figure 3. As a result, daily cooling energy use in this office building is strongly positively correlated, hence, it is important to include it in the correlation structure account when creating machine learning models for daily cooling energy use.

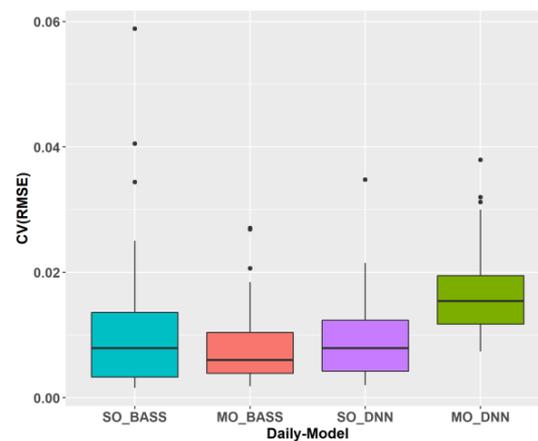


**Figure 3.** Correlation coefficient of daily cooling energy use in July (1, 1 July; . . . ; 31, 31 July).

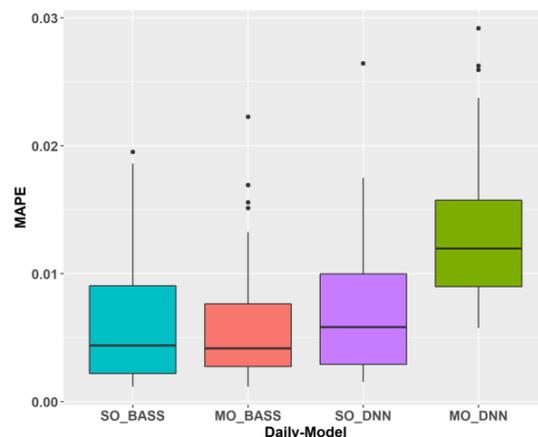
Figure 4 shows the predicted performance of the multi-output daily energy cooling model. Figure 4a demonstrates that the four models of SO-BASS, MO-BASS, SO-DNN, and MO-DNN have median CV(RMSE) of 0.008, 0.006, 0.008, and 0.015, respectively. For the day with the worst prediction accuracy of the four models (SO-BASS, MO-BASS, SO-DNN, and MO-DNN), the largest CV(RMSE) values are 0.059, 0.027, 0.035, and 0.038, respectively. The median and maximum values of the CV(RMSE) from the four models indicate that the MO-BASS model performs the best in this case study. Figure 4b shows that the MO-BASS model has the smallest median MAPE, followed by SO-BASS and SO-DNN. The MO-BASS model has better predictive performance due to the small variations of MAPE in comparison with the SO-BASS although the median MAPE values of the two BASS models are similar. This is because the MO-BASS model considers the output correlation to avoid the extreme errors from the single-output models. Hence, the MO-BASS model has the best performance in terms of the MAPE. Figure 4c indicates that the coefficients of determination ( $R^2$ ) are very high, except for three data points in SO-BASS. Most of the  $R^2$  values are 1 greater than 0.9 to indicate that all these four models have good performance. Among them, the two models with the largest  $R^2$  are MO-BASS and SO-DNN. Further comparison of the median and interquartile range shows that the  $R^2$  from the MO-BASS model is larger than that of SO-DNN. From the above analysis, the MO-BASS model is the best daily cooling energy model in this case study.

Table 3 lists the computational time for single-output and multi-output models of cooling energy for BASS and DNN algorithms at different time scales models. The comparison of computational time for daily cooling energy use will be described in this section. The discussion of computational time for monthly and multi-time scales would be presented in Sections 3.2.2 and 3.2.3, respectively. The computation time for the multi-output models is less than the computation time for the single-output models for daily cooling energy use. For the BASS models, the modeling time of the multi-output BASS model for daily cooling energy is 19.6 times that of the single-output BASS model. This is due to the fact that the output number in the BASS multi-output model would be same as the principal component

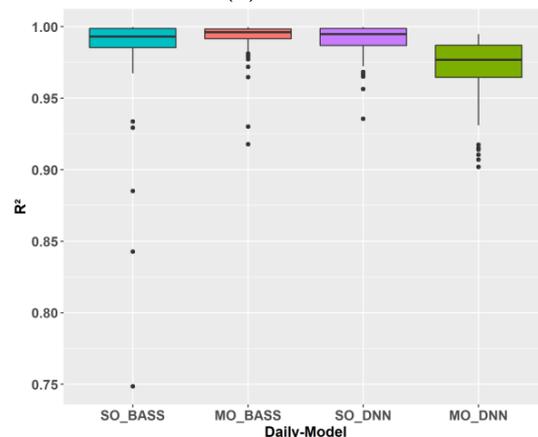
number, which is much less than the original correlated outputs in the single-output models. The modeling time of the multi-output DNN model is 27.7 times that of the single-output DNN model for daily cooling energy. Hence, adopting the multi-output models can significantly reduce modeling time and increase the model's efficiency, presuming the accuracy of the model is acceptable. The DNN models require more computational cost compared with the BASS models. For the single-output models, the computational time for the DNN model is almost two times that of the computational cost for the BASS model.



(a) CV(RMSE)



(b) MAPE

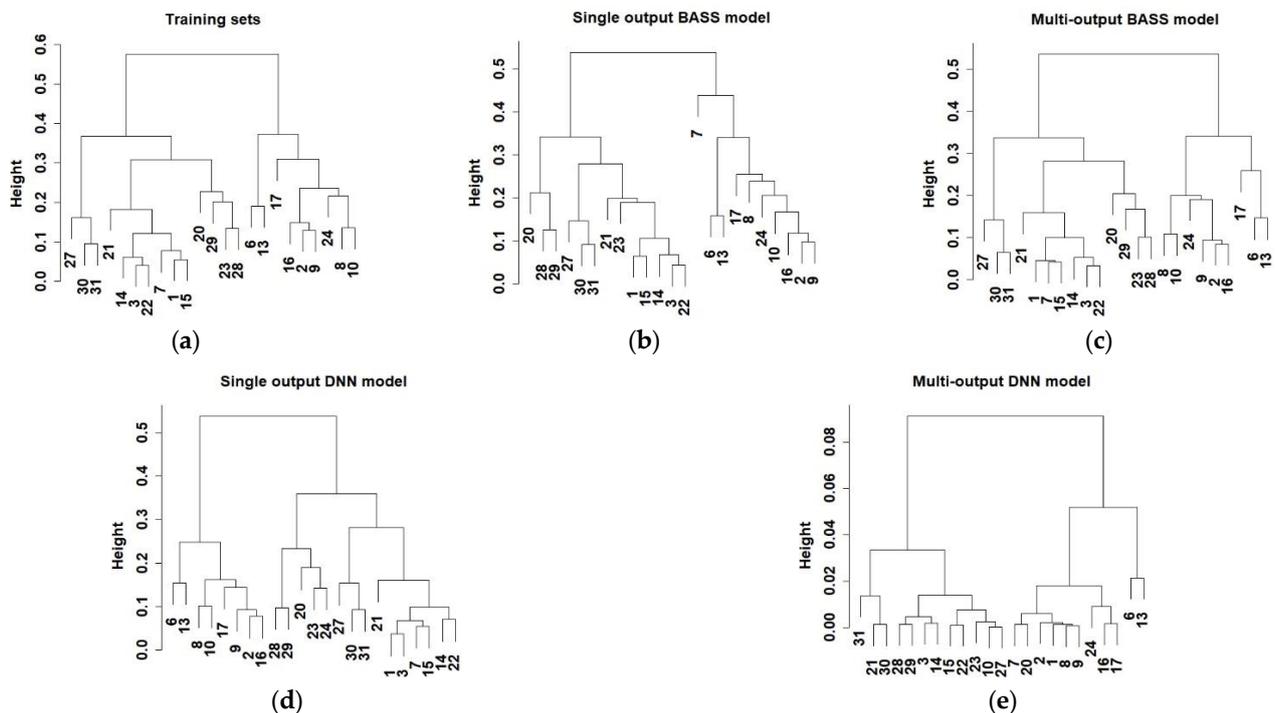
(c) R<sup>2</sup>

**Figure 4.** Prediction performance of machine learning models for daily cooling energy (SO, single output; MO, multiple output).

**Table 3.** The computational time of different time-scale cooling energy models (unit: s).

Models	Daily	Monthly	Multi-Time Scale
SO-BASS	2032.2	98.7	-
MO-BASS	103.9	81.1	197.2
SO-DNN	3807.2	361.7	-
MO-DNN	137.3	72.2	236.3

Figure 5 shows the correlated structure tree diagram from the training set and the four machine learning models on the daily cooling energy in July. The numbers in Figure 5 represent a specific day in July in which the weekends and the holidays have been excluded since there is no cooling energy use. For example, 1 represents the first day of July. From the top of Figure 5a, there are two groups of data in which the left tree has 14 data and the right tree has nine data. The group can be further divided into sub-groups based on the similarity of cooling energy use. The cooling energy in the 3rd and 22nd are the most similar as the height of the link that joins them together is the smallest in Figure 5a. By comparing Figure 5a with the other four dendrograms, Figure 5a–c have the most similar clustering distributions. Thus, the MO-BASS model can properly maintain the output correlation of the original training set from the EnergyPlus models. Figure 5b,d,e show that there are significant differences from the training set. Therefore, the MO-BASS model can be more in line with the training data set from the engineering-based building energy models from the perspective of the correlation structure of building cooling energy use.

**Figure 5.** Dendrogram of July daily cooling energy for the training set and four machine learning models.

### 3.2.2. Monthly Cooling Energy Models

The correlation coefficients of monthly cooling energy are shown in Figure 6. All correlation coefficients are above 0.98, indicating that there is a strong correlation between monthly cooling energy. Therefore, it is of great significance to create multi-output machine learning models for monthly cooling energy.

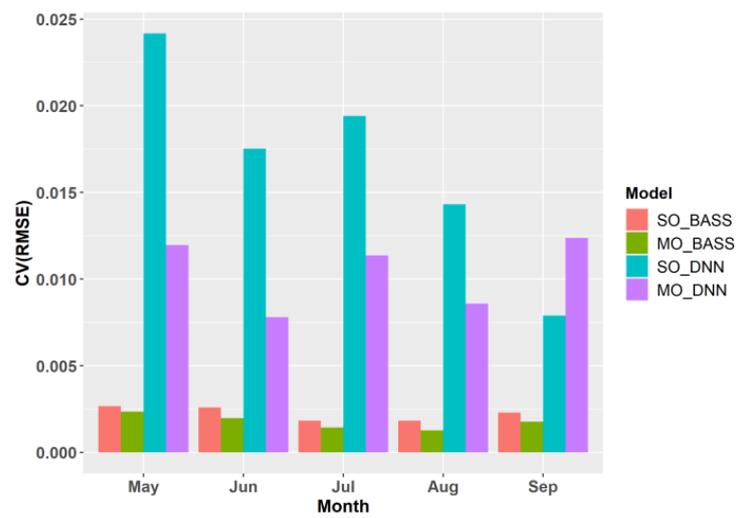


**Figure 6.** Correlation coefficient of monthly cooling energy.

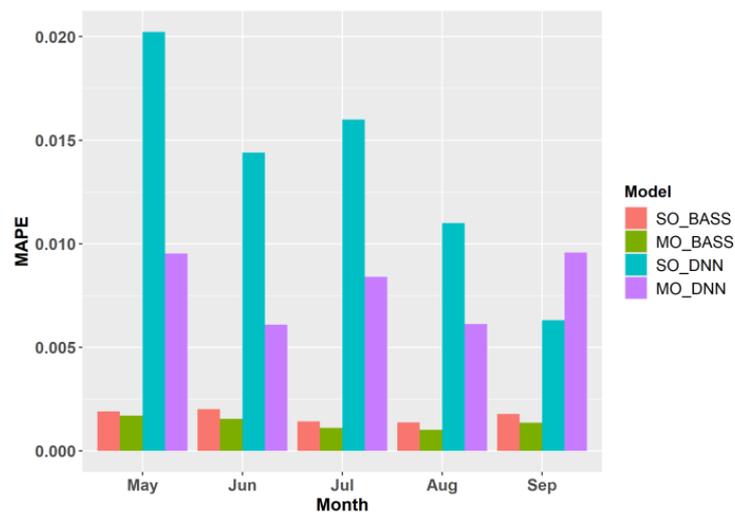
Figure 7 shows the predicted performance of four monthly cooling energy machine learning models. The MO-BASS model has the smallest CV(RMSE), followed by the SO-BASS model. The CV(RMSE) of both BASS models is below 0.005, indicating that these two models have good performance. The CV(RMSE) values of the SO-DNN models are between 0.075 and 0.025, slightly larger than the CV(RMSE) of the two BASS models. From Figure 7b, similar conclusions can be drawn as in Figure 7a. Finally, it can be discovered from Figure 7c that all four models have  $R^2$  values higher than 0.93, with two BASS models having  $R^2$  values close to 1. In summary, the best performance models for monthly cooling energy prediction are the MO-BASS in this case study.

The modeling time of the four monthly cooling energy machine learning models is listed in Table 3. The modeling time of the two single-output models (SO-BASS and SO-DNN) is higher than that of the two multi-output models. For the monthly cooling energy model, the modeling time of the BASS single/multi-output model has no significant differences since the number of monthly cooling energy data is only 5. There is still a marked difference in the modeling time of the two DNN models, and the multi-output model is around five times that of the single-output model. Overall, the multi-output monthly cooling energy models are more time-saving compared to the single-output monthly cooling energy models, especially for the DNN models.

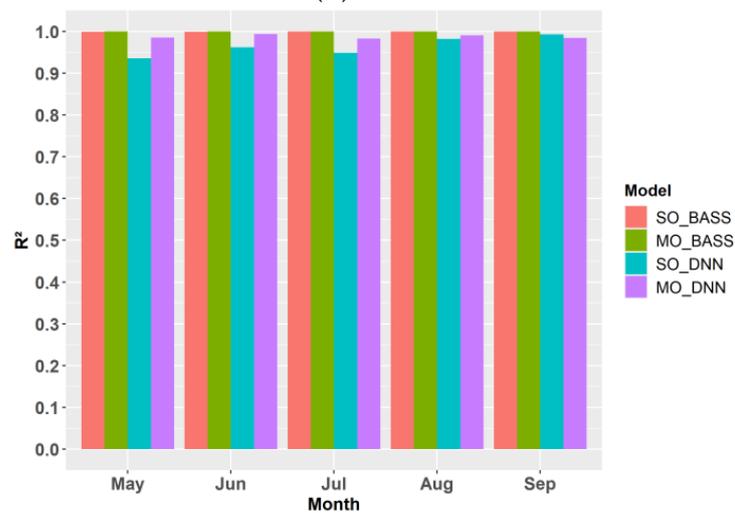
Figure 8 shows the correlation structure tree diagram of five monthly energy use models. Figure 8a shows the dendrogram of engineering-based building energy models, and the adjacent months of July and August have the highest correlation. Hence, the hottest two months have similar patterns of cooling energy from the air-conditioning system in this office building. The two transitional months, May and September, are closely correlated because they have similar climatic conditions including outdoor temperature and solar radiation. The two BASS models for monthly cooling energy (Figure 8b,c) can both retain the same correlation structure as the engineering-based model. In contrast, the two monthly cooling DNN models (Figure 8d,e) are unable to maintain the same correlation as the patterns illustrated in Figure 8a. As a result, the multi-output BASS models can maintain inter-output correlation when predicting both month-by-month and day-by-day cooling energy in this office building as shown in Figures 5 and 8.



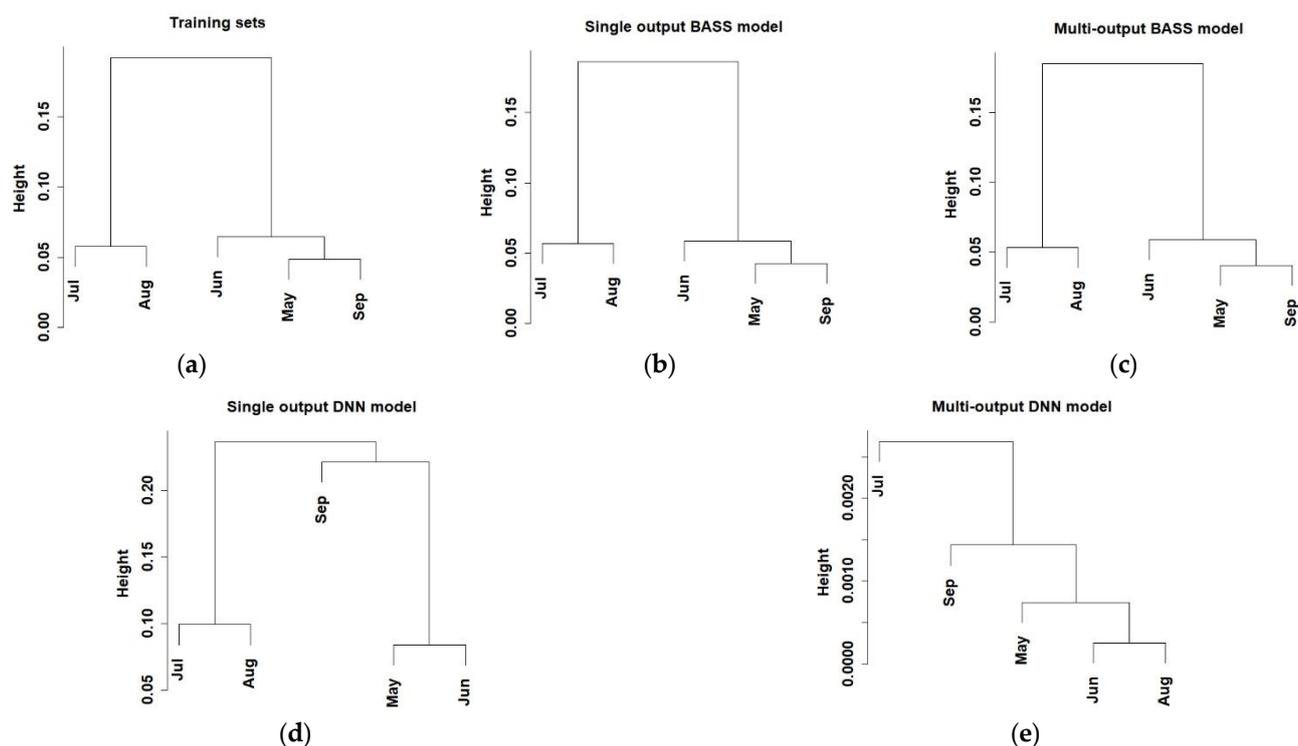
(a) CV(RMSE)



(b) MAPE

(c)  $R^2$ 

**Figure 7.** Prediction performance of four machine learning models for monthly cooling energy (SO, single output; MO, multiple output).



**Figure 8.** Dendrogram of monthly cooling energy for the training set and four machine learning models.

### 3.2.3. Multi-Time Scale Cooling Energy Models

The prediction performance of the multi-time scale cooling energy models is shown in Figure 9. The features of these multi-time scale models can predict the cooling energy at daily, monthly, and annual time scales, simultaneously. First, the performance of daily cooling energy prediction from the multi-time scale models is analyzed. The medians of CV(RMSE), MAPE, and  $R^2$  of MO-BASS and MO-DNN models are 0.007 and 0.02, 0.005 and 0.016, 0.99 and 0.97, respectively. Therefore, the multi-time scale MO-BASS model has a better performance in predicting daily cooling energy. Note that the performance of MO-BASS and MO-DNN models is similar in the case of individual prediction of daily cooling energy as described in Section 3.2.1. Second, the analysis of the multi-time scale model is focused on the monthly cooling energy forecast. The CV(RMSE) and MAPE values from the MO-BASS models are smaller than those from the MO-DNN models, and the  $R^2$  values of MO-BASS are larger than those from the MO-DNN models. Hence, the MO-BASS model still performs better than the MO-DNN model for the monthly cooling energy. Third, the model performance of annual cooling energy is analyzed. The CV(RMSE) and MAPE values from the MO-BASS models are close to 0, and the corresponding  $R^2$  is close to 1. The performance of MO-DNN is worse than that of MO-BASS. It is worth noting that the prediction accuracy of the multi-time scale model increases with the increase in time scale. This is because the data complexity decreases with the increase in time scale. In summary, the MO-BASS model performs better in predicting multiple time scales compared to the MO-DNN model.

Table 3 lists the computational time for multi-time scale models of BASS and DNN. The MO-DNN model requires 236.3 s, while the MO-BASS model needs 197.2 s. The computation time MO-DNN is 20% more than that of the MO-BASS model. As a result, the MO-BASS offers both high accuracy and low modeling costs in comparison with the MO-DNN model for the simultaneous prediction of daily, monthly, and annual cooling energy.

### 3.2.4. Performance Analysis of 10 Models for Monthly and Annual Cooling Energy

This section compares the predictive performance and accumulative characteristics of monthly and annual cooling energy from 10 models as listed in Table 4. The model

names in Table 4 include three parts. The first two letters denote the single-output (SO) or multi-output (MO) model. The third letter denotes the machine learning algorithm, in which the B is the Bayesian adaptive spline surface (BASS) and the D is the deep neural network (DNN). The letters after the hyphen denote the time scale of energy models. The D is the daily model, the M is the monthly model, and the Mu is the multi-time (daily, monthly, and annual) scale model. For example, the MOB-Mu is to simultaneously predict the daily, monthly, and annual energy from the multi-time scale multi-output BASS model. Note that there are two ways to obtain monthly or annual energy: direct and accumulation. For instance, the MOD-M or SOD-M models can directly predict the monthly energy based on the deep neural network models. In contrast, the SOB-D model needs to sum the daily energy from the single-output daily BASS models to obtain the monthly (or annual) energy since there is unavailable for monthly (or annual) energy use from the daily BASS model.

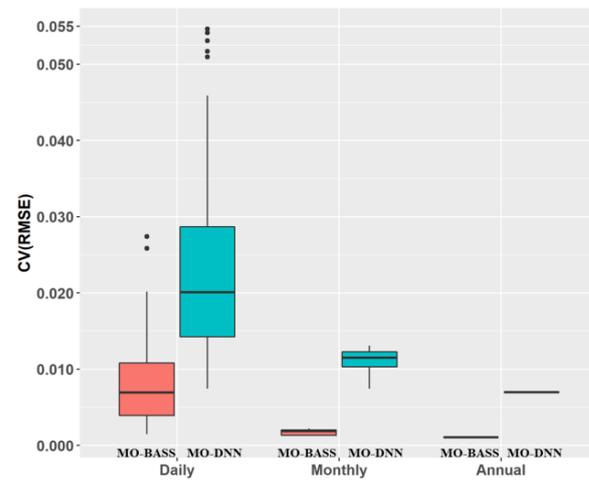
The predictive performance of 10 models for monthly cooling energy is shown in Figure 10. The performance of monthly cooling energy from the direct prediction is better than those from the accumulative models for the BASS approaches in terms of CV(RMSE), MAPE, and  $R^2$ . The SOB-M performs better than the SOB-D and the MOB-M has better predictive capability compared to the MOB-D. The multi-time scale BASS model has very good performance compared to the monthly and daily BASS models for cooling energy prediction. As for the DNN models, the best prediction model of cooling energy is the monthly summation from the daily SOD-D models. The multi-time scale MOD-Mu model has moderate performance compared to the other four DNN models. The BASS models have better prediction performance in comparison with the DNN models for monthly cooling energy. All the  $R^2$  values for the BASS models are above 0.99, which indicates that the BASS has very high predictive capability for monthly cooling energy.

**Table 4.** Description of 10 models for monthly or annual energy prediction.

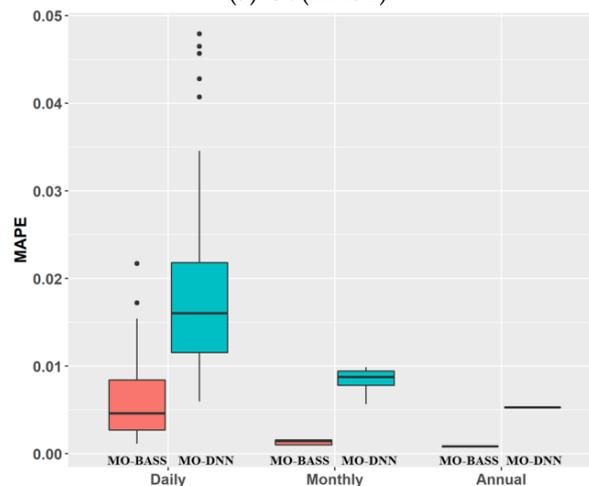
Machine Learning	Model	Description
BASS	SOB-D	Sum the daily energy from the single-output daily BASS models to obtain the monthly or annual energy
	MOB-D	Sum the daily energy from the multi-output daily BASS models to obtain monthly or annual energy
	SOB-M	Monthly predictions or annual prediction (sum of monthly predictions) from the single-output monthly BASS models
	MOB-M	Monthly predictions or annual prediction (sum of monthly predictions) from the multi-output monthly BASS models
	MOB-Mu	Monthly or annual predictions from the multi-output multi-time scale BASS models
DNN	SOD-D	Sum the daily energy from the single-output daily DNN models to obtain the monthly or annual energy
	MOD-D	Sum the daily energy from the multi-output daily DNN models to obtain monthly or annual energy
	SOD-M	Monthly predictions or annual prediction (sum of monthly predictions) from the single-output monthly DNN models
	MOD-M	Monthly predictions or annual prediction (sum of monthly predictions) from the multi-output monthly DNN models
	MOD-Mu	Monthly or annual predictions from the multi-output multi-time scale DNN models

It is also interesting to compare the accumulative features for the different time scale models. As might be expected, the results from the monthly summation of daily cooling energy models are different from the direct prediction of monthly cooling energy models, even for the same machine learning algorithm. The computational results from this case study confirm this statement. For the BASS models, there are five methods to obtain monthly energy use, which may make it confusing regarding which monthly results should be selected in the multiple time scale analysis. For instance, if the best daily model is from the daily BASS single-output models and the best monthly model is from the monthly BASS single-output model in terms of prediction accuracy, then the monthly energy use

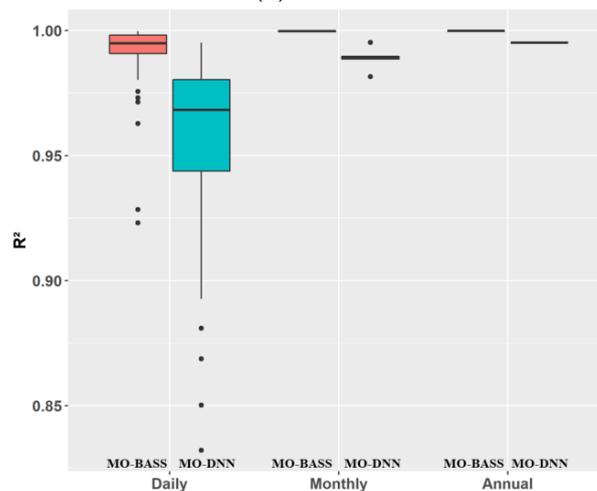
would not be the same as the summation of daily energy use in a specific month. The only model that can maintain good accumulative features is the multi-time scale BASS model (MOB-Mu), due to the processing of the principal component analysis. The monthly cooling energy in July would equal the summation of the July daily prediction of cooling energy for the MOB-Mu model.



(a) CV(RMSE)



(b) MAPE

(c)  $R^2$ 

**Figure 9.** Prediction performance of two machine learning models for multi-time scales cooling energy.

Figure 11 demonstrates the predictive performance of 10 models for the annual cooling energy use. The meaning of these 10 models is described in Table 4. The three best models are MOB-Mu, MOB-M, and MOB-D, which all belong to the multi-output BASS models. The next three models are SOB-M, SOB-D, and SOD-D, which all belong to the single-output models. All these six models have very high  $R^2$  values, above 0.997. The worst model is the single-out monthly DNN models (SOD-M), in which the CV(RMSE), MAPE, and  $R^2$  values are 0.016, 0.013, and 0.970, respectively. The remaining nine models' CV(RMSE) values are less than 0.008, MAPE values less than 0.0078, and  $R^2$  values greater than 0.995, indicating that these models have good predictive performance.

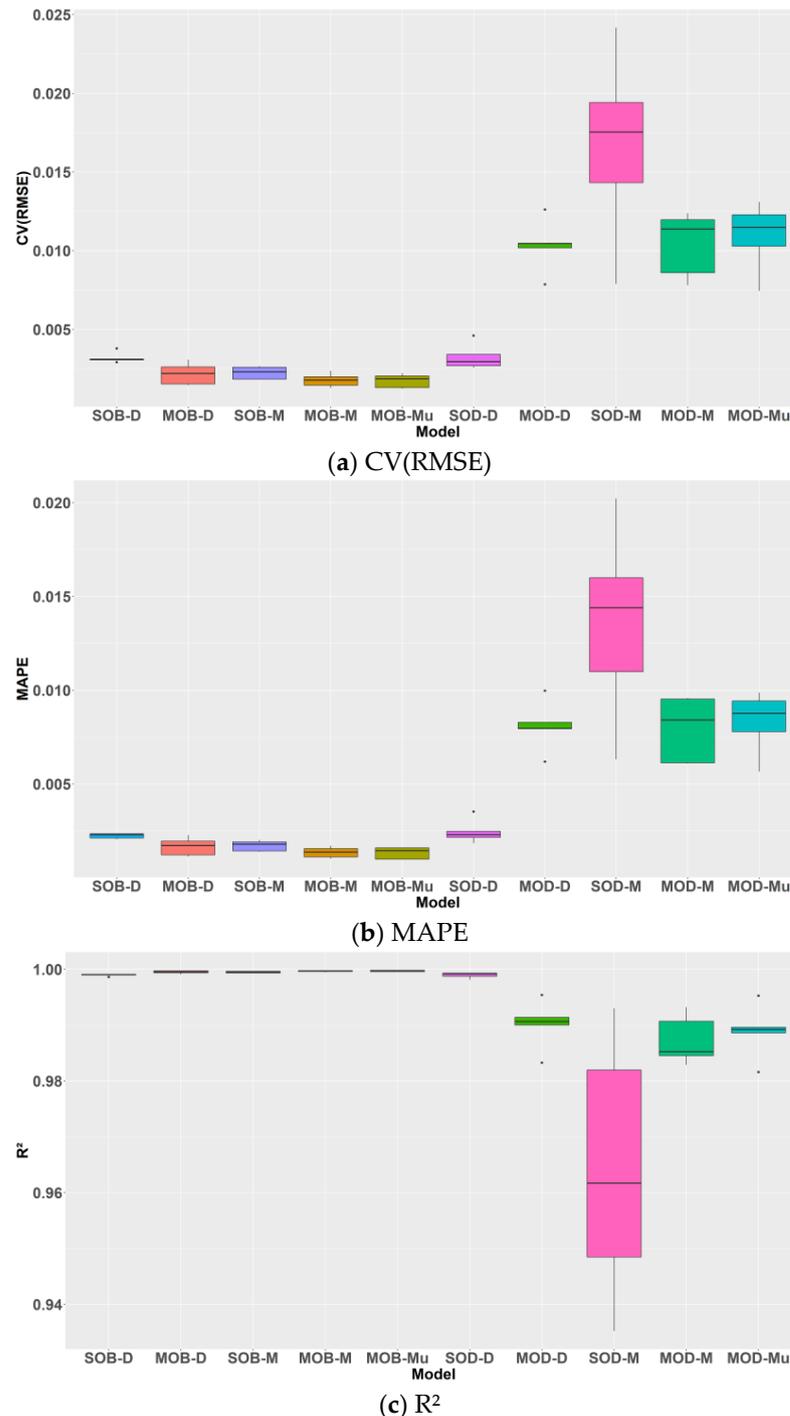
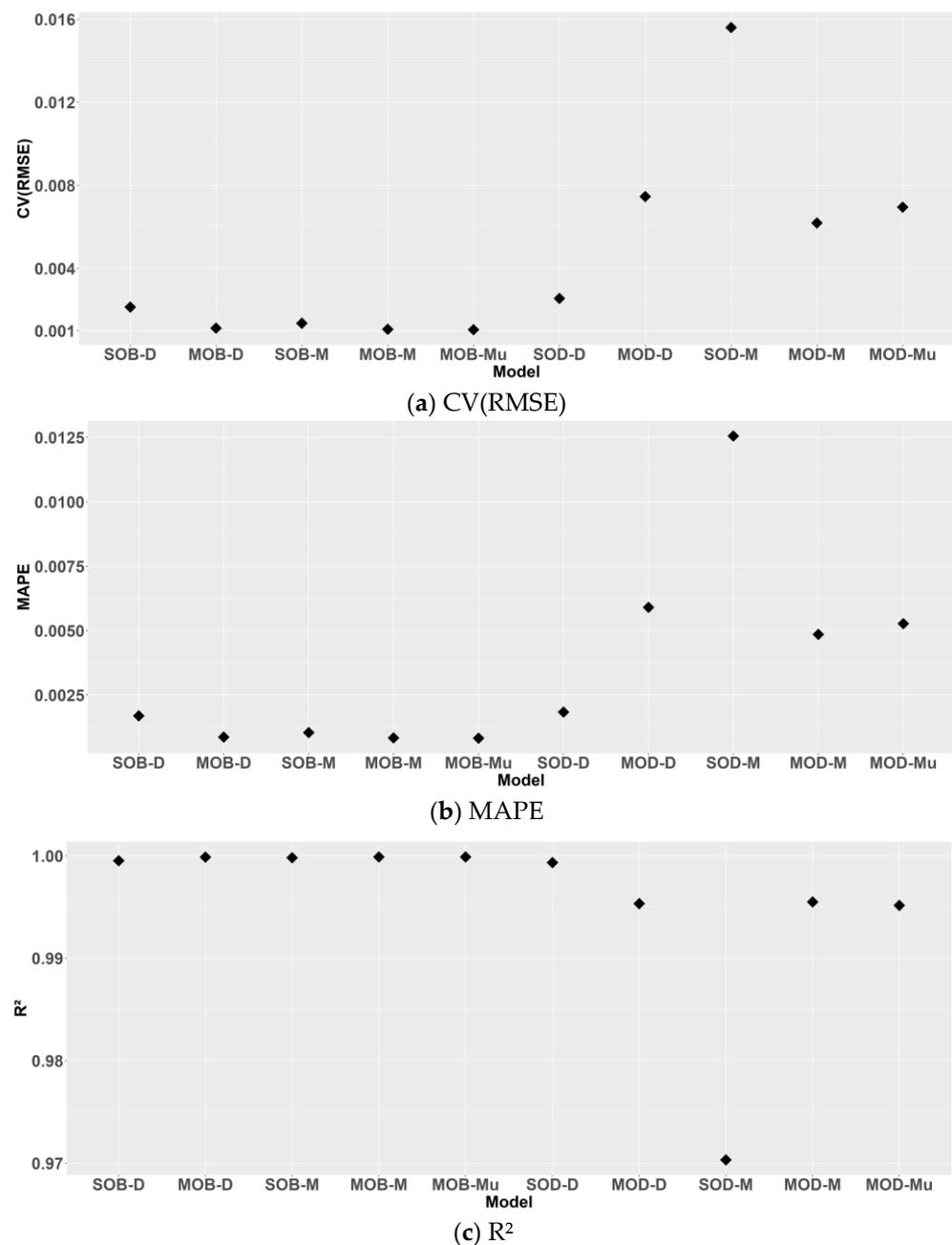


Figure 10. Performance of 10 models for monthly cooling energy (refer to Table 4 for model explanation).



**Figure 11.** Performance of 10 models for annual cooling energy (refer to Table 4 for model explanation).

For the sake of accumulative features, the energy use prediction at a larger scale can be obtained from the summation of small-scale energy use for the single-output or multi-output models. However, this simple accumulation is not necessarily a good method in predicting the multi-time scale energy use in buildings. There are at least three reasons for this. The first reason is that the residual errors from the small scale models may lead to uncontrolled errors. If all the residual errors from the small scale models are positive or negative, the final residuals at a large scale can be added to larger residual errors. If there are negative and positive errors for the small time scale models, the residual errors at a larger scale may be small due to the offsetting of negative and positive values. However, the fundamental logic for the small scale energy models may be unreasonable in the first place, which likely leads to unexpected errors in estimating energy use at a larger scale. The second reason is that the model accuracy would be easily improved at a larger time scale. This is because the efforts of creating machine learning models at a larger time scale would

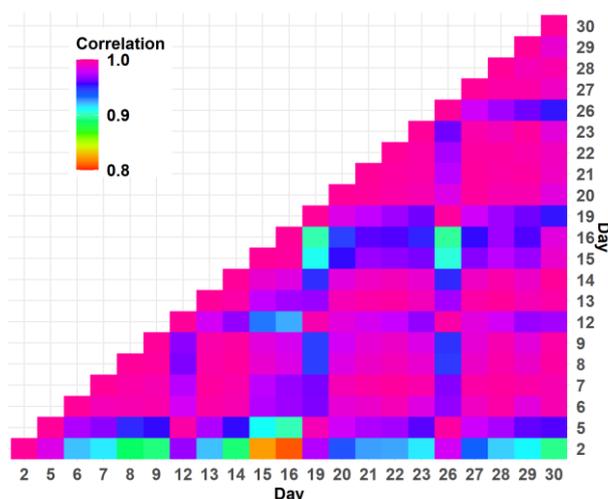
be much less compared to the small time scale models due to the decrease in model number. The third reason is the computational cost. The increase in computational cost by adding large time scale models is not significant since there are much fewer models with an increase in time scale. Moreover, the multi-time scale multi-output machine learning models that can maintain the correlation structure of energy data show very good performance, as discussed in this section.

### 3.3. Results of Multi-Output Heating Energy Models

Section 3.3.1 discusses the results of daily heating energy models. Section 3.3.2 presents the results of monthly heating energy models. Section 3.3.3 discusses the results of multi-time scale models to simultaneously predict daily, monthly, and annual heating energy. Section 3.2.4 presents the results from the models in which monthly and annual energy can be obtained by using the direct simulation of energy models or the summation of energy use at a smaller time scale.

#### 3.3.1. Daily Heating Energy Models

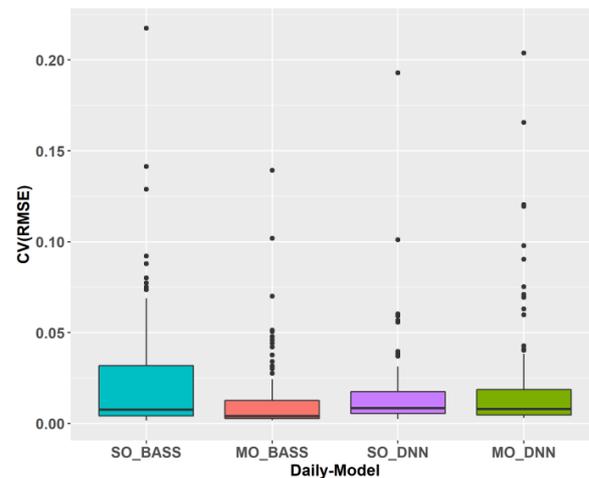
The correlation coefficient between the daily heating energy in January is shown in Figure 12. The daily heating energy is selected in January as a representative cold month for analysis, which shows the correlation coefficients between the energy use of all working days in January. The correlation coefficient between 0.8–0.85 is orange and brown, which only exists between the 15th and 16th of January. The color labels for seven correlation coefficients are in green, representing the correlation coefficients between 0.85–0.9. The correlation coefficients between all the remaining days are above 0.9. Therefore, there is a strong correlation between daily heating energy.



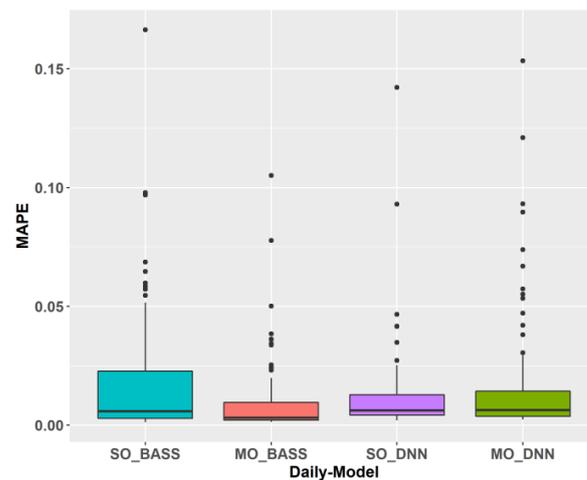
**Figure 12.** Correlation coefficient of daily heating energy in January (2, 2 January; . . . ; 30, 30 January).

Figure 13 shows the predictive performance for daily heating energy from the single-output (SO) and multi-output (MO) models based on the BASS and DNN approaches, respectively. Figure 13a indicates that the median CV(RMSE) values of the boxplots for the four models are not significantly different—all below 0.02. Hence, the four models all have good performance regarding daily heating energy prediction. However, there are outliers for these models as illustrated in Figure 13a. The three models (SO-BASS, SO-DNN, and MO-DNN) have the CV(RMSE) values greater than 0.15, and two points are even greater than 0.2. This shows that the model performs poorly on certain days of heating energy prediction. In contrast, the CV(RMSE) values of the MO-BASS models are all less than 0.15, indicating good prediction capability. Moreover, the interquartile range from the MO-BASS model is the smallest, indicating that the MO-BASS model is the most stable in terms of model errors. The trends of MAPE values shown in Figure 13b are similar to those shown

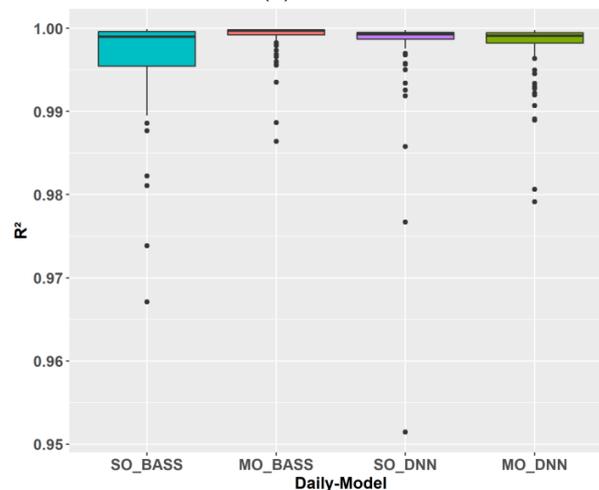
in Figure 13a, which verifies the accuracy results from the CV(RMSE). The  $R^2$  values shown in Figure 13c are above 0.95 for the four models, indicating that the model performance is high. Among them, the  $R^2$  values of the MO-BASS model are greater than 0.985. Therefore, the MO-BASS model is the most accurate daily heating energy model in this case study.



(a) CV(RMSE)



(b) MAPE

(c)  $R^2$ 

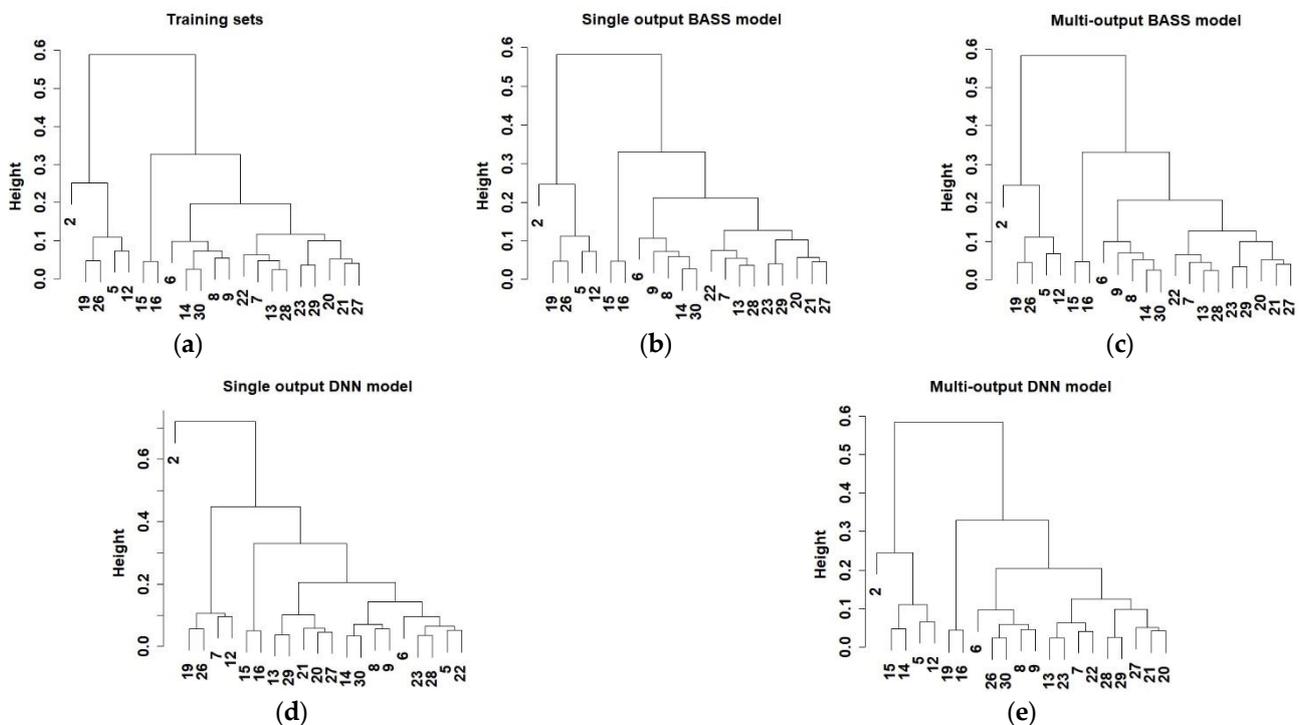
**Figure 13.** Prediction performance of four machine learning models for daily heating energy (SO, single output; MO, multiple output).

The modeling time of the four daily heating energy models is shown in Table 5. The most time-saving model is the MO-DNN model, which requires only 69.7 s. The most time-consuming model is the SO-DNN, which spends 3722.6 s (more than 1 h)—around 53 times that of the SO-DNN model. The MO-BASS model is also significantly more time efficient compared to the SO-BASS model. Therefore, the multi-output models save computational time compared to the single-output models.

**Table 5.** The computational time of different time-scale heating energy models (unit: s).

Model	Daily	Monthly	Multi-Time Scale
SO-BASS	2293.0	128.8	-
MO-BASS	216.6	101.6	150.4
SO-DNN	3722.6	373.6	-
MO-DNN	69.7	65.4	165.9

The dendrograms for daily heating energy from the training set and the four models are shown in Figure 14. The numbers in Figure 14 represent dates in a specific month, such as 2 for 2 January. Figure 14a reflects the dendrogram of the correlation between the outputs from the engineering-based EnergyPlus energy models. It can be seen that 2, 5, 12, 19, and 26 belong to one subgroup, and these days are the first working days after the holidays or weekends. The reason for the aggregation is that the office building is not heated during non-working days, resulting in a low temperature inside the office building. The heating system starts to work on the first working day, which consumes more energy than the following days. Figure 14d,e have significantly different correlation structures from Figure 14a since the non-working days do not belong to one subgroup. In contrast, the correlation structures from Figure 14b,c are similar to that from Figure 14a.



**Figure 14.** Dendrogram of January daily heating energy for the training set and four machine learning models.

### 3.3.2. Monthly Heating Energy Models

The correlation coefficient of monthly heating energy is shown in Figure 15. The smallest correlation coefficient is 0.969 between November and March, and the largest correlation coefficient is 0.999 between January and December. As a result, there is a significant correlation between the five months of heating energy data. Therefore, it is of great significance to construct heating energy models by considering the correlation structure of energy use.

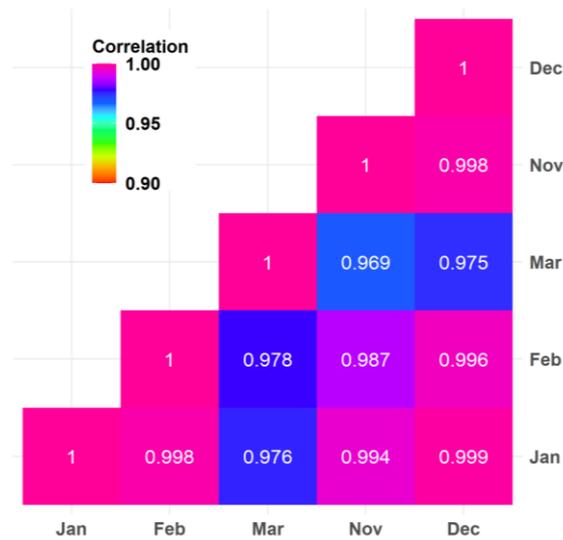
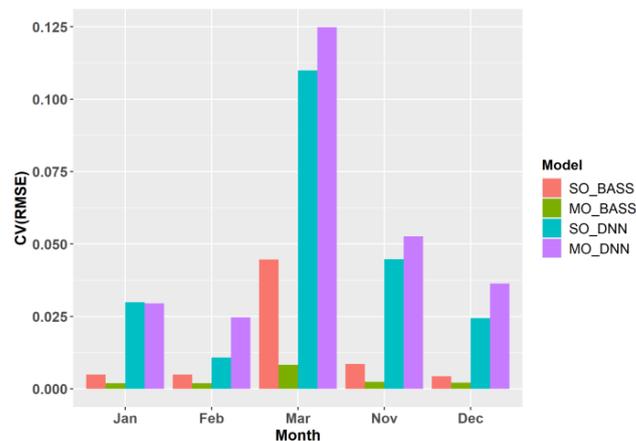


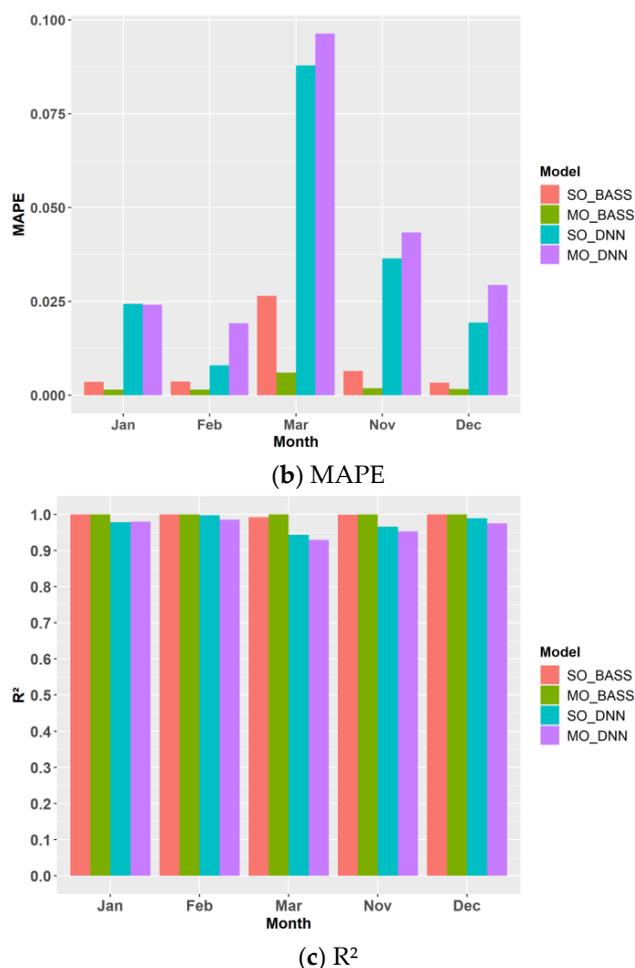
Figure 15. Correlation coefficient of monthly heating energy.

Figure 16 shows the performance indicators for the monthly heating energy models. The MO-BASS model has the smallest CV(RMSE) and MAPE with the largest  $R^2$ , indicating that the MO-BASS outperforms the other three models. The other three models also have CV(RMSE) less than 0.13, MAPE less than 0.1, and  $R^2$  greater than 0.92, indicating that they also have good predictive performance. Note that the month with the worst prediction performance for each model is March, followed by November. What these two months have in common is that they are both transitional months. Climatic conditions such as temperature can vary considerably during the transition months, which can add to the complexity of the heating energy and render it more difficult to predict the heating energy. Prediction accuracy is improved significantly when there are stable changes in climate conditions such as outdoor temperatures, for example in February and December.



(a) CV(RMSE)

Figure 16. Cont.



**Figure 16.** Prediction performance of four machine learning models for monthly heating energy.

The modeling time of the four monthly heating energy models is listed in Table 5. The difference between the two BASS models is similar since there are only five months for heating energy. The computing time of the SO-DNN model is about 5.7 times that of the MO-DNN model. In general, the modeling of the multi-output models is faster compared to the single-output models.

The dendrograms of monthly heating energy for the training set and the four machine learning models are shown in Figure 17. The adjacent months of January and December are grouped together, as is shown in Figure 17a, due to their similar weather conditions. The two transition months, November and March, are classified into one category, indicating that there is a large gap between the weather conditions and the rest of the months. The single-output and multi-output BASS models can maintain the same correlation with the EnergyPlus models, while the single-output and multi-output DNN models are different from the EnergyPlus models. As shown in Figure 16, these correlation structures are helpful to improve the prediction capability of learning models.

### 3.3.3. Multi-Time Scale Heating Energy Models

Figure 18 shows the prediction performance of the heating energy model on multiple time scales. The model accuracy increases as the time scale increases. This is directly related to the complexity of the energy data in different time scales. Figure 18a demonstrates that the median CV(RMSE) values of the daily heating energy of the two models are less than 0.05, although there are a few outliers. In all three time scales, the BASS models perform better than the DNN models in terms of the CV(RMSE). The same conclusions can be obtained from Figure 18b in terms of the MAPE. Figure 18c shows that there are a few

points in  $R^2$  between 0.875 and 0.95 for the DNN daily models, indicating that the model performance becomes worse on some days. Note that the interquartile ranges of the BASS model are smaller than that of the DNN model, indicating that the MO-BASS model is more stable. Therefore, the multi-time scale BASS model has better predictive capability in comparison with the DNN model.

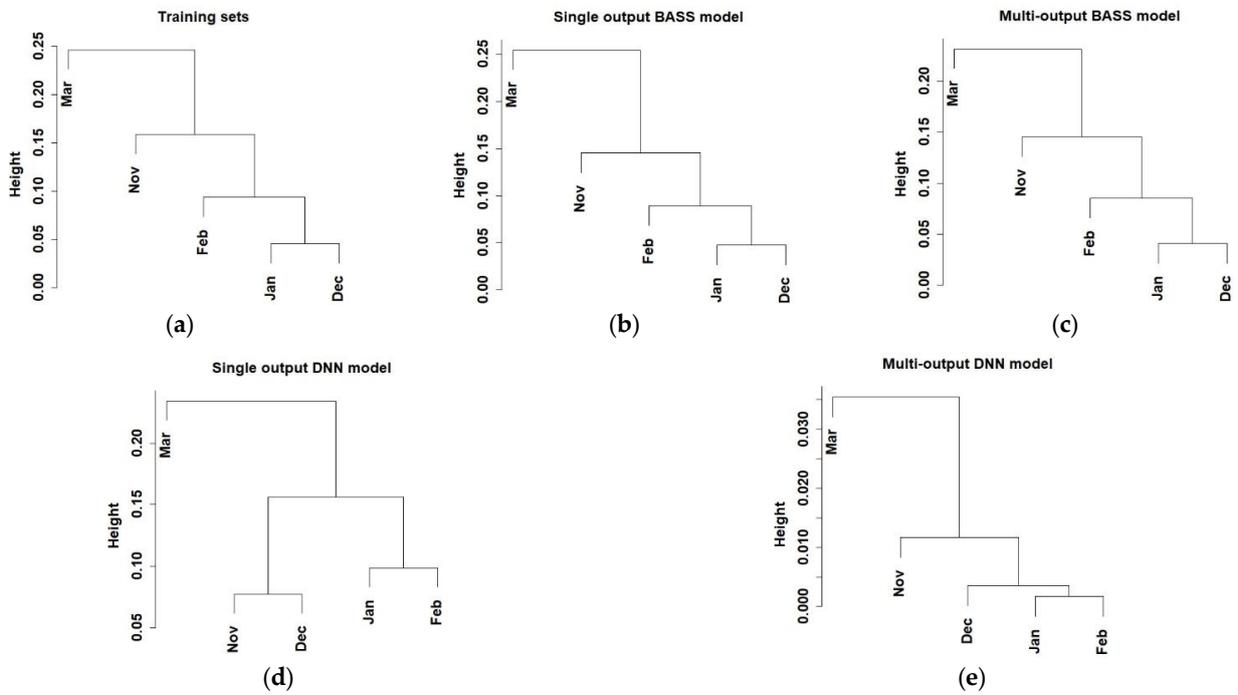


Figure 17. Dendrogram of monthly heating energy for the training set and four machine learning models.

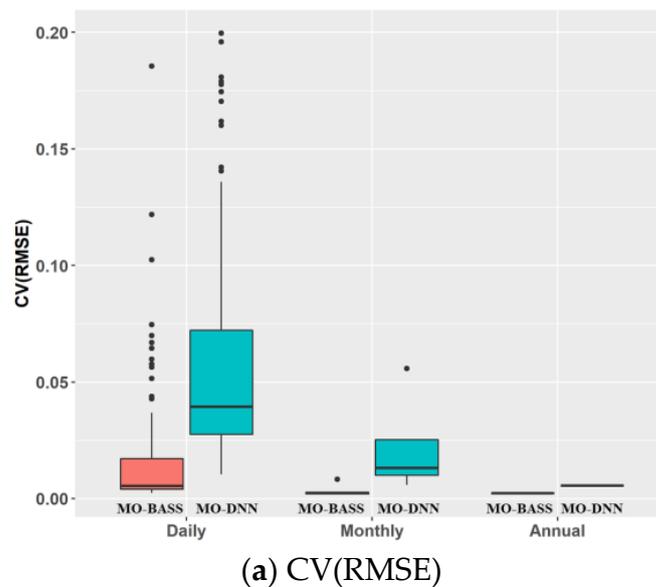
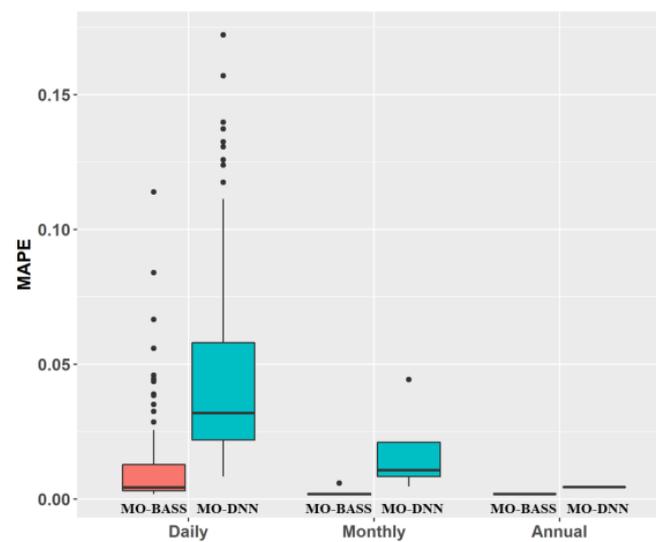
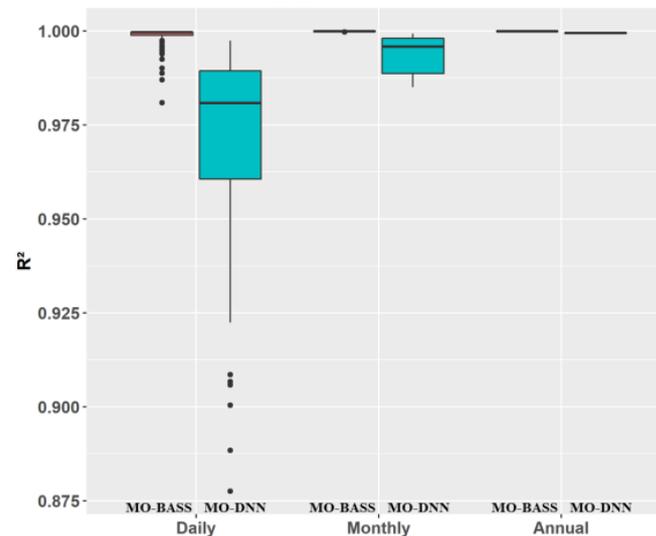


Figure 18. Cont.



(b) MAPE

(c) R<sup>2</sup>

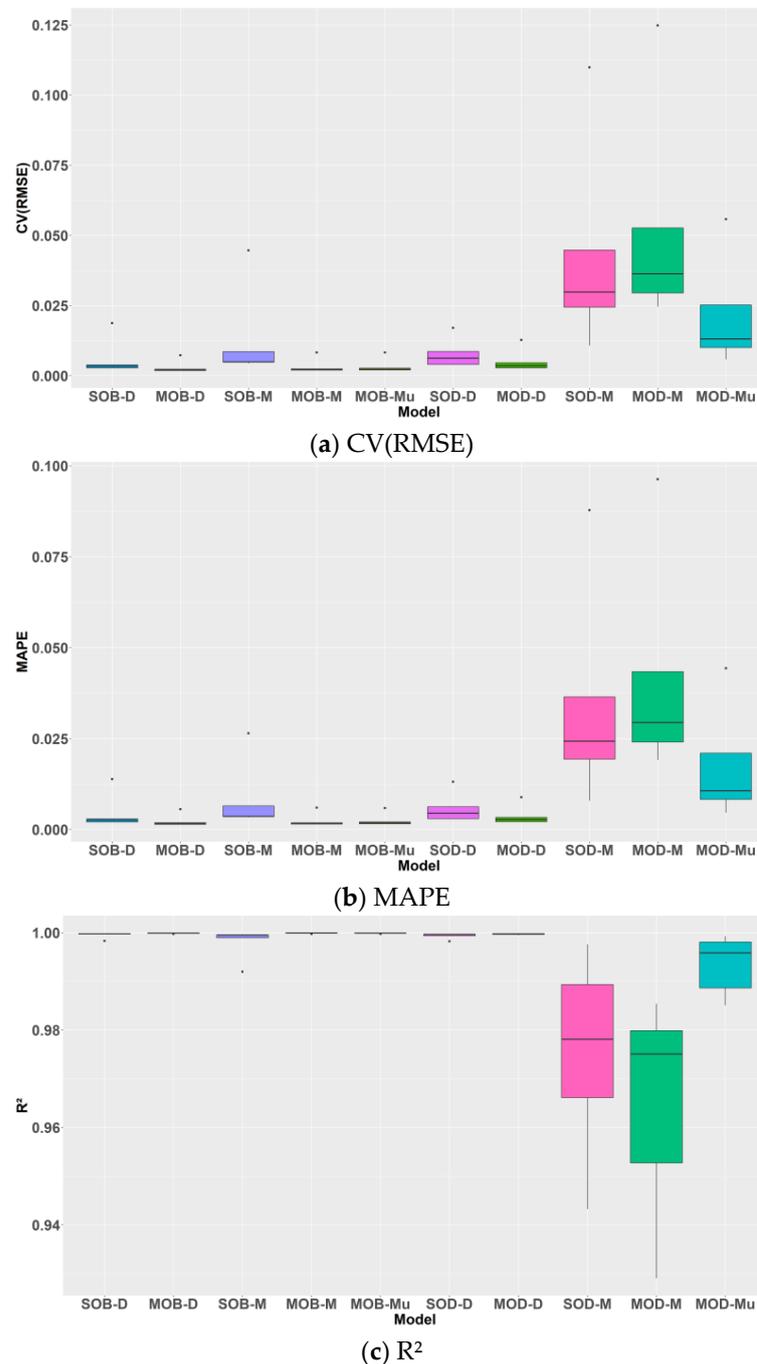
**Figure 18.** Prediction performance of two machine learning models for multi-time scales heating energy.

Table 5 shows the computation time of multi-time scale heating energy models using the BASS and DNN approaches. The computational time is quite similar for these two models. Considering the model accuracy as shown in Figure 18, the multi-time scale BASS models would be a better candidate for the multi-time scale prediction of building energy use.

### 3.3.4. Performance Analysis of Ten Models for Monthly and Annual Heating Energy

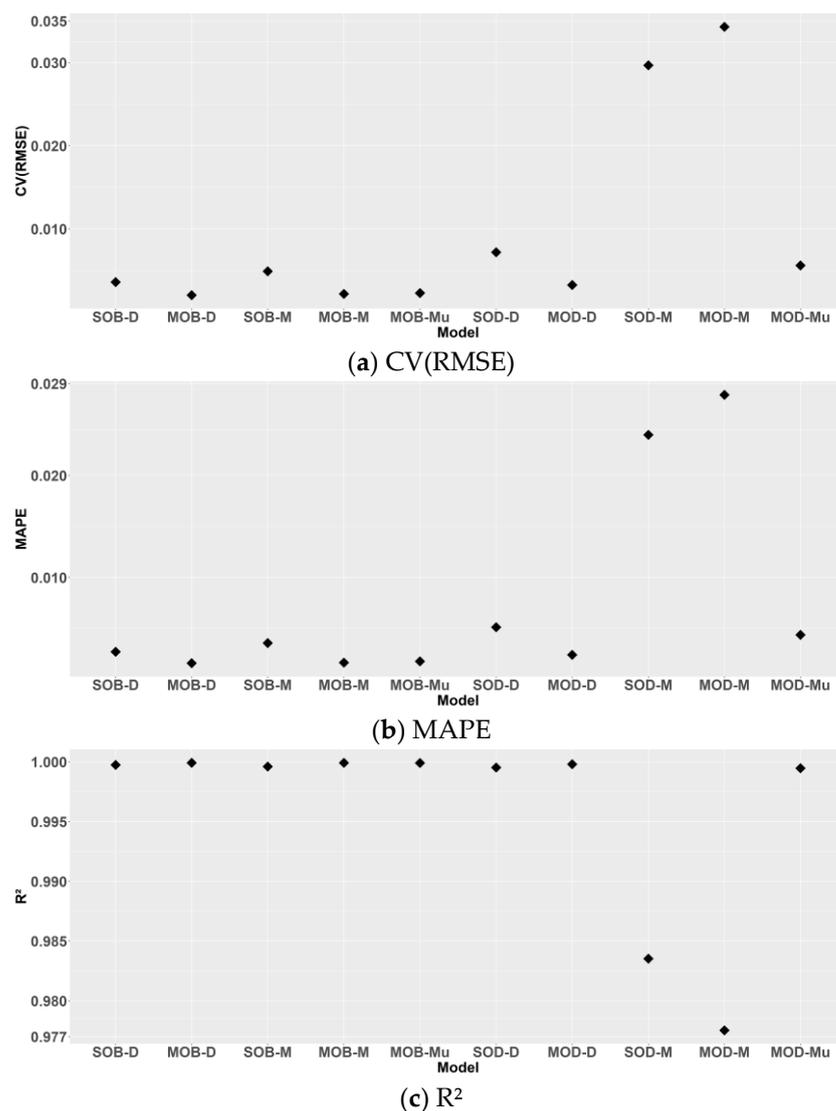
Figure 19 shows the performance of 10 models that can obtain monthly heating energy. The meanings of these 10 models are available in Table 4. The first three best models are the daily multi-output models (MOB-D), monthly multi-output models (MOB-M), and multi-time scale model (MOB-Mu) based on the BASS algorithm. These three models have very low CV(RMSE) and MAPE values. The next two models are the single-output daily BASS models (SOB-D) and the multi-output daily DNN models (MOD-D). These two models still have very high R<sup>2</sup> values—close to 1. The following two models are single-output monthly BASS models (SOB-M) and single-output daily DNN models (SOD-D). The remaining three models do not have good predictive performance compared to the other seven models. The

multi-output models have better performance compared to the single-output models for monthly heating energy.



**Figure 19.** Performance of 10 models for monthly heating energy (refer to Table 4 for model explanation).

Figure 20 shows the performance of 10 models that can obtain annual energy. The three multi-output BASS models (MOB-D, MOB-M, MOB-Mu) are the first three best models in which the CV(RMSE) values are about 0.003, the MAPE values are about 0.002, the R<sup>2</sup> values are about 0.999. The performance of the SOB-D and SOB-M single-output BASS models is slightly worse than that of the above three multi-output BASS models. Among the five DNN models, the MOD-D models and the MOD-Mu models have similar performance to the SOB-M models. The SOD-M and MOD-M models do not have good predictive performance for annual heating energy.



**Figure 20.** Performance of 10 models for annual heating energy (refer to Table 4 for model explanation).

### 3.4. Guide and Application of Building Multi-Output Energy Models

This section discusses the implications, guidelines, and applications of the results presented in Sections 3.1–3.3 from four aspects. The first aspect discusses the choice of machine learning models between the single-output and multiple-output models in building energy assessment. The second aspect is the choice of multiple-output models with or without considering output correlation in the building energy analysis. The third aspect is related to the choice of multi-output models for various time scales by taking the additive or accumulative features into account. The fourth aspect is focused on the application of the multi-output models used in this paper.

The multi-output learning models are preferred in the case of multiple outputs of building energy analysis. This is because the computational cost can be dramatically reduced in comparison with the single-output models, as discussed in Sections 3.2 and 3.3. The benefits of using the multi-output models are more significant when there are a large number of building energy outputs—for instance, over 10 outputs. Moreover, model accuracy from the multi-output models is not necessarily less than that of the single-output models in building energy assessment. Therefore, it would be recommended that the multi-output learning models should be used when applying the machine learning techniques to building energy assessment, especially for a larger number of model outputs.

There are two types of machine learning methods depending on whether to consider the correlation of building energy performance. One example of learning models is the Bayesian adaptive spline surface in which the correlation of outputs can be taken into account, while one example of learning models without considering output correlation is the deep neural network. If there are almost no or very weak correlations among building energy use, both two types of learning models can be used to create multi-output models. If the correlation of outputs is significant, it is necessary to implement the learning methods by considering the correlation of outputs. The model accuracy can be improved by maintaining the correlation of building energy use.

It is necessary to consider the additive or accumulative features when constructing the learning models for predicting building energy use in various time scales. This is because the building energy use at a larger time scale is the sum of building energy at a smaller time scale. For instance, the monthly electricity use from a building would be the annual electricity for this building. For single-output models at a smaller time scale, it is natural to sum up all the energy use at a smaller time scale to obtain total energy use at a larger time scale. However, the residuals for this simple summation would be hard to manage. If there are all positive (or negative) residuals from the smaller time scale models, the total residuals would be very large compared to the single-output model at a larger time scale. Another issue is the inconsistency between the summation of smaller time scale models and the larger time scale model, which confuses the estimation of building energy use at a larger time scale. By considering the correlation of building energy use, the multi-output learning models can create fast and consistent models at various time scales for building energy analysis. Therefore, multi-output learning models that can maintain the correlation of outputs are recommended in multiple time scale predictions of building energy analysis.

The method proposed in this paper can also be applied to performance indicators of buildings (such as energy, loads, and carbon emissions) in different scenarios although this paper is concentrated on building energy consumption. It would be interesting to explore building load profiles at various time scales (hourly, daily, and monthly), which can provide useful information on the design of district heating and cooling systems. When designing net-zero emission buildings, it is necessary to provide the matching of demand and supply at various time scales to guide for the design of PV systems. The multi-output models at various time scales would be very useful to provide accurate building demand data at various time scales to compute solar fraction, self-sufficiency, and other indicators. These results can be used for optimizing the ratio of PV rated power and battery. Moreover, the method used here can be used in uncertainty analysis, sensitivity analysis, and model calibration in building energy assessment. For instance, the multi-output models at daily periods can be used as mathematical models to calibrate these daily models using the Bayesian analysis. The computational loads would be reduced significantly since there are only two multi-output learning models for heating and cooling energy use in buildings.

#### 4. Conclusions

This research investigates the predictive performance of multi-output machine learning models at three time scales (daily, monthly, and annual) for building energy assessment using two algorithms (Bayesian adaptive spline surface and deep neural network). The results indicate that the machine learning models, which could consider the correlation of energy use, can have high model accuracy with fast computation and accumulative features at different time scales in building energy analysis. The multi-output learning models for building energy prediction would significantly reduce the computational time for creating learning models in comparison to the single-output learning models. The ability to maintain the correlation structure of energy data for the multi-output learning models is the key to providing accurate results and accumulative features. The deep neural network models can simulate the multiple outputs without taking into account the correlation of energy use. Hence, the deep neural network cannot have accumulative features. In contrast, the multi-time scale Bayesian adaptive spline surface models can have the same or similar

correlation structure of energy data from the engineering-based building energy models. This means that the predicted energy data at a larger time scale would equal the summation of energy data at a smaller time scale in building energy computation. Moreover, it is found that the simple summation of energy results from smaller time scale learning models do not necessarily have good predictions of energy use at a larger time scale due to the uncontrolled offset of large and small residuals, or the final accumulative large errors.

Three guidelines can be obtained for creating machine learning models in building energy assessment based on this research. The multi-output learning models are preferred compared to the single-output model in dealing with multiple outputs, especially in the case of the number of outputs over 10. Multi-output learning models that can consider output correlation are recommended when the multiple outputs are correlated in building energy analysis. Compared to the direct summation from the single-output models at the smaller time scale, multi-output models with a consistent accumulative feature that can simultaneously predict energy use at various time scales are preferred when dealing with multiple time scale building energy predictions.

The multi-out learning models used in this paper are only applied to an office building. More research should be implemented on other building types, for example, residential buildings, schools, and hospitals, to understand the suitability of these multi-output methods. More studies should also be conducted to explore the predictive performance of multi-output models for building energy analysis using different learning algorithms, such as random forest, support vector machine, and Gaussian process.

**Author Contributions:** Conceptualization, G.L. and W.T.; methodology, G.L. and W.T.; software, G.L.; investigation, H.Z. and B.C.; writing—original draft preparation, G.L.; writing—review and editing, W.T.; visualization, B.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the National Natural Science Foundation of China (No. 51778416) and the Key Projects of Philosophy and Social Sciences Research, Ministry of Education of China (No. 16JZDH014).

**Data Availability Statement:** Not applicable.

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

## Abbreviations

ANN	artificial neural network
BASS	Bayesian adaptive spline surface
BIPV	building integrated photovoltaic
BMARS	Bayesian multivariate adaptive regression splines
CSP	cooling set-point
CSWD	Chinese standard weather data
CV(RMSE)	coefficient of variation of the root mean square error
DNN	deep neural network
DT	decision tree
ENMIM	ensemble model named evolutionary neural machine inference model
EPD	equipment power density
EWU	exterior wall U-value
GA-NMM	genetic algorithm-based numerical moment matching
GB	gradient boosting
HSP	heating set-point
HVAC	heating, ventilation, and air conditioning
INF	infiltration rate
KNN	K-nearest neighbor
lightGBM	light gradient boosting machine

LPD	lighting power density
LR	linear regression
LSSVR	least squares support vector regression
MAPE	mean absolute percentage error
MARS	multivariate adaptive regression splines
MIMO	multi-input multi-output
MO	multiple outputs
OPD	occupancy density
PCA	principal component analysis
PCC	Pearson's correlation coefficient
PV	photovoltaic
R <sup>2</sup>	coefficient of determination
RBFNN	radial basis function neural network
RF	random forest
RU	roof U-value
SARIMA	seasonal autoregressive integrated moving average
SHGC	solar heat gain coefficient
SO	single output
SVM	support vector machine
WU	window U-value

## References

- UN Environment Programme (UNEP). *2021 Global Status Report for Buildings and Construction*; UNEP: Nairobi, Kenya, 2021.
- Zhang, Y.; Teoh, B.K.; Wu, M.; Chen, J.; Zhang, L. Data-driven estimation of building energy consumption and GHG emissions using explainable artificial intelligence. *Energy* **2023**, *262*, 125468. [[CrossRef](#)]
- Al-Shargabi, A.A.; Almhafdy, A.; Ibrahim, D.M.; Alghieth, M.; Chiclana, F. Buildings' energy consumption prediction models based on buildings' characteristics: Research trends, taxonomy, and performance measures. *J. Build. Eng.* **2022**, *54*, 104577. [[CrossRef](#)]
- Guo, Y.-Y. Revisiting the building energy consumption in China: Insights from a large-scale national survey. *Energy Sustain. Dev.* **2022**, *68*, 76–93. [[CrossRef](#)]
- Mohapatra, S.K.; Mishra, S.; Tripathy, H.K.; Alkhayyat, A. A sustainable data-driven energy consumption assessment model for building infrastructures in resource constraint environment. *Sustain. Energy Technol. Assess.* **2022**, *53*, 102697. [[CrossRef](#)]
- Fathi, S.; Srinivasan, R.; Fenner, A.; Fathi, S. Machine learning applications in urban building energy performance forecasting: A systematic review. *Renew. Sustain. Energy Rev.* **2020**, *133*, 110287. [[CrossRef](#)]
- Zhang, L.; Wen, J.; Li, Y.; Chen, J.; Ye, Y.; Fu, Y.; Livingood, W. A review of machine learning in building load prediction. *Appl. Energy* **2021**, *285*, 116452. [[CrossRef](#)]
- Lei, L.; Chen, W.; Wu, B.; Chen, C.; Liu, W. A building energy consumption prediction model based on rough set theory and deep learning algorithms. *Energy Build.* **2021**, *240*, 110886. [[CrossRef](#)]
- Liu, Y.; Chen, H.; Zhang, L.; Wu, X.; Wang, X.-J. Energy consumption prediction and diagnosis of public buildings based on support vector machine learning: A case study in China. *J. Clean. Prod.* **2020**, *272*, 122542. [[CrossRef](#)]
- Alobaidi, M.H.; Chebana, F.; Meguid, M.A. Robust ensemble learning framework for day-ahead forecasting of household based energy consumption. *Appl. Energy* **2018**, *212*, 997–1012. [[CrossRef](#)]
- Wang, Z.; Hong, T.; Li, H.; Piette, M.A. Predicting city-scale daily electricity consumption using data-driven models. *Adv. Appl. Energy* **2021**, *2*, 100025. [[CrossRef](#)]
- Jetcheva, J.G.; Majidpour, M.; Chen, W.-P. Neural network model ensembles for building-level electricity load forecasts. *Energy Build.* **2014**, *84*, 214–223. [[CrossRef](#)]
- Ferrantelli, A.; Kuivjogi, H.; Kurnitski, J.; Thalfeldt, M. Office Building Tenants' Electricity Use Model for Building Performance Simulations. *Energies* **2020**, *13*, 5541. [[CrossRef](#)]
- Wang, W.; Hong, T.; Xu, X.; Chen, J.; Liu, Z.; Xu, N. Forecasting district-scale energy dynamics through integrating building network and long short-term memory learning algorithm. *Appl. Energy* **2019**, *248*, 217–230. [[CrossRef](#)]
- Tran, D.-H.; Luong, D.-L.; Chou, J.-S. Nature-inspired metaheuristic ensemble model for forecasting energy consumption in residential buildings. *Energy* **2020**, *191*, 116552. [[CrossRef](#)]
- Tian, W.; Yang, S.; Li, Z.; Wei, S.; Pan, W.; Liu, Y. Identifying informative energy data in Bayesian calibration of building energy models. *Energy Build.* **2016**, *119*, 363–376. [[CrossRef](#)]
- Zhu, C.; Tian, W.; Yin, B.; Li, Z.; Shi, J. Uncertainty calibration of building energy models by combining approximate Bayesian computation and machine learning algorithms. *Appl. Energy* **2020**, *268*, 115025. [[CrossRef](#)]
- Koschwitz, D.; Frisch, J.; van Treeck, C. Data-driven heating and cooling load predictions for non-residential buildings based on support vector machine regression and NARX Recurrent Neural Network: A comparative study on district scale. *Energy* **2018**, *165*, 134–142. [[CrossRef](#)]

19. Jahani, E.; Cetin, K.; Cho, I.H. City-scale single family residential building energy consumption prediction using genetic algorithm-based Numerical Moment Matching technique. *Build. Environ.* **2020**, *172*, 106667. [[CrossRef](#)]
20. Lin, Q.; Liu, K.; Hong, B.; Xu, X.; Chen, J.; Wang, W. A data-driven framework for abnormally high building energy demand detection with weather and block morphology at community scale. *J. Clean. Prod.* **2022**, *354*, 131602. [[CrossRef](#)]
21. Olu-Ajayi, R.; Alaka, H.; Sulaimon, I.; Sunmola, F.; Ajayi, S. Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. *J. Build. Eng.* **2022**, *45*, 103406. [[CrossRef](#)]
22. Tian, W.; Zhu, C.; Sun, Y.; Li, Z.; Yin, B. Energy characteristics of urban buildings: Assessment by machine learning. *Build. Simul.* **2020**, *14*, 179–193. [[CrossRef](#)]
23. Tian, W.; de Wilde, P.; Li, Z.; Song, J.; Yin, B. Uncertainty and sensitivity analysis of energy assessment for office buildings based on Dempster-Shafer theory. *Energy Convers. Manag.* **2018**, *174*, 705–718. [[CrossRef](#)]
24. Wang, L.; Lee, E.W.M.; Hussian, S.A.; Yuen, A.C.Y.; Feng, W. Quantitative impact analysis of driving factors on annual residential building energy end-use combining machine learning and stochastic methods. *Appl. Energy* **2021**, *299*, 117303. [[CrossRef](#)]
25. Singh, M.; Sharston, R. Quantifying the dualistic nature of urban heat Island effect (UHI) on building energy consumption. *Energy Build.* **2022**, *255*, 111649. [[CrossRef](#)]
26. Luo, X.J.; Oyedele, L.O.; Ajayi, A.O.; Akinade, O.O. Comparative study of machine learning-based multi-objective prediction framework for multiple building energy loads. *Sustain. Cities Soc.* **2020**, *61*, 102283. [[CrossRef](#)]
27. Liu, J.; Zhang, Q.; Dong, Z.; Li, X.; Li, G.; Xie, Y.; Li, K. Quantitative evaluation of the building energy performance based on short-term energy predictions. *Energy* **2021**, *223*, 120065. [[CrossRef](#)]
28. Li, G.; Li, F.; Ahmad, T.; Liu, J.; Li, T.; Fang, X.; Wu, Y. Performance evaluation of sequence-to-sequence-Attention model for short-term multi-step ahead building energy predictions. *Energy* **2022**, *259*, 124915. [[CrossRef](#)]
29. U.S. Department of Energy. *EnergyPlus V22.1.0*; U.S. Department of Energy: Washington, DA, USA, 2021.
30. MOHURD (Ministry of Housing and Urban-Rural Development). *China, Energy-Saving Design Standards for Public Buildings*; China Building Industry Press: Beijing, China, 2015.
31. MOHURD (Ministry of Housing and Urban-Rural Development). *China, Technical Standards for Near-Zero Energy Buildings*; China Building Industry Press: Beijing, China, 2019.
32. Francom, D.; Sansó, B. BASS: An R Package for Fitting and Performing Sensitivity Analysis of Bayesian Adaptive Spline Surfaces. *J. Stat. Softw.* **2020**, *94*, 1–36. [[CrossRef](#)]
33. An, N.; Zhao, W.; Wang, J.; Shang, D.; Zhao, E. Using multi-output feedforward neural network with empirical mode decomposition based signal filtering for electricity demand forecasting. *Energy* **2013**, *49*, 279–288. [[CrossRef](#)]
34. Cohen-Addad, V.; Kanade, V.; Mallmann-Trenn, F.; Mathieu, C. Hierarchical Clustering: Objective Functions and Algorithms. *J. Acm.* **2019**, *66*, 1–42. [[CrossRef](#)]
35. Kaminsky, R.; Shakhovska, N.; Kryvinska, N.; Younas, M. Dendrograms-based disclosure method for evaluating cluster analysis in the IoT domain. *Comput. Ind. Eng.* **2021**, *158*, 107402. [[CrossRef](#)]
36. Varshney, A.K.; Muhuri, P.K.; Lohani, Q.M.D. PIFHC: The Probabilistic Intuitionistic Fuzzy Hierarchical Clustering Algorithm. *Appl. Soft Comput.* **2022**, *120*, 108584. [[CrossRef](#)]