An agent-based modeling approach combined with deep learning method in Simulating household energy consumption

Shanjun Tian [a,b,*], Yi Lu [a], Yuanjie Zheng [a]

[a] School of Information Science and Engineering, Shandong Normal University, JiNan, China

[b] Department of Energy and Power Engineering, Tsinghua University, Beijing, China

[*] Corresponding author. Email: tianshanjun@sdnu.edu.cn

Abstract: Household has been one of main energy consumption sectors in China, which brings about more challenges for the control of air pollutants and greenhouse gases. Due to the heterogeneities among household individuals, it is much more difficult to regulate the household energy consumption. So, it becomes necessary to understand and guide household energy consumption behaviors. In this paper, a multi-level agent-based model was constructed to aggregate the macro-level energy consumption (such as regions) from the micro-level entities (such as devices), and deep learning models were conducted and embedded in to predict the household energy consumption behaviors with factors as inputs and behavior parameters as outputs, meanwhile an energy substitution mechanism was designed. A series of scenarios were simulated based on actual survey samples, and the results indicated the great potential of our proposed method in the household energy simulation. The simulation fit the actual statistics well in the validation step. Under scenario analysis, the whole coal-to-electricity conversion was deeply determined by low-income individuals, and the effect pattern of certain factor was distinct and could be enhanced by others, while the marginal utility changed non-linearly with peak point effected by income growth rate. It is inspired that the local economic and socio-demographic situations should be comprehensively considered to achieve a trade-off between costs and expected results of regional energy strategies. Our research proposes a novelty study framework to simulate the household energy consumption from the micro individual perspective, and it is a valuable attempt to understand and quantify complex household energy consumption behaviors.

Keywords: Agent-Based Model, Deep Learning, Household Energy Consumption

Nomenclature

AAIGR	Average annual income growth rate of household group	Mon_{heat}	Annual heating months
ABM	Agent-based model	P_n	Predicted equipped probability of device n
AUM	Annual used month of device	P'_n	Effected equipped probability of device n
BHDPO	Building heating demands percentage of the original	$Price_n$	Price of energy type n
$Budget_e$	Budget for coal-to-electricity substitution	$potential_{ce}$	Potential of coal-to-electricity substitution

CE_n	Conversion coefficient of energy type n	PW_{coal}	Original heating provided by coal
$Cost_n$	Annual cost of energy type n	PW_{heat}	Original annual heating demand
$DE^i{}_n$	Annual consumption of energy type n for device equipped by household i	PW_n	Annual Heating provided by energy type n
		RCC	Coal consumption relative to the original condition
DEP	Device equipped probability	RE_n	Annual consumption of energy type n for a region
DUH	Daily used hour of device	SB	Subsidy provided by government
ECB	Energy consumption behavior	SV	Building heating demands percentage of the original
ECF	Energy consumption feature	TEED	Thermal efficiency of electric devices
ECH	Energy consumption habit	TR'_c	Thermal efficiency of new coal device
$Energy_m$	Energy consumption of device m, with values of central heating, electric, coal, biomass	TR_m	Thermal efficiency of device m, c means coal, e means electricity
$HE^i{}_n$	Annual consumption of energy type n for household i	UEC	Unit time energy consumption of device
IC	Annual income	UE_m	Energy consumption of device m per unit area
IS	Average annual income growth rate of household group	α	Controller of threshold value at which the family is affected

1 Introduction

The household sector accounted for 13.2% of the total final energy consumption of China in 2017, making it one of the main energy consumption sectors [1]. As a large developing country, energy demands from the household sector will continue to grow due to continuous economic development [2]; thus, it becomes an important source of air pollutants and greenhouse gases [3]. Furthermore, bulk coal used by households exhibit much higher pollutant emission intensity than centralized industrial coal and is more difficult to be supervised due to its diverse and widespread usage and the lack of desulfurization, denitrification, and dust removal treatments [4]. Therefore, it is necessary to study the household energy consumption to save energy, reduce emissions, and control greenhouse gases.

While compulsory measures can be applied to the industrial sector, it is much more difficult to regulate the energy consumption of household sector, due to the heterogeneity among different households [5]. “Coal-to-electricity” and “coal-to-gas” policies are facing a series of problems such as excessive heating expenditures for residents and unsustainable subsidies paid by the government [6]. Household energy consumption is determined by individual behaviors, which is influenced by multiple factors including geographical location, climate, household income, etc. [7]. Whereas improving the device efficiency is generally proposed by numerous policies, understanding and guiding household behavior may be equally important for reducing household energy consumption [8].

The research methods related to household energy consumption behavior (ECB) can be generally divided into two categories: “top-down” and “bottom-up” [9]. The “top-down” method regards households as a whole, then studies the relationship between aggregated energy consumption and energy-related variables such as macroeconomic indicators, climatic conditions, and household structure. Conversely, the “bottom-up” method typically focuses on energy consumption at the individual scale, then calculates regional or national energy consumption based on representative weights [10,11].

The input data of “top-down” method are macro statistical variables obtained from household groups, such as the gross regional product, heating season temperature, and energy prices [12]. It can illustrate the relationship between different factors and energy consumption. However, due to the “black box” characteristics, it is a bit difficult to explain the physics behind this relationship. Energy measures formulated by the “top-down” method at the family-level may ignore atypical individuals [13]. In contrast, the “bottom-up” method separates the overall energy consumption process into spatial and temporal dimensions. It calculates the energy consumption of end devices, and aggregates them to macro-level, so it provides a more detailed overall representation [14]. Therefore, the “bottom-up” method is better to evaluate the effect of various energy measures and can be used to predict future trends or make medium and long-term energy supply strategies in the household sector. Multidisciplinary knowledge and technology should be integrated to quantify the inherent uncertainties and obtain in-depth knowledge of ECB.

An agent-based model (ABM) is a typical “bottom-up” method. The households with various characteristics and the complex relationships between each other can be simulated by a series of heterogeneous entities [15]. The aggregation process of energy consumption occurred in different scales can also be simulated, such as regions and devices.

At present, ABMs are predominantly applied to household energy consumption to: (1) simulate the energy consumption of a specific household or building [16,17]; (2) measure the physical transformation of entities in energy system, then the energy supply, management, and regulation [18,19]; (3) consider the supply-demand balance from the perspective of energy-economic system with energy as commodities, then energy strategies under various market conditions [20,21]; (4) describe the abstract relationship between households based on methods such as social networks, then the diffusion of energy-saving technologies or awareness using technology diffusion theory [22,23].

Existing ABMs typically mainly deal with electricity and natural gas [24]. However, coal, firewood, and straw are still widely used in China, and are major causes of interior and outdoor air pollution. The government is prompted to vigorously implement “coal-to-electricity” and “coal-to-gas” policies [6]. Therefore, these energy sources should be incorporated into ABMs to get a deeper understanding of clean energy substitution. In addition, the entities involved in the energy consumption process can be divided into several levels, such as energy types, devices, occupants, households/buildings, communities, and cities/regions, and the applicable research methods are different for different levels. To analyze the overall impact of different factors on household energy consumption and evaluate the expected effects of various policy measures, it is essential to integrate multi-levels of entities into the model and combine different research methods.

The expression of household ECB is an important step of the constructing of household energy consumption ABM, as the stochastic nature of the behavior makes its prediction and expression highly challenging [25]. Existing uncertainty quantification methods of household ECB predominantly include the Guide to the Expression of Uncertainty in Measurement, probabilistic and non-probabilistic approaches [26]. The probabilistic approach is a classic method [27,28] involving statistical analyze and data mining based on monitoring or questionnaire data, which determines the probability of specific ECB under different environmental conditions [29,30].

Under the above background, our main objective is extending our previous research work, which successfully combined the probabilistic approach with the agent-based modeling method [31], to deeply understand and guide the energy consumption behaviors of household individuals. The deep learning method is introduced into the model framework, which aims at incorporating much more internal, external, subjective, objective, and other related factors to predict the household energy consumption behaviors more realistically. A multi-level agent-based model of household energy consumption is proposed to manage multi-level entities which are abstracted from objects in real world, and simulate the aggregation process of energy consumption from devices to regions via households. We hope that the model can not only simulate household group's energy consumption patterns in detail, but also provide reference for the evaluation of specific energy substitution measures.

Section 2 presents the structure of the model, Section 3 provides and discusses the results of experiments, and Section 4 conclude the whole study work briefly.

2 Methods

2.1 Design principles

The model design mainly involves the expression of household ECBs at the micro-level (with hours or days as time scale and devices as spatial scale) and the aggregation of the energy consumption of household groups at the macro-level (with years as the time scale and regions as the spatial scale). The ECBs of an individual household depend on its living habits and are accompanied with some randomness [32]. The parameters of ECB respectively indicating the equipping and using of various devices are determined by the energy consumption habits (ECH) of the household. Meanwhile, due to the inherent randomness, the ECB parameters should be described in the form of probability distributions. It is believed that the ECH of each household are determined by the comprehensive effect of some characteristics of the household, i.e., the ECF [33]. Therefore, households with similar ECFs will always have similar ECH, then the ECB parameters will follow similar probability distributions.

The quantification of ECB parameters can be summarized as the construction of the mapping relationship from various ECFs to ECB parameters. The ECB parameters mainly include the device equipped probabilities (DEPs), the annual used months (AUMs) and daily used hours (DUHs) of devices, and the unit time energy consumption of devices (UEC).

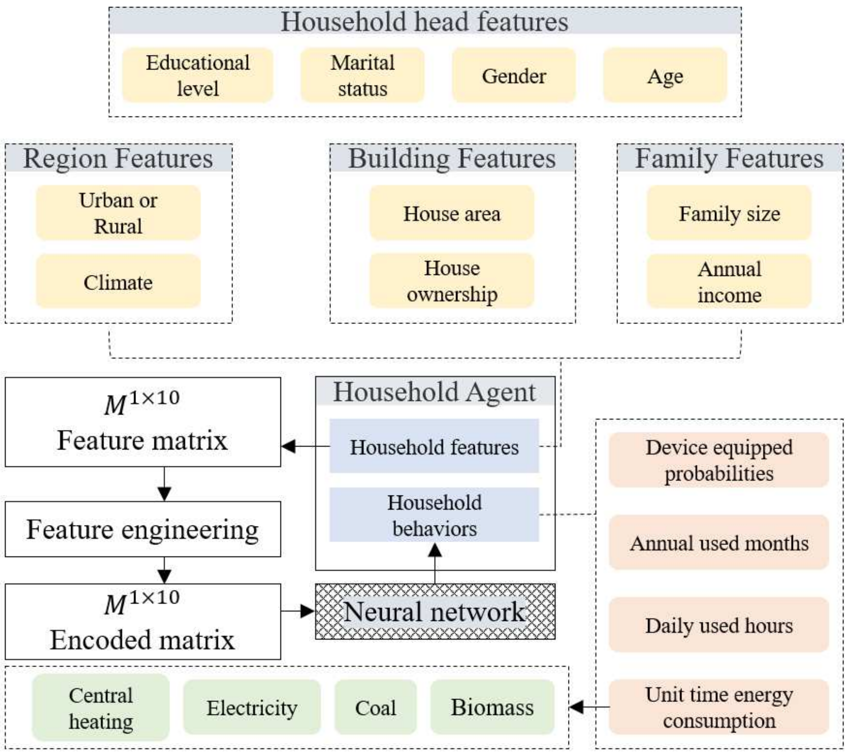

Figure 1 The energy consumption behavior determination according to energy consumption features

The ECFs can be divided into subjective and objective factors [34,35,36,37], or demographic factors, cognitive psychological factors, and external situational factors. These ECFs generally cover three dimensions, i.e. time, user, and space [38]. Among them, 10 typical factors were chosen as the considered ECFs, such as climate, whether the located region is urban or rural, house area, house ownership, family size, annual income, education level and marital status and gender and age of the head of household. These features were further classified into Region features, Building features, Family features, and household head features. The features formed a 1×10 matrix, and further formed a 1×10 encoded matrix by means of feature engineering. The ECB of each household can be quantified by establishing the mapping relationship from ECFs to ECB parameters. The deep learning model is a feasible method by taking the encoded matrix as input and ECB parameters as output (Figure 1).

In the agent-based model, the consumption of an energy type by a household on a day is first simulated. Then, the energy consumption of the household in a month and a year can be obtained through aggregation of days and months. Finally, the energy consumption of all households in a region can be obtained through aggregation of household group located in the region.

2.2 Data collection and pretreatment

The data used in the study came from the Chinese Residential Energy Consumption Survey (CRECS 2014) and Chinese Family Panel Studies data (CFPS 2014). The CRECS questionnaire was designed by the Department of Energy Economics at Renmin University of China and included household demographic characteristics, dwelling characteristics, heating area and appliances, frequency and duration of appliance use, heating energy costs, and electricity bills. The CRECS data covered 3863 households among 28 mainland provinces. The CFPSs was conducted by the Institute of Social Science Survey (ISSS) at Peking University, and mainly focused on the social, economic, population characteristics at the individual, household and community levels. The CFPS data covered 16000 households in China.

All the abnormal part of the data was rectified or abandoned manually and all the missed data fields were replaced by the average values, before the data was used to conduct the study. Finally, 3446 households among CRECS 2014, and 13946 households among CFPS 2014 were chosen as samples in this study.

Otherwise, as an important input of substitution mechanism, the historical energy consumption of each household agent needed to be stored in data files and imported into the model. The temperature of each region agent was another environmental variable that changed over time and needed to be imported externally. The highest and lowest temperature of each month in 2014 were recorded and the monthly temperature of each region agent was set as a random number that varied continuously between the highest and lowest temperature. Only when the temperature was lower than the preset threshold, a household agent was possible to start heating behavior.

2.3 The usage of deep learning approach

2.3.1 Feature engineering

Of the ECFs, climate, urban or rural, house ownership, gender, marital status, education level were categorical variables, whereas house area, annual income, family size, age were numerical variables. To be dimensionless, all ECFs needed to be pretreated by feature engineering before input into the deep learning model. Categorical variables were converted into real numbers through target encoding [39], whereas numerical variables were taken as the logarithm (base 10) to eliminate magnitude differences. Then, all features were processed through Min-Max Scaling and Variance Scaling (Standardization) to achieve normalization [40]. Finally, the Synthetic Minority Oversampling Technique (SMOTE) was conducted to improve the deep learning model results for small family samples. Sklearn [41] and imblearn [42] were the main toolkits in feature engineering.

2.3.2 Deep learning model

The input variables of the deep learning model were ECFs processed by feature engineering, which were denoted as $Input_1$, $Input_2$, $\cdots$, $Input_{10}$. The output variables of the deep learning model were EPDs, AUMs and DUHs of four energy device types (biomass, coal, electric, and central heating as energy sources). The EPDs predicted by the deep learning model were all decimals ranging from 0 to 1. A device was equipped if the probability was greater than 0.5 by default. The parameters of deep learning models were then optimized by comparing the simulated and actual results of device equipment.

Therefore, 12 back-propagation (BP) artificial neural networks were trained, corresponding to 3 ECB parameters of 4 device types (Figure 2). There were 10 nodes in the input layer, which are ECFs processed by feature engineering. The output layer contains one node, which respectively indicates EPD, AUM and DUH of different devices. For EPD parameters, a sigmoid activation function was added to the end of output layer, and for AUM and DUH, the sigmoid function was not used. After repeated trial, the number of hidden layers and nodes of each hidden layer were set to be 3 and 128 for all deep learning model.

A household agent could equip only one device one time. For the four device types, the predicted equipped probabilities were assumed to be P_1, P_2, P_3, and P_4 respectively, and the generated random number was P_h. If $i, j \in [1,2,3,4]$ exists, $P_j - P_h = \max_i(P_i - P_h)$, and $P_j - P_h > 0$, then device j would be equipped by the target household; otherwise, none device would be equipped.

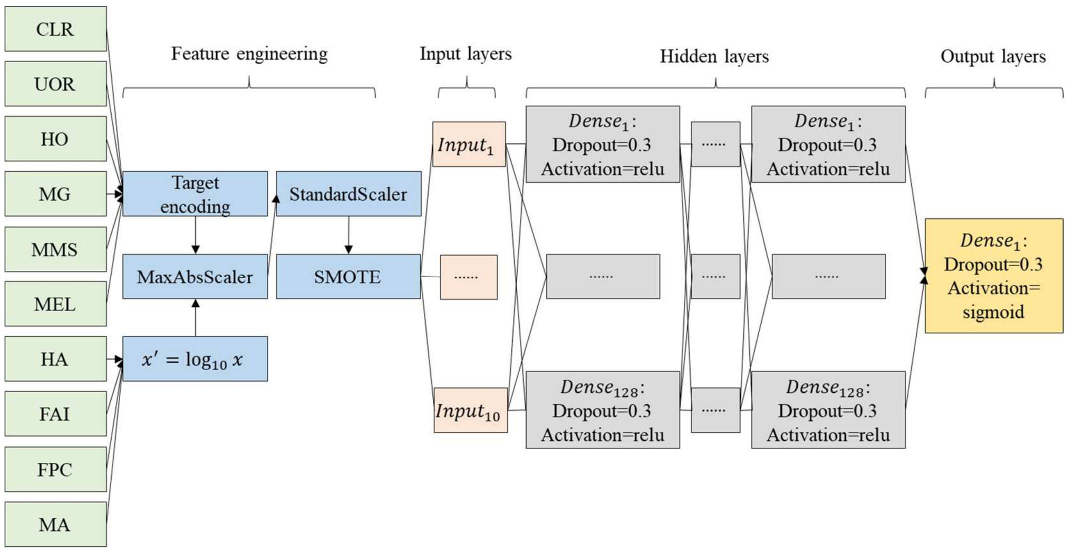

Figure 2 The structure of deep learning models for prediction of device equipped probability

2.3.3 Substitution mechanism based on device equipped probability

For a household agent used coal for heating in the last year: the annual income, the annual heating provided by coal, the annual heating months were respectively defined as $IC\ (CNY)$, $PW_{heat} = PW_{coal} * TR_c\ (kgce)$, $Mon_{heat}(month)$. PW_{coal} was the annual consumed coal and TR_c was the thermal efficiency of coal device.

In current year, the EPD of coal and electric device were P_c and P_e, the thermal efficiency of coal device and electric device were TR'_c and TR_e, the conversion coefficient of coal and electricity from weight units to coal equivalent units were CE_c and CE_e, and the annual heating demand was SV times of the original amount (i.e., PW_{heat}) due to energy-saving technology adopt by the household.

Then, the coal demand was $PW_c = PW_{heat} * SV/TR'_c \cdot CE_c\ \ (kg)$ if coal device was equipped, whereas the electricity demand was $PW_e = PW_{heat} * SV/TR_e \cdot CE_e\ \ (kwh)$ if electric device was equipped (calorific value calculation).

If the prices of coal and electricity were $Price_c\ (CNY/kg)$ and $Price_e\ (CNY/kwh)$, respectively, the annual cost of coal and electricity were as follows:

$$Cost_c = PW_c * Price_c\ \ (CNY) \tag{1}$$

$$Cost_e = PW_e * Price_e \ \ (CNY) \tag{2}$$

Assuming that the average annual income growth rate of households was IS, and 8% of the increased income was used to substitute the coal with electricity (according to the Chinese Household Energy Consumption Report 2016, energy expenditure accounts for approximately 8% of the annual income), the direct subsidy paid by the government for per household was SB, then the budget for coal-to-electricity substitution would be:

$$Budget_e = Cost_c + SB + IC * IS * 0.08 \tag{3}$$

Note that, for users used coal in the last year, the coal cost $Cost_c$ would also be part of the budget. The potential for a household whether substituted the coal with electricity was evaluated through the following formula:

$$potential_{ce} = \begin{cases} 0.0 & , \quad Cost_e > \alpha * Budget_e \\ 1.0 - Cost_e/(\alpha * Budget_e), & \quad Cost_e \leq \alpha * Budget_e \end{cases} \tag{4}$$

Then, the affected equipped probabilities of coal and electric device would be P'_c and P'_e.

$$P'_c = P_c - P_c * effect_{ce} \tag{5}$$

$$P'_e = P_e + (1.0 - P_e) * effect_{ce} \tag{6}$$

where the range of P'_c would be $[0.0 \ \ , P_c]$ and the range of P'_e would be $[P_e, 1.0]$.

Then the affected probabilities were used to model the ECBs of household agents.

The core idea of this substitution mechanism was that the smaller was the cost / budget, the greater was the possibility of coal-to-electricity substitution. By setting the value of α, the threshold value at which the family was affected can be adjusted. In this study, the value of α was a probability that followed a Poisson distribution with average values of 1.0.

2.4 The structure of agent-based model

2.4.1 Agents, variables, behaviors and scales

An agent-based model is composed of various agents which are composed of variables and behaviors. The agents in the model are described as follows from Table 1 to 4.

Model: A complete code or program that can execute all preset simulation processes.

Table 1 Model

	Remarks
Variables	
Data file	Detailed path and file name of the data file
Schedule	Container of Region entities
Time	Time in the model

NN models	Deep learning models to predict energy consumption behavioral parameters
Data collector	Energy data of entities that need to be collected during the simulation
Behaviors	
Model step	Execute a simulation step for the Model and Region contained in the Model
Region data aggregation	Aggregate the energy data of Regions contained in the Model

Region: The geographical space where the household entities locate. The region agent can be further separated into several sub-regions according to administrative divisions, climatic conditions, economic levels, etc.

Table 2 Region agent

	Remarks
Variables	
Climate	Region climate, with values including severe cold, severe cold-cold, cold, cold-hot summer cold winter, hot summer and cold winter, mild, and hot summer warm winter.
Urban	Rural or urban.
Schedule	Container of Household entities
Income increase	Average annual income growth rate of all households in the Region
Subsidy	Subsidies for households to use clean energy
Coal price	Price of coal in the Region
Electricity price	Price of electricity in the Region
Behaviors	
Initialization	Initialize the variables
Feature engineering	The pretreatment of ECFs before input into the deep learning model
Region step	Execute a simulation step for the Region
Data aggregation	Aggregate the energy data of Households contained in the Region

Household: The entity representing the household, including members contained in the household.

Table 3 Household agent

	Remarks
Variables	
Region	Region to which the Household belongs
House area	Area of house
House owner	Ownership of house, with values including own rights and renting
Annual income	Total annual income of all family members
Family size	People count of the family
Age	Age of household head
Gender	Gender of household head, the values include male and female
Marital status	Marital status of household head, the values include married, unmarried, and once married
Educational level	

	Educational level of household head, the values include none, elementary school,
Power of last year	junior high school, high school, university, graduate, doctorate
Devices	Energy consumed in the previous year
Behaviors	Equipped devices, including biomass, coal, electric, and central heating devices
Initialization	
Equip devices	Initialize the variables
Activate AUMs	Determine whether to equip certain energy devices according to a specific
Activate DUHs	probability
Household step	Determine the annual used months of devices according to probability distributions
Data aggregation	Determine the daily used hours of devices according to probability distributions
	Execute a simulation step for the Household agent
	Aggregate the energy data of Devices contained in the Household

Device: The entity representing the device; devices consuming the same energy type are combined.

Table 4 Device agent

	Remarks
Variables	
Energy type	Energy type the device can use. The values Include biomass, coal, electric, and central heating
Unit power	Energy consumption per unit time
Behaviors	
Device step	Execute a simulation step for the Device
Calculation	Calculate the energy consumption of the current device

Scale: One simulation step represents one day and one simulation epoch represents one year (365 days). In order to obtain the simulation results of households in a certain year, the simulation process is repeated for several epochs to obtain the average value.

2.4.2 Process overview and scheduling

In the model, the solo Model entity contained several Region agents, each Region agent contained several Household agents, and each Household agent contained several Device agents; thus, a multi-level spatial structure was developed (Figure 3, shown as sequence diagram drawn with Enterprise Architect).

An epoch was a complete simulation cycle, corresponding to one year in the real world. In each epoch, every entity performed the preset operations. An epoch was composed of an initialization phase and several consecutive steps. In the initialization phase, the parameters of entities were set, including the initialization of model, regions, households, and devices. Each step corresponded to one day in the real world and started by executing the model step. Then, the step functions of lower-level entities were called by higher-level entity, i.e. the model-step called the region-step, the region-step called the household-step, and household-step called the device-step. The action order of lower-level entities was determined by a simple random method, that is, the accessed probability of each lower-level entity followed a uniform distribution.

In addition to calling the steps of lower-level entities, the upper-level entities also completed their own tasks in each step, as follows: the temperature of environment was set according to the preset input data during the region-step (if a new season began); the equipped device and the annual used months (if a new year began) or the daily used hours (if a new month began) were determined during the household-step; the annual energy consumption of central heating device (if a new year began) or daily energy consumption of other device types (if a new day began) were calculated during the device-step.

When it was determined whether a device type would be equipped by a household agent, equipped probabilities was predicted by deep learning models and a random number was generated. Only if the random number was smaller than the probability, the corresponding device was possible to be equipped, and finally the most possible device would be equipped. Annual used months and daily used hours were predicted by deep learning models; however, they were regarded as the average values of repeated actions. So, for a certain device-step, the annual used month or daily used hour was randomly picked from a Poisson distribution with predicted value as mean. In some scenarios, the annual income growth rates of households were needed, and they were also assumed to follow Poisson distribution with preset value as mean.

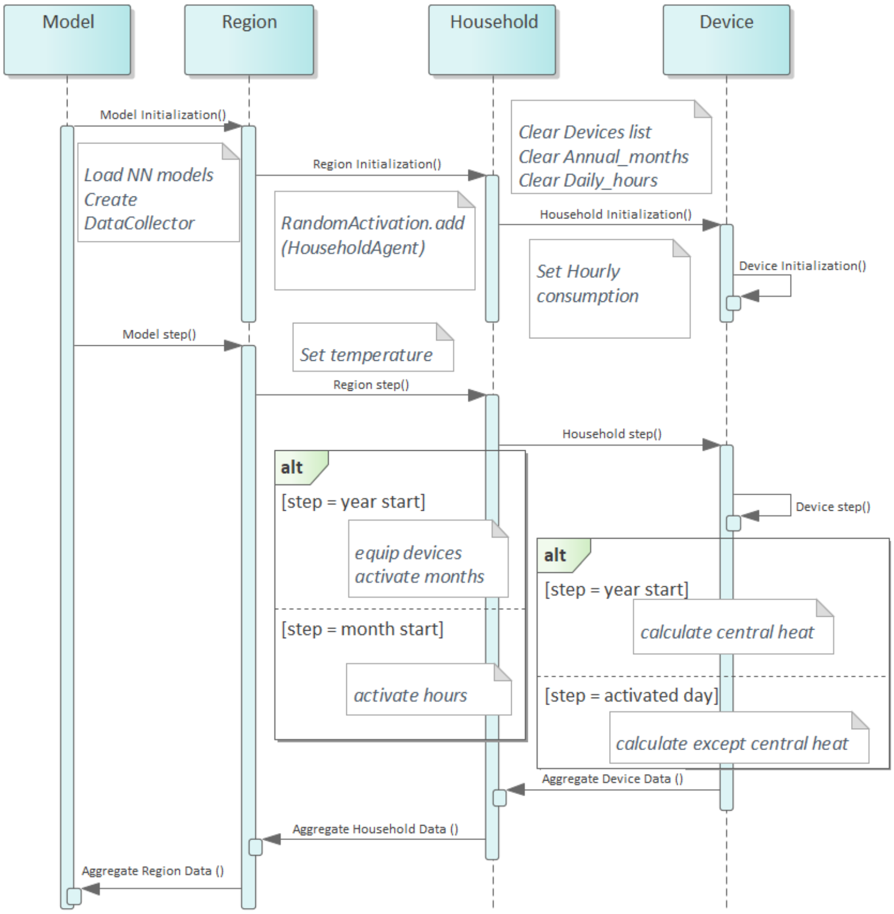

Figure 3 Overview and scheduling of the agent-based model

After the step was completed, each agent fed back its consumption of various energy types to the upper-level agent and the upper-level agent aggregated the feedbacks to obtain its own energy consumption. Once an epoch was completed, the energy consumption of the entire household group could be obtained through the model entity.

2.4.3 Calculation and aggregation of energy consumption

The calculation methods of energy consumption were different for different devices (Table 6) [43]. The site energy of a central heating device was calculated as follows:

$$Energy_{central\ heating} = UE_{central} * HeatingArea * SHSs \quad (7)$$

where $UE_{central}$ represented the unit area energy consumption, and SHSs represented the count of standard heating seasons (SHSs), i.e. (the annual heating months) / (months of an SHS).

Table 6 The parameters of energy consumption calculation

Device type	Parameters	Value	Unit
All	Default heating area	50	m^2
Central heating	$UE_{central}$	12.50	$kgce/(m^2 \cdot SDE)$
	Months of SDE	4.15	month
Electric	Default HourlyPower	1,200	W
	PowerFactor	1,200 W for	

Coal	UE_{coal}	50 m^2	$kg/m^2 \cdot day$
Biomass	$UP_{biomass}$	0.1	$kg/m^2 \cdot hour$
	UP factor	2.0	
		2 kg for 50 m^2	

The energy consumption of electric device was calculated as follows:

$$Energy_{electic} = HourlyPower * PowerFactor * AnnualHeatHours * CE_e \tag{8}$$

Where HourlyPower is the common value of unit time electric consumption and the real power was adjusted by PowerFactor which was positively related to the heating area.

The energy consumption of coal device was calculated as follows:

$$Energy_{coal} = UE_{coal} * HeatingArea * AnnualHeatingDays * CE_c \tag{9}$$

where UE_{coal} was the daily unit area coal consumption.

The energy consumption of biomass device was calculated as follows:

$$Energy_{biomass} = UP_{biomass} * UP\ \ Factor * AnnualHeatingHours * CE_w \tag{10}$$

where $UP_{biomass}$ was the common value of unit time biomass consumption and the real value was adjusted by UP Factor which was positively related to the heating area.

The annual consumption of energy type n for household i (${HE^i}_n$) equaled to that of its equipped device (${DE^i}_n$), which was recorded as

$${HE^i}_n = {DE^i}_n\ \ , n \in \{firewood, coal, electricity, thermal\} \tag{11}$$

The annual consumption of energy type n in a region (RE_n) was the sum of energy consumption of each household (${HE^i}_n$) in the region (where N was the count of households), which was recorded as:

$$RE_n = \sum_i^N {HE^i}_n, i \in [1,2,\cdots,N] \tag{12}$$

The average energy consumption can be obtained by the ratio of total energy consumption and the household count.

3 Results

3.1 Model initialization

The model initialization was aimed at loading the pre-trained deep learning models. The region initialization was aimed at setting the attributes of the region agents. A total of 14 region agents were generated in the model. The household initialization was aimed at setting the attributes of the household agents. The total number of household agents, the count of household agents assigned to each region, and the attributes of each household agent were set according to the actual survey samples, i.e. CRECS and CFPS. The multi-valued attributes of household agents such as the equipped device, annual months, and daily hours were stored by matrixes, and the elements of matrixes were all set to zero during the initialization phase. CRECS was used as the test group to model the behavioral pattern of different household classification and CFPS was used as the reference group. The device initialization was aimed at setting the attribute values of the device, including the used energy type and the unit time energy consumption. The unit time energy consumption was set according to the Chinese Household Energy Consumption Report 2016, which was written based on CRECS 2014 data. Only heating devices were considered in the experiments, with kilograms of coal equivalent (kgce) as the energy consumption unit.

3.2 Training of deep learning models

The loss and the area under the curve (AUC) were used to estimate the deep learning models' predicting ability of the equipped probabilities of four device types. Smaller was the loss, and closer was the AUC to 1.0, better was the deep learning models. CRECS 2014 data were used as input. After 200 training epochs, the deep learning models finally achieved satisfied loss (Figure 4) and AUC (Figure 5).

The loss and AUC curves with training epochs as x-axis of biomass, coal, electric, and central heating devices were shown in Figure 4 and Figure 5, respectively. The loss values declined to close to zero and the AUC values increased to close to 1 as training epochs increased.

For the annual used months and daily used hours of devices, regression was used instead of classification to train the deep learning models. As the mean square error reached minimum, the curves of predicted and actual data with household index as x-axis were compared (Figure 6). The averages of annual months and daily hours in the sample data were 1.78 (Figure 6 a) and 7.22 (Figure 6 b); the corresponding predicted averages were 1.81 and 7.55. The predicted parameter distributions were close to the actual sample data, indicating that the deep learning model was capable to predict the household ECB parameters.

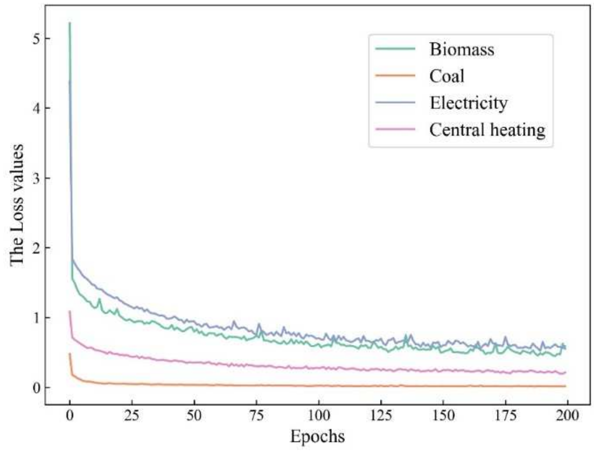

Figure 4 The Loss curves of deep learning models

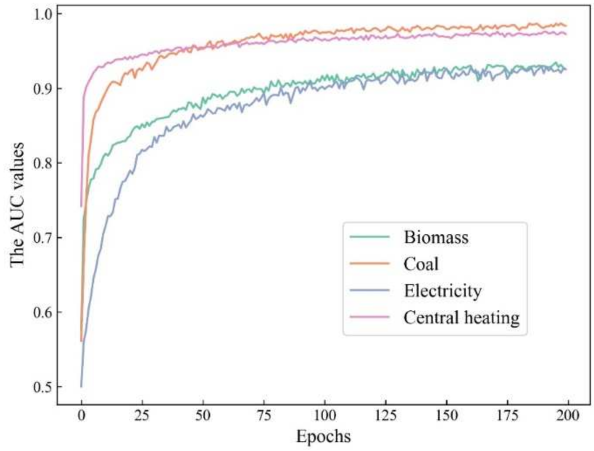

Figure 5 The AUC curves of deep learning models

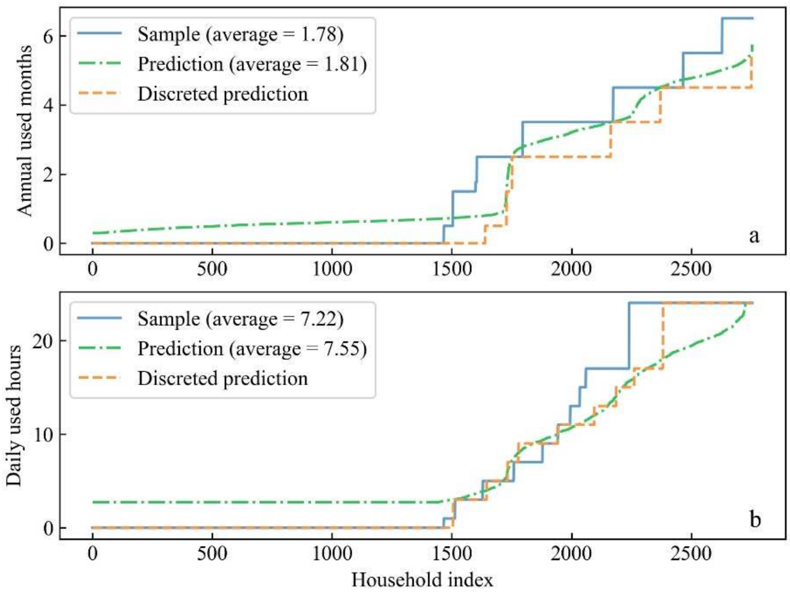

Figure 6 Curves of actual and predicted annual months and daily hours

3.3 Validation of agent-based model

The ABM framework was constructed as mentioned in section 2.4, and the deep learning models were embedded. The parameters of the agents were initialized with the sample data of CRECS 2014, and the cumulative average of the outputs of 200 times repeated simulation were collected. The total energy consumption and the proportions of different energy types were mainly observed (Figure 7). The proportions reached a stable state after approximately 20 simulation times.

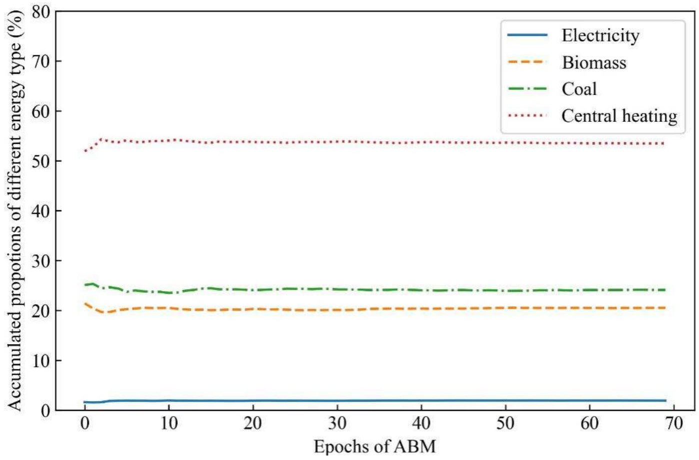

Figure 7 The curves of accumulated proportions of different energy types

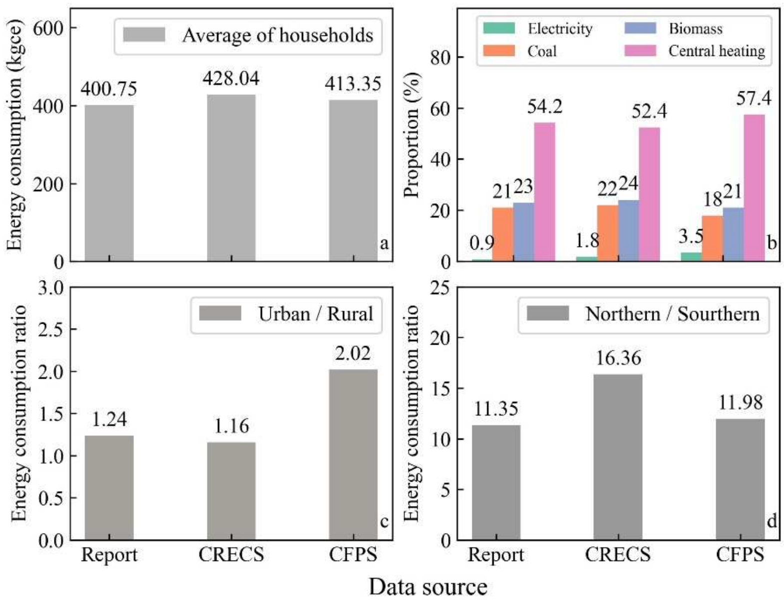

Figure 8 Comparative analysis of simulation results and reported data

Then, the agents were initialized with the sample data of CRECS 2014 and CFPS 2014, and the average results of 20 times simulation were compared with the statistics data of the Chinese Household Energy Consumption Report 2016 (Figure 8). In the report, the average annual energy consumption per household for heating was approximately 400.75 kgce, whereas the simulated results of CRECS 2014 and CFPS 2014 were respectively 428.04 kgce and 413.35 kgce (Figure 8 a). In the report, electricity, biomass, coal, and central heating respectively account for 0.97%, 20.65%, 23.35%, and 54.16% of the total energy consumption; whereas the simulated results were 2.77%, 18.76%, 29.23%, 49.24% for CRECS 2014 and 3.58%, 17.98%, 21.08%, 57.37% for CFPS 2014 (Figure 8 b). In the report, the ratios of average annual energy consumption for heating in urban and rural (Figure 8 c), northern and southern (Figure 8 d) regions were respectively 1.24 and 11.35; the corresponding simulated ratios were 1.33, 16.79 for CRECS 2014, and 2.02, 11.98 for CFPS 2014. The simulations were good fit with the actual statistics which indicated the effectiveness of the model.

3.4 Scenario analysis

Based on the successful verification of the model, a series of scenarios were designed to observe the substitution of coal with electricity under adjustment of certain factors. For convenience, the simulation results with original data as input were considered as 1.0, and the other simulations with factors modified were taken relative values.

3.4.1 Effect of lowest annual income of household group

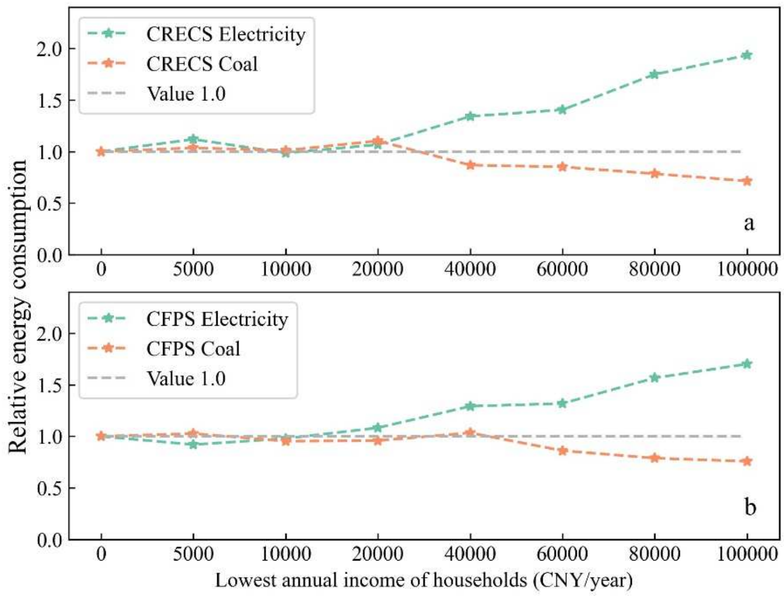

Figure 9 Coal and electricity consumption of households with different lowest annual incomes

To study the energy consumption intention of low-income households, the lowest annual income levels of household group were set to $0.5 * 10^4$, $1 * 10^4$, $2 * 10^4$, $4 * 10^4$, $6 * 10^4$, $8 * 10^4$, and $10 * 10^4$ $(CNY/year)$ based on CRECS 2014 (Figure 9 a) and CFPS 2014 (Figure 9 b) data. The relative electricity and coal consumption were mainly observed (Figure 9). The results showed that the households tended to use more electricity and less coal as the lowest annual incomes increased, and the trend appeared to be non-linear. Moreover, the coal-to-electricity conversion became obvious only when the lowest annual income exceeded a certain threshold (about 20000 CNY/year).

3.4.2 Effect of various single factors

Based on the substitution mechanism proposed in Section 2.3.3, the effect of factors such as average annual income growth rate of households (AAIGR), subsidies provided by government for households substituted their coal devices with electric devices, thermal efficiency of electric devices (TEED), and building heating demands percentage of the original (BHDPO) were simulated (Figure 10). The BHDPO assumed that, the building heating demand was declined, benefiting from the improved envelope structure or the isolation of cold air. The coal consumption relative to the original condition (RCC) were mainly observed. The AAIGR, subsidies, TEED, BHDPO were assumed to be 0–30% with intervals 5% (Figure 10 a), 0–2800 CNY with intervals 400 CNY (Figure 10 b), 1.0–4.0 with intervals 0.5 (Figure 10 c), 100–20% with intervals 10% (Figure 10 d), respectively. The simulated results showed that except for the building heating demand, optimizing of other factors all lead to obvious decreases of RCC, and the downward trends were different.

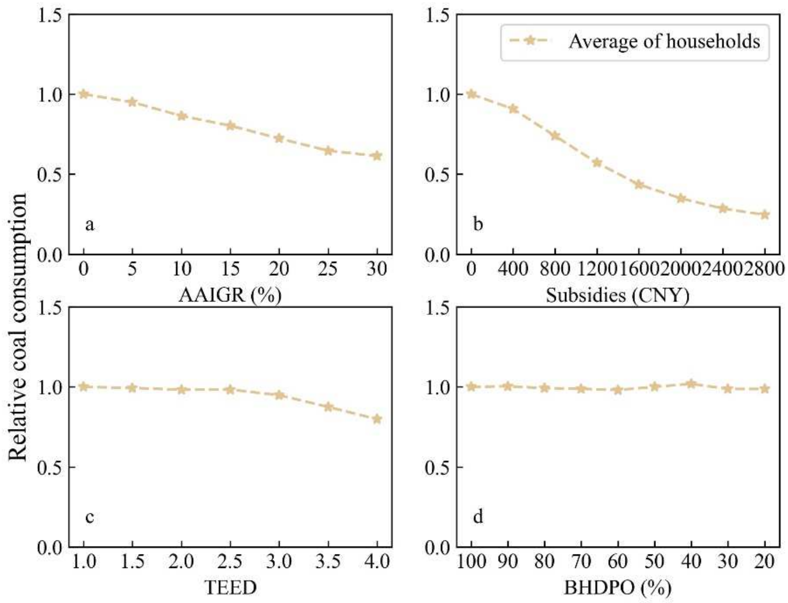

Figure 10 Effects of single factors on household energy consumption

3.4.3 Comprehensive effects of multiple factors

In this section, the comprehensive effects of multiple factors on coal consumption for household heating were simulated. The factor combinations included TEED & AAIGR (Figure 11 a), TEED & subsidies (Figure 11 b), BHDPO & TEED (Figure 11 c), BHDPO & AAIGR (Figure 11 d), and BHDPO & TEED & AAIGR (Figure 13). The RCC were then mainly observed. The TEED, AAIGR, subsidies, BHDPO were assumed to be 1.0–3.0 with intervals 0.5, 0–20% with intervals 5%, 0–2000 CNY with intervals 400 CNY, 100–20% with intervals 10%, respectively.

The results indicated that the decreased RCC caused by factor A could be further decreased by factor B, which can be defined as the enhancement of factor B based on factor A, such as the enhancement of AAIGR (Figure 11 a) and subsidies (Figure 11 b) based on TEED . However, the enhancements of different factors showed big difference. Notably, it was uncertain whether the BHDPO could result in slight decreasing trend of RCC (Figure 10 d), but the TEED and AAIGR could bring about obvious enhancement based on BHDPO, but with different modes. The enhancements of TEED could not change the curve shape of RCC declining trend over BHDPO (Figure 11 c), but the enhancements of AAIGR obviously changed the curve shape (Figure 11 d) which seemed to be non-linear.

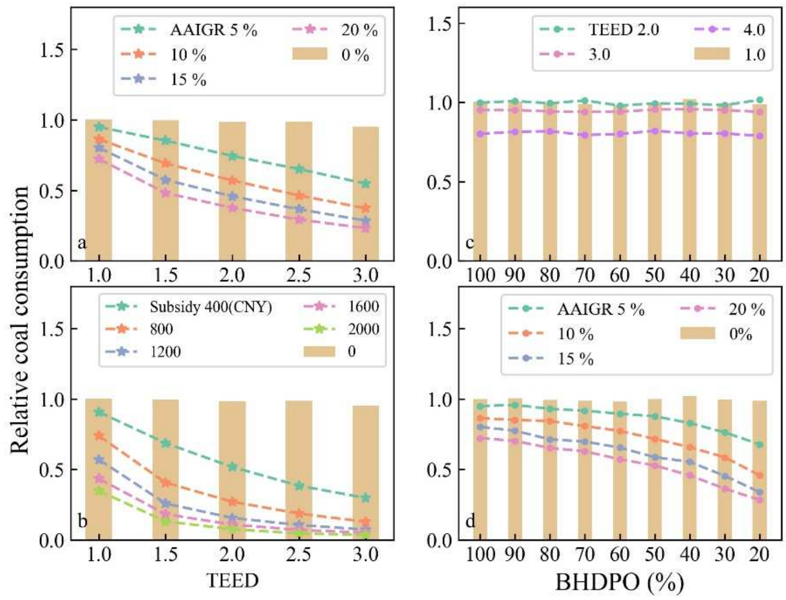

Figure 11 Comprehensive effects of AAIGR / Subsidy & TEED, and TEED / AAIGR & BHDPO

In Figure 12 a and b, the TEED was respectively assumed to be 1.5 and 2.0, the BHDPOs and AAIGRs were both assumed to be 100–20% with intervals 10%, and 0–30% with intervals 10%. The enhancement of AAIGR based on BHDPO was further enhanced by TEED.

To further analyze the reduction of RCC, the TEDD was assumed to be 1.5, and the reduced RCC brought about by different BHDPOs relative to BHDPO 100% under different AAIGR were calculated (Figure 12 c). As the AAIGR was introduced, the reduced RCC brought about by BHDPO became more obvious. Moreover, the increment of AAIGR could further enhance the RCC reduction, but it seemed that the enhancement of AAIGR emerged only after the BHDPO exceeded a threshold (about 60%).

In Figure 12 d, the TEDD was assumed to be 1.5, and the reduced RCC brought about by higher AAIGRs (10%, 20%, 30%) relative to lower AAIGRs (5%, 15%, 25%) were calculated under different BHDPO s, i.e. the marginal utility of income growth. The enhancement of RCC reduction brought by higher AAIGRs first increased then decreased as the BHDPO decreased. Furthermore, the peak value of enhancements occurred earlier with a higher AAIGR.

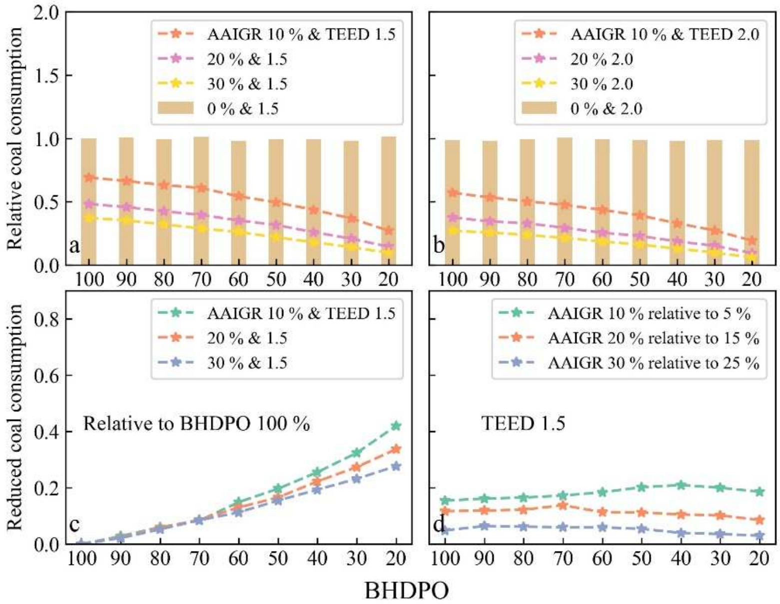

Figure 12 Comprehensive effect of three factors on household energy consumption

3.5 Discussion

The simulation results indicate that the proposed methods have substantial potential for exploring the household energy consumption behaviors.

Firstly, the model was validated. After hundreds of trainings, the deep learning model achieved satisfactory results in predicting various household energy consumption behavior parameters. Then, the trained deep learning model was embedded in the ABM framework to predict the heating behavior of the household agent. The comparison between the simulation results and the actual data revealed that the proposed model exhibits good performance.

Secondly, a series of scenario experiments were conducted to observe the effect of various factors on the household energy consumption behaviors.

1. The lowest annual income levels of CRECS 2014 and CFPS 2014 data were adjusted to initialize the model parameters. As the lowest annual income increased, households tended to use more electricity and less coal. This is consistent with the conclusions of many previous studies, such as high-income groups are inclined to use cleaner energy sources [44]. The trend was non-linear and seemed to exist a threshold. These findings may be useful for understanding the intention of certain household group to use clean energy, and the better formulation of energy substitution strategies for certain regions according to the income structure.

2. Testing the effect of a single factor. Within the scope of the simulations, higher income growth rates led to a greater reduction of coal use. Even a small amount of government subsidies significantly reduced the coal proportion; however, as the subsidies increased, the enhancement brought about by the increment part subsidies became gradually smaller. With lower thermal efficiency of electric devices, the coal proportion was reduced insignificantly; however, continuous improvement of thermal efficiency brought about an increasingly significant coal reduction. Moreover, there was no obvious effect on the proportion of coal use as the building heating demand of original was reduced lonely under the model environment. It was likely because both coal and electricity consumption was reduced as the lower heating demand, then there was no advantage for electric devices to replace coal devices as the costs of both were while down.

3. Testing the comprehensive effect of two or more factors. Generally, multiple factors led to a more pronounced decline in coal use compared to a single factor. The shorten demand of building heating could only slightly reduce coal consumption lonely, but could lead to obvious coal reduction when combined with the improvement of thermal efficiency and income growth. The results of further experiments showed the important roles of income growth in the comprehensive effect of multiple factors. The marginal utility of income growth rose first and then fell as household heating demand declined, and the peak point of marginal utility curves could arrive earlier with a higher income basis.

These results suggested that, when implementing energy strategies aimed at household level, it is necessary to learn the conditions of target group, and evaluate the overall effects of various measures comprehensively, then find the right time to achieve a trade-off between costs and expected results. The economic development level and socio-demographic characteristics of the target region should be key considerations to achieve a multiplier effect with less effort while avoiding ineffective resource inputs.

4 Conclusion

In this paper, an agent-based modeling framework for the simulation of household energy consumption was proposed. The aggregation process of energy consumption from micro-level entities to macro-level entities were modeled. Several deep learning models and a substitution mechanism was conducted to predict the complex energy consumption behaviors based on various affecting factors. Mesa and keras were taken as the main Python development toolkits. Then the results of validation and scenario simulation based on actual survey sample data, CRECS 2014 and CFPS 2014 indicated the effectiveness of our study methods.

Our research work provides high extendibility for the agent-based modeling of household energy consumption. The introduction of deep learning method can help researchers to cover much more factors to predict the household energy consumption behaviors more realistically. Various energy strategies can be conveniently simulated in the model with the help of substitution mechanism based on predicted probabilities. It is a valuable attempt to understand and quantify complex household energy consumption behaviors from the micro individual perspective.

In the future work, up to date data will be collected by various ways to get dynamic understanding of household energy consumption. Besides, social science should be introduced to model the household behaviors more comprehensively together with data-driven method used in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China [grant numbers 71673165].

We thank Professor Xinye Zheng from the Department of Energy Economics, School of Economics, Renmin University of China for the data from the China Residential Energy Consumption Survey 2014.

References

[1] Department of Energy Statistics, National Bureau of Statistics, People's Republic of China. China Energy Statistical Yearbook 2018. China Statistics Press, Beijing, 2019.

[2] Ding YX. Analysis of Spatiotemporal Variations of Household Energy Consumption in China. Lanzhou University, 2017.

[3] Natural Resources Defense Council. China Dispersed Coal Governance Report 2017, http://coalcap.nrdc.cn/. accessed 2 Jul 2021.

[4] H. Molin, Z. Jia, X. Chao, D. Baoguo, J. Dexiang, China Scattered Coal Consumption Map and Influence Factors, Zhongguo Dianli. 51 (2018) 139–146.

[5] S. Hori, K. Kondo, D. Nogata, H. Ben, The determinants of household energy-saving behavior: Survey and comparison in five major Asian cities, Energy Policy. 52 (2013) 354–362. https://doi.org/10.1016/j.enpol.2012.09.043.

[6] Natural Resources Defense Council. China Dispersed Coal Governance Report, 2019. http://nrdc.cn/Public/uploads/2018-09-04/5b8e2ed43720b.pdf. accessed 2 Jul 2021.

[7] D.B. Rahut, B. Behera, A. Ali, Household energy choice and consumption intensity: Empirical evidence from Bhutan, Renewable and Sustainable Energy Reviews. 53 (2016) 993–1009. https://doi.org/10.1016/j.rser.2015.09.019.

[8] K. Gram-Hanssen, Efficient technologies or user behaviour, which is the more important when

reducing households' energy consumption?, Energy Efficiency. 6 (2013) 447–457. https://doi.org/10.1007/s12053-012-9184-4.

[9] L.G. Swan, V.I. Ugursal, Modeling of end-use energy consumption in the residential sector: A review of modeling techniques, Renewable and Sustainable Energy Reviews. 13 (2009) 1819–1835. https://doi.org/10.1016/j.rser.2008.09.033.

[10] Z. Wang, Z. Zhao, B. Lin, Y. Zhu, Q. Ouyang, Residential heating energy consumption modeling through a bottom-up approach for China's Hot Summer–Cold Winter climatic region, Energy and Buildings. 109 (2015) 65–74. https://doi.org/10.1016/j.enbuild.2015.09.057.

[11] L.G. Swan, V.I. Ugursal, Modeling of end-use energy consumption in the residential sector: A review of modeling techniques, Renewable and Sustainable Energy Reviews. 13 (2009) 1819–1835. https://doi.org/10.1016/j.rser.2008.09.033.

[12] A.J. Summerfield, R.J. Lowe, T. Oreszczyn, Two models for benchmarking UK domestic delivered energy, Building Research & Information. 38 (2010) 12–24. https://doi.org/10.1080/09613210903399025.

[13] D.R. Carlson, H. Scott Matthews, M. Bergés, One size does not fit all: Averaged data on household electricity is inadequate for residential energy policy and decisions, Energy and Buildings. 64 (2013) 132–144. https://doi.org/10.1016/j.enbuild.2013.04.005.

[14] M. Kavgic, A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, M. Djurovic-Petrovic, A review of bottom-up building stock models for energy consumption in the residential sector, Building and Environment. 45 (2010) 1683–1697. https://doi.org/10.1016/j.buildenv.2010.01.021.

[15] M. Moglia, S. Cook, J. McGregor, A review of Agent-Based Modelling of technology diffusion with special reference to residential energy efficiency, Sust. Cities Soc. 31 (2017) 173–182. https://doi.org/10.1016/j.scs.2017.03.006.

[16] G. del C.N. Guerrero, G. Korevaar, H.H. Hansen, Z. Lukszo, Agent-Based Modeling of a Thermal Energy Transition in the Built Environment, Energies. 12 (2019) 856. https://doi.org/10.3390/en12050856.

[17] H.C. Putra, C.J. Andrews, J.A. Senick, An agent-based model of building occupant behavior during load shedding, Build. Simul. 10 (2017) 845–859. https://doi.org/10.1007/s12273-017-0384-x.

[18] M. Hu, J. Wen, F. Li, M. Haghnevis, Y. Khodadadegan, L.M. Sanchez, S. Wang, X. Zhuang, T. Wu, An Agent Based Simulation for Building Energy System Modeling, in: ASME 2010 Dynamic Systems and Control Conference, Volume 2, ASMEDC, Cambridge, Massachusetts, USA, 2010: pp. 247–254. https://doi.org/10.1115/DSCC2010-4176.

[19] Z. Wang, R. Paranjape, Agent-Based Simulation of Home Energy Management System in Residential Demand Response, in: 2014 Ieee 27th Canadian Conference on Electrical and Computer Engineering (Ccece), Ieee, New York, 2014.

[20] J. Sachs, Y. Meng, S. Giarola, A. Hawkes, An agent-based model for energy investment decisions in the residential sector, Energy. 172 (2019) 752–768. https://doi.org/10.1016/j.energy.2019.01.161.

[21] C. Wang, K. Zhou, L. Li, S. Yang, Multi-agent simulation-based residential electricity pricing schemes design and user selection decision-making, Nat. Hazards. 90 (2018) 1309–1327. https://doi.org/10.1007/s11069-017-3096-8.

[22] M. Moglia, S. Cook, J. McGregor, A review of Agent-Based Modelling of technology diffusion with special reference to residential energy efficiency, Sust. Cities Soc. 31 (2017) 173–182. https://doi.org/10.1016/j.scs.2017.03.006.

[23] M. Moglia, A. Podkalicka, J. McGregor, An Agent-Based Model of Residential Energy Efficiency Adoption, JASSS. 21 (2018) 3. https://doi.org/10.18564/jasss.3729.

[24] M. Zheng, C.J. Meinrenken, K.S. Lackner, Agent-based model for electricity consumption and storage to evaluate economic viability of tariff arbitrage for residential sector demand response, Applied Energy. 126 (2014) 297–306. https://doi.org/10.1016/j.apenergy.2014.04.022.

[25] C. Wang, D. Yan, H. Sun, Y. Jiang, A generalized probabilistic formula relating occupant behavior to environmental conditions, Building and Environment. 95 (2016) 53–62. https://doi.org/10.1016/j.buildenv.2015.09.004.

[26] H. Shi, Q. Ma, N. Smith, F. Li, Data-Driven Uncertainty Quantification and Characterization for Household Energy Demand Across Multiple Time-Scales, IEEE Transactions on Smart Grid. 10 (2019) 3092–3102. https://doi.org/10.1109/TSG.2018.2817567.

[27] H. Shi, Q. Ma, N. Smith, F. Li, Data-Driven Uncertainty Quantification and Characterization for Household Energy Demand Across Multiple Time-Scales, IEEE Transactions on Smart Grid. 10 (2019) 3092–3102. https://doi.org/10.1109/TSG.2018.2817567.

[28] C. Chatfield, “Model uncertainty,” in Encyclopedia of Environmetrics. New York, NY, USA: Wiley-Blackwell, 2006.

[29] D. Yan, W. O’Brien, T. Hong, X. Feng, H.B. Gunay, F. Tahmasebi, A. Mahdavi, Occupant behavior modeling for building performance simulation: Current state and future challenges, Energy and Buildings. 107 (2015) 264–278. https://doi.org/10.1016/j.enbuild.2015.08.032.

[30] S. D’Oca, T. Hong, J. Langevin, The human dimensions of energy use in buildings: A review, Renewable & Sustainable Energy Reviews. 81 (2018) 731–742. https://doi.org/10.1016/j.rser.2017.08.019.

[31] S. Tian, S. Chang, An agent-based model of household energy consumption, Journal of Cleaner Production. 242 (2020) 118378. https://doi.org/10.1016/j.jclepro.2019.118378.

[32] J. Widén, E. Wäckelgård, A high-resolution stochastic model of domestic activity patterns and electricity demand, Applied Energy. 87 (2010) 1880–1892. https://doi.org/10.1016/j.apenergy.2009.11.006.

[33] E. Azar and C. Menassa, A conceptual framework to energy estimation in buildings using agent based modeling, Proceedings of the 2010 Winter Simulation Conference, 2010, pp. 3145-3156, doi: 10.1109/WSC.2010.5679007.

[34] S. Barr, A.W. Gilg, N. Ford, The household energy gap: examining the divide between habitual- and purchase-related conservation behaviours, Energy Policy. 33 (2005) 1425–1444. https://doi.org/10.1016/j.enpol.2003.12.016.

[35] G. Branco, B. Lachal, P. Gallinelli, W. Weber, Predicted versus observed heat consumption of a low energy multifamily complex in Switzerland based on long-term experimental data, Energy and Buildings. 36 (2004) 543–555. https://doi.org/10.1016/j.enbuild.2004.01.028.

[36] V. Fabi, R.V. Andersen, S. Corgnati, B.W. Olesen, Occupants’ window opening behaviour: A literature review of factors influencing occupant behaviour and models, Building and Environment. 58 (2012) 188–198. https://doi.org/10.1016/j.buildenv.2012.07.009.

[37] A.-L. Lindén, A. Carlsson-Kanyama, B. Eriksson, Efficient and inefficient aspects of residential

energy behaviour: What are the policy instruments for change?, Energy Policy. 34 (2006) 1918–1927. https://doi.org/10.1016/j.enpol.2005.01.015.

[38] K. Zhou, S. Yang, Understanding household energy consumption behavior: The contribution of energy big data analytics, Renewable & Sustainable Energy Reviews. 56 (2016) 810–819. https://doi.org/10.1016/j.rser.2015.12.001.

[39] Target Encoding Done the Right Way. https://maxhalford.github.io/blog/target-encoding/. accessed 2 Jul 2021.

[40] Zheng A, Casari A. Feature Engineering for Machine Learning. O'Reilly Media, Inc. 2018.

[41] Scikit-learn: Machine Learning in Python. https://scikit-learn.org/. accessed 2 Jul 2021.

[42] Imbalanced-learn 0.7.0: Toolbox for Imbalanced Dataset in Machine Learning. https://pypi.org/project/imbalanced-learn/. accessed 2 Jul 2021.

[43] Zheng XY, Wei C, Yu YH, et al. Chinese Household Energy Consumption Report 2016. Science Press [M], 2017.

[44] D.B. Rahut, B. Behera, A. Ali, Household energy choice and consumption intensity: Empirical evidence from Bhutan, Renewable and Sustainable Energy Reviews. 53 (2016) 993–1009. https://doi.org/10.1016/j.rser.2015.09.019.