**City-level household energy consumption and a clustered typology in China: A machine learning based approach**

Yixuan ZHANGa, Yifan Lib, Chenrui XUc, Xi LIANGd, Yi WUd,\*

a

b College of Agriculture and Life Sciences, Cornell University, Ithaca, New York, US

c

d The Bartlett School of Sustainable Construction, University College London, London WC1E 7HB, UK

\* Corresponding author: *Yi WU, University College London, y\_wu.21@ucl.ac.uk*

**Abstract**

[background]

[approach]

[implications]

**Keywords**: household energy consumption, energy consumption inequality, clustering approach, household typology

# 1. Introduction - 1000 words

- [ ] Cities account for a majority of emissions & household energy/lifestyle emissions

- [ ] Understanding of Chinese cities' emission and HCEs

- [ ] What approaches have been applied to explore HCEs and what findings regarding urban/rural HCEs

- [ ] What gap exists in the current researches? -- highly dependent on a fixed range of factors, and lack of deep understanding of lifestyle and behaviors of Households

# 2. Methods and Data Description - 1200 words

## 2.1 Survey samples

The data used in this paper were collected from the Chinese Residential Energy Consumption Survey 2014 (CRECS 2014), which formed the energy data section of the Chinese General Social Survey 2015 (CGSS 2015). The dataset has the advantages of containing detailed sociodemographic characteristics of households, energy use activities and appliance types. The survey data cover 3,863 households from 85 cities and 28 provinces in China including relatively even sampels from urban and rural areas (55% for rural and 45% for urban). The data also cover multiple dimensions of household characteristics, including habit of appliance use and details (type, quantity, purchase year, purchase subsidies, fuel type, energy efficiency labelling, power, time, frequency, effective area, etc.), family members’ personal information (number of family members, age, family relations, house ownership, etc.), household economic status inlcuding annual income and expenditure and so on. Eventually, in this study, variables are categorised into six major types covering the household characteristics of our interest to be included in the modelling procedures where geographic, demographic, economic, living, family-relationship, and energy-consumption variables are included.

## 2.2 Household energy consumption estimation and inequality analysis

家庭能源消费的测量是本文分析的重点，因此准确估计家庭能源消费是关键基础。根据已有文献和相关学术报告，并且通过对survey原始资料的细致整理，我们参考Li, Liu, and Zhang (2022) and Wu, Zheng, and Wei (2017)的做法并且refer to the parameters in the research report published by the National Academy of Development and Strategies, Renmin University of China，通过改进部分测量依据，对不同类别来源的家庭能源消费活动所产生的排放量进行测算，并将所有能源消费活动的结果统一转换为standard coal equivalent 单位。总体而言，单个家庭的能源消费估算可以分为三步，第一是估算能耗单位或者设备功率。以cooker为例，由于中国家庭，尤其是农村家庭的炊具在不同地区间存在较大的差异，不光体现在炊具的类别，还包括炊具所使用能源类型的差异，因此在本研究中不得不假设家庭用炊具的热容量是固定的（一般是2kw），和当时常见炊具的一般水平保持一致。相反地，电器的用能情况更容易获取，只需要按照家庭披露的功率进行换算，然而对于部分在问卷中没有披露家电功率的情况，我们通过能效标识进行辅助判断，对于两者都缺失的情况，登记为缺值；第二是估算活动量。除了供暖的情景，其他用能活动都可以通过活动频率和活动时间估算能源活动量，然而对于供暖，问卷收集了用户采暖持续的天数和供暖方式，可以进行换算。值得注意的是，对于日常生活中频繁使用的电器，例如照明、冰箱、电视机等，我们以用户每年在家的天数作为设备实际使用天数进行估算；第三是识别每一类能源活动涉及的能源类型和实际对应的折标煤系数。我们综合参考了中国国家统计局官方发布的能源统计年鉴所列的折标煤系数(NBS, 2023)以及在其它学术报告(Li et al., 2022; Zheng et al., 2016)中引用的参数值，更新了截至调查数据发布年份之前的实际参数予以计算。

(1)

where represents household 's total energy consumption with the unit kgce, i.e. kilograms of coal equivalent). denotes household 's energy consumption activity of energy type from activity and represents the coal equivalent coefficient of energy type . Additionally, taking the example of electric appliances, the estimation is simplified to be the product of unit power (energy consumption), use time, use frequency and coal equivalent coefficient of electricity, which is as follows:

(2)

where is the sum of electric appliances' energy consumption. denote unit power or unit energy consumption of appliance , use time, use frequency and coal equivalent coefficient of electricity respectively. Nevertheless, the study tried hard to retain relatively accurate data estimation by including every detailed differences inherent in each energy consumption activity and the full description about the method for household energy consumption estimation is available in the **S1 Supplementary Material**.

Notably, to simplify the analysis of category-specific energy consumption modes and merge data with relatively small values, we compress the orginal ten categories of energy consumption activities into six categories, which means energy consumptions from air conditioners, laundry machines, televisions, computers, and lighting equipments are summed up and the new category is labelled as *appliance*.

In addition, inequalities related to energy consumption have long been a concern (Dou, Zhao, Dong, & Dong, 2021; Ma, Xu, Li, Zhang, & Sun, 2021; Shi, 2019). We are interested in the energy consumption inequality between the urban and rural, and in different energy consumption activities. Following the idea of using the Gini coefficient and Lorenz curve to measure energy consumption inequality (Wu et al., 2017), we calculate the Gini coefficient of energy consumption by taking household as the unit. Besides, we calculate the Gini coefficient seaparately for different energy consumption activities to illustrate the potential inequality.

## 2.4 Machine learning approaches - ZYX

Following the six major categories of variabels in the survey, they will be transformed into categorical or numeric variables for machine learning processing.

[insert a table showing the categories and variables]

**Table 1**. Variables applied in the modelling processing

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variables** | **Count** | **Mean** | **Min** | **Max** | **Converted** |
| **Geographic** |  |  |  |  | × |
| prefecture |  |  |  |  |  |
| region (north/south) |  |  |  |  |  |
| **Demographic** |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| **Economic** |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| **Living** |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| **Family-relationship** |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| **Energy-consumption** |  |  |  |  | √ |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

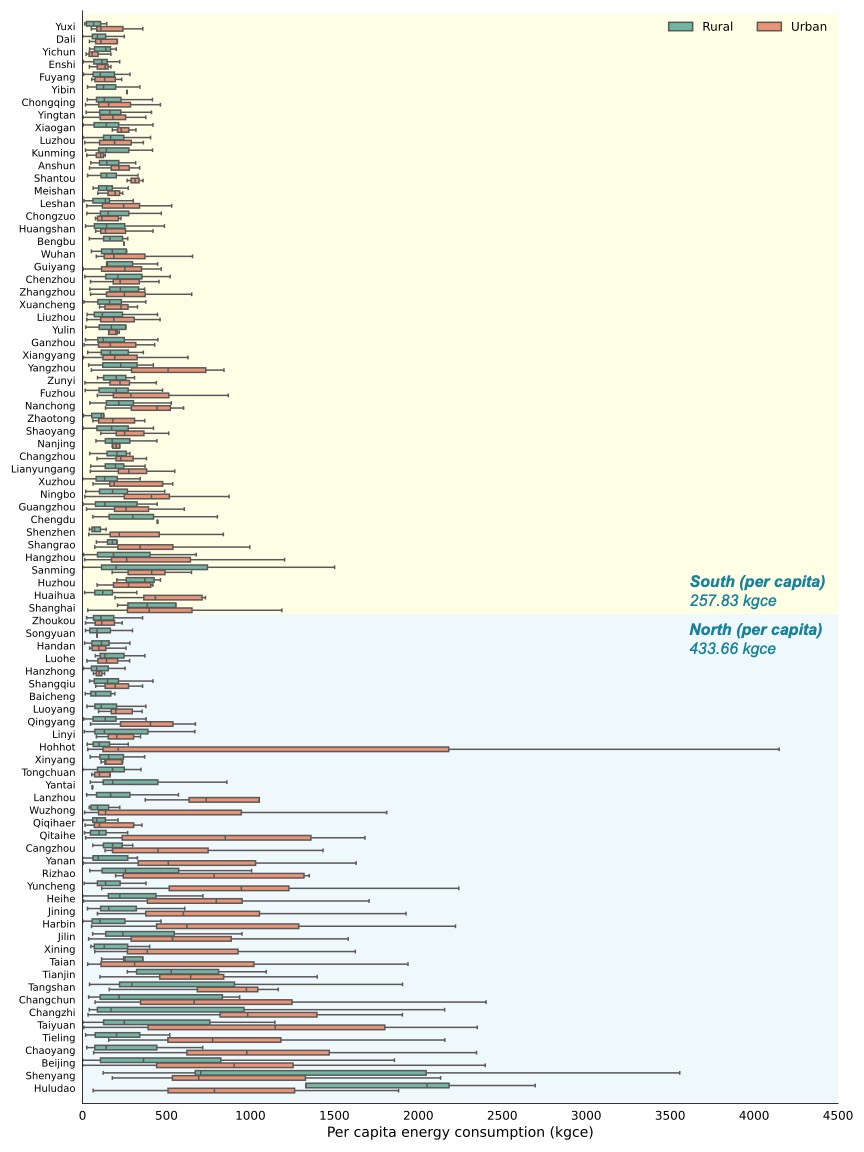
**Source**: the authors.

其中，哪些变量做了处理...

# 3. Results

## 3.1 Household energy consumption disparity across cities

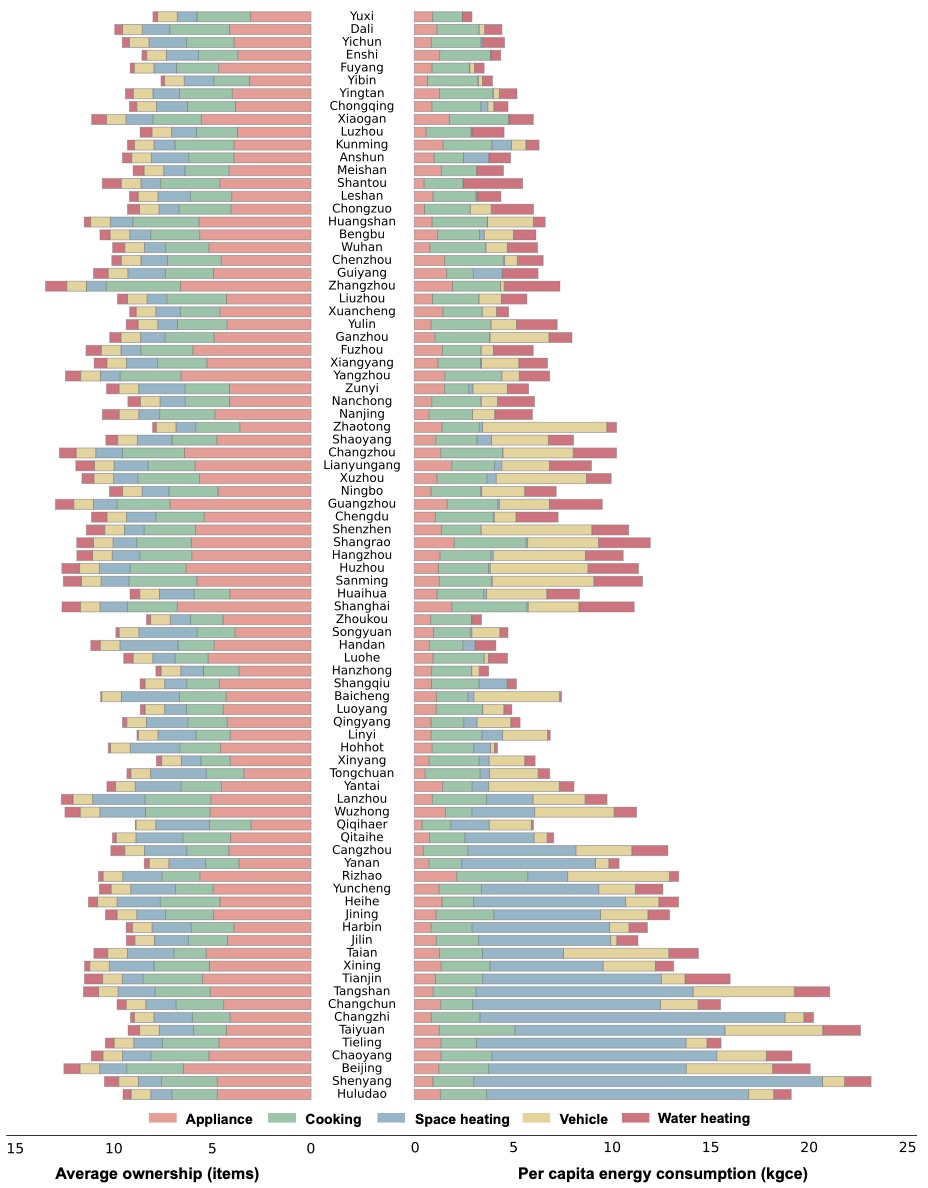
在这一节中，我们强调家庭能源消费在城市层面的差异情况。图1描述了每个城市中家庭人均能源消费的分布情况，以及在区分了城乡两个组别之后，家庭人均能源消费的分布差异。总体上，中国南方和北方的家庭人均能源消费



**Fig. 1**. **Per capita energy consumption across cities ordered by regions.** The boxplot shows the statistics within cities. Cities in the yellow area are southern cities while those in the blue area are northern cities.

**Source**: the authors.

首先，图2保持着和图1一样的城市顺序，为了方便我们对比南北城市之间潜在的能耗类别的差异。



**Fig. 2. Category specific per capita energy consumption and average ownership of appliances across cities.** Colors in the chart represent a specific category of energy consumption in household's daily life and there are six categories in total (appliance for lightred, cooking for lightgreen, space heating for blue, vehicle for yellow, water heating for darkred).

**Source**: the authors.

## 3.2 Household energy consumption inequality

已有的研究认为，家庭能源消费的不平等是广泛存在的， 不仅在不同的城市间、小镇间，包括在不同分组之间。



**Fig 3. Lorenz curves by regions and energy consumption activities**. **Fig. (a)**, Lorenz curve of HEC by urban/rural and north/south regions. **Fig. (b)**, Lorenz curve of HEC by energy consumption acitivities including appliance, cooking, space heating, vehicle and water heating. The diagonal is the line of perfect equality. The numbers in the parentheses are the Gini coefficients.

## 3.3 A machine learning-based HCE typology - ZYX



**Fig. 4. Lorenz curves by clustered groups. Fig. (a)**, Lorenz curve of HEC by clustered groups in all samples. **Fig. (b)**, Lorenz curve of HEC by clustered groups in rural samples. **Fig. (c)**, Lorenz curve of HEC by clustered groups in urban samples. The diagonal is the line of perfect equality. The numbers in the parentheses are the Gini coefficients. The legends represent the clustered groups including: Low-income cooking-demand eldelry family (LCDE), Southern temperature-demand family (STD), Northern heating-demand family (NHD), Higher income younger stayout family (HYS), Northern elderly family (NE), High-income car-owner family with children (HCC).

# 4. Discussion

本文通过重新测算了家庭层面的能源消费情况，并且细致整理了家庭能源消费活动的细粒度变量，构造了一个两步聚类的建模框架用于

[Policy implication] 根据分组之后整体组别的GINI系数有效改善的情况来看，按照生活方式和家庭本身的能源消费特征进行分组的逻辑是正确且有效的，这也表明“一城一策、分类施策”的政策倾向可能是更加有效的。在考虑分类施策的过程中，值得注意的是不仅需要关注趋势相同的家庭类型，更要关注有离群趋势的家庭类型，因为这一类家庭中可能包含着潜在的暴露在能源贫困风险中的家庭。

[Limitation] 和现有采用CGSS和估算家庭能源消费的研究相似，本研究也面临一些局限性，包括一些有限的结论基于比较weak和没有广泛论证的estimation paramter selection。In general, 局限性可以被归纳为3个方面，第一数据采样的准确度可能存在瑕疵，导致估算得到的家庭能源消耗可能有遗漏；第二由于关键数据的缺失，比如更具体的能源种类和用量，使得从家庭用能活动去推算能源的消耗量存在挑战；第三中国南北地区有着明显的自然条件差异、经济发展水平差异和生活习惯的差异，尽管我们的建模框架已经能够处理部分偏差，但仍可能存在遗漏的变量使得聚类的结果有偏误。

# References

Dou, Y., Zhao, J., Dong, X., & Dong, K. (2021). Quantifying the impacts of energy inequality on carbon emissions in China: A household-level analysis. *Energy Economics, 102*. doi:10.1016/j.eneco.2021.105502

Li, Z., Liu, Y., & Zhang, Z. (2022). Carbon Emission Characteristics and Reduction Pathways of Urban Household in China. *Frontiers in Environmental Science, 10*. doi:10.3389/fenvs.2022.896765

Ma, S., Xu, X., Li, C., Zhang, L., & Sun, M. (2021). Energy consumption inequality decrease with energy consumption increase: Evidence from rural China at micro scale. *Energy Policy, 159*. doi:10.1016/j.enpol.2021.112638

NBS. (2023). *China Energy Statistical Yearbook*. Retrieved from

Shi, X. (2019). Inequality of opportunity in energy consumption in China. *Energy Policy, 124*, 371-382. doi:10.1016/j.enpol.2018.09.029

Wu, S., Zheng, X., & Wei, C. (2017). Measurement of inequality using household energy consumption data in rural China. *Nature Energy, 2*(10), 795-803. doi:10.1038/s41560-017-0003-1

Zheng, X., Wei, C., Song, F., Xie, L., Qin, P., Ma, B., . . . Fu, J. (2016). *Research Report on China’s Household Energy Consumption 2015*. Retrieved from