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

Yixuan ZHANGa, Yifan Lib, Xi LIANGc, Yi WUc,\*

**Abstract**

**Keywords**: household energy consumption, clustering approach, urban-rural divide, 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 - YW

To estimate household energy consumption, this paper takes three steps to do the calculation. First of all,

值得注意的是，为了便于下文的分析，我们简化了家庭能耗的类别通过将来自空调、洗衣机、电视机、个人电脑和照明器具的能耗合并为家电，因此原本10个类别被减少为6个类别。

Detailed description about the method for household energy consumption estimation is available in the **S1 Supplementary Material**.

## 2.3 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** |  |  |  |  | √ |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

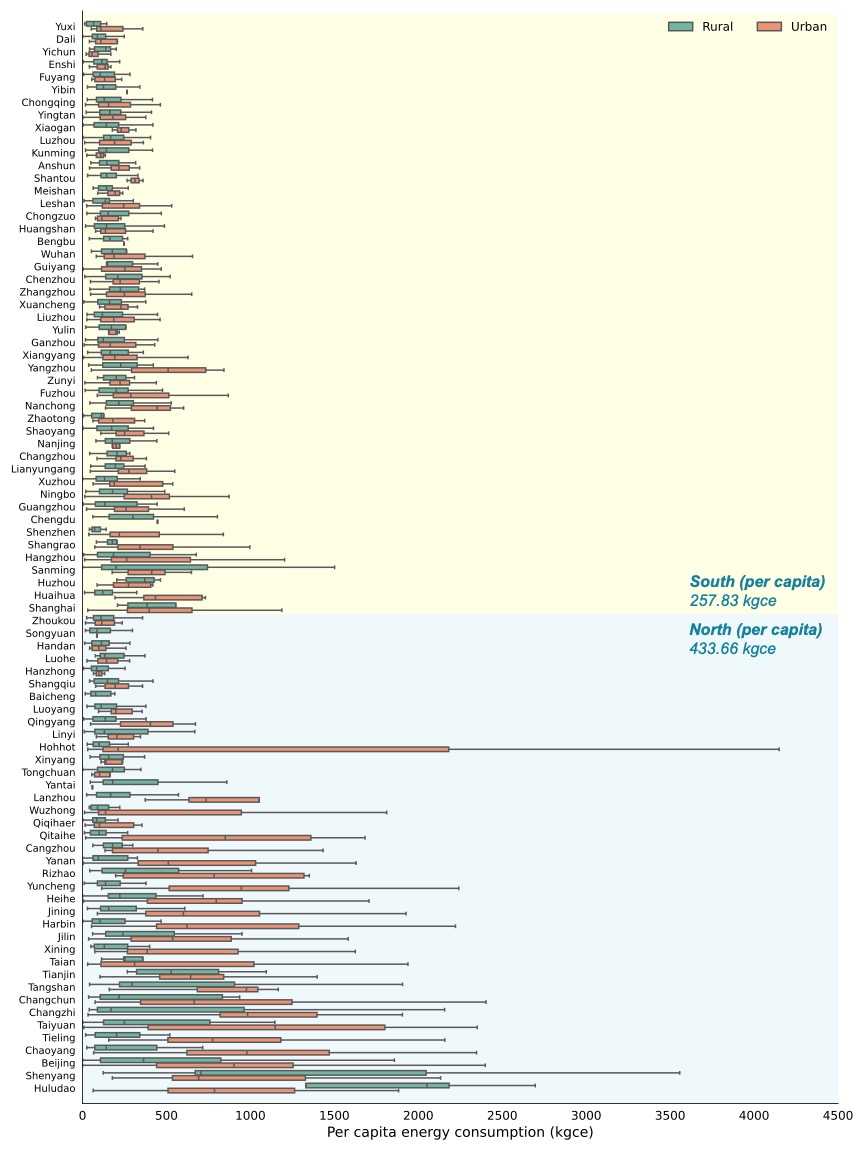
**Source**: the authors.

其中，哪些变量做了处理...

# 3. Results

## 3.1 Household energy consumption disparity across cities

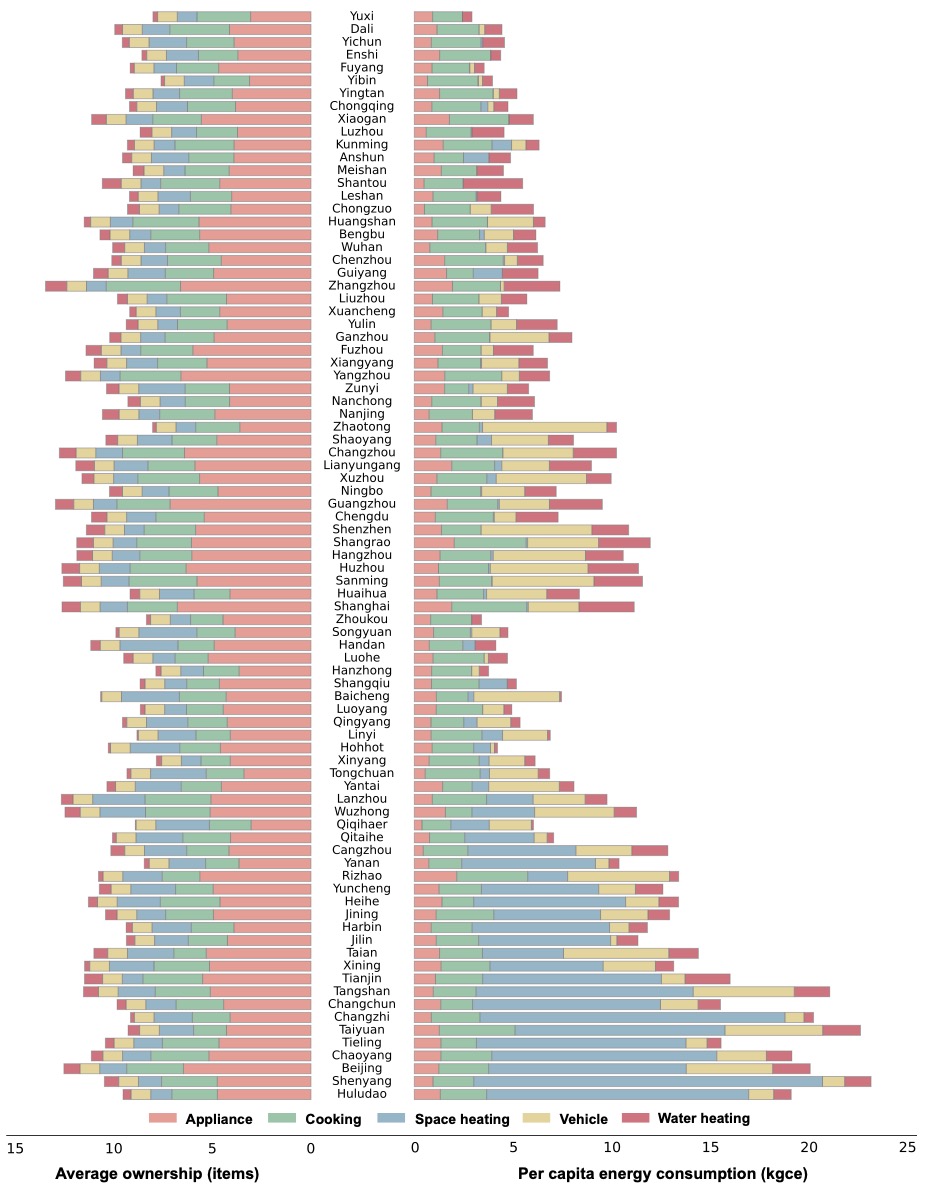
在这一节中，我们计算了



**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 - YW & ZYX

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

# 4. Discussion

[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

# References