Big Data to support sustainable urban energy planning: The EvoEnergy project

Moulay Larbi CHALAL (✉), Benachir MEDJDOUB, Nacer BEZAI, Raid SHRAHILY

School of Architecture, Design, and the Built Environment, Nottingham Trent University, Nottingham, UK
E-mail: Moulay.Chalal@ntu.ac.uk

Moulay Larbi CHALAL, Benachir MEDJDOUB, Nacer BEZAI, Raid SHRAHILY

Big Data to support sustainable urban energy planning: The EvoEnergy project

Abstract   Energy sustainability is a complex problem that needs to be tackled holistically by equally addressing other aspects such as socio-economic to meet the strict CO_2 emission targets. This paper builds upon our previous work on the effect of household transition on residential energy consumption where we developed a 3D urban energy prediction system (EvoEnergy) using the old UK panel data survey, namely, the British household panel data survey (BHPS). In particular, the aim of the present study is to examine the validity and reliability of EvoEnergy under the new UK household longitudinal study (UKHLS) launched in 2009. To achieve this aim, the household transition and energy prediction modules of EvoEnergy have been tested under both data sets using various statistical techniques such as Chow test. The analysis of the results advised that EvoEnergy remains a reliable prediction system and had a good prediction accuracy (MAPE ; 5%) when compared to actual energy performance certificate data. From this premise, we recommend researchers, who are working on data-driven energy consumption forecasting, to consider merging the BHPS and UKHLS data sets. This will, in turn, enable them to capture the bigger picture of different energy phenomena such as fuel poverty; consequently, anticipate problems with policy prior to their occurrence. Finally, the paper concludes by discussing two scenarios of EvoEnergy development in relation to energy policy and decision-making.

Keywords   urban energy planning, sustainable planning, Big Data, household transition, energy prediction

1 Introduction

The UK residential sector accounts for the second largest share of the UK's total energy (30%) and emits around 14% of the country's carbon dioxide (Department for Business, Energy & Industrial Strategy (BEIS), 2017a). This justifies why this sector is playing a central role in the UK decarbonisation framework. In particular, the UK government has recently placed a special emphasis on building retrofit due to the aged (75% built before 1975) and leaky nature of the existing dwelling stock (Edwards and Townsend, 2011). As around 68% of the UK dwelling stock benefited from retrofitting measures, the implementation of physical strategies will be less effective, challenging, and expensive in the long term (BEIS, 2017b). Thus, local authorities should explore alternative measures based on other aspects such as behavioral and socio-economic, which are also responsible for 4%–30% of the variation in domestic energy consumption patterns (Brounen et al., 2012; Jones et al., 2015). However, before developing such strategies (e.g., socio-economic), it is important to determine their degree of influence on residential energy usage. This will not only facilitate the decision-making process with regards to the suitability of available policies but also help estimate their impact prior to implementation.

The effect of socio-economic, behavioral, physical, psychological factors on domestic energy consumption has been extensively reviewed in the literature (Druckman and Jackson, 2008; Abrahamse and Steg, 2009; Frederiks et al., 2015; Longhi, 2015). In our previous study (Medjdoub and Chalal, 2017),

we contributed to this body of literature by investigating the impact of household demographic transitions on their energy consumption patterns. Our findings suggested that household transition patterns from one family type to another have a significant effect on their domestic energy usage. This has led to questioning the non-consideration of this concept in urban energy planning, especially after knowing that it constitutes an important determinant of consumer purchasing behavior in other disciplines (e.g., marketing). By knowing the future transitions probabilities of consumers to different family structures over their lifecycle, marketers could anticipate their needs and determine the services and products that are suitable for them and in a proactive manner (Du and Kamakura, 2006; Hitesh, 2018). In response to the lack of energy prediction tools supporting household lifecycle transition patterns, we have previously developed a 3D urban energy prediction system (EvoEnergy) (Medjdoub and Chalal, 2017). EvoEnergy has the ability to predict domestic energy at the urban scale in function of (1) household transitions from one family type to another (e.g., from single to couple without children, etc.), and (2) the variation in the household socio-economic and demographic factors. The predicted energy figures are then mapped onto a given area in a color-coded manner with the help of a GIS (Geographic Information System) module.

Although incorporating the concept of lifecycle household transitions seems promising as it promotes proactive energy planning, it could be argued that tools relying on such concept (e.g., EvoEnergy) might be prone to validation issues associated with their long-term usage. This could be partially attributed to the fact that algorithms of such tools, which have been developed using specific data sets (e.g., the British household panel data survey (BHPS)), may not reflect current demographic and socio-economic changes of an evolving population. To this end, the research presented in this paper builds upon our previous work (Medjdoub and Chalal, 2017) on the effect of household transition patterns on domestic energy consumption by evaluating the performance of EvoEnergy under the new UK household longitudinal study (UKHLS). To achieve this aim, the following objectives have been addressed:

- The validation of household transition models from Medjdoub and Chalal (2017) under a new UKHLS.

- The comparison of the effect of household transition on domestic energy usage based on both data sets (BHPS and UKHLS).

- The validation of the performance of the energy prediction algorithm using both data sets.

Ensuring the validity and reliability of our 3D urban energy prediction system (EvoEnergy) under the new UKHLS facilitates its adoption in urban energy planning. By doing so, EvoEnergy can assist energy decision-makers to: (1) Predict the long-term variation in the residential energy consumption of urban districts in function of changes in the household demographic and socio-economic profiles; (2) enable proactive management of the energy grid to meet the demand levels; (3) facilitate the development of policies that target specific groups (e.g., low-income lone parents); and (4) estimate the long-term impact of a particular policy on a population segment prior to its implementation.

In addition to the above benefits, making our 3D urban energy prediction system accessible to consumers in the near future (on-going research) would help them: (5) Understand their actual and future energy usage patterns; (6) raise their pro-environmental awareness; and (7) engage in energy saving activities.

The rest of the article is structured as follows. First, section 2 gives an overview of studies that have analyzed the impact of various factors on residential energy consumption. This section ends with a brief insight into EvoEnergy published previously by Medjdoub and Chalal (2017) to help the reader acquaint with the topic. Section 3 discusses the research methodology. Sections 4 and 5 analyze the research findings. Section 6 concludes the article and Section 7 gives future recommendations.

2 Previous work

Please note that physical factors affecting residential energy are outside the scope of this paper. For more information, please refer to our previous extensive review (Chalal et al., 2016).

2.1 The effect of psychological factors

While socio-economic factors have a prominent role in predicting the household energy usage patterns, a number of psychological factors were found to have a direct effect on a person's energy-related behavior

(Yang et al., 2016). These factors include environmental awareness, beliefs, culture, values and attitudes, preferences, subjective norms, and intentions and goals (Huebner et al., 2016; Guo et al., 2018). For example, Abrahamse and Steg (2009) advised that psychological factors are related to energy savings but not the actual energy consumption. Similarly, Vringer et al. (2007) did not find a significant relationship between the variable values and domestic energy usage. As for environmental awareness, few studies (Barr et al., 2005; Steg and Vlek, 2009) advised that superior environmental awareness levels are associated with higher energy savings and lower energy consumption levels. However, such relationship is usually either weak or insignificant. Other scholars such as Khosrowpour et al. (2018) concentrated on developing feedback strategies to tackle the knowledge and environmental awareness gap; consequently, induce changes in household energy consumption. For example, Faruqui et al. (2010) suggested that implementing direct energy feedback in the form of in-home display (IHD) can contribute to 14% and 7% savings in the electricity usage of households who are on prepayment and direct debit schemes, correspondingly.

In addition to the above, many studies suggested that a person's attitudes, values, and intentions to engage in pro-environmental behavior have an impact on domestic energy usage. However, such an effect does not necessarily translate to a conforming change in energy consumption or savings (Bamberg and Möser, 2007; Huebner et al., 2013). For instance, Kavousian et al. (2013) surprisingly discovered that households who expressed an interest in purchasing energy efficient appliances consumed high levels of daily minimum energy consumption.

Likewise, personal comfort can have a significant impact on residential energy usage. More precisely, any possible decrease in personal comfort could reduce the likelihood of engaging in energy conservation activities (Gatersleben et al., 2002). For example, Barr et al. (2005) found that 40% of householders with "good pro-environmental" behavior were not willing to sacrifice their comfort to save energy. On the other hand, the percentage of those who unwilling to compromise their comfort to save energy among the "non-environmentalist" group was more than 75%.

2.2 The effect of socio-economic factors

Longhi (2015) used the UK household longitudinal study (UKHLS) to analyze the change in the household energy expenditure in function of various socio-economic and demographic factors. The study suggested that socio-economic factors explain 11% of the variation in domestic energy use. Similarly, Huebner et al. (2016) analyzed the energy follow up survey (EFUS) which encompasses a sample size of 845 English households and discovered that socio-economic variables explained around 21% of the variability in electricity consumption. Brounen et al. (2012) Advised that demographic and socio-economic factors are responsible for 17% and 5% of the variation in gas and electricity energy consumption, respectively. The above studies advised that household size was the most influential factor on domestic energy consumption. In particular, Longhi (2015) found that one additional member in the household contributes to 33%–35% decrease in the per capita energy expenditure. However, many empirical studies (Bedir et al., 2013; Jones and Lomas, 2015), which included household size as a continuous variable in their prediction models, showed that there is a positive relationship between this variable and the amount of energy consumed in the dwelling. Other factors such as age of household reference person (HRP), income, presence of children, level of education, and tenure mode, were also found to influence the variation in energy consumption (Pereira et al., 2019). However, their significance and magnitude are still inconsistent in the literature. For example, Nair et al. (2010) and BRE (2013) suggested that the energy usage of household reference persons (HRPs) aged between 50 and 65 is high, whereas the ones aged above 65 is low. On the other hand, Tiwari (2000) showed that householders aged below 45 are usually associated with lower energy consumption. Other studies such as Abrahamse and Steg (2009), Poortinga et al. (2004), and Bedir et al. (2013) found the effect of age on residential energy usage not significant. Recently, we have found that household transition patterns from one family to another does have a significant effect on their energy consumption patterns (Medjdoub and Chalal, 2017). For example, on average, a single non-elderly household has a 53.3% chance of moving to different household types after 5 years, where the possibility of becoming a couple with children is 12.1%. The chance of consuming more than 4000 kWh of electricity annually for a single non-elderly making a transition to a couple with children over five years is 35.29%. Based on the findings of our previous study (Medjdoub and Chalal, 2017), we have developed a 3D urban energy prediction system (EvoEnergy) which will be briefly described in the below sub-section.

2.3 Overview of our 3D urban energy prediction model (EvoEnergy)

EvoEnergy was developed at the Creative and Virtual Technologies Laboratory at Nottingham Trent University in collaboration with Nottingham Energy partnership. The main intention behind its development was to provide energy planners with a smart platform that assists their sustainable energy planning decision-making. A future goal of this project is to help consumers better engage in pro-environmental behavior to reduce their home energy usage (on-going research). Since EvoEnergy prediction algorithm relies on the British household panel data survey (BHPS), it can estimate future residential energy consumption for up to 10 years. These predictions are primarily dependent on (1) household transition possibilities to other household structures and (2) the variation in their socio-economic circumstances (e.g., income and age).

EvoEnergy system architecture:

As shown in Fig. 1, the architecture of EvoEnergy comprises four distinct modules. First, the game-based environment represents the 3D platform where it is possible to import and interact with any 3D semantic model via the user-interface module. The 3D semantic model database module stores the different components of the CityGML (3D GIS) models in a hierarchically structured manner to ensure stable and reliable data management, and moreover, to permit data exchange (e.g., export, modify, and save) with the Game-based environment module. On the other hand, the energy related prediction modules estimate the annual energy consumption of different households based on their socio-economic module and transition probabilities to other family types. The inputs (e.g., socio-economic characteristics) and outputs of the energy related prediction modules are stored in and loaded from the 3D semantic model database via the user-interface module. Finally, the Game-based environment enables the visualization/mapping of outputs from the energy related prediction modules.
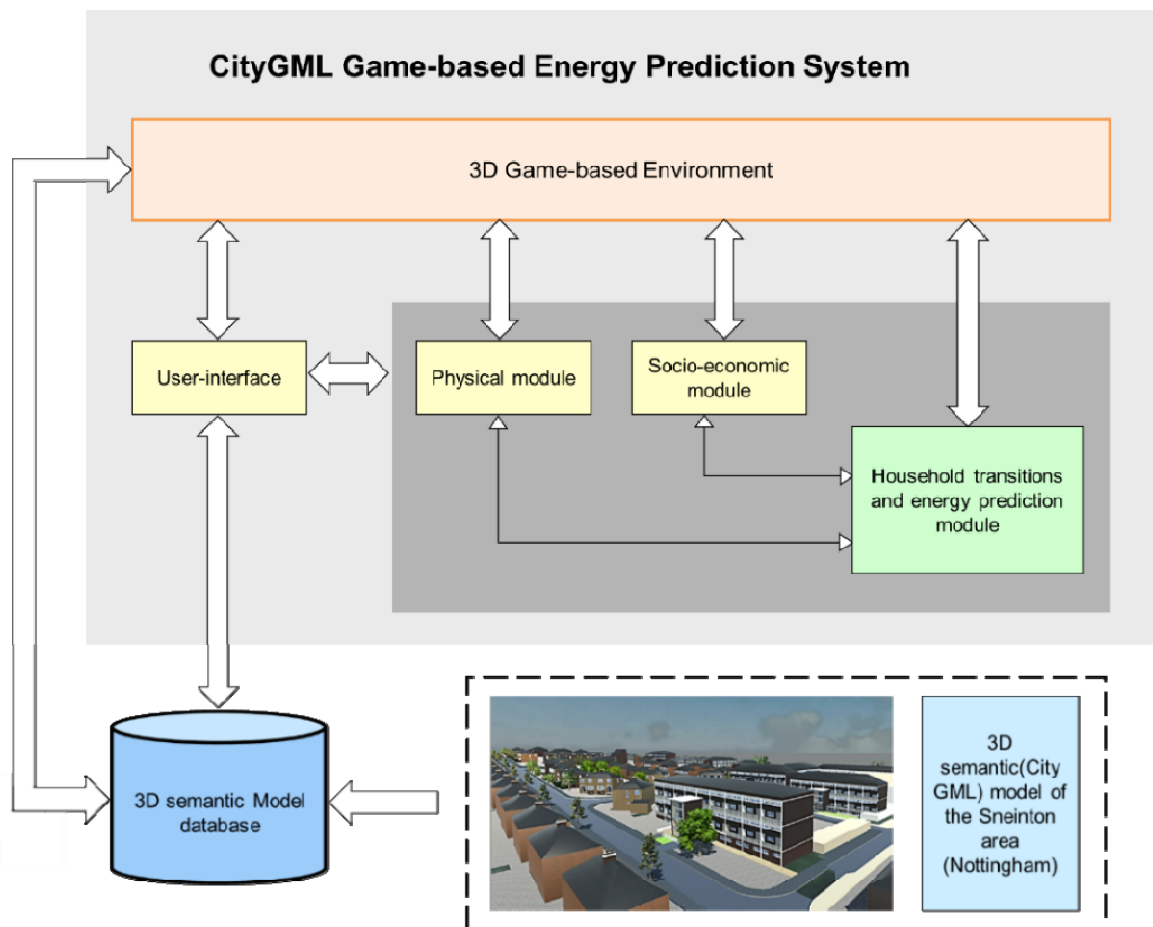

Fig. 1 Architecture of EvoEnergy.

Modus operandi:

Upon launching EvoEnergy, users can navigate through the 3D model of a particular urban area (Fig. 2) and view its energy consumption in a 2D fashion. Moreover, they can trigger a summary of energy history and socio-economic profile pertaining to a given dwelling on mouse hover (Fig. 3). To select a particular house, the user can either right-click on it or search for it using a valid address and postcode. To perform energy predictions, users are required to access the main menu and fill all the input fields in the physical and socio-economic modules. After that, they need to select the target household transition (e.g., to couple without children) and set the timeline (e.g., next two years) as shown in Fig. 4. The prediction module also allows performing meaningful comparisons between the transition patterns and energy usage patterns of different households (Fig. 5).

Fig. 2 3D model of the Sneinton area in Nottingham in EvoEnergy.

Fig. 3 Summary of a household energy history and socio-economic profile on mouse hover.
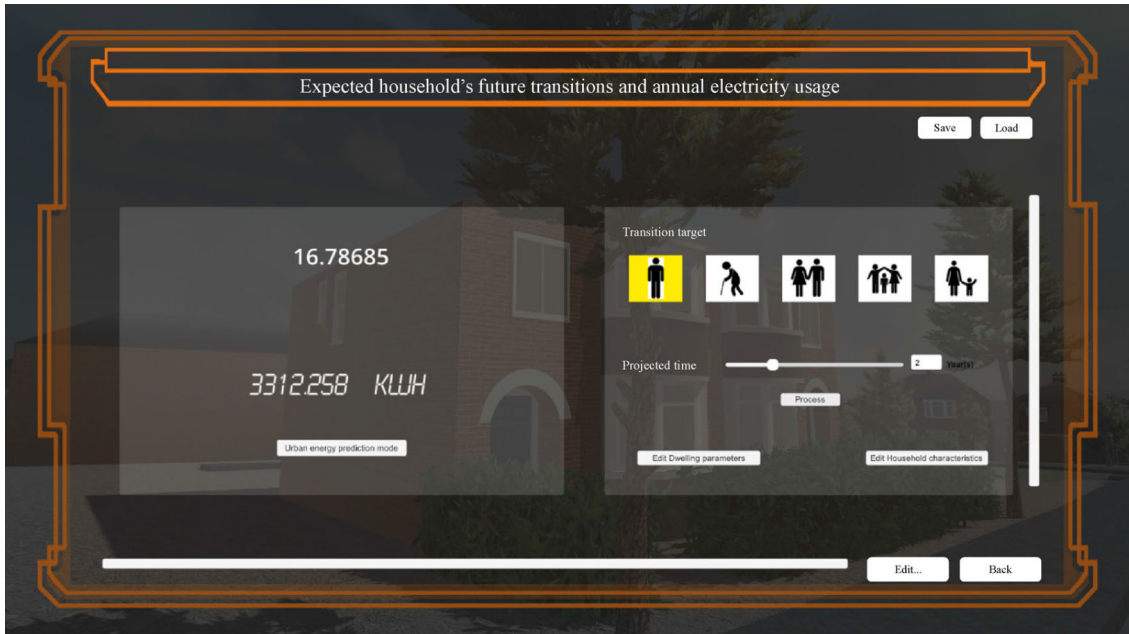
Fig. 4 Household energy prediction module.
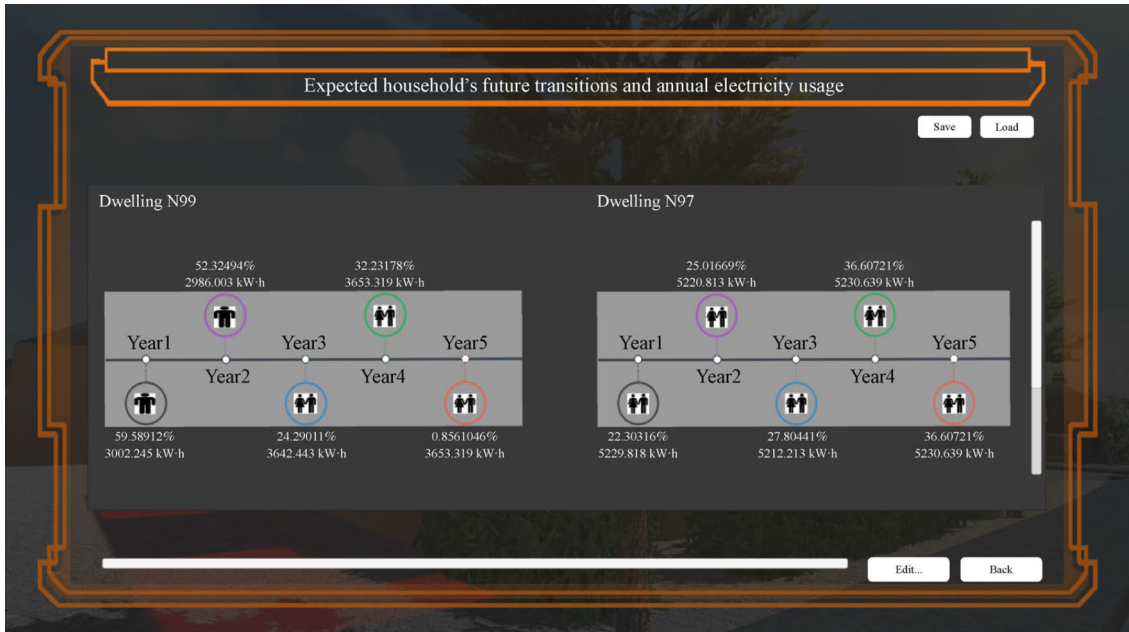
Fig. 5 Comparison of two households' transition patterns and energy consumption figures.

3 Methodology

Figure 6 represents the methodology diagram of the current study in regard to our previous research (Medjdoub and Chalal, 2017). The undertaken study embraces a mix-methods research methodology with a multi-level triangulation design. Overall, there are seven stages of implementation in total, two of which belong to the present study (orange box, Phase VI and Phase VII). However, to allow the reader to understand the link between the current and previous research, phases belonging to our previous work are briefly described below.

Fig. 6 The methodology flowchart of this research in relation to our previous work (Medjdoub and Chalal, 2017).
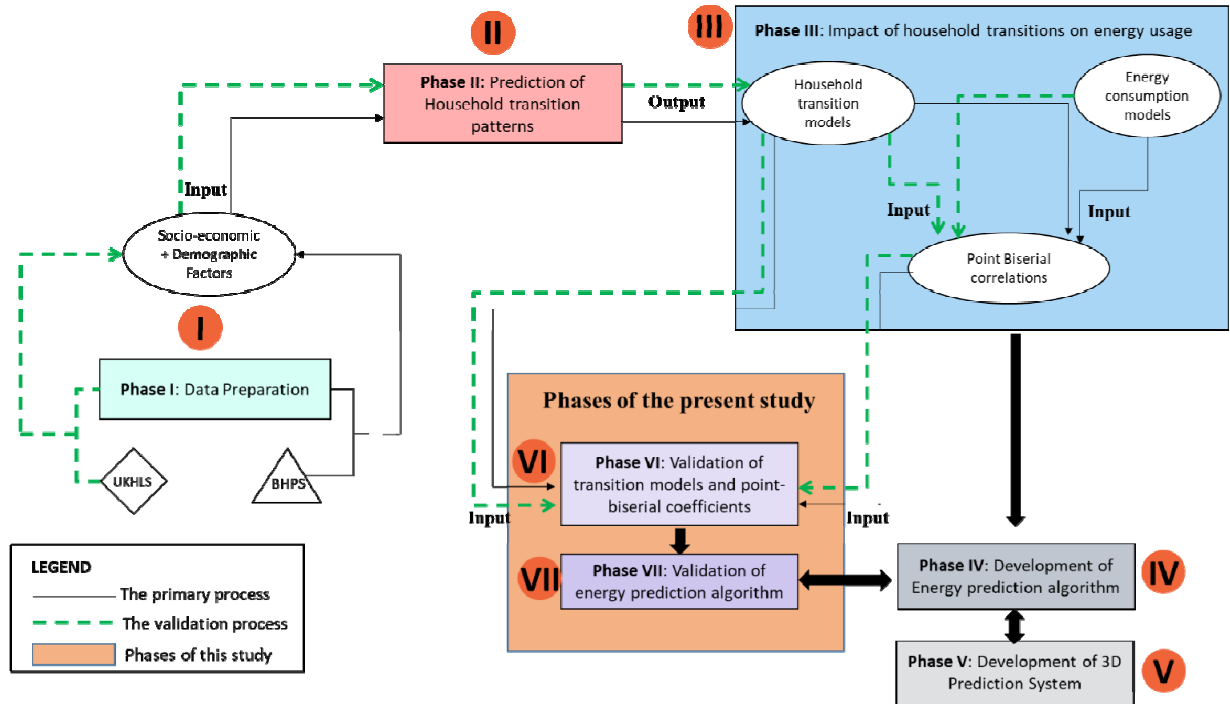
First, Phase I entails the comparison and manipulation of two distinct UK household panel data sets, namely, the British household panel data survey (BHPS) and UK household longitudinal study (UKHLS). The purpose of the manipulation is to prepare both data sets in a format, quality, and structure suitable for further analysis in phases II, IV, and VI. Phase II includes predicting household transition models using fixed and random effects binary logistic regression based on the BHPS and UKHLS data sets. As shown in Fig. 1, the prediction models resulting from the UKHLS data set will be only used for validation purposes. Phase III consists of analyzing the effect of household transition on energy consumption variables using point-biserial correlation. Conversely, Phase IV includes the development of an energy prediction model based on (1) the household demographic transition variables and (2) different socio-economic factors. The developed energy prediction model from phase IV was used to create a 3D urban energy prediction model (EvoEnergy) in Phase V (see Section 2.1). Phase VI entails comparing the prediction models and point-biserial correlation coefficients developed from the BHPS data set against the ones created based on the UKHLS data set. In phase VII, the accuracy of the energy prediction model resulting from BHPS data set will be first evaluated against the one developed based on UKHLS. In addition to comparing both prediction algorithms to each other, they will be evaluated against existing EPC (energy performance certificate) data. Details about the phases of implementation in the present study and their findings are presented in Sections 4 and 5.

3.1 Data preparation—Data description and comparison

The analyzed and compared panel data sets in this study are the British household panel data survey (BHPS) and the UK household longitudinal study (UKHLS). Both are longitudinal data surveys that encompass random UK households annually interviewed on their demographic and socio-economic circumstances in addition to other aspects such as energy expenditure (Institute for Social and Economic Research, 2016). The BHPS tracked more than 5000 households of different structures (e.g., lone parents) over 18 years between 1991 and 2008. On the other hand, the UKHLS, which is the successor of BHPS, has a significantly larger target sample size of 40000 despite starting in 2010 (Understanding Society, 2017). This, in turn, allows for a high-resolution analysis of different time-dependent events such as household demographic transitions. However, UKHLS has only 7 waves which limit the capturing of household transition patterns for more than 2–3 years.

From running several statistical including Levene test, it was found that the socio-economic and demographic profile of the BHPS and UKHLS samples were completely different from each other. Considering that age is a determinant of several socio-economic factors (e.g., income), this difference was

mainly attributed to a significant change in the sample age profiles (Fig. 7). For more information, please consult Table 4 in Appendix A.
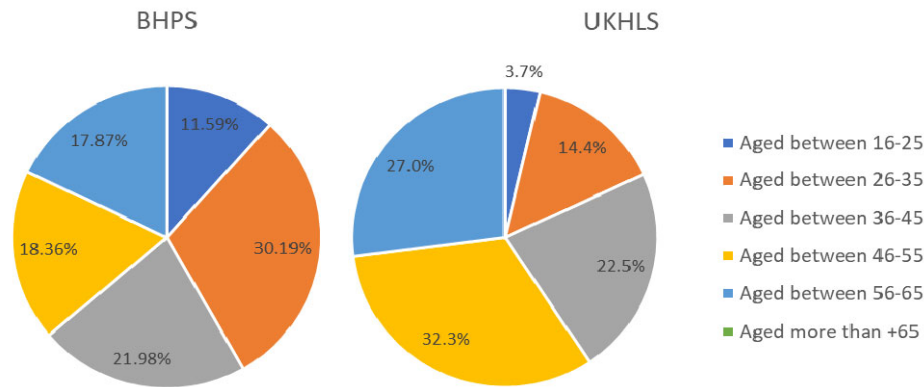

Fig. 7   Pie charts showing the distribution of age groups over the BHPS and UKHLS data sets.

3.2 Data preparation—Data manipulation

As depicted in Fig. 6, BHPS was used as the main data source in this study, whereas the UKHLS was utilized to validate the research findings. Since this work is part of our research project on Nottingham city, which has a high proportion of single non-elderly (Office for National Statistics, 2018), households who were not single non-elderly in wave 1 were omitted from both data sets. As a result, the final sample size of the BHPS was 7038 after merging all waves, except wave 6 that lacked energy expenditure variables. Conversely, the final sample size of UKHLS was 8750. The percentage of missing data in the BHPS and UKHLS data sets were only 2.35% and 4.73% of all values, respectively.

To meet the assumptions of the used statistical tests, the following data screening procedures have been applied. First, energy expenditure for gas and electricity were converted into quantities in kWh. Secondly, variables with inconsistent coding and/or number of categories across both data sets (e.g., marital status) have been recoded. After that, income, expenditure, and energy consumption variables have been normalized using log10 transformation. Finally, outliers were checked for and deleted.

4 Validation of transition models and point-biserial coefficients

To compare the household transition model coefficients, a likelihood ratio test resembling the Chow test in Stata (Eq. (1)) was used (Chow, 1960; Stata, 2015). In particular, the test will compare the coefficients and intercepts of the pooled model (combined BHPS and UKHLS) against the model comprising interaction effects between covariates and the data set dummy variable (BHPS or UKHLS). The statistics of the likelihood-ration test are defined in Eq. (1).

$$LR = -2\left(L_1 - L_0\right). \quad (1)$$

Let L_0 and L_1 be the log-likelihood values related to the pooled model (containing both data sets) and constrained models (model with interaction and main effects), correspondingly. If the constrained model is true, LR is approximately χ^2 distributed with $d_0 - d_1$ degrees of freedom, of which d_0 and d_1 are the degree of freedom pertaining to the pooled and constraints models, correspondingly (Greene, 2002).

Due to the limitation of the UKHLS in capturing household transition patterns beyond 2–3 years, the following procedures have been implemented to support the validation process. First, we compared the line graphs showing the change in the proportion of single non-elderly households over the BHPS and UKHLS waves. Moreover, the Mann–Whitney U and Kolmogorov–Smirnov Z tests were adopted to test the null hypothesis that discrepancies in the transition rates of single non-elderly over both data sets are not significantly different (McCrum-Gardner, 2008).

Findings of the validation of transition models and point-biserial coefficients:

Table 1 represents the statistics of the likelihood ratio Chow test, which compare the regression coefficients of transition models from BHPS and UKHLS. Overall, it is evident that the p-values of this

test overall all models were greater than 0.05. This signifies that there is no significant difference between the constants and coefficients of the compared household transitions models. To overcome the UKHLS limitation in capturing household transitions beyond 2–3 years, the authors have analyzed the decline in the proportion of single non-elderly households across the BHPS and UKHLS as a result of them becoming other family types such as couples without children (Fig. 8). In general, it is evident that the decrease in the proportion of single non-elderly followed the same trend across the first seven waves of both data sets, although there were minor discrepancies of 8.8% on average. In addition to that, it is expected that the decline in the percentage of single non-elderly households in the future waves of UKHLS will follow the same trend of the discontinued BHPS (1991–2008) but with minor discrepancies. To verify these findings, a Mann–Whitney U and Kolmogorov–Smirnov Z tests have been conducted (Table 2). The p-values of both tests were greater than 0.05, which means that there is no significant difference in the distribution of single non-elderly transitions over both data sets. Therefore, we can conclude that both BHPS and UKHLS are reliable data sets for predicting household transition models. However, we recommend employing either the BHPS or a combined BHPS and UKHLS data set as both scenarios allow the capturing of transition patterns for a period of at least 10 years.

Table 1 The results of the Chow likelihood test comparing the coefficients of transition models resulting from BHPS and UKHLS

Transition target	Year of transition	Model	Goodness of fit (McFadden's R^2)	obs	ll (null)	ll (model)	df	AIC	LR chi2 (6)	Prob. > chi2
Couples with children	1	Pooled model	0.626	480	−188.064	−70.40656	6	152.81	7.68	0.2625
		Model with interaction effects	0.646	480	−188.064	−66.56636	12	157.13		
	2	Pooled model	0.514	269	−104.314	−23.03278	6	58.065	9.91	0.1284
		Model with interaction effects	0.527	269	−104.314	−18.07673	12	60.15		
Couples without children	1	Pooled model	0.361	1092	−413.4964	−264.3681	7	542.73	8.18	0.3167
		Model with interaction effects	0.371	1092	−413.4964	−260.2763	14	548.55		
	2	Pooled model	0.468	693	−259.5675	−138.1586	7	290.31	10.18	0.1788
		Model with interaction effects	0.487	693	−259.5675	−133.0703	14	294.14		
Lone parents	1	Pooled model	0.683	4166	−1609.361	−510.5721	8	1037.14	6.12	0.6342
		Model with interaction effects	0.685	4166	−1609.361	−507.5138	16	1047.02		
	2	Pooled model	0.514	2690	−1015.016	−260.9618	8	537.92	3.46	0.9021
		Model with interaction effects	0.534	2690	−1015.016	−259.2309	16	550.46		

Table 2 The results of the Mann–Whitney U and Kolmogorov–Smirnov Z test statistics[a]

Mann–Whitney U Test		Kolmogorov–Smirnov Z Test		
	Transition rates			Transition rates
Mann–Whitney U	15.500	Most Extreme Differences	Absolute	0.429
Wilcoxon W	43.500		Positive	0.429
Z	−1.151		Negative	0.000
Asymp. Sig. (2-tailed)	0.250	Kolmogorov–Smirnov Z		0.802
Exact Sig. [2*(1-tailed Sig.)]	0.259[b]	Asymp. Sig. (2-tailed)		0.541

Note: a. Grouping variable: DATA; b. Not corrected for ties.

Notes: **A Wave represents the year when the panel data survey was conducted. For example, wave 1 of the BHPS was in 1991, whereas wave 1 of the UKHLS, occurred in 2009.

**The British houshold panel data survey, which was conducted from 1991 to 2008, has 18 waves. On the other hand, UKHLS, which was introduced in 2010, has currently 7 waves. Hence why, the line graph belonging to the UKHLS terminates in wave 7. However, it is expected that the decrease in the porportion of single non-elderly in future waves of the UKHLS will follow a similar trend to the one of the BHPS but with minor discrepancies.

**Wave 6 was omitted from both datasets because it lacked information about household energy expenditure.
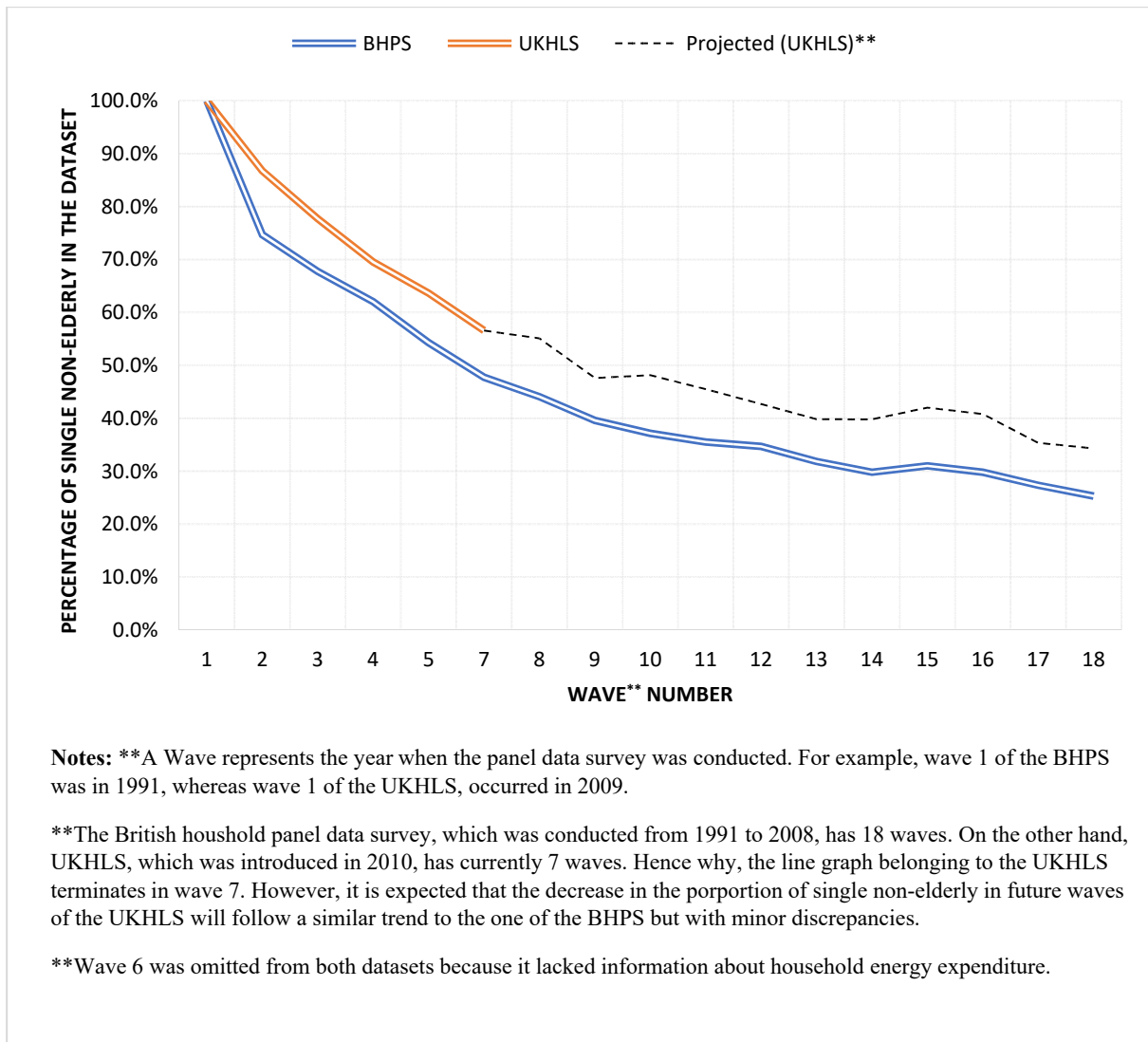
Fig. 8 The decrease in the percentage of single non-elderly households over different waves of BHPS and UKHLS as a result of them moving to other household types such as couple without children.

Following the above discussion, it was expected that the impact of household transitions on energy consumption would remain consistent under the UKHLS data set. To reinforce this claim, the point-biserial correlation coefficients resulting from BHPS and UKHLS have been compared (Table 3). From analyzing Table 3, it is evident that the point-biserial coefficients were in good agreement despite minor discrepancies of approximately 0.01 on average. The direction and significance of the point-biserial coefficients also remained consistent over both data sets. For those reasons and in line with the above recommendation on transition models, BHPS is still a reliable data set for predicting domestic energy consumption in function of household transitions. Nevertheless, using a combined UKHLS and BHPS data set represents also a viable option.

Table 3 Comparison of impact of household transition on energy consumption across BHPS and UKHLS

		LP 1 year	LP 2 years	CN 1 year	CN 2 years	CWC 1 year	CWC 2 years
Log10 annual electricity usage	BHPS	0.11**	0.12**	0.11**	0.093**	0.16**	0.13**
	UKHLS	0.114**	0.117**	0.1157**	0.1043**	0.142**	0.152**

Square root of annual gas usage	BHPS	0.008	0.005	0.114**	0.091**	0.160**	0.135**
	UKHLS	0.01	0.009	0.129**	0.0835**	0.148**	0.115**

Note: LP: Lone parent; CN: Couple without children; CWC: Couple with children.
* Significance at the 95% level; ** Significance at the 99% level

5 Validation of energy prediction algorithms

Reporting the regression coefficients of the developed energy prediction models is outside the scope of this paper. For more information, please refer to our previous work (Chalal, 2018).

As shown in Fig. 9, the accuracy of the energy prediction models developed from BHPS and UKHLS is compared to existing EPC energy data. The EPC data belongs to householders who made at least one transition from a single non-elderly family to different household structures (e.g., Couple without children) over the last 2–3 years. It is worth mentioning that the socio-economic and demographic profiles of the selected householders are distinct from each other. In this way, it is possible to test the accuracy of the prediction models at different input values.
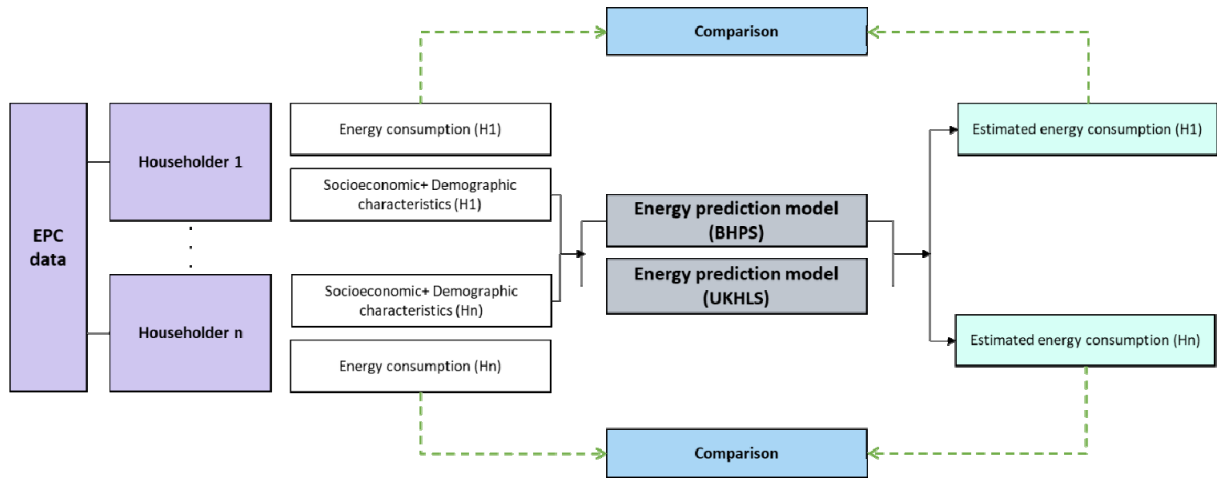

Fig. 9 The validation process of the developed energy prediction model.

The validation process starts by inputting the socio-economic and demographic characteristics of the chosen householders, including household transition possibilities, into the energy prediction algorithms developed from BHPS and UKHLS. The predicted energy values are then compared to each other and then against the actual EPC energy data (Fig. 9). Any discrepancies between the predicted and EPC energy data are reported using the mean absolute percentage error (MAPE) and mean percentage error (MPE) described below in Eqs. (2) and (3).

$$\text{MAPE} = \left(\frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \right) \times 100 \,, (2)$$

$$\text{MPE} = \left(\frac{1}{n} \sum_{t=1}^{n} \frac{A_t - F_t}{A_t} \right) \times 100 \,, (3)$$

where A_t and F_t are actual and predicted energy consumption values, respectively.

Findings of the validation of energy prediction algorithm:

Figure 10 illustrates the predicted and actual annual electricity energy figures of the selected householders. Overall, it is evident that there were some discrepancies between the estimated and actual values. More precisely, the mean absolute percentage errors (MAPE) for the BHPS and UKHLS energy models were 5.47% and 5.15%, respectively.
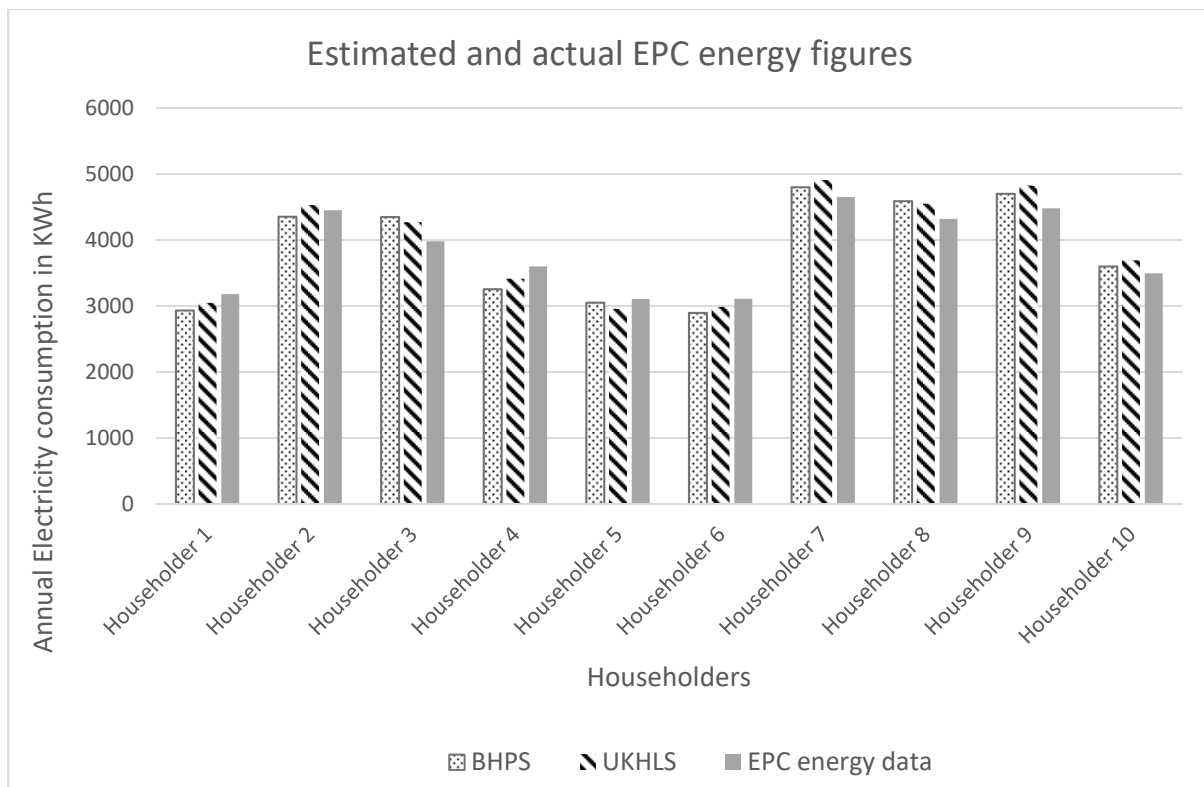
Fig. 10 Clustered bar graph representing the estimated and actual EPC energy figures of the chosen householders.

As shown in Figs. 11 and 12, the minimum and maximum mean percentage errors (MPE) for the BHPS energy model were −1.74% and −9.58%, correspondingly. On the other hand, the lowest and highest MPE values for the UKHLS energy model were 1.74% and 7.71%, respectively. This leads to the conclusion that the UKHLS energy prediction model had a superior accuracy, although there were minor discrepancies between its outputs and the ones of the BHPS energy prediction model (MAPE ; 3%). This was in line with the literature where the goodness of fit of the UKHLS electricity prediction model reported by Longhi (2015) (0.369) was superior to the BHPS one stated by Berkhout et al. (2004) (0.11) and the one we previously reported (0.25) (Chalal, 2018).

After further investigations, we found that the mean absolute percentage errors (MAPE) for householders who made a transition to lone parent, couple with children, and other family structures were higher than 6%. In particular, it seems that both energy prediction models over predict the energy usage associated with couple with children transitions. Moreover, they underestimate the consumption of those moving to a lone parent family and other family structures (e.g., 2 unrelated households). This could be due to the low representativeness of those household types in both data sets. Indeed, we found out that the MPE values associated with transitions to couple without children households, who have better representativeness in the data sets, were below 5% (Figs. 11 and 12).

In addition to the above, we have discovered that the number of transitions made by the householder negatively correlates with the prediction accuracy of the BHPS and UKHLS energy models. For example, the householder who made two transitions in which the last one was to a lone parent household had the mean percentage error (MPE) values of −9.58% (Fig. 11) and −5.23% (Fig. 12). Similarly, a householder who first moved to couple without children and then to a couple with children household had 9.30% and 7.29% mean percentage errors as depicted in Figs. 11 and 12. Surprisingly, the number of steps followed in the prediction process had a significant effect on the estimation accuracy. More precisely, following a multifold prediction process, where the annual electricity usage is predicted at each transition stage, the accuracy improved by up to 7% in comparison to a one-fold approach. The reasons behind this improvement are unknown and are currently under investigation.
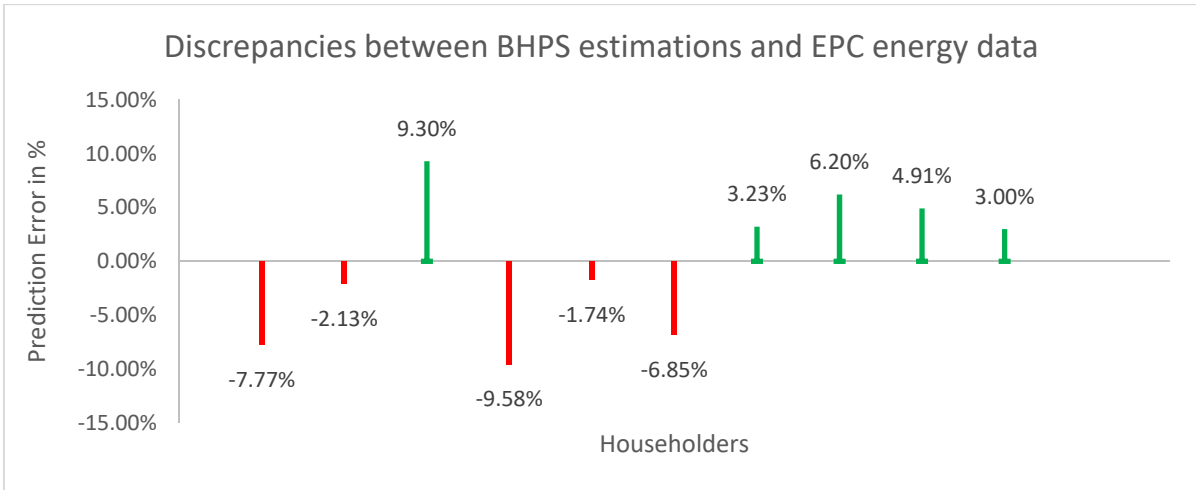
Fig. 11 Discrepancies between the estimated energy figures using BHPS and the EPC energy data reported using the mean percentage error index (MPE).
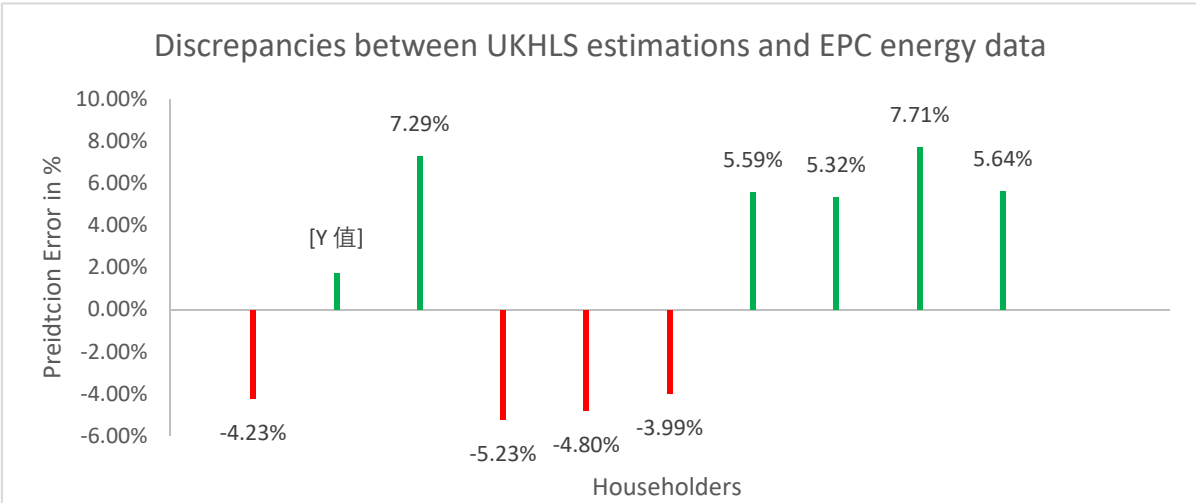

Fig. 12 Discrepancies between the estimated energy figures using UKHLS and the EPC energy data reported using the mean percentage error index (MPE).

6 Discussion and conclusions

In 2009, the British household panel data survey (BHPS) was replaced by its successor, the UK household longitudinal survey (UKHLS) (Understanding Society, 2017). Our comparison suggested that the socio-economic and demographic profiles of the BHPS and UKHLS were distinct, which indicates that the UK society has undergone some important changes from 1991 to the present. An example of these changes includes transformations in the age structure of the population, its educational attainment profile, and home ownership levels.

For the above reasons and considering that our 3D urban energy prediction system (EvoEnergy) was partly developed using the old data set (BHPS), the present research aimed to evaluate the validity and reliability of EvoEnergy under the new household panel data survey (UKHLS). To attain the study aim, we have first evaluated the transition module of EvoEnergy by comparing the coefficients of household transition models generated from the BHPS and UKHLS data sets. Following this, the energy prediction module of EvoEnergy has been tested by first comparing the impact of household transition on energy usage over both data sets. After that, the accuracy of energy prediction algorithms resulting from BHPS and UKHLS has been evaluated against existing energy performance certificate data (EPC).

The analysis of findings advised that there were no significant differences between the coefficients of transition models of both data sets. This suggests that BHPS and UKHLS are reliable sources for analyzing and forecasting dynamic relationships including household demographic transitions. In addition to that, the

analysis of point-biserial correlations over the BHPS and UKHLS data sets proved that the effect of family transition on domestic energy consumption remained consistent across the two data sets. Finally, the BHPS and UKHLS energy prediction models had a good estimation accuracy when compared to the actual EPC data (MAPE ; 5%). However, the UKHLS energy prediction algorithm had superior accuracy.

While the examination of study findings confirmed the validity and reliability of EvoEnergy as it stands, it has opened the doors to new scenarios related to its future development. The first scenario consists of utilizing the BHPS data set as the basis for EvoEnergy household transition module. Moreover, it entails employing a UKHLS based Energy prediction algorithm. Even though this would certainly improve the energy prediction accuracy of EvoEnergy, relying on BHPS data set makes EvoEnergy unable to predict the transition patterns of households with low representativeness such as lone parents beyond 7 years. Furthermore, it does not permit an adequate analysis of the effect of cultural factors on residential energy demand. This is because the BHPS data set has low representativeness of ethnic minority groups in the BHPS sample (McFall and Garrington, 2011). This would pose a problem especially if the focus of policy-makers is placed on monitoring and determining the effectiveness of policies geared toward minor ethnicities and lone parent families. An example of this includes analyzing the change in the fuel poverty gap of minor ethnicity groups in function of government schemes, their CO_2 emissions, and any changes in their socio-economic and demographic factors. From a policy point of view, using EvoEnergy under this scenario limits the monitoring, design, and adjustment of pro-environmental behavior policies and measures that target specific households over different stages of their family lifecycle.

In contrast to the above, a better scenario involves using a combined BHPS and UKHLS data set to inform the development of EvoEnergy's household transition and prediction modules. In this way, it is possible to monitor more households over a period of at least 25 years. This, in turn, permits to increase the prediction period of the transition module to 15 years. Furthermore, it would overcome the limitations of BHPS by allowing for better handling of certain household types and ethnic minority groups. Similarly, adopting a joined BHPS–UKHLS energy prediction algorithm would help correct the over- and under-estimations of couple without children and lone parent households (Section 5.1), respectively. Based on that, it is argued that using EvoEnergy under this scenario would support policy- and decision-making by addressing certain phenomena while taking account of socio-economic and demographic changes occurring over the household lifecycle. This will, in turn, enable the development of proactive measures. For example, one of the challenges facing UK policy-makers in identifying fuel poverty is the change in the socio-economic factors of the households (ADECOE, 2016). Examples of such changes include varying income levels, change in household size, deterioration of housing conditions, and change in the fuel price. Using EvoEnergy in this situation could possibly help anticipate the likelihood of being a fuel pauvre in function of scenarios of change in the household socio-economic circumstances.

7 Recommendations for future work

The analysis of the research findings has highlighted few limitations, which should be addressed in the future. These can be summarized as follows:

• The process of combining the BHPS and UKHLS is challenging and time consuming. Therefore, there is a need for tools that could automate or facilitate this process for other scholars, especially those with little/basic statistical knowledge.

• Since EvoEnergy is only confined to the UK residential sector, one of the possibilities is to extend its socio-economic module to cover different countries such as Germany, Italy, and Spain.

Appendix A

Table 4 represents the summary of the homogeneity of variance analysis of different socio-economic and demographic variables over the BHPS and UKHLS data sets. Overall, it is evident that the variance of most factors across the two data sets was heterogeneous except for the following variables: Gender, aged 36–45, divorced, widowed, never married, separated, living as a couple, A-level, rented from employer, rented from private landlord, living in terraced houses, and living in 3-bedroom dwellings. This implies that the socio-economic and demographic characteristics of both samples are largely significantly different from each other.

Table 4 Summary of the homogeneity of variance analysis of different socio-economic and demographic variables across the BHPS and UKHLS data sets

	Homogeneity of variance	If partially heterogeneous, what category is homogenous/heterogeneous?
Gender	Homogeneous	N/A
Age	Mostly heterogeneous	Aged 36–45 is homogeneous
Marital status	Mostly homogeneous	Married is heterogeneous
Level of education	Mostly heterogeneous	A-Level is homogeneous
Tenure mode	Mostly heterogeneous	Rented from employer and private landlords are homogeneous
Dwelling type	Mostly homogenous	Living in terraced dwelling is heterogeneous
Dwelling size	Mostly heterogeneous	3-bedroom dwelling is homogeneous
Socio-economic class	Mostly heterogeneous	Professional occupations and unskilled workers are homogeneous

References

Abrahamse W, Steg L (2009). How do socio-demographic and psychological factors relate to households' direct and indirect energy use and savings? Journal of Economic Psychology, 30(5): 711–720 doi:10.1016/j.joep.2009.05.006

ADECOE (2016). Is UK's approach to fuel poverty suffering from a poverty of ideas? Available at: adecoe.co.uk/2016/03/10

Bamberg S, Möser G (2007). Twenty years after Hines, Hungerford, and Tomera: A new meta-analysis of psycho-social determinants of pro-environmental behaviour. Journal of Environmental Psychology, 27(1): 14–25 doi:10.1016/j.jenvp.2006.12.002

Barr S, Gilg A W, Ford N (2005). The household energy gap: Examining the divide between habitual- and purchase-related conservation behaviours. Energy Policy, 33(11): 1425–1444 doi:10.1016/j.enpol.2003.12.016

Bedir M, Hasselaar E, Itard L (2013). Determinants of electricity consumption in Dutch dwellings. Energy and Building, 58: 194–207 doi:10.1016/j.enbuild.2012.10.016

Department for Business, Energy & Industrial Strategy (BEIS) (2017a). Energy consumption in the UK. 2017 ed. London: Crown

Department for Business, Energy & Industrial Strategy (BEIS) (2017b). Household Energy Efficiency headline release. 2017 ed. London: Crown

Berkhout P H G, Ferrer-i-Carbonell A, Muskens J C (2004). The ex post impact of an energy tax on household energy demand. Energy Economics, 26(3): 297–317 doi:10.1016/j.eneco.2004.04.002

BRE (2013). Report 3: Metered fuel consumption—Including annex on high energy users. In: Energy Follow Up Survey, 2011. London: Department of Energy and Climate Change

Brounen D, Kok N, Quigley J (2012). Residential energy use and conservation: Economics and demographics. European Economic Review, 56(5): 931–945 doi:10.1016/j.euroecorev.2012.02.007

Chalal M (2018). A Smart Urban Energy Prediction System to Support Energy Planning in the Residential Sector. Dissertation for the Doctoral Degree. Nottingham: Nottingham Trent University

Chalal M, Benachir M, White M, Shrahily R (2016). Energy planning and forecasting approaches for supporting physical improvement strategies in the building sector: A review. Renewable & Sustainable Energy Reviews, 64: 761–776 doi:10.1016/j.rser.2016.06.040

Chow G C (1960). Tests of equality between sets of coefficients in two linear regressions. Econometrica, 28(3): 591–605 doi:10.2307/1910133

Druckman A, Jackson T (2008). Household energy consumption in the UK: A highly geographically and socio-economically disaggregated model. Energy Policy, 36(8): 3177–3192 doi:10.1016/j.enpol.2008.03.021

Du R Y, Kamakura W A (2006). Household life cycles and lifestyles in the United States. Journal of Marketing Research, 43(1): 121–132 doi:10.1509/jmkr.43.1.121

Edwards J, Townsend A (2011) CIOB carbon action 2050: Buildings under refurbishment and retrofit. Bracknell, UK: The Chartered Institute of Building (CIOB)

Faruqui A, Sergici S, Sharif A (2010). The impact of informational feedback on energy consumption—A survey of the experimental evidence. Energy, 35(4): 1598–1608 doi:10.1016/j.energy.2009.07.042

Frederiks E R, Stenner K, Hobman E V (2015). The socio-demographic and psychological predictors of residential energy consumption: A comprehensive review. Energies, 8(1): 573–609 doi:10.3390/en8010573

Gatersleben B, Steg L, Vlek C (2002). Measurement and determinants of environmentally significant consumer behavior. Environment and Behavior, 34(3): 335–362 doi:10.1177/0013916502034003004

Greene W H (2002). Econometric Analysis. Upper Saddle River, New Jersey: Prentice Hall

Guo Z F, Zhou K L, Zhang C, Lu X H, Chen W, Yang S L (2018). Residential electricity consumption behavior: Influencing factors, related theories and intervention strategies. Renewable & Sustainable Energy Reviews, 81(1): 399–412 doi:10.1016/j.rser.2017.07.046

Hitesh B (2018). Family life-cycle. Marketing Management Articles

Huebner G, Shipworth D, Hamilton I, Chalabi Z, Oreszczyn T (2016). Understanding electricity consumption: A comparative contribution of building factors, socio-demographics, appliances, behaviours and attitudes. Applied Energy, 177: 692–702 doi:10.1016/j.apenergy.2016.04.075

Huebner G M, Cooper J, Jones K (2013). Domestic energy consumption—What role do comfort, habit, and knowledge about the heating system play? Energy and Building, 66: 626–636 doi:10.1016/j.enbuild.2013.07.043

Institute for Social and Economic Research (2016). British Household Panel Survey (BHPS)

Jones R V, Fuertes A, Lomas K J (2015). The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings. Renewable & Sustainable Energy Reviews, 43: 901–917 doi:10.1016/j.rser.2014.11.084

Jones R V, Lomas K J (2015). Determinants of high electrical energy demand in UK homes: Socio-economic and dwelling characteristics. Energy and Building, 101: 24–34 doi:10.1016/j.enbuild.2015.04.052

Kavousian A, Rajagopal R, Fischer M (2013). Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior. Energy, 55: 184–194 doi:10.1016/j.energy.2013.03.086

Khosrowpour A, Jain R K, Taylor J E, Peschiera G, Chen J, Gulbinas R (2018). A review of occupant energy feedback research: Opportunities for methodological fusion at the intersection of experimentation, analytics, surveys and simulation. Applied Energy, 218: 304–316 doi:10.1016/j.apenergy.2018.02.148

Longhi S (2015). Residential energy expenditures and the relevance of changes in household circumstances. Energy Economics, 49: 440–450 doi:10.1016/j.eneco.2015.03.018

McCrum-Gardner E (2008). Which is the correct statistical test to use? British Journal of Oral & Maxillofacial Surgery, 46(1): 38–41 doi:10.1016/j.bjoms.2007.09.002 PMID:17961892

McFall S L, Garrington C (2011). Understanding society: Early findings from the first wave of the UK's Household Longitudinal Study. Colchester, Essex: Institute for Social and Economic Research, University of Essex

Medjdoub B, Chalal M L (2017). Impact of household transitions on domestic energy consumption and its applicability to urban energy planning. Frontiers of Engineering Management, 4(2): 171–183 doi:10.15302/J-FEM-2017029

Nair G, Gustavsson L, Mahapatra K (2010). Factors influencing energy efficiency investments in existing Swedish residential buildings. Energy Policy, 38(6): 2956–2963 doi:10.1016/j.enpol.2010.01.033

Office for National Statistics (2018). Population estimates for UK, England and Wales, Scotland and Northern Ireland: Mid-2016

Pereira D S, Marques A C, Fuinhas J A (2019). Are renewables affecting income distribution and increasing the risk of household poverty? Energy, 170: 791–803 doi:10.1016/j.energy.2018.12.199

Poortinga W, Steg L, Vlek C (2004). Values, environmental concern, and environmental behavior: A study into household energy use. Environment and Behavior, 36(1): 70–93 doi:10.1177/0013916503251466

Stata (2015). Longitudinal-Data/Panel-Data Reference Manual, Release 14. College Station, TX: Stata Press

Steg L, Vlek C (2009). Encouraging pro-environmental behaviour: An integrative review and research agenda. Journal of Environmental Psychology, 29(3): 309–317 doi:10.1016/j.jenvp.2008.10.004

Tiwari P (2000). Architectural, demographic, and economic causes of electricity consumption in Bombay. Journal of Policy Modeling, 22(1): 81–98 doi:10.1016/S0161-8938(98)00003-9

Understanding Society (2017). Understanding Society—The UK Household Longitudinal Study. The Institute for Social and Economic Research, University of Essex

Vringer K, Aalbers T, Blok K (2007). Household energy requirement and value patterns. Energy Policy, 35(1): 553–566 doi:10.1016/j.enpol.2005.12.025

Yang S, Zhang Y, Zhao D (2016). Who exhibits more energy-saving behavior in direct and indirect ways in China? The role of psychological factors and socio-demographics. Energy Policy, 93: 196–205 doi:10.1016/j.enpol.2016.02.018