

TECHNICAL ARTICLE

City-wide Building Energy Efficiency Assessment Using EPC Data

Yue Wu, Luke Blunden and AbuBakr Bahaj

The Energy Performance Certificate (EPC) requirements in Europe have provided rich data of energy efficiency status of individual buildings at scale, allowing appropriate assessments to be carried out without the need of conducting additional surveys. By 2012, the EPC data in England and Wales in the UK has reached a coverage of up to 45% of the entire building stock (DECC 2014), and the number of new certificates added every year is over 1 million. The work presented here seeks to understand the potential of using such dataset to deliver systematic analysis of energy efficiency of buildings at a city scale. This study utilises all EPC datasets available for the city of Southampton, UK, and developed an automatic geographic-referencing methodology so the EPC data can be integrated onto geographic information systems (GIS). Such methodology allows large quantity of EPC data to be analysed spatially, facilitating the assessment of energy performance of different regions in a city. The results show clearly areas in the city that have the lowest or highest level of performance in terms of buildings' energy efficiency, and revealed the impact of certain building intervention measures (such insulation, double glazing) on the energy consumption within the considered buildings. The analyses show that factors such as the existence of high streets and new building developments have significant influence on the energy efficiency in the city where these are present. The geographic-referencing methodology developed in this study has the potential to be utilised in a wide range of subjects, bridging the current gap of data availability, allowing building related datasets to be incorporated with GIS analysis.

Keywords: Energy performance certificate; Building; GIS; Energy efficiency; Geographic Information System

1. Introduction

The UK EPC scheme was established in 2007 as a response to the European Union Directive 2002/91/EC (Council of the European Union 2002), which aims to promote improvements on building energy efficiency and requires member states to develop their own scheme of evaluating and labelling the energy performance of buildings (UK Government 2007; Department for Communities and Local Government 2014b). The UK EPC scheme requires dwellings (both domestic and commercial) to have an Energy Performance Certificate (EPC) when constructed, sold or let (Department for Communities and Local Government 2014a). Each of the certificates presents energy system information of an individual dwelling and gives a rating (from G to A+) to indicate the overall energy efficiency of the dwelling as a whole. In addition, the certificate also shows efficiency ratings of building elements and provides recommendation of measures that can be undertaken to improve the dwelling's energy

performance. The recommendations of an EPC includes an economic assessment to make clear to the building owner the impact of any energy saving potential of the recommended interventions and their indicative cost.

By April 2014 around 11.2 million EPC have been completed for UK dwellings, which is equivalent to 44.8% of the total number of households (DECC 2014). During the 12 month period ending June 2015, over 2 million EPCs were lodged into the database (Department for Communities and Local Government 2015) and all existing datasets can be requested to review by the public. Such database provides in-depth knowledge about the energy usage efficiency within individual buildings, and the high coverage of such data enables high-resolution building energy studies to be conducted at a city scale. Using the widely-acknowledged classification method by Swan & Ugursal (2009) in which energy modelling approaches are separated into two different categories – top down and bottom up, Johansson et al. (2016) pointed out that EPC data have the potential of supporting both bottom-up and top-down approaches. Their argument is supported by the work of Fabbri et al. (2012), in which city-wide EPC database was integrated with tools from the geographic information systems (GIS) for the analysis of energy performance of a city in Italy. Similar approach has also been

applied by Dall’O’ et al. (2015) who linked a large quantity of energy performance certificates with the buildings’ geographic location, and so the results can be aggregated to undertake city-scale, or sub-city-scale (e.g. districts), energy analysis. Dall’O’ et al. (2015) also emphasised the importance of managing this ever increasing database, so such data can be associated with other national or regional datasets based on their geographic correlations.

However, the integration of EPC data with GIS models faces practical difficulties as the EPC datasets do not include each dwelling’s geographic references that can be directly recognised by GIS models. In the UK, the only geographic information logged on an EPC is the dwelling’s full address and postcode, and most of such information is manually recorded by EPC assessors, hence susceptible to errors or inconsistency. For example, it has been discovered by the author in numerous cases that the address used by EPC certificates differs to the official way of description for the address of the same property as used by Ordnance Survey (2015). This challenge has also been introduced by Johansson et al. (2016) as that the geographic processing of EPC data requires manual works and is likely to be error-prone. For example, in the work of (Dall’O’ et al. 2012), energy performance information was manually collected from a number of sample buildings and imported into a GIS model. Such process would become even more time consuming and labour intense if the entire building stock in a city is to be analysed. Furthermore, a large number of demographic datasets published in the UK, e.g. national census and fuel poverty statistics, utilise a series of specific geographic scales, namely Super Output Areas (UK Office for National Statistics 2011). The hierarchy of scales of this series is shown in **Figure 1**. It shows that the largest scale – i.e. Middle Super Output Area – encloses around 2000 households, and the smallest scale, namely Output Area, contains around 40 households. These areas, however, cannot be directly incorporated with the geographic reference stated in EPC certificates such as the property’s

address or postcode (IJpelaar 2016). The structure of these two geographic referencing methods is shown in **Figure 1**.

The challenge of recognising the geographic location of large-scale building stocks was discussed in the work of Martin et al. (1994) where the usage of a unique property reference number (UPRN) is proposed as the solution to georeference each property on GIS systems. Such reference numbers use numerical codes to indicate the geographic attributes of each property, including information such as name of county, city, street, and property type. In the UK, the UPRN numbers have been widely used by local authorities for taxation and other administration purposes. However, such number has not yet been recorded in the current EPC database.

In this paper, a model is developed to bridge the gap for linking EPC database with GIS models. This model is aimed to use existing information that can be found on EPC certificates to identify the specific geographic location of each individual dwelling. The performance of this model is tested on 200 randomly selected dwellings in Southampton by comparing the model estimates with the official location of these dwellings given by Ordnance Survey (2015). The approach and results are discussed in the following sections.

2. Methodology and Case study

The model developed in this work uses the *AddressBase* data (also known as gazetteer database) by Ordnance Survey (2015) to obtain the official location of all properties within the UK. This dataset was provided in the format of csv spreadsheet but contains the geographic coordinates (northing and easting) of each data point, enabling the data points to be mapped on GIS systems. The *AddressBase* datasets also contain other attributes such as property type, UPRN, postcode, and full address. However, as introduced in Section 1, a large number of addresses in the OS dataset are found to be written in a different way with their counterpart in EPC database, as

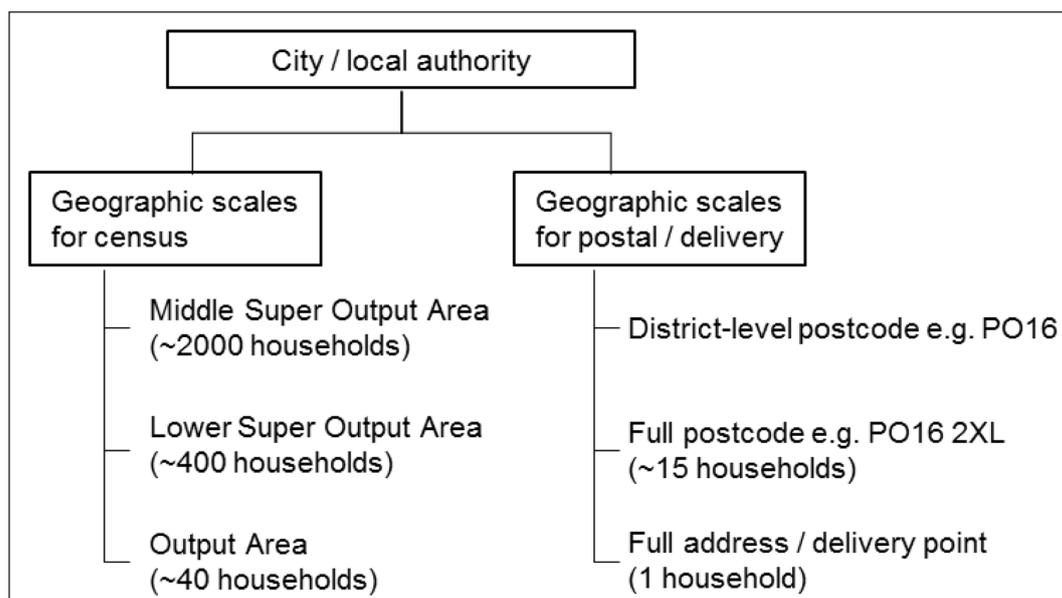


Figure 1: Geographic scales used in the UK for various activities: census versus postal delivery.

some examples are given in **Table 1**. Such difference can be easily distinguished by visual recognition but cannot be directly recognised by computer programs. This challenge is carefully addressed in this model (see Section 2.1.) and a case study on 200 randomly chosen dwellings was undertaken in the city of Southampton to test the performance and accuracy of this model, as shown in Section 2.3.

2.1. Approximate String Matching Method

The *approximate string matching* technique was used to compare a property's full address as shown on EPC certificate (hereinafter referred to as the *EPC address*) with the address records given by the Ordnance Survey database (hereinafter referred to as the *OS address*). The OS address that shares the highest level of string similarity with the EPC address is then selected as the most appropriate match. Its UPRN will then be assigned to the EPC address for georeferencing. A factor is defined in this model to quantify the similarity between an EPC address and OS addresses, namely the String Similarity Coefficient (λ_i^m). This coefficient is calculated using Equation 1.

$$\lambda_i^m = \frac{n_i^m}{N_i} \quad (1)$$

where λ_i^m is the string similarity between an EPC address i and an OS address m ; n_i^m is the length of strings that are

identical in both addresses; and N_i is the length of string in the EPC address.

To shorten the computation time and to enhance the model's accuracy, the postcode of each EPC address is used to create a subset from the Ordnance Survey database, so the string matching process can be completed on a small number of OS addresses. The key steps and processes used in this model are shown in **Figure 2**. Firstly, for each EPC dataset, its postcode will be used to select a subset of OS addresses that have the same postcode. This enables the model to shrink the georeferencing range to a relatively small geographic area, i.e. an area containing approximately 15 households as indicated in **Figure 1**. Secondly, OS addresses within the subset will be compared with the EPC address, and their string similarity coefficient will be calculated by using Equation 1. In the next step, the OS address having the highest string similarity will be identified, and its UPRN and geographic location will be assigned to the EPC address.

2.2. Case Study Area

The city of Southampton is selected in this work as a case study to test the performance of the model. The city is located on the south coast of England, UK, occupying a total land area of 56 km², and accommodating 236,882 residents from 98,254 households (UK Office

Table 1: Example of addresses that are described differently in OS database and on EPC certificates.

Case	Address shown in OS dataset	Address in EPC datasets	EPC reference
#1	Flat 7 Rivendale Court 143-145 Paynes Road	7 Rivendale Court	8190-8625-1120-6506-5083
#2	1 Ted Bates Court The Dell	1 The Dell	8852-7828-0010-8080-7996
#3	4 Edge Hill Road	4 Edgehill Road	8589-6728-8160-4919-4976
#4	Flat 4 315-317 Portswood Road	6 Boston Place 315/317 Portswood Road	0558-2819-6028-0128-4065

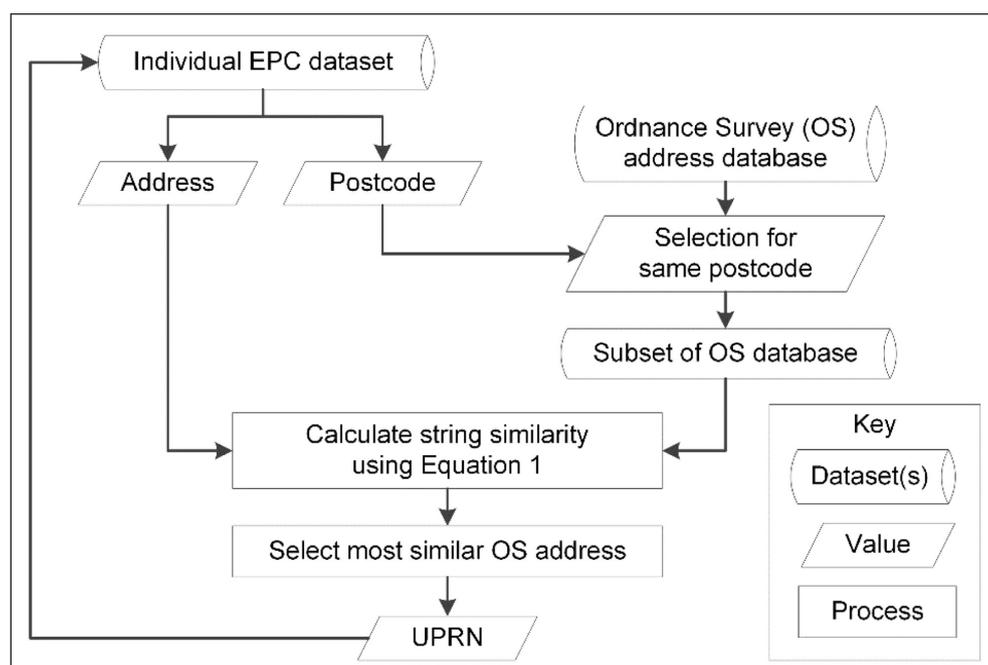


Figure 2: Steps and processes included in the model for matching EPC address with geographic references.

for National Statistics 2012). By August 2013, 47426 dwellings in Southampton have conducted EPC assessments, equivalent to 48% of the number of households in Southampton (98,254 according to Office for National Statistics 2012). For this research, all existing EPC data of Southampton households have been obtained for the purpose of conducting the analysis.

2.3. Sample Selection

Two hundred sample dwellings were randomly selected from the Southampton database to test the performance of the model, and their locations are shown in **Figure 3**. It shows that the samples are well spread over the city, covering all postcode districts including SO14 (n = 23), SO15 (n = 36), SO16 (n = 42), SO17 (n = 24), SO18 (n = 27), and SO19 (n = 48).

Figure 3 also shows that the samples cover a wide range of building types, which are:

- detached houses: houses that do not have a joint wall with other buildings,
- semi-detached houses: houses that have one wall shared with the next building,
- terraced houses: houses sandwiched between the two adjacent properties – both side walls shared, and
- flats: which are also known as apartments.

The number of samples in each type is given in **Table 2**.

3. Results and discussion

The georeferencing results estimated by the model on the 200 samples are manually validated by comparing the property’s EPC address with its actual address, and the results are discussed in the following sections.

3.1. Georeferencing Results and Error Analysis

The validation results and errors are shown in **Table 3**. Only 6 sample properties, out of 200, are mis-located by the model. The table shows that all of these errors are from a particular building type – flats, where the location of the property is misallocated with either another flat or a house sharing the same property number in the same postcode area. This is due to the fact that the address of a flat normally contains 2 numerical references, i.e. the flat number and the building number, causing additional difficulty to the work of the model. This can be found from items No. 1, 2, and 4 in **Table 3**, where the addresses of three flats are mistakenly linked with three houses due to the existence of a same property number.

Figure 4 shows the results of the sample across the city where green dots represent the correct results and red dots are for incorrect results. It shows that all results from the east part of the city (postcode areas of SO18 and SO19) and the area of SO14 are correct, whereas results in SO15, SO16 and SO17 are less accurate. In particular, 4 errors (out of 6) are found to be from the SO17 area, being the least accurate location for this model.

Overall, only 6 properties, out of 200, are found to be misplaced by the model, giving an overall accuracy of 97%. It has been found that flats, especially those containing more than one numerical property reference, are most likely to cause errors. Such error could be avoided in the future by adding the consideration of building type as a criterion in this model.

3.2. EPC Data Coverage in Southampton

The model developed in this work is used to estimate the geographic location of all dwellings that are included in Southampton EPC database (n = 47426). After data cleaning such as deduplication, 39568 dwellings have been

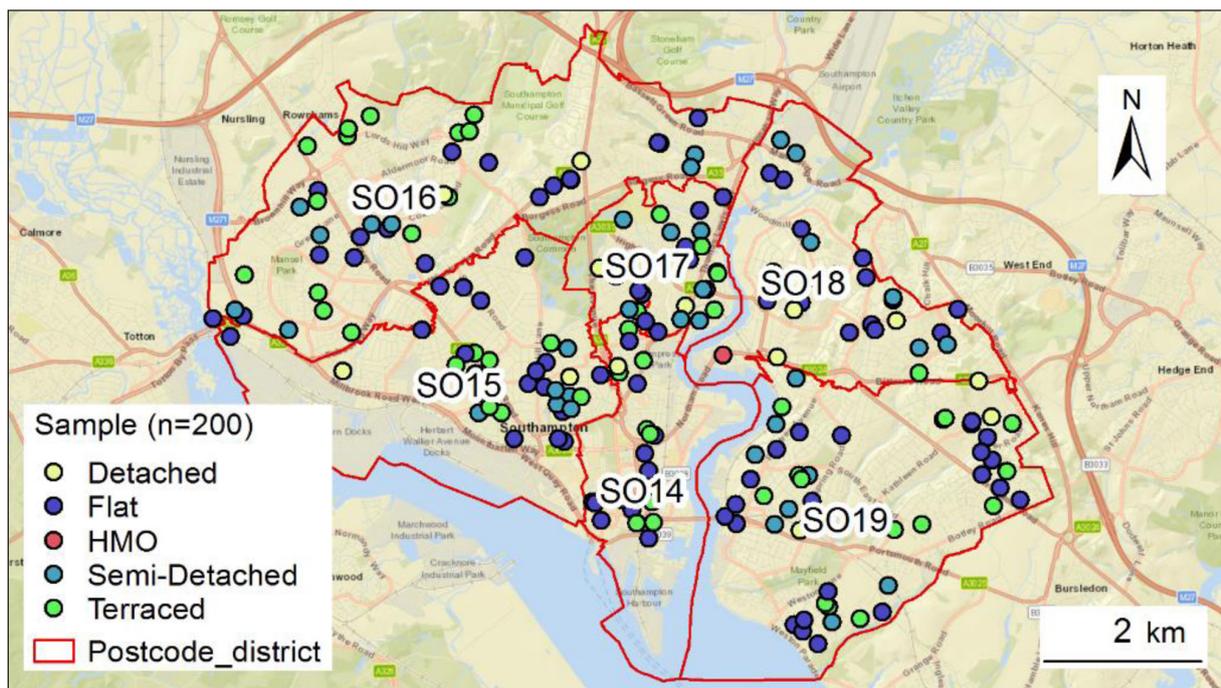


Figure 3: Selected samples for model performance testing.

Table 2: Number of samples in various building types.

Building type	Detached	Semi-detached	Terraced	Flat & HMO*	Total
Number	15	37	52	96	200

*House in Multiple Occupation: a property rented out by at least 3 people who are not from a same family.

Table 3: Errors found in the georeferencing results of the model.

#	Postcode	EPC address	Estimated location (OS address)	SSC**
1	SO17 1NX	Flat 4 25 Highfield Road	4 Highfield Road	0.6
2	SO17 3SF	Flat 3 Westmarch Court 1a Kitchener Road	3 Kitchener Road	0.4
3	SO17 2EX	Flat 11 7 Lawn Road	Flat 7 7 Lawn Road	0.8
4	SO17 2LH	Flat 6 Nelric House Kent Road	6 Kent Road	0.5
5	SO16 5FP	Flat 2 1 Coxford Road	Flat 1 1 Coxford Road	0.8
6	SO15 5QR	Flat 9 85 Anglesea Road	Flat 9 67 Anglesea Road	0.8

**String Similarity Coefficient – see Section 2.1.

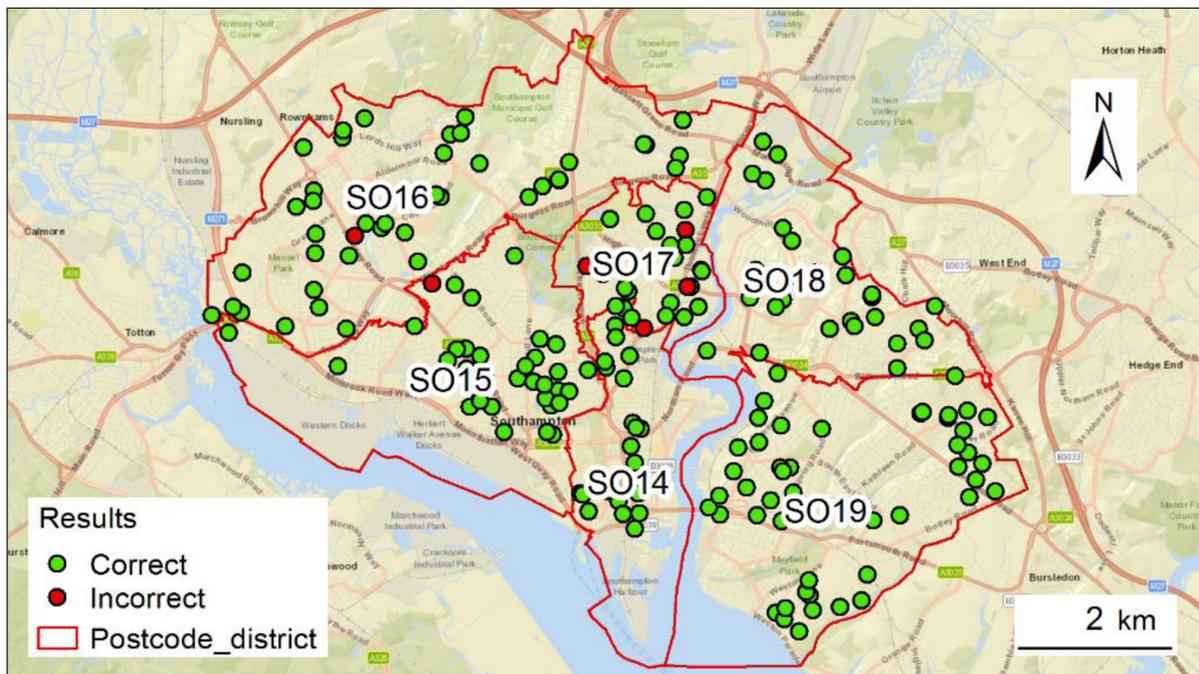


Figure 4: Georeferencing results that are found to be correct (green dots) or incorrect (red).

georeferenced as within the Southampton boundary. **Figure 5** shows the comparison between the number of dwellings having EPC data within each Lower Super Output Area (LSOA, see Section 1) area versus the total number of dwellings (Ordnance Survey 2015), indicating the level of EPC data coverage of each area. Areas marked in red are areas that have the highest EPC coverage, whereas areas in blue are for the lowest coverage. It shows that most areas in Southampton (76 out of 146) have an EPC data coverage between 30 and 40%. Such areas are mostly in SO16, SO18 and SO19 districts. Nineteen areas are found to have a high EPC coverage at over 55%. These areas are concentrated in the middle region of the city, including the south tip of SO14 as well as the SO17 district. Only 3 areas (dark blue) are found to have an EPC data coverage that is lower

than 30%. These areas locate in the west- and east-ends of the city.

The distribution of EPC data shown in **Figure 5** is found to be closely correlated with the demographic features of Southampton areas. The areas marked in red in SO14 in **Figure 5**, namely Bargate, are commonly known as the city centre for Southampton, accommodating a large number of flats that are recently built after 1990's. These properties have very high demand in the housing market due to easy access to commercial and transportation facilities. As a result, a large proportion of dwellings in this region are equipped with EPC certificate as a requirement for entering the housing market.

Similarly, the SO17 district accommodates a major education institute in the UK – the University of

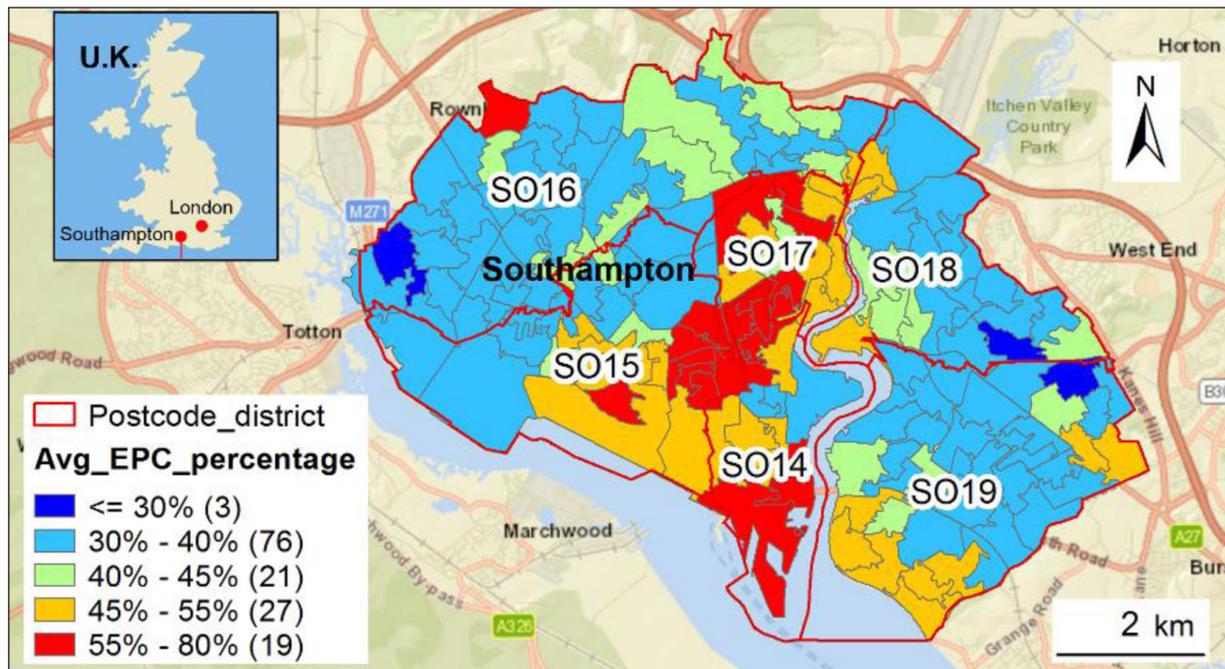


Figure 5: Percentage of households having EPC in LSOA areas of Southampton.

Southampton. Therefore, dwellings in this district, namely the Highfield, face very high demand from students and staff working in the University, hence more likely to possess EPC data.

Conversely, areas away from the city centre are expected to have a more stable residence situation and a lower turnover rate. As a result, households living in such areas are less likely to conduct EPC assessment. This trend is well presented in **Figure 5**.

3.3. Validation Summary

In this section, the 200 randomly selected samples have been tested by our model and the results have been manually validated with their actual location. The validation results show that 97% of the model estimates are accurate. The model has then been applied to all existing EPC data from Southampton, generating results that can be used by GIS systems for further analysis.

4. Conclusion

In this paper, we identified a current gap in the UK where a large number building-level energy performance datasets have not been equipped with geographic references, impeding the effort of connecting such information with other demographic datasets. We addressed this gap by developing a model that is able to automatically estimate the geographic location of all dwellings recorded in EPC database. To test the performance of this model, we applied in to 200 randomly selected sample dwellings in Southampton and the model estimates were manually validated by comparing with the dwellings' actual location. The validation results show a good level of accuracy as 97% of the results are found to be correct. Only 6 (out of 200) samples showed errors, which are caused by the way of how addresses of flats are displayed in the UK (see Section 3.1).

We have then applied our model to existing EPC datasets in Southampton ($n = 47426$). The results are able to reveal the distribution of EPC data within the city (see **Figure 5**). Areas close to the city centre and educational institutions are found to have the highest proportion of EPC datasets. Conversely, areas away from the city centre are found to have less amount of EPC data due to a more stable residence condition.

Overall, the model presented in this paper is able to bridge a current gap in terms of data applicability in the UK, paving the way for building-level energy performance to be widely used for city-scale energy studies.

Acknowledgement

This work is part of the activities of the Energy and Climate Change Division and the Sustainable Energy Research Group at the University of Southampton (www.energy.soton.ac.uk). It also has some support from EPSRC grants: EP/J017698/1, Transforming the Engineering of Cities to Deliver Societal and Planetary Wellbeing, EP/N010779/1, City-Wide Analysis to Propel Cities towards Resource Efficiency and Better Wellbeing and EP/K012347/1, International Centre for Infrastructure Futures (ICIF).

Competing Interests

The authors have no competing interests to declare.

References

- Council of the European Union** 2002 Directive 2002/91/EC of the European Parliament and of the Council of 16 December 2002 on the energy performance of buildings. *Official Journal Of The European Union*, 65–71.
- Dall'O', G**, et al. 2015 On the use of an energy certification database to create indicators for energy planning purposes: Application in northern Italy. *Energy*

- Policy*, 85: 207–217. DOI: <https://doi.org/10.1016/j.enpol.2015.06.015>
- Dall'O', G, Galante, A and Torri, M** 2012 A methodology for the energy performance classification of residential building stock on an urban scale. *Energy and Buildings*, 48: 211–219. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0378778812000515> [Accessed May 25, 2013].
- DECC** 2014 Annex B: Energy Performance Certificate data. June, 16. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/323946/Annex_B_-_Energy_Performance_Certificate_data.pdf.
- Department for Communities and Local Government** 2014a Energy Performance of Buildings Certificates: notes and definitions.
- Department for Communities and Local Government** 2014b Regulatory and administrative amendments made since the implementation of the 2007 regulations.
- Department for Communities and Local Government** 2015 Energy Performance of Buildings Certificates: Statistics Release England and Wales Energy Efficiency. July, 1–20.
- Fabrizi, K, Zuppiroli, M and Ambrogio, K** 2012 Heritage buildings and energy performance: Mapping with GIS tools. *Energy and Buildings*, 48: 137–145. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0378778812000333> [Accessed September 3, 2014].
- Ijpelaar, J** 2016 POSTAL GEOGRAPHY AND GEO-REFERENCING. February, 1–9.
- Johansson, T, et al.** 2016 Energy performance certificates and 3-dimensional city models as a means to reach national targets – A case study of the city of Kiruna. *Energy Conversion and Management*, 116: 42–57. DOI: <https://doi.org/10.1016/j.enconman.2016.02.057>
- Martin, D, et al.** 1994 The use of GIS in the analysis of diverse urban databases, 18(1): 55–56.
- Ordnance Survey** 2015 AddressBase Technical Specification v2.1. June, 1–26.
- Swan, L G and Ugursal, V I** 2009 Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews*, 13(8): 1819–1835. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1364032108001949> [Accessed May 25, 2013]. DOI: <https://doi.org/10.1016/j.rser.2008.09.033>
- UK Government** 2007 2007 No. 991 BUILDING AND BUILDINGS, ENGLAND AND WALES The Energy Performance of Buildings (Certificates and Inspections) (England and Wales) Regulations 2007, 991.
- UK Office for National Statistics** 2011 Super Output Areas (SOAs). Available at: <http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super-output-areas--soas-/index.html>.
- UK Office for National Statistics** 2012 Data release: 2011 Census, Population and Household Estimates for the United Kingdom. Available at: <http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm:77-270247>.

How to cite this article: Wu, Y, Blunden, L and Bahaj, A. 2018 City-wide Building Energy Efficiency Assessment Using EPC Data. *Future Cities and Environment*, 4(1): 3, pp. 1–7, DOI: <https://doi.org/10.5334/fce.10>

Submitted: 19 October 2017

Accepted: 13 December 2017

Published: 26 January 2018

Copyright: © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

] *Future Cities and Environment*, is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 