

High-resolution mapping of residential wood burning heat sources using Energy Performance Certificates: A case study of England and Wales

Laura Horsfall¹, Calum Kennedy¹, Nicola Shelton¹, Jonathon Taylor³,
John R Hurst¹, Matthew Loxham², Adam Dennett¹, and Pia Hardelid^{*1}

¹University College London

²University of Southampton

³Tampere University

April 2025

BACKGROUND: Particulate matter emissions from residential wood burning are rising in many countries. Long-term exposure to fine particulate matter is strongly linked with adverse health effects including cardiovascular and respiratory disease. Policymakers and scientists need accurate tools to identify residential

*Corresponding author: laura.horsfall@ucl.ac.uk

wood burning hotspots at fine scale. However, current methods rely on spatially-misaligned, out-of-date data sources, reducing their practical utility and portability to other contexts. Furthermore, the socio-economic characteristics of residential wood burning in high income countries are poorly understood.

METHODS: We used open data from 26 million Energy Performance Certificates (EPCs) for properties in England and Wales from January 2009 to February 2025 to map the concentration and prevalence of wood burners within small areas. We evaluated our method against the UK national wood burning emissions inventory using national air pollution monitoring networks. We used novel open data linkages to characterise associations between area-level prevalence of wood burners and socio-economic factors including deprivation, ethnicity, and age.

FINDINGS: We identified substantial spatial heterogeneity in the concentration of wood burners, with the highest concentrations in affluent urban areas. Our EPC-based concentration metric was more strongly correlated with peaks in winter $\text{PM}_{2.5}$ at urban monitoring sites than estimates from the UK national emissions inventory. Prevalence of wood burners was positively correlated with age and percentage of residents identifying as ethnically white, and negatively correlated with measures of social deprivation. Prevalence of wood burners in EPCs has increased since 2009.

CONCLUSIONS: EPCs are a valuable data source which policymakers can use to target local interventions or extend existing restrictions on solid fuel burning. Our method is transparent, up-to-date, and portable to other countries where similar EPC data is available. The relationship between social deprivation and prevalence of wood burning heat sources highlights important issues of environmental justice. Epidemiological analyses of wood smoke exposures and health should carefully account for the confounding effects of age, deprivation, and ethnicity.

1 Introduction

Burning solid fuels such as coal, wood and other biomass releases harmful pollutants including carbon monoxide, nitrogen dioxide, and small particulate matter (PM_{2.5}). These pollutants can cause immediate health effects and increase the risk of chronic conditions including lung and cardiovascular disease.^{1,2} The World Health Organisation estimates solid fuel emissions were responsible for 3.2 million premature deaths in 2020.³

In recent decades, high-income countries have focused on reducing particulate emissions from vehicles and industry, with many regions showing dramatic improvements since the 1960s.⁴ However, with efforts concentrated in these sectors, residential emissions from solid fuel - mainly wood fuel - increased in many countries. Consumption of solid biomass in the European Union (EU) has doubled since 1990.⁵ Most of the increase has occurred since 2002, when the EU's first renewable energy policy promoted biomass as a renewable fuel.^{5,6}

Current trends in residential wood burning are concerning due to the potential health impacts of wood smoke exposures. A growing literature highlights possible relationships between wood smoke exposure and adverse health outcomes including lung cancer.^{1,7–9} Residential wood burning has also been linked to climate change and biodiversity loss.^{10,11}

To explore the health and climate impacts of residential wood burning at scale, scientists need to understand where wood burning is most common. High-resolution mapping of residential wood burning can also inform targeted public health interventions or awareness campaigns focused on wood burning. Targeted interventions can be effective. In the Tasmanian city of Launceston, a series of local policy measures during the 1990s and early 2000s including educational campaigns and a wood heater replacement programme reduced the household prevalence of wood stoves from 66% to 30%.¹²

However, high-resolution mapping of residential wood burning using existing data sources is challenging. National air quality monitoring networks are typically biased towards traffic pollution from major roads. Residential wood burning emissions inventories generally integrate multiple data sources at different spatio-temporal scales.^{13–18} Combining these misaligned data sources necessitates modelling assumptions, which can create biases in the high-resolution inventory.¹⁹

Existing residential wood burning maps are also generally uninformative about the socio-economic context of the areas where wood burning is prevalent. More broadly, despite an extensive literature characterising solid fuel usage in developing countries,^{20,21} there is relatively little evidence on the socio-economic distribution of residential wood burning in high-income countries. Such contextual information could play an important role in epidemiological studies and for targeting public health interventions to specific local area characteristics.

The United Kingdom (UK) is an interesting case study for the resurgence of residential wood burning. Following the Great London Smog of 1952, where a combination of residential coal emissions and unusual weather conditions caused an estimated 12,000 deaths, the government established The Clean Air Act. The Clean Air Act included designated Smoke Control Areas (SCAs) in towns and cities. Only smokeless fuels (e.g., anthracite or gas) were permitted for heating homes in SCAs, making the UK the first country to implement legal restrictions on residential emissions.

Today, most wood burning appliances sold in the UK are SCA-compliant. However, residential wood burning remains a major contributor to ambient air pollution, comprising 11% of total PM_{2.5} emissions in 2023.²² A recent survey found that residential wood burning is concentrated in affluent, urban areas, raising issues of environmental justice.²³ In the UK, local governments are responsible for local policies (such as setting SCA boundaries) and awareness campaigns.

Energy Performance Certificates (EPCs) are a potentially valuable data source for high-resolution mapping of wood burning heat sources. EPCs provide information on property-level characteristics including primary and secondary heat source. EPCs can be easily linked to other high-resolution open data sources including local socio-economic characteristics. EPC databases are continuously updated, which means they can be used to track changes over time. Many countries make EPCs openly available, which means that methods based on EPCs are portable to other contexts.

However, EPCs provide an incomplete representation of the overall housing stock, since not all residential properties have an EPC. Uncorrected analyses using EPCs could produce biased estimates due to missing data and selection bias.

To our knowledge, no study has used EPCs to explore the geographic distribution of wood burning heat sources. The key objectives of this study were to:

- Use novel open data linkages to understand the geographic and socio-economic distribution of residential wood burning heat sources in two European countries (England and Wales)
- Develop and benchmark a transparent, portable tool to allow local policymakers to target interventions aimed at reducing emissions from residential wood burning

We used data from 26 million EPCs to map the distribution of wood burners in England and Wales. We evaluated the representativeness of the EPC dataset against a database of all residential properties in England and Wales. We developed a method to correct for missing data in EPCs using Census data, and estimated the prevalence and concentration of wood burners by Lower layer Super Output Area (LSOA). We evaluated our measure against PM_{2.5} measurements from the UK national air pollution monitoring network. We examined the association between LSOA-level prevalence of wood burners and socio-economic factors,

including deprivation, ethnicity, and age. We described changes in the prevalence of wood burners since 2009.

Our novel data linkages highlight for the first time the potential environmental injustice associated with wood burning air pollution at high geographic resolution, particularly within urban areas with existing regulations on wood burning. Moreover, we developed a transparent, portable tool to support policymakers to identify potential wood burning hotspots in England and Wales.

To supplement the main manuscript, we provide a publicly available report featuring interactive maps of the distribution of wood burners in England and Wales, available at: <https://storymaps.arcgis.com/stories/9231a18627b94b3a80e4f33fb3b4a9fd>. All code used to produce the manuscript is available alongside replication instructions at: https://github.com/UCL-Wellcome-Trust-Air-Pollution/EPC_mapping_project_code. All non-EPC open data used to produce the manuscript is available at: <https://zenodo.org/records/14888530>.

2 Methods

2.1 Data

The Energy Performance Certificate dataset

Energy Performance Certificates (EPCs) are used in many countries to monitor the energy efficiency of the housing stock.²⁴ EPCs are an information tool introduced under the 2002 EU Energy Performance of Buildings Directive, and contain information on a building's energy efficiency alongside details on housing characteristics including property type, tenure, and primary and secondary heating source.

In England and Wales, a valid EPC certificate has been legally required at point of build, sale, or private rental since 2008. Accredited energy assessors generate the certificates, which are valid for ten years following inspection. The EPC database is updated monthly in England and Wales. The EPC dataset includes unique property reference numbers (UPRNs), which are unique identifiers for every addressable location in the UK.

Other data sources

We used data from the 2019 Index of Multiple Deprivation (IMD). The IMD is a measure of social deprivation determined at the LSOA level for England²⁵ and Wales.²⁶ The IMD is a combined measure based on weighted indicators of income, employment, education, health, crime, and the environment. Higher IMD scores indicate a higher level of social deprivation.

We used data from the 2021 Census for England and Wales to describe ethnicity at LSOA level, and data from the 2011 Census (the most recent available data) to assign urban/rural status to LSOAs.^{27,28} LSOAs were defined as rural if they were outside settlements with more than 10,000 residents. We used data from the Office for National Statistics' population estimates to derive the mid-2022 estimates for median age by LSOA.

We used data from the Ordnance Survey (OS) AddressBase to evaluate the representativeness of the EPC dataset against the housing stock in England and Wales. The OS AddressBase is a database linking around 29 million residential postal addresses to UPRNs, providing a complete list of active residential addresses in the UK. The OS AddressBase contains granular classification codes for housing type, which can be mapped onto Census 2021 housing type classifications.

We evaluated our high-resolution maps of residential wood burners against existing maps derived from the UK National Atmospheric Emissions Inventory (NAEI) and London Atmospheric

Emissions Inventory (LAEI). The NAEI is a database of annual UK emissions estimates for air pollutants including PM_{2.5}.¹⁸ The NAEI combines residential addresses OS AddressBase data, 2011 Census data on housing and central heating type within small areas, a regional energy consumption model, and a 2020 household survey to generate 1km × 1km grid square estimates of PM_{2.5} emissions from residential wood burning. The LAEI is a separate database of emissions estimates by grid square for Greater London and nearby areas within the M25 motorway.²⁹ The LAEI augments the NAEI methodology to incorporate additional data sources available within London.

We used hourly air quality data from three national air quality monitoring networks in the UK: the Automatic Urban and Rural Network (AURN), Air Quality England, and Air Quality Wales. The AURN is the main network used for compliance reporting against the UK’s Ambient Air Quality Directives.³⁰ Air Quality England and Air Quality Wales are locally-managed networks which are not part of the national monitoring network.^{31,32} We accessed air quality data via the ‘openair’ R package.³³

2.2 Geographical hierarchies in the UK

We selected four geographical levels to report our main findings: LSOAs, Electoral Wards, LADs, and regions. We selected these geographies because they are the main reporting geographies for other data sources such as socio-economic characteristics, and because they are the main geographical units used for policymaking.

LSOAs are UK Census geographies. In 2021 there were 33,755 LSOAs in England and 1,917 in Wales. LSOAs comprise between 400-1,200 households, and have a population between 1,000-3,000 people.

Electoral Wards, LADs, and regions are administrative geographies. In 2021, there were 6,980 Electoral Wards in England and 852 in Wales. Wards were nested within 318 LADs in England and 22 Unitary Authorities in Wales (LADs hereof). In England, LADs are nested within nine former government office regions, which are the highest tier of sub-national division.

2.3 Data cleaning and linkage

We downloaded all 27,491,063 EPCs from January 2009 to February 2025 and removed duplicated entries. We used UPRNs to link EPCs to higher level geographies (LSOA, Ward, LAD, and region) using the National Statistics UPRN Lookup for Great Britain as of May 2022.³⁴ We linked retained records to LSOA-level data including IMD score, ethnicity, median age, and rural-urban status.

The final EPC dataset contained 25,952,810 records, among which there were 18,708,569 unique properties. Further details on the data cleaning procedure are available in Supplementary Appendix A.

2.4 Derived variables

We generated binary variables for whether a property’s primary or secondary heat source was a solid fuel or wood fuel heat source using keyword searches. A list of keyword searches used is available in Supplementary Appendix B.

Properties can have multiple EPCs. We generated a binary variable indicating whether an EPC was the most recent for that property. In our cross-sectional analyses and maps, we used the subset of most recent EPCs.

We re-categorised the property type and tenure variables in the EPC dataset to match the categories defined in the 2021 Census. Property type was defined as one of 'detached', 'semi-detached', 'terrace', 'flat', or 'other accommodation'. Tenure was defined as one of 'new build', 'owner-occupied', 'rented (private)', or 'rented (public)'.

We defined two outcome variables - the prevalence and the concentration of solid fuel and wood fuel heat sources by statistical geography (e.g. LSOA). The prevalence was defined as the percentage of most recent EPCs for houses in each geography which had a solid fuel or wood fuel heat source as the main or secondary heat source. We classified properties defined as 'detached', 'semi-detached', and 'terrace' in the 2021 Census as houses. We restricted the prevalence variable to houses since other accommodation types such as flats generally do not have capacity for wood burners, and had a very low prevalence of wood burners in previous surveys.²³

The concentration variable was defined as the number of most recent EPCs in each geography which had a solid fuel or wood fuel heat source as main or secondary heat source, divided by the total area of the geography in km². The concentration is defined in relation to a geographic area as opposed to a specific population (houses). Therefore, we included all property types including flats. We used our concentration as a proxy variable for monitor-level wood burning emissions exposures in our comparisons with the NAEI methodology.

2.5 Estimating the geospatial prevalence and concentration of wood burners

Correcting for missing data and selection bias

Using EPCs, we estimated the percentage of properties with a wood fuel or solid fuel heat source by property type and LSOA. To mitigate uncertainty due to small sample size, we excluded LSOA-property type pairs where there were fewer than 20 EPCs for that LSOA-property type

pair. This provided an estimate of the prevalence of wood and solid fuel heat sources in EPCs by LSOA and property type.

We used 2021 Census data to derive the total number of properties within LSOAs by property type. We estimated the total number of properties with wood fuel and solid fuel heat sources by property type by multiplying the estimated prevalence from EPCs with the number of properties of each property type from the Census. Finally, we aggregated the estimated number of properties with wood and solid fuel heat sources across property types to derive an estimate for the total number of properties with wood and solid fuel heat sources by LSOA.

To estimate the LSOA-level prevalence of wood fuel and solid fuel heat sources, we divided the estimated total number of properties with wood and solid fuel heat sources by the total number of properties in the LSOA. We derived the estimated concentration of wood and solid fuel heat sources by dividing the estimated total number of properties with a wood or solid fuel heat source by the area of the LSOA in km^2 . We compared our corrected estimates of the concentration and prevalence of wood burners to the uncorrected estimates using only EPCs.

Mathematical details on our methodology are available in Supplementary Appendix C.

Mapping the estimated prevalence and concentration of wood burners

We produced choropleth maps of the estimated prevalence and concentration of wood burners by LSOA for England and Wales. To improve readability of the concentration maps, we truncated low and high extreme values to the 5th and 95th percentile values respectively.

We compared the estimated concentration of wood burners using EPCs in London to existing maps of wood fuel emissions from the LAEI. To generate a measure of wood fuel concentration

comparable to the LAEI, we used the same methodology to correct for missing EPC data with two modifications:

- We aggregated by 1km² grid squares instead of LSOAs.
- We estimated the total number of properties by grid square using the OS AddressBase instead of the Census housing data.

2.6 Evaluating the wood burner concentration metric

We downloaded hourly PM_{2.5} measurements during 2022-2024 from the AURN, Air Quality England, and Air Quality Wales. We excluded 2020 and 2021 since these years may be unrepresentative of normal pollution patterns due to COVID-19. We also excluded years prior to 2020 as data from these years may not reflect the current distribution of wood burners.

We used data from 417 monitoring sites, of which 217 belonged to the Air Quality England network, 164 to the AURN, and 38 to the Air Quality Wales network. For our main analysis, we only used data from the AURN. This is because the AURN sites adhere to strict requirements based on national air quality directives, whereas locally-managed sites vary in data collection and quality practices.³⁰

As in Zalzal et al. [17], we created a circular buffer with radius 1km around each monitoring site and counted the estimated number of wood burners within the buffer radius. We considered two outcome variables - the average monitor-level PM_{2.5} and the difference between average PM_{2.5} at peak vs. non-peak burning times. We defined peak burning times to be between 7pm and 1am and non-peak burning times to be between 5am and 5pm. We restricted our analysis to monitoring sites classified as 'Urban Background' - urban monitoring sites where pollution is an integrated function of all sources upwind of the monitor.

We produced scatter plots of the correlation between the natural logarithm of the number of wood burners around a monitoring site and the two outcome variables. We plotted separate results for summer (June, July, and August) and winter (December, January, and February). We plotted the correlation between wood burner concentration and difference in peak vs. non-peak emissions separately for weekdays and weekends (defined as Saturday and Sunday).

For each scatter plot, we calculated the Spearman rank correlation coefficient (R) between the wood burner concentration variable and the outcome variable separately for the summer and winter seasons. We derived 95% confidence intervals for R using the non-parametric bias and skewness adjusted (BCA) bootstrap procedure with $B = 10,000$ replications³⁵. We also reported bootstrapped confidence intervals for the difference between the Spearman correlation coefficients in summer and winter. Our key assumption in generating confidence intervals using the bootstrap is that the sample of observed urban background monitoring sites is representative of the set of *possible* urban background monitoring sites.

To test the sensitivity of our results to the radius of the spatial buffer, we re-created the scatter plots using buffer radii of 500m and 2km. To further probe the robustness of our results, we plotted the correlation between concentration of wood burners and the outcome variables using urban background sites from all three monitoring networks. To test whether our concentration metric is simply a proxy for housing density, we produced the same plots replacing the x-axis variable with the total number of houses in the EPC data within a 1km radius of the monitor.

We evaluated our concentration metric against 1km^2 grid square wood burning emissions estimates from the NAEI using AURN urban background sites. For each AURN urban background site, we counted the estimated number of wood burners in the site's grid square using the EPC data. We produced the scatter plots for our grid square concentration metric, and compared them to the same scatter plots using the estimated grid square annual wood burning emissions from the current NAEI inventory. Whilst the EPC-based concentration

metric does not attempt to measure actual $\text{PM}_{2.5}$ emissions, the goal of this comparison was to test its utility as a tool to predict daily levels and peaks in $\text{PM}_{2.5}$ emissions relative to the NAEI inventory.

2.7 Associations between prevalence of wood burners and socio-economic indicators

We produced charts displaying the association between wood burner prevalence and LSOA-level IMD score and ethnic composition. Each chart was stratified by region, and averaged the explanatory variable by decile. We summarised LSOA-level socio-economic characteristics by decile of wood burner prevalence.

We built multivariate beta regression models to test the relative strength of the association between socio-economic indicators and prevalence of wood burners. Beta regression is a popular choice for modelling continuous outcomes bounded between zero and one, as with our prevalence metric. We used a logit link function, which means that the model coefficients can be interpreted as the association between a one unit change in the independent variable and the change in the logit of the expected value of the (beta-distributed) outcome variable.

In our first model we included all LSOAs. with independent variables LSOA median age, IMD score, urban/rural classification, SCA status, percentage of residents with a white ethnic background, and region indicators. We ran two further models stratified on the subset of urban and rural LSOAs respectively, including the same independent variables except for urban/rural status.

2.8 Temporal trends in prevalence of wood burners

We calculated the percentage of new EPCs issued each year which indicated presence of a wood burner for detached, semi-detached, and terraced houses. For this analysis, we included all EPCs, including properties with multiple EPCs. This is because the wood fuel status of a property can change over time, for example due to home renovations.

We visualised the change in prevalence of EPCs with wood burners in urban and rural areas stratified by IMD decile of the property's LSOA. We also presented aggregate trends by property type and region.

Observed changes in prevalence could reflect changes to the housing stock (for example changes in building standards for new properties), changes in buying or renting behaviour due to preference shifts, or renovations to existing properties.

To explore whether renovations to existing properties are a relevant factor in driving trends in wood burner prevalence, we used the subset of properties with multiple EPCs. We reported the percentage of these properties which had a wood burner at time of first EPC, and for every subsequent EPC. We excluded results for properties with five or more EPCs due to small sample size.

All analyses were carried out in R version 4.4.1.

2.9 Role of the funding source

The funding source had no role in the study design, data collection and analysis, report writing, or the decision to submit this paper for publication.

3 Results

3.1 Representativeness of Energy Performance Certificates

Out of the 28.6 million properties in the OS AddressBase, 18 million (63.0%) were matched to the EPC data. Flats were over-represented in EPCs, while detached, semi-detached, and terraced houses were under-represented. For example, 28.2% of properties in the EPC data were flats, compared to 23.8% in the OS AddressBase (Supplementary Table 3).

3.2 Characteristics of properties with wood fuel and solid fuel heat sources

Of the 18,708,520 unique EPCs, 1,586,345 (8.5%) had a solid fuel heat source. Of these, 1,368,802 (7.3%) were wood burners. Properties with wood burners were more likely to be detached, owner-occupied properties (Table 1). Most wood burners were classified as secondary heat sources - 0.3% of properties had a wood burner as the primary heat source, whereas 7.1% had a wood burner as secondary heat source.

3.3 Geographic prevalence and concentration of wood fuel heat sources

Estimating the prevalence and concentration of wood burners

The Census-corrected estimated concentration of wood burners was generally higher than that estimated using only EPCs (Supplementary Figure 5 Panel A). The estimated prevalence of wood burners closely matched the uncorrected estimates, suggesting that within housing types EPCs are broadly representative of the OS AddressBase (Supplementary Figure 5 Panel B). Our Census-corrected prevalence estimates were 9.3% for solid fuels and 8.1% for wood fuel.

Mapping the estimated prevalence and concentration of wood fuel heat sources

The prevalence of wood burners in EPCs was lowest in urban centres including Greater London, and Greater Manchester, and highest in rural areas in the South West, North West, and Wales (Figure 1 Panel A, upper). There was a wide variation in the estimated prevalence of wood burners across LADs, ranging from 0.5% in the London borough of Newham to 43.7% in the Isles of Scilly.

Prevalence also varied substantially within urban areas. In Greater London (Figure 1 Panel A, lower), the prevalence of wood burners was highest in the outer boroughs of Richmond upon Thames (6.1%), Kingston upon Thames (6.0%), and Sutton (5.5%), annotated in blue on the map.

The areas with the highest concentration of wood burners were found in smaller urban centres and towns away from major cities (Figure 1 Panel B, upper). The estimated concentration of wood burners was highest in the South Eastern LADs of Worthing (121.4 per km²), Norwich (105.7 per km²), and Reading (104.9 per km²).

There was substantial heterogeneity in the estimated concentration of wood burners within small urban areas (Figure 1 Panel B, lower). The average concentration of wood burners across urban SCAs was 46.3 per km² compared to 87.2 per km² in non-SCA urban areas.

3.4 Evaluating the wood burner concentration metric

Evaluation against air pollution data

The log concentration of wood burners was positively correlated with differences in peak vs. non-peak PM_{2.5} in winter, but not in summer (Figure 2). On winter weekdays, the central

estimate of the Spearman correlation was $R = 0.49$, with a BCA bootstrap 95% confidence interval of $[0.26, 0.67]$. On summer weekdays, the corresponding central estimate for R was 0.09 (95% CI $[-0.2, 0.35]$). The confidence interval for the difference between the winter and summer correlation coefficients was $[0.06, 0.74]$. Our central estimates for the correlation between log wood burner concentration and average $\text{PM}_{2.5}$ were smaller, and both confidence intervals covered zero.

Expanding our data to urban background sites from all three monitoring networks, we found quantitatively similar results to our analysis using only the AURN (Supplementary Figure 8). Correlations between log wood burner concentration and differences in peak vs. non-peak $\text{PM}_{2.5}$ were robust to buffer radii of 500m and 2km (Supplementary Figures 9-10). The correlation between log wood burner concentration and average $\text{PM}_{2.5}$ varied by the choice of buffer radius - we observed a stronger correlation for wider buffer radii.

The EPC-based measure of housing density was weakly correlated with differences in peak vs. non-peak $\text{PM}_{2.5}$ (Supplementary Figure 11). The point estimates for the correlation between log housing density and differences in $\text{PM}_{2.5}$ between peak vs. non-peak burning times were $R = 0.06$ (95% CI $[-0.19, 0.33]$) on winter weekdays and $R = 0.19$ (95% CI $[-0.06, 0.43]$) on winter weekends.

Comparison against NAEI emissions maps

Using grid square data from the NAEI, we continued to find evidence for a positive correlation between our measure of wood burner concentration and differences in measured $\text{PM}_{2.5}$ at peak vs. non-peak burning times in winter (Supplementary Figure 12). The point estimates for the correlation between log wood burner concentration and differences in peak vs. non-peak

average PM_{2.5} in winter were $R = 0.47$ (95% CI = [0.28, 0.63]) on weekdays and $R = 0.35$ (95% CI = [0.14, 0.53]) on weekends.

The implied grid square distribution of wood burning emissions from the existing NAEI inventory was poorly correlated with both average measured PM_{2.5} and differences in peak vs. non-peak average PM_{2.5} (Supplementary Figure 13). In winter, the point estimates for the correlation between NAEI-implied wood burning emissions and differences in peak vs. non-peak PM_{2.5} were 0.24 (95% CI = [0.03, 0.42]) on weekdays and 0.12 (95% CI [-0.09, 0.31]) on weekends.

3.5 Associations between prevalence of wood burners and socio-economic indicators

Low social deprivation was associated with higher prevalence of wood burners. The average IMD score in LSOAs in the lowest wood burner prevalence decile was 36.6, compared to 13.4 in the highest prevalence decile (Table 2). LSOAs with a higher prevalence of wood burners also had older populations. The median age in LSOAs in the lowest wood burner prevalence decile was 33.7 years, rising across every subsequent decile to 51.2 years in the highest prevalence decile. Prevalence of wood burners was also positively correlated with the proportion of LSOA residents identifying as having a white ethnic background.

In urban LSOAs, wood burner prevalence was negatively associated with social deprivation and positively associated with the percentage of residents identifying as ethnically white (Supplementary Figure 14). The relationship between socio-economic characteristics and wood burner prevalence was different in urban and rural LSOAs (Supplementary Figures 15-17). In rural areas, there was no clear univariate relationship between social deprivation and prevalence of wood burners (Supplementary Figure 11).

Results from multivariate beta regression supported the univariate associations between socio-economic characteristics and wood burner prevalence. Urbanicity, SCA restrictions, and IMD score were all negatively associated with prevalence of wood burners, while median age and percentage of residents identifying as white were positively associated (Supplementary Table 5). All coefficients were large relative to their standard error, with confidence intervals excluding zero.

Restricting our model to urban LSOAs, we identified the same directionality of association between wood burner prevalence and SCA restrictions, median age, percentage of residents identifying as ethnically white, and IMD score. However, the strength of the relationship was weaker for median age, and stronger for the ethnicity and social deprivation variables (Supplementary Table 6). In rural LSOAs, the prevalence gradient was stronger with respect to age and weaker with respect to social deprivation and ethnicity (Supplementary Table 7).

3.6 Temporal trends in the prevalence of wood burners

From 2009-2024, the prevalence of wood burners in new EPCs increased from 7.0% to 10.3%. The largest absolute increase was in detached homes, where prevalence rose from 15.2% in 2009 to 20.5% in 2024. In urban LSOAs, prevalence increased across all IMD deciles (Figure 4).

Temporal trends in the prevalence of wood burners were broadly consistent across IMD deciles and property types, though less deprived LSOAs saw a larger absolute increase, due to higher baseline prevalence of wood burners.

In rural areas, wood burners were more prevalent in lower and middle deprivation deciles. As in urban LSOAs, prevalence was highest in detached houses. Compared to urban areas, trends in wood burner prevalence in rural areas remained roughly static over time.

We found evidence that renovations to existing properties are contributing to the increasing prevalence of wood burners (Supplementary Table 8). Among properties with multiple EPCs, prevalence of wood burners increased with each subsequent EPC the property received. For example, among the roughly 700,000 detached properties with two EPCs, 19.4% had a wood burner at the time of first EPC compared to 24.3% at time of second EPC.

4 Discussion

4.1 Summary

This study was motivated by the growing literature linking residential wood burning to adverse health outcomes, the relative lack of evidence on the geographic and socio-economic distribution of wood burning in high income countries, and the unsuitability of existing wood burning inventories for targeting public health campaigns. Using EPCs, we found high variability in the geographic prevalence and concentration of wood burners. Wood burners were more prevalent in rural areas but more concentrated in densely populated urban areas. The concentration of wood burners varied substantially even within small urban areas. LSOA-level prevalence of wood burners was positively associated with age and percentage of residents identifying as ethnically white, and negatively associated with social deprivation and SCA status.

The NAEI wood burning inventory also identifies wood burning hotspots in urban areas.¹⁸ However, the NAEI hotspots are different to the areas with the highest concentration of wood burners identified using EPCs. NAEI maps identify northern English towns with relatively higher levels of social deprivation as emitting the highest quantity of PM_{2.5} from residential wood burning. By contrast, we identified affluent towns in southern England as having the highest concentration of wood burners.

Direct evaluation of our findings against the NAEI is challenging, since the methods target different quantities. The NAEI maps attempt to quantify the spatial distribution of wood burning emissions, whereas our method quantifies the spatial distribution of wood burning heat sources. However, our results are consistent with evidence on the socio-economic characteristics of wood burning from a recent survey ($N = 46,729$) of residential solid fuel burning in the United Kingdom conducted by Kantar.²³ The survey estimated that roughly 8% of UK households burned solid fuels. In the highest 'AB' social grade (a socio-economic classification based on employment type) the prevalence was 13%, compared to 4% in the lowest 'DE' social grade. The survey also found that prevalence of indoor solid fuel burning was higher among respondents identifying as ethnically white, and among homeowners.

Our results highlight clear differences in the socio-economic characteristics of wood burning between urban and rural areas, particularly with respect to social deprivation. These findings are consistent with previous research in Quebec, which found that median household income was positively associated with ownership of fireplaces (predominantly found in urban homes) and negatively associated with ownership of wood stoves (predominantly found in rural homes).¹⁷ More generally, the reasons for burning wood vary between urban and rural areas, with wood more commonly used as a primary heating source in rural areas.²³

Our finding that wood burning heat sources are most prevalent in areas with low social deprivation is interesting in the context of a recent trend analysis of lung cancer in non-smokers (LCINS) in the UK.³⁶ The study found that, contrary to overall trends, LCINS incidence increased for women living in the least deprived areas between 1998-2018. Women in the UK spend more time at home (on average) than men. Therefore, one explanation for rising rates of LCINS for women in affluent areas could be increased exposure to harmful pollutants from residential combustion.

The relationship between prevalence of wood burners and social deprivation in urban areas also

raises important issues of environmental justice. In affluent urban areas, households almost always have alternative heat sources available. In a segmentation analysis of the Kantar survey, fewer than 10% of households were classified as 'necessity' burners, where solid fuels were the main source of heating. The proportion of 'necessity' burners was lowest in the highest social grades and in urban households.

Spatial variability in emissions sources can generate socio-economic inequities in air pollution exposures.^{37,38} Discrepancies in air pollution exposures can have lasting effects - historic air pollution patterns have been linked to the East-West deprivation gradient in formerly industrial cities in England and Wales.³⁹ In urban areas where wood burners are highly concentrated, local populations are potentially being exposed to high levels of air pollution from wood burning which is largely discretionary.

4.2 Comparison against existing methods

High-resolution mapping of residential wood burning is challenging, since wood burning is a diffuse emissions sector. Existing residential wood burning maps are typically derived from national wood burning emissions inventories. In the absence of point source activity data, these inventories typically integrate aggregated activity data (e.g. national emissions estimates) with various spatial proxy data sources including population density, building registries, census data, and household surveys on wood consumption.^{13–18,40,41}

Buildings registries and census data serve a similar function to our EPC data in that they are used to proxy the total number of wood burning heat sources within a specified geography. Household surveys are often used to proxy the spatial intensity of wood burning activity - for example via estimates of the quantity of wood fuel consumed by responding households.^{16,42–44}

Less frequently, surveys have also been used to calibrate disaggregated estimates of wood fuel users against higher-level geographies.¹⁸

Our EPC-based method has several advantages relative to existing inventory-based methods as a tool for local policymakers. Existing emissions maps combine multiple data sources with varying spatial granularity. For example, although census data is typically available at high geographic resolution, survey estimates are typically available at lower resolution due to sample size constraints.

To deal with spatial mismatch in input data, current methods typically present gridded estimates of wood burning emissions based on various smoothing assumptions. Smoothed grid square estimates may be insufficient to capture the high spatial variability in wood burning heat sources, and can be sensitive to the granularity of the input data.¹⁹ Furthermore, grid squares do not directly correspond to administrative or electoral boundaries which are used in local policymaking.

A major benefit of our EPC method is its high spatial granularity. This advantage is particularly pronounced in urban areas - the average area of a London LSOA is 0.32km², compared to the 1km² grid squares available via the NAEI. This increased spatial granularity is possible precisely because we do not rely on low-resolution input data sources, and therefore make minimal spatial smoothing assumptions. EPCs can also be easily linked to administrative boundaries relevant for policymaking.

A further advantage of EPCs is that the dataset is updated continuously as new EPCs are generated. In the UK, the EPC dataset is updated monthly. Many existing methods rely on data sources which are infrequently updated (e.g. Census data) or 'point-in-time' snapshots (e.g. one-off household surveys), which means that estimates may be out of date. By contrast, EPCs provide up-to-date information which can be used to track trends or outliers over time.

It is important to highlight that unlike national emissions inventories, our EPC-based method does not attempt to quantify the geographic distribution of $\text{PM}_{2.5}$ emissions from residential wood burning. Our method relies on the geographic distribution of wood burning heat sources as a proxy variable for emissions. Our key assumption in this analysis is that, at least within small areas, our measure is indeed a good proxy for wood burning emissions.

In our validation exercise, the wood burning concentration metric was correlated with peaks in winter $\text{PM}_{2.5}$ in urban areas, which gives some reassurance that our measure captures meaningful spatial variation in wood burning emissions. Interestingly, our concentration metric was better-correlated with peaks in winter $\text{PM}_{2.5}$ than the existing NAEI maps. High-resolution emissions maps based on granular input data have also outperformed national emissions inventories in other contexts.¹⁷ However, our results should not be interpreted as estimates for the true geographic distribution of wood burning emissions in England and Wales.

4.3 Limitations

Our analysis has several important limitations. Using EPCs, we cannot reliably differentiate between different appliances or fuel types. Emissions factors vary substantially between appliances. For example, open fires emit around nine times as much $\text{PM}_{2.5}$ per gigajoule burned as modern 'Ecodesign' closed stoves, which comply with EU energy efficiency standards.^{45,46} Our concentration maps may over-represent potential emissions in SCAs, since these areas are subject to legal restrictions on the types of appliance and fuels that can be used. In the Kantar survey, 53% of respondents in urban SCAs reported owning an SCA-exempt appliance, compared to 24% of respondents in rural areas.

EPCs are also not informative about usage intensity. Our solid fuel prevalence estimates were generally higher than the percentage of respondents who reported burning solid fuels in the

Kantar survey. For example, we estimated that 12.8% of properties in the South East had a solid fuel heat source, whereas 9.3% of survey respondents in the South East reported using a solid fuel heat source for indoor heating. The discrepancy between our results and the survey estimates could reflect the fact that some households with solid fuel heat sources may choose not to use them. Previous work has found a higher intensity of wood burning in rural areas.^{17,47} Our results could understate the relative importance of rural wood burning for ambient air pollution, since they do not model geographic differences in burning intensity.

EPCs remain valid for ten years, meaning that existing EPC data may be out of date. We found evidence of increasing prevalence of wood burners in properties with multiple EPCs (Supplementary Table 8). Reliance on out-of-date EPCs could lead us to systematically underestimate the prevalence of wood burners, particularly in rural areas where EPCs are updated infrequently. Furthermore, our concentration estimates could be inflated in urban areas undergoing rapid gentrification simply because the data are (on average) more recent.

EPCs are a biased sample of the UK housing stock. Our method to correct for missing EPC data relies on an assumption that within LSOAs and housing types, EPC missingness is uncorrelated with wood burner presence. Whilst we argue that this is a relatively weak assumption given the homogeneity of LSOAs, it is possible that there is still some residual selection bias.

Our validation exercise does not consider weather variables such as prevailing wind direction or wind speed, which could potentially confound the relationship between our proxy concentration measure and monitor-level $\text{PM}_{2.5}$. For example, if the number of wood burners within buffer zones is correlated with their spatial distribution (e.g. in buffers with more wood burners, a higher proportion of those wood burners are located upwind of the monitor) then our method could give biased results by assuming a uniform distribution of heat sources around the monitoring site. However, these biases could only appear if the concentration of wood burners within the buffer was systematically correlated with the weather variables. We argue

that this is unlikely, given that we use a relatively large sample of monitoring sites and restrict our analysis to urban background sites.

Observed correlations with differences in peak vs. non-peak $\text{PM}_{2.5}$ could reflect the influence of confounding pollutants - for example, air pollution from residential cooking. However, the weak relationship between housing density and differences in peak and non-peak $\text{PM}_{2.5}$ provides some evidence that our measure extracts signal independent to the effects of these potential confounders.

4.4 Policy implications and future research

The high spatial heterogeneity in our estimated concentration of wood burners validates a targeted approach to intervention. Policymakers can use EPCs to extend existing SCAs or create new ones. Linkage of EPCs to local socio-economic characteristics can facilitate local interventions and awareness campaigns. Moreover, our EPC-based method is portable to other countries which make similar EPC data openly available.

To-date, there has been relatively little work to quantify the relationship between residential wood burning and socio-economic characteristics in high income countries. It is difficult to compare findings across contexts, since wood burning behaviours can vary substantially across countries.^{47,48} Future work should seek to characterise the relationship between socio-economic characteristics and residential wood burning in other high income contexts.

Local topography and climate will influence how wood burning emissions are distributed, and hotspots for air pollution exposures may differ from hotspots for source emissions. EPC data on the geographic distribution of wood burning and solid fuel heat sources could be a valuable addition into routine air quality modelling. Policymakers should also consider expanding air

quality monitoring networks into diverse urban areas, given the potentially high variation in air pollution within these areas.

Systematic reviews in higher-income countries have identified links between wood smoke exposure and adverse health outcomes, particularly for children.^{7,8} A large prospective US cohort study of 50,226 women reported that using wood for heating for more than 30 days per year increased the incidence of lung cancer by 68% (95% CI: 27-220%).¹ A recent time-series study in London found that short-term exposure to increased levels of wood smoke was associated with an increase in risk of respiratory mortality of 1.70% per interquartile range increase in three-day atmospheric concentration of wood burning carbon (95% CI: 0.64-2.27%).⁹ However, robust evidence on the causal effect of wood smoke exposure on health outcomes is lacking.

Linkage of EPCs with air quality data and national health datasets could advance our understanding of wood smoke’s impact on health. Robust evidence is needed on the health impacts of indoor air pollution exposures, which is particularly pertinent given the high prevalence of wood burners in rural areas.⁴⁹

Our work highlights the potentially important confounding effects of age, social deprivation, urbanicity, and ethnicity on the relationship between wood smoke exposures and health. These factors are not routinely accounted for in existing epidemiological analyses.^{7,50} Future studies seeking to understand the health impacts of wood smoke exposures should develop careful causal inference designs accounting for socio-economic confounding.

5 Conclusion

In this study, we used data from 26 million EPCs to characterise the geographic distribution of wood burners across England and Wales at fine scale. The size and granularity of the EPC data allowed us to detect substantial spatial heterogeneity in the concentration of wood burners, even within small urban areas. Linking EPCs to other open data sources, we described previously-understudied relationships between the prevalence of wood burners and socio-economic characteristics including age, deprivation, and ethnicity.

Our method relied on open data sources, and we generated a publicly-available, reproducible code base which can be used to fully replicate the analysis. Many countries make similar EPC data openly available, which means our method is portable to other contexts.

Public health officials could use EPCs to identify potential wood burning hotspots, target local awareness campaigns, or to extend legal restrictions on solid fuel burning. Our EPC-based concentration measure was better-correlated with winter peaks in urban $\text{PM}_{2.5}$ than existing UK National Atmospheric Emissions Inventory maps, suggesting that EPCs could be a valuable addition to routine air quality modelling.

Our work highlights important issues of environmental justice. Wood burners were most concentrated in affluent urban areas, where households almost always have alternative heat sources available. Moreover, our results show the importance of accounting for socio-economic confounding effects in analyses of wood smoke exposures and health.

6 Data Availability

This study used data on all EPCs made available to registered researchers up to February 2025 under the Open Government Licence v3.0, including Ordnance Survey UPRNs. EPC data is available at: <https://epc.opendatacommunities.org/>.

We used OS AddressBase data under an Ordnance Survey research licence. To facilitate replication of the analysis, we directly provide the summarised OS AddressBase data used to produce the final tables in the manuscript as part of our data files. All other data are from open or public data sources available under the Open Government Licence. All non-EPC open data used in the research are available here: <https://zenodo.org/records/14640852>.

All R code used for the data cleaning, analysis, and reporting is available here: https://github.com/UCL-Wellcome-Trust-Air-Pollution/EPC_mapping_project_code. All analyses and data visualisations were created using R version 4.4.1.

7 Acknowledgements

The authors would like to thank the Ordnance Survey for providing access to the OS AddressBase dataset and to the National Atmospheric Emissions Inventory for providing access to grid square wood burning emissions estimates.

8 Author Contribution Statement

Conceptualization: LJH, PH, and CK; methodology: LJH, CK, JH, JT, AD, NS, ML, and PH; software: LJH, CK, and AD; formal analysis: LJH and CK; investigation: LJH, CK, JH,

JT, AD, NS, ML, and PH; resources: LJH; data curation: CK and LJH; writing—original draft preparation: LJH and CK; writing—review and editing: JH, JT, AD, NS, ML, and PH; visualization: LJH and CK; supervision: LJH; project administration: LJH; funding acquisition: LJH.

9 Funding

This research was funded in whole or in part by the Wellcome Trust [Grant number 225195/Z/22/Z]. For the purpose of open access, the author has applied a CC by public copyright license to any Author Accepted Manuscript version arising from this submission. ML is supported by a David Phillips Fellowship from the Biotechnology and Biological Sciences Research Council (BB/V004573/1) and funding from the NIHR Southampton Biomedical Research Centre.

10 Ethical Approval

This study involved retrospective analysis of primarily open data sources, and did not involve human or animal experiments. Therefore, no ethical approval was required.

11 Competing Interests

The authors declare no competing interests.

References

1. Mehta SS, Elizabeth Hodgson M, Lunn RM, et al. Indoor wood-burning from stoves and fireplaces and incident lung cancer among Sister Study participants. *Environment International*. 2023 Aug 1; 178:108–28. DOI: 10.1016/j.envint.2023.108128
2. Lee KK, Bing R, Kiang J, et al. Adverse health effects associated with household air pollution: a systematic review, meta-analysis, and burden estimation study. *The Lancet Global Health*. 2020 Nov 1; 8. Publisher: Elsevier:1427–34. DOI: 10.1016/S2214-109X(20)30343-0
3. World Health Organisation. Household air pollution. 2024. Available from: <https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health>
4. Butt EW, Turnock ST, Rigby R, et al. Global and regional trends in particulate air pollution and attributable health burden over the past 50 years. *Environmental Research Letters* 2017; 12. DOI: 10.1088/1748-9326/aa87be
5. European Commission. Database - Energy - Eurostat. 2024. Available from: <https://ec.europa.eu/eurostat/web/energy/database>
6. Booth M. Burning up the carbon sink: How the EU’s forest biomass policy undermines climate mitigation, and how it can be reformed. Tech. rep. Partnership for Policy Integrity, 2022. Available from: <https://forestdefenders.eu/wp-content/uploads/2022/11/PFPI-Burning-up-the-carbon-sink-Nov-7-2022.pdf>
7. Guercio V, Doutsis A, and Exley KS. A systematic review on solid fuel combustion exposure and respiratory health in adults in Europe, USA, Canada, Australia and New Zealand. *International Journal of Hygiene and Environmental Health*. 2022 Apr; 241:113926. DOI: 10.1016/j.ijheh.2022.113926

8. Guercio V, Pojum IC, Leonardi GS, et al. Exposure to indoor and outdoor air pollution from solid fuel combustion and respiratory outcomes in children in developed countries: a systematic review and meta-analysis. *The Science of the Total Environment*. 2021 Feb 10; 755(Pt 1). DOI: 10.1016/j.scitotenv.2020.142187
9. Piper R, Tremper A, Katsouyanni K, et al. Associations between short-term exposure to airborne carbonaceous particles and mortality: A time-series study in London during 2010–2019. *Environmental Pollution* 2024 Nov; 360:124720. DOI: 10.1016/j.envpol.2024.124720
10. Savolahti M, Karvosenoja N, Soimakallio S, Kupiainen K, Tissari J, and Paunu VV. Near-term climate impacts of Finnish residential wood combustion. *Energy Policy*. 2019 Oct 1; 133:110837. DOI: 10.1016/j.enpol.2019.06.045
11. Searchinger TD, Beringer T, Holtsmark B, et al. Europe’s renewable energy directive poised to harm global forests. *Nature Communications*. 2018 Sep 12; 9. Publisher: Nature Publishing Group:3741. DOI: 10.1038/s41467-018-06175-4
12. Johnston FH, Hanigan IC, Henderson SB, and Morgan GG. Evaluation of interventions to reduce air pollution from biomass smoke on mortality in Launceston, Australia: retrospective analysis of daily mortality, 1994–2007. *BMJ*. 2013; 346(jan08 12):e8446–e8446. DOI: 10.1136/bmj.e8446
13. Karvosenoja N. Emission scenario model for regional air pollution. Teknillinen korkeakoulu, 2008. Available from: <https://aaltodoc.aalto.fi/handle/123456789/4537>
14. Lopez-Aparicio S, Grythe H, and Vogt M. Model development for high-resolution emissions from residential wood combustion. NILU, 2018
15. Borchers-Arriagada N, Vander Hoorn S, Cope M, et al. The mortality burden attributable to wood heater smoke particulate matter (PM_{2.5}) in Australia. *Science of The Total Environment*. 2024; 921:171069. DOI: 10.1016/j.scitotenv.2024.171069

16. U.S. Environmental Protection Agency. 2020 National Emissions Inventory Technical Support Document: Fuel Combustion – Residential Heating -Wood. 2023. Available from: https://www.epa.gov/system/files/documents/2023-03/NEI2020__TSD__Section27__RWC.pdf
17. Zalzal J, Liu Y, Smargiassi A, and Hatzopoulou M. Improving residential wood burning emission inventories with the integration of readily available data sources. *Science of The Total Environment* 2024; 946:174226. DOI: <https://doi.org/10.1016/j.scitotenv.2024.174226>
18. Tsagatakis I, Richardson J, Evangelides C, et al. UK Spatial Emissions Methodology - A report of the National Atmospheric Emission Inventory 2022. Tech. rep. Ricardo Energy and Environment, 2024. Available from: <https://naei.energysecurity.gov.uk/reports/uk-spatial-emissions-methodology-report-national-atmospheric-emission-inventory-2022>
19. Vedrenne M, Borge R, Lumbreras J, et al. A comprehensive approach for the evaluation and comparison of emission inventories in Madrid. *Atmospheric Environment*. 2016; 145:29–44. DOI: 10.1016/j.atmosenv.2016.09.020
20. McLean EV, Bagchi-Sen S, Atkinson JD, and Schindel A. Household dependence on solid cooking fuels in Peru: An analysis of environmental and socioeconomic conditions. *Global Environmental Change*. 2019; 58:101961. DOI: 10.1016/j.gloenvcha.2019.101961
21. Qiu S, Chen X, Chen X, et al. Solid fuel use, socioeconomic indicators and risk of cardiovascular diseases and all-cause mortality: a prospective cohort study in a rural area of Sichuan, China. *International Journal of Epidemiology*. 2022; 51:501–13. DOI: 10.1093/ije/dyab191
22. Emissions of air pollutants. GOV.UK. 2025 Mar 13. Available from: <https://www.gov.uk/government/statistics/emissions-of-air-pollutants> [Accessed on: 2025 Apr 19]

23. Department for Environment, Food and Rural Affairs. Burning in UK Homes and Gardens: Research Report. 2020. Available from: https://uk-air.defra.gov.uk/library/reports?report_id=1014
24. Department for Levelling Up, Housing & Communities. Energy Performance of Buildings Data England and Wales. 2024. Available from: <https://epc.opendatacommunities.org/>
25. Noble S, McLennan D, Noble M, et al. English indices of deprivation 2019: research report. Tech. rep. Ministry of Housing, Communities and Local Government, 2019. Available from: https://assets.publishing.service.gov.uk/media/5d8b364ced915d03709e3cf2/IoD2019_Research_Report.pdf
26. Welsh Index of Multiple Deprivation 2019: analysis relating to areas of deep-rooted deprivation | GOV.WALES. 2022. Available from: <https://www.gov.wales/welsh-index-multiple-deprivation-2019-analysis-relating-areas-deep-rooted-deprivation-html>
27. Office for National Statistics. Ethnic group, England and Wales - Office for National Statistics. 2022. Available from: <https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/bulletins/ethnicgroupenglandandwales/census2021>
28. Department for Environment, Food and Rural Affairs. Rural Urban Classification. 2021. Available from: <https://www.gov.uk/government/collections/rural-urban-classification>
29. Greater London Authority. London Atmospheric Emissions Inventory (LAEI) 2019 - London Datastore. en-US. 2023. Available from: <https://data.london.gov.uk/dataset/london-atmospheric-emissions-inventory--laei--2019>
30. Department for Environment, Food and Rural Affairs. Automatic Urban and Rural Network (AURN). 2024. Available from: <https://uk-air.defra.gov.uk/networks/network-info?view=aurn>
31. Air Quality England. About Air Quality England. 2020. Available from: <https://www.airqualityengland.co.uk/about>

32. Air Quality in Wales. Maps and Data. (no date). Available from: <https://www.airquality.gov.wales/index.php/maps-data>
33. Carslaw DC and Ropkins K. openair — An R package for air quality data analysis. *Environmental Modelling & Software* 2012; 27–28:52–61. DOI: 10.1016/j.envsoft.2011.09.008
34. Office for National Statistics. National Statistics UPRN Lookup (May 2022). 2024. Available from: <https://geoportal.statistics.gov.uk/datasets/e75da94b5c7642ea920bba5e6a84baa7/about>
35. Efron B. Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association* 1987; 82:171–85. DOI: 10.1080/01621459.1987.10478410
36. Rait G and Horsfall L. Twenty-year sociodemographic trends in lung cancer in non-smokers: A UK-based cohort study of 3.7 million people. *Cancer Epidemiology* 2020 Aug; 67:101771. DOI: 10.1016/j.canep.2020.101771
37. Batisse E, Lloyd M, Cavanaugh A, et al. Examining the social distributions in neighbourhood black carbon and ultrafine particles in Montreal and Toronto, Canada. *Environment International*. 2025; 198:109395. DOI: 10.1016/j.envint.2025.109395
38. Venter ZS, Figari H, Krange O, and Gundersen V. Environmental justice in a very green city: Spatial inequality in exposure to urban nature, air pollution and heat in Oslo, Norway. *Science of The Total Environment*. 2023; 858:160193. DOI: 10.1016/j.scitotenv.2022.160193
39. Heblich S, Trew A, and Zylberberg Y. East-Side Story: Historical Pollution and Persistent Neighborhood Sorting. *Journal of Political Economy*. 2021 May; 129:1508–52. DOI: 10.1086/713101
40. Plejdrup MS, Nielsen OK, and Brandt J. Spatial emission modelling for residential wood combustion in Denmark. *Atmospheric Environment*. 2016; 144:389–96. DOI: 10.1016/j.atmosenv.2016.09.013

41. Kuenen J, Dellaert S, Visschedijk A, Jalkanen JP, Super I, and Denier van der Gon H. CAMS-REG-v4: a state-of-the-art high-resolution European emission inventory for air quality modelling. *Earth System Science Data*. 2022; 14:491–515. DOI: 10.5194/essd-14-491-2022
42. Tian YQ, Radke JD, Gong P, and Yu Q. Model development for spatial variation of PM_{2.5} emissions from residential wood burning. *Atmospheric Environment*. 2004; 38:833–43. DOI: 10.1016/j.atmosenv.2003.10.040
43. Gonçalves C, Alves C, and Pio C. Inventory of fine particulate organic compound emissions from residential wood combustion in Portugal. *Atmospheric Environment*. 2012; 50:297–306. DOI: 10.1016/j.atmosenv.2011.12.013
44. Grythe H, Lopez-Aparicio S, Vogt M, et al. The MetVed model: development and evaluation of emissions from residential wood combustion at high spatio-temporal resolution in Norway. *Atmospheric Chemistry and Physics*. 2019; 19. Publisher: Copernicus GmbH:10217–37. DOI: 10.5194/acp-19-10217-2019
45. Department for Environment, Food and Rural Affairs. Emission Factors for Small Combustion Appliances. 2021. Available from: <https://laqm.defra.gov.uk/air-quality/air-quality-assessment/combustion-emission-factors/>
46. Regulation (EU) 2024/1781 of the European Parliament and of the Council of 13 June 2024 establishing a framework for the setting of ecodesign requirements for sustainable products, amending Directive (EU) 2020/1828 and Regulation (EU) 2023/1542 and repealing Directive 2009/125/EC (Text with EEA relevance). Legislative Body: EP, CONSIL. 2024. Available from: <http://data.europa.eu/eli/reg/2024/1781/oj/eng>
47. Paunu VV, Karvosenoja N, Segersson D, et al. Spatial distribution of residential wood combustion emissions in the Nordic countries: How well national inventories represent local emissions? *Atmospheric Environment*. 2021 Nov 1; 264:118712. DOI: 10.1016/j.atmosenv.2021.118712

48. López-Aparicio S, Guevara M, Thunis P, Cuvelier K, and Tarrasón L. Assessment of discrepancies between bottom-up and regional emission inventories in Norwegian urban areas. *Atmospheric Environment*. 2017; 154:285–96. DOI: 10.1016/j.atmosenv.2017.02.004
49. Department of Health and Social Care. Chief Medical Officer’s annual report 2022: air pollution. GOV.UK. 2022. Available from: <https://www.gov.uk/government/publications/chief-medical-officers-annual-report-2022-air-pollution>
50. Rokoff LB, Koutrakis P, Garshick E, et al. Wood Stove Pollution in the Developed World: A Case to Raise Awareness Among Pediatricians. *Current problems in pediatric and adolescent health care*. 2017; 47:123–41. DOI: 10.1016/j.cppeds.2017.04.001

12 Figures

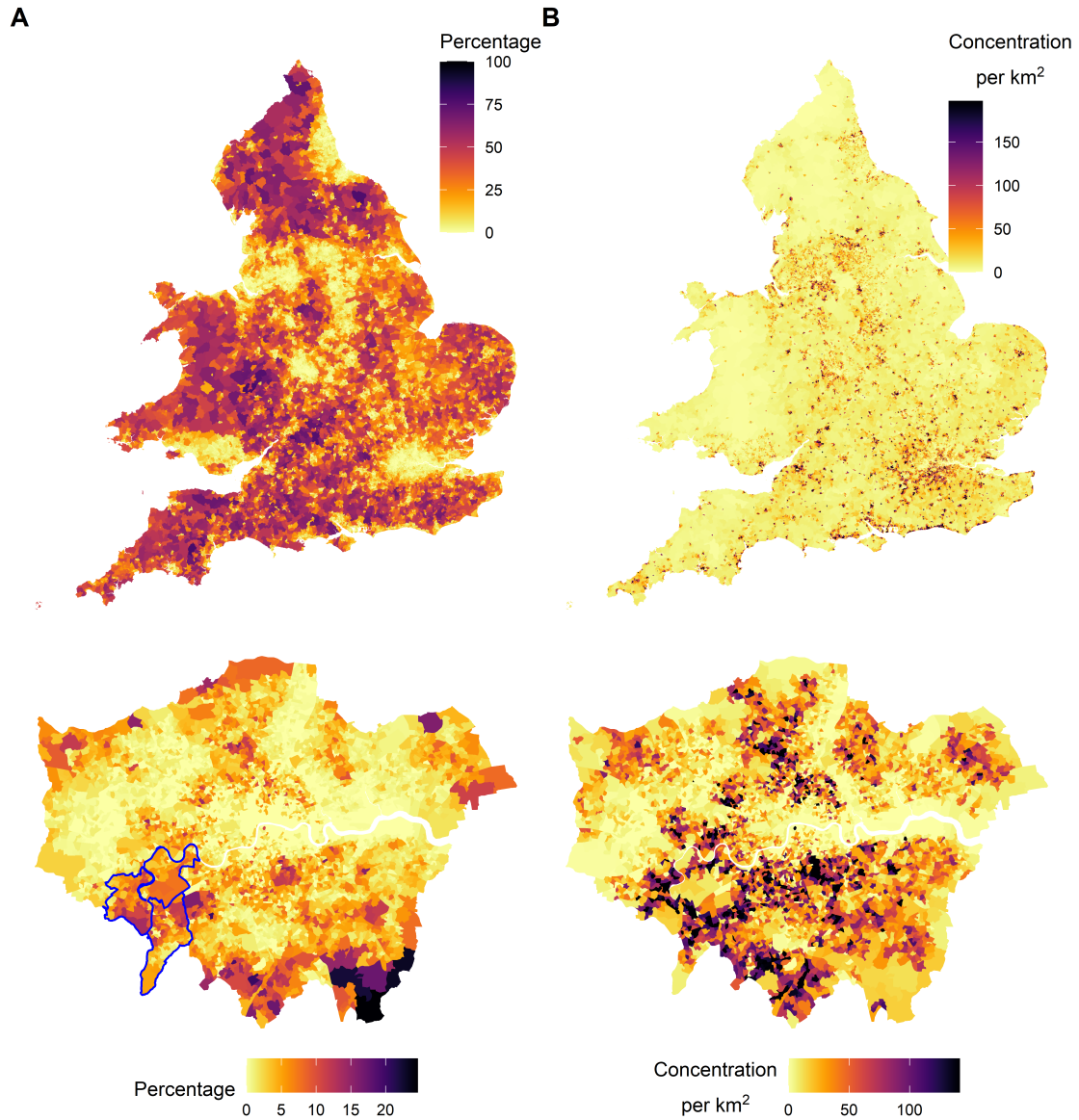


Figure 1: **A.** Top: Estimated prevalence of wood burners by LSOA, England and Wales. Bottom: Estimated prevalence of wood burners by LSOA, London. Local authorities of Richmond upon Thames, Kingston upon Thames, and Sutton highlighted in blue. **B.** Top: Estimated concentration of wood burners per km² by LSOA, England and Wales. Bottom: Estimated concentration of wood burners per km² by LSOA, London. Estimated concentration values were truncated at the 5th and 95th percentiles for mapping.

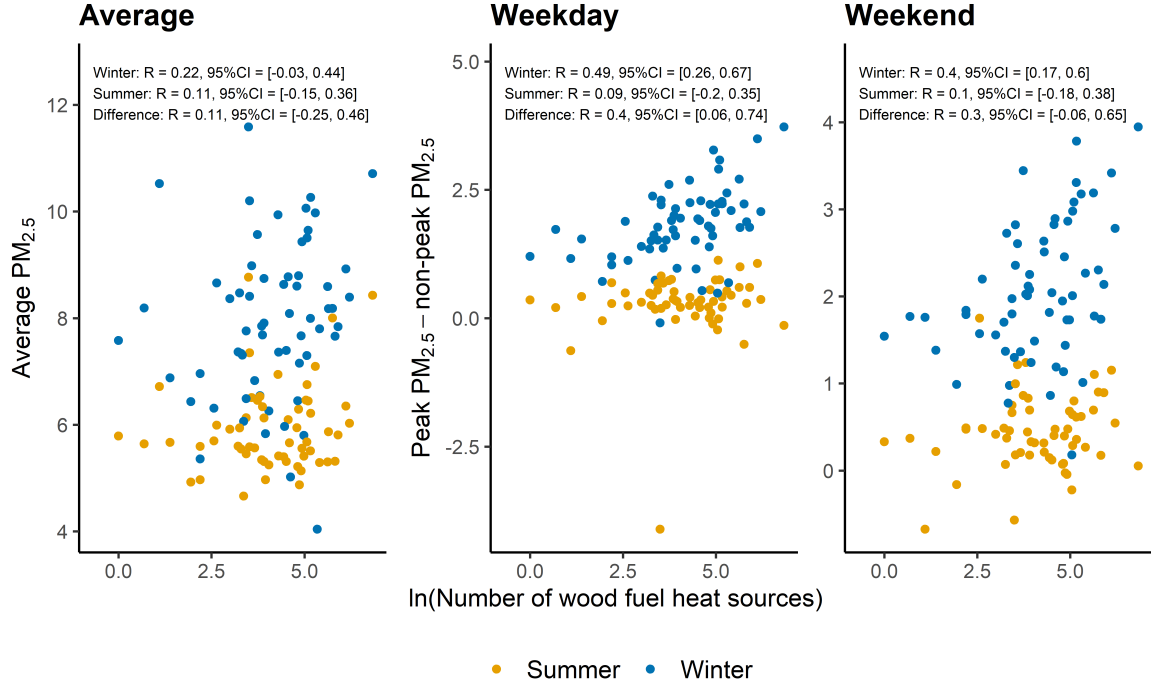


Figure 2: Left: Scatter plot of log estimated wood burner concentration against mean PM_{2.5} at monitoring sites. Centre: Scatter plot of log estimated wood burner concentration against difference in weekday mean PM_{2.5} measurements at peak vs. non-peak burning times. Right: Scatter plot of log estimated wood burner concentration against difference in weekend mean PM_{2.5} measurements at peak vs. non-peak burning times. Peak burning times were defined as 7pm-1am, and non-peak times as 5am-5pm. Weekends were defined as Saturday and Sunday. The concentration of wood burners was estimated as the number of properties with a wood burner in the EPC data within a 1km circular radius of each monitoring site. We used the subset of 'Urban Background' AURN monitoring sites. The scatter plots are stratified into summer and winter observations. Each chart reports the central estimate and 95% confidence interval for the Spearman correlation coefficient (R). We calculated the 95% confidence intervals using the BCA bootstrap method with 10,000 replications.

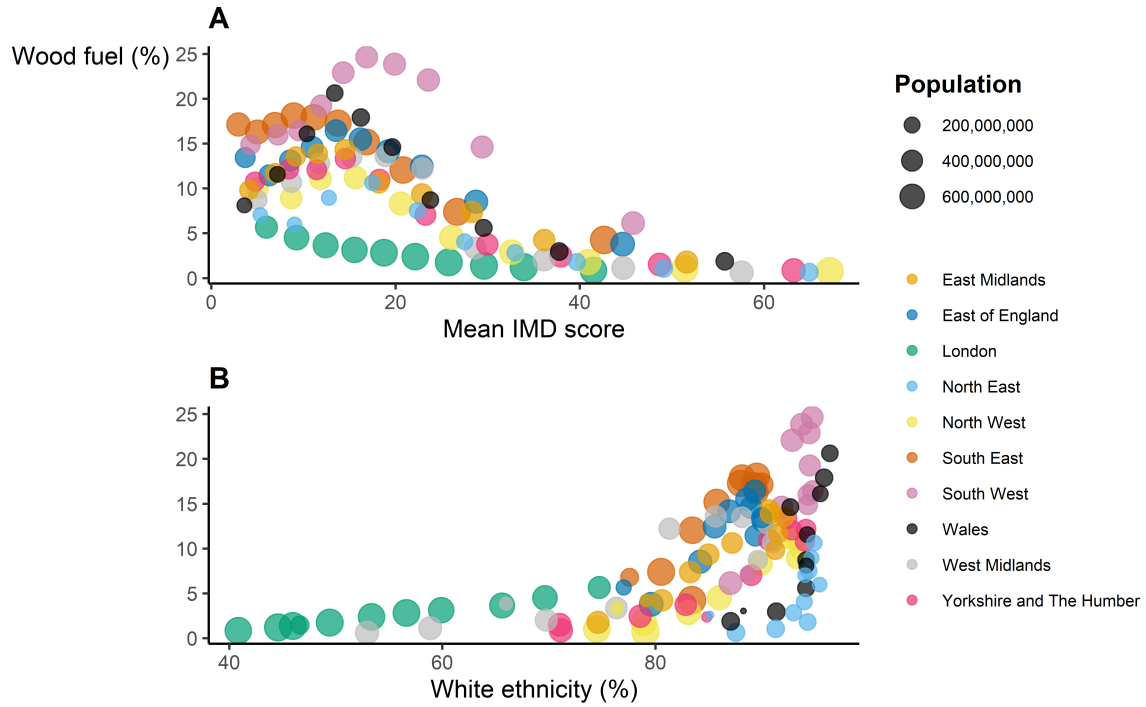


Figure 3: **A:** Association between mean IMD score and prevalence of wood burners. Within regions, we grouped LSOAs by IMD decile and calculated the mean IMD score and prevalence of wood burners for each region-IMD decile pair. Each point on the plot represents the mean IMD score and WF prevalence for a given region and IMD decile. The size of the point corresponds to the total population within that region and IMD decile. **B:** Association between percentage of people from a white ethnic background and prevalence of wood burners. We followed the same method as in **A** to generate the plot.

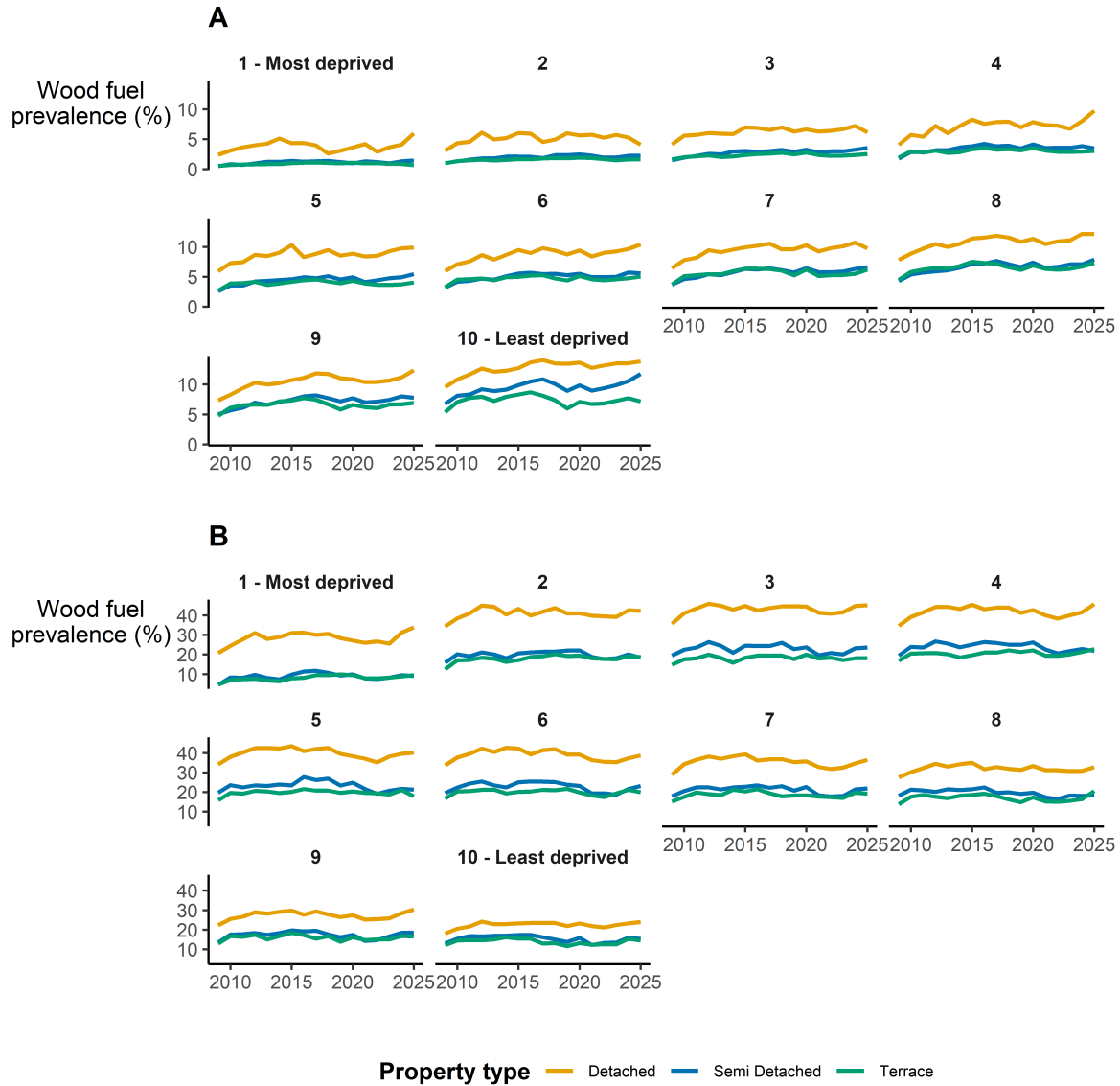


Figure 4: **A:** Prevalence of wood burners in new EPC certificates from 2009 - February 2025 for urban LSOAs, by property type and IMD decile. **B:** Prevalence of wood burners in new EPC certificates from 2009 - February 2025 for rural LSOAs, by property type and IMD decile. The lowest IMD decile (1) corresponds to the most deprived areas, and the highest (10) corresponds to the least deprived. IMD deciles were calculated within urban/rural categories - for example, in **A** deciles were calculated using only the subset of urban LSOAs. For each property type, IMD decile, and year, we calculated the percentage of new EPCs which indicated the presence of a wood burner. We restricted the analysis to the sub-sample of detached, semi detached, and terraced houses, excluding flats and other accommodation types. We used all EPCs available in the dataset, including multiple EPCs for the same property.

13 Tables

Table 1: Property characteristics by presence of wood burner in England and Wales

	Wood burner	
	No (N = 17,339,718)	Yes (N = 1,368,802)
Property type		
Detached	2,890,474 (17%)	679,637 (50%)
Flat	5,308,610 (31%)	26,946 (2.0%)
House Form Missing	53,891 (0.3%)	2,781 (0.2%)
Other accommodation	9,772 (<0.1%)	215 (<0.1%)
Semi Detached	4,295,286 (25%)	377,068 (28%)
Terrace	4,781,685 (28%)	282,155 (21%)
Tenure		
New build	738,240 (4.6%)	18,278 (1.4%)
Owner-occupied	8,707,524 (55%)	1,105,820 (85%)
Rented (private)	3,299,967 (21%)	165,375 (13%)
Rented (social)	3,143,415 (20%)	18,706 (1.4%)
Missing	1,450,572	60,623
Smoke Control Area		
No	8,427,370 (49%)	1,159,221 (85%)
Yes	8,912,348 (51%)	209,581 (15%)
Urban LSOA		
No	2,455,205 (15%)	746,353 (56%)
Yes	14,093,037 (85%)	594,822 (44%)
Missing	791,476	27,627

Note: We used the subset of unique properties in the EPC dataset. We excluded properties which had missing information on wood fuel presence. Each column shows the count of properties in the EPC database with the specified characteristic and wood fuel status, and the percentage of all properties of that wood fuel status with the specified characteristic. Missing values within characteristics are reported.

Table 2: Summary of socio-economic indicators by LSOA wood fuel prevalence decile

Decile	Wood fuel prevalence (%)	IMD score	White ethnic background (%)	Median age	Urban (%)
1	0.1	36.6	56.3	33.7	99.8
2	0.7	34.6	70.6	37.0	99.1
3	1.5	28.6	79.0	39.1	97.5
4	2.5	24.4	82.0	40.4	96.1
5	3.8	21.0	84.2	41.5	94.0
6	5.5	17.9	86.7	42.8	91.9
7	7.9	15.9	89.1	44.2	86.3
8	11.5	13.9	90.9	45.5	79.6
9	18.5	12.2	93.0	47.5	62.8
10	40.1	13.4	96.2	51.2	17.5

Note: The table presents summary statistics on LSOA-level socio-economic indicators grouped by deciles calculated using the estimated prevalence of wood burners by LSOA. The first decile represents LSOAs with the lowest prevalence of wood burners, and the tenth decile represents LSOAs with the highest prevalence of wood burners.

14 Supplementary materials

14.1 Figures

Corrected vs. uncorrected estimated concentration and prevalence of wood burners by LSOA

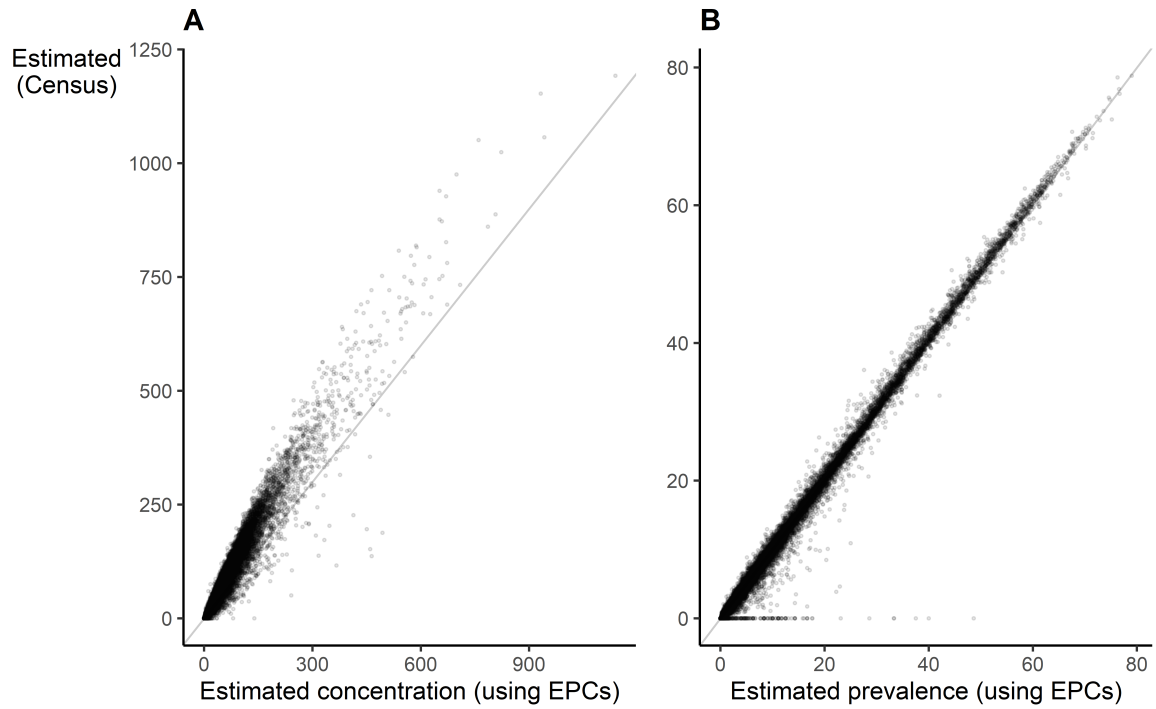


Figure 5: **A** - Scatter plot of estimated concentration of wood fuel heat sources per km² against estimated concentration using Census-based correction method by LSOA. **B** - Scatter plot of estimated prevalence of wood fuel heat sources against Census-based estimated prevalence by LSOA. Solid lines are 45 degree lines.

Maps of estimated prevalence and concentration of solid fuel heat sources by LSOA, England and Wales

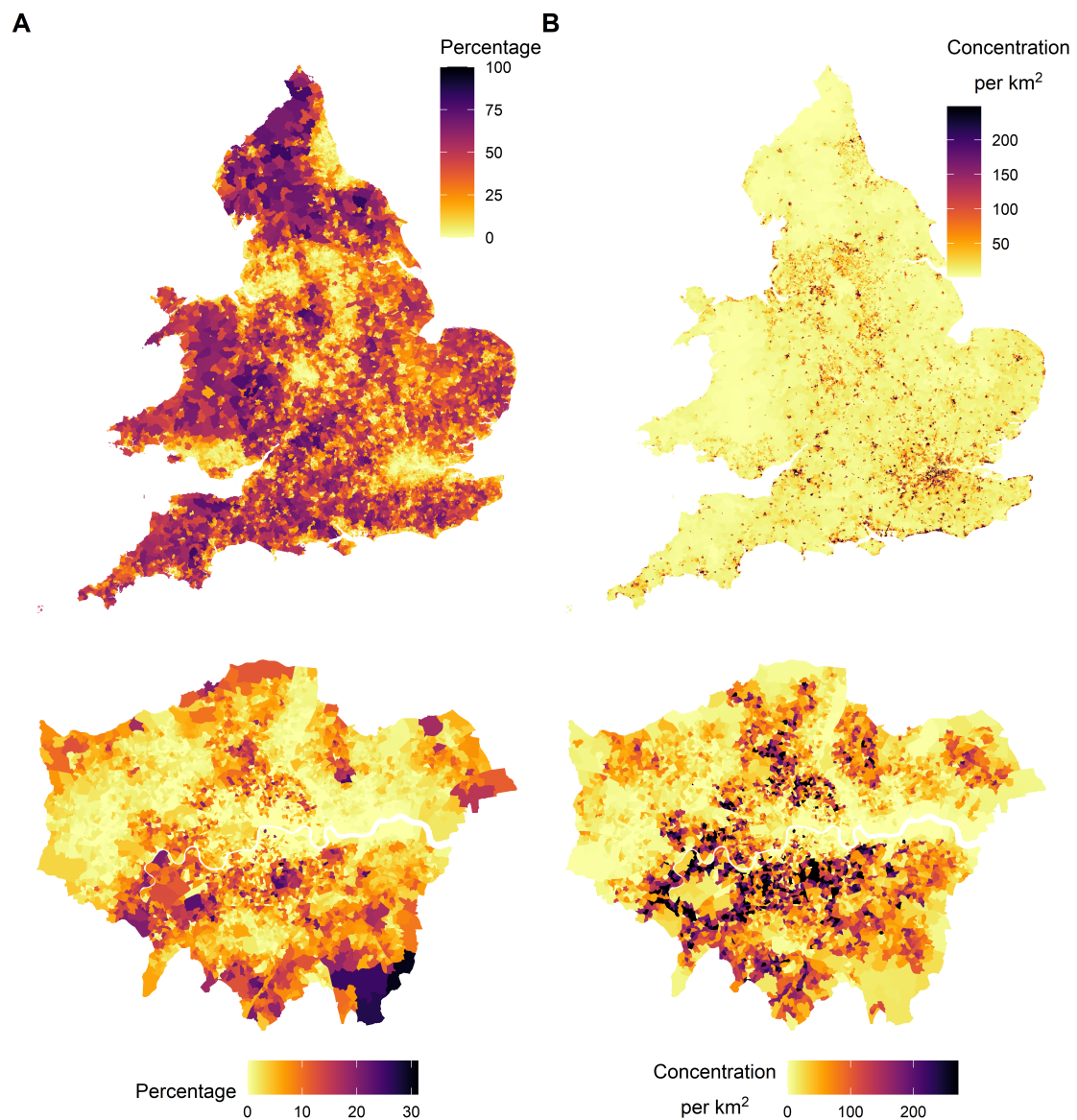


Figure 6: **A** - (Upper) Estimated prevalence of solid fuel heat sources by LSOA, England and Wales. (Lower) Estimated prevalence of solid fuel heat sources by LSOA, London. **B** - (Upper) Estimated concentration of solid fuel heat sources per km² by LSOA, England and Wales. (Lower) Estimated concentration of solid fuel heat sources per km² by LSOA, London. Estimated concentration values were truncated at the 5th and 95th percentiles for mapping.

Comparison of EPC concentration measure against LAEI gridded emissions

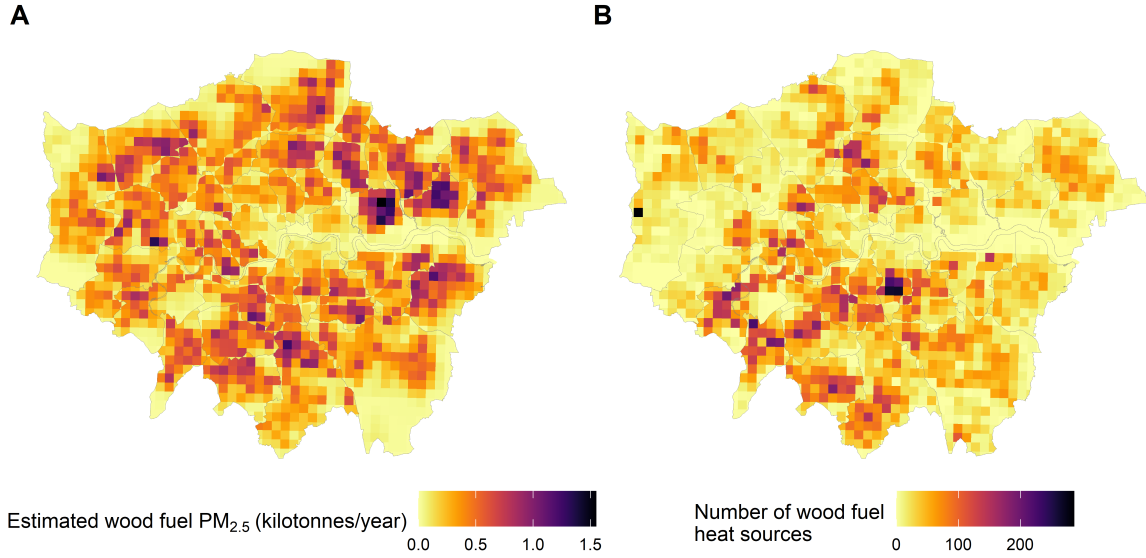


Figure 7: **A:** Estimated wood burning-derived PM_{2.5} emissions in Greater London by 1km x 1km grid square using LAEI methodology (tonnes/year). **B:** Estimated number of wood burners by 1km x 1km grid square, using EPCs. We used OS AddressBase data to estimate the number of residential properties by property type in each 1km x 1km grid square, and used EPCs to estimate the percentage of properties with a wood fuel heat source by grid square and property type. We estimated the total number of wood burners by grid square by taking the product of the estimated number of properties and wood fuel prevalence by property type, before summing across property types.

Scatter plot of log estimated concentration of wood burners against PM_{2.5} at AURN, Air Quality England, and Air Quality Wales monitoring sites

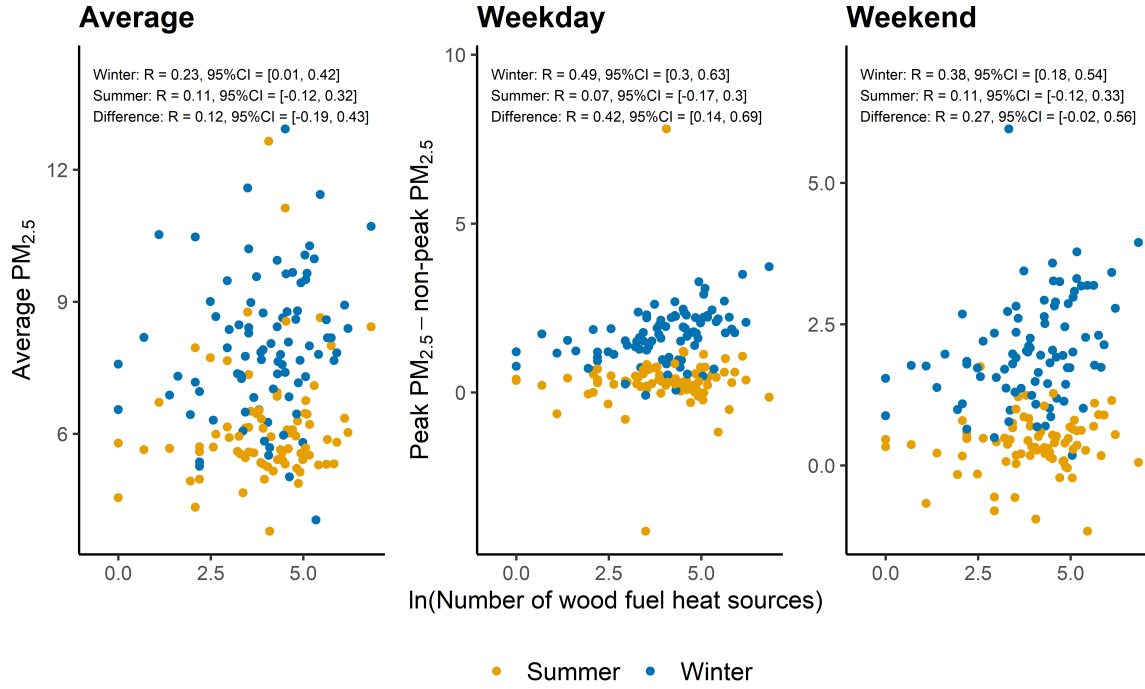


Figure 8: Left: Scatter plot of log estimated wood burner concentration against mean PM_{2.5} at monitoring sites. Centre: Scatter plot of log estimated wood burner concentration against difference in weekday mean PM_{2.5} measurements at peak vs. non-peak burning times. Right: Scatter plot of log estimated wood burner concentration against difference in weekend mean PM_{2.5} measurements at peak vs. non-peak burning times. Peak burning times were defined as 7pm-1am, and non-peak times as 5am-5pm. Weekends were defined as Saturday and Sunday. The concentration of wood burners was estimated as the number of properties with a wood burner in the EPC data within a 1km circular radius of each monitoring site. We used sites classified as 'Urban Background' from the AURN, Air Quality England, and Air Quality Wales monitoring sites. The scatter plots are stratified into summer and winter observations. Each chart reports the central estimate and 95% confidence interval for the Spearman correlation coefficient (ρ). We calculated the 95% confidence intervals using the BCA bootstrap method with 10,000 replications. We also report bootstrap 95% confidence intervals for the difference between the summer and winter correlation coefficients.

Scatter plot of log estimated concentration of wood burners against $PM_{2.5}$ at AURN urban background monitoring sites - 500m buffer

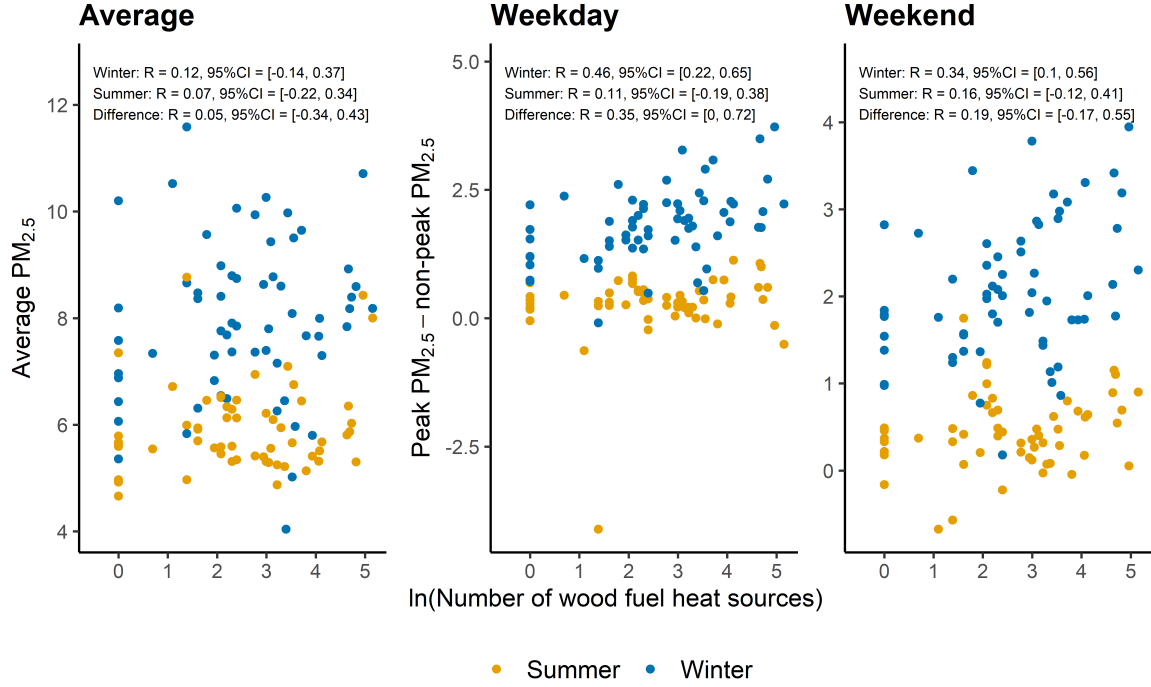


Figure 9: Left: Scatter plot of log estimated wood burner concentration against mean $PM_{2.5}$ at monitoring sites. Centre: Scatter plot of log estimated wood burner concentration against difference in weekday mean $PM_{2.5}$ measurements at peak vs. non-peak burning times. Right: Scatter plot of log estimated wood burner concentration against difference in weekend mean $PM_{2.5}$ measurements at peak vs. non-peak burning times. Peak burning times were defined as 7pm-1am, and non-peak times as 5am-5pm. Weekends were defined as Saturday and Sunday. The concentration of wood burners was estimated as the number of properties with a wood burner in the EPC data within a 1km circular radius of each monitoring site. We used sites classified as 'Urban Background' from the AURN. The scatter plots are stratified into summer and winter observations. Each chart reports the central estimate and 95% confidence interval for the Spearman correlation coefficient (ρ). We calculated the 95% confidence intervals using the BCA bootstrap method with 10,000 replications. We also report bootstrap 95% confidence intervals for the difference between the summer and winter correlation coefficients.

Scatter plot of log estimated concentration of wood burners against $PM_{2.5}$ at AURN urban background monitoring sites - 2km buffer

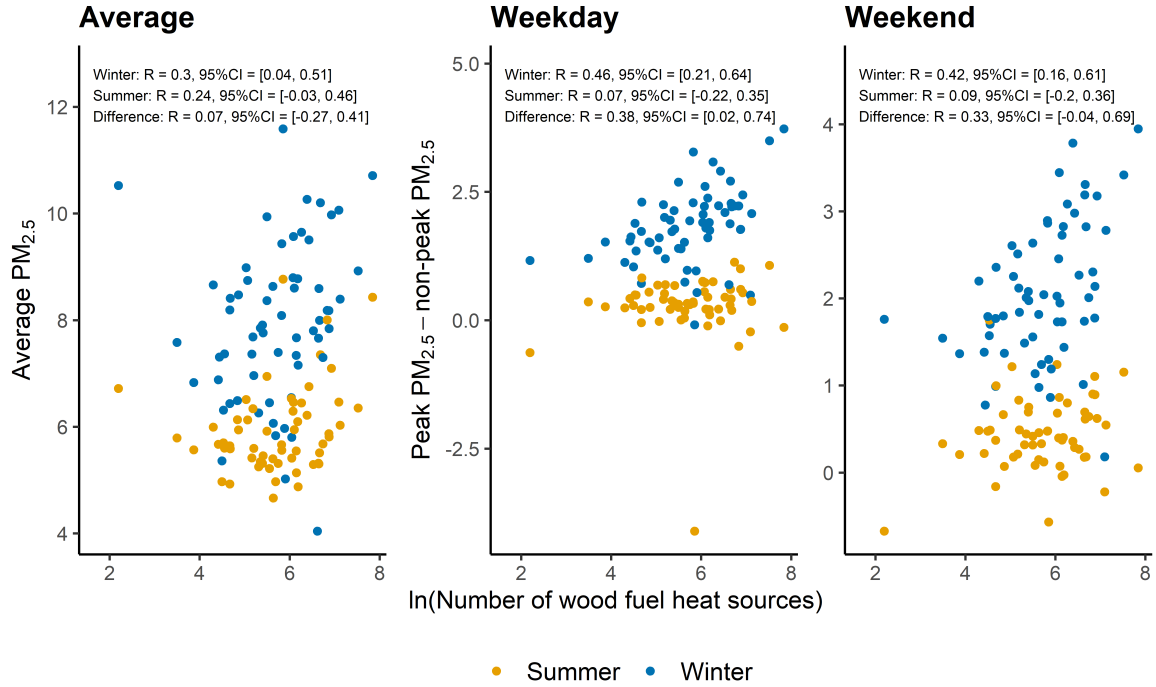


Figure 10: Left: Scatter plot of log estimated wood burner concentration against mean $PM_{2.5}$ at monitoring sites. Centre: Scatter plot of log estimated wood burner concentration against difference in weekday mean $PM_{2.5}$ measurements at peak vs. non-peak burning times. Right: Scatter plot of log estimated wood burner concentration against difference in weekend mean $PM_{2.5}$ measurements at peak vs. non-peak burning times. Peak burning times were defined as 7pm-1am, and non-peak times as 5am-5pm. Weekends were defined as Saturday and Sunday. The concentration of wood burners was estimated as the number of properties with a wood burner in the EPC data within a 1km circular radius of each monitoring site. We used sites classified as 'Urban Background' from the AURN. The scatter plots are stratified into summer and winter observations. Each chart reports the central estimate and 95% confidence interval for the Spearman correlation coefficient (ρ). We calculated the 95% confidence intervals using the BCA bootstrap method with 10,000 replications. We also report bootstrap 95% confidence intervals for the difference between the summer and winter correlation coefficients.

Scatter plot of log estimated housing density against PM_{2.5} at AURN monitoring sites

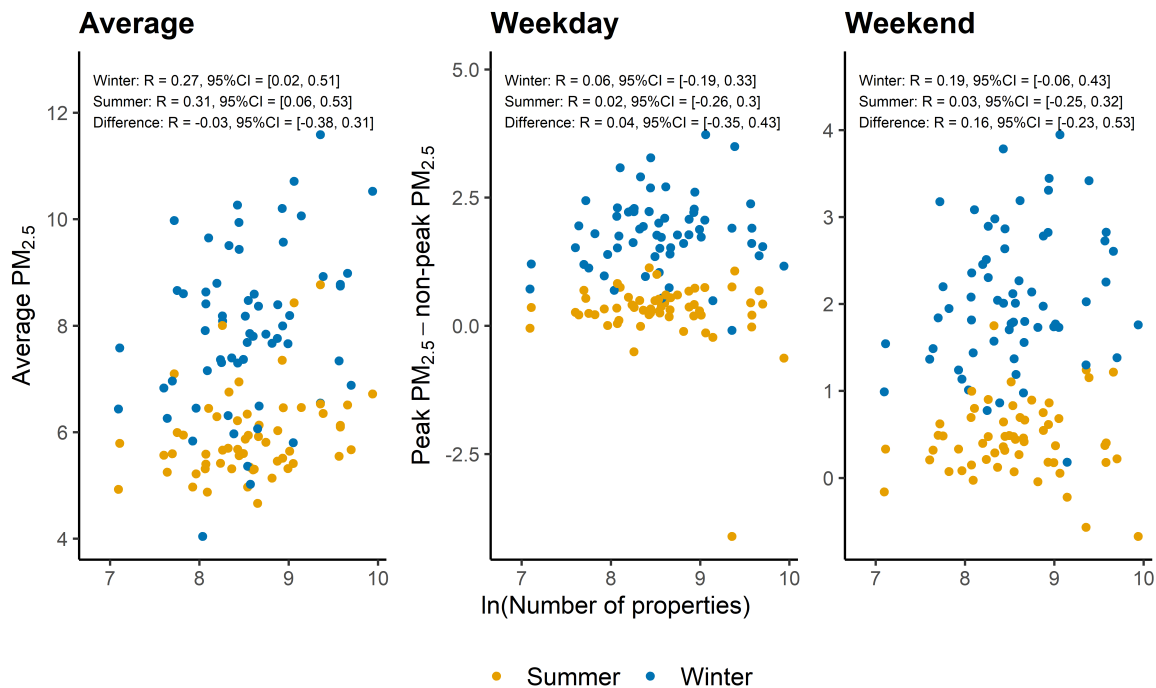


Figure 11: Left: Scatter plot of log estimated housing density against mean PM_{2.5} at monitoring sites. Centre: Scatter plot of log estimated housing density against difference in weekday mean PM_{2.5} measurements at peak vs. non-peak burning times. Right: Scatter plot of log estimated housing density against difference in weekend mean PM_{2.5} measurements at peak vs. non-peak burning times. Peak burning times were defined as 7pm-1am, and non-peak times as 5am-5pm. Weekends were defined as Saturday and Sunday. Housing density was estimated as the number of properties in the EPC data within a 1km circular radius of each monitoring site. We used all AURN monitoring sites classified as 'Urban Background'. The scatter plots are stratified into summer and winter observations. Each chart reports the central estimate and 95% confidence interval for the Spearman correlation coefficient (ρ). We calculated the 95% confidence intervals using the BCA bootstrap method with 10,000 replications. We also report bootstrap 95% confidence intervals for the difference between the summer and winter correlation coefficients.

Scatter plot of log estimated concentration of wood burners against $PM_{2.5}$ at AURN monitoring sites - using NAEI grid squares

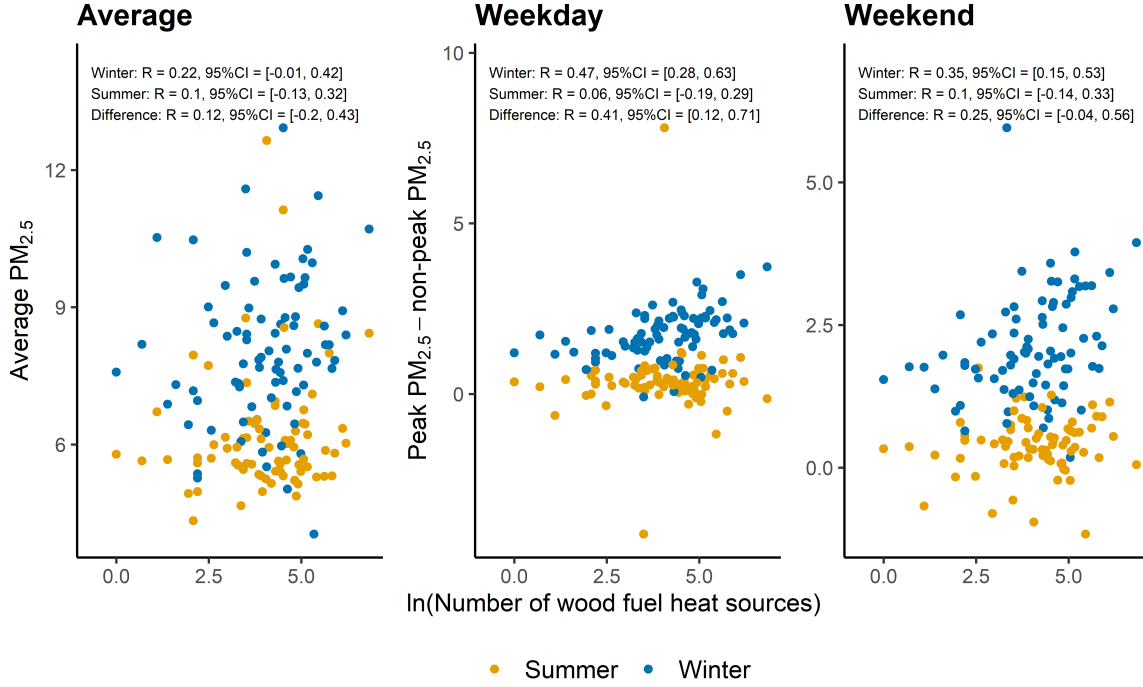


Figure 12: Left: Scatter plot of log estimated wood burner concentration against mean $PM_{2.5}$ at monitoring sites. Centre: Scatter plot of log estimated wood burner concentration against difference in weekday mean $PM_{2.5}$ measurements at peak vs. non-peak burning times. Right: Scatter plot of log estimated wood burner concentration against difference in weekend mean $PM_{2.5}$ measurements at peak vs. non-peak burning times. Peak burning times were defined as 7pm-1am, and non-peak times as 5am-5pm. Weekends were defined as Saturday and Sunday. The concentration of wood burners was estimated as the number of properties with a wood burner in the 1km by 1km grid square containing each monitoring site. We used all AURN monitoring sites classified as 'Urban Background'. The scatter plots are stratified into summer and winter observations. Each chart reports the central estimate and 95% confidence interval for the Spearman correlation coefficient (ρ). We calculated the 95% confidence intervals using the BCA bootstrap method with 10,000 replications. We also report bootstrap 95% confidence intervals for the difference between the summer and winter correlation coefficients.

Scatter plot of NAEI-implied wood burning emissions against PM_{2.5} at AURN monitoring sites - using NAEI grid squares

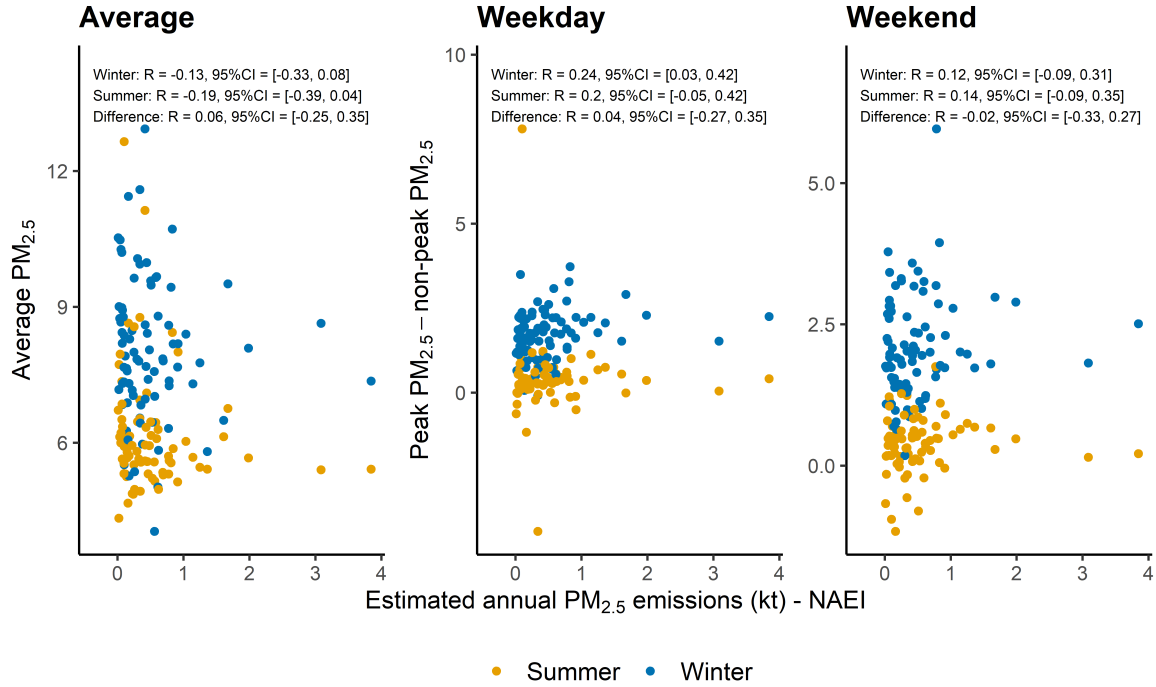


Figure 13: Left: Scatter plot of NAEI-estimated wood burning emissions against mean PM_{2.5} at monitoring sites. Centre: Scatter plot of NAEI-estimated wood burning emissions against difference in weekday mean PM_{2.5} measurements at peak vs. non-peak burning times. Right: Scatter plot of NAEI-estimated wood burning emissions against difference in weekend mean PM_{2.5} measurements at peak vs. non-peak burning times. Peak burning times were defined as 7pm-1am, and non-peak times as 5am-5pm. Weekends were defined as Saturday and Sunday. The NAEI wood burning emissions estimates were reproduced from grid square data downloaded from the NAEI. We used all AURN monitoring sites classified as 'Urban Background'. The scatter plots are stratified into summer and winter observations. Each chart reports the central estimate and 95% confidence interval for the Spearman correlation coefficient (ρ). We calculated the 95% confidence intervals using the BCA bootstrap method with 10,000 replications. We also report bootstrap 95% confidence intervals for the difference between the summer and winter correlation coefficients.

Association between social deprivation, ethnicity, and prevalence of wood burners by region, urban areas

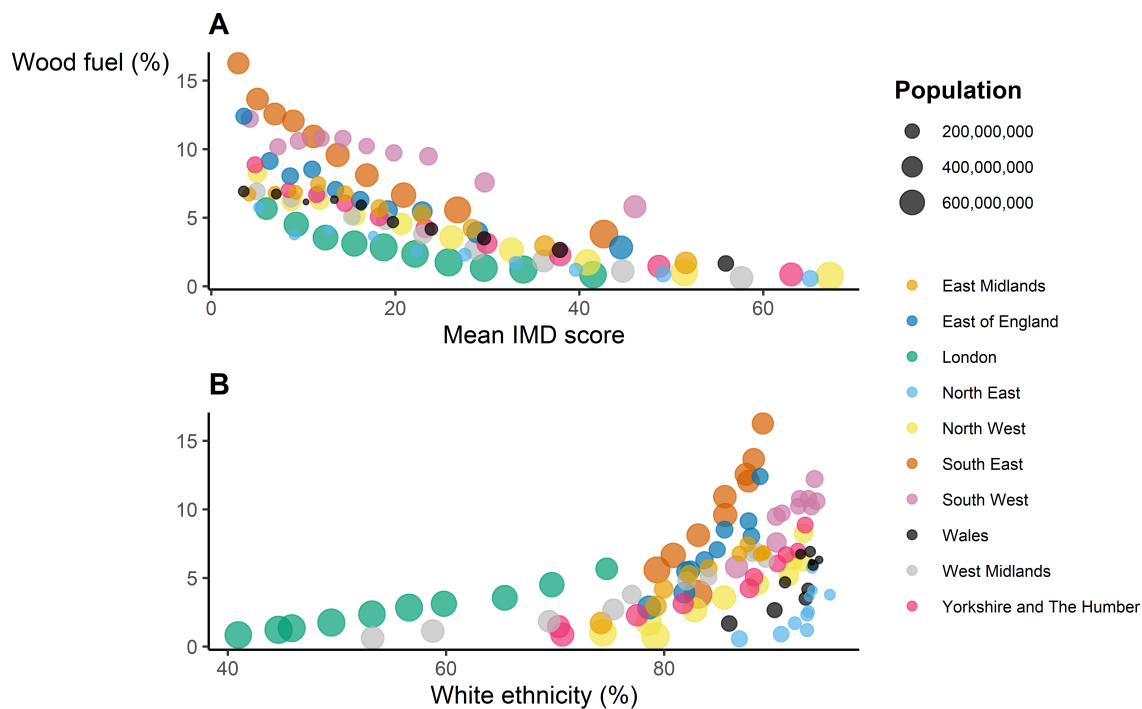


Figure 14: **A** - Association between mean IMD score and prevalence of wood fuel heat sources - urban areas. Within regions, we grouped urban LSOAs by IMD decile and calculated the mean IMD score and prevalence of wood fuel heat sources for each region-IMD decile pair. Each point on the plot represents the mean IMD score and wood fuel prevalence for urban LSOAs within a given region and IMD decile. The size of the point corresponds to the total population within that region and IMD decile. **B** - Association between percentage of people from a white ethnic background and prevalence of wood fuel heat sources - urban areas. We followed the same method as in **A** to generate the plot.

Association between social deprivation and prevalence of wood burners, by LSOA and urban/rural classification



Figure 15: Scatter plot of wood fuel prevalence against IMD score by region. Each point represents an LSOA.

Association between ethnicity and prevalence of wood burners, by LSOA and urban/rural classification

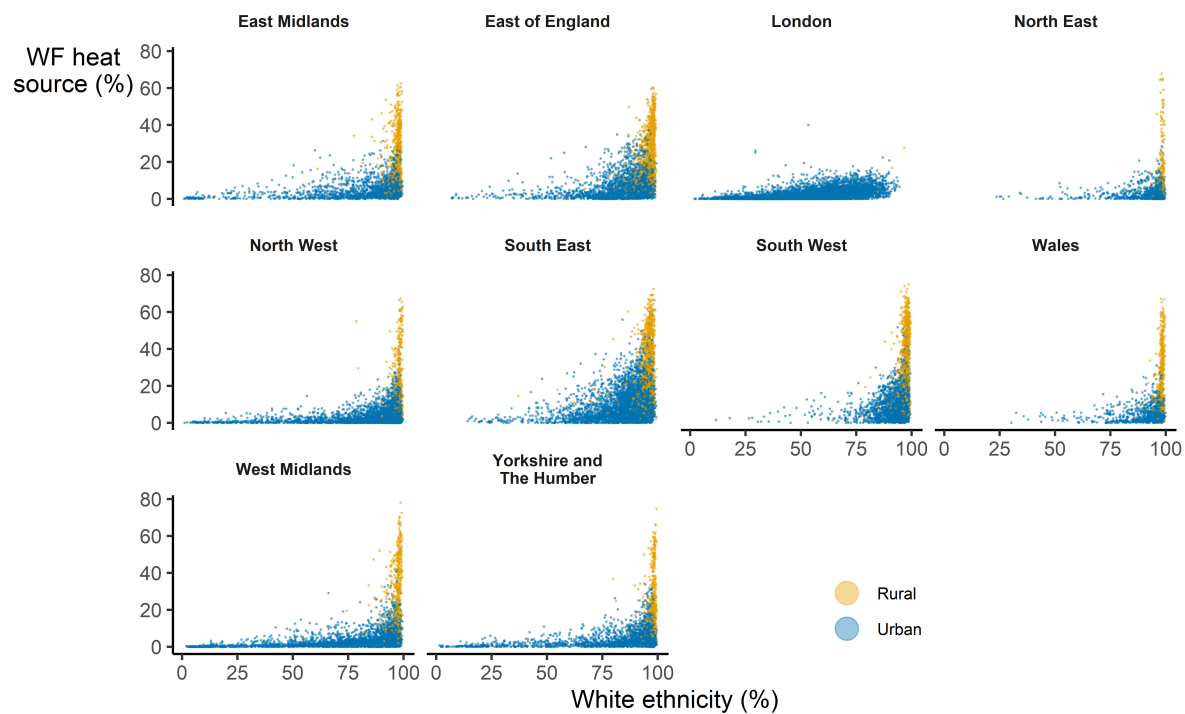


Figure 16: Scatter plot of wood fuel prevalence against percentage of people from a white ethnic background by region. Each point represents an LSOA.

Association between age and prevalence of wood burners, by LSOA and urban/rural classification

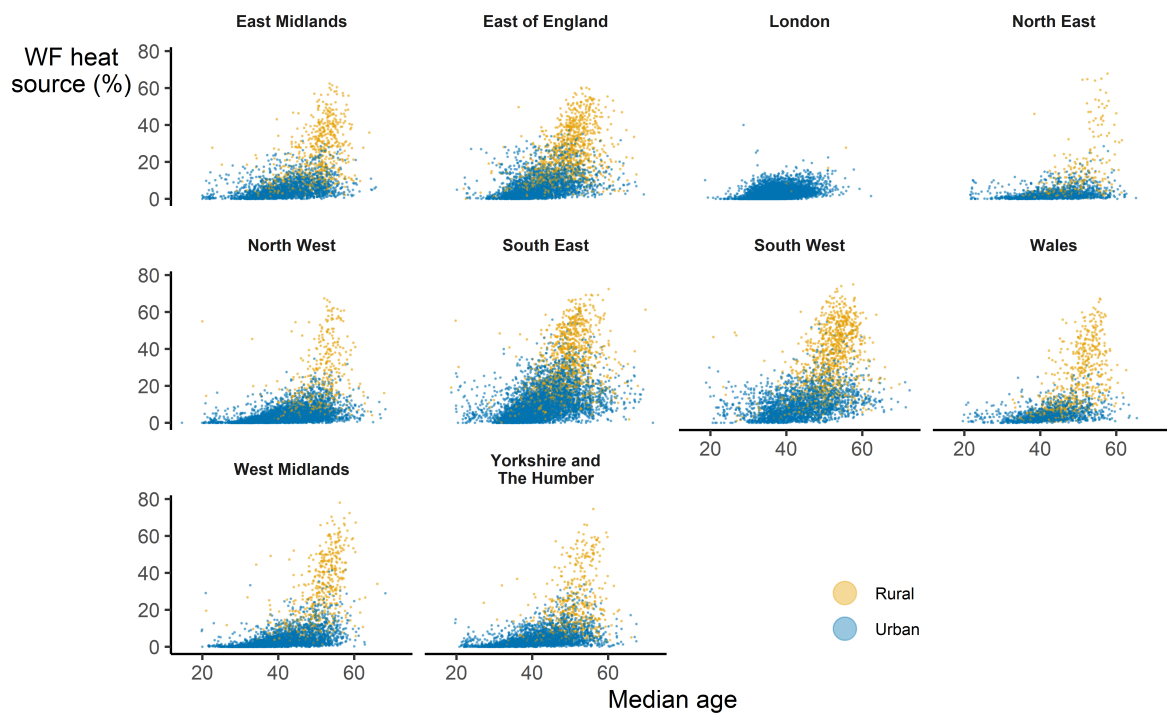


Figure 17: Scatter plot of wood fuel prevalence against median age by region. Each point represents an LSOA.

Temporal trends in prevalence of wood burners in EPCs, by property type and region

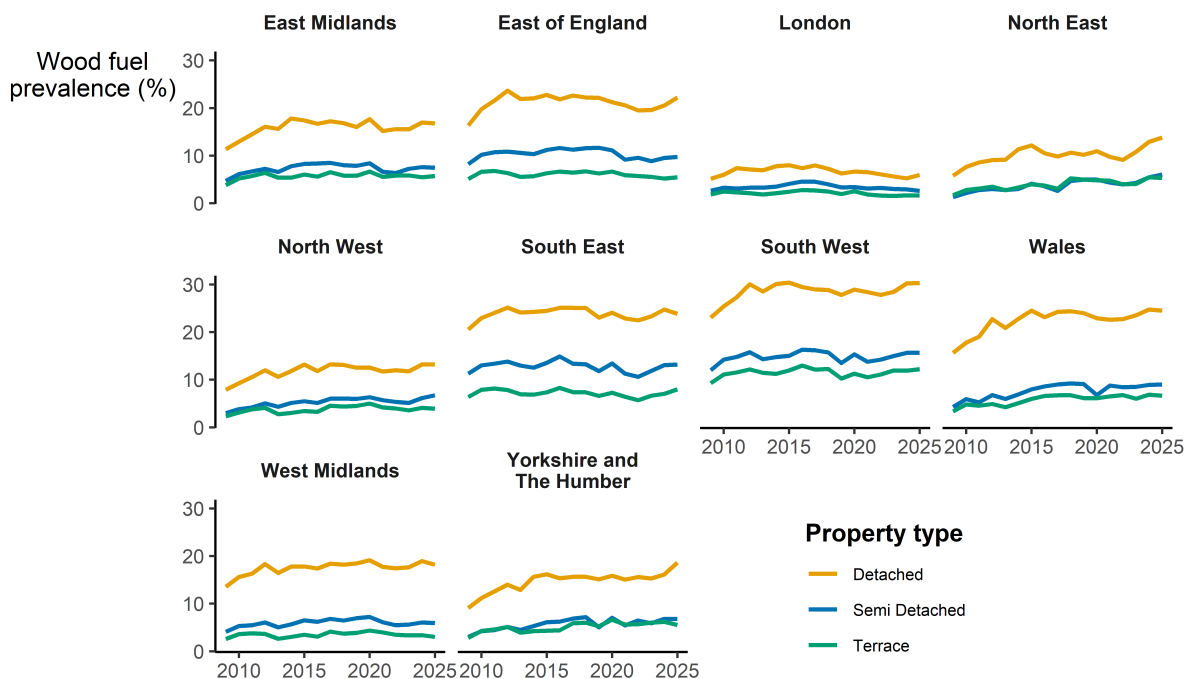


Figure 18: Prevalence of wood fuel heat sources in new EPC certificates from 2009-2025 by property type and region. For each property type, region, and year, we calculated the percentage of new EPCs which indicated the presence of a wood fuel heat source. We restricted the analysis to the sub-sample of detached, semi detached, and terraced houses, excluding flats and other accommodation types. We used all EPCs available in the dataset, including multiple EPCs for the same property.

14.2 Tables

Table 3: Percentage of properties in the OS AddressBase and EPC dataset, by housing type

Property type	OS AddressBase		EPC data	
	N	Percentage	N	Percentage
Detached	5,830,768	21.0%	3,498,984	19.1%
Semi Detached	7,430,980	26.8%	4,590,086	25.0%
Terrace	7,603,717	27.4%	4,997,789	27.2%
Flat	6,606,064	23.8%	5,188,384	28.3%
Other accommodation	307,016	1.1%	9,652	0.1%

Note: Property types in the OS AddressBase and EPC data were re-categorised to match Census 2021 categories. In total, the OS AddressBase contained 28,638,819 properties. This table used data on EPCs up to October 2024.

Table 4: Property characteristics by presence of solid fuel heat source in England and Wales

	Solid fuel heat source	
	No (N = 17,122,175)	Yes (N = 1,586,345)
Property type		
Detached	2,825,767 (17%)	744,344 (47%)
Flat	5,294,287 (31%)	41,269 (2.6%)
House Form Missing	53,890 (0.3%)	2,782 (0.2%)
Other accommodation	9,744 (< 0.1%)	243 (< 0.1%)
Semi Detached	4,222,771 (25%)	449,583 (28%)
Terrace	4,715,716 (28%)	348,124 (22%)
Tenure		
newbuild	737,800 (4.7%)	18,718 (1.1%)
owner-occupied	8,539,885 (54%)	1,273,459 (84%)
rented (private)	3,267,043 (21%)	198,299 (13%)
rented (social)	3,130,467 (20%)	31,654 (2.1%)
Missing	1,446,980	64,215
Smoke Control Area		
No	8,294,583 (48%)	1,292,008 (81%)
Yes	8,827,592 (52%)	294,337 (19%)
Urban		
No	2,377,637 (15%)	823,921 (53%)
Yes	13,956,774 (85%)	731,085 (47%)
Missing	787,764	31,339

Note: We used the subset of unique properties in the EPC dataset. We excluded properties which had missing information on solid fuel presence. Each column shows the count of properties in the EPC database with the specified characteristic and solid fuel status, and the percentage of all properties of that solid fuel status with the specified characteristic. Missing values within characteristics are reported.

Table 5: Beta regression of prevalence of wood burners on LSOA-level characteristics

Characteristic	Coefficient	95% confidence interval
Urban	-0.951	(-0.971, -0.932)
Smoke Control Area	-0.275	(-0.299, -0.252)
Median age	0.039	(0.038, 0.041)
Percentage white ethnicity	0.008	(0.007, 0.009)
IMD score	-0.012	(-0.013, -0.011)
Region:		
East of England	0.061	(0.024, 0.097)
London	-0.117	(-0.164, -0.070)
North East	-0.439	(-0.489, -0.389)
North West	-0.087	(-0.124, -0.050)
South East	0.340	(0.306, 0.374)
South West	0.344	(0.309, 0.380)
Wales	-0.130	(-0.173, -0.087)
West Midlands	0.036	(-0.002, 0.074)
Yorkshire and the Humber	0.003	(-0.035, 0.042)

Note: This table displays results from a beta regression of LSOA-level prevalence of wood burners on geographic and socio-economic factors. The coefficients in the second column represent the the association between a one unit change in the independent variable and the change in the logit of the expected value of the (beta-distributed) outcome variable. Coefficients on region indicators are defined relative to the East Midlands. N = 35,504.

Table 6: Beta regression of prevalence of wood burners on LSOA-level characteristics - urban LSOAs

Characteristic	Coefficient	95% confidence interval
Smoke Control Area	-0.273	(-0.296, -0.250)
Median age	0.023	(0.021, 0.024)
Percentage white ethnicity	0.010	(0.009, 0.011)
IMD score	-0.018	(-0.019, -0.018)
Region:		
East of England	-0.036	(-0.078, 0.006)
London	-0.241	(-0.287, -0.196)
North East	-0.467	(-0.522, -0.412)
North West	-0.162	(-0.201, -0.123)
South East	0.266	(0.228, 0.303)
South West	0.213	(0.171, 0.254)
Wales	-0.285	(-0.337, -0.233)
West Midlands	-0.140	(-0.183, -0.098)
Yorkshire and the Humber	-0.023	(-0.069, 0.015)

Note: This table displays results from a beta regression of LSOA-level prevalence of wood burners on geographic and socio-economic factors among the subset of urban LSOAs. The coefficients in the second column represent the the association between a one unit change in the independent variable and the change in the logit of the expected value of the (beta-distributed) outcome variable. Coefficients on region indicators are defined relative to the East Midlands. N = 28,365.

Table 7: Beta regression of prevalence of wood burners on LSOA-level characteristics - rural LSOAs

Characteristic	Coefficient	95% confidence interval
Smoke Control Area	-0.543	(-0.625, -0.462)
Median age	0.062	(0.059, 0.066)
Percentage white ethnicity	0.008	(0.001, 0.015)
IMD score	-0.001	(-0.004, 0.001)
Region:		
East of England	0.124	(0.052, 0.196)
London	0.401	(-0.141, 0.943)
North East	-0.418	(-0.530, -0.306)
North West	-0.048	(-0.138, 0.041)
South East	0.409	(0.339, 0.479)
South West	0.395	(0.325, 0.466)
Wales	-0.028	(-0.111, 0.054)
West Midlands	0.257	(0.175, 0.339)
Yorkshire and the Humber	-0.003	(-0.088, 0.082)

Note: This table displays results from a beta regression of LSOA-level prevalence of wood burners on geographic and socio-economic factors among the subset of rural LSOAs. The coefficients in the second column represent the the association between a one unit change in the independent variable and the change in the logit of the expected value of the (beta-distributed) outcome variable. Coefficients on region indicators are defined relative to the East Midlands. N = 7,307.

Table 8: Percentage of EPCs with a wood burner by property type and EPC number

Property type	First EPC	Second EPC	Third EPC	Fourth EPC
Number of EPCs: 2				
Detached	696,268 (19.4%)	691,327 (24.3%)		
Semi Detached	1,077,536 (7.2%)	1,070,212 (9.8%)		
Terrace	1,462,561 (4.8%)	1,454,760 (6.1%)		
Number of EPCs: 3				
Detached	104,145 (22.2%)	103,135 (24.8%)	103,022 (28.4%)	
Semi Detached	191,743 (6.6%)	191,687 (8.0%)	189,968 (9.4%)	
Terrace	277,958 (3.7%)	273,437 (4.4%)	275,301 (4.9%)	
Number of EPCs: 4				
Detached	16,250 (24.8%)	16,112 (26.2%)	16,002 (28.2%)	16,028 (31.9%)
Semi Detached	34,642 (6.3%)	34,579 (7.0%)	34,605 (7.8%)	34,404 (8.6%)
Terrace	54,275 (3.0%)	53,451 (3.3%)	53,446 (3.4%)	53,616 (3.8%)

Note: This table shows the percentage of EPCs with a wood burner by property type and EPC number, among the properties in the EPC data which had more than one EPC since 2009. We restrict our analysis to houses (excluding flats, other accommodation, and properties with missing house form) and restrict our analysis to properties with fewer than five EPCs, due to small sample size for higher numbers of EPCs. The total number of properties within rows are not equal due to changes in property classifications in the EPC dataset.

A Data cleaning and linkage

We downloaded all 27,491,063 EPCs from January 2009 to February 2025. We used UPRNs to link EPCs to active property records in the OS AddressBase data. We removed 327,639 (1%) duplicate EPCs for properties with available UPRNs. We retained duplicated entries for properties without UPRNs, since these entries could refer to different properties within the same postcode.

We mapped UPRNs from EPCs to their respective geographies using the National Statistics UPRN Lookup for Great Britain as of May 2022. Of the non-duplicated records in the EPC data, 25,938,702 records could be directly linked to a UPRN. Among the 494,502 EPCs without UPRNs, 479,504 (97%) were in Scotland. For the remaining 14,998 records, we used postcode information to link these records to the National Statistics UPRN Lookup. Next, we linked retained records to LSOA-level data including IMD score, ethnicity, median age, and rural-urban status.

The final EPC dataset contained 25,952,810 records, among which there were 18,708,569 unique properties.

B Search terms for primary and secondary heat source

Using the main EPC dataset, we generated indicator variables for whether a property’s primary or secondary heat source was a solid fuel or wood fuel heat source using keyword searches.

At the time of analysis, the primary heat description contained 1,164 distinct entries, and the secondary heat description included 93 entries. The most common primary heat description was ‘boiler and radiator, mains gas’, with 20.7 million entries. The next most common descriptions were ‘electric storage heaters’ (1.7 million entries) and ‘room heaters, electric’ (1.3 million entries). The most common secondary heat descriptions were ‘none’ (16.2 million entries), ‘room heaters, mains gas’ (4.4 million entries), and ‘room heaters, electric’ (3.4 million entries).

From the primary and secondary heat source data fields of EPCs, we identified entries containing references to solid fuels and wood fuels in English and Welsh. The search terms used to identify solid fuels were ‘wood’, ‘coal’, ‘anthracite’, ‘mineral’, ‘smokeless fuel’ and ‘dual fuel’, and their Welsh language equivalents. The search terms used to identify wood fuels were ‘wood’ and ‘dual fuel’.

C Technical appendix

This section provides further details on the method used to derive the estimated concentration and prevalence of wood fuel heat sources.

Let R denote the number of geographical regions, and let $R_{ir} \in \{0, 1\}$, $i = 1, \dots, N$, $r = 1, \dots, R$ denote the binary indicator that property i belongs to region r . $N_r = \sum_{i=1}^N R_{ir}$ is the number of properties in region r .

Consider a random variable Y defined as follows:

$$Y_i = \begin{cases} 1 & \text{if property } i \text{ has a wood fuel heat source} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Our estimand of interest is the prevalence of wood fuel heat sources by geographic region, π_r :

$$\pi_r = E(Y_i | R_{ir} = 1) \quad (2)$$

Let $S_i \in \{0, 1\}$ denote a binary indicator for whether property i is observed in the EPC data. Then, $N_{rs} = \sum_{i: S_i=1}^N R_{ir}$ is the number of properties in region r observed in the EPC data. The simple within-region sample mean is defined as:

$$\bar{\pi}_r = \frac{1}{N_{rs}} \sum_{i: R_{ir}=1, S_i=1}^N Y_i \quad (3)$$

Assuming that Y_i has finite variance, this estimator is consistent for the expected prevalence of wood fuel heat sources by region, conditional on being observed in the EPC data:

$$\bar{\pi}_r \xrightarrow{P} E(Y_i | R_{ir} = 1, S_i = 1) \quad (4)$$

In general, this quantity will only equal the expectation in Equation 1 if there is no selection bias within regions (i.e. $Y_i \perp\!\!\!\perp S_i | R$).

We calculate an alternative prevalence measure using Census 2021 information on the share of housing types by geographic region. Let $H_{ih} \in \{0, 1\}$, $i = 1, \dots, N$, $h = 1, \dots, H$ denote whether property i belongs to housing type h . $N_{rh} = \sum_{i: R_{ir}=1}^N H_{ih}$ is the number of properties of housing type h in region r , and $\frac{N_{rh}}{N_r}$ is the proportion of housing type h in region r . $N_{rhs} = \sum_{i: R_{ir}=1, S_i=1}^N H_{ih}$ is defined as the number of households of housing type h in region r observed in the EPC data.

First, we estimate the prevalence of wood fuel heat sources within household types and LSOAs using the EPC data:

$$\hat{\pi}_{rh} = \frac{1}{N_{rhs}} \sum_{i: \{R_{ir}, H_{ih}, S_i\}=1}^N Y_i \quad (5)$$

Then, we calculate the estimated prevalence by LSOA, weighted by the Census-derived proportion of housing types by LSOA:

$$\hat{\pi}_r = \frac{1}{N_r} \sum_{h=1}^H \hat{\pi}_{rh} N_{rh} \quad (6)$$

This method relies on the relatively weaker assumption that within regions and housing types, selection is uncorrelated with the probability of having a wood fuel heat source:

$$Y_i \perp\!\!\!\perp S_i | R, H$$

To generalise Equation 6 to higher-level geographies, we construct similar estimates by averaging LSOA-level prevalence measures, weighted by the total number of properties within the LSOA.

We take a similar approach to predict the concentration of wood fuel heat sources by region, except that we take a weighted sum instead of a weighted average in Equation 6, before dividing by the area of the LSOA in km².