Appendix I: WP10 – Stock modelling

Prepared by UCL, EI

BEIS Research Paper Number: 2021/017

Authors: Ian Hamilton, Shih-Che Hsu, Alberto Carron

Section Contents

1. Stock modelling 2
1.1 Scope of WP10 2
1.2 Methodology 2
1.2.1 Predicting exposed brick cavity walls 2
1.2.2 Mapping dwellings in high moisture exposure zones 4
1.3 Results and discussion 6
1.3.1 Predicting dwellings with exposed brick cavity walls 6
1.3.2 Dwellings in high moisture exposure zones 11
1.4 Conclusions 13
References 14
Appendix 14

Stock modelling

1.1 Scope of WP10

The scope of this work package is to a) evaluate the building stock and the number of properties that had untreated, brick faced cavity walls in exposed areas, and b) to estimate the impact of treating those properties in terms of energy and carbon savings.

This scope reflects the concern from BEIS that there were significant challenges in understanding the number of properties with exposed brick cavity walls and how to better evaluate the "700,000 question". In consultation with the project team, it was highlighted that there was a lack of data at any meaningful scale to forecast the actual number of properties that had untreated, brick faced cavity walls in exposed areas. The scope of WP10 was oriented towards the investigation of alternate methods of estimating the number of dwellings with exposed brick cavity walls.

1.2 Methodology

The stock analysis comprised two parts: a) the use of a detailed survey of English dwellings to estimate dwelling features that can predict the probability of having an exposed brick cavity wall; and, b) mapping dwellings in high moisture exposure zones across Great Britain.

1.2.1 Predicting exposed brick cavity walls

Across the stock, lacking direct observations of all dwelling wall types, there is a need to predict the probability of a dwelling having an exposed brick cavity wall (insulated or not). Further, it is important to be able to do this across a geographic scale that allows those dwellings to be allocated to appropriate exposure zones and therefore their risk of requiring specialized treatment when insulating the cavity.

The currently available 'big data' of buildings across the UK are contained within the National Energy Efficiency Data-framework (NEED) and the Energy Performance Certificates (EPC) bulk access database.

NEED, in its published format, is aggregated and anonymised, but the information includes several data points that would be useful for applying a predictive model, relating to the EPC band of the property (EE_BAND), the wall construction (WALL_CON), the presence of Cavity Wall Insulation (CWI) and the year of its installation where known (CWI_YEAR). The full NEED data set held by BEIS goes down to individual property address and postcode details to which a classifier model of exposed brick cavity walls could be applied.

The EPC dataset represents an alternate, and accessible dataset that contains both the property location combined with limited information on the construction of the walls of that dwelling (similar to those in NEED, which assumes this database).

There are limits of the EPC dataset that are important to note, which include the quality of the data when calculating the EPC and the coverage of the data. While there are presently over 7 million EPCs registered, many of these are multiple EPC for the same property. It is estimated there are approximately 2.5 million dwellings with one or more EPCs within the registry. This means that of the 23 million dwellings across England, a little over 10% of dwellings have an EPC. This means that the dataset is limited in terms of its coverage across the country.

However, because the above datasets to not include information as to the nature of the wall type, beyond its insulation, these resources can only provide a base upon which to apply a predictive model. Such a model, however, would need to be able to make use of this available information as input to the prediction of having an exposed cavity wall.

Therefore, this section of the stock modelling focuses on using the English Housing Survey as an existing detailed survey of residential building energy performance and physical characteristics, including wall construction and wall covering, to develop a classifier model for predicting the probability of exposed brick cavity walls. The model construction focuses on using the variables in the NEED and EPC data.

The steps in the analysis were:

1. Compile the English Housing Survey derived datasets (general, physical and interview) for a given survey year (2014/15 in this analysis);
2. Select variables that are likely accessible within NEED and the EPC data for testing (Table 1 below);
3. Sub-select all dwellings with Cavity Wall using the <*construction type*> variable;
4. Develop a variable <*Masonry Type*> for describing the exposure of the wall type, using the <*Predominant type of wall finish*> variable, where all dwellings with [*masonry pointing*] as exposed, and anything else as [*other wall finish*]
5. Use appropriate statistical models to train a model using several different methods for predicting the probability of a cavity wall constructed dwelling in the EHS as having masonry pointing (i.e. exposed).

Table 1 - EHS variables used in modelling

Tenure 2
Dwelling type 3
Dwelling age 3
Useable floor area - original EHS definition
Construction type
Predominant type of roof covering
Predominant type of wall structure
Predominant type of wall finish
Extent of double glazing
Main heating system
Main heating fuel
Loft insulation thickness 3
Energy efficiency rating band (SAP 2012)
Government Office Region EHS version
Rurality classification - morphology (2011 COA)
Nature of area

The modelling method used the SAS 9.2 system to test a selection of model types, these included:

1. Logistic regression to estimate the probability of <Masonry_type> (0=[*masonry pointed*],1=[*Other*]) using all available variables and then a selection of variables with an optimised significance level and AIC.
2. Using GLM regression and LASSO (least absolute shrinkage and selection operator) selection to identify variables and classes with the highest predictive capability to identify dwellings with a higher probability of having masonry pointed walls. LASSO is suitable for modelling where there are high levels of multi-collinearity or when seeking variable selection/parameter elimination.

1.2.2 Mapping dwellings in high moisture exposure zones

A missing component of the stock assessment is the number of properties that are within the different exposure zones (as defined in BS8104:1992) across Great Britain. To address this, GIS was used to construct a layer of those zones and to calculate the number of dwellings using mapped dwelling attribute data for across the country.

There are two parts to this process: (1) creating a projected shapefile (2) calculating the basic residential building information in the four exposure zones of wind-driven rain.

1.2.2.1 Creating a projected shapefile:

The wind-driven rain map derived from a pdf file of BS8104:1992 is a vector image, which was transferred to a shapefile, the commonly used format in the geographic information system (GIS). ArcGIS and QGIS geographic information systems were used to create the wind-driven rain shapefile and conduct the basic statistics. ArcGIS. ArcGIS and QGIS have their own strengths, so we used them according to our tasks.

Both ArcGIS and QGIS have only a limited process that transfers a vector pdf map to raster image, which would lose the precision of a projected mapping, especially on the edge of an image. Therefore, the creationof a projected map from the BS8104:1992 map (Figure 1 A below) used the online transferring system GeoConverter (https://geoconverter.hsr.ch/vector) to directly transfer the pdf format to a shapefile format.

Before transferring, the wind-driven rain map was separated in Adobe Illustrator to make sure there was no overlap in the image. Texts and points were removed from the original pdf image and the vector image was then taken apart into the islands of Great Britain and Northern Ireland and other smaller islands. These parts were then converted to four zones labelled in different colours in the (Figure 1 B).

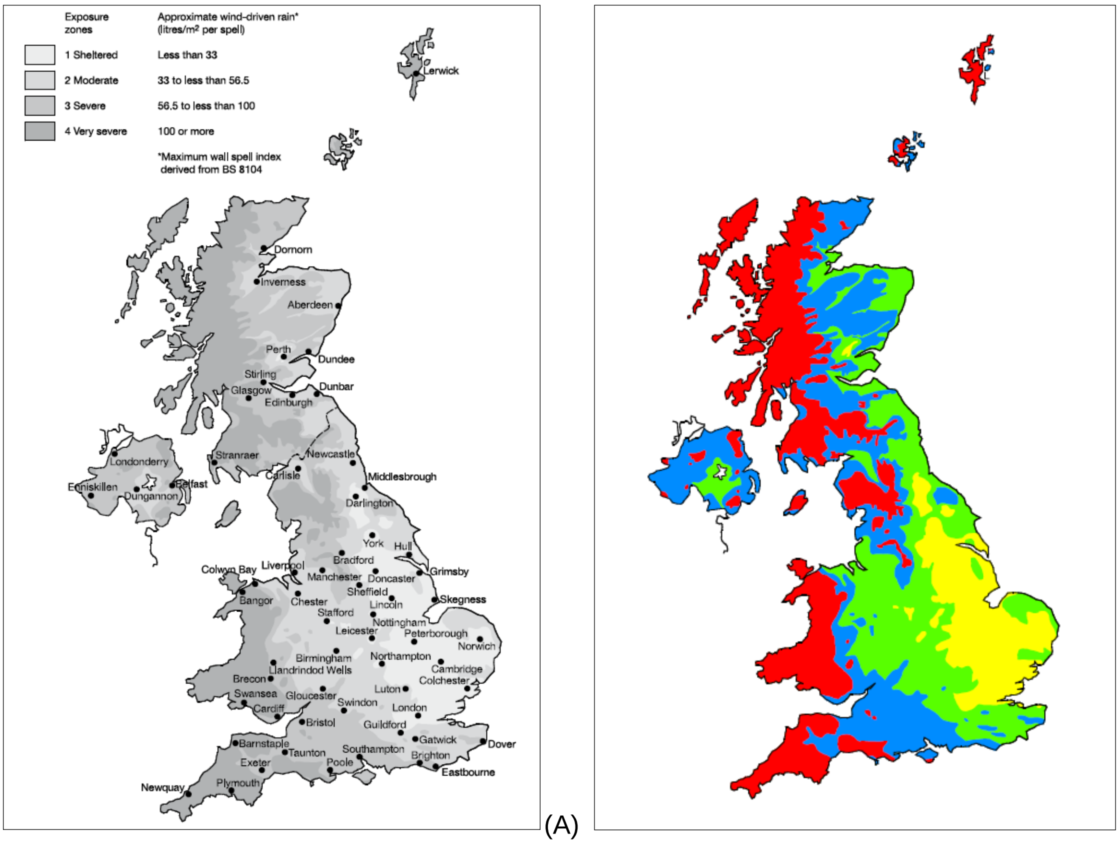

Figure 1 (A) original pdf map from BS8104:1992; and (B) Vectorized map from GeoConverter

Because the original pdf file was not in GeoPDF format it held no projection and coordinates information that could be used when transferred to a .shp file format. Therefore, in order to match the map with the real UK map for the subsequent statistical analysis we adopted the OpenStreetMap (https://www.openstreetmap.org) as the base map to process the georeferencing.

The steps of georeferencing in ArcGIS include using the referencing points to match the image border and resize the area part by part, defining the projection to the WGS84 datum, which is the same with the open street map coordinates system, editing the attribute table for exposure zones and merging all parts.

When creating a projected shapefile image, there is an issue about accuracy that should be mentioned. Because the original image in the pdf file is only a sketch map (i.e. a rough map) the edge of the map features cannot be matched precisely. In addition, the sketch map is not proportional to the real map (i.e. features, topography and curvature are not accounted for), which means we have to cut the map into several sections along with the latitude when matching border and area. This means the image we created would not be 100% the same with the original image in the pdf file.

In order to solve the problem of edge matching, we used the real UK map downloaded from the GADM (https://gadm.org/index.html) to overlay and re-project the image. In Figure 2 (A), which we called the 'real' exposure map, the white area is the area the sketch map cannot match with the real UK map after overlapping two maps. While Figure 2 (B), which we call the simulated exposure map, presents the map after we assign the white area to the closest exposure zone in QGIS using the digitalising tools to cut off the edges and assign them to an exposure zone. The overall effect of this processes is to ensure that all edges and topographic boundary areas (e.g. shorelines and islands) are allocated to their appropriate exposure zone.

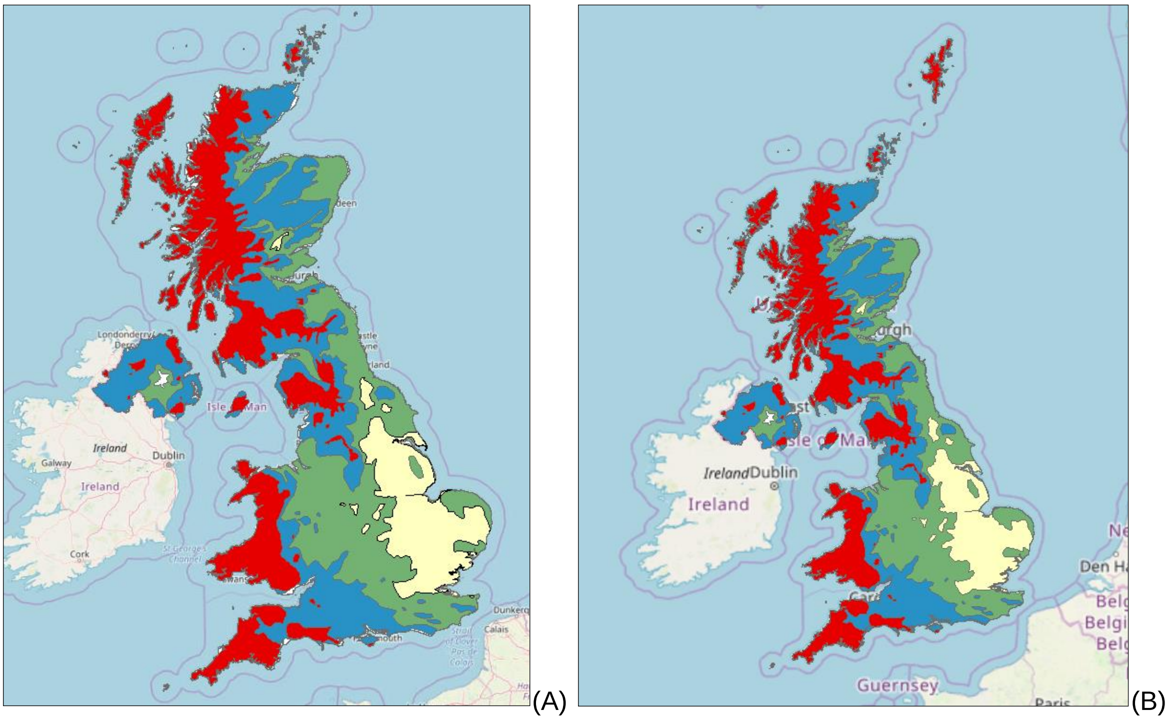

Figure 2 - (A) the calculated exposure map; (B) the simulated exposure map

1.2.2.2 Mapping underlying building stock

The projected exposure map of wind-driven rain shapefiles layers can easily be overlaid to a geo-referenced version of NEED. For illustration in this reporting, we overlapped this map with the OpenStreetMap shapefiles of the UK downloaded from the Geofabrik, which contains building type information (https://www.geofabrik.de/data/shapefiles.html).

1.3 Results and discussion

The following provides results of the stock modelling that focused on using readily available features from NEED and EPCs to predict the probability of having exposed brick cavity walls. It also presents the findings focused on mapping the moisture exposure zones that can subsequently be applied to NEED or EPC data to identify dwellings within high moisture exposure risk.

1.3.1 Predicting dwellings with exposed brick cavity walls

The modelling to predict the probability of dwellings with cavity walls having an exposed masonry used the selection of EHS variables within two main statistical models, standard logistic regression and GLM regression.

1.3.1.1 Method 1: Logistic regression

The logistic regression modelling using the full selection of variables (Table 1) were found to have many variables with low predictive capability. Therefore, using that model, only those variables that had a Pearson Chi-Square value of <0.0001 were retained in selected variables model.

The selected model included:

$$\mathrm{P}(masonry\_type) = dwelling\_age3, government\ region$$

The results shown in Table 2 of the model show that dwelling age and government office region are the only variables that are shown to have any predictive power when fitting a logistic model to estimate the probability of a cavity wall construction having a masonry pointed exposed wall.

Table 2 - Logistic regression model fit and results for all cavity walled dwellings and select variables

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	6534.624	5754.636
SC	6541.602	5852.329
-2 Log L	6532.624	5726.636

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	805.9882	13	<.0001
Score	906.0259	13	<.0001
Wald	715.7412	13	<.0001

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
Dwelling age	5	399.9076	<.0001
Government Office Region	8	457.0777	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter	**Level**	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	1.9349	0.0483	1602.0404	<.0001
Dwelling age	1945 to 1964	1	-0.1941	0.0666	8.4825	0.0036
Dwelling age	1965 to 1980	1	0.5335	0.0755	49.8859	<.0001
Dwelling age	1981 to 1990	1	0.8470	0.1128	56.4103	<.0001
Dwelling age	post 1990	1	0.6359	0.0882	52.0416	<.0001
Dwelling age	pre 1919	1	-0.6462	0.1431	20.3801	<.0001
Government Region	East	1	0.00061	0.0996	0.0000	0.9951
Government Region	East Midlands	1	0.5492	0.1349	16.5701	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter	Level	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Government Region	London	1	-0.2158	0.1239	3.0355	0.0815
Government Region	North East	1	0.4530	0.1302	12.0945	0.0005
Government Region	North West	1	-0.0483	0.0829	0.3400	0.5599
Government Region	South East	1	-0.2468	0.0781	9.9790	0.0016
Government Region	West Midlands	1	0.6536	0.1314	24.7558	<.0001
Government Region	Yorkshire and the Humber	1	0.5097	0.1135	20.1503	<.0001

The above table includes the log parameter estimates for the selected dwelling age and region. Dwellings with cavity walls built between 1965 and 1990 are less likely to have masonry pointed exposed walls compared to dwellings built between 1919 to 1944, with this further decreasing among more newly constructed dwellings. The model also shows that dwellings constructed in London, South East and North West were more likely than dwellings in the South West to have exposed walls. This is illustrated using the odds ratios shown in Figure 3.

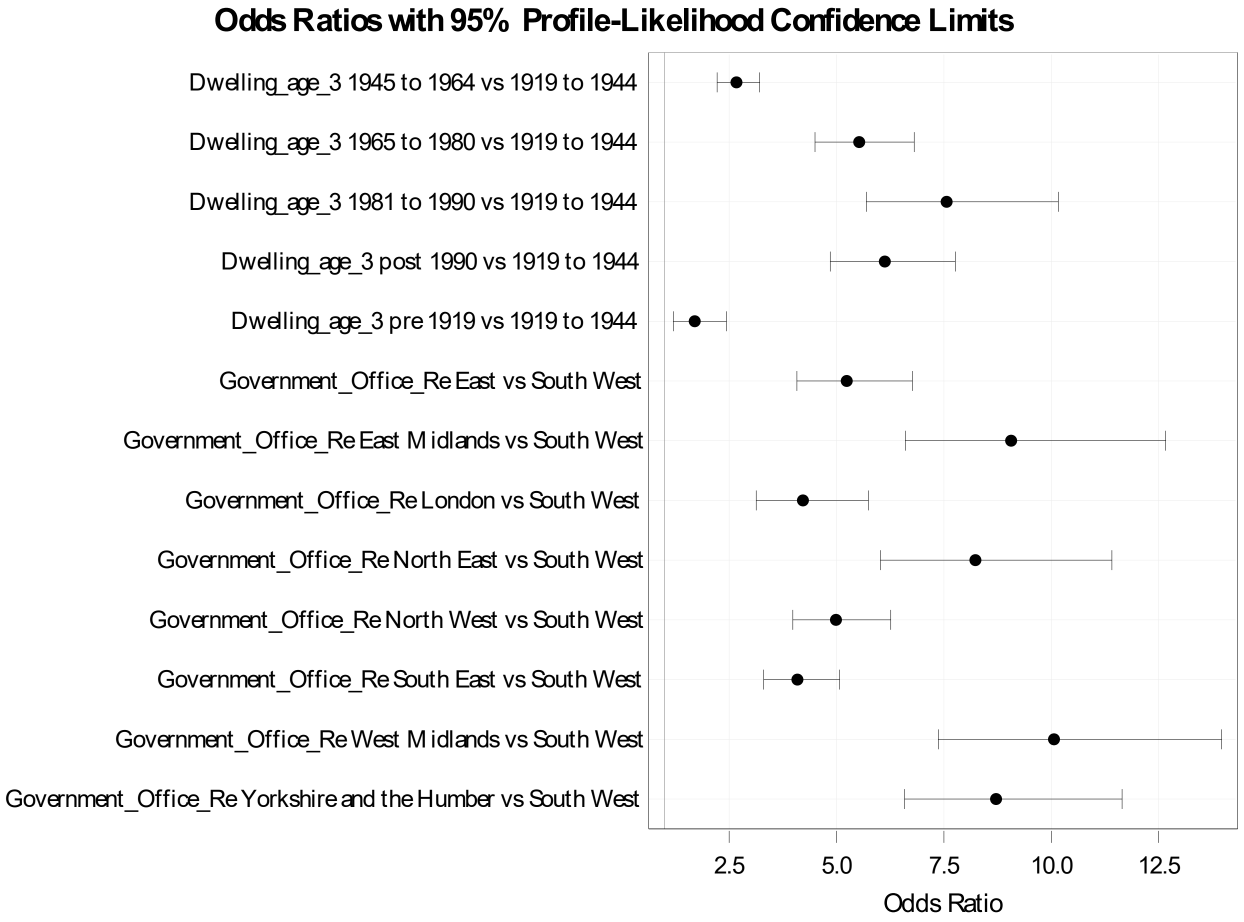

Figure 3 - Logistic regression select model odds ratio results

It is recognized that dwellings in the South West are most likely to already have some form of covering on their walls, and this may also be true for other areas of the ‘very severe’ and ‘severe’ exposure zones. Therefore, a variation of the modelling was to concentrate on only those zones where there existed a high moisture exposure, i.e. the South West, West Midlands and the North West.

Table 3 below shows the logistics regression for only the Western Regions, note that for this analysis, energy performance certificate was retained as an additional feature for the model to fit. The results show that as dwelling age increased, so did the probability of a cavity wall being exposure masonry pointed – compared to 1919-1944 dwellings, and dwellings in the performance bands of C-E had higher probability than those in Level F.

Table 3 - Logistic regression model fit and results for all cavity walled dwellings in English western regions and select variables

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	2794.091	2436.375
SC	2800.043	2513.750
-2 Log L	2792.091	2410.375

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	805.9882	13	<.0001
Score	906.0259	13	<.0001
Wald	715.7412	13	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Dwelling age	5	399.9076	<.0001
Government Office Region	8	457.0777	<.0001
Energy Performance Certificate		13.7710	0.0171

Analysis of Maximum Likelihood Estimates						
Parameter	Level	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	1.2670	0.2052	38.1073	<.0001
Dwelling age	pre 1919	1	-0.1076	0.2290	0.2206	0.6386
Dwelling age	1945 to 1964	1	-0.2962	0.1036	8.1758	0.0042
Dwelling age	1965 to 1980	1	0.2782	0.1082	6.6088	0.0101
Dwelling age	1981 to 1990	1	0.5277	0.1566	11.3498	0.0008
Dwelling age	post 1990	1	0.4645	0.1351	11.8260	0.0006
Region	North West	1	0.2127	0.0778	7.4655	0.0063
Region	West Midlands	1	1.0293	0.1021	101.5331	<.0001
EPC	B	1	-0.3166	0.4905	0.4167	0.5186
EPC	C	1	0.5573	0.2181	6.5291	0.0106
EPC	D	1	0.3571	0.2110	2.8631	0.0906
EPC	E	1	0.1117	0.2430	0.2113	0.6457
EPC	G	1	-0.2185	0.8504	0.0660	0.7972

1.3.1.2 Method 2: GLM regression

The GLM regression using LASSO selection, which included the input of all the available variables (Table 1) were found to have a slightly improved statistical fit (compared to the logistic regression), but more importantly to help identify variables and classes that have explanatory power when estimating the probability of having an exposed brick cavity wall.

The LASSO method works by entering each variable with the aim of shrinking (i.e. reducing) the least absolute coefficient values being evaluated. Figure 4 below shows how the variables are entered with the aim of minimizing the parameters to the point of excluding variables that have no benefit to the estimate and therefore reduces the variance of the model.

The results of the modelling in Figure 4 and Table 4 show that the variable classes of Government Office Region: South West and Dwelling Age: 1919-1944 and pre-1919 had the highest coefficients for dwellings to have non-exposed walls. The other coefficients, dwelling ages built after 1965 and in mid, northern eastern government regions tended to have increased likelihood of exposed brick masonry pointing walls.

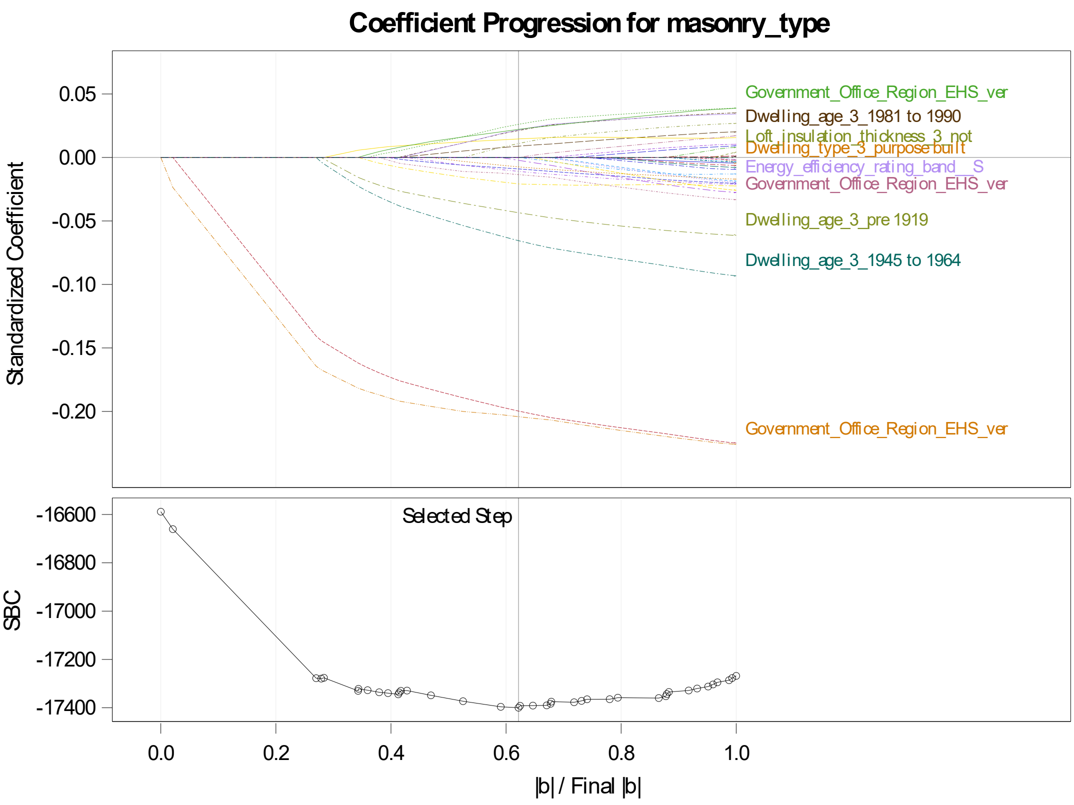

Figure 4 - LASSO regression model variable coefficient tests

Table 4 - GLM Lasso Regression model fit and results for all cavity walled dwellings

Effects:	Intercept Dwelling age 3: pre 1919, 1919 to 1944, 1945 to 1964, 1981 to 1990 EPC Band: C, E Region: East Midlands, London North East, South East, South West, West Midlands, Yorkshire and the Humber Loft insulation thickness: none Rurality classification: hamlets and isolated dwellings, urban > 10k Useable floor area: 110 sqm or more

Analysis of Variance				
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**
Model	17	111.84227	6.57896	60.16
Error	7909	864.92397	0.10936	
Corrected Total	7926	976.76624		

Root MSE	0.33070
Dependent Mean	0.14394
R-Square	0.1145
Adj R-Sq	0.1126

AIC	-9596.38324
AICC	-9596.28712
SBC	-17400

Parameter Estimates		
Parameter	**DF**	**Estimate**
Intercept	1	0.899544
Dwelling age: pre 1919	1	-0.095126
Dwelling age: 1919 to 1944	1	-0.211071
Dwelling age: 1945 to 1964	1	-0.051496
Dwelling age: 1981 to 1990	1	0.009805
EPC Band: C	1	0.010555
EPC Band: E	1	-0.011666
Region: East Midlands	1	0.025565
Region: London	1	-0.003346
Region: North East	1	0.015018
Region: South East	1	-0.012326
Region: South West	1	-0.237106
Region: West Midlands	1	0.030460
Region: Yorkshire and the Humber	1	0.024372
Loft insulation thickness: none	1	-0.031238
Rurality classification: hamlets and isolated dwellings	1	-0.020556
Rurality classification: urban > 10k	1	0.020398
Useable floor area: 110 sqm or more	1	-0.021149

N.B. green colours indicate a decrease in estimate for having a exposed brick wall, while orange denotes an increase

As above, the following model applied the GLM LASSO regression to only those regions in the South West, West Midlands and the North West, in order to identify dwelling features that increased the explanatory power of having exposed brick masonry pointed walls.

Table 5 below shows the GLM Lasso regression for only the Western Regions. The results show that dwellings built post-1990, having an EPC level C, in the West Midlands were shown, and in urban areas >10k. Conversely, the chances of having exposed brick cavity walls was reduced among properties built before 1964 and in the South West.

Table 5 - GLM Lasso Regression model fit and results for all cavity walled dwellings

Effects:	Intercept Dwelling age: 1919 to 1944, 1945 to 1964, post-1990 EPC Band: C Region: South West, West Midlands Loft insulation thickness: none Rurality classification: urban > 10k

Analysis of Variance				
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**
Model	7	62.40278	8.91468	66.27
Error	2833	381.12063	0.13453	
Corrected Total	2840	443.52341		

Root MSE	0.36678
Dependent Mean	0.80641
R-Square	0.1407
Adj R-Sq	0.1386
AIC	-2847.98786

AICC	-2847.92428
SBC	-5643.37257

Parameter Estimates		
Parameter	**DF**	**Estimate**
Intercept	1	0.765399
Dwelling age: 1919 to 1944	1	-0.139957
Dwelling age: 1945 to 1964	1	-0.030373
Dwelling age: post 1990	1	0.006676
EPC Level: C	1	0.012046
Region: South West	1	-0.191134
Region: West Midlands	1	0.053854
Rurality classification: urban > 10k	1	0.119626

N.B. green colours indicate a decrease in estimate for having a exposed brick wall, while orange denotes an increase

1.3.2 Dwellings in high moisture exposure zones

Using QGIS, we allocated the area of the UK into the four exposure zones and the underlying building stock. Using the intersection tool and summarise function of the attribute table in ArcGIS a summary table is created for reference.

Figure 6 shows the England buildings located in the exposure zone 2 (i.e. Moderate) of wind-driven rain and the Table 6 - Part of the summary table of building types (ordered by count) in exposure zone 2 of England shows a part of the summary table of building types in exposure zone 2 of England. If we take a closer look in the exposure zone 2 of England, we chose the place near Sutton train station showed in figure 6 and labelled the residential and house categories of buildings in red.

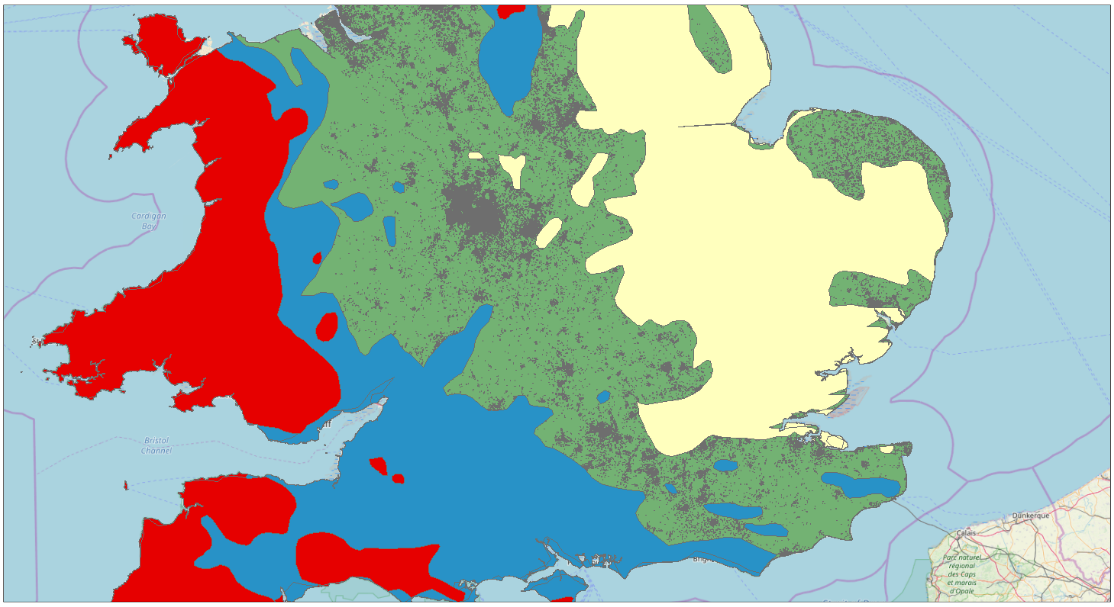

Figure 5 - England buildings (in grey) located in the exposure zone 2

Table 6 and Figure 6 below illustrate how the GIS mapping layer is used to map buildings within the exposure zones. Table 3 shows all building polygons within the Geofabrik and OpenStreetMap database.

Table 6 - Part of the summary table of building types (ordered by count) in exposure zone 2 of England

Type	Count	Type	Count	Type	Count	Type	Count	Type	Count
	1217105	conservato	1742	manufactu	167	transportat	52	bridge	29
residential	897181	greenhous	1741	ruins	162	military	49	residential	29
house	659565	warehouse	1564	supermark	158	farmhouse	48	family	28
garage	74030	hut	1487	offices	157	tower	48	store	28
terrace	40686	pub	1416	hall	144	stadium	47	village_ha	27
industrial	24185	semi-detac	1367	factory	132	car_port	46	cinema	26
retail	19685	hospital	1208	silo	130	club_hous	46	maisonette	26
detached	18281	tank	827	maisonette	127	sports_cen	46	annex	24
semidetached_house	13597	public	684	houses	123	substation	46	restaurant	24
school	13273	flats	637	prison	123	derelict	45	Y	24
garages	12199	college	584	terraced_h	122	library	45	boathouse	23
semi	12011	train_static	560	clubhouse	121	mosque	44	brewery	23
apartments	10464	stable	554	bunker	114	works	44	part	23
commercial	9945	storage_ta	548	beach_hut	100	tower_blo	43	townhouse	23
bungalow	7774	hotel	483	cabin	96	carport	42	block	22
barn	7286	constructic	349	utility	95	place_of_v	40	collapsed	22
church	5185	chapel	312	kindergarte	90	Sheltered	40	fire_statio	22
agricultural	4054	dormitory	303	mobile_hc	89	leisure	38	chimney	21
static_caravan	3653	civic	288	grandstanc	82	retail_outl	37	conveyor	21
farm_auxiliary	3634	pavilion	283	university	76	bus_shelte	36	cottage	21
farm	3189	service	281	canopy	75	heritage	36	governme	21
shed	2874	airport	257	shelter	75	1	33	railway	20
office	2310	hangar	257	chalet	73	parking	32	semidetac	20
shop	2176	no	222	semi-detac	63	grain_silo	31	depot	19
roof	2089	glasshouse	215	bandstand	62	temple	30	kiosk	19
university	2006	stables	169	amenity	57	toilets	30	museum	19

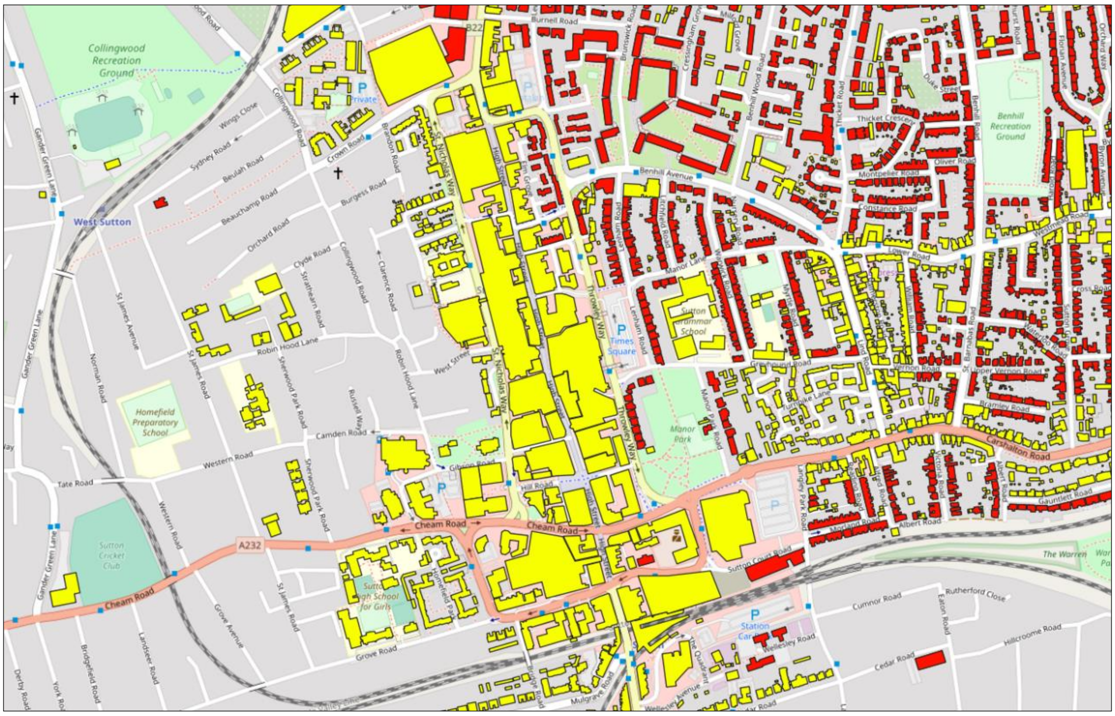

Figure 6 – All types of buildings located near Sutton train station with residential categories of buildings in red

1.4 Conclusions

The resulting outcome of the predictive modelling is that the likely available variables from NEED and EPCs for predicting the probability of a cavity walled dwelling having exposed brick walls (i.e. masonry pointing) will offer only limited power for predicting a dwelling having exposed brick cavity walls.

The models are designed to identify the 'probability' of dwellings having an unfilled cavity exposed brick wall. As the predictive probability of the models are low, they will need to be used to identify neighbourhood areas of 'risk' as compared to predicting any given dwelling's wall composition.

Despite the low predictive power, however, the analysis did identify those variables for data within NEED with the greatest explanatory power and these are related to dwelling location and dwelling age. This means that it would be possible to apply the above logistic or GLM Lasso models to NEED, for example, to identify properties with risk attributes within the severe exposure zones (i.e. zone 4) that might require extra consideration for hygrothermal treatment when filling the cavity with insulation.

When restricting the statistical modelling to only regions in the 'very severe' and 'severe' exposure zones for Wind Driven Rain, which comprise the western regions in England (Scotland and Wales were not modelled in this analysis), those results show a very similar picture to the whole of England in terms of features of the buildings that increase the probability or likelihood of having exposed brick facades. However, in the GLM regression with LASSO selection, the analysis identifies newer dwellings with an EPC C and in urban areas as having a higher likelihood of exposed brick facades.

The mapping of the moisture exposure zones in the form of a GIS shapefile provides an ability to BEIS to geospatially identify all the dwellings in NEED that might be within the moisture risk zones. BEIS can then use the statistical modelling to select dwelling attributes that have a greater/lower probability or coefficients of exposed brick cavity walls.

References

Open Street Map, 2019. https://www.openstreetmap.org/

Geofabrik, 2019. https://www.geofabrik.de/data/shapefiles.html

GADM, 2019. https://gadm.org/index.html

GeoConverter, 2019. https://geoconverter.hsr.ch/vector

Appendix

Exposure map shapefile:

ukexposure_simulation_map.zip

Exposure mapping building data table:

ukwicdata.xlsx