Title

Modelling energy efficiency performance of residential building stocks based on Bayesian statistical inference

Authors names and affiliations

Marta Braulio-Gonzalo[a]*, Pablo Juan[b], María D. Bovea[a], María José Ruá[a]

[a]Department of Mechanical Engineering and Construction, Universitat Jaume I, Av. Sos Baynat s/n, 12071 Castellón, Spain

[b]Department of Mathematics, Universitat Jaume I, Av. Sos Baynat s/n, 12071 Castellón, Spain

Marta Braulio-Gonzalo

Name: Marta (M.)

Surname: Braulio-Gonzalo

E-mail address: braulio@uji.es

Pablo Juan

Name: Pablo (P.)

Surname: Juan

E-mail address: juan@uji.es

María D. Bovea

Name: María D. (M.D.)

Surname: Bovea

E-mail address: bovea@uji.es

María José Ruá

Name: María José (M.J.)

Surname: Ruá

E-mail address: rua@uji.es

***Corresponding author**

Marta Braulio-Gonzalo

Tel. +34 964 729162

E-mail address: braulio@uji.es

Universitat Jaume I, Department of Mechanical Engineering and Construction, Av. Sos Baynat s/n, 12071 Castellón, Spain

Abstract

This paper provides a model based on Integrated Nested Laplace Approximation to predict the energy performance of existing residential building stocks. The energy demand and the discomfort hours for heating and cooling were taken as response variables and five parameters were considered as potentially significant to assess the building energy performance: urban block pattern, street height-width ratio, building class through the building shape factor, year of construction and solar orientation of the main façade. A total of 240 dynamic energy simulations were run varying these parameters, by using the EnergyPlus software with the Design Builder interface, which allowed the response variables to be determined for a set of sample buildings. Simulation results revealed the most and least significant parameters in the energy performance of the buildings. The model developed is a useful decision-making tool in assisting local authorities during energy refurbishment interventions at the urban scale.

Highlights

- A model to predict the energy performance of residential building stocks is developed.
- The energy demand and the discomfort hours were considered as output of the model.
- Bayesian inference based on INLA was the statistical framework of the model
- The model is useful to identify urban areas that require urgent energy refurbishment.

Keywords: energy efficiency; residential building stock; Bayesian inference; INLA

1. Introduction

Knowing the dynamics of a current real system allows values to be predicted considering the effect of certain variables. This can be studied within the framework of inference by using systems of differential equations which enable the values to be forecast taking into account the influence of a set of particular variables. Particularly, this work uses Bayesian inference, which delivers an integrated approach to perform inference, prediction and decision. Modern Bayesian models use simulation methods to generate drawings from the posterior distribution. Markov Chain Monte Carlo (MCMC) combined with the Stochastic Partial Differential Equation (SPDE) approach were the motivation for the Integrated Nested Laplace Approximation (INLA) package for the R software (R Development Core Team, 2011). The library was initiated by Rue and Martino (2007) and subsequently improved through contributions by Rue et al. (2009). The use of INLA in computational time allows the user to work with relatively complex models in an efficient way. The INLA package for the R software was employed to conduct this study.

The aim of this work is to develop a model to forecast the passive energy efficiency performance of the existing residential building stock, according to a set of specific parameters. Some examples of precursor methods aimed at conducting energy assessment of building stocks were found in the literature (Farahbakhsh et al., 1998; Huan and Berkeley, 2000; Shorrock and Dunster, 1997). More recently, some authors have based their studies on these previous ones in order to improve them (Johnston et al., 2005; Boardman, 2007; Natarajan and Levermore, 2007) and others have developed new more complex models that enhance the energy efficiency assessment process (Gouveia et al., 2012; McKenna et al., 2013). In the context of building energy efficiency, many aspects can be assessed, but generally those related to energy use are the most common. For example, Cheng and Steemers (2011) estimated the energy consumption and CO_2 emissions of English building stocks considering the size of buildings, internal and external temperatures, thermal characteristics of the envelope and gas boiler efficiency, as input parameters. Their model also allowed them to determine the influence of different energy efficiency measures on dwellings to assist energy policies at local and national levels. Florio and Teissier (2015) proposed a model to estimate the energy performance certificate (EPC) of the French housing stock based on the building typology and conducted a statistical analysis of the French EPC database. Other items have been assessed by Mauro et al. (2015), who predicted the energy demand (ED) and the discomfort hours (DH) taking into account the building geometry, orientation and physical characteristics of the envelope. At the same time, they analysed the influence of different energy efficiency measures.

Taking into account this framework, this study provides a model based on INLA, aimed at predicting the energy demand and discomfort hours (considered as response variables) for the heating and cooling, respectively, of the existing residential building stock, considering five parameters that affect the energy efficiency (considered as covariates), namely: year of construction, building shape, solar orientation of the main façade, street height-width ratio and urban block pattern. The data set for this study was obtained by modelling the energy performance of sample buildings of an urban district located in Castellón de la Plana (Spain) by using EnergyPlus software (U.S. DOE, 2015).

The work is structured as follows. Section 2 draws the methodological approach of the study based on INLA, broken down into three different stages. Section 3 presents the selection of the response variables and covariates used to conduct the study. Section 4 provides the data set, describes the response variables and covariates, and then presents the configuration of the simulations run. Section 5 deals with the INLA modelling process and provides the configuration of the battery of prediction models and a comparison among them in order to finally choose the most accurate ones. This section also addresses a statistical analysis to outline the significance of the covariates in the models. The paper ends with a final Section 6 offering the conclusions.

2. Methodology

With the aim of obtaining the prediction models (based on INLA) for characterising the energy demand and discomfort hours for the heating and cooling of an existing residential building stock, the stages showed in Figure 1 were applied:

- Stage I: Response variables and covariates selection. Based on a literature review, the response variables and the covariates that characterise the energy performance of an existing building stock needed to be selected.
- Stage II: Data set. After analysing the advantages and disadvantages of different software applications available for simulating the energy efficiency of buildings and building stocks (Machairas et al., 2014), Design Builder Software (2014) was chosen to obtain the values of the response variables against a combination of the selected covariates. After this, a descriptive analysis of the data results was conducted.

- Stage III: Modelling the data set with R-INLA in order to obtain the functions for prediction of the response variables. Firstly, the models that include all the possible combinations of response variables and covariates were obtained. Then, the correlation coefficients and the root mean square error (RMSE) were calculated, which allowed the models to be compared. Subsequently, the significance of the covariates was explored in order to identify the most and least significant ones. Finally, the four models with the best fit for determining the response variables were selected, according to the correlation coefficients and RMSE.

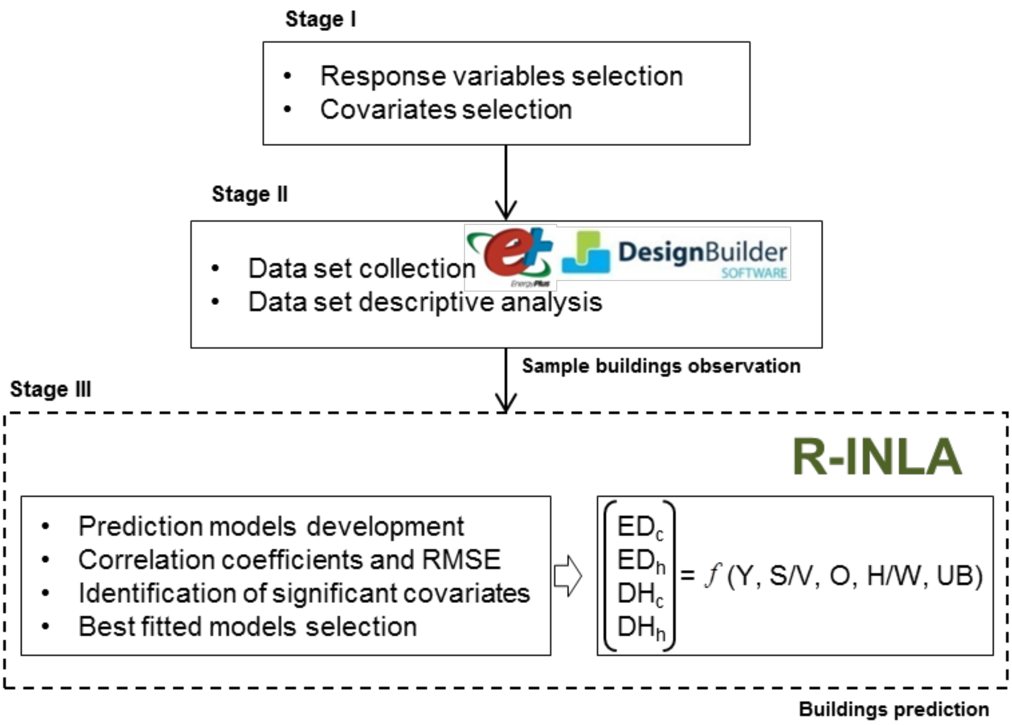


Figure 1. Methodological approach

3. Selection of response variables and covariates

The selection of the response variables and covariates that mainly characterise the passive energy efficiency of an existing residential building stock was based on the literature review. As Table 1 reports, four response variables are the key variables considered to assess the energy performance of the residential building stock:

- Energy demand for heating (ED_h) and energy demand for cooling (ED_c). Both variables measure the amount of energy that the thermal installations of the building have to provide in order to ensure inner comfort conditions according to the building use and climatic zone (CTE, 2013), for heating and cooling, respectively. Both are measured in kWh/m^2·year.

- Discomfort heating hours (DH_h) and discomfort cooling hours (DH_c). Both variables measure the time when the combination of the zone humidity ratio and the operative temperature is not in the ASHRAE 55-2004 summer or winter clothes region (DB, 2015), for heating and cooling, respectively. They are expressed in h/year.

These response variables mainly depend on the performance of five parameters related to the building or city scale that will be considered as covariates in this study. These covariates are described below and the values for each one are presented in Table 2.

- Year of construction (Y). The year of construction of the building implies the typical construction techniques employed for the envelope (façades, roofs, floors and windows). The thermal transmittance of the envelope elements improves over time, when the introduction of thermal insulation materials in the building process presumably contributed to enhance the energy performance of buildings. This covariate involves data at the building scale.
- Building typology through the building shape factor (S/V). The S/V ratio represents the compactness of the building and it is defined as the ratio of the building's total external surface area ($\sum S_i$) to its internal volume (V) (Granadeiro et al., 2013),

$$S/V = \sum \frac{S_i}{V} \left[m^2 / m^3 \right]$$

  Lower S/V values mean more compact buildings. This covariate implies the building typology, which is identified by three aspects: the type of occupancy (multi-family (MF) or single-family (SF)), the number of floors (≤4 or >4), and the type of adjacency (terraced (T) or detached (D)) (Braulio-Gonzalo et al., 2015).Three building typologies were identified in the district under study ($MF_{T(\leq 4)}$, $MF_{T(>4)}$ and $SF_{T(\leq 4)}$) and therefore, three sample buildings were selected to run simulations, which are representative of the building typologies. The dimensional characteristics of these three sample buildings are presented in Table 3. Their floor area is in accordance with the average floor area of each building typology existing in the district. This covariate is related to the building scale.

- Solar orientation of the main façade (O). This covariate, related to the building scale, determines the orientation against the position of geographic North (0º) and is measured in degrees.
- Street height-width ratio (H/W). This represents the relation between the height of the buildings situated in front of the building under study and the street width, and involves the opportunities for solar access in the buildings. Higher H/W ratios imply more shaded streets and, thus, less possibilities of solar access for the surrounding buildings. This covariate is considered a city scale one.
- Urban block pattern (UB). This refers to the geometry of the urban block where the building is situated. The fact of having a big courtyard or not in the middle of the urban block (surrounded by the buildings that make up the block) implies the possibility of solar access in the façades of the buildings. It is thus related to the city scale.

By considering the street height-width ratio, the urban block pattern and the solar orientation, not only the building features but also urban characteristics and the morphology of the surrounding built environment are taken into account in the study.

Table 1. Response variables and covariates

Variables	Description	Reference
Response variables*		
ED_c	Cooling energy demand (KWh/m^2·year)	Snäkin (2000); Natarajan and Levermore (2007); Florio and Teissier (2015); Gouveia et al. (2012)
ED_h	Heating energy demand (KWh/m^2·year)	Snäkin (2000); Natarajan and Levermore (2007); Florio and Teissier (2015); Gouveia et al. (2012)
DH_c	Discomfort hours for cooling (hours/year)	Mauro et al. (2015)
DH_h	Discomfort hours for heating (hours/year)	Mauro et al. (2015)
Covariates		
Y	Year of construction of building	Boardman (2007; Natarajan and Levermore (2007); Hens et al. (2001); Dascalaki et al. (2011); Theodoridou et al. (2011); Dall'O' et al. (2012); Florio and Teissier (2015); Aksoezen et al. (2015)
S/V	Building shape factor (compactness)	Hens et al. (2011); Dall'O' et al. (2012); Penna et al. (2014); Granadeiro et al. (2013); Florio and Teissier (2015); Aksoezen et al. (2015); A. L. Martins et al. (2014)
O	Solar orientation of the main façade	Mauro et al. (2015); Košir et al. (2014)
H/W	Street height-width ratio	artins et al. (2014)
UB	Urban block type	Okeil (2010); Futcher and Mills (2013); Košir et al. (2014); Yezioro et al. (2006) **

* Heating season runs from October 1st to May 31st and cooling season operates from June 1st to September 30th.
** These references use the urban block to conduct the studies but do not develop any model that includes it as a covariate.

This study has been applied to analyse the energy performance in an urban district belonging to Castellón de la Plana (Spain), a medium-sized city with 180,690 inhabitants (INE, 2015). This city is located at a latitude of 39º 59' 11" North and a longitude of 0º 2' 12" East, which means it has a mild climate with temperate winters and warm summers. Table 2 describes the covariates considered in this research, for this specific case study.

Table 2. Description of covariates for the case study

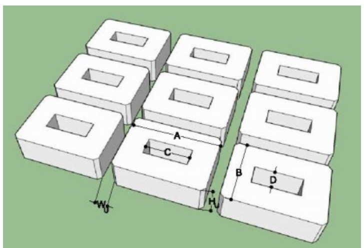
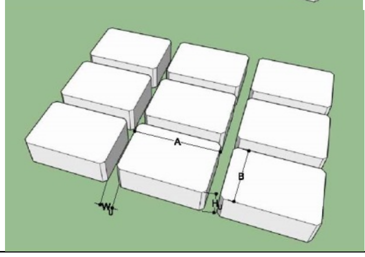
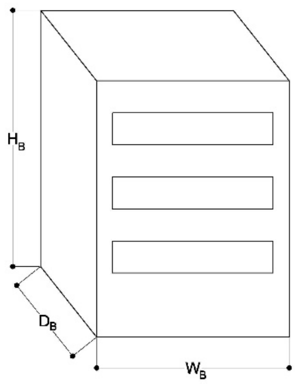
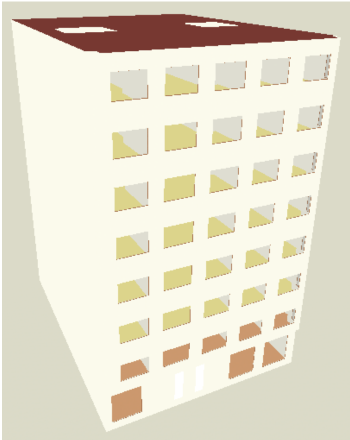
Covariate	Value	Characteristics	Description
Y	1	Before 1940	Absence of thermal insulation
			One layer thick walls with thermal inertia
	2	1940 – 1959	Absence of thermal insulation
			Two layer light walls
	3	1960 – 1979	Absence of thermal insulation
			Two layer light walls
	4	1980 - 2006	Poor thermal insulation
			Two layer light walls
	5	After 2006	Poor thermal insulation
			Two layer light walls
S/V	0.3	$\mathbf{MF}_{T(\leq 4)}$	MF terraced building with 4 or less than 4 floors (see Table 3)
	0.26	$\mathbf{MF}_{T(>4)}$	MF terraced building with more than 4 floors (see Table 3)
	0.4	$\mathbf{SF}_{T(\leq 4)}$	SF terraced building with 4 or less than 4 floors (see Table 3)
O	0º	North	Indirect solar radiation
	90º	East	Direct solar radiation at mornings; low elevation angle
	180º	South	Direct solar radiation at noon; maximum elevation angle
	270º	West	Direct solar radiation at afternoons; low elevation angle
H/W	2.4	H_u=24m; W_U=10m	Narrow streets that imply low solar access
	1.2	H_u=24m; W_U=20m	Wide streets that imply high solar access
UB	1		Big internal courtyard which allows solar gains in South, East and West façades of the buildings with an inward orientation towards the courtyard
	2		No big courtyard, but smaller own light wells as internal element of the building

Table 3. Properties of the sample buildings

Building properties	**MF$_{T(>4)}$**	**MF$_{T(\le4)}$**	**SF$_{T(\le4)}$**
 Building sketch			
Occupancy	MF	MF	SF
S/V (m^{-1})	0.26	0.30	0.40
Floors above ground	8	3	4
H_B Height (m)	25	13.50	14.50
W_B Width (m)	15.90	11.80	4.00
D_B Depth (m)	28.50	14.70	30.50
Conditioned floor area (m^2)	2006.61	337.98	165.24
Conditioned volume (m^3)	4931.69	1628.72	473.05
Building external area (m^2)	1282.83	478.19	160.38
Opaque façade surface (m^2)	741.50	272.90	94.00
Glazing surface (m^2)	293.80	102.90	32.10
Glazing rate (%)	39.60	37.70	34.15

By using a Geographical Information System (GIS) (gvSIG, 2014), it was possible to represent how these covariates are distributed in the neighbourhood under study in Castellón de la Plana. As an example, Figure 2 and Figure 3 show the urban block distribution according to covariates Y and S/V reported in Table 2.

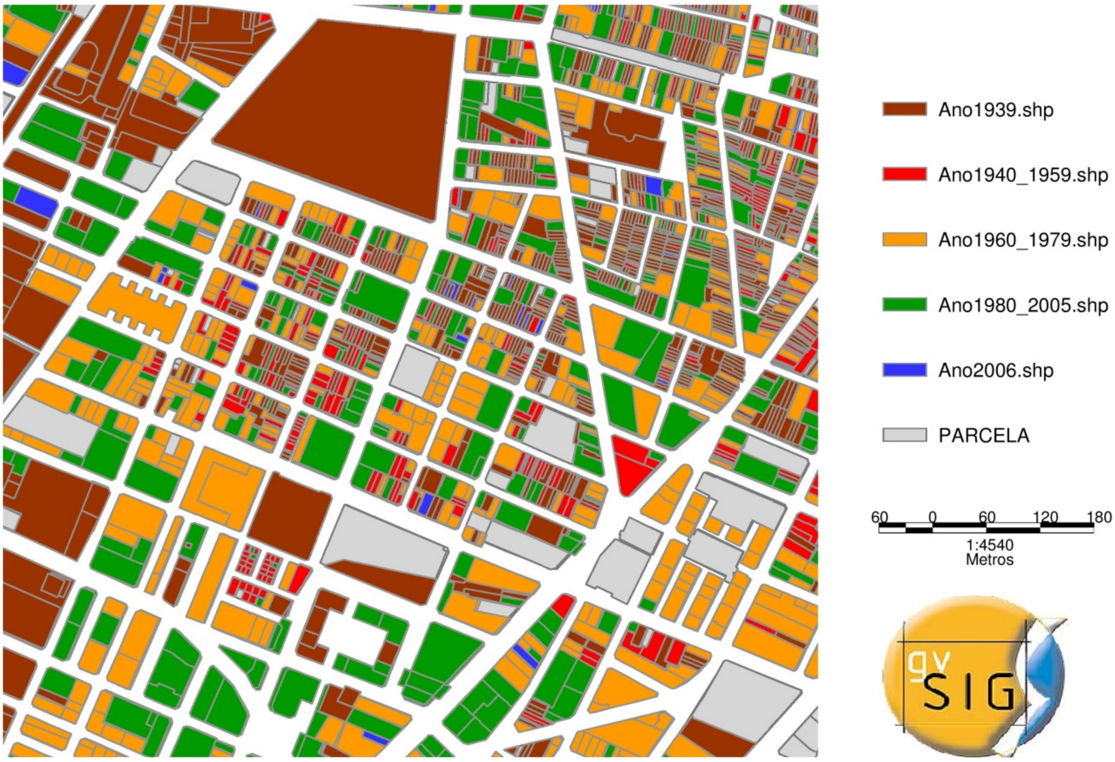


Figure 2. GIS map representing building classification of the district per year of construction (Y)

Figure 3. GIS map representing building classification of the district per shape factor (S/V)

4. Data set

4.1 Data set collection

To obtain the value of the response variables, the EnergyPlus software with the Design Builder interface was applied. This application makes it possible to obtain the response variables (energy demand for heating and cooling – ED_c and ED_h, respectively) and discomfort hours for heating and cooling (DH_c and DH_h, respectively), for a set of sample buildings characterised by a combination of the covariates (year of construction (Y), building shape factor (S/V), solar orientation (O), street height-width ratio (H/W), urban block type (UB)).

In accordance with Figure 4, 240 combinations can be performed by combining the values for the covariates presented in the case study. So, 240 simulations were modelled and processed with EnergyPlus software to obtain the response variables for each one. Table 4 present the boundary conditions assumed for energy simulations. By analysing the behaviour of each variable response against each combination of covariates, it is possible to identify the most and least significant parameters in the energy performance of the buildings, after conducting a statistical analysis.

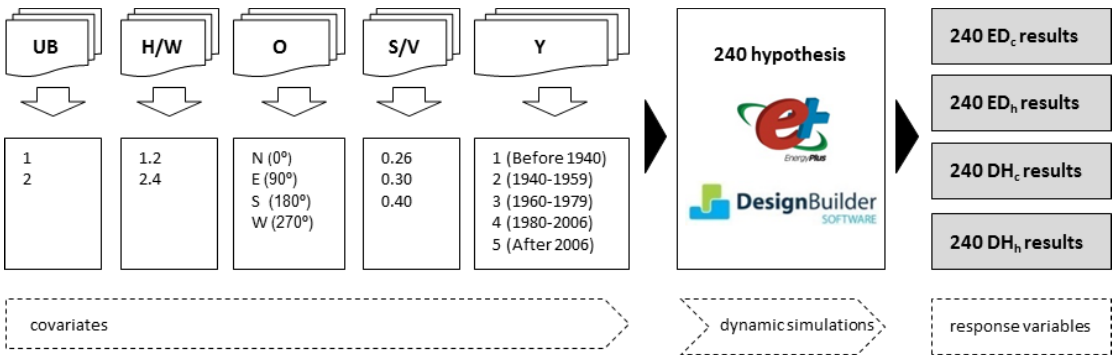


Figure 4. Simulations set-up scheme for determining ED_c, ED_h, DH_c and DH_h

Table 4. Boundary conditions for energy simulations

Parameter (unit)	Value
Occupancy	
Density (person/m^2)	0.03
Schedule pattern	Weekdays: 7:00-15:00h [25%]; 15:00-23:00h [50%]; 23:00-7:00h [100%]
	Weekends 0:00-24:00h [100%]
Metabolic rate (W/person)	117.2
Clothing (clo) winter/summer	1/0.5
Temperatures (ºC)	
Heating set point	20
Cooling set point	25
Natural ventilation set point	24
Internal gains	
Internal loads (W/m^2)	8.8
Lighting (W/m^2 – 100 lux)	4.4
Miscellaneous gains (W/m^2)	4.4
Solar gains	
Shading calculations include all surrounding buildings	
Calculations include modelling reflections and shading of ground reflected solar	
Domestic hot water demand	
DHW demand (l/m^2day)	0.84

4.2 Descriptive analysis of data set

The data obtained from the simulations mark the starting point in this study. Representing the data in boxplots allows us identifying some very high and low values outside the distribution (outliers), which can be removed before conducting the statistical modelling, if necessary. Thus, the observed values for the response variables obtained from the simulations are described numerically and graphically in Table 5. As denoted, the data present strong variability. While ED_c acquires values close to 0, ED_h, DH_c and DH_h adopt notably higher values, substantially high in the case of DH_h. In spite of there being a notable variability of data, the methodology could be implemented in every case, that is, for every response variable.

Table 5. Description of response variables (ED_c, ED_h, DH_c and DH_h)

Variable	Minimun	1st Quartile	Median	Mean	3rd Quartile	Maximun
EDc	0.02	0.65	1.33	1.66	2.45	5.58
ED_h	22.19	64.46	96.54	96.50	122.70	211.90
DH_c	117.60	438.70	637.40	722.60	1029.00	1597.00
DH_h	3830.00	4735.00	4911.00	4859.00	5049.00	5303.00

Figure 5 presents the empirical probability of all the input and output variables. These distributions demonstrate that none of the variables follows the normal distribution, as concluded after performing Shapiro-Wilk. Figure 6 shows the same data broken down into the two existing urban block types. In this case, the data distribution follows the same distribution pattern. It therefore denotes the coherency in the fact of modelling the data all together and separately, in the two urban blocks. Figure 7 displays the scatter plots for each of the covariates with each of the response variables.

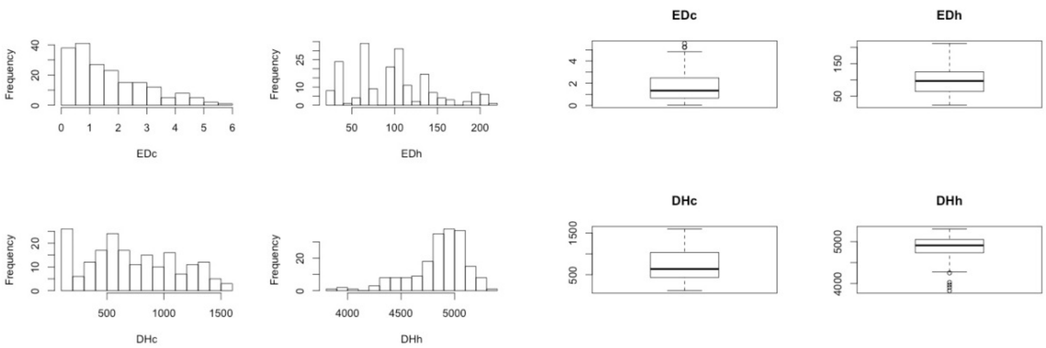


Figure 5. Histogram (left) and boxplot (right) of response variables (ED_c, ED_h, DH_c and DH_h)

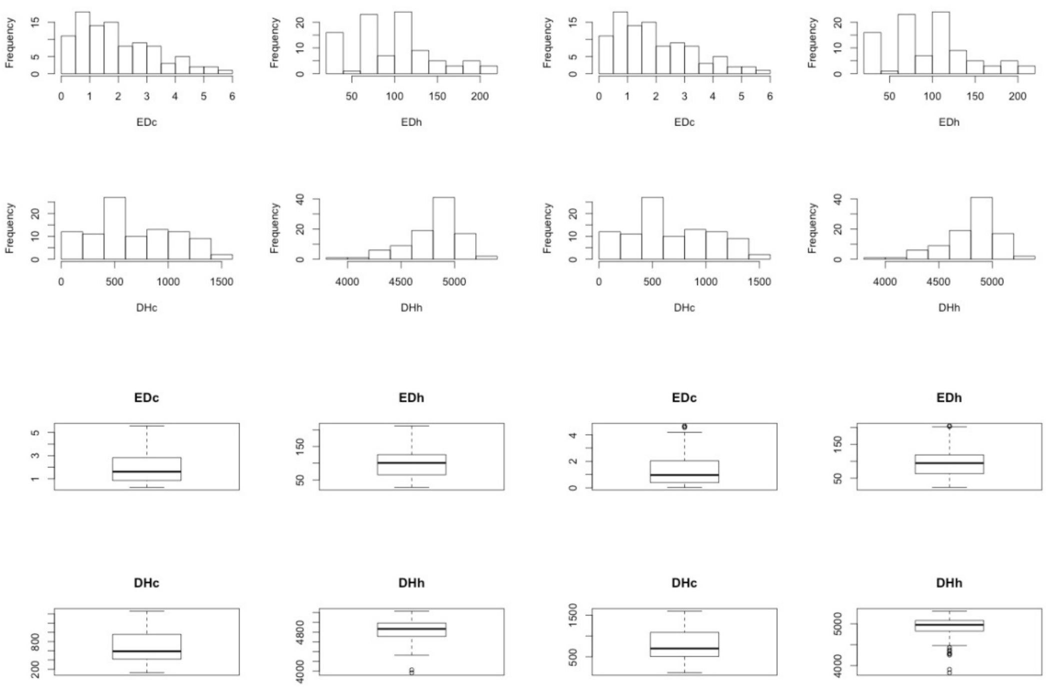


Figure 6. Histogram (top) and boxplot (bottom) of response variables (ED_c, ED_h, DH_c and DH_h) for UB_1=1 (left) and UB_2=2 (right)

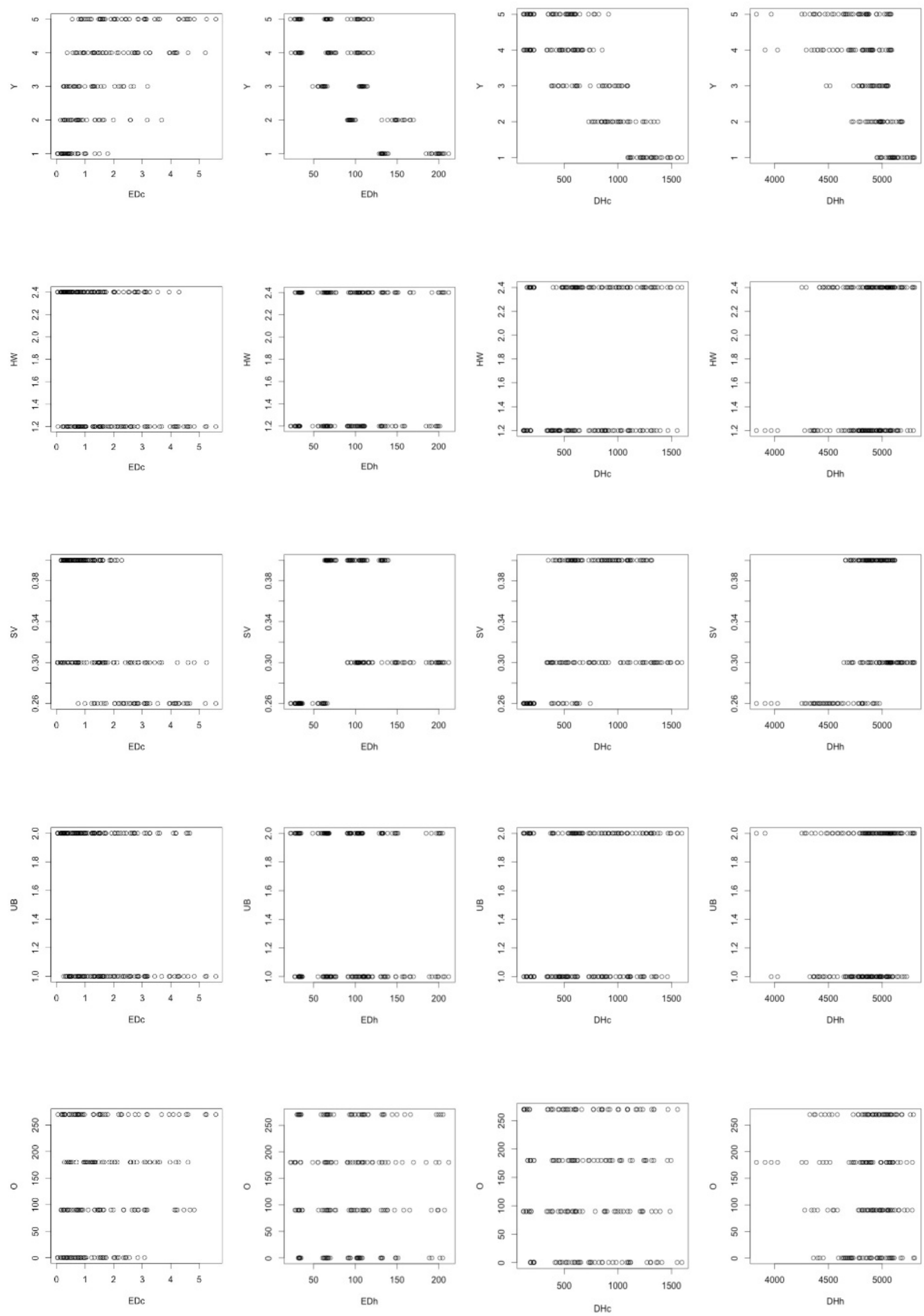


Figure 7. Scatter plot: response variables vs. covariates

5. INLA modelling

5.1 Statistical framework

The data can be idealised as realisations of a stochastic process indexed by

$$Y(\cdot) \equiv \{y(\cdot) \in R \}$$

where $y(\cdot)$ is a temporal subset of R.

The advantages of using INLA over other methods, such as basic statistical methods or more complex ones (like Markov Chain Monte Carlo (MCMC), Generalized Linear Models (GLM)), are the following:

- It works with reasonable computational times, allowing the user to work with complex models quickly and efficiently.
- It allows integrating as many covariates as desired, and also incorporating new covariates in the model in later steps.
- It allows analysing the level of significance of covariates.
- It does not require to work with normal distributions exclusively, due to it is based on Bayesian inference.

The data can be presented by a collection of observations $y = \{y_1,...,y_n\}$(Blangiardo et al., 2013; Cameletti et al., 2013). A temporal correlation structure is a complicated mathematical entity and its practical estimation is very difficult if the covariates are included (Vlad et al., 2015). In statistical analysis, to estimate a general model it is useful to model the mean for the unit using an additive linear predictor, defined on a suitable scale

$$\eta_i = \beta_0 + \sum_{m=1}^{M} \beta_m z_{mi} + \sum_{l=1}^{L} f_l(v_{li})$$

where β_0 is a scalar, which represents the intercept, $\beta = (\beta_1,...,\beta_M)$ are the coefficients of the linear effects of the covariates $z = (z_1,...,z_M)$ on the response, and $f = \{f_1(.),...f_L(.)\}$ is a collection of functions defined in terms of a set of covariates $v = v(v_1,...,v_L)$.

We have considered the parameters of Figure 4 as covariates: Y, S/V, O, H/W, UB.

The first step in defining the structure of the data $y = \{y_1,...,y_n\}$. A very general approach consists in specifying a distribution for y_1 characterised by a parameter ϕ_i (usually the mean $E(y_i)$) defined as a function of a structured additive predictor η_i through a link function $g(.)$, such that $g(\phi_i) = \eta_i$. The additive linear predictor η_i is defined as follows (Blangiardo and Camaletti, 2015):

$$\eta_i = \beta_0 + \sum_{m=1}^{M} \beta_m X_{mi}$$

where β_1 represents the coefficient that quantifies the effect of the covariates in the response X_i.

5.2 Modelling results

All analyses were carried out using the R freeware statistical package (version 3.1) (R Development Core Team, 2011) and the R-INLA package (R-INLA project, 2012). By combining all the covariates, 64 models were obtained. Firstly, in order to compare the results of energy performance in both types of urban block, a first set of 32 models were designed, where the covariate UB is combined with the remaining four covariates (H/W, Y, O and S/V), separately. Secondly, these four covariates were studied all together, obtaining a second set of eight models. Thirdly, UB was considered another covariate (like H/W, Y, O and S/V), which resulted in a third set of 20 models. Finally, the five covariates were studied all together, obtaining the fourth set of the last four models.

Once the battery of competing models has been obtained, we compare them using the deviance information criterion (DIC) (Spiegelhalter et al., 2002), which is a Bayesian model comparison criterion given by

$$DIC = \text{"}goodness\ of\ fit\text{"} + \text{"}complexity\text{"} = D(\bar{\theta}) + 2p_D$$

where $D(\bar{\theta})$ is the deviance evaluated at the posterior mean of the parameters and p_D denotes the *effective number of parameters*, which measures the complexity of the model.

When the model is true, $D(\bar{\theta})$should be approximately equal to the *effective degrees of freedom*, $n-p_D$. DIC may underpenalise complex models with many random effects.

On the other hand, the conditional predictive ordinate (CPO) (Pettit, 1990) is also analysed, which expresses the posterior probability of observing the value (or set of values) of y_i when the model is fitted to all the data except y_i

$$CPO_i = \pi\left(y_i^{obs} \mid y_{-i}\right).$$

Here, y_{-i} denotes the observations y with the i^{th} component removed. This facilitates computation of the cross-validated log-score (Gneiting and Raftery, 2007) for model choice $(-(mean(\log(cpo))))$. Therefore the lowest values of DIC and $(-(mean(\log(cpo))))$ suggest the model with the best fit. A large number of parameters means more complexity. The best models are those with a high level of complexity and a high goodness-of-fit. In general, the model showing the lowest CPO and DIC should be chosen.

Tables 6, 7, 8 and 9 show the summary results related to goodness-of-fit for the battery of models.

Table 6. DIC for the battery of 40 models comparing both UB types11

		1st set of 32 models				**2nd set of 8 models**
		H/W	**Y**	**O**	**S/V**	**H/W; Y; O; S/V**
EDc	UB_1	289.88	253.31	296.38	240.38	126.08
ED_h		1035.85	1059.57	1045.48	1020.27	996.10
DH_c		524.74	396.55	526.40	512.92	381.15
DH_h		445.38	404.37	447.70	431.22	384.18
EDc	UB_2	247.17	203.51	251.64	224.21	151.56
ED_h		1029.79	1047.36	1040.28	1014.11	983.74
DH_c		547.11	442.83	548.57	532.85	425.85
DH_h		-411.06	-446.00	-408.79	-419.93	430.52

Table 7. DIC for the battery of 24 models considering all covariates together

	3rd set of 20 models					4th set of 4 models
	H/W	Y	O	S/V	UB	H/W; Y; O; S/V; UB
ED_c	550.98	483.59	560.92	490.49	558.99	319.82
ED_h	2050.29	1971.48	2049.33	2026.12	2057.42	1942.99
DH_c	1070.93	849.77	1073.99	1045.44	1071.33	804.58
DH_h	920.70	848.46	927.74	899.20	922.71	815.43

Table 8. CPO for the battery of 40 models comparing both UB types

		1st set of 32 models				2nd set of 8 models
		H/W	Y	O	S/V	H/W; Y; O; S/V
ED_c	UB_1	1.541498	1.319262	1.508144	1.250554	0.658676
ED_h		5.394913	5.520670	5.445634	5.313597	5.1877487
DH_c		2.731860	2.063948	2.740482	2.670111	1.9854350
DH_h		2.321519	2.109037	2.333540	2.247051	2.0020945
ED_c	UB_2	1.309396	1.060710	1.286546	1.165767	0.788749
ED_h		5.364257	5.457064	5.419085	5.282162	5.124821
DH_c		2.848420	2.304734	2.856211	2.773782	2.218651
DH_h		2.470266	2.289598	2.480160	2.422938	2.244737

Table 9. CPO for the battery of 24 models considering all covariates together

	3rd set of 20 models					4th set of 4 models
	H/W	Y	O	S/V	UB	H/W; Y; O; S/V; UB
ED_c	1.434107	1.259472	1.459823	1.276742	1.455178	0.833192
ED_h	5.339549	5.135604	5.337008	5.276180	5.358084	5.059606
DH_c	2.788211	2.212131	2.796270	2.721721	2.789244	2.095318
DH_h	2.398959	2.211349	2.417017	2.342444	2.404779	2.124434

From the results obtained, it can be deduced that in every case the inclusion of a higher number of covariates improves the model because DIC and CPO become lower, which indicates that these models enable better prediction. For example, when we analyse the values of DIC and CPO for the four models that relate UB_1 with every single covariate to predict ED_c (first row in Tables 5 and 7), we realise that the best model is $ED_{c(UB1 \leftrightarrow S/V)}$. However, when the model includes all the covariates

($ED_{c(UB1\leftrightarrow H/W-Y-O-S/V)}$, last column in Table 9), the DIC and CPO is even lower, which shows it is a better model.

On the other hand, if we compare those models that integrate UB as a covariate, we can observe the same behaviour. The last four models are better than the previous 20, because of the inclusion of all the covariates; for instance $ED_{c\ (H/W-Y-O-S/V-UB)}$ versus $ED_{c\ (S/V)}$.

The next step is the comparison between the observed and predicted values and this could be performed by the correlation coefficients and the root mean square error (RMSE). The correlation coefficient between the predicted and observed values and the RMSE for all the models has been calculated and is shown in Tables 10 and 11. These tables highlighted the fact that models which include all the covariates are those that best fit the data.

Table 10. Correlation and RMSE for the battery of 40 models comparing both UB types

		1st set of 32 models				**2nd set of 8 models**
		H/W	**Y**	**O**	**S/V**	**H/W; Y; O; S/V**
ED_c	UB_1	0.3272/1.2473	0.5828/1.0727	0.2069/1.2893	0.6624/0.9872	0.90770.5673
ED_h		0.04242/52.3090	-0.7130/59.2891	0.03917/54.8400	0.1555/48.3402	0.5210/41.5530
DH_c		0.1321/3.6207	0.8611/1.8569	0.02381/3.6517	0.3611/3.4065	0.8906/1.6608
DH_h		0.2088/2.3945	0.6129/1.9341	0.1388/2.4237	0.4184/2.2249	0.7242/1.6874
ED_c	UB_2	0.2956/1.1434	0.5739/1.0097	0.1532/1.1867	0.6006/0.9662	0.8303/0.6799
ED_h		0.0436/50.6284	-0.7411/55.6610	-0.0159/53.3735	0.1809/46.8202	0.5689/38.9334
DH_c		0.1335/4.0682	0.8177/2.3630	0.0546/4.0989	0.3905/3.7796	0.8597/2.0965
DH_h		0.1807/2.7722	0.5719/2.3111	0.0981/2.8032	0.3433/2.6493	0.6465/2.1484

Table 11. Correlation and RMSE for the battery of 24 models considering all covariates together

	3rd set of 20 models					**4th set of 4 models**
	H/W	**Y**	**O**	**S/V**	**UB**	**H/W; Y; O; S/V; UB**
ED_c	0.3038/1.2305	0.5562/1.0776	0.1795/1.2712	0.6168/1.0187	0.2239/1.2587	0.8815/0.6336
ED_h	0.0429/49.8644	0.7257/40.6334	0.0119/49.6547	0.1678/46.9283	-0.0557/50.799	0.6266/37.0674
DH_c	0.1316/3.8814	0.8303/2.1820	0.0397/3.9124	0.3731/3.6327	0.1236/3.8854	0.8744/1.8996
DH_h	0.1907/2.6246	0.8303/2.1745	0.0119/2.6732	0.3723/2.4819	0.16197/2.6384	0.6824/1.9539

On observing the first set of models in Table 10, it can be seen that the models that better predict the value of the response variables are those which include the covariates Y (year of construction) and S/V (building shape factor), due to the fact that they have higher correlation coefficients. Nevertheless, it can be observed that the second set of models are clearly the most homogeneous ones and have the highest correlations, because they include all the covariates for each UB type. Thus, these would present a lower RMSE between the observed and the predicted values.

From Table 11, where UB type is included as a covariate, we can draw a similar conclusion as above. The models that include all the covariates (fourth set of models) are better than those which include the covariates separately (third set of models). Consequently, as we add more covariates in the model, the model improves the prediction. Finally, on comparing the second and fourth sets of models we can conclude that the fourth, which includes the UB as a covariate, is the best one, as the four models included have the highest correlation coefficients and the lowest RMSE.

In the same way, Figure 8 represents graphically the correlation between the predicted and observed data for the last four models, which include all covariates. It is noted that the model with the best fit is $ED_{c(H/W\text{-}Y\text{-}O\text{-}S/V\text{-}UB)}$, with a correlation coefficient of 0.88, followed by $DH_{c(H/W\text{-}Y\text{-}O\text{-}S/V\text{-}UB)}$ with 0.87, , $DH_{h(H/W\text{-}Y\text{-}O\text{-}S/V\text{-}UB)}$ with 0.68 and finally $ED_{h(H/W\text{-}Y\text{-}O\text{-}S/V\text{-}UB)}$ with 0.63.

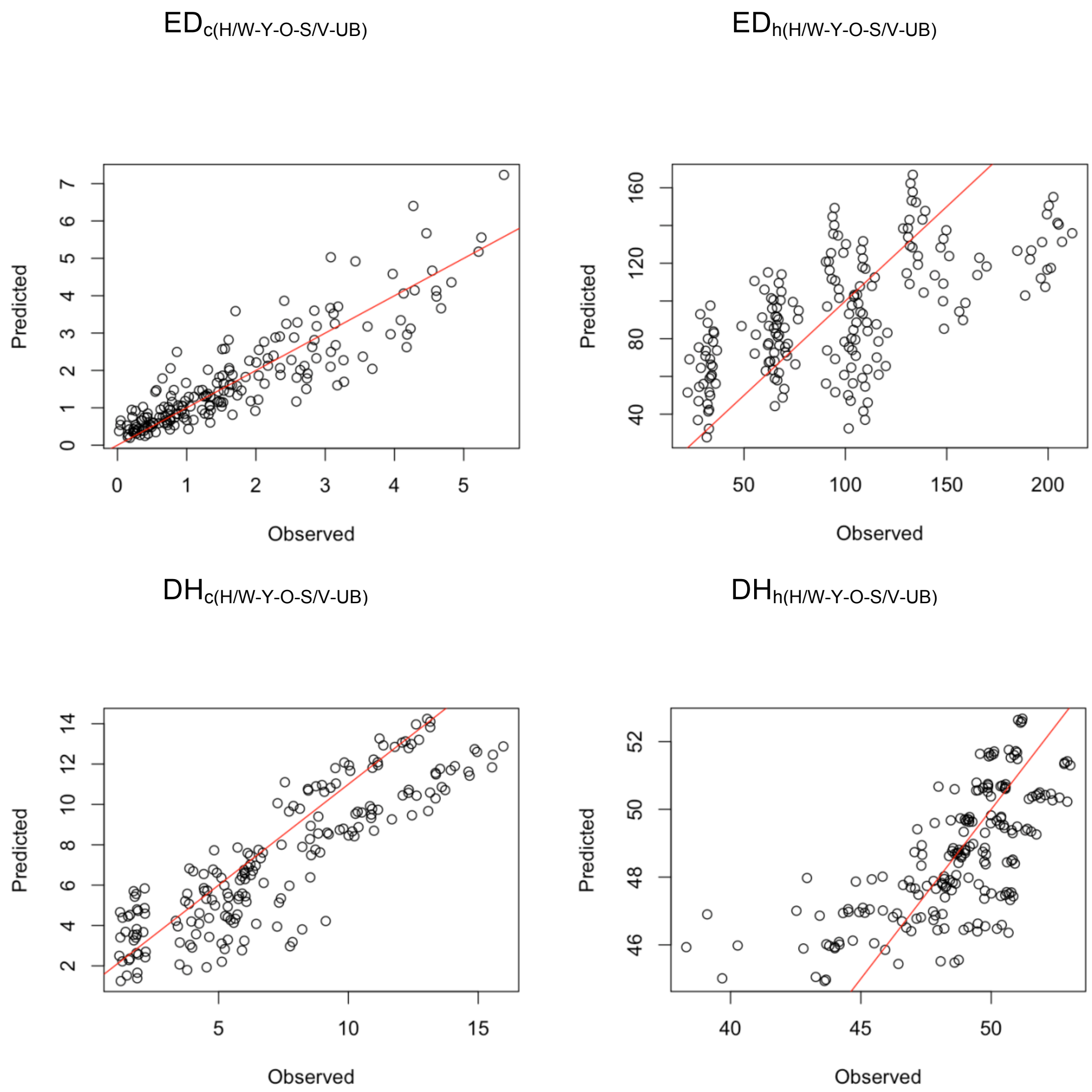


Figure 8. Correlations between the predicted and observed data for the selected models

Figure 9 shows graphically, as an example, the correlation between the predicted and observed data for two of the models designed for the response variable ED_c, comparing both urban block types. We can observe a higher correlation in the model for UB_1, with a correlation coefficient of 0.91 against 0.83.

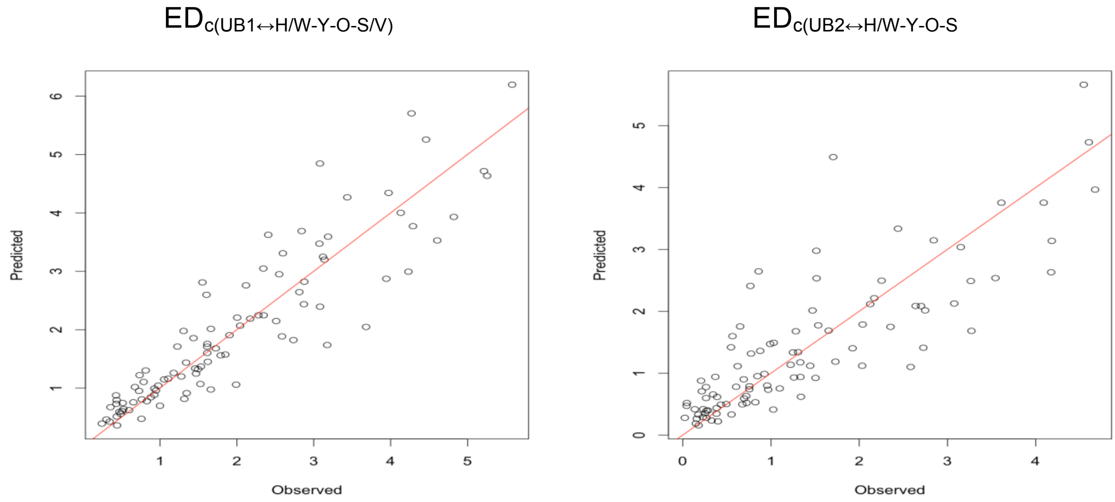


Figure 9. Example of correlations between the predicted and observed data for two of the models that compare UB_1 and UB_2

5.3 Identification of significant covariates

In this section, the impact of every single variable on the results was analysed. Fixed effects for all the models (and denoted by β_i) are expected to have a systematic and predictable influence on data.

Interactions between covariates for each response variable have been tested as suggested Tsanas and Xifara (2012), concluding from the obtained results that any of them was significant. After this analysis, Tables 12, 13, 14 and 15 present the models whose covariates are significant.

Table 12. Fixed effects: β_i (mean [0.025quant, 0.975quant]). 1st set of 32 models

		1st set of 32 models			
		H/W	**Y**	**O**	**S/V**
EDc	UB_1	-0.3745 [-0.5955, -0.1538]	0.3238 [0.2428, 0.4041]	0.0014 [0.0000, 0.0028]	-8.2477 [-10.0081, -6.4818]
EDh		36.8255 [26.7441, 47.0489]	7.5230 [1.5431, 14.0321]	0.2333 [0.1391, 0.3322]	139.2179 [95.5385, 182.7811]
DHc		0.8402 [-0.3677, 2.0482]	-2.2046 [-2.4676, -1.9416]	-0.0007 [-0.0080, 0.0066]	21.3303 [10.2201, 32.4048]
DHh		0.9917 [0.1900, 1.7977]	-1.0224 [1.2959, -0.7478]	0.0038 [-0.0011, 0.0086]	18.8925 [11.5351, 26.3047]
EDc	UB_2	-0.4397 [-0.328, -0.1469]	0.4700 [0.3601, 0.5777]	0.0020 [0.0000, 0.0040]	-7.8284 [-10.3440, -5.3079]
EDh		34.2280 [24.3390, 44.2917]	5.2809 [-0.5471, 11.4892]	0.1986 [0.1060, 0.2956]	134.4105 [90.9687, 177.7394]
DHc		0.9653 [-0.3888, 2.3199]	-2.3467 [-2.6810, -2.0123]	-0.0020 [-0.0102, 0.0062]	25.6614 [13.4154, 37.8582]
DHh		0.0085 [-0.0008, 0.0178]	-0.0113 [-0.0146, -0.0081]	0.0000 [-0.0001, 0.0000]	0.1620 [0.0726, 0.2512]

Table 13. Fixed effects: β_i (mean [0.025quant, 0.975quant]). 2nd set of 8 models

		2nd set of 8 models
		H/W; Y; O; S/V
EDc	UB1	-0.4137 [-0.5125, -0.3149]; 0.2725 [0.2297, 0.3152]; 0.0009 [0.0003, 0.0015]; -7.2562 [-8.2794, -6.2303]
EDh		31.6630 [20.1911, 43.5282]; -13.3457 [-18.5742, -7.9547]; 0.0984 [0.0177, 0.1805]; 100.4968 [48.2215, 152.0663]
DHc		0.8170 [0.2512, 1.3828]; -2.1055 [-2.3496, -1.8613]; -0.0008 [-0.0042, 0.0025]; 11.6464 [5.8796, 17.4134]
DHh		0.9203 [0.3459, 1.4967]; -0.9111 [-1.1589, -0.6623]; 0.0036 [0.0017, 0.0036]; 13.7288 [7.8825, 19.6182]
EDc	UB2	-0.5189 [-0.7086, -0.3293]; 0.4091 [0.3210, 0.4969]; 0.0020 [0.0007, 0.0032]; -5.7667 [-7.6639, -3.8597]
EDh		30.0743 [19.1843, 41.4377]; -14.2590 [-19.1722, -9.1602]; 0.0673 [-0.0081, 0.1445]; 108.4249 [56.8547, 159.093]
DHc		0.9335 [0.2202, 1.6469]; -2.2142 [-2.5219, -1.9065]; -0.0022 [-0.0064, 0.0021]; 15.8751 [8.6411, 23.1090]
DHh		0.9645 [0.2336, 1.6991]; -0.9780 [-1.2932, -0.6609]; -0.0024 [-0.0068, 0.0019]; 13.2601 [5.8608, 20.7335]

Table 14. Fixed effects: β_i (mean [0.025quant, 0.975quant]). 3rd set of 20 models

	3rd set of 20 models				
	H/W	**Y**	**O**	**S/V**	**UB**
ED_c	-0.401 [-0.589, -0.213]	0.377 [0.306, 0.448]	0.002 [0.0004, -0.003]	-8.071 [-9.81, -6.458]	-0.352 [-0.581, -0.123]
ED_h	28.331 [19.987, 36.818]	-8.485 [-12.444, -4.231]	0.126 [0.061, 0.192]	142.972 [102.013, 183.879]	29.941 [19.739, 40.307]
DH_c	0.881 [-0.034, 1.796]	-2.281 [-2.499, -2.063]	-0.0015 [-0.007, 0.004]	23.964 [15.484, 32.427]	0.995 [-0.105, 2.095]
DH_h	0.934 [0.314, 1.555]	-1.073 [-1.290, -0.856]	0.0006 [-0.003, 0.004]	17.791 [11.950, 23.647]	0.969 [0.221, 1.718]

Table 15. Fixed effects: β_i (mean [0.025quant, 0.975quant]). 4th set of 4 models

	4th set of 4 models
	H/W; Y; O; S/V; UB
ED_c	-0.464 [-0.573, -0.354]; 0.334 [0.285, 0.382]; 0.001 [0.0007, 0.002]; -6.581 [-7.693, -5.467]; -0.437 [-0.569, -0.306]
ED_h	19.828 [11.871, 27.959]; -17.633 [-21.094, -14.116]; 0.051 [-0.0009, 0.102]; 118.441 [73.168, 163.323]; 14.505 [4.988, 24.228]
DH_c	0.866 [0.412, 1.320]; -2.164 [-2.360, -1.968]; -0.001 [-0.004, 0.001]; 13.646 [9.002, 18.288]; 0.976 [0.431, 1.521]
DH_h	0.895 [0.429, 1.362]; -0.968 [-1.169, -0.766]; 0.0004 [-0.002, 0.003]; 12.578 [7.816, 17.351]; 0.920 [0.361, 1.480]

Those significant covariates have the mean, 0.025 quant and 0.975 quant without a change of sign. The positive sign implies that the response variable increases when the covariate does, while the negative sign implies that the response variable decreases when the covariates does. For instance, the shape factor (S/V) is always positive, whereas the year of construction (Y) is mostly negative.

On analysing the fixed effects, we note that all covariates become significant at some point. The most significant one is the S/V. Higher values of S/V mean less compactness and lead to a higher heat transfer of the building with the outdoor environment. We can observe that as the S/V is higher, all response variables (ED_c, ED_h, DH_c, DH_h) increase, which means that buildings with lower compactness require more energy to acquire indoor comfort.

Regarding Y, we observe that the discomfort hours per year for heating and cooling decrease when the year of construction is more recent. This is in accordance with the fact that the thermal characteristics of the building envelope elements improved over time, and younger buildings, which include thermal insulation materials with better performance in the façades, roofs, floors and windows, require less energy than the older ones.

In relation to the street H/W ratio, we note that as the H/W becomes higher, the response variables increase. For ED_h, this means that higher H/W ratios lead to a higher energy demand for heating because of the poor solar access that impedes natural solar gains. For ED_c, higher H/W ratios lead to a lower energy demand for cooling (values closer to 0), also due to the lower solar access (especially during the summer season), although, in this case, with a positive effect.

The urban block also has a relationship with solar access. UB_1 refers to a block with a big inner courtyard surrounded by the buildings that make up the block, while UB_2 refers to a block without a big courtyard but where the buildings have their own smaller light wells. UB_1 offers the possibility of bigger solar gains in the inner façades of the buildings than in the case of UB_2, where the smaller and taller light wells considerably reduce the solar access. Regarding the urban block, the fixed effects indicate that this covariate is significant and in the four best models that consider all the covariates, the response variables increase in the UB_2 case.

The least significant covariate is O, which is only significant for ED_c but with values extremely close to 0. On considering the orientation we observe that the models improve and the RMSE is lower. However, a deeper analysis is required to study the effect of orientation in the energy performance of buildings, as it represents a more complex performance than the other covariates.

5.4. Selection of prediction models for response variables

In accordance with the correlation coefficients and the RMSE, the four models with the best fit for predicting the response variables were selected, which are defined by four functions

$$ED_c = 2.6480 + (0.3336 \cdot Y) + (-6.5815 \cdot SV) + (0.0013 \cdot O) + (0.4637 \cdot HW) + (0.4372 \cdot UB)$$

$$ED_h = 46.8932 + (-17.6330 \cdot Y) + (118.4408 \cdot SV) + (-0.0505 \cdot O) + (19.8277 \cdot HW) + (14.5047 \cdot UB)$$

$$DH_c = 6.9152 + (-2.1639 \cdot Y) + (13.6464 \cdot SV) + (-0.0015 \cdot O) + (0.8659 \cdot HW) + (0.9760 \cdot UB)$$

$$DH_h = 44.5077 + (-0.9679 \cdot Y) + (12.5778 \cdot SV) + (-0.004 \cdot O) + (0.8951 \cdot HW) + (0.9202 \cdot UB)$$

By varying the values of the covariates, the energy demand and the discomfort hours for cooling and heating, respectively, can be obtained for every single building in the district under study. The aggregation of these results makes it possible to measure the total energy demand and the total number of discomfort hours at the urban scale.

6. Conclusions

This work has presented the development of four models for predicting the energy performance of the existing residential building stock at an urban scale. The energy demand and the discomfort hours were taken as response variables and five covariates, at both the building and the urban scale, were considered as potentially significant for conducting an accurate prediction. The proposed methodology was applied to an urban district in the city of Castellón de la Plana (Spain) and the specific characteristics of the district were taken as input parameters to conduct the study. A total of 240 dynamic simulations were run with the EnergyPlus software, considering a combination of the five covariates, which allowed the response variables to be determined for a set of sample buildings. The results obtained were taken as input data to develop the prediction models by using the R-INLA package. The statistical analysis also made it possible to identify the fact that the shape factor of the building and the year of construction were the most significant parameters, as they present the highest correlation coefficients. The least significant is the solar orientation of the main façade of the building. However, all covariates are significant at some point and the inclusion of all of them notably improves the accuracy of the models, as the CPO and DIC demonstrated.

The method proposed in this paper allows the energy performance of individual buildings to be determined by taking into account not only the physical features of the building, but also the features of the surrounding area, thereby demonstrating that the urban morphology has a relevant influence on the building's passive energy efficiency. By aggregation of the results obtained for individual buildings, it is possible to estimate the energy performance of buildings at the urban scale for a specific neighbourhood or even an entire city. Thus, the methodology can be extrapolated to any other urban area, by implementing the stages described above.

In addition, the methodology could be extended to integrate additional covariates with the aim of investigating their influence on the response variables (energy demand and discomfort hours). Similarly, new response variables related to the energy performance of residential stocks could be explored, following the approach developed in this study. Moreover, the presented methodology is applicable regardless of the simulation software that generates data collection.

The implementation of the models developed in a city allows the identification of those urban areas that require a higher energy demand and that incur in major thermal discomfort for the buildings' occupants. Thus, the models establish a useful decision-making tool for the local authorities, architects, urban planners and developers to make it easier for them to identify those urban areas that require more urgent intervention and energy refurbishment.

Abbreviations

Nomenclature

MCMC	Markov Chain Monte Carlo
INLA	Integrated Nested Laplace Approximation
SPEDE	Stochastic Partial Differential Equation
RMSE	root mean square error
DIC	deviance information criterion
CPO	conditional predictive ordinate
S	building external envelope surface area (m^2)
V	building conditioned volume (m^3)
S/V	shape factor (m^{-1})
EPC	energy performance certificate
MF	multi-family
SF	single-family
H	Height
W	Width
ED	energy demand (kWh/m^2year)
DH	discomfort hours (h/year)
GIS	Geographic Information System
N	North
S	South
E	East
W	West

Subscripts

$_h$	heating
$_c$	cooling
$_B$	building
$_U$	urban

Acknowledgements

Authors would like to thank the Architectural Construction Area of the Universitat Jaume I for providing the DesignBuilder software, which was used to conduct the data set in this research. Authors would also thank the two anonymous reviewers for their constructive comments, which contributed to improve this work.

References

Martins, T.A.L., Adolphe, L., & E.G. Bastos, L. (2014). From solar constraints to urban design opportunities: Optimization of built form typologies in a Brazilian tropical city. *Energy and Buildings*, *76*, 43–56. doi:10.1016/j.enbuild.2014.02.056

Aksoezen, M., Daniel, M., Hassler, U., & Kohler, N. (2015). Building age as an indicator for energy consumption. *Energy and Buildings*, *87*, 74–86. doi:10.1016/j.enbuild.2014.10.074

Blangiardo, M., Camaletti, M., 2015. Spatial and Spatio-temporal Bayesian Models with R-INLA. John Wiley & Sons, Chichester, West Susex.

Blangiardo, M., Cameletti, M., Baio, G., & Rue, H. (2013). Spatial and Spatio-temporal Epidemiology Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-Temporal Epidemiology*, *4*, 33–49. doi:10.1016/j.sste.2012.12.001

Boardman, B. (2007). Examining the carbon agenda via the 40% House scenario. *Building Research & Information*, *35*(4), 363–378. doi:10.1080/09613210701238276

Braulio-Gonzalo, M., Ruá Aguilar, M., Bovea Edo, M., 2015. Characterisation of urban patterns at the neighbourhood scale as an energy parameter. Case study: Castellón de la Plana. *Proceedings of the II International and IV National Congress on Sustainable Construction and Eco-Efficient Solutions*. University of Seville, Sevilla, pp. 1069-1079.

Cameletti, M., Lindgren, F., Simpson, D., & Rue, H. (2013). Spatio-temporal modeling of particulate matter concentration through the SPDE approach, (D), 109–131. doi:10.1007/s10182-012-0196-3

Cheng, V., & Steemers, K. (2011). Modelling domestic energy consumption at district scale: A tool to support national and local energy policies. *Environmental Modelling & Software*, *26*(10), 1186–1198. doi:10.1016/j.envsoft.2011.04.005

CTE. Código Técnico de la Edificación. , Pub. L. No. Orden FOM/1635/2013 (2013). España. Retrieved from http://www.boe.es/diario_boe/txt.php?id=BOE-A-2013-9511

Dall'O', G., Galante, A., & Torri, M. (2012). A methodology for the energy performance classification of residential building stock on an urban scale. *Energy and Buildings*, *48*, 211–219. doi:10.1016/j.enbuild.2012.01.034

Dascalaki, E. G., Droutsa, K. G., Balaras, C. A., & Kontoyiannidis, S. (2011). Building typologies as a tool for assessing the energy performance of residential buildings – A case study for the Hellenic building stock. *Energy and Buildings*, *43*(12), 3400–3409. doi:http://dx.doi.org/10.1016/j.enbuild.2011.09.002

DB DesignBuilder Software (2014) Available from: www.desingbuilder.co.uk

DB DesignBuilder Help official web page. Available from: http://www.designbuilder.co.uk/helpv4.5/ (last accessed: 25 October 2015)

Farahbakhsh, H., Ugursal, V. I., & Fung, A. S. (1998). A residential end-use energy consumption model for Canada. *International Journal of Energy Research*, *22*(13), 1133–1143. doi:10.1002/(SICI)1099-114X(19981025)22:13<1133::AID-ER434>3.0.CO;2-E

Florio, P., & Teissier, O. (2015). Estimation of the Energy Performance Certificate of a housing stock characterised via qualitative variables through a typology-based approach model: A fuel poverty evaluation tool. *Energy and Buildings*, *89*, 39–48. doi:10.1016/j.enbuild.2014.12.024

Futcher, J. A., & Mills, G. (2013). The role of urban form as an energy management parameter. *Energy Policy*, *53*, 218–228. doi:10.1016/j.enpol.2012.10.080

Gneiting, T., & Raftery, A. E. (2007). Strictly Proper Scoring Rules , Prediction , and Estimation Strictly Proper Scoring Rules , Prediction , and Estimation, *102*(477), 359–378. doi:10.1198/016214506000001437

Gouveia, J. P., Fortes, P., & Seixas, J. (2012). Projections of energy services demand for residential buildings: Insights from a bottom-up methodology. *Energy*, *47*(1), 430–442. doi:10.1016/j.energy.2012.09.042

Granadeiro, V., Correia, J. R., Leal, V. M. S., & Duarte, J. P. (2013). Envelope-related energy demand: A design indicator of energy performance for residential buildings in early design stages. *Energy and Buildings*, *61*, 215–223. doi:10.1016/j.enbuild.2013.02.018

gvSIG Asociación (2015). Available at http://www.gvsig.com/ (accessed 5 April 2015).

Hens, H., Verbeeck, G., & Verdonck, B. (2001). Impact of energy efficiency measures on the CO2 emissions in the residential sector, a large scale analysis. *Energy and Buildings*, *33*(3), 275–281. Retrieved from http://www.sciencedirect.com/science/article/pii/S037877880000092X

Huang, Y. J., & Berkeley, L. (2000). A Bottom-Up Engineering Estimate of the Aggregate Heating and Cooling Loads of the Entire US Building Stock Prototypical Residential Buildings. In *Proceedings of the 2000 ACEEE summer study on energy efficiency in buildings* (pp. 135–148). Pacific Grove.

INE. (2015). Spanish Statistical Office. Available at http://www.ine.es/ (accessed 30 May 2014)

Johnston, D., Lowe, R., & Bell, M. (2005). An exploration of the technical feasibility of achieving CO2 emission reductions in excess of 60% within the UK housing stock by the year 2050. *Energy Policy*, *33*(13), 1643–1659. doi:10.1016/j.enpol.2004.02.003

Košir, M., Capeluto, I. G., Krainer, A., & Kristl, Ž. (2014). Solar potential in existing urban layouts—Critical overview of the existing building stock in Slovenian context. *Energy Policy*, *69*, 443–456. doi:10.1016/j.enpol.2014.01.045

Machairas, V., Tsangrassoulis, A., & Axarli, K. (2014). Algorithms for optimization of building design: A review. *Renewable and Sustainable Energy Reviews*, *31*, 101–

112. doi:10.1016/j.rser.2013.11.036

Mauro, G. M., Hamdy, M., Vanoli, G. P., Bianco, N., & Hensen, J. L. M. (2015). A new methodology for investigating the cost-optimality of energy retrofitting a building category. *Energy and Buildings*, *107*, 456–478. doi:10.1016/j.enbuild.2015.08.044

McKenna, R., Merkel, E., Fehrenbach, D., Mehne, S., & Fichtner, W. (2013). Energy efficiency in the German residential sector: A bottom-up building-stock-model-based analysis in the context of energy-political targets. *Building and Environment*, *62*, 77–88. doi:10.1016/j.buildenv.2013.01.002

Natarajan, S., & Levermore, G. J. (2007). Predicting future UK housing stock and carbon emissions. *Energy Policy*, *35*(11), 5719–5727. doi:10.1016/j.enpol.2007.05.034

Okeil, A. (2010). A holistic approach to energy efficient building forms. *Energy and Buildings*, *42*(9), 1437–1444. doi:10.1016/j.enbuild.2010.03.013

Penna, P., Prada, A., Cappelletti, F., & Gasparella, A. (2014). Multi-objectives optimization of Energy Saving Measures in existing buildings. *49th AICARR International Conference - Historical and Existing Buildings: Designing the Retrofit*, *95*, 57–69. doi:10.1016/j.enbuild.2014.11.003

Pettit, A. L. I., Journal, S., Statistical, R., & Series, S. (1990). The conditional predictive ordinate for the normal distribution, *52*(1), 175–184.

R Development Core Team (2011) R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, 2011

R-INLA project, accessed on 15 September 2015

Rue, H., & Martino, S. (2007). Approximate Bayesian inference for hierarchical Gaussian Markov random field models, *137*, 3177–3192. doi:10.1016/j.jspi.2006.07.016

Rue, H., & Martino, S. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, 319–392.

Shorrock, L., & Dunster, J. (1997). The physically-based model BREHOMES and its use in deriving scenarios for the energy use and carbon dioxide emissions of the UK housing stock. *Energy Policy*, *25*(12), 1027–1037. doi:10.1016/S0301-4215(97)00130-4

Snäkin, J.-P. A. (2000). An engineering model for heating energy and emission assessment The case of North Karelia, Finland. *Applied Energy*, *67*(4), 353–381. Retrieved from http://www.sciencedirect.com/science/article/pii/S0306261900000350

Spiegelhalter, D. J., Best, N. G., & Carlin, B. P. (2002). Bayesian measures of model complexity and fit, 583–639.

Theodoridou, I., Papadopoulos, A. M., & Hegger, M. (2011). Statistical analysis of the Greek residential building stock. *Energy and Buildings*, *43*(9), 2422–2428.

doi:10.1016/j.enbuild.2011.05.034

Tsanas, A., Xifara, A., (2012). Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and Buildings,* 49, 560-567. doi:10.1016/j.enbuild.2012.03.003

U.S. Department of Energy (2015) Energy Efficiency and Renewable Energy. EnergyPlus 8.10: energy simulation software. EnergyPLUS. Available from: http://apps1.eere.energy.gov/buildings/energyplus/

Vlad, I. T., Juan, P., & Mateu, J. (2015). Bayesian spatio-temporal prediction of cancer dynamics. *Computers and Mathematics with Applications*, *70*(5), 857–868. doi:10.1016/j.camwa.2015.06.006

Yezioro, A., Capeluto, I. G., & Shaviv, E. (2006). Design guidelines for appropriate insolation of urban squares. *Renewable Energy*, *31*(7), 1011–1023. doi:10.1016/j.renene.2005.05.015