6th International Building Physics Conference, IBPC 2015

Heat demand estimation for different building types at regional scale considering building parameters and urban topography

Nils Schüler[a,*], Alessio Mastrucci[b], Alexandre Bertrand[a,b], Jessen Page[c], François Maréchal[a]

[a]*Industrial Process and Energy Systems Engineering Group, École Polytechnique Fédérale de Lausanne, Switzerland*
[b]*Environmental Research and Innovation, Luxembourg Institute of Science and Technology, Luxembourg*
[c]*Institut Systèmes Industriels, HES-SO Valais-Wallis, Switzerland*

Abstract

This study aims towards an improved estimation of annual heat demand of the building stock for an entire region. This requires the holistic representation of aspects influencing the heat demand of buildings, namely their geometry, fabric, users and surrounding environment. A large data base for the building stock of the Swiss canton of Geneva was systematically assessed to identify parameters suited for representation of these aspects. Due to the expectable differences in heat demand, the building stock was categorized into 8 building types. For each type a multiple linear regression model was developed to predict the heat demand.
An aspect which has so far been neglected by regression models of buildings' heat demand is the influence of microclimate. Since this aspect is considerably influenced by the surrounding topography, parameters suited for the representation of the urban topography were defined and included in the regression.
The regression analysis revealed that all models were able to explain high shares of the variance ($\overline{R}^2$: 71.2% to 88.9%). The mean average errors for hotel, health-care, educational and office buildings were ranging between 30.2% and 39.8% while the error for residential buildings was 17.8%. The suitability and of the selected parameters for heat demand prediction was analyzed in detail for the residential building model and revealed that almost all chosen parameters were highly suited.
Keywords: energy demand; city-scale; building types; multiple linear regression; density parameters

1. Introduction

The building sector accounts for 40% of the final energy consumption in Europe and thus presents a high potential for energy savings [1]. A reliable prediction of the heat demand of buildings is required to improve the design and operation of energy systems and target demand reduction measurements as e.g. incentives for refurbishment at city or regional scale. A prediction on such a scale in turn requires to regard several building types representing as comprehensive as possible the entire building stock. Such a prediction further requires a holistic representation of aspects influencing the heat demand of buildings, namely their geometry, fabric, users and surrounding environment. The representation of these different aspects needs building specific information. However, the availability of information in terms of both different parameters and a complete parameter set for each building usually limits the detail of the representation and the number of buildings, respectively, for which a heat demand estimation can be made. Thus one objective of this work was to include the mentioned aspects influencing buildings' heat demand into the estimation while taking into consideration the availability of data.

Amongst possible techniques, physical and statistical modeling have been widely used for the heat demand estimation of building stocks [2]. Due to their mathematical nature and diametric to it, both modeling methods have different strengths and weaknesses regarding the representation of the different aspects of buildings [3]. While the representation of statistic or stochastic information as user-related aspects is an issue for physical models, the representation of physical aspects as building geometry or topography is usually less or not considered by stochastic models. Due to their general higher flexibility in incorporated information [4, 5], a stochastic method in form of multiple linear regression modeling was used for this work. In order to overcome the previously mentioned drawbacks

* Corresponding author. Tel.: +41-21-69 36754; fax: +41-21-69 33502.
E-mail address: nils.schueler@epfl.ch

of statistical models and better represent physical aspects, a set of geometric parameters was included in the regression. An aspect so far neglected by regression models of buildings' heat demand is the influence of solar gains and microclimate. This aspect is substantially characterized by the surrounding topography [6]. Thus meaningful parameters were included, which are suited to represent urban topography and are easily derivable from building data.

2. Materials and methods

2.1. Data description

This study was carried out on basis of georeferenced data for the canton of Geneva in Switzerland. Thus next to mostly urban areas also rural areas were regarded in this study. The data were obtained from the territorial information system for Geneva (SITG) [7], which provides a large amount of publicly available data for the entire building stock. The 115 listed building types were sorted into 7 categories according to expectable similarities in heat demand: residential buildings, offices, commercial buildings for wholesale and retail, industries (factories and workshops, excluding storehouses), educational buildings (schools, universities and research institutes), healthcare buildings (excluding sport facilities) and hotels (including guest houses). An eighth building type represents less frequent buildings with yet a considerable heat demand (e.g. museums, libraries, churches, stations, airport). Further buildings with minor or without heat demand as e.g. garages were excluded. Figure 1 shows the resulting number of buildings per building type, the number of buildings for which annual heat demand measurements were available and the number

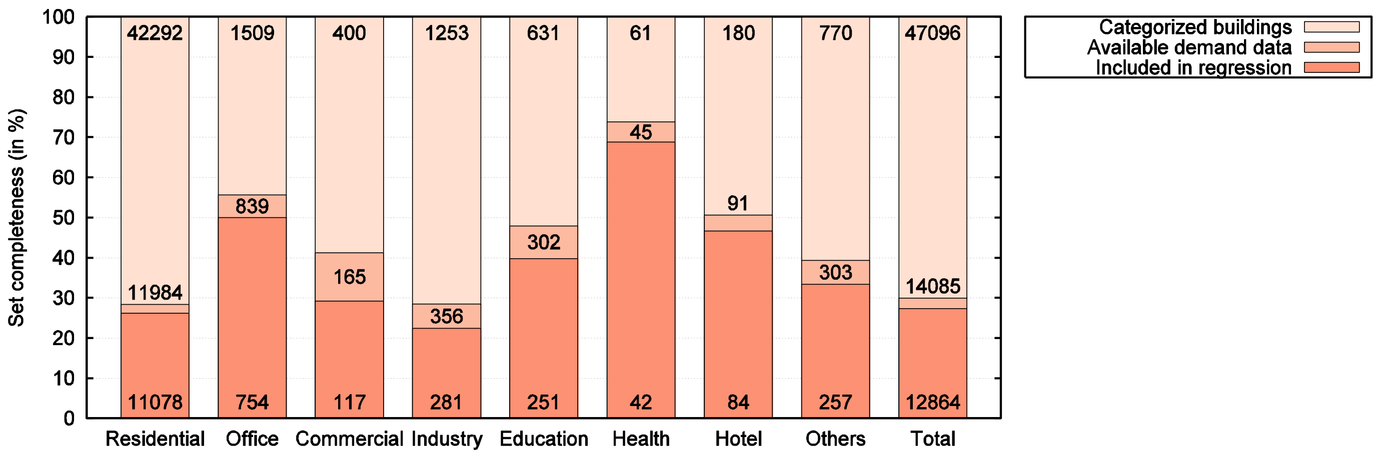

Figure 1. Categorization of Geneva's building stock in building types and regarded data availability (absolute values indicated)

of buildings regarded in the regression analysis excluding records with incomplete parameter sets.

Figure 2 summarizes all used parameters and their availability for the regarded building stock and in terms of their completeness. The annual heat demand of each building was available in form of an average heat demand for the years 2010 to 2012, which included both hot water and weather-corrected space heating energy demand, normalized by the energetic reference floor area (SRE) of the building. However, it was refrained from using the normalized values for two reasons: First, the SRE is almost only available for buildings for which heat demand data are available. Predicting the SRE specific demand would thus not help to predict the total demand for the rest of the building stock. Second, regressing the normalized value does not allow to compare the influence of the floor area against other parameters. Therefore the total consumption was obtained by multiplying the area specific heat demand and the energetic reference floor area (SRE) of each building.

Information about the geometry of buildings are available from both an extensive building cadaster and a three-dimensional city model. The cadaster includes height and number of floors. The gross floor area is estimated by multiplying the number of floors with the footprint area. Analogously buildings' volumes were computed using their footprint areas and height. The three-dimensional model provided further information about the their envelope surface in terms of total, shared and unshared wall surfaces as well as roof surface and average roof pitch. An important further aspect is the buildings' fabric. Since information about e.g. U-values and airtightness is not available, the building's construction period is used as a representing parameter [8]. Reliable information about renovations of buildings was

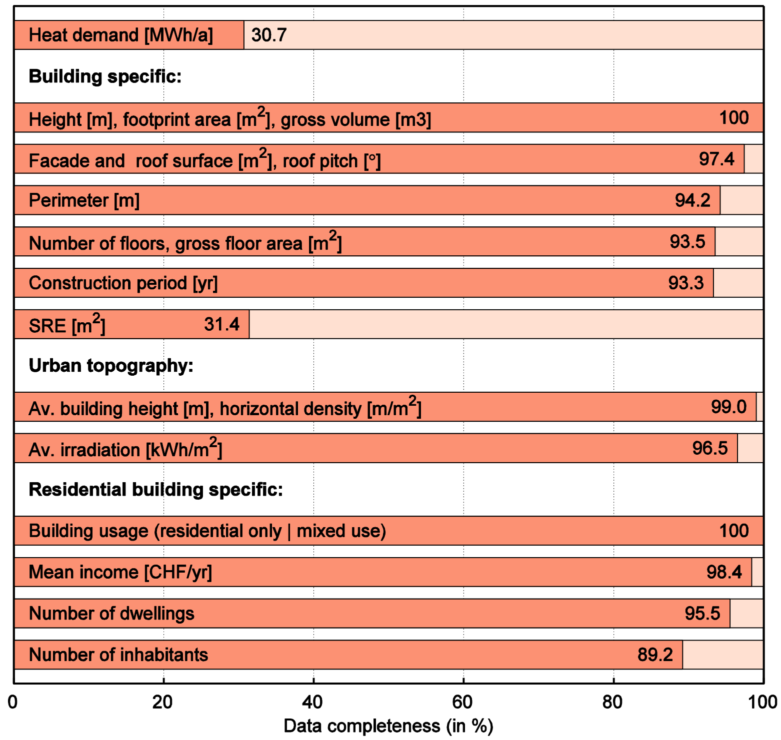

Figure 2. Parameter availability and completeness for the total 47096 categorized buildings and 42292 residential buildings, respectively (percentage values indicated)

not available and thus left out of the analysis since neither is the type of renovation specified nor is a clear distinction possible if a building was not renovated or the entry is missing.

To reflect potential effects of the urban topography on the heat demand, two additional parameters were defined. As a basis for their definition the partition of the canton of Geneva into 474 sectors, defined by the cantonal office of statistics, was used. Two considerations led to their definition: (a) Long and shortwave radiation processes and micro climate are rather influenced by buildings' external dimensions than by their area or volume. (b) Horizontal and vertical topography might have different effects on the demand since they are influencing obstruction to solar irradiation and convective heat transfer in different ways [6]. Thus the vertical topography was represented by the average building height within each of the statistical sectors. The horizontal topography was represented by the ratio of the sum of all buildings' perimeters P_{build} within one sector and the sector's area A_{sect}: $p_{sect} = \sum P_{build} / A_{sect}$. These factors, specific to each sector, were attributed to the respective buildings and together indicate the amount of façade surfaces within each sector normalized by the sectors area. To further specifically represent the effect of solar gains, data of a solar cadaster is used, which gives for every roof surface the area specific, average annual irradiation.

Additional information is available for residential buildings in the form of number of dwellings per building, inhabitants per building and their approximate income. The latter is not available on building level but on the level of the previously mentioned statistical sectors as average income. Also information is available if a building is used only for residential purposes or for other purposes as well (mixed).

2.2. Regression analysis

Multiple linear regression has been previously used by several authors for predicting the annual heat demand of buildings [4, 9]. The suitability of various regression methods has been compared by different authors and it was mostly found that linear regression models have a performance comparable to others and thus are preferable since being computationally more efficient and easier interpretable [10]. Thus linear multiple regression models for heat demand prediction were developed separately for each building type using the statistical software R [11]. Each of these models included all parameters summarized in figure 1 except of the SRE and the parameters specific to residential buildings. The models were fitted using the Ordinary Least Squares method. Kolter and Ferreira [10] showed how a logarithmic transformation of predicted variable and predictors substantially improves the performance of methods for the regression of buildings' heat demand. Thus all parameters whose distributions were clearly skewed on the original scale, were logarithmically transformed (indicated in table 2). Construction periods and information about mixed usage were included as factorial variables. The initial regression model for each building type included the full parameter set as listed in figure 2 except the SRE due to its limited availability. Starting with this parameter set stepwise regression was performed for the residential building model in order to identify variables of less significance for the heat demand estimation of this building type.

The assumptions of multiple linear regression were carefully verified and the prediction error was estimated using the mean average percentage error (MAPE) and the root-mean-square error (RMSE) for both the logarithmically scaled model outputs (log) and the ones transformed back to the original scale. This makes it possible to compare the performance of the models of the different building types against each other and values reported by other authors.

Due to the partly very large sample size, 10-fold cross-validation was chosen since being a non-exhaustive method. However, the inclusion of the building period as factorial variable raises an issue if the entire sample per category contains only one building built within a specific period. Since every building will be once part of the test data, the model will then not be trained on data containing this specific period thus preventing the application of the resulting model for this building. The issue was circumvented by excluding the regarding buildings from the cross-validation. This was the case for the categories of commercial (1 building excluded), educational (1) and healthcare (3) buildings.

3. Results and discussion

3.1. Results of the regression analyses for the whole building stock

Results of the regression analysis for the different building types are shown in table 1. The models were able to account for between 73.1% and 88.9% of the variances (R^2). The difference between R^2 and the adjusted coefficient of determination $\bar{R}^2$, naturally increases with decreasing sample size [12]. Considering the sample size, the highest share of explained variance is achieved by the residential model with 88.9%.

Table 1. Results of the regression analyses for the different building types

Building type	Residential	Office	Commerce	Industry	Education	Health	Hotel	Other
R^2	0.889	0.795	0.800	0.744	0.838	0.915	0.889	0.731
$\bar{R}^2$	0.889	0.789	0.760	0.725	0.824	0.854	0.861	0.712
MAPE [%]	17.8	39.8	47.0	121.6	38.7	36.9	30.2	58.4
MAPE (cross-validated) [%]	17.9	41.6	64.4	167.1	44.0	85.2	38.6	66.2
RMSE (log)	0.243	0.488	0.555	0.714	0.468	0.428	0.374	0.623
RMSE (log, cross-validated)	0.244	0.506	0.666	0.783	0.531	0.765	0.488	0.674
RMSE [MWh/a]	61.3	423.4	458.1	361.5	419.6	460.0	432.1	271.3
RMSE (cross-validated) [MWh/a]	61.6	449.1	674.5	380.2	691.3	1690.3	642.5	306.0

The p-values are for all models close to zero demonstrating that the models are globally significant and the null-hypothesis can be rejected. The scatterplots of the residuals against the predicted values were checked but revealed no pattern for any building type thus showing no heteroscedasticity in the errors.

The lowest errors were achieved for the residential building type with a MAPE of 17.8% and an RMSE of the logarithmically scaled outputs of 0.243. The MAPE for hotel, healthcare, educational and office buildings range between 30.2% and 39.8%.

Three aspects of industrial buildings differ quite considerably from building to building which might explain the very high error for this type (MAPE: 167.1%): (a) The share of gross floor which is heated, (b) the building's operation with set-point temperatures and air exchange rates and (c) the internal heat gains or loads due to installed equipment and people. These arguments might as well apply to commercial buildings (MAPE: 47.0%) although the expectable differences should be lower. Thus a common conclusion for these both categories is that a consideration of the specific type of industry or commerce might reduce the model error. The high model error for last building type representing other buildings (MAPE: 58.4%) is not surprising because of the diversity of expectable demand patterns within this type.

The discrepancy between RMSE of scaled (log) and unscaled [MWh/a] model output stems from the differing average heat demand of the buildings. For example a high RMSE for unscaled outputs of hotel buildings is not a contradiction to a low RMSE for scaled outputs and a low MAPE since the hotel buildings were found to have the highest average heat demand within the measured demand data.

The difference between the errors for models of the entire data sets and the cross-validated errors, i.e. the optimism index, is lowest for the residential building model (ΔMAPE: 0.1%) and considerably low (ΔMAPE: 1.6 to 8.4%) for all other categories apart from industrial and healthcare buildings. The high index for the healthcare building model (ΔMAPE: 48.3 %) hints towards an overfitting of the data due to a small sample size of 42 buildings. For industrial buildings (ΔMAPE: 45.5 %) overfitting is not so likely to be a problem since the sample size of 281 buildings is quite large. It rather points out again the low suitability of statistical models for predicting industrial heat demand without having further information about the type of the industry.

The good model results for residential buildings hint towards a suitability of the chosen parameter set for this category. A further inclusion of building type specific parameters is expected to improve the models for other categories.

3.2. Results of the regression analysis for residential buildings

The influence and significance of the various parameters is assessed more in detail exemplary for the model of residential buildings since it represents the biggest part of the building stock. Table 2 lists the estimated model coefficients, their standard errors, t-values, significances and variance inflation factors. When interpreting the values it has to be noted that the logarithmic transformation of some of the parameters prohibits a direct comparison in terms of value to unscaled parameters.

The interpretation of the regression coefficients requires further the prevention of multi-collinearity [12]. Thus only a limited set of all available geometric parameters was included. Multi-collinearity was assessed by examining the variance inflation factors (VIF). The parameters were selected thus that the VIFs showed values at maximum around 10 indicating no serious multi-collinearity issues [13]. Of all initially used parameters only the number of floors was excluded from the model for residential buildings due to the results of the stepwise regression.

Table 2. Statistical summary of the linear regression model for residential buildings: Factorial coefficients are compared against a reference value which is thus not listed. The reference values are “<1919” for the building period and “Mixed usage” for the usage of residential buildings. Coefficients marked with lt are logarithmically transformed prior to the regression.

Parameter	Est. coeff. [$\cdot 10^{-3}$]	Std. Error [$\cdot 10^{-3}$]	t-value	Signif.	VIF
Intercept	6846.46	48.26	141.85	<0.0001	
Building specific					
Height	22.81	0.66	34.46	<0.0001	5.03
Gross floor area^{lt}	244.61	9.63	25.41	<0.0001	11.04
Shared façade surface^{lt}	-3.22	1.40	-2.30	0.0215	1.34
Roof surface^{lt}	429.54	11.13	38.58	<0.0001	4.12
Roof pitch	-4.07	0.24	-16.90	<0.0001	1.83
1919-1945	17.63	9.45	1.87	0.0621	1.49
1946-1960	-7.62	8.83	-0.86	0.3881	1.82
1961-1970	-0.22	9.12	-0.02	0.9808	2.05
1971-1980	-7.28	9.29	-0.78	0.4332	1.89
1981-1985	-6.67	14.01	-0.48	0.6340	1.24
1986-1990	-43.51	12.00	-3.63	0.0003	1.36
1991-1995	-89.11	11.26	-7.91	<0.0001	1.53
1996-2000	-99.23	11.10	-8.94	<0.0001	1.58
2000-2005	-183.00	13.00	-14.08	<0.0001	1.62
2006-2010	-382.15	20.23	-18.89	<0.0001	1.21

Urban topography					
Average building height	0.18	0.64	0.28	0.7807	2.68
Horizontal density	875.77	116.56	7.51	<0.0001	1.99
Average irradiation	0.15	0.02	6.89	<0.0001	1.37
Residential building specific					
Number of dwellings^{lt}	161.31	6.81	23.70	<0.0001	6.64
Number of inhabitants^{lt}	65.38	5.95	10.99	<0.0001	5.11
Average income	-0.00	0.00	-4.66	<0.0001	1.78
Residential only	-35.17	5.61	-6.27	<0.0001	1.35

Almost all chosen parameters are highly significant for the model of residential buildings. The estimated coefficients for geometric parameters reveal importance of these parameters for the estimation of buildings' heat demand. Both an increasing heat demand with building dimensions and a decreasing demand with increasing area of shared walls is reasonable. The estimated coefficients for construction periods show that the demand of buildings constructed from 1919 on generally increases with building age. Since buildings constructed before 1919 are forming the reference period, the model further shows that those buildings tend to have a lower heat demand than buildings constructed between 1919 and 1945 and buildings constructed only after 1985 tend to have a considerable lower heat demand. This result is highly in-line with the findings of Aksoezen et al. [14] and reveals a non-linear dependency of heat demand on building age. Thus this effect would be missed by linear models e.g. by considering the age as an integer in a linear regression model. The decrease in heat demand of buildings constructed after 2005 is further remarkable.

The horizontal density factor is found to have a high estimated coefficient. An interpretation of the topographic factors, however, is complicated since the building density affects heat demand in different ways. Due to increased obstruction denser districts should correspond to lower solar gains and thus increased demand. At the same time the increased amount of façade surfaces means also an increased long-wave exchange and thus a decreased heat demand. When assessing the significance of the chosen topographic factors, it has to be taken into account, that both factors are not defined per building and thus rough in comparison to most other parameters. A future study shall therefore determine and compare further parameters to account for urban topography. The data from the solar irradiation cadaster did not have a considerable effect. Here again the question is if the effect of solar gains on annual heat demand on a regional scale is generally not so high, which is e.g. quite plausible for older buildings, or if more suited parameters could be identified for a better representation of this effect. In fact the solar irradiation cadaster represents only the amount of incident irradiation on roofs. The effect of solar irradiation on vertical walls is thus not specifically represented. However, the previously discussed topographic parameters represent the amount of façade surfaces within a sector normalized by sector area and should thus incorporate information about the amount of shaded surfaces. Furthermore information about glazing ratios of buildings should be of particular importance for the estimation of the effect of solar gains.

The analysis of the additional parameters for residential buildings reveals that the number of dwellings per building has a remarkable bigger effect on heat demand than the number of inhabitants. The information about average income has almost no effect, which might again be due to the fact that this information was only available at the level of statistical sectors. The negative estimated coefficient for the information that a building is only used for residential purposes means that these buildings generally have a lower heat demand than buildings with mixed usage.

4. Conclusions

In order to predict the annual heat demand of buildings within the canton of Geneva regression models for 8 building categories were developed, which represent about 47000 buildings. The models were fitted using available data for about 13000 buildings. A systematic assessment of available data was performed to identify parameters suited to represent how geometry, fabric, users and topographic environment of buildings affect their heat demand. Adjusted to the sample size the models were able to explain between 71.2% and 88.9% of the total variance. Mean average percentage errors of 30.2 to 39.8% were achieved by the models for hotel, healthcare, educational and office buildings. The lowest error (MAPE: 17.8%) and highest share of explained variance (R^2: 0.889) was obtained with the model for residential buildings, for which also information about inhabitants was respected. Further parameters specific to other building types should improve the other models as well. Moreover, parameters representing the urban topography have been included in the regression analysis. However, only the factor accounting for horizontal building density was found to have a considerable influence. A future identification and calculation of topographic parameters on a building level could improve the representation of urban topography in statistical models of buildings' heat demand.

Acknowledgements

The service of SITG is gratefully acknowledged by the authors. The data has been extracted between August 2014 and February 2015. We further want to thank Jean-Marie Fürbringer for his advice on statistical questions. This work was funded by the European Commission under the FP7-PEOPLE-2013 Marie Curie Initial Training Network "CI-NERGY" project, grant agreement number 606851, and by the Luxemburgish FNR grant agreements AFR - 5775018 "OptiHeat" and AFR "DAEDALUS".

References

[1] The European Parliament and the Council of the European Union. Directive on energy efficiency, (Oct. 2012). eur-lex.europa.eu/-LexUriServ/LexUriServ.do?uri=OJ:L:2012:315:0001:0056:en:PDF
[2] L. G. Swan, V. I. Ugursal. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. Renewable and Sustainable Energy Reviews 13 (8) (2009) 1819–1835. http://www.sciencedirect.com/science/article/pii/S1364032108001949
[3] A. Foucquier, S. Robert, F. Suard, L. Stéphan, A. Jay. State of the art in building modelling and energy performances prediction: A review. Renewable and Sustainable Energy Reviews 23 (2013) 272–288. http://www.sciencedirect.com/science/article/pii/S1364032113001536
[4] A. Mastrucci, O. Baume, F. Stazi, U. Leopold. Estimating energy savings for the residential building stock of an entire city: A GIS-based statistical downscaling approach applied to rotterdam. Energy and Buildings 75 (2014) 358–367. http://www.sciencedirect.com/science/-article/pii/S0378778814001467
[5] O. Guerra Santin, L. Itard, H. Visscher. The effect of occupancy and building characteristics on energy use for space and water heating in dutch residential stock. Energy and Buildings 41 (11) (2009) 1223–1232. http://www.sciencedirect.com/science/article/pii/-S0378778809001388
[6] D. Robinson, U. Wilke, F. Haldi. Multi agent simulation of occupants' presence and behaviour. Proceedings of Building Simulation 2011: 2110–2117. http://infoscience.epfl.ch/record/174524?ln=en
[7] SITG. Le territoire genevois à la carte. http://ge.ch/sitg/
[8] L. G. Swan, V. I. Ugursal. A new hybrid end-use energy and emissions model of the Canadian housing stock. 3rd Canadian Solar Buildings Conference 2008.
[9] B. Howard, L. Parshall, J. Thompson, S. Hammer, J. Dickinson, V. Mod. Spatial distribution of urban building energy consumption by end use. Energy and Buildings 45 (2012) 141–151. http://www.sciencedirect.com/science/article/pii/S037877881100524X
[10] J. Z. Kolter, J. Ferreira. A large-scale study on predicting and contextualizing building energy usage. AAAI, 2011. http://www.aaai.org/ocs/-index.php/AAAI/AAAI11/paper/download/3759/4088
[11] R Core Team. R: A language and environment for statistical computing. (2013). http://www.R-project.org/
[12] S. G. Makridakis, S. C. Wheelwright, R. J. Hyndman. Forecasting: Methods and Applications. 3rd ed. New York: Wiley; 1997.
[13] J.F. Hair. Multivariate data analysis. 6th ed. New York: Pearson; 2006.
[14] M. Aksoezen, M. Daniel, U. Hassler, N. Kohler. Building age as an indicator for energy consumption. Energy and Buildings 87 (2015) 74–86. . http://www.sciencedirect.com/science/article/pii/S0378778814009207