PLEA '15
Architecture in (R)Evolution
Bologna, 9-11 September

THE SENSITIVITY OF PREDICTED ENERGY USE TO URBAN GEOMETRICAL FACTORS IN VARIOUS CLIMATES

Emilie Nault	Interdisciplinary Laboratory of Performance-Integrated Design (LIPID),
Parag Rastogi	Ecole polytechnique fédérale de Lausanne (EPFL), Switzerland.
Emmanuel Rey	Laboratory of Architecture and Sustainable Technologies (LAST), EPFL
Marilyne Andersen	LIPID, EPFL

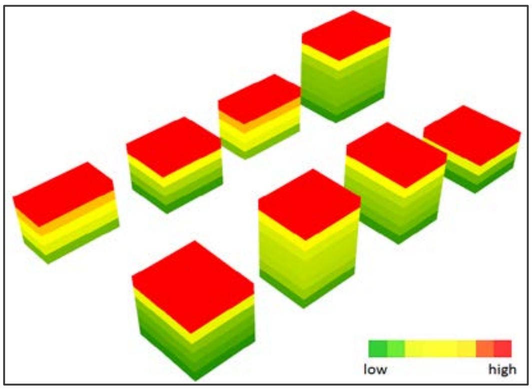

Fig 1: Heating need per floor in an early neighbourhood design

WHICH ARE YOUR ARCHITECTURAL (R)SOLUTIONS TO THE SOCIAL, ENVIRONMENTAL AND ECONOMIC CHALLENGES OF TODAY?

Research summary

Urban morphology, including building typology and layout, has a significant influence on the built environment's access to the sun, which impacts its energy exchange with the environment. This energy exchange is a strong factor in determining the comfort levels of occupants in buildings and the energy consumed to reach comfort. The influence of urban form has been quantified in previous studies for certain building typologies and programs for specific climates (i.e. location-specific case studies). We are interested in taking this further to assess the *variation*, due to climate, of the influence of different urban forms on the urban energy balance. This is part of a larger project to study the interaction between form and climate vis-à-vis energy and comfort in buildings.

In this paper, we explore this issue through simulation, in various climates, of 3D neighbourhood models. These models consist of a series of parametrically generated variations on building typologies like block, L-shaped, and courtyard block. Each neighbourhood alternative is described through a set of geometrical parameters including the form factor, window-to-floor and plot ratio.

We used an extensive database of heating and cooling uses generated by simulating each variant in a representative set of climates to assess the sensitivity of energy use to the geometrical descriptors and climate types. This is done using a regression equation whose input parameters are easily calculable, e.g. form factor, and whose output is an estimate of simulated energy use.

The aim of exploring this relationship is to use it to assess the suitability of different urban forms in a given climatic context. Moreover, it provides a promising route to avoid the necessity of detailed energy simulations in comparing the performance of different early urban design alternatives.

Keywords: climate, neighbourhood-scale, early-design phase, building typology, energy performance

1. Introduction

Urban-scale energy simulation is a complex and computationally expensive undertaking. Setting up any building simulation requires extensive and detailed knowledge of the material properties of the constructions being simulated, which may not be available at the early design phase. The complexity of this information is often unmanageable at the urban scale and the time investment is prohibitive.

To address these issues, some studies have proposed prediction models to replace building simulation. Most techniques use multiple linear regression at the building scale (Foucquier, Robert, Suard, Stéphan, & Jay, 2013), usually focusing on a unique climate (Asadi, Amiri, & Mottahedi, 2014); (Tsanas & Xifara, 2012)). Notable examples using more than one climate are the work of (Hemsath & Alagheband Bandhosseini, 2015), who conducted a sensitivity analysis to quantify the impact of certain geometrical parameters on energy use for four climates in the USA; (Lam, Wan, Liu, & Tsang, 2010), who developed multiple regression models for five climates in China, based on 12 design parameters for an office building; and (Hygh, DeCarolis, Hill, & Ranji Ranjithan, 2012)), who applied multivariate linear regression to predict the heating and cooling need for an office building, based on 27 parameters (including for example orientation, depth, wall U-value), and for four different climates in the USA.

In this paper, we are interested in assessing the variation, due to climate, of the influence of different urban forms on the energy performance. We explore this issue through the analysis of an extensive dataset of simulation results, in various climates, of 3D neighbourhood models. These models consist of a series of parametrically generated variations from base case building typologies arranged according to one of a set of urban layouts.

The energy use for heating and cooling for conditioned buildings is computed for one representative climate in each ASHRAE climate zone type. We then use multiple linear regression with a series of input parameters related to the morphology (geometry-based) and solar exposure level (irradiation-based) of the neighbourhood variants. The resulting prediction models output an estimate of the simulated energy use for heating and cooling.

2. Methodology

2.1 Overview

Fig 2 gives an overview of our approach. First, a series of neighbourhood design variants were generated by varying simple early-design phase geometrical parameters for six base case designs. Then, we carried out thermal and irradiation simulations on each variant for a series of climates. The output data was analysed with respect to the base case designs and the climates. The details of each step are given in the following sections.

2.2 Parametric modelling

A parametric modelling workflow was set up in Grasshopper®, an algorithmic modelling platform for the 3D modelling interface Rhinoceros®, to generate a series of design variants starting from six base case neighbourhood designs. Each base case consists of a replicated building typology according to a certain urban layout. An example variant for each base case is illustrated in Fig 2, along with the corresponding number of total variants. M0 to M2 come from a collaboration with an architecture and urban design firm located in located in Lausanne, Switzerland (Urbaplan), while M3 to M5 were generated as a preliminary dataset and inspired by student projects analysed in the context of a collaborative study (Rey, 2013).

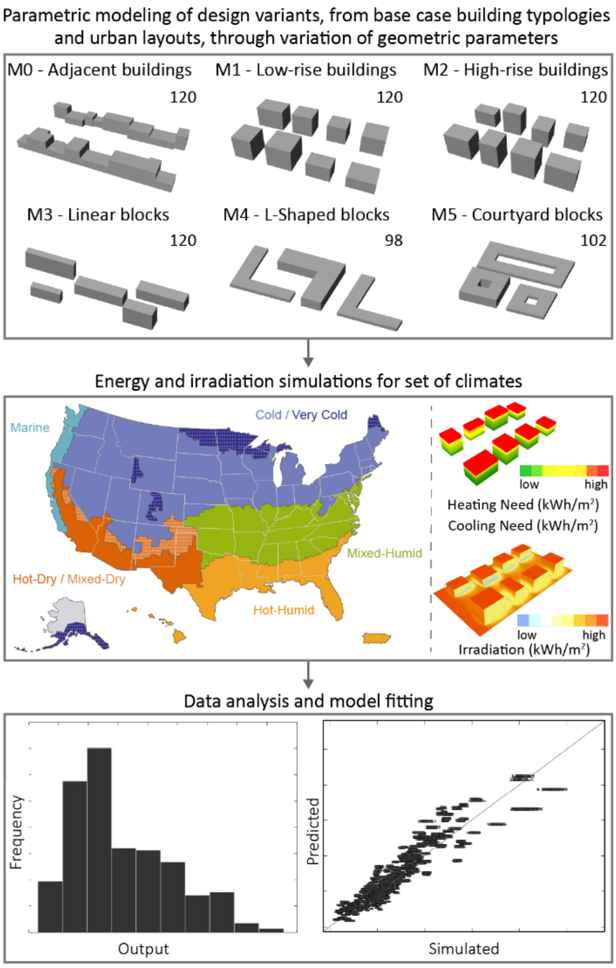

Fig 2: Main steps of the approach

In this paper, we will refer to each base case as typology Mx, as a term inclusive of both the building typology and the urban layout. The variants were generated by varying the depth, width and height of individual buildings, as well as by rotating the grid by 90°. For more details on the generation of the variants, we refer the reader to (Nault, Rey, & Andersen, 2015). Each floor (3m high) was defined as a thermal zone, and windows were modelled equally on each façade, representing between 45% and 50% window-to-wall ratio depending on the typology.

2.3 Simulation

We obtained the heating and cooling need, summed over all buildings for each design variant and normalised by total floor area, using the following steps. The initial IDF (EnergyPlus input) files were generated via the Grasshopper plug-in Archsim (Dogan, 2014). Then they were processed in MATLAB® before simulation with EnergyPlus®.

The final simulation settings are in Table 1. The internal loads are similar to ASHRAE 90.1-2007 values, while the U-values fall between the SIA 380/1 minimum and target values (SIA, 2009, p. 380). The choice of envelope properties and other settings does not imply a recommendation for any of the climates. Neither is it an indication of the prevalent construction in any of the cities. It is an arbitrary choice to test with these particular values only, consistently across all climates.

The internal loads were deliberately kept on the lower end to reduce the overwhelming influence of internal heat gains. This is clearly an arbitrary choice, and one we made after a round of pilot simulations. We realised that due to a simplification, where we apply the internal loads over the entire floor area of the building instead of dividing it up into separate zones by usage, the internal load contribution is extremely strong. So strong, in fact, that even in sub-arctic climates like Alaska (zone 8), our (highly-glazed) buildings have very little heating demand (below 80 $kWh/m^2_{floor\ area}$). It is to be expected that the effect of climate as the overwhelming driver of energy consumption is severely undercut by heavy internal heat gains.

Table 1: Settings for the thermal simulation

Parameter	EnergyPlus settings
Building function	Office
Heating/Cooling set point	20/26°C
Loads Equipment Lighting Occupancy	 12 W/m^2 3 W/m^2 0.05 pers/m^2; 80 W/pers
Ventilation	0.0125 m^3/sec-pers
Infiltration	0.1 ach
U-value Wall Floor Roof Windows (double glazing low-e with argon)	 0.186 W/m^{2}K 0.185 W/m^{2}K 0.194 W/m^{2}K 1.512 W/m^{2}K

2.4 Selection of climates

In order to cover the maximum possible variation in climate types while keeping the experiment size manageable, we chose to work with the ASHRAE climate zone system (ASHRAE, 2007). The Building Performance Database of the US Department of Energy (Richard E. Brown, et al., 2014) uses standard cities to represent each of the fifteen climates, and these weather files are freely available on the EnergyPlus web site. In addition to these cities, we also simulated Geneva, Switzerland, since it is the closest major city to the actual location of the projects on which our base cases are based.

In this paper, we only present the results from a subset of the complete set of climates due to space constraints: Burlington (Zone 6A), Fairbanks (Zone 8), Geneva (Zone 4A), Miami (1A).

2.5 Data analysis and model fitting

We first analyse the simulation results to identify any anomaly and trend in the data before proceeding with the model fitting. Multiple linear regression is then used to fit a model to the acquired data of the form:

$$y = \beta_0 + \sum_{i=1}^{P} \beta_i x_i + \epsilon \qquad (1)$$

where y is the predicted output (heating/cooling need), P the number of inputs, β_0 is a constant, β_i the model coefficients of each input x_i, a value corresponding to a geometry- or irradiation-based parameter, and ϵ the error term.

The model is trained and tested in an iterative sequence as illustrated in Fig 3. To assess the predictive accuracy of the model, we use the root mean square error (RMSE) and percentage error (PercErr) measures to quantify the deviation of the prediction ($Y_{predicted}$) from the reference (in this case simulated, Y_{test}).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(y_i - f(x_i)\right)^2}{N}} \qquad (2)$$

$$PercErr = 100 \times \frac{\sum_{i=1}^{N} \frac{|f(x_i) - y_i|}{y_i}}{N} \qquad (3)$$

where N is the number of training samples.

More details about the fitting algorithm can be found in (Nault et al., 2015). The results from that paper influence our selection of parameters in this study. For example, we discarded those parameters which were strongly correlated, to the point of redundancy. Some other inputs were rejected because the energy demand results showed either a highly non-linear trend to them or the trend was too close to a horizontal line. After this reduction in the number of candidate parameters, we trained the model using various combinations of the remaining parameters. We expect, in the future, to automate the selection of model parameters through optimization.

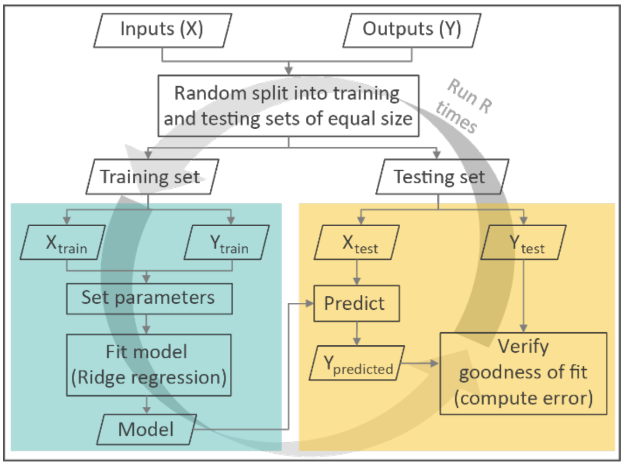

Fig 3: Main steps in the model fitting algorithm

3. Results

3.1 Simulation results

We analyse the simulated heating and cooling need for all typologies (M0 to M5) and four selected climates – Burlington (BUR), Fairbanks (FAI), Geneva (GEN) and Miami (MIA). We begin with models trained on all typologies together, to see if the typology-specific inputs we consider are sufficient to explain the variance of heating and cooling need. The results presented are for the model with the least inputs, among the best performing subset of all the input combinations tested.

The regression procedure shown in Fig 3 uses several iterations to train and test the model on a subset of the total simulation results each time. This means that we have several values of the coefficients for each set of regression parameters tested. Since the randomisation we use is not systematic, there is the possibility that the regression coefficients predicted by one particular iteration are not representative of the population. The iterations solve this by giving us a robust estimate of the ranges of the coefficients. This range is shown in Fig 4 for each climate. In general, the coefficients are not particularly sensitive to the choice of training subset (i.e. the partition of the training/testing used during the fitting phase of the model).

The sign of each coefficient (i.e. positive or negative) indicates the direction of its influence over the output. It helps us see if the model fit makes sense. For example, we know that the form factor (floor area/envelope area) is typically negatively correlated to the heating

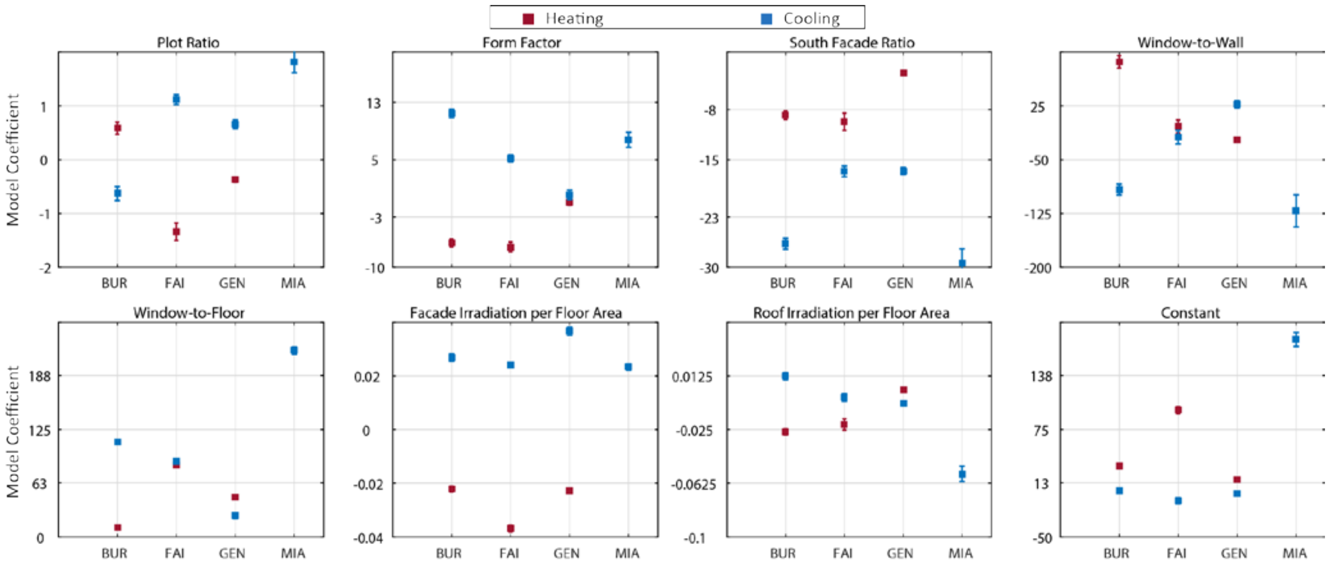

Fig 4: Regression coefficients for the selected factors. The square represents the mean and the error bar shows one standard deviation. β_1 to β_7 from Eq. 1 correspond to the inputs shown in this graph, from top left to bottom right. The constant is β_0.

need in cold climates (less compact, more heat losses and therefore more heating need). Thus, the sign of the form factor's coefficient should be negative, which is the case for all climates. However, not all coefficients are consistent in this regard. For example, the sign of the coefficient for plot ratio varies across climates, for both heating and cooling. This could be indicative of an interaction we have missed or that plot ratio should not have been included in the model. Planned future work in using optimisation to pick model parameters should address this issue. The point is always to pick the most parsimonious model possible: the less coefficients the better. The model coefficients for Geneva often diverge from the others. This seems to have no clear explanation, except perhaps that Geneva is a more moderate climate compared to the other three. Further investigation with the full suite of ASHRAE climate zones should shed more light on whether this behaviour is anomalous or the model is indeed less powerful for moderate climates.

In Fig 5, we show the results from using the coefficients predicted by each iteration of the model training. The graphs presented compare the $Y_{predicted}$ with the Y_{test} of the right-hand side of Fig 3. Some predictions are clearly outliers, vindicating our decision to always train the model over several iterations. Once again, Geneva has the worst fit for both heating and cooling.

Table 2 lists the coefficients from fitting a model, with the selected coefficients (shown in Fig 4), to all available data. This is the 'best' fit, as it were, for a given climate. As such, it cannot be verified against any other simulation data. The signs of the various coefficients tend to be consistent across climates, with a few exceptions. In the formulation of the model we present here, the contributions of total irradiation on façade and roof, normalised by floor area, seem to pale in comparison with those of window-to-wall and window-to-floor ratios. For Miami, the constant seems to have a huge impact. This could be suggestive of a certain 'minimum load' that must be expected in this climate for our building. This probably arises due to a mismatch between the climate type (warm, humid) and the basis for the U-values, which come from a Swiss norm designed for the colder conditions present there.

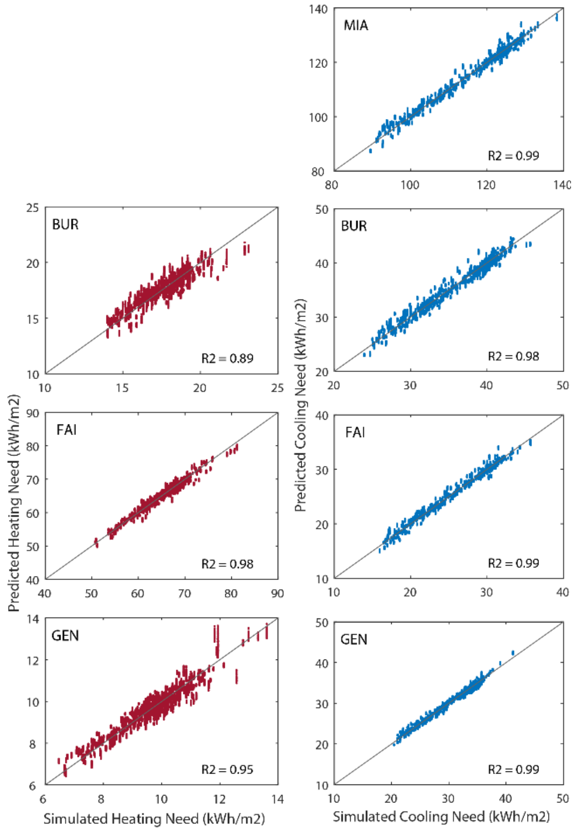

Fig 5: Predicted vs simulated output over 50 training/testing runs for each climate. There are no predictions for heating need in Miami since the heating demand was negligible.

Fig 6 shows a boxplot of the RMSE (Eq. 2) and PercErr (Eq. 3) computed over the 50 iterations for each climate. The RMSE is below 1 kWh/m^2 for all climates except FAI and MIA for the heating and cooling models respectively. However, these models still perform very well, as measured by the PercErr. The PercErr is lowest for these two cases since it is calculated considering the magnitude of the values, which is not the case for the RMSE.

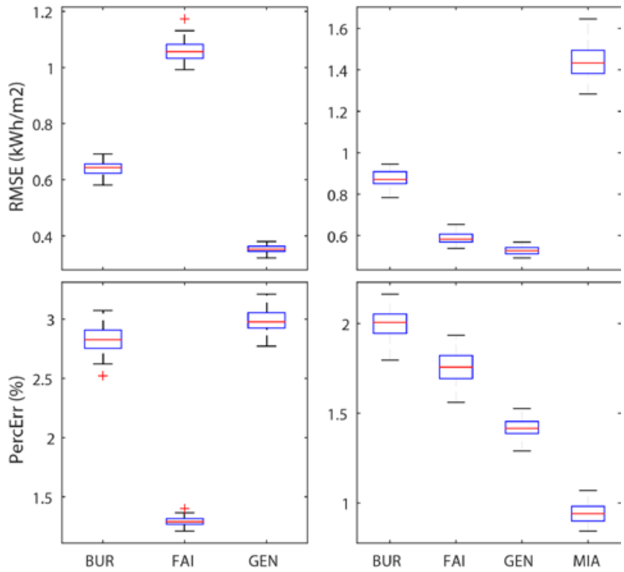

Fig 6: Boxplot of the root mean square error (RMSE) and percentage error (PercErr) over 50 training/testing runs for each climate.

4. Conclusion

The initial work on this project, which is presented in this paper, shows a clear influence of climate on the regression equation. What is interesting to note is that the sets of inputs of the best-performing models in each climate (i.e. the regression inputs) were not very different, if at all. In other words, the considered geometrical or climate inputs that exercise the most influence in one climate also do the same in the others. The only issue with this is that the structures were selected manually to begin with. Whether a future implementation of automated parameter selection yields the same structures remains to be investigated.

The prediction errors of the models are very low; the RMSE is below 1.5 kWh/m^2 for all climates and for both heating and cooling, while the PercErr is below 3% and 2% for the heating and cooling models respectively. These results show this approach to be a promising alternative to full simulation.

An important limitation of the results in this phase of the project is that our training data set (which gives $Y_{predicted}$) and our 'ground truth' simulations (Y_{test}) are both calculated from the same weather file in each climate. To be useful, the predicted equation structures and coefficients should be insensitive to the choice of weather file within a climate. So far, we have assumed that the coefficients of a linear regression equation calculated from any 'typical' weather file will be representative of its climate. This is an assumption we are testing in on-going work. We expect that the estimated

Table 2: Final coefficients for each climate, calculated from the entire dataset. β_1 to β_7 from Eq. 1 correspond to the inputs shown in Fig 4, from top left to bottom right. The constant is β_0.

	Heating				Cooling			
	BUR	**FAI**	**GEN**	**MIA**	**BUR**	**FAI**	**GEN**	**MIA**
β_0	32.70	97.12	16.46	N/A	3.51	-8.82	0.53	178.83
β_1	0.59	-1.36	-0.37	N/A	-0.65	1.12	0.68	1.78
β_2	-6.60	-7.13	-0.96	N/A	11.58	5.24	0.04	7.79
β_3	-8.86	-9.73	-2.83	N/A	-26.85	-16.77	-16.59	-28.96
β_4	86.37	-1.69	-21.94	N/A	-91.96	-16.30	27.95	-119.52
β_5	11.64	84.38	46.42	N/A	110.74	87.92	24.18	217.15
β_6	-0.02	-0.04	-0.02	N/A	0.03	0.02	0.04	0.02
β_7	-0.03	-0.02	0.00	N/A	0.01	0.00	-0.01	-0.06

coefficients will be more robust if we calculate them using more than one weather file per climate.

Any data-based approach has the limitation that it does not account for the underlying physics of the phenomena being studied. In our case, this is further exacerbated by the fact that while a linear model theoretically extends to negative values of output (Y), a negative heating or cooling need is meaningless. Other fitting techniques and model types will be tested in future work.

5. Acknowledgements

This work was supported by the EuroTech Universities Alliance and the SECURE (Synergistic Energy and Comfort through Urban Resource Effectiveness) project funded by the CCEM (Competence Centre for Energy and Mobility). The authors would like to thank Dr Luisa Pastore for her advice and support.

6. References

Asadi, S., Amiri, S. S., & Mottahedi, M. (2014). On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design. *Energy and Buildings*, *85*, 246–255.

ASHRAE. (2007). *International Climate Zone Definitions*. Retrieved from http://www.ashrae.org/File Library/docLib/Public/20081111_cztables.pdf

Dogan, T. (2014). *Archsim Energy Modeling for Grasshopper*.

Foucquier, A., Robert, S., Suard, F., Stéphan, L., & Jay, A. (2013). State of the art in building modelling and energy performances prediction: A review. *Renewable and Sustainable Energy Reviews*, *23*, 272–288.

Hemsath, T. L., & Alagheband Bandhosseini, K. (2015). Sensitivity analysis evaluating basic building geometry's effect on energy use. *Renewable Energy*, *76*, 526–538.

Hygh, J. S., DeCarolis, J. F., Hill, D. B., & Ranji Ranjithan, S. (2012). Multivariate regression as an energy assessment tool in early building design. *Building and Environment*, *57*, 165–175.

Lam, J. C., Wan, K. K. W., Liu, D., & Tsang, C. L. (2010). Multiple regression models for energy use in air-conditioned office buildings in different climates. *Energy Conversion and Management*, *51*(12), 2692–2697.

Nault, E., Rey, E., & Andersen, M. (2015). Towards a predictive model for assessing the passive solar potential of neighborhood designs. Presented at the Building Simulation 2015 (in print), Hyderabad.

Rey, E. (Ed.). (2013). *Green Density*. Presses Polytechniques Universitaires Romandes.

Richard E. Brown, Travis Walter, Laurel N. Dunn, Claudine Y. Custodio, Paul A. Mathew, D. Magnus Cheifetz, … Jessica Knapstein. (2014). *Getting Real with Energy Data: Using the Buildings Performance Database to Support Data-Driven Analyses and Decision-Making* (ACEEE Summer Study on Energy Efficiency in Buildings) (pp. 11–49:11–60). US Department of Energy.

SIA. (2009). *SIA 380/1 Thermal Energy in Buildings*.

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.