Master's Thesis of City Planning

Urban Form Classification under Local Climate Zone Framework for Energy Efficient City Development

- Case Study of Seoul, Korea -

에너지 효율적 개발을 위한 LCZ 기반의 도시형태 분류

- 서울시를 중심으로 -

August 2019

Graduate School of Environmental Studies

Seoul National University

Urban and Regional Planning Major

Parth Bansal

Urban Form Classification under Local Climate Zone Framework for Energy Efficient City Development

- Case Study of Seoul, Korea -

Examiner: Jige Quan

Submitting a master's thesis of city planning
April 2019

Graduate School of Environmental Studies
Seoul National University
Urban and Regional Planning Major
Parth Bansal

Confirming the master's thesis written by
Parth Bansal
June 2019

Chair 김태형 (Seal)
Vice Chair 송영근 (Seal)
Examiner Jige Quan (Seal)

Abstract

With the focus on energy efficient development, the role of urban form as influencer of building energy efficiency has been an area of interest in recent researches. However, the research literature lack common methodology and definition of urban form to measure its efficiency, has been diverse and subjective at best.

Parallelly, Local Climate Zone (LCZ) framework, primarily developed in the urban climatology field to study urban heat island and microclimate, has been gaining traction, both in classification methodology advancements and its applications. The framework defines 17 classes based on urban and natural features with existing research showing that each class has a unique air and surface temperature profile.

Based on the hypothesis that microclimate is indeed the major cause of urban forms impact on building energy consumption, this researcher investigates the relationship between LCZ classes and building electricity and gas consumption. Firstly, the appropriate unit to determine urban form is established with care (to minimize modifiable area unit problem) using LCZ parameter's spatial autocorrelation and distribution curve.

At this point, as has been described by past literature, due to heterogeneous nature, parts of urban area do not fall within any LCZ class. To address it, this research explores the use of machine learning algorithms to identify the closest resembling LCZ class for unidentified areas. Results from algorithms are compared using surface temperature to determine the most suitable classification.

Finally, using this LCZ classification of Seoul and building electricity and gas consumption data (2015-18), the research investigates how energy consumption varies for each LCZ class as compared to sparsely built area which lacks any urban context. For this, two way fixed effect panel data analysis is used so that the effect of both LCZ classes and time can be modeled simultaneously, since the influence

of urban form can vary depending upon broader seasonal conditions. The research indicates that overall, ‘open low rise’ and ‘compact low rise’ are the only urban LCZ classes which are energy efficient. Rest of classes appears to have higher energy consumption (per sq. m of floor space) as compared to sparsely built areas.

Lastly, we compare residential zoning classes with LCZ classes to determine what zoning parameters results in a specific LCZ class. Thus, this research helps urban planners to determine urban form which is energy efficient. So far, the focus on achieving energy efficiency has been limited to improvement in building technology or transportation energy demand, but it is hoped that using this research, it would be possible to identify urban form which is most energy efficient when all the energy consumption aspects are assimilated.

Keywords: Local Climate Zones (LCZ), Building Energy Consumption, Urban Form, Morphology

Student Number: 2017-21289

Table of Contents

List of Tables

List of Figures

I. Introduction

The Urban surface structure has been altered by rapid urbanization in the past century. The guidance to urban form has been limited to a juxtaposition of land-use to keep incompatible uses a far and manage transportation load, while built form regulations focusing on basic health and fire safety. However, the effect of urbanization on climate in urban area has been an area of focus recently, especially the relative increase in urban temperature compared to its rural surroundings, termed as Urban Heat Island (UHI) effect.

Heat wave incidences' intensity and frequency has seen increase due to climate change and number of researches, including Tan et al. (2010) suggesting that UHI amplifies heat wave's impact by further increasing temperature and exposing more population (due to higher density of urban areas). However, UHI studies in the past have been fragmented due to multiple issues, including temperature measurement methods, frequency and accuracy of measurements and methodologies for urban-rural temperature comparison. However, the most prominent issue has been the different meaning of 'urban' since an urban area can have diverse forms within itself and the definition of each typology of form can vary between studies (Oke, 2006). Thus a meta-analysis of UHI reported across literature has been difficult which has hindered development of guidelines for urban design which can minimize UHI and its impact on citizens.

Hence, to facilitate urban climate studies by generating standardized urban form database, Local Climate Zone (LCZ) framework has been developed by Stewart & Oke (2012). The objective of the framework is to develop urban and natural form classification (for UHI studies) which is universal and agnostic to local conditions and differences in perceived definitions of urban form terminologies. It has been deployed over by many cities (Zheng et al., 2018) and different methods for implementing classification have been developed. However urban area being inherently heterogeneous, have been difficult to classify. Users often find that part of their study area do not concur with all the parameters for any of LCZ classes.

Increase in heat wave intensity and frequency has been a major concern during summer season in Korea. According to BBC (2018) substantial financial allocation to minimize its impact has been made, but measures by local bodies have been of short term nature. LCZ framework can help urban planners to characterize the urban form of city and identify areas susceptible to extreme temperature. Thus, the short term interventions to reduce UHI (and thus extremity of a heat wave) could be designed according to urban morphology of the area, while in the long term, urban form of vulnerable area can be molded to UHI minimizing LCZ's form.

Additionally, LCZ framework's application to aspects which are influenced by UHI and micro climate has been proposed by a number of scholars. Quan et al. (2017) have proposed that LCZ can be deployed as an urban form identifier to

measure influence of urban context on building energy consumption (BEC). However, while the relationship between BEC and urban climate has been theorized, numerical relationship between BEC and LCZ has not been explored. Due to change in area's temperature by UHI, building energy consumption could vary, with Ratti et al. (2005) suggesting that urban climate may influence up to 20% of variance observed in BEC for same size and use of building.

1.1 Research Question

Thus, this research aims to develop a methodology for complete LCZ classification of the city and then study how different LCZ class modulates BEC in residential buildings of Seoul. For this, the main objectives are:

1) What is the LCZ classification of Seoul?
2) How is residential building energy consumption dependent on the LCZ?

1.2 Research Range

The research uses Seoul city as the case study, with 2015 as base year for data collection. The lowest unit of data points is individual buildings / road networks and natural bodies whose properties are aggregated to LCZ dimension. LCZ parameters have been taken from Stewart & Oke (2012) and GIS methodology predominantly follows as prescribed by Zheng et al. (2018). For classification, i.e. objective 1, machine learning algorithms for classification, modified K-Means and Random Forest are used and validated using Land Surface Temperature (LST) derived from Landsat 8. For investigating relation between LCZ and residential building energy consumption, i.e. objective 2, fixed effect panel data analysis is used to determine impact of LCZ based urban form on residential energy consumption in terms of difference in electricity and gas consumption as compared to non-urbanized areas.

II. Theoretical Background and Literature Review

This chapter investigates into the reasoning for utilization of LCZ framework and its relationship with building energy consumption. Firstly, the theoretical basis of urban heat island is described, and various methods to measure and characterized it are explored. Then, the relative strengths and weaknesses of LCZ framework are described. Secondly, approaches to LCZ classification are explored and prominent methods for each approach are described. Then the issue of LCZ grid size and location is explored and methodology adopted in this study is described. Thirdly, current literature on urban form and BEC is summarized and strength of LCZ as a tool for measurement of urban form for BEC studies is discussed. Lastly, panel data analysis and its variant used to explore relationship between LCZ and BEC is discussed.

2.1 Urban Heat Island (UHI)

According to Kim (1992), the surface energy balance equation can be stated as:

$$R_{bal} = SRI - OLR\ - H_{sens} - LE$$

Where R_{bal} is radiation balance, SRI is short radiation intake (incoming solar energy), OLR is outgoing longwave radiation, H_{sens} is sensible heat flux (energy exchange between surface and atmosphere), LE is latent heat of evaporation.

According to Oke (1982, 2010) urban form can modify the natural surface energy balance and thus alter the climate within the urban area, by:

1) Altering the surface geometry by increasing the paved surface area and generating enclosed spaces between buildings. This results in greater solar absorption and reduced heat loss at night due to smaller sky view.
2) Altering thermal property of the material, as building materials have higher capacity for latent heat.
3) Loss of vegetation resulting in loss of evapotranspiration.

This phenomenon, by which cities exhibit higher temperature than the surrounding countryside, is termed as UHI. It has been shown to be directly proportional to city size (Oke, 1973), street design (Oke, 1988), amount and relative location of green and blue space (Gunawardena et al., 2017). This difference can exceed 10°C and has been documented over multiple cities including Seoul (Kim & Baik, 2002).

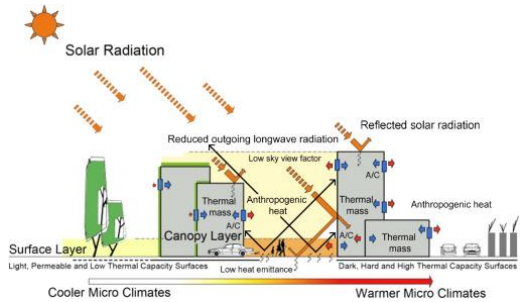


Figure 2-1) Process of UHI. Adopted from (Soltani & Sharifi, 2017)

2.1.1 Local Climate Zones (LCZ)

The conventional approach for measuring UHI has been the comparison of temperature data from urban and rural area i.e.

$$\Delta T = T_U - T_R \text{ (Oke, 1973)}$$

Where, ΔT is UHI, T_U and T_R are temperature measurement at urban and rural site respectively. However, the efficacy of this methodology has been disputed by Stewart & Oke (2012) based upon:

1) Terms urban and rural are subjective in nature. The nature of site varies drastically and an area classified as urban in one context might be classified as rural in another. Thus, it is difficult to construct a broader database of UHI effect based upon the existing literature. Figure 2-2 shows some of the examples of sites delineated as 'urban' in literature.
2) Even when a single region is studied, there is drastically different morphologies within urban and rural areas, all of which have different UHI related properties.

Based upon above issues, Stewart & Oke (2012) have proposed urban form classification called Local Climate Zones (LCZ) containing 17 classes; 10 based on built type and 7 based upon land cover. Figure 2-3 (A & B) contains description of LCZ classes and their defining parameters, respectively.

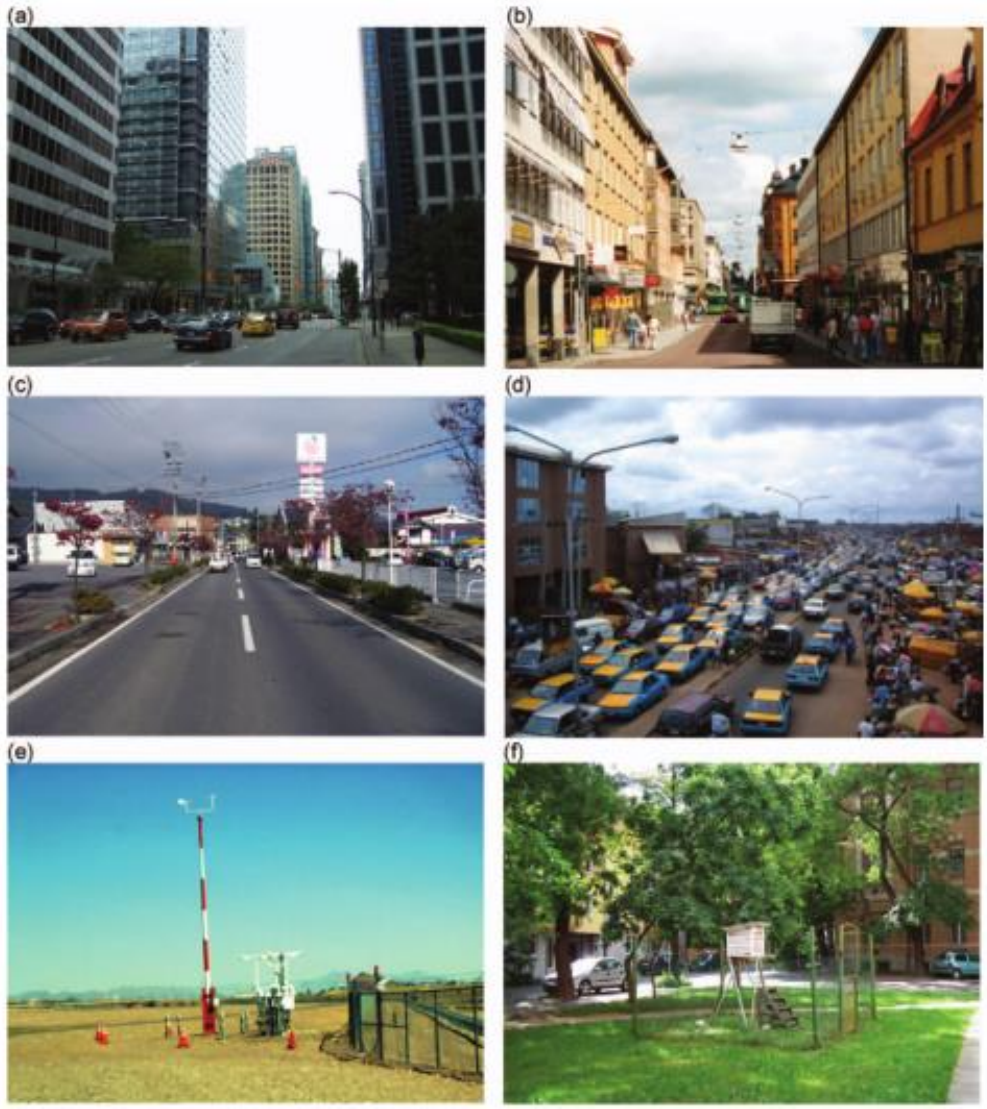


Figure 2-2) Urban settings from different urban heat island literature. Figure compiled by Stewart & Oke (2012). Different meaning of ‘urban’ makes researches incomparable to draw broader conclusions.

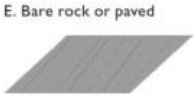
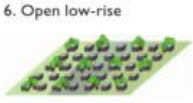
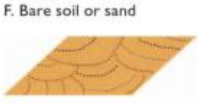
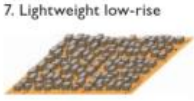
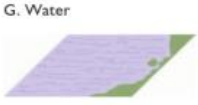
Built types	Definition	Land cover types	Definition
1. Compact high-rise 	Dense mix of tall buildings to tens of stories. Few or no trees. Land cover mostly paved. Concrete, steel, stone, and glass construction materials.	A. Dense trees 	Heavily wooded landscape of deciduous and/or evergreen trees. Land cover mostly pervious (low plants). Zone function is natural forest, tree cultivation, or urban park.
2. Compact midrise 	Dense mix of midrise buildings (3–9 stories). Few or no trees. Land cover mostly paved. Stone, brick, tile, and concrete construction materials.	B. Scattered trees 	Lightly wooded landscape of deciduous and/or evergreen trees. Land cover mostly pervious (low plants). Zone function is natural forest, tree cultivation, or urban park.
3. Compact low-rise 	Dense mix of low-rise buildings (1–3 stories). Few or no trees. Land cover mostly paved. Stone, brick, tile, and concrete construction materials.	C. Bush, scrub 	Open arrangement of bushes, shrubs, and short, woody trees. Land cover mostly pervious (bare soil or sand). Zone function is natural scrubland or agriculture.
4. Open high-rise 	Open arrangement of tall buildings to tens of stories. Abundance of pervious land cover (low plants, scattered trees). Concrete, steel, stone, and glass construction materials.	D. Low plants 	Featureless landscape of grass or herbaceous plants/crops. Few or no trees. Zone function is natural grassland, agriculture, or urban park.
5. Open midrise 	Open arrangement of midrise buildings (3–9 stories). Abundance of pervious land cover (low plants, scattered trees). Concrete, steel, stone, and glass construction materials.	E. Bare rock or paved 	Featureless landscape of rock or paved cover. Few or no trees or plants. Zone function is natural desert (rock) or urban transportation.
6. Open low-rise 	Open arrangement of low-rise buildings (1–3 stories). Abundance of pervious land cover (low plants, scattered trees). Wood, brick, stone, tile, and concrete construction materials.	F. Bare soil or sand 	Featureless landscape of soil or sand cover. Few or no trees or plants. Zone function is natural desert or agriculture.
7. Lightweight low-rise 	Dense mix of single-story buildings. Few or no trees. Land cover mostly hard-packed. Lightweight construction materials (e.g., wood, thatch, corrugated metal).	G. Water 	Large, open water bodies such as seas and lakes, or small bodies such as rivers, reservoirs, and lagoons.
8. Large low-rise 	Open arrangement of large low-rise buildings (1–3 stories). Few or no trees. Land cover mostly paved. Steel, concrete, metal, and stone construction materials.	**VARIABLE LAND COVER PROPERTIES**	
		Variable or ephemeral land cover properties that change significantly with synoptic weather patterns, agricultural practices, and/or seasonal cycles.	
9. Sparsely built 	Sparse arrangement of small or medium-sized buildings in a natural setting. Abundance of pervious land cover (low plants, scattered trees).	*b. bare trees*	Leafless deciduous trees (e.g., winter). Increased sky view factor. Reduced albedo.
		s. snow cover	Snow cover >10 cm in depth. Low admittance. High albedo.
10. Heavy industry 	Low-rise and midrise industrial structures (towers, tanks, stacks). Few or no trees. Land cover mostly paved or hard-packed. Metal, steel, and concrete construction materials.	*d. dry ground*	Parched soil. Low admittance. Large Bowen ratio. Increased albedo.
		w. wet ground	Waterlogged soil. High admittance. Small Bowen ratio. Reduced albedo.

Figure 2-3(A) Local Climate Zone (LCZ) classes defined by Stewart & Oke (2012). The framework defined 10 urban classes and 7 natural classes.

Local climate zone (LCZ)	Sky view factor[a]	Aspect ratio[b]	Building surface fraction[c]	Impervious surface fraction[d]	Pervious surface fraction[e]	Height of roughness elements[f]	Terrain roughness class[g]
LCZ 1	0.2–0.4	> 2	40–60	40–60	< 10	> 25	8
Compact high-rise							
LCZ 2	0.3–0.6	0.75–2	40–70	30–50	< 20	10–25	6–7
Compact midrise							
LCZ 3	0.2–0.6	0.75–1.5	40–70	20–50	< 30	3–10	6
Compact low-rise							
LCZ 4	0.5–0.7	0.75–1.25	20–40	30–40	30–40	>25	7–8
Open high-rise							
LCZ 5	0.5–0.8	0.3–0.75	20–40	30–50	20–40	10–25	5–6
Open midrise							
LCZ 6	0.6–0.9	0.3–0.75	20–40	20–50	30–60	3–10	5–6
Open low-rise							
LCZ 7	0.2–0.5	1–2	60–90	< 20	<30	2–4	4–5
Lightweight low-rise							
LCZ 8	>0.7	0.1–0.3	30–50	40–50	<20	3–10	5
Large low-rise							
LCZ 9	> 0.8	0.1–0.25	10–20	< 20	60–80	3–10	5–6
Sparsely built							
LCZ 10	0.6–0.9	0.2–0.5	20–30	20–40	40–50	5–15	5–6
Heavy industry							
LCZ A	<0.4	>1	<10	<10	>90	3–30	8
Dense trees							
LCZ B	0.5–0.8	0.25–0.75	<10	<10	>90	3–15	5–6
Scattered trees							
LCZ C	0.7–0.9	0.25–1.0	<10	<10	>90	<2	4–5
Bush, scrub							
LCZ D	>0.9	<0.1	<10	<10	>90	<1	3–4
Low plants							
LCZ E	>0.9	<0.1	<10	>90	<10	<0.25	1–2
Bare rock or paved							
LCZ F	>0.9	<0.1	<10	<10	>90	< 0.25	1–2
Bare soil or sand							
LCZ G	>0.9	<0.1	<10	<10	>90	–	1
Water							

[a] Ratio of the amount of sky hemisphere visible from ground level to that of an unobstructed hemisphere

[b] Mean height-to-width ratio of street canyons (LCZs 1–7), building spacing (LCZs 8–10), and tree spacing (LCZs A–G)

[c] Ratio of building plan area to total plan area (%)

[d] Ratio of impervious plan area (paved, rock) to total plan area (%)

[e] Ratio of pervious plan area (bare soil, vegetation, water) to total plan area (%)

[f] Geometric average of building heights (LCZs 1–10) and tree/plant heights (LCZs A–F) (m)

[g] Davenport et al.'s (2000) classification of effective terrain roughness (z_0) for city and country landscapes. See Table 5 for class descriptions

Figure 2-3(B) Parameters for each LCZ class as defined by Stewart & Oke (2012). Some parameters have different meaning or way of measurement depending upon the LCZ class.

2.1.2 LCZ parameters

LCZ classification is based upon 7 criteria to measure urban form for UHI and microclimate investigations. However, most of subsequent literature uses 6 criteria classification and drops '*terrain roughness classes'* (Davenport et al., 2000) due to its subjective nature and lack of a common methodology. These 6 criteria and how they relate with UHI and micro climate are discussed below:

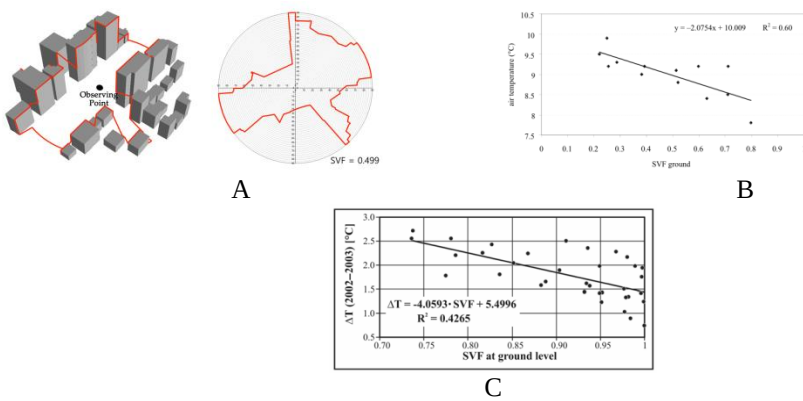
A) Sky View Factor (SVF)

Sky view factor is the geometric ratio of sky visible from a specified viewpoint, i.e. it indicates the degree of open sky space in the hemisphere (Oke, 1981). It varies from 0 to 1 with value approaching 1 for perfect flat terrain and decreases with the increase in obstruction. It is given by

$$\psi_s = 1 - 2\psi_w$$

where, $\psi_w = 0.5(sin^2\theta + cos\theta - 1)(\cos\theta)^{-1}$ *and* $\theta = tan^{-1}(H/0.5W)$
(Oke, 1981)

Ψ_S is SVF, Ψ_W is wall view factor, *H* and *W* are height and width of obstructing feature. Generally, decrease in air and surface temperature is observed as SVF approaches one (Svensson, 2004) (Unger, 2009) (Figure 2-4).



A B

C

Figure 2-4) A) Estimation of Sky View Factor. Adopted from Park et al (2017). B and C) Inverse relationship between SVF and air and surface temperature (Svensson, 2004) (Unger, 2009)

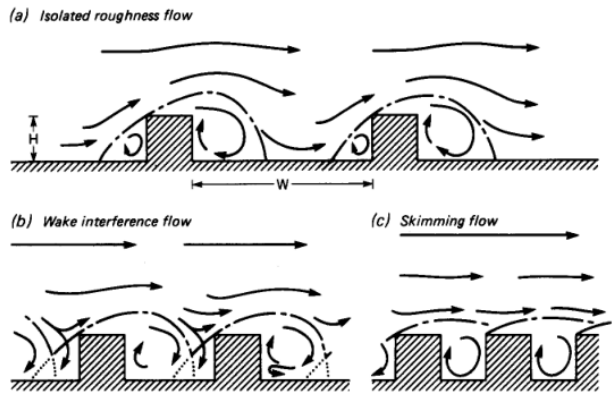
B) Aspect Ratio (AR)

Aspect ratio is defined as the ratio of height to width of street canyon and is highly correlated with SVF under symmetrical street conditions. However, unlike SVF, AR not only influences the incoming and outgoing radiation, but also wind flow regime which controls how fast heat is dissipated back into the atmosphere beyond urban canyon layer. As AR increases, wind flow within the canyon initially

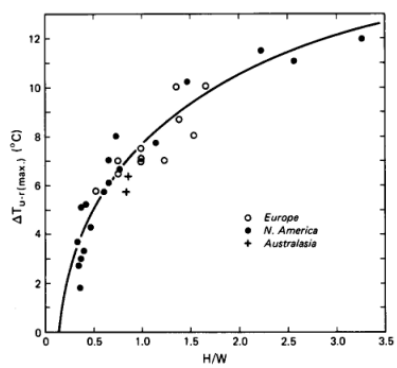
becomes rough (wake interference) but then smoothen out due to skimming of air above the canyon (Oke, 1988) (Figure 2-5 A). Asymmetrical AR is calculated as:

$$AR = \frac{\sum_{i=1}^{n} \frac{BH_i}{SW_i}}{n}$$ (Mohajeri et al., 2016)

Where *BH* is building height and SW is street width. Generally, it is observed that as AR increases, UHI increases (Oke, 1988)(Figure 2-5 B).



A



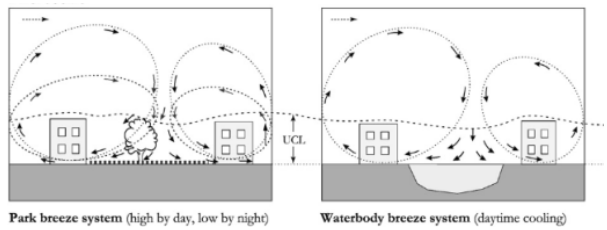
B

Figure 2-5) A) Change in air flow due to canyon's AR. B) Relationship between UHI and AR (Oke, 1988)

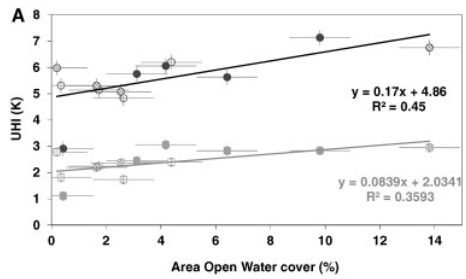
C) Pervious Surface Fraction (PSF)

Pervious surface fraction is defined as percentage of area occupied by the pervious bodies (water, bare soil, trees) compared to the total area. Generally, pervious area have lower specific heat value and thus heat up slowly as compared to urban built up, hence causing cooling of the surround areas (Gunawardena et al., 2017) (Figure 2-6 A). Thus, an increase in the pervious area reduces the UHI magnitude (Nastrana et al., 2018) (Figure 2-6 B). However, type, size and relative

juxtaposition with other features has been shown to dictate if and how much UHI reduction is achieved (Steeneveld et al., 2014).



A



B

Figure 2-6) A) Process of air cooling due to presence of pervious surface (Gunawardena et al., 2017). B) Relationship between amount of open water area and UHI indicating that the association is not always same as theorized (Steeneveld et al., 2014) .

D) Building and Impervious Surface Fraction (BSF / ISF)

Building / impervious surface fraction are defined as the percentage of area occupied by building / impervious features compared to the total area. It is calculated as:

$$BSF = \frac{\sum_{i=1}^{n} BF_i}{Grid\ Area}$$

$$ISF = 100 - BSF - PSF \quad \text{(Zheng et al., 2018)}$$

Where *BF* is building footprint area. Increase in combination of BSF and ISF percentage is generally associated with an increase in temperature (Yaun & Bauer, 2007) (Figure 2-7).

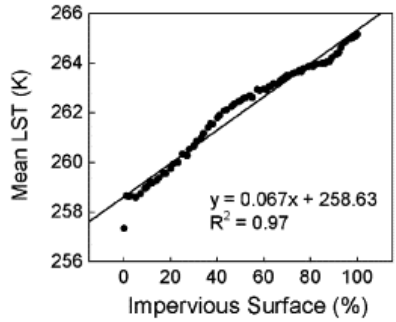


Figure 2-7) Relationship between impervious surface and land surface temperature (Yaun & Bauer, 2007).

E) Height of Roughness Element

Height of the roughness element has different definition for urban built type and land cover type classes. For built type classes, it is defined as geometric mean of building heights. However in recent literature, researchers prefer to use building footprint weight mean of height defined as:

$$BH = \frac{\sum_{i=1}^{n} BF_i * BH_i}{\sum_{i=1}^{n} BF_i}$$ (Zheng et al., 2018)

Where, *BF* is building footprint area and *BH* is building height. On the other hand, for land cover classes, it is defined as the mean height of tree/plant.

2.1.3 Critic of LCZ framework

However, the reliance on fixed and limited parameters has often been criticized. For example, mountains in Seoul are a major modulator of local climate. However, in its current status, the framework is unable to capture its influence. Additional critic come from its agnostic nature towards material in urban LCZs, since the albedo of construction material has large variations. Also, since LCZs do not have any parameter which can account for street pattern, the directional aspect is not captured in the framework.

Another problem related with LCZ is the problem associated with any study involving grids. Since LCZ are not continuous but categorical, meaning that a grid is independent of its surrounding. This is obviously not the case in the real cities, where urban form is connected and it is reasonable to assume that microclimate within a grid is influenced by what LCZ surrounds it. See section 2.3 for further discussion on this and possible solution investigated in this research.

2.1.4 LCZ classification

LCZ framework aims to develop relatively homogenous morphological classification which is independent of local conventionality (Hammerberg et al.,

2018). However, earlier attempts at LCZ classification by Emmanuel & Kruger (2012) & Leconte et al. (2015) lacked procedural consistency and various methodological approaches have been proposed. These can broadly be classified into:

A) Supervised pixel-based classification

Currently the popular and promoted under World Urban Database and Access Portal Tools (WUDAPT) (Bechtel et al., 2015) supervised pixel based classification uses satellite imagery from Landsat-8 (or comparable sensor) which is used for classification based on user defined region of interests (ROIs). The user defines ROI using high resolution Google Earth imagery and then acquires moderate resolution (30 to 90M) Landsat 8 data. These ROIs are used to train Random Forest classifier, which then classifies Landsat 8 imagery.

Pixel based classification suffers from a number of critical flaws:

1) ROI training is still subjective to the user's interpretation of the LCZ class.
2) The coarse resolution of Landsat means that small features are not identifiable. Additionally, spectral signature of features smaller than the pixel size mixes together, creating distorted output. Although mixed pixel classification techniques exist, they require hyper-spectral imagery which as of now, are limited and not freely available.

B) GIS based classification

With above flaws and the availability of high quality vector data for cities, parallel research has also begun to explore GIS based LCZ classification (Geletic & Lehnert, 2016), (Zheng et al., 2018). Using GIS, the study area is divided into grids and LCZ parameters are calculated for each grid. LCZ class is then assigned based upon these parameters. In case of grids failing to satisfy criteria for any of LCZ classes, various approaches have been deployed:

1) Zheng et al. (2018) used land use map to assign LCZ class based upon their subjective interpretation.
2) Geletic & Lehnert (2016) tried to implement Random Forest classification suggested under WUDAP but has found poor agreement (51%) between satellite and GIS based classification.

A common issue of fuzziness i.e. presence of mixture of LCZ (Bechtel et al., 2019) is still unresolved. Currently no methodology exists to identify what mix exists and how to classify them.

2.2 Classification Algorithms

Classification in statistics and machine learning is defined as a process of identifying categories of observations or detecting which observations are similar to each other. Broadly, two types of classification methodology: supervised and unsupervised exist. Supervised learning is used when data with known classes (training data) is available and the objective is to derive boundaries for each class. These boundaries are then used to assign classes to unknown observations. Random Forest classification used under WUDAP and Naïve Bayes used by Hammerberg et al. (2018) are the example of supervised classification algorithms.

On the other hand, unsupervised learning is used when all the observations are unlabeled, i.e. there classes are unknown. The algorithm tries to find which observations are similar to each other and assigns them a class, which is then investigated by the user to determine what the class represents. K-means is an example of unsupervised classification. Semi-supervised clustering can be thought of as unsupervised learning when certain aspects of the observation's class and its relation to other observations are known, but are not enough to generate classification boundaries. COP K-Means is an example of semi-supervised clustering.

In this study, we deploy: 1) Supervised clustering algorithm: Random Forest. 2) Semi-supervised clustering algorithm: Constrained K-Means with Background Knowledge (COP K-Means)[1] (Wagstaff et al., 2001). We used these algorithms because:

1) Under WUDAPT, Random Forest has been the choice of algorithm due to its speed, ability to work with different units of measurement, and simplistic implementation in GIS and remote sensing software.
2) COP K-Means provides ability to integrate the most popular centroid based classification method (K-Means) and insufficient classification samples.

By constraining parameters in unsupervised classification, we are able to generated better classification results using limited information available. Next we discuss theoretical background of classification algorithms discussed above.

2.2.1 Random Forest

As discussed previously, Random Forest has been most popular algorithm in remote sensing based LCZ classification. Random Forest consists of group of tree classifiers generated from random samples that are selected independently. Each tree makes a unit vote for determine the most likely class (Breiman, 1999). User

[1] Another algorithm, PCK-Means (Pairwise Constrained K-Means) (Basu et al., 2002) has been developed further using COP-Kmeans. Some literature user term PCK-Means even when additional inputs for PCK-Means are not used and thus algorithm is effectively same as COP K-Means in such scenario.

provides two parameters: a) number of variables in each node and b) number of trees in the forest.

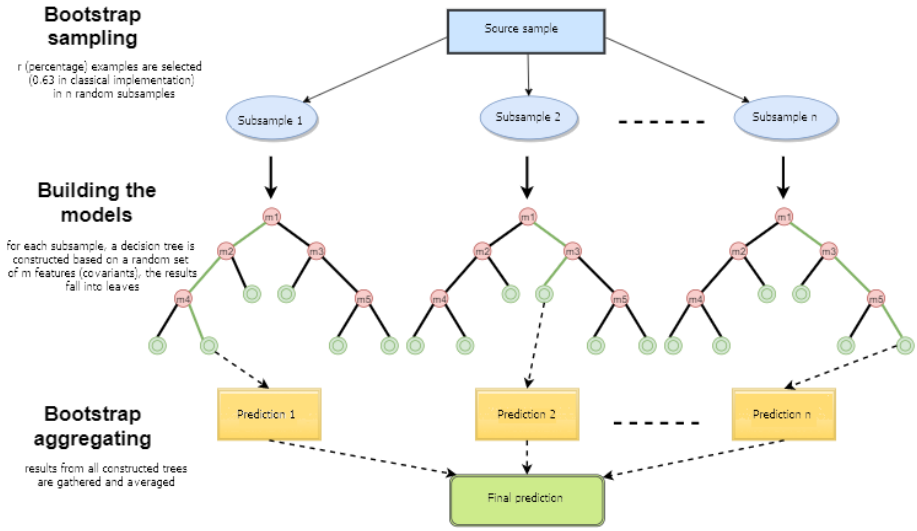


Figure 2-8) Process of Random Forest classification. (Source: https://www.mql5.com/en/articles/3856)

2.2.2 K-Means

Consider a dataset X = (x_1,x_2,…x_n) with V = (v_1,v_2,…v_c) cluster center in p dimensions (LCZ parameters), where n is the number of observations and k is the number of clusters. A cluster is described by its member observation and its center. The centers are selected such that the sum of the distance between center and all the member observation is minimized and the distanced between centers are maximized.

K-Means clustering (MacQueen, 1967) is an unsupervised classification algorithm which clusters the observations into K number of groups based upon the observation's attributes in p dimensions. The user specifies K along with optional parameters of initial centroid, convergence margin and distance method. In step 1, each observation is assigned to the closest centroid; in step 2, cluster's centroid is changed to mean of observations assigned to the cluster. In mathematical terms, the algorithm minimized the function:

$$K = \sum_{i=1}^{c}\sum_{j=1}^{n} D_{ij}^2$$

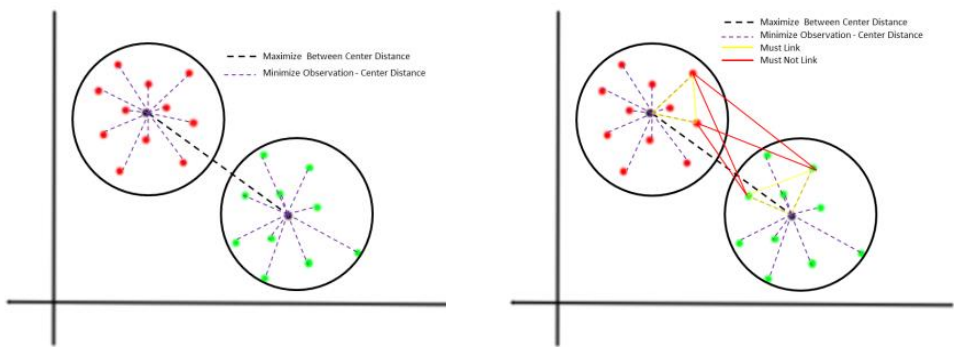
$$D_{ij} = ||x_{ij} - v_i||^2 (\text{Arthur \& Vassilvitskii, 2007})$$

Such that $1 \leq i \leq c$. Where n_i is the number of observations in i^{th} cluster and the centroid in step two is updated as:

$$v_i = \sum_{i=1}^{n_i} \frac{x_{ij}}{n_i}$$ (Cebeci & Yildiz, 2015)

Such that $1 \leq j \leq n_i$. The algorithm repeats these steps until the difference between new and old centroid is less than the specified threshold value.

Constrained K-Means clustering (COP-KMeans) (Wagstaff et al., 2001) is a modification of K-Means used when certain background knowledge about centroid and categories of some observation are known. Specifically, the algorithm allows for specifying: 'must link', i.e. observation that must belong to same categories and 'cannot link', i.e. observation which must not belong to the same categories. An observation after initial assignment to cluster will not change its assignment if its violets specified constraints and the algorithm checks for next best non-constrain violating cluster. If no such cluster is found, the observations are put into additional empty partition. Additionally by specifying must-link between observation and centroid, we limit the movement of centroid within the LCZ parameter's range.



A B

Figure 2-9) A) K-Means. B) Constrained K-Means.

2.3 Boundary Issues

Unlike remote sensing based methods, where grid location and size are predetermined based upon data, GIS based method offer flexibility in selecting the location and size depending upon nature of study and study area itself. Using Hong Kong city as a case study Zheng et al. (2018) have proposed spatial autocorrelation & Kolmogorov-Smirnov test based method for determining grid size and location. Before this, grid size has often been left to user discretion since the Stewart & Oke (2012) have stated LCZ size to be flexible of 'hundreds of meters to kilometers' and grid location rarely received any attention.

2.3.1 Grid size

To determine the appropriate grid size we use spatial autocorrelation, which refers to the tendency of a variable (especially those with spatial context) to be

more similar to another observation in its vicinity than expected for random pair of observation (Legendre, 1993). Spatial autocorrelation can be investigated by comparing the distance and square of difference between observations using semivariogram (ArcGIS, 2017). In case of spatial autocorrelation, semivariance rises as distance increases and at range (a) reaches the maximum level (Sill γ(h)) (Curran, 1988). This range is a strong indicator of maximum distance till which spatial autocorrelation exists and pairs beyond the range are considered uncorrelated (ArcGIS, 2017).

Additionally, to confirm the finding from semivariogram, Moran's I (ArcGIS, 2017b) which evaluates the pattern as clustered / dispersed / random for given feature's location and value, can be used. Moran's I is calculated as:

$$I = \frac{n}{S_0}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}z_i z_j}{\sum_{i=1}^{n} z_i^2} \text{ (Zhang et al., 2008)}$$

$$S_0 = \sum_{i=j}^{n}\sum_{j=1}^{n} w_{i,j}$$

Where, z_i is the difference between a feature and its mean, $w_{i,j}$ is the spatial weight between i and j, S_0 is the sum of all spatial weights and n is the number of features.

$$Z_I = \frac{I - E[I]}{\sqrt{V[I]}}$$

$$E[I] = \frac{-1}{n-1}, V[I] = E[I^2] - E[I]^2 \text{ (ArcGIS, 2017b)}$$

Z_I is calculated for I at various maximum correlation distance, with max Z_i depicting range till which spatial autocorrelation exist.

2.3.2 Grid location

Grid location determines the distribution of LCZ parameters and thus can influence the classification result. Kolmogorov-Smirnov test's application in climate science to test empirical distribution of sampled datasets has been studied by Knutson et al. (1998) & Orlowsky (2008). The method calls for identifying parameters whose distribution changes significantly under alternative grid location and then choosing grid location in which this identified parameter has least variance. In LCZ classification, Zheng et al. (2018) recommends 9 possible grid geo-location including base and 8 possible shifts in each direction, shown in figure 2-10.

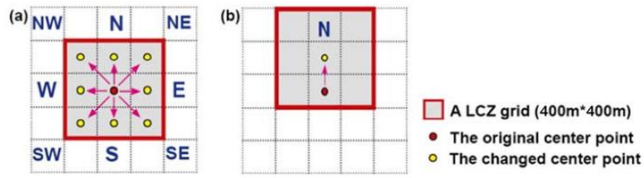


Figure 2-10) Variations in grid location. Figure adopted from Zheng et al. (2018)

2.4 Urban Form and Building Energy Consumption (BEC)

With the increasing focus on reducing our carbon footprint, reduction of energy consumption has been emphasized area of research. According to Song & Choi (2012), buildings consume 25.3% of total energy in Korea. For Seoul and other highly urbanized area, this figure is expected to be much higher, such as in the case of Hong Kong, where the value stood at 64% in 2017 (EMSD, 2018).

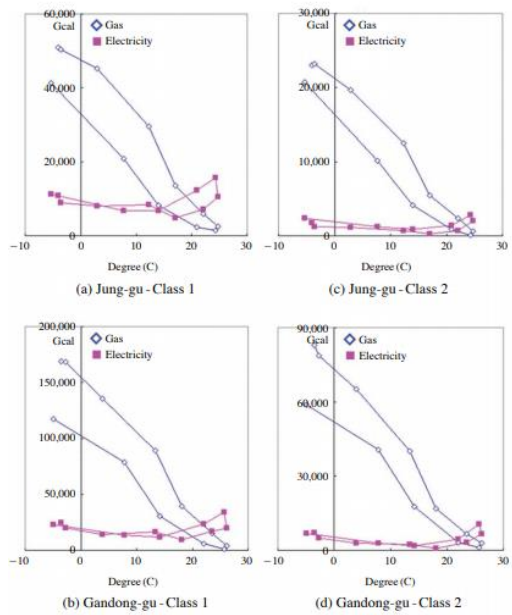


Figure 2-11) Energy pattern consumption influenced by overall morphology and land use in Seoul. (Lee et al., 2014)

When trying to explain building energy consumption (BEC) based upon building design / heating, ventilation and air conditioning system efficiency / occupant behavior, Baker & Steemer (2000) found that these factors explain up to 10 fold variations in BEC. However, if the function and size of building in held constant, BEC can vary up to 20 folds. Ratti et al. (2005) have hypothesized that the remaining variance can be explained by urban form by the pathway of mutual shading and surface area exposed to the environment, however experiments suggests that these factors explain only a small portion of the remaining variation in BEC. Other scholars including Priyadarsini (2009) & Malys et al. (2015) have shown that urban heat island (UHI) and microclimate as functions of urban form, also impacts BEC.

This phenomenon was also observed by Lee et al. (2014) in their limited study of BEC in Seoul, where two districts with different morphology [Jung-gu (historic center) and Gandong-gu (typical suburb)] had a different pattern of building energy consumption (electricity and gas). The difference in energy consumption pattern

was dependent on the district itself and land use. Figure 2-11 shows this pattern for class 1 and 2 land use (detached and multifamily housing respectively).

Youngsoo & Saehoon, (2018) used structural equation model to explore the complex relation between land use, urban design, architecture and thermal efficiency of buildings. They found that urban form characteristics (location, road width, lot area, etc.) and land use characteristics (FAR, mixture and ownership) not only have a direct impact on building's thermal performance but also indirect effects via architectural elements of the building.

2.4.1 LCZ as measure of urban form for BEC studies

A number of measurements, indicators and indices have been developed in the literature to measure or quantify urban form. However, as Lynch & Rodwin (1958) have said, the measurement of urban form should be in relation to the objective of such measurement. As discussed above, in case of BEC studies, the combination of variables deployed to define urban form varies for each study. The choice has largely been driven by data availability, 'what fits the model best' approach and judgment of the researcher.

Thus, there is need for an urban form measurement framework which capture parameters that affects BEC and is also universal to ensure comparability of studies. Thus, Quan et al. (2017) have explored the possibility of linking energy resilience with LCZ. The argument for LCZ as measure of urban form for BEC studies can be thought as:

A) Urban form impact microclimate**:** As discussed in section 2.1.2, theoretical, measurement and simulation studies including Middel et al. (2014) have demonstrated that urban form exerts influence on its surrounding environment's climatic conditions.
B) Microclimate impacts BEC**:** As theorized by Ratti et al. (2005) and empirically shown by the number of studies including Li et al. (2014), outdoor climate has significant impact on the building's energy consumption pattern. In fact, some researchers have been able to reversely identify climatic zones based upon the electricity consumption pattern of the area (Vu & Parker, Preprint).
C) LCZ as a *suitable* measure of urban form for its microclimate**:** Using simulation and empirical measurement (figure 2-12 A & B respectively), it has been shown that each LCZ class has its own unique and distinct temperature profile. Additionally, LCZ framework provides stability to urban form variable to ensure cross study comparability and is relatively straightforward to estimate.

Thus, using the LCZ classification described in the previous section, this study investigates the relation between LCZ and residential building energy consumption.

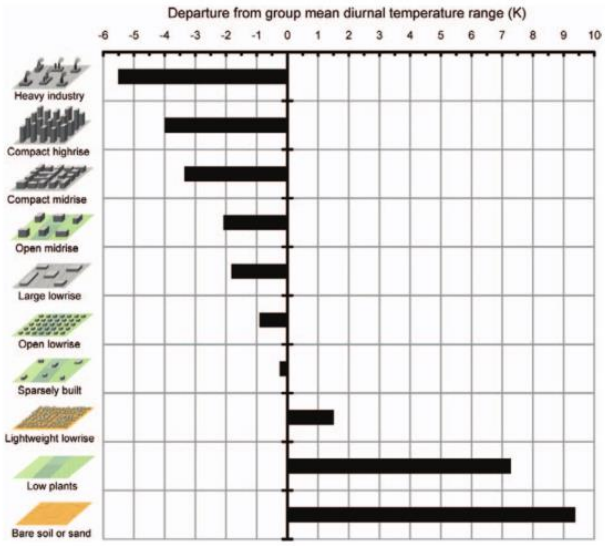


Figure 2-12 A) Difference in temperature profile of LCZ classes as estimated by Town Energy Balanced Model (TBDM) (Stewart & Oke, 2012)

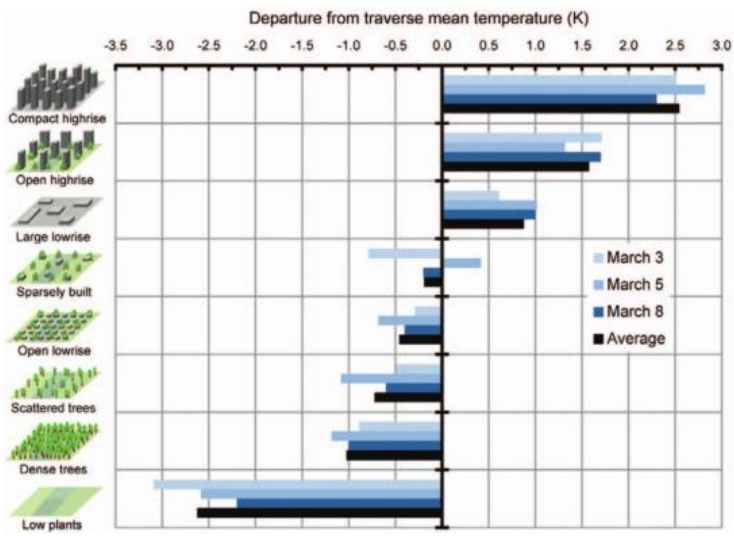


Figure 2-12 B) Difference in temperature profile in Vancouver City (2010) recorded using automobile traverser. (Stewart & Oke, 2012)

2.5 Statistical Analysis

The data under study exist as observation for each grid (with membership of one LCZ class) and for each month, i.e. L categories with L_n grids, measured on t instances. Thus, firstly, to visualize if any of these factor (LCZ and time) has impact on energy consumption, time series analysis is used. Then to model these factors, fixed effect panel analysis is used where other socio-economic variables are included as control variable so that 'pure impact' of LCZ and time on building energy consumption can be extracted.

2.5.1 Time series

Since multi observation data (for each period of interest) can be hard to decipher, a simple moving average (2 Level) analysis can drastically simplify visualization, i.e.:

$$X_T' = |\mathrm{X}_T - \overline{\mathrm{X}}_{T-1}|$$

$$X_L = \sum_{i=1}^{n} \frac{X'_{T_n}}{n}$$

Where,
X_T = Temperature at time T
n = number of observation/entity/time unit

2.5.2 Panel data analysis

Panel data[2] refers to data structure in which observations spans over time and entity; where an entity can be a person, political entity or any defined geographic boundary. Panel data analysis allows for controlling unmeasured source of individual heterogeneity that varies for each individual, thus minimizing impact of omitted variable bias (Stock & Watson, 2007). Figure 2-13 shows how the regression model changes if intercept is allowed to vary for each firm as compared to results of pooled OLS regression.

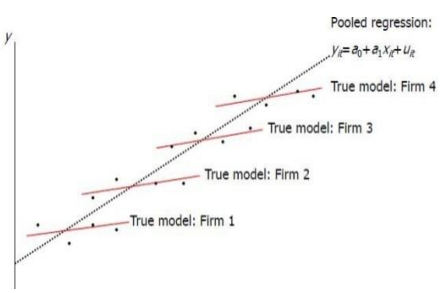


Figure 2-13) Pooled OLS vs. panel data analysis.
(Source:https://towardsdatascience.com/understanding-panel-data-regression-c24cd6c5151e)

Panel data analysis, broadly, can be divided into 3 types:

[2] Also referred as cross sections over time / longitudinal data.

- One Way Fixed Effect

The one way fixed effect is used when only impact of either entity or time is to be modeled. It is expressed as:

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}$$

However, in application, it can also be modeled using n-1 binary variables as:

$$Y_{it} = \beta_0 + \beta_1 X_{1,it} + \gamma_2 E_2 + \ldots + \gamma_n E_n + u_{it}$$

Where:
Y_{it} = Dependent variable for entity i and time t
X_{it} = Independent variable for entity i and time t
α_i = Intercept
E_n = Binary variable (either for time or entity)
γ_n = Coefficient for the binary variables
u_{it} = Between entity error term

- Two Way Fixed Effect

The two way fixed effect is used when impact of both entity and time is to be modeled. It is express as:

$$Y_{it} = \beta_1 X_{it} + \alpha_i + \partial_i + u_{it}$$

Like one way fixed model, in application, it can also be modeled using n-1 and m-1 binary variables as:

$$Y_{it} = \beta_0 + \beta_1 X_{1,it} + \gamma_1 E_1 + \ldots + \gamma_{n-1} E_{n-1} + \varphi_1 T_1 + \ldots + \varphi_{n-1} T_{n-1} + u_{it}$$

Where,
α_i and ∂_i = Unknown intercept for each entity and time respectively.
E_n and T_n = Binary variable for each entity and time respectively.
γ_n and φ_n = Coefficient for the binary variables for entity and time respectively.

However, as discussed by Kropko & Kubinec (2018), this structure for two way fixed effect is often difficult to interpret and derive any meaningful conclusion. While impact compared to an entity or time is often desirable, when both entity and time are together used as a reference point (as in this case); the resulting coefficients have little usable meaning. For example, if the above method is followed, in which the omitted binary variable is entity 1 and time period 1, then the coefficient for entity E_2, denoted by γ_2 , would mean 'impact of E_2' as compared to entity E_1 and T_1. Similarly coefficient of time T_2 denoted by φ_2, would mean 'impact of T_2' as compared to entity E_1 and T_1.

Instead, two possible alternatives are suggested, each with its own merits and demerits.

1) Using binary variables for either entity or time dimension and mean or instance centering the other dimension, i.e.:

$$Y_{it}^{*} = \beta_1 X_{it}^{*} + \alpha_i + \partial_i + u_{it}$$

Where,

$$Y_{it}^{*} = Y_{it} - \bar{Y}_{i\ or\ t} \quad \text{And} \quad X_{it}^{*} = X_{it} - \bar{X}_{i\ or\ t}$$

2) Use (mathematical) combination of n-1 entity and m time as binary variables, i.e.:

$$Y_{it} = \beta_0 + \beta_1 X_{1,it} + \gamma_1 E_1 T_1 + \gamma_2 E_1 T_2 + \ldots + \gamma_{(n-1)*m} E_{n-1} T_m + u_{it}$$

The advantage of this (2) method is easier interpretation and application, i.e. impact of entity as compared to omitted entity in the same time period. However, the number of binary variables is drastically larger and increase multiplicatively.

- Random Effect

In random effect model, unlike fixed effect, the variation between entities is assumed to be random and uncorrelated with independent variables in the model. It is defined as:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \varepsilon_{it} + u_{it}$$

Where, $\varepsilon_{it}\ and\ u_{it}$ are with-in entity and between entity error respectively. However, unlike fixed effect model, random effect is prone to omitted variable bias (Reyna , 2007).

To determine suitability of fixed effect or random effect model, Hausman Test can be used. It tests whether u_{it} is correlated with independent variables with null hypothesis stating that they are not. Hence the null hypothesis can be interpreted as that the random effect model is more suitable with alternative stating better suitability of fixed effect model (Green, 2008).

III. Research Methodology and Case Study

This chapter explains the methodology deployed in this study. Broadly, the methodology can be subdivided into 'identification of LCZs in Seoul' and deploying the LCZs to determine urban forms impact on building energy consumption. Subsequently, the major shortcoming and issues faced while deploying this methodology are briefly discussed. Finally, the case study area of Seoul is concisely introduced and the reasoning behind its selection and challenged posed by it are discussed.

3.1 Methodology

3.1.1 LCZ classification

Figure 3-1 provides an overview of the process followed. Broadly, the process can be divided into three parts.

1) Deciding the appropriate grid size and location such that maximum possible homogeneity of urban form within a grid is achieved.
2) Calculation of LCZ parameter for each grid. Identify grids which satisfy conditions for at least one of LCZ class.
3) Use machine learning algorithm to assign LCZ class for grids which fails to satisfy the criteria for any LCZ class. Two separate procedures, Constrained Over Parameter (COP) K-Means and Random Forest are used.

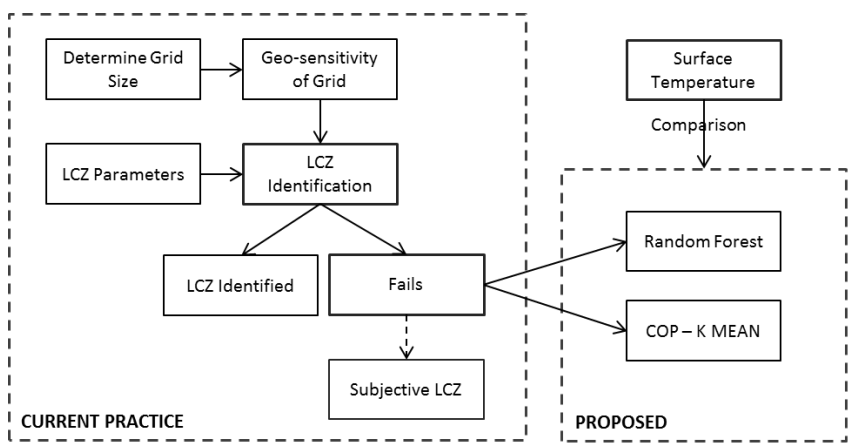


Figure 3-1) LCZ classification methodology

3.1.2 Linking urban form with building energy consumption

Figure 3-2 provides an overview of the process followed. Firstly, categorical variable (LCZ Class) is converted into n-1 dummy variables. Then, other independent variables which can also have an impact on building energy consumption, such as socio-demographic profile, building specific attributes, etc.

are calculated. Dependent variable (electricity and gas consumption per month) was calculated for each grid by taking mean of all the residential buildings (for which data is available) in the grid. Finally, using fixed effect panel data analysis, dependent variables are linked with LCZ-Month dummy variables and other independent variables in two separate models, for electricity and gas respectively.

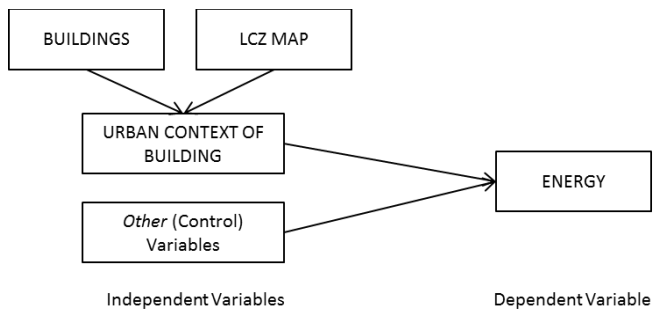


Figure 3-2) Methodology for relating urban form with building energy consumption

3.2 Issues with Methodology

The methodology discussed above posed a number of challenges. Here a brief overview is presented, with details described in relevant sections when each aspect of the methodology is discussed.

Firstly, despite the progress in grid size and location determination process, modifiable areal unit area posed significant issues. While the study used statistical methods to minimize sensitivity to the boundary of grid and loss of small feature in aggregation, because the LCZ requires six parameters with their own ideal grid size and location, it was not possible to have an error free grid system.

Secondly, some assumptions were made during LCZ parameter calculation due to data limitations. This includes the absence of tree related data (height, etc.), building as the cube of their footprint, etc. Additionally, data limitations required the use of surface temperature instead of air temperature for validation of the LCZ classification.

Lastly, in LCZ and BEC modeling, the energy dataset was found to be skewed for building of larger size. Additionally, some of the 'other control variables' in the model were available at a different resolution or grid location, thus requiring resampling, which without recourse induced error into the dataset.

3.3 Case Study

Seoul is the largest city of South Korea and fourth largest metropolitan economy in the world. It has a total land area of 605 KM^2 with an average elevation of 38 M. About 25.29% of land falls under the regulated green belt where limited or no development activity is allowed. The general structure of Seoul is polycentric with high and medium density residential areas and general lack of smooth density pattern. With average density of 16,000 / KM^2 ,the metropolitan area of the city is inhabited by more than half the population of whole South Korea.

Seoul was chosen as potential case study because:

1) The study area contains diverse land-use and morphology, hence it was possible to cover majority of LCZ classes.
2) The large amount of area is covered by water and green body, thus the inclusion of natural classes was possible.
3) The study area was required to be big enough to generate sufficient sample for two way fixed effect models due to a large number of dummy variables deployed in the model.

However, the study area did posed number of issues, some of which were addressed, while attempt was made to minimize the impact of remaining one and are further discussed in limitations (Section 5.3).

1) Like most urban area, many parts of Seoul did not fell under any LCZ class. Thus clustering was used to estimate *closest* LCZ class.
2) The issues discussed regarding limitation of LCZ are amplified in case of Seoul. Not only mountain ranges forms an inseparable part of Seoul's landscape, organic growth means that the *urban material* is diverse. Both of these aspect influences microclimate and BEC, but are not studied in this research.

IV. Analysis and Results

This chapter builds on theoretical discussions from the previous chapter. Firstly, the methods for calculating LCZ parameters for Seoul used in this research are discussed. Data and software used for parameter estimation are described and shortcomings or limitations of a method or data are discussed. Secondly, the results from classifications are shown and compared using surface temperature from four seasons. Thirdly, investigation into mean BEC for each LCZ is shown to describe the idea which was the initial basis and pilot analysis for this research. Lastly, results from panel data analysis are shown to determine the relative BEC efficiency of urban LCZ.

4.1 LCZ Classification of Seoul

4.1.1 Calculations for LCZ parameters

Table 4-1: LCZ parameters discussed in section 2.1.2 were calculated as per description below:

Parameter	Definition	Data Type	Calculation
BH	Footprint area weighted mean of building height	Building Footprint Vector Data	$BH = \frac{\sum_{i=1}^{n} BF_i * BH_i}{\sum_{i=1}^{n} BF_i}$
BSF	Percentage of space covered by building with respect to grid size.	Building Footprint Vector Data	$BSF = \frac{\sum_{i=1}^{n} BF_i}{Grid\ Area}$
PSF^	Percentage of space occupied by water body, parks & forest and number of trees outside of parks/forest * $1m^2$	Natural Feature Vector Data	$ISF = \frac{\sum_{i=1}^{n} WA_i + \sum_{i=1}^{n} GA_i + T}{Grid\ Area}$
ISF	Percentage of space occupied by impervious surface	Building Footprint and Natural Feature Vector Data	$ISF = 100 - BSF - PSF$
AR^^	Mean of asymmetrical aspect ratio measured at 10 meter interval	Building footprint and street vector data	$AR = \frac{\sum_{i=1}^{n} \frac{BH_i}{SW_i}}{n}$
SVF^^^	Mean of Sky View Factor with Grid	Building footprint and tree vector data. DEM.	$SVF = \frac{\sum_{i=1}^{n} SVF_i}{n}$

BH = Building Height, BSF = Building Surface Fraction, PSF = Pervious Surface Fraction, ISF = Impervious Surface Fraction, AR = Aspect Ratio, BF = Building Footprint Area, WA = Water Body Area, GA = Park and Forest Area, T = Number of Trees, SW = Street Width, SVF = Sky View Factor, DEM = Digital Elevation Model

^ In absence of related data, we assume each tree has $1m^2$ of pervious surface around it.

^^ Different approaches have been adopted for the calculation of aspect ratio in previous researches. For example, Zheng et al. (2018) have used a simplified method of sum of the street area divided by street length. Memon et al. (2010) have built a detailed model for aspect ratio in hypothetical symmetrical condition. In this study, we used mean of building height on both sides of street to estimate asymmetrical aspect ratio. The process and required GIS tools have been discussed by Mohajeri et al. (2016)

^^^ This research uses Urban Multi-scale Environmental Predicator (UMEP) (Lindberg et al., 2018) for SVF estimation. SVF was calculated at 2 meter resolution since the finest resolution DEM for Seoul is available at 2 meter. The area occupied by building was removed before SVF aggregation. Additionally, it needs to be mentioned that tree database only covers 'major' trees and we assumed that all trees to be of uniform height of 10 meters, in absence of availability of such data.

4.1.2 Grid size and location

As discussed in section 2.3.1, building heights and sky view factor were used for Moran's I analysis. Other parameters of LCZ could not be reduced to point features. In ArcGIS 10.4. This can be done using 'Incremental Spatial Autocorrelation' tool in spatial statistics toolbox. The results of Euclidean distance based Moran's I are shown in figure 4-1. For building height, we observe maximum Z_I at a distance of 430 M while for SVF, it is 390 M. Consequently, 400 meter was chosen to be the appropriate grid size for LCZ classification of Seoul.

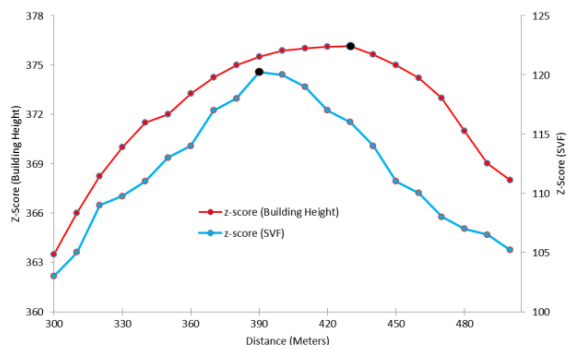


Figure 4-1) Spatial autocorrelation distance for building height and sky view factor in Seoul

As per discussion in section 2.3.2, for all LCZ parameters except SVF, we fail to reject K-S test's null hypothesis at 95% confidence interval. This indicates that the frequency distribution of all parameters except SVF followed a similar curve despite the change in location of grid (figure 4-2). However, for SVF, the p value is

smaller than 0.05, indicating that new grids had a different distribution than the base grid. In such scenario, Zheng et al. (2018) recommends selecting option which has minimum standard deviation of factor failing K-S test with the rational of minimizing variation within grid when comparison of LCZ is the objective. Thus we calculated the standard deviation of SVF for each grid in each geo-location option and chose 'SW' (see figure 2-10), since it has least standard deviation in 7 of 10 LCZ categories.

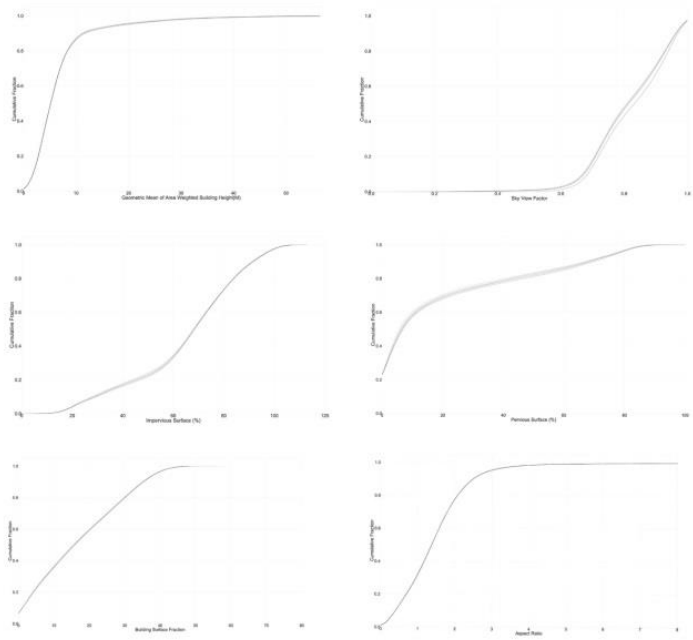


Figure 4-2) Cumulative frequency distribution of LCZ parameters for alternative grid locations.

4.1.3 Classic classification

In case of Seoul, no grid which satisfied LCZ 10's (Heavy Industry) criteria was found. Review of land-use-land-cover (LULC) map confirmed the absence of heavy industry within city boundary. Thus LCZ 10 was not considered in any further classification. Additionally, due to absence of data on green bodies features, LCZ A, B, C, D were considered together as one class 'LCZ Green'. Lastly bare rock and soil occupied less than 5% in any of the grid and thus LCZ E,F were also not considered in classification.

Classic attempt to classify Seoul based upon 11 LCZ categories (LCZ 1-9, LCZ Green, LCZ Water) resulted in only 1121 of 4255 grids be assigned LCZ class (Table 4-2) (Figure 4-3). All other remaining grids failed to satisfy one or more

LCZ criteria for any of the classes. Figure 4-4 shows typical view of LCZs in Seoul from high resolution Google Earth imagery.

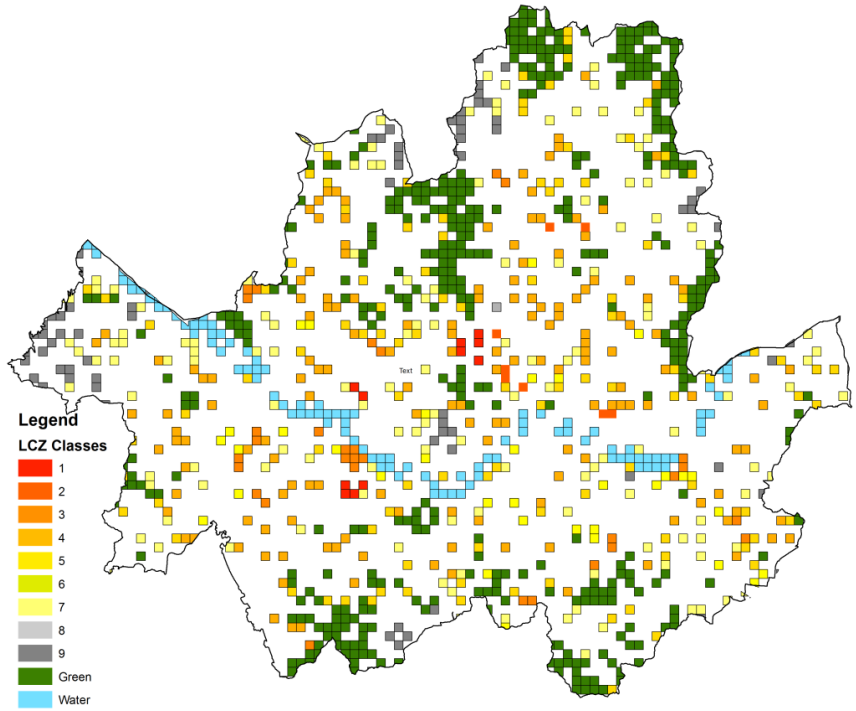


Figure 4-3) LCZs in Seoul which satisfy all the parameters (classic classification).

LCZ 1 Compact High Rise

LCZ 2 Compact Mid Rise

LCZ 3 Compact Low Rise

LCZ 4 Open High Rise

LCZ 5 Open Mid Rise

LCZ 6 Open Low Rise

LCZ 7 Light Weight Low Rise

LCZ 9 Sparsely Populated

Figure 4-4) Typical urban form example using Google Earth images for major LCZ in Seoul.

4.1.4 Algorithm based classification

For remaining unclassified grids, COP K-Means and Random Forest (Section 2.2.1 and 2.2.2) was used. Below, the parameters used for both algorithms are specified.

- COP K-Means
  1) Number of classes (K) = 11
  2) Initial Centers = Z-score of LCZ parameter's center, for each of 11 classes under study.

3) An observation which was *classically* classified was specified as 'must link' other observation in class and must not link with observation in other classes.
4) Iteration = 10,000
5) Tolerance = 0.0001

- Random Forest
  1. No of trees = 2000
  2. Criterion = Gini
  3. Minimum impurity split = 0.01

For COP K-Means, Silhouette scores for classification was 0.715 indicating algorithm has produced a fairly distinct cluster. Silhouette score describes how tightly a cluster is packed and its separation from other clusters. It ranges from -1 to 1, with 1 indicating the observation is most similar to another member of cluster, than to any other observation (Rousseeuw, 1987). Out of bag Error, a similar matrix for Random Forest (varies from 0 to 1) reported score of 0.87.

Figure 4-5 show classification results from COP K-Means and Random Forest respectively, and table 4-2 show the distribution.

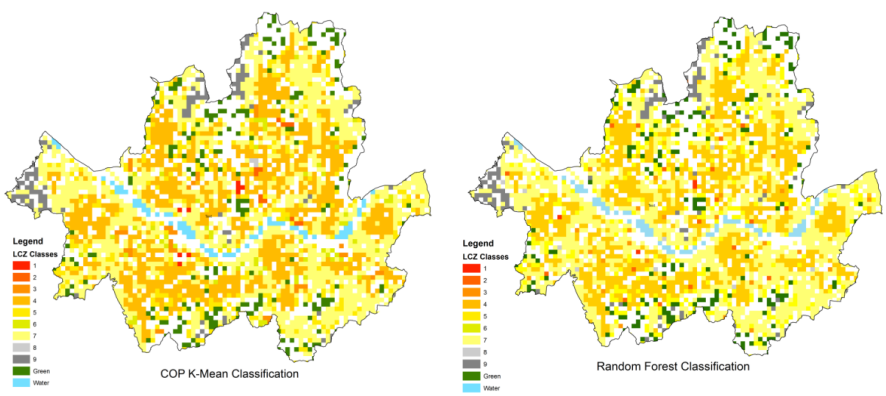


Figure 4-5) LCZ classification using machine learning algorithms

Table 4- 2) LCZ Distribution

LCZ Class	Classic Classification	Random Forest Classification	COP K-Means Classification
1	11	4	9
2	8	4	13
3	31	12	39
4	160	966	973
5	102	399	330

6	34	1009	1065
7	145	307	245
8	1	33	5
9	91	136	178
Green	431	192	211
Water	107	72	72
Total	1121	3134	3134
% Classified	26.35	73.65	73.65
Total Number of Grids = 4255			

4.1.5 Surface temperature profile of LCZs

Ideally, the LCZ classification should be validated against air temperature at various heights. However, in the absence of sufficient air temperature data, we decided to use surface temperature as validation method. Zhao (2018) has shown that surface temperature varies distinctively for most of LCZ categories.

Landsat 8 Band 10 was processed in accordance with the method specified below using ENVI (Harris Geospatial, 2017) and surface temperatures were collected for each grid. TOA brightness temperature is calculated as per USGS manual (2018) :

$$T_A = \frac{K2}{\ln(\frac{K1}{L_\lambda} + 1)}$$

$$L_\lambda = M_L * Q_{cal} + A_L$$

Where,

L_λ = Spectral radiance
M_L = Radiance multiplicative scaling factor for the band
A_L = Radiance additive scaling factor for the band
Q_{cal}= Digitial Number
T_A = TOA Brightness Temperature, in Kelvin.
K1,K2 = Thermal conversion constant for the band

Number of methods to convert TOA to LST has been developed depending upon the number of channels, emissivity information, etc. A comprehensive review and comparison has been done by Li et al. (2013). We use NDVI derived emissivity method (Valor & Caselles, 1996) which uses normalized difference vegetation index (NDVI) derived from band 4 (Red) and band 5(NIR) of Landsat 8.

$$\text{LST} = \frac{T_A}{1 + \frac{\boldsymbol{\lambda T_A}}{p} ln\varepsilon_{\boldsymbol{\lambda}}}$$

$$\varepsilon_{\boldsymbol{\lambda}} = \varepsilon_{\mathbf{v}\boldsymbol{\lambda}} + \varepsilon_{\mathbf{s}\boldsymbol{\lambda}} (1 - P_v)$$

$$P_v = [\frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}}]^2$$

Where,

LST = Land surface temperature
λ = Wavelength of emitted radiance (Landsat 8, Band 10 λ = 10.895μm (USGS, 2018))
P = Planck's constant (1.438 * 10^{-2} mK)
ε_λ / $\varepsilon_{v\lambda}$ / $\varepsilon_{s\lambda}$ = Emissivity of Pixel / Emissivity of Vegetation / Emissivity of Soil
P_v = Vegetation fraction in Pixel
$NDVI_{max}$ /$NDVI_{min}$ = Max/Min NDVI value of study area.

Since "LCZ classification is based on the thermal characteristics of urban area" (Bechtel et al., 2015) it is reasonable to assume that classification algorithm that produces LCZ map with lesser variation in temperature (i.e. box-plot length is smallest) is better in identifying the closest resembling LCZ class of unknown grids. Figure 4-6 shows the surface temperature distribution for LCZ classes from classic classification (grids with parameters in the range of original LCZ classification), Random Forest and COP K-Means. Four different dates for different seasons in Seoul are shows separately. In all urban form related LCZ (1 to 10) except for 8 (Large low-rise), the Constrained K-Means method result have lower variance than Random Forest in all four seasons. LCZ 8 results should be treated with caution due to small number of grids falling in the class.

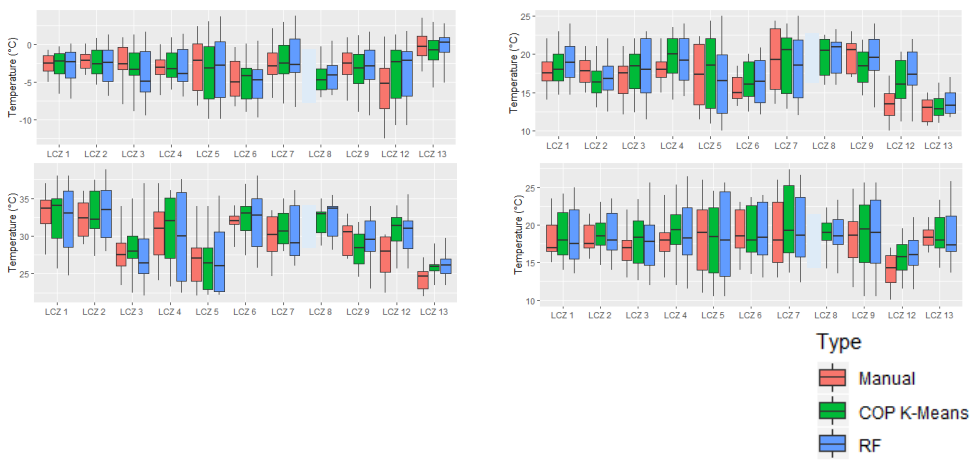


Figure 4-6) Boxplot of surface temperature of LCZs for 14 January (top left), 19 March (top right), 26 August (bottom left), 13 October (bottom right) (2017)

Thus (at least for Seoul) it can be concluded that COP K-means produces better classification results. Combined (manual and COP K-means) LCZ map of Seoul is shown in figure 4-7.

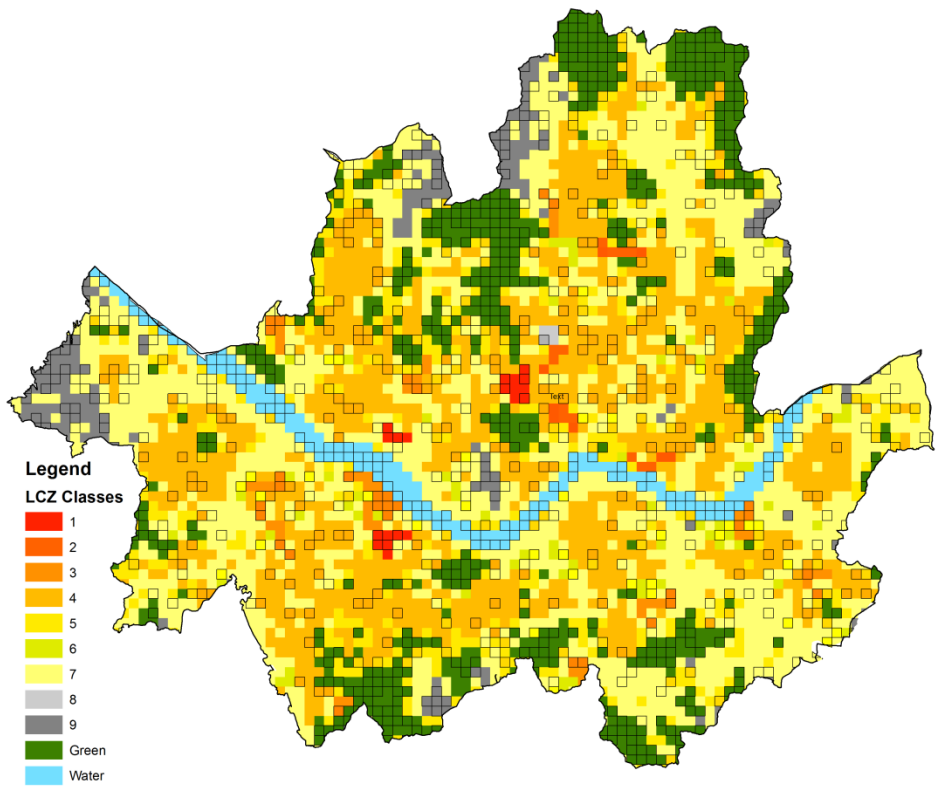
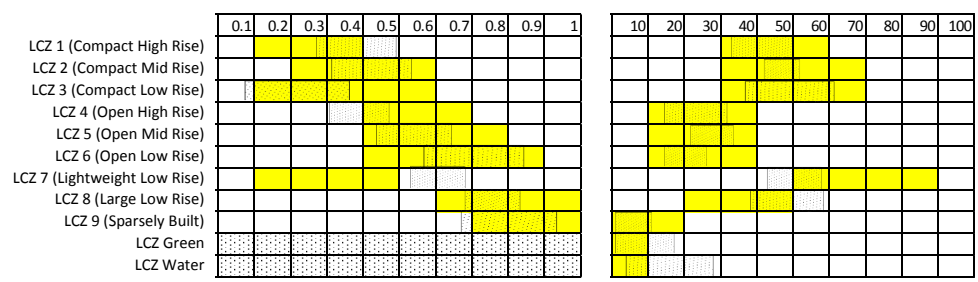


Figure 4-7) LCZ map of Seoul

4.1.6 LCZ parameters for Seoul

Based upon above classification, the LCZ parameter ranges are shown in fig 4-8.



Sky View Factor Building Surface Fraction

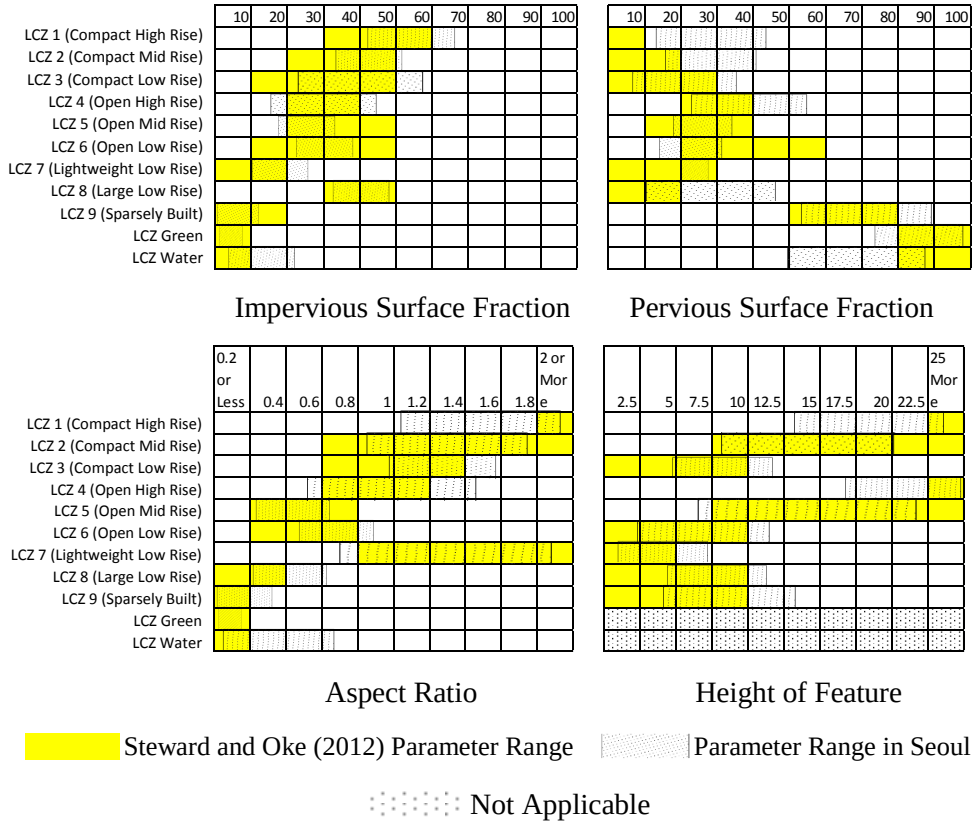


Figure 4-8) LCZ parameter after classification in Seoul.

Pervious Surface Fraction (PSF) and Aspect Ratio (AR) appear to be the biggest cause of partial failure of the classic classification method. For example, for LCZ 1 (Open High Rise), the PSF in Stewart & Oke (2012) parameter table should be less than 10%. However, we see in case of Seoul, the range is 14 to 44 % with right extreme influenced by core area located near to mountains which are predominately green and thus have higher pervious surface percentage.

4.2 LCZ and Building Energy Consumption

As discussed in section 2.4, there are evidence that building energy consumption is influenced by the urban form within which building is located. Appendix 1 shows the trend of electricity and gas consumption for each LCZ. Since it would be difficult to determine any sort of trend from such figure, thus as exploratory analysis, moving average time series analysis was performed as discussed in section 2.5.1. To infer how consumption varies in different urban typology with respect to natural area, the results were normalized w.r.t. LCZ 9 (Sparsely Built Areas), as shown in figure 4-9 and 4-10 for electricity and gas respectively.

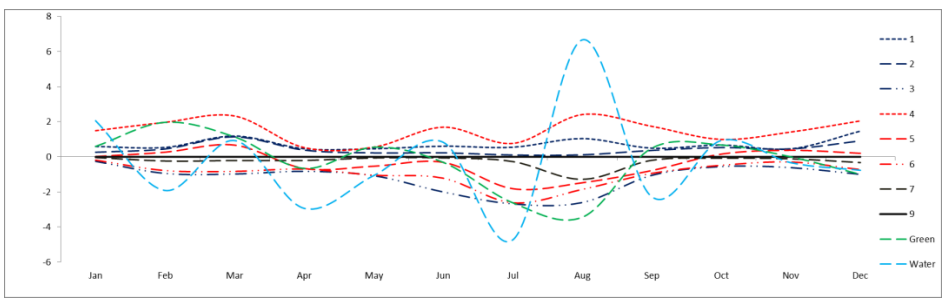


Figure 4-9) Result of moving average analysis for electricity consumption normalized w.r.t. LCZ 9.

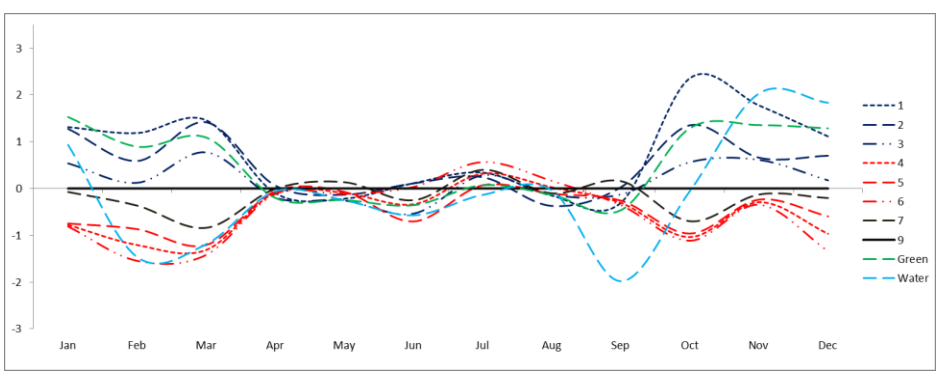


Figure 4-10) Result of moving average analysis for gas consumption normalized w.r.t. LCZ 9.

Figure 4-9 and 4-10 are interpreted as difference in energy consumption per sq. meter of floor space in a given LCZ, as compared to LCZ 9. In case of electricity, it is observed that both compact and open high-rise and compact mid-rise has higher energy consumption than LCZ 9. On the other hand, compact and open low rise has lower energy consumption, while trend for other LCZ varies, depending upon season. For gas, the difference is negligible in summer months (May – Sept). In

winter, however, while all 3 compact LCZs (1- 3) has lower energy consumption, open LCZs (4-6) has higher energy consumption as compared to LCZ 9.

However, building energy consumption depends on a multitude of factors, including socio-economic variables and thus the above results only explain the difference and not the impact of urban form (defined by LCZ) on energy consumption.

4.2.1 Model results

Thus to extract the pure impact of urban form on building energy consumption, two way fixed effect model (section 2.5.2) is used, in which LCZ is used as entity with LCZ 9 acting as the baseline and months are used as time to capture the seasonal nature of impact. Table 4-3 provides the list and description of independent and dependent variables.

Table 4-3) Variables in the model

	Name	Code	Unit	Description
Dependent Variable	Electricity	EL	KWH/M^2	Monthly mean of electricity consumption per sq. meter by residential buildings within a grid
	Gas	GS	KWH/M^2	Monthly mean of gas consumption per sq. meter by residential buildings within a grid
Independent Variables	Population Density	Pop_Den	Person/1000M^2	Number of people per 1000 sq meter within a gird
	Population Above 65	Per_65	Percentage	Percentage of population above 65
	Population Below 19	Per_19	Percentage	Percentage of population below 19
	Building Age	Mean_BLD_Age	Year	Mean of building age within grid
	Real Estate Price	Mean_RE_Price	100000won/M^2	100,000 won /m^2 floor Space
	Spatial Autocorrelation	SA_Control	N/A	Term to control for spatial autocorrelation of error term. See Appendix 4 for details.
	Dummy Variables	Month_LCZ	KWH/M^2	9 x 12 (No. of LCZ -1 * Time Intervals)

Table 4-4 and 4-5 gives the model summary for electricity and gas respectively. Detailed coefficient values are shown in Appendix 2. For better interpretability only major urban form LCZ (1 to 6) and those coefficients which are statistically significant at $P < 0.05$ are shown in figure 4-12.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.827[a]	0.683	0.653	2.778

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	14.330	1.329		0.000
	Pop_Den	9.523	1.712	0.183	0.040
	Per_65	-1.643	1.162	-0.025	0.056
	Per_19	2.543	1.869	0.041	0.020
	Mean_BLD_Age	0.653	0.181	0.010	0.000
	Mean_RE_Price	3.925	1.755	0.047	0.019
	SA_Control	2.524	1.296	0.035	0.000

Table 4-4 Model summary for electricity

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.794[a]	0.630	0.613	2.840

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	11.454	1.534		0.000
	Pop_Den	14.994	1.484	0.183	0.027
	Per_65	2.365	1.297	-0.025	0.006
	Per_19	-1.345	1.990	0.041	0.002
	Mean_BLD_Age	1.843	1.031	0.010	0.035
	Mean_RE_Price	1.532	1.077	0.047	0.006
	SA_Control	1.235	1.564	0.035	0.011

Table 4-5) Model summary for gas

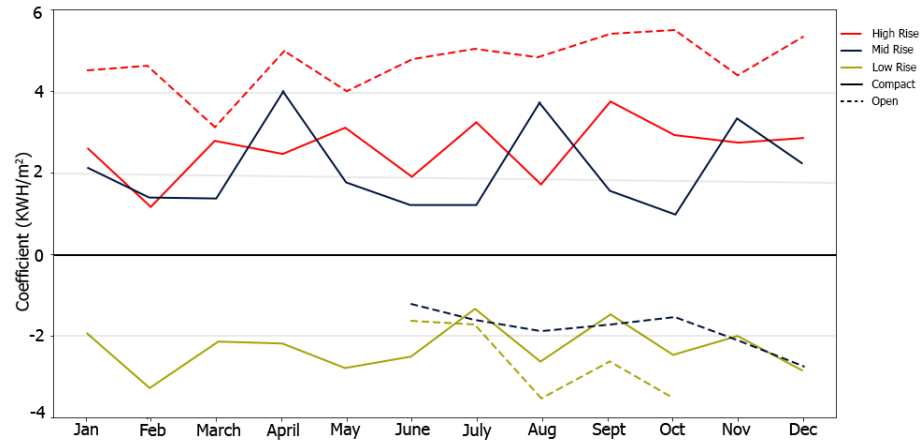
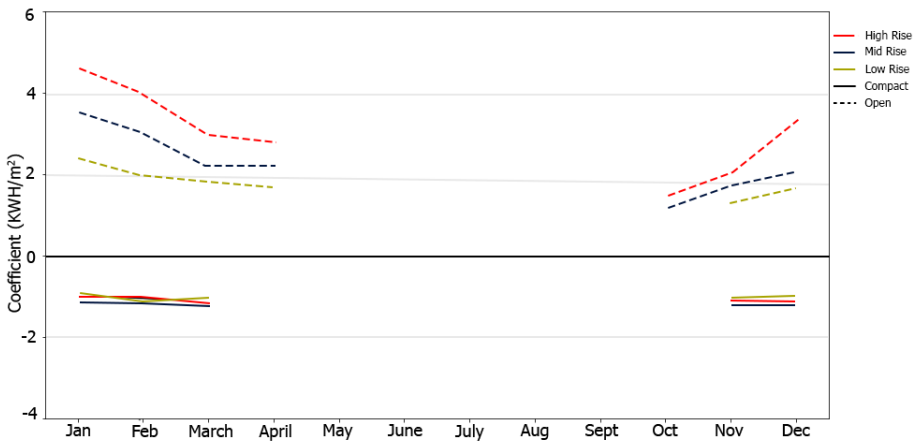


Figure 4-11 Coefficients of LCZ-Month dummy variables for electricity model described in table 4-4



.;

Figure 4-12 Coefficients of LCZ-Month dummy variables for gas model described in table 4-5

In figure 4-11 / 4-12, each coefficient can be interpreted as difference in electricity/gas consumption due to its urban form, as compared LCZ 9 (sparsely built). In case of electricity, LCZ 3 (Compact Low-Rise) has lower electricity consumption throughout the year, with mean of 2.09KWH/m^2 electricity savings due to its urban form. Additionally LCZ 5 & 6 (Open Mid & Low Rise) in summer and LCZ 7 (Lightweight low rise) in winter also have lower electricity consumption. Conversely, LCZ 1, 2 & 4 (Compact High & Mid & Open High Rise) have higher electricity consumption.

In case of gas, urban form seems to have no impact on consumption during summer period. However, for colder period, compact LCZ has have lower and

open LCZ has higher gas consumption. The difference in consumption is higher during extreme winter and reduces during moderate winter months. Annually, this translates into - 4.01/m^2 consumption in compact LCZ and +17.05/m^2 consumption in open LCZ due to urban form, as compared to sparsely built areas.

4.3 Alternative Model

The discussion and the analysis below is based upon the idea that the broader landscape condition would also have an impact on BEC, but have not been incorporated in analysis done in section 4.2. Specifically in case of Seoul, due to its mountainous topography, it would be reasonable to expect that buildings located at higher altitude would have different energy consumption due to differences in solar exposure and wind pattern, as compared to those located in plain areas. Below, it is discussed why this issue exists, how it can be addressed and experimental analysis to determine whether a link to underlying (mountainous) landscape exist. The reason this is an *alternative model* is because one of the main contribution of this research was LCZ classification methodology. Hence, developing a new classification framework or modifying it was considered beyond the research's scope.

4.3.1 Landscape in LCZ

LCZ framework, at least so far, has been agnostic towards broader landscape and context of the study area. This means that effects of features such as mountain, nearby water body, etc. are not considered because none of 6 classification parameter account for them. These issues stem from two reasons, explained using the example of mountain and water body.

- In case of the mountains, there is no elevation parameter. LCZ were developed by *reverse engineering* Town Energy Balance Model (TBDM). Neither initial literature on LCZ development nor subsequent research addresses whether elevation was not found to have significant impact in TBDM and thus excluded or its exclusion was an oversight.

- On the other hand, water body itself is an LCZ class. However, if a small water body is present within a grid, or a grid is surrounded by water body, the grid's class remains agnostic to it. This stems from the broader issue of grid size and location (also called as modifiable area problem in GIS literature). The original framework has no mechanism to decide on size of unit or its boundary shape. Hence, for example, while *surface cover* might be better analyzed X by X square grid, *structure* is better suited for Y by Y hexagonal grid.

Two alternative solutions are possible. Either we can have a fluid grid size (and boundary). Or we can *somehow* assign multiple classes to a single grid. For the

latter case, instead of using rigid clustering used in this research which tries to find the most dominant class in a heterogeneous grid, it could be possible to use fuzzy clustering which finds the constituents of heterogeneity. However this is beyond the research scope.

4.3.2 Mountainous landscape, LCZ and BEC

In this section, the analysis is done to determine if mountains have any impact of BEC. A straight forward approach would be to include elevation as independent variable. However, it is expected that LCZ class in a covariate, i.e. the amount of influence of elevation is dependent on LCZ class of elevation. Thus, LCZ classes (or part of it) which are located in mountainous area were identified using below given map (figure 4-13) and the sample was divided into LCZ_Not_Mountain and LCZ_Mountain. Here, mountain is defined as any area where the elevation exceeds 100 M.

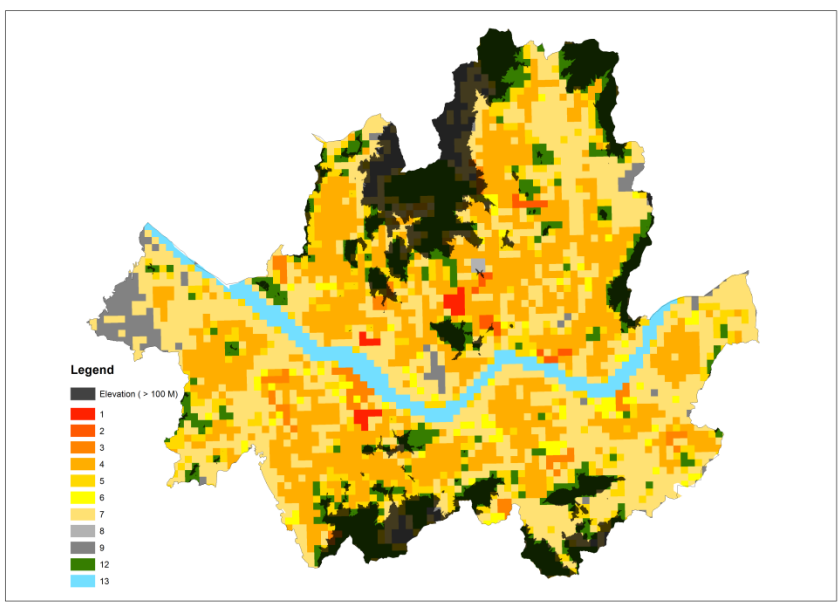


Figure 4-13 LCZ map and mountainous areas

LCZ 1, 2, 8 and water have no grids which lie in mountainous areas. A two way fixed effect panel data analysis, similar to that done in section 4.2, but with (9 x 12) + (5 x 12) (for non-mountain LCZ and mountain LCZ respectively) dummy variables is done. The R Square / Adjusted R Square for electricity and gas for new model are 0.694 / 0.641 and 0.711 / 0.671, respectively. The statistically significant (at 95% CI) coefficients are shown in figure 4-14 and 4-15.

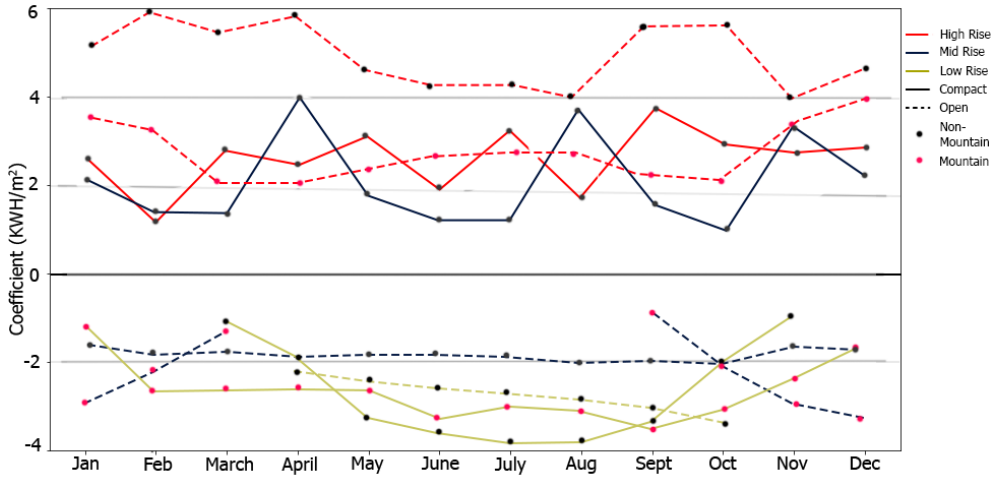


Figure 4-14) Significant coefficient of dummy variables in electricity model.

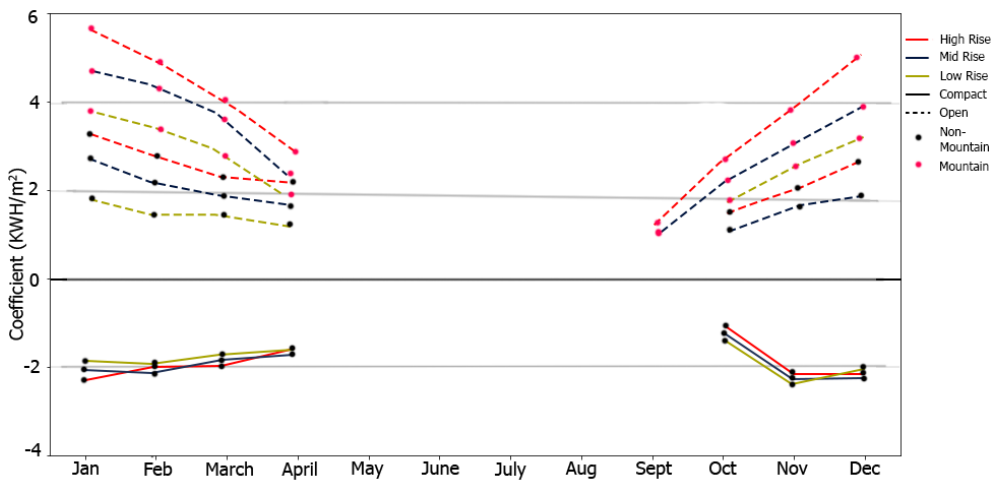


Figure 4-15) Significant coefficient of dummy variables in gas model.

Considering the overall model, as with increase in number of I.V., R-square increases in both cases. However, adjusted R-square which is penalized if the new variable is non-significant or is correlated with other I.V., decreases for electricity but increases for gas. This means that the new model which differentiate between mountain and non-mountain area performs better in case of gas, but slightly worse in case of electricity model (despite the increase in R-square for both).

Considering individual coefficients, for electricity model, the efficient (or inefficient) LCZs stay the same. However, for inefficient LCZ 1 (Open High Rise), mountainous area has much higher consumption as compared to non-mountainous areas. Additionally, for efficient LCZs, the changes in coefficients as compared to

model 3.2.1 are drastic, but the general observation of Open Low Rise being more efficient than Open Mid Rise remains true. In case of gas model, for inefficient LCZs, mountainous class consistently has the higher consumption as compared to their counterpart non mountainous class. For efficient LCZ classes, all mountainous classes have insignificant coefficients and as compared to model 3.2.1, the coefficients for April and October are also significant for some LCZ classes.

In conclusion, for gas, it is clear that the broader landscape factors missing from LCZ framework (such as mountains) does have significant impact on consumption. However, for electricity, a contradictory conclusion is drawn. The decrease in adjusted R-square indicates that newly added dummy variables (or part of them) does not improve overall model strength. However, the coefficients of existing dummy variable changes (and remain significant) and some of new coefficients are also significant.

V. Conclusion, Policy Suggestion and Limitations

This study primarily contributed to urban climate and energy literature by exploring new ways for GIS based LCZ classification and relating LCZ as an identifier of urban form to measure its impact on building energy consumption. The widely reported issue of failure to identify LCZ class based on Stewart & Oke (2012) parameters alone is reiterated, when it was found that only 26.35% of Seoul satisfy all the parameters of atleast one of LCZ class.

Subsequently, we tested COP K-Means as an alternative to widely used Random Forest to identify classes of remaining areas. However, when the two classifications are compared using surface temperature, COP K-Means offers only fringe improvements. While the COP K-Means results have smaller variance in surface temperature as compared to Random Forest based classification, the improvements can best be described as marginal. Additionally, COP K-Means is absent in GIS and image processing software as of now and requires more detailed parameter input. Thus, it is debatable if the researchers will be willing to adopt this newer method. However, the move from Random Forest classifier trained using user defined ROI to centroids (derived from LCZ parameters) based method (whether K-Means or its variants used in this research or others) opens up possibility to deploy a host of different methodologies including investigations in fuzzy algorithms to determine the constituents of urban heterogeneity.

However, the study is far more successful in demonstrating the relationship between urban form and building energy consumption. The fragmented nature of the literature on urban form classification for BEC studies has hampered progress on this sub-field. The fact that relationship between LCZ and BEC is found to be significant helps in establishing LCZ as a legitimate canditate for urban form identification for BEC studies. While better results might be obtained by hand picking indicator of urban form, such studies would continue to exist in a vacuum. However, adopting common urban form identifier would not only help in cross comparisons, but also present a complete picture of a study area, devoid of caveats such as limited to a certain use, size, building type, etc.

Additionally, the study contributes in identifying relatively efficient urban form, in terms of comparative electricity and gas consumption. Our study portrays low rise as efficient, with LCZ 3 (Compact Low Rise) being most efficient in terms of energy saving in a building due to its urban context. However, this study does not claim this to be universal truth, but rather results for the Seoul city only. Further research which can accommodate different urban area and there climatic conditions is required to draw broader conclusion. However, we also see short coming of LCZ framework for BEC related studies, when we found missing parameters such as elevation to be a significant factor in microclimate and building energy consumption pattern. None the less, this study provides starting ground to develop

a common framework for defining urban area, whether it be LCZ in its current or modified form or some other framework, and methodology to relate it with BEC.

5.1 Conclusion

This research demonstrates the implementation of the LCZ classification to Seoul and explorers the widely faced problem of classification failure due to heterogeneity of urban form. Two algorithms, Random Forest and COP K-Means are used due to their popularity in LCZ literature and better theoretical support. The results from two algorithms are compared using surface temperature variance, with COP K-means producing marginally better results. This LCZ map of Seoul then becomes test best for investigation into the relationship between building energy consumption and urban form.

Using LCZ grid as unit of analysis, fixed panel analysis reveals small but statistically significant relation between the LCZ into which given floor space is located and its energy consumption. The consumption differences are normalized to LCZ 9 (sparsely built) to capture the influence of urban form and calculated independently for each month so that the opposite sign of coefficient due to seasons are not neutralized by one another. The study indicates that LCZ 3 (compact low-rise) to be most efficient urban form in terms of energy consumption differences arising from the urban context in which building is located. Additionally, the study suggests that natural features such as elevation also play an integral part in modulating BEC and microclimate.

However, the study requires further consideration on the LCZs applicability into Seoul due to its relief features and green belt regulations. Whether there is potential to modify LCZ such that it is more sensitive towards these missing features or a new classification is required, need to be further investigated. Additionally, bias due to the nature of data should be further investigated to answer how urban context and building features together control building energy consumption.

5.2 Policy Suggestions

Actionable recommendations require that the energy efficiency of urban form should be based on all the seasons and combined electricity and gas consumption. Fig 5-1 shows the sum of monthly coefficients for each LCZ to identify which LCZ offers potential for energy saving.

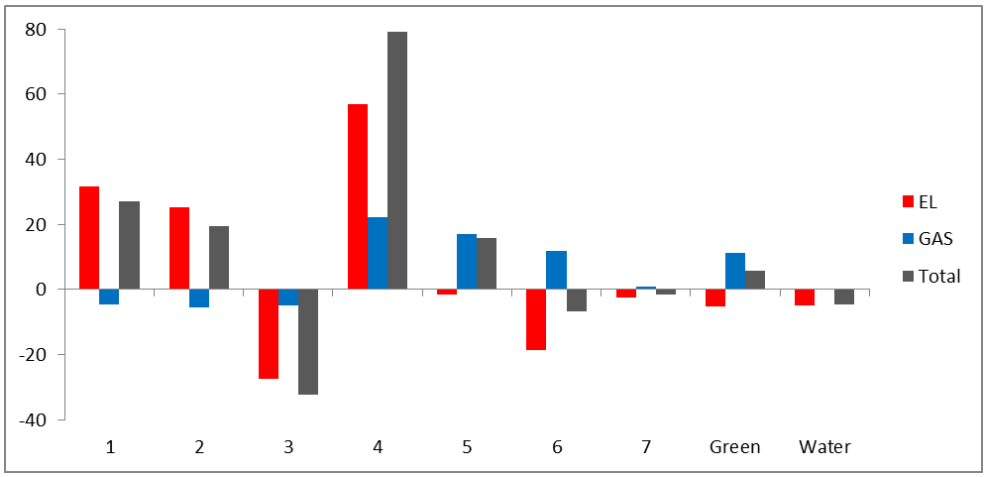


Figure 5-1) Annualized difference in energy consumption with respect to LCZ 9

It is clear from figure 5-1 that LCZ 4 (Open High Rise) has the highest difference in energy consumption as compared to LCZ 9 (Sparsely Built) for both electricity and gas. On the other hand, LCZ 3 (Compact Low Rise) offers highest potential saving in building energy consumptions, with this analysis suggesting that such urban form consumes 32.37 KWH/m^2 less energy annually, as compared to sparsely built area. Other than that, in terms of large scale habitable urban forms, LCZ 6 (open low rise) also have overall positive energy savings.

5.2.1 LCZ and zoning regulations

The local climate zone framework is based on parameters which aim to measure urban form as causal agent of UHI and microclimate differences. However, these parameters are rarely used for regulating urban form and growth management in planning practice. Thus to analyze how LCZ classes relate to residential zoning in Seoul, LCZ classification of Seoul (fig 4-7) was overlaid on the residential zoning map (용도지역 지도) (fig 5-2). The objective was to determine, in Seoul, what residential zoning (or its mixture) generates a given LCZ class. Table 5-1 provides a summary of residential zoning classes in Seoul. It should be noted that zoning classes discussed in Table 5-1 are from 2015[3] and have changed slightly afterward. This is because most of the dataset used in LCZ classification is from

[3] http://blog.naver.com/PostView.nhn?blogId=gm9282&logNo=220296639373

2015. It should also be noted that zoning is not the only urban form determinant in Seoul. Other regulations (federal and district) and superimposed special district ordinances are also frequently used to shape urban form. Additionally, despite being residential zone, some degree of commercial and other activities (depending upon type of residential zones) are generally allowed.

Table 5-1. Residential zoning classes in Seoul

Zoning Class	Zoning Class (Korean)	Maximum Coverage (in %)	Maximum FAR (in %)	Percentage Share in Seoul
General Residential 1	1 종일반주거지역	60	150	20.67
General Residential 2 (less than 7 floors)	2 종일반주거지역	60	200	25.91
General Residential 2A (more then 7 floors)	2 종일반주거지역	60	200	17.20
General Residential 3	3 종일반주거지역	50	250	30.14
Exclusive Residential 1	1 전용주거지역	50	100	1.52
Exclusive Residential 2	2 전용주거지역	40	120	0.25
Semi- Residential	준주거지역	60	400	3.96
Mixed Industrial	준공업지역	*	*	0.14
Green Belt	녹지지역	*	*	0.16
Other	-	*	*	0.018

* Depend upon regulations for *main* zone.

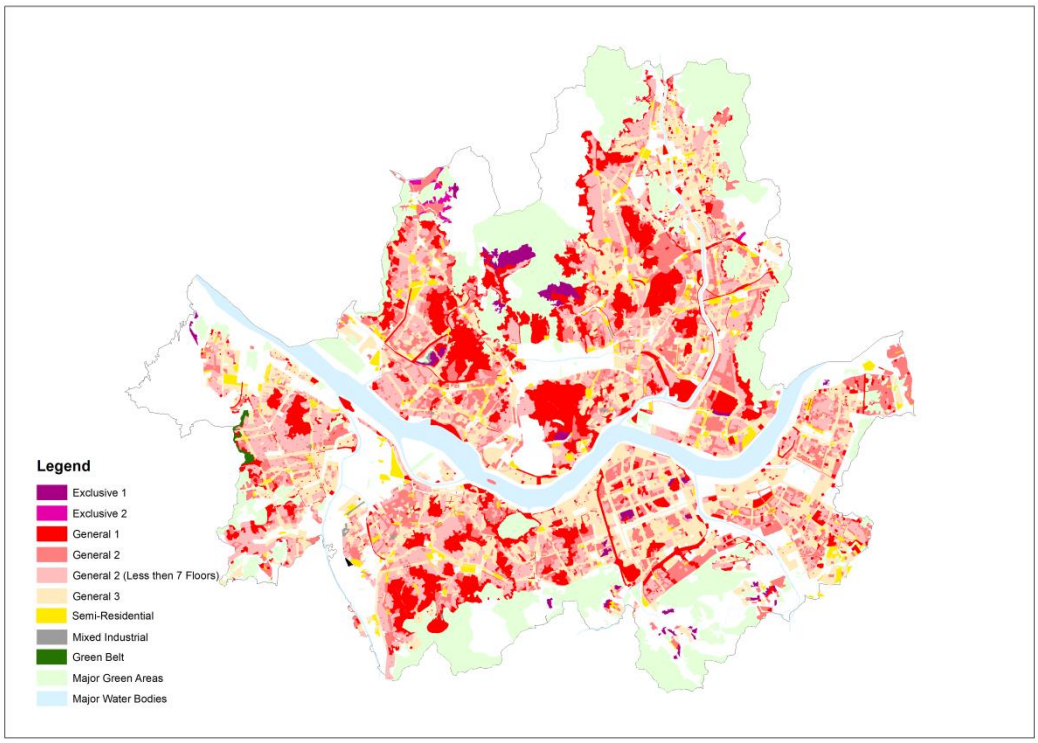


Figure 5-2) Residential zoning in Seoul.

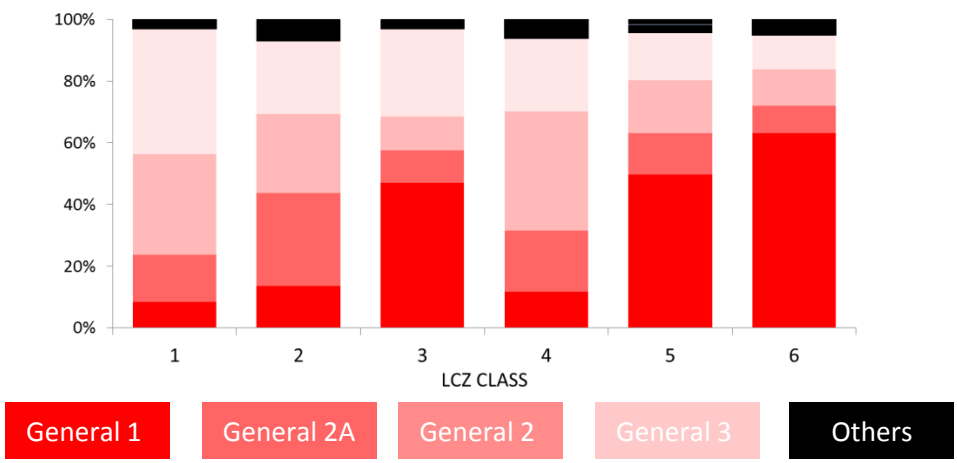


Figure 5-3) Zoning composition in major urban LCZ classes

Figure 5-3 show what the composition of residential zoning exists for major LCZ categories (1 to 6). LCZ 3 and 6, which were found to be most energy efficient, have a higher percentage of ‘general residential 1’ zoning class. LCZ 4, which was least energy efficient, has the highest percentage of ‘general residential

2 and 3'. LCZ 5 which although was slightly inefficient (+ 17.07 KWH/m^2), also has a higher percentage of general residential 1. Thus, given this contradictory results, the research struggles to find conclusive evidence for zoning type's relation with BEC efficiency.

5.3 Limitations

Primarily, the research has following limitations:

1) As discussed before, although LCZ framework has been shown to be suitable for surface temperature, it has been primarily developed for air temperature differences. The relationship between surface and air temperature is still researched upon and certainly do not have a linear association.
2) The research is not a comprehensive review of *best algorithm* for LCZ classification. It is possible that there are other machine learning algorithms, which may produce better results. However, it is unlikely that more complex approaches will see widespread adoption due to implementation difficulty and computational requirements.
3) While LCZ are described by measurable parameters, urban form is controlled by zoning and other land related law. Thus, it can difficult to formulate policies which would result in desired type of LCZ, without studying relation between LCZ and development regulation. While, the linkage between zoning and LCZ has been explored briefly, research to link LCZ with actionable policy is required, given the conflicting results obtained in section 5.2.1.
4) The outcomes, suggesting low rise as 'better' urban form for reducing energy consumption, is based solely on building energy efficiency and do not consider other types of energy expenditure, such as transportation or embodied energy, which may or may not support this conclusion.

Lastly, for fixed effect modeling studied in section 4.2.1, energy consumption data limitation has multiple implications. Generally, more observations are available for *larger* size building. Thus, mean percentage of buildings for which energy consumption data is available varies considerably w.r.t LCZ class. For example, for major urban LCZ grids, LCZ 1 had, on average, 47% of building with data points, while LCZ 6 only has 12% building with data points (figure 5-4). The major implication of this is that the mean energy consumption for each grid box has different error interval and the mean for grid with smaller sample size is more sensitive to outliers.

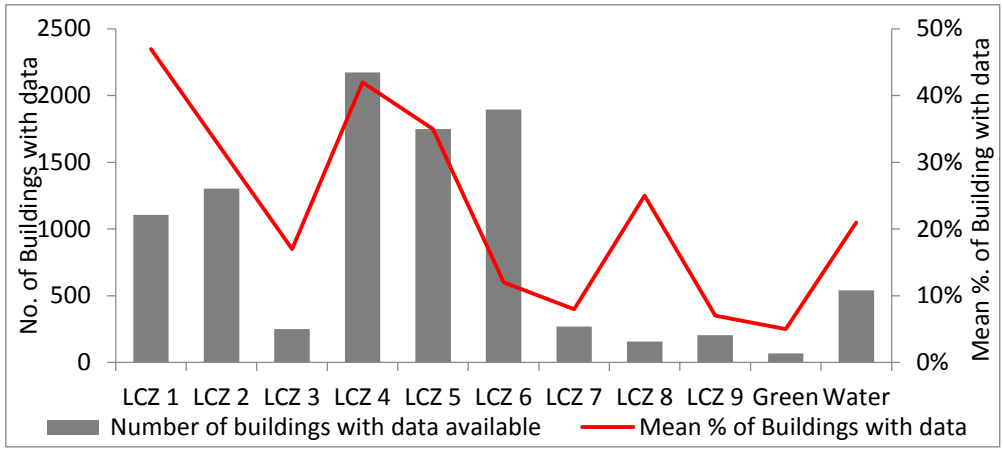


Figure 5-4) Sample size for different LCZs.

Additionally, as shown in Appendix 3, due to bias in data collection, the rate of sampling varies w.r.t to floor area. In Appendix 3, the space between solid line (sample distribution) and dotted line (population distribution) is missing data and it can be inferred that the dataset is biased towards the building of larger size. This raised the question of implacability of the results for building of all sizes. It may indeed be possible that urban form's impact on BEC dependents on building size, i.e. coefficients shown in figure 4-11 and 4-12 could vary if modeled separately for buildings of different sizes.

BIBLIOGRAPHY

ArcGIS, 2017. [Online] Available at: http://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/examining-spatial-autocorrelation-and-directional-variation.htm [Accessed September 2018].

ArcGIS, 2017b. [Online] Available at: http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm [Accessed November 2018].

Baker, N. & Steemer, K., 2000. *Energy and Environment in Architecture*. London: E&FN Spon.

Basu, S., Banerjee, A. & Mooney, R., 2002. Semi-supervised Clustering by seeding. In *Proceedings of the 19th International Conference on Machine Learning*., 2002.

BBC, 2018. *Koreans turn to ice vests and hot soup in record Seoul heat*. News.

Bechtel, B. et al., 2019. Generating WUDAP Level 0 data - Current status of production and evaluation. *Urban Climate*, 27, pp.24-45.

Bechtel, B. et al., 2015. Mapping Local Climate Zones for a Worldwide Database of the Form and Function of Cities. *ISPRS International Journal of Geo-Information* , pp.199-219.

Breiman, L., 1999. *Random forests—random features, Technical Report 567*. Statistics Department, University of California, Berkeley.

Cebeci, Z. & Yildiz, F., 2015. Comparison of K-Means and Fuzzy C-Means Algorithms on Different Cluster Structures. *Journal of Agricultural Informatics*, 6(3), pp.13-23.

Ching, J., 2016. *World Urban Database and Access Portal Tools*. 96th AMS Annual Meeting. American Meteorological Society.

Coll, C., Caselles, V. & Valor, E., 2012. Comparison between different sources of atmospheric profiles for land surface temperature retrieval from single channel thermal infrared data. *Remote Sensing of Environment*, 117, pp.199-210.

Curran, P., 1988. The Semivalqogram in Remote Sensing: An Introduction. *Remote Sensing of Environemnt*, 24(3), pp.493-507.

Davenport, A., Grimmond, S., Oke, T. & Wieringa, J., 2000. Estimating the roughness of cities and sheltered country. In *12th Conf. on Applied Climatology*., 2000.

Dormann, C., McPherson, J., Araujo, M. & Bivand, R., 2007. Methods to account for spatial autocorrelation in thge analysis of species distrubutional data: a review. *Ecography*, pp.609-28.

Dunn, J.C., 1973. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, 3(3), pp.32-57.

Emmanuel, R. & Kruger, E., 2012. Urban heat island and its impact on climate change resilience in a shrinking city: The case of Glasgow, UK. *Building and Environment*, 53, pp.137-49.

EMSD, 2018. *Hong Kong Energy End-use Data*. Govt. of HKSAR.

Geletic, J. & Lehnert, M., 2016. GIS-based delineation of local climate zones:The case of medium-sized Central European cities. *Moravian Geographical Reports*, 24(3), pp.2-12.

Green, W.H., 2008. *Econometric analysis*. 6th ed.

Grira, N., Crucianu, M. & Boujemaa, N., 2005. Active semi-supervised fuzzy clustering for image database categorization. In *Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval*., 2005.

Gunawardena, K.R., Wells, J.M. & Kershaw, T., 2017. Utilising green and bluespace to mitigate urban heat island intensity. *Science of Total Environment*, 584-585, pp.1040-55.

Hammerberg, K., Brousse, O., Martilli, A. & Mahdavi, A., 2018. Implications of employing detailed urban canopy parameters for mesocale climate modelling: a comparison between WUDAPT and GIS databases over Vienna, Austria. *International Journal of Climatology*, pp.1241-57.

Houet, T. & Pigeon, G., 2011. Mapping urban climate zones and quantifying climate behaviors – An application on Toulouse urban area. *Environmental Pollution*, 159(8-9), pp.2180-92.

Kaufman, L. & Rousseeuw, P., 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons.

Kim, H.H., 1992. Urban heat island. *International Journal of Remote Sensing*, 13(12), pp.2319-36.

Kim, Y.H. & Baik, J.J., 2002. Maximum Urban Heat Island Intensity in Seoul. *Applied Meteorology*, 41, pp.651–59.

Knutson, T., Tuleya, & Kurihara, , 1998. Simulated Increase of Hurricane Intensities in a CO2-Warmed Climate. *Science*, 279(5353), pp.1018-21.

Kropko, J. & Kubinec, R., 2018. Why the two way fixed effects model is difficult to interpret, and what to do about it. *Social Science Research Network*.

Leconte, F., Bouyer, J., Claverie, R. & Petrissans, M., 2015. Using local climate zone scheme for UHI assessment: evaluation of the method using mobile measurements. *Building and Environment*, 83, pp.39-49.

Lee, I.H., Ahn, Y.H., Park, J. & Kim, S.D., 2014. District Energy Use Patterns and Potential Savings in the Built Environment: Case Study of Two Districts in Seoul, South Korea. *Asian Journal of Atmospheric Environment*, 8(1), pp.48-58.

Legendre, P., 1993. Spatial Autocorrelation: Trouble or New Paradigm? *Ecology*, 74(6), pp.1659-1673.

Lindberg, F. et al., 2018. Urban Multi-scale Environmental Predictor (UMEP): An integrated tool for city-based climate services. *Environmental Modelling & Software*, 99, pp.70-87.

Li, Z.L. et al., 2013. Satellite-derived land surface temperature: Current status and perspectives. *Remote Sensing of Environment*, 131, pp.14-37.

Li, C., Zhou, J. & Cao, Y., 2014. Interaction between urban microclimate and electric air-conditioning energy consumption during high temperature season. *Applied Energy*, 117, pp.149-56.

Lopes, R., 2011. *International Encyclopedia of Statistical Science*. Berlin: Springer.

Lynch, K. & Rodwin, L., 1958. A Theory of Urban Form. *Journal of the American Institute of Planners*, 24(4), pp.201-14.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Symposium on Math, Statistics,and Probability*., 1967.

Malys, L., Musy , M. & Inard, C., 2015. Microclimate and building energy consumption: study of different coupling methods. *Advances in Building Energy Research*, 9(2).

Memon, R., Leung, D. & Liu, C.-H., 2010. Effects of building aspect ratio and wind speed on air temperatures in urban-like street canyons. *Building and Environment*, 45(1), pp.176-88.

Middel, A., Hab, K. & Brazel, A., 2014. Impact of urban form and design on mid-afternoon microclimate in Phoenix Local Climate Zones. *Landscape and Urban Planning*, 122, pp.16-28.

Mohajeri, N., Kunckler, T., Upadhay, G. & Assouline, D., 2016. How street-canyon configurations affect the potential of solar energy. In *International Conference on Passive and Low Energy Architecture*., 2016.

Nastrana, M., Kobala, M. & Elerb, K., 2018. Urban heat islands in relation to green land use in European cities. *Urban Forestry & Urban Greening*.

NCSS, 2018. *NCSS 12 Documentation*. [Online] NCSS (12) Available at: https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Fuzzy_Clustering.pdf.

Oke, T.R., 1973. City Size and The Urban Heat Island. *Atmospheric Environment*, 7, pp.769-79.

Oke, T.R., 1981. Canyon Geometry and the Nocturnal Urban Heat Island: Comparison of Scale Model and Field Observations. *Journal of Climatology*, 1, pp.237-54.

Oke, T.R., 1982. The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society*, 108(455), pp.1-24.

Oke, T.R., 1988. Street Design and Urban Canopy Layer Climate. *Energy and Building*, 11(1-3), pp.103-13.

Oke, T.R., 2006. *Initial Guidance to Obtain Representative Meteorological Observations at Urban Sites*. World Meteorological Organization.

Oke, T.R., 2006. Towards better scientific communication in urban climate. *Theoretical and Applied Climatology*, 84(1-3), pp.179–90.

Oke, T.R., 2010. Urban Heat Islands. In Douglas, Goode, M. Houck & R. Wang, eds. *The Routledge Handbook of Urban Ecology*. p.120.

Orlowsky, B., 2008. A resampling scheme for regional climate simulations and its performance compared to a dynamical RCM. *Theoretical and Applied Climatology*, 92(3-4), pp.209-23.

Park, C., Ha, J. & Lee, S., 2017. Association between Three-Dimensional Built Environment and Urban Air Temperature:Seasonal and Temporal Differences. *sustainability*, 9, p.1338.

Perera, N.G.R. & Emmanuel, R., 2018. A "Local Climate Zone" based approach to urban planning in Colombo, Sri Lanka. *Urban Climate*, 23, pp.188-203.

ppclust, 2018. *Documentation of ppclust*. [Online] (0.1.1).

Priyadarsini, R., 2009. Urban Heat Island and its Impact on Building Energy Consumption. *Advances in Building Energy Research*, 3(1).

Quan, S.J., Dutt, F., Woodworth, E. & Yamagata, Y., 2017. Local Climate Zone Mapping for Energy Resilience: A Finegrained and 3D Approach. In *The 8th International Conference on Applied Energy*., 2017.

Ratti, C., Baker , N. & Steemers, K., 2005. Energy consumption and urban texture. *Energy and Buildings*, 37(7).

Reyna , O.T., 2007. *Panel Data Analysis Fixed and Random Effects using Stata*. Princeton Univeristy.

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, pp.53-65.

Scikit-Fuzzy, 2016. *Skfuzzy Documentation*. [Online] (0.3) Available at: https://media.readthedocs.org/pdf/scikit-fuzzy/latest/scikit-fuzzy.pdf.

Scikit-learn, 2018. *Scikit-learn user guide*. [Online] (0.20.1) Available at: https://scikit-learn.org/stable/_downloads/scikit-learn-docs.pdf.

Sobrinoa, J., Munoza, J. & Paolini, L., 2004. Land surface temperature retrieval from LANDSAT TM 5. *Remote Sensing of Environment*, 90, pp.434-40.

Soltani, A. & Sharifi, E., 2017. Daily variation of urban heat island effect and its correlations to urban greenery: A case study of Adelaide. *Frontiers of Architectural Research*, 6(4), pp.529-38.

Song, D. & Choi, Y., 2012. Effect of building regulation on energy consumption in residential buildings in Korea. *Renewable and Sustainable Energy Reviews*, 16, pp.1074-81.

Steeneveld, G.J., Koopmans, S., Heusinkveld, B. & Theeuwes, N., 2014. Refreshing the role of open water surfaces on mitigating the maximum urban heat island effect. *Landscape and Urban Planning*, 121, pp.92-96.

Stewart, I.D. & Oke, T.R., 2012. Local Climate Zones for Urban Temperature Studies. *Bulletin of the American Meteorological Society*, 93(12), pp.1879-900.

Stock & Watson, 2007. *Regression with panel data*.

Svensson, M., 2004. Sky view factor analysis – implications for urban air temperature differences. *Meteorological Applications*, 11(3), pp.201-11.

Tan, , Zheng, & Guo, C., 2010. The urban heat island and its impact on heat waves and human health in Shanghai. *International Journal of Biometeorology*, 54, pp.75–84.

Unger, J., 2009. Connection between urban heat island and sky view factor approximated by a software tool on a 3D urban database. *International Journal of Environment and Pollution*, 36, p.59.

USGS, 2015. *TIRS Stray Light Correction Implemented in Collection 1 Processing*. [Online] Available at: http://phenology.cr.usgs.gov/ndvi_foundation.php [Accessed March 2016].

USGS, 2018. *Landsat 8 (L8) Data User Handbook*. Department of the Interior, USGS.

Valor, E. & Caselles, V., 1996. Mapping land surface emissivity from NDVI: application to European, African, and South American areas. *Remote Sensing of Environment*, 57, pp.167-84.

Vu, T. & Parker, D., Preprint. Inferring Microclimate Zones from Energy Consumption Data. *arXiv (Unpublished)*.

Wagstaff, K., Cardie, C., Rogers, S. & Schroedl, S., 2001. Constrained K-means Clustering with Background Knowledge. In *Proceedings of the Eighteenth International Conference on Machine Learning*., 2001.

Yaun, F. & Bauer, M., 2007. Comparison of impervious surface area and normalized difference vegetation index as indicators of surface urban heat island effects in Landsat imagery. *Remote Sensing of Environment*, 106, pp.375-86.

Yin, X., Shu, T. & Huang, Q., 2012. Semi-supervised fuzzy clustering with metric learning and entropy regularization. *Knowledge-Based Systems*, 35, pp.304-11.

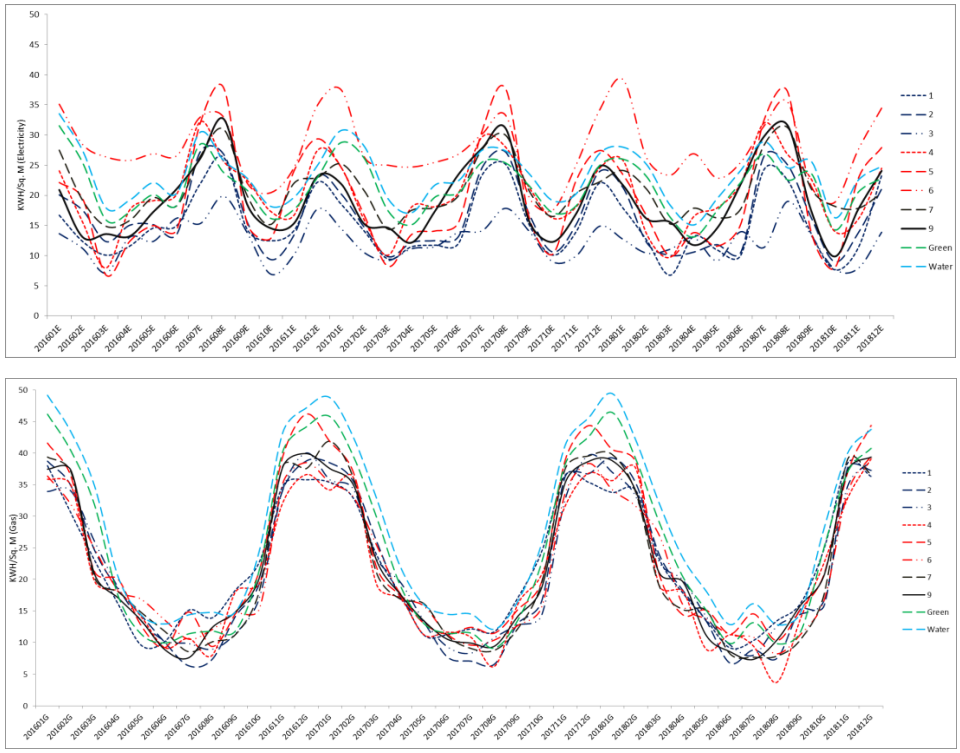
Youngsoo, Y. & Saehoon, K., 2018. Revealing the Mechanism of Urban Morphology Affecting Residential Energy Efficiency in Seoul, Korea. *Sustainable Cities and Society*, 43, pp.176-90.

Zhang, C., Luo, L., Xu, W. & Ledwith, V., 2008. Use of local Moran's I and GIS to identify pollution hotspots of Pb in urban soils of Galway, Ireland. *Science of The Total Environment*, 398(1-3), pp.212-21.

Zhao, C., 2018. Linking the local climate zones and land surface temperature to investigate the surface urban heat island, a case study of san antonio, texas, US. In *ISPRS TC III Mid-term Symposium "Developments, Technologies and Applications in Remote Sensing*., 2018.

Zheng, Y. et al., 2018. GIS-based mapping of Local CLimate Zone in the high-density city of Hong Kong. *Urban Climate*, 24, pp.419-48.

Appendix 1. Trends in electricity and gas consumption between 2016 -18. Mean consumption was calculated for each LCZ for preliminary investigation into the difference in consumption by LCZ classes. This raw data formed the basis of panel data analysis to calculate the relationship between BEC and urban context of the building.



Appendix 2: Coefficients of all dummy variables in the fixed panel data analysis described in section 4.2.1

Month	LCZ	Coef_E	Sig_E	Coef_G	Sig_G
Jan	1	2.642	0.000	-1.051	0.004
	2	2.099	0.000	-1.133	0.002
	3	-1.933	0.028	-1.030	0.023
	4	4.5415	0.020	4.753	0.023
	5	0.4068	0.097	3.643	0.036
	6	-0.4536	0.147	2.523	0.002
	7	-2.3593	0.029	0.124	0.294
	12	-2.2319	0.027	1.943	0.020
	13	0.1877	0.108	1.435	0.014
Feb	1	1.2826	0.022	-1.012	0.004
	2	1.3729	0.028	-1.112	0.002
	3	-3.2352	0.016	-1.093	0.023
	4	4.6457	0.012	4.121	0.040
	5	0.8565	0.182	3.144	0.025
	6	-0.5792	0.171	2.101	0.016
	7	-2.1929	0.015	0.155	0.371
	12	-1.8577	0.010	1.564	0.027
	13	-1.411	0.011	-1.965	0.035
Mar	1	2.7762	0.025	-1.078	0.004
	2	1.4004	0.019	-1.087	0.002
	3	-2.1349	0.022	-1.011	0.023
	4	3.1398	0.025	3.124	0.016
	5	0.844	0.164	2.311	0.002
	6	-0.1244	0.117	1.923	0.008
	7	-1.8223	0.029	0.035	0.181
	12	-2.5812	0.015	1.753	0.005
	13	1.1974	0.019	-1.346	0.009
Apr	1	2.4794	0.027	0.212	0.092
	2	3.9852	0.030	-0.063	0.131
	3	-2.1888	0.015	-0.058	0.295
	4	5.0117	0.022	2.900	0.022
	5	0.3298	0.140	2.323	0.039

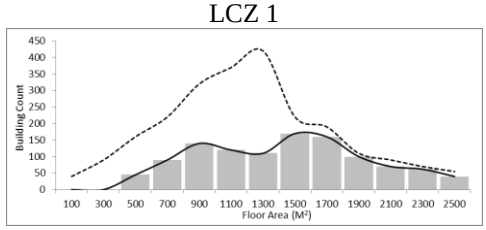
Month	LCZ	Coef_E	Sig_E	Coef_G	Sig_G
July	1	3.257	0.027	0.019	0.131
	2	1.2123	0.020	0.065	0.289
	3	-1.3105	0.025	0.161	0.139
	4	5.0907	0.023	-0.035	0.272
	5	-1.5986	0.016	-0.062	0.171
	6	-1.627	0.017	-0.100	0.246
	7	1.5172	0.028	0.026	0.119
	12	-0.5	0.166	-0.088	0.151
	13	-2.219	0.019	0.033	0.103
Aug	1	1.7758	0.016	-0.015	0.220
	2	3.7619	0.018	-0.002	0.343
	3	-2.647	0.024	-0.060	0.342
	4	4.8401	0.012	0.063	0.229
	5	-1.9215	0.021	0.017	0.266
	6	-3.5418	0.012	0.124	0.138
	7	1.5065	0.024	0.152	0.381
	12	-1.6927	0.022	0.051	0.196
	13	1.8581	0.020	0.076	0.314
Sep	1	3.7226	0.024	0.168	0.118
	2	1.5981	0.020	0.056	0.193
	3	-1.4557	0.027	0.199	0.394
	4	5.4188	0.029	0.021	0.382
	5	-1.7272	0.011	0.127	0.157
	6	-2.6842	0.028	-0.031	0.399
	7	1.8194	0.018	0.092	0.078
	12	0.189	0.085	0.084	0.135
	13	-2.7932	0.018	-1.654	0.012
Oct	1	3.0328	0.021	0.185	0.174
	2	1.04	0.022	0.109	0.129
	3	-2.4648	0.011	-0.020	0.133
	4	5.5555	0.023	1.566	0.006
	5	0.7626	0.144	1.325	0.007

	6	-1.1131	0.049	1.864	0.014
	7	2.4962	0.027	-0.012	0.104
	12	0.5282	0.135	0.5282	0.121
	13	-0.9832	0.014	-0.9832	0.021
May	1	3.1363	0.026	-0.036	0.266
	2	1.7865	0.026	-0.052	0.346
	3	-2.7946	0.030	-0.065	0.266
	4	4.0409	0.015	0.048	0.135
	5	0.8178	0.151	0.194	0.398
	6	-2.8605	0.014	0.042	0.215
	7	1.5712	0.015	0.053	0.342
	12	1.3258	0.027	0.077	0.254
	13	2.3269	0.026	-0.013	0.215
Jun	1	1.9828	0.015	0.146	0.377
	2	1.2538	0.024	-0.078	0.213
	3	-2.5086	0.020	0.165	0.196
	4	4.813	0.021	-0.035	0.283
	5	-1.2426	0.015	-0.045	0.189
	6	-1.6133	0.028	-0.037	0.121
	7	1.3969	0.020	0.073	0.141
	12	-1.7718	0.028	0.188	0.359
	13	-2.53	0.011	0.024	0.275

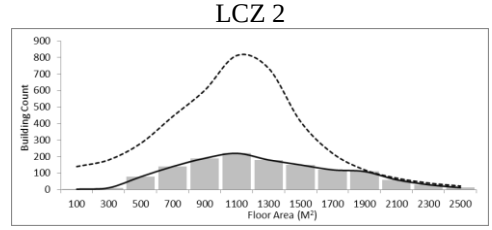
	6	-3.5465	0.030	0.343	0.088
	7	-1.5184	0.018	0.141	0.306
	12	-1.1954	0.029	1.564	0.021
	13	0.2122	0.152	0.106	0.377
Nov	1	2.8092	0.018	-1.073	0.004
	2	3.3559	0.024	-1.129	0.002
	3	-2.0245	0.028	-1.006	0.023
	4	4.4252	0.029	2.145	0.031
	5	0.2332	0.163	1.932	0.037
	6	-0.3117	0.095	1.412	0.040
	7	-2.0725	0.024	-0.091	0.226
	12	1.7437	0.019	1.743	0.019
	13	-1.1127	0.012	2.453	0.024
Dec	1	2.8491	0.029	-1.066	0.004
	2	2.2432	0.024	-1.077	0.002
	3	-2.8499	0.023	-1.008	0.023
	4	5.3742	0.012	3.453	0.035
	5	0.8372	0.107	2.164	0.009
	6	-0.1808	0.075	1.789	0.004
	7	-2.7175	0.019	0.049	0.316
	12	2.727	0.014	1.761	0.005
	13	0.2901	0.175	2.112	0.009

Appendix 3: Distribution of building floor area for each LCZ in comparison with buildings for which electricity and gas data is available. The difference exists because BEC data is only released for *larger* buildings. The graphs indicate that data used in the model described in section 4.2. and 4.3 might only be biased towards larger buildings.

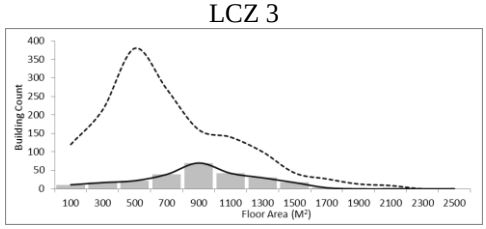
LCZ 1



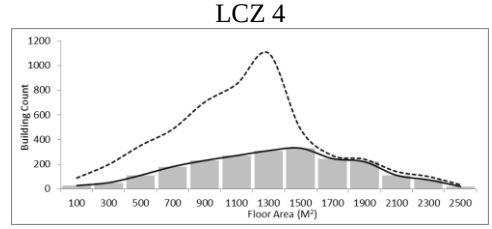
LCZ 2



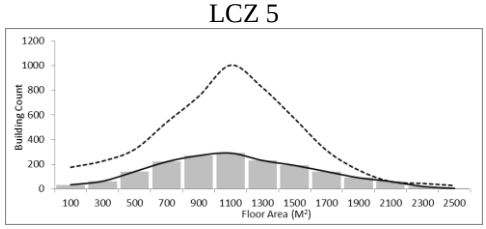
LCZ 3



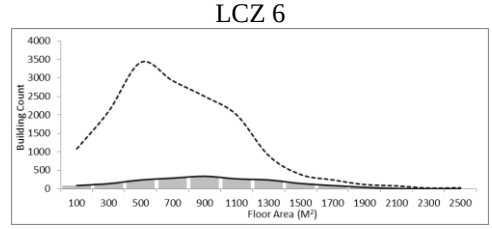
LCZ 4



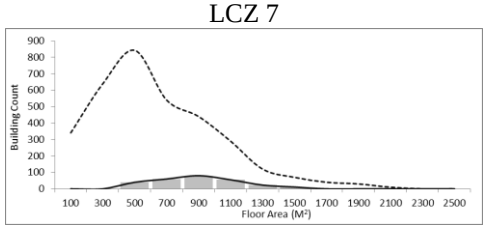
LCZ 5



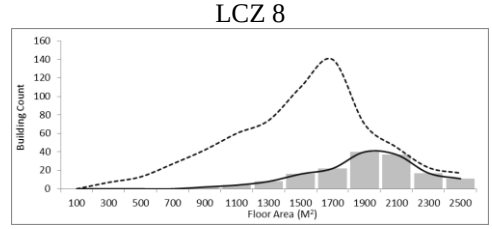
LCZ 6



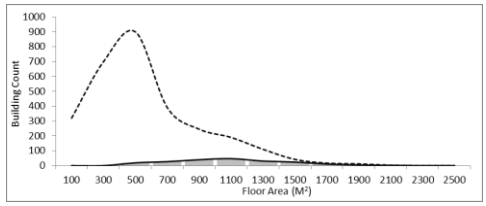
LCZ 7



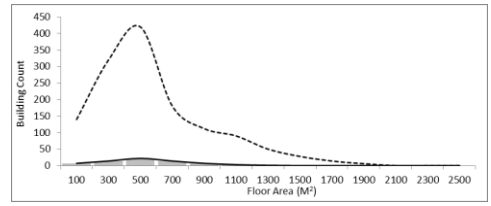
LCZ 8



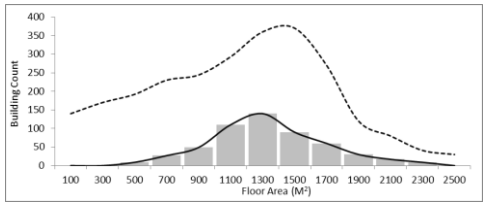
LCZ 9



LCZ Green



LCZ Water



Buildings with BEC data

All Buildings

Appendix 4. Spatial Autocorrelation (SA)

In here, the reasoning for deploying independent variable 'SA_Control' in model 3.2.1 and 3.3.2 is discussed. Firstly, a brief introduction of spatial autocorrelation and its implication in modeling are discussed. Next, it is shown why some type of spatial autocorrelation control measure was required and what options exists for it.

Generalized linear regression (GLM) assumes that each observation is independent and identically distributed (i.i.d). If the assumption holds true, residual after modeling will also have i.i.d. In simpler terms, it is assumed that an observation has no effect on another observation. If and when this assumption is violated, parameter estimates can be biased and type 1 error rate (rejecting the null hypothesis falsely) increases. (Dormann et al., 2007).

However, in geographical datasets, such as in this study, it is often found that geographically closer observations are more similar to each. For example, in model 4.2.1 and figure 4-7 it may be possible that those grids next to each other have similar energy consumption as compared to grid which is far away. This obviously violates the i.i.d assumption described above. To test if spatial autocorrelation plays significant role in modeling in section 4.2 and 4.3, we can use Moran's I, similar to that described in section 4.1.2 to determine grid size. However, instead of using observation data as in the case of section 4.1.2, we use residual of electricity and gas model respectively, before and after spatial autocorrelation term is used. The results are shown in figure below.

It is also important to describe what 'spatial autocorrelation term' / 'SA_Control' means. As summarized by Dormann et al. (2007) if spatial autocorrelation is playing a significant role in GLMs, following alternative exists.

1) Additional Covariates: Add a new variable which can account for SA.
    a. Autocovariate
    b. Spatial Eigenvector
2) Spatially Explicit Generalized Least Square (SE-GLS): Use SE-GLS instead of OLS to fix variance-covariance matrix.
    a. Simultaneous autoregressive models
    b. Conditional autoregressive models
3) Generalized Estimating Equations (GEE): Split data into cluster and then model each cluster independently.

In this study, we used autocovariate term (1a) where a distance-weighted function of neighboring values is added as independent variable. The reason for this is:

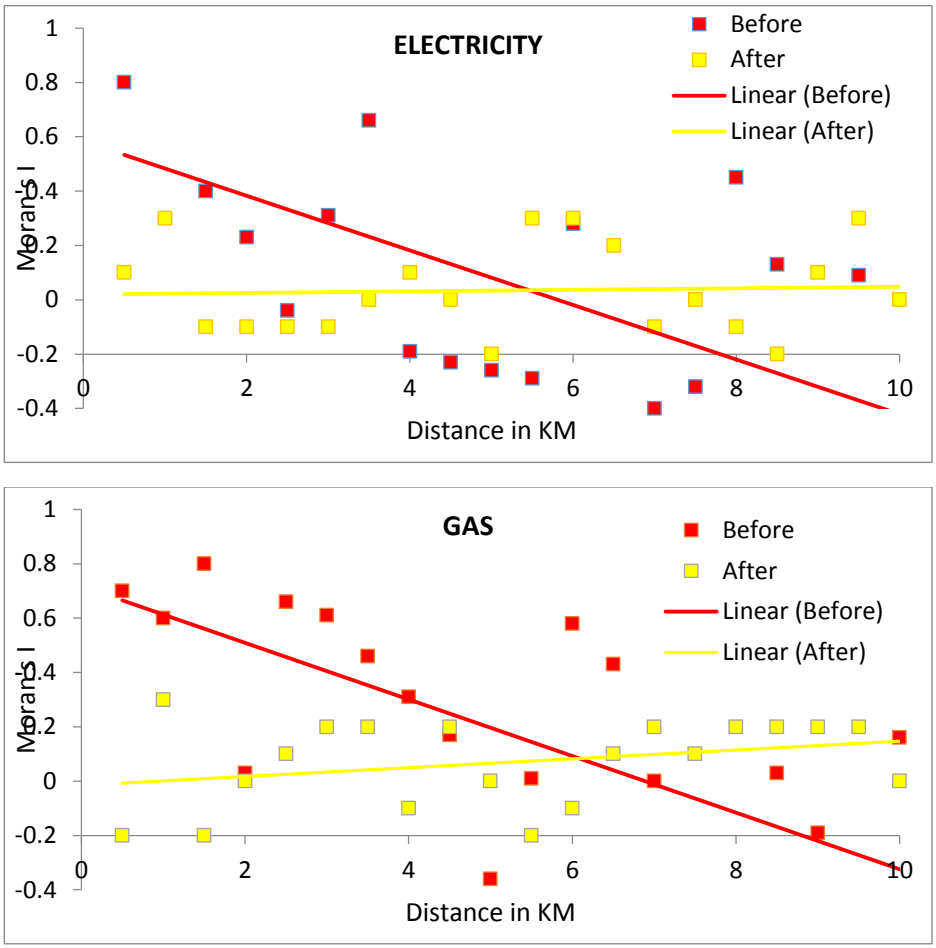
1) Additional covariates can easily be incorporated into simple regression and panel data analysis. Spatially explicit modeling requires specialized

software. Currently, to the author's knowledge, no software or library provide out of the box ability for spatially explicit panel data analysis.

2) Between autocovariate (1a) and Spatial Eigenvector (1b), 1a is easier to compute and can be done using 'generate spatial weights matrix' tool in ArcGIS.

However, it should be noted that Dormann et al. (2007) has shown that this method substantially underestimate the SA bias and other more complex method should be preferred.

From figure below, for both electricity and gas models, it is clearly seen that residual from the model without 'SA_control' are strongly correlated. However, after autocovariate term was added as independent variable, the correlation is drastically reduced and close to *ideal* Moran's I = 0 in both cases.



Spatial Autocorrelation in models residual before and after autocovariate independent variable is added to panel data model.

초 록

에너지 효율적인 개발에 중점을 두면서 건축물 에너지 효율성에 영향을 미치는 도시형태의 역할이 최근 많은 연구자들의 관심 영역이 되었다. 하지만 도시형태의 에너지 효율성에 관한 연구는 명확하게 구분되지 않고 주관적인 정의와 공통적인 방법론에 국한되여 있다.

이와 동시에, 도시 열섬현상과 미기후를 연구에서 도시 기후학 분야에서 개발된 지역 기후 구역(LCZ) 프레임워크는 도시형태의 분류 방법론과 적용 모두에서 주목을 받고 있다. LCZ 프레임워크는 도시와 자연적 특징을 바탕으로 도시형태를 17 개의 클래스로 분류하며, 기존 연구에서는 각 클래스에서 특정한 공기온도와 표면온도를 나타내고 있음을 보여준다.

본 연구에서는 미기후가 건축 에너지 효율성에 미치는 도시형태의 주요 원인이라는 가설을 바탕으로 LCZ 클래스와 건물 전기 및 가스 소비 사이의 관계를 연구한다. 첫째, 도시형태를 결정하는 적절한 단위는 LCZ 매개변수의 공간적 자기 상관관계 및 분포 곡선을 이용하여 (수정 가능한 면적 단위 문제를 최소화하기 위해) 주의하여 설정한다.

다음, 선행연구에서 설명한 바와 같이, 일부 도시형태의 이질적인 특성 때문에 어떤 LCZ 클래스에도 속하지 않는다. 이를 해결하기 위해, 본 연구는 명확하지 않은 도시형태 영역에 대해 가장 유사한 LCZ 클래스를 식별하기 위한 인공지능 기계 학습 알고리즘의 사용하였다. 알고리즘 적용 결과의 적합성을 위하여 표면 온도데이터와 비교한다.

마지막으로, 서울의 이러한 LCZ 클래스와 건물 전기 및 가스 소비 데이터(2015-18)를 사용하여, 도시적 환경이 전혀 없는 낮은 밀도 건축 면적에 비해 LCZ 클래스별로 에너지 소비량이 어떻게 달라지는지 분석한다. 이를 위해 도시형태의 영향은 광범위한 계절 조건에 따라 달라질

수 있으므로 LCZ 클래스와 시간의 효과를 동시에 모델링할 수 있도록 양방향 고정 효과 패널 데이터 분석 기법을 사용한다. 본 연구는 전반적으로, '넓은 부지의 고층건물 지역'과 '넓은 부지의 저층건물 지역'이 에너지 효율이 높은 유일한 도시 LCZ 클래스라는 것을 보여준다. 나머지 등급은 낮은 건설밀도 부지에 비해 에너지 소비량이 높은 것으로 보인다(건축부지 1 ㎡ 당).

마지막으로, 용도지구와 LCZ 클래스과 비교하여 어떤 구역제 매개 변수가 특정 LCZ 클래스를 초래하는지 결정한다. 하여 본 연구는 도시 계획가들이 에너지 효율성이 높은 도시 형태를 결정하는데 도움을 준다. 지금까지 에너지 효율성 달성에 초점을 맞춘 것은 건축 기술이나 교통 에너지 수요의 향상에 국한되어 왔지만, 본 연구를 통해 가장 에너지 효율성이 높은 도시형태를 파악하는 것이 가능하기를 희망한다.

Acknowledgements

Professor Steven, thank you for guidance and listening to my ideas. Most of them never materialized, but I am grateful for your patience and direction.

Professor Tommy Gim and Youngkeun Song, thank you for your feedback. Your opinion and criticism helped me to immensely improve this research.

Lina and all of my lab mates, thank you for your support and feedback.

For Ivy and Oreo, my two little guinea pigs.

None of this would have been possible without anonymous users of Stackoverflow and GIS Stackexchange who berated my questions, but answered them nonetheless.

This work was partially supported by the National Research Foundation of Korea (NRF) grant (NRF-2018R1C1B5043758).

-Parth