

The Pennsylvania State University

The Graduate School

**A DATA MINING MODEL AND A REAL-TIME PREDICTIVE SOFTWARE
PROTOTYPE FOR THE SPATIAL DESIGN AND PLANNING OF HIGH
ENERGY PERFORMANCE SOLAR COMMUNITY MICROGRIDS**

A Dissertation in

Architecture

by

Mina Rahimian

© 2022 Mina Rahimian

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2021

The dissertation of Mina Rahimian was reviewed and approved by the following:

José P. Duarte

Professor of Architecture, Stuckeman Chair in Design Innovation
Dissertation Co-Advisor, Chair of Committee

Lisa D. Iulo

Associate Professor of Architecture
Dissertation Co-Advisor

Guido Cervone

Assistant Professor of Geography

Seth Blumsack

Assistant Professor of Energy and Mineral Engineering

Ute Poerschke

Professor of Architecture, Associate Department Head for Graduate Education

ABSTRACT

With severe natural disasters occurring around the globe, cities are experiencing the consequences of climate change more than before. Frequent power outages attributed to aging equipment of power distribution systems and coupled with natural disasters such as hurricanes or wildfires are threatening everyday lives and businesses of urban dwellers. Communities, especially urban communities, which have experienced frequent blackouts are taking a closer look at adopting microgrid technologies to operate independently from the main power grid during emergencies. Microgrids are local, decentralized power distribution systems involving the use of power sources such as solar panels and diesel engines and storage devices like batteries to provide electricity for a cluster of buildings. Providing resiliency and reliability under unexpected power interruptions, microgrids have typically been used for backup supporting critical loads such as military bases and hospitals. However, with the increasing environmental concerns associated with fossil fuels and the frequency of natural disasters, a growing interest in adopting microgrid technologies is rising in towns and communities in the interest of transitioning into energy-independent urban settlements. Known as community microgrids, these energy-independent urban settlements are generally comprised of various mixes of residential, commercial, agricultural, and industrial loads followed by the local renewable and/or nonrenewable sources of power.

As with any other energy system, the efficiency of a community microgrid's energy performance is evaluated by comparing the energy inputted to the system from the on- and off-site sources of energy, to the energy that is outputted from the system, mostly in the form of useful energy for buildings operation. Current research on improving energy performance in community microgrids has been exclusively advocating technological advances enhancing the limited supplies of local energy and addressing the constantly growing demands of the loads. However, researchers argue that focusing on technological innovations alone wouldn't solve the current energy issues in the built environment. In the case of community microgrids this statement is especially accurate since they are contextualized in cities and urban areas; citing research from the 1960's onwards, considerable attention has been directed towards the impact that spatial structure of urban form has on the energy required for space heating, cooling and lighting as well as the feasibility of adopting on-site renewable energy generators such as Photovoltaic (PV) panels and wind turbines.

Literature today emphasizes the importance of obtaining an energy-conscious point of view when taking actions toward urban design and planning. Architects and urban planners are expected to consider the tradeoffs between the living qualities of an urban context and its potential for high-performance energy systems design and engineering. Despite the evident need for involving architects and urban planners in the development process of urban

energy systems as community microgrids, in practice this engagement is generally neglected. This is possibly due to the complexity of understanding urban form and its impact on energy performance in community microgrids and the unavailability of custom tools for the spatial design and assessment of these energy systems. The intention of this research is to engage architects and urban planners in the process of developing and constructing community microgrids. Benefiting from artificial neural networks, this research adds a spatial dimension to the existing technical discourse of developing high energy performance community microgrids and by surrogate modeling, delivers a real-time energy simulation software prototype that aids architects and urban planners in designing and assessing urban scale energy systems.

This dissertation focuses on solar community microgrids with San Diego county serving as a case study for data production. In this research artificial neural networks are used to uncover the complex relationship between San Diego's urban form and its impact on community scale energy consumption and PV energy production as the two main pillars of evaluating energy performance. Results of the training procedure proves the existence of a strong statistical relationship between urban form and energy performance in communities. By architecturally and spatially interpreting the results of the outputted statistical model, a comprehensive set of design principles were extracted from this analysis that guides architects and urban designers towards spatially designing high energy performance community microgrids in San Diego. For instance, to spatially design a high energy performance community microgrid in the coastal region of San Diego, the distance between the community buildings needs to be maximized to prevent overshadowing adjacent buildings and minimize the community's net energy demand for heating.

A spatially aware design and assessment of solar community microgrids as studied herein brings the need for a new generation of computational modeling, simulation, and evaluation tools for the field. In this regard, an urban scale energy simulation software prototype is developed that predicts the energy performance of any given solar community microgrid design scenario - in real-time - by the virtue of its urban spatial configuration. The real-time prediction feature of this software (which is due to benefiting from the trained machine learning models at its backend rather than hard coding physical laws of energy) is a major contribution to the field and sets an example for future simulation software development, since current existing urban scale energy simulation tools are extremely time and resource consuming.

Table of Contents

List of Figures.....	vii
List of Tables	xi
Acknowledgment.....	xii
Chapter 1 Research Description	1
1.1 Introduction	1
1.2 Problem Statement.....	4
1.3 Hypothesis	5
1.4 Research Questions.....	6
1.5 Methodology Overview	7
1.5.1 Urban Form and Energy Consumption.....	10
1.5.2 Urban Form and PV Energy Production.....	10
1.5.3 Final Deliverables.....	10
1.6 Contributions and Findings	11
1.7 Dissertation Roadmap.....	12
Chapter 2 Microgrids	14
2.1 Microgrid Definition.....	14
2.2 Microgrid Components.....	15
2.3 Different Categorizations of Microgrids	17
2.4 Benefits and Values of Microgrids.....	19
2.5 Microgrid Development Process	20
Chapter 3 A Software Survey.....	23
3.1 Background and Evaluation Criteria	23
3.2 Urban Superstructure Energy Modeling.....	24
3.3 Energy Infrastructure Energy Modeling.....	27
3.4 Discussion.....	31
Chapter 4 Urban Form and Energy in Community Microgrids.....	34
4.1 A Brief Background on Urban Form and Energy in Communities.....	34
4.2 Identification of Urban Form Features Relevant to Energy Performance of Solar Community Microgrids	36
Chapter 5 Community Energy Consumption and Urban Form: A Machine Learning Model	43
5.1 Why Machine Learning and Artificial Neural Networks?	43
5.2 Preparing the Dataset.....	44

5.2.1 Data Collection	44
5.2.2 Data Processing	45
5.3 Training an Artificial Neural Network on the Dataset	49
5.3.1 Data Preparation	50
5.3.2 Model Building Using Artificial Neural Networks	58
5.3.3 Evaluation and Prediction.....	62
Chapter 6 Community Energy Consumption and Urban Form: A Statistical and Architectural Interpretation.....	68
6.1 Statistical Interpretation.....	68
6.2 Spatial and Architectural Interpretation	71
Chapter 7 Community PV Energy Production and Urban Form.....	79
7.1 A Generative Urban Design Algorithm.....	80
7.1.1 Generative Design	80
7.1.2 Decoding Urban Typologies in San Diego: A Shape Grammar Approach.....	81
7.2 Adding PV Energy Simulation Properties to the Generative Tool.....	108
Chapter 8 A Realtime Predictive Software Prototype	110
8.1 Surrogate Modeling for Short-Circuiting Simulation.....	110
8.2 Preparing Surrogate Models for Estimating Monthly Values of Energy Consumption	113
8.3 Implementation and Software Architecture.....	120
Chapter 9 Conclusion.....	124
9.1 Summary of Research.....	124
9.2 Research Discussion, Contributions, and Limitations.....	128
9.3 Next Steps and Future Directions	131
Glossary	133
References.....	135

List of Figures

Figure 1 - A microgrid connected community within the urban context, with clearly defined energy and regional boundaries that can disconnect from the larger grid and operate in island mode.....	3
Figure 1 - Community microgrids comprise of an infrastructure and superstructure.....	5
Figure 3 - A sample output of CitySim (Source: https://plot.ly/~ClaytonMiller/112.embed)	25
Figure 4 - UMI output sample (Source: https://architecture.mit.edu/event/modeling-urban-sustainability-energy-daylight-and-walkability)	26
Figure 5 - DER-CAM output sample (Source: https://building-microgrid.lbl.gov/projects/der-cam)	27
Figure 6 - PV-Design Pro output sample (Source: http://www.mauisolarsoftware.com/)	30
Figure 7 - PV*SOL 3D building modeling sample (Source: http://solarprofessional.com/articles/design-installation/performance-modeling-tools-overview/page/0/2)	31
Figure 8 - PV*SOL 3D visualization of PV energy evaluation (Source: http://www.valentin-software.com/en/products/photovoltaics/57/pvsol-premium)	31
Figure 9 - Measurements of urban form per zip code (dataset header)	47
Figure 10 - Original energy consumption dataset acquired from SDG&E (dataset header)	47
Figure 11 - Preprocessed energy dataset (dataset header)	47
Figure 12 - Synthesized dataset (dataset header)	48
Figure 13 - Pair plot visualizing the relational pattern between every two urban form variables	49
Figure 14 - San Diego cloud coverage from 2012 to 2018	50
Figure 15 - Loading the dataset and dropping unnecessary columns	51
Figure 16 - Initializing the RFE model and using LinearRegression for training on the initial set of features and obtaining the importance of each feature based on rfe.support and rfe.ranking	53
Figure 17 - Code for the loop that retrieves the optimum number of features yielding a model with highest accuracy	53
Figure 18 - Code retrieving the selected features	53
Figure 19 - Finding the best performing features using backward elimination	54
Figure 20 - Selected features by fitting a Lasso model to the dataset	55
Figure 21 - Code for separating the response variables from the predictor variables	57
Figure 22 - Code for normalizing the dataset	57
Figure 23 - Left plot shows the normalized X and the right plot shows the normalized Y...57	57
Figure 24 - Splitting the train data into training and validation	57
Figure 25 - Defining the architecture of the artificial neural network	60

Figure 26 - Compiling the model	61
Figure 27 - Fitting the model with 500 epochs. The result of the first 5 epochs is shown here	61
Figure 28 - Value history of the last 5 epochs	61
Figure 29 - The code for plotting the learning curve. The top left plot shows the performance of model trained on the dataset derived from the RFE feature selection method. Top right is for the dataset from the backward elimination feature selection method. Bottom left is for the dataset from lasso feature selection method	62
Figure 30 - Comparing the predicted values for the validation dataset with its true values. The top left plot belongs to the model trained on the dataset derived from the RFE feature selection method. Top right is for the dataset from the backward elimination feature selection method. Bottom left is for the dataset from the lasso feature selection method...	63
Figure 31 - Comparing the predicted values for the validation dataset with its true values. The left plot belongs to the model trained on the dataset derived from the RFE feature selection method. The middle plot is for the dataset from the backward elimination feature selection method. The right one is for the dataset from lasso feature selection method	63
Figure 32 - Comparing the distribution of errors for the three models. The top left plot belongs to the model trained on the dataset derived from the RFE feature selection method. Top right is for the dataset from the backward elimination feature selection method. Bottom left is for the dataset from lasso feature selection method	64
Figure 33 - Four plots showing the performance of the model training	66
Figure 34 - Plot showing the model's performance on test dataset	66
Figure 35 - Visualization showing the features impact on model output	69
Figure 36 - The magnitude hierarchy that each of the urban form features has on energy consumption in this study	70
Figure 37 - Plot showing the mean absolute value of the SHAP values for each feature....	71
Figure 38 - Different climate zones in San Diego county. Source: waternewsnetwork.com	72
Figure 39 - 3D visualization of the simulated relationship between average SVF, neighborhood's average building height, and average density - Source: Mirzaee et al (2018)	73
Figure 40 - Urban heat island index for San Diego county - Source: California Environmental Protection Agency (https://calepa.ca.gov/climate/urban-heat-island-index-for-california/urban-heat-island-interactive-maps/)	74
Figure 41 - A portion of San Diego's map from 1979. Source: www.sunnycv.com	82
Figure 42 - Samples of identified three urban typologies: (a) organic grid, (b) mixed grid, (c) orthogonal grid	83
Figure 42 - The selected region in this image represents Ls boundary which is a clipped selection of a region which its inner area is flat with minimum steepness suitable for an	

orthogonal grid construction. This means that the immediate area outside of Ls could all be flat (then Ls is entirely made of straight lines) or could all be steeped and surrounded by mountains and valleys (then Ls is entirely made of curved lines) or could be a mix of flat and steeped (then Ls is made of straight and curved lines)	85
Figure 44 - Identify viewsheds	86
Figure 45 - Drawing the main road axis towards directing the viewsheds. AN is the collection of road axes	86
Figure 46 - Assigning major street width to the main axes. The width dimension is adopted from San Diego zoning requirements	86
Figure 47 - Rules for generating the compositional structure of the orthogonal street grid.	87
Figure 48 - The width of horizontal streets depends on the location of Ref	88
Figure 49 – The width of horizontal streets depends on the location of Ref (continued) ...	89
Figure 50 – The width of vertical streets depends on the location of Ref	90
Figure 51 – The width of vertical streets depends on the location of Ref (continued)	91
Figure 52 - Rule for trimming axes outside the Ls boundaries	92
Figure 53 - Contour code 0 representing edge of a steeped slope. Source: getoutside.ordnancesurvey.co.uk	92
Figure 54 - First grammar from top is for finding the concave segment of the t contour line. Second grammar is for finding and measuring the concave area. The bottom two grammars are used for comparing the size of the concave area to the area of one urban block	94
Figure 55 - Grammar for comparing the size of the inscribed circle to the width of one urban block when $\text{Area}_{\text{Concave}} > \text{Block}_{\text{Area}}$	95
Figure 56 - Grammar for comparing the size of the inscribed circle to the width of one urban block when $\text{Area}_{\text{Concave}} \leq \text{Block}_{\text{Area}}$	96
Figure 57 - Rules for inserting blocks to cells	97
Figure 58 - Rules for adjusting blocks that fall outside the cell boundaries	98
Figure 59 - Rules for adjusting blocks that fall outside the cell boundaries (continued) ...	99
Figure 60 - Rule for erasing small blocks	99
Figure 61 - Rules for combining blocks to create larger blocks	100
Figure 62 - Complete rectangular blocks are divided into 24 uniform rectangular lots ...	100
Figure 63 - Mapping a grid of 24 urban lots onto non-rectangular blocks and selecting those lots completely inside the blocks	101
Figure 64 - Dividing non-rectangular urban blocks into smaller ones by recursively offsetting the Ls border or the coincident ax street	102
Figure 65 - Dividing non-rectangular urban blocks into smaller ones by recursively offsetting the Ls border or the coincident ax street (continued)	103
Figure 66 – Process of creating urban lots in non-rectangular urban blocks: Identifying the edge of the urban block adjacent to Ls or ax, identifying the edge facing (in front of) the first identified edge, dividing both edges into segments with a preset length (the width of	

an urban lot) starting from the same end of the edge, connecting the division points one by one	104
Figure 67 - Rules for combining urban lots to make larger ones	104
Figure 68 - Deleting axes after all block and lots have been created	104
Figure 69 - The Grasshopper + Python script of the generative tool which is based on San Diego's urban planning principles and urban shape grammars	105
Figure 70 - Two samples of community designs generated by the generative tool	107
Figure 71 – January's learning curve (left) and prediction accuracy plot (right)	116
Figure 72 - February's learning curve (left) and prediction accuracy plot (right)	116
Figure 73 - March's learning curve (left) and prediction accuracy plot (right)	116
Figure 73 - April's learning curve (left) and prediction accuracy plot (right)	116
Figure 75 - May's learning curve (left) and prediction accuracy plot (right)	117
Figure 76 - June's learning curve (left) and prediction accuracy plot (right)	117
Figure 77 - July's learning curve (left) and prediction accuracy plot (right)	117
Figure 78 - August's learning curve (left) and prediction accuracy plot (right)	117
Figure 79 - September's learning curve (left) and prediction accuracy plot (right)	118
Figure 80 - October's learning curve (left) and prediction accuracy plot (right)	118
Figure 81 - November's learning curve (left) and prediction accuracy plot (right)	118
Figure 82 - December's learning curve (left) and prediction accuracy plot (right)	118
Figure 83 - Plot showing the model's performance on test dataset for the month of September	119
Figure 84 - A picture screen of the output of the software. The urban setting that was created by the tool is shown on the right, its values of urban form is shown on top left, and the predicted values of energy consumption is shown on bottom left	121
Figure 85 - The prototype's software architecture	122

List of Tables

Table 1 – Sumamry table of the conducted software survey	32
Table 2 - Selected attributes of urban form along with their index and metric	42, 46
Table 3 - Comparing the selected high performing features by each of the feature selection methods	56
Table 4 - Comparing the mean squared error regression loss for all three models.....	64
Table 5 - Features common among all three tested datasets	65
Table 6 – Original and modified variables in the ANN architecture after retraining process and its resulting MSE	115

Acknowledgment

I would like to take this opportunity to thank the following persons for being there for me during my PhD journey. These people filled me with their joy, encouragement, support, and generously offered me their advice and help during my time at Penn State.

First and foremost, I would like to thank my two amazing co-advisors which their friendship and encouragement made my doctoral studies among the most memorable years of my life which I will cherish forever. Jose P. Duarte and Lisa D. Iulo, your knowledge and guidance shed light on my research path and provided clarity to my thoughts and words. What I learnt from you was beyond computational design and sustainability; I learnt how to be there for those who rely on you and how to be a better person. You two defined “support” all over again. I could’ve not asked for better co-advisers, ever!

I would like to express my gratitude to Guido Cervone and Seth Blumsack. Knowing and working with you as members on my committee has brought me so many lessons. Your intellectual rigor and your insights and contributions to your fields of expertise has definitely added to my scholarship.

I am very grateful to Mehrdad Hadighi, our department head. Mehrdad, your leadership, graciousness, and drive made the Stuckeman School my second home. Since the first day I started my studies at the Stuckeman School as a master student, I knew that I wanted to continue my PhD there as well. Your thoughtfulness has made our department one BIG family and I’m very grateful for that. I learnt how to be a leader from you!

To my friends at the Stuckeman School, at SCDC, the Hamer Center and the PhD suite, how grateful I am to all of you! I will never forget our friendship and our 12pm lunches that was always accompanied with laughter. Thank you all for your hugs, your smiles, your enthusiasm, and your amazingness in your scholarship. Amongst my homies at SCDC I would first like to name Debora Verniz, which was also my roommate – the best one in fact! Our friendship goes beyond words. We’ve shared so many moments together, both in State College and in Lisbon that I will forever keep close to my heart. See? We made it! My other friends at SCDC and the Stuckeman School (from both my masters and PhD) which I’d like to express my gratitude are Paniz Farrokhsiar (my sister from another mother), Negar Ashrafi, Ali Ghazvinian, Shokufeh Darbari, Nastaran Tebyanian, Mahyar Hadighi, Farzaneh Oghazian, Mona Mirzaei, Rana Zarei, Dhaval Chedda, Shiva Puntanthanbaker, Irem Oz, Elena Vazquez, Julio Diarte, Clarissa Ferreira Albrecht Da Silveira, Lacey Goldberg, Cansu Tari, Aastha Singh, Vernelle Noel, Dario Vanegas-Vargas, Eduardo Castro e Costa, Berfin Evrim, Angelica Rodriguez, Rohini Raghavan, Amruta Saki, and Valerio Dario.

To my colleagues and friends at the Design Computing Group (DCG) at the University of Lisbon, I had one of the best 5 months of my life with you! Thanks for making sunny Lisbon even sunnier. Jose N. Beirão, your supervision and faith in me for enhancing CityMaker was very precious. I’m very proud to have had the opportunity to enhance my

scholarship with your insights and contribution. My dearest friends at DCG, I appreciate your hospitality and making me feel at home. Thank you, Rui de Klerk, Elif Ensari, João Neves, Deborah Macêdo, Rogério Lima, Daniel Mateus, Ronaldo Fiuza, and Stefano Fiorito.

To my family:

Writing this part is bittersweet for me since my family in Iran never got to visit me in the US but their love from miles away always warmed my heart. When my father was studying his PhD, I was a little girl. I could for sure say that seeing you Dad, has been my biggest inspiration for doing a PhD and I'm grateful for you and your hard work and love. Mom, your loving personality, and support made me forget all the hardship of this path. Thank you for letting me be me and doing what I want. I love you both!

Shaya, you're the most amazing brother one can ask for. You cherished me and made me laugh at my lowest moments and I love you for being you!

Saving the best for last - I would like to give my biggest THANK YOU to Ardavan; my friend, my partner in crime, and my husband. I cannot even start writing about how supportive you've been in this journey! Starting our life first as master students and then PhD students had not been easy, but we made it! Your endless love, warmth, kindness, and support has been soothing throughout this journey. Thanks for making me a better person! I love you!

I dedicate this dissertation to my grandparents, both deceased during my studies in the US: Mamani and Baba Abdi, your love lives within me.

Chapter 1 Research Description

1.1 Introduction

Urban areas and the building sector have shown to be a major source of energy consumption and waste. Recent reports argue that as of 2014 more than 50% of the world population are settled in urban areas and this number is expected to increase to 64%-69% by 2050 (Seto, et al., 2014). Such unprecedented speed of urbanization is followed by increasing rates of industrial activities which are highly associated with economic development, growth in the income rate and subsequent increase in energy consumption and greenhouse gas emissions (Jebaraj & Iniyar, 2006; Dodman, 2009; Seto, et al., 2014). Statistics show that urbanization and industrialization account for 75% of global energy use (Bastiononi, Pulselli, & Tiezzi, 2004) and 80% of global greenhouse gas emissions appear in urban areas (Grubler, et al., 2012).

In a paper by Höök and Tang (2013), the authors have reviewed the emission scenarios witnessed throughout history and concluded that high dependence on existing fossil fuel-based power plants, have primarily been one of the main drivers of CO₂ and greenhouse gas emissions in urban areas. These power plants generate electricity from nonrenewable sources of energy which makes them environmentally harmful. However, this is not the only shortcoming associated with regional power plants; due to their centralized architecture electricity is generated far from its point of consumption with long distance interconnected transmission lines and distribution networks. This has resulted in massive amounts of energy to be wasted along the transmission lines before being delivered to customers along with having negative impacts on the landscape. Moreover, with rapid urbanization and continuous changes of the dynamic environment, the traditional power infrastructure has been causing problems in terms of lacking resiliency and flexibility for a continuous service (Amin & Wollenberg, 2005; Farhangi, 2010; Kang, Park, Oh, & Park, 2014; Villareal, Erickson, & Zafar, 2014). This is mainly due to the aging equipment of centralized power distribution systems which, when coupled with natural disasters and/or with increasing rates of energy demand, cause frequent power outages and threaten the everyday lives and businesses of urban dwellers.

The increasing rates of greenhouse gas emissions along with unprecedented demand of electrical power at the users' end and outdated investments in the power infrastructure, have created the need to resolve the existing power grid issues. Respectively, researchers have suggested transitioning from centralized electrical infrastructure where power generators are being built around large regions, to decentralized self-sustained communities which produce and consume their own energy onsite (Amin & Wollenberg, 2005; Farhangi, 2010). In a decentralized scenario, energy is generated by renewable and/or clean energy resources such as solar, wind, and fuel cells. Because of the

intermittency and unpredictability of these energy resources the existing electrical grid needs to be updated with innovative technologies that would handle the fluctuations of energy generation. Additionally, for solving the problem of resiliency in the outdated electrical grid the decentralized power infrastructure needs to be equipped with pervasive control and monitoring systems, data management infrastructure, and communication technologies enabling a two-way exchange of energy and data between the different subsystems. These basic ingredients are set to update the centralized energy infrastructure to become decentralized and for “microgrids” to emerge.

Microgrids are local energy infrastructures which support resiliency in the electrical grid by exercising greater control over production by generating energy close to its point of consumption. Microgrids provide energy to a defined cluster of buildings and thus energy circulates only locally in its infrastructure. Microgrids integrate various techniques of automation, optimization, pervasive control and computation on both the supply and demand side (Sherman, 2007; Paglia, 2011; Rahimian, Iulo & Cardoso Llach, 2015). Microgrids consist of three main components and their subcomponents¹ (Lidula & Rajapakse, 2011; Mariam, Basu, & Conlon, 2013; Fu, et al., 2015):

- *Distributed energy resources (DER)* which include:
 - Distributed energy generators: including renewable (i.e., solar photovoltaic energy generators and wind turbine, biomass) and non-renewable energy resources (i.e., combined cycle gas turbine) which produce energy onsite, meeting the energy demand of the microgrid’s loads.
 - Energy storage devices: such as batteries and fuel cells that take care of balancing onsite energy generation and energy demand in microgrids. Additionally, energy storage devices respond to power disruptions resulting from intermittent onsite energy generation, providing initial energy for a seamless transition between grid-connected and island mode, and allowing different energy generators to work as dispatchable generation units.
- *Loads*: are the customers that the microgrid’s local power infrastructure serves. Loads are mainly grouped based on the demand of high degree power quality and reliability in the event of an outage (Lidula & Rajapakse, 2011). By this, loads are classified into critical loads such as military sites, hospitals, and university campuses where even momentary power interruptions will cause significant financial, physical, and humanistic consequences, and non-critical loads such as residential buildings and urban communities.
- *Distribution facilities*: including wires and transformers for delivering electricity, and pipes to transmit useful steam and hot or chilled water to the loads within the network.

¹ More details on microgrids are discussed in Chapter 2.

In this research, the configuration of the electrical components and distribution facilities of a microgrid is referred to as the “*microgrid infrastructure*” which also outlines the microgrid’s local energy boundaries². Different microgrid installations have different electrical configurations implying different combinations of generators, storages, and loads that they support. The variety of microgrid “*architectures*” depend on a mix of parameters such as resource availability, geographical locations, load demand, and the existing electrical transmission and distribution systems (Mariam, Basu, & Conlon, 2013).

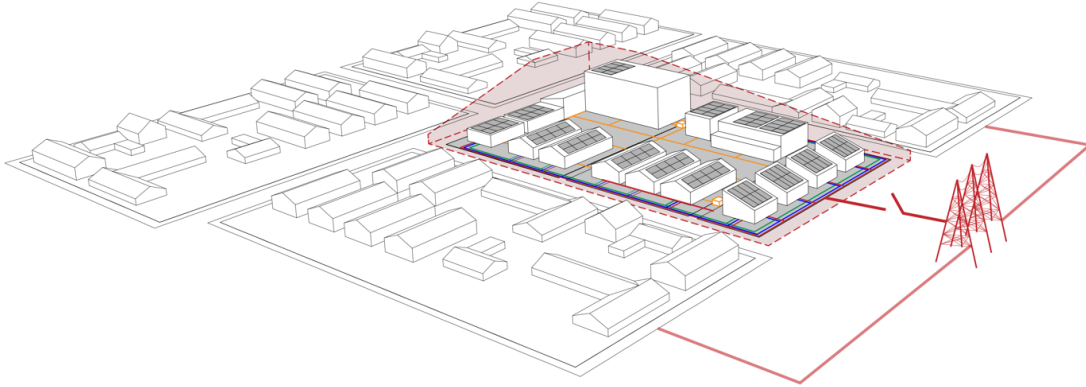


Figure 4- A microgrid connected community within the urban context, with clearly defined energy and regional boundaries that can disconnect from the larger grid and operate in island mode.

The interconnection and clustering of distributed energy resources, loads, and distribution facilities per se does not qualify for being a microgrid unless the infrastructure is able to operate when disconnected from the larger power grid (Stadler, et al., 2016). Microgrids are equipped with controlling capabilities which make them technically able to disconnect from the grid by isolating a group of buildings and self-powering them with onsite energy generators. The state of disconnection from the larger power grid is known as “*islanding*”. The process of islanding prepares microgrids for serving as emergency backups for critical and sensitive loads (as mentioned above), providing security, reliability and resiliency under unexpected environmental conditions of resource depletion (i.e., earthquake, hurricane, etc.), or rise of energy expenses. In addition to critical loads, microgrids are emerging as the local power infrastructure for communities in urban contexts which do not necessarily operate as critical loads. The rise of microgrid-connected communities in cities is specifically geared towards reducing energy costs and greenhouse gas emissions as well as developing energy independent urban settlements. Known as “*community microgrids*”, these energy-independent urban settlements are generally comprised of various mixes of residential, commercial, agricultural, and industrial loads followed by the local sources of power supply.

² The term “*energy boundaries*” is used herein since in some microgrids in addition to electricity, heat is also distributed locally.

This research is concerned with community microgrids and due to the common utilization of solar photovoltaic (PV) technologies in existing community microgrids in the US, the focus is specifically on solar-powered community microgrids (also known as solar community microgrids). Solar community microgrids generate power via solar energy as one of the main resources of clean energy³ harnessed onsite. Unlike typical solar communities, solar community microgrids are designed to physically provide the energy needs of the populous; that is, in solar community microgrids the photovoltaic energy⁴ that is generated onsite is directly used by the interconnected buildings or gets stored for backup in case of power disruptions or for use when production is not feasible⁵.

1.2 Problem Statement

As with any other energy system, the efficiency of a community microgrid's energy performance is evaluated by comparing the energy input to the system from the on- and off-site sources of energy to the energy that is outputted from the system in the form of useful energy for buildings operation, the stored energy in microgrids' storage devices, and the wasted energy during the conversion or distribution process (Fernandez & Blumsack, 2010). In simple terms, improving the energy performance of a community microgrid as used in this research, involves increasing the local energy production and reducing the energy demanded for community buildings operation; a high energy performance community microgrid ensures extended durations of energy self-sufficiency while in island mode.

Current research on improving energy performance in community microgrids has been exclusively advocating technological advances enhancing the limited supplies of local energy and addressing the constantly growing demands of the loads (Wouters, 2015; Siderius, 2004). However, researchers argue that while per capita energy consumption, specifically in residential buildings, has been gradually increasing since the 1980's, focusing on technological innovations alone wouldn't solve the current energy issues in the built environment (Ewing & Rong, The Impact of Urban Form on U.S. Residential Energy Use, 2008). In the case of community microgrids this statement is especially accurate since they are contextualized in cities and urban areas. Citing research from the 1970's onwards, different spatial configurations of urban form changes the local wind pattern and drives urban heat island effect (Steadman, 1977; Owens, 1986; Santamouris,

³ Clean energy does not solely refer to renewable energy resources. i.e. natural gas which is a non-renewable source of energy is called 'clean' because it emits less carbon dioxide than burning coal. Many community microgrid are equipped with different types of clean energy technologies such combined heat and power (CHP) systems using natural gas as their fuel.

⁴ Photovoltaic energy refers to the conversion of solar energy to electricity.

⁵ This is while the buildings involved in a solar community do not physically benefit from the photovoltaic energy, instead they receive financial benefits, often in form of reduced utility bills. Also, in the case of power disruptions in the grid, solar community microgrids can disconnect or island from the grid while in solar communities the grid-connected solar technologies will stop working too (Chaurey & Kandpal, 2010).

Papanikolaou, Livada, & Koronakis, 2001; Reinhert, Dogan, Jakubiec, Rakha, & Sang, 2013; Silva, Oliveira, & Leal, 2017; Chatzidimitriou & Yannas, 2015); these changes influence the buildings' thermal comfort which eventually leads to fluctuations in patterns of energy consumed for building space heating and cooling as well as the feasibility of adopting on-site renewable energy generators such as photovoltaic panels and wind turbines⁶ (Steadman, 1977; Owens, 1986; Grosso, 1998; Steemers, 2003; Cajot, Peter, Bahu, Guignet, & Koch, 2017). Accordingly, literature today emphasizes the importance of obtaining an energy-conscious point of view when taking actions toward urban design and planning; urban planners are expected to consider the tradeoffs between the living qualities of an urban context and its potential for high performance energy systems design and engineering (Cajot, Peter, Bahu, Guignet, & Koch, 2017). Along these lines of studies, this research argues that studying energy performance in community microgrids in isolation from their comprising buildings and urban context is insufficient and limits a holistic understanding of community microgrids as building-integrated energy systems.

1.3 Hypothesis

This research discusses energy performance in solar community microgrids in the interest of expanding the existing discourse by adding a spatial dimension to its definition. Therefore, this study componentizes a community microgrid into two distinct but related parts (Figure 2):

- *The community microgrid infrastructure:* which is the combination of electrical components and distribution facilities conceiving a microgrid's local energy boundaries.
- *The community microgrid superstructure:* which refers to its urban form and in the context of this research refers to buildings and their arrangement and configuration

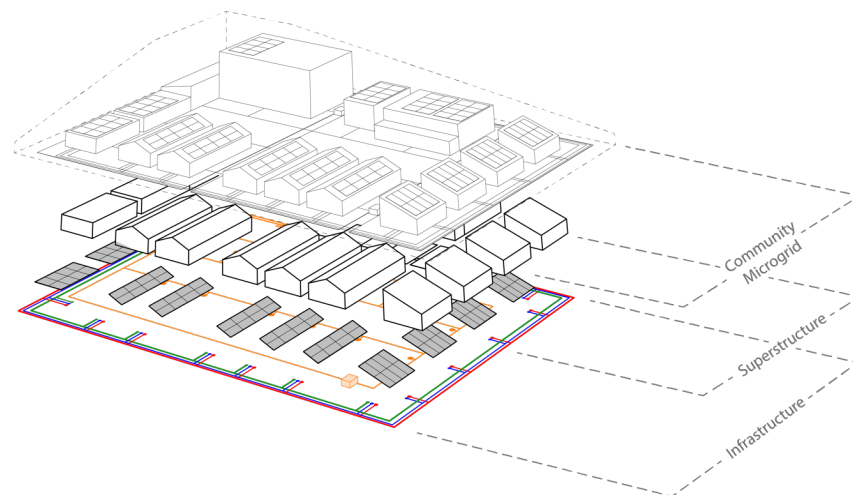


Figure 5- Community microgrids comprise of an infrastructure and superstructure.

⁶ This concept will be fully discussed in the “state-of-the-art” section

within the community, the urban spaces between the buildings and the buildings' relationship to one another.

The hypothesis driving this study is that a community microgrid's energy performance goes beyond the technical assessment of its electrical infrastructure, arguing that there is a relationship between the urban form (superstructure) of a community and how well the microgrid performs in terms of the local supply and demand of energy (infrastructure). It's important to note that in this research the impact that the spatial geometry of urban form has on community microgrids' energy performance is explored without considering the physical characteristics, construction type or age of each individual building in the community⁷.

Considering solar community microgrids and for the purposes of this research, the definition of energy performance is specified by comparing the net amount of solar energy captured onsite with the net amount of electrical energy required for space heating and cooling⁸. To better understand the energy performance of a solar community microgrid there's a need to explore how the urban form of a microgrid-connected community impacts the amount of solar energy captured onsite (energy input) and the amount of energy that is required for building's operation especially for space heating and cooling⁹ (the useful energy outputted). An energy optimized community design for a high performance solar microgrid as pursued in this research, is a design that takes maximum advantage from the site's ambient resources of energy for passive space heating and cooling and PV energy production. Following this principle in the early phases of design and planning of a microgrid-connected community, will downsize the active systems that need to be installed at the infrastructure (such as advanced technical measures and ventilation systems), and therefore reduces the need for reconnecting to the grid for purchasing energy for further supply. This translates into the ability of a community microgrid to sufficiently self-power while in island mode and for longer periods of time. Based on this premise, the urban form of any region or area becomes an undeniable component of the microgrid-connected community that is located there, regarding both the local supply and demand of energy.

1.4 Research Questions

In attempt to examine the presented hypothesis, this research is oriented towards addressing the following research questions:

⁷ In an ideal energy self-sufficient scenario, individual buildings would also be optimized for energy performance.

⁸ This study does not consider heat energy and it is an opportunity for future work.

⁹ Susan Owens (1986) argues that the major source of energy consumption in buildings are for space heating and cooling.

1. The spatial geometry of urban form is a complex entity and is defined by many spatial attributes. Which of these attributes have an energy-relevance regarding the community-wide supply and demand of energy and how are they measured?
2. What is the most appropriate model for describing the desired relational pattern between urban form and energy performance of microgrid-connected communities?
3. Which combination of the attributes of urban form impact a solar community microgrid's energy performance?

1.5 Methodology Overview

Researchers suggest three different methods for developing relational models explaining energy performance in the built environment¹⁰ (Magoules & Zhao, 2016; Silva, Horta, Leal, & Oliveira, 2017):

- *Engineering methods*: where physical principles are used to calculate the energy performance of an entire building (or its sublevel components) relevant to the physicality of the building. The basis of this method is to precisely calculate the thermal dynamics and physical functions of buildings based on their structural and operational characteristics, environmental factors, and sublevel building components. Engineering models are typically associated with a great extent of complexity and detail. The main problem with this method is that in order to achieve accurate simulations, detailed information on building quality parameters is required which are unavailable to many organizations and to the public to study on. The engineering method is known as white-box models (Tardioli, Kerrigan, Oates, O'Donnell, & Finn, 2015). In studies on buildings, white-box models are applied at a scale of a single building or a subset of a building (Silva, Horta, Leal, & Oliveira, 2017). Utilizing white-box models for an entire urban building stock is of high complexity and requires a considerable amount of time and data due to processing the manifold of energy relevant urban-scale variables that needs to be considered (Tardioli et al., 2015).
- *Statistical methods*: also known as grey-box models combine physical and engineering structure with data-based, statistical modelling (Tardioli, Kerrigan, Oates, O'Donnell, & Finn, 2015). Grey-box models usually have very particular analysis methods including linear correlation, regression analysis, stepwise regression analysis, logit models, ANOVA, t-test, factor analysis, panel data, structural equation models, and cross-tabulation (Silva, Horta, Leal, & Oliveira, 2017) with the aim of correlating energy indexes with influencing variables.
- *Data mining methods*: data mining is the computing process of discovering patterns and “extracting implicit, previously unknown, and potentially useful knowledge

¹⁰ Note that explanations in these approaches are based on research on urban energy modeling.

from data” (Tsui, Chen, Jiang, & Aslandogan, 2006). According to Fayyad et al. (1996), data mining consists of “*applying data analysis and discovery algorithms that produce a particular enumeration of patterns (or models) over the data*”. Data mining techniques originated a branch named machine learning (Silva, Horta, Leal, & Oliveira, 2017; Chen, Sakaguchi, & Frolick, 2000). Machine learning is the “*science and art of programming computers so that they can learn from data*” (Géron, 2017) in which “*the ‘machine’ is able to identify and generalize patterns*” from large datasets (Silva, Horta, Leal, & Oliveira, 2017). Arthur Samuel (1959) explains machine learning as giving computers (or ‘machines’) the ability to learn without being explicitly programmed. Machine Learning algorithms are essentially systems that learn and discover relational patterns and unsuspected new trends from collected data (that were not immediately apparent) and make data-driven predictions and decisions. Machine learning methods are known for working as a ‘black box’ where a *machine learning algorithm* is constructed and ‘trained’ to discover the relational pattern and hidden structures in the input dataset that the user provides (Tardioli, Kerrigan, Oates, O'Donnell, & Finn, 2015). A *machine learning model*, is the outputted mathematical model artifact that explains the discovered relational pattern in the input dataset and can be further used to make data-driven predictions and decisions on unseen examples (Tardioli, Kerrigan, Oates, O'Donnell, & Finn, 2015). In other words, machine learning is the construction of algorithms that can learn relationships from data through building mathematical models and make data-driven predictions or decisions accordingly.

Adopting a machine learning method is adequately useful when the problem of interest is multidimensional and complex, and solving it with conventional engineering or statistical models are extremely time and resource consuming. In the case of this study, researchers (Ewing & Cervero, 2010) have marked the importance of studying the *combined* impact of different attributes of urban form on energy performance rather than the influence of each individual spatial attribute. When studying the combined effect of urban spatial attributes, the problem becomes too complex for solving it with closed-form solutions. Gil et al. (2012) mark data mining as a sufficient method for analyzing the multidimensional relational complexity of urban environments. Therefore, due to the high-dimensionality, complexity, and computational intensity of finding the relationship between the combination of all relevant spatial attributes of urban form and energy performance in solar community microgrids, a machine learning approach, more specifically artificial neural networks, is selected for this purpose. Details on artificial neural networks and their operation is explained further in this dissertation.

As mentioned earlier, the prerequisite for working with machine learning models is to possess a large, structured dataset. In this research — where the interest lies in identifying the relational pattern between urban form and energy performance in solar community

microgrids — the dataset needs to have quantifiable measures of urban form as the predictor variable and measurements of energy performance as its response variable. In an ideal situation, the required dataset will be prepared by measuring energy-relevant indices of urban form in thousands of operating solar community microgrids world-wide along with their monthly rates of energy performance. Accessing such dataset is not feasible for several reasons including the limited duration of this study, privacy reasons associated with accessing energy data, and lack of data-rich spatial and energy repositories for existing solar community microgrids to name a few. Therefore, instead of concentrating on existing operating solar community microgrids, a city has been selected as a case study to find the relational pattern between the urban form and energy performance of its communities. Finding the correlation between urban form, net energy consumption and potential net PV energy production in the existing communities of a city, will in turn inform the best practices for spatially designing high energy performance solar community microgrids in that city.

San Diego county of California¹¹ has been selected as a case study for three main reasons: Firstly, after Los Angeles, San Diego is the second highest ranked city in the US for total installed solar capacity as of 2019 (Bradford, Stankiewicz, Sundby, Fanshaw, & Sargent, 2019) with the ambitious goal of generating 100 percent of its electricity from renewable sources by 2035 according to the “City of San Diego Climate Action Plan” (The City of San Diego, 2015). Therefore, the region’s sustainability goals perfectly align with the broader goals of this research. Secondly, San Diego is one of the very few counties that has a rich repository of both spatial and energy data publicly available online. Aside from having a comprehensive set of GIS data representing the different physical and infrastructural layers of the county, San Diego’s main utility company, SDG&E¹², has aggregated and published energy consumption information for the entire county from 2012 onwards. Thirdly, San Diego’s urban planning is representative of American urban planning. The county’s varied topography has resulted in different urban typologies to emerge throughout the region which is generalizable to most cities in the US.

By choosing San Diego the spatial and the energy consumption data of the county, which accounts for a substantial portion of required datasets, were covered. However, data on the county’s PV energy production remained inaccessible. Due to this shortcoming, the methodology obtained for this study has been divided into two distinct but related parts and includes analysis on 1. the impact of urban form on energy consumption using real-world datasets and, 2. the impact of urban form on PV energy production using synthesized datasets. The two subsections below briefly describe the approaches taken in each of these phases.

¹¹ In this dissertation, the term “San Diego” refers to the county (and not the city) otherwise noted.

¹² San Diego Gas & Electric.

1.5.1 Urban Form and Energy Consumption

Urban form can be decodified and measured through a set of spatial indicators and metrics. Quantification of urban form in this research is based on a comprehensive number of spatial indicators along with their metrics of measurement which were extracted from an extensive literature review. The selected nineteen indicators of urban form were measured for all zip codes in San Diego using parametric algorithms and geoprocessing tools. Additionally, monthly values of energy consumption data were obtained through San Diego's main utility company from 2012 till 2018 for each zip code. An artificial neural network was then trained on the dataset to identify the relational pattern between urban form and community-wide net energy consumption and to deliver a predictive model. The outputted predictive model gives the answer to 'what' is being predicted, but 'why' certain predictions are being made is often a challenging question due to the black box quality of artificial neural networks. For statistically and architecturally interpreting the artificial neural network model, Shapley values, a method from coalition game theory, was used which is highly regarded due to its consistency and local accuracy in performing statistical inference on non-linear models.

1.5.2 Urban Form and PV Energy Production

Since real-world data on PV energy production in San Diego were not available, a rule-based method for synthesizing the required dataset has been developed herein. In this method, the different urban typologies in San Diego were identified and using shape grammars the shape rules and logic behind the emergence of these typologies were extracted. The shape rules were then codified into parametric algorithms and used towards developing a *generation-evaluation* urban design tool capable of producing similar urban and community design scenarios accordingly and evaluating the communities' potential of PV energy production during 12 months of a year. The intention of this tool is to create a synthesized dataset of measurements of urban form (from the hypothetical communities generated by the tool) along with their measurement of potential of PV energy production. Similar to the approach taken in section 1.5.1, an artificial neural network is then trained on the final synthesized dataset to deliver a predictive model and to identify the extent in which urban form influences PV energy production in communities.

1.5.3 Final Deliverables

This research is rendered towards two final deliverables:

- First are the predictive models resulting from the training procedure and their spatial interpretation on how urban form influences energy performance in solar community microgrids. Statistically and spatially interpreting the decision making behind the neural network's predictions is crucial in this research because by identifying the degree in which different spatial indicators of urban form influence energy consumption and PV energy production in communities, a set of design

principles and guideline can be curated for architects and urban planners for spatially designing and/or retrofitting urban settlements towards high energy performance solar-powered community microgrids well before their construction.

- Secondly, a multimodal study of solar community microgrids as presented in this study brings the need for a new generation of computational modeling, simulation, and evaluation tools for the field. A literature review as presented in Chapter 3 identifies a gap in existing software tools that simultaneously address the necessary interaction between the superstructure and infrastructure of community microgrids, given the importance of its impact. The second deliverable of this research is an experimental software prototype that bridges this gap by predicting the energy performance of any given solar community microgrid design scenario by the virtue of its urban spatial configuration in '*real-time*'. The word '*real-time*' is used here deliberately to imply the benefit of utilizing the trained predictive models as the backend of the software prototype. The software prototype is developed as an add-on for Rhinoceros and Grasshopper and it's created to assist architects and urban planners in designing high energy performance solar community microgrids by enabling them to view the impact of their community's 3D spatial design scenarios on the microgrid's energy performance.

1.6 Contributions and Findings

The findings and contributions of this research are as following:

- The initial goal of this study was to explore the impact of urban form on energy performance of solar community microgrids, which includes energy consumption and PV energy production. However, this study proceeded with focusing only on energy consumption because in the process of the research it became apparent that simulating PV energy production for numerous community scale urban designs for the purpose of synthesizing the data is a task beyond the capacity of existing tools. Therefore, studying the impact of urban form on PV energy production has been listed as a next step of this research.
- Studying the relationship between urban form and community-scale energy consumption demonstrated that the most influential indicators of urban form are the ones related to the compactness, passivity, shading, and diversity of a community in the context of the case study. To be more specific, the following urban form indicators have been deemed to be the most influential in a descending order: Size Factor, Mixed Use Index, Passivity Ratio, Sky View Factor, Plan Depth, Form Factor, Open Space Ratio, Network Density, Community Building Orientations, Urban Horizon Angle, Volumetric Compactness, Obstruction Sky View, Floor Space Index, and Street Orientation
- By architecturally interpreting the relationship between urban form and community scale energy consumption, this research has set the path for engaging architects in

the technical conversation of developing solar community microgrids and offers a spatial vision on their construction.

- Previous research has overlooked the complexity and multidimensionality of the urban form when studying its impact. The deductive nature of past studies does not offer a rigorous and comprehensive evaluation of the desired interaction. This is essentially due to computational limitations and lack of data rich environments. By the innovative use of machine learning as means for knowledge discovery on the relational complexity between different variables of urban form and energy consumption, this study is in opposition to the previous abstract and deductive modes of analysis and offers new knowledge in the field.
- Previous urban generative design tools did not consider limitations of topography as a design constraint. The developed shape grammar and the urban generation-evaluation design system contributes to the field by creating a set of planning rules that take topography into account. The development of this design system - as part of CItYMaker¹³ - extends the capabilities of an existing tool and pushes forward current research.
- The outputted software prototype takes the first step in introducing a new generation of tools for architects and urban designers for spatially modeling community microgrids and evaluating their energy performance. The energy simulations performed by this tool are in real-time due to predictions being based on a trained machine learning model. This is novel since building and urban scale energy simulations performed by existing tools are extremely time consuming since the backend involves heavy hard coded algorithms.

1.7 Dissertation Roadmap

The dissertation is organized into nine chapters including this introductory chapter:

Chapter 2 describes microgrids and presents current different viewpoints on its definition. This chapter also explains the development of microgrids, and the different stakeholders involved in this process.

Chapter 3 offers a comprehensive review of existing software packages involved in community microgrid design and simulation and identifies a gap in existing tools that simultaneously address the necessary interaction between the superstructure and infrastructure of community microgrids.

Chapter 4 reviews literature from 1970's onwards which have discussed the impact of

¹³ CItYMaker is a rule-based parametric urban design tool based on shape and description grammars (Beirão J. N., CItYMaker: Designing Grammars for Urban Design, 2012). CItYMaker was developed as the generative component of the City Induction project which integrates various design support tools for the formulation, generation, and evaluation of urban design scenarios (Duarte J. P., Beirão, Montenegro, & Gil, 2012).

urban form on community-wide energy consumption and PV energy production. This chapter also identifies a list of all spatial indicators of urban form with energy relevance, describes their importance in a solar community microgrid context, and specifies a metric for measuring them. This list is later used in measuring urban form in the methodology.

Chapter 5 presents the technicality behind studying the impact of urban form on energy consumption in the case study using the real-world data and artificial neural networks.

Chapter 6 presents a statistical and architectural interpretation of the resulting predictive model from Chapter 5. The statistical inference helps with identifying the most influential indicators of urban form on energy consumption. Then, by architecturally explaining the statistical results, a guideline is offered to architects and urban planners for designing high energy performance solar community microgrids.

Chapter 7 describes the developed urban shape grammars associated with San Diego's urban planning principles, elaborates on the generative tool that was codified upon, and explains the existing limitations with current parametric PV simulation tools that prevented from fully developing the intended generation-evaluation algorithm.

Chapter 8 dives into the details of developing the real-time predictive software prototype using the monthly trained models.

Chapter 9 provides a summary of the concluded results and discusses them in the context of the bigger picture, describes the contributions and implications of the research to the field, and elaborates on ideas for future directions and next steps.

Chapter 2 Microgrids

Introduction: different definitions and categorizations of microgrids exist depending on the stakeholders involved. However, the general definition as offered by US department of energy refers to microgrids as a *“group of interconnected loads and distributed energy resources within clearly defined electrical boundaries that acts as a single controllable entity with respect to the grid. A microgrid can connect and disconnect from the grid to enable it to operate in both grid-connected or island-mode”* (Ton & Smith, 2012). This chapter gives a brief explanation on the emergence of microgrids, the different components constructing a microgrid infrastructure, different categorization of microgrids, the benefits and values of these local energy systems, and their development process.

2.1 Microgrid Definition

Electrification in industrialized nations such as in the United States, Western Europe, Chile and Australia, initiated with decentralized, current-based and isolated electrical grids that were privately owned and served a limited number of users (Van Hende & Wouters, 2014). At the end of the 19th century, cities all around the world faced high rates of urbanization and industrialization followed by increasing rates of global energy demand. Such growing demand for electricity upscaled the local electricity network to large, centralized power plants far away from consumers with long-distance interconnected transmission and distribution networks (Wouters, 2015). The emerged utility owned central power plants changed the energy regulatory framework to state-wide controlled monopolies (Wouters, 2015). With the dominance of fossil fuel-based power plants as the main power supplier in urban areas, harmful environmental problems started emerging from cities (Dodman, 2009). The rising transformations and unpredictable changes of the climate, the finiteness of fossil fuel-based energy resources, along with the emerging network congestion and atypical power flows initiated a conversation on updating the traditional centralized power grid system. These updates were proposed towards creating a more reliable and resilient power grid that could keep up with the changes of the dynamic environment and the rising need for resolving the power grid issues (Werbos, 2011). In the late 20th century and early 21st century, the power industry recognized some mundane needs for upgrading the existing grid by placing a layer of computation over the infrastructure through novel automation and control technologies facilitating data and energy management and communication among the different components of the power infrastructure (Folke, Carpenter, Elmqvist, & Gunderson, 2002; Farhangi, 2010; Williams, Gahagan, & Costin, 2010). These basic ingredients accelerated utilities to deploy technologies such as energy metering and feedback system, simple sensors and communication networks ascribed at the distribution operator scale and power operation scale (Werbos, 2011).

The context of these new technologies advanced the outdated electrical grid to a ‘smarter’ grid capable of substituting energy and information among its components. The introduced “*Smart Grid*” is not known as a replacement of the existing electrical grid but a complement to it, coexisting with it, and adding to its functionalities and capacities by deploying a collection of innovative pervasive and control technologies (Farhangi, 2010). These technologies allow the fully integrated networking and communications of all generation, transmission and distribution subsystems and thus support the needs of its stakeholders by the efficient exchange of data, services and transactions (Farhangi, 2010).

The smart grid could be viewed as an ad-hoc integration of small groups of discrete energy systems consisting of distributed command-and-control energy sources and loads, capable of operating in parallel with, or independently from the main power grid (Ton & Smith, 2012; Villarreal, Erickson, & Zafar, 2014). The US department of energy refers to these discrete energy systems as microgrids and defines them as “*group of interconnected loads and distributed energy resources within clearly defined electrical boundaries that acts as a single controllable entity with respect to the grid. A microgrid can connect and disconnect from the grid to enable it to operate in both grid-connected or island-mode*” (Ton & Smith, 2012).

2.2 Microgrid Components

As briefly described in Chapter 1, the electrical infrastructure of microgrids comprise of two main components, their subcomponents, and the distribution facilities among them (Sherman, 2007; Lidula & Rajapakse, 2011; Mariam, Basu, & Conlon, 2013; Fu, et al., 2015):

- *Distributed energy resources (DER)*: microgrids are technically able to support integrating various distributed renewable energy resources. Due to the intermittency of these resources, relying exclusively on them would increase the probability of supply failure. Therefore, in addition to renewable energy sources, existing operating microgrids commonly utilize alternative types of nonrenewable but clean energy resources and are backed up with energy storage systems (NYSERDA, DHSES, & DPS, 2014). Distributed energy resources refer to both distributed energy generators (onsite renewable and nonrenewable) and energy storage technologies (Lidula & Rajapakse, 2011).
 - *Distributed energy generators (DG)*: are renewable and nonrenewable energy generators that produce energy on site to meet the energy demand of the microgrid’s loads. The choice of distributed energy generators mainly depends on the climate and topology of the microgrid’s site (Mariam, Basu, & Conlon, 2013). Some popular distributed energy generators include (NYSERDA, DHSES, & DPS, 2014): (1) nonrenewable: combined cycle gas turbine, internal combustion engines, combustion turbine, micro

turbines, stirling engine, reciprocating engine, and (2) renewable: micro hydro, wind turbine, solar electric, solar thermal, biomass, geothermal, ocean energy.

One of the popular forms of distributed energy generations is the use of conventional fuel-burning engines designed for operating as combined heat and power (CHP) systems. These systems capture the wasted heat resulted from electricity production in order to provide steamer hot water for space heating, cooling and other processes (Hampson, Bourgeois, Dillingham, & Panzare, 2013). Other key attributes that could distinguish distributed generators from each other are the input fuel, location, size and relationship to the conventional grid.

- Energy storage devices: The utilization of energy storage in microgrid systems goes beyond merely storing energy for future use. Energy storage devices are responsible for balancing power in the network in case of any disturbance and regulating short term power and energy demand with generation. Moreover, these devices provide the initial power for a seamless transition between grid-connected and island mode, ensuring energy balance in the event of a significant load and allowing different energy generators to work as dispatchable generation units (Lidula & Rajapakse, 2011; Ariyasinghe & Hemapala, 2013). Energy storage is necessary to ensure the stability and consistency of interrupted power specially caused by intermittent energy resources (Mariam, Basu, & Conlon, 2013).

Some popularly cited energy storage technologies used in microgrids are batteries, flywheels and capacitors. Flywheels can be used as a central storage system for the entire microgrid, batteries could be treated as central storage as well in addition to the ability of storing power for future use. In comparison to flywheels and batteries, capacitors are a much more expensive choice to be used in the system (Lidula & Rajapakse, 2011; Mariam, Basu, & Conlon, 2013).

- *Loads:* are the customers that the microgrid's local power infrastructure serves. The classification of loads has an important role in setting up a microgrid. Loads are mainly classified based on the demand of high degree power quality and reliability in the event of an outage (Lidula & Rajapakse, 2011; Ariyasinghe & Hemapala, 2013). By this, loads are classified into critical loads such as military sites, hospitals, and university campuses where even momentary power interruptions will cause significant financial, physical and humanistic consequences, and non-critical loads such as residential buildings and urban communities.

- *Distribution facilities:* including wires and transformers for delivering electricity, and pipes to transmission useful steam and hot or chilled water to the loads within the network. In the case of this research the distribution facilities include DC power distribution, battery storage facilities/interconnection to power grid and AC conversion inverters, AC electricity distribution wires, and etc.

2.3 Different Categorizations of Microgrids

Due to the fact that many stakeholders are involved in a microgrid's development process, different perspectives, models, and applications of microgrids are available (NYSERDA, DHSES, & DPS, 2014). For example, engineers working for a developer may distinguish microgrids by the kinds of energy generation technology they utilize. On the other hand, engineers working at a utility company may categorize microgrids by the type of distribution system they connect with. Lawyers and financiers may have the microgrid's ownership structure or financing arrangement as the main point of their operative differences. Below are some sample categories of microgrids adopted from a report prepared for the state of New York after hurricane Sandy (NYSERDA, DHSES, & DPS, 2014):

- *Based on the intermittency of a microgrid's operation*
 - Emergency microgrids: run only in emergency events and are purposed to avoid the negative impact of power outages. The most popular type of energy generator used in such microgrids are diesel-based since they can respond to varying demands in a very short time.
 - Base load microgrids: run frequently and typically are shut down only for maintenance. Since these microgrids operate continuously, they reduce the site's energy purchase from the larger power grid. Popular base load generator types include gas turbines, reciprocating engines, micro-turbines, and steam-turbines.
 - Intermittent microgrid: do not run continuously due to the type of onsite energy generators utilized that normally include renewable sources of energy as solar and wind. Since adequate energy may not be supplied by intermittent energy generators solely, different combinations of renewable and non-renewable energy generators may be used for these types of microgrids along with substantial energy storage facilities.
- *Based on how the microgrid load are affiliated*
 - Loads belonging to the same entity: are sometimes referred to as campus style because they commonly serve single entity markets particularly military, universities, schools and hospitals. These types of loads are usually sited on a single piece of property with clear regional boundaries.
 - Loads belong to different unaffiliated entities: this type of affiliation is rare in the United States due to regulatory barriers of serving multiple customers and

their affiliated parties. The loads served in this type of microgrid are usually not directly clustered together and a cost recovery security is needed through a power purchase agreement¹⁴ (PPA) or a similar mechanism.

- *Based on the microgrid's ownership structure*
 - Utility microgrids: when the distribution facilities used in the microgrid's infrastructure is owned by the region's existing electrical utility company.
 - *Full utility microgrids*: when in addition to the distribution facilities, the generation assets are also owned by the existing electrical utility company.
 - *Hybrid utility microgrids*: when the existing utility company owns the microgrid's distribution facilities, but a non-utility owns the generation assets.
 - Non-utility microgrids: when the distribution facilities used in the microgrid's infrastructure is owned by a non-utility company.
 - *Own use microgrid*: when a non-utility company owns the distribution facilities of the microgrid, and the loads belong to a *single entity*.
 - *Landlord/tenant microgrid*: when a non-utility company owns the distribution facilities of the microgrid, and the loads belong to multiple *affiliated entities*.
 - *Owner/merchant microgrid*: when a non-utility company owns the distribution facilities of the microgrid, the loads belong to multiple non-affiliated entities, and the microgrid serves the owner's loads.
 - *Independent provider microgrid*: when a non-utility company owns the distribution facilities of the microgrid, the loads belong to multiple non-affiliated entities, and the microgrid *does not* serve the owner's loads.

The described different categories of microgrids are a sample of many available ones. It's important to note that depending on the category that a microgrid belongs to, the applied interconnection and net-metering policies as well as the required financial and regulatory contracts may differ.

According to research conducted by GTM Research Group, as of the year 2015, 124 operating microgrids exist in North America. Forty-eight percent of these microgrids are located in the Northeast (29%) and West Coast (19%). The most popular reason for microgrid development has been providing energy resiliency for critical loads such as military bases and university campuses. However, in recent years, an increasing number of microgrids have been emerging as the local power infrastructure for communities in urban contexts that are not necessarily known as critical loads (Powering a New Generation of Community Energy, 2015). Urban communities usually deploy microgrids for becoming

¹⁴ It's a contract between the two entities of the agreement: the one which generates the electricity and the one who purchases the electricity.

energy independent urban settlements, producing their own energy on site via renewable energy resources, reducing greenhouse gas and carbon dioxide emissions, and ultimately cutting on energy costs. In 2016, eighty-three (83) communities in the state of New York were funded to assess their feasibility to become microgrids (Powering a New Generation of Community Energy, 2015). The city of Pittsburgh started exploring building a grid of community microgrids in its downtown area as part of the city's smart city project in 2016 (NETL & DOE, 2016). Lawrence National Berkeley Lab (LNBL) also started partnering with AECOM¹⁵ on developing community microgrids for the city of Berkeley with the goal of reducing reliance on fossil fuel sources, reducing energy costs, and increasing energy reliability and resiliency for the city.

2.4 Benefits and Values of Microgrids

The IEEE Power Engineering Society calls the benefits of microgrids as being economic, environmental-friendly, resilient and reliable, automated and controllable (Paglia, 2011):

- *Economic Value*: the economic benefits of microgrids are threefold:
 - Integrating renewable energy resources in the microgrid infrastructure reduces the need for exploiting fossil fuel-based resources. Reports show how different cities have cut on energy costs by utilizing renewable energy resources generating heat and power. For example, installing microgrid technologies for the Utica College and St. Luke's Hospital and Nursing Home in Burrstone, New York, have resulted in 15%-20% reduction on utility consumption rates (Sherman, 2007).
 - Generating energy close to its point of consumption in microgrids results in shorter transmission lines and less wasted energy. This benefits communities and cities by saving money on energy costs.
 - Since microgrids practice reliability by preventing power outages, consumers and businesses pay less money to offset the any possible damage resulting from power interruptions.
- *Resiliency and Reliability*: an important mission of microgrids is to deliver resiliency and reliability in the event of power interruptions to the larger grid. The controlling capability of microgrids enables them to disconnect/island from the larger grid under unexpected environmental conditions of resource depletion (i.e., earthquake, hurricane, etc.) or the rise of energy expenses. For example, in 2011 two consequent storms in Connecticut resulted in power outage for the duration of nine days which caused serious disruptions in the electric infrastructure inducing high energy costs and congestion issues (Sherman, 2007). This event initiated the conversation on emergency preparedness as a critical issue to resolve in the state of Connecticut followed by a legislation allowing for the establishment of microgrids

¹⁵ AECOM is an American multinational engineering firm.

to support distributed energy generation in critical situations. As another example, in Sendai, Japan, a university zone upgraded to a microgrid between 2005 and 2008. Thus in 2011 after the earthquake and tsunami followed by a service loss, the engine generators of the microgrid supplied the teaching hospital of Tohoku Fukushi University, with both power and heat while other sectors were suffering from a two-day blackout (Hatziaargyriou, Asano, Iravani, & Marnay, 2007; Paglia, 2011).

- *Environmental*: the deployment of a wide variety of renewable energy resources will decrease the use of fossil fuels and lead cities towards cleaner technologies. A project at Fort Collins, Colorado, belonging to the Colorado State University features a 4-megawatt microgrid powered by PV solar, fuel cells and micro-turbines. Another microgrid demonstration project created at the Santa Rita Jail in Alameda County, CA also features solar, wind and fuel cell energy sources and large-scale battery storage (Marnay, 2010).
- *Automation and Control*: by using advanced metering technologies, automated control systems, information and communication software, multiple DERs and loads are controlled by both the supply and demand side. On the supply side, these technologies enable a central operator to optimize the use of each of the DERs according to a variety of environmental, financial and humanistic factors. On the customer's side, the aggregation of advanced energy metering systems and feedback loops inform the users of their consumption rate and expenses, and accordingly help them to acquire greater control over their energy-related consumption patterns (Sherman, 2007).

2.5 Microgrid Development Process

As with any other construction and development projects, constructing a microgrid requires taking concrete steps from setting energy goals to securing project financing. A rough summary is offered below on the steps that need to be taken for constructing a microgrid project, adopted from three departmental and organizational reports (NYSERDA, DHSES, & DPS, 2014; Bourgeois, Gerow, Litz, & Martin, 2015; US Department of Energy, 2011). The intent of this section is to highlight the importance of integrating planning and spatial design considerations of a microgrid before getting into the details of its infrastructure. Note that the summary mentioned below presents a simple step-by-step development process, but more complicity is expected to occur when planning a microgrid in an actual setting due to the different stakeholders involved and the specificities of the chosen site.

- *Goal setting*: the reports show that there are no universal architecture and configuration for microgrids. Thus, the design, implementation and operation of a microgrid highly depends on its outlined purpose. Illustrating the goals and applications of establishing a microgrid is the first step in planning one. Goal setting is merely important for identifying the best matching sites and stakeholders. Microgrids are usually developed to pursue one or more of the goals listed earlier

namely resiliency and reliability, environmental benefits, and economic values (Ariyasinghe & Hemapala, 2013). If a microgrid is developed to offer resiliency and reliability under uncertain conditions, it has to be ensured that reliable onsite energy generators and substantial backup facilities are integrated in the microgrid infrastructure. If the goal of a microgrid is to reduce the carbon footprint of a community, the project should target areas within the site that have significant rooftop or ground space for solar photovoltaic installations or locate the windmills where the wind flow is stronger. If the goal of a microgrid is to spur economic development, the project is apt to focus on industrial and commercial areas.

- *Site selection matching the proposed goals:* Bourgeois et al (2015) identify an ideal site for implementing a microgrid as a site with: high demand energy users and consistent electric and thermal energy needs, existing critical infrastructure, significant exposure to renewable sources of energy particularly solar and wind or long-term access to biomass and biofuels, and concrete future development plans. Some features of a physical location are identified below which their existence helps the site be closer to ‘ideal’ as cited by Bourgeois et al (2015):
 - Presence of anchor energy users: anchor energy users are those who are likely to be at the location for many years in the future. Example of anchor energy users are hospitals, universities, and convention centers. A common feature among all anchor users is that their presence is usually very economical for the microgrid in terms of having high electrical and thermal energy needs. They are also well set for leading financial negotiations for the microgrid system due to their long-term presence as a microgrid load.
 - Presence of complementary energy users: complementary energy users are those whose connection to a microgrid system presents a fairly consistent energy demand. When a microgrid serves a range of complimentary energy users (for example a mix of residential, school, and office buildings) a relatively constant energy demand is being observed over a 24-hour period. For a microgrid this translates into consistency in power demand and economic stability of the system. For instance, when a commercial center with peak hours from 8 a.m. to 5 p.m. is part of the same microgrid serving a residential area with peak hours in the mornings and evenings, such cluster of users provide a combined daily demand profile that is steady throughout the day. In this scenario, the adjacency of the residential area complements the load profile of the [anchor] commercial center.
 - Existence of untapped potential: another feature worth considering when selecting a site for a microgrid is identifying those physical locations with untapped potentials such as existing generation sites that can be cost-effectively reconfigured or expanded to serve a microgrid.
- *Feasibility and screening studies:* After selecting the most suitable site aligned with

the purposes of the microgrid development, the next step is to conduct technical, political, and financial feasibility studies. Some important topics covered in this step involve exploring energy retrofitting options, project's local environmental impacts, creating an engineering financial model, an overview of any permitting concerns that could impact the project, external funding sources, project risks, operating costs and etc. Screening studies play an important role in providing regulatory guidance, establishing financial incentives, and adopting clear rules of the role of the microgrid in the existing electricity system.

- *Organize, engage, and educate core stakeholders:* This step includes identifying and reaching out to the stakeholders that are involved in the project. Different microgrid project might involve different stakeholders - such as local government agencies, utilities, academia, nonprofits, the solar industry, and state government representatives - depending on the type of microgrid, its ownership status and goals are set in the early stages.
- *Conduct audit grade study:* A robust audit grade study is usually necessary to obtain financing for the project development. Type of audits conducted at this step include but are not limited to business models, ownership structure, tax treatment payback period and etc. This stage appoints the ownership structure of the project and operational responsibilities, sources of income and tax advantages, and financial models for the entire life cycle of the project. These studies provide a strong basis for soliciting financing for the microgrid project.
- *Acquire financing:* Different types of financing as well as the role of the existing utility company are identified at this stage. Some different types of financing include equity financing, debt financing, leasing, third party service model, government grant, loans, and tax credits.
- *Acquire necessary approvals and construction:* Some necessary legal requirements before constructing the microgrid include building permits, zoning variances, excavation permits, engineering permit, and other approvals in compliance with state and local regulations, as well as some more specific requirements such as the municipal's approval to run distribution wires across public rights of way. It's important and necessary to acquire all required approvals and permits before starting to construct a microgrid project.

Chapter conclusion: as a literature review, this chapter casts light on the concept of microgrids and briefly describes the policies and technicalities associated with its development. The knowledge gained in this chapter is later used in Chapter 6 as it discusses the spatial aspects of developing solar community microgrids.

Chapter 3 A Software Survey

Introduction: Computational methods are typically used to simulate, evaluate and predict energy performance when designing community microgrid projects. To better understand opportunities and limitations of existing software packages a software survey was conducted and is presented in this chapter. Since the relationship between urban form and community energy performance is the interest of this research, the presented report excludes analysis on tools merely used for urban planning and those specific to designing the technological details of energy systems. Methods of evaluation started with a literature review, exploration of material available from the software development organizations, and in some cases using the software packages. The goal of this software survey is to identify existing software tools, if any, that evaluate energy performance in community microgrids as a result of the interaction between the superstructure and infrastructure of the energy system¹⁶.

3.1 Background and Evaluation Criteria

Existing literature reviews software packages related to community energy modeling and energy systems modeling. A summary of this work informs the assessment criteria for the software survey conducted in this research:

Allegrini et al. (2015) identified a shortage in existing software tools and simulation packages that take into account urban energy systems along with the buildings they serve. This chapter reviews different modeling approaches and multidisciplinary tools that address the supply and demand side of an urban energy system, and their application to district-level design problems. A descriptive analysis of twenty tools with modeling capabilities was conducted in areas relevant to urban energy system design including some general tools such as ‘TRNSYS’ and ‘Modelica’, several specific packages as ‘CitySim’ and ‘SynCity’, and some specialized computational fluid dynamics and geographic information system tools.

Markovic et al. (2011) reviewed various practical tools developed to “*analyze the energy, economic and environmental performance of energy generation systems, buildings and equipment in a community*”. The authors categorized related software tools into three main groups according to their aims and achievements: geographical assessment tools, energy assessment tools, and evaluation assessment tools. The geographic assessment tools such as ArcGIS and Raster Cities, are used to model geographic features of the built

¹⁶ This chapter was published in the Hindawi Journal of Engineering: Rahimian, M., Iulo, L. D., & Duarte, J. M. P. (2018). A review of predictive software for the design of community microgrids. *Journal of Engineering*, 2018.

environment, analyze the availability of renewable energy resources, and locate facilities and infrastructures. Energy assessment tools, such as EnergyPLAN and EnergyPlus, take various distributed energy generation scenarios and analyze the energy consumption rates respectively. The evaluation tools address other aspects of power generation such as life cycle assessment and environmental and socioeconomic analyses. One example of an evaluation tool introduced in this paper is PLACE3S planning for community energy, economic and environmental sustainability.

Connolly et al. (2010) conducted a study on 37 different computer tools that can be used to analyze the integration of renewable energy into various energy systems under different objectives. Their selection of tools span from those used for analyzing a single building energy system to those used for large scale, national energy systems. The analysis criteria used in this paper include the energy sector, accounted technologies, time parameters used, and tool availability. Authors conclude that there is no energy tool addressing all issues related to integrating renewable energies. With their paper, they hope to provide necessary information for decision makers in choosing the most appropriate tool for their purposes.

Gil and Duarte (2010) identified and compared 12 evaluation tools for sustainable urban development suitable for design and analysis at the neighborhood scale. They studied these tools based on their structure, format, and content, focusing on how the evaluation indicators address dimensions of urban form, accessibility, and the neighborhood's spatiality. Some of the tools reviewed by the authors are CityCAD, DPL, ECOCITY, INDEX, and LEED Neighborhood Development.

The missing component in the precedent reviews is the availability of software tools that address the interaction between urban superstructure and the performance of its underlying energy infrastructure. The following section offers an assessment of existing software tools relevant for designing and evaluating this interaction. The assessment is done from a perspective of requiring a tool that addresses how the urban form of a community impacts the energy performance of its distributed energy system. Consistent with the software survey assessments discussed above, the criteria used for evaluating software tools are:

- Functional purpose of the tool, is the tool used for modeling, or simulation/prediction, or evaluation/analysis purposes,
- Type of data input,
- Output format,
- Target users; architects and urban designers or engineers.

3.2 Urban Superstructure Energy Modeling

CitySim: *Developed by Solar Energy and Buildings Physics Laboratory (LESO-PB), Ecole Polytechnique Federale de Lausanne (EPFL) – Lausanne, Switzerland*

CitySim is a large-scale building energy simulation tool based on simplified thermal models (Walter & Kämpf, 2015). This is a Java-based graphical user interface (GUI) supporting the decision-making process of sustainable urban planning. The goal of this tool is to simulate and optimize building-related resource flows (energy, water and waste) and their interrelationships, as well as study their dependence on the urban climate (Robinson, et al., 2009). The simulation input in CitySim is a manual process based on CityGML geometrical databases including site location, associated climatic data, type and age categories of buildings, 3D forms of buildings, definition of energy supply and storage systems. CitySim conducts parsing of data in XML format from the GUI to C++ solver for hourly simulation of the resource flows, and then streams back the analysis results to the GUI in format of graphs, bars, and tables (Figure 3).

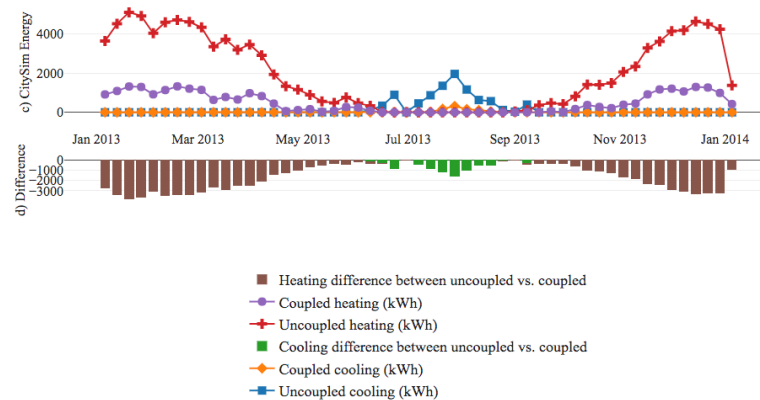


Figure 3- A sample output of CitySim (Source: <https://plot.ly/~ClaytonMiller/112.embed>)

SUNtool: *Developed by Solar Energy and Buildings Physics Laboratory (LESO-PB), Ecole Polytechnique Federale de Lausanne (EPFL) – Lausanne, Switzerland*

SUNtool (Sustainable Urban Neighborhood modeling tool) is the predecessor of CitySim, a Java-based GUI with thermal simulation engines. As a decision-making tool, SUNtool utilizes modelling techniques to predict the performance of various energy generation technologies within the urban context of approximately 50 to 500 buildings. SUNtool is counted as the first of a new generation of simulation tools which supports sustainable master planning (Robinson, et al., 2007) based on simulating resource flows as energy, waste, and water. In the interface, the designer models a 3D geometry of the buildings and defines their type of use. Then, the corresponding dataset is specified for location, climate, occupancy schedule, appliances, glazing ratio, construction, systems, and etc. After selecting the simulation setting and running the simulation, the software outputs four

classes of model: microclimate¹⁷, thermal, stochastic¹⁸, and plant¹⁹. The accuracy of the conducted simulations enables the designer to choose the best design solution in relation to site-specific urban microclimate as well as to human behavior.

UMI (Urban Modeling Interface): *Developed by Sustainable Design Lab at the Massachusetts Institute of Technology (MIT) – Cambridge, Massachusetts, USA*

UMI is a plug-in for the Rhinoceros 3D modeling environment, performing environmental analysis for neighborhoods and cities. This analysis is based on operational and embodied energy use (using EnergyPlus' thermal simulation engines), walkability evaluations (using customized Python scripts), and daylighting potential (via Daysim daylighting simulations) (Reinhert, Dogan, Jakubiec, Rakha, & Sang, 2013). Designers start modeling their building geometries and massing models in Rhinoceros' CAD environment and then assign their models in UMI. These elements may include building envelopes, trees, shading objects, streets and other infrastructures. After inputting models in UMI, material specificities may be defined as well as building usage schedules, construction, and amenity types. As the output, UMI runs individual annual simulations for each modeled building mass, calculates the annual daylight availability for each story in each building, and calculates walkability score based on the defined streets, pathways, and selected urban amenities. The outputs are both reflected back into the 3D model of the Rhinoceros scene, and as generated tables and user-friendly graphical reports (Figure 4).

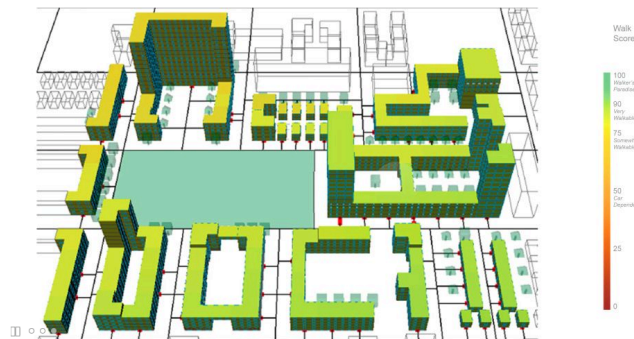


Figure 4 - UMI output sample (Source: <https://architecture.mit.edu/event/modeling-urban-sustainability-energy-daylight-and-walkability>)

¹⁷ Microclimate modeling includes radiation modelling and temperature prediction

¹⁸ Stochastic modeling includes occupant presence, window openings, lights and shading devices, electrical and water appliances, and waste

¹⁹ SUNtool supports a comprehensive plant modeling technique, but it's worth noting that energy resource management, testing and optimizing control strategies, and plant configurations is not one of SUNtools objectives. Also, SUNtool does not support energy storage.

3.3 Energy Infrastructure Energy Modeling

DER-CAM: *Developed by Berkeley Lawrence National Lab – Berkeley, California, USA*

DER-CAM is a decision-support tool for decentralized energy systems. The primary objective of DER-CAM is to run techno-economic evaluation on defined on-site energy generation technologies, CHP, or microgrids, and optimize²⁰ the DER selection and operation through linear programming techniques. The inputs for this tool include customers' end use load profiles, customers' default electricity tariffs and natural gas price, initial investment capital, operating and maintenance cost and specificities, basic physical characteristics of alternative generating, heat recovery and cooling technologies, as well as carbon emission constraints and sensitivity parameters (Lawrence Berkeley National Laboratory (LBNL), 2015). DER-CAM outputs the optimal distributed generation and CHP technologies to be installed with their appropriate level of capacity, the optimal distribution strategy, and the total cost of producing electric and thermal energy (Lawrence Berkeley National Laboratory (LBNL), 2015). Outputs are in the format of tables and graphs (Figure 5).

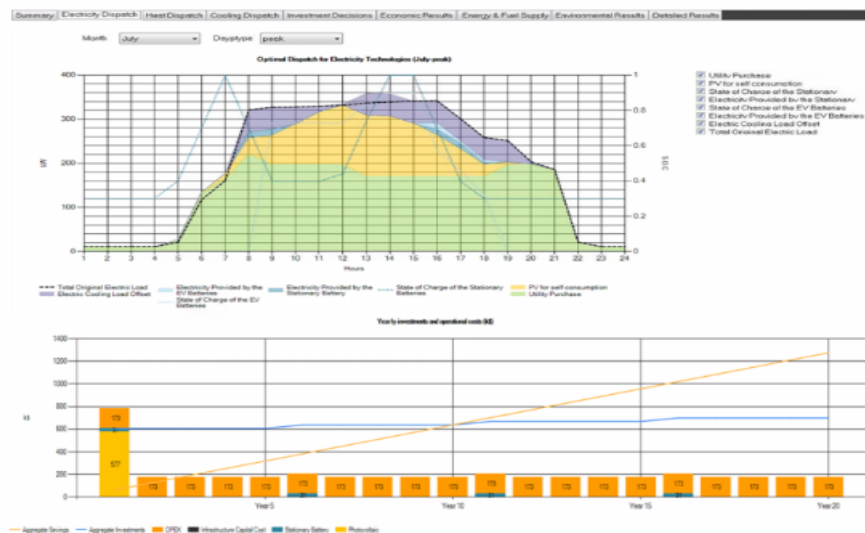


Figure 5 - DER-CAM output sample (Source: <https://building-microgrid.lbl.gov/projects/der-cam>)

²⁰ The output objective of DER-CAM is minimizing the operating costs of on-site energy generations, CHP, or microgrids.

HOMER: *Developed by National Renewable Energy Laboratory (NREL)*

HOMER is developed with the purpose of evaluating grid-connected and off-grid energy systems from an economic and engineering point of view. Simulating the performance of any particular energy system configuration is HOMER's main capability. However, the software is also adequately able to run economic optimization and sensitivity/uncertainty analysis on defined systems. It's worth noting that optimization is done on variables that the designer has control over. Sensitivity analysis is on variables that are subject to uncertainty or change which are out of designer's control such as wind speed and fuel price (Lambert, Gilman, & Lilienthal, 2006). The input data of HOMER include customer's load profiles for electric and thermal energy, any resources and fuel used by the system to generate electric and thermal power, energy system components (generation, distribution, storage, etc.), electric and thermal load curve with up to 1 minute resolution, technical efficiencies, operation and maintenance costs, emission constraints, and sensitivity parameters (Lambert, Gilman, & Lilienthal, 2006). HOMER outputs the evaluation and analysis results in the format of graphs and detailed data reports.

LEAP: *Developed by Stockholm Environment Institute (SEI)*

LEAP (Long Range Energy Alternatives Planning) is an integrated scenario-based energy modeling tool which primarily operates as an accounting system with a capacity for simulation and econometric modeling (LEAP – long range energy alternatives planning, 2017). In these scenarios, a group of input data are set including the electrical and thermal load profiles, the availability of energy resources, operation and maintenance costs, technical efficiencies, and emission constraints. Some assumptions are also being specified on the population growth rate of the energy system, its futuristic economic development, and the interest rate of the energy system. Based on the set assumptions, LEAP evaluates the energy scenario by conducting a physical computing of natural and environmental resources, sensitivity parametric analysis, and integrated energy/environment analysis. The outcome of these analysis is in form of detailed reports and graphs of the energy scenario (Commend – community for energy, environment and development, 2017).

EAM (Economic Evaluation of Microgrids):

EAM is used to evaluate microgrids' economic viability. EAM's objective to optimize the sizing of any microgrid scenario in relation to the equipment unit selection and its corresponding power capacity. The input measures include a 24-hour energy profile of the energy system, selection of microgrid equipment components, their initial costs, as well as utility prices (Mendes, Ioakimidis, & Ferrao, 2011). The analysis output of this software package includes tables, graphs, and charts.

MARKAL/TIMES (Market allocation model and the integrated MARKAL/EFOM system): *Developed by International Energy Agency's Energy Technology Systems Analysis Program*

TIMES and MARKAL share the basic modeling paradigms by being technology explicit and representing equilibrium models of energy markets. However, the development history of MARKAL/TIMES indicates that TIMES is the developed version of MARKAL featuring new analytical capabilities. MARKAL/TIMES is an energy, environmental, and economic evaluation tool that analyzes user-defined energy-environment systems at the global, national, state/province, or community level and over a long period of time (up to 100 years) as a representation of their evolution (Connolly, Lund, Mathiesen, & Leahy, 2010). General input data include demand curve, renewable energy resources, energy station capacities, cost and number of different regulation strategies emphasizing import/export and excess electricity production. Policy scenarios can also be included as input data with measures on cutting emissions, promoting energy efficiency, improving energy security, and reducing new technical costs. The input data should represent the evolution of the system over time in any user-defined time resolution such as seasonal, monthly, weekly, and hourly time measures (Mendes, Ioakimidis, & Ferrao, 2011). With this data, MARKAL/TIMES models the entire energy system in the defined policy scenarios and depicts all possible flows of energy in different phases of extraction/generation, transformation, distribution, and consumption in the format of tables, charts, and graphs. The software then finds the “best” energy system technical mix at each period of time that meets the demand at minimum cost within the limits of all imposed policies and physical constraints (Mendes, Ioakimidis, & Ferrao, 2011).

RETScreen Clean Energy Project Analysis Software: *Developed by the Natural Resources of Canada*

RETScreen is a decision-support tool which performs comparisons between any given different energy system scenarios. The software requires three sets of data to input: 1- project specific general information including site location and energy system specificities (generation types, demand load, technical features), 2- project economics such as initial, annual, and periodic costs, 3- user specific financial parameters including greenhouse gas emission reduction credits, incentives, inflation, discount rate, taxes, and etc. (Leng, 2000). After inputting the projects' specificities, the software runs comprehensive identification, assessment, and optimization of the technical and economic viability of the proposed scenarios, as well as measuring and verifying the resultant greenhouse gas emission reductions (Allegrini, et al., 2015). The output is in form of summary worksheets, and statistical data visualizations such as graphs, charts, and tables

PV-Design Pro: *Developed by Maui Solar Energy Software Corporation – Haiku, Hawaii*

PV Design Pro is designed to simulate photovoltaic energy system operation with the purpose of evaluating PV designs more effectively and maximizing their performance. This software requires modeling and laying out the energy system's components and configuration by inputting climatic loads for different time resolutions, specifying number of panels and electrical characteristics of the PV arrays, designing the wiring configuration, and setting the battery and inverter qualities (Maui Solar Energy Software Corporation, 2017). The software evaluates the modeled solar energy system, by computing the monthly percentage of electricity generated by the PVs, monthly battery states-of-charge, annual energy costs analysis, life cycle financial analysis, and hourly analysis of the system loads and battery state-of-charge. The outputs are in form of detailed reports, charts, and graphs (Figure 6).

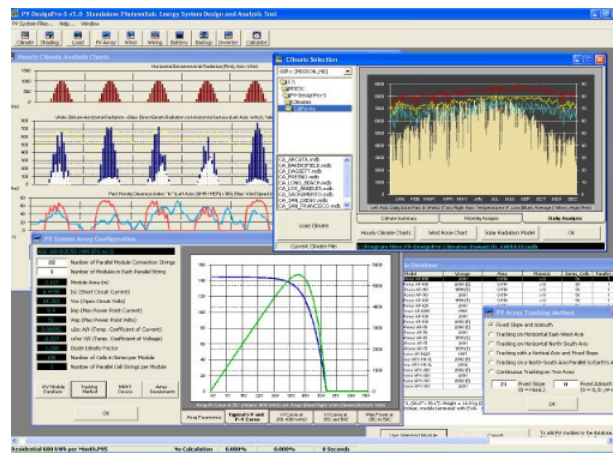


Figure 6 - PV-Design Pro output sample (Source: <http://www.maui-solar-software.com/>)

PV*SOL: Developed by Valentin Software - Berlin, Germany

PV*SOL is a dynamic PV simulation program with 3D visualization and detailed shading analysis of photovoltaic energy systems (PV*SOL Premium, 2017). As input data, PV*SOL enables simple 3D design and modeling of buildings (Figure 7) along with their featured arrays of photovoltaic panels, modeling and circuit design of the grid along with its mechanical components, dimensioning AC and DC wirings, and implementing the costs of PV components. By adding climatic data of the site location and defining the system features, PV*SOL is enabled to simulate the energy produced by the photovoltaic panels and optimize the energy system's configuration while performing economic analysis

(PV*SOL Premium, 2017). The output of this analysis includes 3D visualization (Figure 8), detailed reports, graphs and charts.

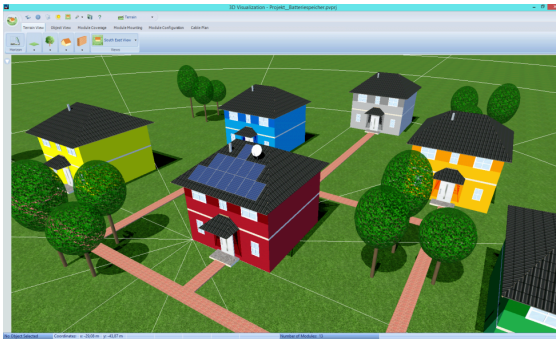


Figure 7 - PV*SOL 3D building modeling sample (Source: <http://solarprofessional.com/articles/design-installation/performance-modeling-tools-overview/page/0/2>)

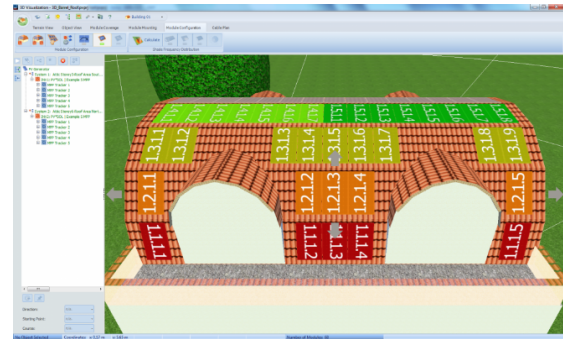


Figure 8 - PV*SOL 3D visualization of PV energy evaluation (Source: <http://www.valentin-software.com/en/products/photovoltaics/57/pvsol-premium>)

3.4 Discussion

This software survey indicates two distinct categories of modeling, simulation, and evaluation tools that respond to a multidimensional and building-integrated definition of community microgrids (Table 1): (1) energy infrastructure modeling, (2) urban superstructure energy modeling. Software tools from the first category are mostly concentrated on evaluating and optimizing the cost, selection, and operation of individual components of generation, distribution, and storage technologies of distributed energy systems as microgrids. The second category of software tools are used to model, predict, and simulate operational energy use for groups of buildings at the urban level.

Based on this review there is a lack of software tools and simulation packages that simultaneously address the necessary interaction between building in communities and energy performance. Software tools like SUNtool use the 3D modeling of buildings to predict and simulate the performance of photovoltaic energy generators. The limitation of this tool is that it predicts the performance of only one components of an energy system, the energy input. Analysis on how urban form impacts energy output including energy consumption and energy loss is not taken into account. CitySim on the other hand considers energy output in addition to energy supply but does not consider energy loss. The analysis of CitySim is on weather data, and not on how the urban structure of a design impacts the flow of energy. Since these tools are designed for architects and urban planners, the input of the software is building geometries.

	Software	Input	Output Type	Functional Purpose	Target User
Superstructure (buildings in communities)	CitySim	Site location Climate data Building geometry Building type Building age Energy supply and storage systems	Graphs Bars Tables	Simulation/Prediction	Architects Urban designers
	SUNtool	Site location Climate data Building geometry Building type Building age Selection of renewable energy and water processing technologies Occupancy schedule	XML documents 3D false color plots 2D graphs	Simulation/Prediction Optimization	Architects Urban designers
	UMI	Building geometry Massing models (building envelopes, trees, shading objects, other infrastructure, and streets) Building material Occupancy schedule Building construction types	3D visualizations Graphs Charts	Simulation/Prediction	Architects Urban designers
Infrastructure (electrical installation)	DER-CAM	Load profiles Electricity price data Operating costs Maintenance costs Fuel cost Physical characteristics of alternative generating, heat recovery and cooling technologies parameters. Emission constraints Sensitivity parameters	XML and HTML documents Reports Graphs	Evaluation/Analysis Optimization	Engineers
	HOMER	Load profile Energy input resources type and details Electrical components type and details Emission constraints Operating costs Maintenance costs Sensitivity parameters	XML and HTML documents Reports Graphs	Simulation/Prediction Evaluation/Analysis Optimization	Engineers
	LEAP	Load profile Energy input resources type and details Electrical components type and details Operating costs Maintenance costs Interest rate Population growth rate Emission constraints	Report	Evaluation/Analysis	Engineers
	EAM	Load profile Representative values of equipment performance, Equipment initial cost Utility price	Tables Graphs Charts	Evaluation/Analysis	Engineers
	MARKAL/TIMES	Load profile Renewable energy input resources type and details Energy station capacities, costs	Tables Graphs Charts	Evaluation/Analysis	Engineers
	RETScreen	Site location Load profile Energy input resources type and details Electrical components type and details Energy system costs Financial parameters related to the avoided cost of energy production credits GHG avoided cost of energy, production credits, GHG emission reduction credits, incentives, inflation, discount rate, debt, and taxes.	Tables Graphs Charts	Evaluation/Analysis	Engineers
	PV-DesignPro	Climate data Load data (weekday and weekend, AC and DC), PV modeling (number of panels, electrical characteristics) Electrical modeling	Reports Graphs Charts	Evaluation/Analysis	Engineers
	PV*SOL	Climate data Buildings geometry PV modeling Modeling and circuit design of the grid with all its electrical components	3D visualization Reports Graphs Charts	Simulation/ Prediction Optimization	Engineers

Table 1- Summary table of the conducted software survey

The other category of software studies, the energy system modeling tools, are precisely focused on the operation of the entire energy infrastructure, that is the mix of all electrical components. Tools as HOMER and DER-CAM optimize the mix of different electrical components in the infrastructure based on overall cost of operation and maintenance. This is while as seen in Table 1, most of these software packages only report an analysis of performance and don't generate solutions. While the main users of these tools are engineers, unlike the first software category introduced, the input for the system are not building geometries but are load energy consumption profiles. Thus, the impact of urban form in predicting the performance of the energy system is completely ignored in such tools.

The software survey studied in this section shows that although some urban superstructure energy modeling tools attempts to address the impact on energy performance, a comprehensive comparison of energy input to energy output has been neglected. Also, the impact of superstructure in the software is limited to individual building geometry and occupancy schedules, and not on the effect of the entire community's urban form on energy performance. Although the energy infrastructure modeling tools thoroughly evaluate the energy performance of the infrastructure, they clearly do not input the superstructure as an affecting factor in the analysis. The coinciding impact of urban form on energy demand, energy supply, and energy waste is not simultaneously addressed in the software tools evaluated.

As discussed in Chapter 1, studying the energy performance of community microgrids in isolation from their comprising buildings is insufficient. A specific understanding needs to be gained on how the features of urban form can simultaneously maximize onsite renewable energy generation (specifically photovoltaic energy due to its popularity in existing settings) in community microgrids and minimize the community's energy demand. Such a multimodal study of community-microgrids requires a new generation of computational modeling, simulation, and evaluation tools.

Chapter conclusion: The reviewed software survey identifies a gap in existing software tools that simultaneously address the necessary interaction between the superstructure and infrastructure of community microgrids, given the importance of its impact. The lack of such tools in the field is perhaps one of the reasons why architects and urban planners have not been actively involved in community microgrid development processes. The presented software investigation serves as the basis for developing an experimental software prototype (as discussed in Chapter 1) that bridges this gap by predicting the energy performance of any given community microgrid design scenario by the virtue of its urban spatial configuration useable for the building and community design and planning sector. The developed software prototype is described and elaborated upon in Chapter 8.

Chapter 4 Urban Form and Energy in Community Microgrids

Introduction: Chapters 2 and 3 presented an overview of what microgrids are as well as the existing tools for modeling and simulating their performances. Through these chapters we understood that no studies have pointed out to the spatial aspect of community microgrids and how it affects its energy performance, hence the unavailability of related modeling and simulation tools for this matter.

In this chapter, we set aside the literature and background given on the current state of microgrids and start reviewing studies that researched the effect of different indicators of urban form on urban scale energy performance, both onsite solar energy capture and energy demand for buildings' operation. The goal of this chapter is to answer the first research question: *“The spatial geometry of urban form is a complex entity and is defined by many spatial attributes. Which of these attributes have an energy-relevance regarding the community-wide supply and demand of energy and how are they measured?”*.

Towards answering this research question and through studying the literature, this chapter identifies and gathers all energy-relevant attributes of urban form that previous researchers have studied their influence on urban scale energy performance along with their metrics of measurement for quantifying the urban form.

4.1 A Brief Background on Urban Form and Energy in Communities

Literature verifies the need for a multidimensional vision on community microgrids, relating the spatial structure of urban form to its energy performance in different cities and communities (Steadman, 1977; Owens, 1986). This literature comes in two categories; some assess the feasibility of utilizing various renewable energy resources as a derivative of urban form (Amado & Poggi, 2014; Compagnon, 2004; Lobaccaro & Frontini, 2014; Robinson, et al., 2007; Sarralde, Quinn, Wiesmann, & Steemers, 2015), and the other category studies the impact of urban form on the amount of energy required for building operations specifically for space heating and cooling as main source of energy consumption in buildings (Ewing & Cervero, 2010; Fayyad, Piatetsky-Shapiro, & Smyth, 1996; Lariviere & Lafrance, 1999; Pont & Haupt, 2005; Silva, Oliveira, & Leal, 2017; Ratti, Baker, & Steemers, 2005; Reinhert, Dogan, Jakubiec, Rakha, & Sang, 2013; Silva, Horta, Leal, & Oliveira, 2017).

Solar energy is one of the main sources of renewable energy available. Therefore, studies considered herein explore the relationships between different attributes of urban form and the potential to harvest solar energy within the urban context and communities to generate photovoltaic energy. For example, a study by Sarralde et al. (2015) on different neighborhoods in London, shows how optimizing a combination of nine spatial attributes of urban form (including share of *semi-detached houses*, *average building height*, *share of*

area covered by private gardens, site coverage, average building perimeter, average distance between buildings, standard deviation of building heights, plot ratio, average distance between buildings) could increase the solar irradiation of roofs by 9% while that of facades by up to 45%. Research by Robinson et al (2007) examines the effect of *urban morphology* and indicators of radiation availability examining *the sky view factor, mean canyon height to width ratio*, and the *urban horizon angle* of three Swiss districts. Compagnon (2004) looks at how the *orientation* of 61 buildings in Perolles area of Fribourg (Switzerland) challenges the potential for capturing solar energy by the building facades and roofs. Moreover, a paper by Lobaccaro and Frontini (2014) examines attributes of building *densification* and *shading* in urban environments as effecting factors for solar availability and therefore potential for utilizing photovoltaic panels in certain communities and neighborhoods.

Significant changes in land use pattern and urban form follows urbanization (Bettencourt, Lobo, Helbing, Kühnert, & West, 2007), and consequential transformations to the patterns of energy consumption in cities (Bastiononi, Pulselli, & Tiezzi, 2004; Dodman, 2009; Grubler, et al, 2012; Jebaraj & Iniyan, 2006). The ‘Contribution of Working Group III of the Fifth Assessment Report of the Intergovernmental Panel on Climate Change’ (Seto, et al., 2014) ranks urban form as the fifth contributor²¹ to greenhouse gas emissions in cities, due to the impact of urban form on patterns of mobility and the energy required for space heating and cooling in individual buildings (Owens, 1986).

Newman & Kenworthy (1989) assert that research on urban form and energy demand for traveling purposes in cities has been largely investigated throughout the years. But to date not much literature has been dedicated to exploring the effects of urban form on energy demand in buildings and communities. Analysis in this area started in the 1960s at the Centre for Land Use and Built Form Studies at the University of Cambridge (Ratti, Baker, & Steemers, 2005). While form is not the only driver of energy demand in the built environment, there is evidence suggesting that the significance of its impact is mainly associated with urban heat island effect, change in local wind pattern, building thermal comfort, and energy conservation (Chatzidimitriou & Yannas, 2015; Reinhert, Dogan, Jakubiec, Rakha, & Sang, 2013; Santamouris, Papanikolaou, Livada, & Koronakis, 2001; Silva et al, 2017).

Urban form has a manifold of energy relevant spatial attributes. Researchers have been attempting to find the effect that each individual attribute of urban form has on energy demand in buildings. Depending on the attribute, a number of research studies have concluded that there is a low impact on energy demand in buildings. But researchers such as that by Silva et al. (2017) and Ewing and Cervero (2010) emphasize the importance of

²¹ This report ranks economic geography and income, sociodemographic factors, technology, and infrastructure as first four contributors to greenhouse gas emissions in urban areas.

a comprehensively studying the combined effect of all relevant urban attributes in order to understand the relationship between urban form and energy demand in communities. They argue that the isolated effect of each individual attribute of urban form has a relatively small effect on the overall energy demand in cities and communities. A multidimensional study of this relationship affirms the complexity associated with considering all energy relevant attributes of urban form (Owens, 1986). Nonetheless, according to Silva et al. (2017) research on the relationship between urban form and energy demand has been increasingly attracting attention during the last decades.

Density has been the most cited attribute of urban form affecting energy demand due to its influence on urban heat island (Silva, Oliveira, & Leal, 2017). Owens (1986) emphasizes *siting* and *orientation* as two important attributes since they can be adjusted to benefit from the site's microclimatic factors and free ambient energy resources. Owens (1986) also determines the importance of a community's *overall surface area (volume ratio)* as an indicator for energy demand where communities with lower surface areas, volume ratio, tend to consume less energy. Moreover, a community's *orientation* and *layout* can change regional wind patterns impacting the rate of passive cooling and natural ventilation in individual buildings. In addition to density, layout, siting, and orientation, Silva et al. (2017) find other urban attributes - including *diversity*, *green areas*, *passivity*, and *shading* - related to energy demand in buildings. In another paper authors identify the *number of floors*, *mix of uses*, and *floor area*, as the most energy relevant features of urban form (Silva, Horta, Leal, & Oliveira, 2017). Ratti et al. (2005) studies the effect of *urban geometry* which resulted in having a low impact on energy demand, but since it has been studied in isolation from other attributes, it's important to explore its significance within an inclusive framework.

4.2 Identification of Urban Form Features Relevant to Energy Performance of Solar Community Microgrids

The purpose of this research is to discuss the impact of different configurations of urban form on energy performance in solar community microgrids, which can be accomplished without detailed building characteristics. It's important to note that in this study, the assumption is that the community's energy performance is primarily dependent on the urban form and not the construction type or age of each individual building. In the past research, the complex effects of urban form on net PV energy generation and energy consumption have not been rigorously and comprehensively evaluated due to computational limitations and the lack of data rich environments. A quantitative translation of urban form is essential to find its contribution to patterns of energy consumption and solar energy capture in communities. From all the attributes suggested by researchers, the ones that will impact the supply and demand of energy in solar community microgrids have been selected and will be further studied. The bullet points below suggest a definition of these attributes based on previous literature along with relevant metrics of measurement

for the purpose of quantifying the urban form in community microgrids. The selected attributes range from those which previous researches have proved significant for energy demand and those that have not been deemed influential. The reason for selecting all energy-relevant spatial attributes is to understand their interaction and influence when they are in confluence as opposed to studying each attribute in isolation:

- *Density*: density has been the most researched urban attribute influencing energy demand and solar energy capture in communities (Silva, Oliveira, & Leal, 2017). Multiple descriptions of density exist depending on the intent of the research. These definitions vary from the density of the physical built environment to the density of people living or working in a given area (Ko, 2013; Silva, Oliveira, & Leal, 2017). Density in this research deals with land-use intensity and is measured per unit of area. Density is treated as a driver of energy demand by influencing the urban heat island effect and wind flow in the urban context. Denser urban environments increase the local temperature and thus increase cooling loads. Depending on the geographical location and site-specific weather conditions the urban heat island effect may be constructive or disruptive for energy demand in buildings (Taha, 1997). In the context of community microgrids, communities with higher densities facilitate the introduction of combined heat and power (CHP) systems in particular contexts (Owens, 1986).

Different density measurements of the physical built environment exist that take the quotients of a fraction problem where the denominator is the total ground area of the land being measured. The numerator of a density fraction could vary from gross floor area and gross building footprint area, to number of rooms and number of buildings. The measurements of density used herein are adopted from the Spacemate research (Pont & Haupt, 2005) and the literature cited by Silva et al (2017):

- Floor Space Index (FSI): gross floor area/total ground area
 - Ground Space Index (GSI): gross area of building footprint on the ground/total ground area
 - Open Space Ratio (OSR): gross area of unbuilt ground/total ground area
 - Layer (L): average number of floors/total ground area OR [total number of floors/number of buildings]/total ground area
 - Network Density (N): length of the network/total ground area
- *Compactness*: density and compactness have very close definitions and sometimes they are used interchangeably. Ko (2013) describes compactness as how tightly buildings stand on site. The main difference between compactness and density is that measurements of density are performed in comparison to the total area of study

while compactness considers the width of the street, distance between buildings, and the height of buildings.

Compactness directly impacts solar access and wind flow pattern in an urban environment and therefore accounts for the thermal comfort in buildings. For example, urban areas with wider streets welcome more solar access as well as provide natural ventilation, while narrow streets create a wind tunnel effect. Depending on the climatic zone of the urban area, the compactness of a community or neighborhood can have different consequences for a microgrid's operation. For example, a compact form in a cold region can increase the heating demand in buildings as it blocks the solar access, while also being possibly limiting for onsite PV energy generation as one building overshadows the adjacent ones. In such cases, other aspects of community design and layout should be carefully considered in order to manage solar access for maximum passive heating and potential PV energy generation. The compactness of a neighborhood could be quantified by measuring its aspect ratio (Ko, 2013):

- Aspect ratio (AR): average building height/average street width

Compactness is also used as an indicator of building geometry (Silva, Oliveira, & Leal, 2017). The building geometry is an important feature for energy demand since buildings' exposed surfaces directly impact heat flows between the inside and outside, as well as access to natural daylight. Researchers claim that the optimum shape for a building to minimize heat loss and maximize daylight gain is a cube (Ratti, Baker, & Steemers, 2005) and deviations from a cubic shape results in increasing heating loads. Three different but related measures of building compactness are suggested by researchers that can be applied at the community scale as well (Ratti, Baker, & Steemers, 2005; Bourdic, Salat, & Nowacki, 2012):

- Volumetric compactness (STV): envelope surface area/building volume
 - Size factor (SF): building volume^{1/3}
 - Form factor (FF): envelope surface area/(building volume)^{2/3}
- *Diversity or land use mix*: is the second most cited attribute of urban form impacting energy demand in urban areas. However, most literature citing diversity study its effect on traveling demand in urban areas. A diverse neighborhood or community is claimed to decrease the need for motorized travels as it brings urban activities closer to the residential context. This research is not concerned with energy demanded for traveling purposes but considers diversity as the mix of different land uses in a community. Diversity is an important feature when planning a community microgrid, since the diversity of the building load types is likely to

regulate the energy consumption peak hour, so the microgrid infrastructure does not face high energy consumption periods of time. For example, a decentralized shared energy system with a complementary mix of land uses is suggested to be more economically beneficial due to balancing the peak hour of energy consumption (Bourgeois, Gerow, Litz, & Martin, 2015). Diversity as land use mix is measured by the mixed-use index as proposed by van den Hoek (2008):

- Mixed-Use Index (MXI): gross residential floor area/gross floor area
- *Green areas*: the existence of green areas is not particularly a spatial feature of urban form but they are advocated to affect the urban microclimate by avoiding the urban heat island effect and consequently resulting in less energy demanded particularly for space cooling²² (Silva, Oliveira, & Leal, 2017). Depending on the geographical region of the urban area and the location where the green areas are planted, the presence of trees could also be beneficial by providing shading and solar gains and blocking unwanted wind in certain seasons. In the scale of a community microgrid, simple metrics adopted from Silva et al. (2017) and Vaz Monteiro et al. (2016), may be useful for quantifying green areas incorporating their width, size, and geometry:
 - Green Space Density (GSD): gross green space area/total ground area
 - Green Area Geometry (GAG): gross green space perimeter/gross green space area
- *Orientation*: is an easily addressed building design feature that can help with buildings' solar gains specifically for potential PV energy generation onsite and passive solar heating. In northern latitudes, south-facing facades are generally the most desirable position to maximize solar access. North-facing facades have the lowest solar gains, whilst east and west orientations are exposed to direct solar gains in the morning and late afternoons. In research conducted by Hemsath (2016), the annual energy use and cost of 7000 typical Midwestern suburban homes were simulated in four different climate regions. The study's analysis show that the cost implications of an individual home's orientation is not noticeable, whereas in the community scale regardless of the climatic region, important saving are conceived in the aggregated energy usage and costs. Basing upon Hemsath's (2016) research results, optimizing the solar orientation of a community microgrid during the planning and design phase, could potentially bring considerable reductions in the community's net energy usage and cost, that could possibly lead to longer periods

²² The type of vegetation in green areas also play an important role in mitigating the urban heat island effect but since vegetation is not a feature of urban form it won't be discussed herein.

of islanding.

A single building's orientation is measured by determining the building's longest axis and calculating the azimuth (Wilson, 2013). When considering the dominant orientation of a community, in addition to measuring the average orientation of all buildings, considering the street orientation is also important as it is claimed to influence various local conditions such as the urban heat island effect, shading, and ventilation of urban canyon (Coseo & Larsen, 2014). A street's orientation is determined by verifying its direction. Measurements of a community's orientation used in this research is as following:

- Community buildings' orientation: sum of all buildings' orientation/number of buildings
- Street orientation: sum of all streets' orientations/number of streets
- *Shading*: is an indicator to evaluate the effects of overshadowing by adjacent buildings as it significantly impacts the energy requirements in buildings as well as the placement and potential for PV energy generation onsite. Baker, Hoch, and Steemers (1992) and Ratti, Baker, and Steemers (2005) have done extensive research on quantifying shading in urban environments. Urban horizon angle (UHA) and obstruction sky view (OSV) are two indexes resulted from their explorations. UHA is the average elevation of the skyline from the center of the façade being considered and OSV quantifies the luminance of the obstructing facades and are measured as below. Moreover, another important factor to consider when calculating shading is to identify *"the ratio of radiation received by a planar surface from the sky or that received from the entire hemispheric radiating environment"* (Jie, Yufeng, & Qinglin, 2013), also known as the measurement of the sky view factor (SVF). The value of SFV is important as it defines if the radiation that is released by building surfaces is blocked by obstructions or received by the sky; $SVF=1$ means that the radiation released by a surface is totally received by the sky and $SVF=0$ means the radiation is totally blocked by surrounding obstructions (Jie, Yufeng, & Qinglin, 2013). Therefore, the value of SFV is impacted by the urban form and distribution of building in the community. Mirzaee et al (2018) discovered a mathematical model that calculates the average SVF for an urban area as opposed to a specific point. This mathematical model is a factor of average building density and average building height of a neighborhood. The average SVF in an area increases when average height of buildings increases as well as when the building density of an area increases.
 - UHA: average height of the opposite skyline/canyon width = $\tan(\text{UHA})$
 - OSV: average height of the opposite skyline/canyon width = $\cos(\text{OSV})$
 - Sky View Factor (SVF): is the value of radiation received by building

surfaces. SVF value of 1 means that the radiation released by a surface is totally received by the sky and SVF value of 0 means the radiation is totally blocked by surrounding obstructions

- if average building height > 25m then $SVF = 1.56 - (0.00572 * H * D)$;
- if average building height < 25m then $SVF = 0.9502 + (0.00042 * H) + (0.0198 * D) - (0.0065 * H * D)$

where H is the average height of the community buildings and D is the ground space index

- *Passivity*: is a condition of urban form that benefits from the site's ambient energy resources (solar and wind) to naturally light, ventilate, and heat building spaces. The measurement of passivity indicates the rate of passive zones (parts of a building which can be naturally lit, ventilated, and heated) in a neighborhood (Ratti, Baker, & Steemers, 2005). A simple rule of thumb for identifying the passive zones in a building is to identify those perimeter parts of each building (each floor) that lie within 6 meters (or twice the ceiling height) from the façade that can let in natural daylight and airflow for ventilation (Ratti, Baker, & Steemers, 2005). In order to calculate the passivity of a neighborhood, the perimeter of all passive zones for each floor of each building needs to be considered. Additionally, Steadman, Evans, and Batty (2009) argue that complementary to the passivity ratio, measurements of the building depth is also an important indicator for air conditioning requirements in buildings. Therefore, for calculating passivity of a community in this research two formulas are used:
 - Passivity ratio: net perimeter of all passive areas in a community/ net perimeter of all non-passive areas in a community
 - Plan depth: net volume of all buildings in a community/net area of all exposed walls in a community

Chapter conclusion: In this chapter we learned that urban form can be decoded and measured through a set of spatial attributes, indices, and metrics. By reviewing past literature, a comprehensive number of spatial attributes were selected and the reasoning behind their selection as well as their relevance to energy performance in community microgrids were discussed. The selected attributes along with their indices ranged from those which previous researchers have proved significant for energy demand and PV energy production and those that were not deemed influential. The reason for selecting all energy-relevant spatial attributes was to understand their interaction and influence when they are in confluence as opposed to studying each attribute in isolation. Attributes of urban form considered along with their indices and metrics of measurement, selected from a comprehensive literature review are outlined in Table 2. Measuring and combining these nineteen indices for each community microgrid project or for any community-scale urban setting for that matter, will give it a unique urban form and spatial footprint identifier which

is specific to that setting. The content of this chapter and its summary outlined in Table 2 answers the first part of the first research question outlined earlier in the dissertation: “*The spatial geometry of urban form is a complex entity and is defined by many spatial attributes. Which of these attributes have an energy-relevance regarding the community-wide supply and demand of energy and how are they measured?*”

Attribute	Index	Metric
Density	Floor Space Index	$(\text{Gross floor area}) / (\text{Total ground area})$
	Ground Space Index	$(\text{Gross area of building footprint}) / (\text{Total ground area})$
	Open Space Ratio	$(\text{Gross area of unbuilt ground}) / (\text{Total ground area})$
	Layer	$(\text{Average number of floors}) / (\text{Total ground area})$
	Network Density	$(\text{Length of the street network}) / (\text{Total ground area})$
Compactness	Aspect Ratio	$(\text{Average building height}) / (\text{Average street width})$
	Volumetric Compactness	$(\text{Envelope surface area}) / (\text{Building volume})$
	Size Factor	$\sqrt[3]{\text{Building volume}}$
	Form Factor	$\text{Envelope surface area} / \sqrt[3]{\text{Building volume}^2}$
Diversity	Mixed Use Index	$(\text{Gross residential floor area}) / (\text{Gross floor area})$
Green Areas	Green Space Density	$(\text{Gross green space area}) / (\text{Total ground area})$
	Green Area Geometry	$(\text{Gross green space perimeter}) / (\text{Gross green space area})$
Orientation	Community Building Orientation	$(\text{Sum of all buildings}^{\wedge} \text{orientation}) / (\text{Number of buildings})$
	Street Orientation	$(\text{Sum of all streets}^{\wedge} \text{orientation}) / (\text{Number of buildings})$
Shading	Urban Horizon Angle (UHA)	$(\text{Average height of the opposite skyline}) / (\text{Canyon width}) = \tan (\text{UHA})$
	Obstruction Sky View (OSV)	$(\text{Average height of the opposite skyline}) / (\text{Canyon width}) = \cos (\text{OSV})$
	Sky View Factor	$\text{if average building height} > 25 \text{ then } \text{SVF} = 1.56 - (0.00572 * H * D)$ $\text{if average building height} < 25 \text{ then } \text{SVF} = 0.9502 + (0.00042 * H) + (0.0198 * D) - (0.0065 * H * D)$ <i>where H is the average height of the community buildings and D is the ground space index</i>
Passivity	Passivity Ratio	$(\text{Net perimeter of all passive zones in community buildings}) / (\text{Net perimeter of all nonpassive zones in community buildings})$
	Plan Depth	$(\text{Net volume of all community buildings}) / (\text{Net area of all exposed walls in community})$

Table 2 - Selected attributes of urban form along with their index and metric.

Chapter 5 Community Energy Consumption and Urban Form: A Machine Learning Model

Introduction: As described in Chapter 4, the main shortcoming with past research on studying building-related energy use at urban scale lies in overlooking the multidimensionality of urban form and excluding the convolutional nature of its spatial indicators. The main objective of this chapter is to offer a comprehensive framework to examine the impact of urban form — as a set of intertwined spatial variables — on the amount of energy that is consumed at urban scale. The suggested framework benefits from artificial neural networks to factor in the many dimensions associated with the spatial structure of urban form. The use of artificial neural networks as a powerful tool handling nonlinear relationships and non-normal data distributions for describing the relationship between urban form and energy consumption, in addition to the specific architecture and hyperparameters used for structuring the artificial neural network model, answers to the second research question of this study: *“what is the most appropriate model for describing the desired relational pattern between urban form and energy performance of microgrid-connected communities?”*. The comprehensive analysis offered herein is reached by studying the combined impact of nineteen different spatial attributes of urban form on energy consumption rather than the influence of each individual spatial attribute²³.

5.1 Why Machine Learning and Artificial Neural Networks?

Machine learning is a sub-branch of artificial intelligence that studies algorithms that can accomplish a given task by observing a series of examples of how that task is being solved rather than explicitly programming the solution. Machine learning algorithms rely on statistical methods to gradually improve their performance through an iterative cycle of experiences. Experience, in this context, may refer to a dataset of unlabeled data (unsupervised learning), a dataset of labeled data (supervised learning), a series of simulations (reinforcement learning), or a series of demonstrations (learning from demonstration).

Machine learning methods are known for working as a ‘black box’ where a machine learning algorithm is constructed and ‘trained’ to discover the relational pattern and hidden structures in the input data that the user provides (Tardioli, Kerrigan, Oates, O'Donnell, &

²³ This chapter along with the section 6.1 of Chapter 6 has been published in the Design Computing and Cognition 2020 conference as: Rahimian, M., Cervone, G., Duarte, J. P., Iulo, L. D., (2020). A Machine Learning Approach for Mining the Multidimensional Impact of Urban Form on Community Scale Energy Consumption in Cities. In *Design Computing and Cognition '20*. Springer, Cham.

Additionally, an extended version of this chapter along with Chapter 8 is currently in process for submission to Artificial Intelligence for Engineering Design, Analysis and Manufacturing (AIEDAM) journal.

Finn, 2015). A machine learning model is the outputted mathematical artifact that explains the discovered relational pattern in the input dataset and can be further used to make data-driven predictions and decisions on unseen examples (Tardioli, Kerrigan, Oates, O'Donnell, & Finn, 2015). In other words, machine learning is the construction of algorithms that can learn relationships from data through building mathematical models and make data-driven predictions or decisions accordingly.

Adopting a machine learning method is adequately useful when the problem of interest is multidimensional and complex, and solving them with conventional engineering or statistical models are extremely time and resource consuming. In the case of this study, researchers (Ewing & Cervero, 2010) have marked the importance of studying the *combined* impact of different attributes of urban form on energy consumption rather than the influence of each individual spatial attribute. When studying the combined effect of urban spatial attributes, the problem becomes too complex for solving it with closed-form solutions. Gil et al. (2012) mark data mining as a sufficient method for the analysis of multidimensional relational complexity of urban environments. Therefore, due to the high-dimensionality, complexity, and computational intensity of finding the relationship between the combination of all relevant spatial attributes of urban form and energy consumption in communities, a machine learning approach, more specifically artificial neural networks, is selected for this purpose.

Artificial neural networks are a widely used machine learning technique for modeling and forecasting building energy consumption (Magoules & Zhao, 2016). According to Goh (1995) “*a neural network is a computer model whose architecture essentially mimics the knowledge acquisition and organizational skills of the human brain.*” Artificial neural networks are deemed as an effective approach to a complex application as this study since they can handle non-normal data distributions and non-linear relationships. Details on the operation of artificial neural networks is explained further in this chapter.

5.2 Preparing the Dataset

A structured dataset is the first requirement working with any machine learning problem, including artificial neural networks, in order to explore the desired relational pattern. Preparing the dataset requires 1. collecting raw data and, 2. processing and structuring it towards the designated research purposes.

5.2.1 Data Collection

As mentioned in Chapter 1, San Diego county of California has been selected as a case-study mainly due to the availability of county-wide energy consumption data. San Diego's main utility company, SDG&E, has aggregated energy consumption information for the entire county from 2012 onward and has allocated them per zip code and customer type (residential, commercial, industrial, and agricultural). Moreover, after Los Angeles, San

Diego is the second highest ranked city for total installed solar capacity as of 2019 (Bradford, Stankiewicz, Sundby, Fanshaw, & Sargent, 2019) with the ambitious goal of generating 100 percent of its electricity from renewable sources by 2035 according to the “City of San Diego Climate Action Plan” (The City of San Diego, 2015).

In order to create the dataset, the first step is to collect the raw data which for this study includes:

- Spatial data: obtained through San Diego’s online regional GIS data source (<http://www.sangis.org/download/>) including shapefiles on the county’s zip codes, parcels, building footprints, street network, parks, topography, and zoning.
- Energy consumption data: which has been published online by SDG&E (<http://www.energydata.sdge.com/>), as mentioned above.

5.2.2 Data Processing

The collected raw data then need to be processed and structured in a way that best suits the intended analysis. As the first step in this phase, the spatial unit of analysis needs to be specified. Since energy consumption data is aggregated per zip code, the spatial unit of analysis is set to zip code; this means that the spatial data and urban form need to be granulated and measured per the defined unit respectively (i.e., buildings footprints per each zip code, parks in each zip code). The procedure of data processing is broken down into three steps of 1. preprocessing the spatial data, 2. preprocessing the energy data, and finally 3. combining the two preprocessed datasets into one and cleaning it.

(i) Preprocessing spatial data

Urban form can be decodified and measured through a set of spatial indices and metrics. This research is based on a comprehensive number of spatial attributes extracted from past literature. The selected attributes range from those which previous researches have proved significant for energy demand and those that have not been deemed influential. The reason for selecting all energy-relevant spatial attributes is to understand their interaction and influence when they are in confluence as opposed to studying each attribute in isolation. To remind from Chapter 4, attributes of urban form considered along with their metrics of measurement, selected from a comprehensive literature review, are outlined in the table below (Table 2).

Attribute	Index	Metric
Density	Floor Space Index	$(\text{Gross floor area}) / (\text{Total ground area})$
	Ground Space Index	$(\text{Gross area of building footprint}) / (\text{Total ground area})$
	Open Space Ratio	$(\text{Gross area of unbuilt ground}) / (\text{Total ground area})$
	Layer	$(\text{Average number of floors}) / (\text{Total ground area})$
	Network Density	$(\text{Length of the street network}) / (\text{Total ground area})$
Compactness	Aspect Ratio	$(\text{Average building height}) / (\text{Average street width})$
	Volumetric Compactness	$(\text{Envelope surface area}) / (\text{Building volume})$
	Size Factor	$\sqrt[3]{\text{Building volume}}$
	Form Factor	$\text{Envelope surface area} / \sqrt[3]{\text{Building volume}^2}$
Diversity	Mixed Use Index	$(\text{Gross residential floor area}) / (\text{Gross floor area})$
Green Areas	Green Space Density	$(\text{Gross green space area}) / (\text{Total ground area})$
	Green Area Geometry	$(\text{Gross green space perimeter}) / (\text{Gross green space area})$
Orientation	Community Building Orientation	$(\text{Sum of all buildings' orientation}) / (\text{Number of buildings})$
	Street Orientation	$(\text{Sum of all streets' orientation}) / (\text{Number of buildings})$
Shading	Urban Horizon Angle (UHA)	$(\text{Average height of the opposite skyline}) / (\text{Canyon width}) = \tan (UHA)$
	Obstruction Sky View (OSV)	$(\text{Average height of the opposite skyline}) / (\text{Canyon width}) = \cos (OSV)$
	Sky View Factor	$\text{if average building height} > 25 \text{ then } SVF = 1.56 - (0.00572 * H * D)$ $\text{if average building height} < 25 \text{ then } SVF = 0.9502 + (0.00042 * H) + (0.0198 * D) - (0.0065 * H * D)$ where H is the average height of the community buildings and D is the ground space index
Passivity	Passivity Ratio	$(\text{Net perimeter of all passive zones in community buildings}) / (\text{Net perimeter of all nonpassive zones in community buildings})$
	Plan Depth	$(\text{Net volume of all community buildings}) / (\text{Net area of all exposed walls in community})$

Table 2 - Selected attributes of urban form along with their index and metric

The preprocessing of spatial data starts with decomposing the county-wide GIS data into zip code units²⁴. Using the QGIS software and relevant geoprocessing tools the following spatial attributes are compiled for the regional boundaries of each zip code: building footprints and maximum allowed height for each building according to zoning, street network and street width, elevation of each building from the sea level, existing parks and green areas, and the location and shape of the urban lots. These shapefiles are then imported to Grasshopper and with an algorithm written in Grasshopper and Python, the nineteen selected indicators of urban form are measured for each zip code, compiled and exported into .csv format. The final output of this phase is a single .csv file with 110 rows each dedicated to one zip code and including nineteen set of numbers representing the zip code's

²⁴ As of spring 2019, the shape files acquired for the county of San Diego includes complete geographic information of 110 zip codes.

unique urban form (Figure 9). Using numbers with 19 significant digits to represent a zip code allows to numerically differentiate between various urban forms and spatial layouts.

Zipcode	Floor Space Index	Ground Space Index	Open Space Ratio	Layer	Network Density	Aspect Ratio	Volumetric Compactness	Size Factor	Form Factor	Mixed Use Index	Green Space Density	Green Area Geometry	Community Bldg Orientation	Street Orientation	Urban Horizon Angle	Obstruction Sky View	Sky View Factor	Passivity Ratio	Plan Depth
91901	0.001205	0.000432	0.999568	5.420000e-10	0.000359	4.875631e+02	0.991402	420.056228	416.444609	0.164667	0.002889	0.001874	41.276349	40.132687	12.106701	85.789267	0.954447	22.226934	1.008673
91905	0.000187	0.000071	0.999929	7.230000e-10	0.000315	7.566314e+02	1.235895	202.790793	250.628217	0.111340	0.000000	0.000000	40.648904	37.704011	24.731619	81.596596	0.954319	29.361111	0.809130
91906	0.000382	0.000135	0.999865	7.660000e-10	0.000343	2.125329e+03	1.120745	254.245908	284.944750	0.271891	0.037659	0.000728	39.813278	38.403403	22.455879	80.442707	0.954408	25.521491	0.892264
91916	0.000207	0.000071	0.999929	7.590000e-10	0.000177	1.392649e+03	1.224450	210.173666	257.347246	0.333521	0.459409	0.000359	41.914261	40.408719	9.953372	83.235151	0.954559	48.167192	0.816693
91917	0.000111	0.000037	0.999963	9.880000e-10	0.000161	3.527497e+02	1.209093	157.857038	190.863831	0.065425	0.010174	0.001465	41.627784	46.507177	22.967965	79.879306	0.954679	46.441537	0.827066
91931	0.000259	0.000088	0.999912	9.040000e-09	0.000122	6.044074e+02	1.327863	99.379660	131.962539	0.744359	0.000000	0.000000	40.894040	38.529412	15.790532	75.696154	0.954587	43.940705	0.753090

Figure 9 - Measurements of urban form per zip code (dataset header)

(ii) Preprocessing energy data

As mentioned above, the original energy consumption data provided by SDG&E is aggregated per zip code and customer type (residential, commercial, industrial, and agricultural) for each month from 2012 onwards (Figure 10).

	ZipCode	Month	Year	CustomerClass	Combined	TotalCustomers	TotalkWh	AveragekWh
0	91901	1	2012	A	Y	0	0	0
1	91901	1	2012	C	Y	267	84928	318
2	91901	1	2012	I	Y	0	0	0
3	91901	1	2012	R	Y	328	93407	285
4	91902	1	2012	A	Y	0	0	0
5	91902	1	2012	C	Y	0	0	0
6	91902	1	2012	I	Y	0	0	0
7	91902	1	2012	R	N	162	24536	151

Figure 10 - Original energy consumption dataset acquired from SDG&E (dataset header)

Preprocessing the energy dataset herein includes calculating the net energy consumed by each zip code, regardless of the existing building types and reordering them in a way that

	Zipcode	Year	kwh-01	kwh-02	kwh-03	kwh-04	kwh-05	kwh-06	kwh-07	kwh-08	kwh-09	kwh-10	kwh-11	kwh-12	kwh-total
0	91901	2012.0	178335	572185	1180728	1628495	1696036	1883073	2525349	3142081	2818045	2358746	2573347	3387368	23943788
1	91901	2013.0	3417028	2892997	3007450	3003970	3354659	3620557	4408138	4648182	4520802	5704204	6063564	6933314	51574865
2	91901	2014.0	5989979	5428413	5641444	5842737	8013901	8571344	10996885	11028397	12137559	9076297	8353464	9802152	100882572
3	91901	2015.0	9193923	7537123	8076293	7582986	7945450	9317607	9713128	11549307	10527516	2430824	2291552	2716079	88881788
4	91901	2016.0	10485777	1848118	1804789	907845	946058	1196945	11337912	11261558	8806578	9327471	8418730	9901345	76243126
5	91901	2017.0	10349153	8200408	7701083	6895892	7608524	8983617	11717806	11301569	9665343	8782920	9192398	8680018	109078731
6	91901	2018.0	13877547	22839151	8059726	3237940	3386157	3695892	12251954	11544396	9038465	3845338	4310910	5756585	101844061
7	91906	2012.0	13127	108819	204512	299827	395143	494300	655571	745822	678509	691468	770363	911910	5560725
8	91906	2013.0	986662	970031	992714	985781	1085703	1192636	1349686	1356730	1253925	1180198	1164000	1392257	13910323
9	91906	2014.0	1369587	1300480	1363786	1474397	2358526	2547569	3105988	2879323	2850134	2421660	2386241	2721728	26779419
10	91906	2015.0	2680298	2229704	2346160	2230584	2350101	2590918	2666769	3044565	2632324	290320	681825	876035	24619603
11	91906	2016.0	2957209	650163	615125	157876	157614	198580	3083187	3037779	2346355	2231644	2312282	2778007	20525821
12	91906	2017.0	2955117	2404648	3275609	2022732	1258960	2582025	3220134	3031542	2441364	2297132	2302473	2417964	30209700
13	91906	2018.0	2454826	2365070	2408589	2001318	2072283	2306570	3259322	3096450	2623918	2242138	2716944	3060164	30607592

Figure 11 - Preprocessed energy dataset (dataset header)

for each zip code the net value of energy consumption is arranged per month for each year (Figure 11).

(iii) *Synthesizing and cleaning the final dataset*

After preprocessing, the spatial and energy datasets are combined and structured into one and cleaned from any errors and inconsistencies. In the synthesized dataset, every 7 rows contain information on the values of urban form for each zip code, along with its dedicated monthly values of net energy consumption for each year (Figure 12). The assumption here is that each zip code's specific urban form has not significantly changed over the course of seven (7) years from 2012 to 2018.

	Zipcode	Floor Space Index	Ground Space Index	Open Space Ratio	Layer	Network Density	Aspect Ratio	Volumetric Compactness	Size Factor	Form Factor	...	kwh-04	kwh-05	kwh-06	kwh-07	kwh-08	kwh-09	kwh-10	kwh-11	kwh-12	kwh-total
0	91901	0.001205	0.000432	0.999568	5.420000e-10	0.000359	487.563149	0.991402	420.056228	416.444609	...	1628495	1696036	1883073	2525349	3142081	2818045	2358746	2573347	3387368	23943788
1	91901	0.001205	0.000432	0.999568	5.420000e-10	0.000359	487.563149	0.991402	420.056228	416.444609	...	3003970	3354659	3620557	4408138	4648182	4520802	5704204	6063564	6933314	51574865
2	91901	0.001205	0.000432	0.999568	5.420000e-10	0.000359	487.563149	0.991402	420.056228	416.444609	...	5842737	8013901	8571344	10996885	11028397	12137559	9076297	8353464	9802152	100882572
3	91901	0.001205	0.000432	0.999568	5.420000e-10	0.000359	487.563149	0.991402	420.056228	416.444609	...	7582986	7945450	9317607	9713128	11549307	10527516	2430824	2291552	2716079	88881788
4	91901	0.001205	0.000432	0.999568	5.420000e-10	0.000359	487.563149	0.991402	420.056228	416.444609	...	907845	946058	1196945	11337912	11261558	8806578	9327471	8418730	9901345	76243126
5	91901	0.001205	0.000432	0.999568	5.420000e-10	0.000359	487.563149	0.991402	420.056228	416.444609	...	6895892	7608524	8983617	11717806	11301569	9665343	8782920	9192398	8680018	109078731
6	91901	0.001205	0.000432	0.999568	5.420000e-10	0.000359	487.563149	0.991402	420.056228	416.444609	...	3237940	3386157	3695892	12251954	11544396	9038465	3845338	4310910	5756585	101844061

Figure 12 - Synthesized dataset (dataset header)

The next step is to clean the dataset by “removing errors and inconsistencies from data in order to improve the quality of data” (Rahm & Do, 2000). Errors and inconsistencies in datasets normally include incorrect, incomplete, or improperly formatted or duplicated data. The cleaning process for the synthesized dataset is done following the steps below:

- Missing yearly energy values: some zip codes in the dataset miss at least one year of energy data. To remain consistent in terms of the structure of the dataset, zip codes with at least one year of null or 0 energy values are removed.
- Missing monthly energy values: some zip codes in the dataset miss monthly energy values. If the missing monthly value is in between two existent ones, it is replaced by the average of the two existing values. If there's more than one consequent missing monthly value in a year, an average of all the existing monthly values of that year is used to replace the missing ones.

After cleaning, the dataset is then plotted and visualized in order to gain insight on its distribution (Figure 13). By eyeing the pair plot some relationships are easy to detect and interpret. For instance, the larger the network density the lower the open space ratio. Network density is the comparison of the network's length and the total ground area, whereas open space ratio compares the total area of unbuilt ground to the total ground area. Obviously when there are more streets in a certain boundary the area of unbuilt ground decreases. On the other hand, some relationships are not as obvious, and they only make sense in the presence of other variables.

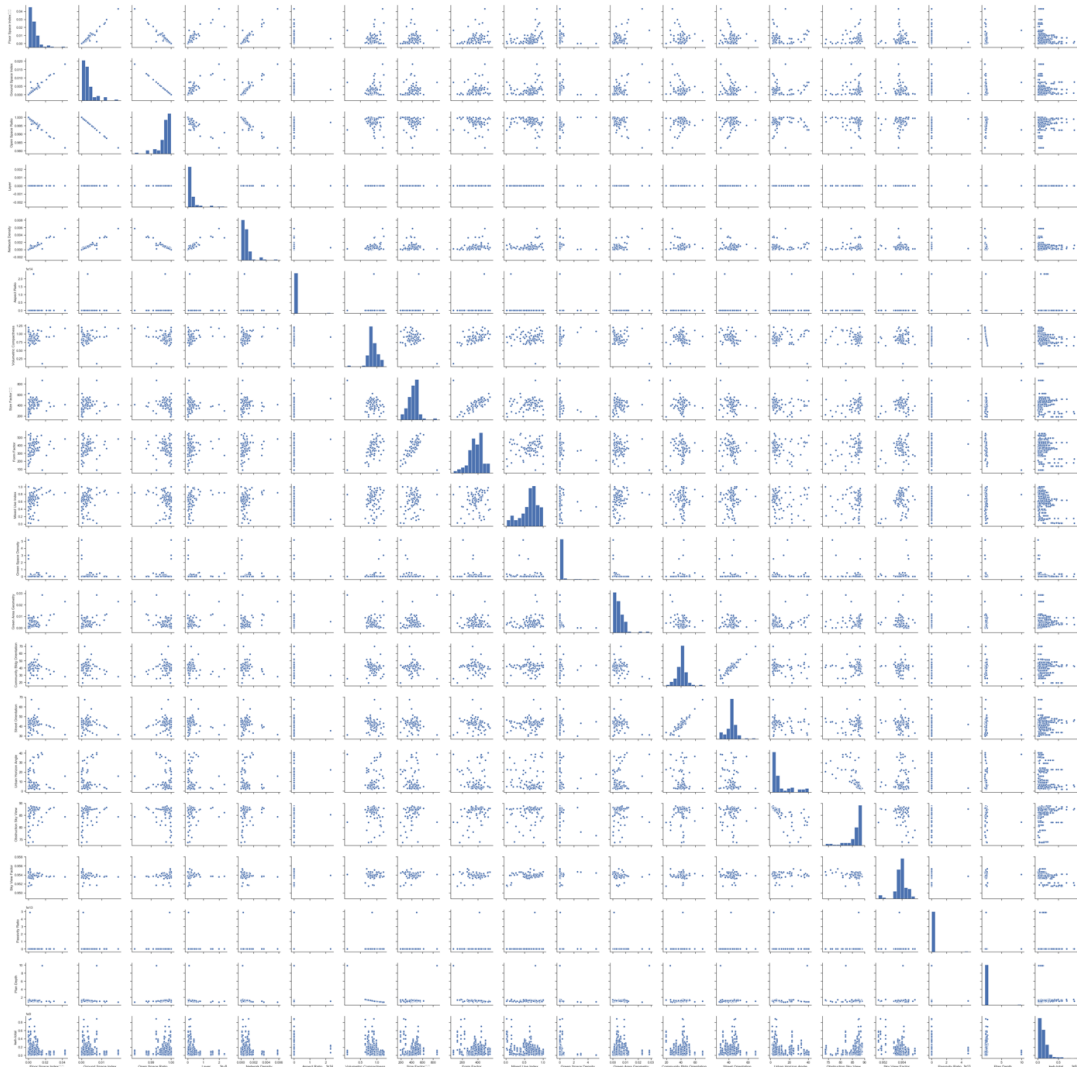


Figure 13 - Pair plot visualizing the relational pattern between every two urban form variables

In addition to gaining insight on how the dataset is distributed, visualizing it helps with detecting outlier data points which significantly differ from other observations. For example, the pair plot in Figure 5 shows that there's one zip code with a plan depth value which doesn't follow the pattern of the overall observations. The same condition applies to one zip code with an extreme passivity ratio and two zip codes with extreme aspect ratios. These four outlier zip codes are identified as outliers and are removed from the dataset since their inclusion could impact the training process.

5.3 Training an Artificial Neural Network on the Dataset

After cleaning the data and ensuring its quality the dataset is ready for analytical purposes. The final dataset includes complete information for 86 zip codes resulting in $86 * 7 = 602$ (number of zip codes * years of energy data) rows of data. This dataset has $19 + 12 + 1 =$

32 columns (19 variables of urban form + 12 months of energy consumption + 1 value of total energy consumed in one year). In this dataset, the nineteen indicators of urban form present the features or predictor variables, while values of energy consumption serve as the response variables.

The mathematical process that a neural network conducts on data for exploring its optimum machine learning model is known as the *training* or *learning process*. The approach used for conducting the learning process is explained in the three categories outlined below along with their relevant subcategories:

- *Data Preparation*
 - Loading data
 - Feature selection
 - Normalizing data
 - Splitting the data into train and validation
- *Model building using Artificial Neural Networks*
 - Define
 - Compile
 - Fit
- *Evaluation and Prediction*

5.3.1 Data Preparation

The data used to build the final machine learning model comes from multiple subsets of the original dataset that are adopted in different stages of the learning process: training, validating, and testing datasets (Mitchell, 1997). In this study, one year of data for each zip code is separated and put aside as the test dataset (86 * 1 rows of data). The portion of the dataset for year 2018 is selected for testing. The remaining data 86 * 6 rows of data (years 2012 to 2017) is used for both training and validating. To assure that the test dataset is not biased by significantly different weather conditions during the timeframe used in the

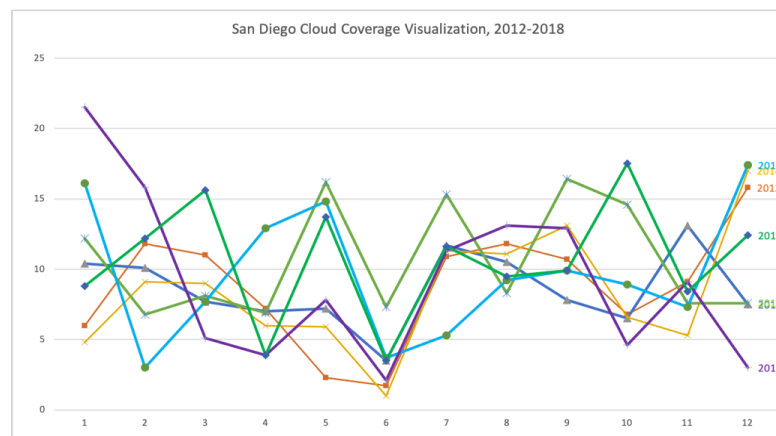


Figure 14 - San Diego cloud coverage from 2012 to 2018

training dataset, a brief examination of the cloud coverage from 2012 to 2018 is conducted. The cloud coverage data is downloaded from Columbia University's online portal²⁵. Figure 14 shows that the cloud coverage during the years of study weren't significantly different from one another, therefore it's assured that the training and test datasets have been aggregated under similar weather conditions.

(i) *Loading the data*

The training dataset is loaded using necessary libraries in Python and in subsequent steps a portion of it is used for validation purposes (Figure 15).

```
#import library for data preparation
import pandas as pd

#Loading the train data
Dataset=pd.read_csv('All Zips Data.csv')

#Dropping the monthly energy values from the dataset and keeping the total consumption for one year
Dataset = Dataset.drop([
    'Zipcode',
    'kwh-01',
    'kwh-02',
    'kwh-03',
    'kwh-04',
    'kwh-05',
    'kwh-06',
    'kwh-07',
    'kwh-08',
    'kwh-09',
    'kwh-10',
    'kwh-11',
    'kwh-12'], axis=1)
```

Figure 15 - Loading the dataset and dropping unnecessary columns

The employed dataset in this stage has 19 features of urban form measured for 110 zip codes in San Diego county as the predictor variable and for the response variable the total energy consumed for each zip code for the years 2012 to 2017 are considered. The main reason for using the annual value of energy consumption instead of monthly values in this dataset is due to the goal of this chapter which is interpreting the extent in which urban form impacts community-wide energy consumption regardless of each month's specific weather conditions, and accordingly to identify the most influential urban form features. However, for developing the final software prototype separate artificial neural networks are trained on datasets with energy consumption values for each month of the year. This process is profoundly described in Chapter 8.

²⁵ <https://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP-NCAR/.CDAS-1/.MONTHLY/.Diagnostic/.low-cloud/.cld/T/616.5/708.5/RANGE/X/239.1875/244.8125/RANGEEDGES/Y/32.64763/34.55237/RANGEEDGES/X/%28117.17W%29%28117.16W%29RANGEEDGES/T/%28Jan%202012%29%28Dec%202018%29RANGEEDGES/Y/%28333.6N%29%28333.7N%29RANGEEDGES/datafiles.html>

(ii) *Feature Selection*

Feature selection is one of the most important steps while performing any machine learning task. Not all independent variables in a dataset impact the response variable. By using relevant feature selection methods, the right subset of features is selected ensuring the accuracy of the learnt model. By omitting irrelevant independent variables, potential noise in the dataset is reduced and chances of the model being overfitted are prevented. Three different algorithmic methods of feature selection are applied to the dataset in order to compare the results and conclude the final set of independent variables. The utilized algorithmic methods encompass two wrapper methods, Recursive Feature Elimination and Backward Elimination, and one embedded method, Lasso model. These methods are selected because of their applicability to regression problems which require both independent and dependent variables to be continuous in nature. The three different processes for selecting relevant features are described in the following.

- *Recursive Feature Elimination*: wrapper methods in general, employ a machine learning algorithm for feeding the entire features to the model and then based on its performance certain features are added or removed. In Recursive Feature Elimination (RFE) method, the algorithm first takes the model to be used and the number of required features as input. It then recursively removes attributes and builds a model on those attributes that remain. Accuracy metric is used herein to rank features according to their importance with ‘1’ being the most important. Ultimately, RFE selects the combination of attributes which contribute the most to the response variable.

Scikit Learn is used herein for importing the RFE module. For the purposes of this study, *LinearRegression* is selected as an external estimator to assign weights to features. The estimator is trained on the initial set of nineteen features. In the code displayed in Figure 16, the estimator is set to select 7 features which have the highest impact on the model (note that the selection of number ‘7’ is random at this stage). Since the goal is to find the optimum number of features that attain a model with the highest accuracy, a loop is used that starts with 1 feature and goes up till 19 features. As seen in the code of Figure 17, the optimum number of features is 16. Now 16 is being fed as the number of features to RFE in order to retrieve the final set of features (Figure 18).

```
#Seperating the feature and target variables
Y = Dataset.pop('kwh-total')
X = Dataset

from sklearn.linear_model import LinearRegression
from sklearn.feature_selection import RFE

model = LinearRegression()
#Initializing RFE model
rfe = RFE(model, 7)
#Transforming data using RFE
X_rfe = rfe.fit_transform(X,Y)
#Fitting the data to model
model.fit(X_rfe,Y)
print(rfe.support_)
print(rfe.ranking_)

[ True True True True True False False False False False True
 False False False False True False False]
[ 1 1 1 1 1 13 3 12 11 4 5 1 7 8 9 10 1 6 2]
```

Figure 16 - Initializing the RFE model and using LinearRegression for training on the initial set of features and obtaining the importance of each feature based on rfe.support and rfe.ranking.

```
import numpy as np
from sklearn.model_selection import train_test_split

#no of features
nof_list=np.arange(1,19)
high_score=0
#Variable to store the optimum features
nof=0
score_list=[]
for n in range(len(nof_list)):
    X_train, X_test, Y_train, Y_test = train_test_split(X,Y, test_size = 0.3, random_state = 0)
    model = LinearRegression()
    rfe = RFE(model,nof_list[n])
    X_train_rfe = rfe.fit_transform(X_train,Y_train)
    X_test_rfe = rfe.transform(X_test)
    model.fit(X_train_rfe,Y_train)
    score = model.score(X_test_rfe,Y_test)
    score_list.append(score)
    if(score>high_score):
        high_score = score
        nof = nof_list[n]
print("Optimum number of features: %d" %nof)
print("Score with %d features: %f" % (nof, high_score))

Optimum number of features: 18
Score with 18 features: 0.482077
```

Figure 17 - Code for the loop that retrieves the optimum number of features yielding a model with highest accuracy

```
cols = list(X.columns)
model = LinearRegression()
#Initializing RFE model
rfe = RFE(model, 16)
#Transforming data using RFE
X_rfe = rfe.fit_transform(X,Y)
#Fitting the data to model
model.fit(X_rfe,Y)
temp = pd.Series(rfe.support_,index = cols)
selected_features_rfe = temp[temp==True].index
print(selected_features_rfe)

Index(['Floor Space Index\r\r\n', 'Ground Space Index', 'Open Space Ratio',
       'Layer', 'Network Density', 'Volumetric Compactness', 'Mixed Use Index',
       'Green Space Density', 'Green Area Geometry',
       'Community Bldg Orientation', 'Street Orientation',
       'Urban Horizon Angle', 'Obstruction Sky View', 'Sky View Factor',
       'Passivity Ratio', 'Plan Depth'],
      dtype='object')
```

Figure 18 - Code retrieving the selected features

- *Backward Elimination*: In backward elimination all possible features are first given to the model. By checking the performance of the model, the worst performing features are removed until the model reaches an overall acceptable performance. P-value is selected herein as the performance metric for evaluating feature performance; if p-value is above 0.05 the feature is removed and if it's below 0.05 the feature is kept in the model. For this, Ordinary Least Squares (OLS) method is used towards performing linear regression on the dataset and obtaining the p-values of each feature. Since backward elimination is an iterative process it's coded as a loop; by this in the first iteration the model is being fit on the initial set of features and after removing the weakest features, the model runs on the remaining features. This process is done until the final set of best performing features are obtained (Figure 19).

```
#Backward Elimination
cols = list(X.columns)
pmax = 1
while (len(cols)>0):
    p= []
    X_1 = X[cols]
    #Adding constant column of ones, mandatory for sm.OLS model
    X_1 = sm.add_constant(X_1)
    #Fitting sm.OLS model
    model = sm.OLS(Y,X_1).fit()
    p = pd.Series(model.pvalues.values[1:],index = cols)
    pmax = max(p)
    feature_with_p_max = p.idxmax()
    if(pmax>0.05):
        cols.remove(feature_with_p_max)
    else:
        break
selected_features_BE = cols
print(selected_features_BE)

['Floor Space Index\r\r\n', 'Open Space Ratio', 'Network Density', 'Aspect Ratio', 'Volumetric Compactness', 'Size Factor\r\r\n', 'Form Factor', 'Mixed Use Index', 'Urban Horizon Angle', 'Obstruction Sky View', 'Sky View Factor', 'Passivity Ratio', 'Plan Depth']
```

Figure 19 - Finding the best performing features using backward elimination

- *Lasso model*: is an example of embedded methods of feature selection. Similar to wrappers, embedded methods are also iterative since in each iteration of the model training process those features that contribute the most to the training of each particular iteration are selected. Lasso is a regularization method that in each iteration it penalizes irrelevant features in comparison to a given coefficient threshold by changing their coefficient to 0. By this, Lasso reduces the degree of overfitting or variance of a model by adding more bias. The code below (Figure 20) demonstrates the logic used for fitting Lasso to the dataset and the features selected accordingly.

Table 3 shows the different sets of high performing features as outputted by each feature selection method. In order to identify the final set of features, one particular artificial neural

```

from sklearn.linear_model import LassoCV
reg = LassoCV()
reg.fit(X, Y)
print("Best alpha using built-in LassoCV: %f" % reg.alpha_)
print("Best score using built-in LassoCV: %f" % reg.score(X,Y))
coef = pd.Series(reg.coef_, index = X.columns)

```

Best alpha using built-in LassoCV: 14291635.483876
Best score using built-in LassoCV: 0.185444

```

print("Lasso picked " + str(sum(coef != 0)) + " variables and eliminated the other " + str(sum(coef == 0)) + " variable

```

Lasso picked 7 variables and eliminated the other 12 variables

```

imp_coef = coef.sort_values()
import matplotlib
import matplotlib.pyplot as plt
matplotlib.rcParams['figure.figsize'] = (8.0, 10.0)
imp_coef.plot(kind = "barh")
plt.show()

```

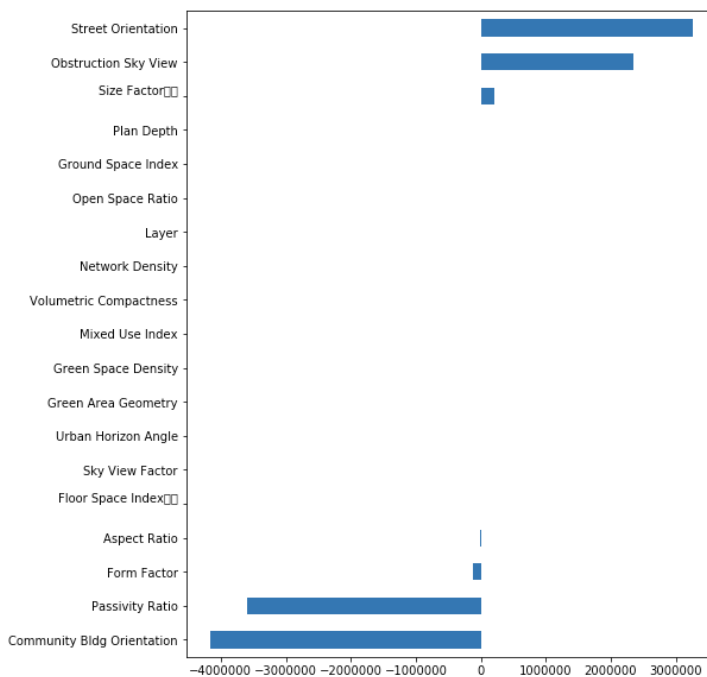


Figure 20 - Selected features by fitting a Lasso model to the dataset

network with a defined architecture is trained on these three datasets separately. By comparing the prediction accuracy of the three trained models, the final dataset with the final set of features is selected.

The following sections describe the steps required for building an artificial neural network to be trained on the three datasets with the three different set of features. These steps include normalizing data, splitting the dataset into train and validation, and then building the artificial neural network. Note that the codes conducting the following procedure

remains consistent when applied to the three datasets.

Method	Recursive Feature Elimination	Backward Elimination	Lasso Model
Selected Features	• Floor Space Index • Ground Space Index • Open Space Ratio • Layer • Network Density • Volumetric Compactness • Mixed Use Index • Green Space Density • Green Area Geometry • Community Building Orientation • Street Orientation • Urban Horizon Angle • Obstruction Sky View • Sky View Factor • Passivity Ratio • Plan Depth • Size Factor • Form Factor	• Floor Space Index • Volumetric Compactness • Mixed Use Index • Passivity Ratio • Form Factor • Urban Horizon Angle • Obstruction Sky View • Sky View Factor • Network Density • Plan Depth • Aspect Ratio • Open Space Ratio • Size Factor	• Community Building Orientation • Street Orientation • Aspect Ratio • Size Factor • Passivity Ratio • Obstruction Sky View • Form Factor
Number of selected features	18	13	7

Table 3 - Comparing the selected high performing features by each of the feature selection methods

(iii) Normalizing Data

The final step into preparing the dataset for training purposes is normalizing the data between 0 and 1. Normalizing is important for data integrity as it eliminates the units of measurement for data and enables an easier comparison between different instances of data. The code below (Figure 21) is used to separate response variables or targets (Y) from the predictor variables or features (X). The *MinMaxScaler* module from *Scikit Learn* is used for normalizing X and Y separately with their own assigned scalers as they have different minimum and maximum values (Figure 22). As an example, Figure 23 shows the output of plotting the normalized X and Y for the dataset derived from the RFE feature selection method.

```
dataset = pd.read_csv('EliminatedTrainData.csv')
dataset_stats = dataset.describe()
dataset_stats = dataset_stats.transpose()

Y = dataset.pop('kwh-total')
X = dataset
```

Figure 21 - Code for separating the response variables from the predictor variables

```
from sklearn.preprocessing import MinMaxScaler
Xscaler = MinMaxScaler()
Yscaler = MinMaxScaler()

X = Xscaler.fit_transform(X)
Y = Yscaler.fit_transform(Y)

#the normalized data is what we use to train our model

plt.plot(X)
plt.show()
print(type(dataset))
plt.plot(Y)
plt.show()
print(type(dataset))
```

Figure 22 - Code for normalizing the dataset

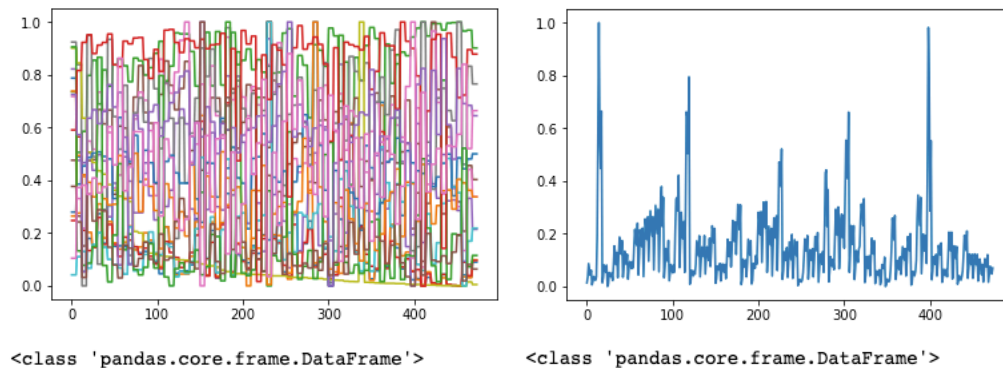


Figure 23 - Left plot shows the normalized X and the right plot shows the normalized Y

(iv) *Splitting the Dataset into Train and Validation*

As mentioned earlier, the test dataset - containing data for year 2018 - is set aside earlier in the process. The dataset used for training and validation purposes contains data for years 2012 to 2017 and includes $82 * 6$ rows of data. The *train-test-split* module is used to divide the train dataset in training and validation using 80:20 ratio where the validation is used for evaluating the trained model's performance before exposing it to the test dataset. Figure 24 shows the code for this procedure.

```
from sklearn.model_selection import train_test_split
seed = 7
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, shuffle=True)
```

Figure 24 - Splitting the train data into training and validation

5.3.2 Model Building Using Artificial Neural Networks

An artificial neural network consists of several neurons that are organized in three types of layers named the input, hidden, and output layers. The input layer is used to introduce the dataset to the network with no computation performed. The shape of the dataset is defined in this layer by identifying the number of inputs or features of the dataset; the neurons placed and defined in this layer just pass on the dataset's input information to the hidden layers. Hidden layers are placed between the input and output layer which their quantity and the number of neurons in each of them can be as many as desired. Defining the hidden layers are manual and do not follow any specific rule. Usually, different structures are tested and the one which yields the best results for the desired problem of study is selected. These layers are named 'hidden' because the outputs of their computation remain in the network and are not available outside the neural network. The 'black box' perception of neural networks is due to the abstraction of the computations happening in the hidden layers. The hidden layers are responsible for performing all sorts of computation on the features entered from the input layer and then transferring the results to the output layer. The output layer then puts forward the information learnt by the network to outside the neural network. The number of neurons in the output layer directly correspond to the number of outputs or response variables of the dataset.

Artificial neural networks are normally fully connected implying that the neurons of the adjacent layers are fully connected to each other and each connection has an associated weight (Ciaramella, Staiano, Cervone, & Alessandrini, 2015). These neurons operate in correspondence with their associated weight, bias, and activation function in the following procedure: each neuron sums the result of multiplying each input by its associated weight of the input connection and after adding a bias, it applies a function to the result. So, if a neuron is considered to be:

$$X = \Sigma (weight * input) + bias$$

a function is applied to the value of X which its functionality is to decide whether a neuron should be activated or not. Referred to as the "activation function", its purpose is to add non-linearity to the output of the neurons assuring its learning capabilities. Different variants of activation functions exist including but not limited to Sigmoid, Rectified Linear Unit (Relu), Tanh, and Leaky Rectified Linear Unit (Leaky Relu). Knowing certain characteristics of the problem of interest can help with choosing appropriate activation functions that lead to faster training and more accurate results. For instance, the sigmoid function is a good choice for the output layer of a binary classification. The different structures and "architectures" of artificial neural networks are identified by "*the different topologies adopted for the connections and by the choice of the activation function*" (Ciaramella, Staiano, Cervone, & Alessandrini, 2015).

An artificial neural network is first trained on the train dataset and then its performance is validated and tested against the validation and test datasets respectively. In the training procedure, the weights and biases of the of the neurons are updated based on the error of the output. This is done through a stipulated number of iterations where in each iteration or epoch the connection weights are updated and adjusted with the goal of minimizing the error between the dataset's real output and the output as provided by the neural network. To perform this iterative procedure, training algorithms are used “*for minimization of an error function, with adjustments to the weights being made in sequence of steps*” (Bishop, 1995). Back-propagation training algorithm is the widely used learning algorithm in the field chiefly due to its computational efficiency in computing the adjustments made to the weights (Bishop, 1995).

In this research the artificial neural network is defined and trained using Python programming language and relevant machine learning libraries namely Keras. Keras is an open-source library implemented for the fast calculations with deep neural networks, running on top of other libraries including TensorFlow, Theano, and Deeplearning4j. The steps for building an artificial neural network in Keras are:

Define → Compile → Fit.

- *Define*: For this study the *Sequential* model from Keras is used to build a fully connected artificial neural network model. The name of the model is *sequential* because the layer instances are sequentially passed to the constructor. As the first step in building the artificial neural network, the input layer is defined by identifying the shape of the input dataset. This is done via the *input_dim* function which its value corresponds to the number of features or predictor variables the dataset has. For instance, if the model is trained on the dataset derived from the RFE feature selection method, *input_dim* is set to 17 for the 17 predictor variables.

Next, is to define the hidden layers by deciding the number of hidden layers along with the number of neurons in each hidden layer and their activation functions. As mentioned earlier there's no specific rule for designating the number of hidden layers and their consisting number of neurons. 5 hidden layers are finalized for this study's neural network starting from 128 neurons in the first hidden layer and converging to 8 neurons in the last hidden layer (the number of neurons in each hidden layer is a power of 2). The activation function for the input and hidden layers are set to *relu*, *kernel_initializer* is set to *VarianceScaling*, and *bias_initializer* is set to *RandomNormal*. The final architecture of the artificial neural network is achieved through multiple iterations and measuring the final network's performance.

The last layer to define is the output layer. The output layer herein is constructed with one neuron corresponding to the one response variable in the dataset - the energy consumed in one year ('kwh_total'). Figure 25 shows the code used for defining the architecture of the artificial neural network.

```
import keras
from keras.layers import Dropout
from keras.models import Sequential
from keras.layers import Dense
from keras import optimizers
```

Using TensorFlow backend.

```
#dropout ratio
dp = 0.2
```

```
model = Sequential()
model.add(Dense(256, input_dim=17, activation='relu', kernel_initializer="VarianceScaling", bias_initializer="RandomNormal"))
model.add(Dropout(dp))
model.add(Dense(128, activation='relu', kernel_initializer="VarianceScaling", bias_initializer="RandomNormal"))
model.add(Dropout(dp))
model.add(Dense(64, activation='relu', kernel_initializer="VarianceScaling", bias_initializer="RandomNormal"))
model.add(Dropout(dp))
model.add(Dense(32, activation='relu', kernel_initializer="VarianceScaling", bias_initializer="RandomNormal"))
model.add(Dropout(dp))
model.add(Dense(16, activation='relu', kernel_initializer="VarianceScaling", bias_initializer="RandomNormal"))
model.add(Dropout(dp))
model.add(Dense(8, activation='relu', kernel_initializer="VarianceScaling", bias_initializer="RandomNormal"))
model.add(Dropout(dp))
model.add(Dense(1)) #output Layer
model.summary()
```

Model: "sequential_1"

Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 256)	4608
dropout_1 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 128)	32896
dropout_2 (Dropout)	(None, 128)	0
dense_3 (Dense)	(None, 64)	8256
dropout_3 (Dropout)	(None, 64)	0
dense_4 (Dense)	(None, 32)	2080
dropout_4 (Dropout)	(None, 32)	0
dense_5 (Dense)	(None, 16)	528
dropout_5 (Dropout)	(None, 16)	0
dense_6 (Dense)	(None, 8)	136
dropout_6 (Dropout)	(None, 8)	0
dense_7 (Dense)	(None, 1)	9
Total params: 48,513		
Trainable params: 48,513		
Non-trainable params: 0		

Figure 25 - Defining the architecture of the artificial neural network.

- *Compile*: Compiling configures the structured model for training. This is done by setting the optimizer to change the weights and biases, and the loss function and metric to evaluate the model's performance. The optimizer used here is *rmsprop*. Since the output variable is a real number the loss function and the metrics are set to *mean squared error* and *mean absolute error* (Figure 26).

```
model.compile(loss='mse', optimizer='rmsprop', metrics=['mae', 'mse'])
```

Figure 26 - Compiling the model

- *Fit*: The final step in building the model is fitting the model on the training dataset (which 10% of it has been separated for validation). Figure 27 shows the code for fitting the model on the X_train and Y_train values with 500 training epochs and the output of the 5 first epochs.

```
history = model.fit(X_train, Y_train, validation_data=(X_test, Y_test), epochs=500, batch_size=10)
```

```
Train on 379 samples, validate on 95 samples
Epoch 1/500
379/379 [=====] - 0s 972us/step - loss: 0.0297 - mae: 0.1189 - mse: 0.0297 - val_loss: 0.0304 - val_mae: 0.1079 - val_mse: 0.0304
Epoch 2/500
379/379 [=====] - 0s 189us/step - loss: 0.0202 - mae: 0.0944 - mse: 0.0202 - val_loss: 0.0284 - val_mae: 0.1045 - val_mse: 0.0284
Epoch 3/500
379/379 [=====] - 0s 179us/step - loss: 0.0189 - mae: 0.0875 - mse: 0.0189 - val_loss: 0.0290 - val_mae: 0.1055 - val_mse: 0.0290
Epoch 4/500
379/379 [=====] - 0s 171us/step - loss: 0.0169 - mae: 0.0836 - mse: 0.0169 - val_loss: 0.0238 - val_mae: 0.0978 - val_mse: 0.0238
Epoch 5/500
379/379 [=====] - 0s 174us/step - loss: 0.0156 - mae: 0.0833 - mse: 0.0156 - val_loss: 0.0223 - val_mae: 0.0938 - val_mse: 0.0223
```

Figure 27 - Fitting the model with 500 epochs. The result of the first 5 epochs is shown here.

Figure 28 shows the value history of the last 5 epochs. Back-propagation is done automatically in Keras when fitting the model to the dataset. The value of back-propagation here, besides minimizing the error that the network has over all the training set examples, is that the model is learning from the examples that it's fed to it instead of specifying the mathematical relationship between various variables (Goh, 1995).

```
hist = pd.DataFrame(history.history)
hist['epoch'] = history.epoch
hist.tail()
```

	val_loss	val_mae	val_mse	loss	mae	mse	epoch
495	0.010839	0.067729	0.010839	0.007935	0.056724	0.007935	495
496	0.011089	0.069208	0.011089	0.008387	0.054164	0.008387	496
497	0.010642	0.067396	0.010642	0.008231	0.056428	0.008231	497
498	0.010556	0.066077	0.010556	0.006511	0.053987	0.006511	498
499	0.010626	0.067115	0.010626	0.007604	0.055809	0.007604	499

Figure 28 - Value history of the last 5 epochs

5.3.3 Evaluation and Prediction

When building a machine learning model, it's important to keep the error as low as possible. Learning curves are widely used to diagnose the performance of the training process or learning over experience in every repetition or epoch (Anzanello & Fogliatto, 2011). A learning curve monitors the evolution of two error scores belonging to the training and validation datasets where the x-axis shows the number of epochs and the y-axis shows the improvement rate of the learning procedure. Since the learning rate of the trained model herein is evaluated by mean squared error, the y-axis of the line plot shows improvements on this metric. A model that improves over time is demonstrated by a gradually descending mean squared error during the training period. Figure 29 shows the code for plotting the learning curve and compares the outputs of the code when ran on the three different datasets with the different set of features. The overall shape of the three learning curves are almost the same representing a good fit: the training and validation loss in all three decrease to a point of stability with a small gap between the two curves. Table 4 shows the resulting mean squared error from the three models.

```
plt.plot(history.history['mean_squared_error'])
plt.plot(history.history['val_mean_squared_error'])
plt.title('mean_squared_error')
plt.ylabel('mean_squared_error')
plt.xlabel('epoch')
plt.legend(['train', 'validation'], loc='upper left')
plt.show()
```

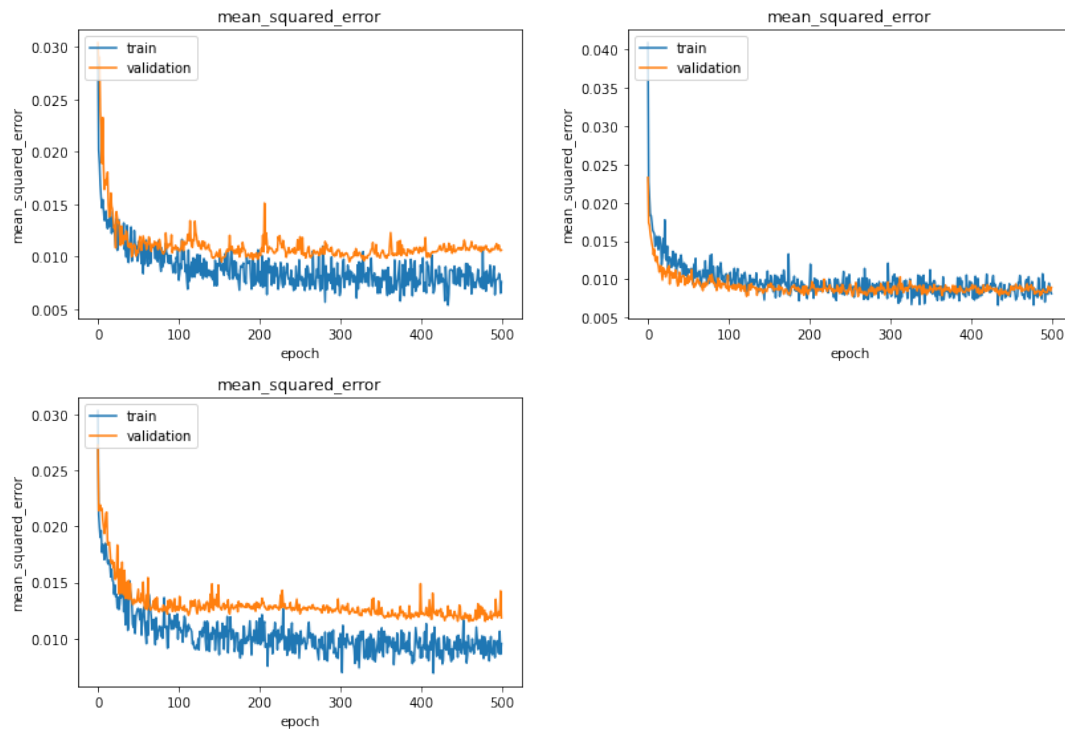


Figure 29 - The code for plotting the learning curve. The top left plot shows the performance of model trained on the dataset derived from the RFE feature selection method. Top right is for the dataset from the backward elimination feature selection method. Bottom left is for the dataset from lasso feature selection method.

```

prediction = model.predict(X_test).flatten()
plt.plot(prediction, label='Pred')
plt.plot(Y_test, label='True')
plt.legend()
plt.show()

print (prediction)
print(Y_test)

realPrediction = Yscaler.inverse_transform(prediction)

```

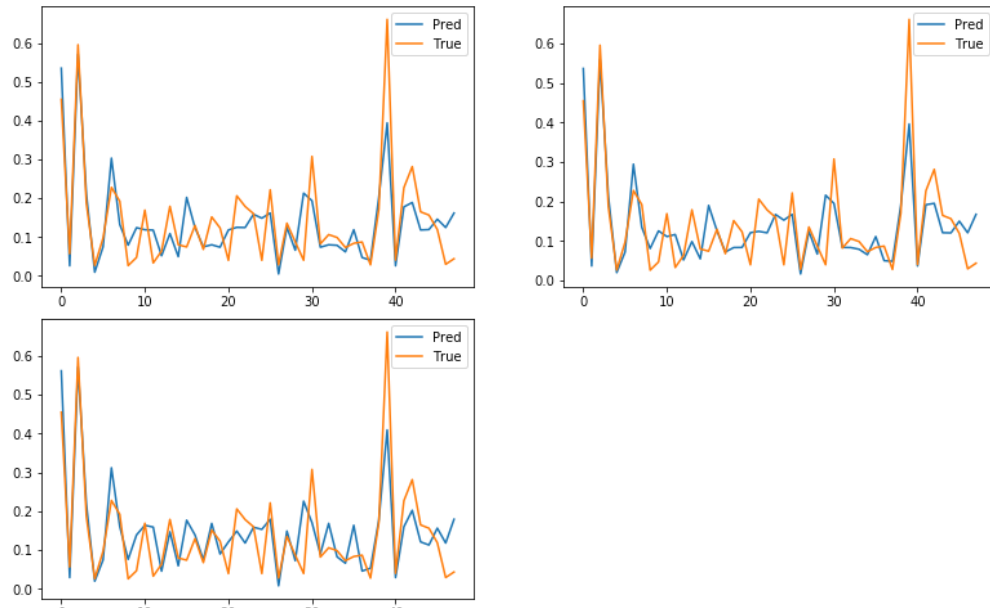


Figure 30 - Comparing the predicted values for the validation dataset with its true values. The top left plot belongs to the model trained on the dataset derived from the RFE feature selection method. Top right is for the dataset from the backward elimination feature selection method. Bottom left is for the dataset from the lasso feature selection method.

```

plt.scatter(Y_test, prediction)
plt.xlabel('True Values')
plt.ylabel('Predictions')
plt.axis('equal')
plt.axis('square')
plt.xlim([0, plt.xlim()[1]])
plt.ylim([0, plt.ylim()[1]])
_ = plt.plot([-1, 1], [-1, 1])
plt.show()

```

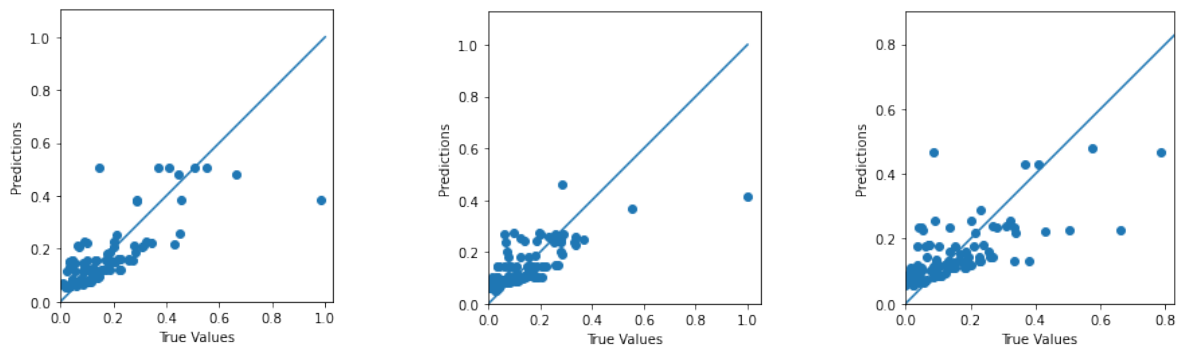


Figure 31 - Comparing the predicted values for the validation dataset with its true values. The left plot belongs to the model trained on the dataset derived from the RFE feature selection method. The middle plot is for the dataset from the backward elimination feature selection method. The right one is for the dataset from lasso feature selection method.

	Dataset from RFE feature selection method	Dataset from Backward Elimination feature selection method	Dataset from Lasso Model feature selection method
MSE Regression Loss	0.0106	0.008	0.0118

Table 4 - Comparing the mean squared error regression loss for all three models.

To further compare the performance of the three models, predictions are made for the validation dataset and the results are compared against the validation true values (Figure 30-31). All three model plots show a reasonably good prediction. Figure 32 shows the normal distribution of errors for each of the three models indicating that the trained models are true to their specified probabilistic interpretations. Note that since the dataset is comparatively small the mean squared error between training and validation might infer a bit high.

```
error = prediction - Y_test
plt.hist(error, bins = 25)
plt.xlabel("Prediction Error")
_ = plt.ylabel("Count")
plt.show()
```

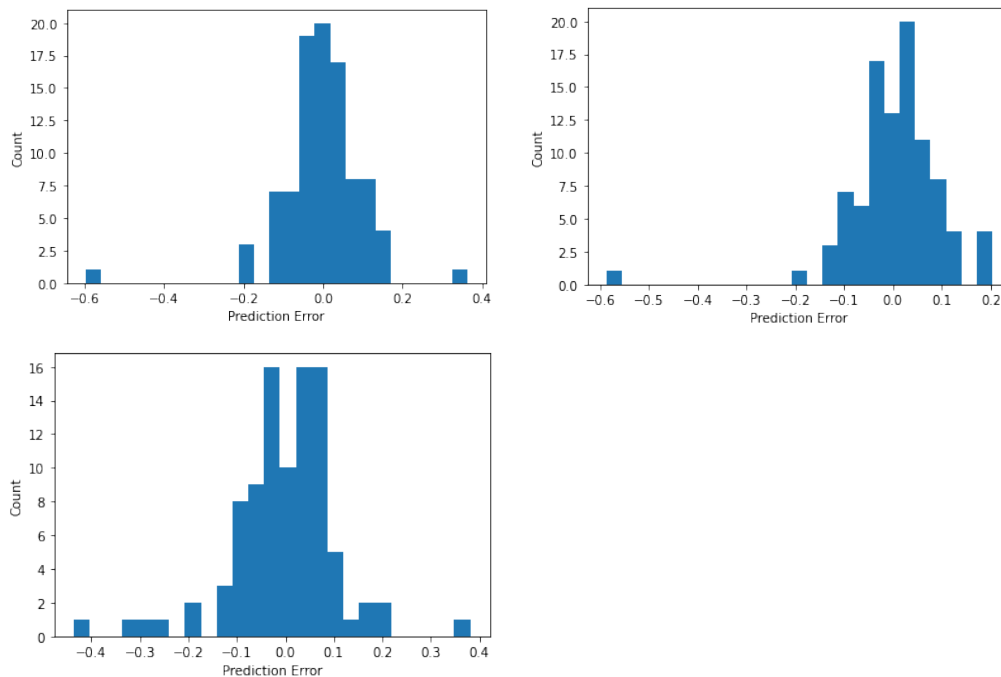


Figure 32 - Comparing the distribution of errors for the three models. The top left plot belongs to the model trained on the dataset derived from the REF feature selection method. Top right is for the dataset from the backward elimination feat selection method. Bottom left is for the dataset from lasso feature selection method.

As observed in the evaluation plots, all three trained models almost have the same performance. This means that the selected features in all three datasets are contributing to the training process more or less at the same rate. Therefore, there's no 'one' best

performing model to chose. In order to find the final set of contributing features, the features common in all three datasets or at least in two of them are chosen (Table 5). These features are floor space index, volumetric compactness, mixed use index, passivity ratio, form factor, urban horizon angle, obstruction sky view, network density, plan depth, open space ratio, size factor, community building orientation, street orientation, and sky view factor.

	Recursive Feature Elimination	Backward Elimination	Lasso Model	Common Features
Selected Features	• Floor Space Index • Ground Space Index • Open Space Ratio • Layer • Network Density • Volumetric Compactness • Mixed Use Index • Green Space Density • Green Area Geometry • Community Building Orientation • Street Orientation • Urban Horizon Angle • Obstruction Sky View • Sky View Factor • Passivity Ratio • Plan Depth • Size Factor • Form Factor	• Floor Space Index • Volumetric Compactness • Mixed Use Index • Passivity Ratio • Form Factor • Urban Horizon Angle • Obstruction Sky View • Sky View Factor • Network Density • Plan Depth • Aspect Ratio • Open Space Ratio • Size Factor	• Community Building Orientation • Street Orientation • Aspect Ratio • Size Factor • Passivity Ratio • Obstruction Sky View • Form Factor	• Floor Space Index • Volumetric Compactness • Mixed Use Index • Passivity Ratio • Form Factor • Urban Horizon Angle • Obstruction Sky View • Network Density • Plan Depth • Open Space Ratio • Size Factor • Community Building Orientation • Street Orientation • Sky View Factor
#of selected features	18	13	7	14

Table 5 - Features common among all three tested datasets.

A new dataset is created afterwards with the selected common set of features. A neural network with similar architecture as described earlier is trained on the dataset and its performance is evaluated. Figure 33 shows the performance plots of the training procedure. The learning curve (Figure 33 – top left) demonstrates a pretty good fit with decreasing values of mean squared error during the entire learning experience and with a small gap between training and validation curves. By comparing the predicted values for the validation dataset with its true values (Figure 33 – top right and bottom left), these plots indicate that reasonably good predictions are made by the trained model. This is also asserted by the almost-bell-shaped distribution of errors in the bottom right plot of Figure 33. The computed mean squared error regression loss for this model is around 0.007.

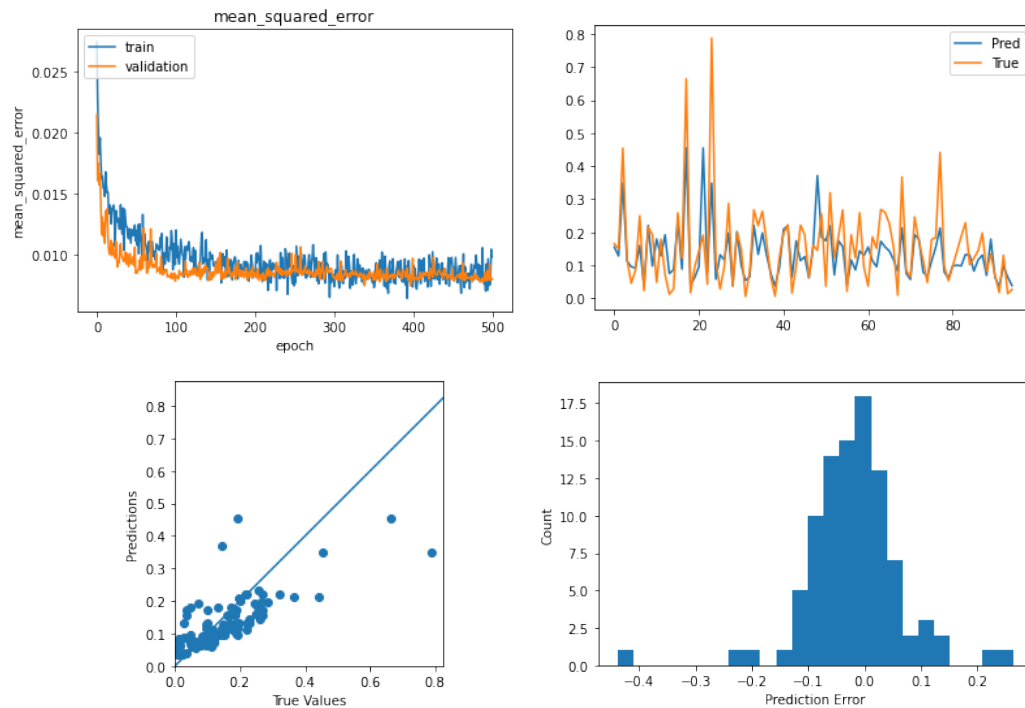


Figure 33 - Four plots showing the performance of the model training.

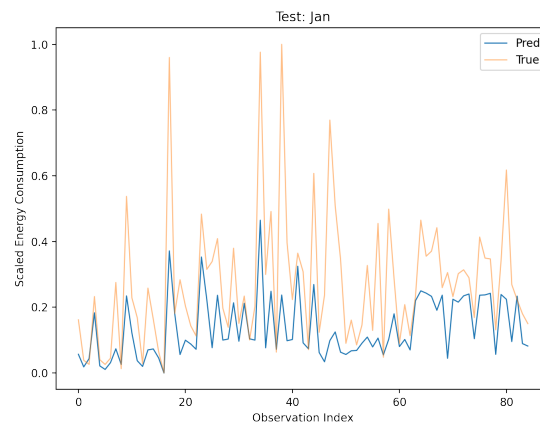


Figure 34 - Plot showing the model's performance on test dataset.

These analyses show limited difference between the performance rates of the four trained models with four different sets of features. Therefore, the final model selected is the one trained on the dataset whose features are common between the previous three datasets. The trained model was then tested on the test dataset resulting in a mean squared error of 0.04 (Figure 34). Discussion on the performance of the testing procedure is elaborated on in Chapter 8.

Chapter conclusion: This chapter focuses on the technicality of training an artificial neural network on the dataset. By selecting, configuring, and structuring an artificial neural network to study the complex relationship between urban form and community scale energy consumption, this chapter has responded to the second research question of this study: *“what is the most appropriate model for describing the desired relational pattern between urban form and energy performance of microgrid-connected communities?”*

For the purposes of this research, the resulting predictive model is not complete until it is being statistically and architecturally interpreted to find the most influential indicators of urban form on community scale energy consumption. Having this objective in mind, the inference and interpretation of this chapter’s predictive model is discussed in chapter 6

Chapter 6 Community Energy Consumption and Urban Form: A Statistical and Architectural Interpretation

Introduction: In Chapter 5, the selected nineteen indicators of urban form from Chapter 4 were measured for all zip codes in San Diego and their monthly values of energy consumption were acquired through the county's utility company, SDG&E. An artificial neural network was then trained on the dataset with urban form as its predictor variable and energy consumption as its response variable. In this chapter, a statistical inference is performed on the predictive model from Chapter 5 using Shapley values. The results of the statistical inference indicates that the most influential indicators of urban form on energy consumption are related to the compactness, passivity, shading, and diversity of a community in the context of the case study. These fourteen most influential indicators of urban form (in order of magnitude of impact from high to low) include: Size Factor, Mixed Use Index, Passivity Ratio, Sky View Factor, Plan Depth, Form Factor, Open Space Ratio, Network Density, Community Building Orientations, Urban Horizon Angle, Volumetric Compactness, Obstruction Sky View, Floor Space Index, and Street Orientation. Focusing on energy consumption in this chapter as one of the components of measuring energy performance, an answer is offered to the third research question of this research: *“which combination of the attributes of urban form impact a solar community microgrid's energy performance?”*²⁶

After identifying the most influential indicators of urban form on community wide energy consumption, the next step is to architecturally interpret the results of the statistical inference which is extensively described in section 6.2 of this chapter. The objective of the architectural interpretation is to set the path for engaging architects and urban planners in the technical conversation of developing solar community microgrids and to offer a spatial vision on their construction.

6.1 Statistical Interpretation

The power of neural networks lies in their predictive ability within the context they're trained. When using predictive modeling as such we get the answer to 'what' is being predicted, but 'why' certain predictions are being made is often a challenging question to answer. This is because neural networks are normally seen as a black box *“whose unimaginably complex inner workings somehow magically transform inputs into predicted outputs”* (Garson, 1991). Understanding the 'why' behind neural networks' decision making in predictions can be crucial in certain studies. In this research particularly, the interest lies in understanding which features of urban form and which correlations have the

²⁶ This chapter is currently in process for publication in the “Smart Cities and Climate- Resilient Urban Planning” special issue of *Environment and Urban Planning B* journal.

most influence in estimating the community's net energy consumption. With this knowledge, a set of design principles and frameworks can be developed and offered to architects and urban planners for designing new and retrofitting existing energy efficient urban settlements. However, unraveling how neural networks make certain predictive decisions has been a challenge and is still a developing field in machine learning.

Among different methods proposed for interpreting neural networks, Shapley regressions are highly regarded due to their consistency and local accuracy in model inference (Štrumbelj & Kononenko, 2014; Lundberg & Lee, 2017). Shapley regressions offer a framework for statistical inference on non-linear machine learning models where inference is achieved based on Shapley values – a method from coalition game theory (Joseph, 2019). A Shapley value is the average marginal contribution of a feature value across all possible coalitions. In simple terms, Shapley values calculate the importance of a feature by comparing what the model predicts with and without the feature in every possible order that a model sees the features so that all of them are fairly compared. As explained by Lundberg et al (2020) “...shapley values are computed by introducing each feature, one at a time, into a conditional expectation function of the model's output, $f_x(S) = E[f(X)|do(X_s = x_s)]$, and attributing the change produced at each step to the feature that was introduced, then averaging this process over all possible feature orderings. Note that S is the set of features we are conditioning on, X is a random variable representing the model's M input features, x is the model's input vector for the current prediction, and we follow the causal *do*-notation formulation which improves on the motivation of the original SHAP feature perturbation formulation. Shapley values represent the only possible method in the broad class of additive feature attribution methods that will simultaneously satisfy three important properties: local accuracy, consistency and missingness²⁷.”

Another advantage of this framework, also known as Shapley Additive Explanations (SHAP), is that unlike other methods of model interpretability it is a unified framework capable of explaining any machine learning model (Lundberg & Lee, 2017). The SHAP

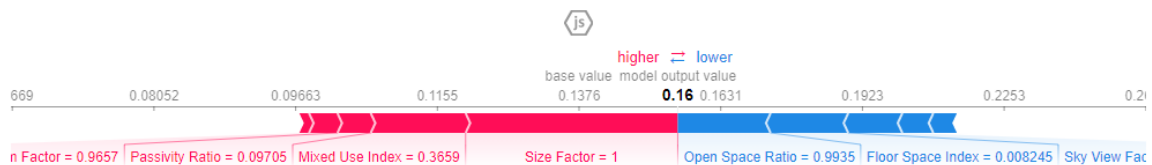


Figure 35 - Visualization showing the features impact on model output.

²⁷ Lundberg et al (2020) describe: Local accuracy equivalent to efficiency in game theory states that when approximating the original model f for a specific input x , the explanation's attribution values ϕ_i for each feature i should sum up to the output $f(x)$. Consistency equivalent to monotonicity in game theory, states that if a model changes so that some feature's contribution increases it stays the same regardless of the other inputs, that input's attribution should not decrease. Missingness equivalent to null effects in game theory.

framework has been implemented in the SHAP library²⁸ for Python which makes it feasible to compute all possible feature combinations. To uncover the magnitude of impact that different features of urban form have on communities' net energy consumption, SHAP has been used on the final dataset with the final set of selected features. Figure 35 shows features each contributing to push the model output from the base value (the average model output over the training dataset) to the model output. Features pushing the prediction higher are shown in red and those in blue push the prediction lower.

Figure 36 shows the impact that each urban form feature or variable has on the output which is energy consumption by plotting the SHAP values of every feature for every sample. In this plot variables are ranked in descending order according to the magnitude of impact on community-wide net energy consumption, the horizontal location shows whether the effect of that value is associated with a higher or lower prediction, and color shows whether that variable is high (in red) or low (in blue) for that observation. Different correlations are observed from this plot which can be spatially and architecturally interpreted. From a high-level point of view, indicators related to the compactness of buildings within community boundary, the diversity of building functions, the shading and overshadowing resulted from building placement, and the passivity associated with community buildings have a high impact on community scale net energy demand. This is

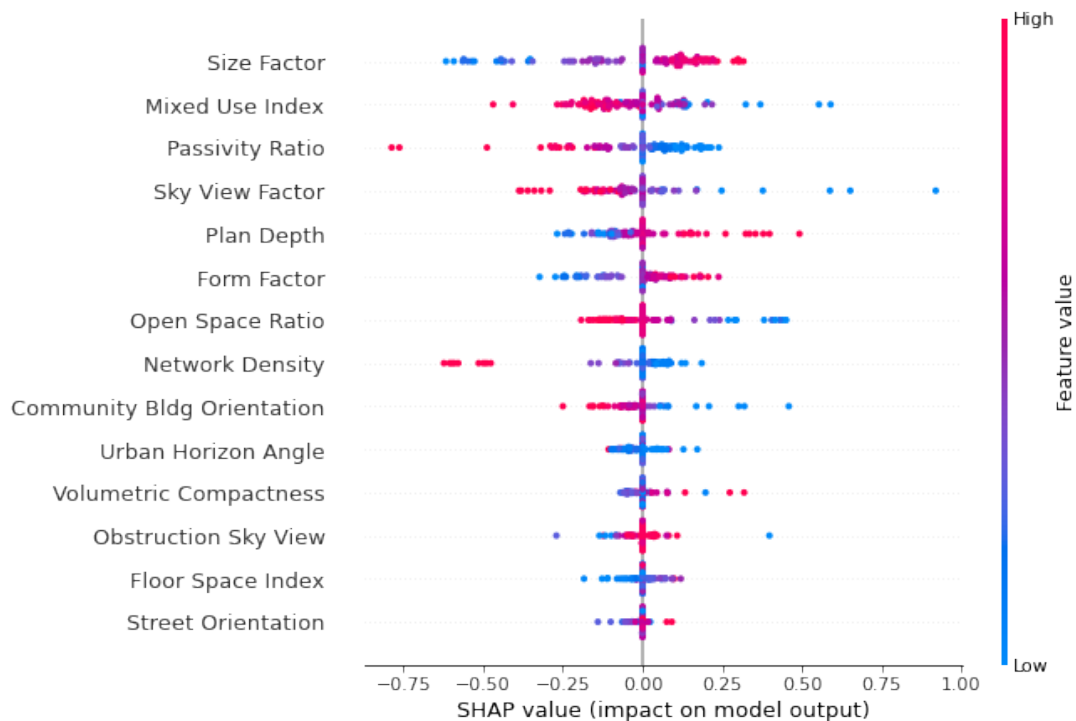


Figure 36 - The magnitude hierarchy that each of the urban form features has on energy consumption in this study.

²⁸ <https://github.com/slundberg/shap>

while indicators related to the orientation of a community (streets and buildings) in San Diego has minimal impact on net energy consumption. Figure 37 also shows the mean absolute value of the SHAP values for each feature to get a standard bar plot.

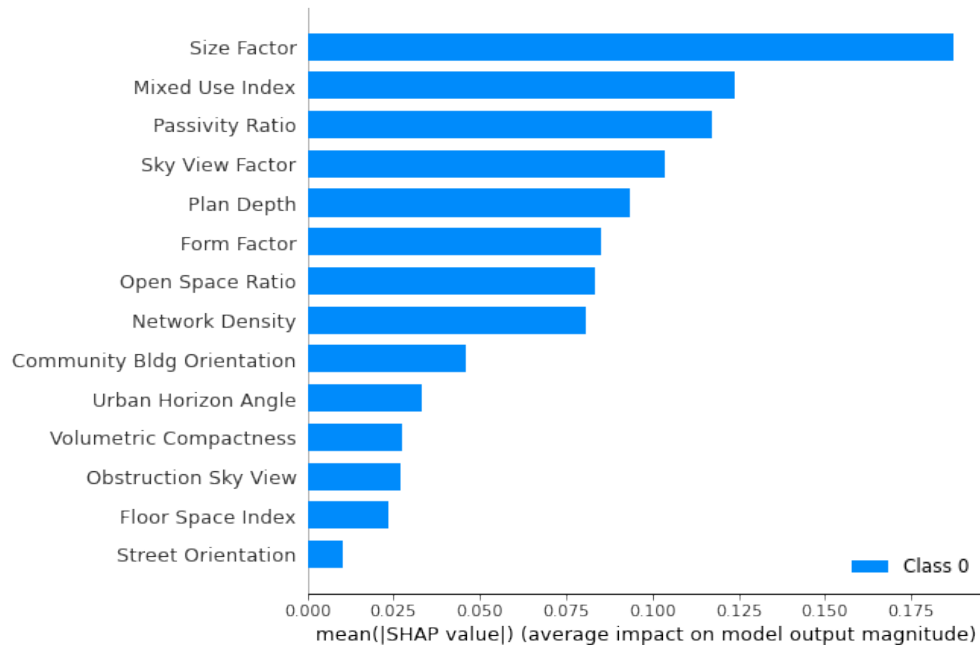


Figure 37 – Plot showing the mean absolute value of the SHAP values for each feature.

6.2 Spatial and Architectural Interpretation

The plot in Figure 36 shows San Diego’s most influential combination of urban form indices that impact community scale energy consumption. One important point to note here is that the resulting combination impacts the amount of energy demand in *San Diego* with its specific urban form characteristics and climatic conditions. Therefore, the results of this research cannot be generalized to any other city and/or climate, however, the introduced framework can be utilized to discover each region’s specific combination of urban form indices that have the highest impact on community scale energy demand.

San Diego county is home to varied climate zones under the California Irrigation Management Information System (CIMIS) (Figure 38):

- Zone 1 – Coastal: mild maritime climate where winters are mild, and summers are cool with year-long moisture in the air.
- Zone 4 – Coastal Inland: weather conditions are close to coastal with higher temperatures and less humidity.
- Zone 6 – Upland Central: higher elevations with moist coastal air and dry interior air and moderate humidity and wind flows.

- Zone 9 – Transition: marine to desert transition zone which is combination of warmer thermal belts and cold air basins with occasional marine influence.
- Zone 16 – Mountain: with variations in sun and wind exposure and more rainfall.
- Zone 18 – Desert: dry and hot days and cold night, low levels of humidity and very low rainfalls.

According to this map most urbanized areas of the county belong to zones 1, 4 and 6, therefore the majority of urban form calculations in this research as well as the following spatial assessments belong to these three climatic zones.

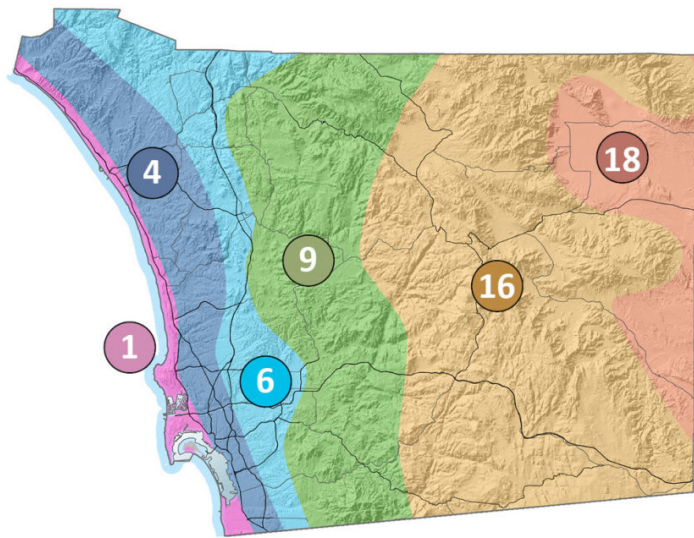


Figure 38 – Different climate zones in San Diego county. Source: waternewsnetwork.com

Sky view factor (SVF), urban horizon angle (UHA), and obstruction sky view (OSV) are geometric parameters of urban form that relate to the degree of obstruction or of access to the sky. The energy absorbed by any given building facade comes from the sky and the radiation reflected from opposite building facades. Available daylight in buildings as well the amount of radiation absorbed by building facades are measured by SVF and UHA. Sky view factor is the value of radiation received by building surfaces defined by the ratio of the amount of sky visible with obstructions to the amount of sky visible without obstruction. SVF value of 1 means that the radiation released by a surface is totally received by the sky and SVF value of 0 means the radiation is totally blocked by surrounding obstructions (Mirzaee, Özgun, Ruth, & Binita, 2018). UHA determines the effect of overshadowing by adjacent buildings and is a function of the mean elevation of the skyline from a building façade (Ratti, Baker, & Steemers, 2005). Larger UHA means more obstruction by surrounding buildings resulting in more overshadowing. In order to estimate the radiation reflected from obstructing buildings, we need to know the amount of radiation that falls on the obstructing building facades through their angle of obstruction by measuring the OSV (OSV is the primarily the same as UHA for the obstructing facades).

The formula used for calculating SVF in this research is at an urban scale (rather than a specific point) and is a function of neighborhood's average height and density (ground space index); this is while the measurements for UHA and OSV have been carried out for several points in a neighborhood and averaged out. This is perhaps the reason why in the plot the impact of UHA and OSV on neighborhood scale energy consumption is not quite clear.

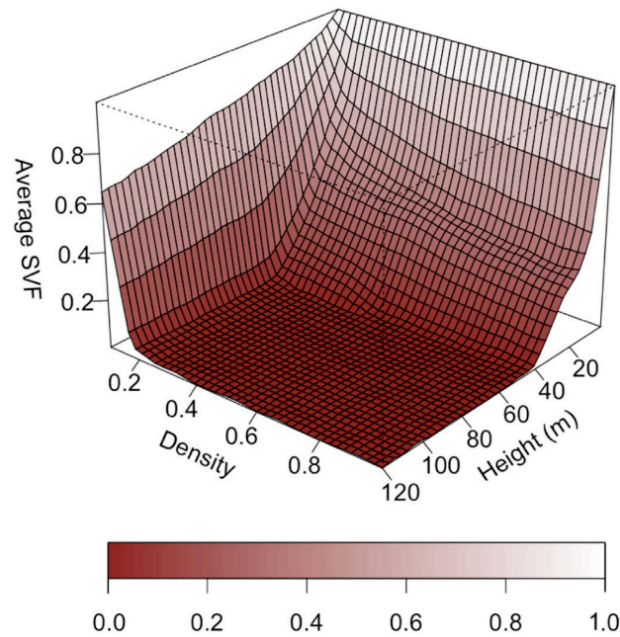


Figure 39 - 3D visualization of the simulated relationship between average SVF, neighborhood's average building height, and average density - Source: Mirzaee et al (2018)

Mirzaee et al (2018) have portrayed the relationship between SVF, height, and density not to be quite linear (Figure 39), however, it could generally be interpreted that neighborhoods with shorter and more scattered buildings have more sky visibility than denser neighborhoods with taller buildings. In neighborhoods with higher values of SVF where development dominates the natural landscape, direct radiation from the sky or reflected radiation from building facades gets trapped in the urban fabric, magnifying urban heat island effect. From this point of view, San Diego county's most urbanized areas such as the city of San Diego and its downtown area should be experiencing high temperatures due to UHI effect. Theoretically this is true but in practice the effects of UHI in San Diego county are highly influenced by the coastal local wind patterns. These westerly winds are blown from the ocean and help disperse heat from the coastal regions to the inland areas of the county. As demonstrated in Figure 40 the UHI effect is manifested in the southern part of the county and incrementally increases in the east of the county. This means that the rising UHI-related temperatures of the coastal regions of San Diego county – which

incorporates denser urban developments and thus has higher SVF – gets moderated due to the coastal winds while higher temperatures get blanketed in zones 9, 16, and 18.

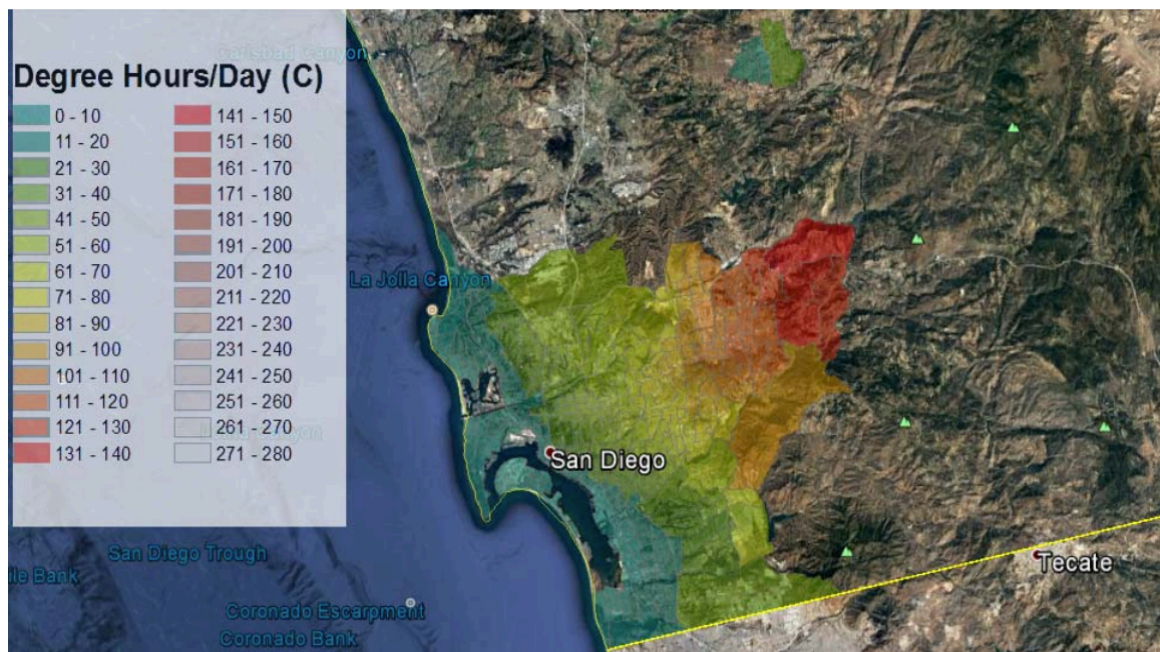


Figure 40 - Urban heat island index for San Diego county - Source: California Environmental Protection Agency (<https://calepa.ca.gov/climate/urban-heat-island-index-for-california/urban-heat-island-interactive-maps/>)

This explains the negative correlation between SVF and energy consumption where higher values of SVF have lesser values of neighborhood-scale energy consumption. The urbanized areas of San Diego county typically experience year-long cool, breezy, mild, and pleasant weather conditions and therefore most buildings in that area, specifically residential buildings, do not have air-conditioning systems (Wang & Chen, 2014). This means that the major part of energy is consumed for space heating in winter. However, when the SVF of a neighborhood is high means the buildings are less obstructing each other and therefore building surfaces receive more direct radiation from the sky. When building surfaces absorb more radiation than the building's interior spaces are passively heated leading to minimized rates of energy consumption for space heating on colder weathers.

Denser urban areas with low values of SVF result in higher values of UHA which lead to overshadowing adjacent buildings. According to the plot in Figure 36, in this scenario more energy is consumed for space heating buildings in a neighborhood. This can be justified in the context of San Diego which most operational energy use in buildings is oriented towards space heating purposes. Overshadowing passively cools the buildings and therefore increases the community's net energy demand for heating. Thus, in the case of developing a solar based community microgrid in San Diego, one has to determine that dense and compact neighborhoods where the distance between buildings are minimal, not only increases rates of energy demand but also due to the effect overshadowing, the

potential for installing PV panels on building and land surfaces are decreased. While currently PV panels are mostly installed on building roofs, not being overshadowed by adjacent buildings is an important point to account for. Even in scenarios where building facades are used for mounting PV cells, it's important to keep south facing façade (in case of buildings in San Diego which are in the northern hemisphere) clear from any shadows from surrounding building to maximize the amount of produced solar energy. This is when the orientation of communities become important. Unlike for community scale energy consumption, community scale PV energy production is highly dependent on the orientation of buildings and having the larger building facades oriented towards the south for maximum potential of PV cell installations and therefore maximum solar energy production.

Another important factor when considering energy consumption in communities is the amount of building envelope/surface exposed to the outside environment. The building envelope is the medium between inside and outside of a building where the transfer of energy between these two spaces happens. To minimize energy loss at the building scale, architects usually chose the materials used at the building envelope in a way that prevents unnecessary transfer of energy between inside and outside of a building. However, at an urban scale, researchers have found that different metrics considering the relationship between building surface and volume can play a significant role studying urban scale energy consumption (Ratti, Baker, & Steemers, 2005; Bourdic, Salat, & Nowacki, 2012). Some indicators of this relationship as selected herein are plan depth, volumetric compactness, size factor, and form factor, all of which have shown to impact community scale energy consumption in different extents. The fact that the combined effect of all four of these indicators impact energy consumption at urban scale conveys the importance that building volume and exposed building surfaces play in reducing energy consumption in San Diego communities. Volumetric compactness²⁹, size factor³⁰, form factor³¹ and plan depth³² all have a high and positive relationship with community net energy consumption. This implies that larger building volumes demand more energy, and with a fixed volume those buildings with more surfaces exposed to the outside environment demand higher rates of energy consumption. Additionally, the plot implies that passivity ratio – all perimeter parts of a building falling within 6 meters (19.68 ft) or twice the ceiling height from the façade that can potentially be naturally lit and ventilated - have a high and negative impact on energy consumption. This implies that buildings with more passive zones demand less energy. Therefore, for lesser energy consumption, spaces within a building need to be approximately 12 meter (or less) in depth, and if a space (or a building floor

²⁹ community wide net envelope surface area-to-building volume

³⁰ building volume^{1/3}

³¹ envelope surface area / (building volume)^{2/3}

³² net volume of all buildings in a community/net area of all exposed walls in a community

area) has more depth and thus is larger, shorter heights can help with reducing the energy demand of that space or building.

According to the plot, open space ratio of a community has a negative relationship with net energy consumption. This means the more unbuilt area in a community, the lesser rates of energy demand in that community. On the other hand, floor space index which is the ratio of gross floor area to total ground area has a positive relationship with energy consumption implying that buildings with larger floor areas or higher number of stories demand more energy. The reasoning behind these two relationships is clear since in any specific regional boundary, as the total number of buildings or units within a building increases, the energy demanded by the entire community increases respectively. However, when it comes to developing a community microgrid, there is an optimal total number of buildings or operational units operating under the same microgrid infrastructure. There's no universal rule for how many operational units or buildings need to be clustered as a community microgrids. This is because the number of buildings or operational units which use energy to function, is highly associated to their type of use in a microgrid context. Mixed use index is the metric for measuring the diversity of buildings inside a community and is the ratio of net residential floor area to net ground floor area. The plot suggests a high and negative relationship between mixed use index and energy consumption. This indicates that residential communities or communities with less non-residential buildings, have lower rates of energy use since non-residential buildings demand more operational energy use than residential ones. But when considering the operation of a community microgrid, it's important to have an appropriate mix of building types with complementary energy use profiles in order to maintain a fairly consistent energy demand throughout the system. When a microgrid serves a range of complimentary energy users (for example a mix of residential, school, and office buildings) a relatively constant energy demand is being observed over a 24-hour period. For a microgrid, this translates into consistency in power demand and economic stability of the system. For instance, when a commercial center with peak hours from 8 a.m. to 5 p.m. is part of the same microgrid serving a residential area with peak hours in the mornings and evenings, such cluster of users provide a combined daily demand profile that is steady throughout the day. Therefore, in order to develop a community microgrid with the optimal number of operating units or buildings, one has to first determine the combination of load types in the regional boundary of the community microgrid including the presence of any anchor loads. As described in Chapter 2, anchor energy users are those who are likely to be in that location for many years in the future, such as hospitals and universities. Identifying the anchor energy users in a site is important because by knowing the pattern of their energy consumption, one can determine the number of complementary loads (such as residential and retail) needed to create a consistent energy demand in the community microgrid given the amount of energy that is potentially produced on site.

To summarize these analyses, some general rules and principles for spatially designing communities in San Diego that would maximize the energy performance of their underlying solar microgrid can be outlined as below:

- Dense and compact neighborhoods where the distance between buildings is minimal, not only increases rates of energy demand but also due to the effect overshadowing, the potential for installing PV panels on building and land surfaces are decreased. Specially in scenarios where building facades are used for mounting PV cells, it's important to keep south facing façade clear from any shadows from surrounding building to maximize the amount of produced solar energy. This is when the orientation of communities become important. Unlike for community scale energy consumption, community scale PV energy production is highly dependent on the orientation of buildings.
- Larger building volumes demand more energy.
- For a fixed volume, buildings with less surfaces exposed to the environment require lesser energy for space heating and cooling.
- For lesser energy consumption, spaces within a building need to be approximately 12 meter (or less) in depth,
 - If a space (or a building floor area) has more depth and thus is larger, shorter heights can help with reducing the energy demand of that space or building.
- More buildings or more operational units within a community result in more energy consumed by the community which in turn means the need for more PV energy to be produced in the community,
 - Note that there is no universal rule for how many operational units need to be clustered as a community microgrids as it's highly associated to their type of use in a microgrid context.
 - A community microgrid with a fairly consistent energy demand needs to have an appropriate mix of building types with complementary energy use profiles in order to maintain a fairly consistent energy demand throughout the system.

Chapter conclusion: The content of this chapter answers the third research question of this study: *“which combination of the attributes of urban form impact a solar community microgrid’s energy performance?”*

Focusing on community scale energy consumption, this chapter has identified those features of urban form which have the most impact in determining the amount of energy that a community consumes. The combination of artificial neural networks and Shapley values has been selected as the most appropriate model for handling the complexity and non-linearity of the desired relationship as well as statistically describing it. The architectural interpretation of the statistical results magnifies the importance of considering

the combined effect of all energy-relevant attributes of urban form when studying its impact on community wide energy consumption in the context of community microgrids.

Transforming or developing new urban areas require a holistic understanding of the tradeoffs between the impact that every single design decision has on the operational energy demand of community buildings. Although the results of this part of the research are context-specific, the proposed methodological framework can be applied to any city worldwide given the availability of the required datasets. Moreover, the findings of this chapter are specifically useful as it engages the design sector in the technical conversation of planning solar community microgrids, leading to energy self-sufficient spatial designs of microgrid-connected communities well before their development.

Chapter 7 Community PV Energy Production and Urban Form

Introduction: San Diego county has been selected as case study due to several reasons outlined in Chapter 1, one of them being the county's rich repository of GIS spatial data as well as monthly energy consumption data. Unlike energy consumption data, accessing measured PV energy production data from all existing PV panels in San Diego was not feasible³³ and therefore necessitated to find a way to synthesize the required dataset. The focus of this chapter is on the development process of a *generation-evaluation* algorithm in Grasshopper and Python environments for synthesizing PV energy production data along with their corresponding urban forms. This algorithm is intended to generate numerous different spatial community design scenarios based on San Diego's principles of urban planning, and then to evaluate the communities' potential of PV energy production during 12 months of a year.

The *generative* part of this algorithm – responsible for generating different random spatial community design scenarios - was developed by extracting San Diego's urban shape grammar and then translating the grammars into a generative code. However, the *evaluation* part of the algorithm - responsible for evaluating the PV energy production capacity of each generated community and creating the synthetic dataset – was not developed due to the technical limitations of existing PV energy evaluation plugins used in Grasshopper. The occurred technical difficulty consequently resulted in impeding the creation of the synthetic dataset as originally planned, but it was an affirmation of the limitations of existing simulation tools for urban scale energy analysis and an emphasis on requiring energy simulation tools capable of simulating urban scale energy performances.

Although the *evaluation* part of the *generation-evaluation* algorithm was not developed, but the *generative* part, referred to as the *urban generator*, has had another purpose to its development. Since the urban generator algorithm is developed based on the local zoning standards and urban planning principles of San Diego, any community designs generated by this algorithm will be similar to the existing urban fabric of that city. This is particularly important and useful when developing the real-time predictive software prototype of this research that will be extensively explained in Chapter 8.

This chapter describes the developed urban shape grammars associated with San Diego's urban planning principles, elaborates on the generative tool that was codified upon, and explains the existing limitations with current parametric PV simulation tools that prevented from fully developing the intended generation-evaluation algorithm. The urban shape grammars formulated and described herein contributes to the field of shape grammars by capturing urban planning rules that take topographic constraints into consideration.

³³ Due to the unavailability of such data accessible to the public.

Moreover, the subsequent generative tool has been developed as an extension to CltyMaker - an existing rule-based parametric urban design tool based on shape and description grammars – and therefore is pushing forward existing research³⁴.

7.1 A Generative Urban Design Algorithm

7.1.1 Generative Design

Generative design is a design automation technology based on different artificial intelligence techniques ranging from shape grammars and procedural rule-based systems to genetic algorithms and advanced machine learning based processes (Gu, Singh, & Merrick, 2010). Although overlaps and similarities exist among the different generative design techniques, some appear to be more suitable for certain design and automation tasks (Gu, Singh, & Merrick, 2010). For example, rule-based systems like shape grammars tend to be used when there is strong domain knowledge and stochastic systems like genetic algorithms are used when there is weaker domain knowledge. To elaborate, methods of generative design can be classified into being explicit or implicit depending on the availability and complexity of data:

- *Explicit* - teaching an AI by feeding it human-readable information related to what the programmer/designer thinks the generative system needs to know for generating design options. Rule based systems are an example of explicit methods.
- *Implicit* - where raw data is fed into an AI so the algorithm can analyze and construct its own implicit knowledge about the design such as different machine learning based generative models i.e., autoregressive models, variational autoencoders and generative adversarial neural networks.

The content of this chapter is constructed upon explicit methods of generative design where a common procedure towards design is practiced: instead of laboriously designing a single artifact, a design system is programmed which starts with a set of design goals, constraints, and variables - often stemmed from the designer's past experiences and the environment where the design is situated - and then innumerable possible permutations of a solution are explored.

In the following several sections the development of a generative urban design tool based on shape grammars is described. This tool was developed with the purpose of synthesizing the required datasets by generating numerous community design scenarios for any given district boundaries. Focusing on San Diego as this research's case study, shape grammars

³⁴ Part of this chapter was published in eCAADe 37/ SIGraDi 23 conference proceedings: Rahimian, M., Beirão, J.N., Duarte, J.M., Iulo, L.D., (2019), A Grammar-Based Generative Urban Design Tool Considering Topographic Constraints: The Case of American Urban Planning. eCAADe SIGraDI 2019 Proceedings, Volume 3 (pp. 267-276). Porto, Portugal.

were used to extract the rules and patterns forming the urban structure of each urban typology. The extracted urban shape grammar was then used as the basis of a generative design tool producing various urban design scenarios considering the limitations and potentials of the site's topology. The novelty of this tool lies in considering the topographic constraints of the site and generating various alternatives of urban design scenarios accordingly.

7.1.2 Decoding Urban Typologies in San Diego: A Shape Grammar Approach

The use of shape grammars has gained popularity in design of various scales ranging from urban design (Beirão, 2012) to housing (Benros, Duarte, & Hanna, 2015) and product design (Garcia & Barros, 2015). Shape grammars is a method of shape generation theorized by Stiny and Gips (1972). A specific shape grammar is an algorithmic process defined by a set of shape rules where the shape generation process starts with an initial shape and then the shape rules are recursively applied leading to the generation of indefinite various design options. Shape grammars can be *analytical* - used to describe an existing design style – and/or can be *synthetic*, used towards generating new styles and designs (Li, 2001). In this chapter, an analytical shape grammar is developed to describe the principles and logic of urban planning in San Diego and then is transformed into a synthetic grammar used to generative new spatial configurations of urban and community design scenarios accordingly. The synthetic shape grammar is then codified into a generative tool using Grasshopper and Python to generate various spatial urban design scenarios.

The reason for shape grammars' popularity is conceivably for its fourfold benefits as stated by Beirão and Duarte (2018): (1) analytical purposes where design rules and processes are understood, (2) synthetic purposes where new design alternatives are generated, (3) regulatory purposes where the limitations of new design alternatives are controlled, and (4) prediction purposes where simulation algorithms are developed accordingly. A rich body of literature has used shape grammars for unveiling patterns of urban design. The first urban grammar was proposed by Catherine Teeling (1996) for describing the generation of urban form in the docklands of Friedrischafen, Germany. Duarte et al. (2007) presented an urban grammar for the Marrakesh Medina, an ancient fabric that grew spontaneously over time, resulting in a complex urban form that expressed social conventions and physical conditions. Following these analytical grammars, Duarte and Beirão (2011) proposed the use of shape grammars for supporting flexible urban planning. Then, in 2013, Barros et al. (2013) presented a grammar for describing a Mozambican slum, Cidade dos Caniços, an informal settlement where land is divided prior to occupation. More recently, Verniz and Duarte (2017) developed a grammar to describe the spontaneous growth of favelas in steep terrains and Ena (2018) proposed a shape grammar to describe the development of favelas in flat terrains, both in Rio.

The part of the research as presented in this chapter adds to the current conversation on

urban grammars by capturing planning rules that take topographic constraints into consideration. In the work of previous researchers, the site's topography did not play a pivotal role in the development of the urban grammar. The only precedent which explores the role of topography is the work of Verniz and Duarte (2017) in which topography acts as the main stimulus for shaping the form of urban structure in informal settlements. However, unlike the work of Verniz and Duarte, our investigations do not see topography as the primary force impacting urban structures but explored the different ways topography distorts and changes the formal stream of planned urban structures.

(i) *San Diego Urban Typologies*



Figure 41 - A portion of San Diego's map form 1979. Source: www.sunnycv.com

Orthogonal grids form the structural core of the majority of American cities and towns (Southworth & Owens, 1993). An orthogonal grid is made of a series of parallel streets which run at right angles to each other forming a grid. The popularity of orthogonal grid plans lies in creating an efficient pattern for maximum land use where the emerging square or rectangular blocks build an interconnected variety of routes that are not only vehicle-oriented, but also are readily expandable for potential urban growth scenarios. The New York City grid is perhaps the most recognized grid plan in the American history. Houston, Texas, is an example of a city constituted of several grids oriented in different directions.

Other cities, such as Chicago, Salt Lake City, San Francisco, and Charlotte to name a few, follow the same principle of an orthogonal grid plan.

Orthogonal grids are especially used for structuring flat, non-steep areas. American cities tend to follow an orthogonal grid pattern until the pattern gets distorted due to topography resulting in organic or mixed grid arrangements. This structure is especially apparent in San Diego due to its varied topography (Figure 41) which has caused for different urban typologies to emerge throughout the region (Figure 42):

- *Organic grids*: are often created where the topology fluctuates considerably within the region limits and close-to-narrow ridges and alleys are observed. In organic grids, buildings are constructed on the ridges and alleys where minimum construction standards are met.
- *Orthogonal grids*: which as described before are normally planned and constructed where a massive area of the site is flat or has a limited steepness. According to San Diego's GIS data, a flat area is where the slope is less than 15% making it suitable for orthogonal urban grids.
- *Mixed grids*: are the combination of the previous two typologies. Mixed grids emerge where the orthogonal grid meets the edge of a steep slope and gets distorted, organically, in consequence.

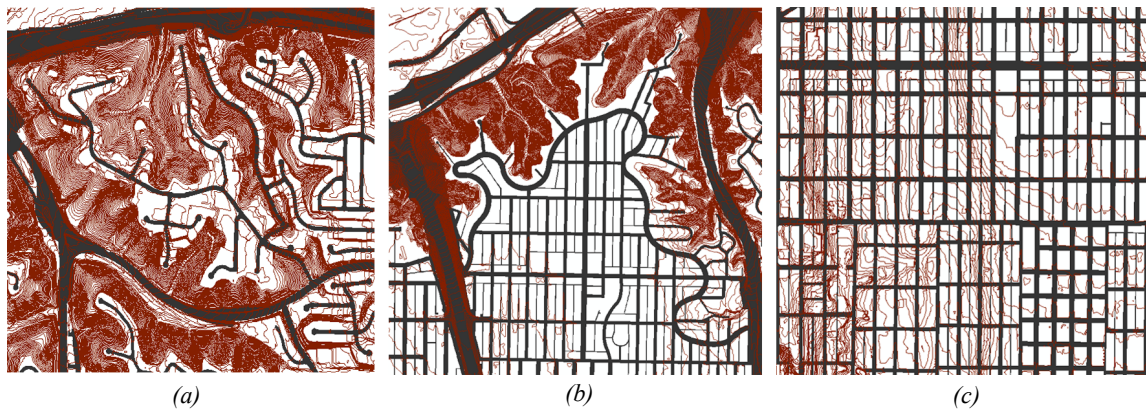


Figure 42 - Samples of identified three urban typologies: (a) organic grid, (b) mixed grid, (c) orthogonal grid

(ii) Development of an Urban Grammar for San Diego

As stated earlier, most regions in San Diego (as well as in any other American city) tend to follow an orthogonal grid pattern until the pattern gets distorted due to topography resulting in an organic or a mixed grid arrangement. This principle forms the backbone of the urban shape grammar developed for San Diego in this research; rather than recommending three separate set of grammars for each one of the identified urban typologies, one general set of grammars is developed that formalizes the described principle and therefore covers the underlying shape rules of all three urban typologies.

Following the lines of the above-mentioned principle, the urban grammar introduced in this study is divided into two parts: the first part is the grammar for generating orthogonal grids, and the second part is the grammars explaining the logic behind grid distortions when reaching the edge of steeped slopes. Note that the grammar is developed in an abstract way such that the underlying design procedure is not strictly specific to San Diego and could be used for recurrent design purposes fitting many urban design scenarios in American contexts.

- *Urban shape grammar part I: Orthogonal grid*

The grammar proposed for orthogonal grids in this section has been adopted from the underlying grammar of CIttyMaker with minor changes. CIttyMaker is a rule-based parametric urban design tool based on shape and description grammars (Beirão J. N., 2012). CIttyMaker was developed as the generative component of the City Induction project which integrates various design support tools for the formulation, generation, and evaluation of urban design scenarios (Duarte J. P., Beirão, Montenegro, & Gil, 2012). The adopted urban grammar generates the compositional axes of the urban plan leading to the generation of orthogonal grids, and patterns to generate urban blocks.

The design of an orthogonal grid starts with creating the main perpendicular axes (a_v = main vertical axis, a_h = main horizontal axis) directed towards the identified major viewsheds or major gateways of the region. The main axes are then assigned major road widths based on San Diego's street design requirements (Figure 46). Before starting the generation process the initial attributes of the shapes need to be labeled: site limitations as *Ls* and references as *Ref*. The intervention site limit is always a closed polyline and references are elements inside *Ls* which have been either selected by the designer or contextualized inside *Ls* as regional landmarks (Figure 44 and 45). Note that the *Ls* boundary is a clipped selection of a region which its inner area is flat with minimum steepness suitable for an orthogonal grid construction. This means that the immediate area outside of *Ls* could all be flat (then *Ls* is entirely made of straight lines) or could all be steeped and surrounded by mountains and valleys (then *Ls* is entirely made of curved lines) or could be a mix of flat and steeped (then *Ls* is made of straight and curved lines). This means that if *Ls*, as a closed polyline comprised of straight and curved lines, is placed in the context of an area, outside its straight lines the area would still be flat, while the curved lines represent the edge of steep slopes (Figure 43).

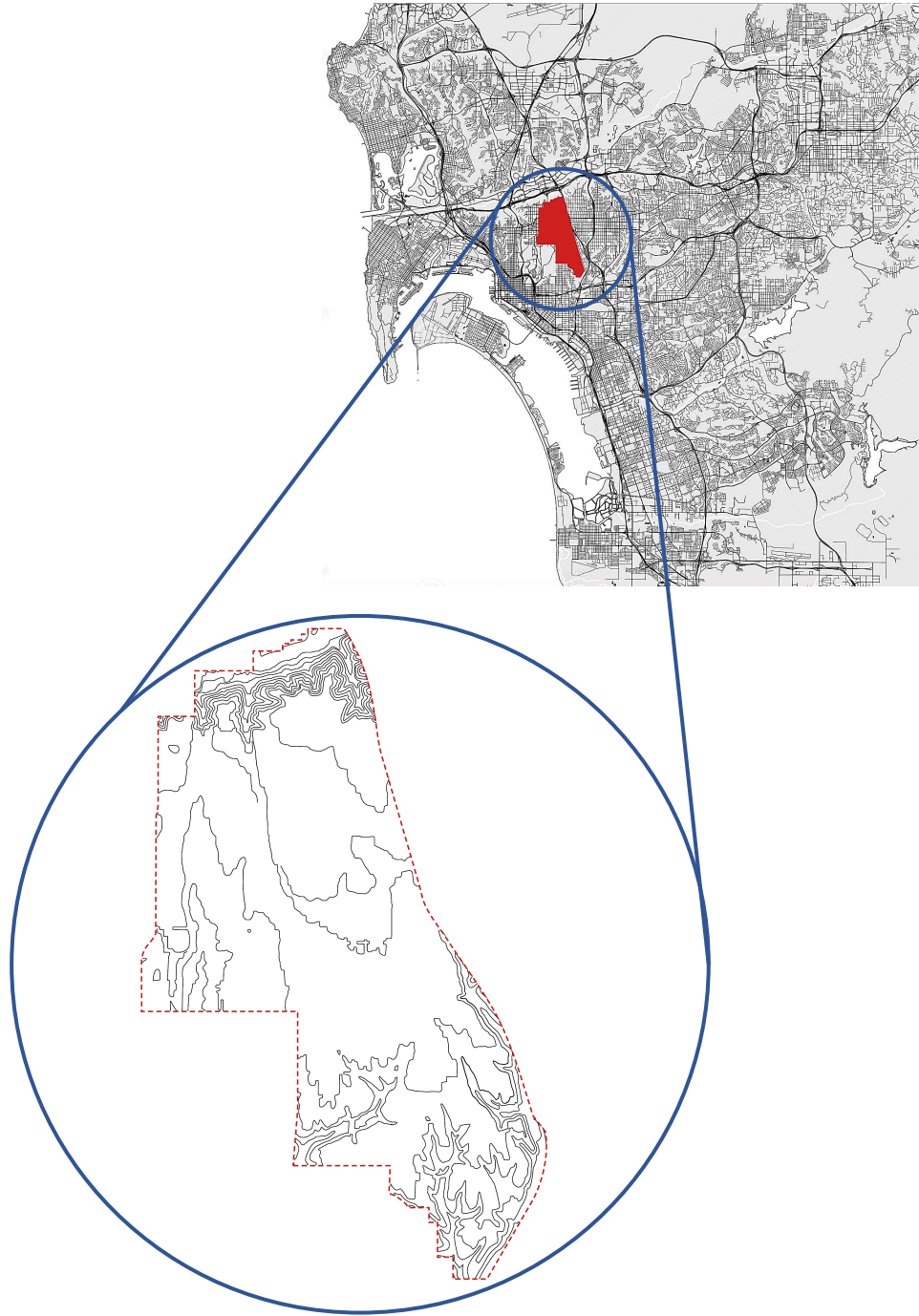


Figure 46 – The selected region in this image represents Ls boundary which is a clipped selection of a region which its inner area is flat with minimum steepness suitable for an orthogonal grid construction. This means that the immediate area outside of Ls could all be flat (then Ls is entirely made of straight lines) or could all be steeped and surrounded by mountains and valleys (then Ls is entirely made of curved lines) or could be a mix of flat and steeped (then Ls is made of straight and curved lines).

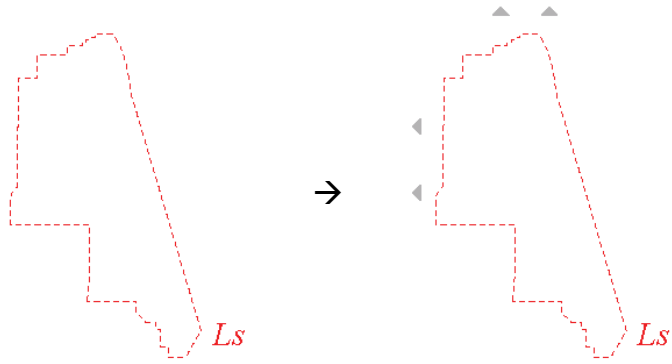


Figure 44 - Identify viewsheds

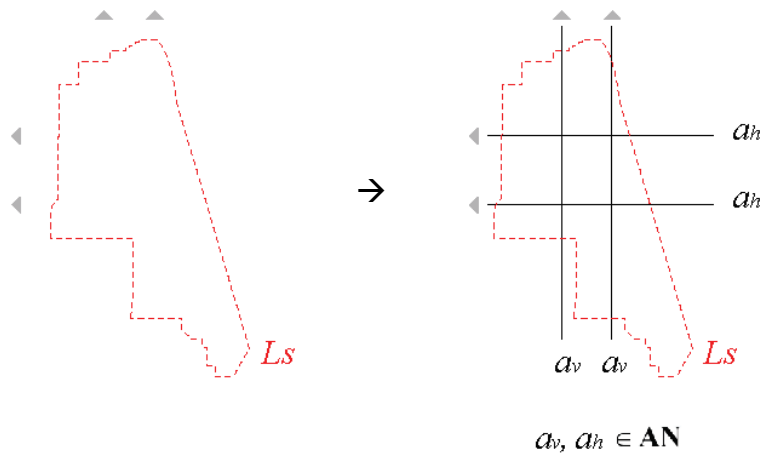


Figure 45 - Drawing the main road axis towards directing the viewsheds. $\mathbf{AN}$ is the collection of road axes.

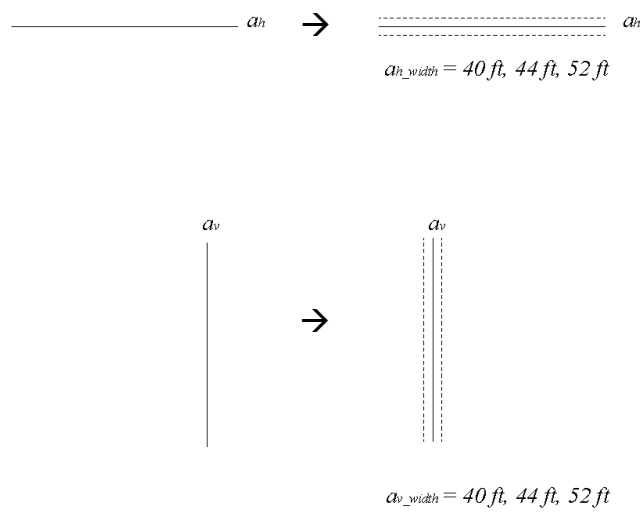


Figure 46 - Assigning major street width to the main axes. The width dimension is adopted from San Diego zoning requirements.

The rest of the streets, responsible for the compositional structure of the grid, are generated in parallel to a_v and a_h . The default distances between the set of horizontal and vertical streets are the size of the predefined block length and width. In the case of San Diego and according to the city's planning framework, block sizes diverse spanning from 200-300 ft (approx. 60-80 m) in width to 300-750 ft (approx. 90-230 m) in length. The rule application can be better understood in Figure 47. The upward and downward arrows labels are used for the recursive application of rules and indicate the direction for applying the rule. The rule applies recursively until it falls outside L_s . The width of the streets depends on the proximity of a *Ref* point to it. For instance, if the distance of *Ref* point from the generated street equals the sum of block length and half of the reference street's width, then the generated street becomes a major street by changing its width to the width of a major street. The widths of major and minor streets are specified according to the county's zoning street design guidelines. When the *Ref* is in between two streets, depending on its distance to each of those the rules illustrated in Figures 48, 49, 50, and 51 apply which draw all possible vertical and horizontal axes in relation to any specified reference point (Beirão, Duarte, & Stouffs, An Urban Grammar for Praia: Towards Generic Shape Grammars for Urban Design, 2009).

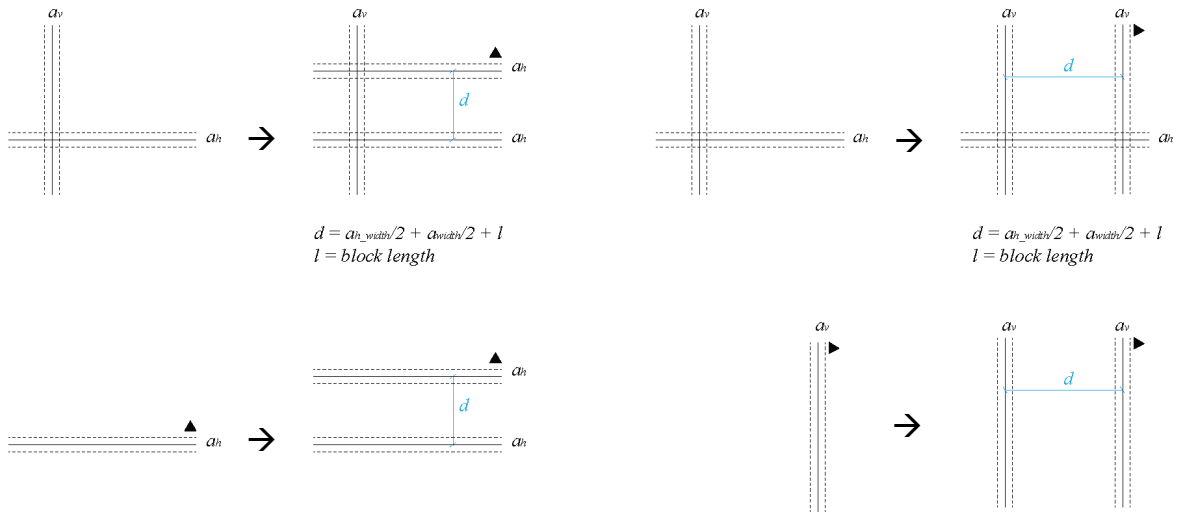


Figure 47 - Rules for generating the compositional structure of the orthogonal street grid.

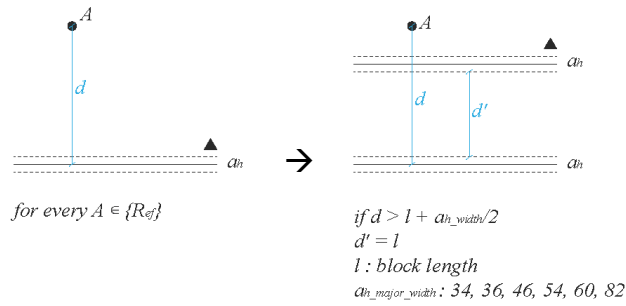
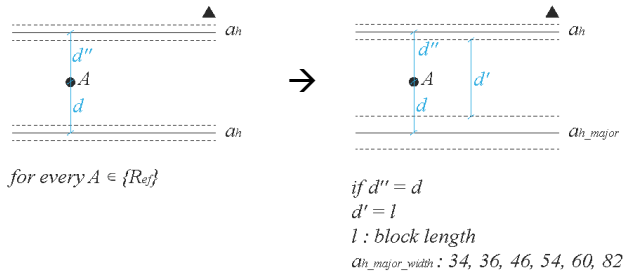
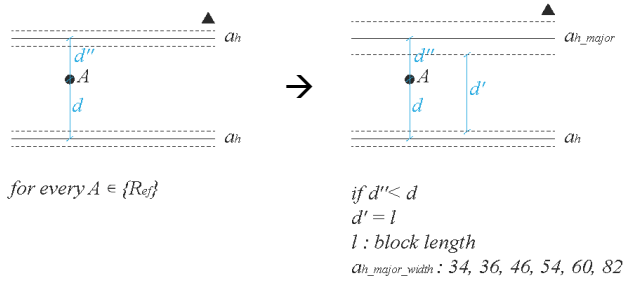
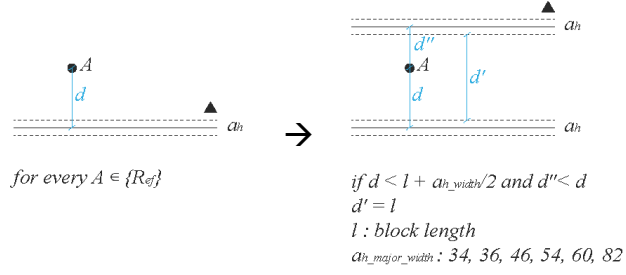
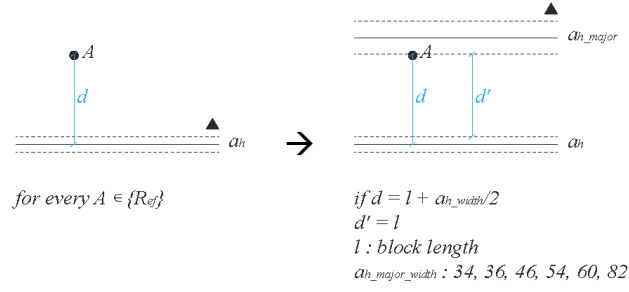


Figure 48 - The width of horizontal streets depends on the location of Ref.

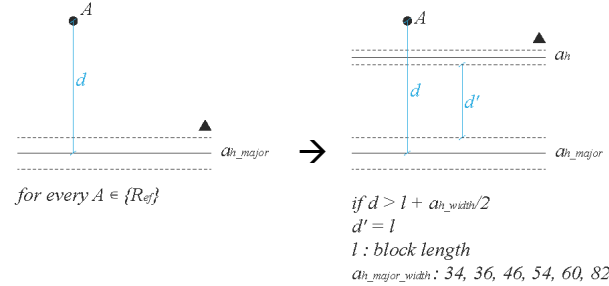
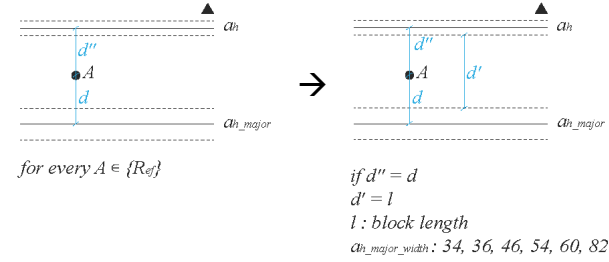
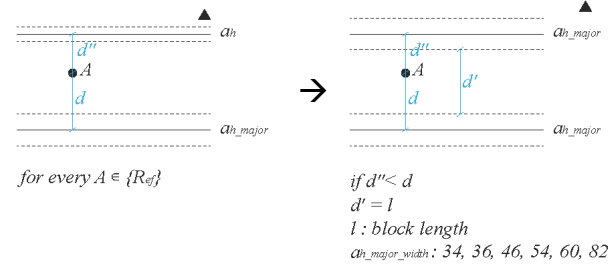
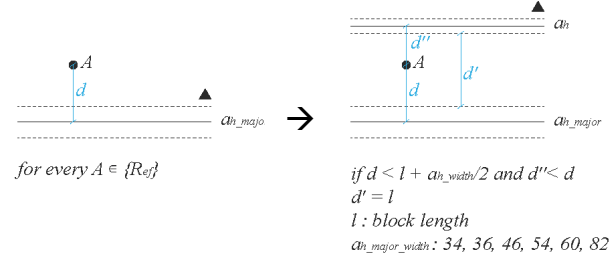
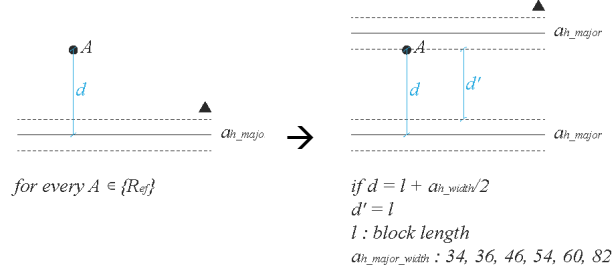


Figure 49 – The width of horizontal streets depends on the location of Ref (continued).

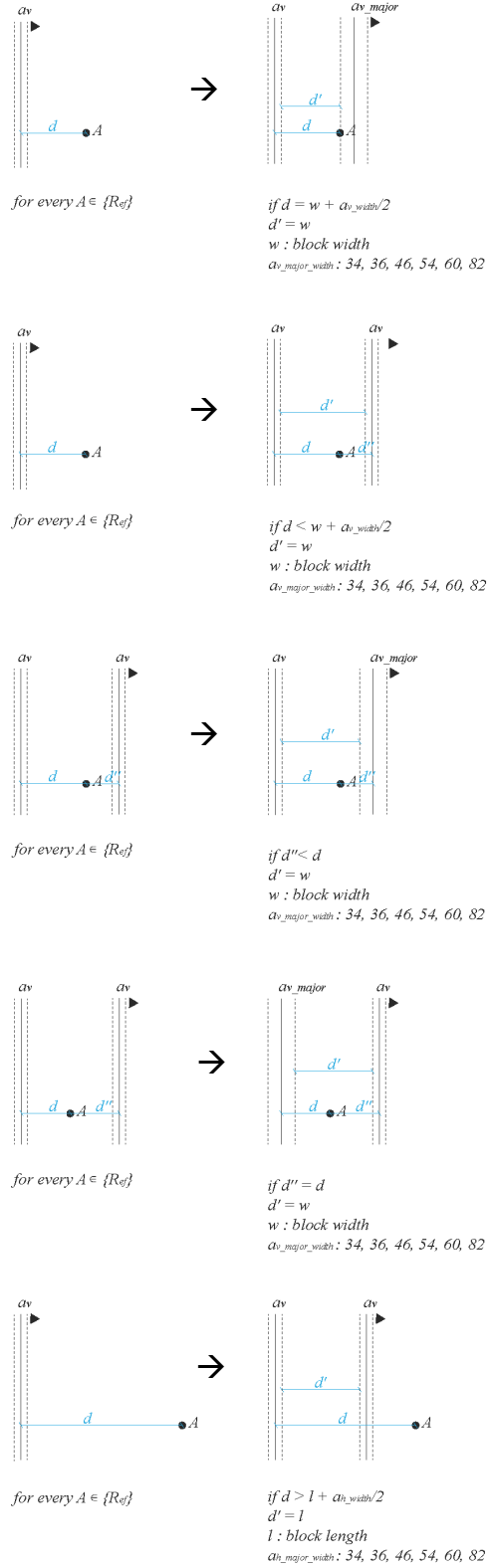


Figure 50 – The width of vertical streets depends on the location of Ref.

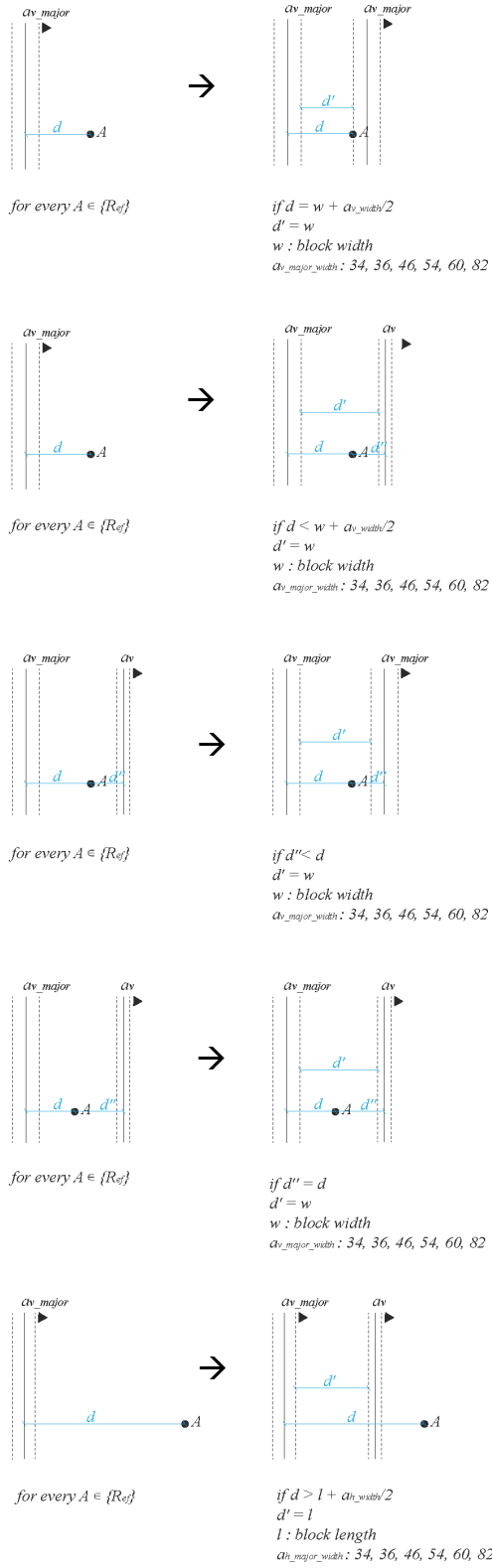


Figure 51 – The width of vertical streets depends on the location of Ref (continued).

After all orthogonal axes are drawn in the L_s boundaries, a rule for trimming parts of the axes outside L_s is applied (Figure 52). By applying these set of rules, an orthogonal grid pattern is constructed within the boundaries of L_s . But L_s is composed of straight and curved lines. As described earlier outside the straight lines of L_s the area is flat while the curved lines represent the edge of a steep slope where grid distortions appear. Next section describes the modification applied to the orthogonal grid where it meets steeped slopes.

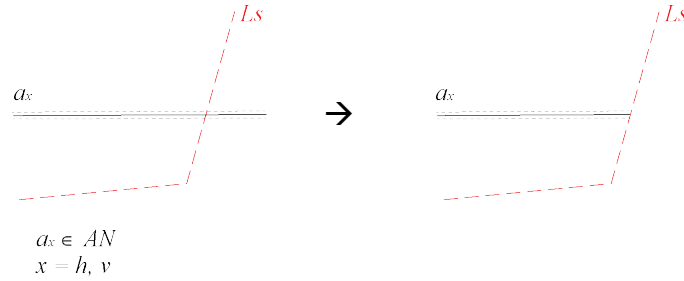


Figure 52 - Rule for trimming axes outside the L_s boundaries

- *Urban shape grammar part II: Grid distortions*

Hills, slopes and mountains are represented on a topography map using contour lines. In a topography map, areas where the contour lines are close together the slope is steeper and where the contour lines are further apart the slope is shallower. As mentioned above, according to San Diego's topography map and GIS data³⁵, construction is done in areas where the steepness is less than 15%. Accordingly, the urban grid follows orthogonal arrangements in such areas and distorts when it reaches

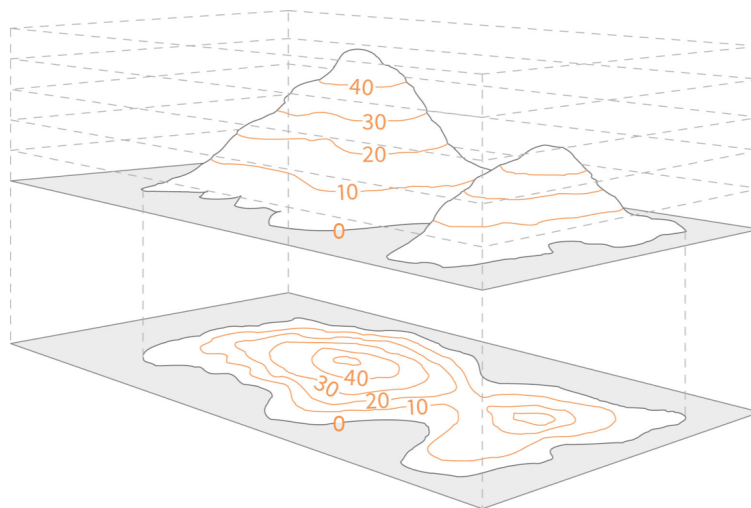


Figure 53 - Contour code 0 representing edge of a steeped slope. Source: getoutside.ordnancesurvey.co.uk

³⁵ <http://www.sangis.org/download/>

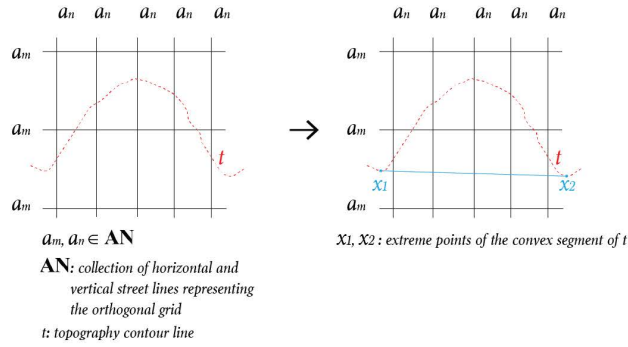
the edge of steep slopes where the incline exceeds 15%. The contours of interest for developing the grammars of grid distortions are the contours representing the edge of a steep slope, labeled as t (with elevation code 0 in Figure 53).

The findings of this research show that patterns of grid distortions vary depending on: (1) the size of the concave areas created by the t contour line, as well as (2) the narrowness of the concave segments of the t contour line itself. The size of the concave area is compared against the area of one urban block and the narrowness of the concave segment is measured by the diameter of a circle inscribed in the concave segment of t and compared against the length of an urban lot. Therefore, the distortion grammars are categorized into four classes representing four different regional conditions that trigger change in the orthogonal grid:

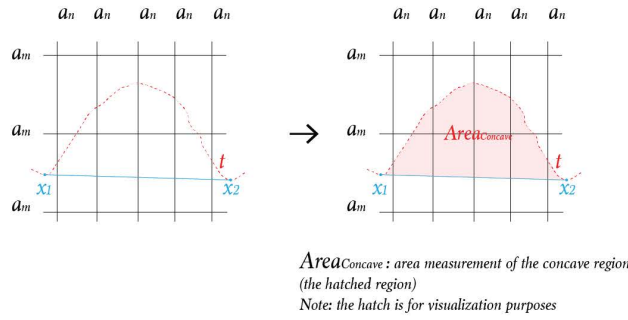
- Large concave area with wide concave segment:
 $Area_{Concave} > Block_{Area}$ AND $d > w$;
- Large concave area with narrow concave segment:
 $Area_{Concave} > Block_{Area}$ AND $d \leq w$;
- Small concave area with wide concave segment:
 $Area_{Concave} \leq Block_{Area}$ AND $d > w$;
- Small concave area with narrow concave segment:
 $Area_{Concave} \leq Block_{Area}$ AND $d \leq w$.

Where $Area_{Concave}$ = area of concave region, $Block_{Area}$ = area of one block, d = diameter of the inscribed circle c , and w = distance between two consecutive street lines or block width. Thus, for any given area of interest that includes the t contour line, the orthogonal grid is first created using the grammar described previously, then for creating the grid distortions, the grammars as described in Figures 54, 55 and 56 are applied.

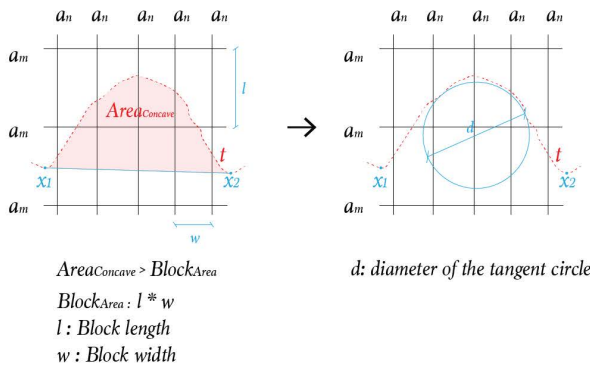
Find the concave segment of contour



Find the area of concave segment



$Area_{Concave} > BlockArea$



$Area_{Concave} \leq BlockArea$

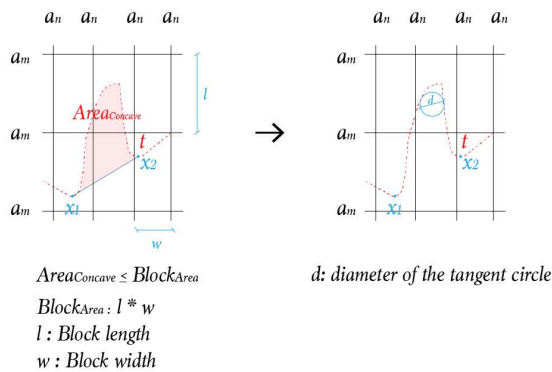


Figure 54 - First grammar from top is for finding the concave segment of the t contour line. Second grammar is, finding and measuring the concave area. The bottom two grammars are used for comparing the size of the concave area to the area of one urban block.

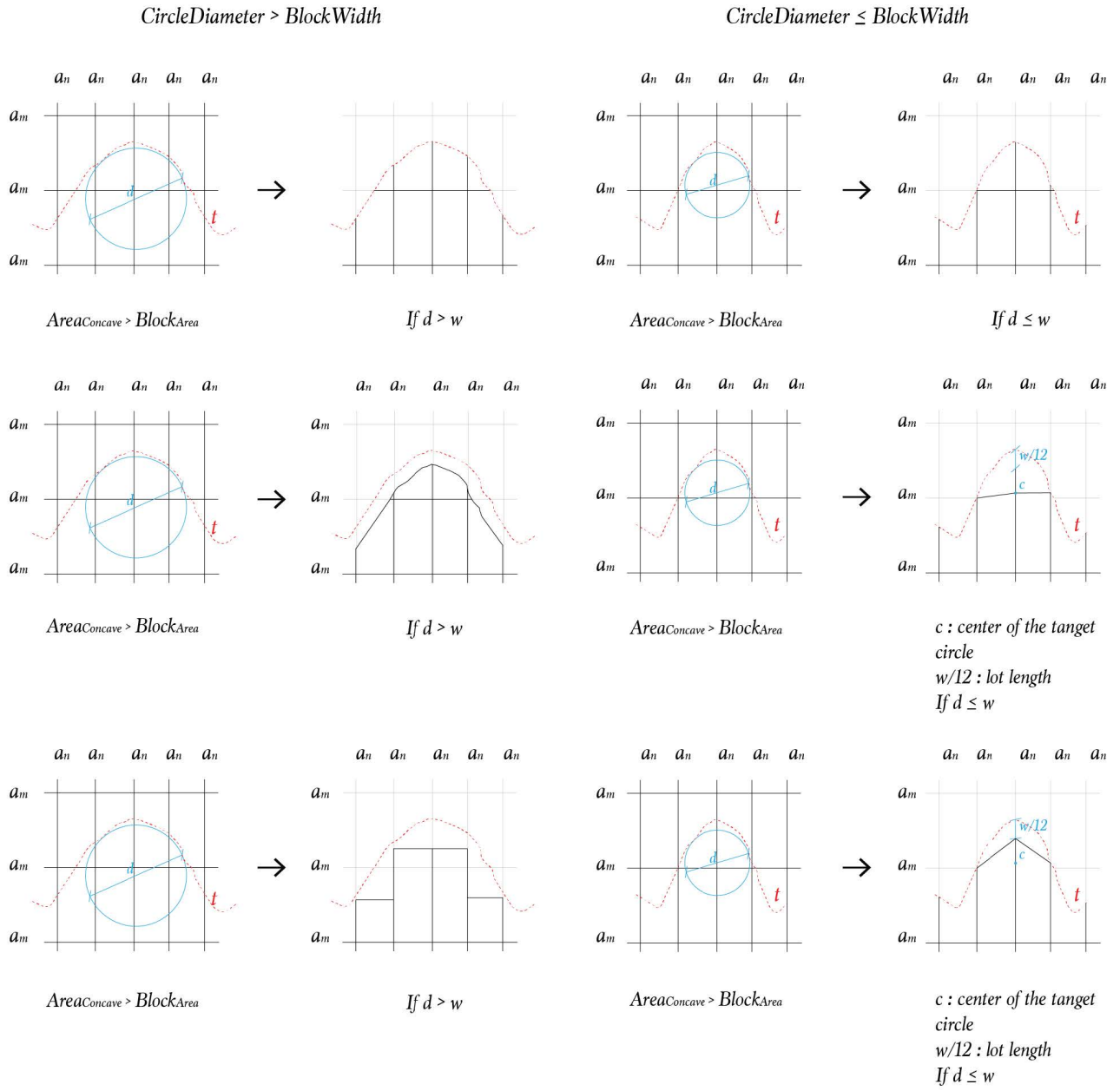


Figure 55 - Grammar for comparing the size of the inscribed circle to the width of one urban block when $Area_{Concave} > BlockArea$

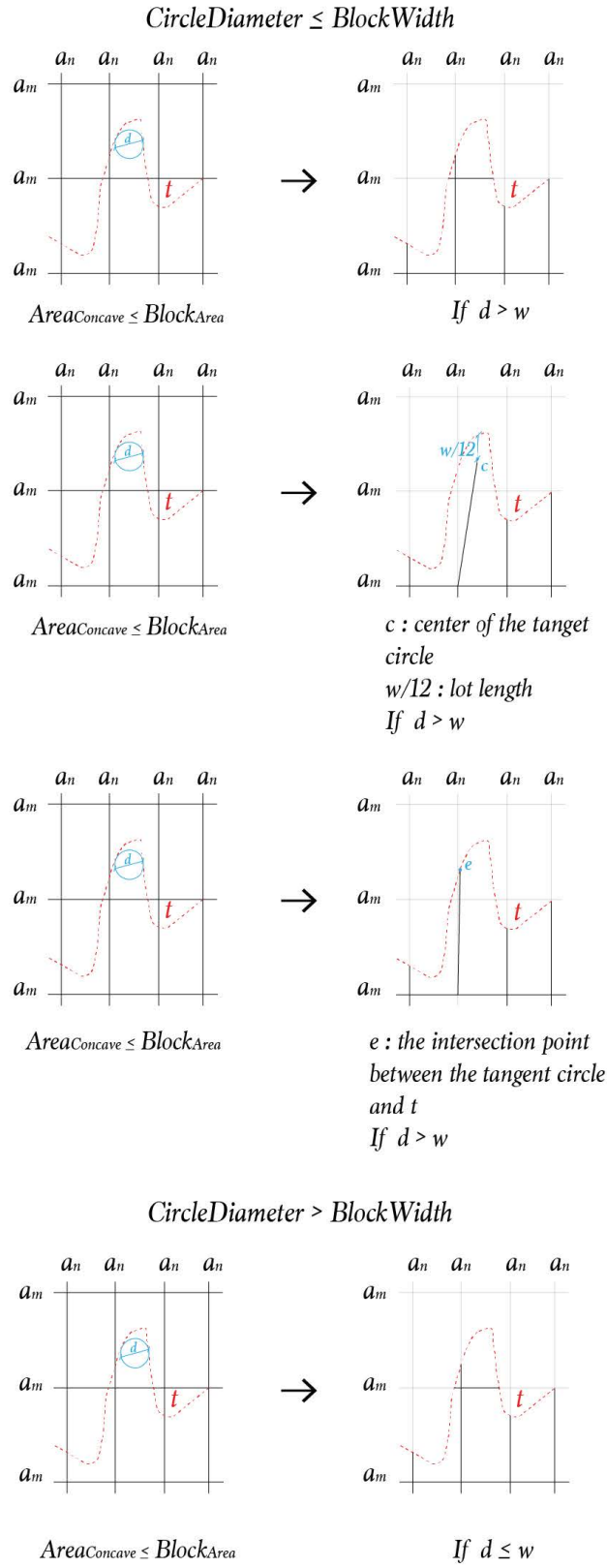


Figure 56 - Grammar for comparing the size of the inscribed circle to the width of one urban block when $\text{Area}_{\text{Concave}} \leq \text{BlockArea}$

With the proposed set of grid distortion grammars, all orthogonal axes that meet the edge of steep slopes of the Ls boundary get modified accordingly. The next step is to add blocks to the cells that are created from the different axes and Ls intersection. Figure 57 shows the different intersection and cell conditions and rules for adding blocks to those cells.

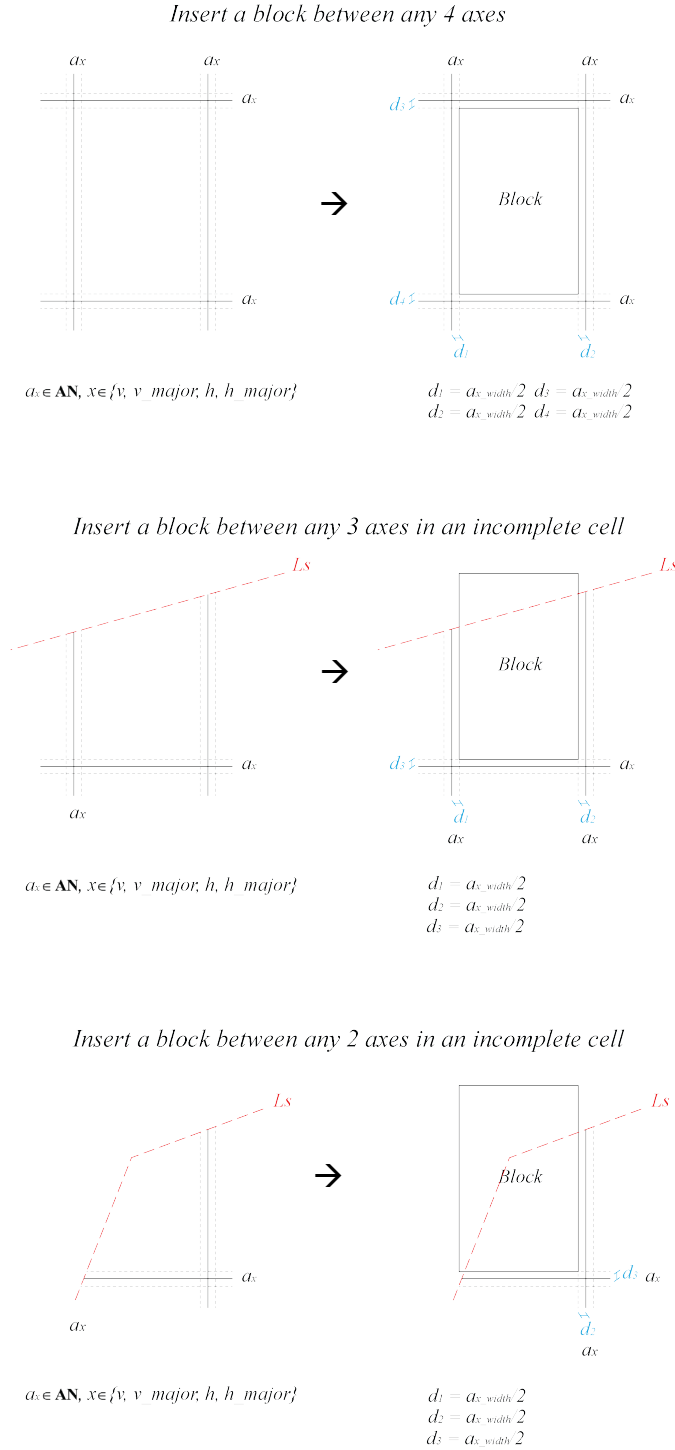
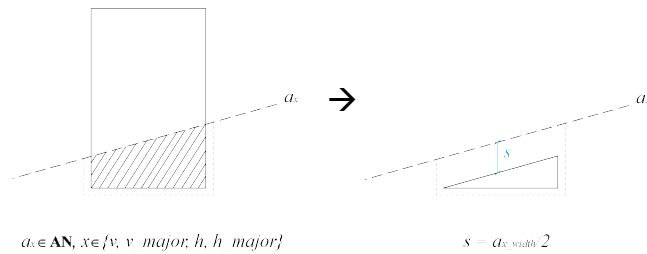
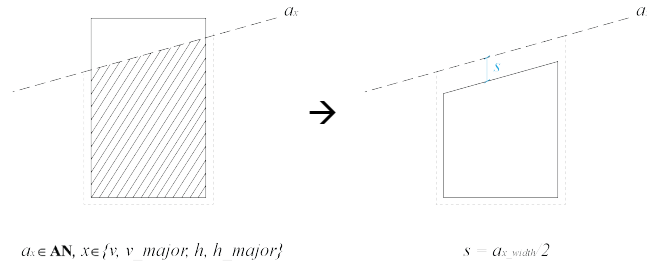


Figure 57 - Rules for inserting blocks to cells.

Figure 58 and 59 shows rules for adjusting those blocks that fall outside cell boundaries. The rules laid out in Figures 57, 58, and 59 are adopted from “*Monitoring Urban Design*

Reduces the size of a block so that it fits inside the bounded area giving an additional buffer area corresponding to the width of the street a_x which is coincident with the border



Reduces the size of a block so that it fits inside the bounded area. Ls is the edge of a steep slope and no street is coincident with the border line

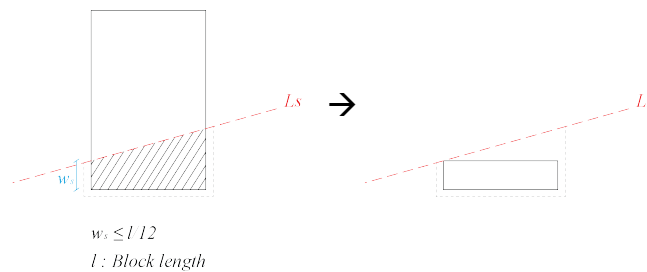
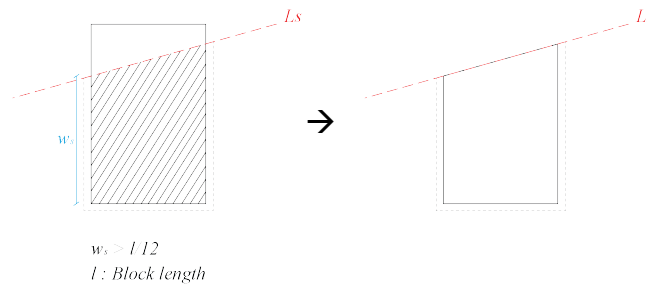


Figure 58 - Rules for adjusting blocks that fall outside the cell boundaries.

Through Generative Design Support Tools: A Generative Grammar for Praia” (Beirão J. N., Duarte, Montenegro, & Gil, 2009) and modified with respect to the urban planning requirements of San Diego.

Reduces the size of a block so that it fits inside the bounded area and creates an a_x street.

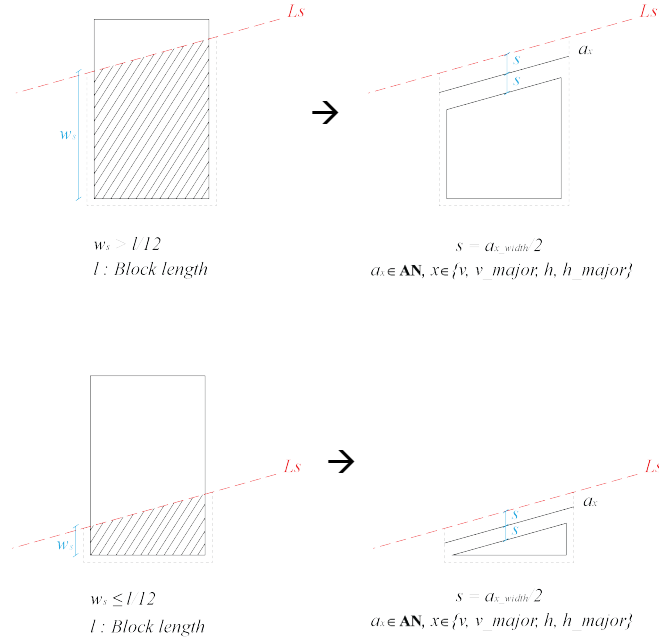


Figure 59 - Rules for adjusting blocks that fall outside the cell boundaries (continued).

After placing the initial blocks inside the cells and adjusting them, the next iteration of rules are applied that either erases the small blocks (Figure 60) or combines them with adjacent blocks to create larger ones (Figure 61).

Erasing small blocks if the block area is smaller than 10% of the regular block size or smaller than a prefixed minimum block area or $w_s \leq l/12$

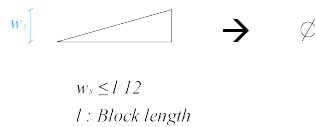
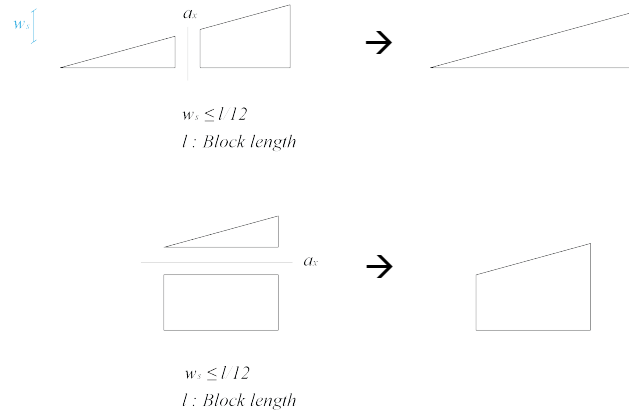


Figure 60 - Rule for erasing small blocks.

Instead of erasing them, small blocks can be joined with others through rules:
 Joining below minimum block with another incomplete block
 if the block area is smaller than 10% of the regular block size or smaller than a
 prefixed minimum block area or $w_s \leq l/12$



Incomplete blocks can be joined together even when none of them are below the minimum block size.

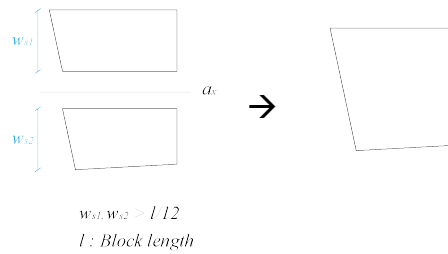


Figure 61 - Rules for combining blocks to create larger blocks.

After creating the urban blocks, the next step is to divide the blocks into urban lots. Depending on the shape of urban block (rectangular or non-rectangular) and its location relative to the L_s boundary different rules are applied for generating the lots. For complete

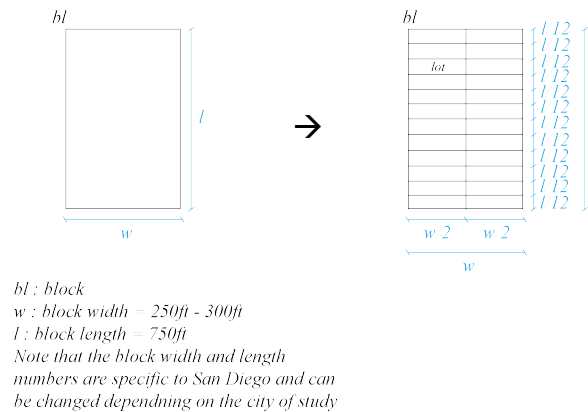


Figure 62 - Complete rectangular blocks are divided into 24 uniform rectangular lots.

rectangular blocks, regardless of their location inside L_s , they will be divided into 24 urban lots by dividing the block width to two and the block length to 12 (Figure 62).

However, the urban blocks adjacent to the L_s border or to the a_x streets which are coincident to the L_s border, often do not have rectangular shapes. For these blocks, two different set of rules may be applied for generating urban lots: one is to map a grid of 24 urban lots on these non-rectangular blocks and only choose those lots which are completely inside the block (Figure 63), the second strategy is to divide the urban block into smaller

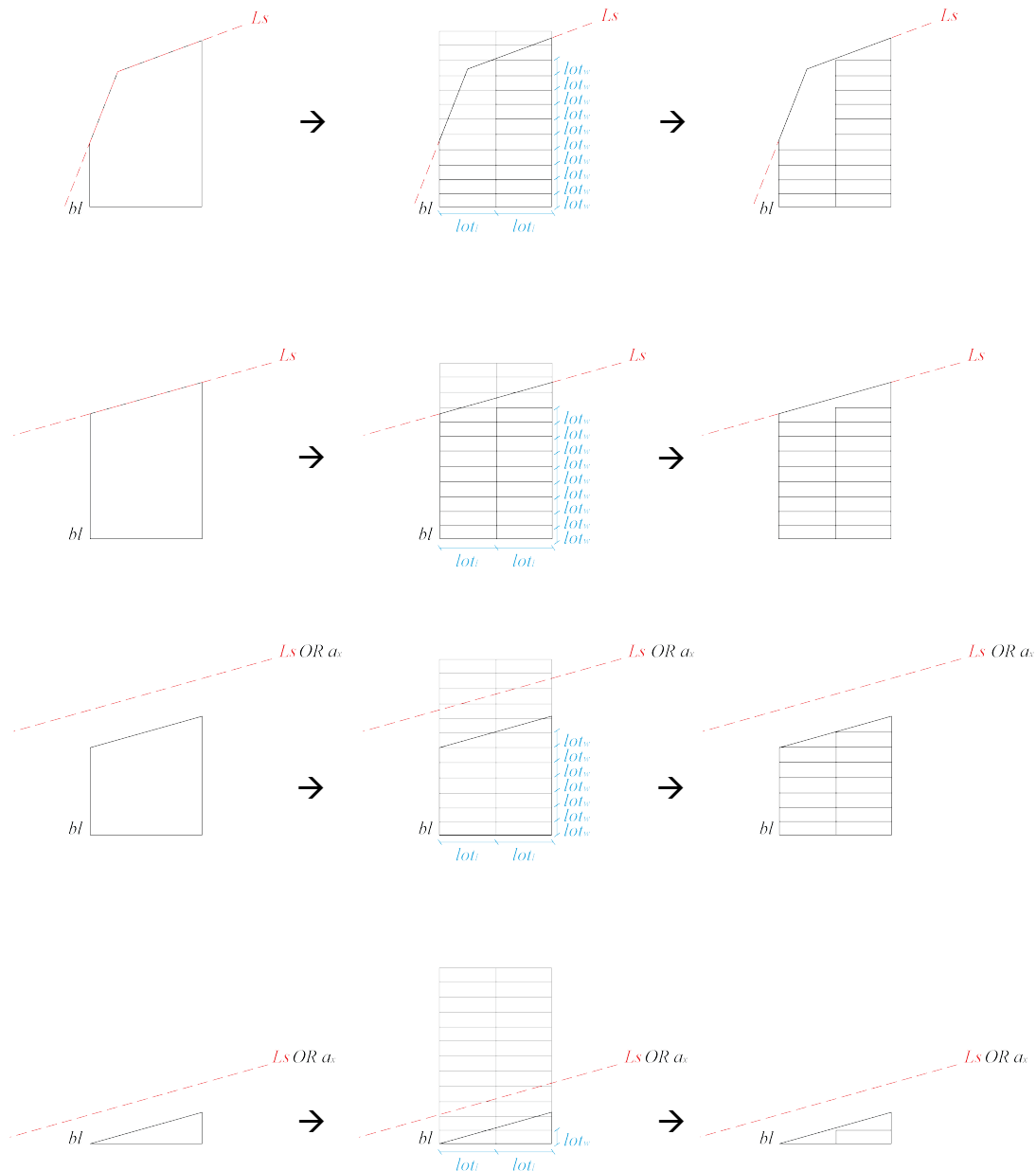


Figure 63 - Mapping a grid of 24 urban lots onto non-rectangular blocks and selecting those lots completely inside the blocks.

blocks by recursively offsetting the Ls border or the coincident ax street in the distance of the length of an urban lot, until the offset goes beyond that specific block (Figure 64-65).

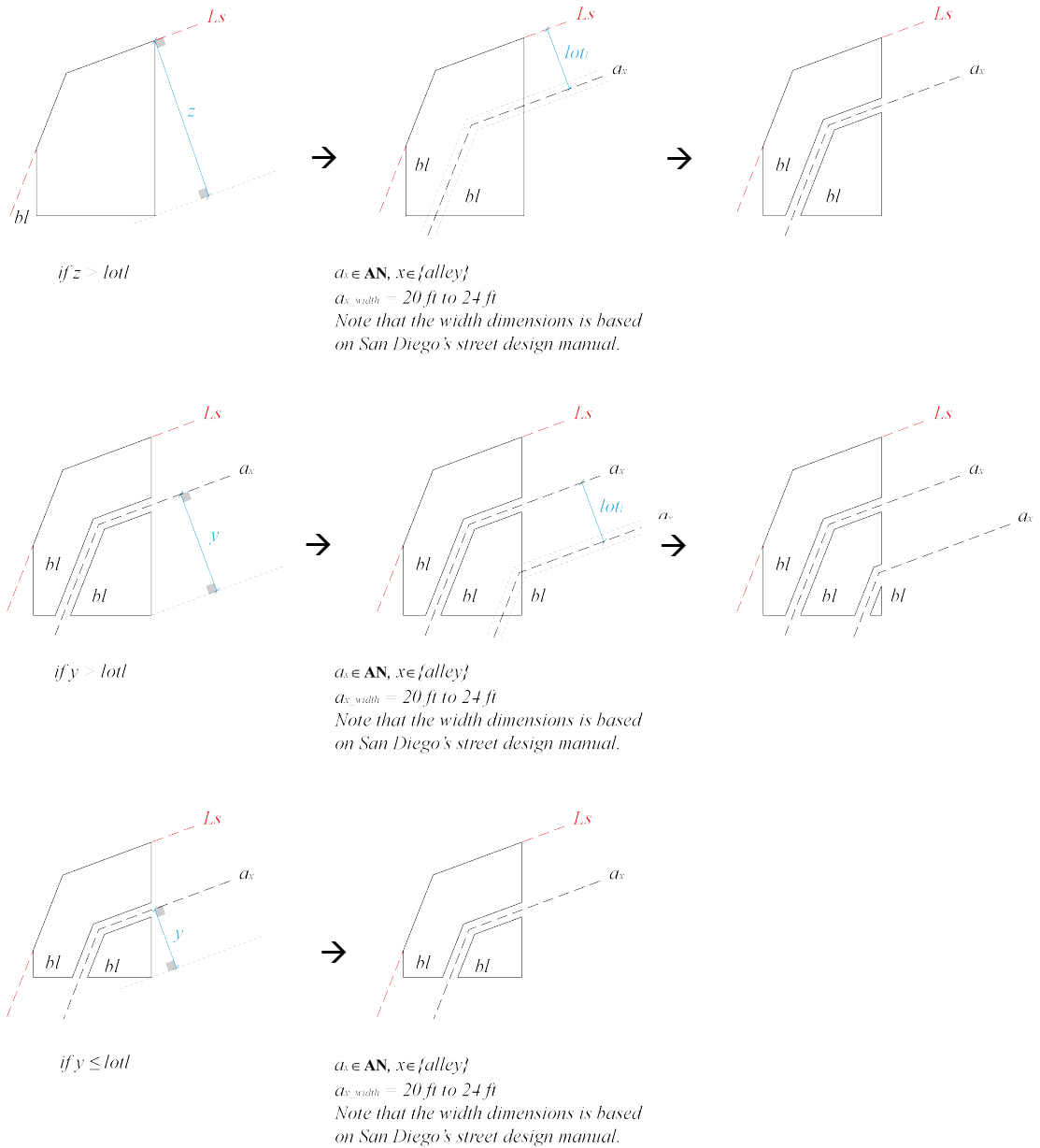


Figure 64 - Dividing non-rectangular urban blocks into smaller ones by recursively offsetting the Ls border or the coincident ax street.

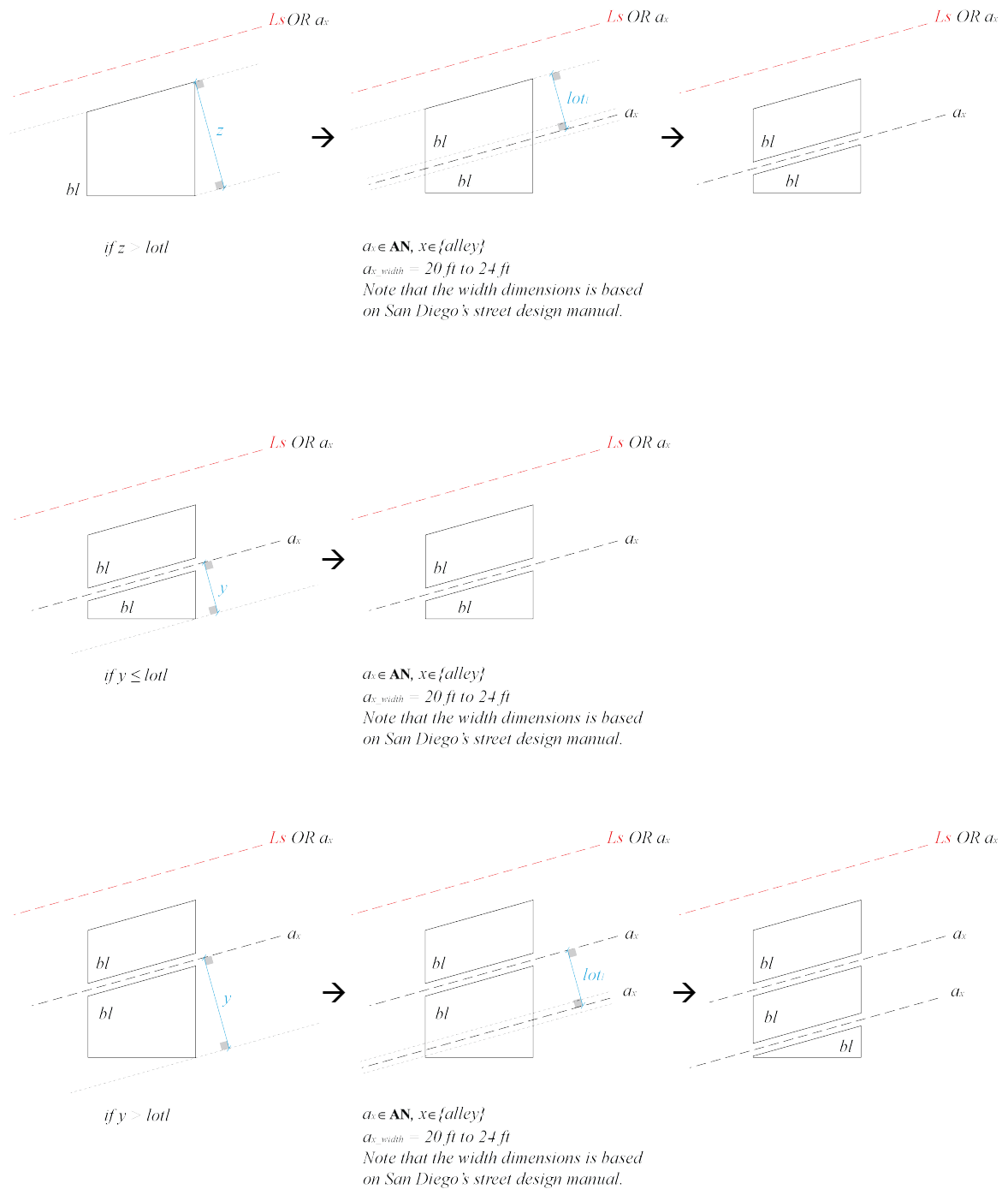


Figure 65 - Dividing non-rectangular urban blocks into smaller ones by recursively offsetting the Ls border or the coincident ax street (continued).

For the second strategy, another set of rules are applied that divide the generated blocks into lots: identify the edge of the urban block adjacent to L_s or a_x , identify the edge facing (in front of) the first identified edge, divide both edges into segments with a preset length (the width of an urban lot) starting from the same end of the edge, then connect the resulting division points of the two facing edges one by one (Figure 66). In this process the two facing edges might have different number of division points and the divided segments might have different shapes resulting in irregular shaped urban lots in different sizes.

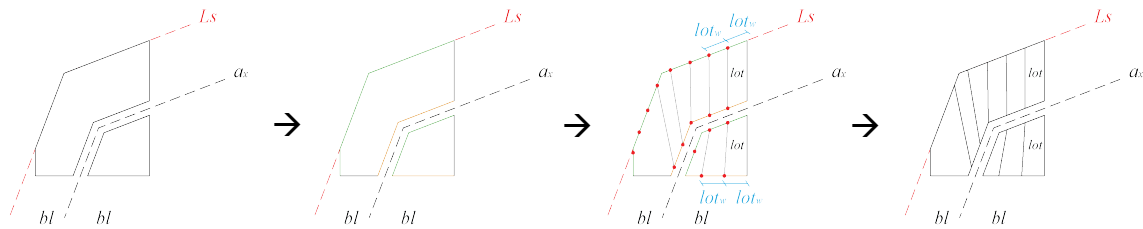


Figure 66 – Process of creating urban lots in non-rectangular urban blocks: Identifying the edge of the urban block adjacent to L_s or a_x , identifying the edge facing (in front of) the first identified edge, dividing both edges into segments with a preset length (the width of an urban lot) starting from the same end of the edge, connecting the division points one by one.

Similar to the urban blocks' grammars, after creating urban lots some rules are applied where lots could be combined with other lots to make larger ones (Figure 67). After all urban blocks and urban lots are created, the axes lines are erased (Figure 68).

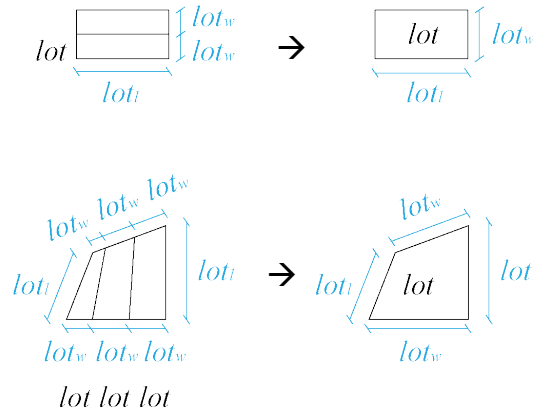


Figure 67 - Rules for combining urban lots to make larger ones.

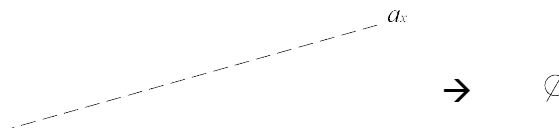


Figure 68 - Deleting axes after all block and lots have been created.

(iii) Codifying the Urban Grammar

The described shape grammars are used for descriptive and generative purposes. Given any desired *Ls* boundary, by following the grammars one would be able to create an orthogonal grid of urban blocks and urban lots which is responsive to the nuances of the topography. The next step in this process was to translate the grammars into code. Grasshopper (a visual programming tool built on top of Rhinoceros®) and Python were used to codify the grammars into a generative tool (Figure 69).

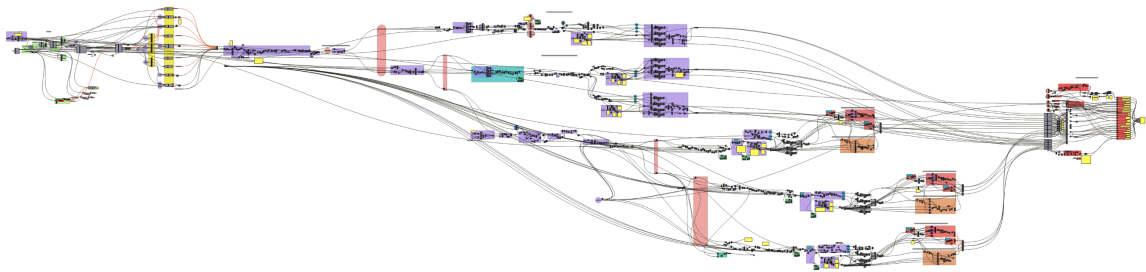


Figure 69 - The Grasshopper + Python script of the generative tool which is based on San Diego's urban planning principles and urban shape grammars.

The end goal of the generative tool is to mimic exactly the behavior of the shape grammars towards creating various urban block and lot layouts for any given *Ls* boundary and populate them with green areas and buildings based on the zoning restrictions of San Diego. This section will not go into the details of the code but rather explains its behavior from a high-level point of view which entails evaluating the site, creating the urban grid, and distorting it where needed. The developed generative tool takes the following steps for any given region that includes a polygon representing the *Ls* boundary:

- *Checking the status of the inputted *Ls* polygon:* this step is to check whether the polygon is closed or open, and to close the polygon if it is open. The generative tool only works with closed polygons.
- *Smoothing *Ls* boundary:* since *Ls* is a clipped selection of a region consisting of flat areas and edge of steep slopes, the inputted or drawn polygon is expected to be a large collection of small segments and control points. This step simplifies the polygon by removing the unnecessary control points and joining the segments to create larger ones while retaining the overall shape of *Ls*. The output of this step is a smoothened closed planar curve.
- *Finding the concave and convex segments of *Ls*:* as mentioned earlier, the inner area of *Ls* is flat with minimum steepness suitable for an orthogonal grid construction. This means that the immediate area outside of *Ls* could all be flat (then *Ls* is entirely made of straight lines) or could all be steeped and surrounded by mountains and valleys (then *Ls* is entirely made of curved lines) or could be a mix of flat and steeped (then *Ls* is made of straight and curved lines). The goal of this part of the code is to evaluate the concavity and convexity of different segments

of L_s in order to identify which parts are a continuation of a flat area and which parts are the edge of steep slopes. The respective code places numerous points on L_s and evaluates the concavity and convexity of the curve at each of these points. By identifying all consequent series of points that have the same curvature, the algorithm outputs segments of the curve which are concave and segments that are convex.

- *Creating the orthogonal grid*: the orthogonal grid grammar as described earlier is used at this stage to generate the orthogonal grid inside L_s . For this, the code inputs the orientation of the grid as well as the rectangular urban block dimensions (width and length). These inputs can be altered at any time in the design generation process in order to output different design scenarios.
- *Finding the areas where grid distortions occur*: the next step is to find the areas where the orthogonal grid distorts due to topography. For this, the extreme points of the convex segments of L_s are identified and connected to create the boundary limits of the orthogonal grid. By finding the region difference between L_s and the boundary limits of the orthogonal grid, the concave areas where grid distortions happen are created.
- *Distorting the orthogonal grid as a response to topography*: at this stage, the concave areas – in which the grid distortions happen - are categorized into two groups: larger than a block size and smaller than a block size. For each group, a circle is inscribed into the concave segment of the curve to further categorize the areas into ones with narrow concavity and ones with wide concavity. The narrowness of the curvature is measured by its inscribed circle and is compared against d as described in the distortion grammars. Ultimately the concave areas are categorized into four groups:
 - Large concave areas with wide concave segment:
 $Area_{Concave} > Block_{Area}$ AND $d > w$.
 - Large concave areas with narrow concave segment:
 $Area_{Concave} > Block_{Area}$ AND $d \leq w$.
 - Small concave areas with wide concave segment:
 $Area_{Concave} \leq Block_{Area}$ AND $d > w$.
 - Small concave areas with narrow concave segment:
 $Area_{Concave} \leq Block_{Area}$ AND $d \leq w$.

Where $Area_{Concave}$ = area of concave region, $Block_{Area}$ = area of one block, d = diameter of the inscribed circle c , and w = distance between two consecutive street lines or block width. According to the distortion grammars, each of these four groups have a set of grammars and grid distortion rules that have been codified and are applied to them.

- *Applying grammar for creating urban blocks and lots:* all steps until now have been towards generating the grid pattern of the district. At this stage, the grammars related to creating urban blocks and lots, adjusting and modifying them are codified and applied respectively.
- *Populating urban lots with buildings or converting to green areas:* After all lots are generated and finalized, they could be either labeled as a green area or a building could be placed in the lot following the site setback requirements of San Diego. Lots larger than 20,000 sqft are labeled as green area and the rest of the lots have a zoning code randomly assigned to them. The utilized zoning codes, following the zoning handbook of San Diego³⁶, are labeled as B, C, E, F, G, H, I, J, L, M, N, O, P, R which allow for mixed use, residential, and non-residential construction. Depending on the lot's assigned zoning code different height and setback distances are applied to it³⁷. After all lots (except for green areas) have a zoning designation

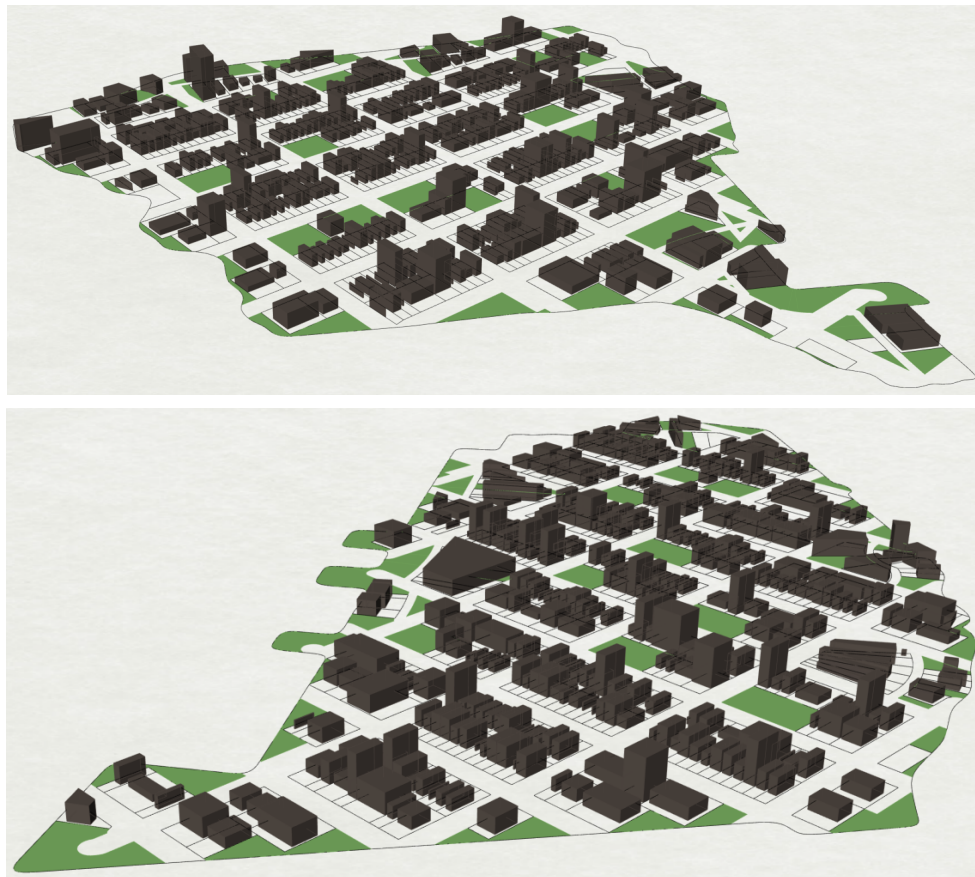


Figure 70 - Two samples of community designs generated by the generative tool.

³⁶ <https://www.sandiegocounty.gov/content/sdc/pds/zoning.html>

³⁷ Zoning code = ['B', 'C', 'E', 'F', 'G', 'H', 'I', 'J', 'L', 'M', 'N', 'O', 'P', 'R']
 Height (ft) = [20, 25, 30, 30, 35, 35, 35, 40, 45, 50, 55, 60, 60, no-limit]
 Rear setback (ft) = [50, 25, 15, 25, 40, 25, 25, 25, 25, 25, 25, 15, 15]
 Side setback (ft) = [15, 15, 1, 5, 10, 10, 7.5, 5, 5, 5, 5, 1, 1, 1]
 Front setback (ft) = [60, 60, 60, 60, 50, 50, 50, 50, 50, 50, 50, 50, 50]

assigned to them, the code identifies the front, rear, and side edges of each lot depending on their adjacency to the streets, offsets the edges according to their specific setback distance, and places a building with max footprint and max height on each urban lot. Figure 70 shows a sample of the different community design generations by this tool.

7.2 Adding PV Energy Simulation Properties to the Generative Tool

The next step after developing the generative part of the tool, was to add PV energy evaluation features to it. That is for each of the generated community design scenarios the tool would evaluate the communities' potential of PV energy production during 12 months of a year. The planned strategy was to identify and categorize all south, north, east, and west facing facades and to evaluate their PV energy production potential using relevant environmental analysis plugins in Grasshopper.

Existing stable tools for simulating PV energy production in the Grasshopper environment, such as DIVA³⁸ and Ladybug tools³⁹, operate based on the EnergyPlus engine. EnergyPlus is one of the main engines for running building energy simulations developed by the US Department of Energy (DOE) and is used as the operating engine behind many building energy simulation programs. In a paper authored by the DOE and some national laboratories (Crawley, Hand, Kummert, & Griffith, 2008), the authors provide an extensive comparison of the features and capabilities of major building energy simulation programs including EnergyPlus. All the reviewed simulation programs in this paper, which as mentioned are still⁴⁰ the major ones in the industry, utilize complex hard coded mathematical models and physics-based computation for running simulations. The pros and cons of physics-based computation will be extensively described in Chapter 8. The main downside of these programs is the extremely time-consuming process for running simulations specifically at an urban scale which numerous building energy models need to be created for buildings across the region of study.

The generative tool of this chapter was tested and combined with DIVA and Ladybug tools in two separate attempts. While theoretically time was the only expected constraint when running urban scale energy simulations but practically, the program raised more extreme geometry-related and technical difficulties when attempting to create numerous energy models for running PV energy simulations at scale. With these attempts, it became apparent early in the process that simulating PV energy production for numerous community scale urban designs for the purpose of synthesizing the data is a task beyond the capacity of existing tools. Therefore, studying the impact of urban form on PV energy production has

³⁸ <http://solemma.net/Diva.html>

³⁹ <https://www.ladybug.tools/>

⁴⁰ As of 2020.

been listed as a next step of this research.

Chapter conclusion: This chapter explains the development of a generative urban design tool based on shape grammars. The novelty of this tool lies in considering the topographic constraints of the site and generating various alternatives of urban design scenarios accordingly. The shape grammars obtained herein are based on a case-study which is an appropriate urban representative of most American cities. With the use of shape grammars, the rules and patterns forming the urban structure of each typology have been decodified. Since the inferred design language has generative properties, it is important to note that the tool generates urban design scenarios within the same design language but for different topological contexts. The extracted urban shape grammar is then used as the basis for a generative design tool producing various urban design scenarios considering the limitations and potential of the site's topography. This chapter describes the extracted urban shape grammars and how that informs the development of the presented generative urban design tool. Since the urban generative algorithm is developed based on the local zoning standards and urban planning principles of San Diego, any community designs generated by this algorithm will be similar to the existing urban fabric of that city. This algorithm is therefore used as part of the real-time predictive software prototype for the reasons described in Chapter 8.

Another important conclusion of this chapter is affirming the limited technical capacity of existing energy simulation tools for urban scale energy analysis and emphasizing on the necessity for having energy simulation programs that do not rely on complex mathematical models and physics-based computations (more on this is discussed in Chapter 8).

The study presented in this chapter contributes to the field of shape grammars by creating a set of rules that take topography into account. By implementing the rules into a tool, it also is extending CItYMaker and is pushing forward existing research.

Chapter 8 A Realtime Predictive Software Prototype

Introduction: Chapter 3 identified an existing gap in the current software tools that evaluate the energy performance of any given community microgrid design scenario by the virtue of its urban spatial configuration. It was concluded that perhaps it's the lack of tools as such that has prevented architects and urban planners to be actively involved in community microgrid design development processes. Additionally, in Chapter 6 it was realized that a strong relationship exists between urban form and energy consumption of communities. This relationship was the described in term of what it meant for developing solar community microgrids and the value of engaging the design sector in the development process. A spatially aware design and assessment of solar community microgrids as described in Chapter 6 brings the need for a new generation of computational modeling, simulation, and evaluation tools for the field that currently does not exist.

This chapter discusses the development of an urban scale energy simulation software prototype that predicts the monthly value of energy consumed of any inputted community design scenario by assessing its urban form. Unlike existing building energy simulation tools, this prototype does not operate on a hard-coded backend that takes *hours* to run; it rather uses predictive trained models to estimate the monthly value of energy consumption for any designed community scenario in *real-time*.

An imprecise simulation of energy performance provided in real-time while designing an urban area is much more practical and useful for an architect and urban designer than a software that takes hours to run a simulation but provides a more accurate number. With a real-time tool as such, designers would be able to change their designs and to see the simulation results instantaneously while with the latter, the designs are rather not to be changed to hit the energy targets. Therefore, a real-time tool based on surrogate models as the one prototyped herein, would be useful for architects and urban planners for spatially modeling solar community microgrids and evaluating their performances. Moreover, for urban scale energy planning and spatial solar community microgrid designs, an approximate number is sufficient to guide the designs towards more energy aware ones. This prototype and its development process shed light for future researchers to develop similar tools that can run more comprehensive urban scale energy simulations with the purpose of designing and constructing any types of community microgrids ⁴¹.

8.1 Surrogate Modeling for Short-Circuiting Simulation

The adoption of computer simulation codes and programs have been extensively used in many science and engineering fields as a flexible way to study complex and real-world

⁴¹ This chapter along with parts of Chapter 7 are currently in process for publication in the "Smart Cities and Climate- Resilient Urban Planning" special issue of *Environment and Urban Planning B* journal.

phenomena under controlled environments (Gorissen, Couckuyt, Demeester, Dhaene, & Crombecq, 2010). Compared to expensive physical experimentations, computer simulations are economically more efficient and improve the quality of engineered products and services. However, the downside of simulation activities, especially when the problem of interest reaches a certain level of complexity, is the high cost of computing a simulation that may take hours and days to perform (Forrester, Sobester, & Keane, 2008). Calculating and simulating the energy behavior of one or a group of buildings, is a good example of a complicated phenomenon that simulating its performance could easily become extremely time and resource consuming.

Energy performance in buildings is influenced by many different environmental factors such as weather conditions, building specific characteristics such as materials and structure, building operational components such as HVAC system, and occupants and their behaviors. To accurately calculate and simulate the energy behavior of a building, all physical functions and thermal dynamics related to these parameters need to be factored in the computing process which ultimately results in complex mathematical models (Magoules & Zhao, 2016; Seyedzadeh, Pour Rahimian, Glesk, & Roper, 2018). Over the past 60 years, hundreds of building energy programs have been developed that utilize these mathematical models with different levels of complexity depending on the number and type of parameters they incorporate for running energy performance measurements (Crawley, Hand, Kummert, & Griffith, 2008). Common to all existing energy performance tools - whether it runs analysis at the building scale or city scale, it's stand-alone such as TRNSYS⁴² or integrated into computational design workflows such as Ladybug tools and DIVA - is that their backends are all hardcoded physical laws for the derivation of building energy performance (Crawley, Hand, Kummert, & Griffith, 2008; Magoules & Zhao, 2016; Tamke, Nicholas, & Zwierzycki, 2018). Although the simulation results of these tools are effective and accurate, but in practice they bear some difficulties: firstly, they require lots of input parameters about the building and its environmental context which information on these parameters might not be accessible to all users. Secondly, the hardcoded backend results in extremely time-consuming computing processes making running simulations a tedious task to perform. This is specifically accurate when simulating buildings' energy performance at a city scale. In such cases enormous amount of time and resources need to be dedicated to creating building energy models for hundreds or thousands of buildings across a city in order to run urban scale energy simulations effectively.

Due to these reasons, researchers have utilized methods other than physics based and engineering methods to estimate building energy performance in a less time and resource consuming way. One way to deal with this problem (and one complex simulation problem in other field of science and engineering), is to construct simpler approximation models to

⁴² <http://www.trnsys.com/>

develop a relationship between input and outputs and predicting performance. When the “simpler approximation model” is properly constructed, it could mimic the behavior of a simulation program quite accurately while being extremely cheaper computationally to run evaluations. Different methods exist for constructing such approximation models. The concentration of this chapter is the use of compact surrogate models, also known as metamodels (Simpson, Toropov, Balabanov, & Viana, 2008), which are data-driven approaches that incorporate either statistical analysis or machine learning methodologies capable of mimicking the behavior of a simulation program.

Reminder from Chapter 1 of this dissertation, in a statistical approach buildings’ historical data is used to run statistical analysis to correlate energy performance with oversimplified variables and predict future performance. In a machine learning approach, which some argue may fall under statistical methodology (Seyedzadeh, Pour Rahimian, Glesk, & Roper, 2018), computer algorithms are trained to learn from data without being explicitly programmed. In the paper “*Machine learning for estimation of building energy consumption and performance: a review*”, authors (2018) provide a sufficient review of the application of different machine learning techniques in forecasting building energy performance such as the use of artificial neural networks, support vector machines, Gaussian-based regression and clustering. The authors conclude that traditional building energy modeling and forecasting using engineering methods are not fast enough to meet the demands of decision-makers and therefore are not as frequently used by professionals as it’s expected. This is while machine learning models have shown great potential for predicting building or city scale energy performance fast and accurate.

Most research on using machine learning for predicting building energy performance have been carried on in engineering as reviewed by Seyedzadeh et al. (2018). However, the field of architecture has also been observing the benefits of machine learning not only for design but also for advancing performance-based decision making in design. In the paper “*Machine Learning for Architectural Design: Practices and Infrastructure*” (2018), Martin Tamke et al. discuss the different potentials that machine learning can bring into architectural practices throughout the design and fabrication process. Currently most research on the application of machine learning in architecture have been focusing on image-based design generation and shape recognition such the work of Hu et al. (2020), Huang and Zheng (2018) and Chaillou (2019). The main drawback of these studies is that the outputs are not usable in architectural practice since the data they work with are images and pixels. For example, the work of Stanislas Chaillou (2019) uses GANs (Generative Adversarial Neural Networks) to produce new images of architectural plans. Although this work is valuable due to its novelty in time of publication, but the results are not useful in architectural practice because of generating pixel-based images of architectural documents rather than vectors. The value and importance that Tamke et al. (2018) bring are suggesting emergent practices in architecture that can benefit from machine learning in creative use

and synthesis that goes beyond current design generation utilization. Five novel and practical streams of applications are suggested by the authors which one of them is using “*machine learning for short-circuiting simulation*”.

As the term suggests, trained machine learning based surrogate models can be used for simulating building energy performance rapidly in a time frame very close to real time. Short-circuiting buildings’ performance simulation - whether it’s energy, structural, mechanical, thermodynamics or other – can help with understanding the behavior of an architecture well before its construction. Since current energy performance simulation tools are based on physical laws, running these simulations are time consuming and computationally intensive and therefore are not often used within the design process and as a tool to drive design decisions. Using trained surrogate models as the backend of such tools for predicting simulation results, in a very short amount of time can aid with integrating simulation into the design process and advancing performance-based design methodologies.

One of the final deliverables of this dissertation is to provide a proof-of-concept software prototype that has trained machine learning models in its backend. Benefiting from machine learning, this software prototype simulates urban scale energy consumption in fraction of a second, given a neighborhood’s three-dimensional design. The next two sections provide an overview of the software architecture and its implementation.

8.2 Preparing Surrogate Models for Estimating Monthly Values of Energy Consumption

Chapter 6 went deep into the technicality of training an artificial neural network on San Diego’s urban form and energy consumption dataset. In Chapter 6, the dataset used had 19 features of urban form measured for 110 zip codes in San Diego county as the predictor variable, and for the response variable the total energy consumed for each zip code for the years 2012 to 2017 were considered. The reason for using the annual value of energy consumption (instead of monthly values) in that dataset was due to the goal of interpreting the extent in which urban form impacts community-wide energy consumption regardless of each month’s specific weather conditions, and to identify the most influential urban form features. In this chapter the goal is to develop a software prototype which uses 3D community design scenarios as input and estimates its monthly values of energy consumption accordingly. This requires having artificial neural networks trained on monthly values of energy consumption and use the resulting predictive models as the backend of the software prototype. The approach used here is to train separate neural networks for each month of the year; the reason behind this logic is to avoid unnecessary complexity by training one neural network over a dataset with 12 months as outputs.

As described previously, finding the best performing artificial neural network architecture is an empirical process. Usually, different architectures are tested and the one which yields the best accuracy is selected ⁴³. Empirically finding the best performing artificial neural network for each month of the year is a time-consuming procedure. A solution for simplifying this process was to code a grid search operation that for each month's dataset the algorithm tests different artificial neural network architectures and selects the ones with highest performance (in here highest performance means lowest mean squared error). The variables instituting the different permutations of the grid search were:

- *Optimizer*: adam, nadam
- *First layer size*: 512, 1024, 2048
- *Number of layers*: 6, 7, 8

Some parameters along with their values were also selected to be used throughout all the neural network variations. For example, the activation function for all neural networks were set to Relu, dropout layers were added with the rate of 0.2 ⁴⁴, number of epochs were set to 300, and patience value ⁴⁵ was set to 100. Note that the choice of the static parameters as well as the different permutable variables were all based on the empirical training experience gained in Chapter 5. With this setting for each of the 12 datasets, $2*3*3 = 18$ different artificial neural network architectures were trained, and the top 5 best performing models were returned. For each month, the top 5 returned models were compared against each other based on the value of their mean squared error, the behavior of their learning curve plot, and the accuracy of each predictive model in predicting the values for the observations of the validation dataset. The human intuition in selecting the best performing artificial neural network [among the top 5] deemed necessary since there's no universal rule on structuring best performing artificial neural networks, also since it's a problem highly dependent on the nature of any dataset and the problem of interest.

The downside of performing multiple trainings in a grid search run one after another, is that the computer's GPU memory does not clear after each run and there's a possibility that the training information of one model may leak to another training procedure. Therefore, another layer of investigation and fine-tuning were performed on the twelve selected artificial neural networks. In this investigation phase, each of the neural network architectures were re-built and re-trained one-by-one on their designated datasets and the resulting learning curves were plotted in Tensorboard. One of the advantages of Tensorboard is that metrics, such as loss and accuracy, could be tracked and visualized for

⁴³ This topic has been vastly discussed in Chapter 5.

⁴⁴ Dropout was used on the first hidden layer as a regularization method to avoid potential overfitting due to the small size of the datasets.

⁴⁵ Patience value is the number of epochs to wait before early stopping if no progress on the validation set was made.

each model. This is specifically useful in case of comparing the performance of different neural networks trained on the same dataset. When a model's learning curve was not performing as expected, a fine-tuning and slight modification was performed on the code and the new model's performance was plotted on Tensorboard ready for comparison and ultimately selecting the best performing one. In this process, some of the selected original neural network architectures were modified and some architectures remained the same. Table 6 shows the variables in the selected model architectures for each month from the grid search procedure, as well as the final variables after fine-tuning the selected architectures. Figures 71 to 82 show each month's learning curve on the left (x-axis: number of epochs, y-axis: epoch loss), and prediction accuracy plot on the right where predictions are made for the validation dataset and the results are compared against the validation true values.

	Variables in the selected model architectures from grid search			Final variables after re-training			Final returned MSE value
	# of layers	1 st layer size	Optimizer	# of layers	1 st layer size	Optimizer	
January	6	512	Adam	6	512	Nadam	0.009350555156575199
February	6	512	Adam	6	512	Adam	
March	8	512	Nadam	8	512	Adam	0.008406761498070537
April	7	512	Adam	7	512	Adam	0.008348045845998105
May	8	512	Nadam	8	512	Nadam	0.011946498677007852
June	7	2048	Nadam	7	512	Adam	0.008610573696633138
July	6	1024	Adam	7	512	Nadam	0.008045464576725394
August	7	512	Nadam	7	512	Adam	0.00773768380128509
September	7	512	Nadam	7	512	Adam	0.00706806445608078
October	6	512	Nadam	6	512	Adam	0.007876764878022083
November	7	2048	Nadam	7	2048	Nadam	0.009284132136650806
December	6	1024	Nadam	7	512	Adam	0.008146409050588875

Table 6 – Original and modified variables in the ANN architecture after retraining process and its resulting MSE.

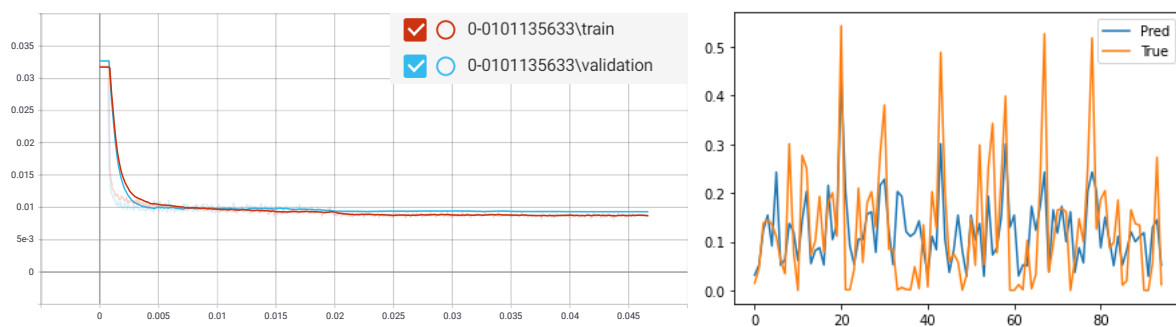


Figure 71 – January's learning curve (left) and prediction accuracy plot (right)

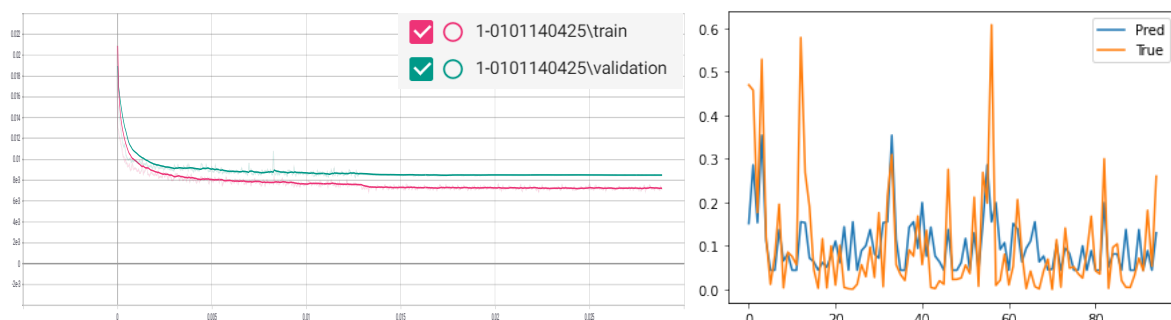


Figure 72 - February's learning curve (left) and prediction accuracy plot (right)

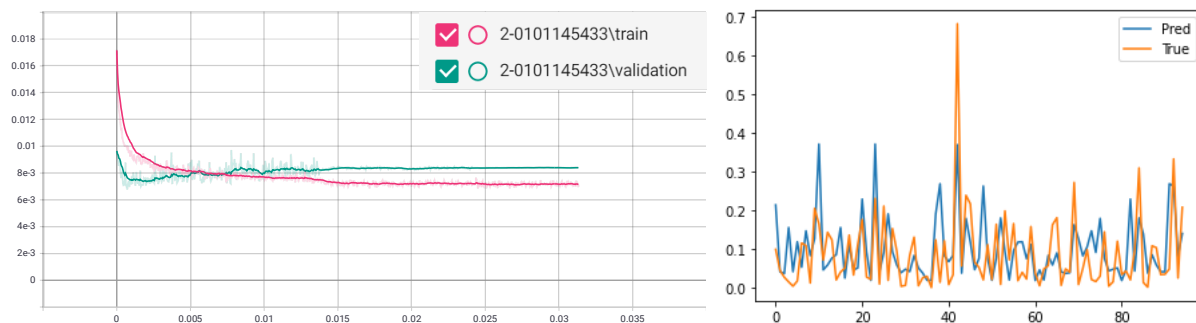


Figure 73 - March's learning curve (left) and prediction accuracy plot (right)

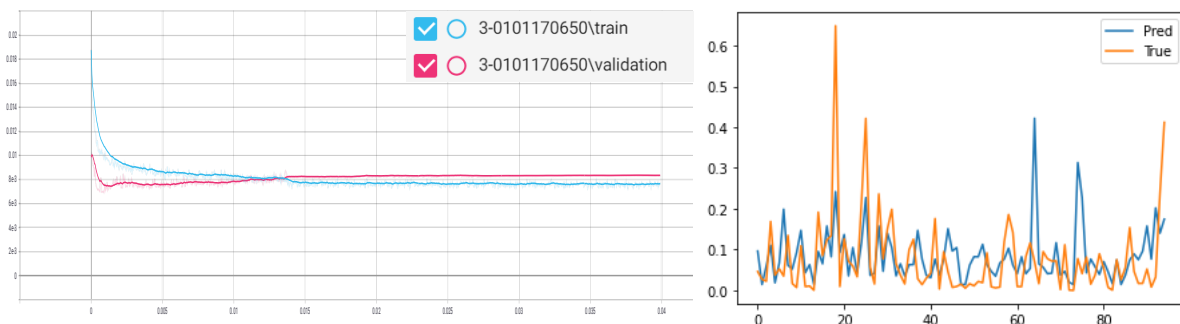


Figure 77 - April's learning curve (left) and prediction accuracy plot (right)

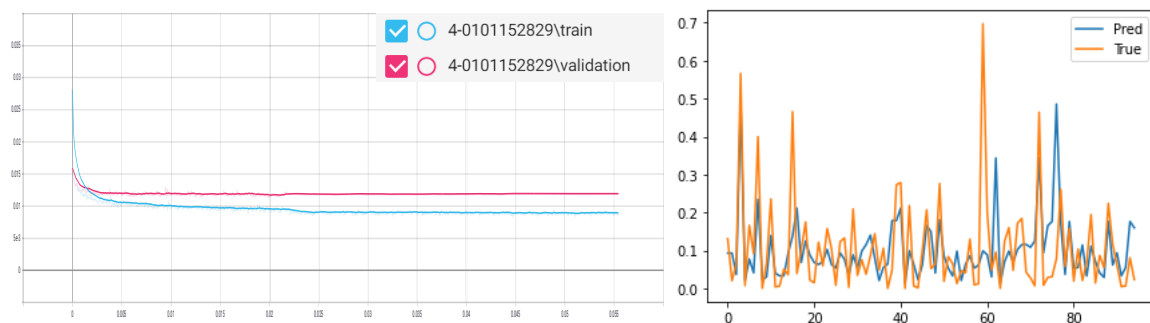


Figure 75 - May's learning curve (left) and prediction accuracy plot (right)

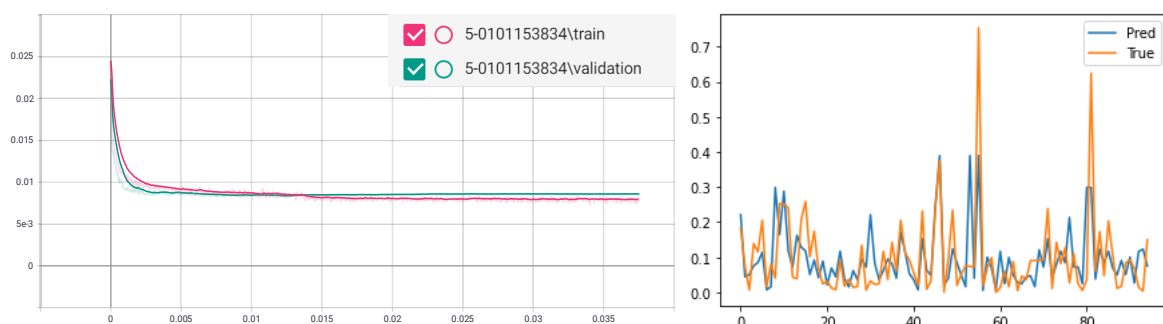


Figure 76 - June's learning curve (left) and prediction accuracy plot (right)

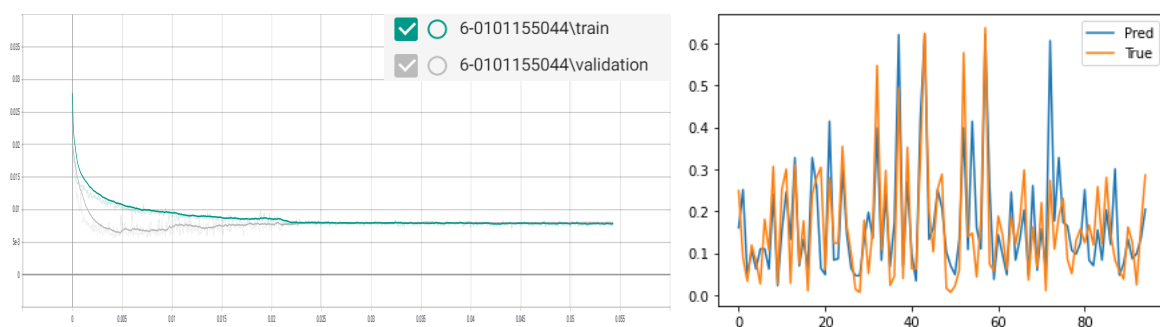


Figure 77 - July's learning curve (left) and prediction accuracy plot (right)

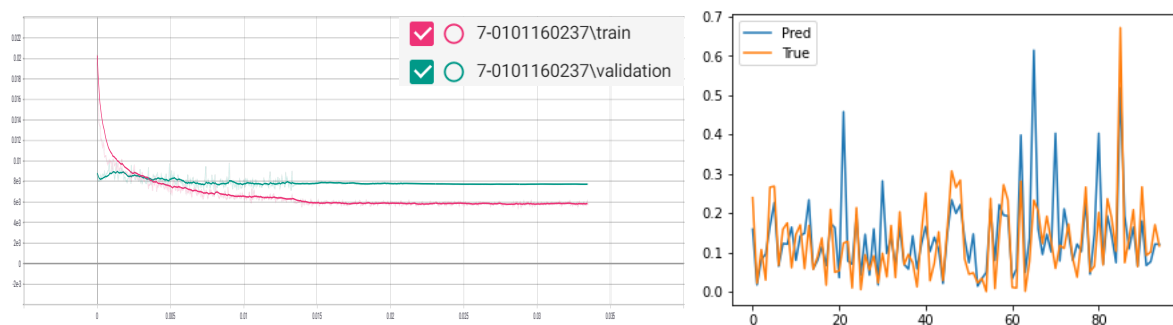


Figure 78 - August's learning curve (left) and prediction accuracy plot (right)

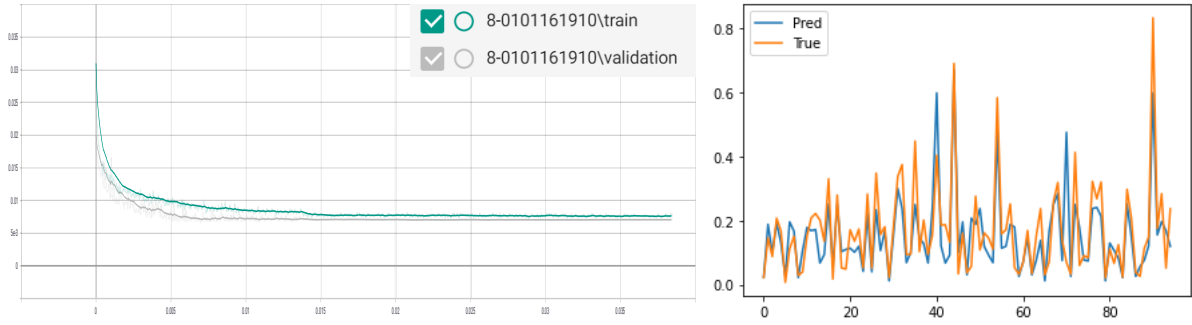


Figure 79 - September's learning curve (left) and prediction accuracy plot (right)

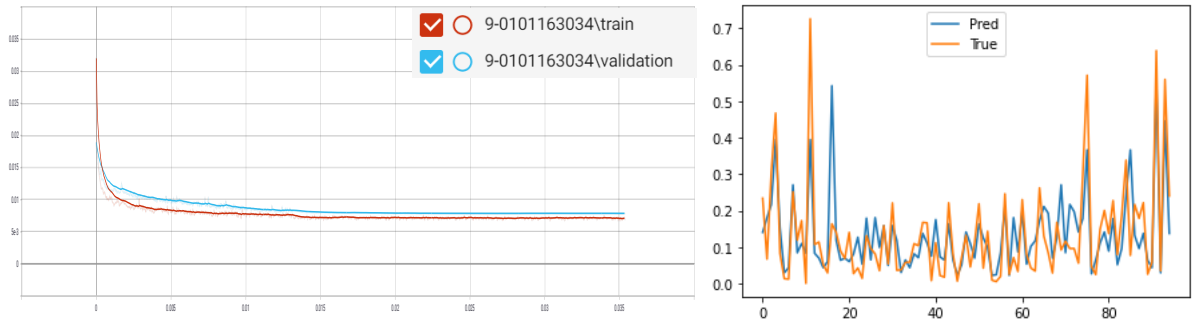


Figure 80 - October's learning curve (left) and prediction accuracy plot (right)

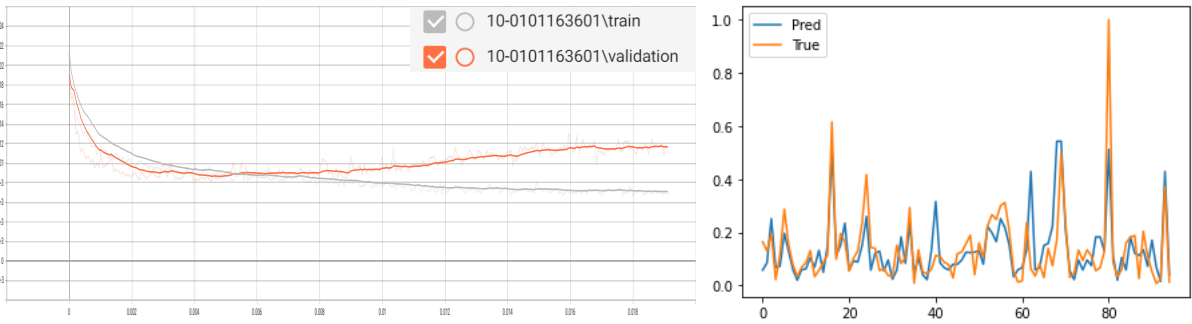


Figure 81 - November's learning curve (left) and prediction accuracy plot (right)

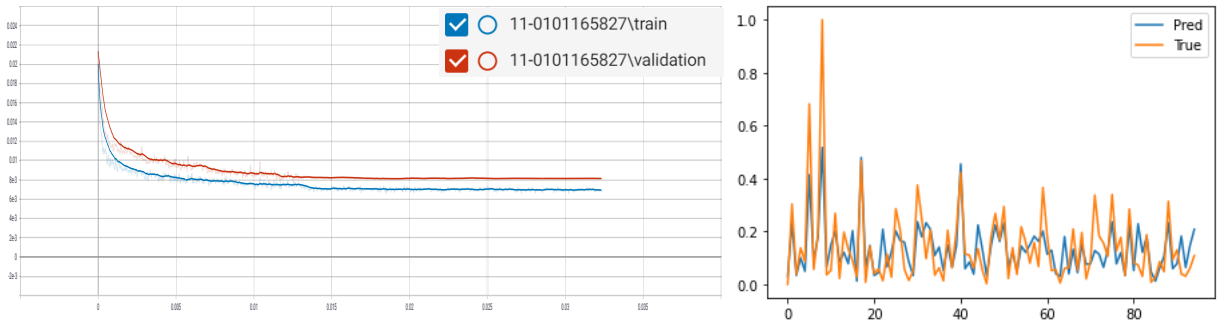


Figure 82 - December's learning curve (left) and prediction accuracy plot (right)

After fitting and evaluating the neural network models on each dataset, the finalized models are run on held-back test datasets - which the models have never seen before - in order to verify the models' performance. The test results of the month of September are shown here as an example. The resulted mean squared error from the testing procedure is reported as 0.046819390610493866. In Figure 83 the orange line shows the true energy consumption values from the test dataset, and the blue line shows the predicted value of energy consumption for each of the observations in the test dataset.

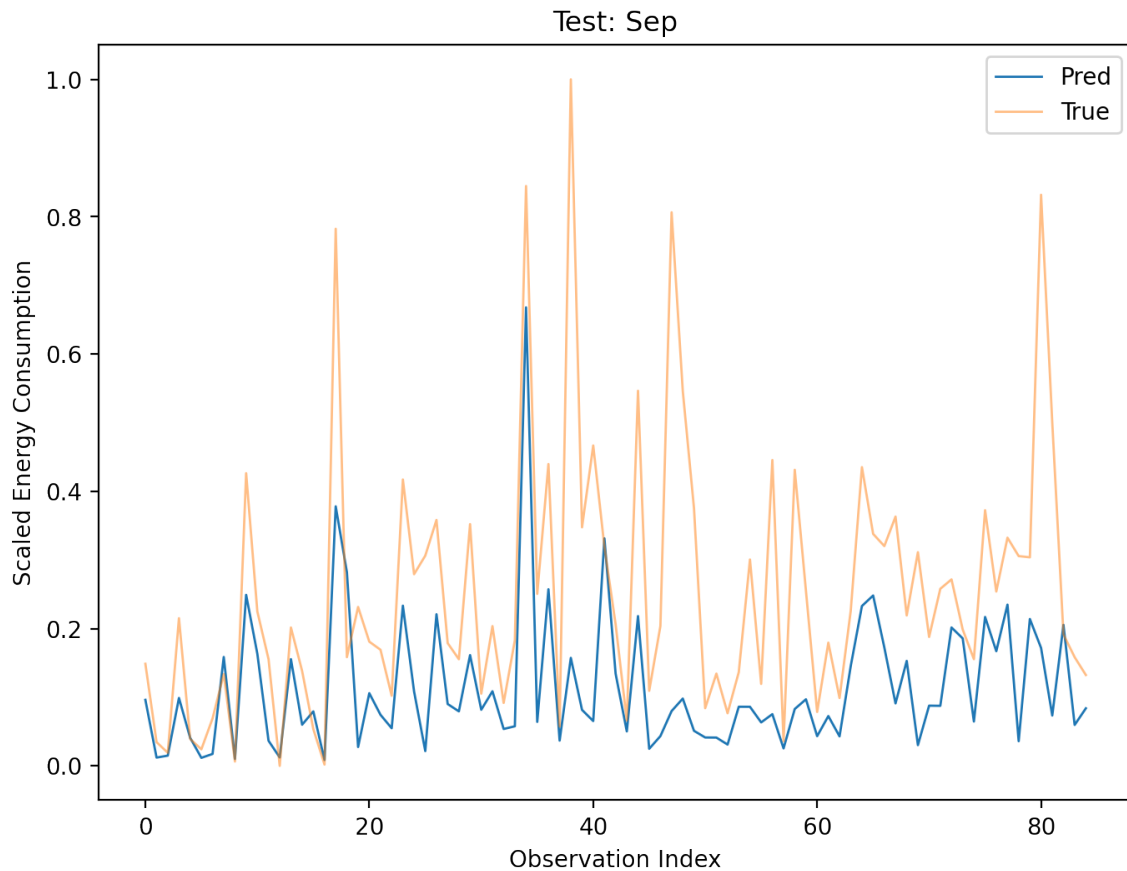


Figure 83 - Plot showing the model's performance on test dataset for the month of September

Comparing the model's performance on the test dataset with the model's performance on the train dataset, a performance mismatch is observed (similar to the test results of Chapter 5); there's a promising performance when evaluating the model on the training dataset and a poor performance when evaluating it on the test dataset. One of the reasons behind model performance mismatch is training on a small and unrepresented dataset. This means that the examples in the training set do not effectively cover the cases observed in the broader domain. Another reason is the stochastic nature of machine learning algorithms resulting from the random initial weights in an artificial neural network, the shuffling of data, and etc. This means that with the same artificial neural network architecture ran on the same

dataset, different sequences of random numbers are used which in turn return models with different performances. This could potentially be problematic in small datasets, as the ones used in this research, where each data point or observation counts towards training a model; there might be hard-to-learn observations which sometimes are in train and sometimes are in the validation or test datasets as a result of shuffling. The remedy is often to enrich the dataset to become larger and more representative, a solution which is not possible in the course of this research. Therefore, while being cognizant of this problem⁴⁶, the final trained models are used in the next section towards developing the predictive software prototype.

8.3 Implementation and Software Architecture

The monthly trained models were each saved in .h5 format and saved locally. All trained models are capable of predicting values of energy consumption [for that specific month] for any unseen and new values of urban form as long as the unseen data follows the generalization principle. The generalization principle indicates that trained machine learning models can provide valid predictions for unseen and new data as long as they are drawn from the same distribution as the original dataset that was used for training. Honoring this principle and the fact that our trained models are limited to San Diego, it's important to have in mind that the trained models are unable to estimate valid values of energy consumption for 'any' designed urban form scenario; the unseen values of urban form that will be given to the trained models for prediction purposes should fall in the same range of urban form values of the training dataset, representing hypothetical communities as if they were constructed in San Diego.

Since the trained models of this research are used as the backend of the energy simulation software, it's crucial assuring the validity⁴⁷ of community design scenarios that will be inputted to the software. This assertion is made by adding the real-time energy simulation software prototype to the urban generator algorithm. This is because the urban generator algorithm was developed based on the local zoning standards and urban planning principles of San Diego as described in Chapter 7. By this, any community designs generated by this algorithm will be similar to the existing urban fabric of San Diego, therefor their derived measurements of urban form follow the same distribution as the original dataset.

Based on the aforementioned reasonings, the proposed real-time energy simulation software is composed of three main parts:

- *The urban generator in Grasshopper*
- *The Python communicator*

⁴⁶ To enrich the dataset in addition to making it larger, another solution (offered as next step in Chapter 9) is to add more features such as weather data to the dataset.

⁴⁷ A valid community design in this context is a design that follows the generalization principle as indicated in this research.

- *The backend server*
- *The visualization component*

The urban generator scripted in Grasshopper and Python has been extensively described in Chapter 7. Following the prescribed urban form measurements (refer to Table 3 in Chapter 5), a number of equations and formulas have been added to the generator which takes in any generated community design scenario and computes the 14 most impactful indicators of urban form⁴⁸ and outputs as a data tree. These 14 indicators of urban form are: floor space index, open space ratio, network density, volumetric compactness, size factor, form factor, mixed use index, community building orientation, street orientation, urban horizon angle, obstruction sky view, sky view factor, passivity ratio, and plan depth. The important point to note here is that these 14 indicators need to be sorted and fed into the trained models in the declared order reflecting the order initially used for training the artificial neural networks (in Chapter 5). A Python communicator is scripted in Grasshopper which takes the generated 14 values of urban form as a data tree, converts it into a list, and sends it to the server⁴⁹. The server is a virtual computer scripted in Python language which uses



Figure 84 - A picture screen of the output of the software. The urban setting that was created by the tool is shown on the right, its values of urban form is shown on top left, and the predicted values of energy consumption is shown on bottom left.

⁴⁸ Initially 19 indicators of urban form were hypothesized to have an impact on community scale energy consumption, however in Chapter 5 - through an extensive feature selection methodology - 5 least impactful indicators were omitted leaving the study with 14 indicators of urban form which were proven to impact community scale energy consumption. The selected 14 indicators of urban form were then used for training the artificial neural networks. Therefore for any prediction purposes the input data needs to have 14 values reflecting those of urban form.

⁴⁹ A server is a computer designed to process requests and deliver data to another computer over the internet or a local network

websockets and Tensorflow (Keras) to load all the 12 trained machine learning models from the database which they are stored in. By loading the models, the server is capable of taking any set of 14 numbers and input them to the trained models as urban form values and use that to predict and output 12 monthly values of energy consumption. The urban form values are inputted to the server via the Python communicator. By receiving these values, the server uses the loaded trained models to *instantly* estimate its relevant 12 monthly values of energy consumption and to send the outputs back to the Python communicator through a local network. A visualization script is added to the Python communicator which takes the received predicted values from the Python communicator and visualizes them as a bar chart on Rhino viewport. The urban generator takes 1-3 minutes to generate a community design scenario but when a design is generated it takes only a fraction of a second to simulate and visualize the design's monthly value of energy consumption. A screenshot of the output of the software prototype is shown in Figure 84.

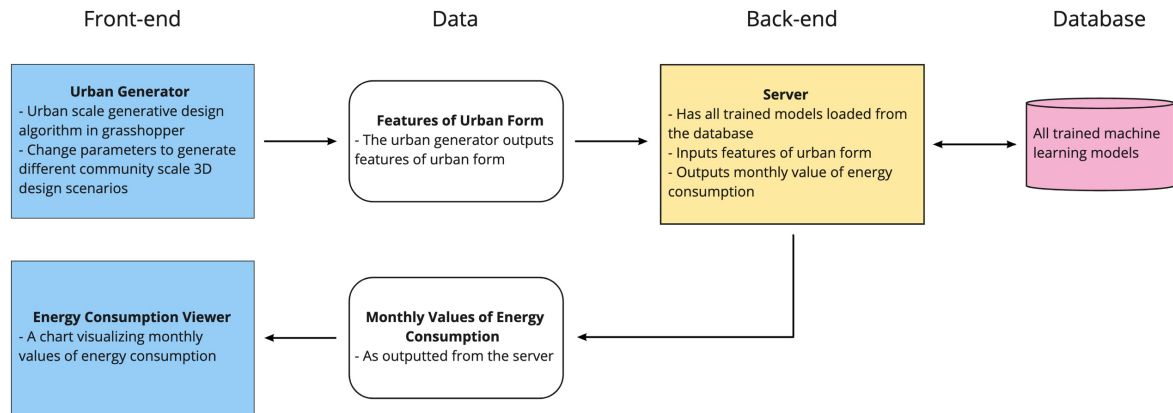


Figure 85 - The prototype's software architecture

A diagram is provided in Figure 85 which shows the prototypes architecture for clarification. To summarize the diagram:

1. Database: is a folder on computer storing all trained machine learning models as in .h5 format.
2. Back-end is a server responsible for calling and loading the trained machine learning models from the database, receiving and computing requests and delivering data to other computing programs.
3. Front-end: is the interface the user works with. In this prototype the front-end is the urban generator in Grasshopper where the user can change certain parameters and have the algorithm generating various different 3D community scale urban scenarios based on San Diego's principles of urban planning. The front-end is also used to visualize the energy simulation results.

4. Data: are the values being inputted to the server (14 values of urban form) and the data outputted from the server (12 monthly values of energy consumption). When the user selects its custom community design through the front end, the generator algorithms automatically extract 14 features of urban form and through a python code sends that to the server. The server immediately predicts the monthly values of energy consumption as outputted by the trained models and sends it back to Grasshopper. The front-end instantly visualizes the predicted energy values as a bar chart in the Rhino environment.

Chapter conclusion: The content of this chapter describes the technicality of implementing the trained models as the backend of a community scale energy simulation software prototype that is capable of running simulations in real time. Besides the technical aspect, this chapter sets a good example of how surrogate models can make predictions of energy consumption in a much computationally cheaper way than usual engineering-based simulations.

Chapter 9 Conclusion

Introduction: This final chapter reflects back on the intentions and goals of this research, provides a summary of the concluded results and discusses them in the context of the bigger picture. By this, the original research questions are responded to, contributions and implications to the field are discussed, and ideas for future directions and next steps are propounded⁵⁰.

9.1 Summary of Research

This dissertation started with the motivation of adding a spatial dimension to the technical procedure of developing high energy performance community microgrids since the exclusive nature of current processes deemed insufficient. Chapter 1 described a high energy performance community microgrid as a microgrid that ensures extended durations of energy self-sufficiency while in island mode. By this, improving energy performance in community microgrids involves increasing the rates of local energy production and reducing the energy demanded for community buildings' operation. We also learned that current research and practice on improving energy performance in community microgrids have been solely focused on enhancing the technological aspect of their development with the goal of solving the limited supplies of local energy and increasing rates of consumption. We understood that such mere technical approach is insufficient to solve the energy problems associated with a community microgrid's performance especially since they are contextualized in cities and urban areas. The hypothesis supporting these claims suggests that a community microgrid's energy performance goes beyond the technical assessment of its electrical infrastructure, proposing an extant relationship between the urban form of a community and how well the microgrid performs in terms of the local supply and demand of energy. This hypothesis was backed up in Chapter 4 by a literature review verifying that different spatial configurations of urban form changes the local wind pattern and drives urban heat island effect. These changes influence the buildings' thermal comfort which eventually leads to fluctuations in patterns of energy consumed for building space heating and cooling, as well as the feasibility of adopting on-site renewable energy generators such as photovoltaic (PV) panels and wind turbines.

For this research, solar powered community microgrids (also known as solar community microgrids) were selected as the main focus. Accordingly, the purpose and hypothesis of this research could be rephrased and specified as following: to spatially understand the variations of energy performance in solar community microgrids, there's a need to explore how the urban form of a microgrid-connected community impacts the amount of solar energy captured onsite and the amount of energy that is required for building's operation

⁵⁰ An overview of this dissertation is currently in process as a book chapter for submission to "*Artificial Intelligence in Urban Planning and Design*" book for Elsevier.

especially for space heating and cooling. Supporting this claim the three following research questions were proposed:

1. The spatial geometry of urban form is a complex entity and is defined by many spatial attributes. Which of these attributes have an energy-relevance regarding the community-wide supply and demand of energy and how are they measured?
2. What is the most appropriate model for describing the desired relational pattern between urban form and energy performance of microgrid-connected communities?
3. Which combination of the attributes of urban form impact a solar community microgrid's energy performance?

Answering these research questions started with reviewing past literature with the purpose of unpacking those features and variables of urban form that have an energy relevance in the context of community microgrids. In Chapter 4 these spatial attributes of urban form along with their index and metric of measurement were identified and disclosed, giving a response to the first part of the first research question - *which of these attributes [of urban form] have an energy-relevance regarding the community-wide supply and demand of energy and how are they measured?* The selected nineteen indices ranged from that which previous research proved significant for community scale energy consumption and PV energy production and those that were not deemed influential. The reason for selecting all energy-relevant spatial attributes was to understand their interaction and influence when they are in confluence as opposed to studying each attribute in isolation. These indices include floor space index, ground space index, open space ratio, layer, network density, aspect ratio, volumetric compactness, size factor, form factor, mixed use index, green space density, green area geometry, community building orientation, street orientation, urban horizon angle, obstruction sky view, sky view factor, passivity ratio, plan depth.

Studying the combined effect of different spatial attributes on community scale energy performance alludes to a level of complexity which cannot be solved by using conventional engineering or simple statistical methodologies; based on literature, data mining specifically machine learning, was chosen as the main approach for unraveling the implied relationship. By choosing artificial neural networks and then Shapley regressions for the statistical inference of the trained model, an answer for the second research question was given: *“What is the most appropriate model for describing the desired relational pattern between urban form and energy performance of microgrid-connected communities?”*

The main requirement when working with machine learning algorithms are the availability of large datasets. San Diego county was selected as case study in this research due to several reasons outlined in Chapter 1, the main one being the availability of energy consumption data for the entire county. Despite the accessible energy consumption data, information on county's PV energy production remained inaccessible. Due to this

shortcoming, the adopted methodology was divided into two distinct but related parts and included analysis on 1. the impact of urban form on energy consumption using real-world datasets and, 2. the impact of urban form on PV energy production using synthesized datasets. With this arrangement, each methodological part would have had its own separate final dataset with measurements of urban form as the predictor variable and monthly values of energy performance (consumption or production) as the response variable prepared for machine learning training procedures; first part based on real-world and second part based on synthetic information.

In Chapter 5 the process of preparing and analyzing the real-world dataset was discussed. The final dataset was prepared by measuring the nineteen values of urban form for 110 zip codes in San Diego county (predictor variable) and aggregating monthly values of energy consumption of those zip codes for seven consecutive years (response variable). Studying the impact of all nineteen indices of urban form on energy consumption is multidimensional and complex and cannot be performed using conventional engineering or statistical methods. In this regard, datamining was used as the main methodology to solve the problem. Working with datamining techniques requires large amount of data about the problem of interest so different algorithms could be trained on the dataset towards mining different patterns from it. After processing and cleaning the dataset an artificial neural network was trained on it to identify which combination of urban form features contribute the most to community scale energy consumption.

In Chapter 6 the trained model of Chapter 5 was then statistically interpreted using Shapley regressions and the magnitude of impact of the different predictor variables and their correlational pattern with the response variable were discovered. In this Chapter a response to the third research question - *“which combination of the attributes of urban form impact a solar community microgrid’s energy performance?”* - was offered where fourteen indicators of urban form were proven to have the most influence on community scale energy consumption. These indicators include: Size Factor, Mixed Use Index, Passivity Ratio, Sky View Factor, Plan Depth, Form Factor, Open Space Ratio, Network Density, Community Building Orientations, Urban Horizon Angle, Volumetric Compactness, Obstruction Sky View, Floor Space Index, and Street Orientation. The results of the statistical inference were then elaborately discussed from an architectural point of view. The architectural interpretation helps creating a novel and comprehensive knowledge of the tradeoffs between the impact that every single design decision has on the net energy demand of community buildings in a microgrid [in the context of San Diego].

In Chapter 7 the process for creating the synthetic dataset was described. The goal in this stage was to develop a *generation-evaluation* urban design tool that creates numerous different 3D community design scenarios, measures their nineteen selected features of urban form, evaluates each generated community’s capacity of producing PV energy, and

finally for each design option aggregates the measured values of urban form and monthly values of PV energy production into a .csv file, ready for processing and training purposes. The *generation* part of this tool – responsible for generating different random 3D community design scenarios - was developed in Grasshopper and Python by translating San Diego’s urban shape grammar into code. Details on the urban grammar describing the rules and patterns forming different urban typologies in San Diego have been thoroughly discussed in Chapter 7. However, the *evaluation* part of the tool - responsible for evaluating the PV energy production capacity of each generated community – was not developed due to the technical limitations of existing PV energy evaluation plugins used in Grasshopper. The occurred technical difficulty consequently resulted in impeding the creation of the synthetic dataset as originally planned, but it was an affirmation of the limitations of existing simulation tools for urban scale energy analysis and an emphasis on requiring energy simulation tools capable of simulating urban scale energy performances.

As mentioned, in Chapter 6 a new knowledge was possessed on spatially designing energy conscious communities and ultimately high-performance community microgrids [in the context of San Diego]. A spatially aware design and assessment of solar community microgrids as described in this research brings the need for a new generation of computational modeling, simulation, and evaluation tools for the field. In Chapter 3 a literature review was presented in which a gap was identified in the existing software tools that simultaneously address the necessary interaction between the superstructure and infrastructure⁵¹ of community microgrids. Given the proven importance of this interaction as seen in Chapter 6, in Chapter 8 an experimental software prototype was delivered that bridges the identified gap by predicting the energy consumption of any given solar community microgrid design scenario by the virtue of its urban spatial configuration in *real-time*. The developed tool predicts monthly energy consumption values for any given community design scenario in *real-time*; this has been possible by using surrogate models – the predictive trained models as described in Chapter 8 - as the backend of the prototype instead of utilizing a hard-coded engineering-based backend.

Creating the software prototype provides a preliminary example of a tool used by architects and urban planners for designing *energy conscious* microgrid-connected communities well before their construction. Initially the goal was to create a tool that aided in designing *energy self-sufficient* community microgrids, but since we ended up with a tool predicting only *energy consumption*, the tool guides the designer towards creating *energy conscious* community microgrids. With this tool and the knowledge developed for designing high performance community microgrids leveraging its urban form given its significant impact, a proof of concept is provided for engaging the design sector in the technical conversation of developing community microgrids.

⁵¹ Superstructure and infrastructure in a community microgrid has been defined in Chapter 1.

9.2 Research Discussion, Contributions, and Limitations

Johnathan Barnett, in “*Redesigning the Metropolis the Case for a New Approach*” (1989), points out to the dynamic evolution of cities and urban areas and emphasizes the necessity of updating urban design and planning techniques as cities face constant changes in their environmental, developmental, and political settings. Taking the environmental aspect of built environments into consideration along with the fact that the climatic conditions around us are changing in a rapid speed (Andreson & Bausch, 2006), researchers suggest that reaching energy self-sufficient built environments requires moving past building-scale analysis and adopting up-to-date and innovative energy-related measurements for urban design and planning (Gossop, 2011; Davila, Reinhart, & Bemis, 2016).

Addressing energy issues at urban scale brings more complexity than those of building scale mainly because of having a large number of stakeholders involved in urban-scale projects. With this comes extended and obscurant power relations making urban issues ill-defined and multi-faceted ones specially when it comes to energy-related issues and its inherently political nature. In an era where the causes and effects of climate change have been a topic of dispute among politicians and scientists, the goal of reaching low carbon and energy self-sufficient communities and cities has become more urgent than before⁵². Both the research and planning communities have come to the conclusion that a new understanding of how urban planning impacts energy dynamics in cities and communities is required (Cajot, Peter, Bahu, Guignet, & Koch, 2017). Reaching a low carbon and energy self-sufficient community entails reducing fossil fuel consumption and combating greenhouse gas emissions by taking actions in pursuit of building resilient communities and cities which are less pollutant and less energy demanding. A main action item for reaching this goal is the development of community microgrids which support the local supply and demand of clean energy (Amin & Wollenberg, 2005; Farhangi, 2010). To develop low carbon and energy self-sufficient community microgrids, we grounded this research based on past studies and spatially assessed the role that urban planning plays in the energy performance of these power-grid-independent territories.

What motivated this study was the practical disengagement of the design sector in the development process of community microgrids while the importance of this involvement was marginally highlighted among research and academic communities. One reason behind this disengagement is the illiteracy among architects and urban planners on how urban form and urban geometry impacts the energy performance of cities at large. Normally in practice, the energy implications of individual buildings are considered, however those of urban scale are easily neglected. This could be because of the spatial complexity associated with urban form and the fact that studies to date were not able to

⁵² EU COM 112/2011

capture the multidimensional influence of urban form on community scale and urban scale energy consumption and PV energy production. Measuring and understanding the energy implication of an urban area, unlike common knowledge, goes beyond the sum of energy implications of individual buildings in a certain area, hence the associated complexity with urban scale energy studies. As described in Chapter 1, urban form and geometry has a major impact on changing the local wind patterns and trapping heat in urban areas and therefore creating micro-climates. The existence of several urban micro-climates in a city in addition to the energy performance of individual buildings, all in all need to be studied when considering the energy implications of urban areas. Adding the feasibility study of accessing renewable energy to this relational pattern, brings the understanding of how urban form impacts energy performance in community microgrids to another level of complexity which has not previously been offered to the building and urban design communities.

The second reason behind the above-mentioned disengagement is the lack of urban scale energy modeling and simulation tools that capture the presented complexity. As explained in the software survey of Chapter 3, the quantity outputted from existing urban scale energy simulation tools are the summed amount of each building's energy performance in an area, which is not an accurate reflection of the real-world energy performance of cities. Moreover, the hardcoded backends of these tools make running urban scale energy simulations a tedious task and of expensive computation, practically impossible to use for real-world projects.

This research attempted to solve this disengagement by studying the multidimensional impact of urban form, as a complex phenomenon made up of a set of intertwined variables, on the energy performance of communities and cities at large. The result made it evident that urban context impacts the implementation of efficient and effective high energy performance energy systems. More importantly, it not only proves that architects and urban planners are in a key position to coordinate and manage the development of these energy systems, specifically community microgrids, but also provides them with necessary tools and design guidelines. Current practices in improving energy efficiency and energy self-sufficiency in the built environment are mostly focused on a single building scale as self-defined entities without considering the concatenation of events that happen at the urban scale. This research addresses energy efficiencies in the built environment from a morphological point of view where all energy related spatial parameters related to urban form were taken into account - an undertaking which is novel to the field. Therefore, the results of this study have captured the energy implication of all aspects related to urban form including the energy performance of individual buildings and the urban spaces among them.

The results of this study underline how urban form can be designed to reduce the energy demanded by a community or a neighborhood as a whole, but also facilitate access to renewable energy most specifically benefit from onsite PV energy production. Planning the urban form in a way that facilitates the development of high energy performance community microgrids can significantly contribute to shaping energy self-sufficient communities and neighborhoods, and at a larger scale, energy-efficient and low-carbon city. This research has not only configured a design framework for designing energy self-sufficient community microgrids but also developed a tool for architects and urban designers that helps them reach the required energy targets for their designs within a much lesser time compared to any similar and commercially existing tool.

By setting the path to take ownership of designing energy self-sufficient community microgrids, the results of this study enable architects and urban planners to carefully and profoundly address the pressing issues related to the local supply and demand of clean energy within their profession. This means that planners could be more and more involved in the technical conversation of developing community microgrids by not only handling the aesthetics and quality of life in urban communities but also to get quantitatively concerned with energy system design and engineering. In this regard, this research has reached its main goal of adding a spatial dimension to the development of community microgrids where architects and urban planners get more involved in the development process of these local energy systems.

While conducting this research, a valuable lesson was drawn from Chapter 7 and that is the need and urgency of having urban scale energy simulation programs that do not rely on complex mathematical models and physics-based computations. Because current PV energy simulation tools on Grasshopper environment were unable to run community scale PV energy simulations, this research was unable to create the synthetic dataset. This inability alone proves the deficiency associated with current urban scale energy simulation tools. In order to have architects and urban planners to design energy self-sufficient communities, the first step is to provide tools that enable them to do so. This research provides a prototype of a software that runs urban scale energy simulations in real-time which can be used for community designs in the context of San Diego. This prototype does not provide exact numbers of energy consumption in any designed community but rather offers a ‘predictive’ number. An imprecise simulation of energy performance provided in real-time while designing an urban area is much more practical and useful for an architect and urban designer than a software that takes hours to run a simulation but provides a more accurate number. With the real-time tool the designer would be able to change their designs and see the simulation results instantaneously while with the latter the designs are rather not to be changes to hit the energy targets. Moreover, for urban scale energy planning and design an approximate number is sufficient to guide the designs towards more energy

aware ones. This prototype and the process of developing it can shed light for future researcher to develop a similar tool that can run urban scale energy simulations for any climatic region.

In addition to the general contributions described above, the shape grammars presented in Chapter 7 has also added knowledge to the field by capturing urban planning rules that take topographic constraints into consideration. The presented shape grammars although has been extracted from San Diego but have been developed in an abstract way such that the underlying design procedure is not strictly specific to San Diego and could be used for recurrent design purposes fitting many urban design scenarios in American contexts. An urban grammar capturing American urban planning principles along which include the impact of topography, is a discourse novel to the field of shape grammars. The developed shape grammars were then used to develop a generative urban design tool which has extended CItMaker and has pushed forward existing research.

9.3 Next Steps and Future Directions

This doctorate dissertation has paved the way for the following next steps to extend this research in multiple different ways:

- *PV energy production:* as mentioned earlier and in Chapter 7, this study was supposed to study the impact of urban form on PV energy production in addition to energy consumption. However, due to the limitations of existing tools this attempt could not be completed. An immediate next step for this research is to study how different configurations of urban form can impact access to solar energy, either with real-world data if available or synthetic datasets as proposed earlier.
- *Expanding the dataset:* an appropriate next step is to enrich the dataset and expand it in ways that includes measurements of the climatic and environmental condition of the case study such as humidity, radiation, perception, temperature, geographic coordinates, and etc. This inclusion could possibly lead to discoveries around understanding the correlation between urban form, energy performance, and climate which could consequently lead to more generalizable results.
- *Comparative study:* to conduct the same methodology to couple of different cities in different climate zones and compare their most energy effective variables urban form. By architecturally interpreting the results, one could assess how the spatial design of a community microgrid could be dissimilar in different parts of the world and that each region has an optimal configuration of urban form leading to its best performing community microgrids.
- *Energy implications of the built environment as a whole:* when designing climatically responsive and energy conscious cities, sustainability measures need to be implemented at the building and transport level. While this research looks at

energy implications at an urban scale, a valuable next step is to add energy implications of transport and those of building scale architecture and retrofitting to the study. This way a comprehensive understanding of energy performance of cities could be offered. Suitable urban forms can positively affect energy demand in buildings and access to renewable energy next to other factors such as occupant behavior and building design and systems efficiency, moderating traffic, enhancing pedestrian comfort and outdoor activities. Such comprehensive approach results in optimal urban configurations for different climate zones that decrease CO₂ emissions and therefore could be used as important mitigation measures for global warming.

- *Effect of climate change:* this study was undertaken with the assumption that the weather in San Diego wouldn't change due to global warming in the future. This is not a true assumption and studies show that by 2080's climate change would extensively impact the weather in major cities including San Diego (Wang & Chen, 2014). Unlike this research, future measures and principles of urban planning need to be considerate of the ever-changing and sometimes uncertain changes of weather and climatic conditions in urban areas.

Glossary

Urban form: the collection of urban attributes that make up built areas.

Urban attribute: spatial attributes of an urban built environment with energy relevance i.e., density, layout, orientation, land use mix, geometric pattern, and size.

[Urban] Attribute index: each urban attribute can be identified by different indices. For example, density (as an attribute) can be identified by various indices including floor space index, ground space index, open space ratio, layer, and network density.

[Urban] Index metric: the formula for measuring each urban form index is its metric. For example, gross floor area / total ground area is the metric for measuring floor space index.

Urban spatial configuration: the set of spatial relations holding between different attributes of an urban built environment.

Microgrid: group of interconnected loads and distributed energy resources within clearly defined electrical boundaries that acts as a single controllable entity with respect to the grid. A microgrid can connect and disconnect from the grid to enable it to operate in both grid-connected or island-mode

Community microgrid: a microgrid that serves as local energy infrastructure for communities in urban areas.

Microgrid infrastructure: the distributed energy resources and loads that are within microgrid's defined electrical and regional boundaries. A microgrid's infrastructure forms its electrical boundaries.

Microgrid superstructure: the buildings that are within microgrid's electrical and regional boundaries are tied together via microgrid's infrastructure. A microgrid's superstructure forms its regional boundaries.

Microgrid boundaries: refers to both regional and electrical boundaries.

Photovoltaic or PV energy: photovoltaic energy is the conversion of sunlight into electricity.

Solar community microgrid: a community microgrid that has PV panels as its onsite energy generator.

Energy performance: combination of energy input and energy output of an energy system. In this research, the energy performance of a solar community microgrids means the net

PV energy generated on site and the net energy consumed by the buildings at the community scale.

Machine learning algorithm: the algorithm that is constructed and used to discover the relational pattern and hidden structures in the input dataset.

Machine learning model: the outputted mathematical model that explains the discovered relational pattern in the input dataset and can be further used to make data-driven predictions and decisions on unseen examples.

Training or learning process: the mathematical process that a machine learning algorithm conducts on the inputted data that explains its embodied relational pattern and hidden structures.

References

- Allegrini, J., Orehounig, K., Mavromatidis, G., Ruesch, F., Dorer, V., & Evins, R. (2015). A review of modelling approaches and tools for the simulation of district-scale energy systems. *Renewable and Sustainable Energy Reviews*, 1391-1404.
- Amado, M., & Poggi, F. (2014). Solar urban planning: a parametric approach Miguel. *Energy Procedia*, 1539-1548.
- Amin, M., & Wollenberg, B. (2005). Toward a Smart grid. *IEEE Power and Energy*, 34-41.
- Andreson, J., & Bausch, C. (2006). Climate change and natural disasters: Scientific evidence of a possible relation between recent natural disasters and climate change. *Policy Department Economic and Scientific Policy*, 2.
- Anzanello, M. J., & Fogliatto, F. S. (2011). Learning curve models and applications: Literature review and research directions. *International Journal of Industrial Ergonomics*, 573-583.
- Ariyasinghe, M. N., & Hemapala, K. (2013). Microgrid Test-Beds and Its Control Strategies. *Smart Grid and Renewable Energy*, 11-17.
- Štrumbelj, E., & Kononenko, I. (2014). Štrumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and information systems*, 41(3), 647-665. *Knowledge and information systems*, 41(3), 647-665.
- Baker, N., Hoch, D., & Steemers, K. (1992). The LT Method, Version 1.2: An Energy Design Tool for Non Domestic Buildings. Commission of the European Communities. *Directorate-General for Science, Research and Development and Directorate-General for Telecommunications, Information Industries and Innovation*.
- Barnett, J. (1989). Redesigning the Metropolis - The Case for a New Approach. *Journal of the American Planning Association*, 131-135.
- Barros, P., Beirão, J., & Duarte, J. P. (2013). The language of Mozambican slums. *eCAADe* (pp. 715-724). Delft, The Netherlands: Proceedings of the 31st International Conference on Education and research in Computer Aided Architectural Design in Europe.
- Bastiononi, S., Pulselli, F., & Tiezzi, E. (2004). The problem of assigning responsibility for greenhouse gas emissions. *Ecological Economics*, 253-257.
- Beirão, J. N. (2012). *CltyMaker: Designing Grammars for Urban Design*. TU Delft : PhD Dissertation.
- Beirão, J. N., Duarte, J. P., Montenegro, N., & Gil, J. (2009). Monitoring urban design through generative design support tools: a generative grammar for Praia. *Proceedings of the APDR Congress*.

- Beirão, J., & Duarte, J. P. (2018). Generic grammars for design domains. *AI EDAM*, 225-239.
- Beirão, J., Duarte, J. P., & Stouffs, R. (2009). An Urban Grammar for Praia: Towards Generic Shape Grammars for Urban Design. *eCAADe* (pp. 575-584). Proceedings of the 27th eCAADe conference.
- Benros, D., Duarte, J. P., & Hanna, S. (2015). The inference of generic housing rules: a methodology to explain and recreate Palladian Villas, Prairie Houses and Malagueira Houses. *Design Computing and Cognition'14* (pp. 401-419). Springer, Cham.
- Bettencourt, L., Lobo, J., Helbing, D., Kühnert, C., & West, G. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the national academy of sciences*, 104(17), 7301-7306.
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford University Press.
- Bourdic, L., Salat, S., & Nowacki, C. (2012). Assessing cities: A new system of cross-scale spatial indicators. *Building Research & Information* ISSN:, 592-605.
- Bourgeois, T., Gerow, J., Litz, F., & Martin, N. (2015). *Community Microgrids: Smarter, Cleaner, Greener*. New York: Pace Energy and Climate Center.
- Bradford, A., Stankiewicz, T., Sundby, J., Fanshaw, B., & Sargent, R. (2019). *Shining Cities 2019 - The top U.S. Cities for Solar Energy*. Environment America Research & Policy Center.
- Cajot, S., Peter, M., Bahu, J., Guignet, F., & Koch, A. (2017). Obstacles in energy planning at the urban scale. *Sustainable cities and society*, 223-236.
- Chaillou, S. (2019). *AI + Architecture - Towards a New Approach*. Master's Thesis. Graduate School of Design, Harvard University.
- Chatzidimitriou, A., & Yannas, S. (2015). Microclimate development in open urban spaces: The influence of form and materials. *Energy and Buildings*, 156-174.
- Chaurey, A., & Kandpal, T. C. (2010). A techno-economic comparison of rural electrification based on solar home systems and PV microgrids. *Energy Policy*, 38, 3118-3129.
- Chen, L. D., Sakaguchi, T., & Frolick, M. (2000). Data Mining Methods, Applications, and Tools. *Information systems management*, 67-68.
- Ciaramella, A., Staiano, A., Cervone, G., & Alessandrini, S. (2015). A Bayesian-Based Neural Network Model for Solar Photovoltaic Power Forecasting. *International Workshop on Neural Networks* (pp. 169-177). Springer, Cham.
- Commend – community for energy, environment and development. (2017, February). Retrieved from [http:// www.energycommunity.org/](http://www.energycommunity.org/)
- Commission, E. (2011). *A Roadmap for moving to a competitive low carbon economy in 2050*. Brussel: European Commission .
- Compagnon, R. (2004). Solar and daylight availability in the urban. *Energy and Buildings*, 36, 321-328.

- Connolly, D., Lund, H., Mathiesen, B., & Leahy, M. (2010). A review of computer tools for analysing the integration of renewable energy into various energy systems. *Applied Energy*, 1059-1082.
- Coseo, P., & Larsen, L. (2014). How factors of land use/land cover, building configuration, and adjacent heat sources and sinks explain Urban Heat Islands in Chicago. *Landscape and Urban Planning*, 125, 117-129.
- Crawley, D. B., Hand, J. W., Kummert, M., & Griffith, B. T. (2008). Contrasting the capabilities of building energy performance simulation programs. *Building and Environment*, 661-673.
- Davila, C. C., Reinhart, C., & Bemis, J. (2016). Modeling Boston: A workflow for the generation of complete urban building energy demand models from existing urban geospatial datasets. *Energy*, 237-250.
- Dodman, D. (2009). Blaming cities for climate change? An analysis of urban greenhouse gas emissions inventories. *Environment and Urbanization*, 185-201.
- Duarte, J. P., Beirão, J. N., Montenegro, N., & Gil, J. (2012). City Induction: a model for formulating, generating, and evaluating urban designs. *Digital Urban Modeling and Simulation* (pp. 73-98). Springer Berlin Heidelberg.
- Duarte, J. P., Rocha, J., & Ducla-Soares, G. (2007). Unveiling the structure of the Marrakech Medina: A shape grammar and an interpreter for generating urban form. *AI EDAM*, 21(04), 317-349.
- Duarte, J., & Beirão, J. (2011). Towards a methodology for flexible urban design: designing with urban patterns and shape grammars. *Environmental and Urban Planning B: Planning and Design*, 38, 879-902.
- Ena, V. (2018). De-coding Rio de Janeiro's Favelas. *eCAADe* (pp. 429-438). Lodz, Poland: Proceedings of the 36th eCAADe Conference .
- Ewing, R., & Cervero, R. (2010). Travel and the Built Environment. A Meta-Analysis. *Journal of the American Planning Association*, 76(3), 265-294.
- Ewing, R., & Rong, F. (2008). The Impact of Urban Form on U.S. Residential Energy Use. *Housing policy debate*, 19(1), 1-30.
- Farhangi, H. (2010). The Path of the Smart Grid. *IEEE Power & Energy*, 18-28.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37.
- Fernandez, A., & Blumsack, F. (2010). *Distributing Electric Energy in Rural America Efficiently and Economically: The Micro-grid Option*. The Pennsylvania State University. University Park: The Pennsylvania State University.
- Folke, C., Carpenter, S., Elmqvist, T., & Gunderson, L. (2002). Resilience and Sustainable Development: Building Adaptive Capacity in a World of Transformations. *AMBIO: A Journal of the Human Environment*, 31(5), 437-440.
- Forrester, A., Sobester, A., & Keane, A. (2008). *Engineering Design Via Surrogate Modelling: A Practical Guide*. Wiley.

- Fu, Q., Nasiri, A., Solanki, A., Bani-Ahmed, A., Weber, L., & Bhavaraju, V. (2015). Microgrids: Architectures, Controls, Protection, and Demonstration. *Electric Power Components and Systems*, 1453-1465.
- Garcia, S., & Barros, M. (2015). A grammar-based system for chair design: from generic to specific shape grammars. *eCAADe* (pp. 427-436). Vienna, Austria: Proceedings of the 33rd eCAADe Conference.
- Garson, D. (1991). Interpreting neural-network connection weights. *AI expert*, 46-51.
- Géron, A. (2017). *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*. O'Reilly Media.
- Gil, J., & Duarte, J. (2010). A Review of Urban Design Sustainability Evaluation Tools. *10th International Conference on Design & Decision Support Systems in Architecture and Urban Planning*. Eindhoven University of Technology, Eindhoven.
- Gil, J., Nuno Beirão, J., Montenegro, N., & Duarte, J. P. (2012). On the discovery of urban typologies: data mining the many dimensions of urban form. *Urban Morphology*, 16, 27-40.
- Goh, A. (1995). Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering*, 143-151.
- Gorissen, D., Couckuyt, I., Demeester, P., Dhaene, T., & Crombecq, K. (2010). A Surrogate Modeling and Adaptive Sampling Toolbox for Computer Based Design. *Journal of Machine Learning Research*, 2051-2055.
- Gossop, C. (2011). Low carbon cities: An introduction to the special issue. *Cities*, 495-497.
- Grosso, M. (1998). Urban Form and Renewable Energy Potential. *Renewable Energy*, 331-336.
- Grubler, A., Bai, X., Buettner, T., Dhakal, S., Fisk, D., Ichinose, T., . . . Weisz, H. (2012). Chapter 18: Urban Energy Systems. In *Global Energy Assessment. Toward a Sustainable Future* (pp. 1307-1400). Cambridge, UK and New York, NY, USA and the International Institute for Applied System Analysis, Laxenburg, Austria: Cambridge University Press.
- Gu, N., Singh, V., & Merrick, K. (2010). A framework to integrate generative design techniques for enhancing design automation. *New Frontiers: Proceedings of the 15th International Conference on Computer-Aided Architectural Design in Asia (CAADRIA)* (pp. 127-136). Hong Kong: Association of Computer-Aided Architectural Design Research in Asia (CAADRIA).
- Höök, M., & Tang, X. (2013). Depletion of fossil fuels and anthropogenic climate change- A review. *Energy Policy*, 52, 797-809.
- Hampson, A., Bourgeois, T., Dillingham, G., & Panzare, I. (2013). *Combined Heat and Power: Enabling Resilient Energy Infrastructure for Critical Facilities*. Washington DC: No. ORNL/TM--2013/100. ICF International.

- Hatziargyriou, N., Asano, H., Iravani, R., & Marnay, C. (2007). Microgrids: An overview of ongoing research, development, and demonstration projects. *IEEE Power & Energy*, 78-94.
- Hemsath, T. L. (2016). Housing orientation's effect on energy use in suburban developments. *Energy and Buildings*, 122, 98-106.
- Hu, R., Huang, Z., Tang, Y., van Kaick, O., Zhang, H., & Huang, H. (2020). Graph2Plan: Learning Floorplan Generation from Layout Graphs. *arXiv preprint arXiv:2004.13204*.
- Huang, W., & Zheng, H. (2018). Architectural Drawings Recognition and Generation through Machine Learning. *ACADIA* (pp. 156-165). ACADIA.
- Jebaraj, S., & Iniyan, S. (2006). A review of energy models. *Renewable and sustainable energy reviews*, 10(4), 281-311.
- Jie, W., Yufeng, Z., & Qinglin, M. (2013). Calculation Method of Sky View Factor Based on Rhino-Grasshopper Platform. *13th Conference on International Building Performance Simulation Association* (pp. 2658-2666). Chambéry, France: Proceedings of BS2013.
- Joseph, A. (2019). Shapley regressions: A framework for statistical inference on machine learning models. *arXiv preprint arXiv:1903.04209*.
- Kang, S. J., Park, J., Oh, K., & Park, H. (2014). Scheduling-based real time energy flow control strategy for building energy management system. *Energy and Building*, 239-248.
- Ko, Y. (2013). Urban form and residential energy use: A review of design principles and empirical findings. *CPL bibliography*, 28(4), 327-351.
- Lambert, T., Gilman, P., & Lilienthal, P. (2006). Micropower system modeling with HOMER. *Integration of alternative sources of energy*, 1(15), 379-418.
- Lariviere, I., & Lafrance, G. (1999). Modelling the electricity consumption of cities: effect of urban density. *Energy Economics*, 53-66.
- Lawrence Berkeley National Laboratory (LBNL). (2015). *DER-CAM User Manual*. Berkeley, California: LBNL.
- LEAP – long range energy alternatives planning. (2017, February). Retrieved from <http://www.scib.org/leap/>
- Leng, G. (2000). RETScreen™ International: A decision support and capacity building tool for assessing potential renewable energy projects. *INDUSTRY AND ENVIRONMENT-PARIS*, 22-23.
- Li, A. (2001). *A shape grammar for teaching the architectural style of the Yingzao Fashi*. Massachusetts Institute of Technology: PhD Dissertation.
- Lidula, N., & Rajapakse, A. (2011). Microgrids research: A review of experimental microgrids and test systems. *Renewable and Sustainable Energy Reviews*, 186-202.
- Lobaccaro, G., & Frontini, F. (2014). Solar energy in urban environment: how urban densification affects existing buildings. *Energy Procedia*, 1559-1569.

- Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *31st Conference on Neural Information Processing Systems (NIPS 2017)*, (pp. 4765-4774). Long Beach, CA.
- Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., . . . Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2, 56-67.
- Magoules, F., & Zhao, H.-X. (2016). *Data Mining and Machine Learning in Building Energy Analysis: Towards High Performance Computing*. John Wiley & Sons.
- Mariam, L., Basu, M., & Conlon, M. (2013). A Review of Existing Microgrid Architectures. *Journal of Engineering*, 1-8.
- Markovic, D., Cvetkovic, D., & Masic, B. (2011). Survey of software tools for energy efficiency in a community. *Renewable and Sustainable Energy Reviews*, 4897-4903.
- Marnay, C. (2010). Introduction to Microgrids. *The CADER Conference 2010, Community Energy Systems, the Rise of Local Energy*. La Jolla, CA.
- Maui Solar Energy Software Corporation. (2017, February). Retrieved from <http://www.maisolarsoftware.com/>
- Mendes, G., Ioakimidis, C., & Ferrao, P. (2011). On the planning and analysis of Integrated Community Energy Systems: A review and survey of available tools. *Renewable and Sustainable Energy Reviews*, 4836-4854.
- Mirzaee, S., Özgün, O., Ruth, M., & Binita, K. C. (2018). Neighborhood-scale sky view factor variations with building density and height: A simulation approach and case study of Boston. *Urban Climate*, 95-108.
- Mitchell, T. M. (1997). *Machine Learning*. MIT Press and WCB/McGraw-Hill.
- NETL, & DOE. (2016). *NETL partners with the city of Pittsburgh and regional stakeholders to help design and implement a 21st century energy infrastructure*. Pittsburgh, PA: National Energy Technology Laboratory.
- Newman, P. W., & Kenworthy, J. R. (1989). Gasoline consumption and cities: a comparison of US cities with a global survey. *Journal of the american planning association*, 55(1), 24-37.
- NYSERDA, DHSES, & DPS. (2014). *Microgrids for Critical Facility Resiliency in New York State*. Albany, NY: New York State Legislature.
- Owens, S. (1986). *Energy, Planning, and Urban Form*. London: Pion Limited.
- Paglia, T. K. (2011). *Energy Improvement Districts and Local Energy Production*. Ithaca, NY: Master of Regional Planning Dissertation, Cornell University.
- Pont, M. B., & Haupt, P. (2005). The Spacemate: density and the typomorphology of the urban fabric. *Nordic Journal of Architectural Research*, 55-68.
- Powering a New Generation of Community Energy*. (2015). Retrieved 2017, from New York State Energy Research and Development Authority: <https://www.nyserda.ny.gov/All-Programs/Programs/NY-Prize>

- PV*SOL Premium*. (2017, February). Retrieved from Valentin Software: <http://www.valentin-software.com/en/products/photovoltaics/57/pvsol-premium>
- Rahimian, M., Domenica Iulo, L., & Cardoso Llach, D. (2015). The Case for a Collaborative Energy Sharing Network for Small Scale Community Microgrids . *Architecture and Resilience on the Human Scale* (pp. 215-222). Sheffield: The University of Sheffield.
- Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. *IEEE Data Eng*, 4(23), 3-13.
- Ratti, C., Baker, N., & Steemers, K. (2005). Energy consumption and urban texture. *Energy and Buildings*, 37, 762–776.
- Reinhert, C., Dogan, T., Jakubiec, J., Rakha, T., & Sang, A. (2013). UMI - AN URBAN SIMULATION ENVIRONMENT FOR BUILDING ENERGY USE, DAYLIGHTING AND WALKABILITY. *13th Conference of International Building Performance Simulation Association*, (pp. 476-483). Chambéry, France.
- Robinson, D., Campbell, N., Gaiser, W., Kabel, K., Le-Mouel, A., Morel, N., . . . Stone, A. (2007). SUNtool – A new modelling paradigm for simulating and optimising urban sustainability. *Solar Energy*, 1196-1211.
- Robinson, D., Haldi, F., Kämpf, J., Leroux, P., Perez, D., Rasheed, D., & Wilke, U. (2009). CitySim: Comprehensive Micro-Simulation of Resource Flows for Sustainable Urban Planning. *Eleventh International IBPSA Conference* (pp. 1083-1090). Glasgow, Scotland: Building Simulation.
- Samuel, A. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 210-229.
- Santamouris, M., Papanikolaou, N., Livada, I., & Koronakis, I. (2001). On the impact of urban climate on the energy consumption of buildings . *Solar Energy*, 201-216.
- Sarralde, J. J., Quinn, D. J., Wiesmann, D., & Steemers, K. (2015). Solar energy and urban morphology: Scenarios for increasing the renewable energy potential of neighbourhoods in London. *Renewable Energy*, 10-17.
- Seto, C., Dhakal, S., Bigio, A., Blanco, H., Delgado, G., Dewar, D., . . . Ramaswami, A. (2014). Human Settlements, Infrastructure and Spatial Planning. In O. Edenhofer, R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, . . . J. Minx, *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA.: Cambridge University Press.
- Seyedzadeh, S., Pour Rahimian, F., Glesk, I., & Roper, M. (2018). Machine learning for estimation of building energy consumption and performance: a review. *Visualization in Engineering*, 6(5).
- Sherman, G. R. (2007). Sharing Local Energy Infrastructure - Organizational Models for Implementing Microgrids and District Energy Systems in Urban Commercial

- Districts. Cambridge, MA: Master of City Planning, Massachusetts Institute of Technology.
- Siderius, H.-P. (2004). The End of Energy Efficiency Improvements = The Start of Energy Savings?! . *2004 ACEEE Summer Study on Energy Efficiency in Buildings* (pp. 11.165 - 11.176). Pacific Grove, CA: American Council for an Energy-Efficient Economy.
- Silva, M. C., Horta, I. M., Leal, V., & Oliveira, V. (2017). A spatially-explicit methodological framework based on neural networks to assess the effect of urban form on energy demand. *Applied Energy*, 386-398.
- Silva, M., Oliveira, V., & Leal, V. (2017). Urban Form and Energy Demand: A Review of Energy-relevant Urban Attributes. *Journal of Planning Literature*, 0885412217706900.
- Simpson, T., Toropov, V., Balabanov, V., & Viana, F. (2008). Design and analysis of computer experiments in multidisciplinary design optimization: a review of how far we have come or not. *Proceedings of the 12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*. MAO, Victoria, Canada.
- Southworth, M., & Owens, P. (1993). The Evolving Metropolis: Studies of Community, Neighborhood, and Street Form at the Urban Edge. *Journal of American Planning Association*, 271-287.
- Stadler, M., Cardoso, G., Mashayekh, S., Thibault, F., Nicholas, D., & Agarwal, A. (2016). Value Streams in Microgrids: A literature Review. *Applied Energy*, 162, 980-989.
- Steadman, P. (1977). Energy and Patterns of Land Use. *Journal of Architectural Education*, 62-67.
- Steadman, P., Evans, S., & Batty, M. (2009). Wall area, volume and plan depth in the building stock. *Building Research & Information*, 37, 455-467.
- Steemers, K. (2003). Energy and the city: density, buildings and transport. *Energy and Buildings*, 3-14.
- Stiny, G., & Gips, J. (1972). Shape Grammars and the Generative Specification of Painting and Sculpture. *IFIP Congress (2)*, 2, pp. 125-135.
- Taha, H. (1997). Urban climates and heat islands: albedo, evapotranspiration, and anthropogenic heat. *Energy and Buildings*, 99-103.
- Tamke, M., Nicholas, P., & Zwierzycki, M. (2018). Machine learning for architectural design: Practices and infrastructures. *International Journal of Architectural Computing*, 123-143.
- Tardioli, G., Kerrigan, R., Oates, M., O'Donnell, J., & Finn, D. (2015). Data driven approaches for prediction of building energy consumption at urban level. *Energy Procedia*, 78, 3378-3383.
- Teeling, C. (1996). Algorithmic design: generating urban form. *Urban Design Studies*, 89-100.

- The City of San Diego. (2015). *City of San Diego Climate Action Plan*. San Diego: The City of San Diego.
- Ton, D. T., & Smith, M. A. (2012). The US Department of Energy's microgrid initiative. *The Electricity Journal*, 25(8), 84-94.
- Tsui, K. L., Chen, V., Jiang, W., & Aslandogan, Y. (2006). Data Mining Methods and Applications. *Springer handbook of engineering statistics*, 651-669.
- US Department of Energy. (2011). *Solar Powering Your Community: A Guide for Local Governments*. US Department of Energy.
- van den Hoek, J. (2008). The MXI (Mixed-use Index) as Tool for Urban Planning and Analysis. *Corporations and Cities: Envisioning Corporate Real Estate in the Urban Future*.
- Van Hende, K., & Wouters, C. (2014). The Regulation of Microgrids in Liberalized Electricity Markets in the EU and East Asia. *European Networks Law and Regulation Quarterly (ENLR)*, 2(3), 190-206.
- Vaz Monteiro, M., Doick, K. J., Phillip, H., & Peace, A. (2016). The impact of greenspace size on the extent of local nocturnal air temperature cooling in London. *Urban Forestry & Urban Greening*(16), 160-169.
- Verniz, D., & Duarte, J. P. (2017). Santa Marta Urban Grammar - Towards an understanding of the genesis of form. *eCAADe* (pp. 477-484). Rome, Italy: proceedings of the 35th International Conference on Education and Research in Computer Aided Architectural Design in Europe.
- Villareal, C., Erickson, D., & Zafar, M. (2014). *Microgrids: A Regulatory Perspective*. San Francisco: CPUC Policy & Planning Division.
- Walter, E., & Kämpf, J. (2015). A verification of CitySim results using the BESTEST and monitored consumption values. *Proceedings of the 2nd Building Simulation Applications conference* (pp. 215-222). Bozen-Bolzano, Italy: Bozen-Bolzano University Press.
- Wang, H., & Chen, Q. (2014). Impact of climate change heating and cooling energy use in buildings in the United States. *Energy and Buildings*, 428-436.
- Werbos, P. J. (2011). Computational intelligence for the smart grid-history, challenges, and opportunities. *IEEE Computational Intelligence Magazine*, 6(3), 14-21.
- Williams, B., Gahagan, M., & Costin, K. (2010). Using Microgrids to Integrate Distributed Renewables Into the Grid. *Innovative Smart Grid Technologies Conference Europe (ISGT Europe)* (pp. 1-5). Gothenburg, Sweden: IEEE PES.
- Wilson, B. (2013). Urban form and residential electricity consumption: Evidence from Illinois, USA. *Landscape and Urban Planning*, 115, 62-71.
- Wouters, C. (2015). Towards a Regulatory Framework for Microgrids - The Singapore Experience. *Sustainable Cities and Societies*, 22-32.

VITA - Mina Rahimian

EDUCATION

The Pennsylvania State University, Department of Architecture, University Park, PA
Doctoral Candidate in Architecture, Computational Design and Sustainability (PhD) 2015 - 2021
Master of Science in Architecture, Computational Design and Sustainability (MS) 2013 - 2015
The Pennsylvania State University, Department of Statistics, University Park, PA
Graduate Certification in Applied Statistics 2016
University of Tehran, Department of Architectural Engineering, Tehran, Iran 2008 - 2013
Bachelors in Architectural Engineering

WORK EXPERIENCE

Outer Labs, Remote Spring 2021 - Present
Senior Technical Project/Product Manager
The Pennsylvania State University, Remote Fall 2021 - Present
Adjunct Lecturer
iBuilt (previously known as Deluxe Modular), New York, NY Winter 2019-Spring 2021
Director of Generative Design
Autodesk, San Francisco, CA Spring 2019-Winter 2019
Sustainable AEC Generative Design Researcher
Design Computing Group at University of Lisbon, Lisbon, Portugal Spring 2018-Spring 2019
Visiting Researcher, Affiliate Researcher
Autodesk, San Francisco, CA Summer 2017-Fall 2017
Sustainability Technology Analyst Intern

PUBLICATIONS

- **Rahimian, M.**, Duarte, J., & Iulo, L. (2021). A real-time predictive software prototype for simulating urban-scale energy consumption based on surrogate models. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 35(4), 353-368.
doi:10.1017/S0890060421000184
- **Rahimian, M.**, Cervone, G., Duarte, J.P., Iulo, L.D., (2020), A Machine Learning Approach for Mining the Multidimensional Impact of Urban Form on Community Scale Energy Consumption in Cities. in *Design Computing and Cognition 2020 (DCC'20)*, (pp. 627-646)
- **Rahimian, M.**, J. Beirao., J.P. Duarte., L.D. Iulo. (2019) A Generative Urban Design Tool Considering Topographic Constraints: The Case of American Urban Planning. *Proceedings of eCAADe+SIGraDi'19*
- **Rahimian, M.**, Iulo, L.D., Duarte, J.P.,(2018). A Review of Predictive Software for the Design of Community Microgrids. (pp. 1-13). *Hindawi Journal of Engineering*
- **Rahimian, M.**, Iulo, L.D., (2017). A Feedback Delivery System for Communal Energy Consumption Practices. *Springer Lecture Notes in Computer Science and Artificial Intelligence* (Invited paper submission)
- **Rahimian, M.**, Cardoso Llach, D., & Iulo, L. D. (2015). Participatory Energy Management in Building Networks. *Sustainable Human-Building Ecosystems* (pp. 27-35). Pittsburgh: American Society of Civil Engineers.
- **Rahimian, M.**, Iulo, L. D., & Cardoso Llach, D. (2015). The Case for Collaborative Energy Sharing Network for Small - Scale Community Microgrids. *Architecture and Resilience on the Human Scale* (pp. 200 - 193). Sheffield: The University of Sheffield.
- Iulo, L. D., Kimel, A., Fadaei, S., Macht, K., **Rahimian, M.**, & Riley, D. (2014). Innovate, Evaluate, Learn: The Importance of Iterative Research in Housing Design-Build. 2014 ACSA Fall Conference: Working out - Thinking while Building. Halifax: ACSA Press.