A Machine Learning-Based Approach to Predict and Optimize the Performance of Zero Energy Building (ZEB): A Case Study for Florida

Benjamin Kubwimana: bkubwimana2016@my.fit.edu

MSc Mechanical Engineering

Advisor: Dr. Hamidreza Najafi

Committee Members: Dr. Troy Nguyen, Dr. Aldo Fabregas, Dr. Ryan White

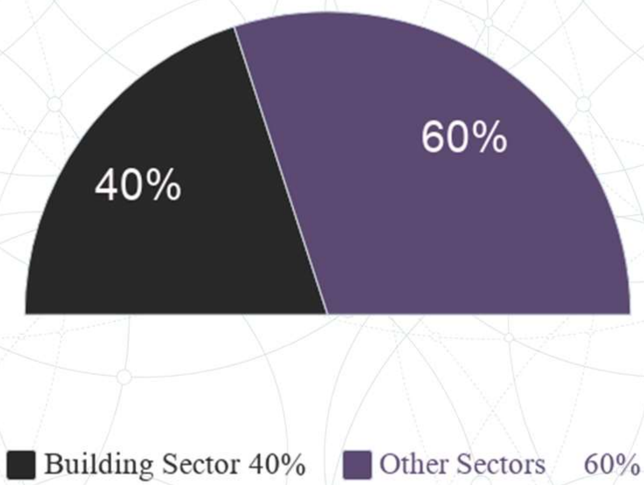
December 3rd 2021

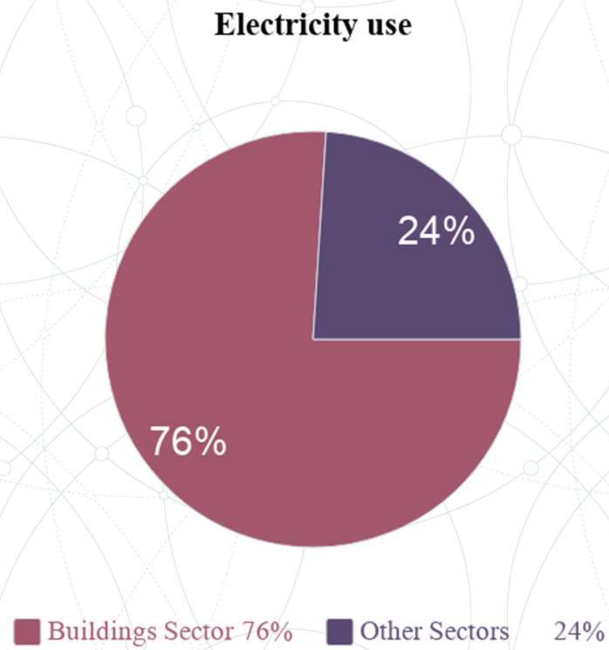
Presentation Outline

- **Motivation**
- **Problem Statement**
- **Literature Review**
- **Methodology**
  - Building energy model: EnergyPlus
  - Neural Network based Surrogate Model
  - Surrogate Model Optimization
- **Results and Discussion**
- **Conclusion**
- **Future Consideration**

Motivation

More than 40% of all U.S. energy use and greenhouse gas (GHG) emissions are associated with the building sector



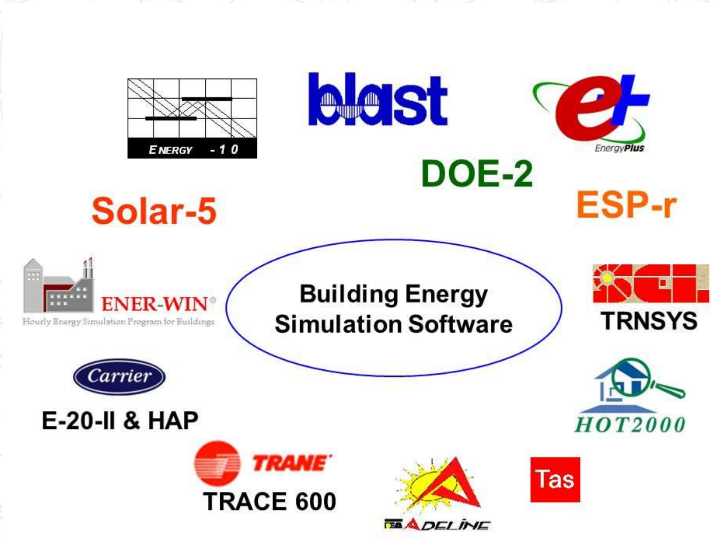


Motivation

- Building energy evaluation is a complex problem to solve
- Several interconnected subsystems impact overall energy performance of a building:
  - Architectural design
  - Envelope materials
  - Energy end-users
  - Operation, control, and maintenance

Motivation

- Current simulation tools include:
    - EnergyPlus,
    - eQuest,
    - DesignBuilder, etc.
- Developing a physics-based model for energy simulation of buildings
- Facilitates simulation of multiple scenarios (parametric study)



https://www.slideshare.net/hamzaaaaaah/energy-simulation-analysis

Problem Statement

- Building energy simulation tools, such as EnergyPlus, are very powerful in predicting the hourly energy consumption of buildings
- They allow conducting various parametric studies to understand the impact of each parameter on the performance of the building,
- However, they do not facilitate the optimization process for achieving a global optimal building design

- **My study attempts to develop an effective approach that is both computationally efficient and facilitates the global optimization of building parameters to produce an efficient building design**

FLORIDA TECH

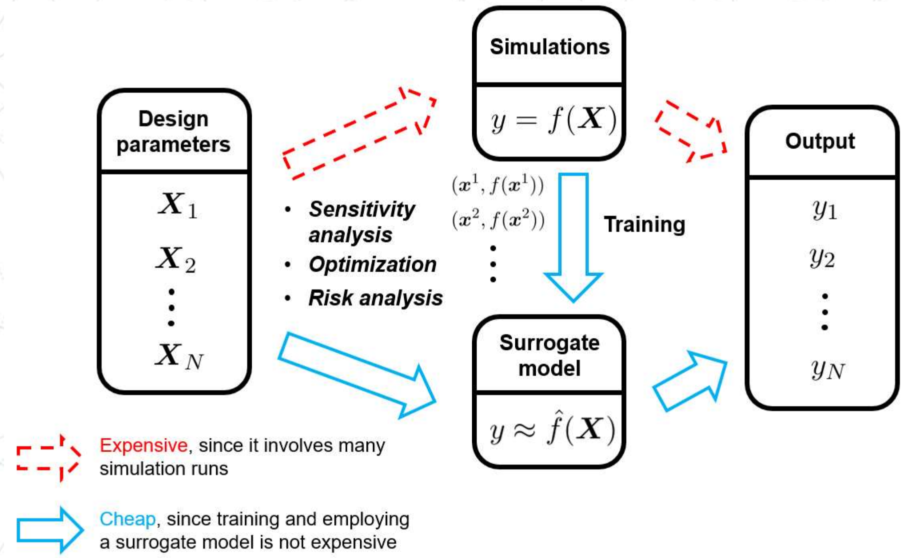
Introduction

Simplified models instead of detailed building models, namely "surrogate model", have become popular

A surrogate model (meta-model) is a prediction model of the original simulation model, and it reliably represents its behavior

There are several methods used to construct a surrogate model based on data sets:

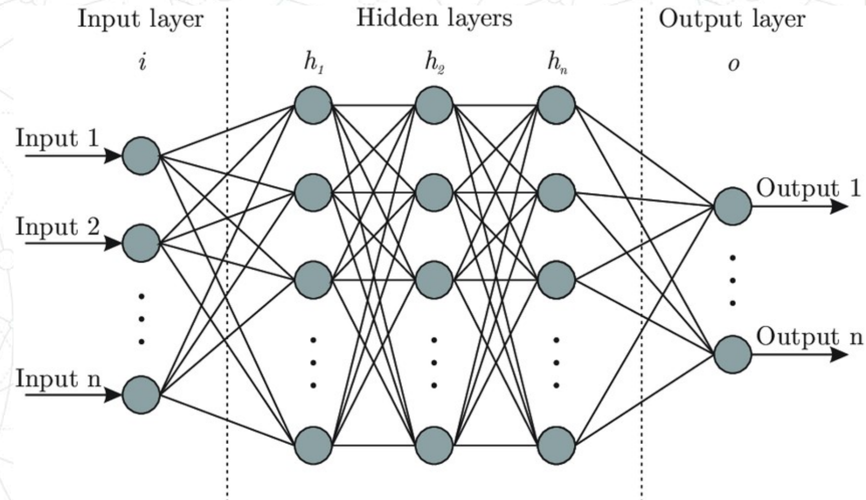
- Artificial neural network (ANN),
- Support vector machine (SVM) and
- Gaussian process regression



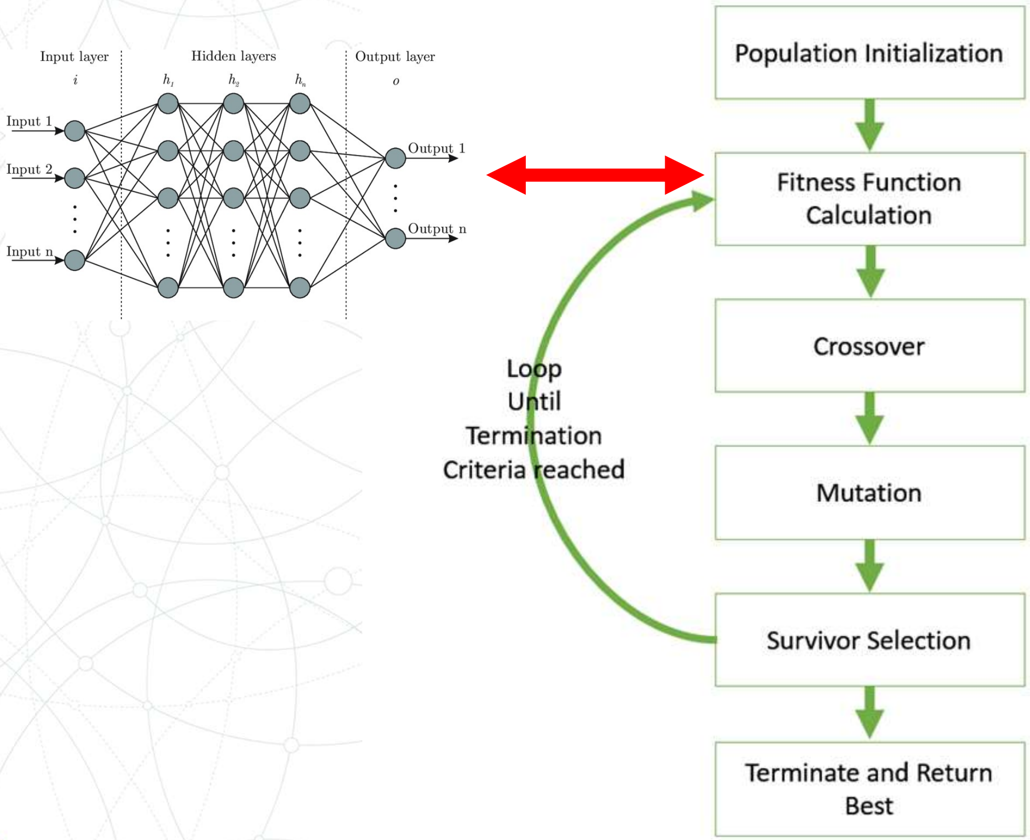
https://towardsdatascience.com/an-introduction-to-surrogate-modeling-part-i-fundamentals-84697ce4d241

Introduction

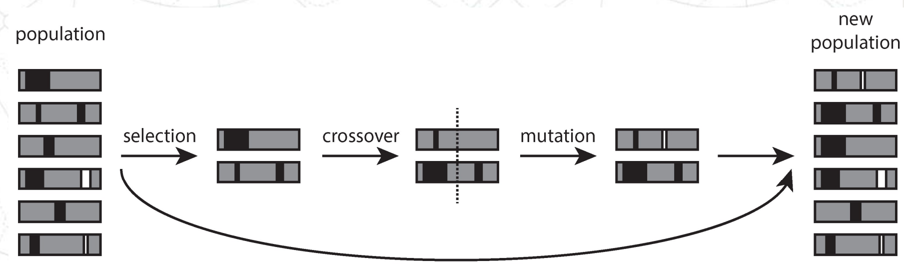
The so-called “artificial” neural networks, as opposed to the biological neurons in our brain, imitate the functioning of the human brain



Introduction



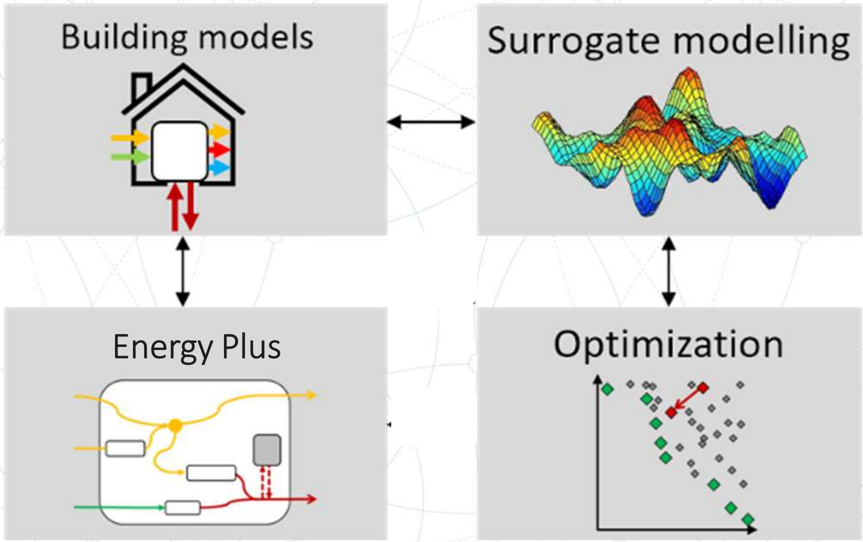
Genetic Algorithm for Optimization



Surrogate Models can then be coupled with Optimization algorithms such as the conventional Genetic optimization Algorithm

Guariso, Giorgio, and Matteo Sangiorgio. 2020. "Improving the Performance of Multiobjective Genetic Algorithms: An Elitism-Based Approach" *Information* 11, no. 12: 587. https://doi.org/10.3390/info11120587

Introduction



- Optimizing surrogate models vs. physics-based model
- Developing an accurate surrogate model is a challenge to the process

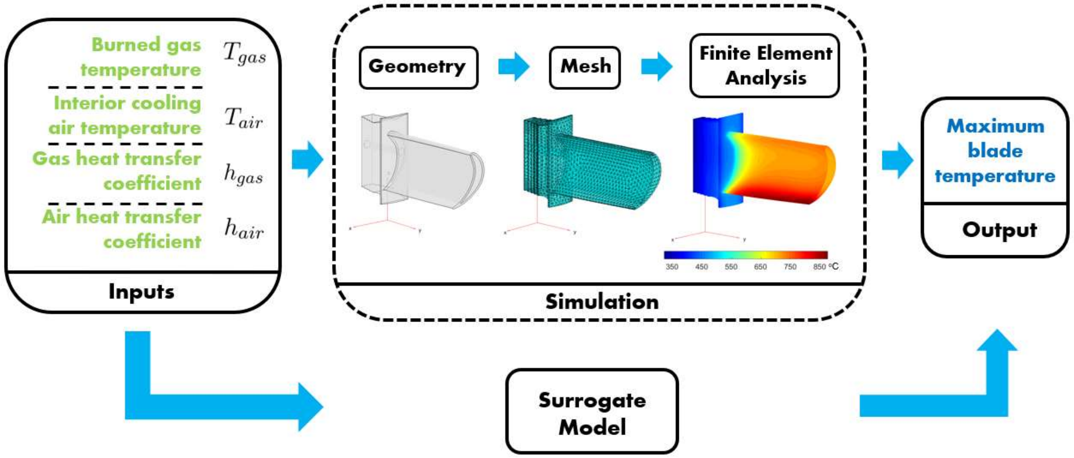
https://besos.readthedocs.io/en/stable/

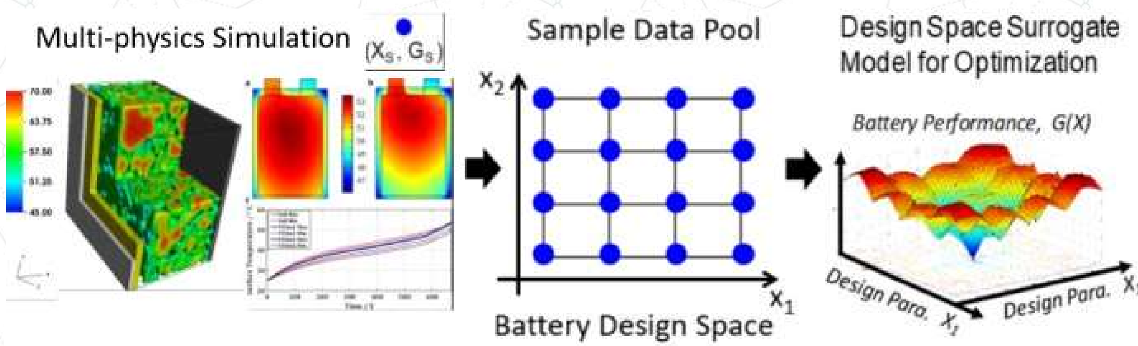
Literature Review

Surrogate Modeling

Given the benefits of Surrogate Models; they are being used for various applications:

1. Turbine blade models
2. Battery design models
3. Fluid Boiling heat flux Curves
4. Biomedical model

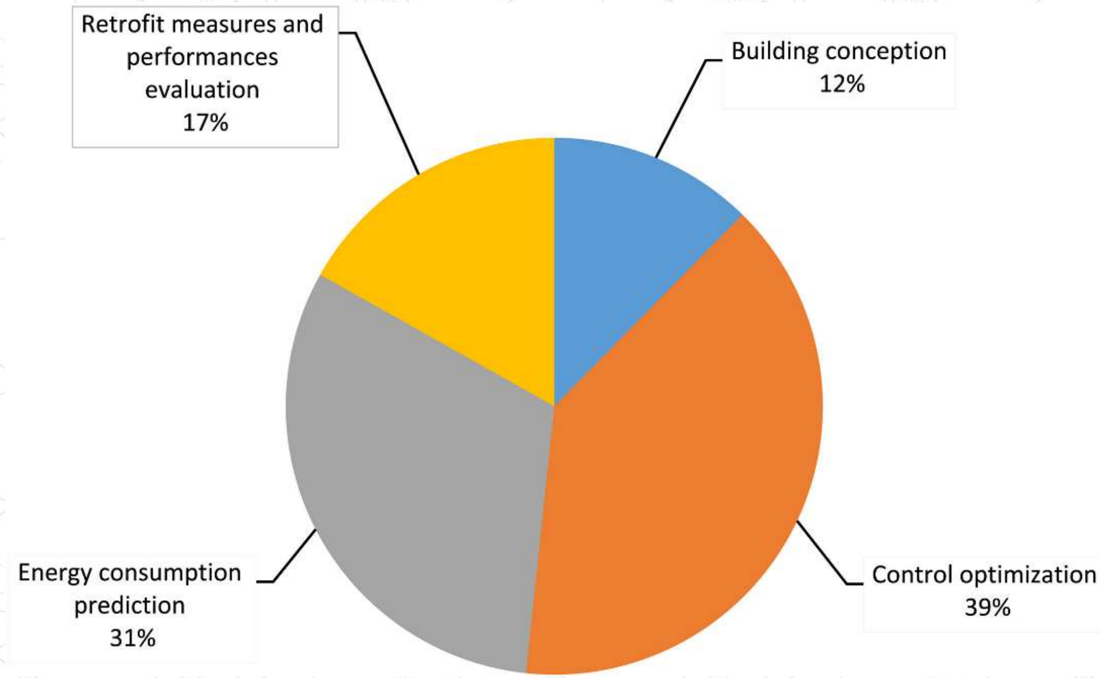




https://towardsdatascience.com/an-introduction-to-surrogate-modeling-part-ii-case-study-426d8035179e

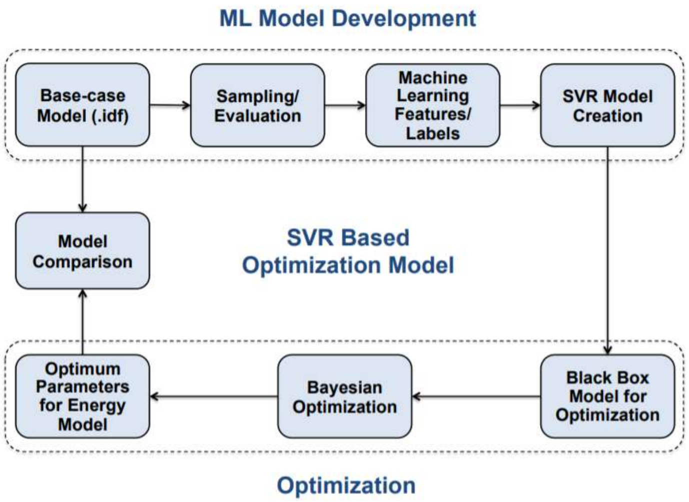
Literature Review

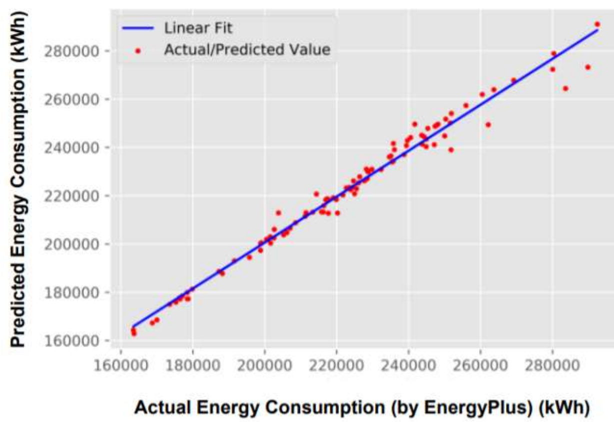
- The applications of neural networks in the building sector are numerous and diverse
- Neural Networks are implemented in the building sector
- A total of 89 published articles from 1998 to 2018 were divided into four categories ranging from design to renovation of a building

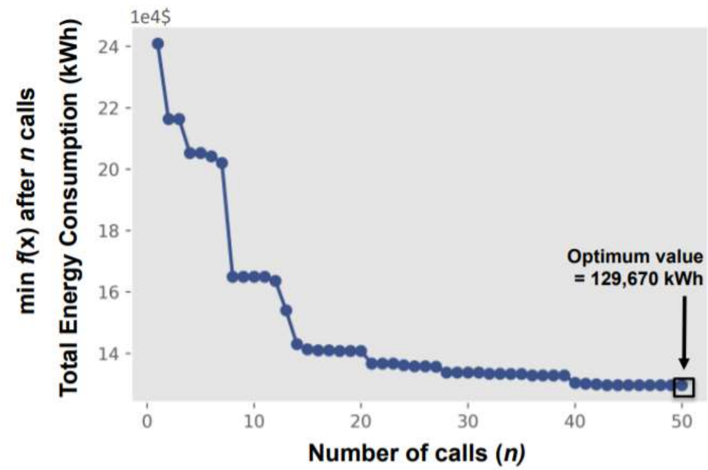


Literature Review

- Approximation accuracy: R2 of 0.9813
- Energy consumption reduced from average of 220,000 kWh to 129,000 kWh per year







- Machine Learning (ML) Model based on Support Vector Regression algorithm to predict energy consumption was created
- Bayesian optimization was performed on the ML model to determine the optimum design parameters

https://www.ashrae.org/file%20library/conferences/specialty%20conferences/2020%20building%20performance/papers/d-bsc20-c009.pdf

FLORIDA TECH

Research Questions

1. Is it possible to build an accurate Neural Network Based Surrogate Model from a physics-based model?

2. Can we use the Genetic Algorithm to optimize building energy consumption of Surrogate Models?

3. What are the benefits of Genetic Algorithm compared to Bayesian Algorithm when optimizing surrogate models?

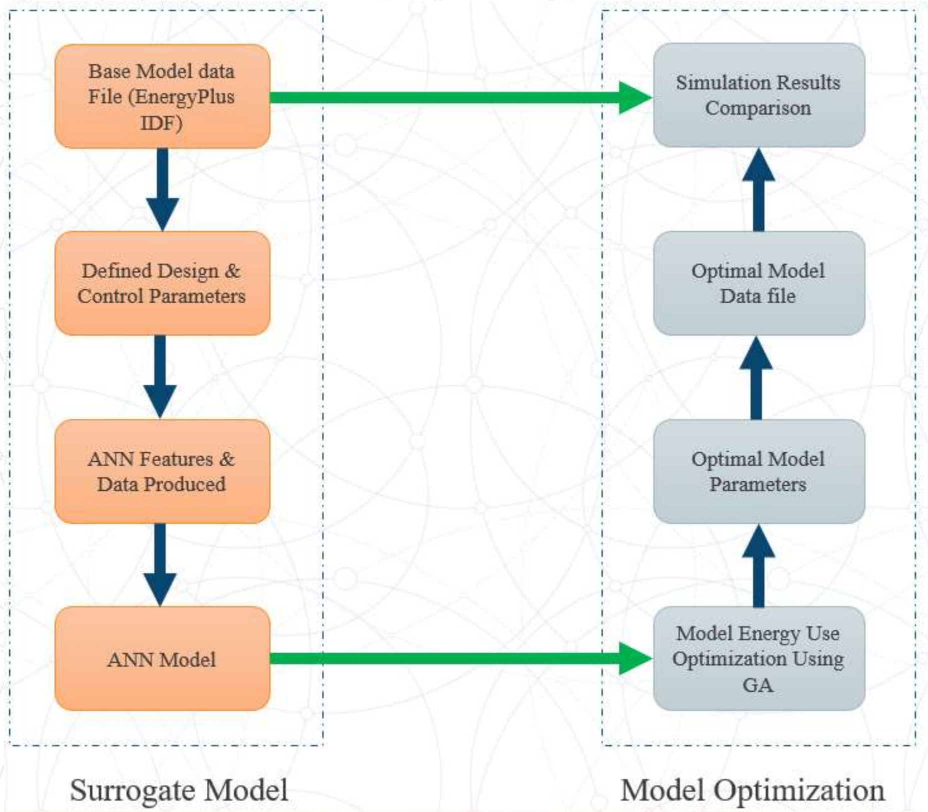
Methodology

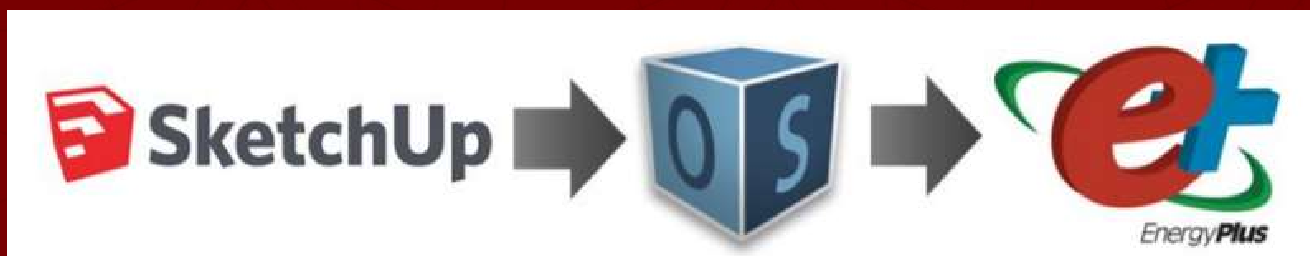
Stage 1: Develop a physics-based model of the building using EnergyPlus

Stage 2:Develop a surrogate model using artificial neural networks . The generated dataset is loaded to the predictive model to train it to predict energy consumption, production, and net site energy

Stage 3: Optimization of the surrogate model. The neural network model is then used as a black-box for the Genetic optimization algorithm.

Methodology





Building Energy Modeling

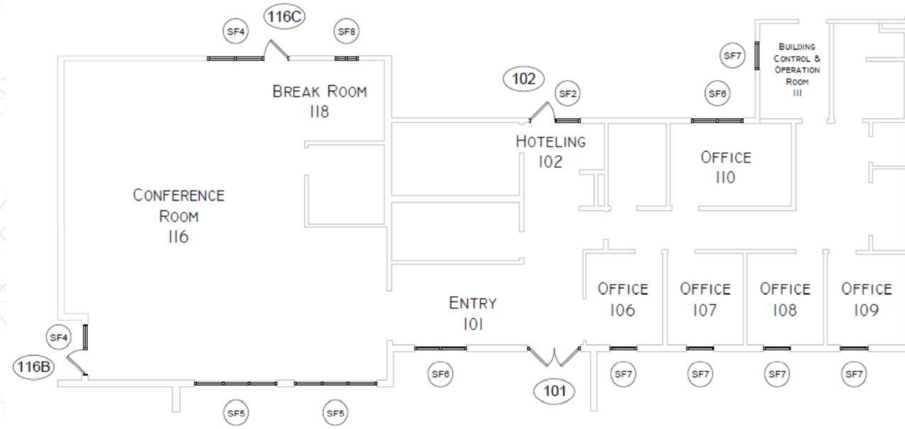
Physics Based Model

FLORIDA TECH

Building Geometry

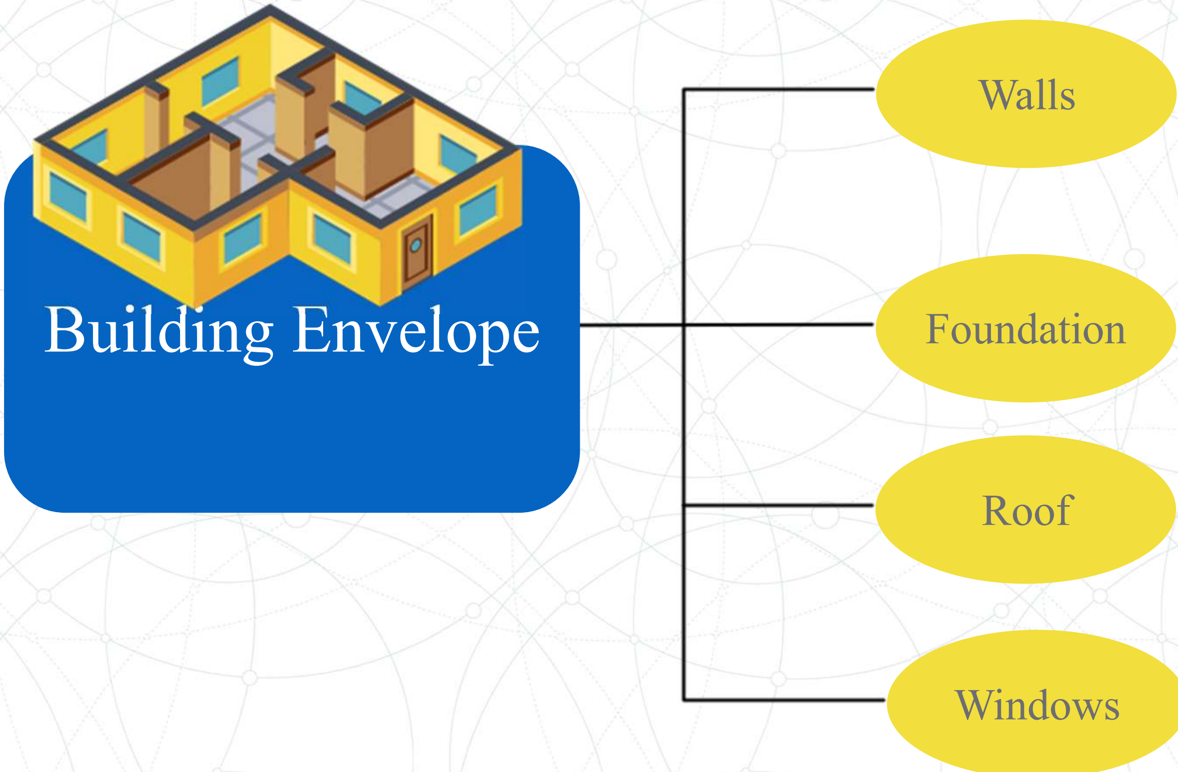
Folliard alumni center is used as a case study

- CAD modeling was performed using SketchUp with an OpenStudio Plugin
- Some of the designed components at this stage include:
  - Building Envelope
  - Offices, restrooms, storage & conference room
  - Specify thermal zones & space types
  - Windows
  - Set material thermal values
  - PV system



Structural Modeling

SketchUp Modeling

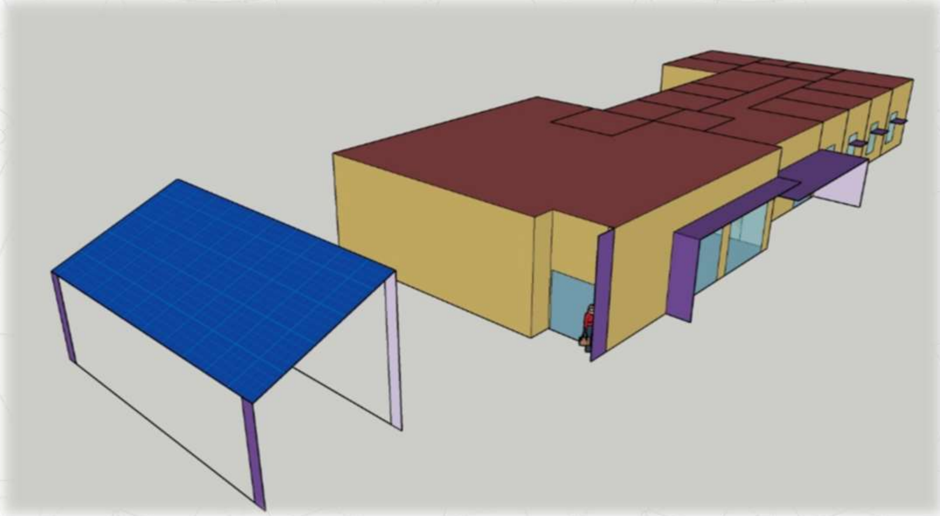


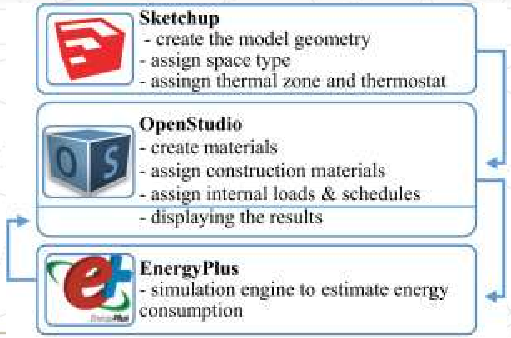
Structural designs are done based off the dimensions provided by the architectural design by BRPH.

OpenStudio Modeling

The geometry model was then exported to OpenStudio to complete the modeling process by adding:

1. Internal loads: people & equipment
2. HVAC system loops
3. Water system
4. Occupancy schedules
5. Lighting schedules
6. HVAC schedules
7. Rest of modeling was performed with E+





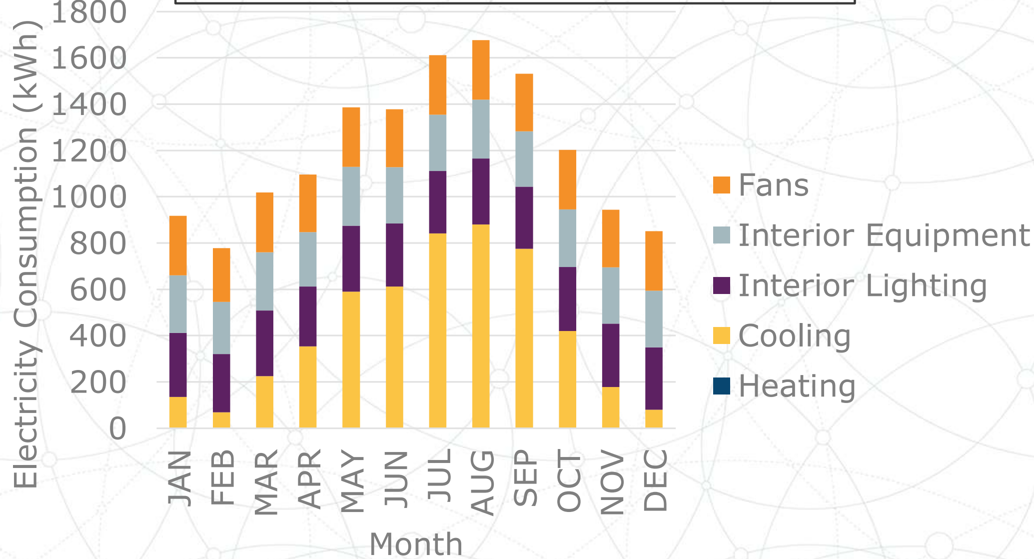
Simulation Results

- Once all building components modeling was complete, a simulation was performed using the EnergyPlus engine taking an average of 3 mins to complete

- Two models were created and simulated to represent:

  1. Building operations during pandemic
  2. Building operations during non pandemic times (regular operations)

Monthly energy consumption for regular building operations



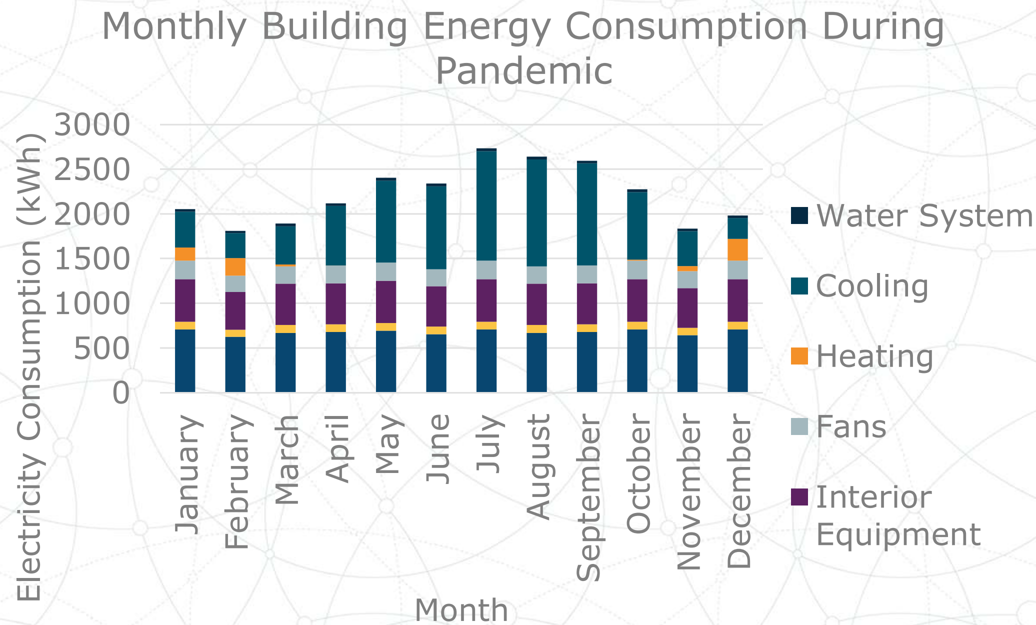
Electricity consumption	14,243 kWh
Total Annual electricity peak demand	10.15 kW
EUI (Energy Use Intensity)	15 kBtu/ft^2/year

Simulation Results

Pandemic operations cause high building consumptions because:

1. HVAC system is set to continuous operations with cooling setpoint of 24
2. Conference room occupancy scheduled is changed to office activity schedule from 8 to 5pm

Higher energy consumptions are expected during these times



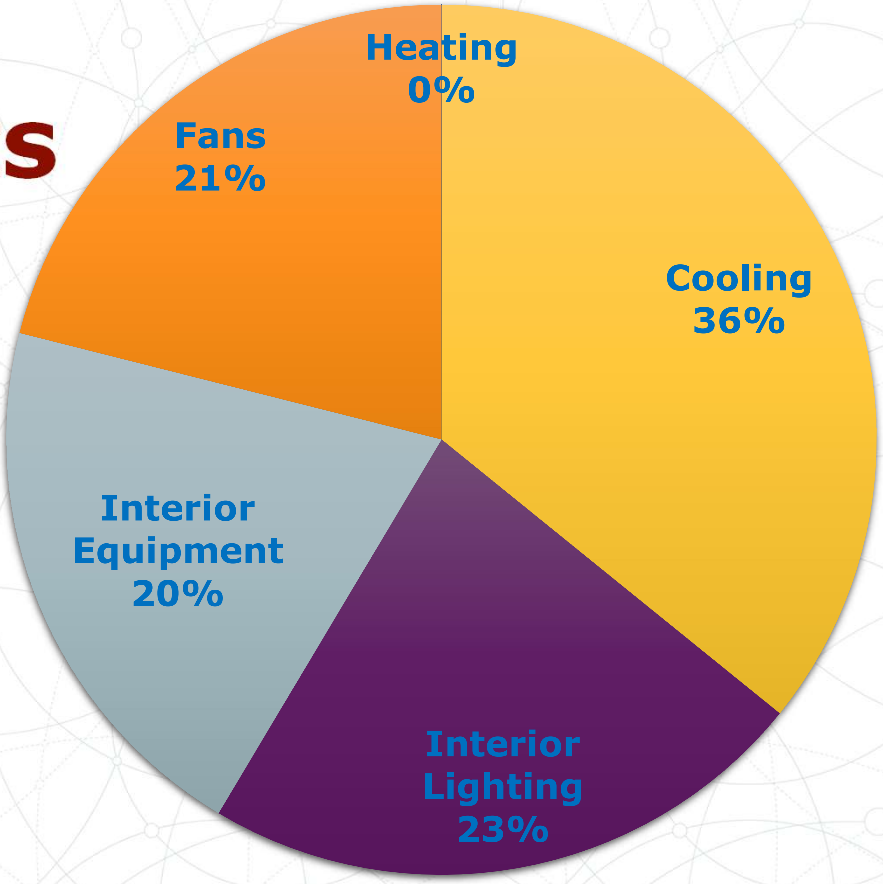
Electricity consumption	26,674 kWh
Total Annual electricity peak demand	20 kW
EUI (Energy Use Intensity)	25 kBtu/ft^2/year

FLORIDA TECH

Simulation Results

Energy end uses have ranging consumptions:

- Cooling consumes the highest
- Followed by lighting
- Heating consumes the least amount of energy
- Each of the end uses were accurately selected to improve the energy efficiency of the Folliard Alumni Center



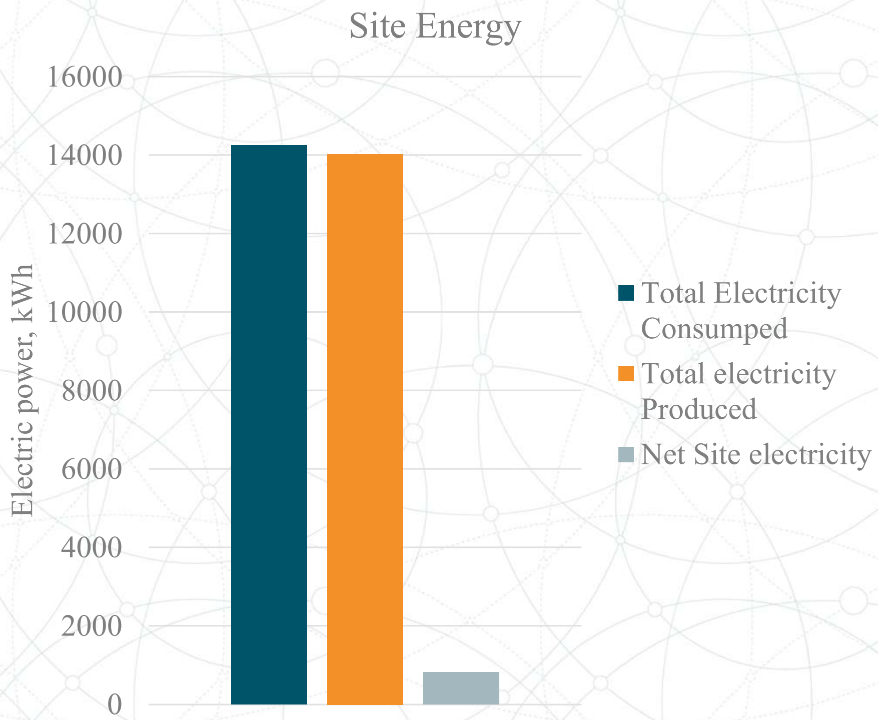
Regular building operations end use consumptions

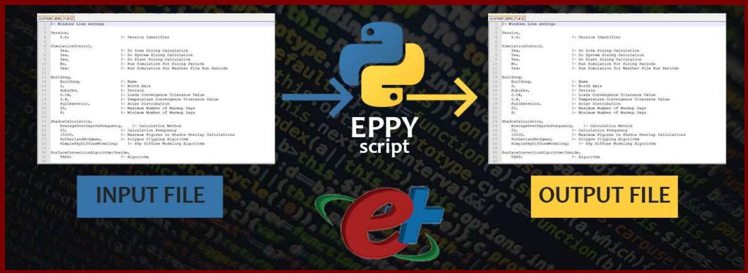
Simulation Results

Data is sampled by selecting annual data pertaining to:

1. Energy consumption
2. PV power production
3. Net site energy (depicts cost)

A similar approach is done for the model simulating the pandemic times



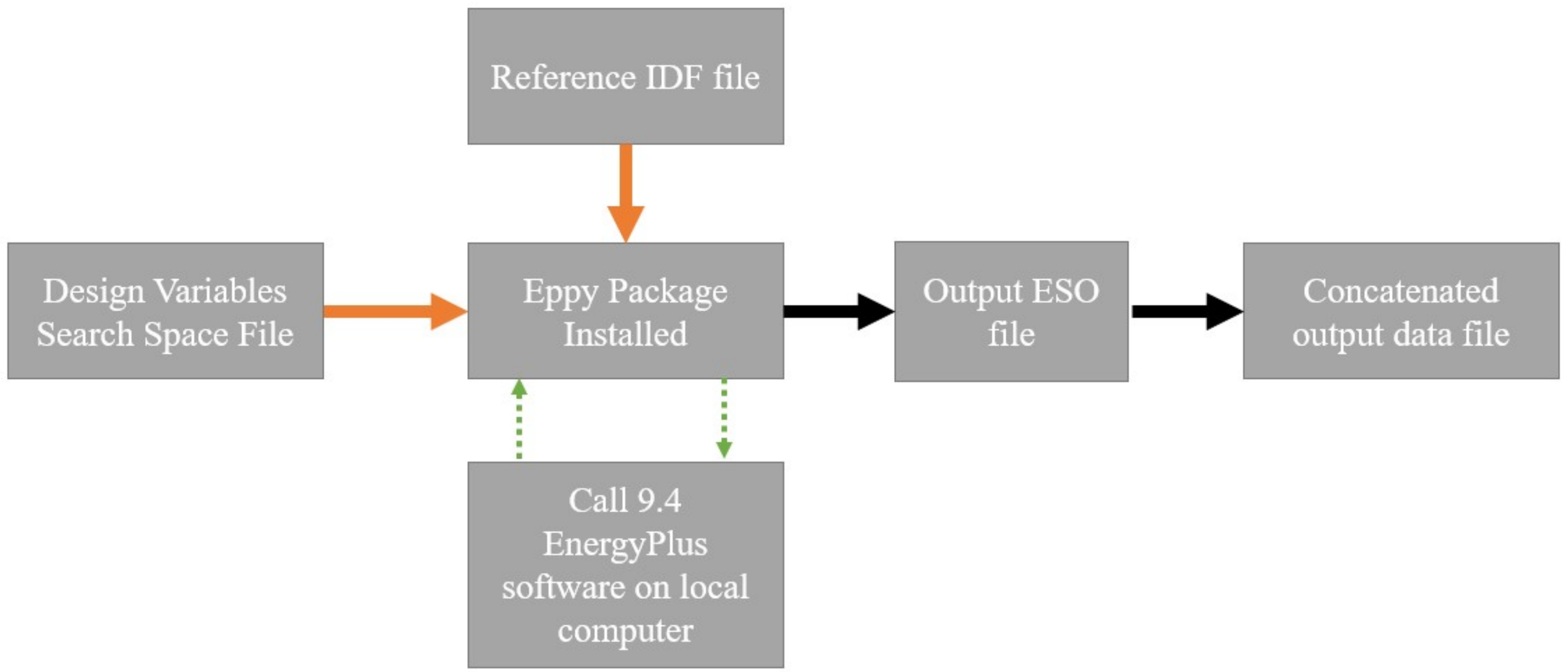


Auto-Simulation Program

EnergyPlus python API, Database generation

https://edsglobal.com/projects/index/Y2F0aWQ9NSZjYXRwb3NpdGlvbj0z/

Program Flow Chart



Parameter Search Space

- 8 design and control parameters are changed 200 times
- Generate subsequent energy consumption, production, and net site energy
- A total of 200 different building designs are generated
- Parameter values are randomly selected between the ranges provided in table for each variable

Parameter	Minimum	Maximum
Wall R-value	8	25
Roof R-value	25	45
Window U-value	1	7
Window SHGC	0.2	0.7
Cooling Setpoint	21	26
HVAC Cooling COP	2	5.5
Light power density	2	15
PV Tilt Angle	10	35

Target parameters can be easily changed based on a designer’s needs.

FLORIDA TECH

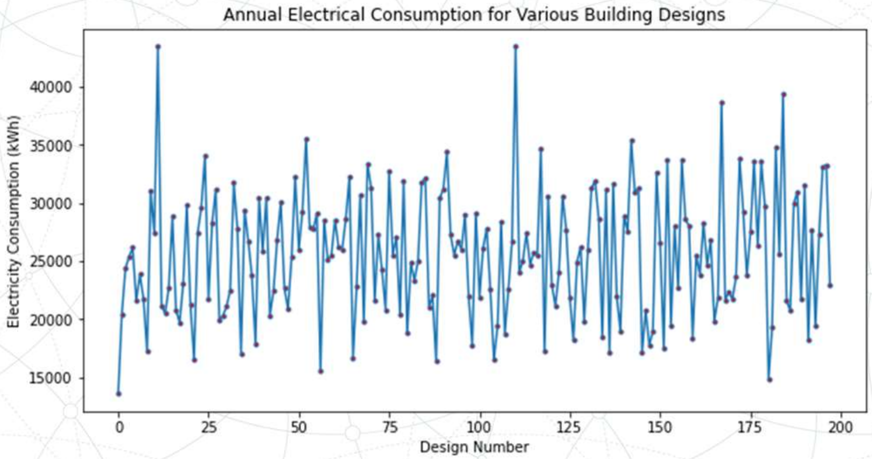
Parameter Search Space

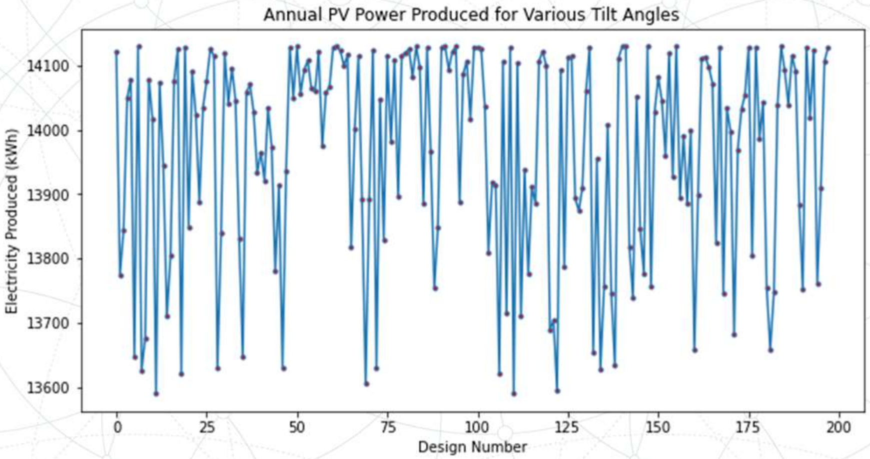
- Eppy API was given the variables of interests the reference IDF file
- Once variables are successfully mapped, they are continually changed to generate new designs
- The new model containing new design values is simulated to provide annual energy data
- Energy data is added to the database until generation is complete

```
def get_mapped_vals(col):
    if col=='Wall R value':
        return 'Thermal_Resistance'

    elif col=='Roof R value':
        return 'Thermal_Resistance'

    elif col== 'HVAC COP':
        return 'Gross_Rated_Cooling_COP'

    elif col=='Window U value':
        return 'UFactor'

    elif col== 'Window SGHC':
        return 'Solar_Heat_Gain_Coefficient'

    elif col== 'Cool setpoint':
        return 'Value_Until_Time_1'

    elif col=='Light Pwr Density':
        return 'Watts_per_Zone_Floor_Area'

    elif col== 'PV Tilt angle':
        return 'Tilt_Angle'

def make_mappings(obj_num_df, obj_name_df):
    mapping= dict()

    for col in obj_num_df.columns:
        mapping[col]= [obj_name_df[col].dropna().tolist(),(obj_num_df[col]
.dropna().astype(int)-1).tolist(),get_mapped_vals(col)]
    return mapping

mappings= make_mappings(obj_num_df,obj_name_df)
```

Dataset Distribution

Pandemic Building Operations



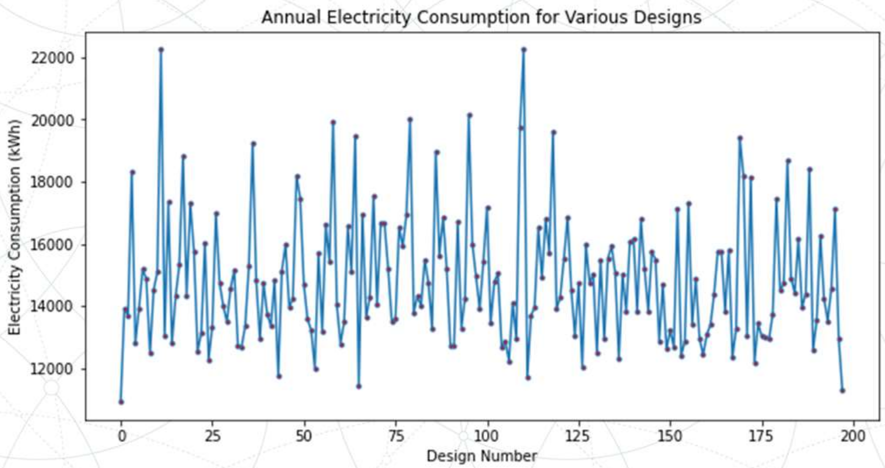


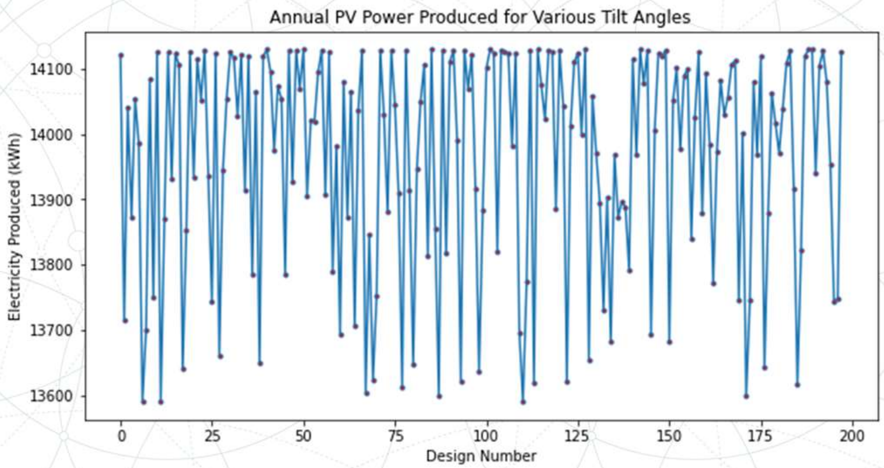
- Across all designs the distribution of energy consumptions and PV power production varies around the baseline model

- The PV power production is only affected by one variable in the designs: tilt angle
- The tilt angle selections are random for each design

Dataset Distribution

Regular building Operations





- Similarly, the energy consumption for the non-pandemic operations varies around 14,000 kWh per year for each building design
- PV power production is not different from the pandemic model because the same variable is changed under same constraints

Artificial Neural Network Based Model

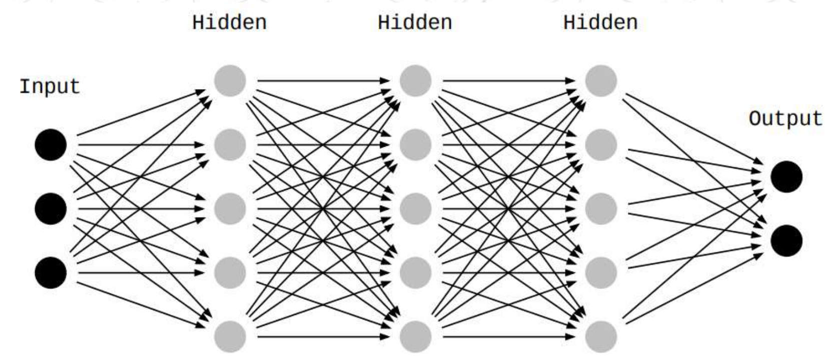
Surrogate Model

Research question 1

Model Architecture

- ANNs are made up of layers an input layer, hidden layers, and an output layer
- These layers are made up of interconnected neurons. Each neuron has an associated weight and threshold
- An individual neuron is activated if its output is above the specified threshold value; this sends data to the next layer
- Otherwise, data is not passed along to the next network layer. Different activation functions are used at each layer

$$a = \sum_{i=1}^{m} (W_i X_i + b)$$



$$output = f(x) = \frac{1}{1 + e^{-a}}$$

$$MSE = \frac{1}{m} \sum_{i=1}^{m} (\hat{y} - y)^2$$

Model Architecture

- There are a total of 8 features and 3 outputs, and each feature has 200 data points.
- The input layers and output layers contain 8 and 3 neurons, respectively.
- Data preprocessing is also included during NN model development, but since this study involved automated data generation, the python co-simulation program was also used to do data preprocessing before network development.
- before data was fed to the neural network, it was standardized to have the mean

Layer number	Neurons	Activation function
1: input layer	8	ReLU
2	dropout (0.2)	
3	128	ReLU
4	dropout (0.2)	
5	64	ReLU
6	32	ReLU
7	16	ReLU
8	8	ReLU
9	4	ReLU
10: output layer	3	

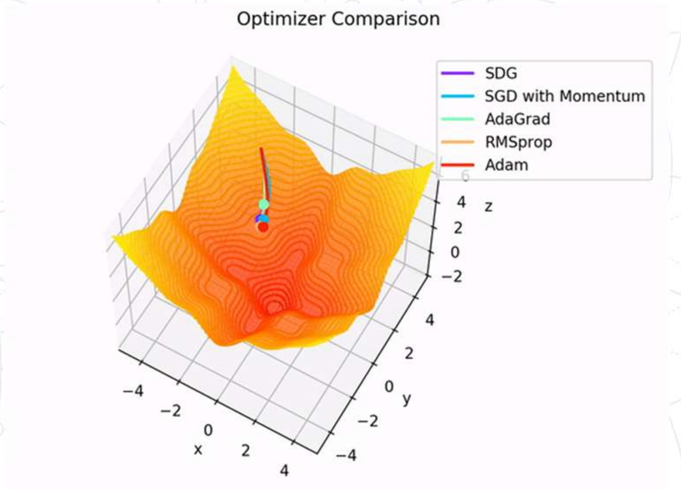
FLORIDA TECH

Model Architecture

- Training dataset accounts 80%
- Validation dataset accounts for 20%

- The model is modified to increase the prediction accuracy as measured by the R2 coefficient of determination and RMSE. A few things that are made in this modification include:

  - Using standard scaling
  - Increased the number of neurons in the dense layers.
  - Using Adam optimizer
  - I am using the Uniform initialization since it was allowing better model performance
  - ReLU activation provided better convergence with lower loss values



Results and Discussion

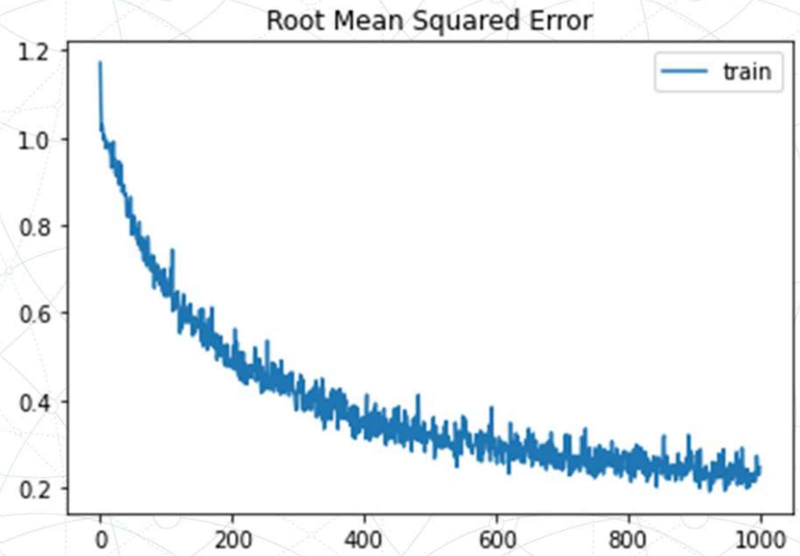
Pandemic Operations

On the testing, the model converges to an RMSE loss of 0.151, which is a good performance

However, this is only the first model version with minimal tuning implemented

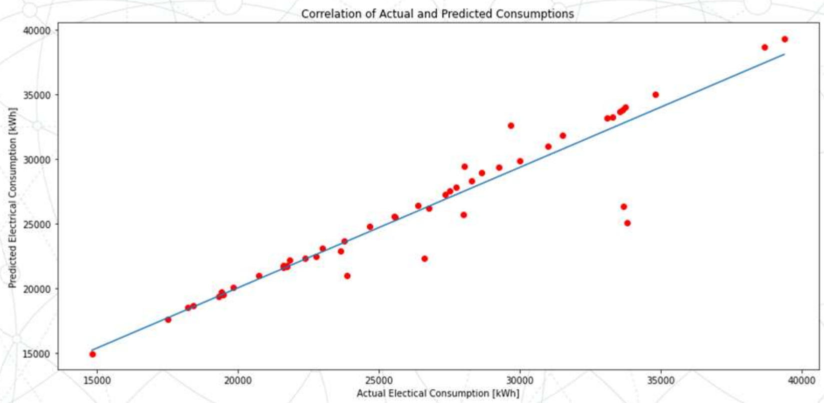
Further tuning is done to have a well-generalized model

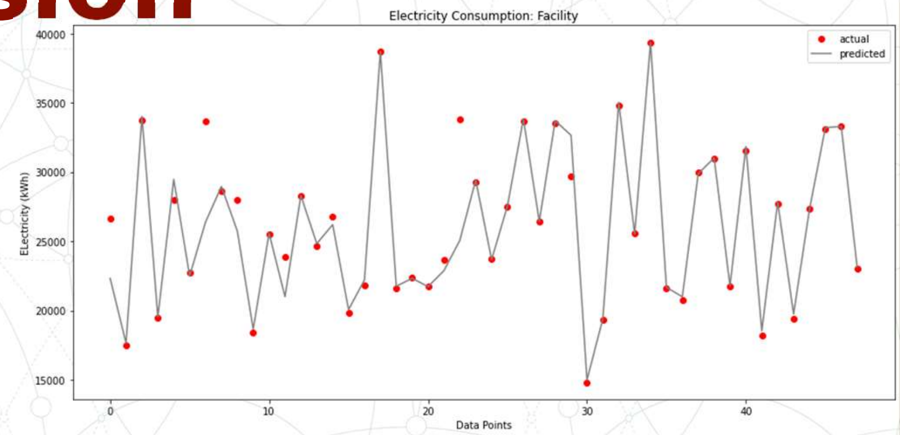
This is evaluated using a reserved test set to determine how well the model represents the FALC under pandemic operations

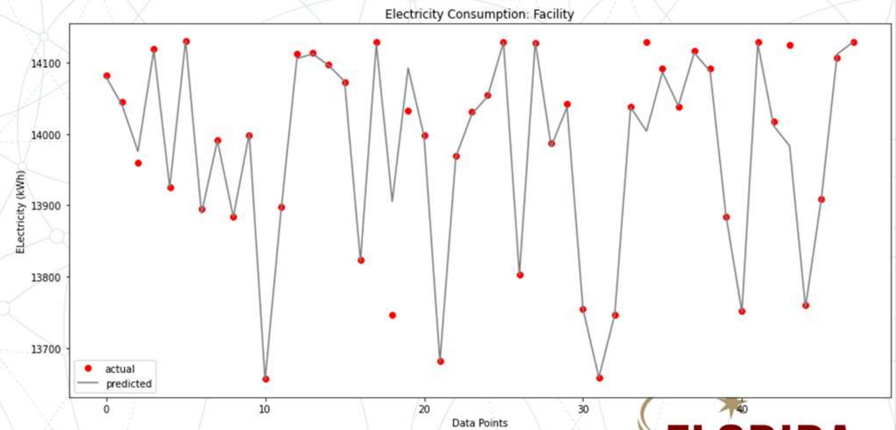


Results and Discussion

Pandemic Operations



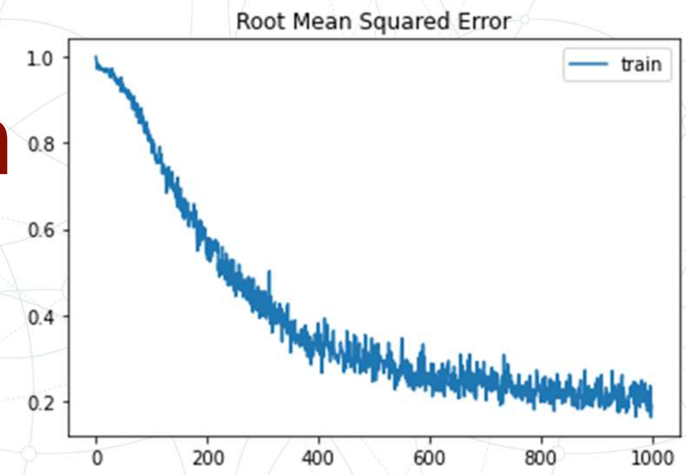




- An accurate surrogate model is developed to represent the pandemic operations energy consumption
- After observing a good correlation trend, the model is saved as a surrogate model

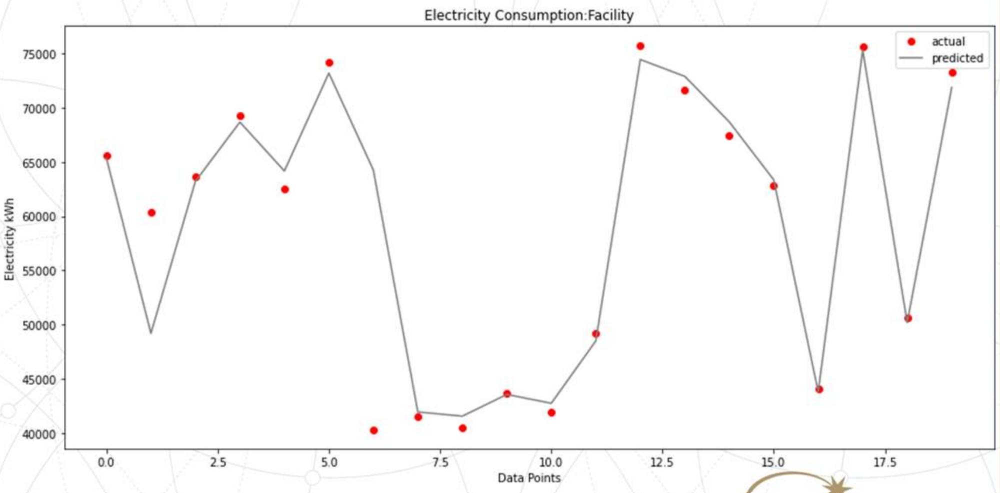
Results and Discussion

Pandemic Operations



K-Fold validation

Train Loss: RMSE	Test Loss: RMSE
0.108	1.269



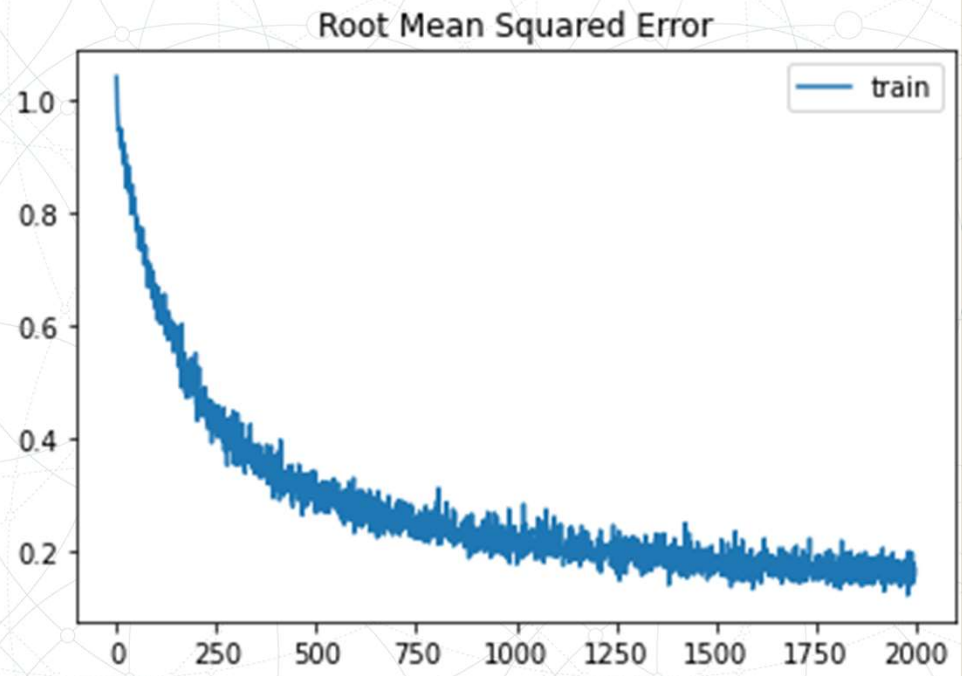
- Using K-fold cross-validation does not show a significant difference with the initial model
- The minimum loss values for training and validation sets remain close to the initial model at 0.108 RMSE for training and 1.269 for the validation set.

Results and Discussion

Regulations Operations

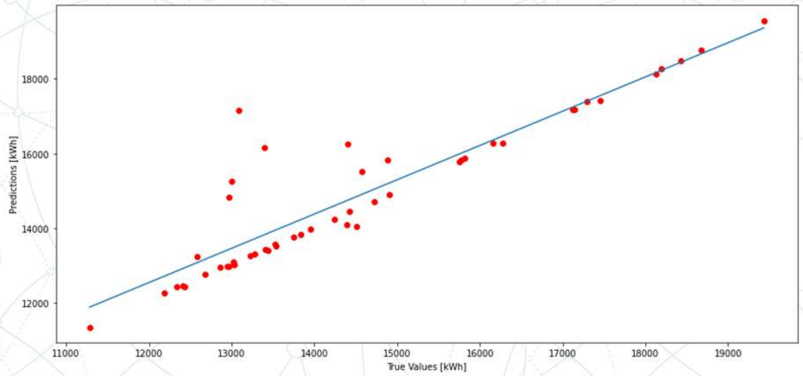
A convergence plot for the model:

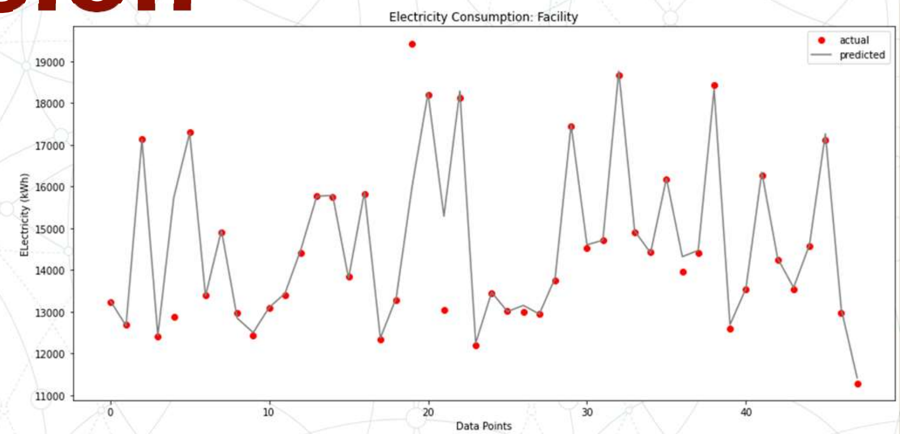
- The minimum loss values for training and validation sets remain close to the initial model at 0.079 RMSE for training and 1.070 for the validation set.
- The performance is much better than the pandemic model

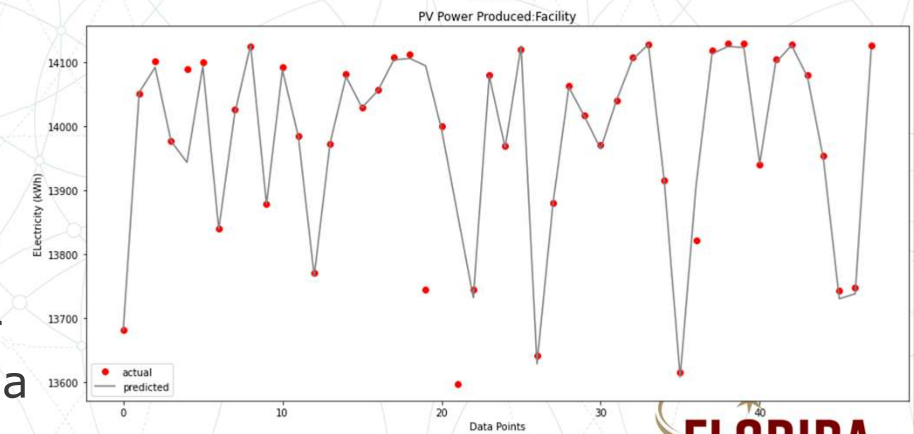


Results and Discussion

Regulations Operations







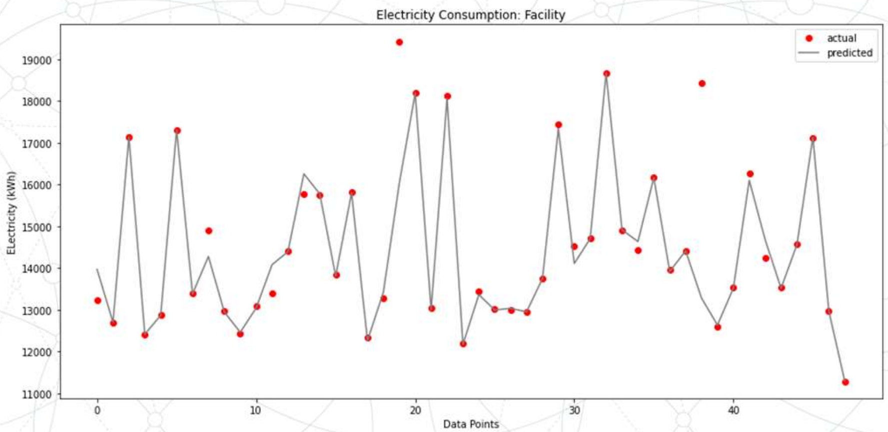
- The correlation graph plotted illustrates the similarity between NN model predicted data and simulated data from the test set

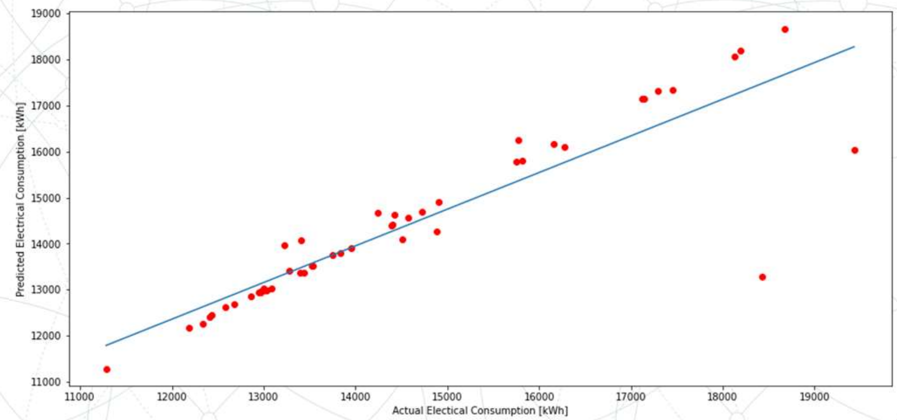
- Most of the data points show a perfect linear relation, indicated by the fit line, apart from a few data points away from the best fit line

FLORIDA TECH

Results and Discussion

Regulations Operations **K-Fold validation**





- Using K-fold cross-validation does not show a significant difference with the initial model

- The K validation shows an even better performance and a better correlation
- The model at this point was saved and used as a surrogate model

Surrogate Model Optimization

Genetic and Bayesian Optimization Algorithms

Research question 1 & 2

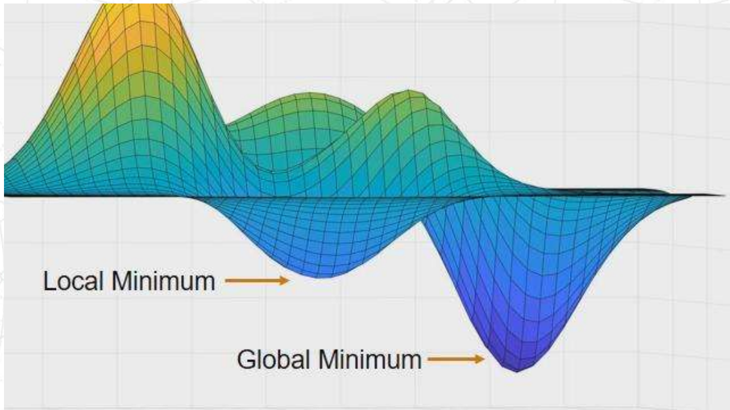
Genetic Algorithm

Objective function: neural network surrogate model developed to represent the regular building operations. The saved model is the objective function of the GA: it takes inputs of 8 variables and outputs 3 variables.

One output variable is minimized: energy consumption

Optimization variables: Roof and Wall R values, HVAC Cop, Cooling setpoint, Window U value and SHGC, and light power

GA parameters: mutation rate, crossover rate, number of generations, selection probability,

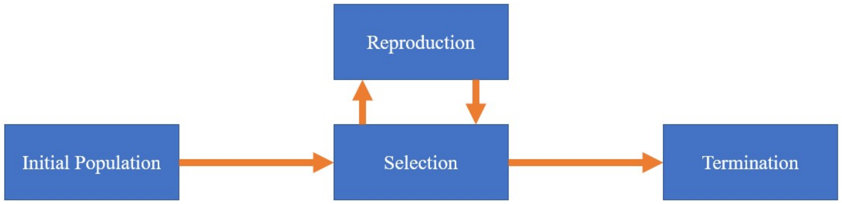


https://www.mathworks.com/videos/surrogate-optimization-1535996187824.html

FLORIDA TECH

Algorithm flowchart

1. Initialization
2. **Select** parents & **crossover**
3. **Mutate** offspring
4. Merge main population and offspring
5. Evaluate, sort and **select**
6. Back to step 2 if needed otherwise terminate



GA Parameters

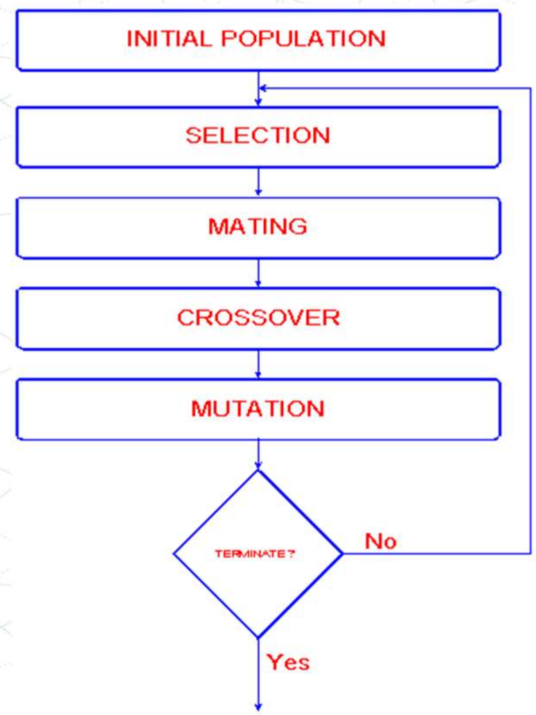
Selection is a strategy for selecting individuals from the existing population to breed a new generation. There are numerous selection methods, such as roulette wheel selection, tournament selection, ranking selection

Crossover is a recombination operation that two selected parents are exchanged to produce two new design solutions.

Mutation is used for maintain genetic diversity by altering one or more gene values in a parent individual.

Termination is a major part for the determination of an appropriate point in time to terminate the search

Genetic Algorithm for Optimization



https://www.cs.ucdavis.edu/~vemuri/classes/ecs271/Genetic%20Algorithms%20Short%20Tutorial.htm

Search Space

Constrained Variables

A higher crossover rate may lead to premature convergence of GA, yet a higher mutation rate may result in the loss of good solutions

- The Genetic algorithm is constrained to the search space provided for each individual variable
- The search space is specific to the input variables for the surrogate model
- The GA minimizes the function representing the energy consumption only however can be expanded to be a multi objective optimization

Optimization Search Space

$$\min f(x) = min\ of\ surrogate\ model$$

$$constained\ to$$

$$10 \leq x_1 = Wall\ R\ Value \geq 25$$

$$25 \leq x_2 = roof\ R\ Value \geq 45$$

$$2 \leq x_3 = HVAC\ COP \geq 5.5$$

$$1 \leq x_4 = Window\ U\ value \geq 7$$

$$0.1 \leq x_5 = Window\ SHGC \geq 1$$

$$21 \leq x_6 = Cool\ setpoint \geq 26$$

$$2 \leq x_7 = Light\ power\ density \geq 15$$

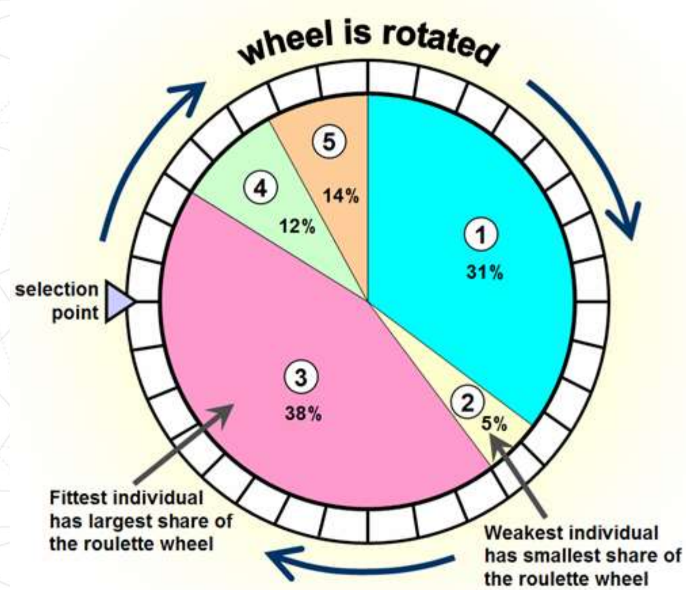
FLORIDA TECH

Introduction

Selection is very important with GA: using the **Roulette wheel selection** ensures that from the population, individual designs with lower energy consumption have the highest probability of selection

Termination is done when solution meets the pre-set minimum requirement, or termination after reaching a plateau with no better results can be produced

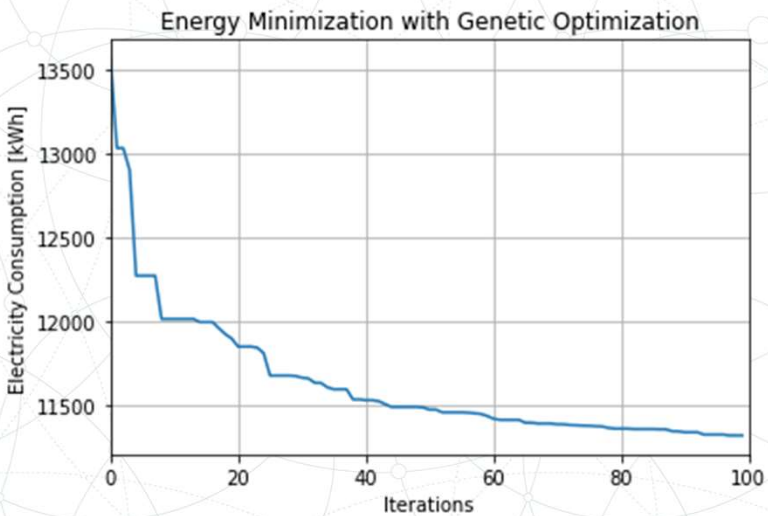
Roulette wheel selection



http://www.edc.ncl.ac.uk/highlight/rhjanuary2007g02.php/

Results and Discussion

Genetic Algorithm



Parameter	Best solution	Units
Wall R-value	24.4	W/m^{2}K
Roof R-value	28.94	m^{2}K /W
HVAC Cooling COP	4.33	-
Window U value	4.75	W/m^{2}K
Window SHGC	0.40	-
Cool Setpoint	25.6	°C
Light power density	2.01	W/m^{2}K
Lowest Annual Energy Consumption	11,300.3	kWh

- The GA minimizes the energy consumption of the regular BEM from 14MWh per year to 11.3MWh per year.
- A significant improvement in terms of energy consumption.
- Design model with more efficient design variables.

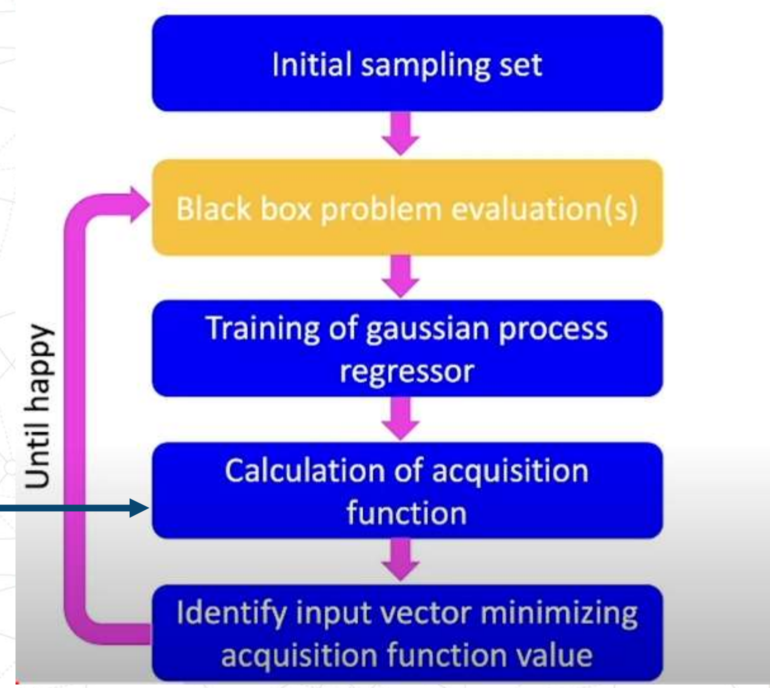
Bayesian Algorithm

We are interested in solving: $x^* = arg\min_x f(x)$

Constrained to

- **f** is a black box for which no closed form is known (nor its gradients);
- **f** is expensive to evaluate;
- and evaluations of **y =f(x)** may be noisy.

Lower Confidence Bound



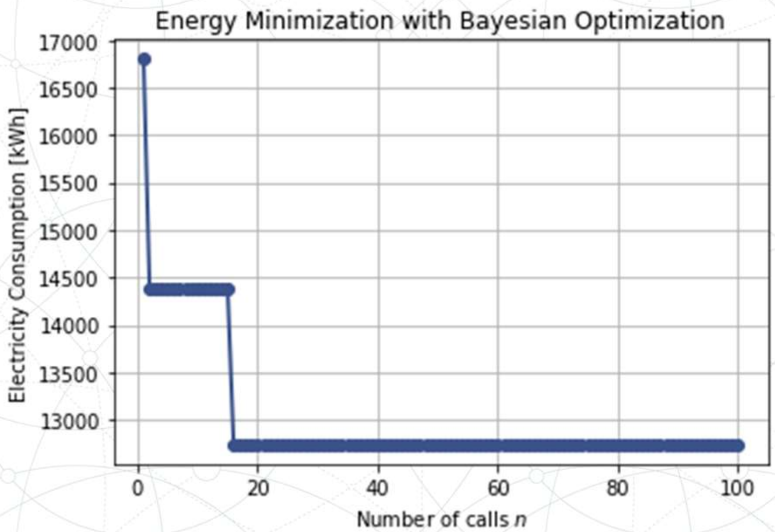
https://www.mathworks.com/videos/surrogate-optimization-1535996187824.html

Bayesian Algorithm

```
res = gp_minimize(f,                  # the function to minimize
                  [(10, 25),(25,45),(3.5,5.5),(1,7),(0.2,0.7),(23,26),(2,15),(10,30)],      # the bounds on each dimension of x
                  acq_func="EI",      # the acquisition function
                  n_calls=100,         # the number of evaluations of f
                  n_random_starts=8,  # the number of random initialization points
                  noise=0,       # the noise level (optional)
                  random_state=1234)   # the random seed
```

Results and Discussion

Bayesian Optimization



Parameter	Best solution	Units
Wall R-value	23	W/m²K
Roof R-value	27	m²K /W
HVAC Cooling COP	3.6	-
Window U value	4	W/m²K
Window SHGC	0.44	-
Cool Setpoint	26	°C
Light power density	3	W/m²K
Lowest Annual Energy Consumption	12,738.2	kWh

FLORIDA TECH

Results and Discussion

- The Bayesian optimization converges after 20 iterations and takes a shorter computational time
- The total time for all 100 iterations was 59.5s; still, a faster computational time than that taken by the GA optimization
- The Bayesian optimization can minimize the total building design energy consumption from 14MWh to 12.7MWh
- A minimization that is not as low as the one achieved by the GA but does provide lower consumption than the actual BEM energy consumption estimates

Conclusion

- A computational efficient approach has been designed and tested for building energy modeling and optimization
- Physics based model was used to generate data in order to develop the surrogate model using neural networks
- The performance of the ANN based model was successfully tested
- The surrogate model was optimized using the GA and Bayesian optimization approaches and their results were compared
- The developed approach bridges between physics-based building energy models and strong optimization tools available in python which can allow achieving global optimization

Future Considerations

- Improve the design of experiment which involves having an experiment that provides sufficient training data for the neural networks

- Investigating techniques for sampling and training the surrogate model to decrease prediction errors and

- Establish more design variables and the target output to help the optimization algorithm find practical optimum designs

- Additionally, the GA should be improved to allow multi-objective optimization

FLORIDA TECH

Acknowledgement

- My advisor, Dr. H. Najafi
- Committee members: Dr. Fabregas, & Dr. Nguyen, Dr. White
- Department of Mechanical and Civil Engineering at Florida Tech
- Florida Tech Evan's library
- Folliard Alumni Center Staff
- Florida Tech student's admin. assistant: Mrs. J. Nessmith, Mrs. A. Harris, Mrs. Susan Allison

Thank you.

FLORIDA TECH
FLORIDA'S **STEM** UNIVERSITY™

Weather Based Model

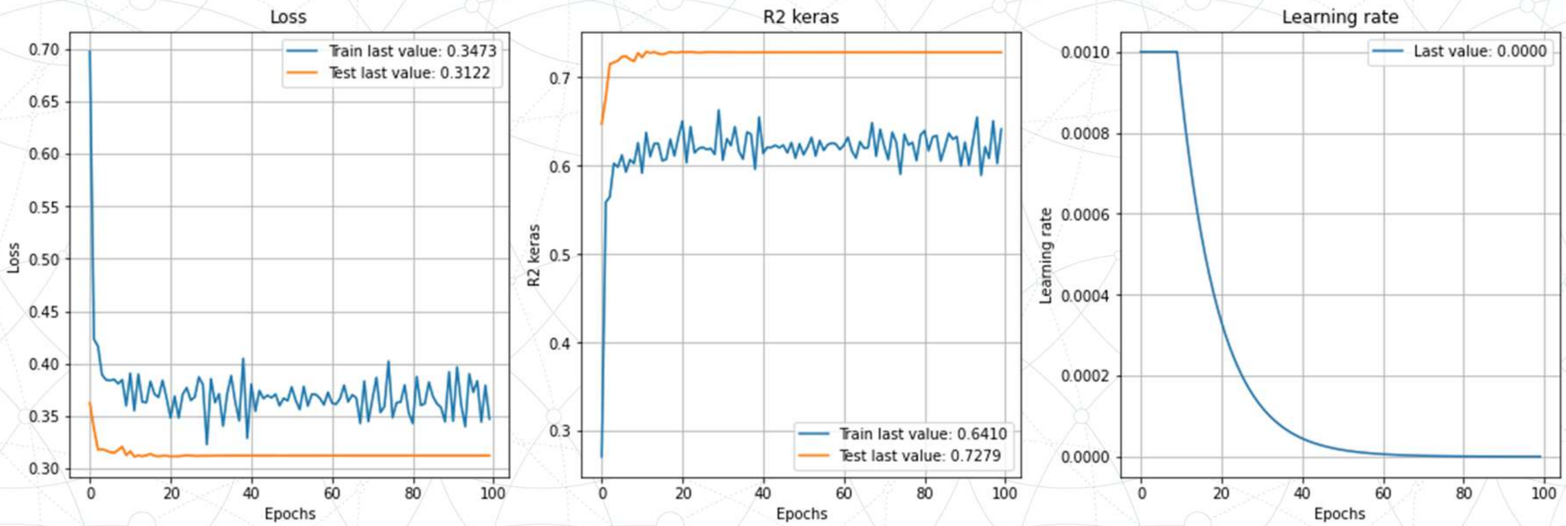
This project is done based on simulation data obtained from EnergyPlus

1. 7 weather features
2. 1 output feature for building energy consumption
3. 1 building occupancy feature
4. 2190 datapoints equivalent to daily time steps on 5 years simulation.

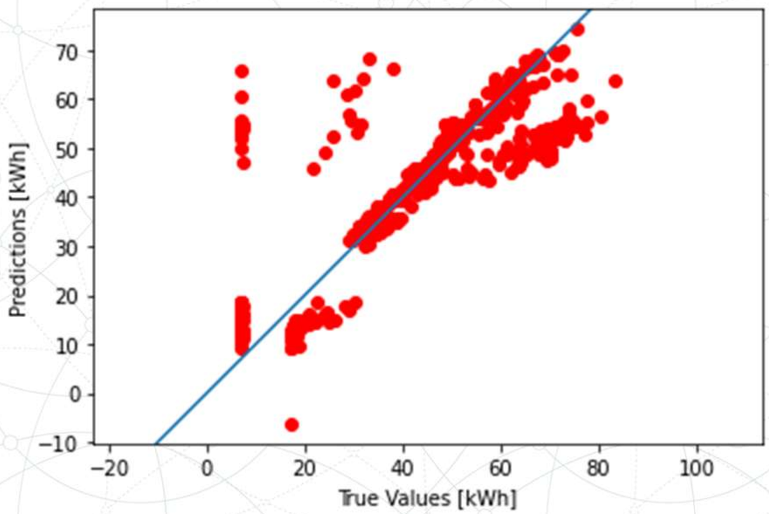
1. outdoor temperatures,
2. relative humidity,
3. rain status,
4. ground temperatures,
5. Windspeed

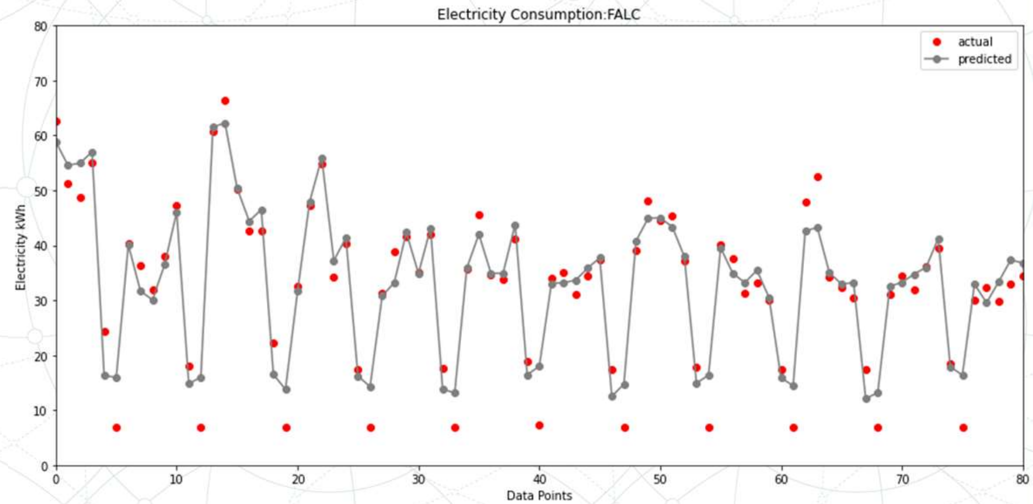
Weather Based Model

Train Loss	Train R^2 Coeff	Test Loss	Test R^2 Coeff
0.350	0.697	0.306	0.74



Weather Based Model





Train Loss	Train R^2 Coeff	Test Loss	Test R^2 Coeff
0.350	0.697	0.306	0.74

Best NN model

Best Model

- 10 layers
- 8 dense layers
- Adam optimizer
- HeNormal Initialization
- Adaptative Learning

Results of best model

Training

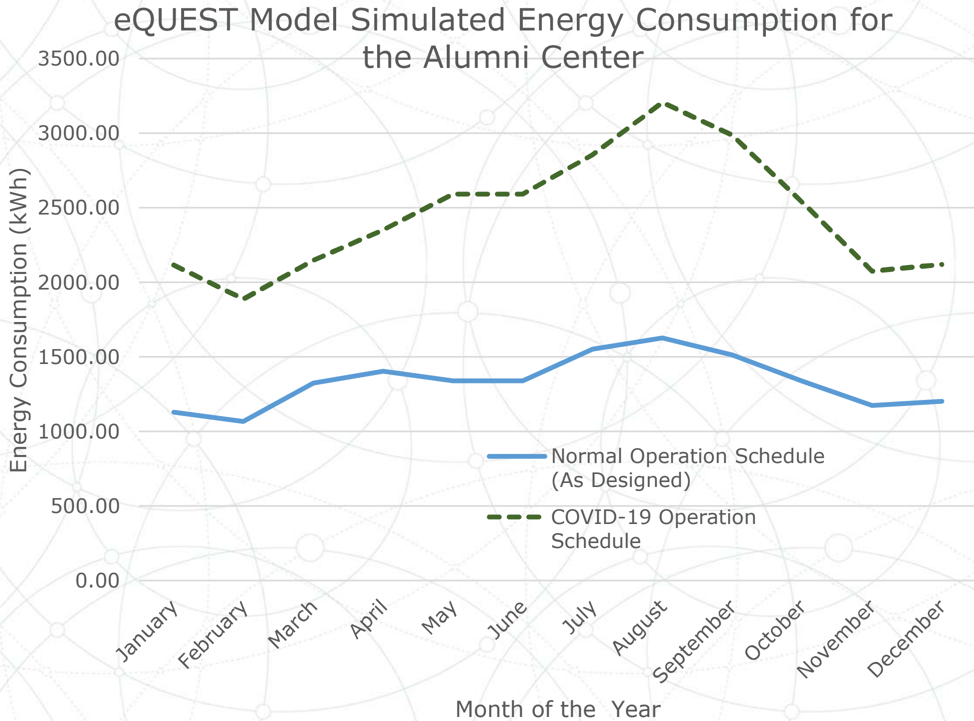
- MSE loss: 0.350 higher than Adam optimizer
- R2 coefficient: 94%

Test

- MSE loss: 0.306
- R2 Coeff: 91%

FLORIDA TECH

Weather Based Model



	Designed Use		Pandemic use	
	Ventilation Auto	Ventilation On	Ventilation Auto	Ventilation On
Designed Annual energy consumption (kWh/year)	16,227	28,749	17,986	29,969
Energy Use Intensity (kBtu/ft2/year)	15.4	27.3	17.1	28.5

FLORIDA TECH