METHODOLOGY TO DEVELOP ARTIFICIAL NEURAL NETWORK BASED CONTROL STRATEGIES FOR MULTIPLE AIR-COOLING UNITS IN A RAISED FLOOR DATA CENTER

By

VIBIN SHALOM SIMON

THESIS

Presented to the Faculty of Graduate School of
The University of Texas at Arlington in partial
fulfillment of the requirements for the Degree of

MASTER OF SCIENCE IN MECHANICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2019

ACKNOWLEDGEMENT

I express my deepest gratitude to my advisor Dr. Dereje Agonafer for providing me abundant opportunities to work on projects and his overwhelming support, guidance and enthusiasm throughout my research. Working with him has been my greatest privilege and fortune. Sincere gratitude is extended to my dissertation committee: Dr. Abdolhossein Haji-Sheikh and Dr. Andrey Beyle for their thoughtful discussions and valuable comments.

I would like to thank the National Science Foundation (NSF) Center for Energy Smart Electronics Systems (ES2) for providing me with a great research opportunity. A special thank goes to NSF I/UCRC ES2 industry mentors Mark Seymour and Kuroush Nemati for their productive Inputs and guidance in all the projects.

I would like to thank my colleagues in the EMNSPC lab for their support and making this journey a memorable one My sincere thanks to Ashwin Siddarth for providing ample amount of resources, time, support and patience. His comments were valuable in each step of the research.

I would like to thank my father Mr. Francis Simon and my mother Mrs. Srividhya Simon for their love and support throughout my graduate study. I would like to thank All EMNSPC Team members & Friends for supporting me.

ABSTRACT

METHODOLOGY TO DEVELOP ARTIFICIAL NEURAL NETWORK
BASED CONTROL STRATEGIES FOR MULTIPLE AIR-COOLING UNITS
IN A RAISED FLOOR DATA CENTER

Vibin Shalom Simon, M.S.

The University of Texas at Arlington, 2019

Supervising Professor: Dereje Agonafer

A data center consists of a hierarchy of systems with dedicated control algorithms dictating their operational states. We could say, there exists an ensemble of dynamical systems, each executing a control task, while the global objective is to drive the overall system to an optimum i.e. minimum total operational power at desired rack inlet temperatures. The dynamics of the Information Technology Equipment (ITE) workload and the cooling provisioning is non-linear in spatial and temporal parameter space. Data-driven modelling is one method to realistically model such non-linear dynamics and make predictions that are necessary for improved control design of the cooling system.

In this study, the data center non-linear dynamics are approximated by well-defined operational scenarios. Multiple ACU's are to be optimally controlled in provisioning a varying rack-level workload within the data center. CFD tool 6SigmaRoom is used to model and simulate a raised-floor data center with multiple Air-Cooling Unit (ACU)s. The input parameter space and boundary conditions to be applied to the simulations are sampled on a random basis using the Latin Hypercube Sampling technique. Artificial Neural Network (ANN) is a suitable data driven technique that has the ability to capture the non-linear relationships between the dynamic operation of ACU's setpoints and the ITE's workload. Training data is obtained from the results and observations of a large number of parametric CFD simulations. An objective function is defined. ANN model is developed to predict the parameters that are necessary in designing a data-driven control framework.

LIST OF ILLUSTRATIONS

LIST OF TABLES

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION TO DATACENTER FACILITY

Data centers are buildings built primarily for housing Information Technology Equipment (ITE) and provide power and cooling. As computing power skyrockets, modern data centers are beginning to experience higher concentrated heat loads. The demand on data center facilities is ubiquitous with the increasing requirement for record databases, internet-based services, and cloud-based services from workspace storage to on-demand media streaming. Technological advancement and price erosion enabled high growth rate of electronic packaging [1]. Recent study claims that in the year 2014, 2% of the total energy consumption in United States are due to data centers (estimated 70 billion kWh) [2]. A large data center at an industrial-scale operation uses as much electricity as a small town in the United States.

Figure. 1.1 Facebook Datacenter room

“If we could find a way to address the data center cooling design issues, we would be helping the environment and providing millions in savings for companies” said Dr. Dereje Agonafer.

Management of the data center cooling infrastructure is essential to ensure IT system reliability, up-time, and operating conditions for optimum performance. The magnitude of electrical power consumption by the cooling infrastructure varies widely, depending on operating conditions, workload, data center design and their age. Cooling power consumption in the region of 30 – 50 % of total facility power have been reported [3]. Effective air distribution and provisioning of ITE will have a significant impact on energy consumption and equipment reliability.

The energy used by a typical rack of state-of-the art servers, consuming 20 KW of power at a cost of 12 cents per kWh is more than $21,000 per year in electricity. Data centers holding hundreds of such racks constitute an energy-intensive building. Essentially, efforts to improve energy efficiency in data centers can pay big dividends [4].

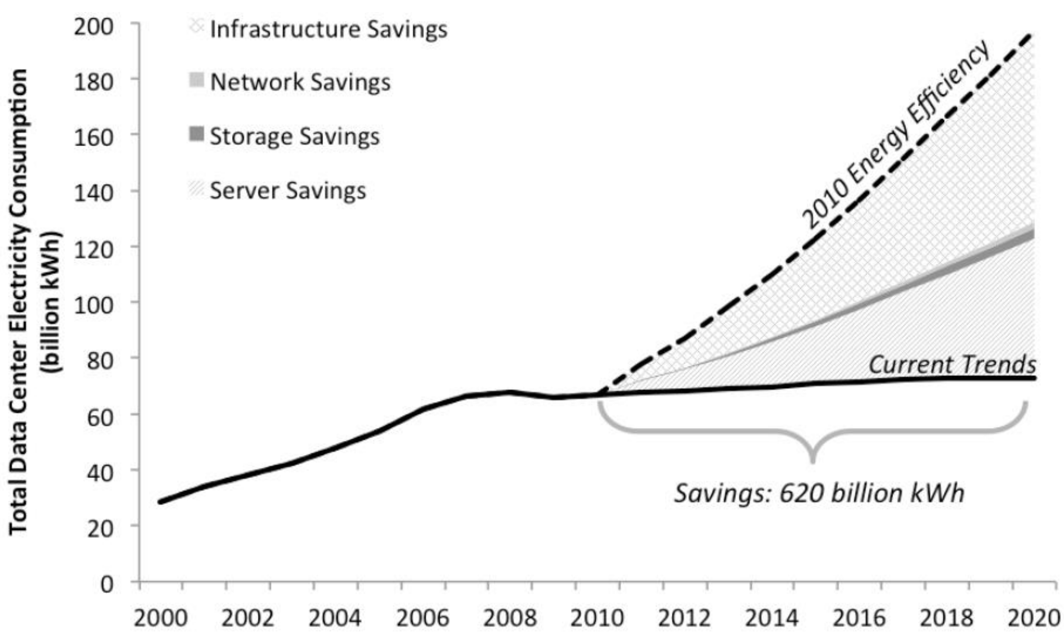

Figure 1.2 Datacenter power consumption

‘You can't manage what you don't measure’ is a common mantra that is still accurate today. However, the cooling is generated and distributed by various systems, but airflow management is a key to optimum cooling of ITE and corresponding energy consumption. Optimizing the delivery of cool air and the removal of heat generated by the ITE can involve many design and operational practices. The general goal is to minimize or eliminate inadvertent mixing between supplied air to the ITE and hot air removed from it. A simple approach for addressing this problem is to use "hot aisle and cold aisle" arrangements where racks of computers are stacked with the cold inlet sides facing each other and similarly the hot outlet sides facing each other. Hot-aisle containment is one method to maintain the cold air supplied to the racks as generated by the cooling unit so that they are evenly distributed throughout the ITE without significant change in the temperature or humidity [4].

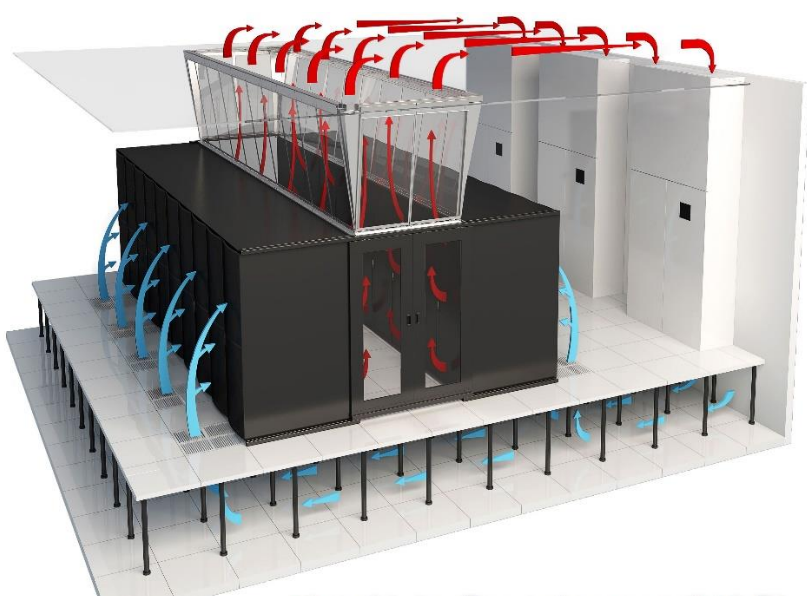

Figure 1.3 Design and layout of Datacenter room

To the author’s knowledge, multiple Computer Room Air Control (CRAC) units respond altogether increasing cooling power to provision a localized hotspot which result s in excessive cooling for other ITEs. The unnecessary cooling expenditure can be cut down by establishing a new strategy that could involve a particular CRAC unit or a combination of those to provision any localized hotspot. CRAC power dependency on operating temperature may be considered as a characteristic of the cooling strategy architecture. Knowing the fact that ITE workload fluctuation results in time-based temperature variation, the cooling unit also requires a certain time interval to respond to the scenario and make changes to address the situation.

Data center Facilities are of various types in terms of design, layout, ITE workload distribution and cooling strategies. Our data center design is chosen based on the literature survey from small-scale raised-floor data centers having indoor Computer Room Air Control (CRAC) Unit and hot aisle containment. The necessity of this design is to intentionally use the CRAC unit to provision localized hotspots created by specific ITE so that the cooling power consumption can be optimized based on the need. The Data center room does not involve Power distribution units, cables, pillars, exhaust vents and other supplementary equipment used since they are considered insignificant in this study.

To optimize the power usage of the data center we predict the configurations which leads to optimum operation and implement efficient power usage strategies. Commercial CFD software 6SigmaRoom provided by Future Facilities is used to model and simulate the operation of a data center. The main drawback of these softwares is that they need expertise in modeling and take ample amount of time to produce the results for steady state/transient conditions. Since data centers are dynamic, CFD cannot be used to produce real-time results to improve the power usage. Instead, machine learning tools like Artificial Neural Network

(ANN) is used to learn and mimic the behaviour of the data center operations and predicts results in short span of time thereby implementing power usage strategies in real-time scenarios. In our case the training dataset is generated by the CFD model having various configurations of cooling strategies for multiple Air-cooling units. The data driven model learns the non-linearity of physics-based systems and predicts parameters used to frame a cooling strategy. The ANN is trained until it delivers the least error such that its prediction can be validated using the existing data center facility.

After observing various configurations and types of data center we came up with a model that is predominantly used in a regular sized data center. To explain the model, we purposefully arranged the racks and Air- cooling units in such a way that provisioning the ITE is visualized and quantified for various hotspot scenarios.

The description of the data center modeled in CFD software is given below in table 1.

Raised Floor Hot Aisle Contained Data Center	
Total Room Capacity	300 KW
No. of racks per row	12
No. of Rows	3
Power density per rack	8.4 (max.)
Cooling system	3 Air cooling Units (ACU) of 120 KW max. sensible cooling

Table. 1.1 Datacenter Design & Room specifications

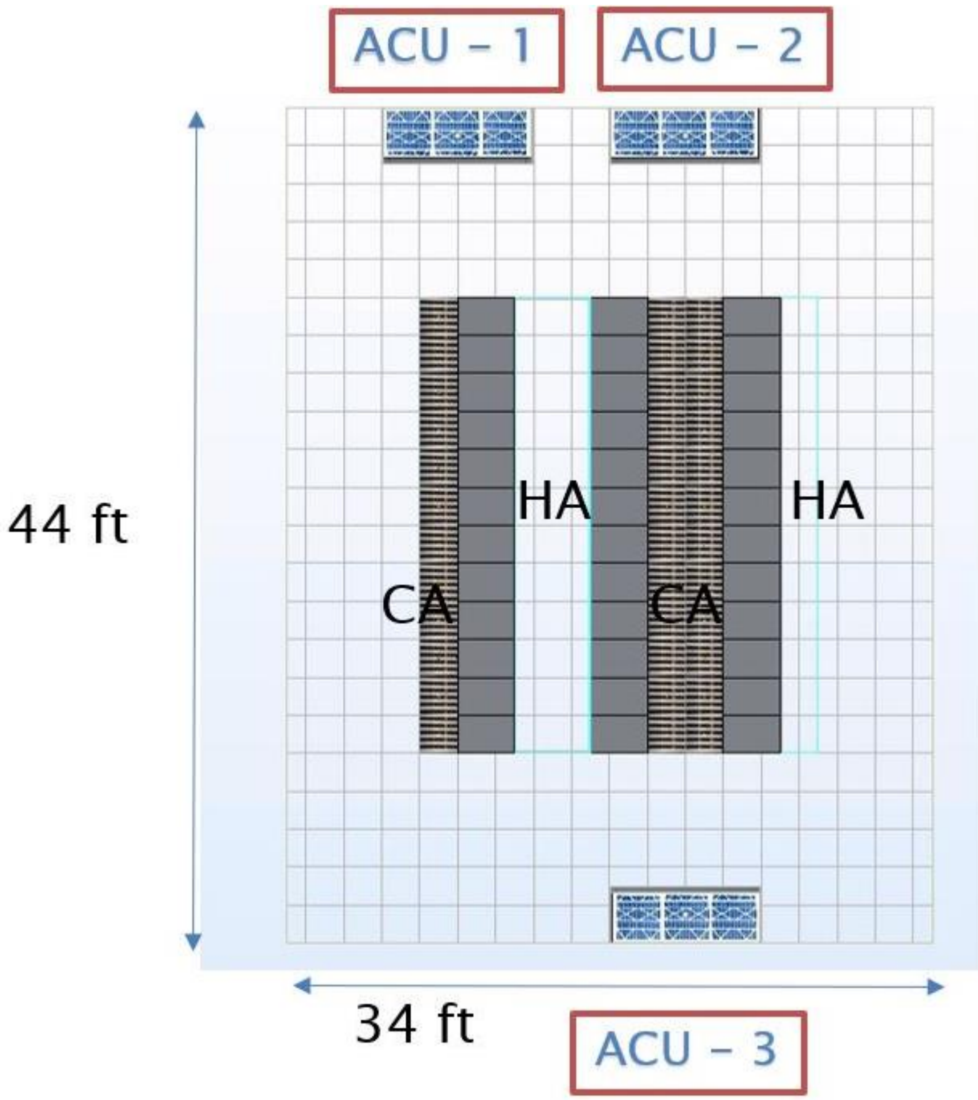

Figure 1.4 Design and layout of Datacenter room

The data center is modeled using the CFD software provided by Future Facilities, 6SigmaRoom. It is a robust and powerful tool to perform room level and component level analysis where it considers ITE as a whole component and has basic operational parameters required for room level cooling and a detailed model of ITE to perform intricate analysis when designing at component level. Various features apart from the above mentioned, helps us a lot to minimize the complexity of the data center model. In the next chapter, I will be discussing the CFD analysis, results and generation of datasets.

CHAPTER 2

COMPUTATIONAL FLUID DYNAMICS ANALYSIS & RESULTS

CFD analysis is carried out using 6SigmaRoom with the model showed in Figure.1.2. Boundary conditions, initial conditions and the types of analysis carried out is discussed in this chapter. Initially the datacenter room layout is created followed by the components namely, ceiling, raised-floor, cabinet, ITE, air-cooling unit (ACU), hot aisle containment, sensors.

The Datacenter is designed with a 2ft raised-floor design containing 36 racks, 12 racks per row provisioned by 3 ACUs. Ceiling is built at 14ft from the floor for the hot air to escape from the hot aisle containment to the return duct of the ACU. Solid obstructions are built from the hot-aisle containment to the vents in the ceiling to direct the hot air upwards. Similarly, the obstructions are built to direct the air from the ceiling to the return duct of the ACU. Floor grills of size (2 x 2) ft^2 with 50% dampers are arranged in-line in front of the rack inlet to direct the cold-air upward.

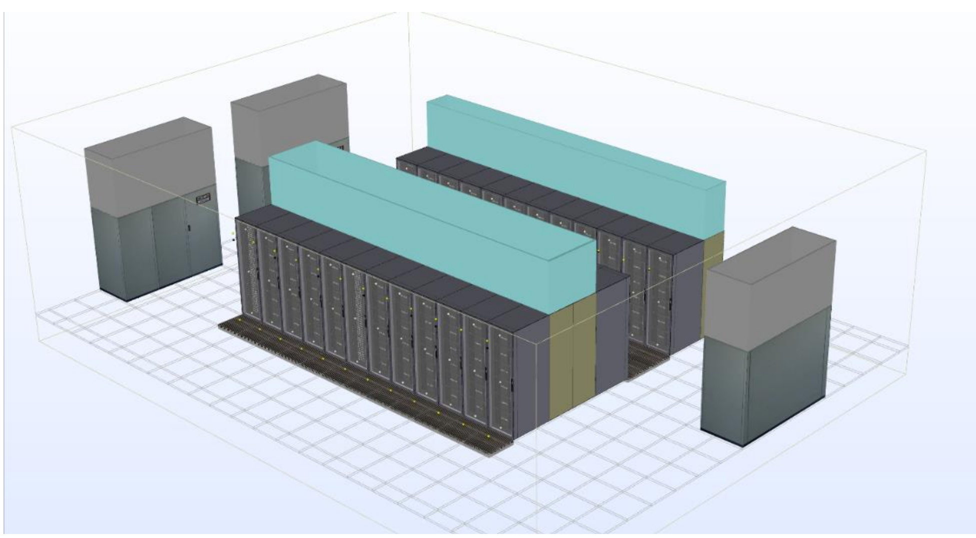

Figure 2.1 Datacenter room model

ITE DISTRIBUTION & SPECIFICATION

Each rack is filled to its capacity with 1U servers of 200 W, a total of 42 servers per rack. Typical air leakages of 5% is set to all the racks. Workload is distributed equally through all the servers, typical load of 40 W per ITE is given at idle conditions and 180 W per ITE is given at peak usage conditions. The above parameter is one of the boundary conditions used in the simulation. The server used for modeling and analysis is HP SE1120 which has an outflow pressure curve measured using experimental analysis using Air-Flow bench.[BG]

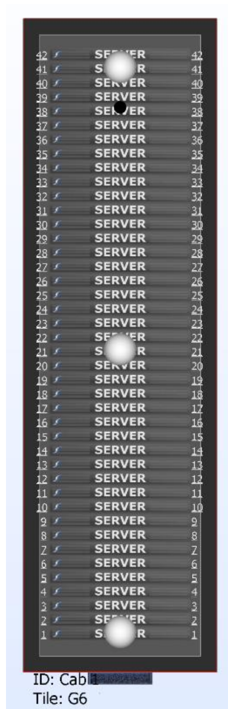

Figure 2.2 Rack with 1U servers and 3 temperature sensors (white sphere)

The outflow pressure curve denotes the pressure difference across the ITE during its operation at various modes or workloads. It has an indirect relation with the cooling system performance. ITE power is time dependent and is set to a range of values for different scenarios.

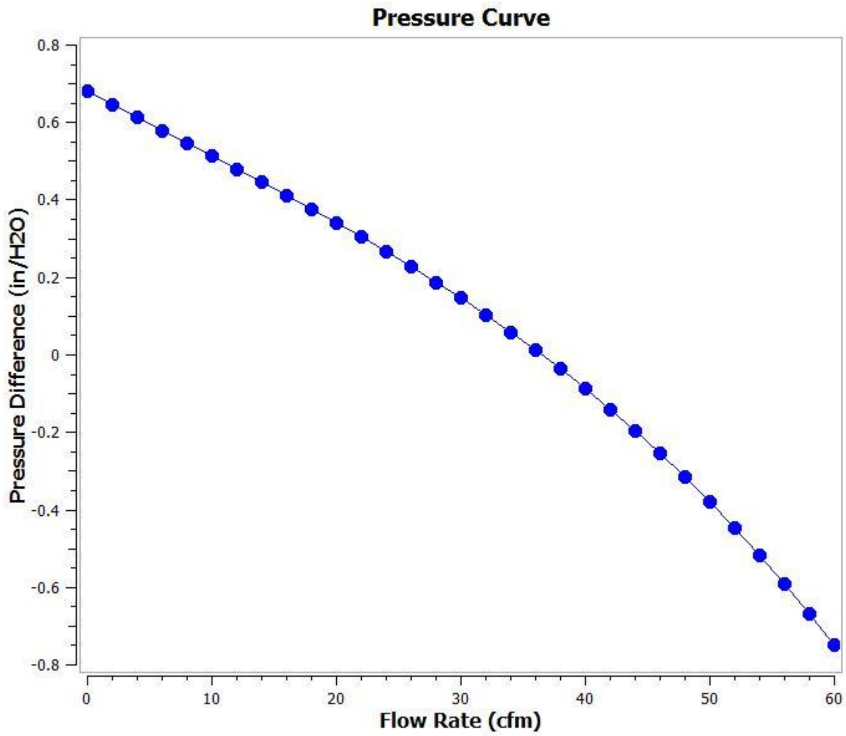

Figure 2.3 Outflow Pressure Difference vs Air flow rate

AIR-COOLING UNIT CONTROL STRATEGIES

ACU uses chilled water-cooling system where the primary coolant is the water supplied from the chiller. The reference air i.e. Return air from the ITE is passed through the cooling coils to cool down to the required temperature. Supply temperature and flowrate variation is determined based on the thermal energy consumption equation embedded in the CFD software.

Q = m*Cp*dT

Q – Energy consumed by the Air-cooling unit

m – mass flowrate of the air

Cp – specific heat capacity of the air

dT – change in temperature of the air (T_{return} – T_{supply})

Weighted average of ITE outlet temperature sensor values based on % of influence is taken as T_{return} for the power consumed by the ACU. Supply temperature setpoint is set at 16^{o}C such that the temperature rise depends on the given ITE power. The blower speed for the ACU fans are varied for every iteration for capturing different scenarios for same boundary conditions.

CRAC unit dimensions and specifications are modeled according to Liebert CW 114, ACU built by Vertiv cooling technologies. Major ACU design parameters are listed below:

Total sensible cooling capacity: 114 KW

Max. coolant flow rate: 6 Gallons/minute

Supply air flowrate range: 0 – 17,300 CFM

Figure 2.5 Liebert CW-114 Air-Cooling Unit

Two CFD analysis were carried out, one to visualize the zone of influence of the CRAC units in provisioning the ITEs and the other to calculate the power consumed by the CRAC units in various scenarios.

ZONE OF INFLUENCE ANALYSIS

ACUs supply air to the room through underfloor plenum and reaches the racks in a random fashion. To visualize and understand the influence of an ACU over the rack we try to simulate a steady state analysis with set boundary conditions given below.

Boundary conditions:

ACU Blower Speed: 90 % (15,570 CFM)

ACU Supply temperature: 15^0C

ITE Power/server: 180 W

Once the spatial locations of the racks are found having maximum influence (75–100%) of every ACU those racks are kept as targets. Power given to the targeted ITE is ramped up for a certain interval of time and back to idle condition. To address the increased load on ITE the corresponding ACU responds based on the change in outlet temperature sensor. Since all the ACUs respond together for minimal change in the workload, we forcibly allow only one ACU to respond to the hotspot.

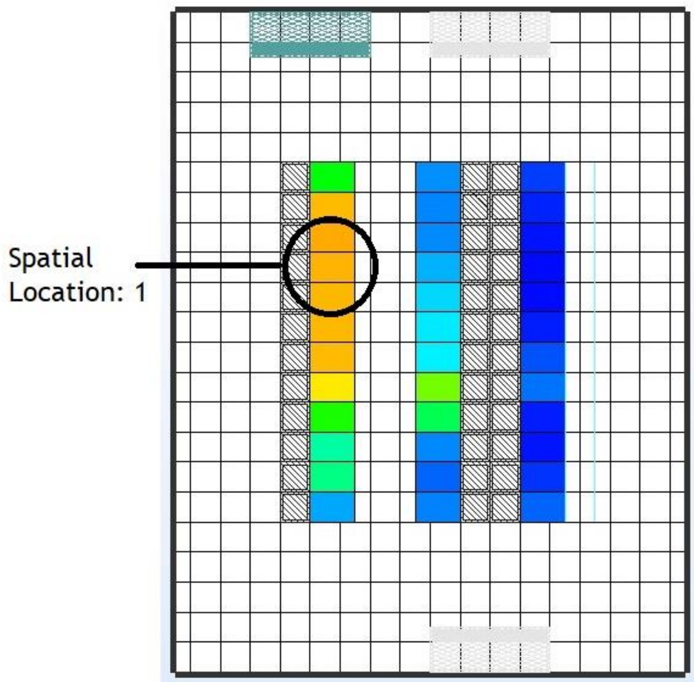

Figure 2.6 Zone of Influence of ACU-1

From steady state analysis, ACU-1 has the highest influence on the racks at spatial location-1. The racks at spatial location-1 is given increased workload for a certain interval of time and ACU-1 is forced to respond to the scenario while the other ACUs are constantly working at providing cooling to other racks. By doing so, we can find the energy consumed by ACU-1 to provision the targeted rack.

Similarly, the model is analyzed for a set of scenarios according to the corresponding influence of ACUs. To address the variability several combinations of hotspot scenarios were created to generate training datasets for the neural network model. The following two figures show the influence of ACU-2 and ACU-3 on the respective racks.

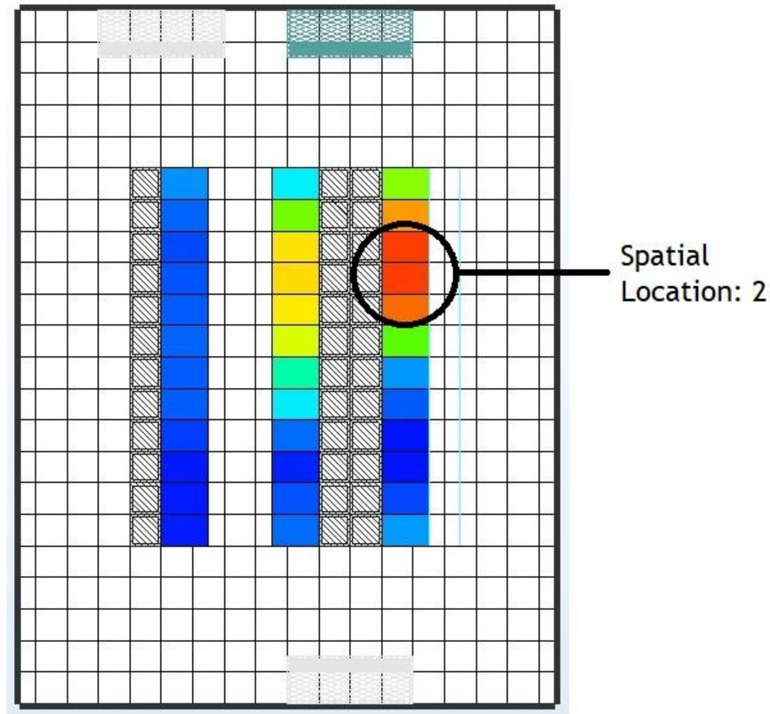

Figure 2.7 Zone of influence of ACU-2

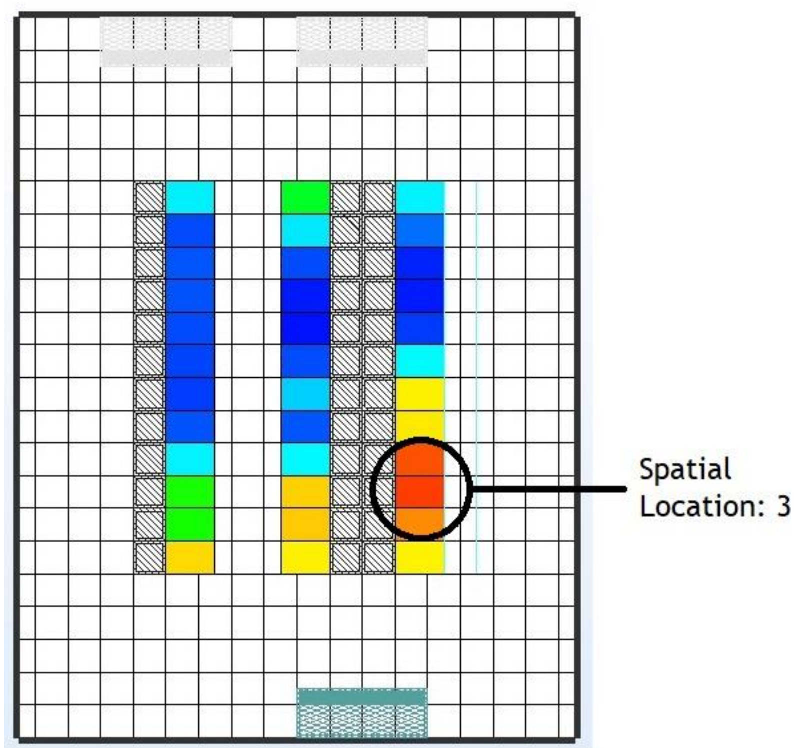

Figure 2.8 Zone of influence of ACU-2

Ramping up the IT load and using temperature dependent control for the ACU is one such strategy modeled. Datasets are generated by collecting data from a set of simulations run using PAC study in 6SigmaRoom.

DESIGN OF EXPERIMENTS

The objective function of this study is to capture the non-linearity of the physical phenomenon in the data center. A methodology for choosing the input parameter space out of 'N' number of measurable parameters is carried out using Latin hypercube sampling(LHS), a statistical technique where the domain of interest is filled with samples portraying the variability shown in original data.

The multi-dimensional parameter space should be space-filling and non-collapsing [5]. Computational simulations are deterministic in nature and are not prone to the uncertainties inherent in experimental methods; hence, it is critical that the input parameter space is determined using a random sampling technique to avoid any bias and introduce required variability in the training data for neural networks. Latin hypercube sampling (LHS) [6], a statistical method for generating a near-random sample of parameter values from a multidimensional distribution, ensures that the ensemble of random numbers is representative of the real variability.

INPUT PARAMETER SPACE

Boundary conditions

ACUs: Fixed blower speed for all ACUs depending on the ITE load (50%,70%,90%)

IT Equipment:

Set the range for IT Load Factor: 50-90%

Power ratio (CFM/W) based on max. temperature rise of 10^0 C: 0.30 CFM/W

Rack: Location of the targeted rack (1,2,3) & ITE load step (2 KW)

Using Latin hypercube sampling from a range of parameters, the input parameter space is generated to provide the maximum variability in the CFD simulation. The CFD model is run for different combinations of inputs to generate training datasets. Python is used to generate the type of parameter, bounds and constraints to have a space filling design. Based on Latin hypercube sampling from known parameter values and resolutions, the input parameter space is defined for different combinations for 27 CFD simulations.

Inputs for the CFD simulations given are: Time based ITE Load, Spatial Location of targeted ITE, power ratio set for the ITE.

Time varying Outputs from the CFD simulations obtained are cooling power consumption of all three ACUs in different scenarios, ACU blower speed.

Input Variables	Bounds	Constraints
ACU Blower speed	50 – 90 %	Interval of 20 (50%, 70%,90%)
Total ITE Workload / rack	4 – 8 KW	Interval of 2 (4,6,8)
Spatial Location	1 - 3	Interval of 1 (1,2,3)
Temperature rise of the IT load	5 – 10^0C	Interval of 1 (5,6,7,…10)

Table 2.1 Input parameter space

The two functions that govern the control parameters are as follows:

1. ACU Blower Speed = f(ACU number , Spatial Location, Outlet Temperature, IT Load)

2. ACU Cooling power = f(ACU number , Spatial Location, Supply temperature, IT Load)

These two equations are captured mathematically by neural networks to predict the ACU number, blower speed and cooling power consumption.

CHAPTER 3

INTRODUCTION TO MACHINE LEARNING AND ITS TOOLS

Machine learning is well-suited for the DC environment given the complexity of plant operations and the abundance of existing monitoring data. The modern largescale DC has a wide variety of mechanical and electrical equipment, along with their associated setpoints and control schemes. The interactions between these systems and various feedback loops make it difficult to accurately predict DC efficiency using traditional engineering formulas [7]. For example, a simple change to the cold aisle temperature setpoint will produce load variations in the cooling infrastructure (chillers, cooling towers, heat exchangers, pumps, etc.), which in turn cause nonlinear changes in equipment efficiency. Ambient weather conditions and equipment controls will also impact the resulting DC efficiency. A machine learning approach leverages the plethora of CFD simulation data to develop a mathematical model that understands the relationships between operational parameters and the holistic energy efficiency.

The sheer number of possible equipment combinations and their setpoint values makes it difficult to determine where the optimal efficiency lies [10]. Using standard formulas for predictive modeling often produces large errors because they fail to capture such complex interdependencies. Data driven models are the best methods that can completely represent a non-linear physics-based scenario in mathematical form so that we can train neural network models to learn and predict the desired parameters to meet the needs.

A neural network is selected as the mathematical framework for training DC energy efficiency models. Neural networks are a class of machine learning algorithms that mimic cognitive behavior via interactions between artificial neurons [8]. They are advantageous for modeling intricate systems because neural networks do not require the user to predefine the feature interactions in the model, which assumes relationships within the data. Instead, the neural network searches for patterns and interactions between features to automatically generate a best-fit model.

ARTIFICIAL NEURAL NETWORKS

ANN is a machine learning tool which works on the same principle as the neurons in the human brain does. It acts as a black box model and learns the input & output data using mathematical tools and uses a percentage of the same dataset to train, validate and test. ANN comprises of a large network of neurons or nodes similar to the neurons in the brain which stores the intermediate values to calculate data which are not present in the training dataset.

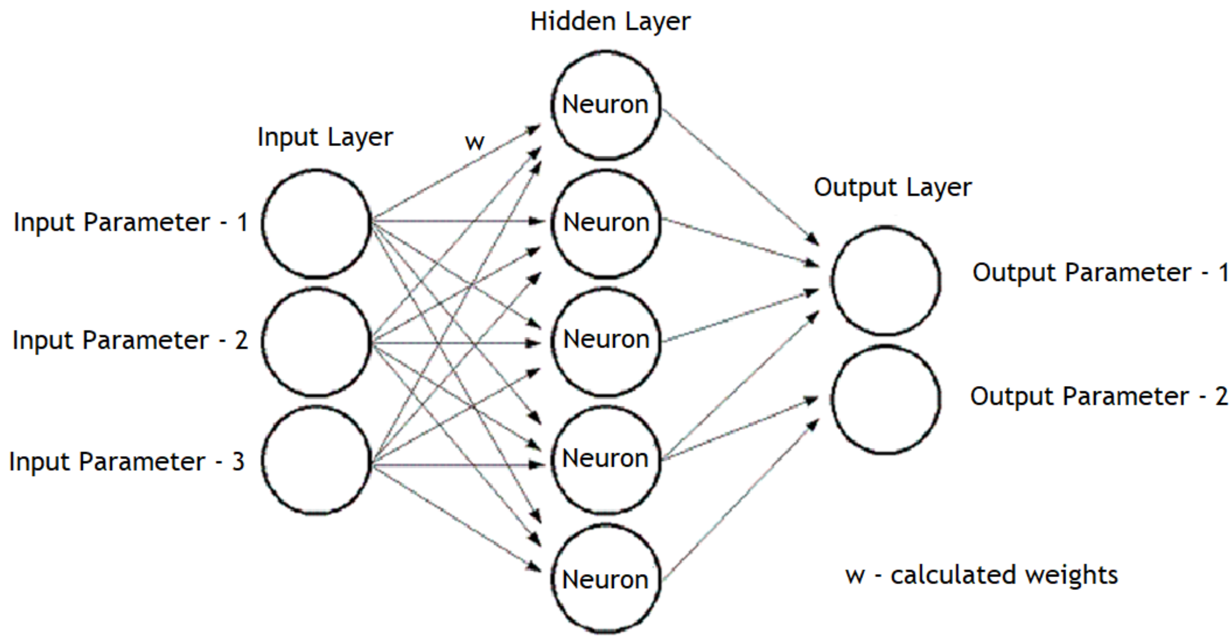

Figure 3.1 Structure of ANN

ANN ALGORITHM

Artificial neural network model training can be broken down into five steps:

1. Initialize random input parameters
2. Implement the forward propagation algorithm
3. Compute the cost function $J(\theta)$
4. Implement the back-propagation algorithm
5. Repeat steps 2-4 for several iterations until the test data produces least error [8].

RANDOM INITIALIZATION

Random initialization is the process of randomly assigning θ values between [1,1] before starting model training. The inputs into each successive layer in the neural network would then be identical, since they are multiplied by θ . Furthermore, since the error is propagated backwards from the output layer through the hidden layers, any changes to the model parameters would also be identical [8]. We therefore randomly initialize θ with values between [1,1] to avoid the formation of unstable equilibriums [8].

FORWARD PROPAGATION

Forward propagation refers to the calculation of successive layers, since the value of each layer depends upon the model parameters and layers before it. The purpose of the activation function ***f(x)*** is to mimic biological neuron firing within a network by mapping the nodal input values to an output within the range (0, 1).

It is given by the sigmoidal logistic function:

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

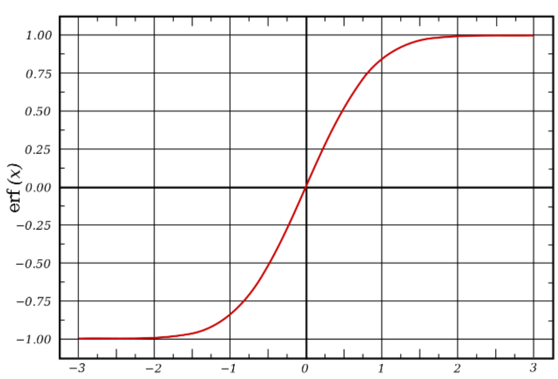

Figure 3.2 Sigmoid Activation Function

It calculates the errors between calculated output and sample output data and uses this to create an adjustment to the weights, thus implementing a form of gradient descent.

COMPUTING COST FUNCTION

The cost function ***J***(θ) serves as the quantity to be reduced with each iteration during model training. It is typically expressed as the square of the error between the predicted and actual outputs. For linear regression problems, the cost function can be expressed as:

$$J(\theta) = \frac{1}{2m}[\sum_{i=1}^{m} (h_\theta(x^i) - y^i)^2 + \lambda \sum_{i=1}^{L-1} \sum_{j=1}^{n} \theta_{ij}^2]$$

where $\boldsymbol{h}_\theta(\boldsymbol{x})$ is the predicted output, ***y*** is the actual data corresponding output variable of interest, m is the number of training examples per feature, ***L*** is the number of layers, and n is the number of nodes [8]. The regularization parameter λ controls the tradeoff between model accuracy and overfitting [9].It calculates the errors between calculated output and sample output data and uses this to create an adjustment to the weights, thus implementing a form of gradient descent

BACK PROPAGATION

After computing the cost function ***J***(θ) , the error term δ is propagated backwards through each layer to refine the values of θ. The error for the output layer is defined as the difference between the calculated output $\boldsymbol{h}_\theta(\boldsymbol{x})$ and the actual output ***y***

ANN TRAINING & VALIDATION

Using datasets from the CFD simulations ANN is trained to predict the desired outputs that can allow us to frame a control strategy for the provisioning the ITE running under various workload scenarios. 70% of the dataset is used for training the neural network, the remaining 30% used for cross validation and testing. Data pre-processing such as sampling and data filtration is done using Python (PyCharm by JetBrains) in conjunction with the NumPy module. MATLAB R2019a was used for neural network model training, validating, testing and post-processing.

In our study, the input parameters for the ANN are chosen in such a way that they can be measured directly from the data center facility. The probability of error is negligible. The list of input parameters for the ANN model are listed as follows:

1. ACU number
2. Spatial location of the targeted rack
3. Pressure difference across the rack (Pa)
4. Temperature at the outlet of the rack (^{0}C)
5. Rack IT Load (W)

The output parameters of the ANN model are considered in a way such that it is used to develop a control strategy that will help in regulating the cooling power of each ACU and the blower speed.

The output parameters are as follows:

1. Blower Speed (CFM) for all 3 ACUs
2. Sensible cooling load on (Power consumed – KW) all 3 ACUs

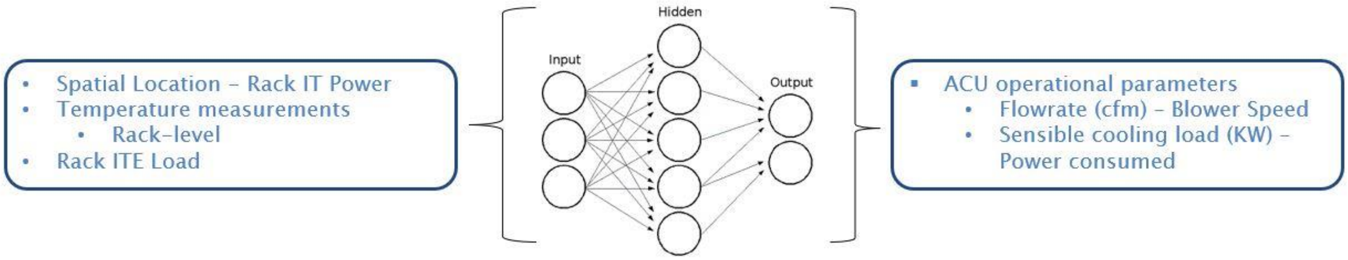

Figure 3.3 ANN model

Non-Autoregressive network with exogenous input (NARX) model

Number of neurons required to achieve minimal error in training the ANN is calculated with a set of values defined by the thumb rule. Thumb rule is stated as follows [7]:

1. The total number of neurons must be more than the summation of number of inputs and outputs to the ANN model.

   No. of Neurons > Input parameter size + Output parameter size

2. The total number of neurons must be equal to summation of 2/3rd of the inputs size and the output size

   No. of neurons = (2/3)*Input parameter size + Output parameter size

3. The total number of neurons can be less than twice the input size.

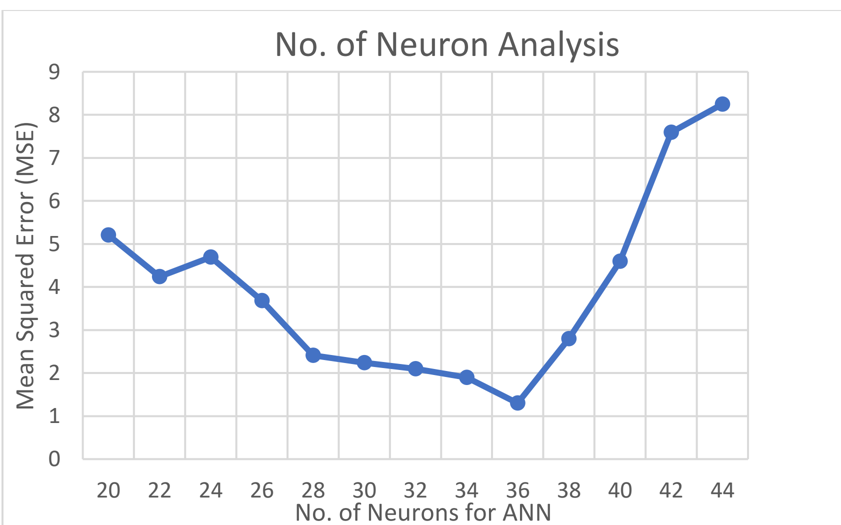

Figure 3.4 No. of Neuron Analysis

Similar to grid sensitivity analysis, number of neurons used in ANN should also be precise in order to avoid overfitting and underfitting. In our case 36 neurons in the hidden layer is optimum for the ANN to predict results with minimal error. If we have neurons more or less than 36, we may end up having the predicted results with large error values.

The model is tested from a sample data that is not in the training dataset to see the accuracy of the prediction. The Mean Squared Error (MSE) refers to the difference in original value and the predicted value and if it is low then the dataset has high resolution and vice versa. Accuracy is improved by generating training dataset having higher resolution by optimizing the grid control in CFD.

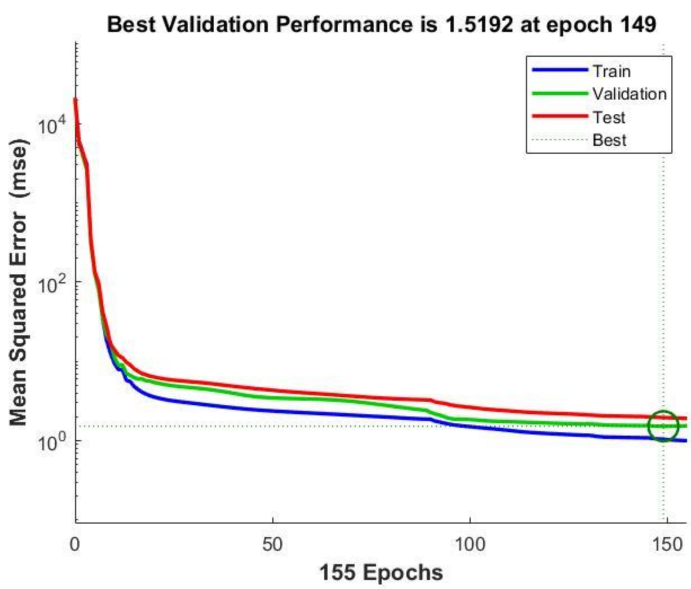

Figure 3.5 ANN Training Performance

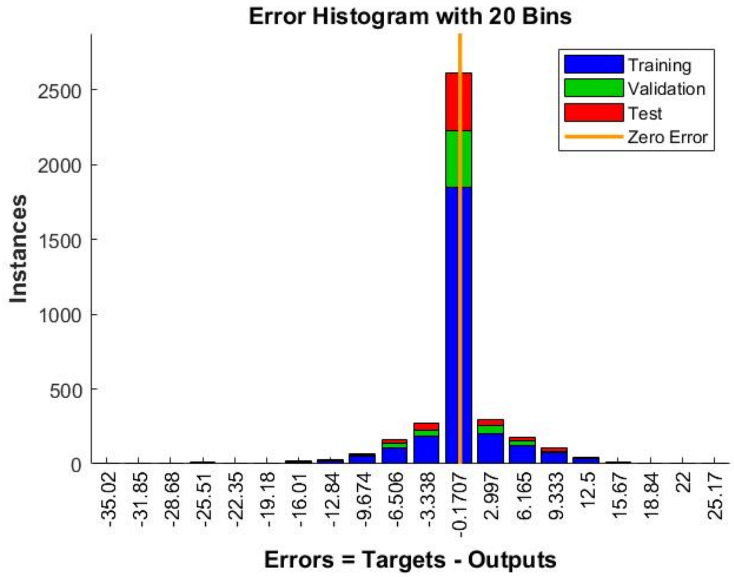

Figure 3.6 ANN Training Error

From these results we can conclude that the ANN can predict the values of ACU control parameters having MSE in the order of 10^-1.

CHAPTER 4

RESULTS AND DISCUSSION

Table 4.1 below shows the test inputs given to the ANN model to predict the ACU power consumed and flowrate to be used to resolve the hotspot

ANN Test Inputs			
Time (mins)	ITE Load (KW)	Spatial Location of targeted ITE	Temp. rise at rack-level (^{0}C)
50	6KW	3	8
26	8KW	2	10
35	4KW	2	7

Table 4.1 ANN Test Input parameters

Table 4.2, 4.3, 4.4 shows the predicted results for the given inputs. It shows which ACU to be operated and in what operating conditions.

ANN Test Output 1		
	Power (KW)	Flowrate (cfm)
Time (mins)	50	50
ACU 1	37.0409	8254
ACU 2	46.7322	8356
ACU 3	51.1894	8449

Table 4.2 ANN Test Output parameters (First Case)

ANN Test Output 2		
	Power (KW)	Flowrate (cfm)
Time (mins)	26	26
ACU 1	46.8162	11355
ACU 2	110.6364	11669
ACU 3	48.6988	11568

Table 4.3 ANN Test Output parameters (Second Case)

ANN Test Output 3		
	Power (KW)	Flowrate (cfm)
Time (mins)	50	50
ACU 1	33.5233	8135
ACU 2	72.4583	8796
ACU 3	49.0487	8678

Table 4.4 ANN Test Output parameters (Third Case)

In the first case, the ITE load of 6 KW is given at spatial location at 3 which creates a hotspot. To address this scenario, power consumed by ACU-3 increases while other two ACUs remain in normal operation. Similarly, power consumed by ACU-2 increases in other two cases while the other two remains in normal operation. This provisioning strategy will help in saving a lot of energy in large data centers.

CHAPTER 5

CONCLUSION

The predicted values by the ANN can be used to design a framework that can allow the operation of multiple CRAC units operating in tandem to provision the ITE loads. ANN can be used to frame a control strategy based on the hotspot in a typical raised floor data center with chilled water-cooling system. This ANN model can be implemented in real time live data center provided; it is trained based on in-house sensor values. This research shows that non-linearity in the behavior of an operating data center can be captured and ACU operational parameters can be altered based on ANN prediction. Cooling strategy can be essentially based on the ANN predicted results thereby reducing power consumed by the cooling units.

FUTURE WORK

High resolution CFD models and a greater number of hotspot scenarios can be considered to address cooling provisioning. Increasing the resolution of input parameter space to get more variability in the dataset. Cooling Unit failure analysis can be included in predicting ACU operational parameters. Step ahead prediction of parameters is one such method where we will be able to have a better response from the cooling unit and also suitable for real-time implementation. Reinforcement learning, a new technology in machine learning that can learn non-linear operations of the data center by computing the error between the actual value and the predicted value in a working environment without getting input from the user. The algorithm is designed in such a way that it learns the input and output required to solve a specific scenario on the fly.

REFERENCES

[1] Scaramella, J., 2008, "Next-Generation Power and Cooling for Blade Environments," IDC Technical Report No. 215675

[2] J. Koomey. Growth in data center electricity use 2005 to 2010. Analytics Press. Oakland, CA. 2011[Online]. Available: http://www.analyticspress.com/datacenters.html

[3] EPA. ENERGY STAR Rating for Data Centers: Frequently Asked Questions [Online]. Available: https://www.energystar.gov/ia/partners/prod_development/downloads/DataCenterFAQs.pdf?acac-cbed.

[4] Steve Greenberg, Evan Mills, and Bill Tschudi, Lawrence Berkeley National Laboratory Peter Rumsey, Rumsey Engineers Bruce Myatt, EYP Mission Critical Facilities "Best Practices for Data Centers: Lessons Learned from Benchmarking 22 Data Centers".

[5] D.C. Montgomery, "Design and Analysis of Experiments", John Wiley and Sons (2011)

[6] Stein, M., "Large Sample Properties of Simulations using Latin Hypercube Sampling", Journal of Technometrics, Vol. 29 (2), pp. 143-151, 1987

[7] Jim Gao, Google, "Machine Learning Applications for Data Center Optimization"

[8] Andrew Ng. "Neural Networks: Representation (Week 4)." Machine Learning. Retrieved from https://class.coursera.org/ml2012002.2012. Lecture.

[9] Andrew Ng. "Regularization (Week 3)." Machine Learning. Retrieved from https://class.coursera.org/ml2012002.2012. Lecture.

[10] Jayati Athavale, Yogendra Joshi, Minami Yoda, "Artificial Neural Network Based Prediction of Temperature and Flow Profile in Data Centers"