2022

Physics Informed Machine Learning Based Reduced Order Model of Unitary Equipment

Shahzad Yousaf

Craig Bradshaw

Rushikesh Kamalapurkar

Omer San

Informed Machine Learning to Develop a Reduced Order Model of Unitary Equipment

SHAHZAD YOUSAF[1*], CRAIG R. BRADSHAW[2], RUSHIKESH KAMALAPURKAR[3], OMER SAN[4]

[1,2,3,4] Center for Integrated Buildings Systems, Oklahoma State University,
Stillwater 74075,Oklahoma, United States

[1] shahzad.yousaf@okstate.edu,[2] craig.bradshaw@okstate.edu,[3]rushikesh.kamalapurkar@okstate.edu,
[4]osan@okstate.edu
* Corresponding Author

ABSTRACT

Combining machine learning tools with conventional reduced order modeling approaches produces the potential for an enormous increase in the ability to select suitable models. This paper presents a machine learning based approach for predicting the cooling capacity of a fixed speed unitary air-conditioner. Experimental data from a 10-ton Roof Top Unit (RTU) is simplified by utilizing a feature selection methodology, Elastic Net (EN), to accurately record and reduce the parameters while simultaneously preserving the physics of the system. The simplified RTU data set resulting from the EN is fed into a novel method for model order reduction employing Principal Component Analysis coupled with a supervised Artificial Neural Network (ANN). Preliminary results show that proposed technique is able to predict the equipment cooling capacity within ±2% of experimental results. Furthermore, the current work has also been compared to the recent models in the literature and has been found to be superior.

1. INTRODUCTION

Energy generation and its use in buildings is on the rise. In the US, 38% of the energy delivered to households is being used for the air conditioning purposes, a much larger share than any other end use. It has also been estimated that energy use for residential system would increase by 59% by 2050 (AEO, 2022). This level of energy use would also be accompanied by global warming. It has been noted that existing systems may be optimized without investing in costly new solutions. This will result in increased efficiency and decreased energy use. (Širokỳ et al., 2011) .In-depth studies aimed at optimizing Heating, Ventilation, and Air Conditioning (HVAC) systems are necessary for this purpose, and major research is being conducted to improve the operation and efficiency of HVAC systems and their components. Previous studies have shown that the building energy efficiency can be significantly improved if one considers the whole building as an integrated system rather than optimizing individual building components such as lighting and heating, ventilation and air conditioning (HVAC) systems (e.g., see Borggaard et al. (2009)).

The compressor plays an important role in the estimation of cooling capacity of air-conditioning equipments. Tanveer and Bradshaw (2021) proposed a model to scale the scroll compressor for different cooling capacities for using low-GWP refrigerants and reports that the model can predict the mass flow rate and power with 03% of mean absolute error (MAE) with leakage area as major contributor towards the drop in compressor performance. Belman-Flores et al. (2015) modeled compressor using both physical and artificial neural network (ANN) model for comparison purposes. The physical model predictions are in ±10% while the ANN model predictions are in range of ±1%.

Additionally, in central air conditioning systems, Z. Ma and Wang (2009) implemented pressure drop model for various water networks and offered an optimal pump control strategy resulting in energy savings.Similarly, Wang et al. (2022) presented a nonlinear predictive control strategy for chiller-air handling units and reported a significant amount of energy saving. However, it is important to note that every system offers a different behavior, and it should be programmed as per its own conditions and components. A small residential HVAC system has numerous components and offers a complex circuitry with numerous variables. The complexity further increases for large systems which are affected by many other variables such as location, orientation, type of construction and dimensions. Inclusion of such variable increases the complexity of the systems and for this reason such systems are modelled using Building Energy Models (BEM) like TRNSYS and EnergyPlus using system parameters. From a modeling perspective, such systems can be considered as highly complex multi-scale and multi-physics dynamical systems, operating under wide varieties of uncertainties and are computationally expensive.It is also worth noting that computationally expensive models may not be practical in control applications requiring speedy responses.Such control issues can be handled if systems are described in a reduced order manner. Performance of Reduced Order Models (ROM) are expected to be close to the high fidelity models with lower computational costs. This is achievable if the ROM inputs are selected based on an algorithm that would relate the inputs of the model with the expected outcomes and evaluate their relevance to with respect to each other.

Cheng et al. (2021) presented a model for estimation of cooling capacities that would take the indoor and outdoor temperatures along with the building load. They presented a genetic algorithm based strategy to reduce the number of experiments for model training. The model could predict the equipment performance with ±10% of accuracy.

Shi and O'Brien (2018) proposed a reduced order model utilizing clustering techniques for selection of redundant zones and Principal Component Analysis (PCA) for model order reduction. Energy plus was used to model the building annual energy and it was noted that the ROM could estimate the yearly energy consumption within 5% of error margin and with 95% reduction in simulation time.

Chen et al. (2021) employed a physics informed neural network model for high-fidelity problem solutions using proper orthogonal decomposition (POD). POD, a specific type of PCA , efficiently reduced the model dimensions coupled with Galerikin projection. Geometrical parameters were mapped onto the ROM using feedforward neural network. J. Ma et al. (2021) presented a reduced order modelling approach for dynamic vapor compression cycles using POD and Trajectory Piece Wise Linear (TPWL) methods. A clustering algorithm first selects the linearization points and then each full order linear piece is reduced using POD resulting in reduced order TPWL model. The full space has been projected onto the reduced space using POD-Galerikin approach. They concluded that POD-Galerikin approach performs better in cases where the stabilization of nonlinear reduced order is a challenge.

Yoon and Lee (2010) proposed ANN based model for the dynamic simulation of vapor compression cycle. A generalized radial basis function is used for prediction of air to water heat pump current states by taking inputs from previous states. They employed Levenberg-Marquardt (LM) and Scaled Conjugate Gradient (SCG) algorithms and compare them with Generalized Radial Basis Function (GRBF) for both transient and steady-state conditions. They recommended to use GRBF due its dynamic modeling ability with lesser understanding of the target plant, fast learning, and user friendly implementation.

Zhao et al. (2014) investigated a single hidden layer feed-forward network to model the vapor compression system. They used 6 input neurons, one for each input of compressor, evaporator,condenser fan speeds and expansion valve opening area along with outdoor and indoor temperatures. With a single hidden layer comprising of 50 neurons, they train the network with over a 1000 experimental data points and made output predictions for condensing pressure, evaporator pressure, sub-cool, super-heat and power consumed by the system.Neurons in the network were linked using sigmoidal function.

Among numerous ANNs, the feed-forward network is special type of Neural Network (NN) wherein the information propagates in the forward direction. The ANN structure includes an input layer with a neuron for every input parameter, hidden layers, and output layer for the desired output predictions. All these neurons are activated using different activation function. It has been shown in numerous studies that the performance of the ANN is highly influenced by input parameters, number of hidden layers, the activation functions, training algorithms and the amount of training data as well as the data pre-processing(Seo et al. (2020) ,Yousaf et al. (2019),Nawi et al. (2013), Shafi et al. (2006)) All these factors, together, plays an important role in the performance evaluation of the ANN. For this purpose, an extensive amount of trial-and-error based experimentation is usually carried out until an ANN architecture with an acceptable performance is achieved.

Moreover, training of fully linked ANN necessitates solving non convex optimization problem, so there is a chance that the outcome may represent local solutions instead of global ones. This behavior is also shown with respect to the features in the training data as well (Sildir & Aydin, 2022). So, to address this problem, extensive work is present in literature regarding the selection of relevant features. Accurate selection of the features also addresses the well-known curse of dimensionality, pronounced more with larger number of features than the number of observations. Along with over-fitting reduction,feature selection also helps in improving the accuracy by showing only the relevant features to the network during training process. Hence, not only improving accuracy but also decreasing computational cost.

Many methods are used in literature to create functions which could identify relevant features in the data without loss of data quality. Huang et al. (2020) categorized the feature selection in the similarity, information theoretical and the statistical based groups. The similarity-based methods make use of a similarity matrix, the information theoretical are applicable to continuous data while the statistical based feature methods analyze the features individually for selection of redundancy and then a subset is selected. Subset selection can be done based on either wrapper or embedded techniques.Wrapping techniques calculate models using a predetermined selection of features and assess the significance of each feature. Then, they iterate and test various subsets of features until the best subset is determined. This approach has two disadvantages: a lengthy computation time for data with numerous features and a potential to overfit the model when the number of data points is small. As part of the process of building models, embedded methods pick features. One such method is lasso proposed by Tibshirani (1996), who compared its performance to ridge and bridge regressions. Both these methods perform similar. Lasso was found favorable due to its sparse representation, but it has limitation in selection of variables when the number of features are greater than the observations in the data set. Also, if there is a group of variables where in the correlation is high, the lasso does not care which one is selected. To address these problems, elastic net (EN) was proposed by Zou and Hastie (2005) and reported that it outperforms lasso and ridge regression.

EN is being used in forecasting applications due to its nature of shrinking the coefficients for correlated features, hereby decreasing the model complexity and increasing the model predictive abilities.Nikodinoska et al. (2022) used EN to estimate the solar and wind power generation by removing the highly correlated input parameters. They noted that employing EN reduced the error by removing the highly correlated inputs. Fan et al. (2017) employed unsupervised deep learning algorithms for the estimation of cooling load. They noted that the model predictions improved when feature extraction techniques were utilized. They report that developing models based on multiple linear regression and EN are computationally efficient, and the model produced are simple to understand.

The current work presents a reduced order model for a unitary air-conditioner. Using experimental data from a 10-ton RTU, feature selection has been performed using EN. Before implementation of EN on the real data set, it has been employed on benchmark optimization problem and a known model of a cooling capacity estimation to see its performance. The second step in model order reduction is the implementation of PCA. Finally, an ANN model, informed by EN and PCA, for a unitary air-conditioner is presented and the results of current model are compared against existing models from the literature.

2. METHODOLOGY

The current work is aimed at presenting a low dimensional ANN model for prediction of various performance parameters for a unitary air conditioner. The dimensionality of the system is reduced in two steps. Initially, high fidelity data is provided to EN for feature reduction. EN has been tested on benchmark functions of Hartmann 6-dimensional function as well as Energy plus's model for cooling load and EIR predictions for feature screening. The reduced data is then provided to the PCA for a second fold model order reduction. PCA transforms the data into low dimensions and points out new dimensions that preserves the data behavior. The whole methodology has been shown in the flowchart as shown in Figure1. Before implementing the proposed method, an evaluation of the EN capability to adequately capture feature generation is explored.

2.1 Elastic Net (EN) and Benchmark Functions

Experimental data sets for unitary equipment are typically small as it is generally time intensive to collect large quantities of data. Furthermore, they often have large numbers of input and measured features. For example, the experimental data set utilized in this work has 40 features with only 31 experimental instances. To address the problem of high dimensional data with few samples, EN is used for feature selection. EN is a linear regression model that is trained with both coefficients for LASSO and ridge regularization. The benefit is that EN allows a balance of both penalties, which can result in better performance than a model with either one or the other penalty on some problems (Zou & Hastie, 2005).

Suppose we have a response variable y and a set of predictor vector x for N observations pairs. All the entries are necessary to be normalized in the interval [0,1]. Then, EN would solve the minimization problem as given by Zou and Hastie (2005).

$$\hat{\beta} = argmin(\left\|y - x\beta\right\|^2 + \lambda_2 \left\|\beta^2\right\| + \lambda_1 \left\|\beta\right\|) \tag{1}$$

where β is the regression coefficient and λ's are the $L1$ and $L2$ regularization coefficients. To see the performance of

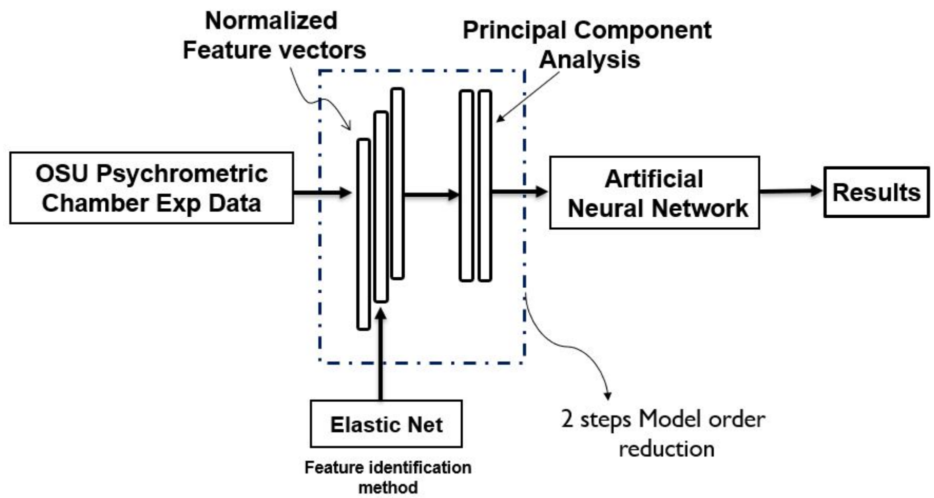


Figure 1: Proposed method of machine learning based model order reduction.

EN on filtering out the noisy variables, a benchmark optimization function, the Hartmann 6-dimension function (Cho et al., 2014), has been selected for data generation and then EN has been implemented. The Hartmann 6-D function is evaluated in the interval of [0-1] with a functional form,

$$f(x) = \sum_{i=1}^{4} \alpha_i exp(-\sum_{j=1}^{6} A_{ij}(x_j - P_{ij})^2) + 3.3082, \tag{2}$$

where,

$$\alpha_i = [1.0, 1.2, 3.0, 3.2]^T, \tag{3}$$

$$A = \begin{bmatrix} 10 & 3 & 17 & 3.5 & 1.7 & 8 \\ 0.05 & 10 & 17 & 0.1 & 8 & 14 \\ 3 & 3.5 & 1.7 & 10 & 17 & 8 \\ 17 & 8 & 0.05 & 10 & 0.1 & 14 \end{bmatrix} \tag{4}$$

$$P = \begin{bmatrix} .1312 & .1696 & .5559 & .124 & .8283 & .5886 \\ .2329 & .4135 & .8307 & .3736 & .1004 & .9991 \\ .2348 & .1451 & .3522 & .2883 & .3047 & .6650 \\ .4047 & .8828 & .8732 & .5743 & .1091 & .3810 \end{bmatrix} \tag{5}$$

The input random variables $x_1, x_2, x_3, x_4, x_5, x_6$ uses distribution types of normal, log-normal, gamma, weibull, normal and weibull, respectively. While x_5 and x_6 are correlated with each other using Clayton copula with a Kendall's tau of 0.5. To introduce irrelevant and noisy variables, three more variables, a_1, a_2, a_3 have been also added. a_1 and a_2 have distribution of type normal while a_3 is a linear combination of the first two noisy variables.

Another well-known model to the HVAC community is the Energy Plus's bi-quadratic curve fitted model (*EnergyPlus*, Accessed: 4/12/2022) for estimation of cooling capacity Q for a unitary equipment.

$$Q_{tot} = Q_{rat} * f_{temp} * f_{airflow}, \tag{6}$$

whereas:

$$f_{temp} = a_0 + a_1 T_{ODB} + a_2 T_{ODB}^2 + a_3 T_{IWB} + a_4 T_{IWB}^2 + a_5 T_{ODB} T_{IWB}, \tag{7}$$

and

$$f_{airflow} = b_0 + b_1 \frac{V}{V_{rat}} + b_2 \frac{V}{V_{rat}}^2. \tag{8}$$

While, T_{ODB},T_{IWB},and V are outdoor dry bulb temperature, indoor wet bulb temperature and air mass flow rate respectively.

Using regression analysis, model coefficients computed using the rated cooling capacity , gross cooling capacity, compressor power consumption at varying outdoor and indoor temperature conditions of the unitary equipment. To segregate the noisy and useful variables, furthermore, four additional random variables Var_1, Var_2, Var_3 and Var_4 have been included in the data set to look for the effectiveness of EN.

2.2 Principal Component Analysis

PCA is utilized to decrease the complexity of multidimensional data while keeping its trends and patterns (Jolliffe & Cadima, 2016). This is achieved by compressing the data into fewer dimensions that act as feature summaries. The purpose of transforming data into smaller dimensions, Principal Components (PC) is to summarize the data with minimal loss. First PC is chosen in such a way that the total distance between the data and its projection is as low as possible resulting in increase in the variance as well. Rest of the PCs are selected in the similar manner but with an added obligation of not being correlated to the previously chosen components. It has an overall effect of increasing the variance in the transformed data with low dimensions. The PCs can be thought of as linear combination of the variables in original data and are stored in PCA loading matrix. This loading matrix can be explained as a rotation matrix for the data that would align the projection vector with the highest variance with the first axis and so on for the rest of the vectors.

It is essential to note that PCA is distinct from linear regression. In linear regression analysis, the goal is to decrease the distance between a true response and its predicted value, but in principal component analysis, the goal is to minimize the orthogonal distance between a data point and its principal component.

2.3 Artificial Neural Network

An ANN is a biologically inspired computing approach that may adapt to a wide variety of issues for which a mathematical model is unavailable or solving the model is impractical owing to its complexity. ANN is composed of tiny units called as neurons that are interconnected with other neurons in the neighborhood. A set of neurons combines and form a layer. An ANN has an input layer through it takes data from the user and provides data through an output layer. The flow of information from input to output happens through intermediate layers known as hidden layers. ANNs having multiple hidden layers are knows as multi-layer perceptron (MLP).

For the sake of simplicity, a single hidden layer is often recommended, although several hidden layers may be necessary if the problem at hand cannot be described accurately with a single hidden layer. The number of neurons in the input and output layers is determined by the model's input and output. Neurons are connected to each other using numeric values known as weights. For every neuron,inputs are multiplied with weights, summed and the result is provided by an activation function, which transmits it to the following neuron or provides it as an output. Among numerous activation functions, Sigmoid function is well known function and is given as:

$$f(x) = logsig(y) = \frac{1}{1 + e^{-y}}. \tag{9}$$

To increase the learning flexibility of neuron, a constant term known as bias is also introduced, which allows the summation value to be varied according to the needs of the model.ANNs must be trained using data, which entails adjusting weights for the intended outputs. Training data set refers to the set of data used to train the network. It contains training features that serve as network inputs for predicting the target data set. Target data set consists of the output, and ANN modifies the weights to discover a pattern in the data relative to the outputs. To avoid over-fitting of network, the training data is split into groups of training and testing in ratio of 80%-20% respectively. One more important thing to be considered is that the performance of the ANN varies with respect to the number of neurons and layers. Minimum number of neurons and layers should be used, else there is a risk that network might get over-fit on greater number of neurons and layers (Yousaf et al., 2019). There is no such rule of thumb for the specification of number of neurons and layers, so a trial-and-error method is mostly followed (Shafi et al., 2006). Following ANN training comes testing, which needs a distinct data set known as the testing set. Testing evaluates the performance of an ANN on an unknown data set. This study use relative percentage error (RPE) as its assessment measure. Although evaluation may be conducted using a variety of performance metrics, this study employs RPE. It is as follows:

$$RPE = \frac{(Q_{exp} - Q_{pred})}{Q_{exp}} * 100. \tag{10}$$

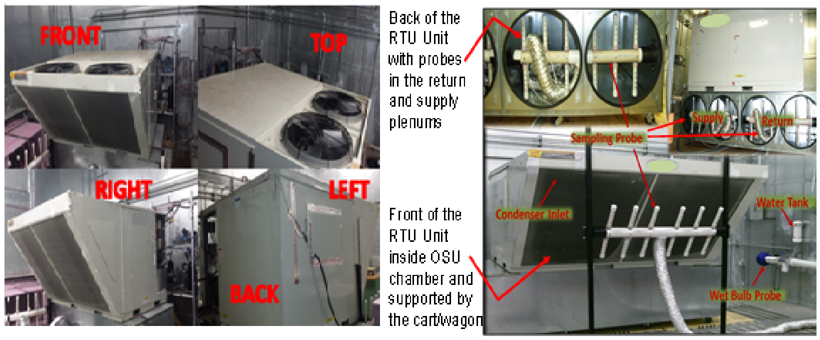


Figure 2: Example RTU under testing at OSU psychrometric chambers.

Whereas, Q_{exp} and Q_{pred} are the experimental and predicted values of cooling capacities for the unitary equipment.

2.4 Experimental Methodology

To present a machine learning based ROM for unitary air-conditioning system, we requires high fidelity experimental data. Use of EN along with PCA would reduce the dimensions and then the reduced dimensioned data shall be used for ANN training.Experiments were conducted in state-of-the-art psychrometric chambers at Oklahoma State University (OSU) on a 10-ton rooftop air conditioning unit (RTU) similar to what is shown in Figure 2. Air temperatures along with the refrigerant temperatures and pressures were noted at various critical points of the system as described in Figure 3. For each experiment, the collected data had a total of 40 feature vectors for both air and refrigerant sides. The recorded outputs were sensible and latent cooling capacities along with the coefficients of performance for steady state conditions. Thus, the data matrix obtained had both the refrigerant side as well the air side measurements for total of 31 experiments for varying indoor and outdoor temperature conditions. The tests were run in accordance with ASHRAE (2009)

3. RESULTS AND DISCUSSION

Before proceeding towards the next step, the variable screening, we tested the EN on the benchmark optimization function, the Hartmann 6-dimensional function. As described in Section 2.1, additional noise was introduced in the data set generated from the function itself for randomly selected inputs and data set was fed to the EN. It was observed that EN filtered out noisy variables that were introduced into the dataset as a part of the process of evaluating the EN. Input variables for the functions had a considerable higher score of relevance as compared to the made-up noisy variable which had almost zero score of relevance. EN was made to pass through another test to identify and hence enable us to segregate the noise and the useful variables from a well-known model of EnergyPlus, described in Section 2.1, for estimation of the cooling capacity of the unitary air conditioning system. Similar to the earlier process of introducing some made up variables, we added four additional random parameters to the data.Figure 4 presents the results of these benchmark experiments where the scores presented represents the correlation inputs to the output variable. In figure 4a, the 6 input variables have been identified while the random variables are identified as non- useful. For the EnergyPlus model, in figure 4b, EN identifies the temperatures as relevant while the CFM, which was a constant throughout the experimentation has not been identified as useful, which is in accordance to the real applications for a fixed speed equipment. These results demonstrates that EN is able to detect that the random variables having no effect on the output and can accurately forecasts the relative relevance of the input variables for the supplied functions. This provides assurance that EN will be beneficial for choosing features from a large data source.

Since EN has proved its capacity to filter noisy and relevant variables, the experimental data obtained for the aim of estimating the cooling capacity is sent to EN. and it has been determined that some of the features can be disregarded. For air side, EN has identified all the air side temperatures to be important. For a fixed compressor speed unit, the compressor speed was found to be not important. Complete list of variables recorded and feed to the EN along with their relevance with respect to cooling capacity is as shown in Figure5.The EN resulted in a 50% reduction in the dimension of the data set compared to the original feature set.

In an effort to further reduce the dimensions of the data, we utilize PCA to further compress the data without loss.

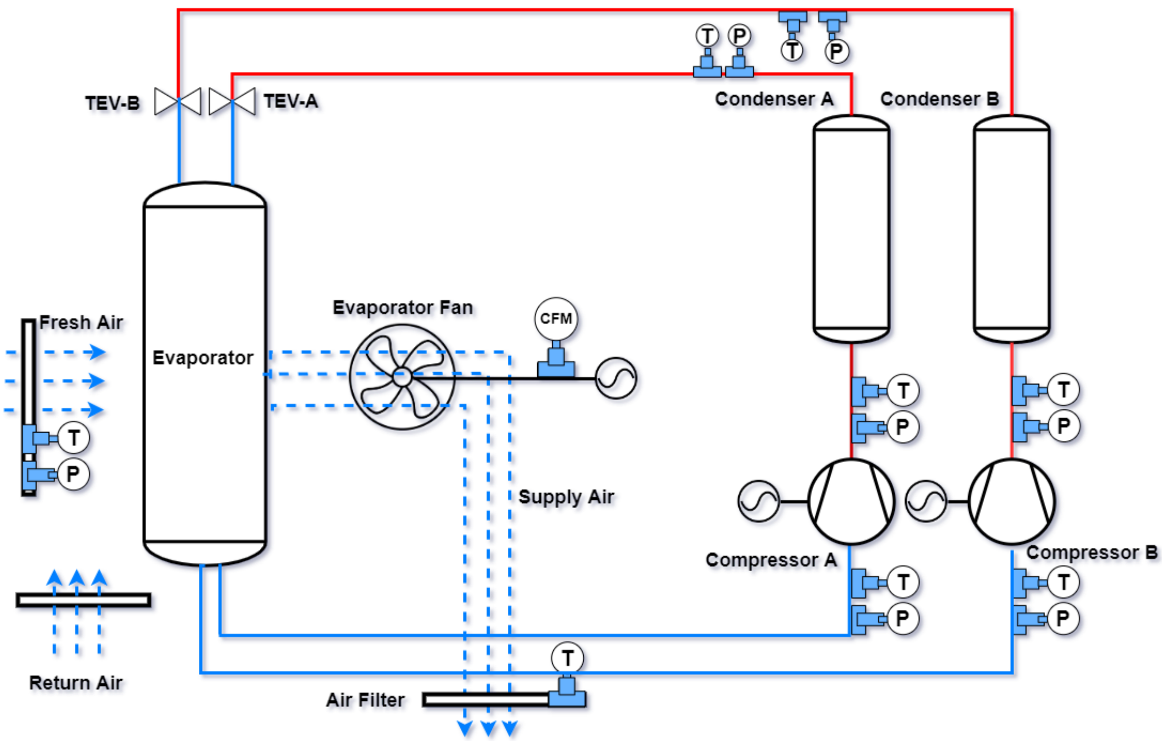
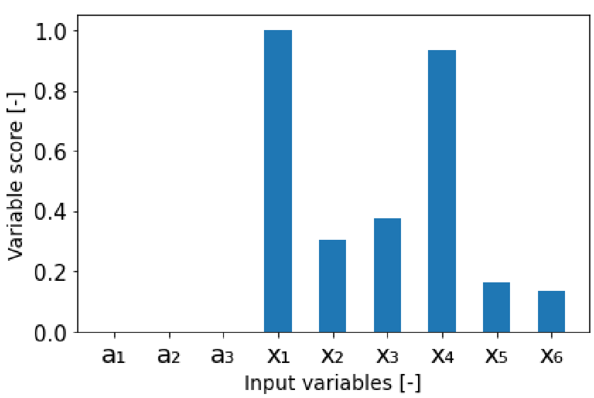
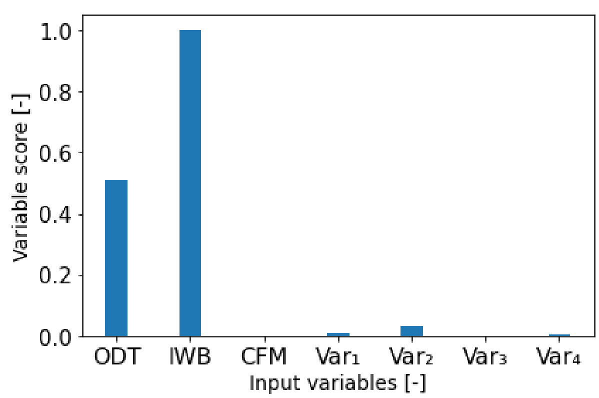


Figure 3: Experimental schematic showing critical locations of measurements.



(a) Relevance score for Hartmann 6D function



(b) Relevance score for Energy Plus model for cooling capacity

Figure 4: Bench marking of EN as feature selection method.

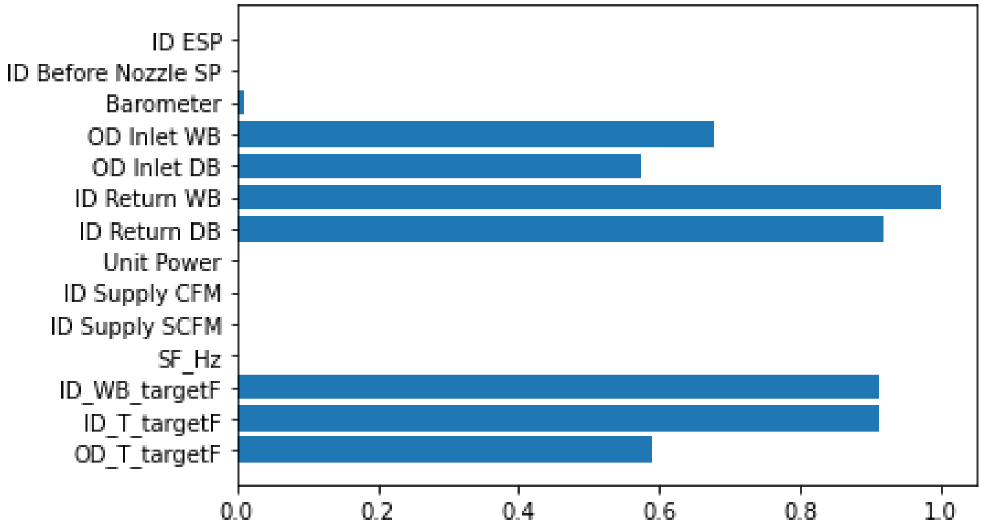
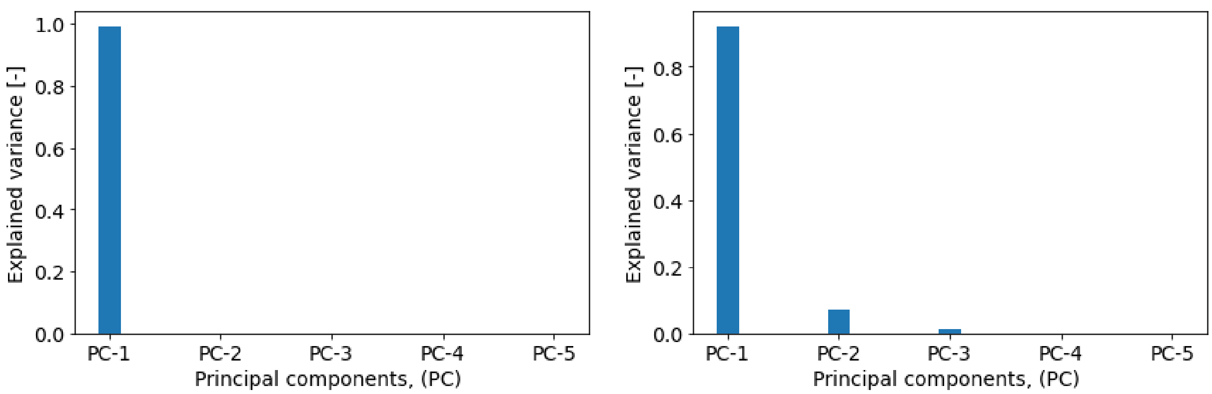


Figure 5: Elastic net based calculated score for various input features.



(a) No. of PCs with PCA and EN

(b) No. of PCs with PCA only

Figure 6: Use of PCA with and without EN.

Figure 6(a) demonstrates the results of the PCA. It has been demonstrated that PCA can turn data into a single Principal Component (PC) that captures 99% percent of the energy of the data.As a benchmark for comparing the 2-step model order reduction, PCA was similarly applied to the unfiltered raw data set. As a basis of comparison of the 2-step model order reduction, PCA was similarly applied to the raw data set with no EN filtering. The results of this analysis, as shown in Figure 6(b), depicted that without employing the EN, the data can be represented in 3 PCs which means that the 2-step model order reduction is more effective than a standalone PCA. Now that the dimensionality has been lowered using a two-step model order reduction approach, the ANN model must be trained on this data. For ANN, it is crucial to divide the data into training and testing subsets before training the network using the training data set. For the current problem at the hand, we split the data set into 80-20% of training-testing data set. The topology of ANN has a significant impact on training outcomes. Different numbers of neurons in the hidden layers were examined for this purpose. For ANN, it is advised to begin with a simple architecture and, based on its performance, add more hidden layers and neurons to achieve the appropriate accurateness. In accordance with this proposal, originally 5 neurons were chosen from the single hidden layer, and this number was later raised to 10.The performance of ANN on the test data improved to the desired level of within ±2% and hence no further combinations were tried, with architecture 2 being chosen. We could have increased the number of neurons in the hidden layer but the improvement in the prediction might not be significant and furthermore, there is a chance that the ANN might get over-fit as well. The results of the

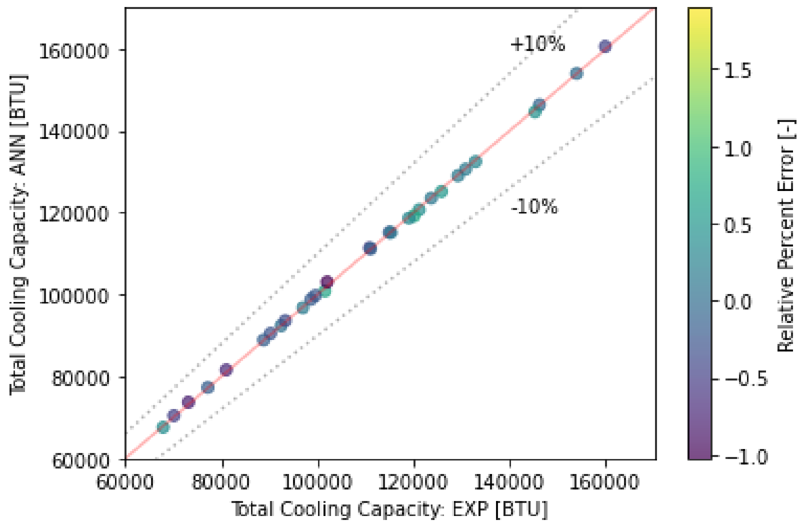


Figure 7: 2-step ROM for total cooling capacity prediction

two architectures are shown in the table1. The decision was made based on results with least amount of relative percent error in the test results. Because PCA condensed the data into a single PC, the final architecture selected comprises an input layer with one neuron and a single hidden layer with 10 neurons. Since we are predicting a single output, the output layer, like the input layer, included a single neuron.The result from the model is shown as in Figure 7.It can be observed that the model predicts the cooling capacity very accurately and the results are almost within 2% for lower and higher capacities.

With the final results from the proposed method, it should be compared to the existing models of EnergyPlus and Cheng et al. (2021).Cheng et al. (2021) is an extension of EnergyPlus model and both are bi-quadratic polynomials based models. Cheng et al. (2021) model for the estimation of the total cooling capacity is shown as follows.

$$Q_{t,max} = Q_{rat}[a_0 + a_1 \frac{B_{ID}}{B_{IDrat}} + a_2 \frac{T_{OD}}{T_{ODrat}} + a_3 \frac{B_{ID}}{B_{IDrat}}^2 + a_4 \frac{T_{OD}}{T_{ODrat}}^2 + a_5 \frac{B_{ID}}{B_{IDrat}} \frac{T_{OD}}{T_{ODrat}}]. \tag{11}$$

$Q_{t,max}$ is the total cooling capacity at specific conditions while $Q_{t,rat}$ is the rated cooling capacity at rated conditions. B_{ID}/B_{IDrat} are the normalized indoor wet bulb temperatures and T_{OD}/T_{ODrat} are the normalized outdoor dry bulb temperatures. $a_0, a_1, a_2, a_3, a4$ and a_5 are the model coefficients obtained using nonlinear optimization.

We compared two existing models from the literature as shown in Figure 8. It can be seen that for the given data set, 2-step ROM and EnergyPlus model are performing within ±10% while the Cheng's model performance closer to ±10%. The best performing model among them is 2-step ROM with RPE of ±2% while the EnergyPlus model predicts the cooling capacities in range of -6% to 4%. Considering the Mean Absolute Percentage Error (MAPE), it can be further analysed that MAPE's for 2-step ROM, EP and Cheng models are 0.20, 2.32 and 3.56, demonstrating the effectiveness of the proposed method.

Table 1: Neural Network architecture optimization.

Architecture no.	Hidden layers	No. of neurons in hidden layer	Relative percent error
1	1	5	-1.5 to 2.20
2	1	10	-1.02 to 1.89

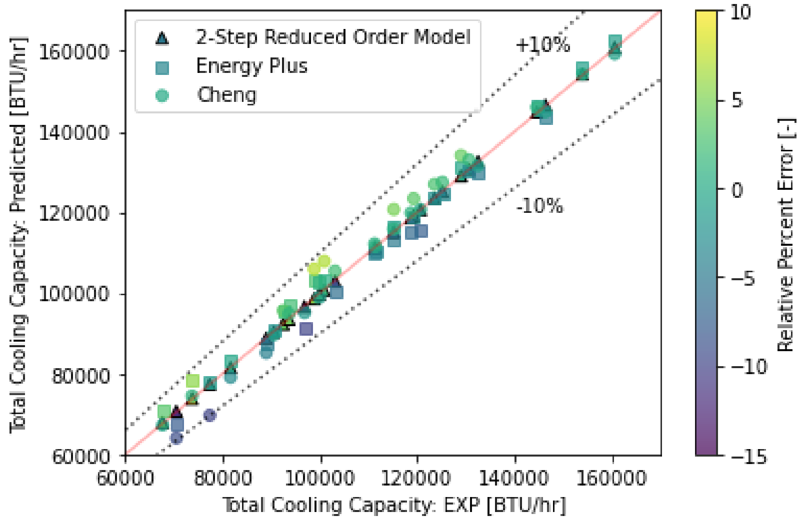


Figure 8: Comparison of two step reduced order machine learning model with existing models from literature.

4. CONCLUSIONS

The paper presents a model order reduction for a unitary air conditioning system that takes place over the course of two steps.In the first step, EN was used to separate the redundant features and was able to cut the number of input features by 50%, and in the second phase, PCA was able to reduce the dimensionality of the data even more. The proposed approach of model order reduction required just 6.25% of the initial data dimension to train the ANN. Furthermore, the current work has been compared to the EnergyPlus and Cheng *et. al.* (2021) and has been found to be superior in terms of cooling capacity estimation. The ANN results are shown to be in the ±2% of the experimental results and with these results, it can be concluded that an informed machine learning based reduced order model can be used for predictive control strategies in building energy optimization.

REFERENCES

ASHRAE. (2009). Methods of testing for rating electrically driven unitary air-conditioning and heat pump equipment (ashrae/ansi approved).

Belman-Flores, J., Ledesma, S., Barroso-Maldonado, J., & Navarro-Esbrí, J. (2015). A comparison between the modeling of a reciprocating compressor using artificial neural network and physical model. *International Journal of Refrigeration*, *59*, 144–156.

Borggaard, J., Burns, J. A., Surana, A., & Zietsman, L. (2009). Control, estimation and optimization of energy efficient buildings. In *2009 american control conference* (pp. 837–841).

Chen, W., Wang, Q., Hesthaven, J. S., & Zhang, C. (2021). Physics-informed machine learning for reduced-order modeling of nonlinear problems. *Journal of Computational Physics*, *446*, 110666.

Cheng, L., Braun, J. E., & Horton, W. T. (2021). Load-based testing using a thermostat environment emulator. *International Journal of Refrigeration*, *126*, 109–122.

Cho, H., Bae, S., Choi, K., Lamb, D., & Yang, R.-J. (2014). An efficient variable screening method for effective surrogate models for reliability-based design optimization. *Structural and multidisciplinary optimization*, *50*(5), 717–738.

Energyplus. (Accessed: 4/12/2022). Retrieved from `https://energyplus.net/assets/nrel_custom/pdfs/pdfs_v22.1.0/EngineeringReference.pdf`

Fan, C., Xiao, F., & Zhao, Y. (2017). A short-term building cooling load prediction method using deep learning algorithms. *Applied energy*, *195*, 222–233.

Huang, Y., Jin, W., Yu, Z., & Li, B. (2020). Supervised feature selection through deep neural networks with pairwise connected structure. *Knowledge-Based Systems*, *204*, 106202.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical*

Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, *374*(2065), 20150202.
Ma, J., Kim, D., & Braun, J. E. (2021). Proper orthogonal decomposition for reduced order dynamic modeling of vapor compression systems. *International Journal of Refrigeration*, *132*, 145–155.
Ma, Z., & Wang, S. (2009). Energy efficient control of variable speed pumps in complex building central air-conditioning systems. *Energy and Buildings*, *41*(2), 197–205.
Nawi, N. M., Atomi, W. H., & Rehman, M. Z. (2013). The effect of data pre-processing on optimized training of artificial neural networks. *Procedia Technology*, *11*, 32–39.
Nikodinoska, D., Käso, M., & Müsgens, F. (2022). Solar and wind power generation forecasts using elastic net in time-varying forecast combinations. *Applied Energy*, *306*, 117983.
Seo, Y. M., Luo, K., Ha, M. Y., & Park, Y. G. (2020). Direct numerical simulation and artificial neural network modeling of heat transfer characteristics on natural convection with a sinusoidal cylinder in a long rectangular enclosure. *International Journal of Heat and Mass Transfer*, *152*, 119564.
Shafi, I., Ahmad, J., Shah, S. I., & Kashif, F. M. (2006). Impact of varying neurons and hidden layers in neural network architecture for a time frequency application. In *2006 ieee international multitopic conference* (pp. 188–193).
Shi, Z., & O'Brien, W. (2018). Building energy model reduction using model-cluster-reduce pipeline. *Journal of Building Performance Simulation*, *11*(5), 553–567.
Sildir, H., & Aydin, E. (2022). A mixed-integer linear programming based training and feature selection method for artificial neural networks using piece-wise linear approximations. *Chemical Engineering Science*, *249*, 117273.
Širokỳ, J., Oldewurtel, F., Cigler, J., & Prívara, S. (2011). Experimental analysis of model predictive control for an energy efficient building heating system. *Applied energy*, *88*(9), 3079–3087.
Tanveer, M. M., & Bradshaw, C. R. (2021). Performance evaluation of low-gwp refrigerants in 1–100 ton scroll compressors. *International Journal of Refrigeration*, *129*, 317–331.
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.
Wang, W., Zhou, Q., Pan, C., & Cao, F. (2022). Energy-efficient operation of a complete chiller-air handing unit system via model predictive control. *Applied Thermal Engineering*, *201*, 117809.
Yoon, Y.-J., & Lee, M. H. (2010). Dynamic simulation of vapor-compression cycle using neural networks. *International Journal of Control, Automation and Systems*, *8*(6), 1241–1249.
Yousaf, S., Shafi, I., Din, S., Paul, A., & Ahmad, J. (2019). A big data analytical framework for analyzing solar energy receptors using evolutionary computing approach. *Journal of Ambient Intelligence and Humanized Computing*, *10*(10), 4071–4083.
Zhao, L., Cai, W.-J., & Man, Z.-H. (2014). Neural modeling of vapor compression refrigeration cycle with extreme learning machine. *Neurocomputing*, *128*, 242–248.
Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), 301–320.

ACKNOWLEDGMENT

This work has been funded by the Oklahoma Center for the Advancement of Science and Technology (OCAST) with grant number AR-042 and the Center for Integrated Buildings Systems (CIBS), and Industry/University Cooperative Research Center at Oklahoma State University.