International Conference on Applied Energy 2019
Aug 12-15, 2019, Västerås, Sweden
Paper ID: 0671

RELATIVE HUMIDITY ESTIMATION: MACHINE LEARNING APPROACH–RANDOM FOREST-BASED PREDICTION MODEL

Kinza Qadeer[1], Ashfaq Ahmad[2], Muhammad Abdul Qyyum[1], Moonyong Lee[1*]

1 School of Chemical Engineering, Yeungnam University, Gyeongsan 712-749, Rep. of Korea

2 Department of Computer Science, COMSATS University Islamabad (CUI), Lahore Campus, Defense Road, Off Raiwind Road, Lahore, Pakistan

ABSTRACT

Relative humidity (ϕ) is considered a major parameter during the designing of HVAC (Heating, ventilation, and air conditioning) systems. Generally, HVAC engineers use a psychrometric chart to observe and estimate the air quality parameters. Nevertheless, high skills are required to make rigorous and accurate reading from the psychrometric chart and the "human error" is an added factor that can lead to big disasters. Therefore, rigorous and user-friendly estimation of air quality parameters is still an ongoing issue. In this context, we are going to implement the state-of-the-art "Machine learning" technique to develop a simple, robust, and rigorous predictive tool for the estimation of relative humidity. A well-proven approach i.e., the random forest (RF) is employed to train the model for robust estimation. It was found that the mean absolute deviation was 54.3% lower than that of well-known ordinary least square (OLS) regression method.

Keywords: Relative humidity, random forest, prediction, ordinary least square, HVAC engineers

NONMENCLATURE

Abbreviations	
HVAC	Heating, ventilation, and air conditioning
IB data	In-bag data
OOB data	Out of bag data
ML	Machine learning
RF	Random forest
PC	Psychrometric chart
MSE	Minimum squared error
OLS	Ordinary least-square method
Symbols	
T_{wb}	Wet-bulb temperature
T_{db}	Dry-bulb temperature
T_{wbd}	Wet-bulb depression
ϕ	Relative humidity

1. INTRODUCTION

The accurate estimation of relative humidity (ϕ) as a function of T_{wb} and T_{db} is of critical importance from domestic to industrial scale. Many processes involve air dependent operations such as air-cooled LNG plants [1], proton exchange membrane fuel cell [2], air conditioning [3], drying, weather forecasting, petrochemical industries, and cooling towers. Figure 1 is showing the some major air-involved industrial operations.

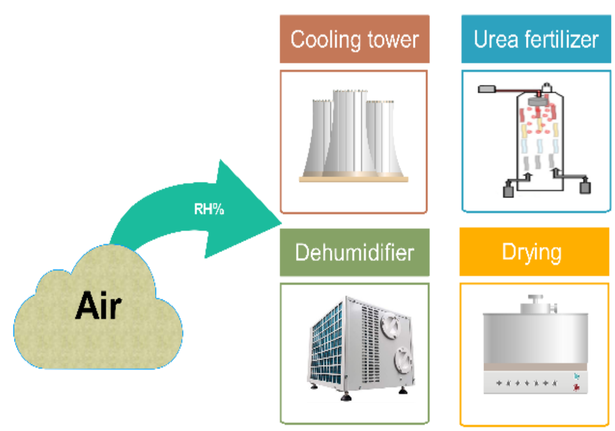

Fig 1 Air dependent industrial operations

*Corresponding author: mynlee@yu.ac.kr, Telephone: +82-53-810-2512

Hence, a rigorous and robust estimation of ϕ has an immense importance from daily life to industrial life. Several estimation tools and models [4–7] have been presented to predict the moist air properties including relative humidity, wet bulb temperature, and dew point. Most of them are so complicated due to number of parameters involved and thus difficult for process engineers to apply and get appropriate results. Eccel [8] estimated air humidity and tested the application of algorithms. They did not present any robust model to estimate the ϕ under different T_{db} and T_{wb}. Martínez et al. [9] estimated temperature and ϕ in the tobacco drying using a system based on Artificial Neural Network (ANN). However, some models are robust for example, Bahadori et al. [10] used wet-bulb depression (T_{wbd}) and T_{db} as independent variables to predict ϕ at standard atmospheric conditions using an Arrhenius-type asymptotic exponential function. They tested their model for ranges of T_{db}: -10°C to 50°C and T_{wbd}: 1°C to 45°C. The major issue associated with their model is the complexity, mainly due to many (≥16) unknown constants; making it computationally expensive.

This study proposes a state-of-the-art machine learning based predictive tool for rigorous and robust estimation of a major air property i.e., ϕ as a function of T_{wb} and T_{db}. A well-proven training algorithm i.e., random forest is adopted to train a model. The proposed predictive tool is developed in MATLAB 2018b. The performance analysis of the proposed estimation tool is compared with existing latest estimation models as well as standard psychrometric chart.

2. MACHINE LEARNING BASED ESTIMATION TOOL

Machine learning is a subset of artificial intelligence in which computers use statistical techniques to dig out essential information from the data without being explicitly programmed. In general way, machine learning (ML) helps to process the large massive data and make predictions within least amount of time.

2.1 Proposed predictive tool

Supervised learning random forest methodology is used for the prediction and causal inference. A brief structure of RF algorithm is shown in Figure 2.

Random forest (RF) algorithm is one of the most promising machine learning algorithms to its simplicity, flexibility (used for both classification and regression), easy to use (even without hyper-parameter tuning) and give great results most of the time. As it already clears from its name that it creates forest to somehow random.

The forest that is developed by it using the ensemble of decision trees (also known as Regression trees "CART"), usually trained by the "bagging" method. Bagging is a combination of bootstrap aggregating. The general concept behind the "bagging" or obviously behind the random forest algorithm is that it combines several decision trees and merge them together to produce a more rigorous and accurate prediction than a single classifier. The prediction phenomenon in RF algorithm is a combination of training and testing phases. In the training phase, a huge number of decision trees are constructed using multiple bootstrap samples (N) from the training data set. For each sample a regression tree (CART) is allocated which consist of node, branches and leaves. Objective function attribute is located on each node of the tree which is selected using a random subset from the data. Then the values of objective function are passed from branches to leaves using the logical principle "IF-THEN".

To build the proposed Random forests algorithm a sample dataset of 2616 observations were taken. The algorithm is then trained and tested in two different types of experiments. In first type of experiment, the algorithm is trained on whole sample data of 2616 observations and tested on user defined test data and then validated with psychrometric chart. In second type of experiment, Out of sample data, 70% of data was randomized drawn with replacement for training purpose. Picking training data with replacement ensures the occurrence of variety of outputs with different results. The trained model was then tested on remaining 30% of data.

3. PERFORMANCE ANALYSIS

Performance of RF algorithm mainly depends on the number of leaves grown on random number of trees. To find an optimal leaf size with minimum squared error (MSE) is a one critical step during the prediction adopting RF. In this study, a sensitivity analysis was performed to find out an optimal no. of leaves against various size of random trees corresponding to MSE. Figure 3 shows the analysis for the selection of optimal size of leaves. According to Figure 3, it can be seen that small size of leaves gives lower value of MSE upon increasing the number of trees. Therefore, the proposed study uses 5 no. of leaves for the estimation of ϕ. It is found that MSE value is least if number of trees built on training data are

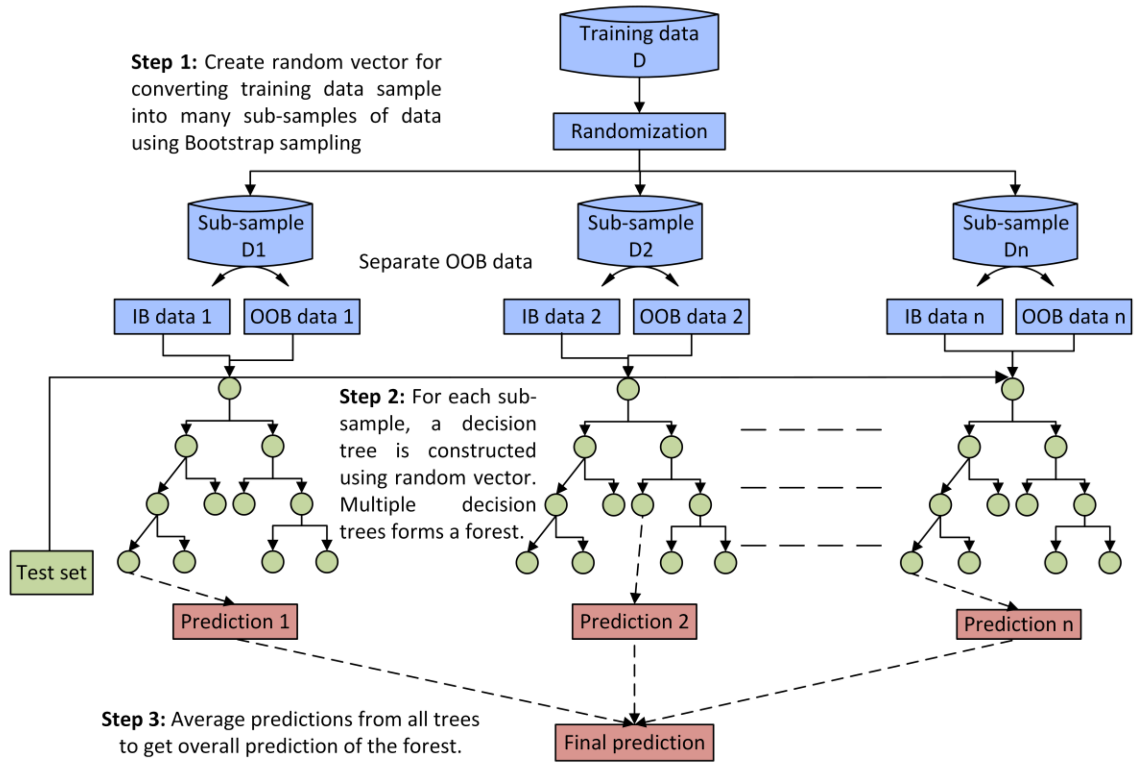

Fig 2 A brief structure of the Random Forest algorithm

greater than 10. The red curve gives lowest MSE value that actually verifies our claim explained in above experiments. It is also obvious that the prediction accuracy curve remains smoothen with increasing number of trees in a forest.

However, by increasing the number of trees in forest we develop maximum combination of observations in training data that affects the final prediction.

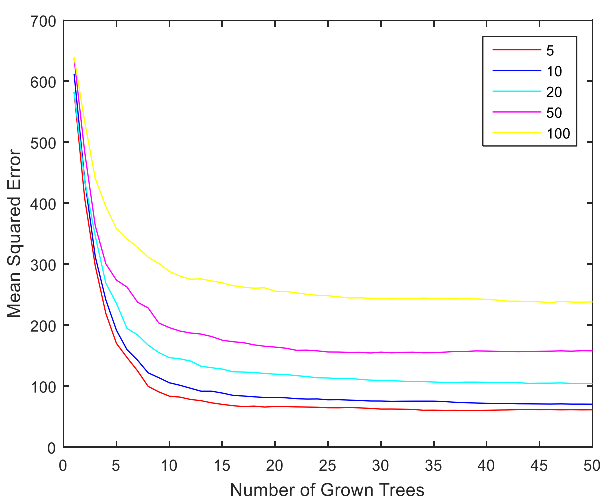

Fig 3 Verification of optimal leaf by comparing mean squared error value for various no. of leaf

Furthermore, the performance analysis of the proposed random-forest-based prediction was compared with standardized psychrometric chart and OLS regression method. Table 1 lists ϕ values calculated from RF and OLS in comparison with PC [11].

Table 1 Comparison of calculated values with psychrometric chart and ordinary least sqaure method

T_{db} (°C)	T_{wb} (°C)	ϕ_{PC} %	ϕ_{OLS} %	ϕ_{RF} %	ε* %	ε** %
8	3	40.0	39.4	42.0	1.4	5.0
-1	-5	25.0	32.2	25.7	28.6	2.9
10	3	25.0	34.6	19.7	38.4	21.3
19	13	50.0	47.7	50.1	4.6	0.1
23	18	62.0	55.7	57.8	10.1	6.8
39	30	52.0	54.0	54.1	3.8	4.0
26	15	29.0	36.3	26.7	25.1	8.0
29	19	39.0	42.2	41.1	8.3	5.3
13	8	50.0	45.1	49.1	9.8	1.8
Mean absolute deviation					13.4	6.1

$\varepsilon^* = |((\phi_{OLS} - \phi_{PC})/\phi_{PC})|$; $\varepsilon^{**} = |((\phi_{RF} - \phi_{PC})/\phi_{PC})|$

According to Table 1, the highest absolute deviation was observed at same conditions (T_{db} = 10°C and T_{wb} = 3°C). At these conditions, the absolute deviations were found 38.4% and 21.3% for OLS and RF models, respectively. Nevertheless, the minimum absolute deviation was not found at same conditions. For OLS case, the minimum absolute deviation was 1.4% at 8°C dry bulb temperature and 3°C wet-bulb temperature. Whereas, in the case of RF, 0.1% minimum absolute deviation was observed at T_{db} = 19°C and T_{wb} = 13°C.

4. CONCLUSIONS

This study has presented the machine-learning based estimation model for the prediction of relative humidity and wet-bulb temperature. A random forest algorithm has examined to train the model for rigorous estimation of ϕ. The ϕ has estimated using classical OLS method and proposed RF-based prediction model. OLS method gives ϕ values with 13.4% mean absolute deviation that is 54.3% higher than that of estimated from RF-based estimation model. It has observed that by increasing the training dataset can also affect the prediction performance but the addition of incorrect training sets in the training data may affect the performance, reversely. The examination of the correlation of each independent and the dependent variables helps in deciding the preferential predictor that further directs for the consequence of the inclusion of the datasets of the respective predictors. This work provides help to process engineers in designing of different industrial operations such as drying, cooling, and heating.

ACKNOWLEDGEMENT

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2018R1A2B6001566) and the Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2014R1A6A1031189).

REFERENCE

[1] Qyyum MA, Minh LQ, Ali W, Hussain A, Bahadori A, Lee M. Feasibility study of environmental relative humidity through the thermodynamic effects on the performance of natural gas liquefaction process. Appl Therm Eng 2018;128. doi:10.1016/j.applthermaleng.2017.08.090.

[2] Jeon SW, Cha D, Kim HS, Kim Y. Analysis of the system efficiency of an intermediate temperature proton exchange membrane fuel cell at elevated temperature and relative humidity conditions. Appl Energy 2016;166:165–73. doi:https://doi.org/10.1016/j.apenergy.2015.12.123.

[3] Lim DK, Ahn BH, Jeong JH. Method to control an air conditioner by directly measuring the relative humidity of indoor air to improve the comfort and energy efficiency. Appl Energy 2018;215:290–9. doi:https://doi.org/10.1016/j.apenergy.2018.02.004.

[4] Stull R. Wet-Bulb Temperature from Relative Humidity and Air Temperature. J Appl Meteorol Climatol 2011;50:2267–9. doi:10.1175/jamc-d-11-0143.1.

[5] Sadeghi S-H, Peters TR, Cobos DR, Loescher HW, Campbell CS. Direct Calculation of Thermodynamic Wet-Bulb Temperature as a Function of Pressure and Elevation. J Atmos Ocean Technol 2013;30:1757–65. doi:10.1175/JTECH-D-12-00191.1.

[6] Zhang W, Ma H, Yang SX. An Inexpensive, Stable, and Accurate Relative Humidity Measurement Method for Challenging Environments. Sensors (Basel) 2016;16:398. doi:10.3390/s16030398.

[7] Brooker DB. Mathematical Model of the Psychrometric Chart. Trans ASAE 1967;10:558–60. doi:https://doi.org/10.13031/2013.39729.

[8] Eccel E. Estimating air humidity from temperature and precipitation measures for modelling applications. Meteorol Appl 2012. doi:10.1002/met.258.

[9] Martínez-Martínez V, Baladrón C, Gomez-Gil J, Ruiz-Ruiz G, Navas-Gracia LM, Aguiar JM, et al. Temperature and relative humidity estimation and prediction in the tobacco drying process using artificial neural networks. Sensors (Switzerland) 2012. doi:10.3390/s121014004.

[10] Bahadori A, Zahedi G, Zendehboudi S, Hooman K. Simple predictive tool to estimate relative humidity using wet bulb depression and dry bulb temperature. Appl Therm Eng 2013;50:511–5. doi:https://doi.org/10.1016/j.applthermaleng.2012.07.033.

[11] Kutz M. Handbook of Environmental Engineering. John Wiley & Sons; 2018.