Evaluating Stack Effect Impact of Thermal Comfort in High Rise Office Towers

Gary Chang[1], Karim El Mokhtari[1], Russell Richman[1], J.J. McArthur[1]
[1]Ryerson University, Toronto, Canada

Abstract

Stack effect adversely affects energy consumption and thermal comfort in high-rise buildings due to the excess infiltration caused by the resultant pressure differentials. Mitigating the effects of this phenomenon is crucial to achieve ambitious energy targets such as Canada's commitment to COP21. A novel data-driven modelling approach is leveraged using return air temperature as a proxy for thermal comfort issues caused by stack effect. A case study investigated the prediction algorithms, where it was determined that random forest was the most suited for stack effect modelling. The research demonstrates the utility of return air temperature data to identify situations where stack effect is most detrimental to occupant comfort.

Key Innovations

- Application of return air temperature as a proxy for stack effect related thermal comfort issues.
- Development of novel data-driven approach for predicting thermal comfort issues related stack effect.

Practical Implications

Buildings typically include return air temperature, and this paper develops a new opportunity to use it as a key indicator of thermal comfort issues without costly infrastructure.

Introduction

Stack effect is a phenomenon found predominantly in high-rise buildings due to the buoyancy of heated air moving upwards (Wilson & Tamura, 1968).

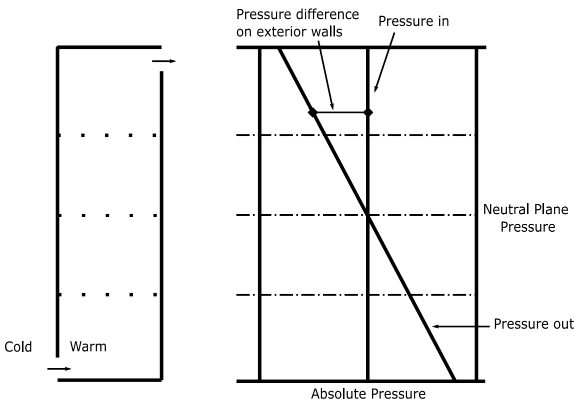


Figure 1 Stack effect during heating season

As illustrated in Figure 1, the effect is driven by pressure differentials across building components leading to increase air infiltration, which may be beneficial in cooling seasons in mild climates but adverse in heating seasons (Jung-yeon, Song, & Dong-woo, 2017) and in cooling in humid climates. This is also known as the chimney effect since it causes heat to rise in buildings, causing uneven temperature distributions. In the heating season, temperatures tend to increase with elevation due to stack effect (Yang, Du, Peng, & Li, 2013). Stack effect is detrimental to both energy efficiency, and occupant comfort. It is estimated stack effect contributes to 10.27% of the heating load in high-rise residential buildings (Yoon, Song, Kim, & Lim, 2019), and the excessive air infiltration leads to inefficient heating/cooling strategies (Mijorski & Cammelli, 2016); it's mitigation is thus an important energy-conservation measure. Further, the uncontrolled air movement from the stack effect has also reported to adversely affect occupant comfort (Lovatt & Wilson, 1994), and cause unwanted movement of pollutants (Lim, Cho, & Kim, 2011) and smoke (Zhang, Ji, Huo, Yuan, & Yang, 2006). Traditional modelling methods for stack effect consisted of using the multi-zone model, CFD, and coupled models (Chen, 2009). The need to acquire cumbersome data such as air leakage, large computational processing power required, and model complexity are limitations for the current approach. Data-driven modelling present a novel approach to model stack effect, through statistical learning (Djenouri, Laidi, Djenouri, & Balasingham, 2019). A lack of literature investigating this approach for modelling stack effect suggest there is an opportunity to develop a novel predictive model using a data-driven approach. This paper presents the results of using return air temperature as a proxy for thermal comfort issues caused by stack effect in the heating season. The results from this paper are the first step in a larger framework to develop a data-driven predictive stack effect model, where future work will investigate the pressure differential aspect of stack effect. This paper is organized as follows: (1) literature review regarding stack effect and data-driven modelling; (2) methodology; (3) results of the data-driven model; (4) discussion of limitations, and future potential for the research; (5) conclusion.

Literature Review

Traditionally stack effect has been modelled using a physics-based approach, which could be split into the multi-zone, CFD, and coupled models. Yu et al. (2017) utilized the multi-zone model to simulate stack effect in their case study building, and running various pressurization schemes to determine the most effective method in combating stack effect. The model was successfully reduced stack effect in the case study building, which was validated with field results but encountered difficulties with calibration due to inaccurate estimates of air leakage CFD models. Wong and Heryanto (2004) modelled stack effect using CFD to explore potential opportunities to leverage to enhance natural ventilation, but the model was confined to one room due to the large computational power required to simulate a full scale building. Coupled models combine the multi-zone method with either CFD or data-driven algorithms to leverage the strengths of both, while mitigating their respective weaknesses but raises the overall model complexity. Yoon et al. (2015) created a novel modelling approach, which coupled the multi-zone model with genetic algorithms to optimize the leakage area. The results were more accurate than the multi-zone model, and were significantly less computationally-intensive than CFD. When using any of these methods, the airtightness of the building, zone temperatures, zone volume, leakage areas, weather data, and HVAC characteristics (Yoon, Seo, Cho, & Song, 2015) have been identified as key parameters to capture in order to model stack effect. ASHRAE standard 55 (ASHRAE, 2017) recommends to maintain thermal comfort indoor temperatures must be within 20°C to 28°C, consistent with the Fanger model (Fanger, 1970). The optimal temperature is dependent on occupant clothing, relative humidity, and air speed. Return air temperature can serves as an indication of zone air temperature.

Previous research investigated methods to mitigate stack effect in buildings. Lim et al. (2020), conducted a comprehensive study to investigate the available countermeasures presently, which is split into architectural solutions, and mechanical solutions. Architectural solutions focuses on improving air tightness, and compartmentalization (Li J. , 2018) of the building to decrease the overall air leakage area, which effectively minimizes the stack effect problem. Mechanical solutions leverage the HVAC system to pressurize the entire building, or portions in order to reduce the pressure differential between the interior and exterior. The benefit of a mechanical system is it does not require costly changes to the building layout, or envelope since a HVAC system is already in place for high-rise buildings.

Data-driven modelling is rapidly growing in since it can provide accurate prediction and classification models while significantly reducing model complexity and computational cost. Machine learning is split into four categories: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning (Molina-Solana, Ros, Ruiz, Gómez-Romero, & Martin-Bautista, 2017). In the context of smart buildings, supervised learning is predominantly used since it maps the inputs to the outputs from the training dataset, building a relationship from known parameters. The literature review discovered a gap in the current research, as no studies have yet to explore modelling stack effect using a purely data-driven method. Data-driven models have seen widespread use in smart building research, for applications such as energy optimization (Seyedzadeh, Rahimian, Glesk, & Roper, 2018), and fault detection (Shohet, Kandil, Wang, & McArthur, 2020). Machine learning consists of many different algorithms that each have their own strengths and weaknesses, therefore, the best algorithm changes depending on the problem. Linear discriminate analysis (LDA) has been used to classify faults in chiller operation (Li, Hu, & Spanos, 2016). Shohet et al. (2020) explored leveraging machine learning for fault detection in non-condensing boilers, comparing the performance of decision trees, random forest (RF), K-nearest neighbour (KNN), and support vector machines (SVM). Capozzoli et al. (2015) found success in using artificial neural networks (ANN) for fault detection in energy consumption of a cluster of smart office building. Based on the literature review, these five algorithms have been identified with potential to model stack effect due to past success in the building context, and the different learning approaches.

Methods

The study was done in the following steps: (1) data collection; (2) data visualization; (3) data pre-processing; (4) model training and analysis. These steps are visualized in Figure 2.

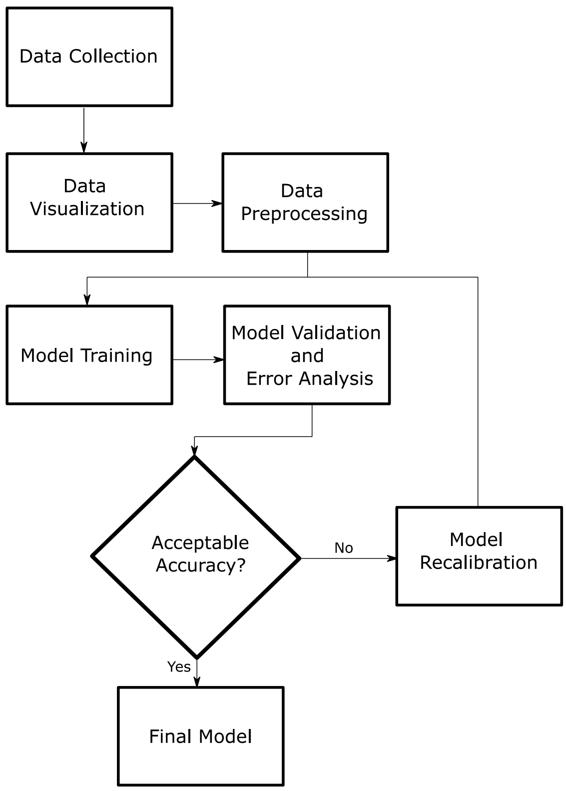


Figure 2 Workflow Diagram

Data Collection

Data was collected from 223m case study building: a high-rise office tower 223m tall, with 56 above-grade floors and a total area of 123,422 m², located in Toronto, Canada. The 14th and 43rd floor serve as the mechanical rooms. HVAC zoning is split into east and west for both the perimeter and core zones. Each zone is further split into a north and south subzone. The plaza is an expectation to rule as it only has the west and east zone designation. The AHUs located on the 14th floor mechanical room service floors 2 to 28, while AHUs on the 43rd floor service floors 29 to 55.

The key parameters to capture were derived from Yoon et al. (Yoon, Seo, Cho, & Song, 2015) who identified the critical data points required to create a physics-based model for stack effect. Based on their work, the following parameters were chosen for data collection: weather data, HVAC trend data, and interior temperature. These are the critical variables to consider when creating a physics-based stack effect model, and did not require labour-intensive testing. Building envelope air tightness, and leakage rates were also identified key parameters to create a multi-zone model but they were omitted as it was infeasible to measure these variables in such a large structure, particularly due with Covid-19 restrictions in place. Pressure differentials were another key parameter omitted for this research due to delays in sensor instalment related to the current Covid-19 lockdown in Toronto, Canada, but will be implemented in future work. The HVAC trend data entailed the supply, and return air temperatures of the fans, supply temperature set points, and if the fan was on or off. Return fan temperature was used to estimate the interior temperature due to the lack of granular temperature sensor data available. The weather data consisted of the outdoor air temperature, relative humidity, wind speed, and wind direction which were taken from a local weather station via Tomorrow.io (Tomorrow.io, 2020). Data collection took place between November to December 2020, as the focus of the research is for stack effect during the heating season. The sensors, and support programs used for data collection are detailed in Table 1.

Table 1. Data Collection Sensors and Support Programs

Element	Sensor/Support Program
Weather	Tomorrow.io
HVAC Trend	Metasys
Interior Temperature	Metasys

Data Pre-processing

Data pre-processing is an intermediate step required to refine the dataset of any imperfections, inconsistencies, and redundancies (García, Ramírez-Gallego, Luengo, Benítez, & Herrera, 2016). The focus of pre-processing in this context was to convert all the streamed data into the same time intervals, remove outliers, and interpolate missing values in the dataset. The data was stored on excel files, while the pre-processing workflow was done through python using the *pandas* library (McKinney, 2011) for data manipulation. All the data was resampled to five-minute intervals to keep the time series uniform. Outliers were filtered based on observations from the data visualizations, and missing values were linearly interpolated.

Return air temperature was used as an instrumental variable to evaluate potential for stack effect related thermal comfort issues. Labels were added to each return fan based on their temperature using ASHRAE 55 to set the bare requirement for thermal comfort. Based on the standard, the minimum air temperature required was set at 21ºC assuming normal conditions for an office space. When the return air temperature was below 21ºC the fan was flagged at that instance, which meant the zone had failed to meet the set thermal comfort standard. Supply air temperature was not included in the labelling process to limit potential contamination from other factors such as boiler performance.

The fans were further classified based on the zone, by adding a green or red label based on the comparison between the upper and lower return fans. Green meant there was no potential for stack effect related thermal comfort issues while red meant stack effect had potential to impact thermal comfort. Cases where both the lower and upper return fans passed the thermal comfort check, both failed, or when only the upper fan failed were classified as green. In these instances, stack effect had little impact on thermal comfort based on the proxy and was therefore evaluated as a non-issue. When the lower return fan failed but the upper return fan passed it implied stack effect was potentially moving the heated upwards causing this discrepancy, therefore it was classified as red. Binary classification was utilized instead of multi-class to reduce the class imbalance problem. The zoning is illustrated in Figures 3 and 4.

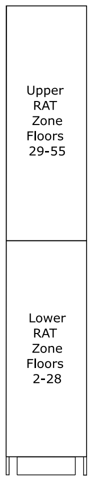


Figure 3 Upper and Lower RAT Zoning Diagram

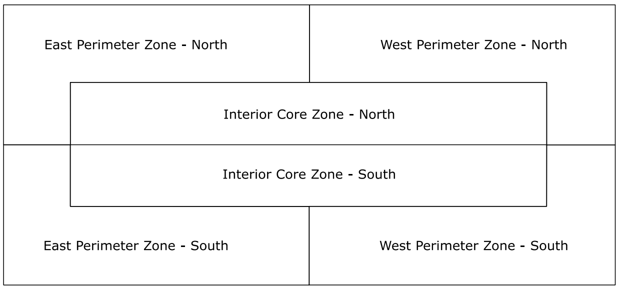


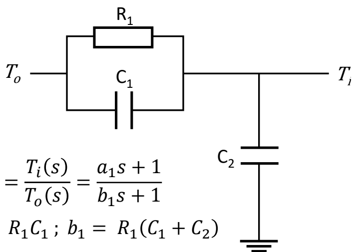
Figure 4 HVAC Zoning for Typical Floor

Data Visualization

The data discussed in data collection was plotted onto scatter plots, and histograms to understand the underlying relationships of each variable. The variables were plotted in various configurations to investigate correlations in the data, and develop a baseline understanding of the building characteristics. Matplotlib, a library package in python, was leveraged to visualize the data (Hunter, 2007). The predictive model will be incorporated into a digital twin, for real time tracking.

Feature Selection - RAT as a Proxy for Stack Effect

The model presented uses return air temperatures as a proxy for stack effect. A simplified thermal circuit model demonstrates how the return air (average space) temperature is impacted by infiltration caused by stack effect. Consider the half-building (block) depicted in Figure 5. R_1 and C_1 correspond to the envelope thermal resistance and thermal mass, respectively, and C_2 represents the total thermal mass of the block, including air. Assuming each block as a linear, time-invariant system, we consider the simplified transfer function F(s) that approximates the block's interior temperature T_i response to the outside temperature T_o variations. We assume that both blocks have similar thermal properties R_1, C_1 and C_2.



$$F(s) = \frac{T_i(s)}{T_o(s)} = \frac{a_1 s + 1}{b_1 s + 1}$$

$$a_1 = R_1 C_1 \; ; \; b_1 = R_1(C_1 + C_2)$$

Figure 5 Simplified thermal circuit of a half-building

We simulate the interior temperature (considered equal to RAT) evolution of each block during the first weekend of December 2020. The initial simulated temperature is made equal to the RAT initial value. We adjust the values of R_1, C_1 and C_2 so that the model response matches the real interior temperature of the lower half of the building as shown in Figure 6b.

When considering the same model, Figure 6a shows that the upper half temperature decreases much less significantly than expected. Given that the envelope thermal properties R1, C1 are constant, matching the simulated and the observed responses can only be achieved by increasing C2, representing an additional thermal mass that slows the building response- as in Figure 6a. Because the building halves have the same number of levels and nearly the same interior space, this increase in thermal mass (ΔC_2) is explained by the rising (warm) air from the lower half of the building to the upper, due to stack effect; or conversely by considering the lower half to have less effective thermal mass due to infiltration from the outdoors, which speeds the thermal response.

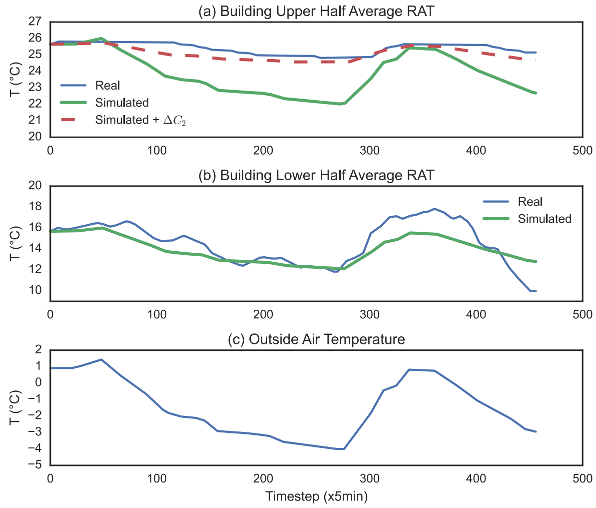


Figure 6 Real and simulated temperature of the lower and upper halves of the building on a weekend

Model Training

The model leveraged outdoor air temperature, relative humidity, supply air temperature, fan status, and supply air set point as the features for the model. Green and red acted as the labels for the model. The dataset was split into 70% for training and validation, and 30% for testing. The training and validation datasets were combined in order to perform a sliding window cross validation to tune the hyperparameters of the model since the temporal order is crucial for time series data (Lu, Cheng, Ma, & Hu, 2020). The window length was set at seven, where the data was further divided into 65% for training, and 35% for testing. Final testing was performed using the same splits but substituting the test data with the remaining 30% of the data for each model. Each algorithm used the same dataset to eliminate bias in the results. Model training was implemented using python through the scikit-learn package that simplifies the workflow (Pedregosa, et al., 2011). As identified in the literature review, the following algorithms are best suited for modelling stack effect: LDA (Li, Hu, & Spanos, 2016), KNN, RF, SVM (Shohet, Kandil, Wang, & McArthur, 2020), and MLP (Capozzoli, Lauro, & Khan, 2015).

Model Validation and Error Analysis

The models were validated using fundamental classification metrics (Géron, 2017), as shown in equations (1) to (4). No information rate (NIR), also known as null accuracy, was also used in the analysis to gauge the statistical significance of the model accuracy (McArthur, et al., 2018). The NIR is the accuracy of the model if the largest class label was always picked, therefore, an accuracy higher than the NIR implied a model with statistical significance.

$$Recall/Sensitivity = \frac{TP}{P} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$f1\_Score = \frac{2}{\frac{1}{precision}+\frac{1}{recall}} \quad (3)$$

$$Accuracy = \frac{TP+TN}{P+N} \quad (4)$$

Based on the metrics, each algorithm was recalibrated to tune their hyperparameters in order to optimize the cost function. Hyperparameter optimization was done through the grid search technique, where every possible set of parameters is tested to determine the combination. The parameter combinations are listed in Table 2. As the training leveraged the sliding window method the metrics from each frame was averaged for the final result. After every algorithm has been optimized, the best suited algorithm is chosen based on the accuracy, and the statistical significance to the model.

Table 2 Hyperparameters Tested in Grid Search

Algorithm	Hyperparameter	Range
LDA	Solver	{svd, lsqr}
KNN	n_neighbors	{3,5,7}
RF	n_estimators	{25,50,100}
	max_depth	{1,2,3}
	min_samples_split	{2,5,7}
	min_samples_leaf	{1,2,5}
MLP	Solver	{lbfgs, sgd, adam}
SVM	Kernel	{linear, rbf, poly, sigmoid}
	C	{1,10,50}
	Gamma	{scale, auto}
	Degree (for poly)	{1,2,5}

Results

This section presents the results from the data mining to develop an understanding of return air temperature as a proxy for stack effect. Data mining was done from November 2020 to December 2020 to reflect the building operation under normal heating conditions. The model results are presented in the subsequent subsection.

Data Visualization of Case Study Building

As shown in Figure 7, the average return air temperature (RAT) of the fans located in the lower of the building (floor 14 AHU) followed a typical heating schedule for a high-rise building, where the building was conditioned based on the weather, and occupancy. The outdoor air temperature, shown in Figure 7, illustrated good correlation with the average lower RAT as both trended upwards during the day when the building is heated, and outdoor air is warmer due to the presence of solar radiation. The same principles applied for the night when the average outdoor air temperature dropped, and heating is minimum due to the lack of occupancy in the building resulting in a downward spike. Figure 7 illustrates the typical RAT behaviour of the fans from the upper half the building (floor 43 AHU) which also exhibited small spikes based on occupancy and weather, but showed much less variance than the average lower RAT. The lack of variance indicated that floors 29 to 55 were heated throughout the day due to the presence of a constant heat source, as shown by the almost constant temperature. As heating is reduced at night, this unexplained heating is evidence of stack effect. Stack effect causes the warm air to rise due to the increased air infiltration rate caused by the pressure differential. Heated air is likely rising from the lower levels causing the upper levels to have a constant source of heat, and consequently lower temperatures in the lower levels. This is further reinforced by the trends in the outdoor air temperature since it is lower during the night, it intensifies stack effect due to the larger pressure differentials.

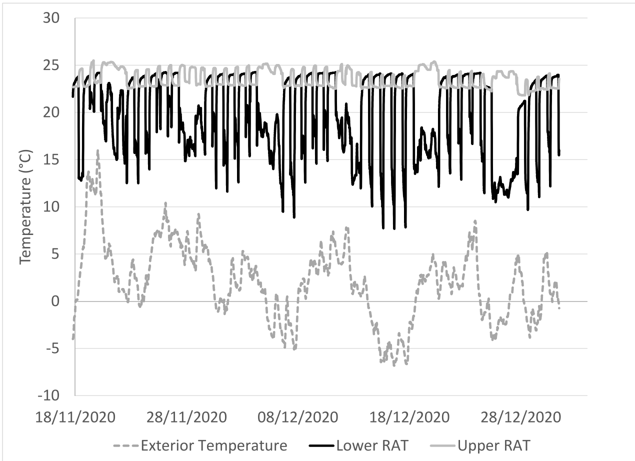


Figure 7 Average RAT for upper half, lower half of the building, and exterior temperature

Model Results

A sliding window analysis was implemented for cross validation to reduce overfitting of the model due to the seasonality of the data. A window length of 7, with 65% of each window used for training, and 35% for testing was found to be the best length as it reduced the seasonality of the data, while maintaining class balance between the outputs. Using the grid search method the best hyperparameters for each algorithm are listed in Table 3. Using the classification metrics, each model was evaluated using the test set with the results presented in Table 4. Each model showed statistical significance as the average accuracy surpassed the NIR. RF performed the best with 93% prediction accuracy, with high values for precision, recall, and f1-score indicating good overall model performance. LDA, MLP, and SVM performed similarly, with the largest difference being the accuracy between LDA and SVM at 1.8%. KNN showed the worst performance out of all the models, with poor accuracy and high variance for the predictions.

Table 3 Optimal Hyperparameters for each Model

Algorithm	Hyperparameter	Optimal Value
LDA	Solver	lsqr
KNN	n_neighbors	3
RF	n_estimators	25
	max_depth	2
	min_samples_split	5
	min_samples_leaf	2
MLP	Solver	adam
SVM	Kernel	poly
	C	1
	Gamma	scale
	Degree (for poly)	1

Table 4 Classification Results

Algorithm	LDA	KNN	RF	MLP	SVM
Accuracy	0.870	0.671	0.930	0.861	0.852
Precision	0.878	0.753	0.931	0.870	0.876
Recall	0.884	0.684	0.932	0.854	0.865
f1-Score	0.869	0.636	0.929	0.849	0.846
NIR	0.593	0.593	0.593	0.593	0.593

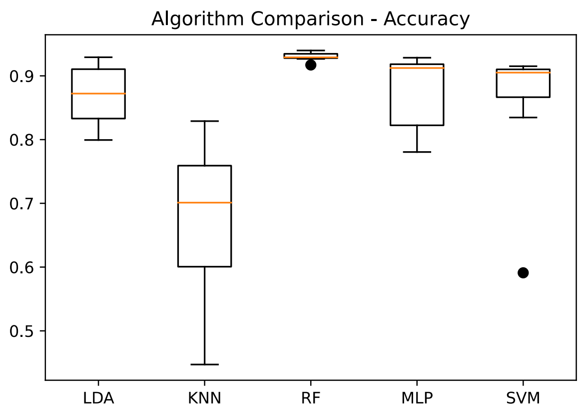


Figure 8 Boxplot of Accuracy

This poor performance is reinforced with Figure 8, which plotted the accuracy for all iterations of the sliding of each algorithm. In certain iterations, the KNN dropped below the NIR indicating no statistical significance from the model. The boxplot also illustrates RF had good generalization as the accuracy stayed almost identical for each iteration, while LDA, MLP, and SVM had average generalization due to the variance in the accuracies. As depicted in Figure 9, is the typical confusion matrix for random forest showing high classification accuracy for both the red and green labels. Shown in Figure 10, was an interesting error which was observed for LDA where it was unable to successfully predict red classes. This was consistent for many of the windows, but was alarming the same errors were occurring even as the training data window moved closer to the dates of the test data.

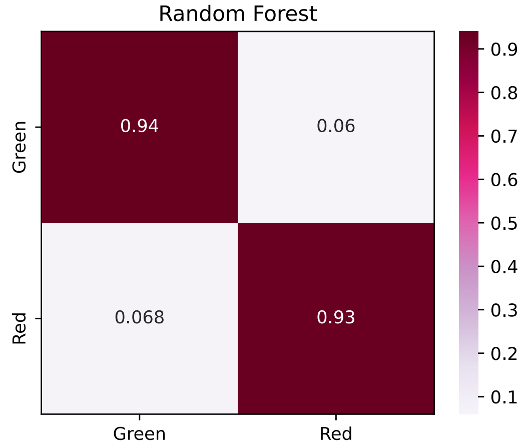


Figure 9 Random Forest Average Confusion Matrix

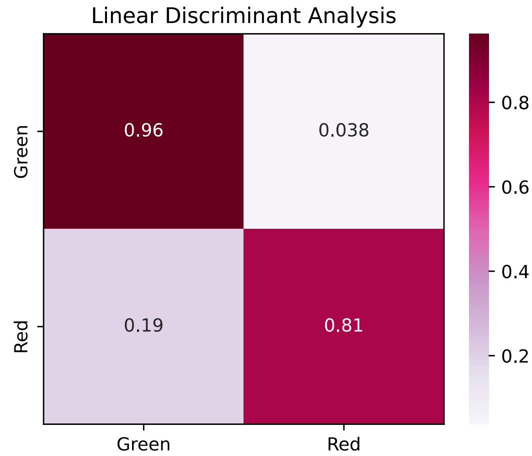


Figure 10 LDA Average Confusion Matrix

To better understand the impact of the features in the learning process, the *feature_importances_* function for the RF model from scikit-learn was utilized to visualize the weights of each feature. The weights of the most important features are depicted illustrated in Figure 11. From the feature importance, it is shown the model is heavily learning from characteristics of the core zone supply fans. With the expectation of one, all the fans listed in Figure 11 serve the core zone for either the north or south side for both the upper and lower portion of the building. This is an indication the model is learning based on the conditions of the core zones, which is another piece of evidence that RAT as a proxy for stack effect is a valid assumption because the core zones contain the elevator shafts. The shafts act as the major highway for stack effect movement as they span the entirety of the building height allowing for uninterrupted air movement. By using the core zones as a major criterion in the prediction of thermal comfort issues related to stack effect, it implied the model is building a correlation between RAT and stack effect. *Low_SF_NW_Per_O/F_27* (lower half supply fan serving the northwest perimeter on/off status for 27 floors; similar nomenclature used for all fans, where O/F = on/off;

SAT=supply air temperature and SP = setpoint) is the outlier in this data, as it indicated the on/off status of the western perimeter supply fan for the lower half building.

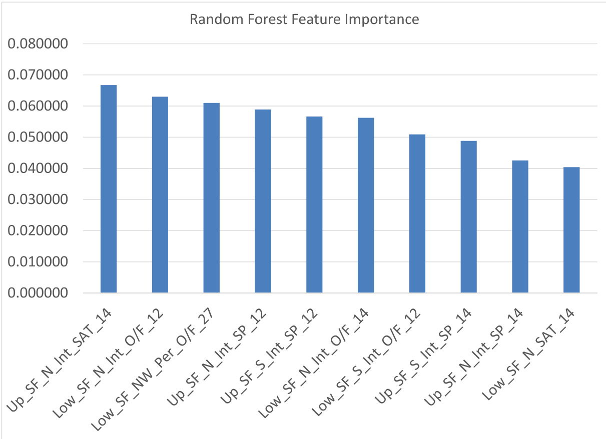


Figure 11 Feature Importance from Random Forest

Discussion

Based on the results, a novel data-driven modelling approach for stack effect using return air temperature as a proxy was developed, where random forest showed the greatest potential to predict thermal comfort issues related to stack effect. Random forest achieved the highest average accuracy, while maintain low variance in the results which indicated good generalization of the model. KNN was not effective at predicting stack effect related thermal comfort issues due to low accuracy and high variance, attributed to the lazy learning approach the algorithm utilizes which is not suited for this problem. LDA, MLP, and SVM performed similarity well but were outperformed by random forest on all metrics. It was noted that the green class was misclassified more frequently than the red class for all models. This was explained by a small class imbalance problem found in the testing dataset as it contained 41% green classes, and 59% red classes, while the training data had a distribution closer to 50%.

Collecting only two months of data for model development resulted in a seasonal solution. Further, extreme weather events were not captured, which reduces the generalizability of the model. This was largely overcome by using the sliding window method, but ideally a larger dataset would be used to improve model training, and generalization. The building floor temperatures were also evaluated to compare their trends with the RAT. The floor temperatures followed a similar behaviour to the RAT but showed less variation in the temperature. This is potentially due to the verticality of the return air duct, which are more susceptible to stack effect versus the horizontal layout of the floors. Future research will look at co-evaluating the RAT with the building floor temperature to develop a better understanding of using RAT as a proxy for stack effect.

The scope of this research was also limited to the heating season as the focus was evaluating thermal comfort issues related to stack effect by using RAT as a proxy. In the cooling season, stack effect is reversed which could be interesting to observe for future work, and potentially beneficial to thermal comfort due to the additional exfiltration. The lack of pressure differential data is a limiting factor in this research, as they are a crucial piece of data to validate the assumption of using RAT as a proxy for thermal comfort issues related to stack effect. The original scope of this research included the pressure differential data as part of the model but was delayed due to the Covid-19 lockdown in Toronto, Canada. This data will be incorporated into future iterations of this research in order to create a more robust predictive model that evaluates stack effect impact on thermal comfort.

Conclusion

This research provided a framework for future data-driven stack models. Using return air temperature as a proxy to predict thermal comfort issues related to stack effect was proven feasible as the data from the upper and lower parts of the fans indicated influence due to stack effect. This was reinforced by the learning weights of the prediction model, since the core zone parameters were a pivot factor in classification, but pressure differential sensors are the final component required to validate this method. The study concluded that random forest is the most suited algorithm for predicting thermal comfort issues related to stack effect as it achieved 93% accuracy while maintaining good generalization. Future research will leverage targeted testing in order to acquire data during various conditions in order to expand the captured conditions to further improve the generalizability of the predictions. The end goal of this research is to create a recommender algorithm to predict instances where stack effect is detrimental to thermal comfort, and energy consumption in order to output corrective actions to mitigate the impact. The recommender algorithm will be visualized via a digital twin to aid facilities with an effective tool for stack effect management.

Acknowledgements

This study was funded by the Natural Science and Engineering Research Council of Canada (NSERC) under the Engage program [Grant # EGP 543443-19]. In addition, Mr. Chang has been supported under the NSERC CREATE Program [Grant # CREATE 510284-2018]. The authors also gratefully acknowledge the support and contribution of our industry partners, H.H. Angus and Associates and the Cadillac Fairview Corporation.

References

ASHRAE. (2017). ANSI/ASHRAE Standard 55-2017 Thermal Environmental Conditions for Human Occupancy. Altanta, Georgia, United States of America: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc. (ASHRAE).

Capozzoli, A., Lauro, F., & Khan, I. (2015). Fault detection analysis using data mining techniques for a cluster of smart office buildings. *Expert Systems with Applications, 42*(9), 4324-4338.

Chen, Q. (2009). Ventilation performance prediction for buildings: A method overview and recent applications. *Building and Environment, 44*(4), 848-858.

Djenouri, D., Laidi, R., Djenouri, Y., & Balasingham, I. (2019). Machine Learning for Smart Building Applications: Review and Taxonomy. *ACM Computing Surveys, 52*(2), 1-36.

Fanger, P. O. (1970). Thermal Comfort—Analysis and Applications in Environmental Engineering. *Danish Technical Press*.

García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J., & Herrera, F. (2016). Big data preprocessing: methods and prospects. *Big Data Analytics, 1*(1), 1-22.

Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn Concepts, Tools, and Techniques to Build Intelligent Systems.* Sebastopol, California : O'Reilly Media, Inc.

Hunter, J. (2007). Matplotlib: A 2D graphics environment. *Computing in Science Engineering, 9*(3), 90-95.

Jung-yeon, Y., Song, K.-d., & Dong-woo, C. (2017). Resolving Stack Effect Problems in a High-Rise Office Building by Mechanical Pressurization. *Sustainability, 9*(10), 1731.

Li, D., Hu, G., & Spanos, C. (2016). A data-driven strategy for detection and diagnosis of building chiller faults using linear discriminant analysis. *Energy and Buildings, 128*, 519-529.

Li, J. (2018). Using Compartmentalization to Mitigate the Impacts of Stack Effect in Tall Residential Buildings. Toronto: University of Toronto.

Lim, H., Seo, J., Song, D., Yoon, S., & Kim, J. (2020). Interaction analysis of countermeasures for the stack effect in a high-rise office building. *Building and Environment, 168*, 106530.

Lim, T., Cho, J., & Kim, B. (2011). Predictions and measurements of the stack effect on indoor airborne virus transmission in a high-rise hospital building. *Building and Environment, 46*(12), 2413-2424.

Lovatt, J. E., & Wilson, A. G. (1994). Stack effect in tall buildings. *ASHRAE Transactions, 100(2)*, 420-431.

Lu, H., Cheng, F., Ma, X., & Hu, G. (2020). Short-term prediction of building energy consumption employing an improved extreme gradient boosting model: A case study of an intake tower. *Energy, 203*.

McArthur, J., Shahbazi, N., Fok, R., Raghubar, C., Bortoluzzi, B., & An, A. (2018). Machine learning and BIM visualization for maintenance issue classification and enhanced data collection. *Advanced Engineering Informatics, 38*, 101-112.

McKinney, W. (2011). pandas: a Foundational Python Library for Data Analysis and Statistics. *Python High Performance Science Computer, 14*(9).

Mijorski, S., & Cammelli, S. (2016). Stack Effect in High-Rise Buildings: A Review. *International Journal of High-Rise Buildings, 5*(4), 327-338.

Molina-Solana, M., Ros, M., Ruiz, M., Gómez-Romero, J., & Martin-Bautista, M. (2017). Data science for building energy management: A review. *Renewable and Sustainable Energy Reviews, 70*, 598-609.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Cournapeau, D. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*, 2825-2830.

Seyedzadeh, S., Rahimian, F., Glesk, I., & Roper, M. (2018). Machine learning for estimation of building energy consumption and performance: a review. *Visualization in Engineering, 6*(1), 1-20.

Shohet, R., Kandil, M. S., Wang, Y., & McArthur, J. (2020). Fault detection for non-condensing boilers using simulated building automation system sensor data. *Advanced Engineering Informatics, 46*, 1-16.

Tomorrow.io. (2020). *Tomorrow.io API*. Retrieved from Tomorrow.io: https://www.tomorrow.io/weather-api/

Wilson, A., & Tamura, G. (1968). Stack effect in buildings. *Canadian Building Digest, 104*.

Wong, N., & Heryanto, H. (2004). The study of active stack effect to enhance natural ventilation using wind tunnel and computational fluid dynamics (CFD) simulations. *Energy and Buildings, 36*(7), 668-678.

Yang, D., Du, T., Peng, S., & Li, B. (2013). A model for analysis of convection induced by stack effect in a shaft with warm airflow expelled from adjacent space. *Energy and Buildings, 62*, 107-115.

Yoon, S., Seo, J., Cho, W., & Song, D. (2015). A calibration method for whole-building airflow simulation in high-rise residential buildings. *Building and Environment, 85*, 253-262.

Yoon, S., Song, D., Kim, J., & Lim, H. (2019). Stack-driven infiltration and heating load differences by floor in high-rise residential buildings. *Building and Environment, 157*, 366-379.

Yu, J.-Y., Song, K.-D., & Cho, D.-W. (2017). Resolving Stack Effect Problems in a High-Rise Office Building by Mechanical Pressurization. *Sustainability, 9*(10), 1731.

Zhang, J., Ji, J., Huo, R., Yuan, H., & Yang, R. (2006). A Comparison of Simulation and Experiment on Stack Effect in Long Vertical Shaft. *Journal of Fire Sciences, 24*(2), 121-135.