Assessment of Micro-Organism Growth Risk on Filters with Machine Learning

Andreas Hantsch^{1,*} and Sabine Döge²

¹CLOUD&HEAT Technologies GmbH, Königsbrücker Straße 96, 01099 Dresden, Germany

²Institute of Air Handling and Refrigeration gGmbH, Bertolt-Brecht-Allee 20, 01309 Dresden, Germany

Abstract. Modern buildings usually have a practically air-tight envelope. Therefore, mechanical ventilation is very often necessary. A crucial part of the system is the filter which allows to create an atmosphere which is free of dust, aerosols, and pollen. As organic material accumulates on the filter surface, the risk of microorganism growth rises. This may yield health issues especially for the occupants of buildings in humid regions. For this purpose, a test filter with electrodes has been designed which allowed to measure its electro-magnetic properties, such as resistance, capacitance, and impedance as an indicator for the micro-organism growth risk. After some preliminary tests, electrodes of stainless steel and the electrical capacitance have been selected due to their best durability and signal-to-noise-ratio. The test filter has been implemented in the HVAC system of the institute in order to aggregate data for different abnormal and normal operation data. A machine learning algorithm has been trained successfully to detect anomalies of the filter behaviour and therefore provided more insight than pressure drop measurement alone. Finally, the change intervals of the filter could be adapted to the real degree of pollution without the requirement for visual observation in order to provide best air conditions.

1 Introduction

The Paris Agreement forces the signing states to reduce their carbon dioxide emissions [1]. For this purpose, the consumption of fossil fuels has to be reduced dramatically that can be achieved by increasing both the renewable share of the energy supply and the energy efficiency [2]. One measure is to improve the insulation of building that leads, however, to practically air-tight building envelopes. Therefore, mechanical ventilation systems are required in order to assure a healthy ventilation rate. An important task of a mechanical ventilation system is to prevent pollutants from outside entering the building, for which purpose filters are employed. These filters are prone to fouling and, if the conditions are humid enough, to microorganism growth, yielding severe health issues. Thus, filters are usually changed on a regular basis. This procedure is, however, most likely too often or too seldom.

Methods of machine learning might help solving this optimisation problem. They are widely used for internetbased services such as search engines, news feeds, image classification, etc. They became also quite popular in conjunction with the so-called internet of things, where devices are connected over internet in order to provide some benefit to the user [3]. First academic studies investigate the utilisation of machine learning for HVAC systems [4–11].

In particular, algorithms for anomaly detection gained recently some interest. There are some publications with the intention to detect malfunctions in HVAC systems, such as Ref. [12-17].

Based on this earlier work, the objective of this contribution is to provide a set-up and machine learning algorithm to assess the micro-organism growth risk of filters. The paper is organised as follows: Section 2 presents the experimental set-up with the HVAC and measurement systems as well as the measurement series. Moreover, the machine learning algorithm with the model and the workflow is presented. Section 3 provides the results and discussion.

2 Experimental investigation

2.1 HVAC system

The experimental investigation is carried out in one of two parallel operable HVAC systems of the room air-flow laboratory of the institute (see Fig. 1 for an

Corresponding author: andreas.hantsch@cloudandheat.com



Figure 1. Schematic drawing of the experimental set-up: HVAC system of the room air-flow laboratory (top) and detail of the test section (bottom) with the measurement devices for temperature (T), relative humidity (φ), velocity (ν), pressure drop (Δp), electrical capacity C, and the function generator for the voltage U(t)

illustration) [18]. It consists of the following major components: air intake (A), flap (B), intake filter (C), recirculation air flap (D), cooler (E), heater (F), supply fan (G), humidification (H), silencer (I), air exhaust (J), flow-rate control (K), fire protection flap (L), air emission (M) into the room air-flow laboratory and the air extraction (N) from there. The test section is placed just before the treated air enters the distribution channel (see magnified part of Fig. 1).

2.2 Measurement system

The subject under consideration is the filter highlighted in Figure 1 that is prepared with electrodes made of stainless steel oriented perpendicularly to the flow direction. The electrodes are connected to a function generator providing sinusoidal voltage signals. As a measure for the humidity of the filter, the electrical capacitance exhibits the best signal-to-noise ratio compared to resistance, inductance, and impedance, as obtained from preliminary tests.

The interrelation between electrical capacitance and the humidity of the filter via the permittivity ε is provided by [19]:

$$C = \frac{\pi \varepsilon L}{\operatorname{arcosh}\left(\frac{S}{D}\right)}.$$
 (1)

Herein, L, D, and S are the length, diameter and distances of the electrodes, respectively. In this study, the electrical capacitance is measured and depends upon the humidity

of the filter via the permittivity ε . Moreover, the pressure difference between the ambient and the inlet of the filter $\Delta p_{\rm fa}$, the pressure drop over the filter $\Delta p_{\rm f}$, the ambient pressure $p_{\rm amb}$, the temperature *T*, relative humidity φ , and velocity of air v at the inlet of the filter are measured (see Figure 1 for details and Table 1 for the measurement devices and their uncertainties).

| Table 1. Measurement devices and their uncertaintie |
|---|
|---|

| Quantity | Device | Uncertainty |
|--------------------|------------------------|-------------|
| $\Delta p_{ m fa}$ | Ashcroft, 0100 Pa | ±0,25 Pa |
| $\Delta p_{ m f}$ | Ashcroft, 0500 Pa | ±1,25 Pa |
| $p_{ m amb}$ | Vaisala PMB 100 | ±0.3 hPa |
| Т | Vaisala HMT 100 | ±0,3 K |
| φ | Vaisala HMT 100 | ±3%-points |
| v | Hot-wire anemometer | ±5% |
| С | LCR measurement bridge | ±<0.1 % |

It is convenient to non-dimensionalise the pressure drop Δp as it depends strongly upon the air velocity v. This yields the pressure drop coefficient [20]:

$$\zeta = \frac{2\Delta p}{\varrho v^2}.$$
 (2)

Herein, ρ is the density of moist air (Index ma) that is calculated with [21]:

$$\varrho = \frac{p}{R_{\rm ma}T}, \text{ and}$$
(3)

$$R_{\rm ma} = \frac{R_{\rm da}}{1 - \left(\frac{\varphi p_{\rm sat}}{p}\right) \left(1 - \frac{R_{\rm da}}{R_{\rm m}}\right)}.$$
(4)

Herein, *p* is the absolute pressure, *T* is the temperature, φ is the relative humidity, $p_{sat}(T)$ is the saturation pressure, and $R_{da} = 287.2 \text{ J/(kg K)}$ and $R_{m} = 461.4 \text{ J/(kg K)}$ are the specific gas constants of dry air (Index da) and moisture (Index m), respectively [21].

2.3 Measurement series

The filter mounted in the HVAC channel was employed in the normal operation of the institute's room air-flow laboratory in 2017. The measurements were carried out in two periods (May–August, November). During these periods, the afore-mentioned data points were recorded with one second time intervals. The time interval is not necessary for this investigation but has been determined due to secondary experiments not presented here. Moreover, distinguished experiments were carried out in which the filter has been humidified manually by sprays of water droplets.

2.4 Machine learning algorithm

Machine learning algorithms can be classified into two major groups. They are, firstly, supervised learning algorithms – where training data has to be manually labelled – and, secondly, unsupervised learning algorithms – where the algorithm detects patterns in the data on its own. The algorithm employed here belongs to the second group and is in particular a so-called anomaly detection algorithm [22, 23].

The data set obtained in the measurements is split into the following three subsets with their respective use cases:

- 1) **Training set**: train the algorithm (60% of the data)
- 2) **Cross validation set**: fine-tune the algorithm (20% of the data)
- 3) **Test set**: assess the quality of the algorithm (20% of the data)

The major intention is to use the test set for the assessment of the quality only. In order to manage the data, Python 3 libraries are employed, in particular Pandas [24] and Scikit-learn [25,26] since they provide most flexibility for an automated workflow in which the anomaly detection algorithm was implemented. The data visualisation is carried out with the library matplotlib.

The work flow is as follows:

- 1) Data acquisition
- 2) Data preparation (i.e., calculating derived quantities according to Eq. (2))

- 3) Data normalisation with the mean value (Eq. (6)) and variance (Eq. (7)) of the training set
- 4) Training algorithm with the training data set
- 5) Tuning the algorithm with a cross validation data set
- 6) Testing the tuned algorithm with a test set
- 7) Visualisation of the results

The probability of a set of *n* parameters $x_i \mid i \in [1, n]$ to be abnormal is determined with [25,26]:

$$P(x_i, \mu_i, \sigma_i^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right].$$
 (5)

Herein, x_i is the parameter, μ_i and σ_i^2 are the mean value and the variance of the set, respectively. These are defined as follows:

$$u_i = \frac{1}{m} \sum_{j=1}^m x_{i,j},$$
 (6)

$$\sigma_i^2 = \frac{1}{m} \sum_{j=1}^m (x_{i,j} - \mu_i)^2.$$
(7)

The values for these variables are determined for each parameter in the training step. A set of parameter values is abnormal, if

$$P(x_i, \mu_i, \sigma_i^2) < P_{\text{crit}}, \tag{8}$$

whereby $P_{\rm crit}$ is a critical probability. Evaluation of the cross-validation set revealed that the critical probability shall be $P_{\rm crit} = 10^{-5}$ allowing to find all anomalies and to prevent false-positives. In order to improve the accuracy of the machine learning algorithm, it is convenient to normalise the quantities as follows:

$$\tilde{x}_{i,j} = \frac{x_{i,j} - \mu_i}{\sigma_i}.$$
(9)

It shall be stressed that the algorithm actually struggles if the electrical capacitance is employed in Farad. For the remainder of the paper, normalised quantities are considered.

3 Results and discussion

Figure 2 visualises the density distribution of both normalised quantities: electrical capacitance and pressure drop coefficient. It can be observed that the normalised electrical capacitance has a range of [-2.5, 2] with three moderate peaks in the proximity of the mean value. Contrary, the normalised pressure drop coefficient has a very narrow peak slightly below the mean value and a smaller peak approximately at 2.5. Moreover, both quantities have maximum values at 10.3 (capacitance) and 5.2 (pressured drop coefficient).



Figure 2. Kernel density [26] of the normalised electrical capacitance (capacitance_norm) and pressure drop coefficient (zeta_norm)

Figure **3** presents the results of the anomaly detection machine learning algorithm applied on the training set. Herein, the crosses denote normal values, whilst the dots denote abnormal ones. There are two areas with abnormal parameter values. The first one is approximately at (1, 5) and the second one at (7, -0.5). It can be learned hereby that the first area corresponds to a dry but polluted filter and the second one belongs to a strongly humidified filter. Whilst the first area can be found for almost any state-of-the-art filter system with a pressure drop and air-flow velocity measurement, the second one is presently not detectable. Here, the proposed electrodes come into action by providing the capacitance signal and allow some conclusion about the humidity of the filter. Henceforth, the risk of micro-organism growth can be assessed.



Figure 3. Diagram of the normal (x) and abnormal (o) data of the normalised training set

This diagram illustrates a problem of all data-driven algorithms: If the data basis has some bias or is somehow incomplete, the algorithm can be unable to work properly. Figure **3** clearly proofs that there are almost no data in the bottom left quadrant. If there were data, it would be considered, although the parameters obtained with the (incomplete) training set fit not perfectly. However, there is no data available since the investigations were carried out in two stages. I.e., experiments had been finished before the data analysis started due to some organisational reasons. Therefore, an online learning algorithm can solve this problem as it adjusts the parameter set from time to time.

Figure 4 provides some idea how this algorithm decides for a complete test set. In addition to the measured data, a synthetic data is generated randomly and discriminated in normal and abnormal parameter combinations.



Figure 4. Diagram of the normal (x) and abnormal (o) data of a normalised synthetic test set

4 Conclusion

Since modern buildings have practically air-tight envelopes, mechanical ventilation becomes more and more important. Filters are required in this context in order to prevent pollutants from outside to enter the system. However, filters are prone to fouling, especially in humid environments, and, therefore, have to be changed on a regular basis.

In order to substitute this rule-of-thumb-like process by a data-driven approach, a test set-up in the HVAC system of the room air-flow laboratory of the institute was designed and an air filter was equipped with electrodes made of stainless steel. The pressure drop over the filter and the electrical capacitance were measured and employed as indicators for the level of pollution and the micro-organism growth risk.

Extensive measurement series were carried out in 2017, providing a lot of data for the machine learning algorithm. A data analysis and machine learning pipeline were implemented in Python 3 language that yields an efficient approach towards the problem.

Finally, the following statements can be made:

- Correlation of humidity and electrical capacitance in the filter
- New quality of micro-organism growth risk assessment by electrical capacitance measurement: the higher the capacity, the higher the risk
- Efficient data analysis and machine learning algorithm

Future work shall address an important point: The results of machine learning algorithms are only as good as the data is. It is planned to develop the machine learning algorithm towards an online learning algorithm in order to carry out the analysis during the operation. Moreover, it is intended to apply this system also to other HVAC components.

This work has been supported by the Federal Ministry for Economic Affairs and Energy of the Federal Republic of Germany under Contracts no. MF 140105 (SD, experimental investigations) and no. MF 170011 (AH, machine learning algorithm).

References

- 1. United Nations Framework Convention on Climate Change (ed.) *Adoption of the Paris Agreement*, number FCCC/CP/2015/L.9/Rev.1
- 2. Millar, D., Tonolo, G., Ziebinska, U. *Energy efficiency indicators HIGHLIGHTS* (International Energy Agency, 2016)
- 3. Lenzen, M. Künstliche Intelligenz, Was sie kann & was uns erwartet (C.H. Beck, 2018)
- 4. Adey, R., Sriram, D. *Applications of artificial intelligence to engineering problems*, report 5113597, (1987)
- Kalogirou, S. A. Renew. Sust. Ener. Rev., 5, 373–401 (2001)
- Magnier, L., Haghighat, F. Build. Env., 45, 739–746 (2010)
- 7. Kusiak, A., Xu, G. Energy, 42, 241–250 (2012)
- 8. Mohanraj, M., Jayaraj, S., Muraleedharan, C. Renew. Sust. Ener. Rev., **16**, 1340–1358 (2012)
- 9. Nassif, N. Build. Sim., 7, 237-245 (2014)
- 10. Ahmad, M. W., Mourshed, M., Yuce, B., Rezgui, Y. Build. Sim., Springer Nature, 9, 359–398 (2016)
- 11. Afram, A., Janabi-Sharifi, F., Fung, A. S., Raahemifar, K. Ener. and Build., **141**, 96–113 (2017)
- 12. Liang, J., Du, R. Int. J. Refrig., 30, 1104–1114 (2007)
- 13. Najafi, M. Fault Detection and Diagnosis in Building HVAC Systems (University of California, Berkely, PhD thesis, 2010)
- West, S. R., Guo, Y., Wang, X. R., Wall, J. *Automated Fault Detection and Diagnosis of HVAC Subsystems Using Statistical Machine Learning*. Proceedings of Building Simulation. 12th Conference of International Building Performance Simulation Association, Sydney, 14-16 November., 2011, 2659-2665
- Narayanaswamy, B., Balaji, B., Gupta, R., Agarwal, Y. Data Driven Investigation of Faults in HVAC Systems with Model, Cluster and Compare (MCC). BuildSys'14, November 5–6, 2014, Memphis, TN, USA, 2014, 1-10
- Du, Z., Fan, B., Jin, X., Chi, J. Build. Environ., 73, 1–11 (2014)
- Araya, D. B., Grolinger, K., ElYamany, H. F., Capretz, M. A., Bitsuamlak, G. Ener. Build., 144, 191–206 (2017)

- Hantsch, A., Mai, R. Einfluss instationärer Raumluftströmung auf die thermische Behaglichkeit im Aufenthaltsbereich. CEGA-Congress für Experten der TGA, 27.–28.11.2018, Baden-Baden, 2018
- 19. Kuchling, H. *Taschenbuch der Physik*, 16th Ed. (Fachbuchverlag Leipzig, 1996)
- 20. Aris, R. Vectors, *Tensors and the Basic Equations of Fluid Mechanics* (Dover Publication Inc., 1990)
- 21. Baehr, H. D. Thermodynamik (Springer-Verlag, 1992)
- 22. Bowles, M. *Machine Learning in Python* (John Wiley & Sons Inc., 2015)
- 23. Müller, A. C., Guido, S. Introduction to Machine Learning with Python: A Guide for Data Scientists (O'Reilly Media Inc., 2016)
- 24. McKinney, W. *Datenanalyse mit Python* (Dpunkt.Verlag GmbH, 2015)
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. J. Mach. Learn. R., **12**, 2825–2830 (2011)
- 26. Géron, A. Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to build intelligent systems (O'Reilly Media Inc., 2017)