

Preliminary assessment of a model to predict mold contamination based on
microbial volatile organic compound profiles

Ryan F. LeBouf, Environmental Science and Engineering, Clarkson University, Potsdam, NY

Stephanie A. Schuckers, Electrical and Computer Engineering, Clarkson University, Potsdam,
NY

Alan Rossner, Environmental Health Sciences Program, Department of Biology, Clarkson
University, Potsdam, NY

Affiliations:

Clarkson University
8 Clarkson Ave., Box 5805
Potsdam, NY 13699

Corresponding Author:

Ryan F. LeBouf
Clarkson University
8 Clarkson Ave., Box 5805
Potsdam, NY 13699
W: 304-285-6287
F: 304-285-6321
rlebouf@cdc.gov

Abstract

Identification of mold growth based on microbial volatile organic compounds (MVOCs) may be a viable alternative to current bioaerosol assessment methodologies. A feed-forward back propagation (FFBP) artificial neural network (ANN) was developed to correlate MVOCs with bioaerosol levels in built environments. A cross-validation MATLAB script was developed to train the ANN and produce model results. Entech Bottle-Vacs were used to collect chemical grab samples at 10 locations in northern NY during 17 sampling periods from July 2006 to August 2007. Bioaerosol samples were collected concurrently with chemical samples. An Anderson N6 impactor was used in conjunction with malt extract agar and dichloran glycerol 18 to collect viable mold samples. Non-viable samples were collected with Air-O-Cell cassettes. Chemical samples and bioaerosol samples were used as model inputs and model targets, respectively. Previous researchers have suggested the use of MVOCs as indicators of mold growth without the use of a pattern recognition program limiting their success. The current proposed strategy implements a pattern recognition program making it instrumental for field applications. This paper demonstrates that FFBP ANN may be used in conjunction with chemical sampling in built environments to predict the presence of mold growth.

Keywords: indoor air, volatile organic compounds, artificial neural network, mold

1. Introduction

Mold growth is ubiquitous in indoor and outdoor environments. Elevated airborne mold concentrations may lead to occupant exposures and increased risk of associated adverse health effects (Burge, 2001). A number of studies have identified increased levels of asthma, allergic rhinitis, and adverse respiratory symptoms associated with elevated levels of mold spore exposure (Bünger et al., 2004; Savilahti et al., 2000; Spengler et al., 2004; Stark et al., 2005). One researcher found an association between moldy odor and allergic symptoms among children (Hägerhed-Engman et al., 2009). The last decade has seen a significant increase in air sampling activity for bioaerosols in indoor environments. This rise in the number of exposure assessments for bioaerosols has also raised interesting questions as to the effectiveness of the current sampling protocols. While the current methods, visual mold inspections and bioaerosol monitoring, may suitably assess the presence of mold growth in some circumstances, the ability to consistently identify mold growth behind walls or barriers may not be effectively

accomplished. Thus, the exploration of new methods proposed herein may allow for the refinement of current sampling protocols to fully characterize mold infestation of indoor environments.

Development of a numerical model that correlates a volatile organic compound (VOC) profile with an airborne fungal concentration of a residence could significantly enhance mold screening assessments. Two important concerns associated with current bioaerosol sampling techniques are the variability in airborne spore concentrations and short sampling times that provide only a snapshot into the bioaerosol levels (Burge, 2002; LeBouf et al., 2007). Viable bioaerosol sampling onto agar nutrient media via an Anderson N6 impactor will only provide the fungal ecology of the mold that grows most readily on the agar used; only 0.001 to 4% of soil microorganisms may grow on organic growth media (Colwell and Grimes, 2000). While viable and non-viable bioaerosol sampling methods have shortcomings, they are currently the preferred method.

Bioaerosol sampling has most recently been dominated by the use of non-viable slit impactors, which has reduced the need for lengthy growth times and given a more representative view of the fungal ecology by spore counts. Inter- and intra-analyst variability, however, can dramatically influence the spore concentrations and spore type distributions (Hung et al., 2005). Variability arises from both different laboratory standard operating procedures and analysts with diverse experience levels (Hung et al., 2005). Wessen et al. (2001) formally recognized the need for development and testing of new analytical techniques to assess microbial impact in buildings (Wessen et al., 2001). Collection of a passive whole-air chemical sample using evacuated containers is a viable, low-cost technique that can be collected in seconds by a building occupant or homeowner and shipped out for analysis (LeBouf et al., 2010). Even though chemical

sampling is also a snapshot in time, the variability associated with chemical source emissions in a building is less influenced by a shorter sampling time than bioaerosol monitoring due to the residence time and emission characteristics of chemicals versus fungal spores in the air (Elke et al., 1999; Górny et al., 2001; Wilkins et al., 2000).

Other researchers have also recognized that VOCs can be used as an indicator of mold growth in the laboratory (Wady et al., 2003), but field settings contribute confounders that have not yet been addressed in the literature. Elke et al. (1999) studied volatile metabolite profiles in damp and moldy dwellings; they discovered elevated levels of a majority of the microbially generated volatile organic compounds (MVOCs) investigated in contaminated dwellings (Elke et al., 1999). Nilsson et al. (1996) recognized that the volatile metabolite pattern will change depending on the growth stage of mold at the time of sampling (Nilsson et al., 1996). Claeson et al. (2002) identified that single specific volatiles can not be used to track mold growth, but volatile patterns may be identified (Claeson et al., 2002). While the individual compounds are not specific enough to fingerprint a mold contaminated room or building, the VOC profile can be used to identify elevated levels of mold. Wilkins et al. (2000) identified volatile patterns from moldy building materials in the laboratory; they found that volatile patterns were dramatically affected by changing substrate (Wilkins et al., 2000). Kuske et al. (2005) in an interesting study with regard to “hidden” mold growth purported that MVOCs can be used to detect unseen mold growth as well as mold growth before the evolution of spores occurs (Kuske et al., 2005). These researchers also state that evaluation of MVOCs in the field may be difficult due to low-MVOC production and multiple volatile organic compound sources that may interfere with the detection of mold generated VOCs (Kuske et al., 2005). The aforementioned researchers, however, did not examine the profile of the MVOC with modeling techniques such as artificial neural

networks that incorporate background correction or eliminate the need for detailed prior knowledge of interfering compounds. ANNs have successfully been used in the recognition of cancer on the basis of urinary nucleosides, in the identification of *Campylobacter* species based on Fourier transform infrared spectroscopy, and in non-linear QSAR modeling (González-Arjona et al., 2002; Mouwen et al., 2006; Seidel et al., 2007)

The purpose of this research was to develop a set of MVOC profiles with corresponding bioaerosol measurements, then design an ANN that can be used to predict the presence of mold contamination in built environments. The method proposed is based on passive whole-air chemical samples to estimate mold growth in indoor environments using an artificial neural network. Through exploitation of the paramount MVOCs in the chemical signature of a residence, the presence of fungal growth may be predicted from the chemical profile using an associated model algorithm. Through pattern recognition, models can be built to guide decision making processes based on training with input-output pairs. The goal of this research was to develop a relatively rapid, inexpensive prediction method using the chemical signature that mold emits coupled with a numerical model to indicate the presence of mold growth in indoor residential environments.

2. Materials and Methods

2.1 SAMPLE COLLECTION

Bioaerosol and chemical sampling data were collected from a series of indoor environments in northern NY for the purpose of training the ANN to classify a residence with potential mold growth based on a chemical signature. The chemical sampling data is used as the

input to compare with the bioaerosol data as the reference. Sample collection, data analysis and model development is subsequently described.

Sampling campaigns began with a visual mold inspection of the dwelling followed by chemical and biological sampling. Sampling sites were originally classified into three discrete categories based on New York City Department of Health guidelines: control (no visible mold growth), low-medium (<10 sq. ft. visible mold growth), and high (>10 sq. ft. visible mold growth) (New York City, 1995). Although human activity can produce considerable chemical interferences from cooking and cleaning activities, occupants were asked to abstain from smoking, cooking and cleaning for 12 hours prior to and during the sampling events. No smoking, cooking or cleaning activities occurred at any of the sampling locations 12 hours prior to and during the sampling events. Qualitative exposure assessments tools (i.e. occupant questionnaires) provided insight into potential classes of compounds in the residence.

Evacuated glass bottles, Entech Bottle-Vacs, were used to collect whole-air grab samples at a total of 10 locations in northern NY during 17 different sampling periods from July 2006 to August 2007 with the majority of samples collected in the summer.

Prior to sampling, a subset of bottles was checked with an analog pressure gauge to ensure proper vacuum was maintained. Samplers were shipped to Entech Instruments, Inc. in Simi Valley, CA for analysis. A chain of custody form accompanied all samplers to ensure proper sample handling. Samples were analyzed on a 7500 Autosampler attached to a 7100A Extended Cold Trap Dehydration Preconcentrator (Entech Instruments Inc., Simi Valley, CA) coupled with an Agilent 6890/5973N GC/MS (Agilent Technologies, Santa Clara, CA). The GC column was a DB-1, 60 meter by 0.32 mm ID with a film thickness of 1 μ m. Calibration standards were both internal and external standards. Internal Standards consisted of

bromochloromethane, 1,4-difluorobenzene, and chlorobenzene-d5. A surrogate, bromofluorobenzene, was also included. Relative response factors were calculated for MVOCs of interest. A majority of samples were analyzed within one month of sampling: only two samples were analyzed 31 and 32 days after sampling. Samples were analyzed in SIM and Scan mode; 22 chemicals were chosen as representative MVOCs from a range of chemicals known to be emitted by mold during metabolic activity. All concentrations were blank corrected.

This chemical MVOC list was chosen due to their prevalence in the literature; chemical profiles were used to create input vectors for the numerical model. Linear regressions were used to reduce the dimensionality of the input vector to 18 influential chemicals, which can reduce the probability of chance correlations (Yi et al., 2007). These chemicals include the following: 2-methylfuran, 2-butanone, 3-methylfuran, 2-methyl-1-propanol, 3-methyl-2-butanol, 2-pentanol, 1,4-dioxane, 3-methyl-1-butanol, 2-methyl-1-butanol, 1-pentanol, 2-hexanone, 2-heptanone, 1-octen-3-ol, 3-octanone, 2-pentylfuran, 3-octanol, 2-ethyl-1-hexanol and 1-octanol. When redundant or non-influential metrics are used in the development of an ANN, computational efficiency is lost and there will be unused or low-weighted connections that are unnecessary to the correlation. The following 4 chemical concentrations were removed from the profile: 1-butanol, 2-methylisoborneol, geosmin, 2-methyl-2-butanol, and 2-isopropyl-3-methoxypyrazine. A total of 59 chemical profiles were produced as input for the artificial neural network.

Bioaerosol samples were collected concurrently with the chemical samples; bioaerosol data were used as the model target values. An Anderson N6 impactor was used in conjunction with malt extract agar (MEA) and dichloran glycerol 18 (DG18) in petri dishes to collect viable mold samples. MEA is a mesophilic agar while DG18 is a xerophilic agar; by using two agar types, a broader range of fungi can be cultured giving a better representation of the fungal

ecology. Six samples of each type of agar were collected with an associated field blank for each sampling event. These were cultured for six days and counted at Clarkson University. Colony counts were blank corrected and a positive hole correction was applied (Andersen, 1958). Non-viable samples were collected with Air-O-Cell cassettes and analyzed at Clarkson University. These biological characterizations were converted into target values between zero and one via a continuous function approximation after rank-ordering the sampling events according to indoor-outdoor ratios; these target values were then used for training the ANN. By developing the target values in this manner, the hit ratio result may be used as a direct comparison method between chemical characterization and relative bioaerosol concentrations in the residence. The output of the model is a subjective value ranging between zero and one indicating the level of airborne mold in a residence. A measure of model performance is the hit ratio, which is the ratio of correctly classified observations to all observations (Walde et al., 2004).

2.2 MODEL DEVELOPMENT

Back propagation algorithm evaluates the derivatives of the error function and is used in adjusting the weights backwards through the network to minimize the error function (Bishop, 1995a). The configuration of the neural network layers, neurons and connections is referred to as a neural network architecture; these are currently developed through a trial-and-error procedure (Rocha et al., 2007). ANN performance is affected by such things as network architecture, initial weight value, learning rate and momentum term (Yang et al., 2002). Feed-forward back propagation (FFBP) ANNs were originally created in a graphic user interface, MATLAB's Neural Network Toolbox (The MathWorks, Natick, Massachusetts), to assess model performance with varying network architectures. The FFBP ANN was created with one hidden layer consisting of a number of log-sigmoid transfer function neurons that was fed into a

single linear output layer. The inputs are the MVOC profiles from chemical sampling. The training references are the target values based on the indoor-outdoor ratios of airborne fungal concentrations. The final chosen network architecture consisted of three layers: input, 40 log-sigmoid neurons, and one linear output neuron.

A MATLAB script was developed to create an iterative cross-validation of the optimum neural network architecture. Using the MATLAB cross-validation script, the influence of adjusting network parameters was determined. A mean square error goal (i.e. training parameter goal) of zero with a maximum of 100 epochs was used for training the network with a Levenberg-Marquardt algorithm. No validation set was used in the creation phase of the network construction. Validation sets were subsequently used for early stopping throughout the iterative training. Unlike previous research by others who use the validation set as a measure of model performance, the proposed modeling techniques used the validation set only for early stopping and the test set for a true test of the model's predictive ability.

To increase the generalizing ability of the network, expanded data sets were created from the original 59 profiles by adding in a random amount of noise to each chemical concentration (Bishop, 1995b; Sietsma and Dow, 1991). Noise was introduced because it is particularly useful for expanding small data sets and increasing the generalizing ability of the network. Six profiles were removed from the data set to be used at the end of cross-validation as a true test of the model's predictive ability. Then, a random matrix was called in the script and multiplied by a standard deviation of 0.05. These discrete values were added to the original data set creating a random perturbation of the value. This process was repeated ten times to produce 530 noisy profiles; the total training and validation profile population consisted of 583 profiles. 58 of these profiles (~10%) were randomly drawn out of the population to be used as the validation set as

was done in a previous study (Walde et al., 2004). The remaining 525 profiles were used to train the neural network. At each of the 100 iterations of the cross-validation program, new training (525 profiles), validation (58 profiles), and test sets (6 profiles) were randomly drawn from the input data and the network weights were reinitialized. Hit ratio is the ratio of correctly classified observations to all observations based on some subjective cutoff criteria such as a 95% match with the target value. Where “95% match” is defined as the target value being within 5% of the FFBP ANN model output. Hit ratios were recorded in a matrix along with linear regression parameters (slope, intercept, Pearson correlation coefficient) of the model output versus target values. Figure 1 displays an example linear regression of output versus target values for a single iteration of the cross-validation model.

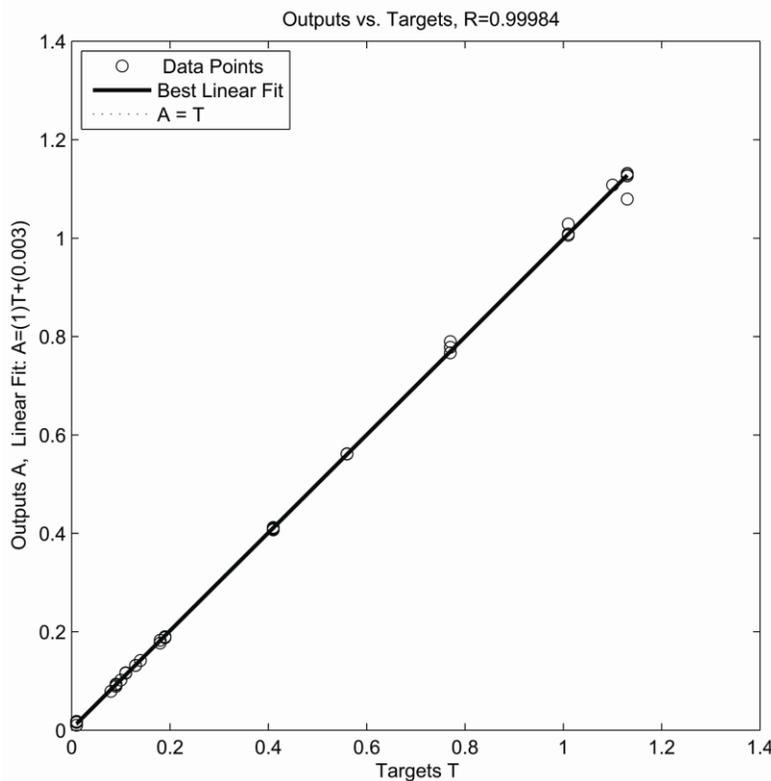


Figure 1: Example linear regression of model outputs versus target values for a single training iteration

Development of target values based on visual mold inspection yielded inadequate correlations with chemical profiles. Model target values were created from the indoor to outdoor ratios of MEA. MEA-based target values were used for model sensitivity analysis. When target values were based on DG18 and AOC, similar model performance was achieved. The visible mold values correspond to control (0.0), low-medium (0.5) and high (1.0).

3. Results

In Table 1, it can be seen that 17 sampling events were used to create a total of 59 chemical profiles and 17 indoor/outdoor ratios from MEA, DG18 and AOC sampling. The number of chemical profiles per location indicates the number of chemical samples taken at the location; each location was a unique sampling event and no location was repeated on alternate days. Seasonal distribution of sampling events consisted of the following: 24% in the autumn months (September to November) and 76% in the summer months (June-August). Discrepancies are prevalent between target values based on bioaerosol characterization and those based on visible mold growth in 13 of 17 sampling events. Visible mold categories did not accurately reflect the observed I/O ratios from the bioaerosol concentrations. The MEA target values presented in Table 1 are an example of the target values used as a reference for the model. Target values based on DG18 and AOC were also produced. DG18 and AOC based target values did not produce an appreciable difference in model performance.

Table 1: Field sample location summary

Date	Location	Sample ID	# chemical profiles	Visible mold	DG18	AOC	MEA	MEA Target Value
7/4/2006	1	Z	2	1.0	1.64	4.82	1.48	0.52
7/5/2006	2	Y	2	1.0	0.62	2.06	0.74	0.26
7/6/2006	3	A	2	1.0	1.04	0.70	0.62	0.22
7/7/2006	4	B	4	1.0	0.83	nd	1.02	0.35
7/11/2006	5	C	4	0.5	0.02	1.41	0.04	0.02
7/12/2006	6	D	2	0.5	0.61	1.88	0.99	0.34
7/13/2006	7	W	2	0.5	0.55	0.96	0.69	0.25
10/20/2006	8	X	3	0.0	5.18	6.10	4.80	0.98
11/9/2006	9	U	3	0.0	1.60	0.80	1.40	0.46
11/10/2006	10	R	2	0.5	3.80	3.50	1.00	0.35
6/7/2007	11	G	6	0.5	4.70	2.20	3.30	0.83
6/11/2007	12	E	6	1.0	6.90	0.90	18.30	1.00
7/2/2007	13	V	2	0.5	6.20	3.60	2.50	0.70
7/3/2007	14	T	9	1.0	21.20	5.30	11.00	1.00
7/10/2007	15	I	2	1.0	13.30	21.10	13.40	0.97
7/11/2007	16	H	4	1.0	6.90	3.90	7.20	1.05
8/14/2007	17	F	4	0.0	0.90	1.10	0.80	0.28

*nd= not determined, outside AOC measurement lost

Note: DG18 = Dichloran Glycerol 18 indoor/outdoor ratios, MEA = Malt Extract Agar indoor/outdoor ratio, AOC = Air-O-Cell indoor/outdoor ratio

Figure 2 displays the change in model response to varying the network architecture. Each data point is an average hit ratio of 100 iterations of the cross-validation script. The number of log-sigmoid neurons in the first layer was adjusted to find the optimal number of neurons. Regardless of the cut off criteria (i.e. 75%, 90%, 95%, or 99% of the target value), 40 log-sigmoid neurons in the 1st layer produced the best hit ratio. As the cut off criteria becomes more stringent, the hit ratio generally decreases as would be expected. A comparison between graph A (>95%) and graph B (>99%) displays a declination in hit ratio with increasing from 95% to 99% cut off criteria.

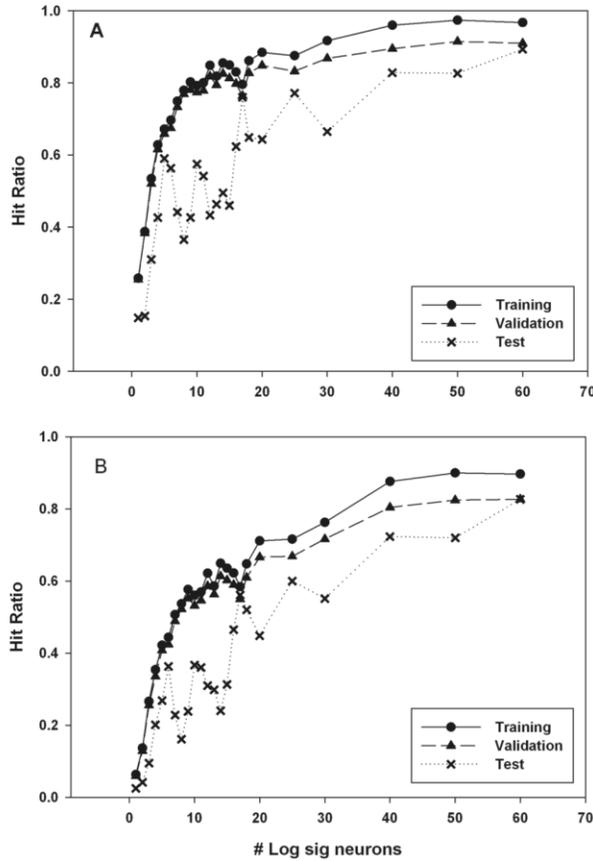


Figure 2: Average model response to varying the number of log-sigmoid neurons in the first layer (A) >0.95 of target value and (B) >0.99 of target value

Model results from the cross-validation script produced hit ratios of 0.960 ± 0.071 (mean \pm sd), 0.895 ± 0.069 , 0.828 ± 0.165 for training, validation and test sets based on a 95% match of the MEA target value. In the case of the test set, 83% of the time the model output was within 95% of the target value: a correct classification is produced when the model output is 95 to 105% of the target value. When DG18 target values were used, model results from the cross-validation script produced hit ratios of 0.923 ± 0.092 (mean \pm sd), 0.885 ± 0.099 , 0.800 ± 0.144 for training, validation and test sets based on a 95% match. When AOC target values were used, model results from the cross-validation script produced hit ratios of 0.982 ± 0.051 (mean \pm sd), 0.930 ± 0.048 , 0.823 ± 0.228 for training, validation and test sets based on a 95% match. Summary

statistics of model output vs. MEA target values linear regressions are displayed in Table 2 for 100 iterations of the cross-validation program.

Table 2: Linear regression summary of training and validation sets (n = 100 iterations)

	Training			Validation		
	m	b	r	m	b	r
Mean	0.999	0.000	1.000	0.999	0.000	0.998
SD	0.003	0.001	0.001	0.008	0.004	0.003
Min	0.976	-0.010	0.988	0.966	-0.022	0.984
Max	1.007	0.007	1.000	1.025	0.010	1.000

Note: m=slope, b=intercept, r=Pearson correlation coefficient

Figure 3 displays the effect of the training parameter goal on the model's hit ratio. Training and validation hit ratios were virtually identical due to the iterative cross-validation procedure in the MATLAB script. A decrease in hit ratio is seen as criteria MSE increases above 1E-5.

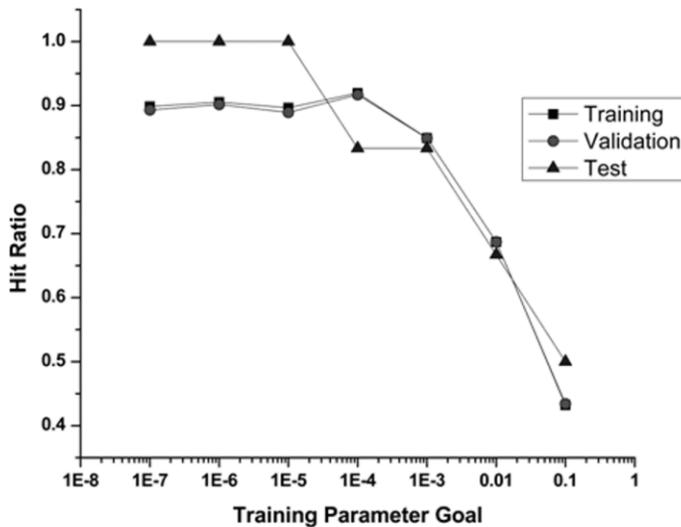


Figure 3: Sensitivity analysis of varying training parameter goal on model's hit ratio (75% of target value)

4. Discussion

In this study, a FFBP ANN was developed that can predict the presence of mold growth based on chemical profiles. While species-specific chemicals and chemical profiles have previously been used in laboratory experiments to identify mold growth, this study used a novel approach by combining a mathematical model with a non-unique chemical signature of indoor air to predict the presence of mold growth in indoor environments. Looking at bioaerosol monitoring in a completely different way provides alternative assessment tools to current practices that can enhance indoor air quality investigations. This model is in the preliminary stage of development, yet the data presented here suggest a promising foundation for an effective bioaerosol assessment.

The initial approach was to develop target values for low, medium and high levels of contamination based on visual mold inspection and compare the model output to these target values. However, this technique did not yield adequate correlations with chemical profiles presumably due to similar chemical signatures in overlapping discrete categories. This suggested that the visible mold growth assessment did not accurately reflect the MVOC evolution or mold sporulation at the time of sampling. When one considers the phase of mold growth at the time of sampling, MVOC patterns and relative individual chemical contributions can change. A visually identified mold-contaminated house, for example, may not have significant MVOCs emitted if there is no actively growing mold. In this situation, the chemical profile would indicate limited contamination while the visual mold inspection would indicate significant contamination. The importance of using bioaerosol levels as the means of target value development is the fact that chemical sampling could be used as a quick precursor to, or replacement for, bioaerosol sampling.

Multiple target values were developed based on varying mold assessment strategies since there is currently no universally accepted mold sampling standards. By using each of these mold assessment strategies, the most useful correlation could be elucidated. The output of the model is a subjective value ranging between zero and one indicating the level of airborne mold in a residence. Model results indicate an excellent correlation was established between MVOC profiles and MEA-based target values as well as DG18- and AOC-based target values. Model results from the cross-validation script produced hit ratios of 0.828 ± 0.165 for new input data based on a 95% match of the MEA target value: 66-99% of the time the model can predict the bioaerosol characterization from the chemical profile with 95% confidence. For this initial evaluation of the model's predictive ability, a 17% misclassification rate was identified. While understanding the nature of the misclassification is important, it was not possible to divide the misclassification rate into false-positive and false-negative rates. False-positive, or Type I error, rates delineate the frequency with which the model would over predict the level of fungal growth in a home based on a chemical signature; this could cause undue alarm to occupants if the presence of mold growth were incorrectly identified in their dwelling. False-negative, or Type II error, rates delineate the frequency with which the model would under predict mold growth. An analysis of false-positive and false-negative rates for the model will be useful for the future to understand the boundaries in which the model can be applied.

The small data set (59 chemical profiles) from 17 sampling periods did not adversely affect model performance given the addition of randomly perturbed chemical signatures to the input data. By incorporating random noise to the training and validation data sets, model performance on new data was vastly improved which has been previously reported by others (Sietsma and Dow, 1991).

Hit ratios for training, validation and test set data were very close to one (i.e. 100% classification) as can be seen in Figure 2. In this fashion, use of an ANN produced excellent results for a semi-continuous approximation of multiple output categories. Sensitivity analysis of changing model architecture and training parameter goals indicates a useful approach to developing network structures and optimizing network parameters in a systematic manner.

The diagnostic specificity and sensitivity of MVOCs as indicator compounds has been brought into question (Schleibinger et al., 2008). Schleibinger et al. (2008) found weak correlations of individual MVOCs with mold infestation in apartment dwellings. The utility of MVOC profiles coupled with ANN as opposed to individual compounds as predictors lies in its ability to withstand concentration changes in individual constituents, as well as allow for compensation for non-MVOC peaks, such as VOCs from cleaning agents. A change in one chemical concentration in the profile, for example, will not adversely affect the chemical profile characterization of mold growth due to the intrinsic nature of the ANN correlation development. In addition to the cleaning agents, changes in one or several chemical concentrations may be due to volatile interferences from building material emissions or occupant activities.

Although volatile interferences in indoor air quality sampling are unavoidable, no explicit knowledge of the background signal should be required when modeling with an ANN. These background signals are implicitly modeled in the chemical profiles such that small changes to the chemical profiles due to interferences from non-microbial sources will not adversely affect the model's predictive ability. To the best of our knowledge, this is the first application of ANN to be used in classifying potential mold growth in built environments based on a chemical signature.

5. Conclusions

Ease of sampling and expediting analysis of chemical samples make MVOC sampling for mold contamination an exciting alternative to current bioaerosol sampling techniques. The ability of VOCs to infiltrate a living space makes them an attractive choice for evaluating hidden mold growth based on chemical signatures. By eliminating the inter- and intra-analyst variability of mold sampling from the propagation of error, chemical sampling for mold coupled with an appropriate ANN may provide useful and expedient information not currently available from conventional IAQ assessments. By developing a FFBP ANN model, the chemical signature that mold emits may be associated with bioaerosol levels at the time of sampling allowing for a more rapid decision making process for mold remediation cases.

Use of this model has only been assessed on data gathered from one geographic location. While the model worked well with the data set collected, as with all models, the predictive limitations must be recognized. As an example, the misclassification rate that was assessed using this model could be broken into false-positive and false-negative rates. Additional testing of the model can be done with new chemical and biological characterizations of dwellings across seasons and geographies. Laboratory-developed chemical profiles of water-damaged building materials could also be generated and could be examined by the model. While the ANN does not recognize environmentally enhanced chemical signatures due to volatile interferences, development of background correction techniques via laboratory generated chemical profiles of potential interferences may prove useful as an alternative means of controlling these influences. As an example, if we generated profiles with fairly large concentrations of common household cleaners that contain interfering compounds such as limonene or pinene, we could further test the predictive capabilities of the model in regards to common interferents. In the future, incorporation of chemical and biological data sets from various geographical regions would

significantly enhance the applicability of this indoor air quality technique for the elucidation of mold presence. Development of this cross-validation MATLAB script may prove helpful to other researchers interested in using ANNs for pattern recognition and classification problems.

Acknowledgements

Entech Instruments, Inc. provided samplers and sample analysis. The authors would like to thank Clarkson University and specifically the Center for Air Resources Engineering and Science as well as the Center for the Environment.

References

- Andersen A. New sampler for the collection, sizing, and enumeration of viable airborne particles. *Journal of Bacteriology* 1958; 76: 471-484.
- Bishop CM. Neural networks for pattern recognition. New York: Oxford University Press, 1995a.
- Bishop CM. Neural networks for pattern recognition. New York: Oxford University Press, 1995b.
- Bünger J, Westphal G, Mönnich A, Hinnendahl B, Hallier E, Müller M. Cytotoxicity of occupationally and environmentally relevant mycotoxins. *Toxicology* 2004; 202: 199-211.
- Burge HA. Fungi: toxic killers or unavoidable nuisances? *Annals of Allergy, Asthma, & Immunology* 2001; 87: 52-56.
- Burge HA. An update on pollen and fungal spore aerobiology. *Current Reviews of Allergy and Clinical Immunology* 2002; 110: 544-552.
- Claeson AS, Levin JO, Blomquist G, Sunesson AL. Volatile metabolites from microorganisms grown on humid building materials and synthetic media. *Journal of Environmental Monitoring* 2002; 4: 667-672.
- Colwell RR, Grimes DJ. Nonculturable microorganisms in the environment. ASM Press, Washington, D.C., 2000.
- Elke K, Begerow J, Oppermann H, Krämer U, Jermann E, Dunemann L. Determination of Selected Microbial Volatile Organic Compounds by Diffusive Sampling and Dual-Column Capillary GC-FID -- a New Feasible Approach for the Detection of an Exposure to Indoor Mould Fungi? *Journal of Environmental Monitoring* 1999; 1: 445-452.
- González-Arjona D, López-Pérez G, González AG. Non-linear QSAR modeling by using multilayer perceptron feedforward neural networks trained by back-propagation. *Talanta* 2002; 56: 79-90.
- Górny RL, Reponen T, Grinshpun SA, Willeke K. Source strength of fungal spore aerosolization from moldy building material. *Atmospheric Environment* 2001; 35: 4853-4862.
- Hägerhed-Engman L, Sigsgaard T, Samuelson I, Sundell J, Janson S, Bornehag CG. Low home ventilation rate in combination with moldy odor from the building structure increase the risk for allergic symptoms in children. *Indoor Air* 2009; 19: 184-192.
- Hung LL, Miller JD, Dillon HK. Field guide for the determination of biological contaminants in environmental samples. AIHA, Fairfax, 2005.
- Kuske M, Romain AC, Nicolas J. Microbial volatile organic compounds as indicators of fungi. Can an electronic nose detect fungi in indoor environments? *Building and Environment* 2005; 40: 824-831.
- LeBouf RF, Casteel C, Rossner A. Evaluation of an air sampling technique for assessing low-level volatile organic compounds in indoor environments. *Journal of Air and Waste Management Association* 2010; 60: in press.
- LeBouf RF, Yesse L, Rossner A. Seasonal and diurnal variability in airborne mold from an indoor residential environment in Northern New York. *Journal of Air and Waste Management Association* 2007; 58: 684-692.

- Mouwens DJM, Capita R, Alonso-Calleja C, Prieto-Gómez J, Prieto M. Artificial neural network based identification of *Campylobacter* species by Fourier transform infrared spectroscopy. *Journal of Microbiological Methods* 2006; 67: 131-140.
- New York City DoH. Guidelines on Assessment and Remediation of Fungi in Indoor Environments. Bureau of Environmental and Occupational Disease Epidemiology, 1995.
- Nilsson T, Larson TO, Montanarella L, Madsen JØ. Application of head-space solid-phase microextraction for the analysis of volatile metabolites emitted by *Penicillium* species. *Journal of Microbiological Methods* 1996; 25: 245-255.
- Rocha M, Cortez P, Neves J. Evolution of Neural Networks for Classification and Regression. *Neurocomputing* 2007; 70: 2809-2816.
- Savilahti R, Uitti J, Laippala P, Husman T, Roto P. Respiratory morbidity among children following renovation of a water-damaged school. *Archives of Environmental Health* 2000; 55: 405-410.
- Schleibinger H, Laussmann D, Bornehag CG, Eis D, Rueden H. Microbial volatile organic compounds in the air of moldy and mold-free indoor environments. *Indoor Air* 2008; 18: 113-24.
- Seidel P, Seidel A, Herbarth O. Multilayer perceptron tumour diagnosis based on chromatography analysis of urinary nucleosides. *Neural Networks* 2007; 20: 646-651.
- Sietsma J, Dow RJF. Creating Artificial Neural Networks that Generalize. *Neural Networks* 1991; 4: 67-79.
- Spengler JD, Jaakkola JJK, Parise H, Katsnelson BA, Privalova LI, Kosheleva AA. Housing characteristics and children's respiratory health in the Russian Federation. *American Journal of Public Health* 2004; 94: 657-662.
- Stark PC, Celedón JC, Chew GL, Ryan LM, Burge HA, Muilenberg ML, et al. Fungal levels in the home and allergic rhinitis by 5 years of age. *Environmental Health Perspectives* 2005; 113: 1405-1409.
- Wady L, Bunte A, Pehrson C, Larsson L. Use of gas chromatography-mass spectrometry/solid phase microextraction for the identification of MVOCs from moldy building materials. *Journal of Microbiological Methods* 2003; 52: 325-332.
- Walde JF, Tappeiner G, Tappeiner U, Tasser E, Holub HW. Statistical aspects of multilayer perceptrons under data limitations. *Computational Statistics and Data Analysis* 2004; 46.
- Wessen B, Ström G, Palmgren U, Schoeps KO, Nilsson M. Chapter 4.3: Analysis of microbial volatile organic compounds. In: Flannigan B, Samson RA, Miller JD, editors. *Microorganisms in home and indoor environments: Diversity, health impacts, investigation and control*. CRC Press LLC, Boca Raton, 2001, pp. 267-274.
- Wilkins K, Larsen K, Simkus M. Volatile metabolites from mold growth on building materials and synthetic media. *Chemosphere* 2000; 41: 437-446.
- Yang J, Xu G, Kong H, Zheng Y, Pang T, Yang Q. Artificial neural network classification based on high-performance liquid chromatography of urinary and serum nucleosides for the clinical diagnosis of cancer. *Journal of Chromatography B* 2002; 780: 27-33.
- Yi QX, Huang JF, Wang FM, Wang XZ, Liu ZY. Monitoring rice nitrogen status using hyperspectral reflectance and artificial neural network. *Environmental Science and Technology* 2007; 41: 6770-6775.