Implications of Error Rates Associated with Numerical Criteria for Airborne Fungal Data

R. Christopher Spicer[1*] and Harry J Gangloff[1]

[1] WCD Consultants, Pennington, New Jersey

*Corresponding email: RCSpicer@wcdconsultants.com

SUMMARY

The criterion used to evaluate airborne fungal data is the underlying foundation for any interpretation of differences in data from a suspect environment and background or reference conditions. Utilizing the bootstrap version of Monte Carlo (BMC) analysis, data presented herein are consistent with previous published studies which demonstrate high error rates with commonly utilized numerically based criteria. Conversely, BMC analysis has demonstrated significant differences in detection frequency relative to the combined median concentration (Δf_d) for each fungal type, to be a valid descriptor. Use of Δf_d as a validated data criterion provides for a quantification of fungal populations in the absence of fixed numerical health based criteria, and its use provides the valid mathematical foundation for assessment of airborne fungal data.

IMPLICATIONS

Verification of Δf_d as a criterion for evaluating environmental fungal data implies the necessity for a fundamental change in the way numerical concentrations of airborne fungi are customarily reported and evaluated, and defines the underlying sampling necessary to demonstrate significant differences in fungal data sets.

KEYWORDS

Bootstrap
Detection frequency
Simulation
Statistical significance

INTRODUCTION

A thorough building inspection and measuring residual moisture in building materials is often sufficient to diagnose many building microbial (fungal) problems. However, "mold testing" of air and/or surfaces is often conducted or requested as an intuitive response, often for potential health concerns to building occupants. With no fixed numerical health based standards, the evaluation of environmental fungal data is done on a relative basis, in which an investigator tests the hypothesis of similarity in fungal populations of an indoor test environment and a suitable reference, such as the general outdoor air at that location. Due to the relative standard model required to evaluate indoor and reference zone fungal populations, and the fact that none of the various sampling and analysis methodologies can detect all airborne fungi present, the mathematical criterion by which to compare the respective data sets is a critical component for a building investigation when air sampling is conducted.

A variety of numerically based criteria (fixed concentrations or various ratios derived from airborne concentrations) for both spore trap and culturable samples have been proposed as

indicators of problematic indoor environments. (Mahooti-Brookes et al, 2004; Baxter et al, 2005) Using bootstrap Monte Carlo techniques (BMC), the frequency by which a given criterion judges data from outdoor air as problematic ("false positive") can be estimated. Similarly, "false negatives" for a given criterion can be (conservatively) estimated, by generating the frequency that data from environments assumed to have degraded air quality (e.g., water damage, visible mold growth) is identified as acceptable (Efron and Tibshirani, 1998). Using this approach, both spore trap (Air o Cell^{TM}) and culturable (Andersen N-6) data from several sites demonstrated highly variable error rates (0.00 – 1.00) for all numerical criteria evaluated (Spicer and Gangloff, 2000, 2008, 2010).

Alternatively, significant differences in frequency of detection (Δf_d) relative to a suitable reference concentration compares each detected fungal type across the test and reference zones, categorizing numerical concentrations by whether they exceed the combined median (reference) concentration. Under a hypothesis test, the test and reference zones are assumed to be the same; axiomatically there cannot be a significant difference in Δf_d if the a priori assumption of similarity holds. (The median concentration of combined indoor and reference zone data in most cases is 0, in which the data is predominantly evaluated relative to detection/non-detection.) A discrete probability value can be directly calculated for all possible detection frequencies in each zone using the binomial random function, while the probability for each possible Δf_d is the product of the probability of the underlying occurrences. Statistical significance is demonstrated when the sum of the probabilities associated with all Δf_d's equal to and greater than what is produced by the data is at the target probability/significance (i.e., 0.95 probability analogous to 0.05 significance) of the investigator (Eudey et al, 1995; Spicer and Gangloff, 2003, 2005). The "error rate" for Δf_d is estimated in an analogous manner as for the numerical criteria by comparing the calculated probability/significance associated with the observed Δf_d to the frequency that Δf_d greater than what is observed in the data is generated from several thousand random BMC resamples.

Spore trap and culturable data from several problematic indoor environments revealed the same probability value approximated by BMC and the directly calculated significance for Δf_d for each of the fungal types detected. The data presented herein is an extension of earlier reporting of error rates for both numerically derived criteria and Δf_d from culturable and spore trap data (Spicer and Gangloff, 2008, 2010).

METHODS

Culturable fungal spore sampling at four sites with visible mold growth and a history of intrusion was conducted utilizing an N-6 impactor with malt extract agar (MEA) as the growth media. Approximately fourteen to eighteen samples from the suspect indoor environment and from the reference (outdoor air or a remote/unaffected area of the respective building) were collected at approximately twenty to thirty minute intervals throughout the same day. Sampling time for all samples was four minutes at 28.3 liters per minute. Laboratory analysis of all samples from any one site was performed by one of three experienced environmental microbiology laboratories.

Raw data (CFU/m^3) from each indoor or reference zone was transformed into frequency of detection greater than the combined median of the reported concentration for each detected fungal species/type. The probability that levels for each fungal species/type were greater indoors than outdoors as reflected by differences in Δf_d was then calculated as previously described. Estimation of the probability/significance of Δf_d was then determined via BMC

for each fungal type, tabulating the frequency that Δf_d greater than actually observed from the field data appeared from 10000 random simulations of sample sets of N-1 for each zone. The theoretical "error rate" was then evaluated as the difference between the probabilities (significance) derived by the two methods.

Two numerically based criteria were similarly evaluated using BMC from the same data sets. Criteria tested were indoor/outdoor total (culturable) spore levels (IA/OA>1.0), and "dominance" of indoor non phylloplane (NP) fungi to phylloplane (P) fungi (total of *Aspergillus sp.* and *Penicillium sp.*/total of *Cladosporium sp*., *Alternaria sp*., and *Epicoccum sp*.), expressed as NP/P>1.0. For each site, the frequency that the respective criterion accurately identified the building as problematic was estimated via 2000 BMC simulations for sample sets of N = 2, 5, 10, and 15 in each zone. .

RESULTS

The data collected from all sites reflects the highly variable and sparse nature of airborne fungal spore data. The most frequently reported data point (approximately 50 – 70%) is "none detect, " while airborne concentrations of detected fungal types may vary by orders of magnitude within minutes at the same location. Table 1 indicates the performance of $\Delta f_{d,}$ displaying direct calculation and BMC estimates of significance/probability for representative fungal types detected. P values are interpreted as the probability that indoor levels of the respective fungal type are greater than the reference zone level. As can be seen, significant differences between indoor and reference zone fungal populations are consistently identified by Δf_d using both a direct calculation of probability or an estimate via the BMC random simulation.. The majority of differences between the exact calculation and the BMC estimates are due to normal small variation inherent in all Monte Carlo based techniques, and the nature of fungal spore data. The relatively large difference in probability value for *S. chartarum* (~0.13) identified at HAM occurs as a result of the extreme sparse nature of the data, while the two values do not conflict with regard to identifying statistical significance. However, statistical significance for such fungal types such as *S. chartarum, Chaetomium,* and others require special consideration in their interpretation (further discussed herein).

The difference between the two probability values for Δf_d is analogous to the implied false negative rates for the two criteria displayed in Tables 2 and 3 (1 – displayed probability value). The performance of indoor/outdoor ratio of total spores (Table 2) and indoor phylloplane/nonphylloplane ratio (Table 3) can be seen to be erratic, being "accurate" in some instances, and "missing" a (likely) problematic environment in others. Poor performance is not necessarily improved by increasing sample number.

DISCUSSION

Random simulation techniques such as BMC are at the core of modern probability theory, and have consistently demonstrated poor performance (high false positive and negative error rates) of numerically based criteria for environmental fungal data The significant issue revealed through BMC is the influence that the extreme and sparse distribution of fungal data plays upon airborne fungal data analysis, and the potential for misevaluation inherent in numerically based criteria. Despite their conformance with standard industrial hygiene and environmental/public health models for contaminant "levels," the distribution of the data does

Table 1. Δf_d Criterion Performance

Site - BET: Northeast metropolitan apartment building with visible mold and history of water intrusion						
Sampling: February 2004				Total Types Detected: 30		
Fungal Type	Med	Detected >/= Med		Δf_d	$P_{IA>OA\ CALC}$	$P_{BMC\ (10000)\ OA=14\ IA=14}$
		OA (15)	IA (15)			
A fumigatus	0	1	3	0.133	0.80	0.786
A.versicolor	0	0	6	0.400	0.99	0.999
P. chrysogenum	0	1	7	0.400	0.99	0.993
Site - ANNEX: Northeast metropolitan government building under construction with periodic water infiltration and visible mold growth						
Sampling: October 2003				Total Types Detected: 32		
Fungal Type	Med	Detected >/= Med		Δf_d	$P_{IA>OA\ CALC}$	$P_{BMC\ (10000)\ OA=17\ IA=14}$
		OA (18)	IA (15)			
Acremonium sp.	0	1	7	0.411	1.00	0.999
A. flavus	0	0	2	0.133	0.90	0.855
A. fumigatus	0	4	5	0.111	0.75	0.748
A. sydowii	0	2	4	0.156	0.86	0.864
A. versicolor	0	2	7	0.356	0.99	0.989
Site - RAR: 4 floor zone in northeast commercial high rise building subjected to partial flooding and inadequate drying						
Sampling: March 2002				Total Types Detected: 48		
Fungal Type	Med	Detected >/= Med		Δf_d	$P_{IA>OA\ CALC}$	$P_{BMC\ (10000)\ OA=14\ IA=17}$
		OA (15)	IA (18)			
P. chrysogenum	0	0	3	0.167	0.94	0.956.
P. citrinum	0	0	2	0.111	0.87	0.866
P. decumbens	0	0	5	0.278	0.99	0.996
Site – HAM: 3 story modern northeast hotel subject to flooding and inadequate drying						
Sampling: January 2003				Total Types Detected: 51		
Fungal Type	Med	Detected >/= Med		Δf_d	$P_{IA>OA\ CALC}$	$P_{BMC\ (10000)\ OA=16\ IA=13}$
		OA (17)	IA (14)			
A. versicolor	0	5	6	0.135	0.77	0.767
P. variable	9	6	11	0.433	0.99	0.993
R. glutinis	0	4	7	0.267	0.93	0.930
S. chatarum	0	0	1	0.071	0.76	0.632

Table 2 Total Spore Criterion Performance

Simulated Sample Size	Frequency Total IA/OA > 1.0 Site			
	BET	ANNEX	RAR	HAM
2	0.134	0.829	0.960	0.114
5	0.222	0.944	0.99	0.031
10	0.004	0.988	1.00	0.004
15	0.000	0.997	1.00	0.001

Table 3 Non Phylloplane/Phlloplane Ratio Criterion Performance

Simulated Sample Size	Frequency NP/P > 1.0 Site			
	BET	ANNEX	RAR	HAM
2	0.535	0.999	0.394	0.999
5	0.574	1.00	0.448	1.00
10	0.570	1.00	0.469	1.00
15	0.598	1.00	0.457	1.00

not favor fixed airborne levels, ratios, rank order analysis, or similar parametrically based descriptors. Current licensing of mold investigators permit, and/or require, environmental mold data to be incorporated into building investigations or final clearance (post remediation) acceptance. However, numerically based criteria cannot consistently identify differences in indoor and outdoor data sets, and therefore are not adequately protective of public health, despite their intuitive appeal. Similarly, potentially misleading indicators of a problematic indoor environment through high false positive rates for numerical criteria potentially drive additional sampling and/or remediation that otherwise would not be indicated.

Within the context of the short term sampling times (e.g. minutes) permitted by current air sampling devices, the mathematical validation of Δf_d as a criterion carries important implications for sample size, as well as bridging the oft blurred distinction between data analysis and data interpretation. Many fungal species associated with problematic indoor environments (i.e., *Aspergillus sp*. and *Penicillium sp*.) occur in the outdoor air at a 0.05 or greater detection frequency. This requires approximately fifteen samples in the test zone and reference zones being necessary to generate statistical significance (Spicer and Gangloff, 2003, 2005). For most of the potentially problematic or "indicator" fungal species this equates to at least three detections in the test zone with no detection in the reference, or a Δf_d of 0.2 (or greater). Whether statistical significance demonstrated in the data is of practical importance (i.e., building diagnosis or possibly contributing to occupant complaints) in a given circumstance requires interpretation. However, the determination of statistical significance exhibited in data sets, which is an inherent component of hypothesis testing, is a mathematical calculation, not "interpretation." There are circumstances in which statistical significance can reasonably be assumed, even though the calculations may not reach a target P value. A good example can be seen in Table 1 (Site: HAM) for *S. chartarum*. This organism and several others are recognized to be rarely detected in the outdoor air (i.e. perhaps 0.01 or less). A single detection of *S. chartarum* in fourteen indoor air samples against no detections in seventeen reference (outdoor) samples does not represent a calculated significant difference (Table I). However, based upon the recognized outdoor air distribution, the detection of *S. chartarum* indoors could reasonably be compared to one hundred samples in the outdoor air. In this case, the probability that indoor levels of *S. chartarum* are greater than outdoors is ~0.94, and the fact that actual statistical significance was not demonstrated is more an artifact of the (insufficient) number of samples that can reasonably be collected. (This could be argued as an "interpretation;" however, there still remains an underlying mathematical/statistical rationale.) This implies that fungal data should be evaluated in two tiers. Statistical significance using Δf_d as a criterion is relevant for many species, while *any* detection of *S. chartarum* (and others with similar distribution in outdoor air) regardless of sample size is an indication of a problematic environment. The fact that problematic buildings characteristically exhibit elevations in several fungal species still defines Δf_d as the most discriminatory and accurate criterion for evaluation of environmental fungal data.

CONCLUSION

Differentiating indoor and reference zone fungal populations with a mathematically valid basis ultimately depends upon the criterion utilized. Numerically based criteria are widely applied, based upon their "intuitive" appeal and historical use in industrial hygiene and environmental/public health. However, their usefulness has been repeatedly demonstrated via BMC to be very limited due to the mathematical nature (distribution) of fungal data. (A relevant historical note is that Karl Pearson, who developed the chi square goodness of fit

test in the early 20th century, admonished the scientific community for conducting analysis without an understanding of data distributions and the underlying limitations.) On the other hand, determination of significant differences via Δf_d is an axiomatic mathematical calculation. With some basic (and conservative) assumptions, Δf_d is descriptive of differences in indoor and reference zone data sets as verified via BMC analysis. Implementing Δf_d as a criterion prescribes the number of samples in each zone necessary to demonstrate significance, and implies that for data analytical purposes, fungal species can be evaluated on a two tiered basis depending upon anticipated frequency of detection in the outdoor air.

In the absence of health based standards for environmental fungi, a valid mathematical criterion to differentiate data from two comparative zones must be the foundation for professional opinions regarding the performance of a suspect indoor environment. Otherwise, evaluation of fungal proliferation in buildings will continue to be characterized as "unscientific," providing an avenue for subjective data "interpretation" by those with inadequate training and experience, or those with inherent biases (Johnson et al, 2008). Bradley Efron, credited with popularizing BMC simulation techniques that have been incorporated into data analysis in a wide variety of scientific fields has most applicably stated:

> "Left to our own devices, we are not very good at picking out patterns from a sea of noisy data . . . we are all too good at picking out non-existent patterns that happen to suit our purposes" (Efron and Tibshirani, 1998).

REFERENCES

Baxter D. M., Perkins J. L., McGhee C.R., and Selzer J.M. 2005. A regional comparison of mold spore concentrations outdoors and inside "clean" and "mold contaminated" Southern California buildings. J. Occup. Environ. Hyg., Vol 2, pp. 8-18.

Eudey L., Su H. J., and Burge H.A. 1995. Biostatistics and bioaerosols. In: *Bioaerosols* Burge, HA editor. Bioaerosols. CRC Press; 1995, pp. 290 – 307.

Efron B and Tibshirani R. 1998. Hypothesis testing and the bootstrap. In: *Introduction to the bootstrap*. New York. pp. 202 – 219.

Johnson D, Thompson D, Clinkenbeard R, Redus J. 2008 Professional judgment and interpretation of viable mold air sampling data. J. Occ. and Env. Hygiene. Vol. 5 pp. 656-663.

Mahooti-Brooks N, Storey E, Yang C, Simcox N. J, Turner W., and Hodgson M. 2004. Characterization of mold and moisture indicators in the home. J. Occup. Environ. Hyg.; Vol 1, pp. 826-839.

Spicer R.C. and Gangloff, H.J. 2000. Limitations in application of Spearman's Rank Correlation to bioaerosol sampling data. Am. Ind. Hyg. Assoc. J. Vol. 61 pp. :361-366.

Spicer R C and Gangloff H.J. 2003. Bioaerosol data distribution: probability and implications for sampling in evaluating problematic buildings. Appl Occup Environ. Hyg. Vol. 18 pp.584-590.

Spicer R.C. and Gangloff H.J. 2008. Verifying interpretive criteria for bioaerosol data using (bootstrap) Monte Carlo techniques. J. Occup. Environ. Hyg. Vol. 5 pp. 85-92.

Spicer R.C. and Gangloff H.J. 2010. Differences in detection frequency as a bioaerosol data criterion for evaluating suspect fungal contamination. Building and Environment. Vol. 45 pp.1304-1311.