PREDICTING RESPIRATORY HEALTH SYMPTOM OCCURRENCE IN OFFICE BUILDING ENVIRONMENTS

A Dissertation in
Architectural Engineering
by
Vladimir Vuković

© 2009 Vladimir Vuković

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2009
The dissertation of Vladimir Vuković was reviewed and approved* by the following:

Jelena Srebrić  
Associate Professor of Architectural Engineering  
Adjunct Professor of Mechanical and Nuclear Engineering  
Dissertation Advisor  
Chair of Committee

Stanley A. Mumma  
Professor Emeritus of Architectural Engineering

Bohumil Kasal  
Hankin Chair of Residential Building Construction  
Professor of Civil and Environmental Engineering  
Professor of Architectural Engineering

Leonard J. Peltier  
Adjunct Professor of Mechanical and Nuclear Engineering

Zhengmin Qian  
Adjunct Professor of Public Health Sciences

Chinemelu J. Anumba  
Professor of Architectural Engineering  
Head of the Department of Architectural Engineering

*Signatures are on file in the Graduate School.
Abstract

Providing healthy indoor environments for building occupants is important to ensure people’s wellbeing, satisfaction and productivity. Furthermore, healthy indoor environments can reduce health care costs and positively affect the economy. Currently used techniques to establish relations between indoor environmental parameters and occurrence of health related symptoms among occupants cannot deal with real time concurrent changes of multiple simultaneously monitored parameters, nor provide predictions of real building impacts on occupants’ health. The dissertation is advocating usage of innovative data analyses tools for interpretation of building sensor data and occupants’ perceptions of indoor environments to predict respiratory health symptom occurrence in buildings. Such data analyses establish relations between indoor air quality perceptions and health.

The research goal was development of an artificial neural network (NN) based computational methodology for fast prediction of respiratory health related symptoms among office building occupants. For the purpose of NN training, Environmental Protection Agency’s (EPA) Building Assessment Survey and Evaluation (BASE) study provided measurements of indoor building parameters and occupant survey data within 100 office buildings in the U.S.A. Trained networks had an output indicating occupants’ health symptoms based on measurements and occupants’ perceptions of indoor environments. The method was tested and experimentally validated using on-site measurements and occupants’ survey for a LEED certified “green” building environment. Additional multivariate statistical regression of the BASE data was used for the purpose of comparing results to presently available data analyses tools.

The results showed NN methodology can be applied to predict a number of respiratory health symptoms among office building occupants. High significance of occupants’ perceptions of indoor environments was confirmed by multivariate statistical analyses. Experimental study in a green building revealed better indoor
environmental quality, healthier indoor conditions and higher occupants’ satisfaction compared to an average BASE office building.

The developed methodology could be incorporated in the future design procedures to specify optimal combination of indoor environmental parameters and prevent possible adverse impacts on occupants’ respiratory health. Last, but not the least, the results encourage building researchers and scientific community to initiate applications of NNs, as innovative and powerful data interpretation tools for indoor environments.
# Table of Contents

List of Figures viii

List of Tables xx

List of Abbreviations xxii

Acknowledgments xxv

Chapter 1
Introduction 1

Chapter 2
Literature Review 3
  2.1 BASE Study Data Analyses ...................... 4
  2.2 Analyses Associated with Other Experimental Studies ........ 7
  2.3 Summary of Current Analyses ..................... 9
  2.4 Alternative Data Analyses Methodology ............... 11

Chapter 3
Methods for Determining Health Impacts of Building Indoor Environments 13
  3.1 Multivariate Logistic Regression .................... 13
  3.2 Artificial Neural Networks ....................... 16
    3.2.1 How to Construct a Neural Network? ............. 19

Chapter 4
Building Assessment Survey and Evaluation Benchmark Dataset 24
  4.1 Measurements .................................... 25
4.2 Occupant Questionnaire ........................................... 28
4.3 Building Information ............................................ 31

Chapter 5
Design, Testing and Comparison of Predictive Simulation Models 33
5.1 Multivariate Logistic Regression Design .......................... 33
5.2 Neural Network Design ........................................... 36
   5.2.1 Input Data Set ............................................. 36
   5.2.2 Output Data Set ........................................... 38
   5.2.3 Neural Network Algorithm ................................. 40
   5.2.4 Training, Testing and Validation Scenarios ............... 41
      5.2.4.1 Neural Network Training and Testing ............... 41
      5.2.4.2 Neural Network Validation ......................... 43

Chapter 6
Results and Discussion on BASE Data Analyses 45
6.1 Multivariate Logistic Regression Analysis ....................... 45
   6.1.1 Multivariate Logistic Regression Results ................ 46
   6.1.2 Discussion on Multivariate Logistic Regression Analyses 46
6.2 Neural Network Data Analysis .................................. 50
   6.2.1 Benchmark Predictions .................................. 55
   6.2.2 Discussion on Neural Network Data Analyses ............ 66

Chapter 7
Experimental Setup and Validation 69
7.1 Experimental Setup Location ................................... 69
7.2 Field Planning and Schedule .................................. 72
   7.2.1 Field Operations Schedule ............................... 73
7.3 Instrumentation .................................................. 78
7.4 Practical Recommendations for Similar Studies ................ 83

Chapter 8
Results and Discussion on Experimental Validation 86
8.1 Experimental Results and Comparison to Average Benchmark Building 86
8.2 Validation Predictions .......................................... 95
8.3 Comments on Experimental Measurement Results ............. 96
Chapter 9
Conclusions and Future Work

9.1 Conclusions from Multivariate Regression Analyses ......................................... 99
9.2 Conclusions on Methodology and Usage of NNs .................................................... 100
9.3 Conclusions on Experimental Analyses of a Green Building .................................. 102
9.4 Future Work ........................................................................................................ 103

Appendix A
Programs Used in NN Data Analyses and Detailed Benchmark Prediction Results 105
A.1 Preprocessing ................................................................. 105
A.2 Core Software ............................................................. 106
A.3 Detailed Neural Network Benchmark Results ...................................................... 106

Appendix B
Detailed Comparison of Benchmark and Validation Buildings 147

Appendix C
Detailed Neural Network Validation Results 193

Appendix D
Building Occupants Questionnaire 229

Bibliography 240
List of Figures

3.1 An example of a two-layer neural network .......................... 17

6.1 Neural network input scatter plots categorized with respect to probability for occurrence of cough ................................. 56
6.2 Neural network training error reduction in cough predictions .... 58
6.3 Neural network training performance for cough predictions ...... 58
6.4 Neural network testing performance for cough predictions ........ 59
6.5 Neural network training and testing absolute errors in cough predictions .......................................................... 59
6.6 Neural network actual vs. predicted training probabilities for occurrence of cough .................................................. 61
6.7 Neural network actual vs. predicted testing probabilities for occurrence of cough .................................................. 62
6.8 Neural network actual vs. predicted training and testing probabilities for occurrence of cough .............................................. 63
6.9 Confidence ratios for neural network predictions of cough ......... 64

7.1 PA DEP building floor plan with marked indoor monitoring locations (left) and magnified North-East monitoring location (right) (modified from (Deru et al., 2005)) ................................. 71
7.2 Installed indoor (left, middle) and outdoor (right) measurement equipment .......................................................... 71

8.1 Thermal comfort conditions in PA DEP Cambria “green” building (marked X) compared to average BASE building (marked +) and typical winter (1.0 clo) and summer (0.5 clo) comfort ranges corresponding to 80% occupant acceptability (allowing 10% dissatisfaction for whole body thermal discomfort based on predicted mean vote (PMV Limits) between ±0.5, plus an additional 10% dissatisfaction for partial body thermal discomfort) ................................................. 88
8.2 Occupants’ perceptions of indoor environmental conditions in PA DEP Cambria “green” building compared to average BASE building 89
8.3 Occupant self-reported occurrence of sick building syndrome symptoms in PA DEP Cambria “green” building compared to average BASE building .................................................. 90
8.4 Categories of building occupant symptoms and their prevalence percentage in BASE (gray range covers 50% of the buildings closest to the database mean value, i.e. interquartile range), two NIOSH buildings (solid and dashed lines) and PA DEP Cambria building (dotted line) (modified from (Brightman, 2005)) .................. 92
8.5 Occupant self-reported impact of sick building syndrome symptoms on productivity in PA DEP Cambria “green” building compared to average BASE building (N/A - not applicable category - corresponds to reports of no negative impacts on productivity) ........ 92
8.6 Occupant self-reported diagnosed medical conditions in PA DEP Cambria “green” building compared to average BASE building . . 93
8.7 Measured allergen, bacterial, fungal and particle concentrations in PA DEP Cambria “green” building compared to average BASE building records ......................................................... 94
A.1 Neural network input scatter plots categorized with respect to probability for occurrence of asthma ....................... 107
A.2 Neural network training performance for asthma predictions .... 107
A.3 Neural network testing performance for asthma predictions ..... 108
A.4 Neural network training and testing absolute errors in asthma predictions .................................................. 108
A.5 Neural network actual vs. predicted training probabilities for occurrence of asthma ................................................. 109
A.6 Neural network actual vs. predicted testing probabilities for occurrence of asthma ................................................. 109
A.7 Neural network actual vs. predicted training and testing probabilities for occurrence of asthma ............................. 110
A.8 Confidence ratios for neural network predictions of asthma .... 110
A.9 Neural network input scatter plots categorized with respect to BSI values .............................................................. 111
A.10 Neural network training performance for BSI predictions ....... 111
A.11 Neural network testing performance for BSI predictions ........ 112
A.12 Neural network training and testing absolute errors in BSI predictions 112
A.13 Neural network actual vs. predicted training BSI values ....... 113
A.14 Neural network actual vs. predicted testing BSI values ....... 113
A.15 Neural network actual vs. predicted training and testing BSI values 114
A.16 Confidence ratios for neural network predictions of BSI . . . . . . . 114
A.17 Neural network input scatter plots categorized with respect to prob-

ability for occurrence of chest tightness . . . . . . . . . . . . . . . . 115
A.18 Neural network training performance for chest tightness predictions 115
A.19 Neural network testing performance for chest tightness predictions . 116
A.20 Neural network training and testing absolute errors in chest tight-

ness predictions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 116
A.21 Neural network actual vs. predicted training probabilities for oc-
currence of chest tightness . . . . . . . . . . . . . . . . . . . . . . 117
A.22 Neural network actual vs. predicted testing probabilities for occur-

currence of chest tightness . . . . . . . . . . . . . . . . . . . . . . . 117
A.23 Neural network actual vs. predicted training and testing probabil-

ities for occurrence of chest tightness . . . . . . . . . . . . . . . . . 118
A.24 Confidence ratios for neural network predictions of chest tightness . 118
A.25 Neural network input scatter plots categorized with respect to prob-

ability for occurrence of dust allergy . . . . . . . . . . . . . . . . . 119
A.26 Neural network training performance for dust allergy predictions . . 119
A.27 Neural network testing performance for dust allergy predictions . . 120
A.28 Neural network training and testing absolute errors in dust allergy pr-
edictions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 120
A.29 Neural network actual vs. predicted training probabilities for oc-
currence of dust allergy . . . . . . . . . . . . . . . . . . . . . . . 121
A.30 Neural network actual vs. predicted testing probabilities for occur-

currence of dust allergy . . . . . . . . . . . . . . . . . . . . . . . . 121
A.31 Neural network actual vs. predicted training and testing probabil-

ities for occurrence of dust allergy . . . . . . . . . . . . . . . . . . 122
A.32 Confidence ratios for neural network predictions of dust allergy . . 122
A.33 Neural network input scatter plots categorized with respect to prob-

ability for occurrence of mold allergy . . . . . . . . . . . . . . . . . 123
A.34 Neural network training performance for mold allergy predictions . . 123
A.35 Neural network testing performance for mold allergy predictions . . 124
A.36 Neural network training and testing absolute errors in mold allergy pr-
edictions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 124
A.37 Neural network actual vs. predicted training probabilities for oc-
currence of mold allergy . . . . . . . . . . . . . . . . . . . . . . . 125
A.38 Neural network actual vs. predicted testing probabilities for occur-

currence of mold allergy . . . . . . . . . . . . . . . . . . . . . . . . 125
A.39 Neural network actual vs. predicted training and testing probabil-

ities for occurrence of mold allergy . . . . . . . . . . . . . . . . . . 126
B.6 Validation and benchmark building comparison: desk and chair comfort ...................................................... 151
B.7 Validation and benchmark building comparison: satisfaction with job 151
B.8 Validation and benchmark building comparison: glare frequency at the workstation ..................................... 152
B.9 Validation and benchmark building comparison: workstation location 152
B.10 Validation and benchmark building comparison: workspace cleanliness ...................................................... 153
B.11 Validation and benchmark building comparison: occupants’ education level ................................................... 153
B.12 Validation and benchmark building comparison: room occupancy .............................................................. 154
B.13 Validation and benchmark building comparison: windows in work area ........................................................ 154
B.14 Validation and benchmark building comparison: visibility of windows ......................................................... 155
B.15 Validation and benchmark building comparison: chemical sensitivity ......................................................... 155
B.16 Validation and benchmark building comparison: tobacco smoking status ........................................................ 156
B.17 Validation and benchmark building comparison: type of worn corrective lenses ............................................. 156
B.18 Validation and benchmark building comparison: gender .............................................................................. 157
B.19 Validation and benchmark building comparison: occupants’ age ................................................................ 157
B.20 Validation and benchmark building comparison: workspace change ............................................................ 158
B.21 Validation and benchmark building comparison: computer usage ................................................................. 158
B.22 Validation and benchmark building comparison: presence of carpets ............................................................ 159
B.23 Validation and benchmark building comparison: major responsibilities outside of job .................................... 159
B.24 Validation and benchmark building comparison: photocopier usage ............................................................. 160
B.25 Validation and benchmark building comparison: laser printer usage .............................................................. 160
B.26 Validation and benchmark building comparison: fax usage ....................................................................... 161
B.27 Validation and benchmark building comparison: usage of copy paper .......................................................... 161
B.28 Validation and benchmark building comparison: usage of odorous chemicals ................................................ 162
B.29 Validation and benchmark building comparison: eye irritation frequency ...................................................... 162
B.30 Validation and benchmark building comparison: wheezing frequency .......................................................... 163
B.31 Validation and benchmark building comparison: headache frequency ............................................................ 163
B.32 Validation and benchmark building comparison: sore throat frequency ......................................................... 164
B.33 Validation and benchmark building comparison: unusual tiredness frequency ............................................... 164
B.34 Validation and benchmark building comparison: chest tightness frequency .................................................. 165
B.35 Validation and benchmark building comparison: sinus congestion frequency ................................................. 165
B.36 Validation and benchmark building comparison: cough frequency .............................................................. 166
B.37 Validation and benchmark building comparison: eye tiredness frequency ..................................................... 166
B.38 Validation and benchmark building comparison: tension frequency ............................................................ 167
B.39 Validation and benchmark building comparison: back or shoulder pain frequency ....................................... 167
B.40 Validation and benchmark building comparison: sneezing frequency .......................................................... 168
B.41 Validation and benchmark building comparison: frequency of concentration difficulties ............................... 168
B.42 Validation and benchmark building comparison: dizziness frequency .......................................................... 169
B.43 Validation and benchmark building comparison: depression frequency ......................................................... 169
B.44 Validation and benchmark building comparison: shortness of breath frequency ........................................ 170
B.45 Validation and benchmark building comparison: nausea frequency ............................................................ 170
B.46 Validation and benchmark building comparison: dry skin frequency .......................................................... 171
B.47 Validation and benchmark building comparison: wrist or hand pain frequency ............................................ 171
B.48 Validation and benchmark building comparison: dry eye persistence while not in the building ......................... 172
B.49 Validation and benchmark building comparison: wheezing persistence while not in the building .................... 172
B.50 Validation and benchmark building comparison: headache persistence while not in the building .................... 173
B.51 Validation and benchmark building comparison: sore throat persistence while not in the building ................. 173
B.52 Validation and benchmark building comparison: unusual tiredness persistence while not in the building ......... 174
B.53 Validation and benchmark building comparison: chest tightness persistence while not in the building .......... 174
B.54 Validation and benchmark building comparison: sinus congestion persistence while not in the building ......... 175
B.55 Validation and benchmark building comparison: cough persistence while not in the building ....................... 175
B.56 Validation and benchmark building comparison: eye tiredness persistence while not in the building .............. 176
B.57 Validation and benchmark building comparison: tension persistence while not in the building ......................................................... 176
B.58 Validation and benchmark building comparison: back or shoulder pain persistence while not in the building ................................. 177
B.59 Validation and benchmark building comparison: sneezing persistence while not in the building ...................................................... 177
B.60 Validation and benchmark building comparison: persistence in concentration difficulties while not in the building ............................... 178
B.61 Validation and benchmark building comparison: dizziness persistence while not in the building ....................................................... 178
B.62 Validation and benchmark building comparison: persistence of depression while not in the building .................................................. 179
B.63 Validation and benchmark building comparison: shortness of breath persistence while not in the building ........................................... 179
B.64 Validation and benchmark building comparison: nausea persistence while not in the building ......................................................... 180
B.65 Validation and benchmark building comparison: dry skin persistence while not in the building ......................................................... 180
B.66 Validation and benchmark building comparison: wrist or hand pain persistence while not in the building ........................................... 181
B.67 Validation and benchmark building comparison: frequency of perceived too much air movement ..................................................... 181
B.68 Validation and benchmark building comparison: frequency of perceived too little air movement ....................................................... 182
B.69 Validation and benchmark building comparison: frequency of perceived temperature too high ......................................................... 182
B.70 Validation and benchmark building comparison: frequency of perceived temperature too low ......................................................... 183
B.71 Validation and benchmark building comparison: frequency of perceived air too humid ............................................................... 183
B.72 Validation and benchmark building comparison: frequency of perceived air too dry ................................................................. 184
B.73 Validation and benchmark building comparison: frequency of perceived tobacco smoke odors ....................................................... 184
B.74 Validation and benchmark building comparison: frequency of perceived chemical odors ........................................................... 185
B.75 Validation and benchmark building comparison: frequency of perceived other unpleasant odors ...................................................... 185
B.76 Validation and benchmark building comparison: frequency being asked to do thing that conflict by a person equal in rank .................. 186
B.77 Validation and benchmark building comparison: frequency being
given to do things that conflict to one another by supervisors . . . . . 186
B.78 Validation and benchmark building comparison: frequency being
given to do things that conflict to other work by superiors . . . . . 187
B.79 Validation and benchmark building comparison: frequency job re-
quires very fast work . . . . . . . . . . . . . . . . . . . . . . 187
B.80 Validation and benchmark building comparison: frequency job re-
quires very hard work . . . . . . . . . . . . . . . . . . . . . . 188
B.81 Validation and benchmark building comparison: frequency a little
time left to do things . . . . . . . . . . . . . . . . . . . . . . 188
B.82 Validation and benchmark building comparison: a great deal to be
done frequency . . . . . . . . . . . . . . . . . . . . . . . . . . . 189
B.83 Validation and benchmark building comparison: clear about what
others expect frequency . . . . . . . . . . . . . . . . . . . . . . 189
B.84 Validation and benchmark building comparison: clear on job re-
 sponsibilities frequency . . . . . . . . . . . . . . . . . . . . . 190
B.85 Validation and benchmark building comparison: able to predict
what others will expect frequency . . . . . . . . . . . . . . . . . . 190
B.86 Validation and benchmark building comparison: frequency of work
objectives being well defined . . . . . . . . . . . . . . . . . . . . 191
B.87 Validation and benchmark building comparison: working years in
the building . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 191
B.88 Validation and benchmark building comparison: weekly working
hours in the building . . . . . . . . . . . . . . . . . . . . . . . 192
B.89 Validation and benchmark building comparison: daily computer
usage time . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 192

C.1 Neural network training performance for asthma predictions . . . 194
C.2 Neural network testing performance for asthma predictions . . . . 194
C.3 Neural network training and testing absolute errors in asthma pre-
dictions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 195
C.4 Neural network actual vs. predicted training probabilities for oc-
currence of asthma . . . . . . . . . . . . . . . . . . . . . . . . 195
C.5 Neural network actual vs. predicted testing probabilities for occur-
rence of asthma . . . . . . . . . . . . . . . . . . . . . . . . . . 196
C.6 Neural network actual vs. predicted training and testing probabil-
ities for occurrence of asthma . . . . . . . . . . . . . . . . . . . 196
C.7 Confidence ratios for neural network predictions of asthma . . . . 197
C.8 Neural network training performance for BSI predictions . . . . . 197
C.9 Neural network testing performance for BSI predictions . . . . . 198
C.10 Neural network training and testing absolute errors in BSI predictions 198
C.11 Neural network actual vs. predicted training BSI values . . . . . . . 199
C.12 Neural network actual vs. predicted testing BSI values . . . . . . . 199
C.13 Neural network actual vs. predicted training and testing BSI values 200
C.14 Confidence ratios for neural network predictions of BSI . . . . . . . 200
C.15 Neural network training performance for chest tightness predictions 201
C.16 Neural network testing performance for chest tightness predictions . 201
C.17 Neural network training and testing absolute errors in chest tight-
ness predictions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 202
C.18 Neural network actual vs. predicted training probabilities for oc-
currence of chest tightness . . . . . . . . . . . . . . . . . . . . . . 202
C.19 Neural network actual vs. predicted testing probabilities for occur-
currence of chest tightness . . . . . . . . . . . . . . . . . . . . . . 203
C.20 Neural network actual vs. predicted training and testing probabil-
ities for occurrence of chest tightness . . . . . . . . . . . . . . . . . . 203
C.21 Confidence ratios for neural network predictions of chest tightness . 204
C.22 Neural network training performance for dust allergy predictions . 204
C.23 Neural network testing performance for dust allergy predictions . 205
C.24 Neural network training and testing absolute errors in dust allergy
predictions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 205
C.25 Neural network actual vs. predicted training probabilities for oc-
currence of dust allergy . . . . . . . . . . . . . . . . . . . . . . . . 206
C.26 Neural network actual vs. predicted testing probabilities for occur-
currence of dust allergy . . . . . . . . . . . . . . . . . . . . . . . . 206
C.27 Neural network actual vs. predicted training and testing probabil-
ities for occurrence of dust allergy . . . . . . . . . . . . . . . . . . . 207
C.28 Confidence ratios for neural network predictions of dust allergy . . 207
C.29 Neural network training performance for mold allergy predictions . 208
C.30 Neural network testing performance for mold allergy predictions . 208
C.31 Neural network training and testing absolute errors in mold allergy
predictions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 209
C.32 Neural network actual vs. predicted training probabilities for oc-
currence of mold allergy . . . . . . . . . . . . . . . . . . . . . . . . 209
C.33 Neural network actual vs. predicted testing probabilities for occur-
currence of mold allergy . . . . . . . . . . . . . . . . . . . . . . . . 210
C.34 Neural network actual vs. predicted training and testing probabil-
ities for occurrence of mold allergy . . . . . . . . . . . . . . . . . . . 210
C.35 Confidence ratios for neural network predictions of mold allergy . . 210
C.36 Neural network training performance for shortness of breath pre-
dictions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 211
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.37</td>
<td>Neural network testing performance for shortness of breath predictions</td>
<td>212</td>
</tr>
<tr>
<td>C.38</td>
<td>Neural network training and testing absolute errors in shortness of breath predictions</td>
<td>212</td>
</tr>
<tr>
<td>C.39</td>
<td>Neural network actual vs. predicted training probabilities for occurrence of shortness of breath</td>
<td>213</td>
</tr>
<tr>
<td>C.40</td>
<td>Neural network actual vs. predicted testing probabilities for occurrence of shortness of breath</td>
<td>213</td>
</tr>
<tr>
<td>C.41</td>
<td>Neural network actual vs. predicted training and testing probabilities for occurrence of shortness of breath</td>
<td>214</td>
</tr>
<tr>
<td>C.42</td>
<td>Confidence ratios for neural network predictions of shortness of breath</td>
<td>214</td>
</tr>
<tr>
<td>C.43</td>
<td>Neural network training performance for sinus congestion predictions</td>
<td>215</td>
</tr>
<tr>
<td>C.44</td>
<td>Neural network testing performance for sinus congestion predictions</td>
<td>215</td>
</tr>
<tr>
<td>C.45</td>
<td>Neural network training and testing absolute errors in sinus congestion predictions</td>
<td>216</td>
</tr>
<tr>
<td>C.46</td>
<td>Neural network actual vs. predicted training probabilities for occurrence of sinus congestion</td>
<td>216</td>
</tr>
<tr>
<td>C.47</td>
<td>Neural network actual vs. predicted testing probabilities for occurrence of sinus congestion</td>
<td>217</td>
</tr>
<tr>
<td>C.48</td>
<td>Neural network actual vs. predicted training and testing probabilities for occurrence of sinus congestion</td>
<td>217</td>
</tr>
<tr>
<td>C.49</td>
<td>Confidence ratios for neural network predictions of sinus congestion</td>
<td>218</td>
</tr>
<tr>
<td>C.50</td>
<td>Neural network training performance for sneezing predictions</td>
<td>218</td>
</tr>
<tr>
<td>C.51</td>
<td>Neural network testing performance for sneezing predictions</td>
<td>219</td>
</tr>
<tr>
<td>C.52</td>
<td>Neural network training and testing absolute errors in sneezing predictions</td>
<td>219</td>
</tr>
<tr>
<td>C.53</td>
<td>Neural network actual vs. predicted training probabilities for occurrence of sneezing</td>
<td>220</td>
</tr>
<tr>
<td>C.54</td>
<td>Neural network actual vs. predicted testing probabilities for occurrence of sneezing</td>
<td>220</td>
</tr>
<tr>
<td>C.55</td>
<td>Neural network actual vs. predicted training and testing probabilities for occurrence of sneezing</td>
<td>221</td>
</tr>
<tr>
<td>C.56</td>
<td>Confidence ratios for neural network predictions of sneezing</td>
<td>221</td>
</tr>
<tr>
<td>C.57</td>
<td>Neural network training performance for sore throat predictions</td>
<td>222</td>
</tr>
<tr>
<td>C.58</td>
<td>Neural network testing performance for sore throat predictions</td>
<td>222</td>
</tr>
<tr>
<td>C.59</td>
<td>Neural network training and testing absolute errors in sore throat predictions</td>
<td>223</td>
</tr>
<tr>
<td>C.60</td>
<td>Neural network actual vs. predicted training probabilities for occurrence of sore throat</td>
<td>223</td>
</tr>
</tbody>
</table>
C.61 Neural network actual vs. predicted testing probabilities for occurrence of sore throat ........................................ 224
C.62 Neural network actual vs. predicted training and testing probabilities for occurrence of sore throat ........................................ 224
C.63 Confidence ratios for neural network predictions of sore throat ...................................................... 225
C.64 Neural network training performance for wheezing predictions .............................................................. 225
C.65 Neural network testing performance for wheezing predictions .............................................................. 226
C.66 Neural network training and testing absolute errors in wheezing predictions ...................................... 226
C.67 Neural network actual vs. predicted training probabilities for occurrence of wheezing ........................................ 227
C.68 Neural network actual vs. predicted testing probabilities for occurrence of wheezing ........................................ 227
C.69 Neural network actual vs. predicted training and testing probabilities for occurrence of wheezing ........................................ 228
C.70 Confidence ratios for neural network predictions of wheezing .............................................................. 228
List of Tables

2.1 Summary of reported health impacts of various factors in the existing studies ........................................... 10
3.1 Typically used neural network transfer functions .............................. 22
3.2 Typically used neural network backpropagation training functions .... 23
4.1 Essential measurement variables extracted after data cleaning ....... 28
4.2 Essential measurement variables extracted after data cleaning ....... 30
5.1 Prevalence of the averaged building indoor temperature, relative humidity and CO₂ concentration in BASE buildings .................. 35
5.2 Frequency (percent) of various occupants’ perceptions of indoor environments (CHEMIOD - chemical odors; SMOKEOD - tobacco smoke; OTHEROD - other odors; HUMID - perception of high humidity) ...................................................... 35
5.3 List of input and output parameters used in NN design ................. 37
5.4 Building symptom index (BSI) symptoms used in literature compared to the present study ........................................ 39
6.1 Odds ratios (95% confidence intervals) for statistically significant influences on respiratory health responses (NS - not statistically significant; CHEMIOD - chemical odors; SMOKEOD - tobacco smoke; OTHEROD - other odors; HUMID - perception of high humidity; RH - relative humidity) .................................................. 47
6.2 Average NN target probabilities and absolute prediction errors for each considered output parameter throughout development phases (Pr - probabilities for occurrence of actual symptoms throughout NN development phases; TesEr - average absolute testing error; TrnEr - average absolute training error) ........................................ 65
8.1 Overview of NN validation performance in predicting probabilities for occurrence of building related respiratory symptoms among occupants in PA DEP Cambria “green” building . . . . . . . . . . . . 97
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHU</td>
<td>air handling unit</td>
</tr>
<tr>
<td>ANN</td>
<td>artificial neural network</td>
</tr>
<tr>
<td>ANSI</td>
<td>American National Standards Institute</td>
</tr>
<tr>
<td>ASHRAE</td>
<td>American Society of Heating, Refrigeration and Air-conditioning Engineers</td>
</tr>
<tr>
<td>A&amp;WMA</td>
<td>Air and Waste Management Association</td>
</tr>
<tr>
<td>BASE</td>
<td>Building Assessment Survey and Evaluation</td>
</tr>
<tr>
<td>BMS</td>
<td>building monitoring system</td>
</tr>
<tr>
<td>BRI</td>
<td>building related illness</td>
</tr>
<tr>
<td>BSI</td>
<td>building symptom index</td>
</tr>
<tr>
<td>CHEMIOD</td>
<td>chemical odors</td>
</tr>
<tr>
<td>CI</td>
<td>confidence interval</td>
</tr>
<tr>
<td>CO</td>
<td>carbon monoxide</td>
</tr>
<tr>
<td>CO₂</td>
<td>carbon dioxide</td>
</tr>
<tr>
<td>CoC</td>
<td>chain of custody</td>
</tr>
<tr>
<td>DEP</td>
<td>Department of Environmental Protection</td>
</tr>
<tr>
<td>DNPH</td>
<td>2,4-dinitrophenylhydrazine</td>
</tr>
<tr>
<td>EPA</td>
<td>Environmental Protection Agency</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>HUMID</td>
<td>perception of high humidity</td>
</tr>
<tr>
<td>HVAC</td>
<td>heating, ventilation and air-conditioning</td>
</tr>
<tr>
<td>IAPI</td>
<td>indoor air pollution index</td>
</tr>
<tr>
<td>IESNA</td>
<td>Illuminating Engineering Society of North America</td>
</tr>
<tr>
<td>IEQ</td>
<td>indoor environmental quality</td>
</tr>
<tr>
<td>IRB</td>
<td>Institutional Review Board, The Pennsylvania State University</td>
</tr>
<tr>
<td>ISIAQ</td>
<td>International Society of Indoor Air Quality and Climate</td>
</tr>
<tr>
<td>LBNL</td>
<td>Lawrence Berkeley National Laboratory</td>
</tr>
<tr>
<td>LEED</td>
<td>leadership in energy and environmental design</td>
</tr>
<tr>
<td>MEA</td>
<td>malt extract agar</td>
</tr>
<tr>
<td>MLP</td>
<td>multi-layer perceptron</td>
</tr>
<tr>
<td>MSE</td>
<td>mean squared error</td>
</tr>
<tr>
<td>MTR</td>
<td>environmental measurements part of the BASE dataset</td>
</tr>
<tr>
<td>N/A</td>
<td>not applicable</td>
</tr>
<tr>
<td>NIOSH</td>
<td>National Institute for Occupational Safety and Health</td>
</tr>
<tr>
<td>NN</td>
<td>neural network</td>
</tr>
<tr>
<td>NO\textsubscript{x}</td>
<td>nitrogen oxides</td>
</tr>
<tr>
<td>NS</td>
<td>not statistically significant</td>
</tr>
<tr>
<td>O\textsubscript{3}</td>
<td>ozone</td>
</tr>
<tr>
<td>OR</td>
<td>odds ratio</td>
</tr>
<tr>
<td>ORP</td>
<td>Office of Research Protection, The Pennsylvania State University</td>
</tr>
<tr>
<td>OTHEROD</td>
<td>other odors</td>
</tr>
<tr>
<td>PCA</td>
<td>principle component analysis</td>
</tr>
<tr>
<td>PM\textsubscript{10}</td>
<td>particles with aerodynamic diameters less than 10 µm</td>
</tr>
</tbody>
</table>
PM$_{2.5}$ particles with aerodynamic diameters less than 2.5 $\mu$m

PMV predicted mean vote

ppm parts per million

QSN occupant questionnaire part of the BASE dataset

PEL permissible exposure limit

RH relative humidity

SBS sick building syndrome

SE standard error

SMOKEOD tobacco smoke

SO$_2$ sulfur dioxide

SSE sum of squared errors

SVY building survey part of the BASE dataset

TL team leader

TSA tryptic(ase) soy agar

TVOC total volatile organic compounds

UFAD underfloor air distribution

USGBC United States Green Building Council

VOC volatile organic compound
I would like to express profound gratitude to my academic advisor, Dr. Jelena Srebric, for her support and guidance which went above and beyond regular office hours. In the busy world, she always found time to review my work and provide valuable suggestions that kept me going. Also, I would like to thank all dissertation committee members for their comments and suggestions, and particularly: Dr. Stanley A. Mumma for his emphases on engineering practice and industry experiences, Dr. Bohumil Kasal, for his broad understanding and insight, Dr. Leonard J. Peltier, for his inspiring ideas and perspectives, Dr. Zhengmin Qian, for his willingness to contribute to a multidisciplinary research and the patience he had with engineers. I am also grateful to the members of Dr. Srebric’s research group, Brendon Burley, Paulo Tabares, and Brian Ault for their help with the experimental equipment setup, and Daranee Jareemit for her help with data organization. I would like to thank my wife, Marija, for being beside me, giving me the ultimate moral support and strength in my endeavors. At last but not the least, I would like to thank my friends and family in Serbia, for encouraging my efforts, offering tremendous support and unconditional love.
“Some men see things as they are and ask why. 
   Others dream things that never were and ask why not.”

George Bernard Shaw
Chapter 1

Introduction

People on average spend 90% of their time within buildings, while buildings in the U.S. account for 36% of the total primary energy and two thirds of the total electricity consumption, use 30% of raw materials and 12% of potable water, producing 30% of green house gas emissions and total waste (PA DEP, 2003). This trend is enhanced by modern lifestyle and changing weather conditions. Consequently, it is increasingly important to develop innovative building technologies to significantly decrease the adverse impacts buildings may have on natural environment, while providing healthy indoor conditions for building occupants. Even more as the rising health care costs have negative effects on the currently declining economy. Especially critical are office buildings with limited amounts of fresh air and many building materials emitting pollutants that may affect human health.

The adverse impact of building indoor environments on occupants can be classified as either a Building Related Illness (BRI) or a Sick Building Syndrome (SBS). BRI is associated with occurrence of specific illnesses among occupants, such as diagnosed asthma, influenza or other viral infections, cancer, Legionnaires’ disease. On the other hand, SBS includes non-specific health symptoms, such as: headache; eye, nose, or throat irritation; dry cough; dry or itchy skin; dizziness and nausea; difficulty concentrating; fatigue; and sensitivity to odors. As some of these illnesses and symptoms can be a consequence of exterior influences on building occupants other than the building indoor environments, commonly used criteria to associate them with specific buildings is decreased occurrence of such symptoms when the occupants leave indoor environments. Exposure to adverse in-
door environments is estimated to cause hundreds of thousands respiratory health problems and thousands of cancer deaths in the U.S. alone, every year (U.S. EPA, 2001). Additionally, the "sick" buildings are estimated to cause annual loss in productivity of 15 billion U.S. dollars, and affect up to 30% of all new or remodeled buildings (U.S. EPA, 2007).

Therefore, large research efforts are focused on studying impacts of indoor environments on occupants, while the National Occupational Research Agenda identified indoor nonindustrial work environments as one of the priority research areas (Rosenstock et al., 1998). The most common current methods for investigating health impacts of building indoor environments include multivariate statistical regression analyses that can identify association between multiple parameters. These methods have limiting factors, such as difficulty in dealing with a very large number of concurrently changing parameters and inability to make predictions in real building environments. On the other hand, artificial neural networks (NNs) emerged as a tool for rapid forecasting of indoor and outdoor environmental parameters, such as contaminant concentrations. The intention of the present doctoral dissertation research is to extend applications of NNs to prediction of indoor environmental health impacts on building occupants. The dissertation hypothesis is that there exists a nonlinear relation between building sensor data and occupants’ perceptions of indoor environments on one side and occurrence of respiratory health symptoms on the other side, which can be predicted using non-traditional statistical techniques. Considering the large number of occupants experiencing building related respiratory health problems each year and inadequate ability to predict such influences, the dissertation focuses on predicting respiratory health symptom occurrence in office building environments.
Chapter 2

Literature Review

Various experimental studies were focused on examining multiple buildings and identifying the most influential parameters impacting occupants’ health and performance, primarily in office environments. In Europe, independent experimental studies were performed in England, Denmark, Sweden and the Netherlands between 1987 and 1992, each examining between 28 and 61 office buildings (Wallace et al., 1993). An additional large scale study investigated 56 office buildings in 9 European countries between 1992 and 1994 (Bluyssen et al., 1996). This study was focused on building energy consumption, but also examined indoor air quality.

In the United States, several large scale experimental studies surveyed building occupants and measured indoor environmental conditions trying to establish associations between health symptoms and workplace characteristics. The first such study was administered by the Environmental Protection Agency (EPA) among its employees in three office buildings in the Washington, DC area in 1989 (U.S. EPA 1989; 1990a; 1990b). In 1990, experimental study in 12 public office buildings in California was designed to establish associations between ventilation system types, occupants’ work performance and health related symptoms (Daisey et al., 1990; Fisk et al., 1993).

Using the past experiences, EPA adjusted and expanded the scope of surveyed buildings in a follow-up study, which included 100 offices across United States. The study called Building Assessment Survey and Evaluation (BASE), aimed to provide a comprehensive database about office buildings in the United States and it was the largest such study to date (U.S. EPA, 2003).
2.1 BASE Study Data Analyses

Past research studies related to the BASE study can be classified into 4 groups:

1. studies that describe methodology for collecting BASE data and contents of the dataset;

2. studies that provide statistics of the BASE parameters, such as average, maximum, minimum, median values or percentage quotients, without analyzing impacts on human occupants;

3. comparisons between BASE and other experimental building studies; and

4. BASE data analyses to establish indoor environmental impacts on human occupants.

The most significant for the current dissertation research is the last group of BASE analyses dealing with different parameters impacting building occupants. For example, a study used multivariate logistic regression to link carbon dioxide (CO$_2$) concentration levels to SBS (Erdmann et al., 2002). Using CO$_2$ concentrations as a surrogate indicator of ventilation rates and concentrations of other contaminants, this study found statistically significant dose-response relations for occurrence of sore throat, irritated nose, mucous membrane symptoms, tight chest and wheezing. The exhibited mucous membrane symptoms were subject of an additional analysis including concentrations of volatile organic compounds (VOC) (Apte and Erdmann, 2002). This study applied principle component analysis (PCA) to identify which of the 73 measured VOCs were associated with specific SBS symptoms and observed direct association between these VOCs and mucous membrane/lower respiratory system irritation. The study reported reduction of SBS symptoms up to 80% with increasing ventilation rates, even in buildings complying with the existing ASHRAE ventilation standards. Confirming the association of VOC concentrations to building related symptoms among occupants, an additional study found statistically significant positive associations between outdoor ozone concentrations and indoor concentrations of aldehydes, which were further related to occurrence of upper respiratory and neurological symptoms, headache and dry eyes (Apte et al., 2007). Using multivariate data analyses, this study
suggested a possibility of indoor chemical reactions producing VOCs, caused by the infiltrating outdoor ozone.

Researchers have also shown interest in combining influences from several indoor environmental parameters measured in the BASE study in order to develop indicators of indoor air pollution. Thus, the indoor air pollution index (IAPI) can be used to rank buildings based on their comparison with an average BASE building (Sofuoglu and Moschandreas, 2003). Also, using the available limits for exposure to various contaminants, and applying multivariate regression analyses, IAPI was linearly related with occurrence of three occupant symptom indexes: (1) percentage of occupants with persistent symptoms, (2) maximum number of symptoms experienced by occupants, and (3) average number of occupant symptoms, also known as the building symptom index (BSI). While predicting BSI was the least accurate, IAPI had the best correlation with the percentage of occupants experiencing persistent symptoms, having the coefficient of determination, $R^2 = 0.67$.

Apart from indoor air measurements, other BASE building data were also analyzed. Thus, a study examined heating, ventilation and air-conditioning (HVAC) system characteristics as risk factors for seven occupant symptoms, using multivariate logistic regression (Mendell et al., 2006a). The study found strong association between the height of an HVAC systems’ outdoor air intake and most of the examined symptoms. The investigated symptoms were twice as likely to occur if outdoor air intake was located at the height less than 60m above ground, than due to any of the other reported risk factors. Associations of the occupant symptoms with the other risk factors are given as follows: (1) lower respiratory symptoms - no other associations, (2) upper respiratory symptoms - associated with poor condition of a humidification system or duct liner, and irregular HVAC system inspections, (3) cough - associated with irregular HVAC system inspections, (4) irritated/itching eyes - associated with irregular HVAC system inspections, non-operable windows, and a lack of an outdoor air intake fan, (5) fatigue/difficulty concentrating - associated with poor condition of a humidification system, (6) headache - associated with presence of local cooling coils, and good condition of a humidification system, (7) irritated/itching skin - associated with non-operable windows. Although some of the reported associations might be unexpected, the main relation linking the
outdoor air intake height and occurrence of four of the aforementioned symptoms is confirmed in an additional study (Mendell et al., 2006b). However, contrary to the previous findings, this study reported no associations with the presence of a local cooling system. Besides analyzing design, operation and maintenance of HVAC systems as risk factors, this study performed analyses of the two additional influences: ventilation rates and presence of moisture in HVAC systems. Similar to other reviewed studies, the ventilation rate increase was associated with a decrease in investigated SBS symptoms, while infrequent cleaning of HVAC cooling coils and drain pans were the most influential moisture related factors. Interestingly, the study found no statistically significant associations of occupant symptoms with either condition of the filtration systems, or air handlers. Performing these analyses using multivariate regression methods, the study reported difficulties in discovering underlying relations between different indoor parameters and occupant symptoms due to concurrent changes in values of too many influential parameters.

Contrary to the previous report, another study established relations not only between SBS and the filtration systems, but also investigated influences of different filter materials (Buchanan and Apte, 2006). Multivariate logistic regression analysis from this study indicated 15%-54% reduction in SBS by removing or replacing filters containing synthetic materials or polyester.

In addition to investigations of BASE office workers’ health related symptoms, a recent doctoral dissertation also examined building environmental impacts on productivity (Brightman, 2005). Stating the ineffectiveness of classifying buildings into healthy or unhealthy, due to the continuum of exhibited SBS indicators, the dissertation used work-related occupant symptoms to find four dominant SBS factors. These factors were responsible for 40% of total occupant symptom variance and included (1) tiredness, (2) mucosal irritation, (3) neuropsychological, and (4) lower respiratory factors. At the same time, productivity losses were associated with occupant density above 5 occupants per 1000 ft² (93 m²), presence of operable windows, water or fire damage, recent renovations and temperature outside of 72-73°F (22-23°C) range. Self reported productivity was also affected by building characteristics (Rohr and Brightman, 2003). Reductions in working ability were associated with a decrease in size of office space and building West Coast location.
Additionally, the decrease in size of office space, together with the smaller number of windows per occupant, was related with lost workdays. These factors were the most influential on productivity out of the total of 15 investigated parameters including: occupant density, building age, HVAC inspection frequency and filter replacement, building location and damage, and type of indoor environments with respect to smoking and natural ventilation. The study estimated lost productivity due to all these impacts at over $200,000 U.S. dollars per building per year. Similar to previous BASE related analyses, these studies also used multivariate statistic regression.

2.2 Analyses Associated with Other Experimental Studies

With a large number of experimental studies collecting huge amounts of data about existing building environments in various regions of the world, reviewing the findings from every study around the globe is a tedious task which might not lead to comprehensible information. Therefore, the present review focuses on providing condensed information and guidelines about impacts of indoor environments to occupants’ health in office buildings based on existing review studies. Thus, a panel of European scientists reviewed 105 research publications searching for findings about ventilation impacts on comfort, health and productivity of office workers (Wargocki et al., 2002). The panel agreed upon strong influence of ventilation on comfort and health of the occupants and indicated possible influence on productivity. The authors reported an increased risk for occurrence of SBS with ventilation rates lower than 25 L/s per person, as well as reduced risk from dust mites with ventilation rates higher than 0.5 air changes per hour. Additionally, a higher SBS risk was reported for air-conditioned buildings compared to naturally ventilated buildings. However, the panel did not review health impacts of factors other than the ventilation.

The association of SBS with HVAC systems was confirmed in an additional review of 12 studies covering 467 buildings in Europe and the U.S. (Seppänen and Fisk, 2002). The review revealed 30 to 200% increase in prevalence of at
least one SBS symptom in mechanically ventilated and air-conditioned buildings compared to the naturally ventilated ones. As possible reasons, the review stated poor maintenance, deficiencies in design and construction or inefficient operation of HVAC systems. In addition to ventilation, another review of 21 research studies associated CO\textsubscript{2} concentrations with occupants’ health responses in commercial and institutional buildings (Seppänen et al., 1999). The review reported that half of the examined studies found reductions in SBS with decrease in CO\textsubscript{2} concentrations below 800 ppm.

Additional association of SBS with psycho-sociological factors was reported in a number of studies in a more recent literature review on indoor air quality risk factors (Klotz and Lahm, 2006). Some of the reviewed studies indicated that factors such as workload, office communication, social status, fear of losing a job and private problems may cause similar symptoms in office workers to those typically associated with SBS.

Benefits of increased ventilation rates were confirmed by a study reviewing a variety of reports on different types of buildings, from military barracks and jails, through offices, schools and hospitals to an Antarctic station (Fisk, 2000). The study also pointed that only a small number of findings reported influence of mold on occurrence of respiratory infections. However, all these findings indicated positive association between mold and symptoms of a respiratory illness. Furthermore, strong associations were reported between mold, dust and cat mites, fungi and tobacco smoke and occurrences of asthma and allergies. Finally, this study estimated that increased U.S. workers’ productivity and annual health care savings from improved indoor air quality would range between 20 and 50 billion U.S. dollars. Indoor environmental influences on asthma were a specific focus of an additional literature review which found substantial evidence only for association of asthma with dust mite allergens (Richardson et al., 2005).

Additional indoor air quality influences on occupants’ health and performance that were reported in more than one research publication included: VOC impacts on SBS; moisture impacts on SBS, BRI, allergies and asthma; dust impact on SBS; lighting impact on productivity; thermal conditions impact on SBS and productivity (Kumar and Fisk, 2002). A summary of a variety of possible impacts on SBS was a topic of a study referencing 78 publications covering factors related
to increased prevalence of SBS symptoms (Burge, 2004). The factors were classified into three groups: (1) personal: female gender and lower position in building hierarchy; (2) individual: paper and office dust, cigarette smoke and usage of computers; and (3) building: temperature higher than 23°C, fresh air flow rates lower than 10 L/s per person, poor individual control of temperature and lighting, mechanical ventilation system, poor maintenance, cleaning and water damage. As an indicator of building “sickness”, the study advocates usage of a building symptom index, defined as an average number of SBS symptoms experienced by building occupants.

### 2.3 Summary of Current Analyses

In majority of current analyses, multivariate statistical regression appears as a preferred tool for determining associations between health symptoms and building indoor environmental parameters. A study lists a comprehensive overview of research activities supported by various governmental agencies and specifically mentions the multi-parametric statistical regression as a tool of choice for analyses of multiple simultaneously acting indoor environmental parameters (Fisk et al., 2002). However, no evidence was found of any attempts to establish health predictions for building occupants based on such analyses. The current studies used experimental building data and occupant surveys to examine influences of ventilation, moisture, temperature, lighting, CO₂ and VOC concentrations, dust and animal allergens, mold, fungi, tobacco smoke and HVAC system characteristics, on SBS and BRI, including asthma and allergies. Table 2.1 gives an overview of such studies analyzing BASE data or presenting comprehensive literature reviews. The current BASE data analyses reported no associations of indoor environmental parameters with asthma or allergies, although such relations were established in other studies. Similarly, no BASE studies considered health impacts of allergens, while influence of moisture was investigated only on SBS. Furthermore, the investigated moisture impacts in BASE analyses included only occurrence of condensation in HVAC systems, rather than impacts of relative humidity levels on occupants’ symptoms. Thus, further work is necessary to investigate moisture and allergens impacts on BASE buildings.
Table 2.1. Summary of reported health impacts of various factors in the existing studies

<table>
<thead>
<tr>
<th>Factor</th>
<th>SBS</th>
<th>Asthma</th>
<th>Allergies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CO₂</strong></td>
<td>Erdmann et al. 2002; Apte and Erdmann 2002; Seppänen et al. 1999</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>VOCs</strong></td>
<td>Apte and Erdmann 2002; Apte et al. 2007; Kumar and Fisk 2002</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HVAC system</strong></td>
<td>Mendell et al. 2006a; Mendell et al. 2006b; Buchanan and Apte 2006; Wargocki et al. 2002; Seppänen and Fisk 2002; Burge 2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Allergens</strong></td>
<td>Wargocki et al. 2002; Kumar and Fisk 2002; Klotz and Lahm 2006</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Moisture</strong></td>
<td>Mendell et al. 2006b; Kumar and Fisk 2002; Klotz and Lahm 2006</td>
<td>Kumar and Fisk 2002; Klotz and Lahm 2006</td>
<td>Kumar and Fisk 2002;</td>
</tr>
</tbody>
</table>

plain text - BASE study data analyses; bold - review of other studies

Apart from health, the reported associations of indoor environmental factors, which were not included in Table 2.1, but investigated in the existing studies, cover productivity, and estimates of possible economic effects. However, in spite of the extensive indoor air quality research and findings using multi-parametric statistical regression, most building occupants and managers are not aware of health risks associated with their building environments, nor they are aware how to reduce the risk of severe diseases associated with exposure to adverse indoor environments, such as asthma or cancer (U.S. EPA, 2001). Furthermore, techniques presently used in investigations of indoor environmental health impacts are those typically used for other environments, such as ambient or industrial air, and they are criticized for not being suitable to discover parametric associations indoors. Thus, many causes of building occupants’ health problems remain unclear and as a consequence, the knowledge today is not much greater than it was a century ago (Sundell, 2004). Reported difficulties associated with application of multi-parametric regression include insufficient ability to deal with concurrent changes of values in a
large number of simultaneously acting parameters and lack of predicting capabilities. Therefore, this dissertation will explore opportunities for establishing indoor health and environmental parameter causal relationships using NN analyses able to account for simultaneous dynamic change of multiple indoor parameters.

2.4 Alternative Data Analyses Methodology

Need of a new multi-parametric data analyses method, alternative to previously described statistical regression, is necessary in order to interpret the building sensor information for the benefit of building occupants and managers. Particularly, the goal is to establish a prediction methodology for determining respiratory health symptom occurrence in office building environments. For this purpose, analyses based on artificial neural networks (NN) are considered. The NNs are data mining tools which are increasingly used in medicine and biology. For example, a study found over 800 papers within 18 areas of life science and 25 areas of clinical medicine published during 2000 and 2001 (Robert et al., 2004). Furthermore, usage of NNs for classification and prediction in biological systems showed equal (Jaimes et al., 2005) or better performance (Boulle et al., 2001) than statistical regressions.

Additionally, NNs have been successfully used in predictions of outdoor air quality throughout the world (NIOSH, 2002; Hasham et al., 2004). Attempts of applying NNs in determination and prediction of contaminant concentrations in outdoor urban environments showed that the accuracy of NN predictions depended upon the type of the pollutants and type of NN (Kolehmainen et al., 2000). More recent studies used NNs to forecast concentrations of nitrogen oxides (NO$_x$), sulfur dioxide (SO$_2$) and ozone (O$_3$) in a number of cities in Europe (Dorling et al., 2003), as well as CO$_2$ concentrations along the roads in China (Yang et al., 2007).

Using the experiences from outdoor environments, a study described application of NNs for fast identification of contaminant source positions within buildings (Vukovic and Srebric, 2007). The study used NNs’ non-linear mapping ability to match indoor contaminant source positions to concentration distribution patterns. The method was independent of the contaminant species, source types or background contaminant concentration levels.

Extensive literature review did not reveal any other studies considering the
use of NNs in prediction of indoor environmental parameters. However, existing
NN applications in building systems include: 1) predictions of building energy
consumption and 2) controls of heating, ventilation and air-conditioning (HVAC)
systems. Several papers gave an overview of NN properties, concepts, and de-
scribed applications to building energy consumption predictions (Breekweg et al.,
2000). Another study provided two examples for NN performed estimations of
future building energy use, based on the historical energy usage data. The study
also described an additional case of NN controls application (Curtiss et al., 1996).
Eventually, NNs predictions of building thermal loads can lead to improved con-
trols of energy storage systems and minimize HVAC operating costs (Kawashima
et al., 1996).

The intention of the dissertation research is to extend applications of NNs to
predicting respiratory health symptom occurrence in office building environments.
Methods for Determining Health Impacts of Building Indoor Environments

The most widely used current method for determining health impacts of building indoor environments includes multi-parametric logistic regression. Overview of this method will be presented together with an insight into alternative multivariate data analyses approach based on artificial neural networks. The overview tries to address the need of an interdisciplinary research capable of bringing together science and people, engineering and medicine, in order to provide methodology for predicting health impacts of building indoor environments.

3.1 Multivariate Logistic Regression

Multivariate logistic regression is used to determine relations between a target dependent variable or response and a set of independent variables or predictors. The term logistic is used to indicate a categorical set of values that the target variable can take, as opposite to continuous values. Although there are no theoretical limitations in the number of categories, typically, the target variable is binary or dichotomous, taking the values \{0, 1\}, or \{no, yes\}. Thus, Equation 3.1 models the response variable in a univariate logistic regression, having one independent
variable (Larose, 2006).

\[ Y = \pi(x) + \varepsilon, \quad \pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \] (3.1)

Here, \( Y \) is the response variable, \( x \) is the value of an independent variable, and \( \varepsilon \) is the regression model error. Sigmoidal function \( \pi(x) \) takes values between 0 and 1, and can be interpreted as a probability that binary response variable \( Y \) will be equal to 1. Coefficients \( \beta_0 \) and \( \beta_1 \) are model parameters.

Applying the so called logit or logistic transformation results in a linear function 3.2 connecting logistic with linear regression models.

\[ l(x) = \ln \left( \frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 x \] (3.2)

Now, coefficients \( \beta_1 \) and \( \beta_0 \) can be understood as a slope and an ordinate axis intercept of a linear regression function approximation \( l(x) \). This approach can be extended to include multiple variables \( x_i \) by introducing additional terms and \( \beta_i \) model parameters.

Unlike in linear regression, no analytical solutions exist for \( \beta_i \) coefficients in logistic regression. However, they can be estimated using maximum likelihood estimation, which finds the parameters to maximize likelihood of observing the given data set. To estimate the significance of determining coefficients in such a way, additional tests of statistical significance are performed. Typically, a 95% confidence interval is calculated, specifying the interval in which the actual coefficients will fall with 95% probability. Additionally, calculation of a significance level or p-value allows estimation whether the established relations could be insignificant, i.e. in the case of univariate analyses, whether \( \beta_1 \) could be equal to zero (Larose, 2006). Thus, a typical p-value of 0.05 is used as a threshold between significant \((p \leq 0.05)\) and insignificant \((p \geq 0.05)\) relations, such that p-values less than 0.05 indicate less than 5% chances the established relations are insignificant. Alternatively, such p-values point that there is less than 5% probability for the obtained observations to be achieved in data analyses, although no actual relations exist between independent and response variables. For example, establishing a relation between relative humidity and sinus congestion with p-value of 0.005 (based on
data analyses) means that there is only 0.5% probability that such relation between relative humidity and sinus congestion is insignificant. Another measure, called coefficient of determination, $R^2$, is often used in regression to quantify how well regression approximation represents the actual data. In the case of a linear regression, Equation 3.3 calculates $R^2$ value ranging between 0 and 1, where values above 0.5 (absolute $R$ greater than 0.7) typically indicate good fit to the actual data for research related to human subjects (Larose, 2006).

$$R^2 = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}$$

(3.3)

Here $\hat{y}_i$ and $y_i$ are predicted and actual responses, respectively, while $\bar{y}$ is the actual response mean. Similarly, in logistic regression several pseudo $R$-squared values could be calculated to determine “goodness of fit” for the logistic model approximations. In equation 3.3, a difference between the denominator and nominator is equal to the sum of squared approximation errors (SSE) and it is given by Equation 3.4.

$$SSE = \sum_i (y_i - \hat{y}_i)^2$$

(3.4)

Dividing SSE with the number of degrees of freedom yields a measure of dispersion around the expected (mean) value called variance and equal to the squared standard deviation, as given by Equation 3.5.

$$Var = \sigma^2 = \frac{SSE}{n - 1}$$

(3.5)

Apart from confidence intervals, p- and quasi $R^2$-values, errors and variance, finding the described $\beta_i$ logistic regression model coefficients also allows calculation of odds. An odd is defined as a ratio of probability for occurrence of specific response and probability that such response will not occur. For a binary response variable, the calculated odds of two possible outcomes could be divided to obtain the so called odds ratio (OR). OR is a statistical parameter for comparison of probabilities for occurrence of specific outcomes, widely used in research involving human subjects. For example, in case of examining relations between the human sense of high humidity and occurrence of sinus congestion, OR=2.7 would indicate
that a person sensing high humidity is 2.7 times more likely to experience sinus congestion than the person not sensing high humidity. Equation 3.6 defines the OR in case of a binary target variable.

\[
\text{OR} = \frac{\pi(1)}{1-\pi(1)} \cdot \frac{1-\pi(0)}{\pi(0)} = \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = e^{\beta_1}
\]  

(3.6)

OR values are typically given with confidence intervals and p-values, calculated using the results of statistical tests applied in estimations of significance for previously determined logistic model parameters. Also, unadjusted and adjusted OR values are typically presented to account for possible influences of other variables on specific independent variable of interest in the logistic regression model. Thus, the adjusted value represents an estimate of predictor’s influence on the response variable when all other influences would be held constant.

### 3.2 Artificial Neural Networks

As well as multivariate logistic regression, artificial neural networks are also used for data mining and extracting information from data sets, including predictions, classification, and pattern recognition. As parallel computational architectures, artificial neural networks were originally developed during the 1960’s, inspired by features of biological systems, such as human brain (Simpson, 1990). Hence, the term “artificial” is usually used to distinguish between the biological and developed computational neural systems. The present study, however, is not dealing with biological systems, and will use the term neural network (NN) to describe such artificial computational architectures.

NNs consist of connected computational nodes or neurons, which can be organized in layers. The connections between neurons mark paths for communication and data exchange, which typically does not happen between the neurons within the same layer. Hence, only layers of neurons are connected, and such a structure is known as multi-layer perceptron (MLP). Although other NN organizational structures also exist, MLP is the most common type of NNs (Nørgaard et al., 2000). MLP networks consist of input, output and one or more hidden layers, as shown
in Figure 3.1. As MLP can have one or more hidden layers, the total number of the layers is usually stated in the description of a network, with a custom of not counting the input layer. Hence, a two-layer network will have one input, one hidden and one output layer, three-layer NN will have two hidden layers etc.

Figure 3.1. An example of a two-layer neural network

Apart from organizational structure, different NN types also distinguish how information is exchanged between connected layers of neurons. One of the most common types is feed forward MLP, where data always flow from the input towards the output layer. In this process NN input data are transformed into the values obtained at the output neurons. The entire transformation process depends upon the operation of individual layers of neurons, which can be described as:

$$a = f(w \cdot p + b)$$ (3.7)

where \(a\) is the output vector of the layer, \(w\) is the weight matrix, \(p\) is the input vector, \(b\) is the matrix of biases and \(f\) is the transfer function (typical transfer functions are given in Table 3.1). Biases \(b\) can also be considered as weights applied to a constant input vector \(p\), set to 1. In the MLP NN, Equation 3.7 applies to each layer of neurons, taking the output of the previous layer as a new input vector and using the weights and biases of the current layer. Thus, for a two-layer NN, Equation 3.8 calculates values at the output neurons as:

$$\hat{a}_i = G_i[p, w] = F_i \left[ \sum_{j=1}^{n_h} W_{i,j} f_j \left( \sum_{l=1}^{n_p} w_{j,l} p_l + w_{j,0} \right) + W_{i,0} \right]$$ (3.8)
Here $\hat{a}_i$ represents the output of the $i$-th neuron in the output layer, $p_l$ are network inputs, $w_{j,l}$ and $W_{i,j}$ are network weights for hidden and output layers, $w_{j,0}$ and $W_{i,0}$ are biases for hidden and output layers, $f_j$ and $F_i$ are transfer functions for hidden and output layers, while $n_p$ and $n_h$ present the total numbers of input and hidden neurons, respectively. While the expression within the parenthesis of Equation 3.8 represents the operation of the hidden layer of neurons, the rest of the right-hand side of the equation represents the operation of the output neuron, $i$. When all adjustable parameters are replaced with a vector $w$, function $G_i$ presents the transformation of the input domain into the final NN output.

The important feature of NNs is their ability to train adjustable parameters, weights and biases, such that desirable transformation of input into output domains is established. Thus, network has the ability to adapt to changes in presented sets of input and/or desirable output data. The adaptation process is called training or learning, as it resembles learning in biological systems. Training usually requires several repeated presentations of input and/or output data sets, typically called training sets. Two general training procedures exist: unsupervised and supervised. The unsupervised training is used when no desired NN output data exist. In such cases, only input data are provided to the network, which is able to adjust connections between the nodes to recognize patterns in the input data set. On the other hand, in the supervised training, network uses samples of input data and output (target) results to evaluate connections between the nodes and improve input-output transformation performance.

A possible way to train feed-forward MLP NNs is backpropagation learning algorithm. This algorithm uses gradient of the function indicating performance of the current network, to calculate updates for the network’s weights and biases. The gradient is calculated going backwards through the network, from output to input neurons. The typical performance functions in backpropagation include sum of squared errors (SSE) and mean squared error (MSE) given by Equations 3.9 and 3.10, while the simplest implementation of the backpropagation algorithm is gradient descent given by Equation 3.11 (Hagan et al., 1996).

$$H(w) = \sum_i (t_i - a_i)^2$$  \hspace{1cm} (3.9)
\[ H(w) = \frac{1}{\text{max}\{i\}} \sum_i (t_i - a_i)^2 \]  

(3.10)

\[ w_{k+1} = w_k - \alpha_k g_k \]  

(3.11)

Here, \( H(w) \) is the performance function of network’s weights and biases, \( t_i \) and \( a_i \) are target values from the set of all training data and corresponding NN outputs, \( \text{max}\{i\} \) is the total number of training targets, \( k \) marks the iteration of the backpropagation algorithm, \( \alpha \) is the learning rate which can be constant or adjustable, while \( g \) is the gradient of the performance function \( H \) with respect to the weights and biases \( w \). Notice that Equation 3.9 resembles Equation 3.4 used in regression.

Depending upon the way NN updates the weights and biases, training can be either incremental or batch. In incremental learning, NN parameters are adjusted after presentation of each individual member of the training set. On the other hand, in batch training NN parameters are updated only after all training cases are presented. To distinguish between incremental and batch training, repeated presentations of the training data are termed iterations and epochs, respectively. After a sufficient number of training examples, the network should achieve appropriate values for weights and the training could end. Apart from satisfactory NN performance determined by the specific goal the performance function should reach, several other stopping criteria are related with performance of the training algorithm and can include: (1) exceeding the preset number of training iterations or epochs; (2) reducing the magnitude of the gradient below preset limits; or (3) exceeding the preset training time. Well trained network is able to compute the desired result based upon given input data, which do not need to be part of the training input dataset. Hence, trained NNs have the ability to generalize transformation between the input and output multidimensional domain spaces and predict output values for inputs in between the presented training cases. Important limitation, however, deals with the range of possible NN input values, which should fall within the range of presented training inputs.

### 3.2.1 How to Construct a Neural Network?

To construct a NN, Neural Network toolbox is available within MATLAB software package (MathWorks, 2008). The toolbox offers a variety of built-in functions,
which can be used to automatically preprocess data, design, train and test NNs, as well as postprocess results (Demuth and Beale, 1998). It also offers possibilities to program and implement user defined code, written using MATLAB programming syntax or translated from other programming languages, such as Visual Basic or C++. Additionally, a graphical user interface is available to simplify the NN design and visualize certain NN features.

Alternatively, researchers have also used hardware designed NNs, which more resemble their biological counterparts, in efforts to reproduce some of the processes happening in biological systems (Frye et al., 2007). However, for computational prediction purposes discussed in the current study, MATLAB software package offers sufficient capabilities.

After selecting the software, specific NN design questions related to the type of network need to be addressed. Assuming the selection of feed-forward MLP with backpropagation training algorithm, the remaining design decisions include specifying: (1) number of hidden layers, (2) number of nodes in input and output layers, (3) types of transfer and training functions, as well as (4) choice between batch and incremental training processes.

1. Number of hidden layers
   Total number of layers in a MLP NN, having one input and one output layer, depends upon the number of hidden layers. Additional hidden layers add to the network’s complexity, and should improve ability to make predictions. However, starting with a single hidden layer and going towards more complex NN structures is common practice.

2. Number of nodes
   The number of computational nodes in the NN input layer should equal the number of input parameters or input data dimensionality. Similarly, the number of nodes in the output layer should equal the number of variables whose values are predicted. The number of neurons in hidden layers can vary and impacts the NN complexity and ability to make predictions. For example, a study found significant impacts the number of hidden neurons had on accuracy of NN classifications and nonlinearity, pointing that too large hidden layers led to reduced abilities of NNs to generalize, ie. overfitting.
(Weibao et al., 2009). Rules of thumb are used to determine the size of hidden layers anywhere between the arithmetic mean and sum of input and output neurons. Typically, using less neurons than the arithmetic mean of inputs and outputs yields poor NN performance and insufficient ability to make transformations of input into output domains. On the other hand, too many hidden neurons do not necessarily improve NN performance and impact network’s ability to generalize, i.e. produce viable results with new or “unseen” input data which were not used in the NN training process.

3. Transfer functions

Various transfer functions are used to transform data forwarded from one layer of neurons to another in a MLP NN. Table 3.1 presents an overview of typically used transfer functions. Choosing an appropriate transfer function for solving a particular problem, depends upon the problem itself. For example, utilization of a linear transfer function cannot lead to accurate solutions of nonlinear problems. Also, step functions, such as hardlim, are typically used in data classification. For nonlinear predictions of continuous variables, nonlinear continuous transfer functions are usually applied, such as log-sigmoid or hyperbolic tangent sigmoid. Furthermore, two-layer NNs with sigmoid transfer function in the first layer and linear in the second, can approximate most mathematical functions arbitrarily well (Hagan et al., 1996).

4. Batch or incremental training?

Batch training increases the overall requirements of computational resources, as it requires storing all training dataset performance parameters. This might increase NN training time. On the other hand, incrementally trained NNs are more susceptible to differences between members of the training set, as well as to the order training cases are presented. Thus, whenever computational resources are sufficient and training time is not critical, batch training is a better choice.

5. Backpropagation training functions

Table 3.2 presents an overview of typically used backpropagation training functions (Demuth and Beale, 1998). They are based on mathematical search
<table>
<thead>
<tr>
<th>Function type</th>
<th>MATLAB function and definition*</th>
</tr>
</thead>
<tbody>
<tr>
<td>hard limit</td>
<td>hardlim(n) = { 1, n ≥ 0 \n 0, n &lt; 0 }</td>
</tr>
<tr>
<td>symmetric hard limit</td>
<td>hardlims(n) = { 1, n ≥ 0 \n -1, n &lt; 0 }</td>
</tr>
<tr>
<td>linear</td>
<td>purelin(n) = n</td>
</tr>
<tr>
<td>saturating linear</td>
<td>satlin(n) = { 1, n &gt; 1 \n n, 0 ≤ n ≤ 1 \n 0, n &lt; 0 }</td>
</tr>
<tr>
<td>symmetric saturating linear</td>
<td>satlins(n) = { 1, n &gt; 1 \n n, -1 ≤ n ≤ 1 \n -1, n ≤ -1 }</td>
</tr>
<tr>
<td>positive linear</td>
<td>poslin(n) = { n, n ≥ 0 \n 0, n &lt; 0 }</td>
</tr>
<tr>
<td>hyperbolic tangent sigmoid</td>
<td>tansig(n) = ( \frac{e^n - e^{-n}}{e^n + e^{-n}} )</td>
</tr>
<tr>
<td>log-sigmoid</td>
<td>logsig(n) = ( \frac{1}{1 + e^{-n}} )</td>
</tr>
<tr>
<td>competitive</td>
<td>compet(n) = { 1, \text{for the neuron with } n_{\text{max}} \n 0, \text{for all other neurons} }</td>
</tr>
</tbody>
</table>

* n = wp + b - weighted neuron input

Table 3.1. Typically used neural network transfer functions

Algorithms used to find extreme values of functions. The basic training function, steepest or gradient descent, was described previously in Section 3.2. Other functions listed in Table 3.2 provide faster and improved training performance including ability to pull out of certain local minima of the performance function and search for global minimization. The faster performance is based either on heuristics applied to gradient descent (such as momentum, variable learning rate and resilient backpropagation) or on other numerical optimization techniques (such as conjugate gradient, quaizi-Newton and Levenberg-Marquardt methods). Line search routines are related with application of certain conjugate gradient and quaizi-Newton algorithms (Demuth and Beale, 1998).
<table>
<thead>
<tr>
<th>Function type</th>
<th>MATLAB function</th>
</tr>
</thead>
<tbody>
<tr>
<td>gradient descent</td>
<td>LEARNGD, TRAINGD</td>
</tr>
<tr>
<td>gradient descent with momentum</td>
<td>LEARNGDM, TRAINGDM</td>
</tr>
<tr>
<td>variable learning rate</td>
<td>TRAINGDA, TRAINGDX</td>
</tr>
<tr>
<td>resilient backpropagation</td>
<td>TRAINRP</td>
</tr>
<tr>
<td>conjugate gradient (Fletcher-Reeves, Polak-Ribiére, Powell-Beale and scaled)</td>
<td>TRAINCGF, TRAINCGP, TRAINCGB, TRAINSCG</td>
</tr>
<tr>
<td>line search (golden section, Brent’s, hybrid bisection-cubic, Charalambous’ and backtracking)</td>
<td>SRCHGOL, SRCHBRE, SRCHHYB, SRCHCHA, SRCHBAC</td>
</tr>
<tr>
<td>quasi-Newton (Broyden-Fletcher-Goldfarb-Shanno and one step secant)</td>
<td>TRAINBFG, TRAINOSS</td>
</tr>
<tr>
<td>Levenberg-Marquardt</td>
<td>TRAINLM</td>
</tr>
</tbody>
</table>

*Table 3.2. Typically used neural network backpropagation training functions*
Building Assessment Survey and Evaluation Benchmark Dataset

Building Assessment Survey and Evaluation (BASE) study is a comprehensive database containing information from one hundred U.S. office buildings, including indoor environmental measurements, building characteristics, and questionnaire results of building occupants’ health and performance factors (U.S. EPA, 2003). BASE study protocol includes elaborate information on conducting the study over the period of 10 years. The history of protocol changes and improvements are presented together with the criteria for selection of buildings within major U.S. cities. Additionally, the protocol contains procedures for determination of optimal measurement sites within the selected buildings, description and schedule of all data collection activities during the one week long building field study. Finally, sample and data management procedures are described together with the measures taken to provide quality assurance. As the BASE protocol also contains original documentation used by the field workers to administer the study, it is an indispensable reference for conducting any similar experimental investigations.

BASE data were used to perform statistical analyses and neural network training in the development and evaluation of multi parametric logistic regression models and NN algorithms for predicting respiratory health symptom occurrence in office building environments. In order to be used in such analyses, data from the dataset needed to be preprocessed. The preprocessing, also called “the cleaning phase”, firstly required conversion of the total of 88 database file formats to uni-
versally readable text formats which could be accessed by any computer program, rather than by only specific database software. This allowed greater flexibility in designing the software to process the data, which was essential to enable creation of NN prediction models. Furthermore, the conversion included data coding - replacement of all non-numerical descriptions into unique code sequences. Examples of such coding included alphanumeric building identifiers, comments about monitoring locations, time, date, type of recordings, which were all replaced with numerical values. Information about the code sequences and corresponding original data entries were stored in separate files, called coding keys. After the coding, extraction of significant data and exclusion of unneeded parameters was performed to reduce the amount of data in the analyses out of the original database containing a total of 2033 variables. For example, variables that were excluded contained quality assurance parameters, such as duplicate, field or shipping blank indicators. Furthermore, all non-essential information was stripped from the dataset, such as information on measuring equipment models, names of operators, comments and other data that were not used in the further analyses.

The original dataset contained three types of information: (1) measurement data (MTR) collected by the monitoring equipment at 3 fixed and 2 mobile indoor locations, as well as an outdoor location; (2) questionnaire data (QSN) administered to the building occupants during the study week; and (3) building survey information (SVY) describing overall building characteristics, management and maintenance practices. All three types of data were organized within separate folders containing multiple files. The contents and description of the files was available in the database accompanying information (U.S. EPA, 2008). The cleaning procedure and results of preprocessing of these files will be further described for each of the three groups of data.

4.1 Measurements

In the cleaning process of the BASE measurement data, all mobile location recordings were excluded. Such approach is consistent with other studies examining BASE data (Apte and Erdmann, 2002; Erdmann et al., 2002). Reasons for considering only data from the fixed monitoring locations originate in inconsistencies
between mobile monitoring protocols, which were frequently changed during the course of BASE study, making the collected data from mobile locations more difficult to compare among all study buildings. Generally, the mobile location data were collected by carts equipped with portable sensors and moved between multiple locations, consecutively performing measurements for at least 10 minutes. However, the mobile measurement protocols provided no consistent requirements with respect to the number and order particular monitoring spots were engaged, nor the exact time the measurements were taken. Furthermore, protocol changes affected the type of sensors used in mobile monitoring as well as the number of used mobile monitoring carts. Therefore, only data collected at three fixed indoor and one outdoor location were used in the current analyses.

As a result of the measurement data cleaning process, Table 4.1 contains essential measurement variables extracted from 420 variable long original measurement database.

Table 4.1 includes folder and file location of the variables of interest, together with the description and units of collected measurements. Essentially, the data includes measurements of:

- temperature,
- relative humidity,
- CO$_2$ concentration,
- carbon monoxide (CO) concentration,
- VOC concentrations of 73 chemical compounds,
- aldehyde concentrations (formaldehyde and acetaldehyde),
- particle concentrations (PM$_{2.5}$ and PM$_{10}$),
- bioaerosols (bacteria, fungi, fungal spores),
- allergen concentrations (cat mite and two types of dust mites),
- radon concentration,
- noise level, and
- light intensity.
<table>
<thead>
<tr>
<th>Folder</th>
<th>File Name</th>
<th>Variable Name</th>
<th>Description</th>
<th>Units</th>
<th>Other variables to consider</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTR</td>
<td>aldehyde</td>
<td>CONCENT2</td>
<td>concentration of either formaldehyde or acetaldehyde</td>
<td>ppb</td>
<td>EVENT, CLASS, SITEID, COMPOUND</td>
</tr>
<tr>
<td>MTR</td>
<td>antigen</td>
<td>DER_FI</td>
<td>concentration for dust mites 1 (Dermatophagoides farinae)</td>
<td>µg/g</td>
<td>EVENT, SITEID, LOCATION</td>
</tr>
<tr>
<td>MTR</td>
<td>antigen</td>
<td>DER_PI</td>
<td>concentration for dust mites 2 (Dermatophagoides pteronyssinus)</td>
<td>µg/g</td>
<td>EVENT, SITEID, LOCATION</td>
</tr>
<tr>
<td>MTR</td>
<td>antigen</td>
<td>FEL_D1</td>
<td>concentration for cat mites (Feline domesticus)</td>
<td>µg/g</td>
<td>EVENT, SITEID, LOCATION</td>
</tr>
<tr>
<td>MTR</td>
<td>bactair</td>
<td>CONCENTR</td>
<td>concentration of individual bacterial groups in air (Actinobacteria, Bacillus, gm+ cocci, gm-cocci, gm+ rods, gm- rods, unknown)</td>
<td>CFU/m³</td>
<td>EVENT, CLASS, SITEID, BACT_GRP</td>
</tr>
<tr>
<td>MTR</td>
<td>bactair</td>
<td>TOTAL2</td>
<td>total concentration of bacteria in air</td>
<td>CFU/m³</td>
<td>EVENT, CLASS, SITEID, BACT_GRP</td>
</tr>
<tr>
<td>MTR</td>
<td>bactdry</td>
<td>CONCENTR</td>
<td>concentration of individual bacterial groups in dry bulk (Actinobacteria, Bacillus, gm+ cocci, gm-cocci, gm+ rods, gm- rods, unknown)</td>
<td>CFU/g</td>
<td>EVENT, SITEID, LOCATION, BACT_GRP</td>
</tr>
<tr>
<td>MTR</td>
<td>bactdry</td>
<td>TOTAL2</td>
<td>total concentration of bacteria in dry bulk sample</td>
<td>CFU/g</td>
<td>EVENT, SITEID, LOCATION, BACT_GRP</td>
</tr>
<tr>
<td>MTR</td>
<td>bactwet</td>
<td>CONCENTR</td>
<td>concentration of individual bacterial groups in wet bulk (Actinobacteria, Bacillus, gm+ cocci, gm-cocci, gm+ rods, gm- rods, unknown)</td>
<td>CFU/ml</td>
<td>EVENT, LOCATION, BACT_GRP</td>
</tr>
<tr>
<td>MTR</td>
<td>bactwet</td>
<td>TOTAL2</td>
<td>total concentration of bacteria in wet bulk sample</td>
<td>CFU/ml</td>
<td>EVENT, LOCATION, BACT_GRP</td>
</tr>
<tr>
<td>MTR</td>
<td>co</td>
<td>CO</td>
<td>CO concentration (should be averaged daily and overall with respect to variable DATE)</td>
<td>ppm</td>
<td>EVENT, CLASS, SITEID, DATE</td>
</tr>
<tr>
<td>MTR</td>
<td>co2</td>
<td>CO2</td>
<td>CO₂ concentration (should be averaged daily and overall with respect to variable DATE)</td>
<td>ppm</td>
<td>EVENT, CLASS, SITEID, DATE</td>
</tr>
<tr>
<td>MTR</td>
<td>fungair</td>
<td>CONCENTR</td>
<td>concentration of individual fungal groups in air (52 different fungal groups)</td>
<td>CFU/m³</td>
<td>EVENT, CLASS, SITEID, FUNG_GRP</td>
</tr>
<tr>
<td>MTR</td>
<td>fungair</td>
<td>TOTAL2</td>
<td>total concentration of culturable fungi in air</td>
<td>CFU/m³</td>
<td>EVENT, CLASS, SITEID, FUNG_GRP</td>
</tr>
<tr>
<td>MTR</td>
<td>fungdry</td>
<td>CONCENTR</td>
<td>concentration of individual fungal groups in dry bulk sample (40 different fungal groups)</td>
<td>CFU/g</td>
<td>EVENT, SITEID, LOCATION, FUNG_GRP</td>
</tr>
<tr>
<td>MTR</td>
<td>fungdry</td>
<td>TOTAL2</td>
<td>total concentration of fungi in dry bulk sample</td>
<td>CFU/g</td>
<td>EVENT, SITEID, LOCATION, FUNG_GRP</td>
</tr>
<tr>
<td>MTR</td>
<td>fungspar</td>
<td>CONCENTR</td>
<td>concentration of individual fungal groups in air (29 different fungal spore groups)</td>
<td>spores/m³</td>
<td>EVENT, CLASS, SITEID, FUNG_GRP</td>
</tr>
<tr>
<td>MTR</td>
<td>fungspar</td>
<td>TOTAL2</td>
<td>total concentration of fungal spores in air</td>
<td>spores/m³</td>
<td>EVENT, CLASS, SITEID, FUNG_GRP</td>
</tr>
<tr>
<td>MTR</td>
<td>fungwet</td>
<td>CONCENTR</td>
<td>concentration of individual fungal groups in wet bulk sample (30 different fungal groups)</td>
<td>CFU/ml</td>
<td>EVENT, LOCATION, FUNG_GRP</td>
</tr>
<tr>
<td>MTR</td>
<td>fungwet</td>
<td>TOTAL2</td>
<td>total concentration of fungi in wet bulk sample</td>
<td>CFU/ml</td>
<td>EVENT, LOCATION, FUNG_GRP</td>
</tr>
<tr>
<td>MTR</td>
<td>hvacco2</td>
<td>CO2</td>
<td>CO₂ concentration in HVAC supply and return (specified with SITEID variable)</td>
<td>ppm</td>
<td>EVENT, CLASS, SITEID, DATE</td>
</tr>
</tbody>
</table>
Table 4.1. Essential measurement variables extracted after data cleaning

<table>
<thead>
<tr>
<th>Folder</th>
<th>File Name</th>
<th>Variable Name</th>
<th>Description</th>
<th>Units</th>
<th>Other variables to consider</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTR</td>
<td>light</td>
<td>LIGHT</td>
<td>average illuminance level over instrument averaging period (5 min)</td>
<td>lux</td>
<td>EVENT, CLASS, SITEID, DATE</td>
</tr>
<tr>
<td>MTR</td>
<td>multisorb</td>
<td>CONCENT2</td>
<td>concentration of individual volatile organic compounds (VOCs) (73 different compounds) - multisorbert samples</td>
<td>ppb</td>
<td>EVENT, CLASS, SITEID, COMPOUND</td>
</tr>
<tr>
<td>MTR</td>
<td>particul</td>
<td>CONCENTR</td>
<td>concentration of partulates PM10 or PM2.5 (specified with PARAMETE variable)</td>
<td>μg/m³</td>
<td>EVENT, CLASS, SITEID, PARAMETE</td>
</tr>
<tr>
<td>MTR</td>
<td>radon</td>
<td>CONCENTR</td>
<td>Concentration of radon in air</td>
<td>pCi/L</td>
<td>EVENT, CLASS</td>
</tr>
<tr>
<td>MTR</td>
<td>rh</td>
<td>RH</td>
<td>relative humidity (should be averaged daily and overall with respect to variable DATE)</td>
<td>%</td>
<td>EVENT, CLASS, SITEID, DATE</td>
</tr>
<tr>
<td>MTR</td>
<td>sound</td>
<td>SOUND</td>
<td>average sound level over instrument averaging period (5 min)</td>
<td>dB</td>
<td>EVENT, CLASS, SITEID, DATE</td>
</tr>
<tr>
<td>MTR</td>
<td>temp</td>
<td>TEMP</td>
<td>temperature (should be averaged daily and overall with respect to variable DATE, for each of the values of variable HEIGHT)</td>
<td>°C</td>
<td>EVENT, CLASS, SITEID, HEIGHT, DATE</td>
</tr>
<tr>
<td>MTR</td>
<td>tcounts</td>
<td>B10CC_xx - all</td>
<td>No. of people observed in the space (should be averaged)</td>
<td>-</td>
<td>EVENT</td>
</tr>
<tr>
<td>MTR</td>
<td>voc</td>
<td>CONCENT2</td>
<td>concentration of individual volatile organic compounds (VOCs) (78 different compounds) - canister samples</td>
<td>ppb</td>
<td>EVENT, CLASS, SITEID, COMPOUND</td>
</tr>
<tr>
<td>MTR</td>
<td>weather</td>
<td>DEWPOINT</td>
<td>outdoor weather data</td>
<td>°C</td>
<td>EVENT, DATE</td>
</tr>
<tr>
<td>MTR</td>
<td>weather</td>
<td>PRESSURE</td>
<td></td>
<td>mmHg</td>
<td>EVENT, DATE</td>
</tr>
<tr>
<td>MTR</td>
<td>weather</td>
<td>RH</td>
<td></td>
<td>%</td>
<td>EVENT, DATE</td>
</tr>
<tr>
<td>MTR</td>
<td>weather</td>
<td>TEMP</td>
<td></td>
<td>°C</td>
<td>EVENT, DATE</td>
</tr>
</tbody>
</table>

4.2 Occupant Questionnaire

The occupant questionnaire contained 34 multiple choice questions grouped into 4 sections: (1) workplace information, such as cleanliness of the workspace; (2) health and well-being, such as diagnosed illnesses; (3) workplace conditions, such as occupant perception of workspace temperature; and (4) job characteristics, such as requirements for fast job performance.

Unlike measured data, all questionnaire data of interest for further analyses were contained within a single survey file. Table 4.2 presents 44 variables extracted out of the 161 parameters contained in this file.

Table 4.2 also contains folder, file location and description of the questionnaire
<table>
<thead>
<tr>
<th>Folder</th>
<th>File Name</th>
<th>Variable Name</th>
<th>Description</th>
<th>Other variables to consider</th>
</tr>
</thead>
<tbody>
<tr>
<td>QSN</td>
<td>survey</td>
<td>ODORCHEM</td>
<td>frequency of occupant’s use of odorous chemicals at work</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>MIGRAINE</td>
<td>indication whether occupant reported diagnosed migraine</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>ASTHMA</td>
<td>indication whether occupant reported diagnosed asthma</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>ECZEMA</td>
<td>indication whether occupant reported diagnosed eczema</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>HAYFEVER</td>
<td>indication whether occupant reported diagnosed hay fever</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>DUSTALLE</td>
<td>indication whether occupant reported diagnosed dust allergy</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>MOLDALLE</td>
<td>indication whether occupant reported diagnosed mold allergy</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>TOBACSEN</td>
<td>occupant’s sensitivity to tobacco smoke in the workplace</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>CHEMISEN</td>
<td>occupant’s sensitivity to presence of chemicals in the air in the workplace</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>DRYEYES</td>
<td>frequency of dry, itching, or irritated eyes occupant reported in the workplace</td>
<td>EVENT, DRYEYESA</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>WHEEZIN</td>
<td>frequency of wheezing occupant reported in the workplace</td>
<td>EVENT, WHEEZINA</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>HEADACH</td>
<td>frequency of headache occupant reported in the workplace</td>
<td>EVENT, HEADACHA</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>SORETHR</td>
<td>frequency of sore or dry throat occupant reported in the workplace</td>
<td>EVENT, SORETHRA</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>FATIGUE</td>
<td>frequency of unusual tiredness, fatigue occupant reported in the workplace</td>
<td>EVENT, FATIGUEA</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>CHESTTI</td>
<td>frequency of chest tightness occupant reported in the workplace</td>
<td>EVENT, CHESTTIA</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>SINUSCO</td>
<td>frequency of sinus congestion occupant reported in the workplace</td>
<td>EVENT, SINUSCOA</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>COUGH</td>
<td>frequency of cough occupant reported in the workplace</td>
<td>EVENT, COUGHA</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>TIREDEY</td>
<td>frequency of tired or strained eyes occupant reported in the workplace</td>
<td>EVENT, TIREDEYA</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>TENSION</td>
<td>frequency of tension or irritability occupant reported in the workplace</td>
<td>EVENT, TENSIONA</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>PAINBAC</td>
<td>frequency of pain in back or shoulders occupant reported in the workplace</td>
<td>EVENT, PAINBACA</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>SNEEZIN</td>
<td>frequency of sneezing occupant reported in the workplace</td>
<td>EVENT, SNEEZINA</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>CONCENT</td>
<td>frequency of having difficulty concentrating occupant reported in the workplace</td>
<td>EVENT, CONCENTA</td>
</tr>
<tr>
<td>Folder</td>
<td>File Name</td>
<td>Variable Name</td>
<td>Description</td>
<td>Other variables to consider</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>---------------</td>
<td>-------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>DIZZINE</td>
<td>frequency of dizziness or lightheadedness occupant reported in the workplace</td>
<td>EVENT, DIZZINEA</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>DEPRESS</td>
<td>frequency of feeling depressed occupant reported in the workplace</td>
<td>EVENT, DEPRESSA</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>BREATH</td>
<td>frequency of shortness of breath occupant reported in the workplace</td>
<td>EVENT, BREATHA</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>NAUSEA</td>
<td>frequency of nausea occupant reported in the workplace</td>
<td>EVENT, NAUSEAA</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>DRYSKIN</td>
<td>frequency of dry skin occupant reported in the workplace</td>
<td>EVENT, DRYSKINA</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>HANDPAI</td>
<td>frequency of pain/numbness in hands/wrists occupant reported in the workplace</td>
<td>EVENT, HANDPAIA</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>MUCHAIR</td>
<td>frequency occupant reported too much air movement in the workplace</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>LITTAIR</td>
<td>frequency occupant reported too little air movement in the workplace</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>HOT</td>
<td>frequency occupant reported the workplace being too hot</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>COLD</td>
<td>frequency occupant reported the workplace being too cold</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>HUMID</td>
<td>frequency occupant reported the air being too humid in the workplace</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>DRY</td>
<td>frequency occupant reported the air being too dry in the workplace</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>SMOKEOD</td>
<td>frequency occupant reported tobacco smoke odors in the workplace</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>CHEMIOD</td>
<td>frequency occupant reported unpleasant chemical odors in the workplace</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>OTHEROD</td>
<td>frequency occupant reported other unpleasant odors in the workplace</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>CONVPRIV</td>
<td>occupant’s satisfaction with conversational privacy in the workplace</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>FREENOIS</td>
<td>occupant’s satisfaction with freedom from distracting noise in the workplace</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>CONFLIC1</td>
<td>frequency occupant asked to perform conflicting tasks by people equal in rank</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>CONFLIC2</td>
<td>frequency occupant asked to perform conflicting tasks by supervisors</td>
<td>EVENT</td>
</tr>
<tr>
<td>QSN</td>
<td>survey</td>
<td>CONFLIC3</td>
<td>frequency occupant asked to perform conflicting tasks by superiors</td>
<td>EVENT</td>
</tr>
</tbody>
</table>

Table 4.2. Essential measurement variables extracted after data cleaning

variables of interest. For example, EVENT variable contains a unique building identifier, lumped to an individual code assigned to each questionnaire partici-
pant. All the variables were dimensionless as they represented coded answers to questions in the questionnaire. The questions included 2, 3, 4, 5 and 6 choice answers, as well as several descriptive and numerical responses. However, all the relevant responses extracted in the cleaning procedure and presented in Table 4.2 included either 2, 4 or 5 choice questions.

4.3 Building Information

Building information part of the BASE database contains 10 files covering building physical characteristics, ambient sources, HVAC system description and performance measurements. The data were collected by building on-site inspection and interview with the building manager. Field workers used a comprehensive survey divided into three sections:

1. Whole building description, with information on building ambient sources of pollutants, containing:
   - General building information,
   - Space usage,
   - Building occupancy,
   - Climate and outdoor environment,
   - Building equipment,
   - Operating schedule,
   - Building envelope,
   - HVAC control systems,
   - Outdoor contaminant sources,
   - Smoking policies,
   - Water damage,
   - Fire damage,
   - Building renovations,
   - Cleaning schedules,
   - Trash storage,
   - Cleaning materials,
• Pest control, and
• Special use spaces;

2. Test space description, including test space pollution sources:
   • General test space description,
   • Test space furniture,
   • Test space interior finishes,
   • Test space lighting,
   • Test space ventilation fixtures,
   • Test space smoking,
   • Test space water damage,
   • Test space fire damage,
   • Test space renovations,
   • Test space cleaning,
   • Test space trash storage, and
   • Special use spaces;

3. Test space HVAC system description, covering:
   • Central air handling and distribution systems,
   • Perimeter zone units,
   • Unitary systems,
   • Evaporative cooling systems,
   • Outdoor air intake control,
   • Natural ventilation systems,
   • Air handler specifications,
   • Exhaust fan specifications,
   • Filtration and air cleaning systems,
   • Air washers,
   • Humidification systems,
   • Maintenance, and
   • Inspection.

As the building information part of BASE database was not used in further analyses, it was not subject to data cleaning procedures.
Chapter 5

Design, Testing and Comparison of Predictive Simulation Models

Multivariate logistic regression was used to examine association between indoor environmental parameters, occupants’ perceptions of indoor environments and respiratory health symptoms in BASE study buildings. Following the logistic regression, predictive simulation models for respiratory health symptom occurrence in office building environments were designed using artificial neural networks. This Chapter will describe procedure used for designing multivariate logistic regression and neural network models.

5.1 Multivariate Logistic Regression Design

Multivariate logistic regression analyses were conducted using SAS computer software (SAS, 2007), in collaboration with the Department of Public Health, Milton S. Hershey Medical School, The Pennsylvania State University. The selection of variables for analyses was based on the fact that no previous investigation used BASE data to explore association between reported occurrences of asthma, allergies and SBS respiratory problems on one side, and measured bacteriological, fungal, antigen and moisture levels on the other side, while such findings existed in analyses of other experimental data. Furthermore, occupants’ perceptions of indoor environments were included as additional parameters in the current analysis.

The list of investigated independent regression variables included 3 types of
antigens (2 dust mites and cat mite), 7 categories of bacteria (gram positive and gram negative cocci and rods), 52 types of culturable fungi, 29 types of airborne fungal spores, 3 indoor air parameters (temperature, relative humidity and CO₂ concentration), and 4 categories of perceptions occupants had on indoor environments (regarding humidity, chemical, smoke and other odors). Impact of these independent variables on 10 dependent occupant symptoms was examined using logistic regression. The dependent variables included: (1) upper respiratory SBS symptoms: sneezing, sinus congestion and sore throat; (2) lower respiratory SBS symptoms: chest tightness, shortness of breath, coughing and wheezing; (3) asthma; and (4) dust and mold allergies. Working with such a large number of variables included the following procedure:

1. Calculate prevalence statistics for all variables,

2. Perform univariate logistic regression analysis with each individual independent variable and all dependent variables,

3. Perform multivariate logistic regression analysis with identified significant parameters from univariate analysis and adjust values to control for confounding.

Database prevalence statistics of antigens, bacteria, culturable fungi and airborne fungal spores are available in the literature (Macher et al., 2002; Tsai et al., 2002; Womble et al., 1999; Vukovic et al., 2008), as well as details on prevalence of related respiratory symptoms (Apte and Erdmann, 2002). The prevalence analyses pointed that airborne fungal spore counts were much more comprehensive compared to fungi culture, detecting greater variety of fungal taxa in more BASE investigated buildings.

Table 5.1 presents prevalence statistics of measured indoor temperature, relative humidity and CO₂ concentration in buildings examined by the BASE study. The temperature and humidity levels are within acceptable thermal comfort ranges for office environments (ASHRAE, 2004).

Table 5.2 presents the statistics on examined categories of perceptions occupants have on their indoor environments. These were collected from the BASE study questionnaire administered to the occupants in each investigated building.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indoor temperature, in ºC (ºF)</td>
<td>22.94 (73.3)</td>
<td>0.99 (33.8)</td>
<td>23.13 (73.6)</td>
<td>19.14 (66.5)</td>
<td>25.09 (77.2)</td>
</tr>
<tr>
<td>Indoor relative humidity, RH, in %</td>
<td>38.30</td>
<td>13.06</td>
<td>39.76</td>
<td>9.64</td>
<td>62.84</td>
</tr>
<tr>
<td>Indoor CO₂ concentration, in ppm</td>
<td>545</td>
<td>96</td>
<td>515</td>
<td>391</td>
<td>899</td>
</tr>
</tbody>
</table>

Table 5.1. Prevalence of the averaged building indoor temperature, relative humidity and CO₂ concentration in BASE buildings.

Variables used in the statistics mark frequencies during the last four weeks while at work that the occupants reported particular perception of their work environments, as follows:

- CHEMIOD - unpleasant chemical odors,
- SMOKEOD - tobacco smoke odors,
- OTHEROD - other unpleasant odors, and
- HUMID - air being too humid.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Not in the last 4 weeks</th>
<th>1-3 days in the last 4 weeks</th>
<th>1-3 days per week in the last 4 weeks</th>
<th>Every or almost every workday</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHEMIOD</td>
<td>3136 (80.31)</td>
<td>577 (14.78)</td>
<td>121 (3.10)</td>
<td>71 (1.82)</td>
</tr>
<tr>
<td>SMOKEOD</td>
<td>3480 (89.76)</td>
<td>181 (4.67)</td>
<td>95 (2.45)</td>
<td>121 (3.12)</td>
</tr>
<tr>
<td>OTHEROD</td>
<td>2493 (64.14)</td>
<td>704 (18.11)</td>
<td>320 (8.23)</td>
<td>370 (9.52)</td>
</tr>
<tr>
<td>HUMID</td>
<td>3252 (85.71)</td>
<td>288 (7.59)</td>
<td>123 (3.24)</td>
<td>131 (3.45)</td>
</tr>
</tbody>
</table>

Table 5.2. Frequency (percent) of various occupants’ perceptions of indoor environments (CHEMIOD - chemical odors; SMOKEOD - tobacco smoke; OTHEROD - other odors; HUMID - perception of high humidity).

For each of these choices, majority of the occupants did not report any negative perceptions of their office environments within the last 4 weeks they were at work. However, between 10 and 36%, depending upon a particular perception, did perceive their office environments as odorous or too humid, at least once. While the most common negative perception was about unpleasant non-chemical odors, the least number of times occupants perceived odors from tobacco smoke.

In further analyses, the set of four occupant perception variables was coded binary, such that reported negative perceptions occurring at least once in the last four weeks were coded as 1, while those that did not occur were coded as 0. It
is important to notice that these are not measured indoor odors, but rather the perceived odors by the building occupants.

5.2 Neural Network Design

Following the criteria described in Chapter 3, MLP was a preferred type of NN in the current analyses. As the simplest and the most proven MLP design, one hidden layer NN was selected with log-sigmoid transfer function between input and hidden layers and linear transfer function between hidden and output layers. Number of neurons in the input layer corresponded to the number of input parameters, while the output layer had a single output neuron corresponding to a specific respiratory symptom parameter. Backpropagation training algorithm was used with mean squared error performance function and conjugate gradient training function. Having an improved training performance in comparison to other training functions (Demuth and Beale, 1998), all four conjugate gradient functions available in MATLAB were evaluated to achieve optimal training. BASE data described in Chapter 4 were used in NN training and testing, such that buildings were randomly assigned into training or testing data sets. In order to use the data in NNs, each building in BASE needed to provide only one corresponding combination of input and output parametric values. Therefore, multiple data records collected from occupants and measurements in various indoor locations within each BASE building were averaged to represent entire building conditions with a single value per each of the considered variables. Table 5.3 lists all input and output parameters used in the final NN design.

5.2.1 Input Data Set

The NN input data consisted of selected averaged indoor environmental measurements and occupant perceptions of indoor environments. Results from multivariate logistic regression were used to make selection of input parameters most significant for occurrence of specific respiratory health symptoms. Although separate NNs were trained to predict each particular respiratory symptom, combination of input parameters and NN design were kept the same for each specific output
**Table 5.3.** List of input and output parameters used in NN design

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>Asthma</td>
</tr>
<tr>
<td>Relative humidity</td>
<td>Chest tightness</td>
</tr>
<tr>
<td>CO₂ concentration</td>
<td>Cough</td>
</tr>
<tr>
<td>Perception of unpleasant chemical odors</td>
<td>Dust allergy</td>
</tr>
<tr>
<td>Perception of other odors</td>
<td>Mold allergy</td>
</tr>
<tr>
<td>Perception of high humidity</td>
<td>Shortness of breath</td>
</tr>
<tr>
<td>Total volatile organic compound (TVOC)</td>
<td>Sinus congestion</td>
</tr>
<tr>
<td>concentration</td>
<td>Sneezing</td>
</tr>
<tr>
<td></td>
<td>Sore throat</td>
</tr>
<tr>
<td></td>
<td>Wheezing</td>
</tr>
<tr>
<td></td>
<td>Building symptom index (BSI)</td>
</tr>
</tbody>
</table>

NN design was flexible and allowed addition or removal of individual input parameters, such that influence of particular inputs could be explored. Thus, the original design started with only four most significant input parameters determined in the logistic regression and progressed towards the final list of inputs given in Table 5.3. Generally, addition of inputs allowed better NN prediction capabilities, although particular input factors improved performance more than others. Nevertheless, due to the very large number of available parameters in the BASE database, not all possible input combinations were tested. Rather, selection of input parameters was based on performed logistic regression analysis and available research experience. Furthermore, certain input parameters were not measured in all BASE buildings and their inclusion in the input set would reduce the number of available training and testing cases, as the buildings in which they were not measured would need to be excluded. This would impact the overall prediction accuracy generalization abilities of the NN model. For example, although such parameters as cat and dust mite antigen concentrations may be significant for specific respiratory symptoms, the number of available BASE buildings in which they were measured limited the overall size of available NN training and testing data sets, and reduced prediction accuracy. Therefore, such parameters were not used in the final NN design.

The aforementioned example also illustrated size limitations for inclusion of
larger number of NN input parameters. Namely, although additional input parameters would increase the ability of NNs to establish relations between input and output domain data sets, the NN complexity would also be increased. Consequently additional training cases would be necessary to provide desired NN prediction accuracy. As a rule of thumb the training dataset should include at least the number of cases equal to 10 times the product of input and output neurons. Having 100 building data available in BASE dataset, and considering that part of the data needed to be set aside for testing purposes, NN design with a single output neuron limited the maximum number of input neurons to 9.

5.2.2 Output Data Set

The NN output data consisted of selected occupant respiratory health symptoms. Due to the mentioned NN complexity limitations decision was made to train separate NNs for each particular output symptom. Furthermore, NN design was kept the same for each output symptom in one training cycle. Table 5.3 lists all output symptoms used in NN designs.

Respiratory health symptoms in Table 5.3 were reported by the occupants via questionnaire. In order to average occupants’ responses for each BASE building, probability that occupants would report a particular symptom within each building was calculated. While this was a straightforward procedure for binary choice answers indicating either presence or absence of diagnosed symptoms, such as asthma and allergies, the questions having quadruple choice answers needed additional consideration. Following the previously described approach in multivariate logistic regression, these multiple choice answers were also coded binary, such that presence of symptoms at least once within the last 4 weeks at work was used as a criterion to separate occupants into two groups. This allowed further calculation of probabilities indicating that occupants would experience a particular symptom at work at least once for the 4 week period prior to the study week.

In addition to particular respiratory symptoms, attempt was made to predict an indicator of overall SBS conditions within BASE buildings. One such indicator is building symptom index (BSI), the last parameter listed in Table 5.3. Representing an average number of symptoms occupants reported in the questionnaire, BSI
was selected for the purpose of comparison to other indoor environmental studies (Sofuoglu and Moschandreas, 2003). Since BSI value depends upon the formulation of questions in a particular questionnaire, as well as on a pool of possible symptoms, significant limitations are related to its usage. Nevertheless, attempts were made to consider reported symptoms as similar as possible to comparable building environmental studies. Table 5.4 presents an overview of the reference symptoms used to calculate BSI in literature (Burge et al., 1987) and those used in the present study.

<table>
<thead>
<tr>
<th>Literature BSI</th>
<th>Present study BSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dry eyes</td>
<td>Dry, itchy or irritated eyes</td>
</tr>
<tr>
<td>Itchy eyes</td>
<td></td>
</tr>
<tr>
<td>Runny nose</td>
<td></td>
</tr>
<tr>
<td>Blocked nose</td>
<td>Sinus congestion</td>
</tr>
<tr>
<td>Dry throat</td>
<td>Sore throat</td>
</tr>
<tr>
<td>Lethargy</td>
<td>Unusual tiredness, fatigue</td>
</tr>
<tr>
<td>Headache</td>
<td>Headache</td>
</tr>
<tr>
<td></td>
<td>Cough</td>
</tr>
<tr>
<td>Flu</td>
<td>Sneezing</td>
</tr>
<tr>
<td>Difficulty breathing</td>
<td>Shortness of breath</td>
</tr>
<tr>
<td></td>
<td>Wheezing</td>
</tr>
<tr>
<td>Chest tightness</td>
<td>Chest tightness</td>
</tr>
</tbody>
</table>

*Table 5.4.* Building symptom index (BSI) symptoms used in literature compared to the present study

Having the calculated probabilities for occurrence of each of the symptoms, as previously described, BSI can be obtained by summing up the values of these probabilities, rather than averaging the number of symptoms per each occupant. This approach is possible as probability for occurrence of each symptom contributes the same value to the total number of symptoms experienced by an average occupant. For example, a 20% probability for occurrence of headache, adds 0.2 symptoms to the total number of symptoms experienced by an average occupant. The significance of this approach is in the fact that BSI from various studies can be compared to each other, even when BSI pools of investigated occupant symptoms are inconsistent among the studies. Recalculation of BSI values for the consistent pool of symptoms requires only probabilities for occurrence of individual symptoms, without the need to know the number of occupants per building that responded to the
questionnaire, nor recalculation of personal BSI values for each building occupant. Consequently, predictions of BSI and other respiratory health symptoms reported in Table 5.3 were envisioned through the NN algorithm.

### 5.2.3 Neural Network Algorithm

The following sequence describes a neural network algorithm used to provide predictions of probabilities for occurrence of particular respiratory health symptoms.

**Input**: Data files  
**Output**: NN predictions

1. Define inputs and outputs;  
2. Exclude buildings which should not be part of the dataset;  
3. Normalize data;  
4. Optionally use principle component analysis to reduce number of inputs;  
5. Create input scatter plots (categorize data points with respect to outputs);  
6. Randomize inputs and outputs;  
7. Select testing and training samples;

**foreach** Number of hidden neurons (maximum to minimum) **do**

- Train three neural networks;

**foreach** Neural network **do**

- define type, size, transfer functions, performance function and ratio, number of training iterations, stopping criteria, minimum step, and search function;  
- train network;  
- save training diagram;

**end**

- Test the three trained networks;  
- Choose network with minimum testing error;

**end**

- Scale back to data values before normalization;  
- Write optimal network results to file;  
- Plot network performance figures;

The algorithm uses BASE data, taking into account the data collection protocol changes which have occurred during the BASE study time frame. Consequently,
records from some of the BASE buildings could not be compared to others and were excluded from the analyses. As the collected data contain large variety of parameters, NN algorithm offers possibility to apply principle component analyses to reduce the number of possible inputs only to those which most significantly impact variance of collected respiratory symptom data. Furthermore, with minor adjustments, the algorithm is applicable in analyses of any other health parameters being part of a multi building investigation study. Although only a single NN hidden layer is assumed, optimization is performed with respect to the number of hidden neurons and randomized initialization of NN variables, selecting the best performing NN out of the three differently initialized candidates for each combination of neurons. The NN algorithm additionally includes postprocessing of obtained results and automatically generates NN prediction performance graphs.

5.2.4 Training, Testing and Validation Scenarios

For the purpose of establishing valid predictions of respiratory health outcomes among building occupants, the recorded building data and occupants’ survey responses need to be divided into one of the three categories: training, testing or validation data.

5.2.4.1 Neural Network Training and Testing

Ninety percent of eligible buildings in the BASE study were used to provide input and output training data to adjust NN internal parameters using backpropagation learning algorithm described in Chapter 3. The remaining ten percent of randomly selected eligible buildings from the BASE dataset were used to apply trained NNs and evaluate their predictive capabilities on “unseen” data.

Training was controlled either by specifying the minimum value/minimum gradient of the performance function, or maximum number of training iterations, i.e. epochs. Reaching the specified training criteria resulted in ending the training iterations. Additionally, attempts were made to control for overtraining, which can happen when the same set of training data is presented to the network too many times. Overtrained NNs have the ability to perfectly reproduce the training outcomes, but reduced accuracy of predictions based on “unseen” inputs. In an
attempt to prevent such behavior, two approaches were tested: early stopping and regularization.

The early stopping approach tried to stop NN training before overtraining occurred. The criteria for optimized NN training were based on prediction performance on a control dataset. Namely, part of the BASE dataset was randomly selected and separated from the training data to be used for evaluating the performance of the trained NN after each training epoch. The purpose of this evaluation was to determine when the NN prediction error began to increase in the control dataset, so that NN training would be stopped even before the training criteria were reached. This way the minimum prediction error would be obtained on the control dataset and presumably would result in overall optimal NN training and testing performance. However, the early stopping approach was not successful, as such trained NNs failed to show sufficient trend in following the original target data values. Probable reason for poor early stopping performance is randomized selection of the control dataset which may not be representative of the entire range of training data, particularly taking into account relatively small sizes of the datasets. Consequently, NN performance on the control dataset was not entirely representative of the NN performance on the training data.

Therefore, an alternative overtraining prevention measure was applied using regularization. In regularization, rather than trying to minimize only NN mean squared error, an additional term in the form of mean squared weights is added to the performance function. The resulting performance function, $H(w)$, is given by Equation 5.1.

$$H(w) = \gamma \cdot \frac{1}{\max\{i\}} \sum_i (t_i - a_i)^2 + (1 - \gamma) \cdot \frac{1}{\max\{j\}} \sum_j w_j^2 \quad (5.1)$$

Here, $t_i$ and $a_i$ are target values from the set of all training data and corresponding NN outputs, $\max\{i\}$ is the total number of training targets, while $w_j$ includes all $\max\{j\}$ NN weights and biases. Comparing Equation 5.1 to the typical mean squared error NN performance function in Equation 3.10, notice the weight parameter $\gamma$ defined as a performance ratio ranging between 0 and 1. When $\gamma = 0$, NN training with regularization tries to minimize the mean squared weights only, as the performance function; while $\gamma = 1$ leads to a typical NN mean squared error.
error minimization. Consequently, using the performance ratio between 0 and 1 would optimize NN performance by minimizing its error, as well as values of its weights and biases. Smaller weights and biases result in smoother NN prediction functions, reducing the possibility of NN to ”overfit” the training data and loose generalization capacity. Tradeoff between the desired NN generalization capabilities and needed accuracy would lead towards determining the optimal value of the performance ratio, $\gamma$.

In addition to the possibility of using regularization performance function (MSEREG) and manually set values of the performance ratio, Matlab NN toolbox also offers a Bayesian regularization NN training function with automated calculation of performance ratios (TRAINBR). However, choosing Bayesian regularization function in the current BASE data analyses did not appear useful as NNs would usually lose their prediction accuracy. Possible reason may be randomized NN initialization in combination with noisy data and insufficient robustness of the Bayesian regularization. Therefore, influence of particular selection of initial NN parametric values on Bayesian regularization performance should be further investigated. On the other hand, regularization with manually set performance ratios was implemented successfully. Multiple runs were conducted to determine the optimal performance ratio values depending upon particular prediction scenario. Thus, performance ratios were set to 0.6 for all predicted outputs given in Table 5.3, except in cases of predicting dust allergies and sinus congestion where performance ratio of 0.7 was used, and BSI using 0.9. Such performance ratios indicate that NN generalization performance almost equally benefits from the minimization of NN weights as from the minimization of the training errors in cases of all outputs, except BSI. Consequently, without regularization, trained NNs would be less robust and more sensitive to the small changes in input parameters for all output predictions except BSI, due to the large values of weights and biases.

5.2.4.2 Neural Network Validation

To validate the developed prediction methodology against any possible bias that may exist in the BASE dataset as well as to validate the BASE data collection procedure and applicability in a variety of real building cases, an independent data collection and analysis was performed. These data were used as a validation case
and are fully described within Chapter 7.
Chapter 6

Results and Discussion on BASE Data Analyses

Successful design of multivariate logistic regression models described in Chapter 5 allowed determination of associations between indoor environmental parameters, occupants’ perceptions of indoor environments and exhibited respiratory health symptoms in BASE study buildings. On the other hand, neural networks used significant parameters identified by multivariate logistic regression to make predictions for occurrence of building occupants’ respiratory health symptoms. This Chapter will present results of the conducted analyses and discuss possible limitations of the developed methodology to predict respiratory health symptom occurrence in office building environments.

6.1 Multivariate Logistic Regression Analysis

The univariate regression was initially applied to all of the specified variables to identify the most significant independent parameters, and then the study proceeded with multivariate analyses. During the multivariate regression analyses, range of referent change in values for several variables had been adjusted to correspond to measured data accuracy and overall measured range. For example, we used 10%, rather than a unit change in relative humidity. To control for confounding, OR values were adjusted by introducing the following covariates: age, sex, smoking status, presence of carpets in buildings and volatile organic compound
6.1.1 Multivariate Logistic Regression Results

Table 6.1 presents the final OR results of multivariate logistic regression analyses for statistically significant influences (p-value less than 0.05), assuming 95% confidence intervals as given in the brackets. The Table also shows adjusted OR values and confidence intervals taking into account the aforementioned covariate parameters. The presented results were grouped into 4 categories:

1. Upper respiratory symptoms: Sneezing, Sinus congestion, Sore throat;
2. Lower respiratory symptoms: Chest tightness, Shortness of breath, Coughing, Wheezing;
3. Asthma; and
4. Allergies: Dust allergies, Mold allergies.

Not included in Table 6.1 are additional statistically significant associations established between certain types of fungi and allergies. However, these cases were excluded, because ORs and entire 95% confidence intervals were very close to 1.

6.1.2 Discussion on Multivariate Logistic Regression Analyses

The presented results emphasize relations between occupants’ perceptions of indoor environments and occupant health responses. Based on the results presented in Table 6.1, the most important occupants’ perception parameter is sense of unpleasant chemical odors, which is a statistically significant parameter for all investigated health responses. In most cases, occupants sensing unpleasant chemical odors are almost twice as likely to experience some respiratory response as those unable to sense the odor. Additionally, the most significant indicators for majority of respiratory responses are sense of other odors, sense of high humidity and relative humidity. These results should not be surprising, as the ability of building occupants to sense indoor air quality has been known since the Biblical times (Sundell, 2004). Furthermore, a study found strong association between the sensation of dryness and 3 groups of SBS symptoms: (1) general symptoms including
<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>CHEMIOD</th>
<th>SMOKEOD</th>
<th>OTHEROD</th>
<th>HUMID</th>
<th>RH</th>
</tr>
</thead>
<tbody>
<tr>
<td>unadjusted</td>
<td>1.689</td>
<td>1.352</td>
<td>1.847</td>
<td>1.607</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>(1.386-2.059)</td>
<td>(1.056-1.732)</td>
<td>(1.579-2.161)</td>
<td>(1.302-1.982)</td>
<td>0.855</td>
</tr>
<tr>
<td>adjusted</td>
<td>1.808</td>
<td>NS</td>
<td>1.705</td>
<td>2.015</td>
<td>(0.785-0.932)</td>
</tr>
<tr>
<td></td>
<td>(1.354-2.416)</td>
<td>NS</td>
<td>(1.353-2.149)</td>
<td>(1.496-2.714)</td>
<td>0.859</td>
</tr>
<tr>
<td>Sinus congestion</td>
<td>1.784</td>
<td>NS</td>
<td>1.699</td>
<td>2.140</td>
<td>(0.813-0.908)</td>
</tr>
<tr>
<td>adjusted</td>
<td>1.442</td>
<td>NS</td>
<td>1.778</td>
<td>2.711</td>
<td>0.848</td>
</tr>
<tr>
<td></td>
<td>(1.438-2.212)</td>
<td>NS</td>
<td>(1.440-2.003)</td>
<td>(1.689-2.712)</td>
<td></td>
</tr>
<tr>
<td>unadjusted</td>
<td>1.803</td>
<td>NS</td>
<td>1.922</td>
<td>1.965</td>
<td>(0.777-0.926)</td>
</tr>
<tr>
<td>Sore throat</td>
<td>1.330-2.444</td>
<td>NS</td>
<td>(1.488-2.484)</td>
<td>(1.365-2.831)</td>
<td>0.840</td>
</tr>
<tr>
<td>adjusted</td>
<td>2.070</td>
<td>NS</td>
<td>2.029</td>
<td>NS</td>
<td>(0.767-0.920)</td>
</tr>
<tr>
<td></td>
<td>(1.290-3.323)</td>
<td>NS</td>
<td>(1.355-3.037)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>unadjusted</td>
<td>1.932</td>
<td>NS</td>
<td>1.868</td>
<td>1.832</td>
<td>NS</td>
</tr>
<tr>
<td>Chest tightness</td>
<td>1.630</td>
<td>NS</td>
<td>2.309</td>
<td>1.796</td>
<td>NS</td>
</tr>
<tr>
<td>adjusted</td>
<td>1.155-2.301</td>
<td>(1.677-3.178)</td>
<td>1.718</td>
<td>2.229</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.281</td>
<td>NS</td>
<td>(1.356-2.176)</td>
<td>(1.727-2.876)</td>
<td></td>
</tr>
<tr>
<td>Shortness of breath</td>
<td>2.405</td>
<td>NS</td>
<td>1.924</td>
<td>1.884</td>
<td>NS</td>
</tr>
<tr>
<td>adjusted</td>
<td>1.688-3.428</td>
<td>NS</td>
<td>(1.365-2.712)</td>
<td>(1.295-2.739)</td>
<td></td>
</tr>
<tr>
<td>unadjusted</td>
<td>1.845</td>
<td>NS</td>
<td>1.546</td>
<td>2.009</td>
<td>0.854</td>
</tr>
<tr>
<td>Coughing</td>
<td>1.522-2.236</td>
<td>NS</td>
<td>(1.315-1.817)</td>
<td>(1.626-2.483)</td>
<td>(0.807-0.905)</td>
</tr>
<tr>
<td>adjusted</td>
<td>1.836</td>
<td>NS</td>
<td>1.703</td>
<td>2.135</td>
<td>0.834</td>
</tr>
<tr>
<td></td>
<td>(1.341-2.162)</td>
<td>NS</td>
<td>(1.594-2.858)</td>
<td>2.025</td>
<td></td>
</tr>
<tr>
<td>unadjusted</td>
<td>2.113</td>
<td>NS</td>
<td>1.543</td>
<td>1.541</td>
<td>NS</td>
</tr>
<tr>
<td>Wheezing</td>
<td>2.396</td>
<td>NS</td>
<td>(1.206-1.974)</td>
<td>(1.548-2.649)</td>
<td></td>
</tr>
<tr>
<td>adjusted</td>
<td>1.625-2.747</td>
<td>NS</td>
<td>NS</td>
<td>1.750</td>
<td>0.860</td>
</tr>
<tr>
<td></td>
<td>(1.663-3.450)</td>
<td>NS</td>
<td>(1.184-2.585)</td>
<td>(1.847-2.976)</td>
<td></td>
</tr>
<tr>
<td>unadjusted</td>
<td>1.654</td>
<td>NS</td>
<td>NS</td>
<td>1.542</td>
<td>NS</td>
</tr>
<tr>
<td>Asthma</td>
<td>1.307-2.093</td>
<td>NS</td>
<td>NS</td>
<td>1.185-2.005</td>
<td>NS</td>
</tr>
<tr>
<td>adjusted</td>
<td>1.930</td>
<td>NS</td>
<td>NS</td>
<td>1.201-2.442</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>(1.398-2.664)</td>
<td>NS</td>
<td>NS</td>
<td>(1.067-2.217)</td>
<td></td>
</tr>
<tr>
<td>unadjusted</td>
<td>1.939</td>
<td>NS</td>
<td>NS</td>
<td>2.109</td>
<td>NS</td>
</tr>
<tr>
<td>Dust allergies</td>
<td>2.009</td>
<td>NS</td>
<td>1.415</td>
<td>2.289</td>
<td>NS</td>
</tr>
<tr>
<td>adjusted</td>
<td>1.446-2.600</td>
<td>NS</td>
<td>(1.012-1.977)</td>
<td>(1.522-3.442)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.326-3.044</td>
<td>NS</td>
<td>NS</td>
<td>1.541</td>
<td>NS</td>
</tr>
<tr>
<td>unadjusted</td>
<td>1.528</td>
<td>NS</td>
<td>1.018-2.333</td>
<td>NS</td>
<td>1.538</td>
</tr>
<tr>
<td>Mold allergies</td>
<td>1.013-2.303</td>
<td>NS</td>
<td>NS</td>
<td>(1.067-2.217)</td>
<td></td>
</tr>
</tbody>
</table>

**Table 6.1.** Odds ratios (95% confidence intervals) for statistically significant influences on respiratory health responses (NS - not statistically significant; CHEMIOD - chemical odors; SMOKEOD - tobacco smoke; OTHEROD - other odors; HUMID - perception of high humidity; RH - relative humidity)
fatigue, feeling heavy-headed, headache, nausea/dizziness, and difficulty concentrating; (2) mucous membrane symptoms including itching/burning/irritated eyes, irritated/stuffy/runny nose, hoarse/dry throat, and cough; (3) skin symptoms including dry facial skin, flushed facial skin, and itching/stinging/tight/burning sensation in facial skin (Sundell and Lindvall, 1993). However, contrary to this study which found no associations of measured relative humidity to SBS symptoms, BASE multivariate logistic regression analysis indicated negative association between RH and some of the reported respiratory symptoms. Significantly, such associations occurred only when a 10% change in RH was considered rather than a unit percentage change, as customary in initial logistic regression analyses. The important new findings are also established associations between the examined respiratory symptoms and occupants’ sensations of odors.

In addition to the perceived air quality indicators, several other measured parameters were also used in the multivariate logistic regression analysis, as described in Chapter 5. However, most of these variables were not statistically significant for occurrence of respiratory responses. Again, this is no surprise as instruments used for indoor air quality measurements are typically less sensitive than human nose (Schreiber, 2000). Nevertheless, in several cases where relative humidity measurements appeared as statistically significant and negatively associated with certain respiratory response, occupants’ perception of humidity showed strong positive association. In the biggest discrepancy of this sort, occupants sensing the air as too humid would be 2.7 times more likely to experience sinus congestion than those not sensing high humidity; while relative humidity measurements indicate that an increase in relative humidity by 10% would reduce occurrence of sinus congestion approximately 1.2 times. Similar discrepancies exist between measured and perceived humidity in cases of other respiratory responses.

Such findings confirm the results of other studies indicating that occupants are not able to sense air humidity. Thus, an experiment showed no change in perception of humidity when relative humidity was reduced from 70% to 10% and then back to 70% in controlled laboratory conditions (Andersen et al., 1973). Although the temperature was held constant during the experiment, occupants reported significant changes in sensation of temperature caused by the actual change in humidity. Another study also found no statistically significant associations between measured
humidity and sensation of dryness (Sundell and Lindvall, 1993). Nevertheless, the investigated perception of humidity in BASE buildings is significantly associated with almost all examined respiratory symptoms and presents an important health indicator.

A question what do occupants actually sense when they report humid environment remains unclear, although certain relations may bring additional insight. Namely, the calculated ORs between perceived humidity and unpleasant chemical odors show close agreement, while influence of humidity on human sense of smell has been documented (Schreiber, 2000). Thus, people may not be able to sense varying humidity levels, but only presence of unpleasant odors, which impact their perception of humidity. However, to confirm this hypothesis further research would be necessary.

Limitations of the presented findings are related to intrinsic inconsistencies between the measurements and occupants’ responses. Namely, the occupants answered a question about feeling the indoor air as too humid within a month long period prior to the study week when measurements were taken. Thus, the relative humidity measurements (max measured 63%) might not be representative of the actual building conditions to which the occupants referred answering the questionnaire.

Apart from these inconsistencies, possible limitations of the analyses are related to the insufficient data available in BASE database, which might have prevented establishment of associations between certain parameters or reduced their statistical significance. Namely, several associations were excluded from Table 6.1 as their p-value was slightly above 0.05. For example, this was the case with a number of smoke odor associations to respiratory symptoms, as a result of the fact that most buildings in BASE study were smoke free environments. Additionally, sensitivity limitations of certain measurements in the database may have adversely affected ability of the applied statistical techniques to provide greater resolution of confidence intervals in statistically significant cases. Unfortunately, such limitations are related to the capabilities of used measuring equipment and could not be overcome in the analyses.
6.2 Neural Network Data Analysis

The BASE data analysis using neural networks adopted a phased approach, gradually increasing in complexity.

Phase 1: use probabilities for occurrence of specific respiratory symptoms as outputs, and recorded values of parameters as inputs; vary the number of inputs between 2 and 7 to determine optimal prediction performance.

In this phase the final selection of inputs was determined, while BSI had not yet been considered as an output parameter. Considering that the list of output variables included dust and mold allergies, attempts were made to include measured concentrations of allergens in the list of inputs. However, allergen concentrations were not measured in all BASE buildings and including allergen inputs would require complete exclusion from the training dataset of all buildings not having the allergen measurements. This would significantly reduce the size of available NN training dataset and limit overall NN prediction capabilities dependent upon sufficient number of training cases. Similar limitations impacted inclusion of measured aldehyde (acetaldehyde and formaldehyde) concentrations into the input dataset, while TVOC concentrations excluding aldehydes had been part of the inputs. However, although VOC concentrations were measured in all BASE buildings, during the 10 year long BASE study period there had been a significant change in measurement protocol, affecting the type of measured VOCs. The solution to this problem was either to limit the number of VOCs contributing to TVOC concentrations, such that it was consistent for all BASE buildings, or to exclude from the training dataset a portion of building data containing VOC measurements inconsistent with the majority of BASE buildings. The second option was selected, as the BASE change of protocol happened relatively early into the study, after only 13 buildings had been investigated. On the other hand, choosing to reduce VOC set contributing to TVOC concentrations would require exclusion of 19 VOCs, 6 of which were measured in more than half of the BASE buildings. Finally, the list of NN inputs consisted of relative humidity and occupant perceptions which were statistically significant for occurrence of all considered output parameters in multivariate logistic regression, with addition of temperature (related to the sense of high humidity), CO$_2$ (indicator of ventilation rates found significant in liter-
ature review) and TVOC concentrations, as previously discussed. Due to such input selection and exclusion of certain buildings, the number of BASE buildings used in NN analyses was reduced to 87. Thus, using 7 inputs and one output at a time with 90% of 87 available buildings for NN training and 10% for testing also complied to recommendations mentioned in Chapter 5 that number of NN training cases should equal at least 10 times the product of numbers of input and output parameters.

Phase 2: analyze input interdependence using scatter plots

NN input parameters’ interdependence was analyzed using scatter plots which showed input values for every possible pair of input parameters, grouped with respect to values of each considered output. Consequently, eleven matrices of 7 by 7 scatter graphs were analyzed to discover the underlying structures that may exist between variables. For example, elliptical shapes in scatter plot graphs would indicate linearity and normality of data, while L-shaped plots would point towards curvilinear relationships, which could be transformed to linear normally distributed values applying logarithmic transformations (Mertler and Vannatta, 2005). Such transformations may be desirable to reduce interdependency of input data and ensure better performance of NN output predictions using independent input parameters.

Phase 3: normalize input and output data

Using measured parameters directly as inputs for NN analyses did not yield expected prediction accuracy, in spite of conducted scatter plot analyses to ensure independence of input parameters. The NN predictions were no better than a constant output function equal to an average value of output training data. Therefore, normalization of inputs was essential for NN performance. A built-in MATLAB normalization function, `prestd`, was used. This function transformed the inputs and outputs such that the normalized data would have mean value equal to 0 and standard deviation equal to 1, according to Equation 6.1.

\[
prestd(p) = \frac{p - \bar{p}}{\sigma(p)} \tag{6.1}
\]
Here $\mathbf{p}$ is the vector of inputs or outputs, while $\bar{p}$ and $\sigma(p)$ are the average values and standard deviation of its components, respectively.

In addition to normalization, an attempt was made to apply principle component analysis (PCA) to the input data, using standard MATLAB function \textit{prePCA}. The PCA ensures that original inputs are transformed into a matrix of principle components which are mutually uncorrelated. Therefore, possible interdependence of inputs is eliminated and number of input parameters can be reduced to include only those inputs that significantly contribute to variability of the output dataset (Jolliffe, 1986). The level of desired input significance can be adjusted within the PCA algorithm. However, applying the MATLAB PCA function to 7 previously identified BASE inputs resulted in no significant improvement in prediction performance. Furthermore, reduction in the number of input parameters to 6 principle components only occurred when level of significance was increased to 4\% at a single percentage rate. This would mean that reducing the number of inputs to 6 principle components out of 7 previously selected parameters would account for 96\% of variance of the output data. As PCA did not improve NN prediction accuracy, it was not included in further analyses. Nevertheless, the developed NN prediction algorithm included provisions to add PCA in future studies when larger number of input parameters would be considered.

Phase 4: exclude part of the building data when occupants did not report improvement in their respiratory symptoms outside of the BASE building

Phase 4 in the development of NN analyses was necessary to ensure that observed occupants’ health symptoms were indeed consequence of BASE building indoor environments. Therefore, all cases where occupants did not report improvement in their respiratory health conditions outside of the BASE building environment were excluded from the analyzed dataset. Confidence of such analyses was documented in literature, investigating the ability of occupants to determine building related health symptoms through comparison of their occurrence outside vs. inside of the investigated buildings (Burge et al., 1991). Importantly, BASE building related impact on occurrence of asthma, dust and mold allergies, could not be identified as there were no information available in the dataset concerning the occurrence of such symptoms among occupants inside vs. outside of the BASE
buildings. Rather, asthma and allergy records only indicated presence of diagnosed medical conditions and could not be exclusively linked to a specific BASE office building indoor environment.

Phase 5: determine optimal NN training algorithm

As described in Chapters 3 and 5 NNs were trained using conjugate gradient backpropagation algorithm. The algorithm tries to find the minimum value of the NN performance function using a line search routine in the conjugate gradient direction of the performance function. The algorithm starts by searching the minimum in the steepest descent direction of the performance function and continues in a sequence of conjugate gradient directions computed by combining the current direction with a current gradient of the performance function. Different conjugate gradient algorithms use different ways to compute this combination of search directions and choice between them depends upon particular NN training datasets (Demuth and Beale, 1998). Attempts were made to determine which of the four available MATLAB conjugate gradient algorithms in combination with which of the five available line search routines, given in Table 3.2, would result in optimal NN performance on BASE data. During the training, main control over the number of repeated training iterations laid upon the specified maximum number of training epochs and could be stopped earlier if minimum gradient of the performance function or minimum step size was achieved. In fact, it had been noticed that reaching the minimum training step size would usually prevent further training. Therefore, MATLAB line search routines were tested with respect to their ability to handle smaller step sizes, which would be essential for achieving value of the performance function closer to its global minimum. Using Charalambous' search (SRCHCHA) and Backtracking (SRCHBAC) resulted in very poor training performance, while Brent's (SRCHBRE) and Hybrid Bisection-Cubic search (SRCHHYB) produced non-monotonously decreasing performance error, indicating instability. Therefore, Golden Section search (SRCHGOL) routine was selected and in combination with Powell-Beale (TRAINCGB) version of the conjugate gradient algorithm gave the best NN training performance on the BASE data. It had also been noticed that reducing the initial step size of the search routine within the interval location step below 0.01 would only increase the time necessary to train the network, but not
the prediction accuracy.

**Phase 6: use regularization to improve NN generalization; include prediction confidence intervals**

In phase 6, further steps were undertaken to improve NN generalization performance on the test data as described in Chapter 5. The implemented regularization not only improved NN generalization, but also significantly reduced the required number of NN training epochs, shortening the training time to achieve satisfactory NN performance. Furthermore, both logsig and tansig transfer functions described in Table 3.1 were tested with regularization to determine optimal performance. The hyperbolic tangent sigmoid function had better performance in all cases using regularization with normalized NN input and output parameters.

Additionally, confidence intervals were calculated and included in overall NN prediction graphs. As NN predictions deal with health symptoms of building occupants, it is necessary to evaluate reliability of the averaged occupant responses from only part of the building occupants filling-in the BASE survey, rather than having the data from all the occupants working in each particular BASE building. Confidence intervals take into account the number of occupants responding to the BASE survey, and based on inferential statistics, estimate with certain confidence level the range of possible values for an averaged response of all the occupants. Thus, a 95% confidence interval indicates a range where the actual averaged response from all building occupants will fall with a 95% confidence. Such intervals were calculated for all buildings and target prediction variables $\bar{b}_{art_i}$, following recommendations from the literature and applying Equations 6.2 and 6.3 (Schunn and Wallach, 2008).

$$95\% CI(\bar{t}_i) = (\bar{t}_i - SE(\bar{t}_i) \cdot 1.96, \bar{t}_i + SE(\bar{t}_i) \cdot 1.96) \quad (6.2)$$

$$SE(\bar{t}_i) = \frac{1}{\sqrt{n}} \cdot \sqrt{\frac{1}{n-1} \sum (t_{j,i} - \bar{t}_i)^2} \quad (6.3)$$

Here $\bar{t}_i$ is the probability that an occupant responding to BASE questionnaire in building $i$ expressed presence of a certain health condition or symptom (as previously discussed), $SE(\bar{t}_i)$ is the standard error of such probability, $n$ is the
total number of occupants responding to the questionnaire, while $t_{j,i}$ are responses from individual occupants. Value of the multiplier, 1.96, is used in inferential statistics as a cutoff value for 95% CI when sample size is greater than 30 and assuming the data is normally distributed.

Phase 7: exclude naturally ventilated buildings from the NN analyses

The last phase in NN analyses was aimed at improving the prediction accuracy by excluding naturally ventilated buildings from the dataset, as their ventilation characteristics are physically different from other BASE buildings. Since air is the principle medium impacting occurrence of respiratory symptoms among building occupants, type of ventilation system affects the occurrence of respiratory symptoms (Burge, 2004). Consequently, the final dataset for NN analyses consisted of 85 mechanically ventilated buildings, excluding 2 additional buildings from the dataset formed in phase 1 of the analysis.

6.2.1 Benchmark Predictions

The NN prediction capabilities had been evaluated through multiple runs in aforementioned analyses phases 1 through 7. Although final evaluation of NN prediction performance was based on results obtained in phase 7, it was important to consider prediction capabilities throughout all development stages. NNs predicted all 11 respiratory output symptoms listed in Table 5.3. However, only detailed reports on cough, as one example of NN predictions, will be explained together with a comparison of NN prediction accuracy for occurrence of other symptoms. Detailed graphs presenting the analyses of NN predictions for all symptoms are given within Appendix A.

Following the procedure described in Phases 1 through 7, Figure 6.1 presents scatter plots of NN input data grouped with respect to probability for occurrence of cough among BASE building occupants. Neurons 1 through 7 indicate ranges of preprocessed normalized input parameters as follows:

neuron 1: average measured temperature,
neuron 2: average measured relative humidity,
neuron 3: average measured CO$_2$ concentration,
neuron 4: probability of occupants sensing unpleasant chemical odors,
neuron 5: probability of occupants sensing other odors,
neuron 6: probability of occupants sensing high humidity,
neuron 7: average measured TVOC concentrations,

Percent ranges 0-10%, 10-20% and 20-30% indicate clusters of probability that occupants had building related cough at least once within the 4 weeks period prior to the study week.

Figure 6.1. Neural network input scatter plots categorized with respect to probability for occurrence of cough

Based on the scatter plots presented in Figure 6.1, no consistent interdependence existed between the NN input parameters. Such relationship holds for all
other NN output prediction variables. In cases of predicting occurrence of symptoms other than cough, scatter plots only differ with respect to percent range clusters classifying the input data, while general shape of the scattered graphs and NN input interdependence remain unchanged.

Using the 7 described inputs, 3 MLP NNs with a single hidden layer were trained for any number between 4 and 8 of the hidden layer neurons, following the procedure described in Chapter 3. The network having the highest accuracy on the testing dataset was selected as optimal. For example, in the case of cough predictions second of the three NNs with 5 hidden neurons was the most accurate. Figure 6.2 presents MSE reduction during the training of such a network. In the figure, final MSE is given as performance, approximately equal to 0.37, while the goal is error reduction to 0.0001. Such a goal was obtained from the desired prediction accuracy of 1% (0.01) of the output probability. As the training MSE performance function squared the error, the desired performance goal became as specified. Although the goal was not met, training was stopped after 150 iterations or epochs, due to fulfillment of regularization stopping criteria. It is important to notice that training initialization and choice of training vs. testing cases depend upon random parameters. Therefore, optimal size and NN performance also depend upon random parameters and repeated training may not produce the same results. Nevertheless, the developed algorithm ensures selection of the most accurate network within the given limitations.

Figures 6.3 and 6.4 present training and testing performance of the optimal NN, respectively. Sample count in these figures corresponded to internal building coding - a unique value between 1 and 100 assigned to each BASE building and kept constant during the entire study. Performance value between 0 and 1, presented on the ordinate axis, indicated the probability for building related cough occurrence among occupants.

Based on the performance data presented in Figures 6.3 and 6.4, Figure 6.5 presents absolute NN training and testing prediction errors. These errors are calculated for each building as an absolute difference between predicted and actual probability for occurrence of building related cough and given on a scale of 0 to 100%.

Figures 6.6, 6.7 and 6.8 present statistical evaluations of training, testing and
Figure 6.2. Neural network training error reduction in cough predictions

Figure 6.3. Neural network training performance for cough predictions
Figure 6.4. Neural network testing performance for cough predictions

Figure 6.5. Neural network training and testing absolute errors in cough predictions
combined training plus testing capabilities for prediction of cough, respectively. The axes on the figures mark actual (T) and predicted (A) probabilities for occurrence of building related cough, while the equations above the graphs represent best linear regression fit to the given data points. The shown R-values evaluate how well the best linear regression line fits the data. Ideally, coefficient next to the variable of actual values (T) in the best linear fit equation, i.e. slope of the regression line, as well as the R-value would be equal to 1, while the free coefficient in the equation would equal 0. Such situation would represent a perfect prediction. However, in the case of no perfect prediction, it is important to see how well the data follow the trend of perfect prediction (T coefficient) and how dispersed they are around such a trend (R-value). Following the recommendations from Chapter 3 R-value should be larger than 0.7, while positive coefficient next to the variable of actual values (T) in the best linear fit equation, i.e. positive slope of the regression line, would indicate a trend of increase in the predicted values following the increase of the actual values. This was achieved in training and combined training/testing evaluation, while testing evaluation alone showed lower tendency to follow perfect predictions and greater dispersion of predicted probabilities around the actual values. Reasons for such testing evaluations were relatively small number of testing cases and consequently large influence each individual testing case prediction accuracy had on overall testing performance.

In addition to the statistical evaluation Figures 6.6 through 6.8, Figure 6.9 presents predicted output probabilities for occurrence of building related cough with 95% confidence intervals (CI). The confidence intervals take into account the fact that NN target probabilities were inferred based on a limited number of occupant responses, rather than all building occupants. Thus, each CI, marked with upper and lower bars in Figure 6.9, presents a 95% probable range of values for the actual probability of cough occurrence, calculated based on all building occupants’ responses, rather than just those which participated in the BASE survey. Based on the results from Figure 6.9, only three training and one testing case buildings fell outside of the 95% confidence ranges. Such results indicate that for a majority of buildings NN prediction accuracy for occurrence of cough is satisfactory, compared to the uncertainty contained in the actual data which originates from the fact that not all building occupants participated in the survey.
To compare NN prediction capabilities for all considered output parameters given in Table 5.3 average absolute prediction errors were calculated and compared with average output values. Table 6.2 presents these values with respect to implemented phased NN design adjustments. Thus, rows Pr1, TesEr1 and TrnEr1 present probabilities for occurrence of specific symptoms, average absolute testing and training errors of NN predictions obtained after the described NN development Phase 3. The next three rows indexed with the number 2 contain results obtained after implementation of Phase 5 in NN development. Finally, the third and fourth set of rows in Table 6.2 present probabilities for symptom occurrence and NN prediction errors after implementation of the final NN development Phases.
Figure 6.7. Neural network actual vs. predicted testing probabilities for occurrence of cough

6 and 7, respectively. Note that probabilities Pr2 occur twice in the Table 6.2, as NN adjustments between Phases 5 and 6 have not resulted in changing either the size nor the contents of training/testing datasets, as opposed to other presented stages in NN development. Also notice that all symptom columns in Table 6.2 contain data throughout NN developmental phases, except BSI column. Reason for omitting BSI prediction results is introduction of BSI analyses only after Phase 5 of NN development.
Figure 6.8. Neural network actual vs. predicted training and testing probabilities for occurrence of cough.
Figure 6.9. Confidence ratios for neural network predictions of cough
<table>
<thead>
<tr>
<th></th>
<th>Asthma</th>
<th>Sneez.</th>
<th>Sinus</th>
<th>Sore</th>
<th>Chest</th>
<th>Sh.br.</th>
<th>Cough</th>
<th>Wheez.</th>
<th>Dust</th>
<th>Mold</th>
<th>BSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pr1[%]</td>
<td>12.29</td>
<td>52.11</td>
<td>63.07</td>
<td>43.80</td>
<td>12.24</td>
<td>12.04</td>
<td>35.31</td>
<td>10.70</td>
<td>32.48</td>
<td>25.54</td>
<td>-</td>
</tr>
<tr>
<td>TesEr1</td>
<td>4.25</td>
<td>4.36</td>
<td>5.95</td>
<td>5.37</td>
<td>4.43</td>
<td>7.25</td>
<td>6.91</td>
<td>4.15</td>
<td>7.38</td>
<td>5.41</td>
<td>-</td>
</tr>
<tr>
<td>TrnEr1</td>
<td>1.26</td>
<td>2.67</td>
<td>2.87</td>
<td>4.08</td>
<td>1.58</td>
<td>1.08</td>
<td>2.14</td>
<td>1.76</td>
<td>4.03</td>
<td>2.95</td>
<td>-</td>
</tr>
<tr>
<td>Pr2[%]</td>
<td>12.29</td>
<td>21.02</td>
<td>20.78</td>
<td>16.21</td>
<td>4.86</td>
<td>4.43</td>
<td>11.96</td>
<td>3.86</td>
<td>32.48</td>
<td>25.54</td>
<td>-</td>
</tr>
<tr>
<td>TesEr2</td>
<td>4.84</td>
<td>5.49</td>
<td>6.51</td>
<td>5.01</td>
<td>5.27</td>
<td>3.65</td>
<td>3.51</td>
<td>2.64</td>
<td>12.42</td>
<td>7.79</td>
<td>-</td>
</tr>
<tr>
<td>TrnEr2</td>
<td>1.66</td>
<td>4.75</td>
<td>3.25</td>
<td>4.32</td>
<td>1.16</td>
<td>1.50</td>
<td>2.28</td>
<td>1.20</td>
<td>2.86</td>
<td>2.92</td>
<td>-</td>
</tr>
<tr>
<td>Pr2[%]</td>
<td>12.29</td>
<td>21.02</td>
<td>20.78</td>
<td>16.21</td>
<td>4.86</td>
<td>4.43</td>
<td>11.96</td>
<td>3.86</td>
<td>32.48</td>
<td>25.54</td>
<td>1.69</td>
</tr>
<tr>
<td>TesEr3</td>
<td>4.58</td>
<td>5.72</td>
<td>6.90</td>
<td>3.81</td>
<td>2.30</td>
<td>2.93</td>
<td>3.70</td>
<td>1.84</td>
<td>10.12</td>
<td>5.35</td>
<td>0.60</td>
</tr>
<tr>
<td>TrnEr3</td>
<td>1.81</td>
<td>4.18</td>
<td>4.87</td>
<td>4.03</td>
<td>1.99</td>
<td>2.18</td>
<td>2.47</td>
<td>1.83</td>
<td>4.24</td>
<td>4.36</td>
<td>0.15</td>
</tr>
<tr>
<td>Pr4[%]</td>
<td>12.04</td>
<td>21.03</td>
<td>20.77</td>
<td>16.36</td>
<td>4.88</td>
<td>4.36</td>
<td>11.98</td>
<td>3.90</td>
<td>32.56</td>
<td>25.61</td>
<td>1.69</td>
</tr>
<tr>
<td>TesEr4</td>
<td>5.72</td>
<td>5.85</td>
<td>7.06</td>
<td>5.06</td>
<td>1.99</td>
<td>2.26</td>
<td>3.64</td>
<td>2.03</td>
<td>7.62</td>
<td>6.17</td>
<td>0.31</td>
</tr>
<tr>
<td>TrnEr4</td>
<td>1.91</td>
<td>2.79</td>
<td>3.44</td>
<td>3.61</td>
<td>2.19</td>
<td>1.85</td>
<td>2.63</td>
<td>2.09</td>
<td>4.21</td>
<td>2.91</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Table 6.2. Average NN target probabilities and absolute prediction errors for each considered output parameter throughout development phases (Pr - probabilities for occurrence of actual symptoms throughout NN development phases; TesEr - average absolute testing error; TrnEr - average absolute training error)
6.2.2 Discussion on Neural Network Data Analyses

Based on the presented NN results, the current study shows promising capabilities of NNs to infer probabilities for occurrence of occupants’ respiratory symptoms and predict BSI. NNs showed good performance in most of the predicted building-related health symptoms, while occurrence predictions of asthma and allergies were no better than statistical averaging. Such results are justified by the fact that the available BASE asthma and allergy records only indicated presence or absence of diagnosed medical conditions and significantly differed from other predicted occupant symptoms, as no distinction in asthmatic or allergenic symptom occurrence was recorded between indoor and outdoor environments. Consequently, occurrence of asthma or allergies could not be exclusively linked to specific BASE office building indoor environments, as in the case of other investigated occupants’ respiratory symptoms and BSI.

In the presented analyses several adjustments were made in NN implementation and improved prediction performance was achieved. Such adjustments included preprocessing the data and changing the NN design parameters. These measures generally improved NN performance, but some had greater impact than others. For example, exclusion of the naturally ventilated buildings from the analyzed dataset did not significantly affect the overall NN results, while data preprocessing was essential for successful NN application. The reason for insignificance of natural ventilation may be small relevance of only 3 naturally ventilated buildings present in the 100 building large BASE dataset. Particularly, one of the 3 naturally ventilated buildings had already been excluded from the analyses, having VOC measurements inconsistent with the majority of BASE buildings.

Overall, the presented results offer detailed insight into the used methodology and evaluation techniques, illustrated on an example of predicting building related cough. Similar analyses were applied to other respiratory health symptoms and absolute prediction errors were presented to compare NN predictive capabilities. Based on such comparison throughout phases of NN development, NNs were useful in most prediction scenarios. However, using NNs may not be beneficial in cases when average absolute NN prediction error is consistently too large compared to an average outcome, regardless whether the outcome is probability for occurrence of specific symptoms or a BSI value. To provide recommendation for prospective
application of NNs in evaluation of respiratory health impacts of office building environments, a criterion based on comparison of NN testing errors with actual probabilities may be used. For example, averaged absolute testing error of predicted asthma was consistently above one third of the averaged probability for asthma occurrence. Thus, taking into consideration results from the final NN developmental phase and assuming average asthma prediction with an average error, probability for asthma occurrence would be 12.04%±5.72%, i.e. between 6.32% and 17.76%. On the other hand, applying this “1/3 error approach” to determine applicability of NNs in prediction of other investigated output parameters, the results indicate that NNs can reliably predict cough, sinus congestion, sneezing, sore throat, and BSI, while predictions of probabilities for occurrence of chest tightness, shortness of breath and wheezing were less reliable, regardless of the composition of training/testing data sets.

Consequently, usage of NNs with a current set of seven inputs is not recommended for predicting occurrences of asthma, allergies and lower respiratory symptoms, except cough. Generally, all symptoms with low average probabilities appear more difficult to predict using NNs than symptoms with high probabilities, noticing a borderline probability for successful predictions of around 12%. To increase applicability of NNs in predictions of lower respiratory symptoms, sufficient number of adequate building training cases would be necessary to avoid low symptom occurrence frequencies, resulting in low probabilities.

Importantly, the “1/3 error approach” should not be strictly applied, but rather used as a guideline, as it was observed that NN prediction accuracy highly depended upon a particular selection of training vs. testing building data. Namely, the same NN procedure would result in different prediction accuracies for the same output parameter depending upon the selection of buildings used in training and testing datasets. However, this is a complex phenomenon which does not solely depend on the building data selection, but also on the size of training vs. testing datasets as well as the number of other NN parameters, such as the size of NN hidden layer(s), randomized initialization, type of the used transfer and training functions, and ultimately type of the applied NNs. In order to evaluate significance of selecting particular buildings as part of the training or testing datasets systematic control over all these variables would be necessary. However, such inves-
tigation was not the topic of the current investigation. Rather, the intention was to show how developed NN analyses methodology could be used to provide successful prediction of respiratory health responses among building occupants based on the collected BASE data. Further efforts invested in selecting the most suitable buildings for NN training could result in identification of the representative case building profiles for improved prediction accuracy of respiratory health symptoms among building occupants. Eventually, such representative buildings may bring new insights into good and poor building practices impacting the occurrence of health symptoms.
Chapter 7

Experimental Setup and Validation

In order to validate the developed approach for predicting respiratory health symptom occurrence in office building environments based on the BASE benchmark data, additional real building investigation study was conducted. Although the validation building was not part of the original BASE dataset, experimental data collection methodology followed BASE study protocol as much as possible (U.S. EPA, 2003). The purpose was to show how BASE data collection procedure can be applied in building types which were not part of the benchmark dataset, such as emerging “green” constructions. The collected data would also extend the knowledge about a “green” building environment in comparison to an average US office building within the BASE dataset.

7.1 Experimental Setup Location

For conducting the experimental validation a two-story high “green” building was selected, hosting the offices of Pennsylvania Department of Environmental Protection in Ebensburg (PA DEP), Cambria County, PA. The building was Gold certified as part of the Leadership in Energy and Environmental Design (LEED) program by the United States Green Building Council (USGBC) (U.S. GBC, 2009).

Two previous studies also conducted investigations of environmental conditions in this building in order to evaluate its innovative design and performance. Thus, LBNL conducted a study of the building’s underfloor air distribution (UFAD) system and found 13% better than expected removal efficiency for CO\textsubscript{2}, indicator-
ing reduction in occupants’ exposure to pollutants. At the same time, building occupants were more satisfied with thermal conditions compared to an average building, and despite lower than expected air temperature stratification due to high ventilation rates, evidence was presented to indicate UFAD energy saving benefits (Fisk et al., 2004). Interestingly, such UFAD performance appears as an exception to typical commercial installations and experiences of HVAC consultants (Mumma, 2009). The second study in PA DEP conducted a post occupancy energy consumption measurements and evaluation of building systems over a three year period (Deru et al., 2005). The study used calibrated energy simulations for as-built building and found total annual energy consumption savings of 40% compared to a baseline building (ASHRAE, 2001). Most of the savings originated from reductions of heating and lighting energy. Based on the building energy performance monitoring, the study recommended improvements in building operation and documented measures which could be applied in future low-energy building construction.

Following these efforts, the current study aimed at evaluating the office building’s indoor environmental quality and impact on occupants’ health, particularly related to respiratory symptoms. For this purpose, modified BASE study protocol was applied consisting of three parts: (1) indoor and outdoor environmental quality measurements, (2) survey of occupants’ health symptoms and perceptions of indoor environment, and (3) general building characterization.

The experimental study was conducted between the 25th and 29th of February 2008. Measurements of indoor environmental quality were conducted at the second floor of the building. The three fixed indoor measurement locations in the South-West, South-East and North-East parts of the building were selected in accordance with the BASE study protocol and internally coded as F1 SW, F3 SE and F5 NE, respectively. The floor plan in Figure 7.1 presents positions of the indoor measurement locations together with detailed surroundings of the North-East location.

Close to each of the three indoor measurement locations, internal building monitoring system (BMS) sensors recorded temperatures, relative humidity and \( \text{CO}_2 \) concentrations which were used as a quality assurance measure to confirm accuracy of collected data. A single outdoor monitoring location was selected
Figure 7.1. PA DEP building floor plan with marked indoor monitoring locations (left) and magnified North-East monitoring location (right) (modified from (Deru et al., 2005)) close to the outdoor air intake, as presented in Figure 7.2. The figure also shows two of the three indoor monitoring locations with installed measuring equipment.

Figure 7.2. Installed indoor (left, middle) and outdoor (right) measurement equipment

Additional measurements included building supply air conditions measured in the two mechanical rooms located at the first floor of the building. Parallel with
the measurements, building occupants were invited by the building management to fill-in a comprehensive survey related to the indoor environmental conditions and occupants’ perceptions of indoor environments. Additionally, the investigators collected information about general building characteristics based on their personal observations and inputs received from the building management. The measurements, survey and building data were collected following as much as possible the EPA BASE study protocol (U.S. EPA, 2003). However, unlike other BASE investigated buildings, the intention was to examine a Green LEED certified building, for which purpose certain procedures had to be amended. Additional limitations were related to availability of the measurement equipment and man power. A comprehensive planning and schedule, description of used instrumentation and practical recommendations for similar studies will be further presented.

7.2 Field Planning and Schedule

In order to conduct a study involving human participants it was essential to obtain approval from the Institutional Review Board (IRB) of the Office of Research Protection (ORP), The Pennsylvania State University (PSU ORP, 2009). ORP required detailed study documentation and procedures to be submitted before starting the study. Thus, all study documents and materials were prepared together with a systematic plan of study activities. The documentation included:

- Procedural documentation with all handouts, e-mail correspondence and public announcements for the building occupants,
- Printout of an Internet based questionnaire for building occupants, developed after paper based BASE study questionnaire (given in Appendix D),
- Implied consent form to precede Internet questionnaire, informing occupants about the study and their voluntary participation,
- Signed approval form from the PA DEP building management expressing their willingness to participate in the study,
- Agreement with the EPA to use BASE study results in comparative analyses with the current experimental study.
All documentation was compiled together with an application to IRB and after receiving the approval, study was conducted according to the planned field operations schedule.

7.2.1 Field Operations Schedule

The field operations schedule included planned data collection activities at the field measurements site, consisting of 3 fixed indoor and 1 outdoor measurement locations. Additionally, CO\textsubscript{2} measurements were conducted in the two mechanical rooms. All field operations were divided among the days of the study week and assigned to some of the five team members. Each member’s assignment was indicated at the end of the specific activity description, according to the following coding key:

TL - team leader;
T1 - team member 1;
T2 - team member 2;
T3 - team member 3;
T4 - team member 4;
AT - all team members.

The field team was expected to spend up to 11 hours at the field site during the equipment setup and monitoring days (Monday through Thursday), and conduct up to 9 hour long daily monitoring and data collection activities. However, individual tasks actually required less engagement. Important note was given to the team members to ensure that all measuring equipment, pumps and sensors had a warm-up time of at least 5 minutes.

Additional activities were performed when there was unallocated, “free” time, according to the following list:

- Measure CO concentrations using AQ5000, half a day at one of the fixed sites, and half a day at the outdoor site. T4
- Perform CO\textsubscript{2} mobile measurements using AQ5000, as a measure of quality control. T4
- Ask building manager to fill-in building related information. T3
• Fill-in a list of subjective observations. T1

Daily schedule of field operations was adjusted according to the team members’ availability and included the following activities:

Monday

1. Meet the building manager. T4, TL
2. Provide information leaflets to the occupants. TL
3. Place the recording poles at 3 fixed indoor locations previously identified within 5 m by 5 m (16.5 ft) tiles such that their location is at least 0.5 m (1.6 ft) from corners, windows, walls, partitions and other vertical surfaces such as file cabinets. Poles should be at least 1 m (3 ft) away from printers, copy machines, photocopiers or similar equipment. Poles should not be in the direct sunlight, nor in front of diffusers or fans. If an under floor diffuser is beneath the pole, close it. All measurements except temperature should be performed at the 1.1 m (3.6 ft) height only. Temperature should be measured at the heights of 0.1 m (0.3 ft), 0.6 m (2 ft), 1.1 m (3.6 ft) and 1.7 m (5.6 ft). T4, TL
4. Deploy radon samplers: 3 samplers at 3 fixed monitoring locations, remaining 3 at the possible pathways for radon to come into the test space from lower floors and ground level: stairwells, exits, elevator doors. Samplers will be collected after 96 hours. T4, TL
5. Verify monitoring locations. TL
6. Unpack and prepare instrumentation. T4, TL
7. Assemble instrumentation. T4, TL
8. If there is time, proceed with equipment setup scheduled for Tuesday. T4, TL

Tuesday

AM

1. Measure airflow rates at the room exhaust grilles. T3
2. Continue set up of the 3 fixed sampling locations. Measurements should be averaged over a 5 minute period, meaning that we need more frequent
measurements (e.g. every minute!) if instruments do not provide automatic averaging. Set up loggers to start recording at noon on Tuesday, Feb 26, and end at 5PM on Thursday, Feb 28. Place:

- HOBO loggers at 1.1 m (3.6 ft) height, with additional temperature probe at 1.7 m (5.6 ft) height, TL
- Thermistors at 0.1 m (0.3 ft) and 0.6 m (2 ft) heights, T1, T4
- Sound loggers at 1.1 m (3.6 ft) heights, TL

3. Set up AHU sampling locations for CO₂ measurements. T4
4. Set up outdoor sampling site and start outdoor sampling. TL, T1, T4
5. If there is time, proceed with equipment setup scheduled for Wednesday. TL, T1, T3, T4

PM

1. Begin measurements at the outdoor, and fixed indoor locations as well as at the AHU sampling locations. Outdoor monitoring should start first and end last. TL, T4
2. Measure airflow rates at the AHU room underfloor outlets/diffusers. T3
3. Download data from our loggers and the existing building environmental monitoring system (optional). TL

Wednesday

AM

1. Continue continuous monitoring. TL, T1, T4
2. Set up particle counters at 1 indoor and 1 outdoor location at 1.1 m (3.6 ft) heights. Sampling should be 8 hours. Outdoor monitoring should start first and end last. Additionally set up ultra fine particle counter as a quality assurance measure to confirm readings from the two fixed particle counters. TL
3. Bring VOC SUMMA canisters to 3 fixed indoor and 1 outdoor measurement locations and start sample collection over an 8 hour period, such that sample collection point is at 1.1 m (3.6 ft) height. TL, T1, T2, T4
4. Perform sampling for Formaldehyde and Acetaldehyde using sorbent tubes at 3 indoor and 1 outdoor measurement locations at 1.1 m (3.6 ft) heights. Sampling should last until 15 liters of air passes through the tube. T3
5. If possible, for quality assurance purposes collect mutisorbent VOC sample(s), using RAE PGM-7240 at 1.1 m (3.6 ft) heights, particularly focusing on 1, 2, 4 trimethylbenzene (which accounts for influence of ambient automotive sources), formaldehyde, and acetaldehyde. If possible, also perform measurements of butylated hydroxytoluene, 2-butoxyethanol, 2-ethyl-1-hexanol, Phenol, 4-phenylcyclohexene, 2,2,4-trimethyl-1,3-pentanediol monoisobutyrate, 2,2,4-trimethyl-1,3-pentanediol diisobutyrate. Make sure that these measurements would not interfere with other VOC sample collections, i.e. that no VOCs are released during these measurements!!! T1

10:30-noon Perform bioaerosol sampling during 2 min and 5 min periods at 3 fixed indoor and 1 outdoor locations at 1.1 m (3.6 ft) heights. Flow rate should be 28.3 L/min (1 ft$^3$/min). Repeat the sampling using both TSA and MEA plates. Total number of samples is 4 per location. Additionally make duplicate samples for 1 indoor and outdoor location. Don’t forget field blanks, 1 for each plate. Also 1 of each plate should remain in the refrigerator as shipping blanks. TL, T2

10:30-noon Perform fungal spores sampling during 4 min periods, at 5 L/min (0.18 ft$^3$/min) air flow rate at 3 fixed indoor and 1 outdoor locations. Additionally make duplicate samples for 1 indoor and outdoor location. T4

6. If possible, conduct duct traverse measurements at outdoor air intakes. T3

PM

1. Move ultrafine particle counter at an outdoor location. TL

3:30-4:30 Perform bioaerosol sampling during 2 min and 5 min periods at 3 fixed indoor and 1 outdoor locations at 1.1 m (3.6 ft) heights. Flow rate should be 28.3 L/min (1 ft$^3$/min). Repeat the sampling using both TSA and MEA plates. Total number of samples is 4 per location. Additionally make duplicate samples for 1 indoor and outdoor location. TL, T2

3:30-4:30 Perform fungal spores sampling during 4 min periods at 1.1 m (3.6 ft) heights at 3 fixed indoor and 1 outdoor locations. T4
4:30-5:30 Ship bioaerosol samples for analyses, to be delivered on Thursday morning. Don’t forget to include the 2 field and 2 shipping blanks for bioaerosol samples (total of 52)! Total fungal cartridges=10 TL, T2

2. Download data from data loggers and the existing building environmental monitoring system (optional). T4
3. Ship other collected samples for analyses (optional). TL

Thursday
AM
1. Continue continuous monitoring. TL, T4
2. Ask occupants to complete the questionnaire. (Building manager)
3. Collect bulk biological samples from 3 fixed sites underfloor diffusers and other indoor locations, where noticeable biological contamination is observed. T4
4. Collect carpet dust samples from 3 fixed indoor sites and other indoor locations, where noticeable biological contamination is observed. Each collection area should be approximately 1 m² (10.8 ft²) TL

PM
1. Before the end of the work day remind occupants to fill in the questionnaire. Thank them for participating in the study. TL, T4
2. Pack sampling equipment. Retrieve radon samplers. TL, T4
3. Download data from all installed loggers and the existing indoor environmental monitoring system in the building. TL, T4
4. Ship radon samplers. TL
5. Ship biological samples. TL

Friday
1. Check that all data is collected. TL, T2
2. Quality control check of the data. TL
3. Pack equipment. TL, T2
4. Retrieve the fliers with Frequently Asked Questions. TL
5. Move out. TL, T2
7.3 Instrumentation

Outdoor wind and pressure conditions were measured using the HOBO weather station. Outdoor conditions were averaged over 5 min intervals and then recorded, for the entire duration of the study.

Air temperature was measured at three fixed indoor locations, an outdoor location and two mechanical rooms. At the three indoor locations, temperatures were measured at 4 different heights: 0.1 m (0.3 ft), 0.6 m (2 ft), 1.1 m (3.6 ft) and 1.7 m (5.6 ft). Temperatures at 0.1 m (0.3 ft) and 0.6 m (2 ft) heights were recorded every 10 seconds using thermistors. Temperatures at 1.1 (3.6 ft) and 1.7 m (5.6 ft) heights were recorded every 30 seconds using HOBO U12-012 data logger with a built in thermocouple temperature sensor and a type D HOBO external sensor, respectively. Outdoor temperatures were measured by the HOBO weather station. Outdoor temperatures were averaged over 5 min intervals and then recorded, for the entire duration of the study. In the mechanical rooms, temperatures were measured in the well mixed areas, such that they were representative of the supply air conditions for multiple supply duct intakes, as the entire mechanical rooms behaved as mixing chambers. In the East Mechanical Room, temperatures were recorded using a type D HOBO external sensor and HOBO U12-006 data logger, while in the West Mechanical Room temperatures were recorded using a HOBO H8 sensor and data logger, which had smaller logging capabilities. Therefore temperatures recorded in the West Mechanical Room were downloaded daily, during which period sensor was temporarily removed from the mechanical room. In both mechanical rooms, temperatures were recorded every 30 seconds.

Additional BMS temperatures were recorded for comparison. Measurement points for the BMS sensors were located 0.5 m (1.6 ft) away from the fixed indoor locations, mounted on a vertical wall at 1.55 m (5.1 ft) height. BMS used Vaisala HMW60Y temperature sensors. Originally BMS was set to record data every 15 min, but during the study week these settings were adjusted in order to record data every 5 min.

Relative humidity (RH) was measured at three fixed indoor locations, an outdoor location and West mechanical room. At the three indoor locations, RH was measured at 1.1 m (3.6 ft) height and recorded every 30 seconds using HOBO
U12-012 data logger with a built-in RH sensor. Outdoor RH was measured by the HOBO weather station. Outdoor RH was averaged over 5 min intervals and then recorded, for the entire duration of the study. In the West mechanical room, RH was measured in the well mixed area, such that it was representative of the supply air conditions for multiple supply duct intakes, as the entire mechanical room behaved as a mixing chamber. In the West Mechanical Room RH was recorded every 30 seconds using a HOBO H8 sensor and data logger. Because of the small logging capabilities, RH data recorded in the West Mechanical Room were downloaded daily, during which period sensor was temporarily removed from the mechanical room.

Additional BMS RH was recorded and presented for comparison. Measurement points for the BMS sensors were located 0.5 m (1.6 ft) away from the fixed indoor locations, mounted on a vertical wall at 1.55 m (5.1 ft) height. BMS used Vaisala HMW60Y RH sensors. Originally BMS was set to record data every 15 min, but during the study week these settings were adjusted in order to record data every 5 min.

Sound noise was measured at three fixed indoor locations at 1.1 m (3.6 ft) height and recorded every 10 seconds using MARTEL C322 noise sensor and data logger.

Light intensity was measured at three fixed indoor locations at 1.1 m (3.6 ft) height and recorded every 30 seconds using HOBO U12-012 data logger with a built-in light sensor. Additionally, solar radiation was recorded at the outdoor location by the HOBO weather station. Outdoor measurements were averaged over 5 min intervals and then recorded, for the entire duration of the study.

CO₂ concentration was measured in the two mechanical rooms. Measurements were taken in the well mixed areas using Engelhard Telaire 7001 CO₂ sensor. The concentrations were recorded every 30 seconds using HOBO U12-006 and HOBO H8 data loggers, respectively in the East and West Mechanical rooms. Due to the smaller logging capabilities of HOBO H8, CO₂ concentrations recorded in the West Mechanical Room were downloaded daily, during which period sensor was temporarily removed from the mechanical room. The measured CO₂ concentrations were representative of the supply air conditions for multiple supply duct intakes, as the entire mechanical rooms behaved as mixing chambers.
Additional BMS CO₂ concentrations were recorded in the vicinity of the three fixed indoor monitoring locations. Measurement points for the BMS sensors were located 0.5 m (1.6 ft) away from the fixed indoor locations, mounted on a vertical wall at 1.55 m (5.1 ft) height. BMS used MSA AIROX IAQ MONITOR MODEL 711271 CO₂ sensors. Originally BMS was set to record data every 15 min, but during the study week these settings were adjusted in order to record data every 5 min.

Particle concentration was measured simultaneously at F3 SE indoor and an outdoor location with two different instruments: Lighthouse 3016 IAQ particle counter and TSI Model 8525 P-TRAK ultrafine particle counter. The simultaneous measurements were conducted for the purpose of quality assurance. However, the particle detection range was quite different for the two instruments: 0.3 µm to 10 µm for Lighthouse and 0.02 µm to greater than 1 µm for TSI. Particle concentrations were averaged and recorded every 5 minutes by both instruments. All measurements were conducted on February 28, 2008. In the first half of the day Lighthouse was at the outdoor location, while TSI was at F3 SE indoor location. In the second half of the day the instruments switched positions. This procedure was developed as the researchers did not have sufficient number of instruments to follow BASE protocol for measuring particle concentrations.

Volatile organic compound concentration was measured at three fixed indoor locations and an outdoor location. At the three indoor locations, VOCs were measured at 1.1 m (3.6 ft) height using three types of instrumentation: (1) RAE PGM 7240 VOC meter and data logger, (2) sorbent tubes and (3) TO-15 SUMMA canisters. However, the instruments were used to detect and measure different VOCs compound concentrations. Samples collected with sorbent tubes and TO-15 canisters were analyzed by EMSL Analytical Inc. Samples were collected over different time intervals depending upon the instrumentation that was used. Thus, measurements performed by RAE VOC meter were averaged and collected every 15 seconds. Sorbent tubes sampled 10 L (0.35 ft³) of air and sampling time depended upon a sampling pump airflow rate. Using the 0.2 L/min (0.007 ft³/min) pump, samples were collected over 50 min sampling intervals. TO-15 samples were collected continuously over an 8 hour period. While TO-15 samples were collected simultaneously at all four measurement locations, RAE and sorbent tube samples
were collected one at a time.

Sorbent tubes were used to measure aromatic hydrocarbon concentrations. These measurements were performed twice at two indoor locations, F1 SW and F5 NE, for the purpose of quality assurance.

Acetaldehyde and formaldehyde concentrations were measured at three fixed indoor locations and an outdoor location. At the three indoor locations, aldehydes were sampled at 1.1 m (3.6 ft) height using 2,4-dinitrophenylhydrazine (DNPH) sorbent tubes. For quality assurance purposes measurements of acetaldehyde concentrations were conducted at the same four locations using RAE PGM 7240 VOC meter and data logger. Additionally, duplicate samples were collected at F1 SW indoor location using sorbent tubes. Samples collected with sorbent tubes were analyzed by EMSL Analytical Inc. While RAE VOC meter averaged and collected acetaldehyde concentrations every 15 seconds, sorbent tubes sampled 15 L (0.53 ft\(^3\)) of air over a sampling interval which depended upon a sampling pump airflow rate. Using the 0.3 L/min (0.01 ft\(^3\)/min) pump, samples were collected over the 50 min sampling intervals.

Radon activity was tested at six indoor locations using charcoal based testing device, over a 96 hour period. Test samples were analyzed by EMSL Analytical Inc. Following the BASE protocol, since the study examined indoor environmental conditions at the second floor of the test building, most of the six radon testing locations were at the second floor. Thus, the choice of radon testing locations did not comply to the DEP recommendations for building radon testing, advising radon testing in the lowest part of a building, such as a basement. Three of the testing locations were identical with the already used indoor measuring locations, while another three included pathways for possible indoor radon transport, such as elevator shafts and stairwells.

Bacterial colonies were identified and counted in the air as well as on the surfaces close to the monitoring locations. Air samples were collected at three fixed indoor locations and an outdoor location using Andersen N-6 type VP-400 Impactor kit with Tryptic(ase) Soy Agar (TSA) plates. At these four locations, air was sampled at 1.1 m (3.6 ft) height over 2 and 5 min sampling intervals and 28.3 L/min (1 ft\(^3\)/min) airflow rate. Samples from surfaces close to the three indoor monitoring locations were collected using Mold swabs. Swabbed surfaces were
selected in the airways of under-floor air distribution diffusers closest to the fixed indoor monitoring locations. All samples were analyzed by EMSL Analytical Inc. Although BASE protocol required morning and afternoon air sampling, due to the lack of manpower and equipment, sampling was performed only once, around noontime. Furthermore, unfavorably cold outdoor conditions prevented 5 min outdoor sampling.

For quality assurance purposes, air sampling was doubled at F3 SE indoor location, while an additional swab sample was collected from a possible bacteriological contamination source location in the central stairwell.

Viable fungi colonies and fungal spores were identified and counted in the air as well as on the surfaces close to the monitoring locations. Air samples were collected at three fixed indoor locations and an outdoor location using (1) Andersen N-6 type VP-400 Impactor kit with Malt Extract Agar (MEA) plates and (2) Micro-5 fungal spore trap. At these four locations, air was sampled at 1.1 m (3.6 ft) height. Two and five minute sampling intervals with 28.3 L/min (1 ft³/min) airflow rate were used for MEA agar plates, while four minute sampling with 5 L/min (0.18 ft³/min) flow rate was appropriate for Micro-5 fungal spore counting. Samples from surfaces close to the three indoor monitoring locations were collected using Mold swabs. Swabbed surfaces were selected in the airways of under-floor air distribution diffusers closest to the fixed indoor monitoring locations. Additionally, samples were taken from the central parts of the supply air grilles in the two mechanical rooms, using Bio tape. All samples were analyzed by EMSL Analytical Inc. Although BASE protocol required morning and afternoon air sampling for both MEA agar plates and fungal spore traps, due to the lack of manpower and equipment, MEA agar plate air sampling was performed only once, around noontime.

For quality assurance purposes, MEA agar plate air sampling was doubled at F5 NE indoor location, while additional swab and tape samples were collected from possibly contaminated surfaces in the central and western stairwells. Also, fungal spore air sampling was doubled at F1 SW indoor and an outdoor location.

Mite allergens were identified and measured in the collected dust samples from the carpet floor underneath the three fixed indoor monitoring locations. The samples were collected from 1 m² (10.8 ft²) floor area using a vacuum cleaner mounted
dust collector with filter cassette. All samples were analyzed by EMSL Analytical Inc. for presence of two types of dust mite allergens - Dermatophagoides pteronyssinus (Der p 1) and Dermatophagoides farinae (Der f 1), as well as a cat mite allergen - Felis domesticus (Fel d 1).

## 7.4 Practical Recommendations for Similar Studies

Mock-up experimental setup of all instrumentation and activities was performed prior to the actual study week. This has proved to be very beneficial for all the team members and resulted in timely correction of possible field instrumentation and coordination problems. Such muck-up is suggested at least 2 weeks prior to the actual future study to have enough time to order and purchase supplies that might be necessary for the study, but were not considered in the experimental planning. Nevertheless, detailed plan of activities during the study week for each member of the study team is very beneficial and should be provided at least 3 weeks prior to the field study, such that team members have enough time to prepare for the mock-up.

Compared to the field plan and schedule described in the current study, additional activities should be scheduled for Monday of the study week, including the start of some measurements. This will allow partial relocation of activities planned for Wednesday, as the busiest day of the week, towards Tuesday or Thursday.

Study week should be selected based on the weather forecast to avoid below freezing outdoor temperatures, which may impact the equipment and accuracy of measurements. Precise time synchronization of all equipment, data loggers and personal watches of the team members should be performed before the field study. Access to internet and portable computing capabilities proved essential for resolving on-site information problems. Access to a portable on-site printing/scanning/copying device was also very useful.

During the study week, each measuring location should have a clearly written sign specifying the official study location name/coding, together with a log sheet for members of the study team to indicate date and time of their presence at the
measuring location. Pens which do not emit VOCs should be left at each measuring location and used exclusively throughout the study week.

Close cooperation with the building managers and their willingness to encourage workers to complete the online survey has proven to be very beneficial for successful collection of survey responses. Longer collection time was necessary compared to BASE protocol due to the online nature of the survey. Collection period of 8 days is suggested starting on 9AM Monday of the study week and ending on 5PM Monday of the following week to accommodate any late submissions. Midweek check of the survey collection progress is suggested to overcome any possible errors and/or direct further efforts towards successful survey collection.

Usage of autonomous automatic data loggers similar in accuracy and performance to HOBO U12, together with compatible sensors is recommended. Also, usage of noise sensors and data loggers similar in accuracy and performance to Martel C322 is recommended. These sensors and data loggers have sufficient accuracy, while their portable design makes them easily deployable. As the C322 instruction manual precisely indicates the time when the sensor’s 12V battery will be depleted, at least 2 replacement or batteries with larger capacities are necessary to ensure minimal monitoring interruption.

For the purpose of comparison between indoor light intensity and solar radiation, compatible indoor and outdoor instrumentation is suggested to convert between comparable indoor and outdoor lighting levels.

Usage of particle counters and data loggers similar in accuracy and performance to Lighthouse 3016 is recommended. This type of sensors and data loggers has sufficient accuracy and ability to automatically classify particles with respect to sizes, at the same time being easily deployable. Also, modification of the BASE protocol is recommended to include continuous particle counting during the whole study week, rather than during a single day.

If VOCs are measured with sorbent tubes or devices similar to RAE PGM 7240 VOC meter, effort to collect all VOC samples simultaneously might be worthwhile. Also, sorbent tube sampling with low sampling pump airflow rates is suggested. Such sampling would result in longer sampling times and could provide better IAQ insight. Keeping sampling time records is recommended. However, due to the higher sensitivity of the used VOC meter and TO-15 sampling compared to
sorbent tube analyses, usage of sorbent tubes is not recommended in the future.

Additional protocol changes are suggested regarding the sampling times for bacterial and fungal air samplings. An adjusted study plan is necessary to ensure the collection of both sets of samples, morning and afternoon, as the presented validation study schedule was not met and no afternoon sampling could be collected.

Shipments of collected samples for laboratory analyses are followed by a Chain of Custody (CoC) form. As different analyses required different CoC formats in the current study, a team member should be designated to fill-in the CoCs and ensure consistency of information. Information in the CoC should always contain internal study sample coding together with indicators of location and/or measuring intervals. If present, a distinctive factory sample number should also be included. This will greatly ease the data analyses and eliminate unnecessary code conversions. Also, shipping blanks and field blanks should be clearly indicated in the CoC. Distinction between bacterial and fungal agar plates should be included in the sample code, rather than left to the lab technicians to determine whether the samples are intended for bacterial or fungal analyses. Sample coding should be agreed among the team members prior to the sample collection procedures.
Chapter 8

Results and Discussion on Experimental Validation

This Chapter contains a comprehensive summary results of conducted experimental validation study within PA DEP Cambria office building described in Chapter 7. The data were collected in accordance to EPA BASE guidelines and included occupant survey responses, measurement results, and building information. Apart from the measurements of equipment temporarily installed during the study week, the analyzed data also included records from the permanently installed building monitoring system (BMS). BMS data provided CO\textsubscript{2} concentrations, temperatures and relative humidity measurements used for quality assurance purposes.

The averaged indoor parameters were further used to validate developed NN methodology for predicting respiratory symptoms of building indoor environments. Results of the validation and comments on overall validation procedure provide foundations for future work and improvements of the developed methodology.

8.1 Experimental Results and Comparison to Average Benchmark Building

Following the procedures described in Chapter 7, measurements were conducted at the three fixed indoor and one outdoor location of PA DEP Cambria office building. Additionally, 50 building occupants completed the online survey expressing per-
ceptions of indoor environmental quality, reporting personal health symptoms and answering work related questions. In association with the building management, detailed building information data were collected as part of the building characterization section of the study. Processing and analyses of the numerous data records taken at these monitoring locations resulted in a comprehensive overview of averaged PA DEP Cambria “green” building indoor environmental conditions. The following selection of experimental results, together with supplementary information in Appendix B attempt to present the most significant findings of the conducted analyses.

Thus, Figure 8.1 presents thermal comfort conditions on a psychometric chart together with typical winter and summer temperature and humidity comfort ranges (ASHRAE, 2004), as well as averaged benchmark BASE office building conditions. The results show that thermal comfort within PA DEP Cambria office building was borderline acceptable for winter conditions. These findings were in agreement with a previous study evaluating PA DEP Cambria building system performance (Deru et al., 2005). Additionally, building survey data confirmed that absence of the building HVAC humidification resulted in failure to meet design goals and provide 25% minimum relative humidity.

Consequently, occupants’ perceptions of air dryness in PA DEP Cambria were more often than in the benchmark building, with 55% of the occupants experiencing air as too dry within the 4-week period prior to the study week. Furthermore, none of the occupants experienced the air as too humid. At the same time, all other perceptions of indoor environmental conditions in PA DEP Cambria were better than in the benchmark building, with the highest difference in perceived indoor temperatures. Namely, 20% more PA DEP Cambria occupants perceived indoor temperature as just right, neither too hot nor too cold, compared to an average BASE building. Confirming the thermal comfort findings from the literature (Fisk et al., 2004), and contradictory to the typical performance of commercial UFAD systems (Mumma, 2009), these results are presented in Figure 8.2. Interpreting the results from Figure 8.2, notice that all percentages represent the portion of occupants which did not experience corresponding perceptions during the four week period prior to the study week. Consequently, the higher the percentages are in Figure 8.2 the more satisfied occupants are with indoor environmental conditions.
Figure 8.1. Thermal comfort conditions in PA DEP Cambria “green” building (marked X) compared to average BASE building (marked +) and typical winter (1.0 clo) and summer (0.5 clo) comfort ranges corresponding to 80% occupant acceptability (allowing 10% dissatisfaction for whole body thermal discomfort based on predicted mean vote (PMV Limits) between ±0.5, plus an additional 10% dissatisfaction for partial body thermal discomfort)

Comparing PA DEP Cambria vs. a typical BASE building in Figure 8.2, the most significant advantages and disadvantages of the former are circled with single and double lines, respectively.

Large advantage of examined “green” over benchmark office environment is also visible from the analyses of self-reported SBS symptom occurrence, presented in Figure 8.3. In PA DEP Cambria occupants reported lower occurrence for 15 out of 19 investigated symptom categories with the biggest differences compared to averaged BASE building of 23% lower occurrence of headache and 19% lower occurrence of eye irritation and irritability/nervousness. The rarest symptoms in the investigated “green” building environment were chest tightness, dizziness/lightheadedness and shortness of breath, experienced by less than 7% of the occupants during a 4-week period prior to the study week. Figure 8.3 points out
Figure 8.2. Occupants’ perceptions of indoor environmental conditions in PA DEP Cambria “green” building compared to average BASE building.

these findings, circled with a single line, while double line marks the most often occurring symptom, stuffy/runny nose, and highest reported disadvantages for the PA DEP Cambria “green” building occupants. Namely, 11% more of them experienced dry or itchy skin, while 7% more had cough compared to an average benchmark office building. Higher occurrence of cough may be related to the measured and perceived dryness of indoor air, but such conclusion would require further statistical investigation.

Findings presented in Figure 8.3 were further classified into four symptom groups presented in Figure 8.4. Such classification was based on a previous study using exploratory factor analyses of BASE data to define association between the occupants’ symptoms and the so called “union variables” (Brightman, 2005):

1. tiredness:
   - dry, itching or irritated eyes,
   - headache,
Figure 8.3. Occupant self-reported occurrence of sick building syndrome symptoms in PA DEP Cambria “green” building compared to average BASE building

- unusual tiredness, fatigue or drowsiness,
- tired or strained eyes,
- tension, irritability or nervousness,
- pain or stiffness in back, shoulders or neck

2. mucosal irritation or upper respiratory symptoms:
- sore or dry throat,
- stuffy/runny nose or sinus congestion,
- cough,
- sneezing,
- dry or itchy skin

3. neuropsychological factors:
- tension, irritability or nervousness (also loaded on the tiredness cate-
• difficulty remembering or concentrating,
• dizziness or lightheadedness,
• feeling depressed,
• nausea or upset stomach

4. lower respiratory symptoms:
• wheezing,
• chest tightness,
• shortness of breath

The analyses indicated negligible correlation between the responses from occupants within each BASE building and compared further two office buildings investigated by NIOSH to the BASE dataset. Similar approach was applied in the current study to include PA DEP Cambria building records and compare them with BASE and NIOSH buildings. Consequently, Figure 8.4 shows higher prevalence of mucosal irritation and negligible tiredness among occupants in PA DEP Cambria building compared to BASE, while all four “union” variable values fall within the two NIOSH building study results.

Although PA DEP Cambria building occupants on average experienced less SBS symptoms compared to BASE, and had negligible tiredness, they reported similar impact of these symptoms on productivity, presented in Figure 8.5. Thus, in spite of the reported 4-5% reduction in abstenions from work in PA DEP Cambria “green” building compared to benchmark (circled in Figure 8.5), when asked how often the symptoms reduced their ability to work occupants responded similarly to the benchmark case.

In addition to the reported lower occurrence of SBS, PA DEP Cambria building occupants also reported less medically diagnosed health symptoms for 5 out of 6 investigated conditions, presented in Figure 8.6. Thus, between 5 and 10% less occupants in PA DEP Cambria building reported various diagnosed allergies, migraine and asthma, while only 1.5% reported more diagnosed eczema in comparison to the average benchmark BASE building.

Further advantages for PA DEP Cambria office building environment are related to measurements of 4 to 19 times lower indoor allergen, bacterial and fungal
Figure 8.4. Categories of building occupant symptoms and their prevalence percentage in BASE (gray range covers 50% of the buildings closest to the database mean value, i.e. interquartile range), two NIOSH buildings (solid and dashed lines) and PA DEP Cambria building (dotted line) (modified from (Brightman, 2005))

Figure 8.5. Occupant self-reported impact of sick building syndrome symptoms on productivity in PA DEP Cambria “green” building compared to average BASE building (N/A - not applicable category - corresponds to reports of no negative impacts on productivity)
Figure 8.6. Occupant self-reported diagnosed medical conditions in PA DEP Cambria “green” building compared to average BASE building

Concentrations compared to the average BASE office building, as presented in Figure 8.7. Interestingly, although no dust mite allergens were detected, cat mites were present, which is believed to be the consequence of occupants’ contact with home pets (Ahlbom et al., 1998). Additionally, PM2.5 particle concentrations were 7 times lower than BASE average, while PM10 concentrations measured 60% higher. Comparison of measured indoor and outdoor particle concentrations between PM0.5 and PM10 did not reveal reasons for such high indoor PM10 concentrations, as all outdoor concentrations were above corresponding indoor values. Furthermore, inspection of the building’s air filtration system showed timely maintenance with proper operation, offering no possible explanation for the described indoor PM concentration discrepancies. Consequently, the origin of high indoor PM10 concentrations may be indoor particle resuspension due to UFAD, present in PA DEP Cambria, but not common in BASE buildings (Van Grieken, 2009). Nevertheless, these records should be considered with caution, as indoor and out-

![Diagnosed symptoms](image-url)
door PM measurements were not simultaneous due to the limited measurement resources. Additional spectroscopic analyses of sampled particles may indicate their chemical origin and point to potential sources. Also, BASE measurements did not include PM other than 2.5 and 10 µm, resulting in inability to compare PA DEP Cambria records of PM0.5, PM1, PM5 and total PM concentrations to BASE buildings.

**Figure 8.7.** Measured allergen, bacterial, fungal and particle concentrations in PA DEP Cambria “green” building compared to average BASE building records

Including aldehydes, a total of 75 various VOCs were identified by EMSL Inc. laboratories in the air samples collected in PA DEP Cambria office building. Out of these, 68 were not identified including most of aromatic hydrocarbons, while the only significantly increased concentration was recorded for d-Limonene, a natural part of citrus fruit oil. The recorded d-Limonene concentrations were 5 times higher than BASE building average. Probable reasons for such findings are usage of “green” cleaning products within PA DEP Cambria building, as d-Limonene is used in chemical industry to give orange flavor to cleaning products and paint.
Unfortunately, these indications could not be confirmed as the cleaning products in the building did not list ingredients. Although D-Limonene is considered very safe and has no permissible exposure limit (PEL), it can cause irritation of skin, eyes and respiratory system (OSHA, 2009).

Out of the identified VOCs, only part could be directly compared to VOCs collected in the BASE study due to inconsistencies between BASE and EMSL Inc. compound identification protocols. Additionally, BASE study used two different procedures to analyze VOC samples resulting in two groups of detected VOCs within the buildings, as described in Chapter 6. Consequently TVOC concentrations were not used as a measure of comparison between PA DEP Cambria and BASE data.

Finally, all of the conducted measurements and investigated conditions support the occupants’ reported higher satisfaction with the working environment and job in general, within PA DEP Cambria “green” office building compared to averaged values of the U.S. office buildings in BASE.

8.2 Validation Predictions

The collected experimental data were further used to validate the developed NN methodology for predicting respiratory health symptoms of building indoor environments. However, due to the inconsistencies between the identified VOCs in PA DEP Cambria and BASE, NN input using TVOC concentrations had to be excluded from the validation. Additional justification for omitting TVOC concentrations from the input data set may be found within a study on 12 office buildings in California investigating association of various VOC metrics with SBS (Brinke et al., 1998). Confirming the impact of low level VOC concentrations on occurrence of SBS symptoms using the metrics based on substance irritability, the study discovered that most commonly used VOC metric of calculating TVOC concentrations, was unable to provide statistically significant associations with SBS.

Consequently, exclusion of TVOC NN inputs resulted in using only six of the previously considered seven inputs (Table 5.3). New NNs with six input and single output neurons were designed, trained and tested on the BASE dataset, following the protocol described in Chapters 5 and 6. These NN training and testing
results are contained within Appendix C. The trained networks then used data from PA DEP Cambria “green” office building for the validation purposes. The obtained NN building related symptom prediction results, together with training and testing errors are presented in columns 5 through 7 of Table 8.1 for each of the considered NN outputs. Also, Table 8.1 presents actual probabilities for occurrence of both, building related and overall symptoms, within the third and second columns, respectively. Column 4 in Table 8.1 contains 95% confidence intervals for actual occurrence probabilities of building related symptoms contained in Column 3. In addition to NN output symptoms defined in Table 5.3, second, third and fourth column of Table 8.1 contain probabilities for occurrence of additional 3 SBS symptoms used to calculate BSI as specified in Table 5.4.

Comparison of actual to NN predicted probabilities for occurrence of respiratory health symptoms in PA DEP Cambria building identified NN predictions within 95% CI of the actual records in case of BSI and additional 6 investigated respiratory symptoms: chest tightness, dust allergy, shortness of breath, sinus congestion, sore throat, and wheezing. Out of the remaining 4 respiratory symptom NN predictions 3 were less than 5% away from the 95% CI range, while only occurrence of asthma was significantly overestimated. Such capabilities of NNs to infer respiratory health symptom occurrence probabilities in office building environments and generalize upon variety of building types, offer promising possibilities for future applications in predicting occurrence of health symptoms among building occupants.

8.3 Comments on Experimental Measurement Results

In the beginning of the experimental study week it was discovered that one of the two HVAC systems in the investigated building was not operational. No information was available about the time when this atypical mode of operation occurred, as it did not follow the HVAC system schedule. Since such operation was not representative of the building operating conditions, the system was restarted to enable measurement of representative indoor environmental building conditions.

Measurements of VOCs detected very high concentrations of isopropyl alcohol.
Table 8.1. Overview of NN validation performance in predicting probabilities for occurrence of building related respiratory symptoms among occupants in PA DEP Cambria “green” building

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Total prob. [%]</th>
<th>Building related prob. [%]</th>
<th>95% CI [%]</th>
<th>NN building related prob. [%]</th>
<th>Training error [%]</th>
<th>Testing error [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>asthma</td>
<td>2.04</td>
<td>2.04</td>
<td>3.96</td>
<td>13.54</td>
<td>3.02</td>
<td>3.11</td>
</tr>
<tr>
<td>chest tightness</td>
<td>6.25</td>
<td>4.17</td>
<td>5.71</td>
<td>6.73</td>
<td>2.26</td>
<td>3.59</td>
</tr>
<tr>
<td>cough</td>
<td>41.67</td>
<td>29.17</td>
<td>13.00</td>
<td>11.72</td>
<td>3.14</td>
<td>3.67</td>
</tr>
<tr>
<td>dust allergy</td>
<td>26</td>
<td>26</td>
<td>12.28</td>
<td>23.62</td>
<td>6.72</td>
<td>10.32</td>
</tr>
<tr>
<td>mold allergy</td>
<td>18</td>
<td>18</td>
<td>10.76</td>
<td>30.31</td>
<td>5.75</td>
<td>7.81</td>
</tr>
<tr>
<td>shortness of breath</td>
<td>6.38</td>
<td>2.12</td>
<td>4.17</td>
<td>5.37</td>
<td>1.90</td>
<td>2.85</td>
</tr>
<tr>
<td>sinus congestion</td>
<td>58.33</td>
<td>33.33</td>
<td>13.48</td>
<td>26.52</td>
<td>4.05</td>
<td>8.50</td>
</tr>
<tr>
<td>sneezing</td>
<td>51.02</td>
<td>16.33</td>
<td>10.46</td>
<td>30.07</td>
<td>3.63</td>
<td>7.03</td>
</tr>
<tr>
<td>sore throat</td>
<td>46.81</td>
<td>21.28</td>
<td>11.83</td>
<td>13.13</td>
<td>3.94</td>
<td>9.39</td>
</tr>
<tr>
<td>wheezing</td>
<td>12.77</td>
<td>8.51</td>
<td>8.07</td>
<td>2.68</td>
<td>2.12</td>
<td>3.01</td>
</tr>
<tr>
<td>dry, itchy, irritated eyes</td>
<td>39.58</td>
<td>12.50</td>
<td>9.45</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>unusual tiredness, fatigue</td>
<td>46.67</td>
<td>22.23</td>
<td>12.28</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>headache</td>
<td>45</td>
<td>15.00</td>
<td>11.21</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BSI</td>
<td>3.54</td>
<td>1.65</td>
<td>0.59</td>
<td>1.82</td>
<td>0.28</td>
<td>0.22</td>
</tr>
</tbody>
</table>

One of the reasons for such occurrence may be cross-contamination from concurrently performed bacterial and fungal sampling, where isopropyl alcohol was used to disinfect the instrumentation. Therefore, changes in the study protocol are suggested to ensure that VOC measurements are done before bacterial and fungal samplings. Furthermore, VOC sampling using sorbent tubes had relatively high minimum detection limit resulting in the small number of quantified VOC concentrations. Therefore, the overall number of detected VOCs within PA DEP Cambria building should be considered with caution and reconfirmed using more sensitive detection methods.

The current findings indicate that 5 min bacteriological and fungal air sampling
resulted in identification of greater variety of bacterial and fungal taxa compared to 2 min sampling. Therefore, only 5 min sampling is recommended for similar studies in the future. Other publications based on BASE data confirm these findings and indicate that 5 min sampling time can be even extended to provide sufficient information about the presence of bacteria and fungi in the air. Literature references that only 13.6% of 2 min samples and 28.1% of 5 min samples met the minimum 10 CFU/plate counts suggested for air sampling, while only 2.4% of 2 min samples and 8.8% of 5 min samples exceeded the minimum bacteriological colony counts of 25 CFU/plate suggested for reliable estimation of bacterial concentrations (Womble et al., 1999; Tsai and Macher, 2005).

Neural network training and testing predictions used one input less compared to the previously defined input list in Table 5.3. Such restrictions in availability of input information had negative impact on NN performance with training and testing errors, presented in Table 8.1, being typically higher compared to 7 input NNs showed in Table 6.2. Consequently, NN performance may be increased introducing additional input variables.
Conclusions and Future Work

The current research tried to provide an interdisciplinary approach towards analyses of building indoor environments, attempting to bring together science and people, engineering and medicine. Such approach was demonstrated in analyses of a benchmark Building Assessment Survey and Evaluation dataset, comprised of measurements, occupant questionnaire and building information from 100 U.S. office buildings. Also, additional experimental study of a “green” building environment was conducted following the BASE protocol. The analytical applied methods included multivariate logistic regression to determine statistically significant relations between recorded indoor parameters, as well as artificial neural networks used to predict respiratory health symptom occurrence in office building environments.

9.1 Conclusions from Multivariate Regression Analyses

The current study used multivariate regression analysis to determine the most significant parameters for occurrence of health responses, including upper and lower respiratory SBS symptoms, asthma, as well as dust and mold allergies among occupants in the buildings part of the benchmark BASE database. Environmental parameters that were considered included concentrations of various culturable fungi, airborne fungal spores, bacteria, antigens, temperature, relative humidity and occupants’ perceptions of indoor environments. The fungal spores that were most
often present in buildings, Cladosporium, were also the ones that had the highest measured concentrations. The same fungal taxon was detected in culturable fungal samples from almost all investigated buildings. Data from culturable fungal samples appeared as more comprehensive than airborne fungal spore counts, detecting larger variety of fungal taxa in more buildings. Nevertheless, the most statistically significant parameters were occupants’ perceptions of odors and relative humidity, while the biological contaminants exhibited lower statistical associations with the examined health outcomes. The most common negative perception among occupants was related to unpleasant non-chemical odors, while the least number of times occupants perceived odors from tobacco smoke. At the same time, multivariate analyses showed discrepancies between measured and perceived relative humidity in association with the examined health outcomes, confirming inability of occupants to adequately perceive relative humidity.

Nevertheless, high occupants’ sensitivity to odors implies the need for new types of sensors to be applied in indoor environmental measurements, as current sensor sensitivity is not sufficient. Furthermore, techniques presently used in investigation of indoor environmental health impacts have been criticized for not being suitable to discover parametric associations indoors, as they are designed for other environments, such as ambient or industrial air. Therefore, many causes for building occupants’ health problems remain unclear. It appears that human occupants are still the most sophisticated “sensors”, and, therefore, indoor air quality studies should try to include human subjects’ perceptions whenever possible.

9.2 Conclusions on Methodology and Usage of NNs

Extensive literature review revealed that previous studies almost exclusively used multivariate logistic regression for analyses of association between indoor environmental parameters and occupant experienced respiratory health symptoms. However, such methods were so far unable to identify definite causes of particular respiratory symptoms, provide predictions for occurrence of such symptoms, or ultimately explain the origin of the so called sick building syndrome. Arguably, one
of the reasons for such situation is interdependence of occupant experienced symptoms on a number of concurrently changing indoor environmental influences and inability of traditional techniques to deal with such array of various parameters.

As an alternative approach, the present study introduced NN based methodology which was able to successfully address shortcomings of traditionally used statistical methods. The presented results illustrated promising capabilities of the developed NN methodology to predict respiratory health responses of building occupants. NNs randomly selected 90% of the available BASE buildings for training and 10% for testing, although no strict rules existed governing the number of training and testing cases. While NN prediction results mostly fell within the 95% confidence intervals around actual probabilities for occurrence of respiratory health symptoms among building occupants, NNs also exhibited relatively poor ability to follow the actual trend values in test cases. Coupled with large dispersion of NN results around actual output values, such performance could be attributed to the small number of testing cases and consequently large impact each inaccurate prediction had on overall testing performance evaluation. Thus, alternative distribution of available building data may lead to different training and testing results. The developed NN algorithm tried to optimize NN prediction performance considering differences in randomized initialization and variability in possible sizes of a two layer MLP NN. However, the additional impact on NN performance made by the randomized selection of BASE training/testing buildings was not analyzed. Nevertheless, the presented evaluations of NN predicted probabilities for occurrence of building related respiratory symptoms showed promising capabilities of NNs to infer occurrence of such symptoms. These capabilities were further validated on a certified LEED Gold “green” building experimental case study.

The developed NN prediction methodology was successfully validated using the data collected during an experimental study in a “green” building. The study was conducted in accordance with BASE data collection protocol. However, VOC identification did not correspond to BASE and consequently TVOC concentration inputs could not be included in the NN analyses. Consequently, NN prediction accuracy was lower in the validation case compared to BASE data analyses, due to the reduced number of NN inputs. Nevertheless, NN prediction methodology was
applied and resulted in satisfactory performance predicting 7 out of 11 respiratory conditions within the 95% confidence interval of the actually recorded experimental data. Significantly, such performance was achieved on a “green” building validation dataset, while NNs were trained using BASE “non-green” buildings. The study demonstrated how BASE dataset could be applied to generalize methodology for predicting respiratory health symptom occurrence across the variety of current office building types and input datasets. Inclusion of additional significant input parameters could contribute to the accuracy of used NN prediction methodology. Input significance should be determined by means of multivariate logistic regression analysis and/or research experience from similar studies. On the other hand, number of possible input parameters would be limited by the size of available BASE training dataset, as increased number of inputs would require additional sample cases to perform adequate training. Finally, when applied in accordance to the developed methodology and combination to the traditional statistical techniques, NNs could be a valuable tool to explain SBS and occurrence of occupants respiratory symptoms.

9.3 Conclusions on Experimental Analyses of a Green Building

In addition to validation of the developed NN prediction methodology, data collected within the PA DEP Cambria “green” building were compared to the average BASE U.S. office building. This was possible as BASE data collection protocol was mostly also followed in the “green” building experimental study. The comparison between the collected and average BASE data primarily included measurements and occupant survey responses, resulting in a number of quantifiable advantages for the “green” building compared to average “non-green” BASE office environment. Among other benefits, the “green” building occupants reported reduced occurrence of SBS symptoms and less diagnosed medical conditions, while measurements revealed lower VOC, bacterial, allergen and fungal concentrations. The only significant disadvantage was related to occupant comfort, which was borderline acceptable due to the very low air humidity and unmet design goals. Con-
sequently, the investigated “green” office building was in certain aspects better than BASE buildings, but in other aspects worse. Investigating a single “green” building is insufficient for making any general conclusions about “green” building advantages or disadvantages, but the current study specifies the methodology and opens the path for similar studies in the future.

9.4 Future Work

The collected experimental data offer a wealth of still not investigated information and present valuable resource for future studies. Furthermore, PA DEP Cambria “green” building is only one of the buildings towards a comprehensive database containing information from LEED rated buildings. While the collected data analyses are a significant first step confirming the health related benefits of a “green” building, it is no more than the first step to generalize such findings. Only large scale data collection and analyses on a statistically significant sample would allow researchers to draw valid conclusions applicable to majority of “green” buildings. The author would be happy if his attempt to conduct this study would inspire others to prove that “green” buildings can not only be beneficial to society thanks to energy conservation and sustainability, but also and above all beneficial to occupants, experiencing less health problems compared to average U.S. office buildings.

The current study showed that human perception of indoor environments was an important factor for occurrence of sick building syndrome and illnesses among building occupants. At the same time, techniques presently used in investigation of indoor environmental health impacts are also typically used for other environments, such as ambient or industrial air. They are increasingly criticized for not being suitable to discover associations between indoor environmental parameters, and consequently, many causes for building occupant health problems remain unclear. Furthermore, a study stipulates that today we do not know much more than we did a century ago (Sundell, 2004). Following such reasoning, new investigation methodologies are necessary in order to bring significant breakthrough in indoor environmental quality research. Possibly, such breakthrough technologies may try to mimic the performance of the most sophisticated sensor currently available, human nose. And the closest device that can be currently used to simulate human
nose preceptors is electronic nose. This device is based upon artificial neural networks (ANN) pattern recognitions implemented to interpret information coming from various sensors (Schreiber, 2000). Considering such ANN applications, ANNs may soon show to be tool of choice in identifying health related impacts of building indoor environments, by way of analyzing multiple building sensor data. It is this goal that the current study most strongly advocates.
Appendix A

Programs Used in NN Data Analyses and Detailed Benchmark Prediction Results

A.1 Preprocessing

The programs used in preprocessing were designed to perform coding and generate automatic coding keys as described in Chapter 4. Fortran programming language was used and three different versions of the program were created to deal with very long symbol sequences that needed to be replaced. Namely, in cases of very long sequences, limits on the storage size of string arrays resulted with coding keys not containing the entire numerically replaced string sequences. Consequently, for example certain long comments of field workers were only partially recorded and program adjustments were made to account for such cases. However, this had impacts on the visual perception and clarity of the generated coding key files, so both, improved and earlier versions of the program were used to preprocess files on a case by case basis.
A.2 Core Software

The core code for multivariate logistic analysis was based upon SAS statistical software. The code was developed in collaboration with the Department of Public Health Sciences, Milton S. Hershey Medical Center, The Pennsylvania State University.

The core code for NN analyses was programmed and executed in Matlab. It included all stages towards the final development as described in Chapter 6. Algorithm of the final code is presented in Chapter 5. Additionally designed functions were used in postprocessing of NN results, specifically to calculate and plot confidence intervals.

Data collected during the experimental study were preprocessed and analyzed using MS Excel. Additionally the data collection was computer based and used a number of proprietary data logging software available from corresponding equipment manufacturers.

A.3 Detailed Neural Network Benchmark Results

For each of the NN output parameters figures similar to those describing cough predictions presented in Chapter 6 were generated. All additional NN performance figures were grouped with respect to targeted NN prediction parameters and presented in alphabetical order within the current section.
Figure A.1. Neural network input scatter plots categorized with respect to probability for occurrence of asthma.

Figure A.2. Neural network training performance for asthma predictions.
Figure A.3. Neural network testing performance for asthma predictions

Figure A.4. Neural network training and testing absolute errors in asthma predictions
Figure A.5. Neural network actual vs. predicted training probabilities for occurrence of asthma

Figure A.6. Neural network actual vs. predicted testing probabilities for occurrence of asthma
Figure A.7. Neural network actual vs. predicted training and testing probabilities for occurrence of asthma

Figure A.8. Confidence ratios for neural network predictions of asthma
Figure A.9. Neural network input scatter plots categorized with respect to BSI values

Figure A.10. Neural network training performance for BSI predictions
Figure A.11. Neural network testing performance for BSI predictions

Figure A.12. Neural network training and testing absolute errors in BSI predictions
Figure A.13. Neural network actual vs. predicted training BSI values

Figure A.14. Neural network actual vs. predicted testing BSI values
Figure A.15. Neural network actual vs. predicted training and testing BSI values

Figure A.16. Confidence ratios for neural network predictions of BSI
Figure A.17. Neural network input scatter plots categorized with respect to probability for occurrence of chest tightness

Figure A.18. Neural network training performance for chest tightness predictions
Figure A.19. Neural network testing performance for chest tightness predictions

Figure A.20. Neural network training and testing absolute errors in chest tightness predictions
Figure A.21. Neural network actual vs. predicted training probabilities for occurrence of chest tightness

Figure A.22. Neural network actual vs. predicted testing probabilities for occurrence of chest tightness
**Figure A.23.** Neural network actual vs. predicted training and testing probabilities for occurrence of chest tightness

**Figure A.24.** Confidence ratios for neural network predictions of chest tightness
Figure A.25. Neural network input scatter plots categorized with respect to probability for occurrence of dust allergy

Figure A.26. Neural network training performance for dust allergy predictions
Figure A.27. Neural network testing performance for dust allergy predictions

Figure A.28. Neural network training and testing absolute errors in dust allergy predictions
Figure A.29. Neural network actual vs. predicted training probabilities for occurrence of dust allergy

Figure A.30. Neural network actual vs. predicted testing probabilities for occurrence of dust allergy
Figure A.31. Neural network actual vs. predicted training and testing probabilities for occurrence of dust allergy

Figure A.32. Confidence ratios for neural network predictions of dust allergy
Figure A.33. Neural network input scatter plots categorized with respect to probability for occurrence of mold allergy

Figure A.34. Neural network training performance for mold allergy predictions
Figure A.35. Neural network testing performance for mold allergy predictions

Figure A.36. Neural network training and testing absolute errors in mold allergy predictions
Figure A.37. Neural network actual vs. predicted training probabilities for occurrence of mold allergy

Figure A.38. Neural network actual vs. predicted testing probabilities for occurrence of mold allergy
Figure A.39. Neural network actual vs. predicted training and testing probabilities for occurrence of mold allergy

Figure A.40. Confidence ratios for neural network predictions of mold allergy
Figure A.41. Neural network input scatter plots categorized with respect to probability for occurrence of shortness of breath.

Figure A.42. Neural network training performance for shortness of breath predictions.
Figure A.43. Neural network testing performance for shortness of breath predictions

Figure A.44. Neural network training and testing absolute errors in shortness of breath predictions
Figure A.45. Neural network actual vs. predicted training probabilities for occurrence of shortness of breath.

Figure A.46. Neural network actual vs. predicted testing probabilities for occurrence of shortness of breath.
Figure A.47. Neural network actual vs. predicted training and testing probabilities for occurrence of shortness of breath

Figure A.48. Confidence ratios for neural network predictions of shortness of breath
Figure A.49. Neural network input scatter plots categorized with respect to probability for occurrence of sinus congestion

Figure A.50. Neural network training performance for sinus congestion predictions
Figure A.51. Neural network testing performance for sinus congestion predictions

Figure A.52. Neural network training and testing absolute errors in sinus congestion predictions
Figure A.53. Neural network actual vs. predicted training probabilities for occurrence of sinus congestion

Figure A.54. Neural network actual vs. predicted testing probabilities for occurrence of sinus congestion
**Figure A.55.** Neural network actual vs. predicted training and testing probabilities for occurrence of sinus congestion

**Figure A.56.** Confidence ratios for neural network predictions of sinus congestion
Figure A.57. Neural network input scatter plots categorized with respect to probability for occurrence of sneezing

Figure A.58. Neural network training performance for sneezing predictions
Figure A.59. Neural network testing performance for sneezing predictions

Figure A.60. Neural network training and testing absolute errors in sneezing predictions
Figure A.61. Neural network actual vs. predicted training probabilities for occurrence of sneezing

Figure A.62. Neural network actual vs. predicted testing probabilities for occurrence of sneezing
Figure A.63. Neural network actual vs. predicted training and testing probabilities for occurrence of sneezing

Figure A.64. Confidence ratios for neural network predictions of sneezing
Figure A.65. Neural network input scatter plots categorized with respect to probability for occurrence of sore throat

Figure A.66. Neural network training performance for sore throat predictions
Figure A.67. Neural network testing performance for sore throat predictions

Figure A.68. Neural network training and testing absolute errors in sore throat predictions
Figure A.69. Neural network actual vs. predicted training probabilities for occurrence of sore throat

Figure A.70. Neural network actual vs. predicted testing probabilities for occurrence of sore throat
Figure A.71. Neural network actual vs. predicted training and testing probabilities for occurrence of sore throat

Figure A.72. Confidence ratios for neural network predictions of sore throat
Figure A.73. Neural network input scatter plots categorized with respect to probability for occurrence of wheezing

Figure A.74. Neural network training performance for wheezing predictions
Figure A.75. Neural network testing performance for wheezing predictions

Figure A.76. Neural network training and testing absolute errors in wheezing predictions
Figure A.77. Neural network actual vs. predicted training probabilities for occurrence of wheezing

Figure A.78. Neural network actual vs. predicted testing probabilities for occurrence of wheezing
Figure A.79. Neural network actual vs. predicted training and testing probabilities for occurrence of wheezing

Figure A.80. Confidence ratios for neural network predictions of wheezing
Chapter 8 described the most significant comparison between averaged data from the Building Assessment Survey and Evaluation, referred to as benchmark building, and experimentally collected data from the PA DEP Cambria “green” office environment, used for NN validation purposes. However, certain comparisons performed on these two building cases were not mentioned in Chapter 8. Therefore, such comparisons will be documented within the current section.
Figure B.1. Validation and benchmark building comparison: aldehyde and VOC concentrations
Figure B.2. Validation and benchmark building comparison: radon concentrations

Figure B.3. Validation and benchmark building comparison: carbon dioxide concentrations
Figure B.4. Validation and benchmark building comparison: indoor noise and light intensity

Figure B.5. Comparison between overall and workday indoor light intensity in the validation building case
Figure B.6. Validation and benchmark building comparison: desk and chair comfort

Figure B.7. Validation and benchmark building comparison: satisfaction with job
Figure B.8. Validation and benchmark building comparison: glare frequency at the workstation

Figure B.9. Validation and benchmark building comparison: workstation location
Figure B.10. Validation and benchmark building comparison: workspace cleanliness

Figure B.11. Validation and benchmark building comparison: occupants’ education level
Figure B.12. Validation and benchmark building comparison: room occupancy

Figure B.13. Validation and benchmark building comparison: windows in work area
Figure B.14. Validation and benchmark building comparison: visibility of windows

Figure B.15. Validation and benchmark building comparison: chemical sensitivity
Figure B.16. Validation and benchmark building comparison: tobacco smoking status

Figure B.17. Validation and benchmark building comparison: type of worn corrective lenses
Figure B.18. Validation and benchmark building comparison: gender

Figure B.19. Validation and benchmark building comparison: occupants’ age
Figure B.20. Validation and benchmark building comparison: workspace change

Figure B.21. Validation and benchmark building comparison: computer usage
Figure B.22. Validation and benchmark building comparison: presence of carpets

Figure B.23. Validation and benchmark building comparison: major responsibilities outside of job
Figure B.24. Validation and benchmark building comparison: photocopier usage

Figure B.25. Validation and benchmark building comparison: laser printer usage
**Figure B.26.** Validation and benchmark building comparison: fax usage

**Figure B.27.** Validation and benchmark building comparison: usage of copy paper
**Figure B.28.** Validation and benchmark building comparison: usage of odorous chemicals

**Figure B.29.** Validation and benchmark building comparison: eye irritation frequency
Figure B.30. Validation and benchmark building comparison: wheezing frequency

Figure B.31. Validation and benchmark building comparison: headache frequency
Figure B.32. Validation and benchmark building comparison: sore throat frequency

Figure B.33. Validation and benchmark building comparison: unusual tiredness frequency
Figure B.34. Validation and benchmark building comparison: chest tightness frequency

Figure B.35. Validation and benchmark building comparison: sinus congestion frequency
Figure B.36. Validation and benchmark building comparison: cough frequency

Figure B.37. Validation and benchmark building comparison: eye tiredness frequency
**Figure B.38.** Validation and benchmark building comparison: tension frequency

**Figure B.39.** Validation and benchmark building comparison: back or shoulder pain frequency
Figure B.40. Validation and benchmark building comparison: sneezing frequency

Figure B.41. Validation and benchmark building comparison: frequency of concentration difficulties
Figure B.42. Validation and benchmark building comparison: dizziness frequency

Figure B.43. Validation and benchmark building comparison: depression frequency
Figure B.44. Validation and benchmark building comparison: shortness of breath frequency

Figure B.45. Validation and benchmark building comparison: nausea frequency
Figure B.46. Validation and benchmark building comparison: dry skin frequency

Figure B.47. Validation and benchmark building comparison: wrist or hand pain frequency
Figure B.48. Validation and benchmark building comparison: dry eye persistence while not in the building

Figure B.49. Validation and benchmark building comparison: wheezing persistence while not in the building
Figure B.50. Validation and benchmark building comparison: headache persistence while not in the building.

Figure B.51. Validation and benchmark building comparison: sore throat persistence while not in the building.
Figure B.52. Validation and benchmark building comparison: unusual tiredness persistence while not in the building

Figure B.53. Validation and benchmark building comparison: chest tightness persistence while not in the building
Figure B.54. Validation and benchmark building comparison: sinus congestion persistence while not in the building.

Figure B.55. Validation and benchmark building comparison: cough persistence while not in the building.
**Figure B.56.** Validation and benchmark building comparison: eye tiredness persistence while not in the building

**Figure B.57.** Validation and benchmark building comparison: tension persistence while not in the building
Figure B.58. Validation and benchmark building comparison: back or shoulder pain persistence while not in the building

Figure B.59. Validation and benchmark building comparison: sneezing persistence while not in the building
Difficulties concentrating, while not at work

Dizziness or lightheadedness, while not at work

Figure B.60. Validation and benchmark building comparison: persistence in concentration difficulties while not in the building

Figure B.61. Validation and benchmark building comparison: dizziness persistence while not in the building
Figure B.62. Validation and benchmark building comparison: persistence of depression while not in the building.

Figure B.63. Validation and benchmark building comparison: shortness of breath persistence while not in the building.
**Figure B.64.** Validation and benchmark building comparison: nausea persistence while not in the building

**Figure B.65.** Validation and benchmark building comparison: dry skin persistence while not in the building
Figure B.66. Validation and benchmark building comparison: wrist or hand pain persistence while not in the building

Figure B.67. Validation and benchmark building comparison: frequency of perceived too much air movement
Figure B.68. Validation and benchmark building comparison: frequency of perceived too little air movement

Figure B.69. Validation and benchmark building comparison: frequency of perceived temperature too high
**Figure B.70.** Validation and benchmark building comparison: frequency of perceived temperature too low

**Figure B.71.** Validation and benchmark building comparison: frequency of perceived air too humid
Figure B.72. Validation and benchmark building comparison: frequency of perceived air too dry

Figure B.73. Validation and benchmark building comparison: frequency of perceived tobacco smoke odors
Figure B.74. Validation and benchmark building comparison: frequency of perceived chemical odors

Figure B.75. Validation and benchmark building comparison: frequency of perceived other unpleasant odors
Figure B.76. Validation and benchmark building comparison: frequency being asked to do things that conflict by a person equal in rank

Figure B.77. Validation and benchmark building comparison: frequency being given to do things that conflict to one another by supervisors
Figure B.78. Validation and benchmark building comparison: frequency being given to do things that conflict to other work by superiors

Figure B.79. Validation and benchmark building comparison: frequency job requires very fast work
**Figure B.80.** Validation and benchmark building comparison: frequency job requires very hard work

**Figure B.81.** Validation and benchmark building comparison: frequency a little time left to do things
**Figure B.82.** Validation and benchmark building comparison: a great deal to be done frequency

**Figure B.83.** Validation and benchmark building comparison: clear about what others expect frequency
Figure B.84. Validation and benchmark building comparison: clear on job responsibilities frequency

Figure B.85. Validation and benchmark building comparison: able to predict what others will expect frequency
Figure B.86. Validation and benchmark building comparison: frequency of work objectives being well defined

Figure B.87. Validation and benchmark building comparison: working years in the building
Figure B.88. Validation and benchmark building comparison: weekly working hours in the building

Figure B.89. Validation and benchmark building comparison: daily computer usage time
Detailed Neural Network Validation Results

Neural networks applied in the validation building case study had 6 rather than 7 input neurons used in the benchmark case analyses, as described in Chapter 8. The current section contains detailed NN training and testing results using such reduced number of BASE building inputs. These networks were further used to make predictions in the validation case, “green” office building. As scatter plots of all considered input parameters were already presented within Appendix A, they will be omitted from the current figures. The figures are grouped and presented in alphabetical order with respect to targeted NN prediction parameters.
Figure C.1. Neural network training performance for asthma predictions

Figure C.2. Neural network testing performance for asthma predictions
Figure C.3. Neural network training and testing absolute errors in asthma predictions

Figure C.4. Neural network actual vs. predicted training probabilities for occurrence of asthma
Figure C.5. Neural network actual vs. predicted testing probabilities for occurrence of asthma

![Plot](image)

Best Linear Fit: $A = (0.312) T + (0.0764)$

Figure C.6. Neural network actual vs. predicted training and testing probabilities for occurrence of asthma

![Plot](image)

Best Linear Fit: $A = (0.359) T + (0.078)$

$R = 0.677$
Figure C.7. Confidence ratios for neural network predictions of asthma

Figure C.8. Neural network training performance for BSI predictions
Figure C.9. Neural network testing performance for BSI predictions

Figure C.10. Neural network training and testing absolute errors in BSI predictions
Figure C.11. Neural network actual vs. predicted training BSI values

Best Linear Fit: \( A = (0.413) T + (0.976) \)

R = 0.691

Figure C.12. Neural network actual vs. predicted testing BSI values

Best Linear Fit: \( A = (0.524) T + (0.657) \)

R = 0.807
Figure C.13. Neural network actual vs. predicted training and testing BSI values

Figure C.14. Confidence ratios for neural network predictions of BSI
Figure C.15. Neural network training performance for chest tightness predictions

Figure C.16. Neural network testing performance for chest tightness predictions
Figure C.17. Neural network training and testing absolute errors in chest tightness predictions

Figure C.18. Neural network actual vs. predicted training probabilities for occurrence of chest tightness
Figure C.19. Neural network actual vs. predicted testing probabilities for occurrence of chest tightness

Figure C.20. Neural network actual vs. predicted training and testing probabilities for occurrence of chest tightness
Figure C.21. Confidence ratios for neural network predictions of chest tightness

Figure C.22. Neural network training performance for dust allergy predictions
Figure C.23. Neural network testing performance for dust allergy predictions

Figure C.24. Neural network training and testing absolute errors in dust allergy predictions
Figure C.25. Neural network actual vs. predicted training probabilities for occurrence of dust allergy

```
Best Linear Fit: A = (0.323) T + (0.216)
```

Figure C.26. Neural network actual vs. predicted testing probabilities for occurrence of dust allergy

```
Best Linear Fit: A = (-0.435) T + (0.471)
```
Figure C.27. Neural network actual vs. predicted training and testing probabilities for occurrence of dust allergy

Figure C.28. Confidence ratios for neural network predictions of dust allergy
Figure C.29. Neural network training performance for mold allergy predictions

Figure C.30. Neural network testing performance for mold allergy predictions
Figure C.31. Neural network training and testing absolute errors in mold allergy predictions

Figure C.32. Neural network actual vs. predicted training probabilities for occurrence of mold allergy
**Figure C.33.** Neural network actual vs. predicted testing probabilities for occurrence of mold allergy

**Figure C.34.** Neural network actual vs. predicted training and testing probabilities for occurrence of mold allergy
Figure C.35. Confidence ratios for neural network predictions of mold allergy

Figure C.36. Neural network training performance for shortness of breath predictions
Figure C.37. Neural network testing performance for shortness of breath predictions

Figure C.38. Neural network training and testing absolute errors in shortness of breath predictions
Figure C.39. Neural network actual vs. predicted training probabilities for occurrence of shortness of breath

Figure C.40. Neural network actual vs. predicted testing probabilities for occurrence of shortness of breath
Figure C.41. Neural network actual vs. predicted training and testing probabilities for occurrence of shortness of breath

Figure C.42. Confidence ratios for neural network predictions of shortness of breath
Figure C.43. Neural network training performance for sinus congestion predictions

Figure C.44. Neural network testing performance for sinus congestion predictions
Figure C.45. Neural network training and testing absolute errors in sinus congestion predictions

Figure C.46. Neural network actual vs. predicted training probabilities for occurrence of sinus congestion
Figure C.47. Neural network actual vs. predicted testing probabilities for occurrence of sinus congestion

Figure C.48. Neural network actual vs. predicted training and testing probabilities for occurrence of sinus congestion
Figure C.49. Confidence ratios for neural network predictions of sinus congestion

Figure C.50. Neural network training performance for sneezing predictions
Figure C.51. Neural network testing performance for sneezing predictions

Figure C.52. Neural network training and testing absolute errors in sneezing predictions
Figure C.53. Neural network actual vs. predicted training probabilities for occurrence of sneezing

Figure C.54. Neural network actual vs. predicted testing probabilities for occurrence of sneezing
Figure C.55. Neural network actual vs. predicted training and testing probabilities for occurrence of sneezing

Figure C.56. Confidence ratios for neural network predictions of sneezing
Figure C.57. Neural network training performance for sore throat predictions

Figure C.58. Neural network testing performance for sore throat predictions
Figure C.59. Neural network training and testing absolute errors in sore throat predictions

Figure C.60. Neural network actual vs. predicted training probabilities for occurrence of sore throat
Figure C.61. Neural network actual vs. predicted testing probabilities for occurrence of sore throat

Figure C.62. Neural network actual vs. predicted training and testing probabilities for occurrence of sore throat
Figure C.63. Confidence ratios for neural network predictions of sore throat

Figure C.64. Neural network training performance for wheezing predictions
Figure C.65. Neural network testing performance for wheezing predictions

Figure C.66. Neural network training and testing absolute errors in wheezing predictions
Figure C.67. Neural network actual vs. predicted training probabilities for occurrence of wheezing

Figure C.68. Neural network actual vs. predicted testing probabilities for occurrence of wheezing
Figure C.69. Neural network actual vs. predicted training and testing probabilities for occurrence of wheezing

Figure C.70. Confidence ratios for neural network predictions of wheezing
Appendix D

Building Occupants Questionnaire

The current Appendix contains printout of an internet based questionnaire used to conduct survey among occupants in the validation case PA DEP Cambria “green” building. The electronic questionnaire forms were developed following the original BASE study protocol (U.S. EPA, 2003), as well as the requirements of the Pennsylvania State University Office for Research Protections regarding research involving human participants (PSU ORP, 2009). The current printout does not contain Implied consent form, which participants had to accept before being able to fill in the given questionnaire. Invitation to participate in the validation study occupants received electronically on Monday of the study week, together with an appropriate link to the implied consent form followed by the questionnaire. Considering the nature of the electronic data collection, occupants were given 7 days to fill-in the questionnaire following the provided electronic link. After this period, questionnaire was no longer accessible to the occupants. Such procedure presented modification to the original BASE protocol, which specified questionnaire data collection from the occupants during a single day. However, the electronic questionnaire data collection significantly reduced the time necessary for data processing and analyses.
I. WORKPLACE INFORMATION

This survey is being conducted to determine the environmental quality of your building. This questionnaire asks about how you think your building environment and your work affect you. Please answer the questions as accurately and completely as you can, regardless of how satisfied or dissatisfied you are with conditions in the building.

ALL OF YOUR ANSWERS WILL BE TREATED IN THE STRICTEST CONFIDENCE.

1. How long have you worked in this building, to the nearest year?

1a. If less than one year, how many months have you worked in this building?

2. On average, how many hours a week do you work in this building?

3. During THIS WEEK, including today, how many days did you work in this building?

4. Which best describes the space in which your current workstation* is located?

*For this questionnaire, your "workstation" is the place (desk, cubicle, office, etc.) where you do the majority of your work.

   j1: Single person private office (1)
   j2: Shared private office (2)
   j3: Open space with partitions (3)
   j4: Open space without partitions (4)
   j5: Other (5) (please specify)

4a. How many people work in the room in which your workstation is located (including yourself)?
5. Is there carpet on most or all of the floor at your workstation?
   - Yes (1)
   - No (2)

6. In general, how clean is your workspace* area?
   *For this questionnaire, your "workspace" is the immediate area surrounding your workstation
   - Very clean (1)
   - Reasonably clean (2)
   - Somewhat dusty or dirty (3)
   - Very dusty or dirty (4)

7. Please rate the lighting at your workstation.
   - Much too dim (1)
   - A little too dim (2)
   - Just right (3)
   - A little too bright (4)
   - Much too bright (5)

8. Do you experience a reflection or "glare" in your field of vision when at your workstation?
   - Rarely (1)
   - Occasionally (2)
   - Sometimes (3)
   - Fairly often (4)
   - Very often (5)

9. How comfortable is the chair at your workstation?
   - Very comfortable (1)
   - Reasonably comfortable (2)
   - Somewhat uncomfortable (3)
10. How comfortable is the current set-up of your desk or work table (i.e., height and general arrangement of the table, chair, and equipment you work with)?

- Very comfortable (1)
- Reasonably comfortable (2)
- Somewhat uncomfortable (3)
- Very uncomfortable (4)
- Don't have one specific desk or work table (5)

11. Do you work with a computer or word processor?

- Yes (1)
- No (2)

11a. About how many hours a day do you work with a computer or word processor, to the nearest hour?

11b. If you use a computer or word processor, do you usually wear glasses when you use these machines?

- Yes (1)
- No (2)

11c. Do you use a glare screen on your computer?

- Yes (1)
- No (2)
### 12. Which one of the following statements best describes the windows in your work area?

- [ ] There are no windows in my personal workspace and none in the general area visible from my workspace (when I am either standing or seated). (1)
- [ ] There are no windows in my personal workspace, but I can see one or more windows in the general area. (2)
- [ ] There are one or more windows in my personal workspace. (3)

### 13. If there is a window visible from your workspace, how far (in feet) is the closest window from your desk chair?

### 14. During the PAST THREE MONTHS, have the following changes taken place within 15 feet of your current workstation?

<table>
<thead>
<tr>
<th>Change</th>
<th>YES (1)</th>
<th>NO (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>New carpeting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walls painted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New furniture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New partitions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New wall covering</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water damage</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 15. How often do you use the following at work? (Check the appropriate box for each item.)

<table>
<thead>
<tr>
<th>Item</th>
<th>Several times a day (1)</th>
<th>About once a day (2)</th>
<th>3-4 times a week (3)</th>
<th>Less than 3 times/week (4)</th>
<th>Never (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photocopier</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laser printer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facsimile (FAX) machine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-copying (carbonless) copy paper</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cleanser, glue, correction fluid, or other odorous chemicals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
II. INFORMATION ABOUT HEALTH AND WELL-BEING

1. Have you ever been told by a doctor that you have or had any of the following?

<table>
<thead>
<tr>
<th>Condition</th>
<th>YES (1)</th>
<th>NO (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Migraine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asthma</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eczema</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hay fever</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allergy to dust</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allergy to molds</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. What is your tobacco smoking status?

- never smoked (1)
- former smoker (2)
- current smoker (3)

3. Do you consider yourself especially sensitive to the presence of tobacco smoke in your workspace?

- Yes (1)
- No (2)

4. Do you consider yourself especially sensitive to the presence of other chemicals in the air of your workspace?

- Yes (1)
- No (2)

5. What type of corrective lenses do you usually wear at work?

- none (1)
- glasses (2)
- bifocals (3)
- contact lenses (4)

6. How old were you on your last birthday?

- under 20 (1)
- 20-29 years (2)
- 30-39 years (3)
This page contains questions regarding symptoms you may have experienced while at work during the last 4 weeks.

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Not in Last 4 Weeks</th>
<th>In Last 4 Weeks</th>
<th>Most of Last 4 Weeks</th>
<th>Everyday or Daily</th>
</tr>
</thead>
<tbody>
<tr>
<td>dry, itching, or irritated eyes</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>wheezing</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>headache</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>sore or dry throat</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>unusual tiredness, fatigue, or drowsiness</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>chest tightness</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>stuffy or runny nose, or sinus congestion</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>cough</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>tired or strained eyes</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>tension, irritability, or nervousness</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>pain or stiffness in back, shoulders, or neck</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>sneezing</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>difficulty remembering things or concentrating</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>dizziness or lightheadedness</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>feeling depressed</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>shortness of breath</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>nausea or upset stomach</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>dry or itchy skin</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>
numbness in hands or wrists

9a. In the LAST FOUR WEEKS how often have any of the symptoms listed above reduced your ability to work?

9b. In the LAST FOUR WEEKS how often have any of the symptoms listed above caused you to stay home or leave work?

III. DESCRIPTION OF WORKPLACE CONDITIONS

This page contains questions regarding environmental conditions you may have experienced while at work during the last 4 weeks.

1. During the LAST FOUR WEEKS YOU WERE AT WORK, how often have you experienced each of the following environmental conditions while working in this building? If you select “Not in Last 4 Weeks” (1) – move DOWN the page to the next condition. If you select (2),(3), or (4) move across the page.

- too much air movement
- too little air movement
- temperature too hot
- temperature too cold
- air too humid
- air too dry
- tobacco smoke odors
- unpleasant chemical odors
- other unpleasant odors (e.g., body odor, food odor, perfume)

How satisfied are you with the following aspects of your workstation?

<table>
<thead>
<tr>
<th></th>
<th>Very satisfied (1)</th>
<th>Somewhat satisfied (2)</th>
<th>Not too satisfied (3)</th>
<th>Not at all satisfied (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Conversational privacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Freedom from distracting noise</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
IV. CHARACTERISTICS OF YOUR JOB

1. All in all, how satisfied are you with your job?
   - Very satisfied (1)
   - Somewhat satisfied (2)
   - Not too satisfied (3)
   - Not at all satisfied (4)

2. What is the highest level you completed in school?
   - 8th grade or less (1)
   - Some high school (2)
   - High school graduate (3)
   - Some college (4)
   - College degree (5)
   - Graduate degree (6)

3. Conflicts can occur in any job. For example, someone may ask you to do work in a way that is different from what you think best, or you may find that it is difficult to satisfy everyone.
   HOW OFTEN do you face problems in your work like the ones listed below?
   (Check the appropriate box for each statement.)

<table>
<thead>
<tr>
<th></th>
<th>Rarely or Never (1)</th>
<th>Sometimes (2)</th>
<th>Fairly Often (3)</th>
<th>Very Often (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persons equal in rank and authority over you ask you to do things which conflict</td>
<td>jn</td>
<td>jn</td>
<td>jn</td>
<td>jn</td>
</tr>
<tr>
<td>People in a good position to see if you do what they ask give you things to do which conflict with one another</td>
<td>jn</td>
<td>jn</td>
<td>jn</td>
<td>jn</td>
</tr>
<tr>
<td>People whose requests should be met give you things which conflict with other work you have to do</td>
<td>jn</td>
<td>jn</td>
<td>jn</td>
<td>jn</td>
</tr>
</tbody>
</table>
4. The next series of questions asks HOW OFTEN certain things happen at your job.

(Check the appropriate box for each question.)

<table>
<thead>
<tr>
<th>How often does your job require you to work very fast?</th>
<th>Rarely (1)</th>
<th>Occasionally (2)</th>
<th>Sometimes (3)</th>
<th>Fairly Often (4)</th>
<th>Very Often (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>How often does your job require you to work very hard?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How often does your job leave you with little time to get things done?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How often is there a great deal to be done?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How often are you clear on what your job responsibilities are?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How much of the time are your work objectives well defined?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How often are you clear about what others expect of you on the job?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. In order to better understand your responsibilities outside your normal working day, the next series of questions deals with other significant aspects of your life.

<table>
<thead>
<tr>
<th></th>
<th>YES (1)</th>
<th>NO (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major responsibility for child care duties</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major responsibility for housekeeping duties</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major responsibility for care of an elderly or disabled person on a regular basis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular commitment of five hours or more per week, paid or unpaid, outside of this job (include educational courses, volunteer work, second job, etc.)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
PLEASE USE THIS SPACE TO DISCUSS ANY ASPECTS OF THE BUILDING ENVIRONMENT OR EMPLOYEE HEALTH THAT YOU FEEL APPROPRIATE
Bibliography


Apte, M. G. and Erdmann, C. A.: Indoor carbon dioxide concentrations, VOCs, environmental sensitivity association with mucous membrane and lower respiratory sick building syndrome symptoms in the BASE study: analyses of the 100 building dataset. LBNL Report 51570, 2002


Buchanan, I. and Apte, M. G.: Air filter materials and building related symptoms in the 100-building BASE study, LBNL Report 59663, 2006


Larose, D. T.: Data Mining Methods and Models, John Wiley and Sons, Inc., 2006


Mumma, S. A.: Underfloor Air Distribution (UFAD) Experiences Based upon 10 Million Square Feet of Installations, personal communication, April, 2009

NIOSH: Guidance for protecting building environments from airborne chemical, biological or radiological attacks, Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health (NIOSH), May 2002


PA DEP: Furnishing high performance green buildings, Pennsylvania Department of Environmental Protection (PA DEP), 2003


Sundell, J. and Lindvall, T.: Indoor air humidity and sensation of dryness as risk indicators of SBS, Indoor Air 3, pp. 382-390, 1993


Tsai, F. C. and Macher, J. M.: Concentrations of airborne culturable bacteria in 100 large US office buildings from the BASE study, Indoor Air 15 (Suppl 9), pp. 71-81, 2005


U.S. Environmental Protection Agency (EPA), Indoor Environments Division, Office of Radiation and Indoor Air: A Standardized EPA Protocol for Characterizing Indoor Air Quality in Large Office Buildings, Washington, DC, February 2003


Vita

Vladimir Vuković

Vladimir Vukovic has an interdisciplinary background in the field of numerical simulations. Majoring in Aeronautics, he did his undergraduate and Diploma thesis work at the Department of Mechanical Engineering, University of Belgrade, Serbia. His diploma thesis focused on computational fluid dynamics simulation of fluid flow and heat exchange processes. Prior to starting his graduate studies at Penn State, he briefly attended graduate studies in Fluid Mechanics, Department of Mathematics, University of Belgrade, Serbia. Vladimir Vukovic graduated from the M.Sc. program at the Department of Architectural Engineering, The Pennsylvania State University, focusing on building mechanical systems, indoor air quality and innovative computer simulations, applied to determine indoor pollutant source location in incidental situations. He continued his research at the Department of Architectural Engineering, Penn State, and during his Ph. D. studies, additionally enrolling in a Computational Science Ph.D. minor. His work in the area of architectural engineering has been awarded by the American Society of Heating, Refrigeration and Air-conditioning Engineers (ASHRAE), Air and Waste Management Association (A&WMA) and the Pennsylvania State University. Vladimir Vukovic is author or co-author of 9 research publications, member of ASHRAE, A&WMA, and the International Society of Indoor Air Quality and Climate (ISIAQ). He is currently working as a Project Manager of Strategic Research in Sustainable Building Technologies at the Energy Department of the Austrian Institute of Technology, Vienna, Austria.