

Table 1

Overall SH features used to predict IAQ1 variables. We report the classifier that was used and the number of IAQ variables that are predicted with at least a moderate effect ( $r \geq .3$ ).

Method	number of $r \geq .3$	Total number	Percentage
Random Forest	48	51	94%
Linear Regression	41	51	80%
Support Vector Regression	42	51	82%

Table 2

Overall SH features predict IAQ2 variables

Method	number of $r \geq .3$	Total number	Percentage
Random Forest	50	51	98%
Linear Regression	39	51	76%
Support Vector Regression	30	51	59%

Table 3

Overall SH features predict IAQ1\_2 variables

Method	number of $r \geq .3$	Total number	Percentage
Random Forest	50	51	98%
Linear Regression	31	51	60%
Support Vector Regression	27	51	53%

Table 4

Each IAQ variable that predicted by the complete set of SH by RF  
in the aggregated dataset IAQ1\_2.

Variable	Correlation	Correlation
	Inside	Outside
<b>Higher Correlation inside than outside</b>		
C <sub>3</sub> -benzenes	0.9554	0.3462
C <sub>2</sub> -benzenes	0.9537	0.5457
Temperature	0.9462	0.883
CH <sub>4</sub>	0.9334	NA
methanol	0.9265	0.555
formaldehyde	0.9061	0.5407
methyl ethyl ketone	0.8995	0.6076
methyl vinyl ketone	0.8985	0.5954
styrene	0.895	0.6155
toluene	0.8894	0.218
acetone	0.8779	0.5295
benzene	0.8608	0.5598
CO <sub>2</sub>	0.8465	0.8386
isoprene	0.8338	0.5748
H <sub>2</sub> O (water vapor)	0.8276	0.6539
O <sub>3</sub>	0.8178	0.7971
acetonitrile	0.7706	0.5988
small particle count	0.4471	NA
large particle count	0.4253	NA
<b>Higher Correlation outside than inside</b>		
α-pinene fragment	0.6723	0.7495
C <sub>4</sub> -benzenes	0.502	0.5299
particulate matter	0.4808	0.5121
acetaldehyde	0.4313	0.5536
α-pinene	0.3225	0.6151
wind speed by weather station	NA	0.7596
wind direction about true north	NA	0.7577
pressure	NA	0.733
relative humidity	NA	0.842

