

# **CART SUPPLEMENTARY REPORT (SO<sub>2</sub>, NO, NO<sub>2</sub>)**

Akhil Kadiyala

Ashok Kumar

Department of Civil Engineering



April 28, 2012

## Table of Contents

Overview of Research Study .....	3
A: SULFUR DIOXIDE (SO <sub>2</sub> ) Regression Tree Details (Complete Database) .....	5
Table A1: SO <sub>2</sub> Complete Database – CART Variable Relative Importance .....	5
Table A2: SO <sub>2</sub> Complete Database Tree Summary Report .....	6
Table A3: SO <sub>2</sub> Complete Database Tree Sequence .....	7
Table A4: SO <sub>2</sub> Complete Database Tree Node Information .....	8
Table A5: SO <sub>2</sub> Complete Database Terminal Node Information .....	14
B: NITRIC OXIDE (NO) Regression Tree Details (Complete Database).....	15
Table B1: NO Complete Database – CART Variable Relative Importance.....	15
Table B2: NO Complete Database Tree Summary Report.....	16
Table B3: NO Complete Database Tree Sequence.....	17
Table B4: NO Complete Database Tree Node Information.....	18
Table B5: NO Complete Database Terminal Node Information.....	21
C: NITROGEN DIOXIDE (NO <sub>2</sub> ) Regression Tree Details (Complete Database).....	22
Table C1: NO <sub>2</sub> Complete Database – CART Variable Relative Importance.....	22
Table C2: NO <sub>2</sub> Complete Database Tree Summary Report.....	22
Table C3: NO <sub>2</sub> Complete Database Tree Sequence.....	23
Table C4: NO <sub>2</sub> Complete Database Tree Node Information.....	24
Table C5: NO <sub>2</sub> Complete Database Terminal Node Information.....	26
D: Validation Results .....	27
Table D1: Relative Importance of the Variables for In-Vehicle SO <sub>2</sub> Obtained from CART Runs .....	27
Table D2: Sensitivity Results for In-Vehicle SO <sub>2</sub> Obtained from the ANOVA.....	28
Table D3: Relative Importance of the Variables for In-Vehicle NO Obtained from CART Runs .....	29
Table D4: Sensitivity Results for In-Vehicle NO Obtained from the ANOVA .....	30
Table D5: Relative Importance of the Variables for In-Vehicle NO <sub>2</sub> Obtained from CART Runs.....	31
Table D6: Sensitivity Results for In-Vehicle NO <sub>2</sub> Obtained from the ANOVA .....	31

## Overview of Research Study

- **Instrumentation**

- YES PLUS: Gaseous pollutants, Indoor Temp., and Indoor RH

- **Pollutants Monitored**

- Gaseous pollutants (ppm): NO, NO<sub>2</sub>, and SO<sub>2</sub>

- **Variables Considered**

- Meteorological variables obtained from ULCD NCDC NOAA: Ambient Temp, Ambient MR (computed from Ambient RH) , Wind Speed, Sky Condition, Visibility, Weather Type, and Precipitation
- Indoor Temp. and Indoor MR (computed from Indoor RH) monitored using YES PLUS instrument
- On-Road variables monitored from real-time video monitoring: Passengers, Light Vehicles (Cars/SUVs), Heavy Vehicles (Buses/Trucks), Ventilation Settings (Run/Close, Idle/Close, Idle/Open)
- Time of the day
- Month and Season of the year

- **Definitions for Terms used in this Document**

- Complete Database: Refers to the database having no missing data points for one or more of the variables considered. Such a database had 1453 hourly data points.
- CART<sup>®</sup> software issued by Salford Systems in Salford Predictive Modeling Suite (SPM) has been used in developing the regression trees, and then identify the influential factors affecting vehicular indoor air quality (IAQ).


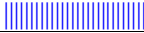










**Codes Used in CART Runs**

<b>Time Code for CART Results</b>	<b>Actual Time</b>
<b>0.250208</b>	<b>6AM</b>
<b>0.291875</b>	<b>7AM</b>
<b>0.333542</b>	<b>8AM</b>
<b>0.375208</b>	<b>9AM</b>
<b>0.416875</b>	<b>10AM</b>
<b>0.458542</b>	<b>11AM</b>
<b>0.500208</b>	<b>12PM</b>
<b>0.541875</b>	<b>1PM</b>
<b>0.583542</b>	<b>2PM</b>
<b>0.625208</b>	<b>3PM</b>
<b>0.666875</b>	<b>4PM</b>
<b>0.708542</b>	<b>5PM</b>
<b>0.750208</b>	<b>6PM</b>
<b>0.791875</b>	<b>7PM</b>
<b>0.833542</b>	<b>8PM</b>
<b>0.875208</b>	<b>9PM</b>
<b>0.916875</b>	<b>10PM</b>
<b>0.958542</b>	<b>11PM</b>

**\*Note:** This report also presents the particulate matter validation results for the two-step approach that involved using the regression trees and the analysis of variance (ANOVA).

## A: SULFUR DIOXIDE (SO2) Regression Tree Details (Complete Database)

**Table A1: SO2 Complete Database – CART Variable Relative Importance**

Variable	Score	
MONTH	100.00	
SKY CONDITION	68.09	
AMBIENT TEMP.	47.16	
PRECIPITATION	33.45	
HEAVY VEHICLES	31.57	
INDOOR TEMP.	28.95	
INDOOR MR	28.95	
RUN CLOSE	25.81	
WEATHER TYPE	23.03	
AMBIENT MR	15.88	
SEASON	14.90	
LIGHT VEHICLES	7.65	
IDLE CLOSE	0.27	

**Table A2: SO2 Complete Database Tree Summary Report**

	Competitor	Split	Improvement	N Left	N Right	N Missing
Main	MONTH	April,	0.00032	962	457	0
		January,				
		July,				
		June,May,				
		November,				
		September				
1	SKY CONDITION	CLR,FEW,	0.00034	21	1398	0
		OVC				
2	PRECIPITATION	0.07500	0.00017	1417	2	0
3	INDOOR MR	30.58180	0.00015	1371	48	0
4	INDOOR TEMP.	89.33508	0.00015	1374	45	0
5	RUN CLOSE	59.50242	0.00014	1393	26	0
6	LIGHT VEHICLES	0.00833	0.00011	217	1202	0
7	SEASON	Spring,	0.00010	583	836	0
		Winter				
8	AMBIENT TEMP.	62.50000	0.00009	557	862	0
9	WEATHER TYPE	FREEZING RAIN MIST,	0.00007	83	1336	0
		HAZE,				
		RAIN,				
		SNOW,				
		THUNDERSTORM MIST,				
		UNKNOWN PRECIPITATION MIST				
10	HEAVY VEHICLES	0.00833	0.00007	303	1116	0
11	IDLE CLOSE	0.93854	0.00005	93	1326	0
12	AMBIENT MR	4.09260	0.00005	207	1212	0

**Table A3: SO2 Complete Database Tree Sequence**

=====  
Tree Sequence – SO2 Complete Tree  
=====

Dependent variable: SO2\_\_PPM\_

Terminal Tree Nodes	Cross-Validated Relative Error	Resubstitution Relative Error	Complexity Parameter	Relative Complexity
1	308	1.20485 +/- 0.09493	0.26773	0.00000
214	22	1.03281 +/- 0.07880	0.69518	0.087812
215	18	1.00873 +/- 0.07540	0.73454	0.119281
216	14	0.98435 +/- 0.07317	0.77568	0.124640
217	13	0.99095 +/- 0.07308	0.78659	0.132231
218**	10	0.97925 +/- 0.07000	0.82239	0.144633
219	5	1.00343 +/- 0.06932	0.89683	0.180465
220	4	0.99954 +/- 0.06899	0.91254	0.190326
221	3	0.99734 +/- 0.06900	0.92916	0.201463
222	2	1.06516 +/- 0.06815	0.96303	0.410542
223	1	1.00001 +/- 0.00007	1.00000	0.448011

Initial mean = 0.15062

Initial variance = 0.00854

**Table A4: SO2 Complete Database Tree Node Information**

=====

NODE INFORMATION – SO2 Complete Tree

=====

```

*****
*                               *
*           Node 1: MONTH$      *
*           N: 1419             *
*                               *
*****

```

```

*****                               *****
*           Node 2                * *           Node 8                *
*           N: 962                 * *           N: 457                 *
*                               * *                               *
*****                               *****

```

Node 1 was split on MONTH\$  
A case goes left if MONTH\$ = ("April", "January", "July", "June",  
"May", "November", "September")

Improvement = 0.000316            Complexity Threshold = 0.448001

Node	Cases	Wgt	Count	Mean	StdDev
1	1419		1419.00	0.15062	0.09242
2	962		962.00	0.13838	0.08888
8	457		457.00	0.17640	0.09439

Surrogate	Split	Assoc	Improvement
1 SEASON\$	s "Fall", "Summer"	0.30197	0.00005
2 AMB_MIXING_RATIO	r 4.38582	0.09628	0.00002
3 AMBIENT_TEMP_F	r 28.50000	0.02407	0.00000
4 WEATHER_TYPE\$	s "HAZE", "Normal", "THUNDERSTORM MIST"	0.01751	0.00000
5 PRECIPITATION	s 0.00250	0.01751	0.00000

Competitor	Split	Improvement
1 SKY_CONDITION\$	"CLR", "FEW", "OVC"	0.00034
2 PRECIPITATION	0.07500	0.00017
3 INDOOR_MIXING_RATIO	30.58180	0.00015
4 INDOOR_TEMP_F	89.33508	0.00015
5 RUN_CLOSE	59.50242	0.00014

```

*****
*                               *
*           Node 2: SKY_CONDITION$      *
*           N: 962                     *
*                               *
*****

```



\*\*\*\*\*

```

=====
= Terminal Node 1 = * Node 3 *
= N: 21 = * N: 941 *
=====
*****

```

Node 2 was split on SKY\_CONDITION\$  
 A case goes left if SKY\_CONDITION\$ = ("CLR", "FEW", "OVC")  
 Improvement = 0.000289 Complexity Threshold = 0.410532

Node	Cases	Wgt Count	Mean	StdDev
2	962	962.00	0.13838	0.08888
-1	21	21.00	0.00010	0.00039
3	941	941.00	0.14146	0.08741

Competitor	Split	Improvement
1 CARS	0.00833	0.00014
2 INDOOR_TEMP__F_	64.22950	0.00013
3 INDOOR_MIXING_RATIO	13.15257	0.00013
4 RUN_CLOSE	59.50242	0.00011
5 TRUCKS	0.00833	0.00010

```

*****
* Node 3: INDOOR_TEMP__F_ *
* N: 941 *
*****

```

```

=====
= Terminal Node 2 = * Node 4 *
= N: 23 = * N: 918 *
=====
*****

```

Node 3 was split on INDOOR\_TEMP\_\_F\_  
 A case goes left if INDOOR\_TEMP\_\_F\_ <= 64.22950  
 Improvement = 0.000123 Complexity Threshold = 0.180455

Node	Cases	Wgt Count	Mean	StdDev
3	941	941.00	0.14146	0.08741
-2	23	23.00	0.22751	0.08451
4	918	918.00	0.13931	0.08639

Surrogate	Split	Assoc	Improvement
1 INDOOR_MIXING_RATIO	13.15257	1.00000	0.00012

Competitor	Split	Improvement
1 INDOOR_MIXING_RATIO	13.15257	0.00012
2 RUN_CLOSE	59.50242	0.00012
3 WEATHER_TYPE\$	"FOG", "FREEZING RAIN MIST", "MIST", "RAIN", "SNOW", "THUNDERSTORM MIST", "THUNDERSTORM RAIN MIST"	0.00008
4 IDLE_CLOSE	0.93854	0.00007

5 AMB\_\_MIXING\_RATIO

4.79021

0.00007

```

*****
*           Node 4: RUN_CLOSE           *
*           N: 918                       *
*****

```

```

*****
*           Node 5                       * = Terminal Node 7 =
*           N: 893                       * =           N: 25   =
*****
=====

```

Node 4 was split on RUN\_CLOSE  
 A case goes left if RUN\_CLOSE <= 59.50242  
 Improvement = 0.000110      Complexity Threshold = 0.181935

Node	Cases	Wgt Count	Mean	StdDev
4	918	918.00	0.13931	0.08639
5	893	893.00	0.14148	0.08590
-7	25	25.00	0.06149	0.06486

Competitor	Split	Improvement
1 WEATHER_TYPE\$	"FREEZING RAIN MIST", "MIST", "RAIN", "RAIN MIST", "SNOW", "THUNDERSTORM MIST", "THUNDERSTORM RAIN", "THUNDERSTORM RAIN MIST", "UNKNOWN PRECIPITATION", "UNKNOWN PRECIPITATION MIST"	0.00009
2 MONTH\$	"July"	0.00008
3 INDOOR_TEMP__F_	87.89691	0.00007
4 INDOOR_MIXING_RATIO	29.69079	0.00007
5 AMB__MIXING_RATIO	4.79021	0.00007

```

*****
*           Node 5: WEATHER_TYPE$       *
*           N: 893                       *
*****

```

```

=====
= Terminal Node 3 = * Node 6 *
= N: 47           = * N: 846 *
=====
*****

```

Node 5 was split on WEATHER\_TYPE\$  
 A case goes left if WEATHER\_TYPE\$ = ("FREEZING RAIN MIST", "HAZE",  
 "RAIN", "SNOW", "THUNDERSTORM MIST", "THUNDERSTORM RAIN MIST")

Improvement = 0.000098      Complexity Threshold = 0.190706

Node	Cases	Wgt Count	Mean	StdDev
5	893	893.00	0.14148	0.08590
-3	47	47.00	0.08865	0.07147

6            846            846.00            0.14442            0.08569

Competitor	Split	Improvement
1 MONTH\$	"April", "January", "July", "June", "May", "November", "September"	0.00010
2 AMB__MIXING_RATIO	4.79021	0.00008
3 INDOOR_TEMP__F_	87.89691	0.00006
4 INDOOR_MIXING_RATIO	29.69079	0.00006
5 AMBIENT_TEMP__F_	88.50000	0.00006

```
*****
*                               *
*           Node 6: MONTH$      *
*                               *
*           N: 846              *
*                               *
*****
```

```
*****
*                               *
*           Node 7              *
*                               *
*           N: 358              *
*                               *
*****
=====
*           Node 7              *   =   Terminal Node 6   =
*           N: 358              *   =   N: 488             =
*                               *
*****
=====
```

Node 6 was split on MONTH\$  
 A case goes left if MONTH\$ = ("January", "May", "September")  
 Improvement = 0.000109            Complexity Threshold = 0.216806

Node	Cases	Wgt	Count	Mean	StdDev
6	846		846.00	0.14442	0.08569
7	358		358.00	0.12863	0.09458
-6	488		488.00	0.15601	0.07647

Surrogate	Split	Assoc	Improvement
1 SEASON\$	s "Fall", "Spring", "Summer", "Winter"	0.38547	0.00001
2 AMB__MIXING_RATIO	s 3.77915	0.21788	0.00005
3 AMBIENT_TEMP__F_	s 57.50000	0.18156	0.00000
4 IDLE_CLOSE	r 13.45455	0.12011	0.00000
5 CARS	r 0.65000	0.10335	0.00003

Competitor	Split	Improvement
1 AMB__MIXING_RATIO	4.79021	0.00009
2 AMBIENT_TEMP__F_	88.50000	0.00007
3 CARS	0.96667	0.00007
4 WEATHER_TYPE\$	"FOG", "MIST", "Normal", "RAIN MIST", "SNOW FREEZING FOG", "THUNDERSTORM", "UNKNOWN PRECIPITATION"	0.00006
5 INDOOR_MIXING_RATIO	31.93421	0.00005

```

*****
*           Node 7: AMBIENT_TEMP_F_           *
*                   N: 358                   *
*****

```

```

=====
=           Terminal Node 4           = =           Terminal Node 5           =
=                   N: 338           = =                   N: 20           =
=====

```

Node 7 was split on AMBIENT\_TEMP\_F\_  
A case goes left if AMBIENT\_TEMP\_F\_ <= 87.50000  
Improvement = 0.000196      Complexity Threshold = 0.278776

Node	Cases	Wgt Count	Mean	StdDev
7	358	358.00	0.12863	0.09458
-4	338	338.00	0.13541	0.09183
-5	20	20.00	0.01391	0.06063

Competitor	Split	Improvement
1 WEATHER_TYPE\$	"FOG", "Normal", "RAIN MIST", "THUNDERSTORM", "THUNDERSTORM RAIN", "UNKNOWN PRECIPITATION", "UNKNOWN PRECIPITATION MIST"	0.00007
2 AMB_MIXING_RATIO	1.38431	0.00007
3 INDOOR_TEMP_F_	90.15275	0.00007
4 INDOOR_MIXING_RATIO	31.85758	0.00007
5 CARS	0.96667	0.00004

```

*****
*           Node 8: PRECIPITATION           *
*                   N: 457                   *
*****

```

```

=====
*           Node 9           * =           Terminal Node 10           =
*                   N: 455           * =                   N: 2           =
=====

```

Node 8 was split on PRECIPITATION  
A case goes left if PRECIPITATION <= 0.07500  
Improvement = 0.000142      Complexity Threshold = 0.201453

Node	Cases	Wgt Count	Mean	StdDev
8	457	457.00	0.17640	0.09439
9	455	455.00	0.17501	0.08711
-10	2	2.00	0.49308	0.45692

Competitor	Split	Improvement
1 TRUCKS	0.94167	0.00013
2 AMB_MIXING_RATIO	4.09260	0.00009
3 INDOOR_MIXING_RATIO		

	14.99130	0.00009
4 AMBIENT_TEMP_F_	84.50000	0.00009
5 INDOOR_TEMP_F_	67.52708	0.00007

```

*****
*                               *
*           Node 9: TRUCKS      *
*           N: 455              *
*                               *
*****

```

```

=====
=           Terminal Node 8     = =           Terminal Node 9     =
=           N: 450             = =           N: 5                 =
=====

```

Node 9 was split on TRUCKS  
A case goes left if TRUCKS <= 0.94167  
Improvement = 0.000134      Complexity Threshold = 0.190316

Node	Cases	Wgt Count	Mean	StdDev
9	455	455.00	0.17501	0.08711
-8	450	450.00	0.17286	0.08407
-9	5	5.00	0.36903	0.12778

Competitor	Split	Improvement
1 AMB_MIXING_RATIO	4.18198	0.00010
2 AMBIENT_TEMP_F_	84.50000	0.00009
3 INDOOR_MIXING_RATIO	14.99130	0.00009
4 PRECIPITATION	0.00250	0.00007
5 INDOOR_TEMP_F_	88.36242	0.00007

**Table A5: SO2 Complete Database Terminal Node Information**

=====

TERMINAL NODE INFORMATION – SO2 Complete Tree

=====

Parent Node	Wgt	Count	Count	Mean	StdDev	Complexity
1	21.00	21.00000	21.00000	0.00010	0.00039	0.41053
2	23.00	23.00000	23.00000	0.22751	0.08451	0.18045
3	47.00	47.00000	47.00000	0.08865	0.07147	0.19071
4	338.00	338.00000	338.00000	0.13541	0.09183	0.27878
5	20.00	20.00000	20.00000	0.01391	0.06063	0.27878
6	488.00	488.00000	488.00000	0.15601	0.07647	0.21681
7	25.00	25.00000	25.00000	0.06149	0.06486	0.18193
8	450.00	450.00000	450.00000	0.17286	0.08407	0.19032
9	5.00	5.00000	5.00000	0.36903	0.12778	0.19032
10	2.00	2.00000	2.00000	0.49308	0.45692	0.20145

## **B: NITRIC OXIDE (NO) Regression Tree Details (Complete Database)**

**Table B1: NO Complete Database – CART Variable Relative Importance**

Variable	Score	
MONTH	100.00	
AMBIENT MR	47.26	
WIND SPEED	31.28	
AMBIENT TEMP.	20.37	
TIME	16.66	
RUN CLOSE	16.06	
IDLE CLOSE	14.81	
IDLE OPEN	7.82	
LIGHT VEHICLES	3.01	
INDOOR TEMP.	1.15	
INDOOR MR	1.15	
SEASON	1.13	
WEATHER TYPE	0.29	

**Table B2: NO Complete Database Tree Summary Report**

	Competitor	Split	Improvement	N Left	N Right	N Missing
Main	MONTH	August,	0.01183	1100	227	0
		July,				
		June,May,				
		November,				
		October,				
		September				
1	TIME	0.95881	0.01070	1320	7	0
2	AMBIENT TEMP.	67.50000	0.00933	605	722	0
3	AMBIENT MR	4.45019	0.00924	176	1151	0
4	RUN CLOSE	51.03191	0.00631	1139	188	0
5	WIND SPEED	8.50000	0.00473	885	442	0
6	IDLE OPEN	6.38406	0.00467	585	742	0
7	INDOOR TEMP.	75.17850	0.00426	419	908	0
8	IDLE CLOSE	4.89776	0.00409	383	944	0
9	INDOOR MR	19.45151	0.00341	489	838	0
10	LIGHT VEHICLES	0.34500	0.00312	940	387	0
11	TRUCKS	0.00833	0.00303	297	1030	0
12	WEATHER TYPE	FOG,HAZE,	0.00236	106	1221	0
		RAIN MIST,				
		SNOW,				
		SNOW FREEZING FOG,				
		THUNDERSTORM,				
		THUNDERSTORM MIST,				
		THUNDERSTORM RAIN MIST,				
		UNKNOWN PRECIPITATION,				
		UNKNOWN PRECIPITATION MIST				
13	SEASON	Fall,	0.00198	1257	70	0
		Spring,				
		Summer				
14	SKY CONDITION	BKN OVC,	0.00192	123	1204	0
		CLR,FEW,				
		OVC				
15	VISIBILITY	2.75000	0.00168	72	1255	0



**Table B3: NO Complete Database Tree Sequence**

=====  
 Tree Sequence – NO Complete Tree  
 =====

Dependent variable: NO\_\_PPM\_

Terminal Tree Nodes		Cross-Validated Relative Error	Resubstitution Relative Error	Complexity Parameter	Relative Complexity
1	212	1.14943 +/- 0.12396	0.62974	0.000000	0.00000
147	26	1.13608 +/- 0.10355	0.68798	2.317973	0.00197
148	25	1.13431 +/- 0.10270	0.68996	2.336121	0.00199
149	24	1.13285 +/- 0.10201	0.69229	2.738787	0.00233
150	21	1.13090 +/- 0.10109	0.69932	2.755701	0.00234
151	20	1.12638 +/- 0.09883	0.70175	2.852394	0.00243
152	13	1.11733 +/- 0.09463	0.72408	3.752486	0.00319
153	11	1.11437 +/- 0.09317	0.73456	6.162641	0.00524
154	7	1.10688 +/- 0.08962	0.75644	6.432156	0.00547
155	6	1.09620 +/- 0.08510	0.76324	8.000082	0.00680
156**	1	1.00068 +/- 0.00000	1.00000	55.688496	0.04735

Initial mean = 0.32970

Initial variance = 0.88625

**Table B4: NO Complete Database Tree Node Information**

=====

NODE INFORMATION – NO Complete Tree

=====

```

*****
*                Node 1: MONTH$                *
*                N: 1327                        *
*****

```

```

=====
=          Terminal Node 1          = *          Node 2          *
=          N: 1100                  = *          N: 227          *
=====

```

Node 1 was split on MONTH\$  
A case goes left if MONTH\$ = ("August", "July", "June", "May",  
"November", "October", "September")

Improvement = 0.011835      Complexity Threshold = 55.688484

Node	Cases	Wgt Count	Mean	StdDev
1	1327	1327.00	0.32970	0.94141
-1	1100	1100.00	0.28028	0.47108
2	227	227.00	0.56916	2.00906

Surrogate	Split	Assoc	Improvement
1 AMB__MIXING_RATIO	r 4.45019	0.65198	0.00924
2 AMBIENT_TEMP__F__	r 39.50000	0.43172	0.00865
3 SEASON\$	s "Fall", "Spring", "Summer"	0.30837	0.00198
4 WEATHER_TYPE\$	s "FOG", "HAZE", "RAIN MIST", "THUNDERSTORM", "THUNDERSTORM RAIN MIST", "UNKNOWN PRECIPITATION MIST"	0.04405	0.00054
5 TIME	s 0.95885	0.00881	0.03027

Competitor	Split	Improvement
1 TIME	0.95881	0.01070
2 AMBIENT_TEMP__F__	67.50000	0.00933
3 AMB__MIXING_RATIO	4.45019	0.00924
4 RUN_CLOSE	51.03191	0.00631
5 WIND_SPEED	8.50000	0.00473

```

*****
*                Node 2: RUN_CLOSE                *
*                N: 227                        *
*****

```

```

*****
*                Node 3                * *                Node 5                *
*                N: 198                * *                N: 29                *
*****

```

Node 2 was split on RUN\_CLOSE

A case goes left if RUN\_CLOSE <= 51.03000  
 Improvement = 0.030210 Complexity Threshold = 65.684204

Node	Cases	Wgt	Count	Mean	StdDev
2	227		227.00	0.56916	2.00906
3	198		198.00	0.40833	1.29643
5	29		29.00	1.66724	4.32861

Surrogate	Split	Assoc	Improvement
1 IDLE_CLOSE	r 4.00550	0.79310	0.02786
2 MONTH\$	s "April", "December", "January"	0.20690	0.00015
3 SEASON\$	s "Fall", "Spring", "Summer"	0.20690	0.00015

Competitor	Split	Improvement
1 IDLE_CLOSE	3.67918	0.02798
2 AMBIENT_TEMP__F_	35.50000	0.02489
3 CARS	0.35000	0.02119
4 AMB__MIXING_RATIO	3.71930	0.02076
5 WIND_SPEED	19.00000	0.01069

```

*****
*           Node 3: AMBIENT_TEMP__F_           *
*                               N: 198          *
*****

```

```

=====
=           Terminal Node 2           = *           Node 4           *
=                               N: 184          = *           N: 14           *
=====
*****

```

Node 3 was split on AMBIENT\_TEMP\_\_F\_  
 A case goes left if AMBIENT\_TEMP\_\_F\_ <= 64.50000  
 Improvement = 0.029672 Complexity Threshold = 58.725285

Node	Cases	Wgt	Count	Mean	StdDev
3	198		198.00	0.40833	1.29643
-2	184		184.00	0.28533	0.47412
4	14		14.00	2.02500	4.24303

Surrogate	Split	Assoc	Improvement
1 AMB__MIXING_RATIO	s 14.20043	0.07143	0.00038

Competitor	Split	Improvement
1 TIME	0.83375	0.01603
2 WIND_SPEED	19.00000	0.01481
3 CARS	0.40833	0.00868
4 AMB__MIXING_RATIO	3.71930	0.00596
5 INDOOR_MIXING_RATIO	22.05912	0.00445

```

*****
*           Node 4: WIND_SPEED           *
*                               N: 14          *
*****

```

```

=====
=           Terminal Node 3           = =           Terminal Node 4           =

```

= N: 9 = = N: 5 =  
 =====

Node 4 was split on WIND\_SPEED  
 A case goes left if WIND\_SPEED <= 7.50000  
 Improvement = 0.058836 Complexity Threshold = 78.075912

Node	Cases	Wgt	Count	Mean	StdDev
4	14		14.00	2.02500	4.24303
-3	9		9.00	0.26481	0.17023
-4	5		5.00	5.19333	5.89423

Surrogate	Split	Assoc	Improvement
1 CARS s	0.62500	0.60000	0.00562
2 IDLE_OPEN r	3.50000	0.40000	0.01470
3 INDOOR_TEMP_F_ r	76.40275	0.40000	0.00216
4 INDOOR_MIXING_RATIO r	19.73196	0.40000	0.00216
5 MONTH\$ s	"April", "December", "January"	0.40000	0.17613

Competitor	Split	Improvement
1 TIME	0.79188	0.05176
2 AMBIENT_TEMP_F_	70.50000	0.04186
3 IDLE_CLOSE	19.50000	0.03459
4 RUN_CLOSE	33.00000	0.02729
5 AMB_MIXING_RATIO	6.37200	0.02532

\*\*\*\*\*  
 \* Node 5: AMB\_MIXING\_RATIO \*  
 \* N: 29 \*  
 \*\*\*\*\*

=====  
 = Terminal Node 5 = = Terminal Node 6 =  
 = N: 5 = = N: 24 =  
 =====

Node 5 was split on AMB\_MIXING\_RATIO  
 A case goes left if AMB\_MIXING\_RATIO <= 5.14414  
 Improvement = 0.079274 Complexity Threshold = 105.197037

Node	Cases	Wgt	Count	Mean	StdDev
5	29		29.00	1.66724	4.32861
-5	5		5.00	5.84000	9.28932
-6	24		24.00	0.79792	0.52898

Surrogate	Split	Assoc	Improvement
1 TIME s	0.35442	0.40000	0.00106
2 CARS r	0.60000	0.20000	0.00004

Competitor	Split	Improvement
1 IDLE_OPEN	4.98579	0.07701
2 RUN_CLOSE	52.98617	0.07078
3 IDLE_CLOSE	3.48157	0.07078
4 VISIBILITY	5.50000	0.06206
5 MONTH\$	"April", "January"	0.04482

**Table B5: NO Complete Database Terminal Node Information**

=====




TERMINAL NODE INFORMATION – NO Complete Tree

=====

Parent Node	Wgt	Count	Count	Mean	StdDev	Complexity
1	1100.00	1100.00000	1100.00000	0.28028	0.47108	55.68848
2	184.00	184.00000	184.00000	0.28533	0.47412	58.72528
3	9.00	9.00000	9.00000	0.26481	0.17023	78.07591
4	5.00	5.00000	5.00000	5.19333	5.89423	78.07591
5	5.00	5.00000	5.00000	5.84000	9.28932	105.19704
6	24.00	24.00000	24.00000	0.79792	0.52898	105.19704

## C: NITROGEN DIOXIDE (NO2) Regression Tree Details (Complete Database)

**Table C1: NO2 Complete Database – CART Variable Relative Importance**

Variable	Score	
TIME	100.00	
AMBIENT MR	6.78	
MONTH	4.80	
IDLE CLOSE	1.16	
SEASON	1.06	
RUN CLOSE	0.02	

**Table C2: NO2 Complete Database Tree Summary Report**

	Competitor	Split	Improvement	N Left	N Right	N Missing
Main	TIME	0.29177	0.00001	9	1318	0
1	RUN CLOSE	58.38235	2.41384E-006	1291	36	0
2	MONTH	April,	1.34845E-006	1098	229	0
		August,				
		January,				
		July,				
		June,				
		November,				
		October,				
		September				
3	SEASON	Fall,	5.16530E-007	906	421	0
		Summer,				
		Winter				
4	AMBIENT MR	10.45390	4.42692E-007	358	969	0
5	IDLE CLOSE	1.64146	4.10609E-007	161	1166	0
6	LIGHT VEHICLES	0.00833	2.95409E-007	217	1110	0

**Table C3: NO2 Complete Database Tree Sequence**

=====  
 Tree Sequence – NO2 Complete Tree  
 =====

Dependent variable: NO2\_\_PPM\_

Terminal Tree Nodes	Cross-Validated Relative Error	Resubstitution Relative Error	Complexity Parameter	Relative Complexity
1 104	1.08941 +/- 0.08460	0.87720	0.000000	0.00000
3 24	1.08460 +/- 0.08281	0.87926	0.000023	0.00017
4 17	1.08165 +/- 0.08254	0.88060	0.000033	0.00025
5 11	1.08280 +/- 0.08351	0.88307	0.000057	0.00043
6 7	1.07850 +/- 0.08147	0.88543	0.000082	0.00062
7 6	1.07694 +/- 0.08015	0.88620	0.000112	0.00084
8 5	1.07707 +/- 0.08079	0.88723	0.000147	0.00111
9 4	1.07518 +/- 0.07920	0.88936	0.000295	0.00221
10 3	1.08415 +/- 0.08399	0.89363	0.000578	0.00434
11 2	1.08415 +/- 0.08399	0.89851	0.000660	0.00496
12** 1	1.00135 +/- 0.00000	1.00000	0.013533	0.10154

Initial mean = 0.01107

Initial variance = 0.00010

**Table C4: NO2 Complete Database Tree Node Information**

=====

NODE INFORMATION – NO2 Complete Tree

=====

```

*****
*                               *
*           Node 1: TIME         *
*           N: 1327              *
*                               *
*****

```

```

=====
=           Terminal Node 1     = *           Node 2           *
=           N: 9                = *           N: 1318          *
=====
*****

```

Node 1 was split on TIME  
A case goes left if TIME <= 0.29177  
Improvement = 0.000010      Complexity Threshold = 0.013523

Node	Cases	Wgt Count	Mean	StdDev
1	1327	1327.00	0.01107	0.01002
-1	9	9.00	0.04970	0.11095
2	1318	1318.00	0.01080	0.00261

Competitor	Split	Improvement
1 RUN_CLOSE	58.38235	0.00000
2 MONTH\$	"April", "August", "January", "July", "November", "October", "September"	0.00000
3 SEASON\$	"Summer", "Winter"	0.00000
4 AMB__MIXING_RATIO	10.45390	0.00000
5 IDLE_CLOSE	1.64146	0.00000

```

*****
*                               *
*           Node 2: MONTH$      *
*           N: 1318              *
*                               *
*****

```

```

=====
=           Terminal Node 2     = *           Node 3           *
=           N: 815              = *           N: 503            *
=====
*****

```

Node 2 was split on MONTH\$  
A case goes left if MONTH\$ = ("April", "August", "January", "July",  
"November", "October", "September")

Improvement = 0.000000      Complexity Threshold = 0.000650

Node	Cases	Wgt Count	Mean	StdDev
2	1318	1318.00	0.01080	0.00261
-2	815	815.00	0.01025	0.00095
3	503	503.00	0.01170	0.00388



Surrogate	Split	Assoc	Improvement
1 SEASON\$	s "Summer", "Winter"	0.69185	0.00000
2 RUN_CLOSE	s 52.08696	0.10934	0.00000
3 AMB__MIXING_RATIO			
	r 4.36390	0.08350	0.00000
4 IDLE_CLOSE	s 25.14050	0.06958	0.00000
5 TIME	s 0.95881	0.00596	0.00000

Competitor	Split	Improvement
1 AMB__MIXING_RATIO		
	3.02864	0.00000
2 SEASON\$	"Fall"	0.00000
3 IDLE_CLOSE	46.87500	0.00000
4 RUN_CLOSE	20.50000	0.00000
5 TIME	0.95878	0.00000

```

*****
*           Node 3: AMB__MIXING_RATIO           *
*                   N: 503                       *
*****

```

```

=====
=           Terminal Node 3           = =           Terminal Node 4           =
=                   N: 39              = =                   N: 464              =
=====

```

Node 3 was split on AMB\_\_MIXING\_RATIO  
A case goes left if AMB\_\_MIXING\_RATIO <= 3.02864  
Improvement = 0.000000      Complexity Threshold = 0.000568

Node	Cases	Wgt Count	Mean	StdDev
3	503	503.00	0.01170	0.00388
-3	39	39.00	0.01536	0.00882
-4	464	464.00	0.01139	0.00293

Competitor	Split	Improvement
1 IDLE_CLOSE	15.50000	0.00000
2 RUN_CLOSE	12.50000	0.00000
3 MONTH\$	"December", "June", "May"	0.00000
4 SEASON\$	"Fall", "Spring", "Summer", "Winter"	0.00000
5 CARS	0.24500	0.00000

**Table C5: NO2 Complete Database Terminal Node Information**

=====

TERMINAL NODE INFORMATION – NO2 Complete Tree

=====

Parent Node	Wgt	Count	Count	Mean	StdDev	Complexity
1	9.00	9.00000	9.00000	0.04970	0.11095	0.01352
2	815.00	815.00000	815.00000	0.01025	0.00095	0.00065
3	39.00	39.00000	39.00000	0.01536	0.00882	0.00057
4	464.00	464.00000	464.00000	0.01139	0.00293	0.00057

## **D: Validation Results**

**Table D1: Relative Importance of the Variables for In-Vehicle SO<sub>2</sub> Obtained from CART Runs**

<b>Variable</b>	<b>Complete Database Score</b>	<b>Test Database (90% Complete Database) Score</b>
MONTH	100.00	100.00
SKY CONDITION	68.09	78.54
AMBIENT TEMP.	47.16	62.11
PRECIPITATION	33.45	60.57
HEAVY VEHICLES	31.57	54.12
INDOOR TEMP.	28.95	52.12
INDOOR MR	28.95	50.77
RUN/CLOSE	25.81	49.98
WEATHER TYPE	23.03	42.15
AMBIENT MR	15.88	41.65
SEASON	14.90	8.65
LIGHT VEHICLES	7.65	3.24
IDLE/CLOSE	0.27	1.84

**Table D2: Sensitivity Results for In-Vehicle SO<sub>2</sub> Obtained from the ANOVA**

Variable	F Value	Sig.	Significant	Rank	Variable	F Value	Sig.	Significant	Rank
<i>Month = Apr. 07 to July 07, Sep. 07, Nov. 07, Jan. 08</i>					<i>Month = Aug. 07, Oct. 07, Dec. 07, Feb. 08, Mar. 08</i>				
Sky condition	13.792, 14.854*	<0.0001, <0.0001*	Yes	1	Sky condition	0.008, 0.004*	0.930, 1.111*	No	-----
Ambient temp.	1.745, 1.875*	<0.0001, <0.0001*	Yes	3	Ambient temp.	1.311, 1.227*	0.049, 0.048*	Yes	4
Heavy vehicles	0.976, 1.254*	0.527, 0.555*	No	-----	Heavy vehicles	1.789, 1.546*	0.001, 0.018*	Yes	3
Indoor temp.	1.162, 1.576*	0.477, 0.645*	No	-----	Indoor temp.	1.427, 1.087*	0.503, 0.645*	No	-----
Indoor MR	0.695, 0.875*	0.769, 0.845*	No	-----	Indoor MR	-----*, -----*	-----*, -----*	-----	-----
Run/close	0.859, 1.133*	0.869, 0.899*	No	-----	Run/close	1.019, 0.987*	0.444, 0.458*	No	-----
Weather type	2.746, 2.831*	0.002, <0.0001*	Yes	2	Weather type	1.193, 1.087*	0.301, 0.364*	No	-----
Ambient MR	1.315, 1.417*	0.007, 0.003*	Yes	4	Ambient MR	1.793, 1.564*	0.003, 0.041*	Yes	2
Season	0.677, 0.992*	0.566, 0.989*	No	-----	Season	2.258, 1.877*	0.106, 0.125*	No	-----
Light Vehicles	0.966, 0.997*	0.572, 0.877*	No	-----	Light Vehicles	0.816, 0.654*	0.868, 0.899*	No	-----
Idle/Close	0.894, 0.997*	0.845, 1.254*	No	-----	Idle/Close	1.276, 1.111*	0.039, 0.048*	Yes	5
Precipitation	1.047, 1.087*	0.402, 0.898*	No	-----	Precipitation	2.938, 2.451*	<0.0001, 0.007*	Yes	1

\* indicates the results obtained with the test database, i.e., 90% complete database

**Table D3: Relative Importance of the Variables for In-Vehicle NO Obtained from CART Runs**

<b>Variable</b>	<b>Complete Database Score</b>	<b>Test Database (90% Complete Database) Score</b>
MONTH	100.00	100.00
AMBIENT MR	47.26	38.46
WIND SPEED	31.28	34.11
AMBIENT TEMP.	20.37	28.56
TIME	16.66	27.65
RUN/CLOSE	16.06	25.22
IDLE/CLOSE	14.81	24.38
IDLE/OPEN	7.82	22.58
LIGHT VEHICLES	3.01	16.25
INDOOR TEMP.	1.15	8.49
INDOOR MR	1.15	2.57
SEASON	1.13	1.18
WEATHER TYPE	0.29	1.12

**Table D4: Sensitivity Results for In-Vehicle NO Obtained from the ANOVA**

Variable	F Value	Sig.	Significant	Rank	Variable	F Value	Sig.	Significant	Rank
<i>Month = May. 07 to Nov. 07</i>					<i>Month = Apr. 07, Dec. 07 to Mar. 08</i>				
Ambient MR	0.925, 0.895*	0.795, 0.811*	No	-----	Ambient MR	0.130, 0.126*	1.000, 1.254*	No	-----
Wind speed	0.602, 0.589	0.919, 0.985*	No	-----	Wind speed	1.508, 1.498*	0.071, 0.097*	No	-----
Ambient temp.	1.710, 1.658*	<0.0001, <0.0001*	Yes	2	Ambient temp.	0.802, 0.758*	0.878, 1.107*	No	-----
Time of the day	0.854, 0.844*	0.624, 0.789*	No	-----	Time of the day	0.841, 0.698*	0.639, 0.894*	No	-----
Run/close	1.401, 1.299*	0.005, 0.014*	Yes	4	Run/close	7.930, 7.812*	<0.0001, 0.036*	Yes	1
Idle/close	1.536, 1.444*	<0.0001, <0.0001*	Yes	3	Idle/close	2.090, 1.898*	<0.0001, 0.041*	Yes	3
Idle/open	1.112, 1.025*	0.212, 0.355*	No	-----	Idle/open	3.748, 3.456*	<0.0001, 0.025*	Yes	2
Light vehicles	0.543, 0.521*	1.000, 1.123*	No	-----	Light vehicles	0.910, 0.889*	0.695, 0.718*	No	-----
Indoor temp.	0.409, 0.356*	0.983, 1.122*	No	-----	Indoor temp.	11.621, 10.786*	0.231, 0.541*	No	-----
Indoor MR	0.354, 0.298*	0.907, 1.111*	No	-----	Indoor MR	-----, -----*	-----, -----*	-----	-----
Season	6.427, 6.222*	0.002, 0.006*	Yes	1	Season	0.811, 0.777*	0.445, 0.687*	No	-----
Weather Type	1.067, 0.999*	0.384, 0.452*	No	-----	Weather Type	0.247, 0.227*	0.981, 1.133*	No	-----

\* indicates the results obtained with the test database, i.e., 90% complete database

**Table D5: Relative Importance of the Variables for In-Vehicle NO<sub>2</sub> Obtained from CART Runs**

Variable	Complete Database Score	Test Database (90% Complete Database) Score
TIME	100.00	100.00
AMBIENT MR	6.78	18.54
MONTH	4.80	16.41
IDLE/CLOSE	1.16	8.56
SEASON	1.06	1.68
RUN/CLOSE	0.02	1.11

**Table D6: Sensitivity Results for In-Vehicle NO<sub>2</sub> Obtained from the ANOVA**

Variable	F Value	Sig.	Significant	Rank	Variable	F Value	Sig.	Significant	Rank
<i>Time of day = 6:00 a.m. to 7:00 a.m.</i>					<i>Time of day = 7:00 a.m. to 11:00 p.m.</i>				
Month	1.123, 1.089*	0.355, 0.478*	No	-----	Month	40.636, 45.452*	<0.0001, 0.007*	Yes	1
Idle/close	1314.106, 1085.457*	<0.0001, <0.0001*	Yes	1	Idle/close	0.164, 0.257*	1.000, 0.877*	No	-----
Ambient MR	215.640, 189.999*	0.005, 0.019*	Yes	2	Ambient MR	1.533, 1.994*	<0.0001, 0.042*	Yes	4
Light vehicles	0.073, 0.064*	1.000, 1.253*	No	-----	Light vehicles	0.333, 0.541*	1.000, 0.756*	No	-----
Season	0.806, 0.754*	0.494, 0.617*	No	-----	Season	16.999, 18.913*	<0.0001, 0.012*	Yes	2
Run/close	102.393, 87.633*	<0.0001, 0.034*	Yes	3	Run/close	0.157, 0.541*	1.000, 0.857*	No	-----
Ambient temp.	0.345, 0.293*	1.000, 1.579*	No	-----	Ambient temp.	1.772, 2.107*	<0.0001, 0.037*	Yes	3
Idle/open	0.145, 0.107*	1.000, 1.289*	No	-----	Idle/open	0.173, 0.251*	1.000, 0.783*	No	-----

\* indicates the results obtained with the test database, i.e., 90% complete database