METHODOLOGY FOR DEVELOPING MODAL EMISSION RATES FOR EPA'S MULTI-SCALE MOTOR VEHICLE & EQUIPMENT EMISSION SYSTEM

H. Christopher Frey, Ph.D.

Associate Professor

NC STATE UNIVERSITY

Prepared for:
Office of Transportation and Air Quality
U.S. Environmental Protection Agency
Ann Arbor, MI

November 6, 2002

NC STATE UNIVERSITY

Acknowledgments

- Sponsored by EPA/OTAQ
- NCSU Project Team

  H. C. Frey, Principal Investigator

  A. Unal

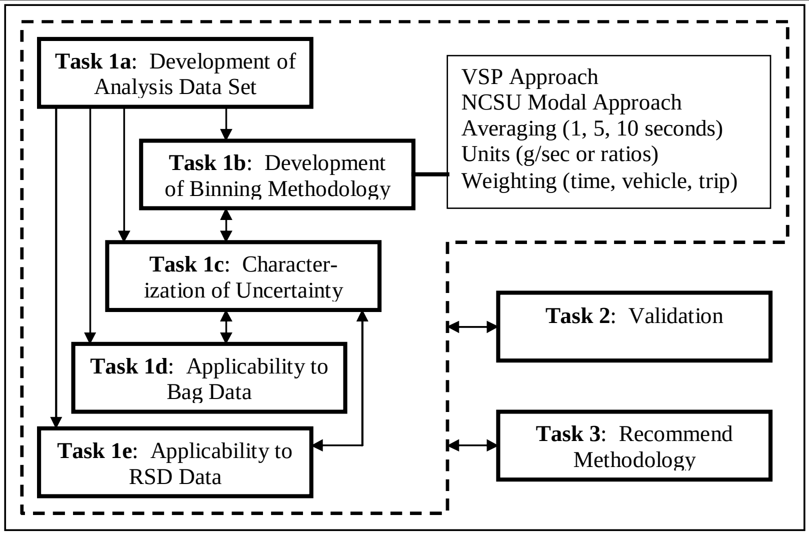
  J. Chen

  S. Li

  C. Xuan
- Final report will be available at EPA OTAQ's web site:
  – Frey, H.C., A. Unal, J. Chen, S. Li, and C. Xuan (2002), "Methodology for Developing Modal Emission Rates for EPA's Multi-Scale Motor Vehicle & Equipment Emissions System," Prepared by NCSU for EPA/OTAQ, August 31, 2002. 287 pages.

NC STATE UNIVERSITY

Objectives



NC STATE UNIVERSITY

Key Questions

1. What dataset should be used for the conceptual model? (Task 1a, Chapter 2)
2. Which binning approach should be used? (Task 1b, Chapter 3)
3. How much detail should be included in the binning approach, in terms of how many explanatory variables and how many strata for each variable? (Task 1b, Chapter 3)
4. What averaging time is preferred as a basis for model development? (Task 1b, Chapter 4)
5. What emission factor units should be used? (Task 1b, Chapter 5)

NC STATE UNIVERSITY

Key Questions

6. What weighting approach should be used ? (Task 1b, Chapter 6)
7. How should variability and uncertainty be characterized? (Task 1c, Chapter 7)
8. How should aggregate bag data be analyzed to derive estimates of modal emission rates? (Task 1d, Chapter 8)
9. What is the potential role and feasibility of incorporating remote sensing data into the conceptual modeling approach? (Task 1e, Chapter 5)
10. How should the conceptual model be validated and what are the results of validation exercises? (Task 2, Chapter 9)

NC STATE UNIVERSITY

Answers to Key Questions

- What hot stabilized Tier 1 Vehicle tailpipe emissions datasets should be used for development of the conceptual approach?
  - EPA Dynamometer;
  - EPA On-Board;
  - NCHRP Dynamometer Datasets
- Which binning approach should be used (VSP, driving modes)?
  - VSP-based approach for Second-by-Second data
  - Driving modes gives comparable results
- How much detail should be included in the binning approach, in terms of how many explanatory variables and how many strata for each variable?
  - Three explanatory variables, including 14 VSP bins with four strata (e.g., Engine Displacement < 3.5 liter and Odometer < 50,000 miles)

NC STATE UNIVERSITY

Answers to Key Questions

- What averaging time is preferred as a basis for model development (1, 5, or 10 seconds)?
  - 1 second
- What emission factor units should be used (g/sec or ratios)?
  - Grams/second
- What weighting approach should be used (time, vehicle, or trip)?
  - Time-weighted average
- How should variability and uncertainty be characterized?
  - Parametric distributions were utilized to characterize variability.
  - Uncertainty was characterized based upon normality assumption in most of the cases, and upon bootstrap in others.

NC STATE UNIVERSITY

Answers to Key Questions

- How should aggregate bag data be analyzed to derive estimates of modal emission rates?
  - Modal emission rates were estimated using Constrained Least-Squares method.
- What is the potential role and feasibility of incorporating RSD into the conceptual modeling approach?
  - Inconsistencies between RSD and Modeling data set
  - Different activity patterns make comparison difficult

NC STATE UNIVERSITY

Answers to Key Questions

- How should the conceptual model be validated and what are the results of validation exercises?
  - The conceptual model was validated against three different datasets:
    - Calibration dataset (verification: consistency check)
    - Independent EPA dynamometer, EPA On-board, and NCHRP dynamometer data
    - CARB dataset
  - Comparisons were made accounting for statistical confidence intervals on predicted and observed emissions

NC STATE UNIVERSITY

Summary of Databases

- 17 on-board vehicles from the “shootout” analysis
- 36 vehicles tested at EPA’s lab for Mobile6
- 311 vehicles from Colorado IM program
- 74 vehicles for development of UC Riverside’s Comprehensive Modal Emission Model
- RSD data on 200,966 Tier 1 LDGVs collected in Missouri
- 150 vehicles tested by California Air Resources board (CARB)

NC STATE UNIVERSITY

Data Post-Processing

- Quality Assurance
  Check for Tier 1 (i.e., Model Year ≥1996)
  Check for zero or negative values
- Correction Factor
  Humidity
  Different HC Measurement Technique
  (i.e., FID HC = 1.652 NDIR HC)
- Estimation of new Parameters
  Acceleration, VSP, Power Demand

NC STATE UNIVERSITY

Organization of the Data for Analysis

- Modeling Dataset
  - 71,699 seconds from 13 on-board vehicles
  - 68,482 seconds from 33 vehicles tested at EPA's lab
  - 92,000 seconds from 49 vehicles tested at UC Riverside
- Validation Dataset 2
  - 19,342 seconds from 3 on-board vehicles
  - 16,967 seconds from 3 vehicles tested at EPA's lab
  - 45,598 seconds from 25 vehicles tested at UC, Riverside
- Validation Dataset 3
  - 228,539 seconds from 17 vehicles tested by CARB

NC STATE UNIVERSITY

Organization of the Data for Analysis

- IM240 data were used for comparisons
- The NCHRP bag data were used in developing modal emission rates from aggregate bag data
- Remote sensing data for 200,966 vehicles in Missouri were selected randomly from approximately 2 million

NC STATE UNIVERSITY

Development of a Modal Emissions Modeling Approach

- Two promising approaches from "Shootout"
  - EPA's VSP-based approach
  - NCSU's driving mode-based approach
- Methodology: Supervised Hierarchical Tree-Based Regression (HTBR)

NC STATE UNIVERSITY

VSP-Based Modal Approach

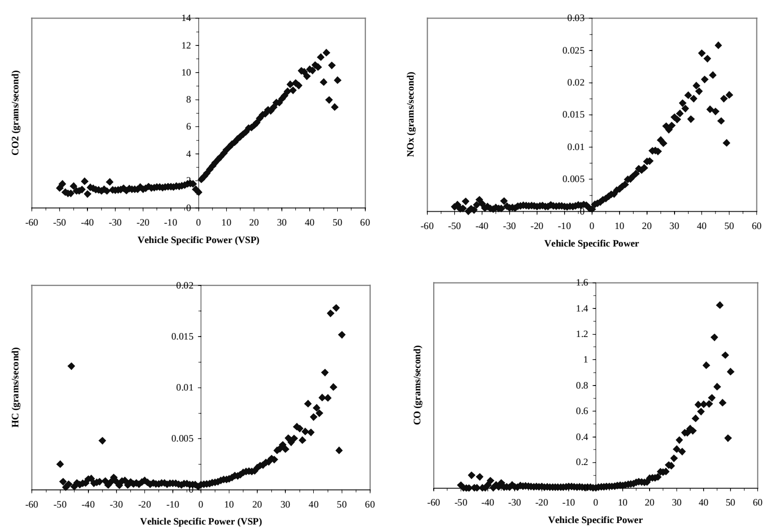
- Utilized EPA's VSP approach:

$$\text{VSP (kW/ton)} = v\left[1.1a + 9.81\left(a\tan\left(\sin\left(\text{grade}\right)\right)\right) + 0.132\right] + 0.000302v^3$$

- Exploratory analysis:
  - Graphical analysis
  - HTBR

NC STATE UNIVERSITY

VSP-Based Modal Approach Exploratory Analysis



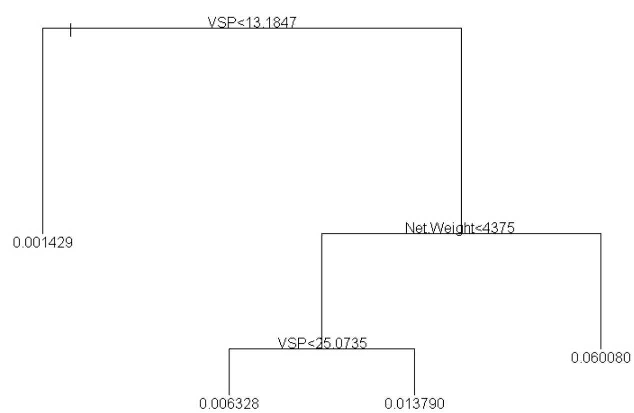
NC STATE UNIVERSITY

VSP-Based Modal Approach Exploratory Analysis

- Parameters used in HTBR:
  - VSP
  - Engine Displacement
  - Odometer Reading
  - Net Weight
  - Speed
  - Acceleration
  - Vehicle Model Year
  - Number of Cylinders
  - Temperature
  - Humidity
  - A/C Usage

NC STATE UNIVERSITY

VSP-Based Modal Approach Exploratory Analysis



Example: HTBR Result for NO_x (g/sec) Emissions

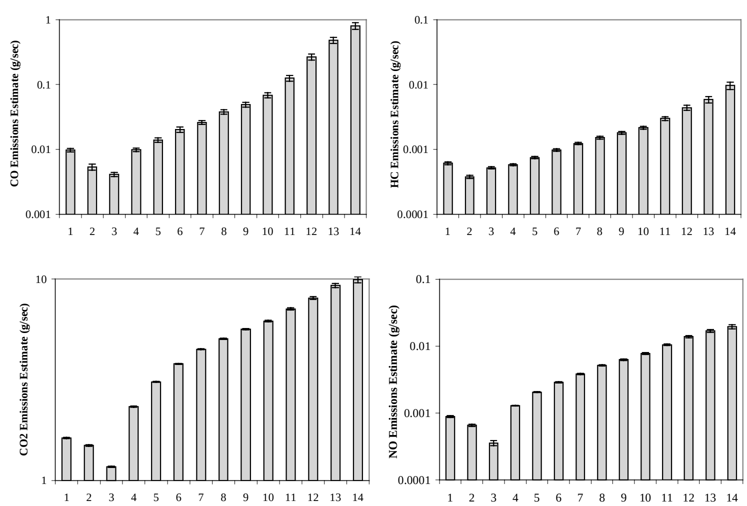
NC STATE UNIVERSITY

VSP-Based Modal Approach Mode Definitions

VSP Mode	Definition
1	VSP<-2
2	-2<=VSP<0
3	0<=VSP<1
4	1<=VSP<4
5	4<=VSP<7
6	7<=VSP<10
7	10<=VSP<13
8	13<=VSP<16
9	16<=VSP<19
10	19<=VSP<23
11	23<=VSP<28
12	28<=VSP<33
13	33<=VSP<39
14	39<=VSP

NC STATE UNIVERSITY

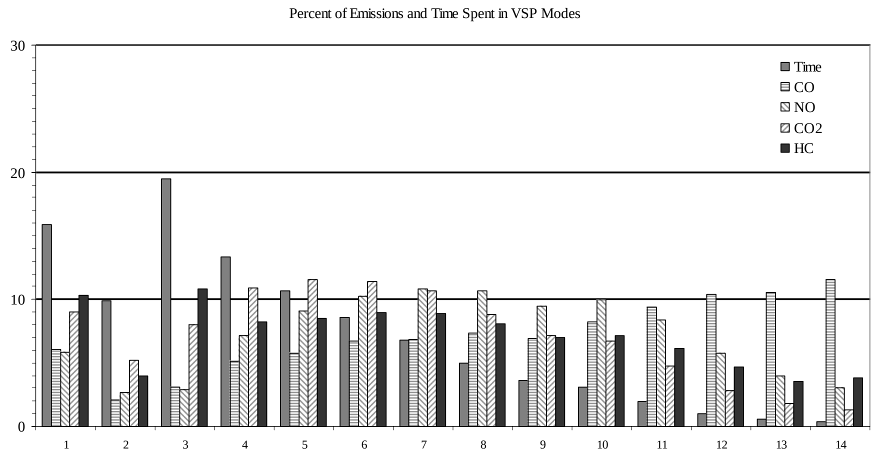
VSP-Based Modal Approach



Average Modal Rates

NC STATE UNIVERSITY

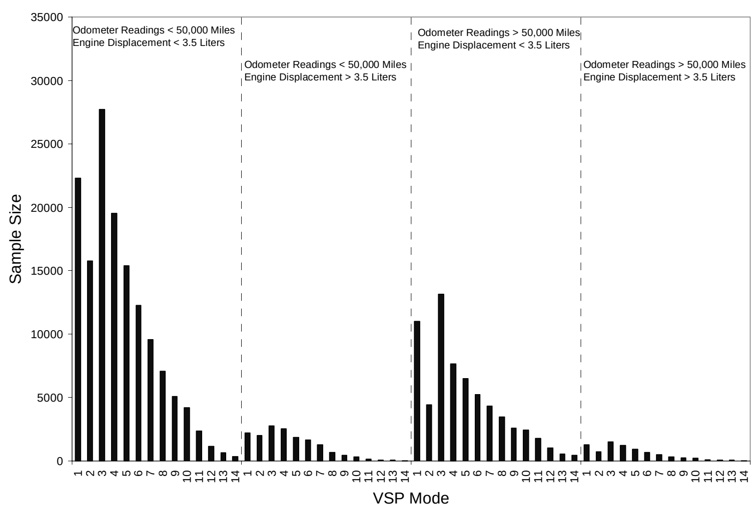
VSP-Based Modal Approach



Percent of Emissions and Time Spent in VSP Modes

NC STATE UNIVERSITY

Sample Sizes for Each Mode



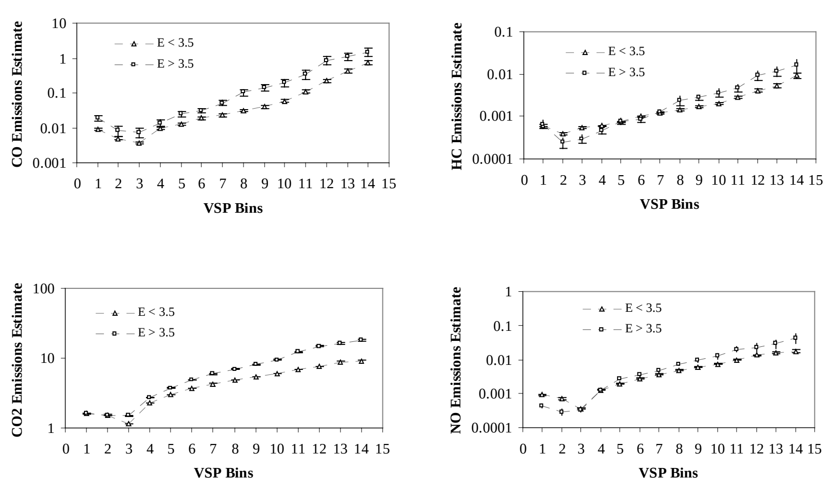
NC STATE UNIVERSITY

VSP-Based Modal Approach
Improving Driving Modes- Vehicle Characteristics

Example: HTBR Result for CO_2			
Mode	**1st Cut point**	**2nd Cut point**	**3rd Cut point**
1	Net Weight 3,200	No of Cylinders 5	Odometer 25,000
2	Net Weight 3,200		
3	No of Cylinders 5	Net Weight 3,200	
4	No of Cylinders 5	Net Weight 2,700	Net Weight 3,600
5	Engine Disp. 2.3	Net Weight 2,800	Net Weight 3,700
6	Engine Disp. 2.3	Engine Disp. 1.95	No of Cylinders 7
7	Net Weight 3,700	Engine Disp. 1.95	Engine Disp. 3.9
8	Net Weight 3,700	Engine Disp. 1.95	Engine Disp. 3.5
9	Engine Disp. 3.5	Odometer 46000	
10	Engine Disp. 3.5	Odometer 44000	
11	Engine Disp. 3.5	Odometer 46000	
12	Engine Disp. 3.5	Odometer 37000	
13	Engine Disp. 3.5	Odometer 23000	
14	Engine Disp. 3.5		Odometer 60,000

NC STATE UNIVERSITY

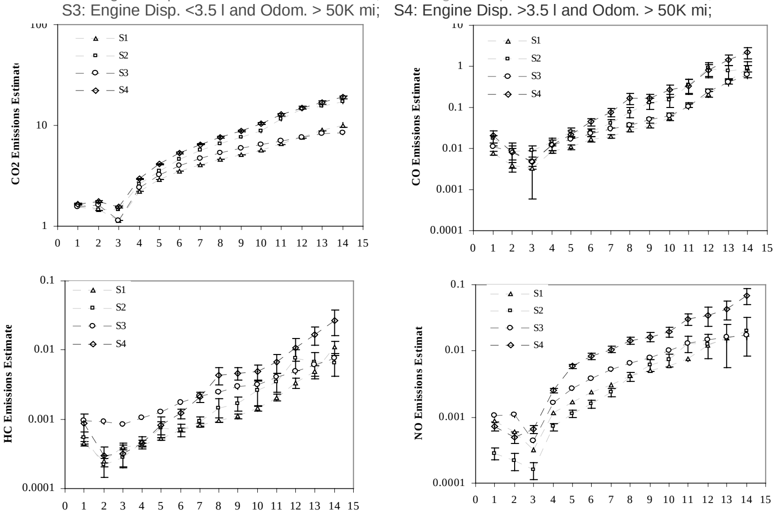
VSP-Based Modal Approach
Improving Driving Modes- Engine Displacement Only



NC STATE UNIVERSITY

Improving VSP-Based Modal Approach based upon Engine Displacement and Odometer Reading

S1: Engine Disp. <3.5 l and Odom. < 50K mi; S2: Engine Disp. >3.5 l and Odom. < 50K mi;
S3: Engine Disp. <3.5 l and Odom. > 50K mi; S4: Engine Disp. >3.5 l and Odom. > 50K mi;



NC STATE UNIVERSITY

NCSU Modal Approach

- Empirical Modal Definitions
  - **Idle:** $v_t = 0 \text{ mph } a_t = 0 \text{ mph/sec}$
  - **Acceleration:**

$$v_t > 0 \text{ mph} \quad a_t > 2 \text{ mph/sec} \quad \frac{\sum_{i=t-3}^{t} a_i}{3} > 1 \text{ mph/sec}$$

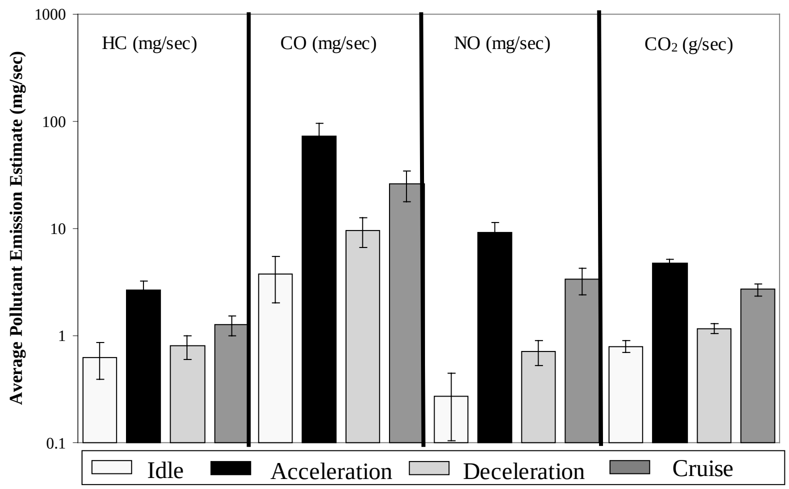
  - **Deceleration:**

$$v_t < 0 \text{ mph} \quad a_t < -2 \text{ mph/sec} \quad \frac{\sum_{i=t-3}^{t} a_i}{3} < -1 \text{ mph/sec}$$

  - **Cruise:** Else

NC STATE UNIVERSITY

NCSU Modal Approach (Shootout Project)



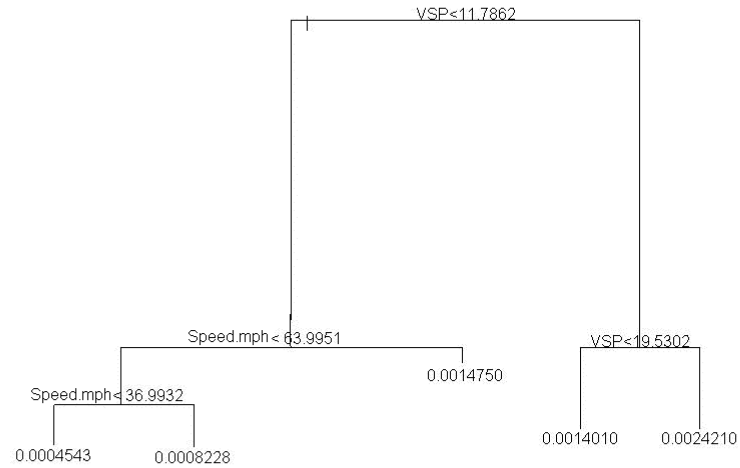
NC STATE UNIVERSITY

NCSU Modal Approach Improving Driving Modes

- Variables used:
  - VSP
  - Engine Displacement
  - Odometer Reading
  - Net Weight
  - Speed
  - Acceleration
  - Vehicle Model Year
  - Number of Cylinders
  - Temperature
  - Humidity
  - A/C Usage

NC STATE UNIVERSITY

NCSU Modal Approach
Improving Driving Modes: Cruise



Example: HTBR Result for NCSU Cruise Mode for NO_x

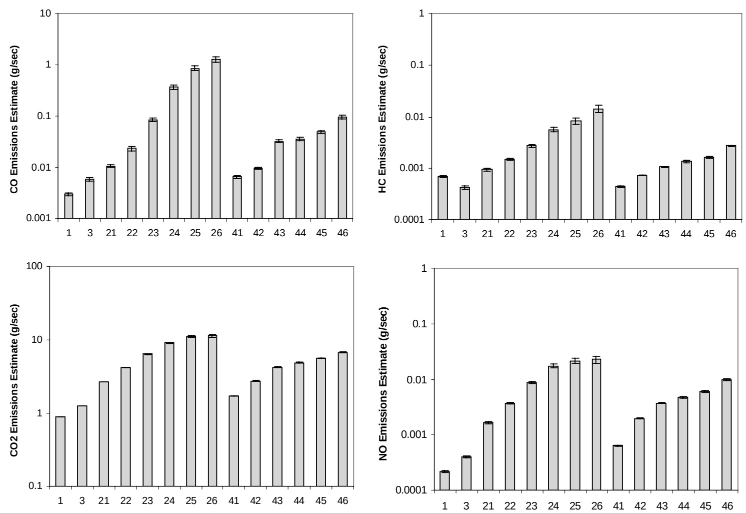
NC STATE UNIVERSITY

NCSU Modal Approach: Modal Definitions

ID	Definition
1	NCSU Idle
3	NCSU Deceleration
21	NCSU Acceleration & VSP<8
22	NCSU Acceleration & 8≤VSP<15
23	NCSU Acceleration & 15≤VSP<25
24	NCSU Acceleration & 25≤VSP<33
25	NCSU Acceleration & 33≤VSP<40
26	NCSU Acceleration & VSP≥40
41	NCSU Cruise & VSP≤12 and Speed≤30
42	NCSU Cruise & VSP≤12 and 30<Speed≤55
43	NCSU Cruise & VSP≤12 and Speed>55
44	NCSU Cruise & 12<VSP≤16
45	NCSU Cruise & 16<VSP≤22
46	NCSU Cruise & VSP>22

NC STATE UNIVERSITY

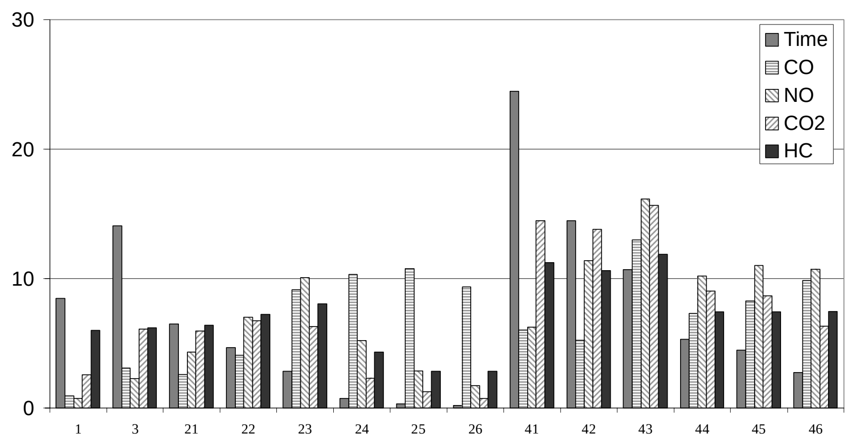
NCSU Modal Approach
Improving Driving Modes



Average Modal Rates

NC STATE UNIVERSITY

NCSU Modal Approach
Improving Driving Modes



Percent of Emissions and Time Spent in NCSU Modes

NC STATE UNIVERSITY

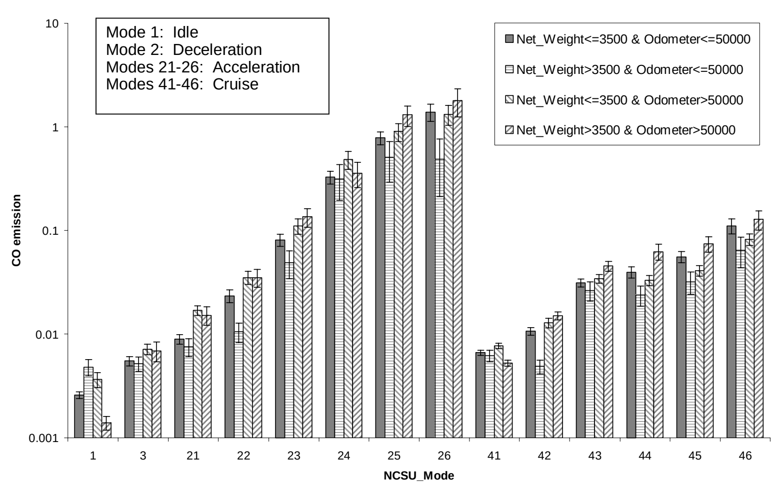
NCSU Modal Approach
Improving Driving Modes- Vehicle Characteristics

Example: HTBR Result for NO_x

Mode	1st Cut point	2nd Cut point	3rd Cut point
1 (Idle)	Net weight 5	Odometer 60,158	Engine Displacement 3.45
3 (Deceleration)	Odometer 8,785		Engine Displacement 3.45
21 (Acceleration)	Odometer 58,057	Odometer 29,057	Engine Displacement 2.75
22	Odometer 66,163	Odometer 38,353	Odometer 45,900
23	Odometer 63,341	Odometer 22,195	Odometer 43,433
24	Odometer 58,560	Odometer 12,800	Net weight 3,486
25	Odometer 58,057	Net weight 2,813	Engine Displacement 2.3
26	Odometer 58,057	Net weight 2,550	
41 (Cruise)	Odometer 71,964	Engine Displacement 0.75	Net weight 3,754
42	Net weight 3,611	Odometer 57,695	Engine Displacement 4.45
43	Odometer 17,220	Engine Displacement 3.05	
44	Odometer 17,220	Odometer 11,493	Net weight 2,531
45	Odometer 38,353	Engine Displacement 3	Odometer 83,491
46	Odometer 83,490	Odometer 61,024	

NC STATE UNIVERSITY

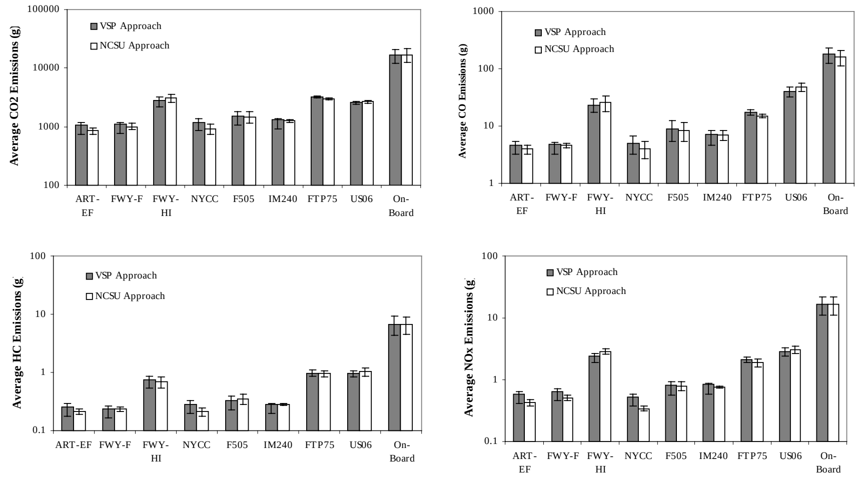
NCSU Modal Approach
Improving Driving Modes- Vehicle Characteristics



Example: Average Rates for Improved Modes for CO

NC STATE UNIVERSITY

Comparison of NCSU and VSP Modal Approaches



NC STATE UNIVERSITY

Selection of an Averaging Time for Model Development

Objective:

- To evaluate the potential benefits of working with data that have been averaged over time when developing bins/modes

Approach:

- 5 and 10 consecutive averaging times were evaluated

NC STATE UNIVERSITY

Selection of an Averaging Time for Model Development

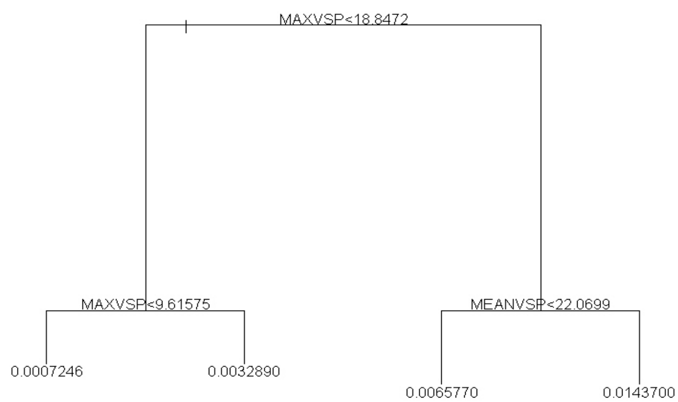
Explanatory Variables:

- Mean, Maximum and Standard Deviation of
  - Speed
  - Acceleration
  - VSP
  - Power demand

NC STATE UNIVERSITY

Selection of an Averaging Time for Model Development

Maximum VSP was found to be most useful explanatory variable



HTBR Result for 5 seconds Averaging NO

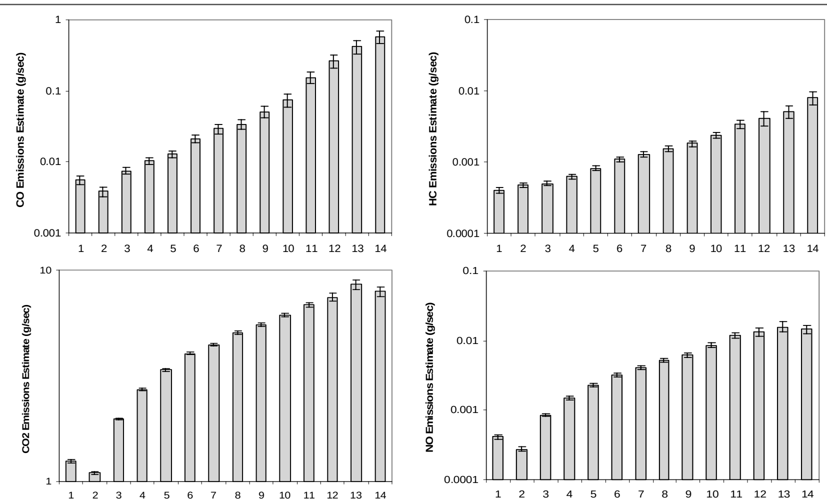
NC STATE UNIVERSITY

Mode Definitions Based Upon 5 Second-Averaged Data

ID	Definition
1	Max VSP <0
2	0 ≤ Max VSP < 2
3	2 ≤ Max VSP < 6
4	6 ≤ Max VSP < 9
5	9 ≤ Max VSP < 12
6	12 ≤ Max VSP < 15
7	15 ≤ Max VSP < 18
8	18 ≤ Max VSP < 21
9	21 ≤ Max VSP < 25
10	25 ≤ Max VSP < 29
11	29 ≤ Max VSP < 34
12	34 ≤ Max VSP < 38
13	38 ≤ Max VSP < 42
14	Max VSP ≥ 42

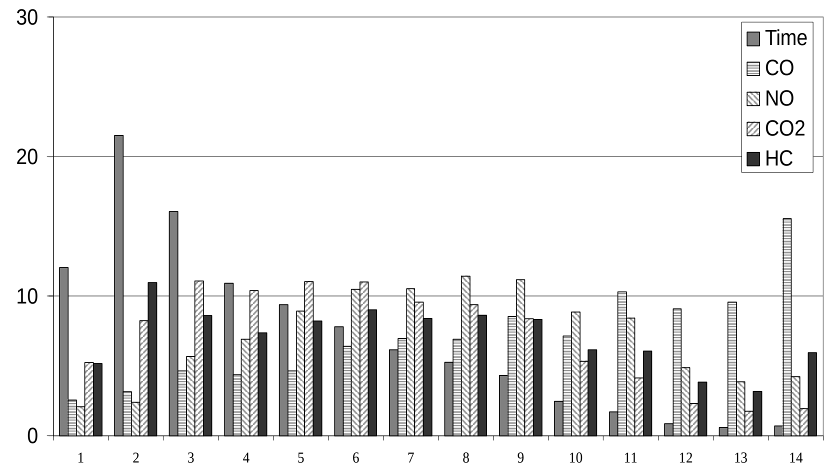
NC STATE UNIVERSITY

Average Modal Emission Rates for Each Mode Based Upon 5 Second-Averaged data



NC STATE UNIVERSITY

Percent of Emissions and Time Spent for Each Mode for 5 Second-Averaged data



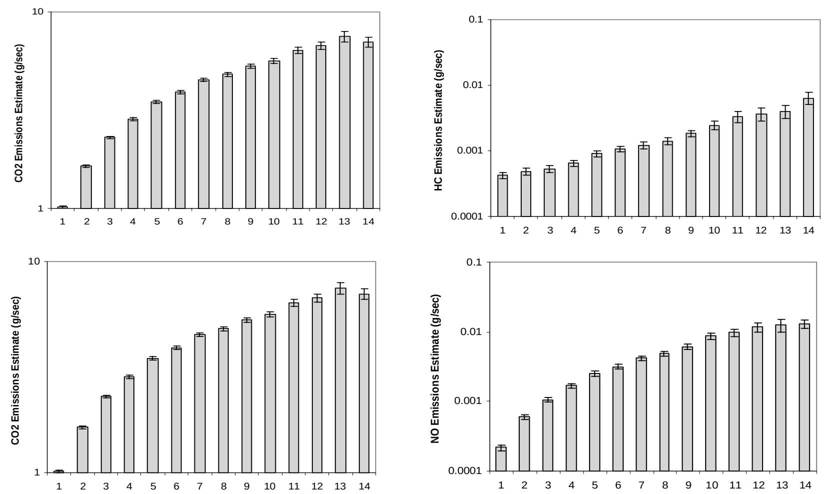
NC STATE UNIVERSITY

Mode Definitions Based Upon 10 Second-Averaged Data

ID	Definition
1	Max VSP < 1
2	1 ≤ Max VSP < 6
3	6 ≤ Max VSP < 9
4	9 ≤ Max VSP < 12
5	12 ≤ Max VSP < 15
6	15 ≤ Max VSP < 18
7	18 ≤ Max VSP < 21
8	21 ≤ Max VSP < 24
9	24 ≤ Max VSP < 27
10	27 ≤ Max VSP < 31
11	31 ≤ Max VSP < 35
12	35 ≤ Max VSP < 39
13	39 ≤ Max VSP < 43
14	Max VSP ≥ 43

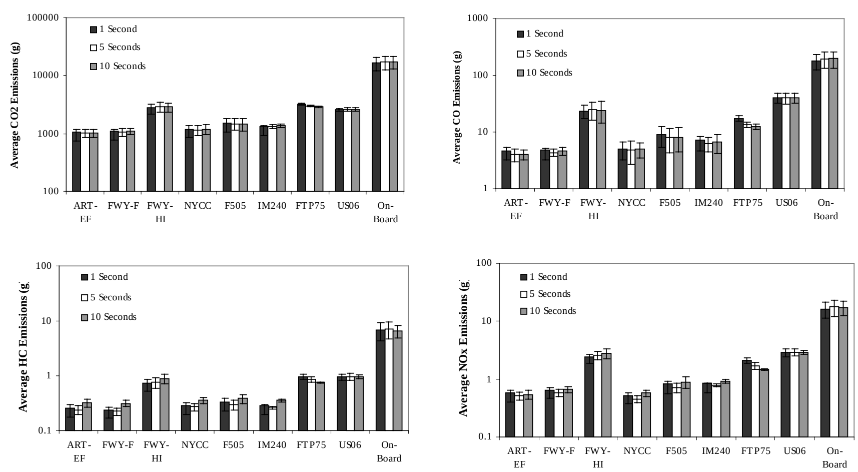
NC STATE UNIVERSITY

Average Modal Emission Rates for Each Mode Based Upon 10 Second-Averaged data



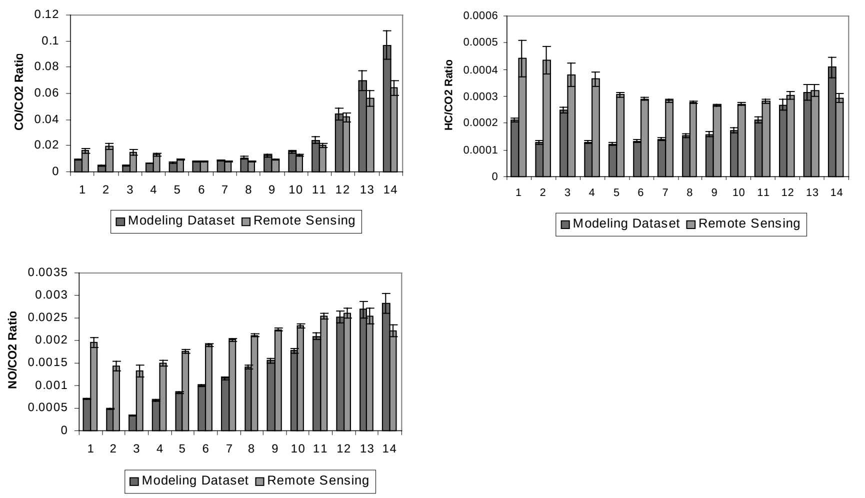
NC STATE UNIVERSITY

Comparison of 1, 5, and 10 Second Averaging Times



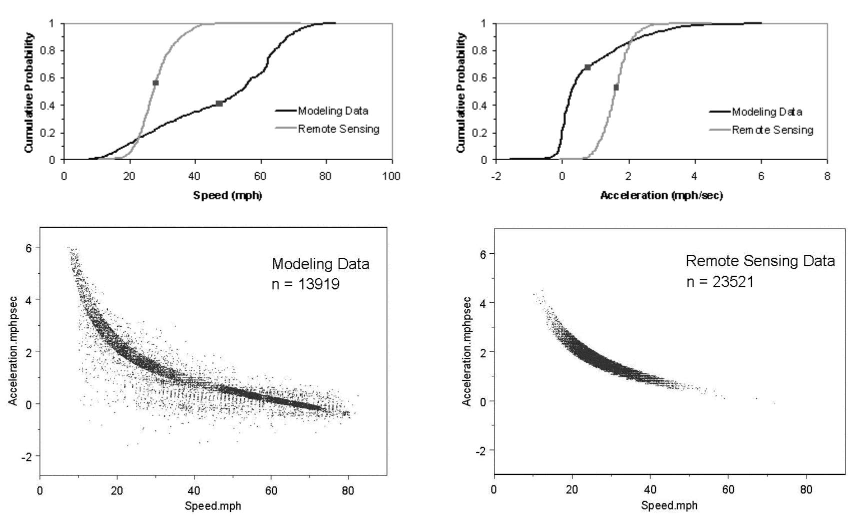
NC STATE UNIVERSITY

Comparison of Emission Ratios Based Upon the VSP Modal Approach: Modeling vs. RSD Data



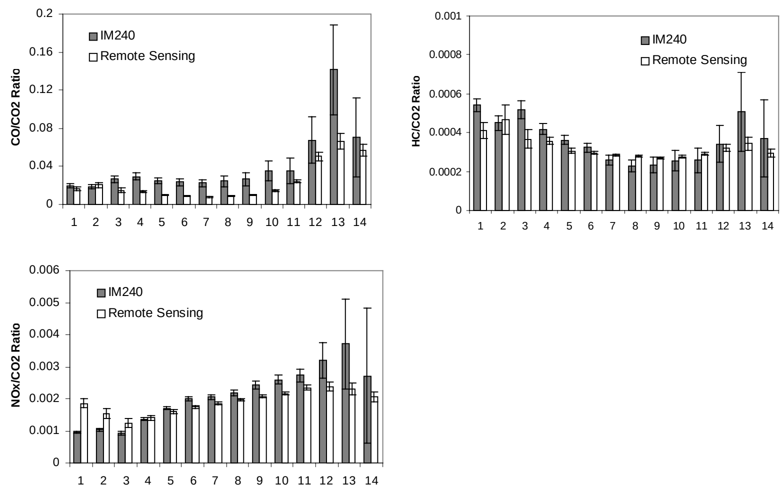
NC STATE UNIVERSITY

Comparison of Vehicle Activity for Modeling and RSD Data for VSP Mode 7



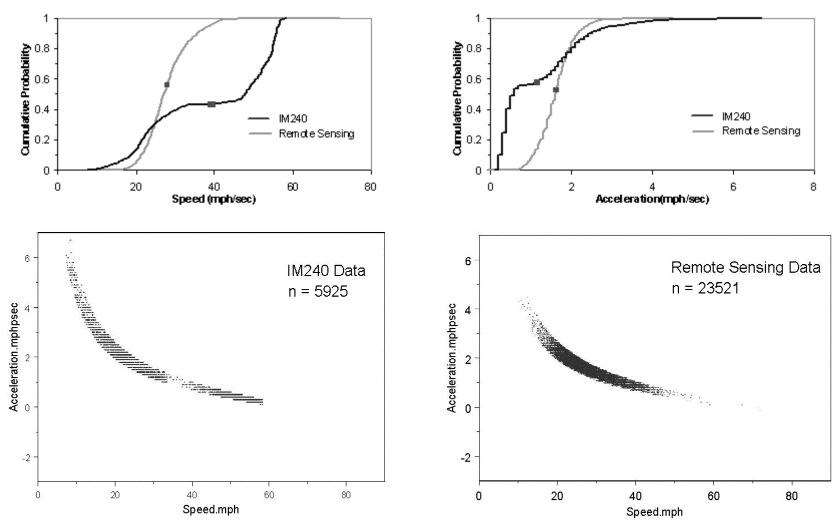
NC STATE UNIVERSITY

Comparison of IM240 and RSD Data



NC STATE UNIVERSITY

Comparison of Activity for IM240 and RSD Data for Mode 7



NC STATE UNIVERSITY

Comparison of Vehicle Activity for Modeling and RSD Data

- Remote Sensing site selection typically emphasizes positive road grades and/or positive accelerations
- RSD measurements are a snapshot at a specific location
- Narrower ranges of speed, and higher accelerations, on average, than for the modeling or IM240 database
- Comparisons between RSD and other types of data should be appropriately stratified to correct for these differences in activity
- Opportunities to improve VSP mode definitions by stratifying with respect to speed and/or acceleration are recommended and are described later

NC STATE UNIVERSITY

Summary of Findings

- NO_x/CO_2 sensitive to all modes
- RSD data produces much higher average ratio estimates for both HC and NO_x for the low VSP modes
- Differences could be because of
  - Fuel
  - vehicle characteristics
  - odometer reading (which is unobservable with RSD technology)
  - Activity (i.e. RSD data has lower speeds and higher accelerations)
  - Proportion of high emitting vehicles

NC STATE UNIVERSITY

Summary of Findings and Recommendations

- For NO_x, need a similar number of modes for the emission ratios and for mass per time units
- Need CO_2 (or fuel use) on a mass per time basis anyway
- RSD data could be used as an aid in recruiting vehicles for on-board measurements

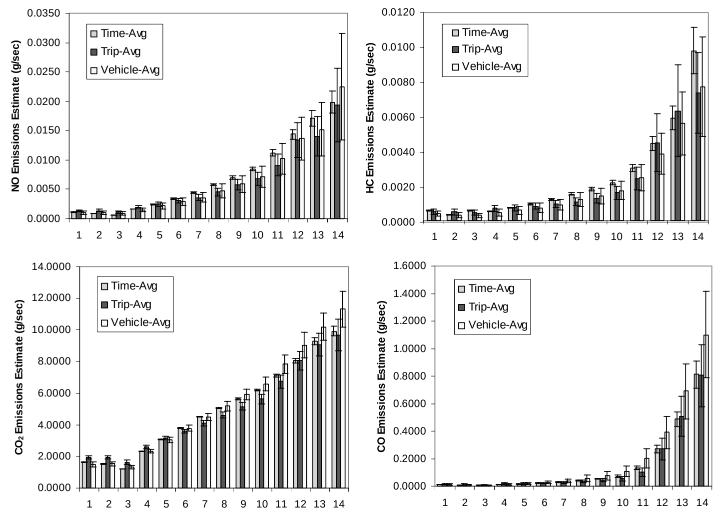
NC STATE UNIVERSITY

Comparison, Evaluation, and Selection of Weighting Approach

- Time Weighted (1 sec data)
- Vehicle Weighted
- Trip Weighted

NC STATE UNIVERSITY

Comparison of Uncertainty in VSP Bins: Time, Trip, and Vehicle Averages



NC STATE UNIVERSITY

Recommendation of Time Weighted Approach over Trip and Vehicle Weighted Approaches

- Compared to time-weighted approach, the means for the trip and vehicle weighted approaches can be either higher or lower.
- The standard deviation increases for the trip weighted approach, and further for the vehicle weighted approach.
- Averaging time varies for both trip and vehicle weighted approaches; there is no standard averaging time
- Uncertainty in the average typically increases with more averaging, because of smaller sample size
- Trip and Vehicle weighted approaches disproportionately give emphasis to trips or vehicles with little data

NC STATE UNIVERSITY

Variability and Uncertainty

- **Variability**: differences in values for different elements of a population over space and/or time
- **Uncertainty**: lack of knowledge regarding the true value of a quantity
- Examples: variability in emissions from one second to another versus uncertainty in the average emissions over all seconds.

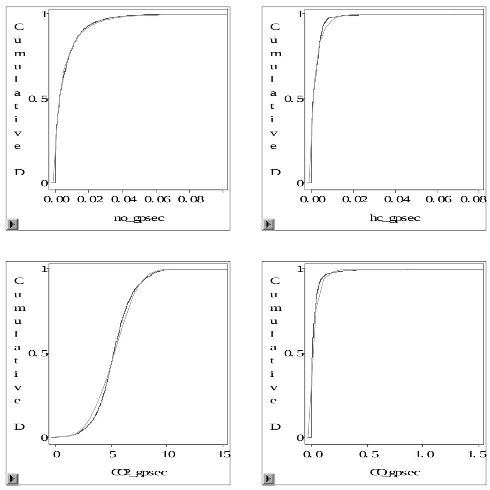
NC STATE UNIVERSITY

Characterization of Variability

- Empirical distributions: flexible, can represent any shape, but require a substantial number of estimates of cumulative probability
- Parametric distributions: compact representation of data, may facilitate software design and data storage, may not always provide a good fit
- Lognormal and Weibull distributions were used
- With large sample sizes, statistical goodness-of-fit tests are very sensitive and typically reject fits that are adequate or very good. Thus, judgment must be used.

NC STATE UNIVERSITY

Example Graph of Fitting Parametric Distributions for Variability in VSP Bins



VSP Bin 8, Time Average, Odometer > 50,000 miles, Engine Size < 3.5 liter

Empirical CDF

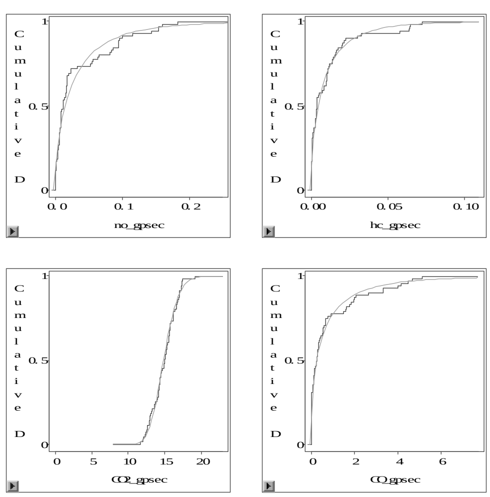
Fitted Parametric Distribution

Pollutant	NO	HC	CO_2	CO
Dist[a]	W	W	W	W

[a] L = lognormal; W = Weibull

NC STATE UNIVERSITY

Example Graph of Fitting Parametric Distributions for Variability in VSP Bins (IV)



VSP Bin 12, Time Average, Odometer > 50,000 miles, Engine Size > 3.5 liter

Empirical CDF

Fitted Parametric Distribution

Pollutant	NO	HC	CO_2	CO
Dist[a]	W	W	L	W

[a] L = lognormal; W = Weibull

NC STATE UNIVERSITY

Typical Parametric Distributions for VSP Bins Based on Different Stratus

Odometer (miles)	Engine Size (Liter)	NO	HC	CO_2	CO
< 50,000	< 3.5	W	L[a]	W[b]	L[c]
< 50,000	> 3.5	W[d]	L[e]	W[f]	L[g]
> 50,000	< 3.5	W	W[h]	W	L, W[i]
> 50,000	> 3.5	W	L, W[j]	L[k]	L, W[l]

L = Lognormal; W = Weibull

Exceptions:

[a] Weibull for Bin 11 and Bin 12; [b] Lognormal for Bin 4; [c] Weibull for Bin 5 to Bin 9; [d] Lognormal for Bin 13 and Bin 14; [e] Weibull for Bin 13 and Bin 14; [f] Lognormal for Bin 4, Bin 12 and Bin 14; [g] Weibull for Bin 12 to Bin 14; [h] Lognormal for Bin 3 to Bin 5; [i] Weibull for Bin 1 and Bin 4 to Bin 8, Lognormal for Bin 2 to Bin 3 and Bin 9 to Bin 14; [j] Lognormal for Bin 1 to Bin 8, Weibull for Bin 9 to Bin 14; [k] Weibull for Bin 1 to Bin 3; [l] Lognormal for Bin 1 to Bin 4 and Bin 10 to Bin 11, Weibull for Bin 5 to Bin 9 and Bin 12 to Bin 14

NC STATE UNIVERSITY

Quantification of Variability

- Goodness-of-fit was evaluated by comparing relative and absolute errors in the mean and standard deviation of the fitted distribution versus the data
- The results suggest that the fits are good for most of the pollutant/VSP mode/strata combinations
- When the fit is not good, it is typically because the data are a mixture of distributions

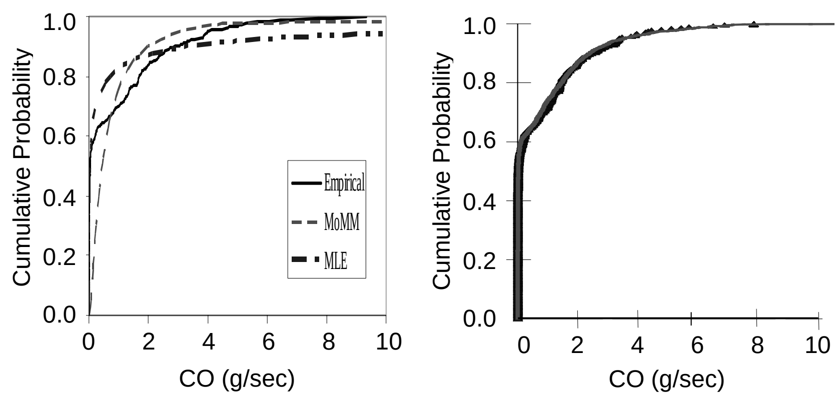
NC STATE UNIVERSITY

Quantification of Variability

- For situations in which the fits are not good, some alternatives include:
  - Use empirical distributions
  - Use mixture distributions
- Lack of good fit when characterizing distributions for variability may affect only a few modes for a given pollutant/strata, and therefore may not contribute substantial error to model predictions
- Estimates of uncertainty can be based directly upon the data instead of upon fitted distributions

NC STATE UNIVERSITY

Example of Mixture Distribution for Variability



Mode 14 CO Emissions, < 3.5 L, < 50,000 miles

NC STATE UNIVERSITY

Quantification of Uncertainty

- If the standard error of the mean (SEM) is small enough, and/or if the sample size is large enough, a normality assumption is a very reasonable approximation for the uncertainty in the mean
- Since the variability in the data is influenced by whatever random measurement errors exist, the estimate of uncertainty in the mean includes both random sampling error and measurement error

NC STATE UNIVERSITY

Quantification of Uncertainty

- Normality is a good assumption for uncertainty in the mean when:
  - Sample size is large enough (e.g., n > 40)
  - Relative standard error of the mean < 0.20
- These criteria were met for 213 of the 224 mode/pollutant emission rates
- Of the 11 cases that did not meet these criteria, there were only 5 cases where normality would not be an adequate approximation
- In situations where normality is not a good approximation, the numerical method of bootstrap simulation was used to estimate a sampling distribution for the mean
- As part of recent work for EPA/ORD, we have developed a software tool, AuvTool, for quantifying uncertainty in statistics, such as the mean for a data set or for a parametric distribution

NC STATE UNIVERSITY

Selected Examples of Use of Numerical Methods to Quantify Uncertainty in VSP Bins

Bootstrap Simulation was used when:

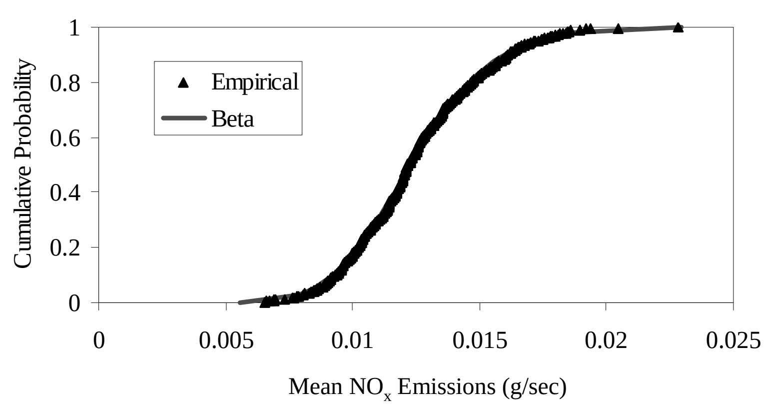
- n < 40
- SEM/mean > 0.2

Examples:

Odometer (miles)	Engine Size (Liter)	Bin	Pollutant	Mean (g/sec)	Relative Lower	Relative Upper
< 50,000	> 3.5	12	NO_x	0.0122	-38%	+46%
< 50,000	> 3.5	13	NO_x	0.0165	-48%	+73%
< 50,000	> 3.5	14	NO_x	0.0272	-36%	+49%
< 50,000	> 3.5	14	HC	0.0066	-33%	+39%
< 50,000	> 3.5	14	CO	0.935	-36%	+44%

NC STATE UNIVERSITY

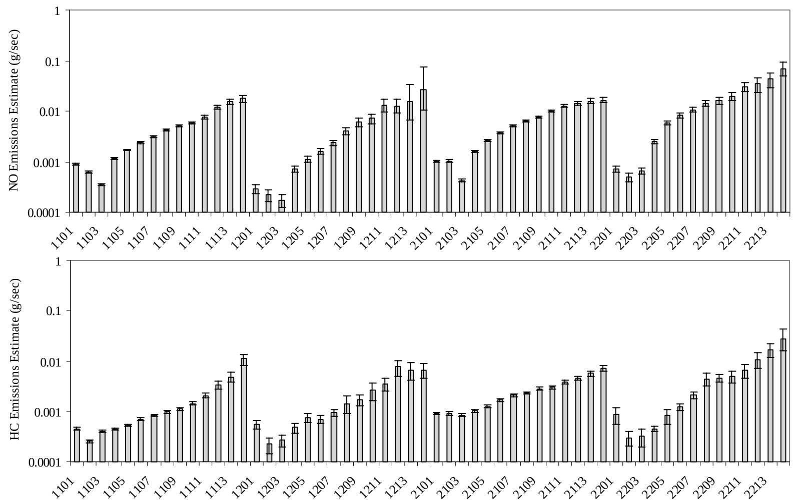
Characterizing Uncertainty When Normality is Not Applicable



Parametric distributions can be fit to the results of bootstrap simulation

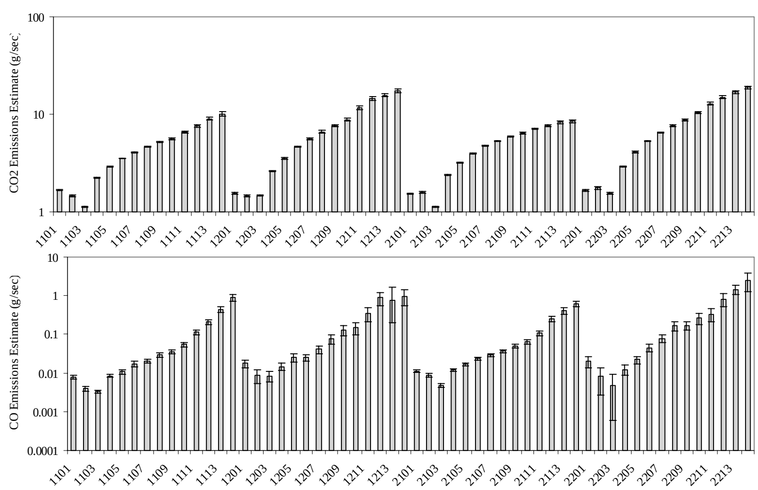
NC STATE UNIVERSITY

Quantified Uncertainty in VSP Bins: NO_x and HC



NC STATE UNIVERSITY

Quantified Uncertainty in VSP Bins: CO_2 and CO



NC STATE UNIVERSITY

Lowest and Highest Range of Quantified Uncertainty in VSP Bins

Odometer reading (miles)	Engine Size (liters)	Range	NO_x (%)	HC (%)	CO_2 (%)	CO (%)
< 50,000	< 3.5	Low	-4 to +4	-4 to +4	-1 to +1	-8 to +8
		High	-16 to +16	-24 to +24	-6 to +6	-18 to +18
< 50,000	> 3.5	Low	-12 to +12	-16 to +16	-2 to +2	-21 to +21
		High	-48 to +73	-39 to +39	-7 to +5	-36 to +44
> 50,000	< 3.5	Low	-4 to +4	-5 to +5	-1 to +1	-8 to +8
		High	-10 to +10	-13 to +13	-3 to +3	-19 to +19
> 50,000	> 3.5	Low	-9 to +9	-11 to +11	-1 to +1	-20 to +20
		High	-32 to +32	-36 to +36	-13 to +10	-87 to +87

NC STATE UNIVERSITY

Uncertainty and Averaging Time

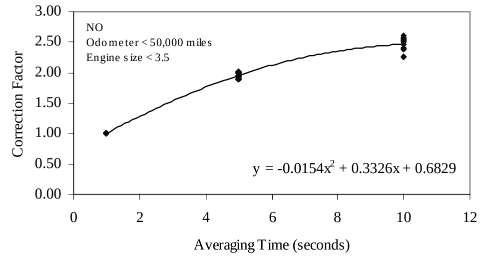
- Uncertainty is a function of averaging time
- 1 sec data - large sample size
- Based upon the 10-second database, it was possible to stratify the data into 5- and 1-second data to permit direct comparisons of averaging times
- With longer averaging times, the sample size decreases
- Typically, the range of uncertainty increases when going from 1-sec to 10-sec averages

NC STATE UNIVERSITY

Definition of Uncertainty Correction Factor for Averaging Time

Define Correction Factor:

$$CF_{t-\text{sec}} = \frac{SEM_{t-\text{sec}}}{SEM_{1-\text{sec}}}$$



where:

$CF_{t\text{-sec}}$: correction factor for t-second time period, no unit

$SEM_{t\text{-sec}}$: standard error of mean for t-second time period, g/sec

$SEM_{1\text{-sec}}$: standard error of mean for 1-second time period, g/sec

NC STATE UNIVERSITY

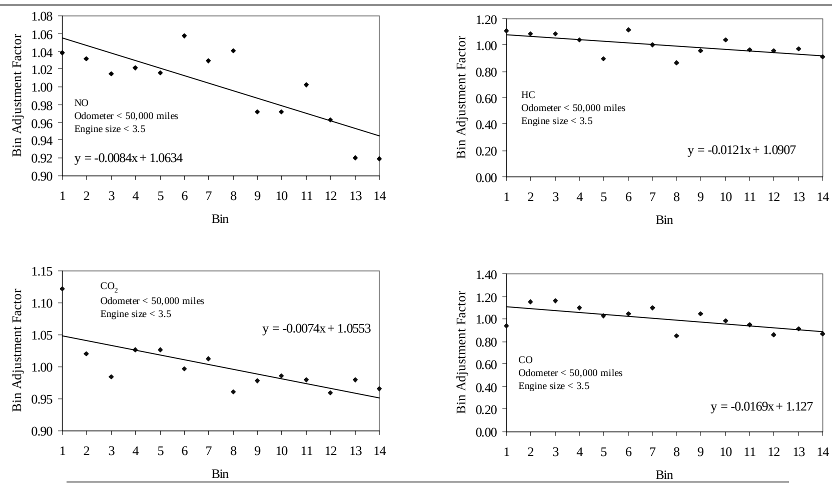
Proposed Uncertainty Correction Factor for Averaging Time

Odometer (miles)	Engine Size (Liter)	NO	HC	CO_2	CO
< 10 seconds					
< 50,000	< 3.5	$y = -0.0154x^2 + 0.3326x + 0.6829$	$y = -0.0167x^2 + 0.3638x + 0.653$	$y = -0.0186x^2 + 0.3608x + 0.6578$	$y = -0.0158x^2 + 0.3441x + 0.6717$
< 50,000	> 3.5	$y = -0.0152x^2 + 0.2935x + 0.7217$	$y = -0.0181x^2 + 0.3654x + 0.6527$	$y = -0.0246x^2 + 0.3802x + 0.6443$	$y = -0.0167x^2 + 0.3948x + 0.6219$
> 50,000	< 3.5	$y = -0.0163x^2 + 0.3479x + 0.6684$	$y = -0.0157x^2 + 0.3607x + 0.6549$	$y = -0.019x^2 + 0.3682x + 0.6508$	$y = -0.017x^2 + 0.3371x + 0.6799$
> 50,000	> 3.5	$y = -0.017x^2 + 0.3496x + 0.6674$	$y = -0.0168x^2 + 0.3687x + 0.6481$	$y = -0.0266x^2 + 0.398x + 0.6286$	$y = -0.0191x^2 + 0.3619x + 0.6572$
> 10 seconds					
< 50,000	< 3.5	2.47	2.62	2.40	2.53
< 50,000	> 3.5	2.14	2.50	1.99	2.90
> 50,000	< 3.5	2.52	2.70	2.43	2.35
> 50,000	> 3.5	2.46	2.65	1.95	2.37

Uncertainty and Averaging Time

- The range of uncertainty in the average emission estimate increases by a factor of 1.95 to 2.90 when comparing 1-second averages to 10-second averages
- Implications:
  - It is important to adjust for averaging time when making predictions of uncertainty
  - There is consistency in the values of these adjustments

Proposed Uncertainty Bin Adjustment Factor



The uncertainty correction factor is different for different bins

NC STATE UNIVERSITY

Propagation of Uncertainty

- Monte Carlo (numerical simulation) methods
  - Flexible: wide variety of distributions can be assumed for any input to the calculation
  - Sensitivity analysis can be applied to identify key sources of uncertainty
  - Potentially but not necessarily computationally intensive (depends on model run time)
- Analytical methods and approximations
  - Okay if normality assumption is reasonable
  - Less flexible
  - Simpler to compute

NC STATE UNIVERSITY

Example: Estimating Uncertainty in Total Emissions for the IM240 Cycle

VSP Mode Number	Seconds in Mode
1	41
2	24
3	16
4	37
5	47
6	19
7	29
8	17
9	4
10	3
11	3
12	0
13	0
14	0

NC STATE UNIVERSITY

Total Emissions for a Driving Cycle

$$TE = 41\times EF_{mode1} + 24\times EF_{mode2} + 16\times EF_{mode3} + 37\times EF_{mode4} + 47\times EF_{mode5} + 19\times EF_{mode6} + 29\times EF_{mode7} + 17\times EF_{mode8} + 4\times EF_{mode9} + 3\times EF_{mode10} + 3\times EF_{mode11}$$

Each Emission Factor is Assigned an Uncertainty Distribution

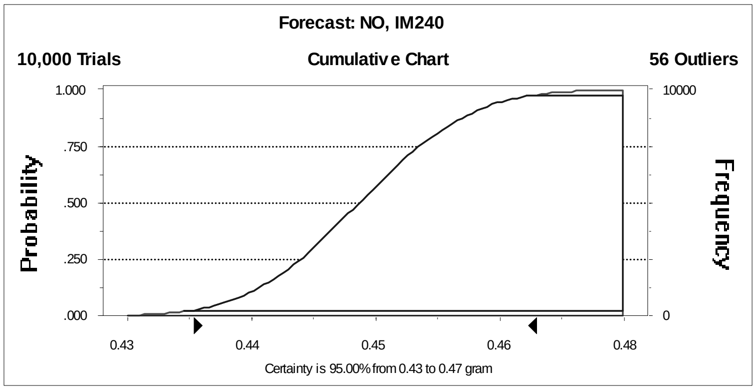
NC STATE UNIVERSITY

Inputs for Uncertainty Analysis Calculation

Mode	Distribution	Mean Emission Rate (g/sec)	Std. Dev. (g/sec)	Correction Factor	Bin Adjustment Factor
1	Normal	0.00090	2.0×10^{-5}	2.47	1.03
2	Normal	0.00063	2.0×10^{-5}	2.47	1.03
3	Normal	0.00035	9.3×10^{-6}	2.47	1.03
4	Normal	0.0012	2.5×10^{-5}	2.47	1.03
5	Normal	0.0017	3.6×10^{-5}	2.47	1.03
6	Normal	0.0024	5.1×10^{-5}	2.47	1.03
7	Normal	0.0031	6.9×10^{-5}	2.47	1.02
8	Normal	0.0042	9.4×10^{-5}	2.47	1.01
9	Normal	0.0051	1.4×10^{-4}	1.77	n/a
10	Normal	0.0059	1.7×10^{-4}	1.54	n/a
11	Normal	0.0076	3.0×10^{-4}	1.54	n/a

NC STATE UNIVERSITY

Example Results: Monte Carlo Simulation



Mean Total Emissions:	**0.45 g NO_x**
Absolute Uncertainty:	**±0.02 g**
Relative Uncertainty:	**±4.2%**

NC STATE UNIVERSITY

Example Results: Analytical Calculation

$$U_{total} = \sqrt{\sum_{i}^{n} (U_i \times W_i)^2}$$

- Where U_{total} = total uncertainty, U_i = uncertainty for mode *i*, and W_i = fraction of time in mode *i*.
- Results: ±0.018g or ±4%
- Similar to results from Monte Carlo simulation
- No uncertainty in weights

NC STATE UNIVERSITY

Summary of Uncertainty Ranges in Total Emissions for Selected Driving Cycles

Numbers in table represent ± percent ranges in mean total emissions

Vehicle Category	NO_x				HC				CO_2				CO			
	IM240	ART-EF	FTP	US06	IM240	ART-EF	FTP	US06	IM240	ART-EF	FTP	US06	IM240	ART-EF	FTP	US06
< 50,000 mi < 3.5 liters	4	4	4	7	6	6	6	13	0.8	0.9	0.8	1.2	11	10	11	16
< 50,000 mi > 3.5 liters	12	12	11	27	25	23	23	24	1.6	1.7	1.5	1.8	24	26	24	28
> 50,000 mi < 3.5 liters	4	4	4	5	6	7	6	6	0.9	1.0	1.0	1.0	8	8	8	15
> 50,000 mi > 3.5 liters	11	10	10	17	26	31	27	26	1.3	1.6	1.4	1.3	25	30	24	26

Uncertainty tends to be greater for:
- smaller sample sizes
- CO and HC

NC STATE UNIVERSITY

Recommendations: Quantification of Variability

- Parametric distributions are feasible and preferred over empirical distributions because they are a more compact summary of variability.
- The Method of Matching Moments is preferred
  - The mean and standard deviation of the fitted distribution will be the same as that of the data
  - distributions fitted using MoMM appear to provide a better fit to the upper tail of the distribution, compared to MLE.
- Single component distributions such as lognormal and Weibull distributions will typically be adequately for most modes.
- Two component lognormal mixture distributions are recommended when single component distributions are not adequate.
- Uncertainty analysis can be based directly upon the data and need not be based upon fitted distributions

NC STATE UNIVERSITY

Recommendations: Quantification of Uncertainty in Emission Factors

- Uncertainty estimates for mean emissions should be based directly upon the data; however, if fits are good, uncertainty analysis can be based either upon the data or upon the fitted distributions for variability;
- A normality assumption will typically be adequate for most modal emission rates;
- The assumption of normality should be tested;
- For cases in which a normality assumption is not valid, bootstrap simulation can be used, and a parametric distribution can be fit to the distribution of the means;
- The range of uncertainty in modal emission rates must be adjusted for different averaging times using an approach such as the correction factor and bin adjustment factor approach demonstrated here.

NC STATE UNIVERSITY

Recommendations: Quantification of Uncertainty in Total Emissions

- A simple analytical approach for estimating uncertainty in total emissions is adequate if:
  - uncertainty in modal emission estimates are approximately normal
  - no need to include uncertainty in vehicle activity in the estimate;
- The analytical method could be used as a quality assurance check on Monte Carlo simulation results;
- Monte Carlo simulation-based methods are recommended if the objective is to include uncertainty in activity;
- Monte Carlo can accommodate situations for normality assumptions are not be valid.
- The range of uncertainty is sufficiently large in many cases that a quantitative uncertainty analysis is well-justified.

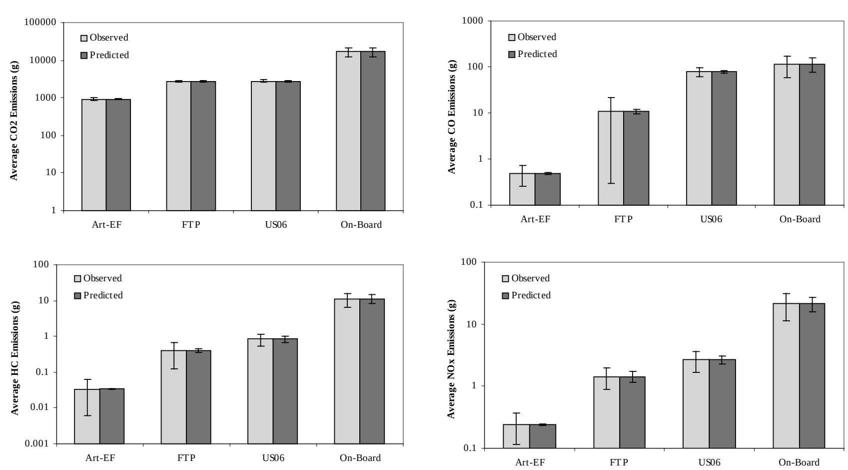
NC STATE UNIVERSITY

Verification and Validation of the Conceptual Model

- Case 1: Consistency Check
- Case 2: Comparison to Withheld Data
- Case 3: Comparison to CARB Data

NC STATE UNIVERSITY

Case 1: Consistency Check



NC STATE UNIVERSITY

Case 2: Comparison to Withheld Data

- Comparisons for Driving Cycles and On-Board Data for sufficiently large sample sizes:
  - FTP
  - US06
  - On-Board
- Different vehicles than those in the modeling database

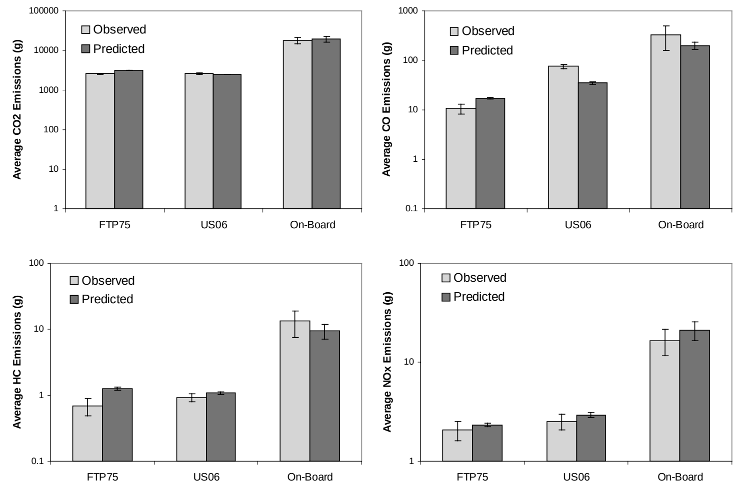
NC STATE UNIVERSITY

Case 2: Validation Dataset

Data Source	Cycle	NO. of Vehicles	No. of Trips	Total Seconds
EPA Dynamometer	ART-AB	2	2	1,471
EPA Dynamometer	ART-CD	2	2	1,255
EPA Dynamometer	ART-EF	3	3	1,507
EPA Dynamometer	FWY-AC	2	2	1,029
EPA Dynamometer	FWY-D	2	2	809
EPA Dynamometer	FWY-E	2	2	909
EPA Dynamometer	FWY-F	3	3	1,321
EPA Dynamometer	FWY-G	2	2	777
EPA Dynamometer	FWY-HI	3	3	1,825
EPA Dynamometer	LOCAL	2	2	1,047
EPA Dynamometer	NONFWY	2	2	2,693
EPA Dynamometer	NYCC	3	3	1,795
EPA Dynamometer	Ramp	2	2	529
NCHRP	FTP75	24	24	32,950
NCHRP	US06	21	21	12,648
On-Board Data	On-Board	3	18	19,243

NC STATE UNIVERSITY

Case 2: Comparison of Predicted and Observed Emissions



NC STATE UNIVERSITY

Case 2: Findings

- Good concordance in terms of the ordinal ranking of which cycles have the highest and lowest emissions.
- The predictions for CO, HC, and NO_x tend to be better when the prediction for CO_2 is also reasonably close.
- A comparison of CO_2 predicted and observed values may be a good diagnostic tool for identifying systematic differences between data sets.
- Validation Data Set 2 is more heavily weighted toward vehicles with smaller engines compared to the calibration data set.
- The systematic differences observed here for CO_2 suggest that additional refinement may be warranted for the engine displacement criteria when binning data
- It may have been better to withhold trips for the same vehicles, rather than validate with different vehicles.

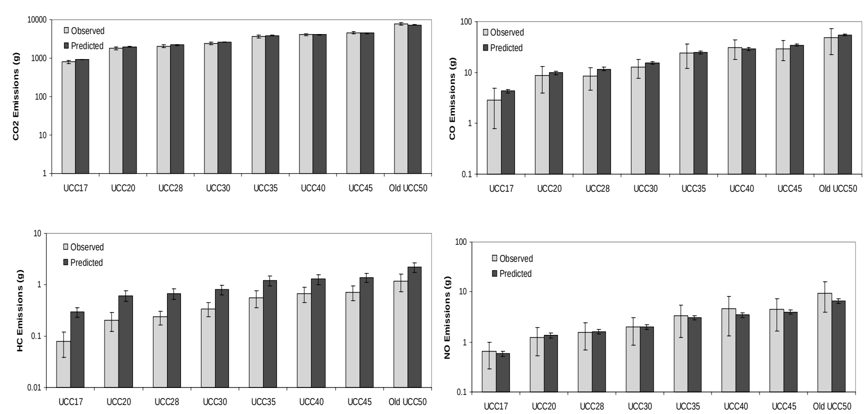
NC STATE UNIVERSITY

Case 3: Validation Dataset

Data Source	Cycle	NO. of Vehicles	No. of Trips	Total Seconds
ARB data	UCC17	17	17	7,174
ARB data	UCC20	17	17	15,048
ARB data	UCC25	17	17	15,372
ARB data	UCC30	17	17	17,712
ARB data	UCC35	17	17	24,318
ARB data	UCC40	17	17	24,012
ARB data	UCC45	17	17	23,472
ARB data	OLD UCC50	15	15	34,663
ARB data	MUC	4	20	46,760
ARB data	UCC50	2	4	8,768
ARB data	UCC60	2	4	11,240

NC STATE UNIVERSITY

Case 3: Comparison of Predicted and Observed Emissions



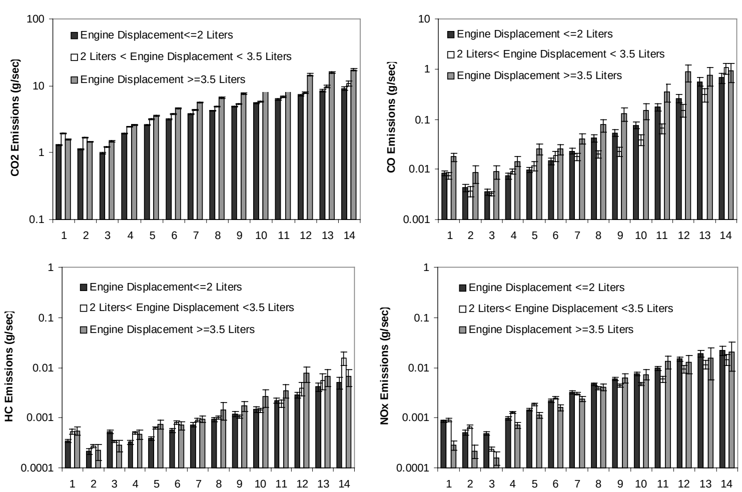
NC STATE UNIVERSITY

Case 3: Findings

- There was excellent agreement between the predicted and observed CO_2, CO, and NO_x emissions.
- There appears to be excellent concordance between the predicted and observed HC emissions.
- Possibility of TLEVs rather than Tier 1 was considered; however, suspected TLEVs had emission rates similar to known Tier 1 vehicles
- HC emissions may differ because of fuel, measurement method, or reporting (e.g., THC vs. NMHC)

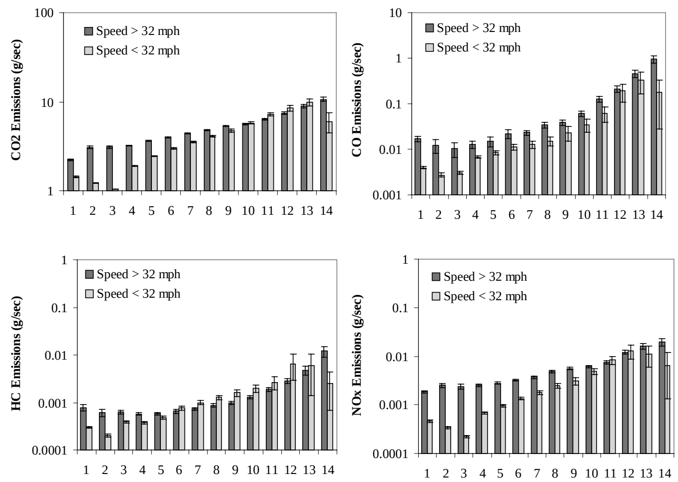
NC STATE UNIVERSITY

Potential Refinements to the VSP-Based Approach: More Engine Displacement Categories



NC STATE UNIVERSITY

Potential Refinements to the VSP-Based Approach: Speed Categories



NC STATE UNIVERSITY

Summary of Key Findings from Validation Case Studies

- The modal modeling approach is internally consistent.
- The model generally performs well for cycles or conditions that are represented by a large portion of the data in the modeling data set.
- The model is highly responsive, predicting a wide range of variability in average emissions.
- A promising approach for refining the modal modeling method is to consider speed as an additional explanatory variable.
- Comparisons of CO_2 emissions appear to be a good method for determining the comparability of two datasets.

NC STATE UNIVERSITY

Findings/Recommendations

- Modal binning approach, based upon VSP, engine displacement and odometer reading
- Speed should be considered as an additional binning criterion
- Road grade is accounted for in VSP
- Approach works for all four pollutants studied
- Supervised hierarchical tree-based regression is a useful modal development technique
- One second average data is useful for predicting total emissions
- Averaging time should be considered in uncertainty analysis
- Time-weighted approach is flexible and simpler; could structure model and data to enable vehicle or trip weighted calculations
- Mass per time emission factors needed for CO_2. Comparable number of modes needed for NO_x as for CO_2.
- For consistency, mass per time emission factors should be used for all pollutants.

NC STATE UNIVERSITY

Findings/Recommendations

- Parametric representations of variability are feasible and recommended
- Uncertainty in average modal emission rates can be treated as normal in most cases
- Bootstrap methods can be used to estimate uncertainty when normality is not valid
- Monte Carlo simulation offers a more flexible framework for incorporating uncertainty in activity and emissions inputs; analytical approach works well if only uncertainty in emissions rates are included
- The range of uncertainty in total emissions is sufficiently large to justify the need for uncertainty analysis
- It is important to have a representative data set, including high emitters. IM240 data may be critical for this purpose.
- Remote sensing data have a different activity pattern than other data considered, thus confounding any comparisons
- Statistical significance must be considered when making comparisons, such as for model development and validation.