CRC Report No. E-119-2

Hager Environmental and Atmospheric Technologies (HEAT) and Denver University (DU) Remote Sensing Device (RSD) Data Mining

August 2018

COORDINATING RESEARCH COUNCIL, INC.

5755 NORTH POINT PARKWAY • SUITE 265 • ALPHARETTA, GA 30022

The Coordinating Research Council, Inc. (CRC) is a non-profit corporation supported by the petroleum and automotive equipment industries. CRC operates through the committees made up of technical experts from industry and government who voluntarily participate. The four main areas of research within CRC are: air pollution (atmospheric and engineering studies); aviation fuels, lubricants, and equipment performance; heavy-duty vehicle fuels, lubricants, and equipment performance (e.g., diesel trucks); and light-duty vehicle fuels, lubricants, and equipment performance (e.g., passenger cars). CRC's function is to provide the mechanism for joint research conducted by the two industries that will help in determining the optimum combination of petroleum products and automotive equipment. CRC's work is limited to research that is mutually beneficial to the two industries involved. The final results of the research conducted by, or under the auspices of, CRC are available to the public.

Final Report for CRC Project No. E-119-2

Hager Environmental & Atmospheric Technologies (HEAT) and Denver University (DU)
Remote Sensing Device (RSD) Data Mining

Prepared for

Coordinating Research Council, Inc.
5755 North Point Parkway, Suite 265
Alpharetta, Ga 30022

Prepared by

Charles L. Blanchard
526 Cornell Avenue
Albany, CA 94706
charles.ld.blanchard@gmail.com
cbenvair@pacbell.net
(510) 525-6231

August 17, 2018

PREFACE

Remote sensing systems have added greatly to knowledge of motor vehicle exhaust emissions in real-world driving conditions. The author thanks Dr. Gary Bishop, Dr. J. Stewart Hager, and the research teams that conducted CRC projects E-106 and E-119 and other studies. It has been a privilege to support their long-term research and development programs. I also thank CRC Deputy Director Amber Leland and Administrative Research Assistant Rebecca Bougher, who greatly facilitated this project, and the members of the CRC Emissions Committee, who initiated this project and provided valuable direction, review, and feedback.

CONTENTS

FIGURES

TABLES

ACRONYMS AND ABBREVIATIONS

BAR	California Bureau of Automotive Repair
CO	carbon monoxide
CO_2	carbon dioxide
CRC	Coordinating Research Council
DiAL	Differential Absorption LiDAR
DU	Denver University
EDAR	emissions data and reporting system
EPA	Environmental Protection Agency
FEAT	fuel efficiency automobile test system
g	gram
HEAT	Hager Environmental & Atmospheric Technologies
HC	hydrocarbons
HDV	heavy duty vehicle
IR	infrared
kg	kilogram
LDT	light duty truck
LDV	light duty vehicle
LiDAR	Light detection and ranging
MOVES	motor vehicle emission simulator model
NH_3	ammonia
NMHC	nonmethane hydrocarbons
NO	nitric oxide
NO_2	nitrogen dioxide
NO_x	oxides of nitrogen ($NO + NO_2$)
SO_2	sulfur dioxide
UV	ultraviolet
VIN	vehicle identification number
VOC	volatile organic compounds
VSP	vehicle specific power

SUMMARY

Remote sensing measurements provide information about vehicle emissions in "real-world" situations, which can be used to evaluate the accuracy and realism of emission inventories or models. Credibility of remote sensing measurements is enhanced by cross-comparisons that demonstrate consistency and reproducibility of measurements made by different measurement systems.

The purpose of this project is to compare and contrast data from on-road vehicle exhaust measurements that were made in Chicago during September 2016 using two ground-level remote-sensing measurement systems: (1) the Hager Environmental & Atmospheric Technologies (HEAT) emissions data and reporting (EDAR) system (CRC E-119) and (2) the Denver University (DU) fuel efficiency automobile test (FEAT) system (CRC E-106). This project report examines the comparability of the EDAR and FEAT measurements of carbon monoxide (CO), nitric oxide (NO), and hydrocarbon (HC) concentrations, emission rates, and ratios of these species to carbon dioxide (CO_2) concentrations. The data are segmented to evaluate the comparability of the measurement systems within data subsets. Measurement differences are considered in relation to statistical significance, measurement accuracy, exhaust plume variability, and sample means. Further comparisons are made to evaluate the comparability of the two measurement systems with respect to vehicle speed and acceleration, sampling completeness (fraction of missed events), and fraction of high-emitting vehicles. Most of the comparisons are blinded to ensure confidentiality. Results published in previous reports are not blinded.

Near-simultaneous measurements were made by the two measurement systems on 4728 vehicles over three days. The measurement systems were located 106 feet from each other and the time delay was ~2 – 3 seconds between same-vehicle exhaust measurements. Since the duration of a plume measurement was less than one second for each system, the two systems did not measure identical exhaust plumes. For vehicles with consistently low emissions, differences of ~2 – 3 seconds between measurements are not expected to introduce large differences, but for other vehicles the observed measurement differences potentially reflect inherent plume variability rather than systematic differences in measurement systems.

The median concentrations obtained by one measurement system in this study were 200 parts per million by volume (ppmv) for CO (0.02%), 16 ppmv for HC, and 5 ppmv for NO. For the second measurement system, the medians were 202 ppmv for CO (0.02%), 24 ppmv for HC, and 12 ppmv for NO. The 90th percentile concentrations obtained by one measurement system were 1930 ppmv for CO (0.19%), 214 ppmv for HC, and 118 ppmv for NO. For the second measurement system, the 90th percentiles were 1688 ppmv for CO (0.17%), 53 ppmv for HC, and 151 ppmv for NO. Comparability of measurements from the two systems was evaluated through consideration of measurement uncertainty, variability of same-vehicle measurements, paired differences, and consistency of speed and acceleration.

Measurement accuracy is established through independent audits using certified standards, which have been reported for both systems in previous studies. No independent audits were conducted

in this study or in CRC E-106 and CRC E-119. Measurement variabilities were instead estimated for CO, NO, and HC concentrations by reviewing published estimates and by re-applying a statistical approach described in CRC E-106 and CRC E-119. The approach yields a measure of precision, or variability, denoted here as σ_0, which is the observed standard deviation from the LaPlace probability density function. Values of σ_0 serve as indicators of instrument noise (Burgard et al., 2006a) and are also indicative of measurement detection limits. The values obtained for σ_0 for one measurement system in this study were 1223 ppmv for CO (0.12%), 214 ppmv for HC, and 30 ppmv for NO. For the second measurement system, the values were 78 ppmv for CO (0.008%), 38 ppmv for HC, and 13 ppmv for NO. These uncertainty estimates are consistent with previously published σ_0 values and are comparable to published detection limits. Median concentrations were less than measurement uncertainties (with one exception) and 90th percentiles were greater than measurement uncertainties (with one exception).

Repeated measurements provide an opportunity to characterize vehicle-specific variations, as previously discussed in Bishop and Haugen (2017), Haugen and Bishop (2018), Hager (2018), and Ropkins et al. (2017). Repeated measurements are reported here to provide a measure of same-vehicle variability as a benchmark for comparison of differences in measurements made by the two instruments. Quantifying same-vehicle variability aids in evaluating the differences between the measurements made by the two sampling instruments, because the instruments did not measure exhaust plumes at identical times. The combined data set, which consisted of 4728 paired measurements, included multiple measurements of 586 vehicles. Pre-2004 model years tended to exhibit the largest differences in repeat-vehicle species concentrations. This result was less pronounced for HC than for CO and NO, and some larger repeat-vehicle differences occurred even on later model-year vehicles. For one measurement system, the concentration ranges that covered 90% of the repeat-vehicle differences were ± 0.25% for CO, ± 93 ppmv for NO, and ± 255 ppmv for HC. For the other measurement system, the concentration ranges that covered 90% of the repeat-vehicle differences were ± 0.19% for CO, ± 170 ppmv for NO, and ± 59 ppmv for HC. There was no evidence that larger repeat differences were associated with longer time differences between measurements. The larger repeat concentration differences were not systematically associated with large differences in speed or acceleration between repeated measurements, although some individual vehicles exhibited both large repeat concentration differences and large acceleration differences. Comparable variability in repeat measurement differences occurred across all observed operating bins of EPA's motor vehicle emission simulator model (MOVES). The highest ranges of repeat measurement concentrations occurred for vehicles with intermittently high emissions, i.e., vehicles having one or more low concentrations and one or more high concentrations. Some of the highest-emitting vehicles were consistently high emitters, and the repeat differences were not as large as for some of the intermittent high emitters. Vehicles with the smallest differences in repeated measurements were consistently low emitters.

The two measurements systems yielded similar statistical distributions of vehicle speed, acceleration, concentrations of CO, NO, and HC, and ratios of concentrations of CO, NO, and HC relative to CO_2. The distributions of paired measurement system concentration differences (near-simultaneous same vehicle measurements) were comparable to the distributions of the

differences in repeated measurements: ~90% of the paired comparisons fell within the ranges ± 0.2% for CO and ± 200 ppmv for NO and HC. Thus, the variability of the differences in paired measurements could have been due to inherent variability in exhaust plume concentrations. The ranges of pairwise measurement differences increased with vehicle age, just as did the ranges of differences in repeat-vehicle measurements.

A paired t-test was used to test the statistical significance of paired instrument measurement differences. The large sample size (N = 4728) ensured that a paired t-test had very high statistical power, identifying statistically significant differences at $p < 0.05$ (95% confidence level) with a probability of 93% when the mean difference between paired measurements was only 5% of the standard deviation of the paired differences. Small measurement differences were therefore both detectable and detected.

Differences in paired measurements of speed, acceleration, and concentrations of CO, NO, and HC were small relative to sample averages, measurement accuracy, and the inherent variability of exhaust plumes. While statistically significant ($p < 0.05$), differences in paired vehicle speed measurements averaged only 1 mph, or ~4% of the average speed of 25 mph. Paired vehicle acceleration differences were also statistically significant and averaged 0.6 mph s^{-1}. This difference corresponds to a modest rate of acceleration.

The mean paired concentration differences between measurement systems (A – B) were -0.002% CO (not significant), -6.9 ppmv NO (significant, $p < 0.05$), and 10.0 ppmv HC (significant, $p < 0.05$). Ranges of ± 0.2% for CO and ± 200 ppmv for NO and HC encompass ~90% of the paired comparisons. For individual measurements, the mean differences would be considered within the measurement limitations of the systems. When averaged over 4728 measurements, the mean NO and HC differences were statistically significant. When expressed as fuel-based emission rates, the mean paired differences between measurement systems were -0.05 g kg^{-1} CO (not significant), -0.87 g kg^{-1} NO (significant, $p < 0.05$), and 0.40 g kg^{-1} HC (significant, $p < 0.05$).

Larger emission differences between measurements made by the two instruments were not systematically associated with large differences in speed or acceleration. The largest average NO differences were associated with vehicles classified as heavy-duty vehicles (HDV, gross vehicle weight range 4578 – 8687 pounds, N = 84). However, the average HDV A – B difference was positive (150 ppm), even though the average A – B NO difference among all 4728 vehicles was negative (-12.2 ppmv).

The two measurement systems consistently indicated that five percent of the vehicles emitted approximately 50 to 70% of the CO and NO mass and about 20 to 30% of the HC mass. Among post-2004 vehicles (N = 3909), four vehicles (0.1%) emitted 3% of the CO mass, 9% of the NO mass, and 2% of the HC mass.

1. INTRODUCTION

1.1 Background

Accurate values of air pollutant emission rates are needed to assess emission source contributions to ambient air pollution, evaluate the effectiveness of past air quality management actions, and predict the effects of new air quality measures. Emission estimates are generally thought to have become increasingly accurate for most major air pollutants, including carbon monoxide (CO), oxides of nitrogen (NO_x= NO + NO_2), sulfur dioxide (SO_2), and volatile organic compounds (VOC, a subset of hydrocarbons [HC]), emitted by point sources and vehicles (Miller et al., 2006). However, emission factors (i.e., emission rates per unit of activity, such as grams per mile of highway travel) are known to vary by season, day, and hour depending on types of emission controls deployed, the intermittency or constancy of controls, and control efficiency (e.g., as a function of temperature or vehicle age). Activity levels also vary by season, day, and hour (e.g., Chinkin et al., 2003), and these temporal variations differ among locations (e.g., geographically, urban versus rural, by terrain). "Bottom-up" calculations used to generate emission estimates are subject to significant uncertainties.

Analyses of ambient measurements have been used to compute emission factors or overall emissions occurring within specified geographical domains over specific periods of time. These analyses have been compared with emission inventories or models to evaluate their accuracy in comparison to "real-world" situations. The types of air quality measurements that have been used for comparison with inventory or emission model estimates include (1) monitoring data from ground sites (e.g., Fujita et al., 1992; Parrish, 2006), (2) data from highway tunnels (e.g., Ingalls, et al., 1989; Lawson, 1990; Pierson et al., 1990; Singer and Harley, 1996; Dallman and Harley, 2010; McDonald et al., 2012; Fujita et al., 2012), (3) ground-based remote sensing data (e.g., Bishop and Stedman., 2008; 2015; Bishop et al., 2010; 2012; 2015; Burgard et al., 2006b), (4) data from aircraft sampling (e.g., Castellanos et al., 2011; Brioude et al., 2013; He et al., 2013; Anderson et al., 2014; Goldberg et al., 2014; Huang et al., 2014; Travis et al., 2016), and (5) satellite data (e.g., Streets et al., 2013; Canty et al., 2015).

Over the years, many of the measurement studies have identified needed improvements in emissions reported in various versions of the U.S. Environmental Protection Agency (EPA) National Emission Inventory (NEI). Whereas older studies using ambient data from the late 1980s and early 1990s often found higher real-world vehicle emissions compared with emission estimates, recent published reports and articles suggest that NEI emissions of NO_x are too high by about 25 to 100% since about 2010 (Appendix). Most recent studies infer that EPA mobile source NO_x emission estimates are too high and suggest that either EPA's Mobile Source Vehicle Emissions Simulator (MOVES) model, or the inputs used in its operation, overestimate mobile source NO_x emissions. It is also possible that other MOVES pollutant emission rates are higher than real-world vehicle emissions across the on-road vehicle fleet.

The initial version of MOVES (versions 2010, 2010a, and 2010b) was substantially modified in 2014. Heiken et al. (2016) reviewed the methods, data, and assumptions of MOVES (version 2014), as well as the emission inventories prepared using MOVES2014, and recommended

adjustments to account for variability of the effectiveness of NO_x control by selective catalytic reduction (SCR), more realistic NO_x start exhaust emission rates, phase-in of new heavy-duty diesel NO_x emission rules, and other adjustments. Some but not all recommendations were implemented in MOVES2014a (Heiken et al., 2016; EPA, 2017). Related work has focused on improving MOVES inputs by developing better spatial and temporal resolution of activity levels (e.g., Lindhjem et al., 2012; DenBleyker et al., 2017). Real-world measurements continue to reveal important considerations; e.g., MOVES assumes that the most rapid change in emissions as vehicles age occurs between 4 and 9 years (EPA, 2015), but Bishop and Haugen (2017) report that model-year emissions of CO and HC only increase after about 10 to 12 years.

Credibility of the measurement studies is enhanced by cross-comparisons that demonstrate consistency and reproducibility of the real-world measurements across different approaches. The present work is intended to evaluate the reproducibility of two real-world measurement systems. Both are ground-based remote-sensing systems.

1.2 Project Objectives

The primary objective of this project is to compare and contrast on-road vehicle exhaust emission measurements made simultaneously using two roadway remote-sensing measurement systems. Results will be of interest for determining the extent to which multiple on-road measurement approaches provide:

- Different types of information,
- Enhanced confidence in the accuracy and representativeness of direct on-road measurements if measurements agree within experimental error and variability,
- Measurements that are useful for evaluating the abilities of emission models and inventories to represent real-world driving situations.

1.3 Measurement Location

On-road vehicle exhaust measurements were made in Chicago by two research groups during September 2016 using two ground-level remote-sensing measurement systems: (1) the Hager Environmental & Atmospheric Technologies (HEAT) emissions data and reporting (EDAR) system (Hager, 2018) and (2) the Denver University (DU) fuel efficiency automobile test (FEAT) system (Bishop and Haugen, 2017). This report examines the comparability of the EDAR and FEAT measurements of carbon monoxide (CO), carbon dioxide (CO_2), nitric oxide (NO), and hydrocarbons (HC). Most of the comparisons are blinded to ensure confidentiality. Results published in previous reports are not blinded.

The measurement site is located ~14 kilometers northwest of Chicago O-Hare Airport along the on-ramp from Algonquin Road to eastbound I-290 / Highway 53 (Figure 1). The FEAT system was situated 106 feet up-ramp from the EDAR location so that the FEAT and EDAR measurement systems would not interfere with each other (Haugen and Bishop, 2018). The time stamps on the measurements indicate that the EDAR measurements were on average 1.5 seconds later than the corresponding FEAT measurements. The EDAR and FEAT clocks were not synchronized, and the average vehicle speed of ~25 miles per hour (mph), or ~ 37 feet per second, implies a time difference of ~3 seconds if the measurement devices were exactly 100 feet apart. Because the systems were not exactly co-located, they did not measure the exact same exhaust plumes. The potential consequences for the comparisons are discussed throughout this report. Here, it is noted that the duration of a measurement is less than one second for each system (Bishop and Haugen, 2017; Hager, 2018). Since exhaust plumes are variable, some differences between measurements likely reflect inherent variability of the plumes, rather than instrumental differences (Ropkins et al., 2017). This effect is discussed in later sections.

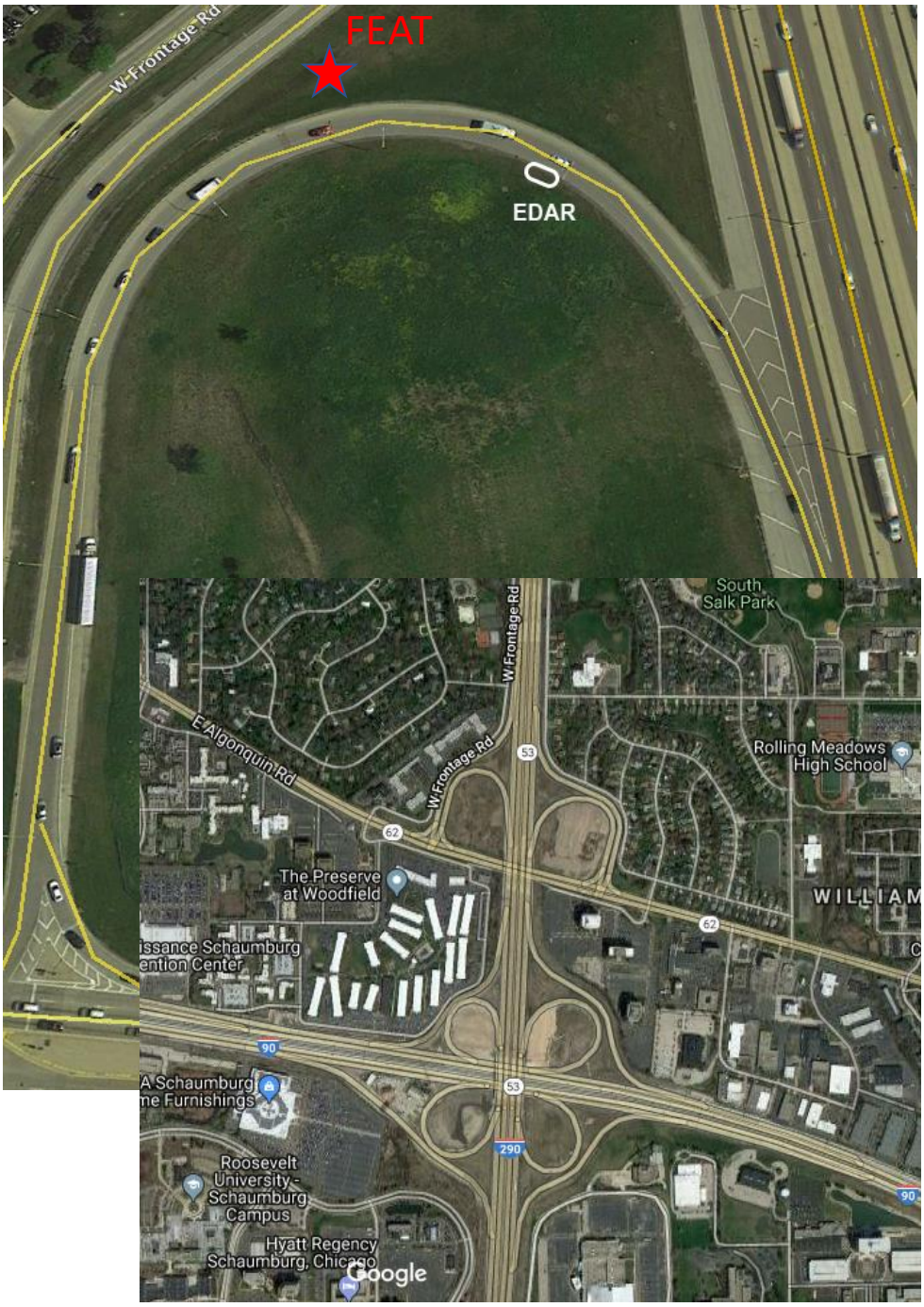


Figure 1. Measurement location. Instrument positions are approximate. Source: Bishop and Haugen (2017), Hager (2018), Haugen and Bishop (2018).

2. APPROACH

2.1 Data Acquisition and Assessment

Data sets were obtained from the principal investigators of CRC projects E-106 and E-119, either directly or through CRC. Each of the research teams had carried out quality assurance and control procedures on their own data sets as described in Bishop and Haugen (2017), Haugen and Bishop (2018), and Hager (2018). In summary, reasons that measurements were considered invalid or were excluded from the data sets by the research teams were:

- Sensing beam did not intercept exhaust plume,
- Measurement error exceeded pre-specified threshold,
- Rain interference,
- Out of state or unmatched vehicle license plates.

We reviewed the summary statistics listed in CRC project reports E-106, E-119, and E-119a to determine the fraction of vehicles having valid measurements compared to the total number of vehicles that passed the measurement locations. Both systems achieved ~90% valid measurements relative to measurement attempts. About 8% of vehicles had license plates from a state other than Illinois (Hager, 2018). A total of 9830 Illinois vehicles out of 13,985 records submitted to the state by EDAR were identified by the Illinois Environmental Protection Agency, Air Quality Planning Section (Hager, 2018) and 9948 Illinois vehicles were identified out of 10,039 records submitted to the Illinois Secretary of State by FEAT (Haugen and Bishop, 2018). The details of the state vehicle identification process are available from the authors of Haugen and Bishop (2018) and Hager (2018). Beginning at 11:33 a.m. on September 19, the FEAT E-119a data set consists of 9631 records. There were therefore 9830 records in the EDAR data set and 9631 records in the FEAT data set available for the comparison (Table 1).

The FEAT and EDAR data did not overlap completely, as indicated by day in Table 1. EDAR measurements began earlier in each day than did FEAT. Rain interferes with measurements and sampling on September 21 was limited by periods of rain (Hager, 2018).

Table 1. Number of records in EDAR and FEAT data sets by period and day.

Date	EDAR N	EDAR Start Time	EDAR Stop Time	FEAT N	FEAT Start Time	FEAT Stop Time
Sep 12-16	**NA**	**NA**	**NA**	**20,431**	**Morning**	**~18:00**
Sep 19-21	**9830**	**NA**	**NA**	**9631**	**Morning**	**18:00**
Mon 9/19	3969	11:33	18:21	3487	11:33	18:00
Tue 9/20	4120	5:40	18:13	4401	9:12	18:00
Wed 9/21	1741	5:11	20:18	1743	10:24	18:00

Before merging the FEAT and EDAR data sets, the data records were checked for entries that were obviously incorrect. Since data quality had previously been evaluated by each research team, extensive additional data validation was not needed. During this check, three pairs of duplicate entries (all values in each member of a pair were identical) were found in one data set and the second member of each duplicate pair was removed. The other data set had 11 pairs of records in which each pair member had the same vehicle identification number (VIN), day, and time but pair members had nonidentical species concentrations. In addition, this data set had 13 records with the same VIN (license plate "RAV"). The first record was retained for each set of duplicates in the second data set (thus retaining 12 of the records). Dropping all but the first member of each set of duplicates had minimal impact on later comparisons, because only two of the duplicated records matched data in the comparison data set; the matched records had consistent data values. After deleting duplicates, the FEAT and EDAR data sets contained 9628 records and 9807 records, respectively.

The FEAT and EDAR data were merged by VIN, day, and time (±2 seconds). Three successive data merges, each with a different time offset centered around the average 1.5 second difference, were used to ensure completeness. The merged data set contained 4728 records, a sufficiently large sample for statistical analysis. The smaller number of records in the merged data set compared with the data sets provided by FEAT and EDAR is due to differences in the number of hours sampled (Table 1) and to hour-by-hour differences in numbers of vehicles with license plates identified by the state of Illinois.

2.2 Measurements

Both data sets include measurements of vehicle speed, acceleration, and vehicle specific power (VSP), which is a measure of engine load that can be computed from the observed speed and acceleration by using the equations in Section 3.2. Units of measurement were converted to provide a consistent basis of comparison and the results are discussed in Section 3.

The air pollutant species measurements included in the FEAT data set were CO, CO_2, ammonia (NH_3), NO, NO_2, and HC, whereas the species measurements included in the EDAR data set were CO, CO_2, NO, and non-methane hydrocarbons (NMHC). The FEAT NH_3 and NO_2 measurements are not discussed further, because comparison measurements are not available in the EDAR data set. Unlike CO and NO, each of which is a single species, the FEAT HC and EDAR NMHC measurements each represent sums of many species. The measurement systems' sensitivities to each of the large number of hydrocarbon species present in the atmosphere differ. The EDAR NMHC data represent hydrocarbon concentrations excluding methane. The FEAT HC data represent hydrocarbon concentrations including methane, but the FEAT instrumental response to methane is limited (Singer et al., 1998) so the FEAT HC concentrations are not equivalent to methane plus NMHC.

FEAT concentrations of CO_2, while reported, are not an independent measurement (Bishop and Haugen, 2017). The reason that FEAT CO_2 concentrations are not an independent quantity is related to the way that FEAT makes measurements. FEAT measures absorption at wavelengths in the ultraviolet (UV) and infrared (IR) portions of the electromagnetic spectrum where absorption is specific to certain molecules. FEAT converts absorption signals to concentrations using the ratios CO/CO_2, NO/CO_2, and HC/CO_2, which are obtained as regression slopes. Data are reported as volume percent concentrations (1 percent = 10^4 parts per million by volume [ppmv]) with a calculation that is intended to generate the value that would be reported by a tailpipe probe, which is carried out through consideration of the stoichiometry of combustion (Bishop and Haugen, 2017). The reported FEAT CO_2 concentrations are constrained by the calculation.

The FEAT data set included measurements of CO, NO, and HC reported as volume percent and as grams of pollutant per kilogram of fuel. The FEAT conversion from ratios of CO/CO_2, NO/CO_2, and HC/CO_2 to grams per kilogram fuel is a simple multiplicative factor derived from molecular weights and a fixed ratio of carbon mass to fuel mass (Bishop and Haugen, 2017).

Ropkins et al. (2017) describe key features and differences between the FEAT and EDAR systems. EDAR uses infrared lasers and differential absorption light detection and ranging (DiAL) to measure gases (Hager, 2018). Whereas light detection and ranging (LiDAR) technology can detect but not quantify, the DiAL method in conjunction with scanning the full exhaust plume allows for the quantification of gases. Due to the absolute nature of EDAR's spectroscopic measurements, EDAR is designed to measure targeted pollutants without explicit field calibration and remain within normal specifications. The EDAR unit is located approximately 5 meters above the roadway and looks down on the plume, so the height of the tailpipe is inconsequential. The unit can detect both heavy and light duty vehicles. The gas

sensor can detect an entire exhaust plume as it exits a vehicle, which allows for high signal to noise ratio (SNR) and measurement of absolute amounts of pollutants. EDAR processes measurements to generate emission rates in mass per unit travelled (grams per mile or grams per kilometer).

The EDAR data set included measurements of CO, NO, and NMHC expressed as parts per million by volume (mixing ratios) and as grams of pollutant per mile. The EDAR data set also included values of CO_2 expressed as grams per mile, which are obtained as an independent measurement. EDAR obtains tailpipe-equivalent concentrations through regression of pollutant masses against CO_2 mass in a manner similar to FEAT.

I converted FEAT and EDAR measurements to comparable concentration units (volume percent and ppmv, which are both used in this report depending on scale). In addition, I determined ratios of CO/CO_2, NO/CO_2, and HC/CO_2 for each data set. For FEAT, I computed ratios from the species concentrations (mixing ratios), yielding molar ratios. For EDAR, I computed ratios from the grams of pollutant per mile, first yielding mass ratios. I then converted EDAR mass ratios to molar ratios for comparability with FEAT molar ratios and then to grams per kilogram fuel using FEAT equations (Bishop and Haugen, 2017).

The concentrations reported in both the FEAT and EDAR data sets included negative values. Negative values result from the measurement systems' conversion of infrared (IR) and ultraviolet (UV) absorption signals to pollutant ratios through regression of the pollutant signal against the CO_2 signal. While physically impossible, negative concentrations are meaningful data. The FEAT system is designed so that repeated measurements of a zero-emission vehicle would randomly yield positive and negative values centered on zero. For computing statistical summaries (e.g., mean concentrations), negative values were retained as reported. For representing the cumulative mass of emissions, negative values were recoded as zero.

Data flags were included in the FEAT data set. All but five species concentrations were flagged as valid. About 9% (896) values for speed were flagged as invalid. These values were listed as zero for both speed and acceleration and were excluded from comparisons involving speed and acceleration. However, the data points were retained for comparisons involving species concentrations.

2.3 Measurement Uncertainties

Measurement uncertainties were estimated for CO, NO, and HC concentrations by reviewing published estimates and by re-applying an approach that was previously utilized by each research team for determining instrument noise.

Past studies report multiple measures of accuracy, detection limits, and precision. A summary is shown in Table 2 (since most of these data have been published, FEAT and EDAR are identified by name). While these measures are generally consistent, distinguishing among them requires consideration of differences in definitions and methods.

Table 2. Measurement accuracy, precision, and detection limits. Units are ppmv or percent of measurement (NA = not available). FEAT measurements are reported as volume percent in which the number of decimal digits varies with species and concentration (2 or 3 decimal places for CO, 3 or 4 decimal places for HC, and 4 or 5 decimal places for NO). The FEAT resolutions correspond to 1, 10, or 100 ppmv. EDAR measurements are reported to the nearest ppmv.

Statistic	CO		HC		NO	
	FEAT	EDAR	FEAT	EDAR	FEAT	EDAR
Reported in previous studies						
Accuracy[1]	±5%	50 or 75	±15%	50 or 125	±5%	10 or 20
Detection limits[2]	NA	50 - 100	NA	100 - 400	25	10 - 30
Uncertainty, σ_0 [3, 4]	1900	28	254	13	59	11
Computed from merged data						
Uncertainty, σ_0 [5]	1223	78	214	38	30	13
$C_0^{0.95}$ [6]	1991	127	348	62	49	21
$3 \times \sigma_{2016}^{LE}$ [7]	1634	640	229	44	37	38

Table notes

1. Ranges not specified (Ashbaugh et al., 1992; Hager, 2017; Hager Environmental & Atmospheric Technologies, https://www.heatremotesensing.com/edar; Ropkins et al., 2017). If two values are shown, both have been reported in the citations.
2. Popp et al. (1999), Ropkins et al. (2017).
3. Standard deviation of double exponential (LaPlace) distribution determined from FEAT negative values (Burgard et al., 2006a).
4. Standard deviation of double exponential (LaPlace) distribution determined from EDAR negative values (Hager, 2018).
5. Standard deviation of double exponential (LaPlace) distribution determined from negative values for merged FEAT and EDAR data.
6. $C_0^{0.95}$, concentration significantly > 0 ($p < 0.05$), determined as the upper 5% value of the LaPlace distribution with parameter σ_0.
7. σ_{2016}^{LE}, 3 times the standard deviation of model year 2016 vehicles excluding any vehicle with a measurement of CO, NO, or HC in the highest 10% of 2016 vehicles. Sample size is N = 184 lower emitting vehicles out of 299 model year 2016 vehicles. All vehicles in the low-emitting subset were gasoline engine vehicles.

Measurement accuracy is established through independent audits using certified standards. Accuracy indicates how close a concentration measurement is to the true value. Accuracy is usually expressed as a percent of concentration within a specified range or as an absolute concentration for concentrations below a threshold value. For example, Hager (2018) reports that EDAR measurements are within the range of the certified gas sample and detector accuracy standards of the California Bureau of Automotive Repair (BAR), which, for CO, are (1) within ± 10% of the certified gas sample, or an absolute value of ± 0.25% CO (whichever is greater) for a gas range less than or equal to 3.00% CO (30,000 ppmv) or (2) within ± 15% of the certified gas sample for a gas range greater than 3.00% CO. HEAT reports accuracies in absolute concentration values, as shown in Table 2 (e.g., 50 or 75 ppmv for CO, equivalent to 0.005 or 0.0075 volume percent, respectively). HEAT also reports that the linear correlation coefficients between measurements and standards exceed $r^2 = 0.99$ for CO, NO, and HC, indicating high accuracy over unspecified concentration ranges (https://www.heatremotesensing.com/edar; Ropkins et al., 2017).

In contrast to accuracy, which indicates how close a concentration measurement is to the true value, precision is a measure of replicability. A measurement may be highly replicable without necessarily being accurate, whereas specified accuracies imply corresponding levels of replicability. Measurement precision is determined from replicate measurements.

Measurement uncertainty (σ_0), as reported in previous studies and recomputed here, is best understood as an indicator of instrument noise (Burgard et al., 2006a). It is also indicative of measurement detection limits. Briefly, negative concentration values are fitted to a double exponential (LaPlace) distribution, which is used to generate an estimate of variance around zero. The probability distribution function of a double exponential (LaPlace) distribution is (Kokoska and Nevison, 1989):

$$f(x) = (1/2\beta) \times e^{(-|x-\alpha|/\beta)}, \beta > 0$$

The double exponential distribution is two exponential distributions extending in opposite directions from $x = \alpha$. Letting $\alpha = 0$ gives a possible distribution of values that are observed when the true concentration is zero; this distribution peaks sharply at zero. The mean and variance of the distribution are α and $2\beta^2$, respectively (Kokoska and Nevision, 1989). The variance can be estimated from negative concentrations, all of which are assumed to be observations whose true value is zero (in contrast, some of the small positive concentrations are not true zero values, so positive values are not used in estimation). The natural logarithm of the function f(x) is a linear function of |x| with slope $1/\beta$ for negative values of x, so the standard deviation ($\beta\sqrt{2}$) of the distribution can be estimated as $\sqrt{2}$ divided by the slope of a regression of the number of values within each small concentration interval versus the midpoint of that interval (Burgard et al., 2006a).

Figures 2 and 3 show the regression plots. The slopes of the regressions were used as described above to generate the values of σ_0 that are listed in Table 2 from the merged FEAT and EDAR data set. Our values of σ_0 deviate somewhat from published values, but modest differences are expected from using different data sets.

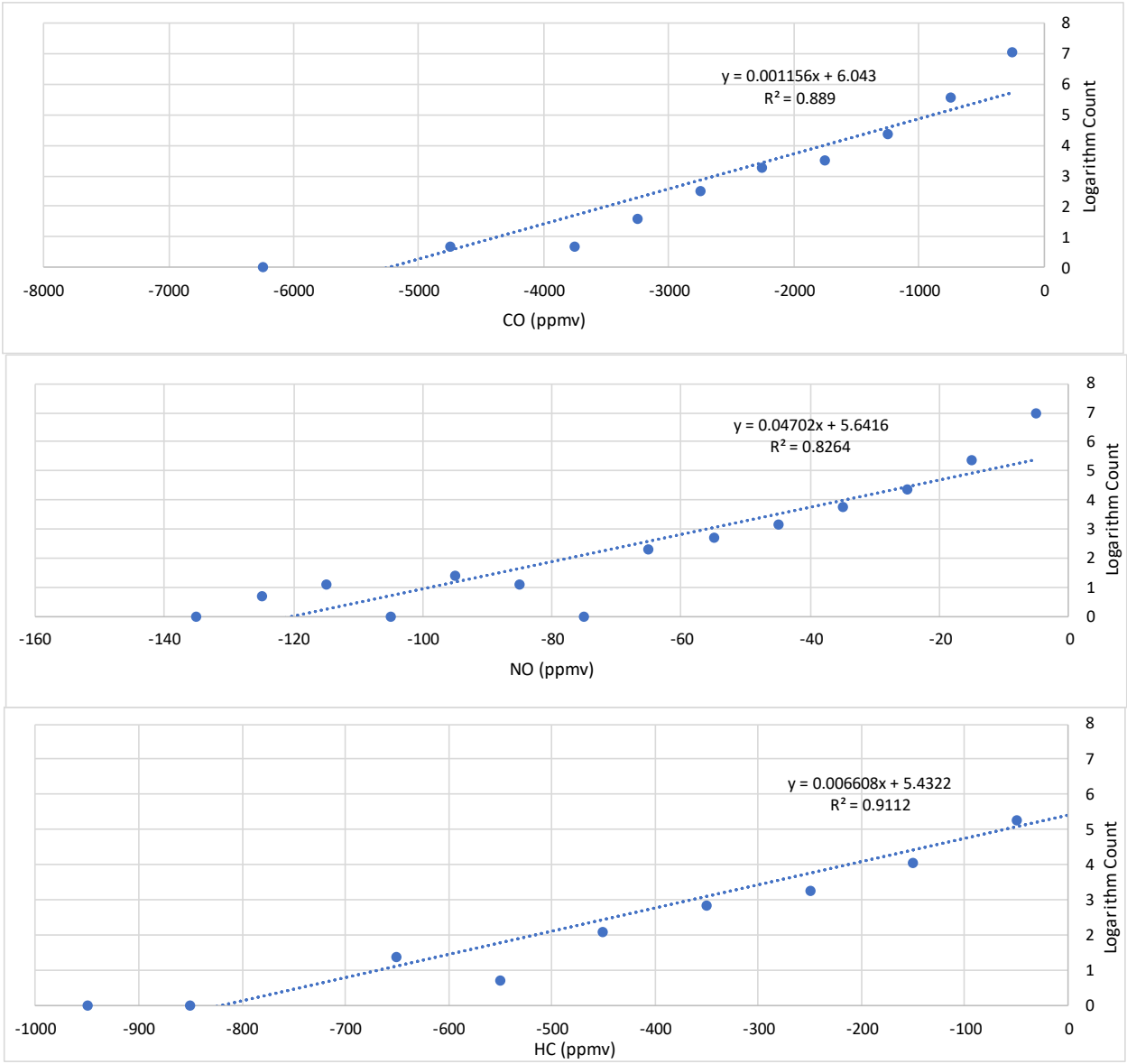


Figure 2. Natural logarithm of binned sample counts versus midpoint of concentrations within each bin range, FEAT data. Bin ranges were defined based on 10 or 15 concentration ranges of equal width depending on the number of bins needed to cover all negative concentrations.

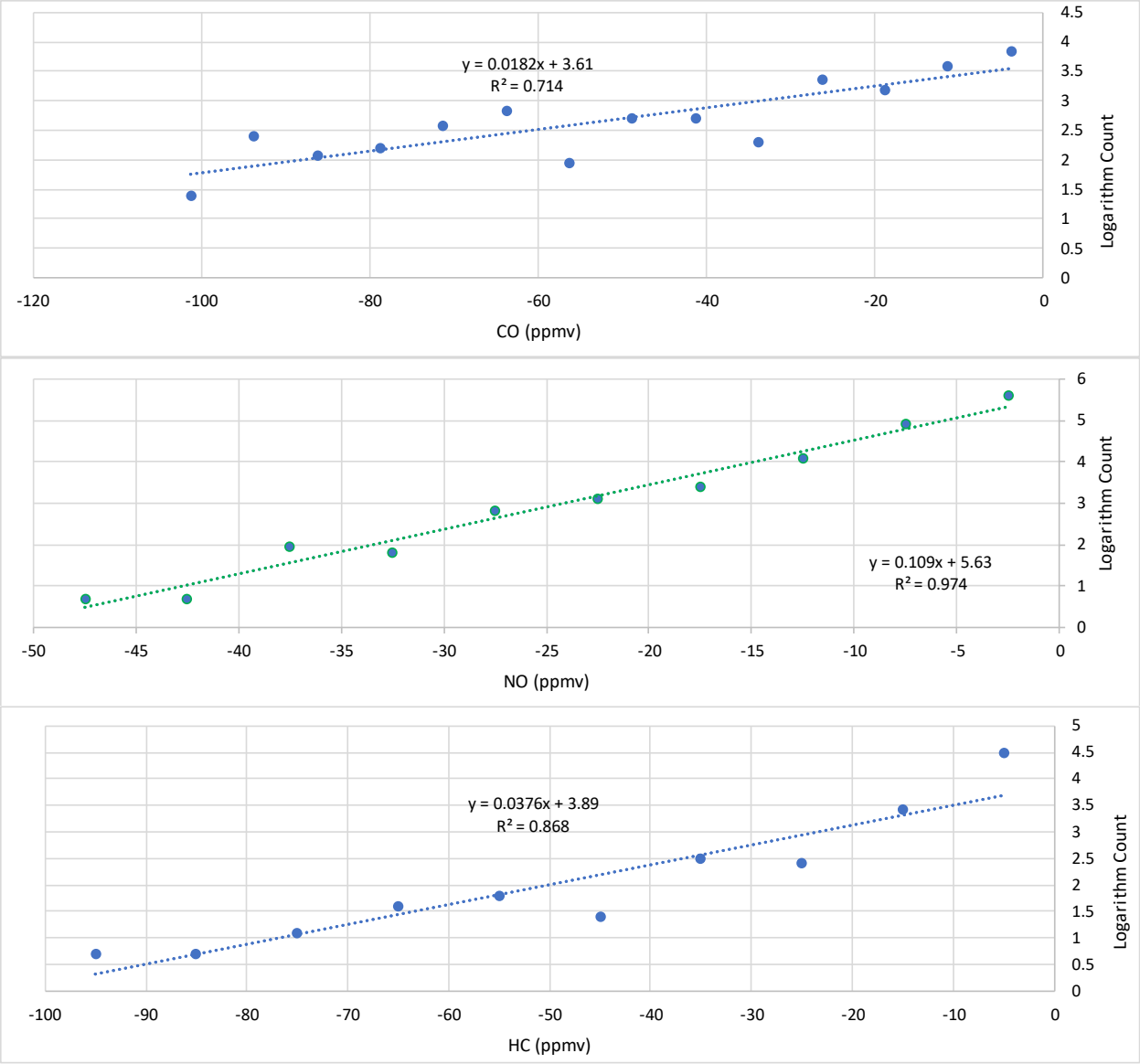


Figure 3. Natural logarithm of binned sample counts versus midpoint of concentrations within each bin range, EDAR data. Bin ranges were defined based on 10 or 15 concentration ranges of equal width depending on the number of bins needed to cover all negative concentrations

I obtained concentration values ($C_0^{0.95}$) that are significantly greater than zero at $p < 0.05$ (95% confidence level) using the probability distribution function of the LaPlace distribution with the values of σ_0 listed in Table 2. The values of $C_0^{0.95}$ tend to be close to detection limits, if the latter have been reported (Table 2).

As an additional comparison, I computed the standard deviations of concentrations for model year 2016 vehicles, excluding those with higher concentrations, because this subset of vehicles is expected to have near-zero or zero emissions. The low-emitting subset consisted only of gasoline-engine vehicles. Table 2 shows values of three standard deviations ($3 \times \sigma_{2016}^{LE}$) of the 2016 vehicle subset (intended to be analogous to three standard deviations of a blank). These values are comparable to the concentration values ($C_0^{0.95}$) that are significantly greater than zero at 95% confidence level based on the LaPlace distribution (full data set). Thus, three assessments of detection limits (those previously reported, those calculated from LaPlace distributions, and those calculated from cleaner 2016 vehicles) yielded comparable values.

Only replicability, not accuracy, can be inferred from the databases provided, since the present study involves no measurement audits with standards. For evaluating differences of individual pairs of measurements, this report uses the computed measurement uncertainties (σ_0) and other metrics to provide benchmark references for computed differences (Section 3).

The standard errors of the regressions of species concentrations versus CO_2 concentrations are used in the FEAT system to determine sample validity (Bishop and Haugen, 2017; Appendix A). These standard errors are not measurement accuracies (as determined by reference to known standards) but can provide another useful uncertainty metric for individual measurements.

2.4 Statistical Methods

Comparisons between measurement systems were made using both graphical approaches and formal statistical tests. Graphical comparisons included scatter plots, box plots, and cumulative distribution plots. The most powerful statistical test of differences is a paired test, provided a logical basis for pairing exists. Since the measurements made by the two measurement systems on the same vehicle at the same time are a logical pairing, a paired t-test was constructed from the exhaust concentrations (e.g., CO) as measured by FEAT and by EDAR on each specific vehicle, then summed (or averaged) over all vehicles within the category. A paired test is more powerful than an unpaired test (an example of an unpaired test is a simple t-test of the difference in the means of two sample sets). Statistical power of the tests is reported.

Box plots and cumulative distribution plots were used to visually compare distributions of measurements. All box plots show the 10th, 25th, 50th, 75th, and 90th percentiles. Some box plots also show individual values below the 10th percentile and above the 90th percentile, but these individual high and low points are not shown when displaying the full range would compress the distribution of the 10th through 90th percentiles into a very narrow band.

Comparisons were determined for the full data set and for data subsets, as described in Section 3.

3. Results and Discussion

3.1 Speed and Acceleration

The ranges of vehicle accelerations were comparable between measurement systems, although graphical time series comparisons suggest that vehicle speeds recorded by the two systems correlated better than did accelerations (Figure 4). Over the full comparison set (N = 4418 vehicles with valid speed and acceleration data), the mean A – B speed difference was -0.95 miles per hour (mph) (statistically significant, $p < 0.0001$) and the 95% confidence interval (CI) for the mean difference ranged from -1.0 to -0.86 mph. The average speed difference (1 mph) was only 4% of the average vehicle speed (23.9 and 24.9 mph, A and B, respectively), which is a small relative difference. Since the A and B measurements were spatially separated by ~100 feet, the small average measured speed differences may reflect true physical differences. Further consideration (below) is given to whether larger speed differences are random or systematic.

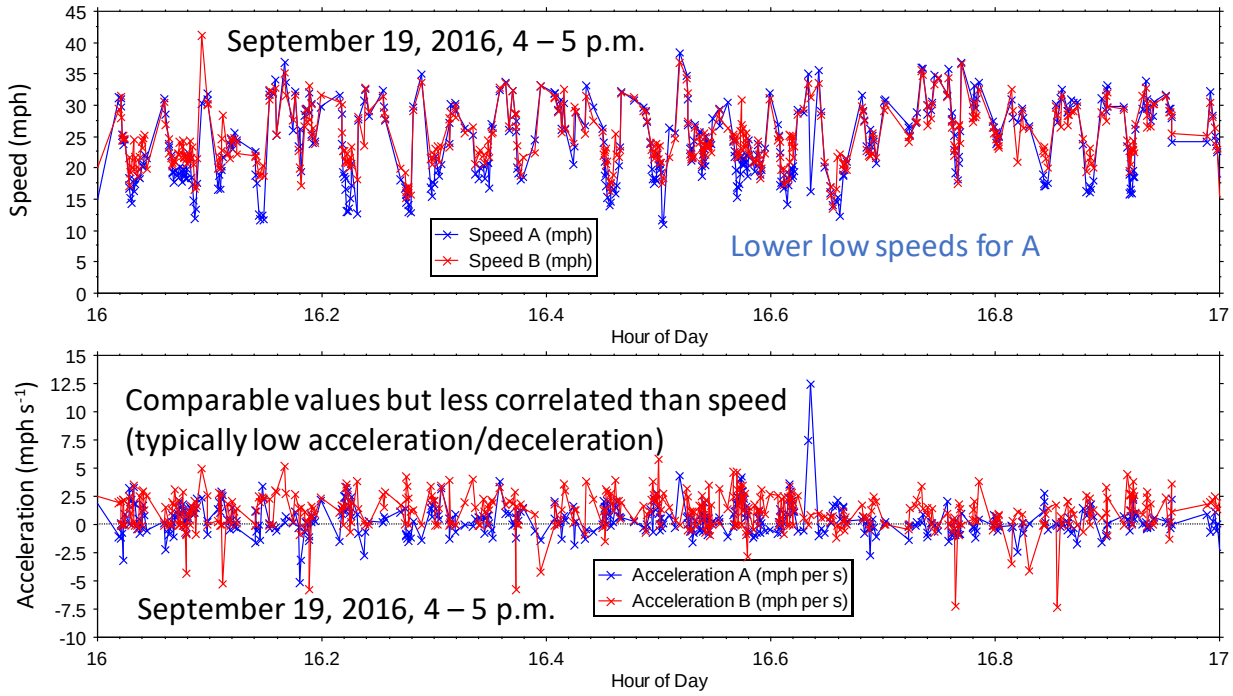


Figure 4. Example time series comparisons of paired vehicle speeds and accelerations.

The sample size (N = 4418) is sufficiently large that a paired t-test has very high statistical power, capable of yielding a statistically significant result at $p < 0.05$ (95% confidence level) with a probability of 91.4% when the mean difference between paired measurements is only 5% of the standard deviation of the paired differences (https://www.anzmtg.org/stats/PowerCalculator/PowerTtest). Since the standard deviation of the paired speed differences was 2.8 mph, the probability of obtaining a statistically significant result at $p < 0.05$ exceeded 90% for differences as small as 0.14 mph (0.05 ×2.8 mph). Therefore, relatively small speed differences were detectable.

For vehicle acceleration, the mean paired A – B acceleration difference was -0.58 mph s^{-1} (statistically significant, $p < 0.0001$) and the 95% CI for the mean acceleration difference ranged from -0.64 to -0.53 mph s^{-1}. Since the standard deviation of the paired acceleration differences was 2.0 mph s^{-1}, the probability of obtaining a statistically significant result at $p < 0.05$ exceeded 90% for differences as small as 0.1 mph s^{-1} (0.05 × 2.0 mph s^{-1}). The mean acceleration difference (~ 0.6 mph s^{-1}) is modest and the average accelerations of 0.065 and 0.65 mph s^{-1} were not large (neither hard acceleration nor deceleration). As for vehicle speed, relatively small acceleration differences were detectable due to the large sample size.

Ninety percent (90%) of the speed differences were less than 5 mph and 88% of the acceleration differences were less than 3 mph s^{-1} (95% were less than 4 mph s^{-1}) (Figure 5). A small number of paired comparisons (< 10%) exhibited larger measurement differences. As noted for speed, some larger differences could reflect real differences due to the offset in the two measurement locations. Other possible explanations for the occurrence of larger measurement differences were considered. To investigate if larger speed or acceleration differences were systematically related to the type of vehicle, the maximum paired speed and acceleration differences were determined for each of 40 body styles recorded in the data set (e.g., 2-door, 4-door, van, etc.). As shown in Figure 6, the maximum paired difference for each body style increased approximately linearly in relation to the logarithm of the vehicle count. This result is consistent with a random effect – the more vehicles in the category, the more likely the occurrence of a larger difference. One outlier is evident in each panel of Figure 6, and each appears to reflect a unique situation (possibly, a measurement artifact). No systematic variation that could be related to specific body styles was otherwise evident.

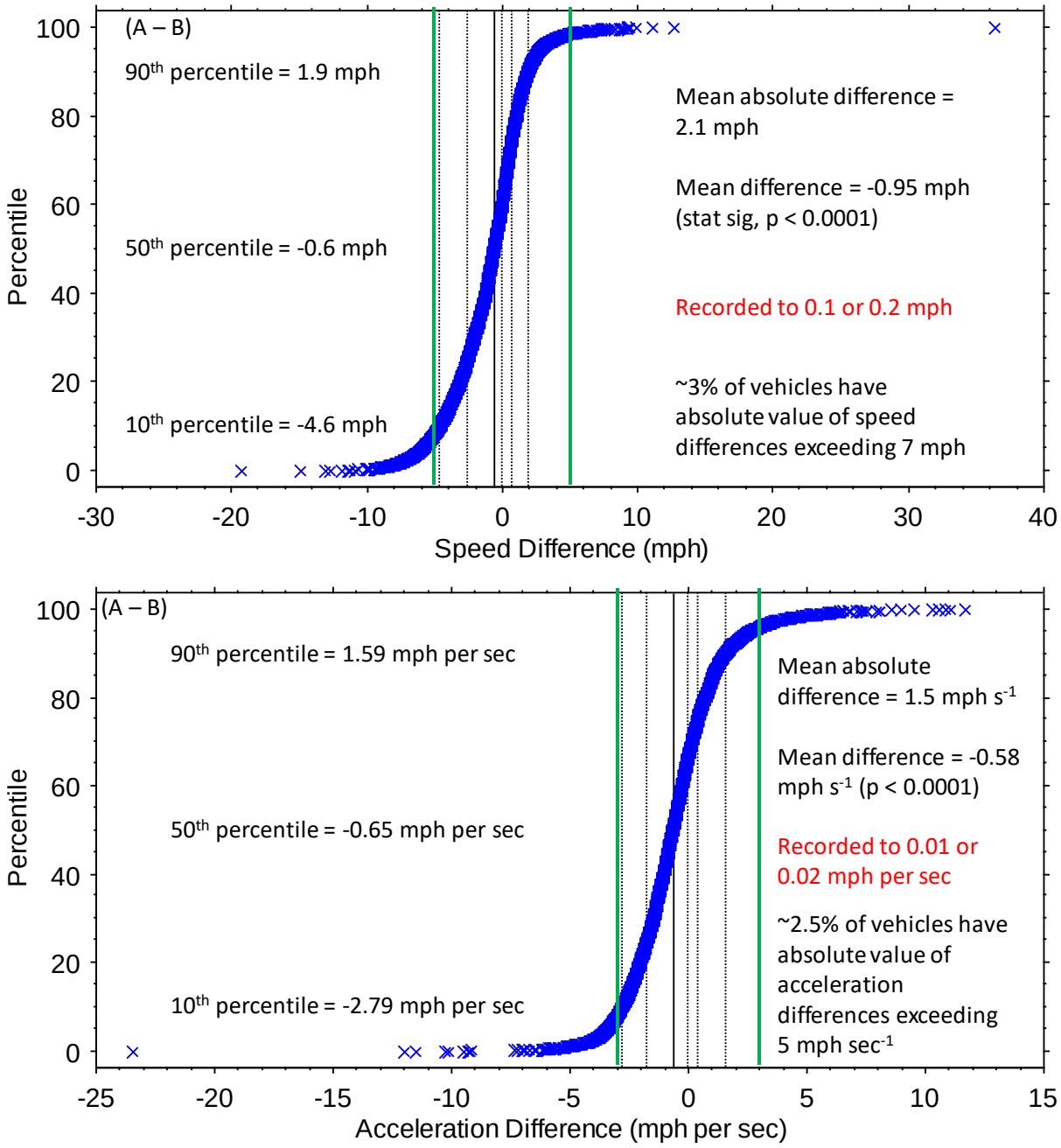


Figure 5. Cumulative frequency distributions of paired speed and acceleration differences (N = 4418).

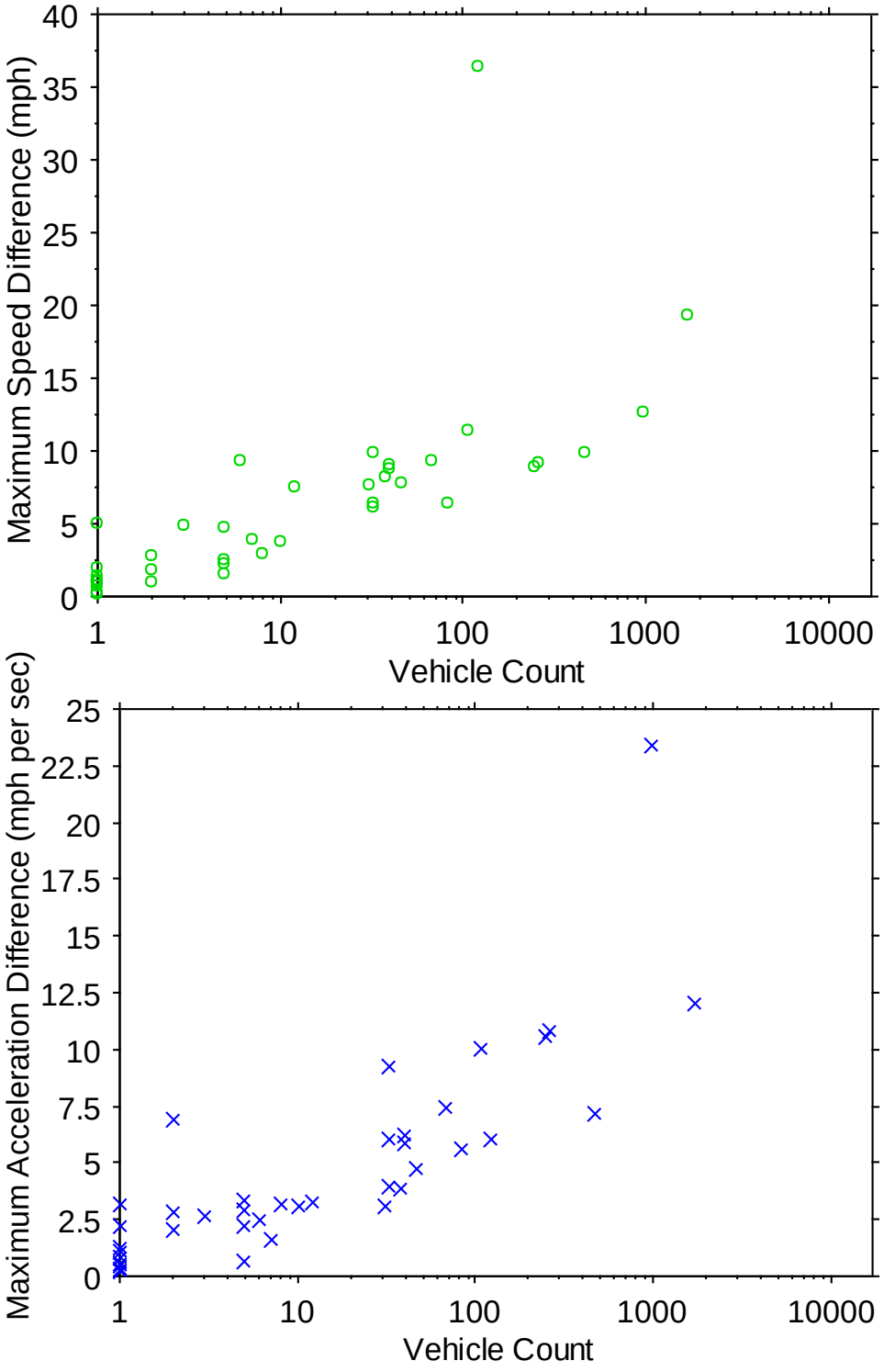


Figure 6. Maximum paired speed and acceleration differences within each category of vehicle body type versus count of vehicles within each category. Each symbol shows results for one body style (40 body styles, e.g., 2-door [N = 41], 2-door hatchback [N = 34], 4-door [N = 1776], van [N = 282], wagon [N = 76], etc.).

3.2 Vehicle Specific Power

Vehicle specific power (VSP) is an empirical estimate of engine load that captures much of the dependence of vehicle emissions on driving conditions and can be calculated from roadside measurements of speed and acceleration (Jimenez, 1999). Default parameterization provides an operational equation (Jimenez, 1999). Hager (2018) uses the original version of the equation to determine VSP for EDAR, in which input units are metric:

$$\text{VSP} = 9.81 \bullet \sin(\alpha) \bullet v + 1.1 \bullet v \bullet a + 0.132 \bullet v + 0.000302 \bullet v \bullet (v + v_w)^2,$$

where v_w = headwind, α = road angle, v = velocity (speed), a = acceleration

Bishop and Haugen (2017) use a different version of the VSP equation for FEAT, in which the input units are mph for speed and mph s^{-1} for acceleration:

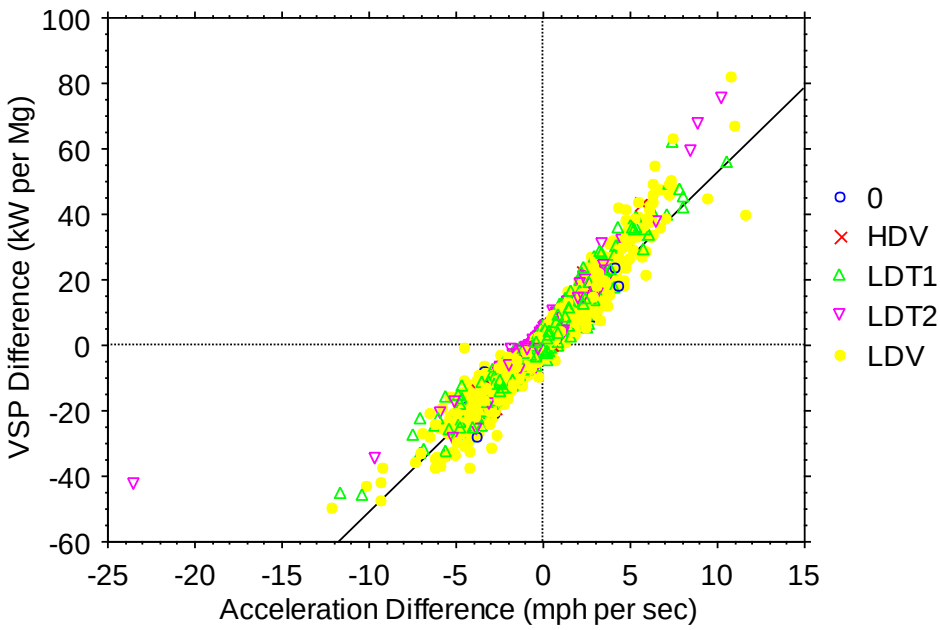
$$\text{VSP} = 4.39 \bullet \sin(slope) \bullet v + 0.22 \bullet v \bullet a + 0.0954 \bullet v + 0.0000272 \bullet v^3,\ slope = \text{road angle}$$

For both equations, the units of VSP are kilowatts per megagram (1 megagram = 1 metric ton or tonne), kW Mg^{-1} or kW $tonne^{-1}$. Road angle (grade) is a 1% upslope (Bishop and Haugen, 2017).

Recalculation of VSP from FEAT measurements of speed and acceleration using the second equation exactly reproduced the FEAT VSP values. Since measurements of v_w (headwind) were not reported, the first equation could not be applied to corroborate the EDAR VSP values. However, application of the second equation using EDAR measurements of speed and acceleration (with appropriate conversion of units) yielded exact agreement with EDAR VSP except for day-specific offsets of 2.69 kW Mg^{-1} on September 19, -0.18 kW Mg^{-1} on September 20, and 2.79 kW Mg^{-1} on September 21 (presumably due to day-specific values of v_w in the internal EDAR calculations).

Larger differences between FEAT and EDAR values of VSP were due to differences in acceleration (Section 3.1), as indicated in Figure 7. These differences occurred across vehicle types. Recalculation of VSP values using the same equation (i.e., the second equation above) changed the regression shown in Figure 7 by only a marginal amount (intercept = -0.42, slope = 5.212, $r^2 = 0.926$). The reported VSP differences are statistically significant ($p < 0.0001$). The mean pairwise difference was -3.47 kW Mg^{-1}, 95% CI = -3.79 to -3.15 kW Mg^{-1}. The mean VSP difference -3.47 kW Mg^{-1} is related to the mean acceleration difference (-0.58 mph s^{-1}). Ninety-four percent (94%) of VSP differences were less than 21 kW Mg^{-1} (Figure 8).

To determine the practical implications of the VSP differences, EDAR and FEAT measurements of speed and VSP were used to assign vehicles to EPA motor vehicle emission simulator (MOVES) model operating modes (U.S. EPA, 2015, Table 1-5). Because most of the modes are defined by VSP ranges having widths of either 3 or 6 kW Mg^{-1}, the observed VSP differences are large enough to lead to different MOVES mode classifications for many vehicles. However, the statistical distributions of vehicles to MOVES modes were similar (Figure 9). Disaggregation of vehicle data into MOVES modes is accomplished later by using a consistent assignment method.



VSP Difference (kW per Mg) = .893 + 5.194 * Acceleration Difference (mph per sec); R^2 = .912

Figure 7. Difference in VSP versus difference in acceleration as reported by the two measurement systems. Vehicle classes are 0 = no information (N = 76), HDV = heavy duty vehicles (N = 58), LDT1 = light duty trucks 1 (N = 1084), LDT2 = light duty trucks 2 (N = 311), and LDV = light duty vehicles (N = 2889).

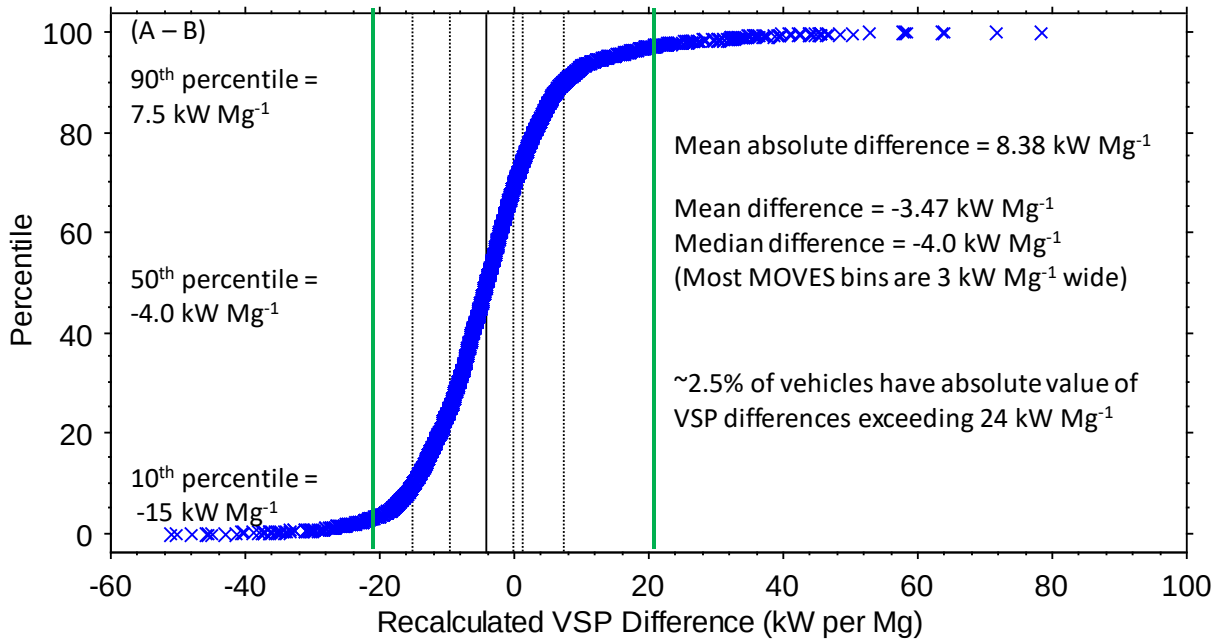


Figure 8. Cumulative frequency distribution of paired VSP differences.

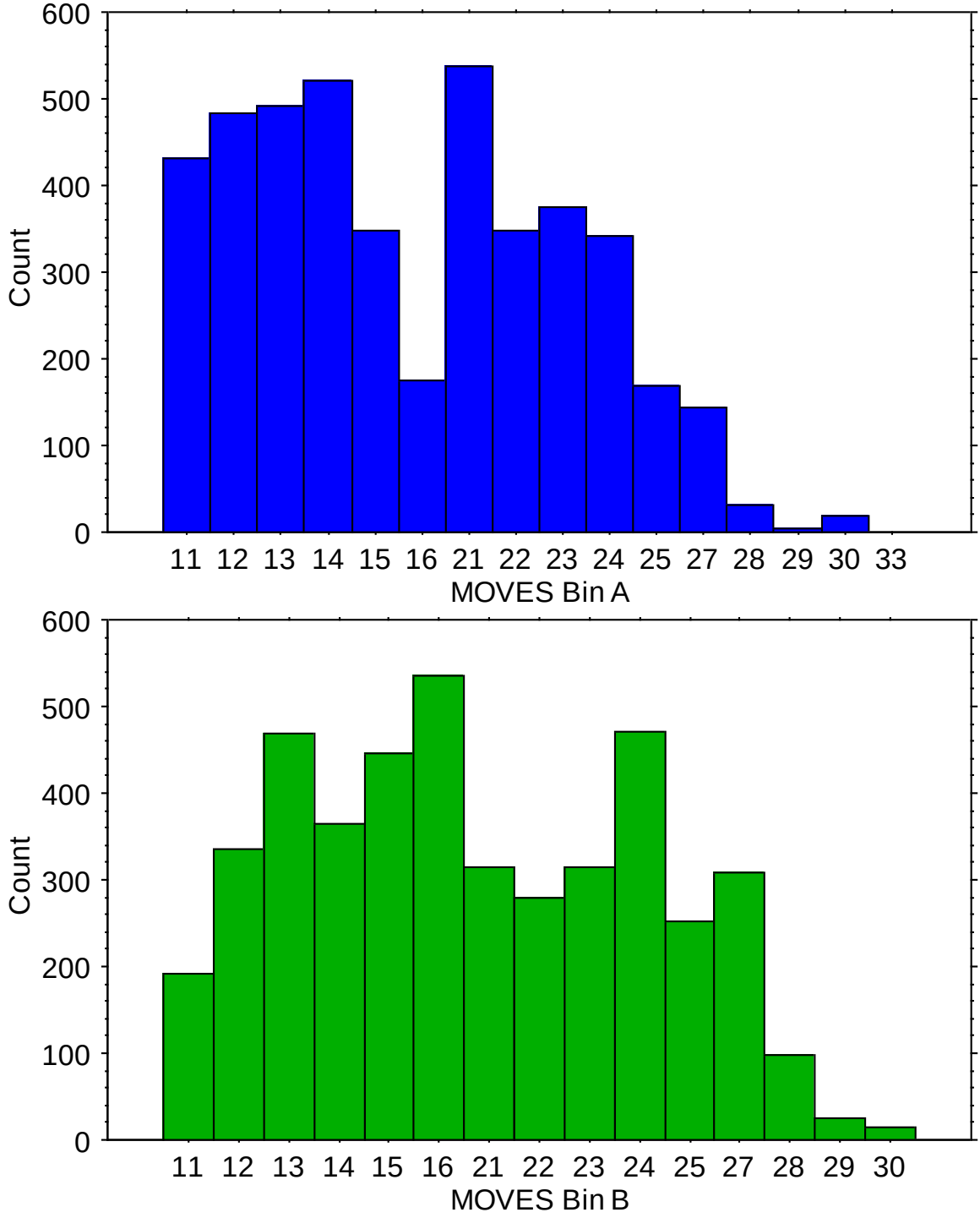


Figure 9. Statistical distributions of vehicle assignments to MOVES operating modes, binned by using speed and VSP data from measurements systems A (top) and B (bottom). Twelve of the 23 MOVES modes are represented by more than 100 vehicles. Modes 11 and 21 are deceleration modes (VSP < 0). MOVES bins 11 – 16: 1 ≤ speed < 25 mph. MOVES bins 21 – 30: 25 ≤ speed < 50 mph. MOVES bins 33 – 40: speed ≥ 50 mph. No idle modes are represented.

3.3 Repeated Measurements and Inherent Variability

Repeated measurements provide an opportunity to characterize vehicle-specific variations, as previously reported in Bishop and Haugen (2017), Haugen and Bishop (2018), and Hager (2018). Plume concentrations vary as a function of multiple factors, especially with variations in exhaust gas temperature. In principle, the DIAL method used in HEAT is not affected by temperature (Hager, 2018). Repeated measurements are discussed here to provide a benchmark of same-vehicle variability. Quantifying same-vehicle variability aids in evaluating the differences between the measurements made by the two sampling instruments, because the instruments did not measure exhaust plumes at identical times. Repeated measurements provide insight into plume variability on time scales ranging from minutes to hours to 24 hours. The shortest time difference between the first and second occasion when a vehicle passed the measurement location was four (4) minutes. Otherwise, typical time differences between repeated measurements were either a few hours or ~24 hours. The numbers of vehicles with repeated measurements and the number of times vehicles passed the measurement location are listed in Table 3. Data set A consisted of 9628 total measurements made on 7831 unique vehicles; 1516 vehicles were sampled more than once. For data set B, there were 9807 total records; 1641 of 7804 distinct vehicles were sampled more than once. The merged data set yielded 4045 unique vehicles with repeated measurements of 586 vehicles.

Table 3. Number of repeat vehicle measurements. The number of passes shows the count of vehicles that passed the measurement location once and once only, exactly twice, exactly three times, exactly four times, and exactly five times.

Number of Passes	Data set A only	Data set B only	Merged A + B
1	6315	6163	3459
2	1252	1304	499
3	249	316	79
4	13	17	6
5	2	4	2
Total unique vehicles[1]	7831	7804	4045
Number of repeat vehicles[2]	1516	1641	586
Total measurement count[3]	9628	9807	4728

Table notes

1. Sum of column
2. Sum of column excluding first row
3. Number of passes times row entry, summed

Repeated measurements provide an opportunity to characterize vehicle-specific variations. As shown in Figure 10, pre-2004 model years tended to exhibit the largest differences in repeat-vehicle species concentrations. This result was less pronounced for hydrocarbons than for CO and NO, and larger repeat-vehicle differences occurred even on post-2004 model-year vehicles. The average same-vehicle mean differences were zero. The concentration ranges that covered 90% of the repeat-vehicle differences were ± 0.25% for CO A, ± 0.19% for CO B, ± 93 ppmv for NO A, ± 170 ppmv for NO B, ± 255 ppmv for HC A, and ± 59 ppmv for HC B. Comparison to the measurement uncertainties in Table 3 indicates that the HC repeat variability did not exceed the corresponding $C_0^{0.95}$, NO repeat variability did exceed the corresponding $C_0^{0.95}$, and CO repeat variability either marginally or definitely exceeded the corresponding $C_0^{0.95}$ values.

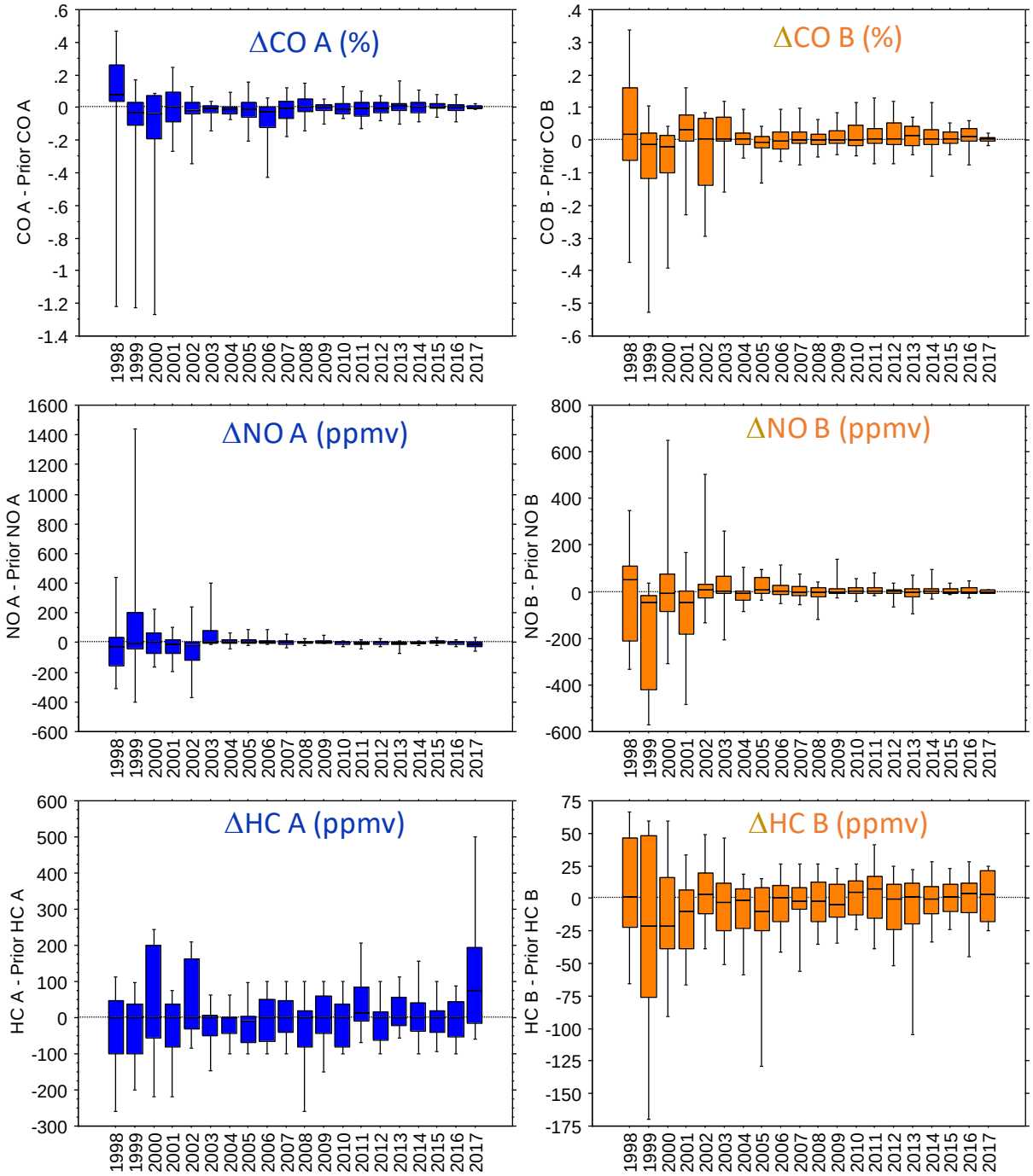


Figure 10. Statistical distributions of same-vehicle measurement differences by model year. Model year 1998 includes all model years from 1987 (earliest in the data set) through 1998. Model year 2017 includes only 5 vehicles.

Since same-vehicle repeated measurements occurred under different driving or ambient conditions (e.g., 24 hours apart), repeated measurement differences exceeding measurement variability were expected for some fraction of the vehicles measured. There was no evidence that larger repeat differences were associated with longer time differences in the measurements. Nor were larger repeat concentration differences associated with larger differences in repeated speed or acceleration values (Figure 11). Therefore, the occurrence of larger differences on a fraction of the repeat vehicles was apparently random. This conclusion is further supported by Figure 12, which shows that larger repeat measurement differences occurred about equally across all MOVES operating bins.

The highest ranges of repeat measurement concentrations occurred for vehicles with intermittently high emissions, i.e., vehicles having one or more low concentrations and one or more high concentrations (Figure 13). Some of the highest-emitting vehicles were consistently high emitters, and the repeat differences were not as large as some of the intermittent high emitters. Vehicles with the smallest differences in repeated measurements were consistently low emitters. As noted, higher repeat measurement differences tended to occur in pre-2004 vehicles and to otherwise occur in an apparently random manner. This result is consistent with the known high variability of vehicle exhaust plumes in vehicles with improperly functioning emission control systems. The time scales of the intermittencies could not be determined with the present data. If the time scales of variability are on the order of seconds for some vehicles, the comparisons of paired measurements that are separated in time by ~2 seconds (as previously described) would be sensitive to emissions variations. The impact of plume variability on the comparison of the two measurement systems is considered further in the next section.

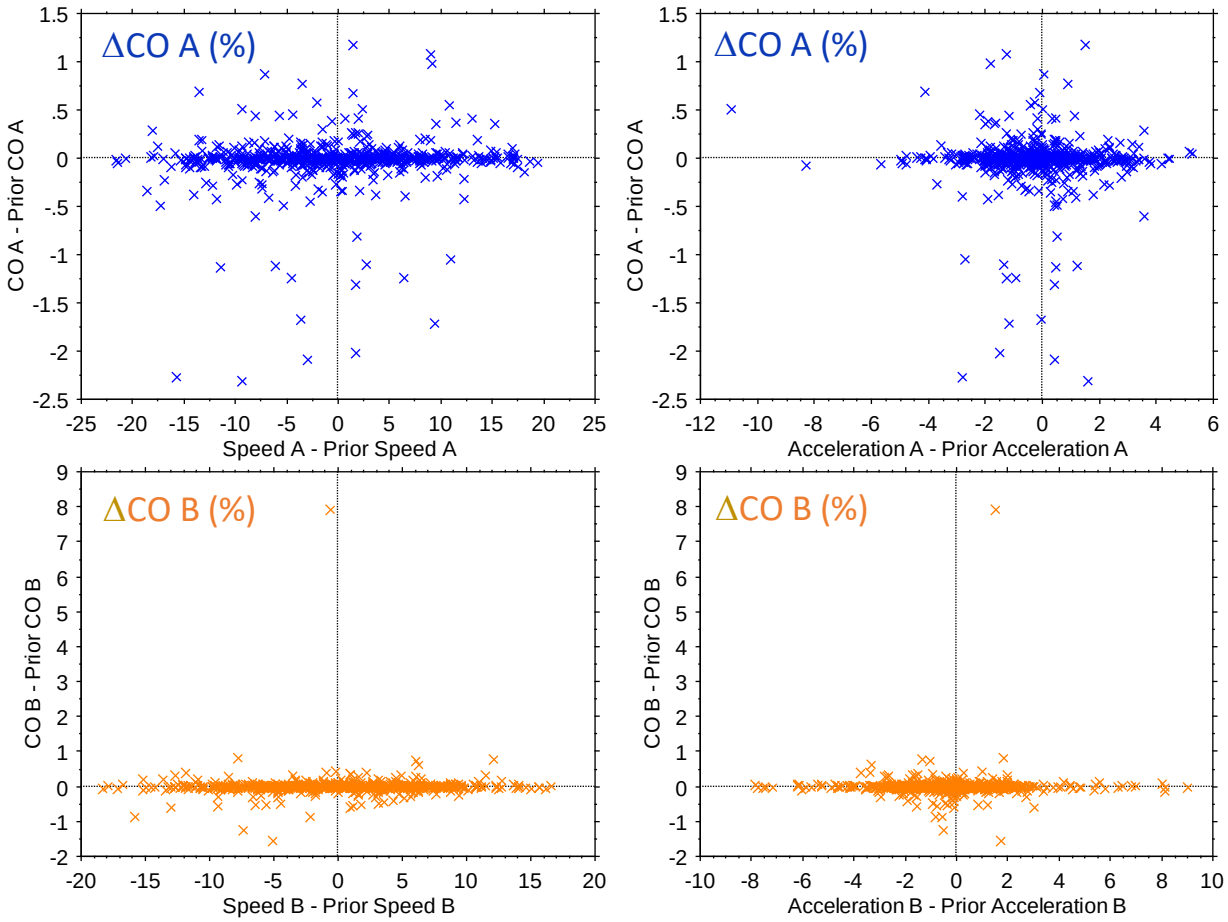


Figure 11. Repeat vehicle CO differences versus speed and acceleration differences.

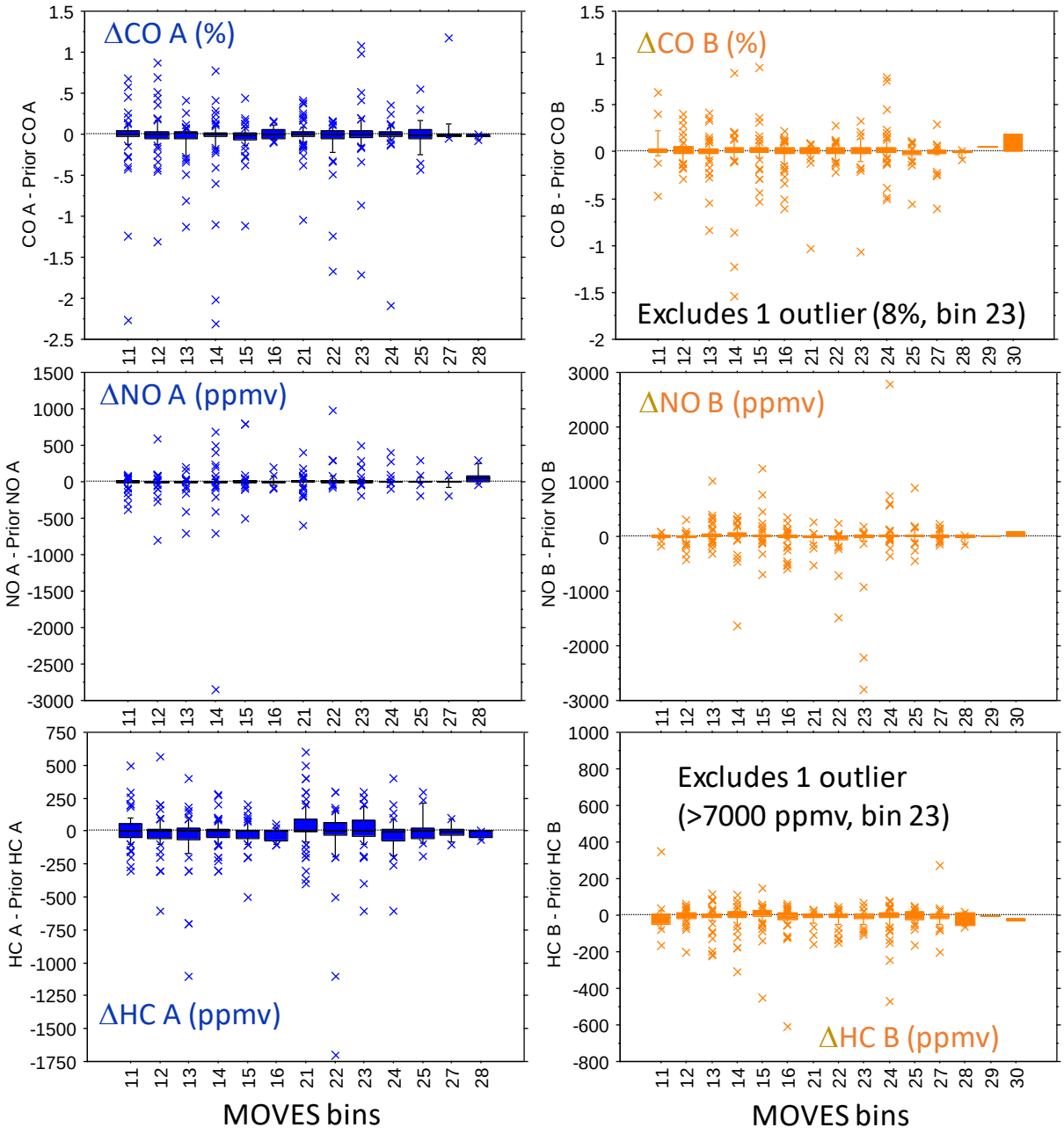


Figure 12. Statistical distributions of repeat vehicle measurements by MOVES operating modes, showing values below the 10th and above the 90th percentiles.

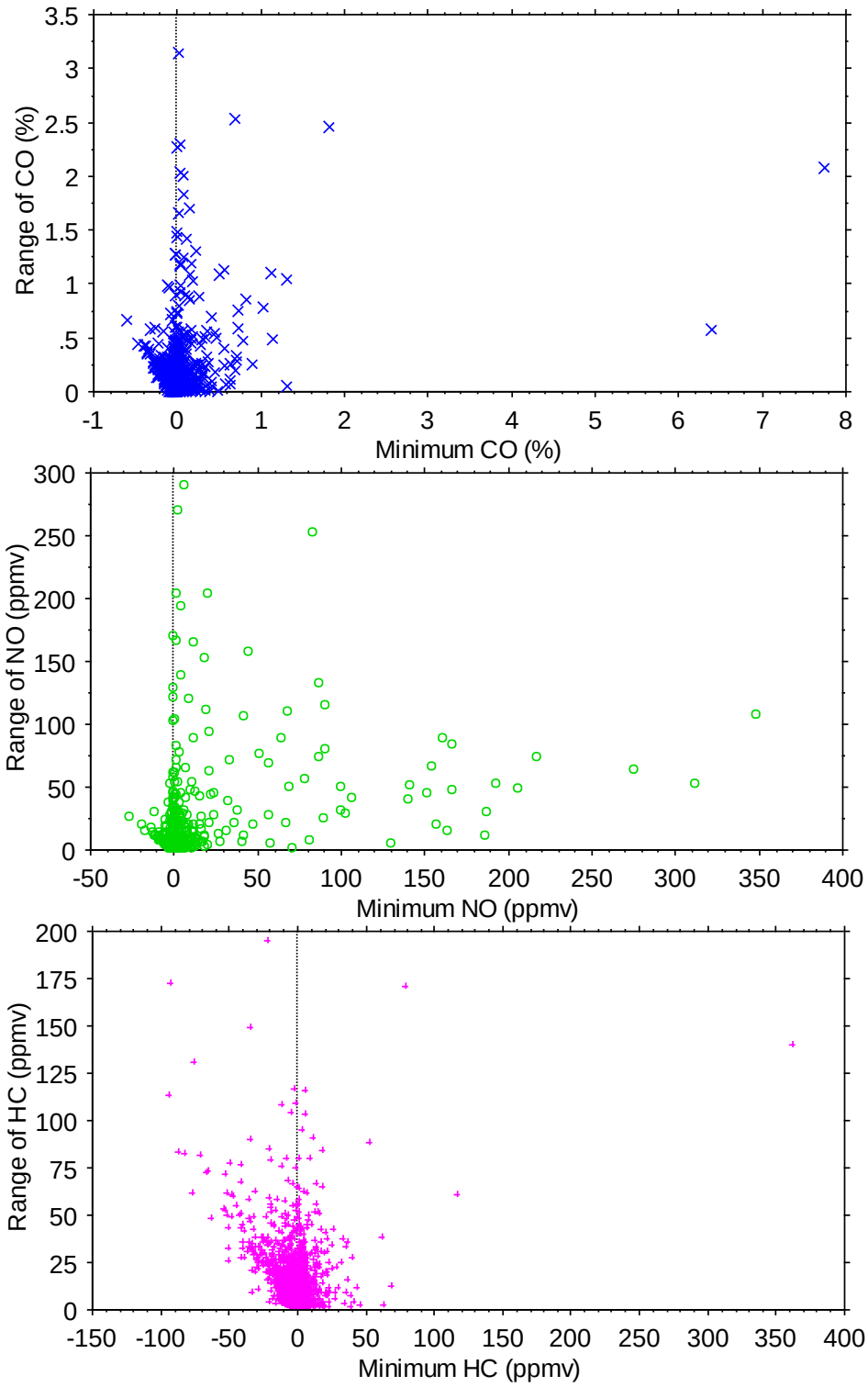


Figure 13. Ranges of concentrations measured on repeat vehicles versus minimum measured concentration. Each point is one vehicle. Measurements are from one of the two data sets. The highest ranges of concentrations occur for vehicles with intermittently high emissions, i.e, repeated measurements showed one or more low concentrations with one or more high concentrations.

3.4 Paired Measurement Comparisons for All Data

The statistical distributions of reported concentrations and molar concentration ratios are shown in Figure 14. For CO and HC, median concentrations are similar for the two measurement systems, as are the molar ratios to CO_2, whereas the systems yielded somewhat different median values of NO. The median concentrations obtained by one measurement system were 200 ppmv for CO (0.02%), 16 ppmv for HC, and 5 ppmv for NO. For the second measurement system, the medians were 202 ppmv for CO (0.02%), 24 ppmv for HC, and 12 ppmv for NO. The 90th percentile concentrations obtained by one measurement system were 1930 ppmv for CO (0.19%), 214 ppmv for HC, and 118 ppmv for NO. For the second measurement system, the 90th percentiles were 1688 ppmv for CO (0.17%), 53 ppmv for HC, and 151 ppmv for NO. Distributions of fuel-based emission factors were also more similar for CO and HC than for NO (Figure 15).

The mean paired concentration differences between measurement systems (A – B) were -0.002% CO (not significant), -6.9 ppmv NO (significant, $p < 0.05$), and 10.0 ppmv HC (significant, $p < 0.05$) (Figure 16). Ranges of ± 0.2% for CO and ± 200 ppmv for NO and HC encompass ~90% of the paired comparisons (Figure 16). For individual measurements, the mean differences would be considered within the measurement limitations of the systems. When averaged over 4728 measurements, the mean NO and HC differences were statistically significant. When expressed as fuel-based emission rates, the mean paired differences between measurement systems were -0.05 g kg^{-1} CO (not significant), -0.87 g kg^{-1} NO (significant, $p < 0.05$), and 0.40 g kg^{-1} HC (significant $p < 0.05$).

As previously discussed for the pairwise comparisons of speed and acceleration, the sample size (N = 4728) is sufficiently large that a paired t-test has very high statistical power, capable of yielding a statistically significant result at $p < 0.05$ (95% confidence level) with a probability of 93.0% when the mean difference between paired measurements is only 5% of the standard deviation of the paired differences (https://www.anzmtg.org/stats/PowerCalculator/PowerTtest). The standard deviations of the paired concentration differences were 0.240% for CO, 225 ppmv for NO, and 208 ppmv for HC. Therefore, the probabilities of obtaining statistically significant results at $p < 0.05$ exceeded 90% for differences as small as 5% of these standard deviations, or 0.012% CO, 11 ppmv NO, and 10 ppmv HC. Thus, small differences in the mean concentrations were both detectable and detected.

For pairwise comparisons of the individual measurements, the $C_0^{0.95}$ values listed in Table 2 and the inherent variabilities revealed by repeated measurements are both useful references. As shown in Figure 17, the distributions of systems A – B concentration differences are comparable in range to the distributions of differences in repeated measurements. This result indicates that measurements from systems A and B do not differ more than two sets of measurements from system A or two sets from system B. The variability of the differences in paired A – B measurements might be attributable to inherent variability in exhaust plume concentrations. As shown in Figure 18, the ranges of pairwise measurement differences increase with vehicle age.

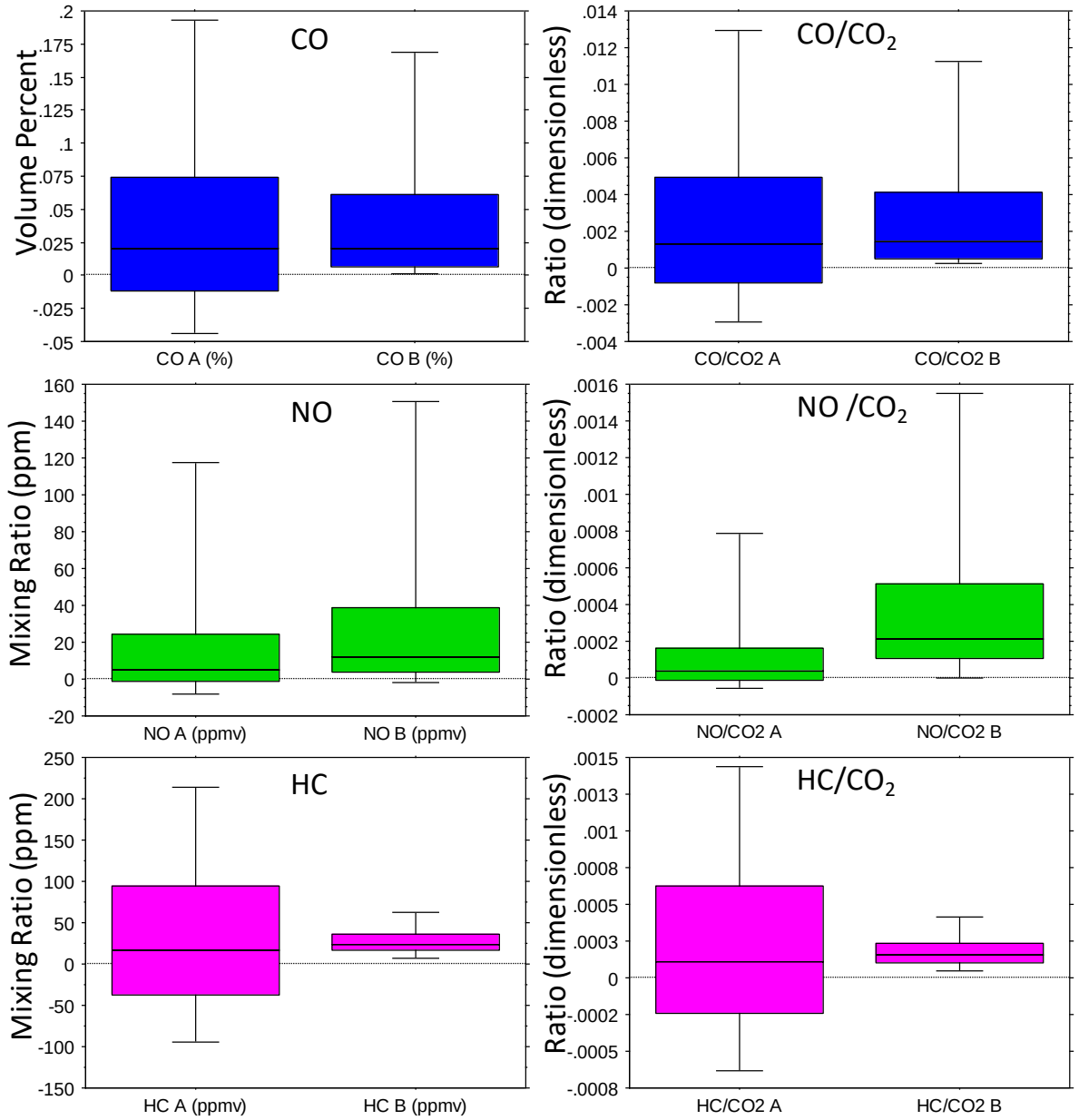


Figure 14. Statistical distributions of concentrations and concentration ratios.

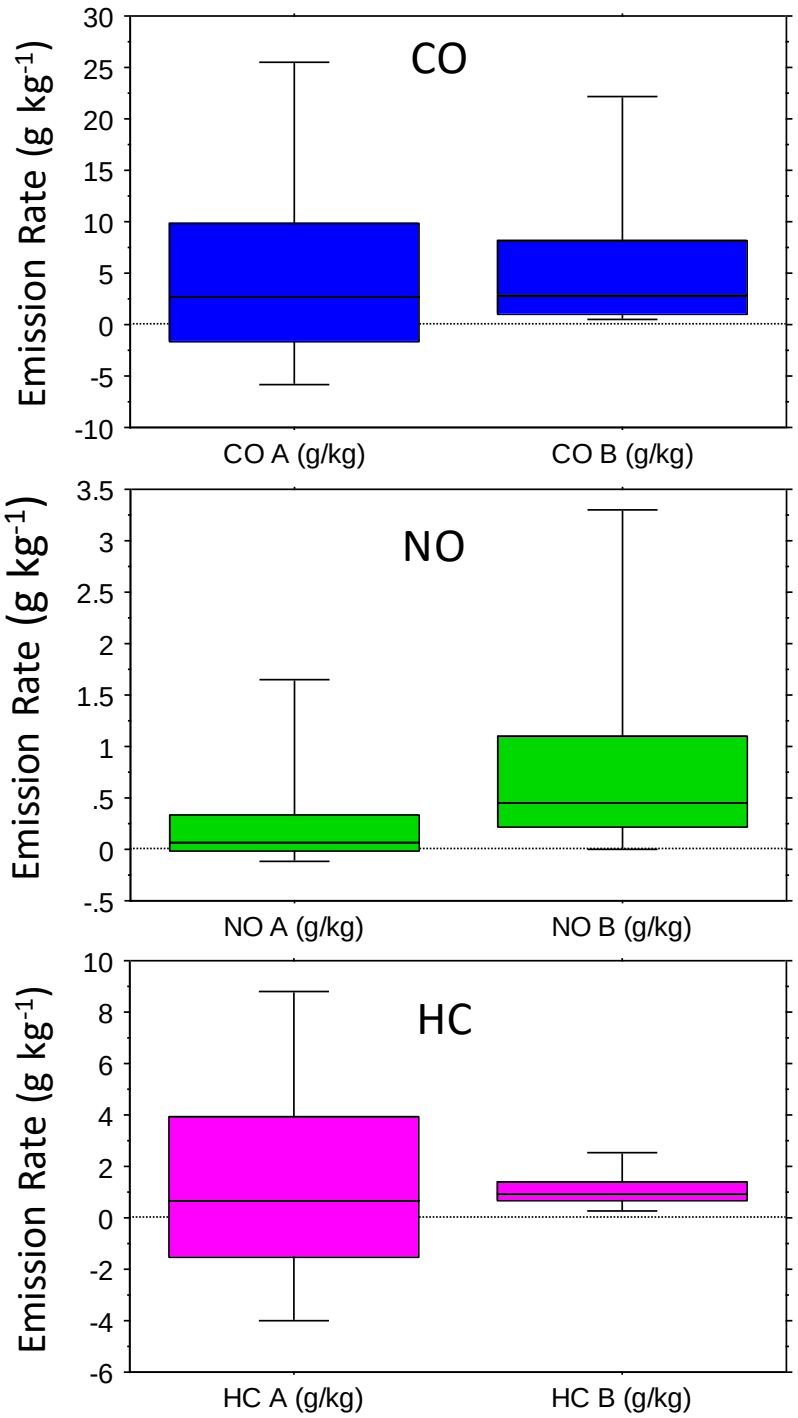


Figure 15. Statistical distributions of fuel-based emission rates. Units are g kg^{-1} fuel.

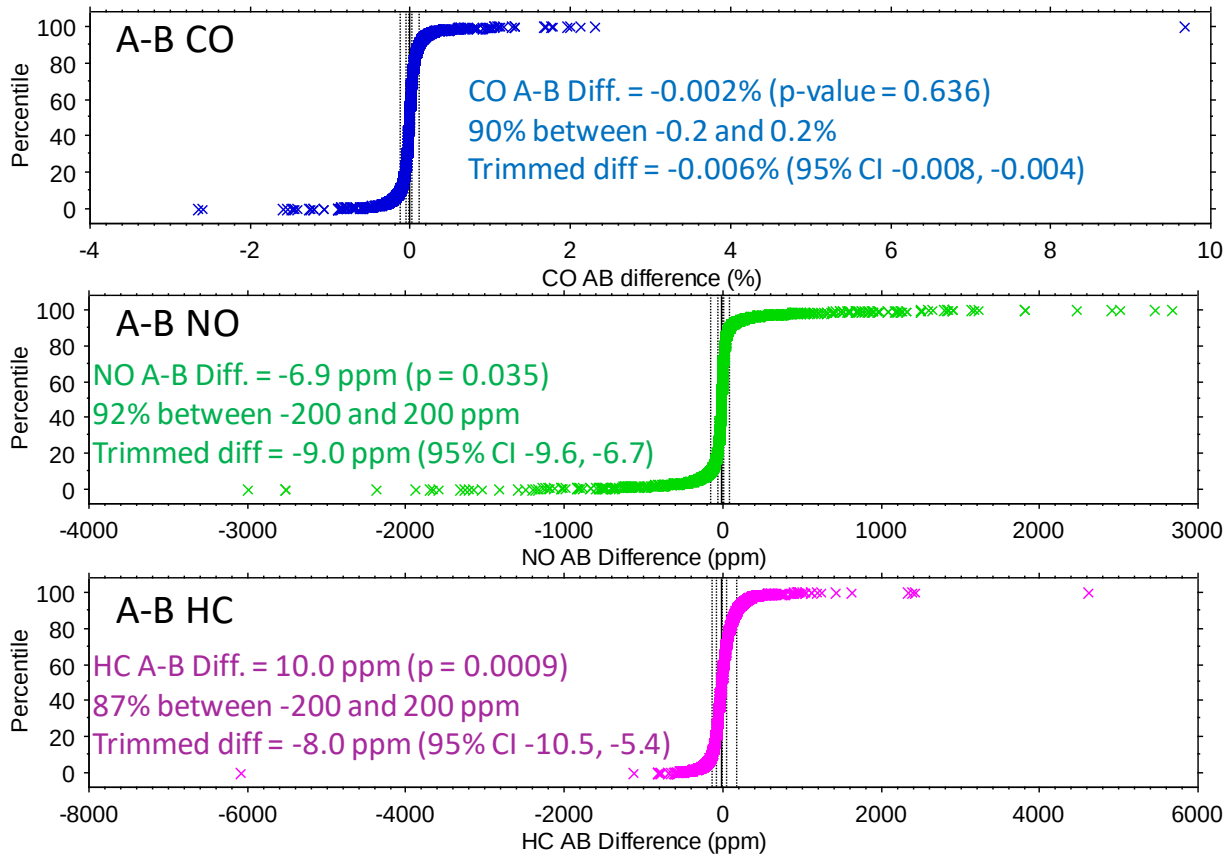


Figure 16. Cumulative distributions of paired concentration differences. The effects of outliers on the mean differences were examined by computing trimmed differences, which were determined by excluding all paired differences falling outside the ranges ± 0.2% for CO and ± 200 ppmv for NO and HC. Values within these ranges encompassed ~ 90% of the data.

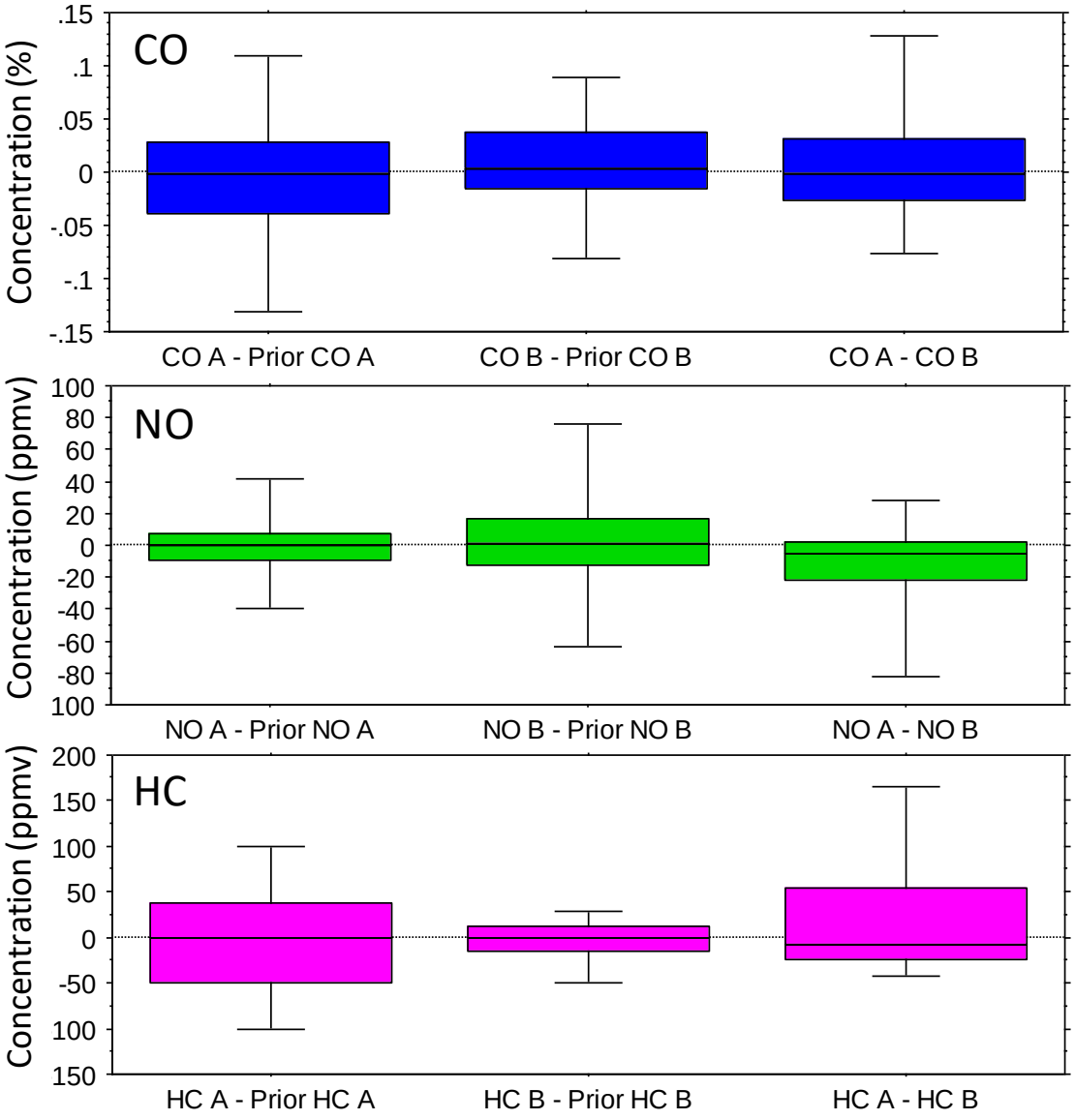


Figure 17. Comparative distributions of system and repeated measurement differences.

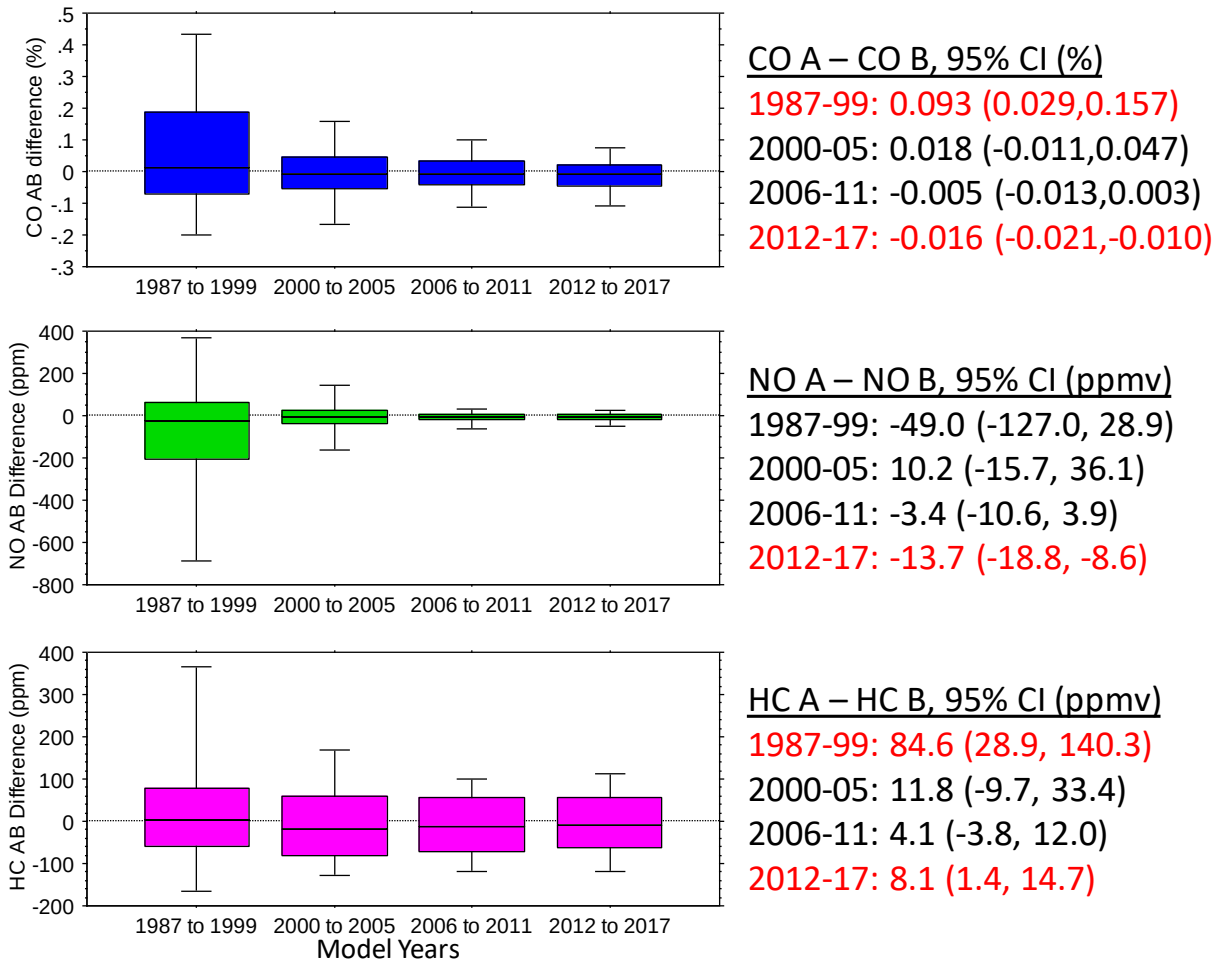


Figure 18. Distributions of paired concentration differences by model year. Text lists group mean A – B differences and 95% CIs (red text indicates statistical significance at $p < 0.05$).

3.5 Paired Measurement Comparisons for Data Subsets

As described in the previous section, the ranges of pairwise measurement differences increased with vehicle age. The goal of this section is to determine if larger measurement differences were associated with specific vehicle types or models of operation.

As shown in Figure 19, the maximum paired concentration differences for each vehicle body style increased approximately linearly in relation to the logarithm of the vehicle count. This result is consistent with a random effect – the more vehicles in the category, the more likely the occurrence of a larger difference. One outlier is evident in the CO and HC panels of Figure 19, which reflects a situation when one system recorded high values but the other didn't. The concentration differences likely represent plume differences rather than instrument differences. This vehicle was decelerating slightly when passing one sampling point and strongly when passing the other sampler. This vehicle was sampled again the next day by both systems, and both systems recorded high concentrations of CO and HC on the repeat measurement. On the second day, the vehicle was accelerating past both samplers, though at different rates. The vehicle in question was a 2003 LDT with 164,000 miles recorded on the odometer. The overall comparisons are insensitive to the inclusion or exclusion of this vehicle.

No systematic differences in the A – B concentration differences were observed across the period of the study, as shown in Figure 20. Differences in weather or time of day therefore did not result in different ranges of the concentration differences.

Scatter plots showed no relationship between measurement system concentration differences and differences in either speed or acceleration. This lack of dependence of concentration differences on operational differences is supported by Figure 21, which shows that the ranges of concentration differences did not vary systematically across MOVES operating modes. A possible exception is that the ranges of HC measurement differences are larger for deceleration modes (11 and 21) than for other modes at the same speeds.

Figure 22 shows that systematic measurement differences were not observed when the data were disaggregated by vehicle class, except that the A – B NO differences were skewed positive for HDVs. The sample size (N = 84) was not large and the average difference for this class was affected by high A – B NO differences recorded on two HDV cabs (model years 2001 and 2004).

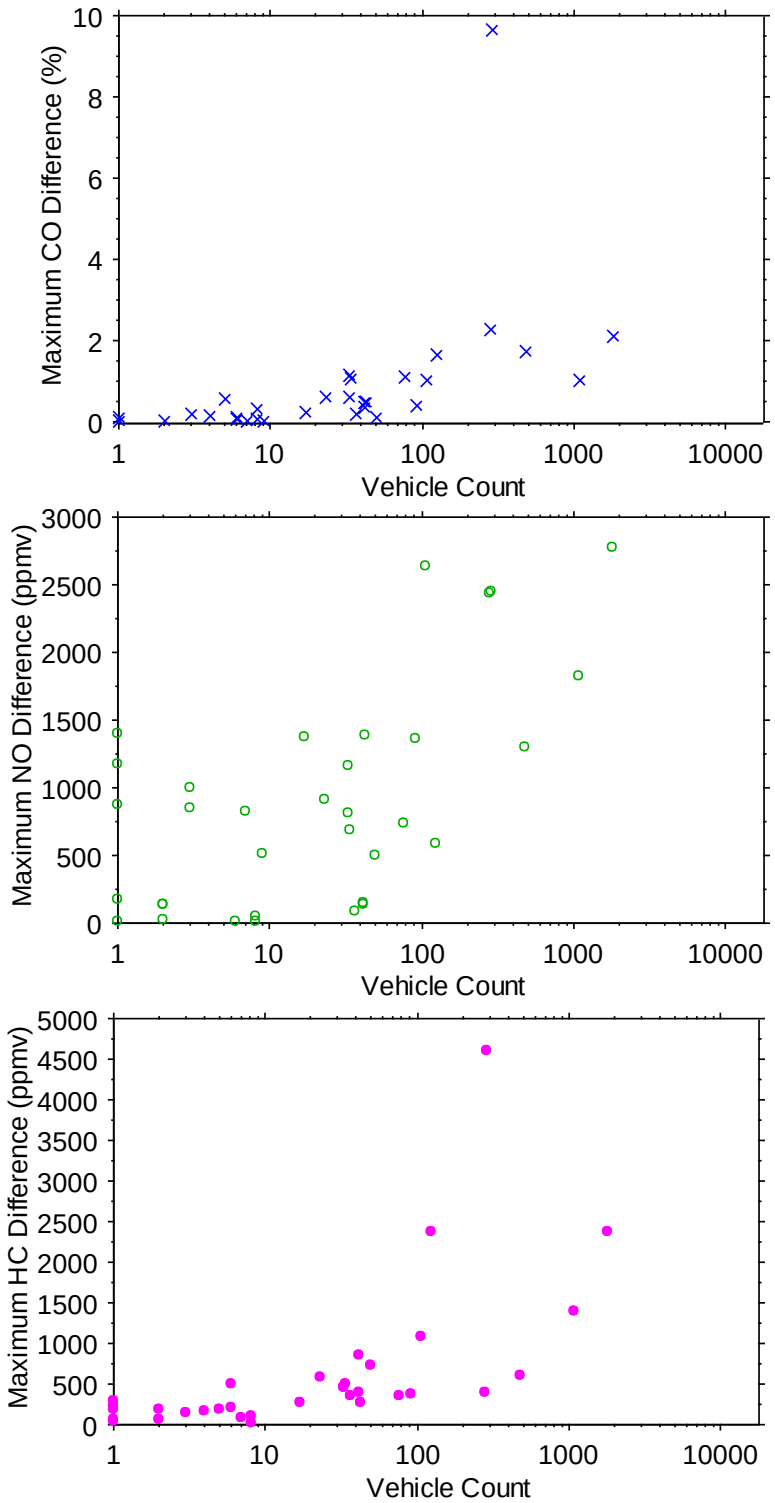


Figure 19. Maximum paired concentration differences within each category of vehicle body type versus count of vehicles within each category. Each symbol shows results for one body style (40 body styles, e.g., 2-door [N = 41], 2-door hatchback [N = 34], 4-door [N = 1776], van [N = 282], wagon [N = 76], etc.).

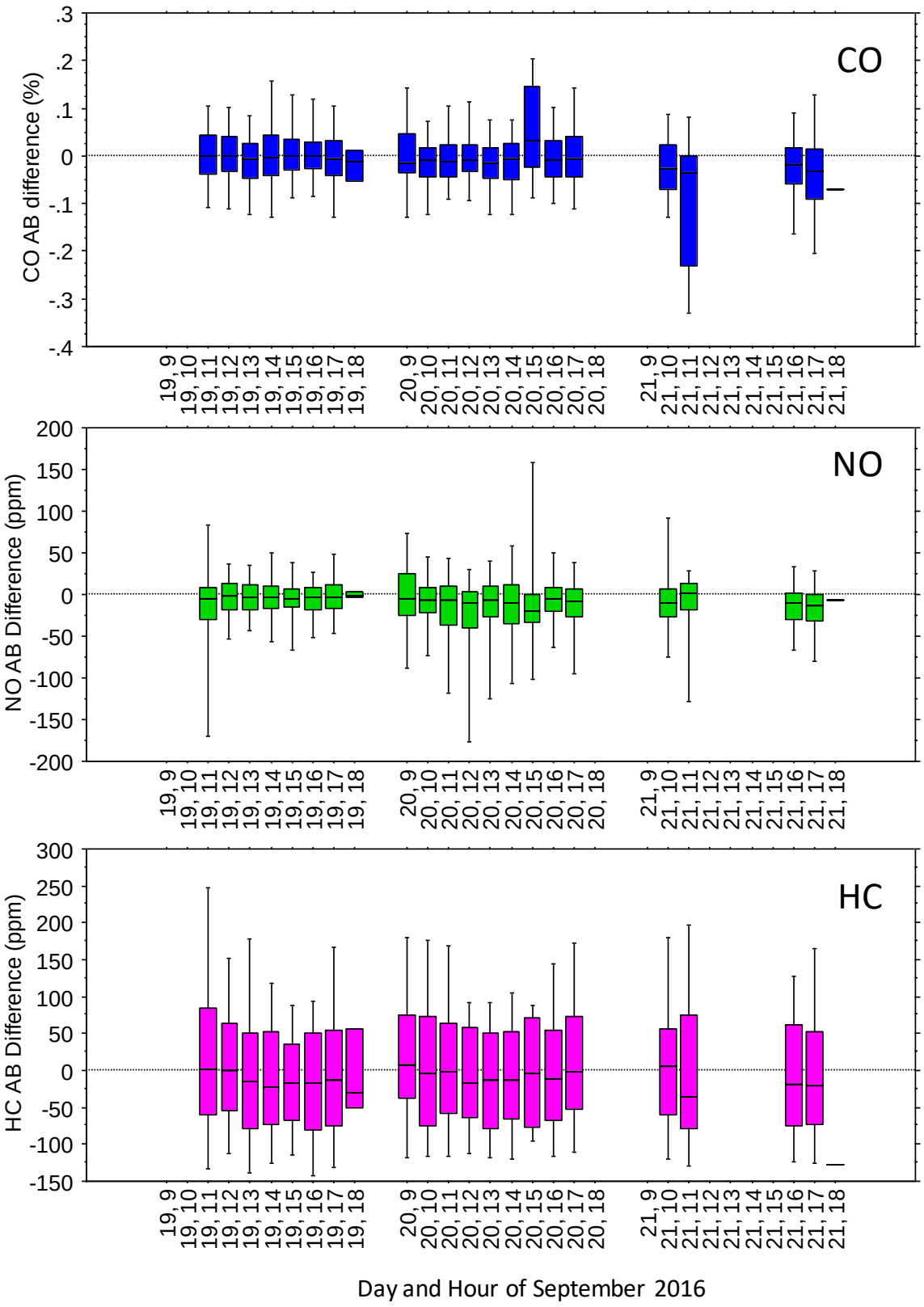


Figure 20. Distributions of concentration differences by day and hour.

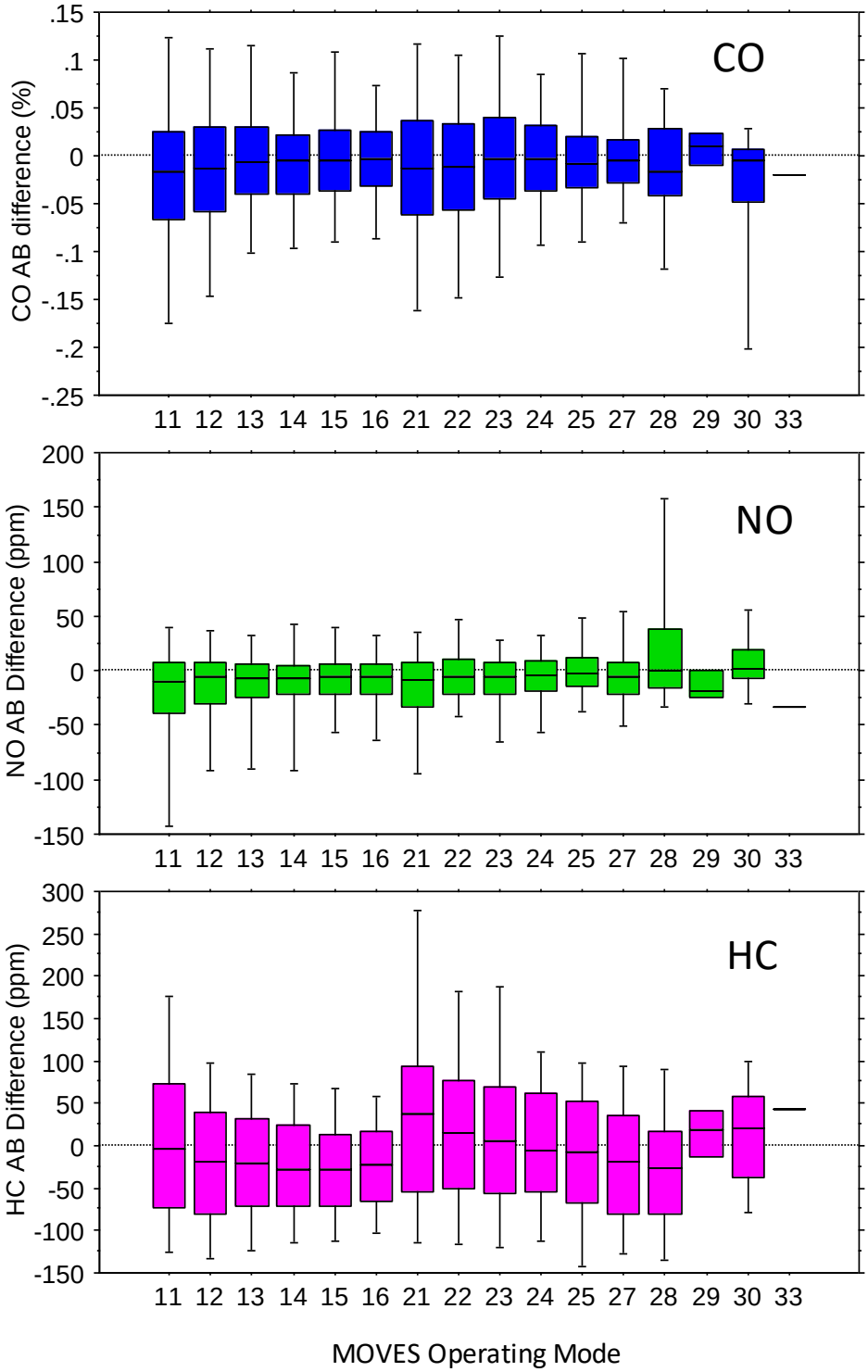


Figure 21. Distributions of concentration differences by MOVES operating mode.

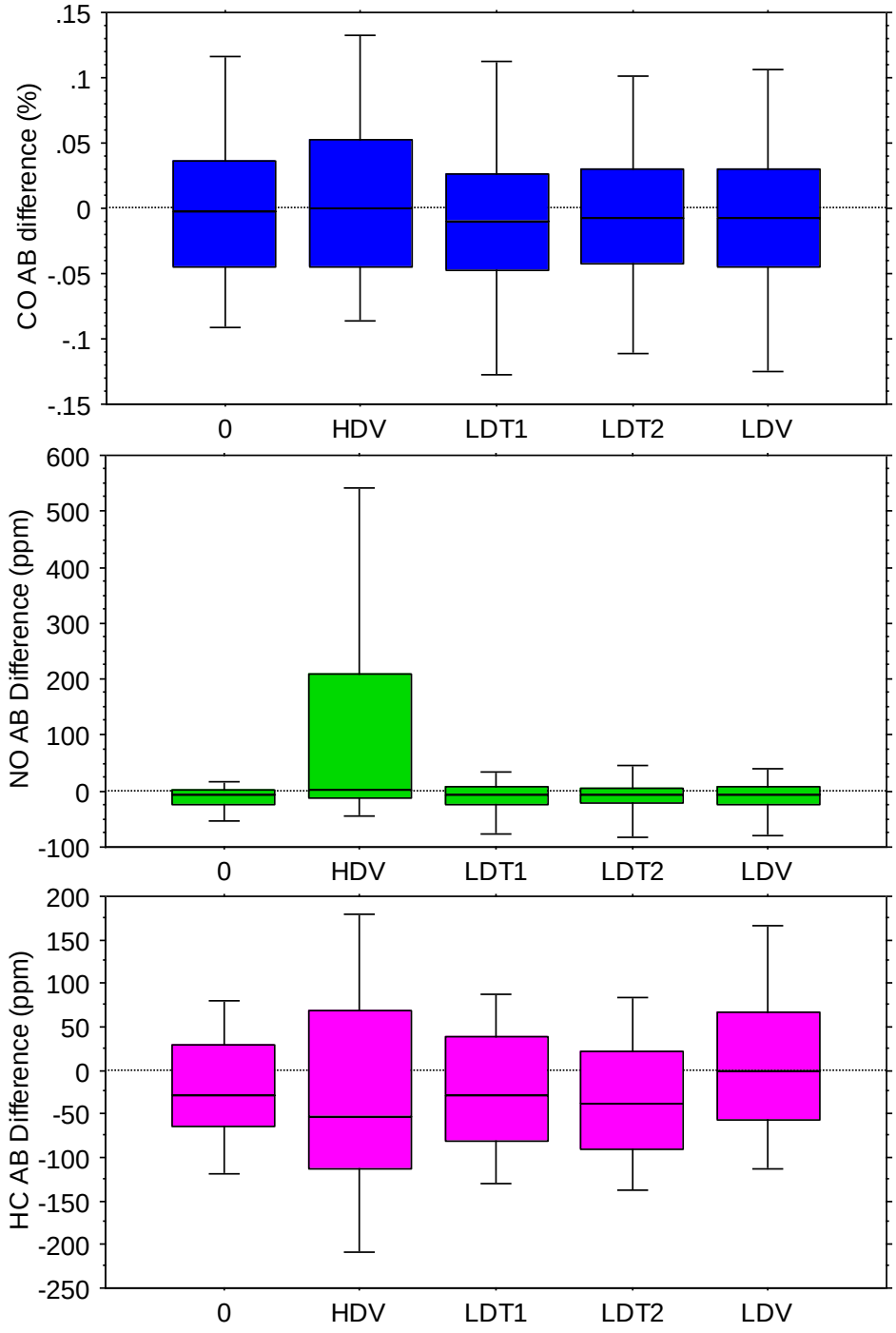


Figure 22. Distributions of measurement differences by vehicle class. Vehicle classes are 0 = no information (N = 86), HDV = heavy duty vehicles (N = 84), LDT1 = light duty trucks 1 (N = 1194), LDT2 = light duty trucks 2 (N = 356), and LDV = light duty vehicles (N = 3008).

3.6 High Emission Vehicles

Higher-emitting vehicles are of interest because they account for a disproportionate share of the total mass of emissions. This section examines high-emitter mass fractions.

The cumulative frequency distributions of CO, NO, and HC concentrations are comparable for the two measurement systems, but not identical (Figure 23). For NO, the differences are consistent with the distributions shown as box plots in Figures 14 through 17 and with the mean paired NO measurement difference of 6.9 ppmv (Figure 16). Given measurement uncertainties (σ_0 = 30 ppmv for one instrument and 13 ppmv for the other, Table 2), individual paired NO measurements differing by only 7 ppmv are not significantly different; when averaged over 4728 measurements, the 7 ppmv mean difference is statistically significant. The measurement systems show similar concentrations associated with the upper quintile (highest 20%).

Mass fractions were determined from the reported mass emissions, which were provided in units of g per kg fuel by FEAT and g per mile by EDAR. I also determined mass fractions for EDAR after computation of g per kg using the FEAT equations; the results were nearly identical to mass fractions determined from the reported g per mile. The EDAR measurements of mass emissions (g per mile) are the more fundamental measurement quantity and all positive. Negative FEAT values of emission mass (g per kg fuel) occurred, corresponding to negative concentrations. FEAT negative emission rates were reset to zero to permit computation of physically realistic cumulative fractions of emissions. Figure 24 shows cumulative fractions of vehicles versus cumulative fractions of emissions mass for the full data set. Differences at low levels are largely due to differences in instrument detection limits. The two systems yield similar results for the fraction of emissions mass attributable to the highest emitters, as shown in Figure 24 and in Table 4. The percentage of emissions from high-emitters is obtained by subtracting tabled values from unity and multiplying by 100%. The five highest-emission vehicles (0.1% of the sample, cumulative frequency 0.999) emitted 4.4% of the CO mass according to system A or 3.1% according to system B (Table 4). The highest five percent of the vehicles emitted 49.2% of the CO mass according to system A or 44.9% according to system B. For NO, the five highest-emission vehicles emitted 5.9% of the mass (A) or 6.8% (B). Five percent of the vehicles emitted 71.4% of the NO mass (A) or 59.3% (B). For HC, five vehicles emitted 2.9% of the mass (A) or 4.6% (B). Five percent of the vehicles emitted 31.0% of the HC mass (A) or 26.9% (B).

Results restricted to model years 2005 and newer are shown in Figure 25 and Table 5. The four highest-emission vehicles (0.1% of the sample, cumulative frequency 0.999) emitted 3.3% of the CO mass according to system A or 3.1% according to system B (Table 5). The highest five percent of the vehicles emitted 56.8% of the CO mass according to system A or 56.2% according to system B. For NO, the four highest-emission vehicles emitted 9.3% of the mass (A) or 8.7% (B). Five percent of the vehicles emitted 72.1% of the NO mass (A) or 55.2% (B). For HC, four vehicles emitted 2.4% of the mass (A) or 1.4% (B). Five percent of the vehicles emitted 30.1% of the HC mass (A) or 22.9% (B).

In summary, the measurement systems consistently indicated that five percent of the vehicles emitted approximately 50 to 70% of the CO and NO mass and about 20 to 30% of the HC mass.

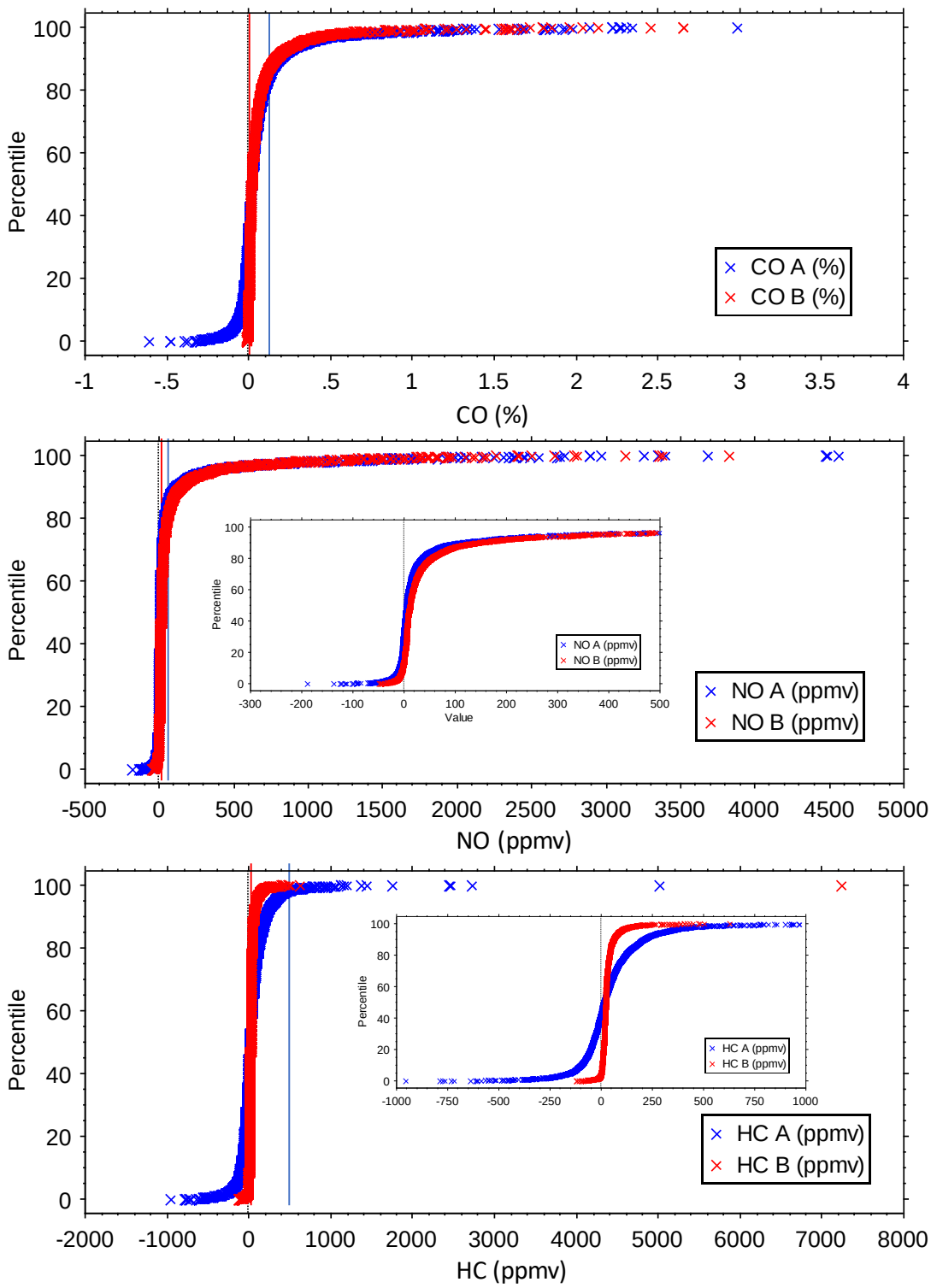


Figure 23. Cumulative frequency distributions of CO, NO, and HC concentrations as reported by the two measurement systems (blue = system A, red = system B). The CO scale excludes three values exceeding 4% (2 from A, one from B). Vertical lines denote one measure of uncertainty, σ_0 (Table 2), color-coded to match the data lines. Reported detection limits are approximately comparable to σ_0 values.

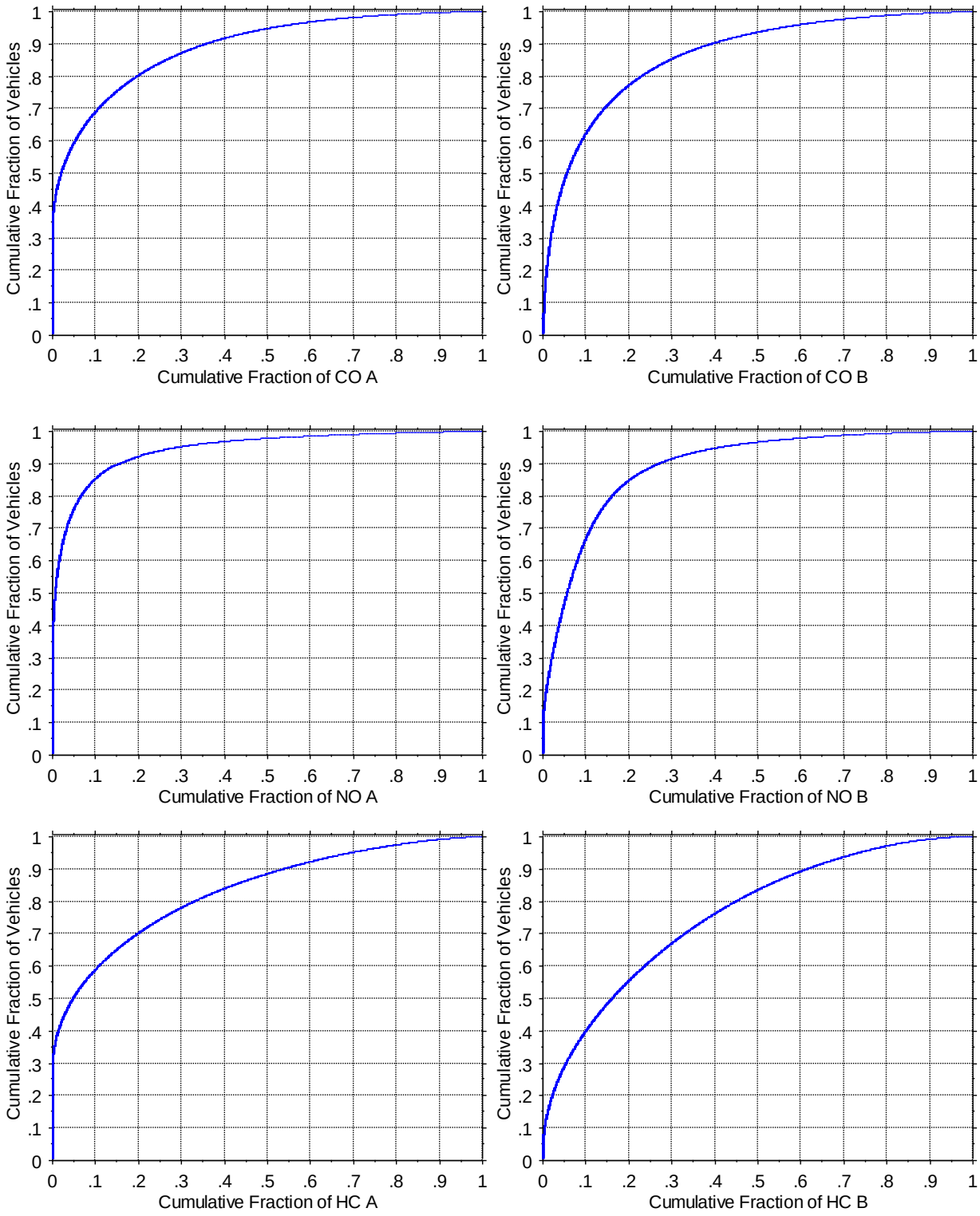


Figure 24. Cumulative fraction of vehicles versus cumulative fraction of emissions mass for all vehicles.

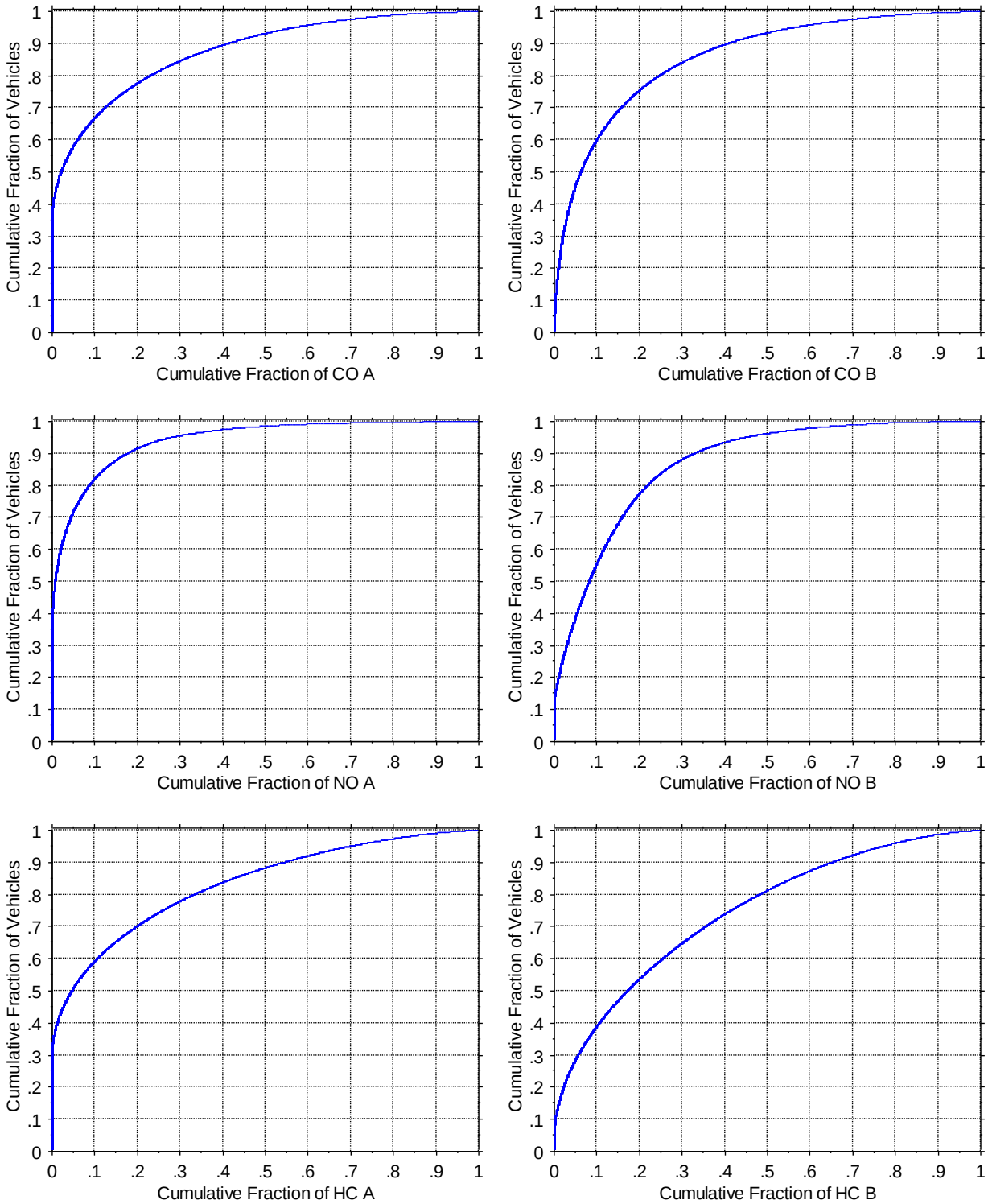


Figure 25. Cumulative fraction of vehicles versus cumulative fraction of emissions mass for model year 2005 through model year 2017 vehicles.

Table 4. Cumulative vehicle counts and emissions mass, full data set.

Vehicle Count	CO A	CO B	NO A	NO B	HC A	HC B
0.1	0	0.002	0	0	0	0.003
0.2	0	0.008	0	0.007	0	0.021
0.3	0	0.017	0	0.02	0	0.053
0.4	0.003	0.032	0.001	0.036	0.012	0.099
0.5	0.019	0.055	0.005	0.055	0.047	0.159
0.6	0.052	0.089	0.014	0.079	0.106	0.235
0.7	0.106	0.141	0.032	0.11	0.196	0.328
0.8	0.195	0.227	0.065	0.16	0.328	0.447
0.9	0.355	0.39	0.154	0.268	0.534	0.612
0.95	0.508	0.551	0.286	0.407	0.69	0.731
0.99	0.785	0.818	0.687	0.727	0.885	0.881
0.995	0.858	0.887	0.0803	0.815	0.925	0.915
0.999	0.946	0.969	0.941	0.932	0.971	0.954

Table 5. Cumulative vehicle counts and emissions mass, post-2004 vehicles.

Vehicle Count	CO A	CO B	NO A	NO B	HC A	HC B
0.1	0	0.003	0	0	0	0.003
0.2	0	0.009	0	0.009	0	0.022
0.3	0	0.02	0	0.029	0	0.056
0.4	0.002	0.037	0.001	0.053	0.011	0.105
0.5	0.02	0.062	0.006	0.081	0.045	0.171
0.6	0.058	0.099	0.015	0.115	0.105	0.253
0.7	0.123	0.156	0.043	0.158	0.197	0.354
0.8	0.229	0.246	0.086	0.217	0.332	0.48
0.9	0.441	0.406	0.175	0.327	0.542	0.65
0.95	0.568	0.562	0.279	0.448	0.699	0.771
0.99	0.816	0.821	0.569	0.712	0.893	0.916
0.995	0.823	0.887	0.702	0.787	0.931	0.948
0.999	0.967	0.969	0.907	0.913	0.976	0.986

4. CONCLUSIONS

Near-simultaneous measurements were made by two measurement systems on 4728 vehicles over three days. The average difference in the recorded times was 1.5 seconds. System clocks were not synchronized and an actual time difference of ~2 – 3 seconds is suggested by the known distance of 106 feet between measurement locations and the average vehicle speed. Since the duration of a plume measurement is less than one second for each system, the two systems did not measure identical exhaust plumes. For vehicles with consistently low emissions, temporal differences of ~2 – 3 s in the measurement of exhaust emissions are not expected to introduce large differences in measurement results. For some vehicles, observed differences may reflect inherent plume variability rather than systematic differences in measurement systems.

Measurement accuracy is established through independent audits using certified standards, which have been reported for both systems in previous studies. No independent audits were conducted in this study or in CRC E-106 and CRC E-119. The values obtained for σ_0 for one measurement system in this study were 1223 ppmv for CO (0.12%), 214 ppmv for HC, and 30 ppmv for NO. For the second measurement system, the values were 78 ppmv for CO (0.008%), 38 ppmv for HC, and 13 ppmv for NO. These uncertainty estimates are consistent with previously published σ_0 values and are comparable to published detection limits. Median concentrations were less than measurement uncertainties (with one exception) and 90th percentiles were greater than measurement uncertainties (with one exception).

Repeated measurements provide an opportunity to characterize vehicle-specific variations and are reported here to provide a benchmark of same-vehicle variability. Quantifying same-vehicle variability aids in evaluating the differences between the measurements made by the two sampling instruments, because the instruments did not measure exhaust plumes at identical times. Repeated measurements of 586 vehicles showed that pre-2004 model years tended to exhibit the largest variability in species concentrations. For one measurement instrument, the concentration ranges that covered 90% of the repeat-vehicle differences were ± 0.25% for CO, ± 93 ppmv for NO and, and ± 255 ppmv for HC. For the other instrument, the ranges were ± 0.19% for CO, ± 170 ppmv for NO, and ± 59 ppmv for HC. There was no evidence that larger repeat differences were associated with longer time differences between measurements, nor were larger repeat concentration differences associated with larger differences in repeated speed or acceleration values. Larger repeat measurement variations occurred across all observed operating bins of EPA's motor vehicle emission simulator model (MOVES). Some of the highest-emitting vehicles were consistently high emitters. Vehicles with the smallest differences in repeated measurements were consistently low emitters.

The two measurements systems yielded similar statistical distributions of vehicle speed, acceleration, concentrations of CO, NO, and HC, and ratios of concentrations of CO, NO, and HC relative to CO_2. The distributions of paired measurement system concentration differences were comparable in range to the distributions of the differences in repeated measurements: the ranges ± 0.2% for CO and ± 200 ppmv for NO and HC encompassed ~90% of the paired comparisons. This result indicates that measurements from systems A and B did not differ more than two sets of measurements from system A or two sets from system B. Thus, the variability of

the differences in paired A – B measurements from the two instrument systems could be due to inherent variability in exhaust plume concentrations, including variations in exhaust gas temperatures.

Small measurement differences were both detectable and detected. The large sample size (N = 4728) ensured that a paired t-test had very high statistical power, identifying statistically significant differences at $p < 0.05$ (95% confidence level) with a probability of 93% when the mean difference between paired measurements was only 5% of the standard deviation of the paired differences.

Differences in paired measurements of speed, acceleration, and concentrations of CO, NO, and HC were small relative to sample averages, measurement accuracy, and the inherent variability of exhaust plumes. While statistically significant ($p < 0.05$), differences in paired vehicle speed measurements averaged only 1 mph, or ~4% of the average speed of 25 mph. Paired vehicle acceleration differences were also statistically significant and averaged 0.6 mph s^{-1}. This difference corresponds to a modest rate of acceleration.

The mean paired concentration differences between measurement systems (A – B) were -0.002% CO (not significant), -6.9 ppmv NO (significant, $p < 0.05$), and 10.0 ppmv HC (significant, $p < 0.05$). Ranges of ± 0.2% for CO and ± 200 ppmv for NO and HC encompass ~90% of the paired comparisons. For individual measurements, the mean differences would be considered within the measurement limitations of the systems. When averaged over 4728 measurements, the mean NO and HC differences were statistically significant. When expressed as fuel-based emission rates, the mean paired differences between measurement systems were -0.05 g kg^{-1} CO (not significant), -0.87 g kg^{-1} NO (significant, $p < 0.05$), and 0.40 g kg^{-1} HC (significant, $p < 0.05$).

The two measurement systems consistently indicated that five percent of the vehicles emitted approximately 50 to 70% of the CO and NO mass and about 20 to 30% of the HC mass. Among post-2004 vehicles (N = 3909), four vehicles (0.1%) emitted 3% of the CO mass, 9% of the NO mass, and 2% of the HC mass.

REFERENCES

Anderson, D. C, C. P. Loughner, G. Diskin, A. Weinheimer, T. P. Canty, R. J. Salawitch, H. M. Worden, Alan Fried, T. Mikoviny, A. Wisthaler, R. R. Dickerson. 2014. Measured and modeled CO and NO_y in DISCOVER-AQ: An evaluation of emissions and chemistry over the eastern US. *Atmospheric Environment.* 96: 78–87. http://dx.doi.org/10.1016/j.atmosenv.2014.07.004.

Ashbaugh, L. L, D. R. Lawson, G. A. Bishop, P. L. Guenther, D. H. Stedman, R. D. Stephens, P. J. Groblicki, J. S. Parikh, B. J. Johnson, and S. C. Huang. 1992. On-road remote sensing of carbon monoxide and hydrocarbon emissions during several vehicle operating conditions. AWMA/EPA Conference on PM_{10} Standards and Nontraditional Particulate Source Controls, January, 1992. http://www.feat.biochem.du.edu/assets/publications/RS_CO_HC_during_several_driving_cond_PM10_AWMA_Conf_1992.pdf (last access, March 15, 2018).

Bishop, G. A. and D. H. Stedman. 2008. A decade of on-road emissions measurements. *Environ. Sci. Technol*. 42 (5): 1651–1656. doi: 10.1021/es702413b.

Bishop, G. A. and D. H. Stedman. 2015. Reactive nitrogen species emission trends in three light-/medium-duty United States Fleets. *Environ. Sci. Technol*. 49 (18): 11234–11240. doi: 10.1021/acs.est.5b02392.

Bishop, G. A. and M. J. Haugen. 2017. On-Road Remote Sensing of Automobile Emissions in the Chicago Area: Fall 2016. Coordinating Research Council, Inc., 5755 North Point Parkway, Suite 265, Alpharetta, Georgia 30022, Contract No. E-106, https://crcao.org/reports/recentstudies2017/E-106/E-106%20Chicago%202016%20report%20v3_June%202017.pdf (last access August 19, 2017).

Bishop, G. A., A. M. Peddle, D. H. Stedman, and T. Zhan. 2010. On-road emission measurements of reactive nitrogen compounds from three California cities. *Environ. Sci. Technol*. 44: 3616–3620. doi: 10.1021/es903722p

Bishop, G.A., B. G. Schuchmann, D. H. Stedman, and D. R. Lawson. 2012. Multispecies remote sensing measurements of vehicle emissions on Sherman Way in Van Nuys, California. *Journal of the Air & Waste Management Association*, 62 (10): 1127-1133. doi: 10.1080/10962247.2012.699015.

Bishop, G.A., D. H. Stedman, D. A. Burgard, and O. Atkinson. 2016. High-mileage light-duty fleet vehicle emissions: their potentially overlooked importance. *Environ. Sci. Technol*. 50: 5405−5411. doi: 10.1021/acs.est.6b00717.

Brioude, J., Angevine, W. M., Ahmadov, R., Kim, S.-W., Evan, S., McKeen, S. A., Hsie, E.-Y., Frost, G. J., Neuman, J. A., Pollack, I. B., Peischl, J., Ryerson, T. B., Holloway, J., Brown, S. S., Nowak, J. B., Roberts, J. M., Wofsy, S. C., Santoni, G. W., Oda, T., and Trainer, M.: Top-down estimate of surface flux in the Los Angeles Basin using a mesoscale inverse modeling technique: assessing anthropogenic emissions of CO, NO_x and CO_2 and their impacts. 2013. *Atmos. Chem. Phys*. 13: 3661–3677. doi:10.5194/acp-13-3661-2013.

Burgard, D. A., G. A. Bishop, R. S. Stadtmuller, T. R. Dalton, and D. H. Stedman. 2006. Spectroscopy applied to on-road mobile source emissions. *Applied Spectroscopy* 60 (5): 135A-148A.

Burgard, D.A., G. A. Bishop, D. H. Stedman, V. H. Gessner, and C. Daeschlein. 2006. Remote sensing of in-use heavy-duty diesel trucks. *Environ. Sci. Technol*. 40: 6938-6942. doi: 10.1021/es060989a.

Canty, T. P., L. Hembeck, T. P. Vinciguerra, D. C. Anderson, D. L. Goldberg, S. F. Carpenter, D. J. Allen, C. P. Loughner, R. J. Salawitch, and R. R. Dickerson. 2015. Ozone and NO_x chemistry in the eastern US: evaluation of CMAQ/CB05 with satellite (OMI) data. *Atmos. Chem. Phys*., 15, 10965–10982. doi:10.5194/acp-15-10965-2015.

Castellanos, P. Marufu, L. T., Doddridge, B. G., Taubman, B. F., Schwab, J. J., Hains, J. C., Ehrman, S. H., and Dickerson, R. R. 2011. Ozone, oxides of nitrogen, and carbon monoxide during pollution events over the eastern United States: An evaluation of emissions and vertical mixing. *J. Geophys. Res.* 116: D16307, doi:10.1029/2010JD01454.

Chinkin, L. R., D. L. Coe, T. H. Funk, H. R. Hafner, P. T. Roberts, P. A. Ryan, and D. R. Lawson. 2003. Weekday versus weekend activity patterns for ozone precursor emissions in California's South Coast Air Basin. *Journal of the Air & Waste Management Association* 53(7): 829-843. doi: 10.1080/10473289.2003.10466223.

Dallman, T. R. and R. A. Harley. 2010. Evaluation of mobile source emission trends in the United States. *Journal of Geophysical Research* 115, D14305, doi:10.1029/2010JD013862.

DenBleyker, A., J. Koupal, T. DeFries, and C. Palacios. 2017. Improvement of Default Inputs for MOVES and SMOKE-MOVES: CRC Project A-100. Coordinating Research Council, Inc., 5755 North Point Parkway, Suite 265, Alpharetta, Georgia 30022. https://crcao.org/reports/recentstudies2017/A-100/ERG_FinalReport_CRCA100_28Feb2017.pdf (last access, August 19, 2017).

Fujita, E. M., B. E. Croes, C. L. Bennett, D. R. Lawson, F. W. Lurmann, and H. H. Main. 1992. Comparison of emission inventory and ambient concentration ratios of CO, NMOG, and NOx in California's South Coast Air Basin. *Journal of the Air & Waste Management Association* 42: 264-276. doi: 10.1080/10473289.1992.10466989

Fujita , E. M. D. E. Campbell , B. Zielinska , J. C. Chow, C. E. Lindhjem, A. DenBleyker, G. A. Bishop , B. G. Schuchmann , D. H. Stedman, and D. R. Lawson. 2012. Comparison of the MOVES2010a, MOBILE6.2, and EMFAC2007 mobile source emission models with on-road traffic tunnel and remote sensing measurements. *Journal of the Air & Waste Management Association* 62: 1134-1149. http://dx.doi.org/10.1080/10962247.2012.699016.

Goldberg, D. L., C. P. Loughner, M. Tzortziou, J. W. Stehr, K. E. Pickering, L. T, Marufu, and R. R. Dickerson. 2014. Higher surface ozone concentrations over the Chesapeake Bay than over the adjacent land: Observations and models from the DISCOVER-AQ and CBODAQ campaigns. *Atmospheric Environment* 84: 9–19. http://dx.doi.org/10.1016/j.atmosenv.2013.11.008

Hager, J. S. 2017. Vehicle Emissions Partnership EDAR Pilot Program. Prepared for: Environmental Health & Trading Standards, West Lothian Civic Centre, Howden South Road, Livingston, West Lothian EH54 6FF, United Kingdom. https://www.heatremotesensing.com/studies-and-reports (last access, March 15, 2018).

Hager, J. S. 2018. On-Road Remote Sensing Of Automobile Emissions in the Rolling Meadows Area: Fall 2016. Prepared for: Coordinating Research Council, Inc., 5755 North Point Parkway, Suite 265, Alpharetta, Georgia 30022, Contract No. E-11. https://crcao.org/reports/recentstudies2018/E-119a/CRC%20Project%20E-119a%20Final%20Report_Mar2018.pdf (last access, March 8, 2018).

Haugen, M. J., and G. A. Bishop. 2018. Fall 2016 On-road Emission Measurements in the Chicago Area: Comparison of two University of Denver Remote Sensing Datasets. Coordinating Research Council, Inc., 5755 North Point Parkway, Suite 265, Alpharetta, Georgia 30022, Contract No. E-106, https://crcao.org/reports/recentstudies2018/E-

119a/CRC%20Project%20E-119a%20Final%20Report_Mar2018.pdf (last access May 31, 2018).

He, H., J. W. Stehr, J. C. Hains, D. J. Krask, B. G. Doddridge, K. Y. Vinnikov, T. P. Canty, K. M. Hosley, R. J. Salawitch, H. M. Worden, and R. R. Dickerson. 2013. Trends in emissions and concentrations of air pollutants in the lower troposphere in the Baltimore/Washington airshed from 1997 to 2011. *Atmos. Chem. Phys*. 13: 7859–7874. doi:10.5194/acp-13-7859-2013.

Heiken, J. G., M. Hixson, and J. Lyons. 2016. Review of EPA's MOVES2014 Model. CRC Report No. E-101. Coordinating Research Council, Inc., 5755 North Point Parkway, Suite 265, Alpharetta, Ga 30022.

Huang, M., K. W. Bowman, G. R. Carmichael, T. Chai, R. B. Pierce, J. R. Worden, M. Luo, I. B. Pollack, T. B. Ryerson, J. B. Nowak, J. A. Neuman, J. M. Roberts, E. L. Atlas, and D. R. Blake. 2014. Changes in nitrogen oxides emissions in California during 2005–2010 indicated from top-down and bottom-up emission estimates. *J. Geophys. Res. Atmos*. 119: 12,928–12,952, doi:10.1002/2014JD022268.

Ingalls, M.N., L.R. Smith, and R.E. Kirksey. 1989. Measurements of on-road vehicle emission factors in the California South Coast Air Basin—Volume I. Regulated emissions. Report no. SwRI-1604, prepared by Southwest Research Institute for the Coordinating Research Council, Atlanta, GA, June 1989.

Jimenez, J. L. Understanding and Quantifying Motor Vehicle Emissions with Vehicle Specific Power and TILDAS Remote Sensing. Ph.D. Thesis, Department of Mechanical Engineering, Massachusetts Institute of Technology, February 1999.

Kokoska, S. and C. Nevison. Statistical Tables and Formulae. Springer-Verlag: New York, 1989.

Lawson, D. R. 1990. The southern California air quality study. *J. Air Waste Manage. Assoc.* 40:156-165. http://dx.doi.org/10.1080/10473289.1990.10466671.

Lindhjem, C. E., A. K. Pollack, A. DenBleyker, and S. L. Shaw. 2012. Effects of improved spatial and temporal modeling of on-road vehicle emissions. *Journal of the Air & Waste Management Association* 62 (4): 471-484. http://dx.doi.org/10.1080/10962247.2012.658955.

McDonald, B. C., T. R. Dallmann, E. W. Martin, and R. A. Harley. 2012. Long-term trends in nitrogen oxide emissions from motor vehicles at national, state, and air basin scales. *Journal of Geophysical Research*, 117, D00V18, doi:10.1029/2012JD018304.

Miller, C. A., G. Hidy , J. Hales, C. E. Kolb, A. S. Werner, B. Haneke, D. Parrish, H. C. Frey, L. Rojas-Bracho, M. Deslauriers, B. Pennell, and J. D. Mobley. 2006. Air Emission Inventories in North America: A Critical Assessment. *Journal of the Air & Waste Management Association*. 56(8): 1115-1129. doi:10.1080/10473289.2006.10464540

Parrish, D. D. 2006. Critical evaluation of US on-road vehicle emission inventories. *Atmospheric Environment* 40: 2288–2300. doi:10.1016/j.atmosenv.2005.11.033.

Pierson, W. R., A. W. Gertler, and Ronald L. Bradow. 1990. Comparison of the SCAQS tunnel study with other on road vehicle emission data. *Journal of the Air & Waste Management Association* 40(11): 1495-1504, doi: 10.1080/10473289.1990.10466799.

Popp, P. J., G. A. Bishop, and D. H. Stedman. 1999. Development of a high-speed ultraviolet spectrometer for remote sensing of mobile source nitric oxide emissions, *Journal of the Air & Waste Management Association*, 49:1463-1468, doi: 10.1080/10473289.1999.10463978.

Ropkins, K., T. H. DeFries, F. Pope, D. C. Green, J. Kemper, S. Kishan, G. W. Fuller, H. Li, J. Sidebottom, L. R. Crilley, L. Kramer, W. J. Bloss, and J. S. Hager. 2017. Evaluation of

EDAR vehicle emissions remote sensing technology. *Sci Tot. Environ*. 609: 1464–1474. http://dx.doi.org/10.1016/j.scitotenv.2017.07.137.

Singer, B. C. and R. A. Harley. 1996. A fuel-based motor vehicle emission inventory. *Journal of the Air & Waste Management Association* 46: 581-593.

Singer, B. C., R. A. Harley, D. Littlejohn, J. Ho, and T. Vo. 1998. Scaling of infrared remote sensor hydrocarbon measurements for motor vehicle emission inventory calculations. *Environmental Science & Technology*, 32: 3241-3248, doi: 10.1021/es980392y.

St. Denis, M. J., R. Budd, S. Hager, Y. Hager. 2015. Connecticut Vehicle Inspection Program 2014 On-Road Vehicle Survey. https://docs.wixstatic.com/ugd/3ee5f0_b6f8746eac594abeacc997db69f23da4.pdf (last access, August 24, 2017).

Streets, D. G., T. Canty, G. R. Carmichael, B. de Foy, R. R. Dickerson, B. N. Duncan, D. P. Edwards, J. A. Haynes, D. K. Henze, M. R. Houyoux, D. J. Jacob, N. A. Krotkov, L. N. Lamsal, Y. Liu, Z. Lu, R. V. Martin, G. G. Pfister, R. W. Pinder, R. J. Salawitch, and K. J. Wecht. 2013. Emissions estimation from satellite retrievals: A review of current capability. *Atmospheric Environment* 77: 1011–1042. http://dx.doi.org/10.1016/j.atmosenv.2013.05.051.

Travis, K. R., D. J. Jacob, J. A. Fisher, P. S. Kim, E. A. Marais, L. Zhu, K. Yu, C. C. Miller, R. M. Yantosca, M. P. Sulprizio, A. M. Thompson, P. O. Wennberg, J. D. Crounse, J. M. St. Clair, R. C. Cohen, J. L. Laughner, J. E. Dibb, S. R. Hall, K. Ullmann, G. M. Wolfe, I. B. Pollack, J. Peisch, J. A. Neuman, and X. Zhou. 2016. Why do models overestimate surface ozone in the Southeast United States? *Atmos. Chem. Phy*s. 16: 13561–13577. doi:10.5194/acp-16-13561-2016.

U.S. Environmental Protection Agency (EPA). 2015. Exhaust Emission Rates for Light-Duty On-road Vehicles in MOVES2014. EPA-420-R-15-00.5 https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P100NNVN.pdf (last access August 22, 2017).

U.S. Environmental Protection Agency (EPA). 2017. U.S. EPA Response to CRC Project No. E-101, Review of EPA's MOVES2014 Model. https://www.epa.gov/sites/production/files/2016-10/documents/420r16012.pdf (last access August 22, 2017).

APPENDIX: Summary of Emission Model and Inventory Evaluations

Table A1. Example measurement-based assessments of emission inventory estimates.

Study	Inventory	Emission Models	Study Dates	Inventory/model vs. measurements
Ground-based monitoring				
Fujita et al. (1992)	Day-specific, hourly, gridded, March 1991 version	EMFAC7E	11 days, Jun – Dec, 1987, Los Angeles	Ambient CO/NO_x and VOC/NO_x ratios were ~1.5 and 2 – 2.5 times higher
Parrish et al. (2006)	NA	MOBILE5 and MOBILE6	1985-2002, national	Agreement for NO_x as of 2000; CO and VOC emissions overestimated
Tunnel studies				
Pierson et al. (1990); Ingalls et al. (1989)	NA	MOBILE4 EMFAC7C	8 days, Oct and Dec 1987, Van Nuys tunnel, Los Angeles	Agreement for NO_x; CO, VOC emissions underestimated ~2X and 3X
Fujita et al. (2012)	Location-specific applications of mobile emission models	MOVES 2010a MOBILE6.2 EMFAC2007	August 2010, Van Nuys tunnel, Los Angeles	MOVES: WD 1.46x, WE 1.38x ambient MOBILE: WD 1.33x WE 1.29x EMFAC: WD 0.84x, WE 0.70x ambient
Fuel-based inventory (emission factors from tunnel studies, ground-based remote sensing, etc.)				
Dallman and Harley (2010)	NEI2005v2	MOBILE6	1996-2006	Agreement for NO_x, measured diesel higher than NEI
McDonald et al. (2012)	NEI1999, 2002, 2005, 2008	MOBILE6, MOVES	1990-2010	Agreement in 2002, 2005, 2008
Ground-based remote sensing				
Bishop et al. (2012) and Fujita et al. (2012)	Location-specific applications of mobile emission models	MOVES 2010a MOBILE6.2 EMFAC2007	August 2010, Van Nuys tunnel, Los Angeles	MOVES emission factor +10% relative to remote sensing
Aircraft studies				
Brioude et al. (2013)	NEI2005v2, CARB 2008	MOBILE, EMFAC; inverse modeling	May – June 2010	NEI: +27%, CARB: <+15%
Anderson et al. (2014)	NEI2008 projected	MOVES2010; CMAQ	July – August 2011	NEI: +51 to +70%
Travis et al. (2016)	NEI 2011 v1 scaled by 0.89 plus soil and fertilizer NO_x	MOVES2010a;	August – September 2013	NEI: +30 to +60% (30% if soil & fertilizer NO_x = zero)
Satellite studies				
Canty et al. (2015)	Study specific	MOVES2010	July – August 2007	NEI mobile 2X high