CRC Report No. E-119-3a

REMOTE SENSING DEVICE (RSD) STATISTICAL ANALYSIS

Final Report

April 2023

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.

LEGAL NOTICE

This report was prepared by Charles Blanchard as an account of work sponsored by the Coordinating Research Council (CRC). Neither the CRC, members of the CRC, Charles Blanchard, nor any person acting on their behalf: (1) makes any warranty, express or implied, with respect to the use of any information, apparatus, method, or process disclosed in this report, or (2) assumes any liabilities with respect to use of, inability to use, or damages resulting from the use or inability to use, any information, apparatus, method, or process disclosed in this report. In formulating and approving reports, the appropriate committee of the Coordinating Research Council, Inc. has not investigated or considered patents which may apply to the subject matter. Prospective users of the report are responsible for protecting themselves against liability for infringement of patents.

Revised Draft Final Report for CRC Project No. E-119-3a

Remote Sensing Device (RSD) Statistical Analysis

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

March 31, 2023

PREFACE

Remote sensing systems have added greatly to knowledge of motor vehicle exhaust emissions in real-world driving conditions. It has been a privilege to support the research and development programs of the teams that conducted CRC project E-119-3 and previous studies through this statistical analysis of their data. I thank all participants for providing their data and supplemental information. I also thank CRC Deputy Director Amber Leland and Administrative Research Assistant Rebecca Kang for facilitating this project, and the members of the CRC Emissions Committee, who initiated the project and provided valuable direction, review, and feedback.

CONTENTS

FIGURES

TABLES

ACRONYMS AND ABBREVIATIONS

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)
ppmv	parts per million by volume
SE	standard error
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. Comparisons that demonstrate consistency and reproducibility of measurements made by different systems are needed to facilitate analyses of vehicle fleet emissions.

This report evaluates the comparability of three ground-level vehicle-exhaust remote-sensing device (RSD) systems. The measurements were made between April 12 and 16, 2021, on the entrance ramp from eastbound U.S. 60 to northbound state highway 101 in Phoenix, AZ, using: (1) the Hager Environmental & Atmospheric Technologies (HEAT) emissions data and reporting (EDAR) system (CRC E-119), (2) the Denver University (DU) fuel efficiency automobile test (FEAT) system (CRC E-106, E-119a, E-123, E-119-3), and (3) the Opus Inspection remote sensing device (https://www.opus.global/vehicle-inspection/remote-sensing/, last access April 18, 2020). The primary focus of the analysis is on measurements of carbon monoxide (CO), nitric oxide (NO), nitrogen dioxide (NO_2), and hydrocarbons (HC), parameters that all three RSD systems report. Further comparisons of vehicle speed and acceleration, sampling completeness (fraction of missed events), and fraction of high-emitting vehicles (model years 2005 and newer) are reported. Additional analyses were conducted for control, calibration, and evaporative emission test vehicles.

Fleet Vehicle Exhaust

Vehicle exhaust measurements were not equivalent among RSDs. However, future studies can add variance to mean HEAT or Opus concentrations to account for instrument differences compared with the historical record provided by past DU studies. Doing so will permit trends to be determined from multiple RSD data sources even if consistent adjustment factors cannot be established. All RSD systems detected high emissions at similar frequencies: ~0.1 to 2 % of vehicle passes, varying with species and concentration threshold.

Mean paired differences in the 2021 Phoenix study were larger than those obtained for a comparison of DU and HEAT measurements made in Chicago in 2016. For CO and NO, the 2021 and 2016 comparisons were not directionally consistent (i.e., HEAT values were not consistently higher or lower than DU). An important caveat noted by the DU principal investigator is that the DU setup, data collection program, and software were different in Phoenix than in DU’s previous standard E-23/116/123 studies due to measurement and estimation of evaporative emissions in Phoenix.

For the full fleet, concentration differences between 2021 DU, HEAT, and Opus paired measurements were statistically significant for 11 of 12 comparisons (Table S-1). The RSD differences varied among species, were larger than measurement uncertainty, were evident throughout ranges of concentrations, and were not uniform across concentrations. Concentration differences were unrelated to differences in vehicle specific power. Opus-HEAT differences in CO, HC, and NO concentrations were consistently smaller than Opus-DU and HEAT-DU. Opus exhibited higher NO_2 and lower NO values than either DU or HEAT. These differences were more extreme for diesel than gasoline vehicles.

Table S-1. Statistical summary of paired DU, HEAT, and Opus differences: CO, HC, NO, and NO_2 ± 1 standard error of the mean excluding values flagged as invalid. Units are ppmv. *** = highly significant difference ($p < 0.0001$), ** = significant ($p<0.01$), * = significant ($p< 0.05$). N = 8631 paired vehicle comparisons.

Species	**N Matched Vehicle Passes, Valid Flags**	**HEAT - DU**	**Opus - HEAT**	**Opus - DU**
CO	8631	-169.8 ± 15.3***	-33.8 ± 11.8**	-203.6 ± 19.3***
HC[a]	8629	-54.7 ± 2.5***	6.2 ± 1.5***	-48.5 ± 1.8***
NO	8630	-20.1 ± 1.8***	-11.5 ± 1.2***	-31.62 ± 2.3***
NO_2	8553	0.05 ± 0.2	6.4 ± 0.7***	6.5 ± 0.7***

a. For DU offset HC, HEAT-DU = -34.7 ± 2.5*** ppmv; Opus-DU = -28.5 ± 1.8*** ppmv

The larger differences between RSD measurements observed in 2021 are not explained by instrumental uncertainties, which were lower in 2021 than in 2016 for both CO and HC (but somewhat higher for NO). Means were determined from large sample sizes (N = 4728 in 2016 and 8631 in 2021), which reduce the instrumental uncertainties of the means by $N^{1/2}$ so that instrumental uncertainties for mean concentrations were smaller than population variability (i.e., the standard errors of the means, Table S-1).

Individual measurements less than detection limits may not be well quantified but are reliably low. Mean concentrations less than detection limits are better quantified because they were determined from large samples. For DU, mean 2016 and 2021 CO, HC, and NO concentrations were lower than the calculated detection limits for individual measurements; for HEAT, mean 2016 and 2021 HC concentrations were lower than the calculated detection limits. For Opus in 2021, mean HC and NO concentrations were lower than the calculated detection limits.

The 2021 average differences between RSD concentrations were smaller than concentration trends observed in previous DU studies between 1998 and 2021. Each RSD system revealed concentration declines in 2021 compared with previous DU studies but substituting 2021 HEAT or Opus for DU concentrations would overestimate the 2021 improvements relative to past years (i.e., 2021 HEAT and Opus concentrations tended to be lower than DU values).

All RSD systems captured a range of motor vehicle emission simulator model (MOVES) operating modes. Average CO and NO concentrations showed little variability across MOVES modes. Average HC concentrations also showed little variability across MOVES operating modes, though DU values tended to be higher in the lowest modes.

Fleet Vehicle Speed, Acceleration, and Specific Power

All pairwise differences in speed, acceleration, and vehicle specific power (VSP) were highly significant ($p < 0.0001$). Many of these differences were not large relative to the averages, though. The speed differences ranged from 4 to 9% of the average speeds. The differences between the DU values compared to those recorded by the other two systems were 19 to 38% higher for DU acceleration and 9 to 22% higher for DU VSP. The speed and acceleration differences need not reflect measurement differences between RSDs, since they could have occurred due to increasing vehicle speeds along the entrance ramp (Opus preceded HEAT, which preceded DU, where distances between systems were as small as possible, ~3 to 8 m).

Evaporative Emissions

In summary:

- Measurements of evaporative emissions were not quantitative as indicated by:
  - Little agreement among RSDs in both the full fleet and test vehicles,
  - Unreliable qualitative distinctions (presence/absence inconsistent among RSDs).
- The Opus ERG evaporative index (but not the Opus Envirotest indicator) overestimated evaporative emissions:
  - Much higher fraction of evaporative detections in full fleet than HEAT or DU,
  - Positive values for blanks in simulated emissions,
  - Positive intercept when measurements regressed against simulated emissions.

For full-fleet evaporative emission measurements (non-test vehicles):

- Rates of detection of evaporative emissions were different among the three RSDs.
  - The Opus Eastern Research Group (ERG) indicator of evaporative emissions identified a much higher percentage of vehicles (19%) having evaporative emissions than did either the Opus Envirotest score (0.2%) or the DU running loss indicator (RLI) (0.4%).
  - HEAT reported that only 3 of 43,205 valid vehicle measurements (0.007%) had detectable evaporative emissions; 1 of these 3 was in the 8631-vehicle merged data set (0.01%) (this vehicle had the maximum DU RLI and Opus ERG index).
- The correlation between the DU RLI and Opus ERG Index was modest ($r^2 = 0.2$).
- Detections of evaporative emissions were reported for nearly all model years by DU and the Opus Envirotest score (low frequencies of detection) and the Opus ERG Index (higher frequencies). (The Opus Envirotest score was recorded as discrete integer values ranging from 0 to 4; values of zero and 1 were reported as non-detections).

Test Vehicles

Test vehicle data were compiled and analyzed by Revecorp Inc. The Revecorp analyses constitute the primary findings for experiments with test vehicles. To facilitate interpretation of the full-fleet comparisons in this report, additional supplementary data analyses were carried out.

For calibration tests and comparison with portable emissions measurement system (PEMS) data:

- The results of the tests were informative but limited by small sample sizes (N = 11 to 15 calibration measurements for each RSD at each of two pollutant concentrations; N = 11 to 25 comparisons with PEMS for each RSD and pollutant) and incomparable units of measurement (e.g., unscaled indices versus g mi^{-1}).
- For CO, CO_2, and NO_x, calibration test results were generally consistent among RSDs and between RSD measurements and nominal test concentrations. Some DU NO_x values were higher than nominal. The mean DU HC response to 255.2 ppmv propane was high and its response to 756.2 ppmv propane was low. Variability was high and the lower calibration point was less than the calculated DU detection limit.
- Statistically significant intercepts and high values for blanks imply overestimation of evaporative emissions by the Opus ERG index but not the Opus Envirotest indicator.
- The calibration tests provided useful supplementary information on measurement uncertainties. Uncertainties as determined from test vehicle calibrations were the same order of magnitude as instrumental uncertainties determined from the full-fleet data set.
- For the Mazda test vehicle, HEAT and PEMS CO emission rates (g kg^{-1} fuel) were highly correlated ($r^2 = 0.94$) but differed in magnitude (regression slope HEAT vs PEMS = 0.57 ± 0.05, PEMS range 0 to 10 g kg^{-1}). DU, Opus, and PEMS each identified one high CO value (1 to 1.4% CO for DU and Opus, 148 g kg^{-1} for PEMS). Otherwise, RSD CO, HC, and NO_x values were not correlated with PEMS emission rates ($r^2 < 0.1$).

For evaporative emission measurements made on test vehicles:

- All RSD systems responded to simulated evaporative emissions. The minimum butane flow rate of simulated emissions was 1 liter per min (L min^{-1}), which is equivalent to 3 to 6 g mi^{-1} butane. This minimum rate is about two orders of magnitude greater than the U.S. EPA running loss standard of 0.05 g mi^{-1} (U.S. EPA, 2016).
- Results were not statistically different for different leak locations.
- Correlations between HEAT measurements and calculated emission rates were high (coefficient of variation, $r^2 = 0.74$ to 0.85) but HEAT values were about a factor of three higher than calculated rates (regression slope = 2.44 ± 0.34 to 4.46 ± 0.43 g mi^{-1}; subject to change because HEAT personnel are reworking the method for calculating g mi^{-1}).
- The Opus and DU measurements were reported as indices, whose conversion to g mi^{-1} would require additional information and data processing that were not part of the study. Correlations (as coefficient of variation) between indices and calculated emission rates were mixed: $r^2 = 0.19$ to 0.53 for DU, $r^2 = 0.36$ to 0.64 for the Opus ERG index (no-intercept regressions yield higher r^2 but the regressions had statistically significant intercepts), and $r^2 = 0.16$ (nondetects only) to 0.55 for the Opus Envirotest indicator.
- RSD measurements of actual evaporative emissions, conducted by removing fuel caps, disconnecting vapor lines, or disconnecting purge valves, were inconsistent (r^2 values of 0 to 0.25). Since the actual emission rates were not known, absolute accuracy could not be assessed.

1. Introduction

1.1 Background

Accurate values of air pollutant emission rates are needed to assess emission source contributions to ambient air pollution and to evaluate the effectiveness of past and proposed air quality measures. Ground-based remote-sensing measurements provide direct information about vehicle emissions in "real-world" situations, can be carried out in many locales, and do not require the types of interpretation needed to convert aircraft or satellite measurements of ambient pollutant concentrations into estimates of emissions. Credibility of the remote-sensing measurements is enhanced by cross-comparisons that demonstrate consistency and reproducibility of the real-world measurements across different approaches. Cross comparisons can also help ensure continuity of past and future remote-sensing data so that a long-term data record can be maintained using measurements from different systems as needed.

Gruening et al. (2019) evaluated the Hager Environmental & Atmospheric Technologies (HEAT) and Opus RSD systems using reference vehicles equipped with portable emission measurement systems (PEMS) and electric vehicles carrying emittable gas standards. Good agreement was found between PEMS or gas standards and the ratios of NO, NO_2, and CO to CO_2 reported by both HEAT and Opus RSDs. Time alignment of PEMS and remote-sensing data was necessary to ensure that PEMS and the remote-sensing devices were not sampling different exhaust plumes corresponding to different engine states (Gruening et al., 2019). For NO, the PEMS vs remote-sensing relationships resulted in slopes of 1.03 ± 0.01 ($r^2 = 0.98$) and 0.92 ± 0.01 ($r^2 = 0.97$) (test results were blinded), indicating strong correlations and slopes near unity. Because NO_2 concentrations were a factor of five lower than NO, relative variability was higher resulting in slopes of NO_2 versus CO_2 of 1.15 ± 0.25 ($r^2 = 0.21$) and 0.82 ± 0.06 ($r^2 = 0.73$) (each system exhibited a consistent high or low bias for NO and NO_2). For CO, the PEMS vs remote-sensing slopes were 0.88 ± 0.02 ($r^2 = 0.96$) and 0.97 ± 0.01 ($r^2 = 0.99$). These comparisons suggest that the remote systems achieved accuracies in the range of 3 to 18% relative to PEMS values under the reported test conditions.

In a previous project (CRC E-119-2), the comparability of DU and HEAT measurements was evaluated through consideration of measurement uncertainty, variability of same-vehicle measurements, paired differences, and consistency of speed and acceleration (Blanchard, 2018). The mean paired concentration differences between measurement systems (DU – HEAT) were -0.002% (-20 ppm) CO (not statistically 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 encompassed ~90% of the paired comparisons. 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, $p < 0.05$).

1.2 Project Objectives

The principal objective of this report is to evaluate the comparability and interchangeability of measurements made using three different ground-level vehicle-exhaust remote-sensing device (RSD) systems. The measurements were made by three research teams between April 12 and 16, 2021, on the entrance ramp from eastbound U.S. 60 to northbound state highway 101 in Phoenix, AZ, using: (1) the Hager Environmental & Atmospheric Technologies (HEAT) emissions data and reporting (EDAR) system (CRC E-119), (2) the Denver University (DU) fuel efficiency automobile test (FEAT) system (CRC E-106, E-119a, E-123, E-119-3), and (3) the Opus International remote sensing device (https://www.opus.global/vehicle-inspection/remote-sensing/, last access April 18, 2020). The primary focus of the analysis is on measurements of carbon monoxide (CO), nitric oxide (NO), nitrogen dioxide (NO_2), and hydrocarbons (HC), parameters that all three RSD systems report. Additional analyses of evaporative (EVAP) and particulate matter (PM) emissions are discussed for systems that record EVAP and PM. Further comparisons of vehicle speed and acceleration measurements, sampling completeness (fraction of missed events), and fraction of high-emitting vehicles (model years 2005 and newer) are reported. Separate analyses are conducted for control, calibration, and EVAP/PM test vehicles.

The technical questions addressed in this report, as listed in the project RFP, are:

1. How well do the measured emissions compare? How does this comparison vary across vehicle types and emissions levels?
2. How well can the three measurement systems detect classes of vehicles, such as high emitters?
3. How do the measurement variabilities of the three systems compare?
4. How well can the systems measure at low levels? What are the limits of detection?
5. What is the fraction of valid readings for both systems?
6. How well do the systems record vehicle speed and acceleration?
7. How effectively can the systems cover a range of vehicle specific power bins (as used in EPAs MOVES model)?
8. What additional information can be ascertained when considering the control/calibration vehicle(s)?

Test vehicle data were compiled and analyzed by Revecorp Inc. (see CRC RW-105 project report). The analyses of test vehicle data in this report are supplementary to the principal analysis by Revecorp. The supplementary analyses of test vehicle data are presented here to assist in the interpretation of the measurements of vehicle exhaust and evaporative emissions recorded for the full vehicle fleet by each RSD system.

2. Approach

2.1 Data Acquisition and Description

The Phoenix field study was originally planned for October 2020 but was delayed to April 2021 due to the Covid-19 pandemic. Measurements for test vehicles were compiled into a data set by Revecorp Inc. in August 2021. Access to the DU, HEAT, and Opus data sets was provided in February 2022, following completion of data compilation and validation by each organization and provision of vehicle information by the State of Arizona.

Each of the research teams carried out quality assurance and control procedures on their own data sets as described in Bishop and Haugen (2017), Haugen and Bishop (2018), Bishop (2022), Hager (2018; 2022), and Klausmeier and Vescio (2022). The principal reasons that research teams invalidated measurements or excluded them from data sets were:

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

Capture rates (ratio of valid to attempted measurements) could not be directly determined from any of the Phoenix data sets because the data consisted of valid vehicle measurements only. However, the RSD E-119-3 project reports provide relevant information. The DU report states that 27,561 measurements were attempted, of which 25,854 (NO_2) to 26,495 (CO) were valid; capture rates were therefore 94% to 96%, depending on the chemical species (Bishop, 2022). HEAT reported 43,205 valid measurements out of 53,063 attempts, or a capture rate of 81% (Hager, 2022). Opus reported 33,434 valid measurements but did not indicate the number of attempted measurements or characterize their capture rates (Klausmeier and Vescio, 2022).

The measurement counts for the individual data sets are listed in Table 1.

The DU and HEAT data sets included the state-provided vehicle information, whereas the Opus data set did not. The Opus project report indicates that the State matched 24,310 Opus measurements, compared with 29,338 vehicles with Arizona license plates (Table 1). I linked Opus data to state vehicle information by merging the three RSD data sets by license plate, day, and time.

The number of vehicle measurements varied among data sets due to differences in hours of operation, capture rates, and the way in which the data had been compiled. All RSD systems operated during the same scheduled hours (7 a.m. through 5 p.m.) on each of five test days but differences in starting and ending times are evident in the data sets. For example, HEAT measurements begin at midnight of the morning of April 12.

Since the state of Arizona provided information only for vehicles registered in Arizona, some differences in the number of records in the three data sets resulted from differing treatment of the information received from the State:

- In the DU data set, measurements had been merged with the state-provided information. The data set excluded vehicles that the State was unable to identify from the submitted license-plate data. Only AZ vehicles with matched plates are reported in the DU data.
- The HEAT data records had also been merged with state-provided information but the data set included all vehicle records, whether matched by the State or not. Data for unmatched AZ vehicles and out-of-state vehicles were included in the data set.
- The Opus data had not been merged with the state-provided information and therefore included all vehicles, AZ and out-of-state.

Table 1. Number of records in individual DU, HEAT, and Opus data sets. NR = not reported.

Species	DU (matched, with valid flags)	HEAT (all records)	HEAT (records matched by State)	Opus (all records)	Opus (AZ plates only)
CO	18,294	43,205	30,510	33,434	29,338
CO_2	18,294	NR	NR	33,434	29,338
HC	18,276	43,205	30,510	33,434	29,338
NO	18,288	43,205	30,510	33,434	29,338
NO_2	18,288	43,205	30,510	33,434	29,338
NH_3*	18,261	NR	NR	NR	NR

*ammonia

Since the vehicle populations differed among individual RSD data sets due to differing treatment of unmatched vehicles, summary tables for individual data sets may exhibit differences in reported mean values due to population variations. Tables that are based on merged data sets, which compare measurements from vehicles that were recorded by all three RSD systems, represent the same vehicle populations for each system. Both are presented in Section 3.

Although the DU data set only includes AZ vehicles matched by the State, the DU project report indicates that the match rate (percent of submitted plates matched by the State) was 86%; the DU report states that this is lower than typical match rates (high 90s%) for unknown reasons. In the HEAT data set, 91% of the vehicles (39,375 out of 43,205) were registered in AZ. Further restriction of the HEAT data to vehicles with AZ license plates that were matched by the State yields 30,510 records (Table 1), which is 77% of the AZ vehicles and 71% of the full HEAT data record. In the Opus data set, 88% of the vehicles (29,338 out of 33,434) were registered in AZ (Table 1); six instances of an AZ-registered test vehicle (Chevrolet Bolt, used for calibration) were found and excluded.

The three data sets were merged by day and license plate. The merged data were then restricted to cases when all three RSD systems recorded the same vehicle (even if some species measurements were flagged as invalid in any data set), ensuring that identical populations were obtained by each system. In addition, data sets were compiled from three two-way merges (DU – HEAT, DU – Opus, HEAT – Opus) to compare two-way and three-way vehicle counts and averages.

Incorrect merges were removed by requiring consistency of time stamps. An incorrect merge could occur, for example, if a vehicle passed the measurement location more than once on a day and was recorded at least once by all three systems (the merging algorithm pairs each instance of the passing vehicle with each other instance).

The three RSD systems were collocated to the extent possible, with initial separation distances of approximately 3 m. However, Opus moved further from HEAT (~ 6 – 8 m ahead of HEAT) on Tuesday afternoon, after experiencing too many optical alignment alarms linked to vehicles with rigid suspensions vibrating the road as they passed over HEAT's raised reflector strip (see discussions in Opus and Revecorp project reports).

On all days, vehicles first passed the Opus system, then HEAT, then DU. However, the mean time differences indicate the reverse order (i.e., earliest times for DU, then HEAT, then Opus) and are too large for the small distances involved (Table 2). The time differences should be approximately one second or less considering inter-system distances and average vehicle speeds. The comparisons therefore indicate that the RSD measurements were independently time stamped using different methods and were not synchronized among systems.

Table 2. Number of three-way co-recorded vehicles (i.e, simultaneously recorded by DU, HEAT, and Opus), mean differences in recorded times for matched vehicles, and start and end times when all systems reported data, by day.

Day	N Vehicles (three-way matches)	Mean Time Difference HEAT-DU (s)	Mean Time Difference Opus-HEAT (s)	Mean Time Difference Opus-DU (s)	Start Time	End Time
All days	8631	50.3 ± 0.09	16. 3 ± 0.02	66.6 ± 0.07	Varied	Varied
Monday	1538	33.5 ± 0.01	20.0 ± 0.01	53.5 ± 0.01	8:50:35	17:11:46
Tuesday	1564	56.1 ± 0.01	17.3 ± 0.01	73.4 ± 0.01	8:34:35	16:31:12
Wednesday	1742	55.4 ± 0.01	15.2 ± 0.01	70.6 ± 0.01	8:35:34	17:01:41
Thursday	1961	53.4 ± 0.01	15.1 ± 0.01	68.5 ± 0.01	8:31:17	16:49:41
Friday	1826	51.4 ± 0.01	14.6 ± 0.01	66.0 ± 0.02	7:28:59	14:48:36

2.2 Measurements

The DU system is fully described in published literature (Bishop and Stedman, 1996; 2008; 2015; Bishop et al., 2010; 2012; 2016; 2020; Burgaard et al, 2006; Popp et al., 1999) and in the DU Phoenix project report (Bishop, 2022). The DU instrumentation consists of a non-dispersive infrared (NDIR) device for detecting CO, CO_2, and HC emissions plus twin dispersive ultraviolet (UV) beams for detecting NH_3, NO, NO_2, and SO_2 emissions (Bishop, 2022). Source and detector units are placed at tailpipe height on opposite sides of a single-lane roadway (e.g., a freeway entrance ramp).

The Opus RSD5300 measurement system is described in the Opus project report (Klausmeier and Vescio, 2022). The 5300 instrument consists of an NDIR component for detecting CO, CO_2, HC, and infrared (IR) smoke and a dispersive UV spectrometer for measuring NO, NO_2, and UV smoke. The source and detector elements are adjacent in a single module. Two 5300 units were utilized in the Phoenix study, one using standard 12-inch beam heights and a second with beam heights of 18 to 23 inches (Klausmeier and Vescio, 2022).

HEAT 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 (Hagar, 2018; Ropkins et al., 2017). The HEAT unit is located approximately 5 meters above the roadway and looks down on the plume, so the height of the tailpipe is inconsequential. The gas sensor can detect an entire exhaust plume as it exits a vehicle, which allows measurement of absolute amounts of pollutants (Hagar, 2018; 2022). The HEAT system is designed to measure targeted pollutants without explicit field calibration based on the absolute nature of the spectroscopic measurements. HEAT processes measurements to generate emission rates in mass per unit travelled (grams per mile or grams per kilometer) but these values were not included in the Phoenix data set.

Ropkins et al. (2017) discuss key features and differences between the DU, Opus, and HEAT systems.

Each RSD data set included 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 Appendix A. 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 DU data set were CO, CO_2, ammonia (NH_3), NO, NO_2, and HC. In the Opus data, species measurements were CO, CO_2, NO, NO_2, HC-propane, HC-hexane, UV smoke, and evaporative index. The species measurements included in the HEAT data set were CO, NO, NO_2, HC, and PM.

The HC measurements represent sums of many hydrocarbon species. The measurement systems' sensitivities to each of the large number of hydrocarbon species present in the atmosphere differ. The HEAT HC data represent hydrocarbon concentrations excluding methane (NMHC). The DU HC data represent hydrocarbon concentrations including methane, but the instrumental response to methane is limited (Singer et al., 1998) so the DU HC concentrations are not equivalent to methane plus NMHC. DU HC concentrations had been corrected for the instrument's different response to a HC mix than to the propane calibration standard and the DU data also include HC measurements that have been offset to better represent total HC (Bishop, 2022).

DU and Opus concentrations of CO_2, while reported, are not an independent measurement (Bishop and Haugen, 2017). The reason that these CO_2 concentrations are not an independent quantity is related to the way that the instruments make measurements. DU and Opus measure absorption at wavelengths in the ultraviolet (UV) and infrared (IR) portions of the electromagnetic spectrum where absorption is specific to certain molecules. They convert 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 CO_2 concentrations are constrained by the calculation.

The DU data set included measurements of CO, NO, NO_2, and HC reported as volume percent and as grams of pollutant per kilogram of fuel. The DU 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).

Opus applies the same approach. The Opus data included ratios (species/CO_2) and mixing ratios (ppm) but not the calculated grams per kilogram fuel. Values of grams of pollutant per kg fuel can be computed using the equations provided in the Opus (pp 13 to 14) and DU (equations 1 to 3, pp 10 to 11) project reports.

The HEAT data included molar ratios (unidentified, but apparently ratios to CO_2) and species mixing ratios (ppmv). HEAT obtains tailpipe-equivalent concentrations through regression of pollutant masses against CO_2 mass comparable to the DU method.

Comparisons among systems were made using mixing ratios, since other units of measurement (e.g., g per kg fuel) were not directly reported by Opus and HEAT. I converted measurements to comparable concentration units (volume percent and ppmv, which are both used in this report depending on scale). Opus and HEAT molar ratios (species to CO_2) can be converted to grams per kilogram fuel using FEAT equations (Bishop and Haugen, 2017).

The concentrations reported in both all 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 represent meaningful data. The DU 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.

Data flags were included in the DU and Opus data sets. Values flagged invalid were handled as discussed in Section 3.

2.3 Determination of Measurement Uncertainty

Measurement uncertainty (σ_0), as reported in previous studies and computed here, is best understood as an indicator of instrument noise (Burgard et al., 2006). Briefly, negative concentration values are fit to a double exponential (LaPlace) distribution, which is used to generate an estimate of variance (or standard deviation) around a zero concentration (for details, see Appendix C). The instrumental uncertainty associated with a concentration average for a sample of size N is $\sigma_0/N^{½}$.

The concentration that is significantly greater than zero at a 95% confidence level is $\mathbf{C_0^{0.95} = 1.6282 \times} \sigma_0$, obtained by integrating the LaPlace distribution from $-\infty$ to 0.95. $C_0^{0.95}$ is one possible estimator of detection limits. A more restrictive estimate of detection limits is three standard deviations of a blank ($\sim C_0^{0.99}$) when blank measurements are made.

I also computed two standard deviations of the lowest calibration concentration released from the test vehicles as a second estimate of measurement uncertainty (Appendix C). No values were reported for calibration blanks, so it was not possible to estimate detection limits as three standard deviations of the blanks.

An additional corroboration of the calculated measurement uncertainties, σ_0, was obtained by analyzing the replicability of measurements made on vehicles passing the test site more than once.

2.4 Test Vehicle Data

Four types of test vehicle data were compiled by Revecorp in their data set: (1) calibration data, (2) simulated evaporative emissions, (3) actual evaporative emissions, and (4) portable emissions monitoring system (PEMS) data. Calibration tests were conducted by releasing calibration gases from an electric vehicle (EV) at approximately 9 a.m., 12:30 p.m., and 4 p.m. These calibrations (audits) were in addition to the usual system calibrations conducted by RSD operators. For each RSD system and each pollutant, 45 calibration tests were recorded (~15 each for a blank, a lower concentration, and a higher concentration). No data were reported for the blanks for any of the three RSDs, however. Some measurements were missing, yielding 11 – 15 measurements of each pollutant at each of two non-zero concentrations for each RSD system.

The results of simulated evaporative emissions were analyzed using regressions of RSD HC values versus simulated mass emission rates (g mi^{-1}), which were calculated from butane volume flow rates (L min^{-1}), temperature, and vehicle speed. In the test vehicle data set, the HEAT data included mass emission rates (g mi^{-1}), which were compared to the simulated mass emission rates. However, the Opus and DU measurements were reported as indices, whose conversion to g mi^{-1} would require additional information and data processing that were not part of the study; therefore, correlations between RSD indices and simulated mass emission rates were checked.

Tests of actual evaporative emissions were conducted by removing fuel caps, disconnecting vapor lines, or disconnecting purge valves. Unlike the calibrations and simulated emissions, the mass or volume emission rates of the actual emissions are unknown. Therefore, comparisons were made between RSD systems.

A PEMS was operated in one vehicle (Mazda 6), yielding 25 measurement records that were matched with data from the RSD systems. Since some measurements were unmatched, the number of comparison points between PEMS and RSD was less than 25 for each system (N = 22, 11, 19, and 24, respectively, for DU, HEAT, Opus1, and Opus2). The test vehicle data file reported PEMS measurements in both g kg^{-1} fuel and g mi^{-1}. HEAT data were reported in g kg^{-1} fuel whereas the DU and Opus measurements were reported as concentrations (% or ppm). These differences in reported measurement units, coupled with small sample size, make comparisons of RSD to PEMS data inconclusive.

Results for test vehicles are discussed in Section 3.5.

2.5 Statistical Methods

Comparisons between measurement systems were made using both graphical approaches and formal statistical tests. The most powerful statistical test of differences is a paired test, provided a logical basis for pairing exists. 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). Since the measurements made by the three measurement systems on the same vehicle at the same time are a logical pairing, paired t-tests were constructed from the exhaust concentrations (e.g., CO) as measured by each RSD system on each specific vehicle, then summed (or averaged) over all vehicles within the category. Statistical power of the tests is reported.

Graphical comparisons included scatter plots, type comparisons, box plots, and cumulative distribution plots. Type comparisons (e.g., by model year, MOVES category) are usually shown as bar charts with standard errors of the means or other measures of uncertainty. Box plots and cumulative distribution plots were used to visually compare distributions of measurements.

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

3. Results

3.1 Speed and Acceleration

Mean speed, acceleration, and vehicle specific power (VSP; Jimenz, 1999) are summarized in Table 3 for the individual data sets and in Table 4 for the merged same-population data. Since the Phoenix vehicle populations differed among the three individual data sets, differences in mean speed, acceleration, and VSP are best considered after restricting data to three-way matches (Table 4). On average, vehicle speed increased as vehicles passed Opus, then HEAT, and then DU, consistent with expectation for a freeway entrance ramp. Average vehicle speed, acceleration, and VSP all increased between the HEAT and DU positions. Some vehicles may have reduced their rates of acceleration between Opus and HEAT, though, as also suggested by field reports. Comparison of Table 4 with Table 3 indicates that the restricted data set in Table 4 is reasonably consistent with the individual full data sets.

DU, HEAT, and Opus use different formulae for calculating VSP, which affects the comparisons. Formulaic differences were removed by recalculating VSP using the same formula for each RSD system (Table 5). Recalculation reduced the largest mean VSP difference from 4.0 to 3.1 kW tonne^{-1} (Table 5). Mean DU VSP was higher than either mean HEAT or Opus VSP because average acceleration and speed were greater for DU than for the other two systems (Table 4) (calculated VSP was more sensitive to acceleration than speed).

Recalculation of DU VSP using the DU formula reproduced VSP as reported by DU, whereas recalculation of HEAT VSP using the HEAT formula did not exactly reproduce HEAT VSP. The latter result might be due to differing values used here for road upslope and road direction compared to those used by HEAT (Table 5). The difference between VSP values reported by HEAT and those recalculated here is not large (average 0.7 kW tonne^{-1}, slope = 1, r^2= 0.999).

If the Opus VSP flag is ignored, 7506 vehicle had valid flags for speed and acceleration. The Opus VSP validity flag eliminated the highest (~15%) and lowest (~5%) VSP values, yielded a different set of vehicle counts for MOVES bins compared with DU and HEAT, and a different VSP distribution compared to DU and HEAT (see next section). Based on results presented in the Opus project report, the Opus VSP validity flags only indicate if VSP values fell within the range defined by the federal test procedure; flags did not describe the validity of the VSP measurements. Using the larger set to recalculate VSP changes the averages (Table 5). For the DU formula applied to 7506 vehicles, the VSP averages are 13.9 ± 0.07 for DU, 10.8 ± 0.06 for HEAT, and 12.6 ± 0.13 for Opus. For the HEAT formula applied to 7506 vehicles, the VSP averages are 12.8 ± 0.07 for DU, 10.2 ± 0.06 for HEAT, and 11.7 ± 0.13 for Opus. Opus averages show the largest difference between the data sets of 5725 and 7506 vehicles, due to the exclusion of the lowest and highest VSP values in the 5725-vehicle data set. For the larger data set (7506 vehicles), average vehicle load was greatest at the DU position and least at HEAT.

Table 3. Statistical summary of unmatched DU, HEAT, and Opus data: speed, acceleration, and vehicle specific power ± 1 standard error of the mean excluding values flagged as invalid.

Species	DU (valid flags, N N = 15,743)	HEAT (AZ records matched by State, N = 30,510)	Opus (AZ plates, excludes test vehicle, $N_{speed} = N_{accel} =$ 29,344, N_{VSP} = 21,784)
Speed (mph)	23.2 ± 0.04	24.9 ± 0.04[a]	21.8 ± 0.04
Speed (m s^{-1})	--	11.1 ± 0.02	--
Acceleration (mph s^{-1})	0.92 ± 0.01	0.37 ± 0.007[a]	1.01 ± 0.02
Acceleration (m s^{-2})	--	0.17 ± 0.003	--
VSP (kW tonne^{-1})[b]	13.6 ± 0.05	9.6 ± 0.04	10.8 ± 0.03

a. Converted from reported units (m s^{-1} and m s^{-2})
b. kilowatts (kW) per metric ton (tonne), 1 tonne = 10^3 kg = 1 Mg

Table 4. Statistical summary of matched DU, HEAT, and Opus data: speed, acceleration, and VSP ± 1 standard error of the mean excluding values flagged as invalid.

Species	DU	HEAT	Opus
Speed (mph)[a]	22.4 ± 0.06	21.2 ± 0.07[c]	20.4 ± 0.07
Speed (mph)[b]	23.3 ± 0.06	22.3 ± 0.06[c]	21.7 ± 0.07
Acceleration (mph s^{-1})[a]	1.05 ± 0.02	0.73 ± 0.01[c]	0.71 ± 0.01
Acceleration (mph s^{-1})[b]	0.96 ± 0.01	0.57 ± 0.01[c]	0.77 ± 0.02
VSP (kW tonne^{-1})[a,d]	13.8 ± 0.08	9.8 ± 0.06	10.5 ± 0.06
VSP (kW tonne^{-1})[b,d]	13.9 ± 0.07	9.5 ± 0.06	12.1 ± 0.13

a. N = 5725 (excluding Opus VSP values flagged as invalid)
b. N = 7506 (excluding DU or Opus invalid speed or acceleration)
c. Converted from reported units (m s^{-1} and m s^{-2})
d. kilowatts (kW) per metric ton (tonne), 1 tonne = 10^3 kg = 1 Mg, as reported by RSDs

Table 5. Statistical summary of matched DU, HEAT, and Opus data for VSP ± 1 standard error of the mean excluding values flagged as invalid. Units are kW $tonne^{-1}$

Metric	DU	HEAT	Opus
VSP reported[a]	13.8 ± 0.08	9.8 ± 0.06	10.5 ± 0.06
VSP reported[b]	13.9 ± 0.07	9.5 ± 0.06	12.1 ± 0.13
VSP recalculated using DU formula[a,c]	13.8 ± 0.08	11.1 ± 0.06	11.0 ± 0.06
VSP recalculated using DU formula[b,c]	13.9 ± 0.07	10.8 ± 0.06	12.6 ± 0.13
VSP recalculated using HEAT formula[a,d]	12.8 ± 0.08	10.5 ± 0.06	10.1 ± 0.06
VSP recalculated using HEAT formula[b,d]	12.8 ± 0.07	10.2 ± 0.06	11.7 ± 0.13

a. N = 5725 (excluding Opus VSP values flagged as invalid)
b. N = 7506 (excluding DU or Opus invalid speed or acceleration)
c. Based on road upslope = 3.8°, as reported by DU.
d. Based on road upslope = 3.8°, as reported by DU, instantaneous wind speed and direction as reported by HEAT, and calculation of headwind from wind speed and direction relative to assumed road direction = 30° east of north.

Statistical significance is evaluated using paired t-tests of differences, which are more powerful than tests of means. The test evaluates the probability that the average of paired differences is significantly different from zero. Results for speed, acceleration, and VSP are summarized in Table 6. All differences are highly significant ($p < 0.0001$).

Although the paired differences in Table 6 are statistically significant, the tested RSD systems are not necessarily incomparable. Even though the RSD systems were located as close to each other as feasible, real speed and acceleration differences could have occurred due to increasing vehicle speeds along the entrance ramp. As noted, the DU system was located closest to the freeway entrance, so vehicles likely accelerated during the short intervals of time that they passed Opus then HEAT then DU. Average speed increased by 0.60 mph between Opus and HEAT and by 1.05 mph between HEAT and DU (Table 6), which is consistent with the average accelerations listed in Table 4 and an approximate time interval on the order of 1 s between RSD systems. Some of the differences are not large relative to the averages. The speed differences range from 4 to 9% of the average speeds. The acceleration and VSP differences for the DU system compared to either of the other two systems were relatively larger: 19 to 38% higher for DU acceleration and 9 to 22% higher for DU VSP.

Table 6. Statistical summary of paired DU, HEAT, and Opus differences: speed, acceleration, and VSP ± 1 standard error of the mean excluding values with speed or acceleration flagged as invalid (valid N = 7506 vehicle passes). ** = highly significant difference ($p < 0.0001$).

Species	HEAT – DU	Opus – HEAT	Opus – DU
Speed (mph)	-1.05 ± 0.02**	-0.60 ± 0.01**	-1.65 ± 0.02**
Acceleration (mph s^{-1})	-0.383 ± 0.013**	0.193 ± 0.023**	-0.19 ± 0.024**
VSP1 (kW $tonne^{-1}$)[a]	-3.02 ± 0.07**	1.75 ± 0.14**	-1.27 ± 0.14**
VSP2 (kW $tonne^{-1}$)[b]	-2.67 ± 0.07**	1.49 ± 0.14**	-1.19 ± 0.14**

a. Calculated using DU formula
b. Calculated using HEAT formula

The power of a statistical test quantifies the probability of detecting a statistically significant difference of any specified magnitude. For a sample of size N = 7506, a statistical test has a high probability (e.g., > 95%) of finding a statistically significant result for a difference that may be considered small in practice or of minor physical importance. For a large sample size (N = 7506) 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 99% when the mean difference between paired measurements is only 5% of the standard deviation of the differences (https://www.statskingdom.com/32test_power_t_z.html). Since the standard deviation of the paired Opus-DU speed differences was 1.769 mph ($7506^{½}$ × standard error, Table 6), for example, the probability of obtaining a statistically significant result at $p < 0.05$ exceeded 99% for differences as small as 0.09 mph (0.05 ×1.769 mph). Therefore, relatively small speed differences are statistically significant. Similarly, the probability of obtaining a statistically significant result at $p < 0.05$ exceeded 99% for an Opus-DU acceleration difference of only 0.105 mph s^{-1}. Since statistical power increases as the square root of the sample size, the detectability of differences within disaggregated data sets (e.g., by day of the study) is smaller than for the full matched data.

3.2 Exhaust Concentrations and Paired Measurement Comparisons

Mean exhaust concentrations are summarized in Table 7 for the individual data sets and in Table 8 for matched data. Since the vehicle populations differed among the individual data sets, the Table 7 means are not strictly comparable and differences in mean concentrations change after restricting data to three-way matches (Table 8). For the population averages listed in Table 7, statistically significant differences (means having nonoverlapping confidence intervals defined by ± 2 standard errors) are evident for CO (DU higher), HC (all pairwise differences), NO (all pairwise differences), and NO_2 (Opus higher). The same pattern appears for the matched data in

Table 8. Opus has updated the UV spectrometer technology for measuring NO_2 since the study was completed, so the NO_2 comparisons are subject to change in the future.

The DU data set included data flags indicating the validity of the individual species measurements. The number of invalid species measurements was small (4 to 18, depending on the parameter) compared to the total vehicle count (N = 18,294). Species averages were determined based on both all valid measurements (N = 18,261 to 18,294, depending on the species), as shown in Table 7, and on vehicles for which all species values were valid (N = 17,944); the latter averages were nearly identical to the values listed in Table 7 (e.g., CO = 885.0 ppmv for 17,944 vehicles, 9 ppmv lower than listed in Table 7).

The Opus data used a single data flag for gases, which indicated that all merged measurements were valid.

Statistical significance of concentration comparisons is formally evaluated using paired t-tests of differences, which are more powerful than tests of means, as previously described. The test evaluates the probability that the average of paired differences is significantly different from zero. Results are summarized in Table 9 for both three-way and two-way matches. Results were not significantly different (i.e., were within two SE of the means) for the smaller number of three-way matches compared with larger numbers of two-way matches with one exception (Opus – HEAT CO). The larger samples of two-way matches tended to show larger differences between paired RSD measurements than the smaller three-way matched data did. Therefore, findings of statistically significant differences based on the three-way matches are robust.

The three-way sample size (N = 8553 – 8631) is sufficiently large that a paired t-test will yield a statistically significant result at $p < 0.05$ (95% confidence level) with a probability of 99.6% when the mean difference between paired measurements is only 5% of the standard deviation (https://www.statskingdom.com/32test_power_t_z.html). Since the standard deviations of the paired CO differences were 1095 ppmv (Opus-HEAT) to 1797 ppmv (Opus-DU), the probabilities of obtaining statistically significant results at $p < 0.05$ exceeded 99% for CO differences as small as 55 – 90 ppmv (e.g., 0.05 ×1095 ppm). Therefore, relatively small CO differences (~10% of mean CO concentrations, Table 9) are statistically significant. Similarly, the probabilities of obtaining statistically significant results at $p < 0.05$ exceeded 99% for HC differences as small as 7 to 11 ppm, NO differences of 6 to 11 ppm, and NO_2 differences of 1 to 3 ppm. Since statistical power increases as the square root of the sample size, the detectability of differences in disaggregated data sets is smaller than for the full matched data.

Table 7. Statistical summary of unmatched DU, HEAT, and Opus data: mean exhaust concentrations ± 1 standard error of the mean. Units are % for CO_2 and ppmv for other species. NR = not reported.

Species	DU (valid flags)	HEAT (AZ records matched by State)	Opus (AZ plates only, excludes test vehicle)
CO	894.0 ± 25.0	614.2 ± 10.8	552.9 ± 15.2
CO_2	15.0 ± 0.002[a]	NR	15.0 ± 0.001[a]
HC	72.1 ± 1.6[b]	17.1 ± 0.6	35.3 ± 0.8[c]
NO	97.6 ± 3.0	63.2 ± 1.4	44.6 ± 1.1
NO_2	2.0 ± 0.1	2.1 ± 0.1	6.5 ± 0.4
NH_3	38.7 ± 0.7	NR	NR

a. CO_2 is not an independent measurement
b. Offset HC = 52.1 ± 1.6 ppm
c. Opus propane HC. Mean Opus hexane HC = 18.0 ± 0.5 ppm.

Table 8. Statistical summary of matched DU, HEAT, and Opus data: mean exhaust concentrations ± 1 standard error of the mean. Units are % for CO_2 and ppmv for other species. NR = not reported.

Species	N Matched Vehicle Passes, Valid Flags	DU	HEAT	Opus
CO	8631	732.7 ± 23.8	562.9 ± 19.7	529.1 ± 17.5
CO_2	8631	15.0 ± 0.002[a]	NR	15.0 ± 0.001[a]
HC	8629	67.7 ± 2.3[b]	13.0 ± 1.1	19.2 ± 1.3[c]
NO	8630	71.9 ± 3.7	51.8 ± 2.8	40.2 ± 2.2
NO_2	8553	1.7 ± 0.2	1.7 ± 0.2	8.2 ± 0.7
NH_3	8625	36.9 ± 1.0	NR	NR

a. CO_2 is not an independent measurement
b. Offset HC = 47.7 ± 2.3 ppm
c. Opus propane HC. Mean Opus hexane HC = 9.8 ± 0.7 ppm.

Table 9. Statistical summary of paired DU, HEAT, and Opus differences: CO, HC, NO, and NO_2 ± 1 standard error of the mean excluding values flagged as invalid. Summaries are shown for three-way (lower counts) and two-way (higher counts) matched data. Units are ppmv. *** = highly significant difference ($p < 0.0001$), ** = significant ($p<0.01$), * = significant ($p< 0.05$).

Species	N Matched Vehicle Passes, Valid Flags[c]	HEAT - DU	Opus - HEAT	Opus - DU
CO	8631	-169.8 ± 15.3***	-33.8 ± 11.8**	-203.6 ± 19.3***
HC[a]	8629	-54.7 ± 2.5***	6.2 ± 1.5***	-48.5 ± 1.8***
NO	8630	-20.1 ± 1.8***	-11.5 ± 1.2***	-31.6 ± 2.3***
NO_2	8553	0.05 ± 0.2	6.4 ± 0.7***	6.5 ± 0.7***
CO	13686/21899/10813	-187.9 ± 12.3***	-108.9 ± 7.8***	-196.3 ± 21.3***
HC[b]	13680/21899/10809	-56.5 ± 1.9***	14.1 ± 0.9***	-45.9 ± 1.8***
NO	13681/21899/10812	-25.2 ± 1.7***	-11.8 ± 0.9***	-36.7 ± 2.2***
NO_2	13525/21899/10712	0.1 ± 0.2	4.3 ± 0.4***	7.0 ± 0.6***

a. For DU offset HC, HEAT-DU = -34.7 ± 2.5*** ppmv; Opus-DU = -28.5 ± 1.8*** ppm
b. For DU offset HC, HEAT-DU = -36.5 ± 1.9*** ppmv; Opus-DU = -26.0 ± 1.8*** ppm
c. Counts for three-way matched data and each set of two-way matches

The RSD concentration differences varied among species (Table 9). Opus exhibited higher average NO_2 and lower average NO than either DU or HEAT.

Differences were not uniform across concentrations (Figure 1). For CO and HC, DU concentrations were evidently higher between about the 20th – 30th and the 80th – 90th percentiles (Figure 1), contributing to higher averages for DU compared with HEAT and Opus (Tables 7 and 8). NO distributions were very similar among RSD systems, despite the concentration differences being statistically significant (Table 9). For NO_2, DU and HEAT distributions visibly differed, but average differences were not statistically significant. The Opus NO_2 distribution visibly differed from DU and HEAT (Figure 1).

Concentration differences were unrelated to differences in vehicle specific power.

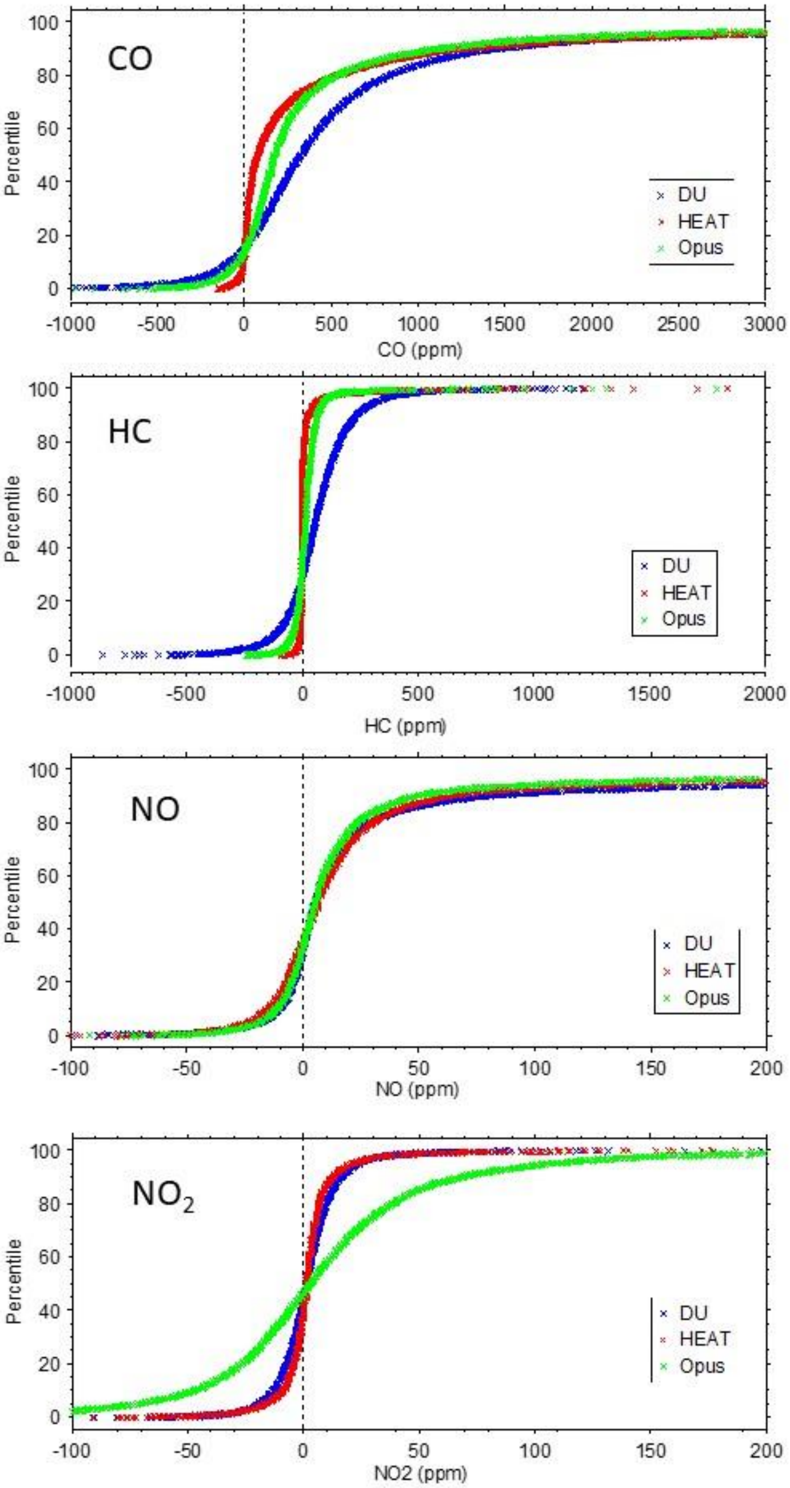

Figure 1. Percentile versus CO, HC, NO, and NO_2 concentrations, by RSD system. The data are restricted to samples flagged "valid." Full data ranges are not shown to improve readability by excluding low and high outliers.

3.3 Comparisons Disaggregated by Vehicle Type and Operating Condition

All RSD systems captured a range of vehicle types (Appendix B) and MOVES operating modes (Table 10). When restricted to vehicles with Opus VSP flags of “valid”, as shown in Table 10, the data set is reduced to 5725 vehicle passes (Table 10). The Opus VSP validity flag eliminates the highest (~15%) and lowest (~5%) of Opus VSP values (Figure 2) and yields a different set of Opus vehicle counts for MOVES bins compared with DU and HEAT (Table 10). Since the Opus VSP validity flag pertains to the federal test procedure, rather than to the validity of the measurements, subsequent analyses are not restricted to the 5725 measurements with valid Opus VSP data flags. The VSP values used for assigning MOVES modes to Opus data were recalculated from valid Opus speed and acceleration measurements using the DU VSP formula. Since VSP was recalculated using the DU formula and since the Opus VSP flag excluded the highest (~15%) and lowest (~5%) VSP values, assignments to all MOVES modes were then possible for Opus records with valid speed and acceleration data.

Table 10. Counts of vehicles within MOVES bins in matched DU, HEAT, and Opus data. Assignment to MOVES bins was determined by speed and acceleration as reported by each RSD and VSP calculated from speed and acceleration using the same formula for each system. Only measurements with valid speed, acceleration, and VSP flags are tabulated.

MOVES Bin	DU	HEAT	Opus
11	15	11	0
12	29	38	0
13	132	153	581
14	421	1119	1248
15	915	1456	1250
16	2728	1687	1462
21	26	71	0
22	24	53	0
23	75	88	110
24	171	217	204
25	297	342	281
27	558	370	402
28	231	92	187
29	68	19	0
30	35	8	0
40	0	1	0
Totals	5725	5725	5725

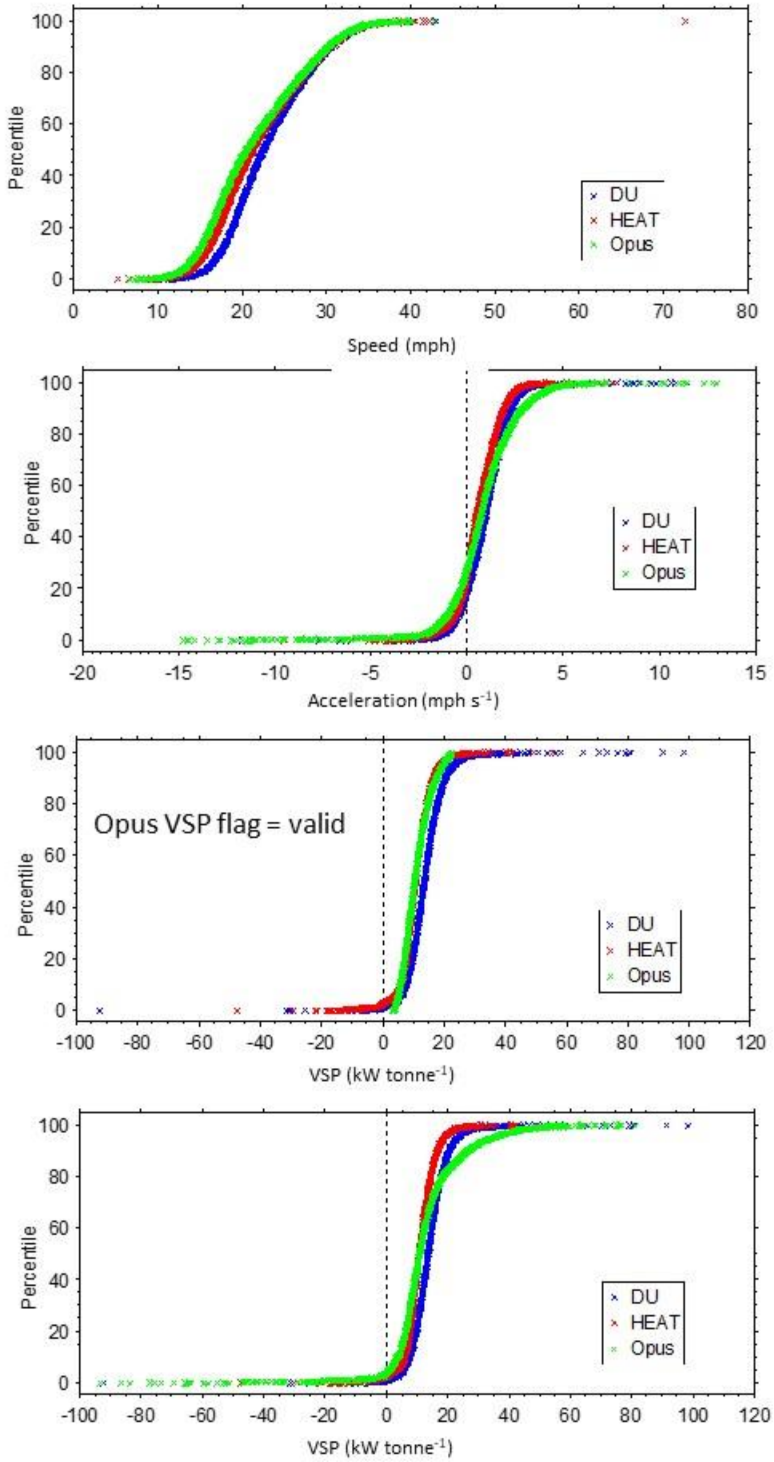

Figure 2. Percentile versus vehicle speed, acceleration, and VSP, by RSD system. The data are restricted to samples flagged "valid" for speed and acceleration except, as noted, an additional restriction in the third panel limits data to Opus samples flagged "valid" for VSP.

Average CO concentrations showed little variability across MOVES operating modes (Figure 3). The DU average CO concentration was higher in mode 11 (VSP < 0), but few (22) vehicles fell in this mode and measurement variability was high.

Average HC concentrations also showed little variability across MOVES operating modes, though DU values tended to be higher in the lowest modes (Figure 4).

The matched data set consists of 140 diesel and 8491 gasoline-fueled vehicles (Appendix B, Table B-1). As previously noted, Opus exhibited higher average NO_2 and lower average NO than either DU or HEAT. These differences were more extreme for diesel than gasoline vehicles (Appendix B, Table B-2).

Of 8631 vehicle records, 5942 (69%) were listed as either 4-door sedans or 4-door station wagons (presumably SUVs, which were not otherwise listed) (Appendix B). The average RSD differences for these two predominant categories therefore largely determined the overall RSD differences (Figure B-1). For NO, RSD differences were more extreme for passenger vans and 4-door pickups than for 4-door sedans or 4-door station wagons (SUVs).

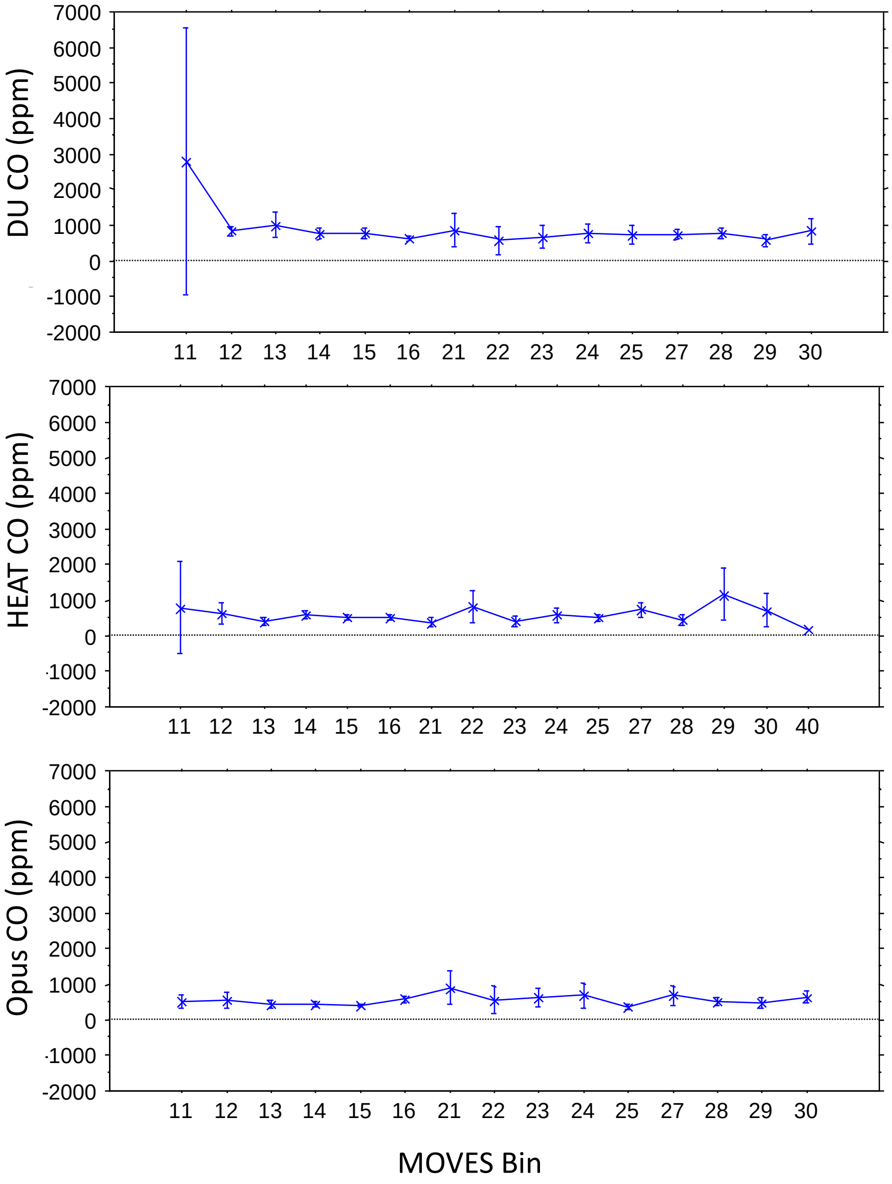

Figure 3. Average CO concentrations disaggregated by MOVES operating mode. The assignments of any vehicle to MOVES modes varied among RSDs depending on vehicle speed and acceleration recorded by each RSD. Data flagged invalid for CO, speed, or acceleration were excluded. Error bars indicate 95% confidence intervals.

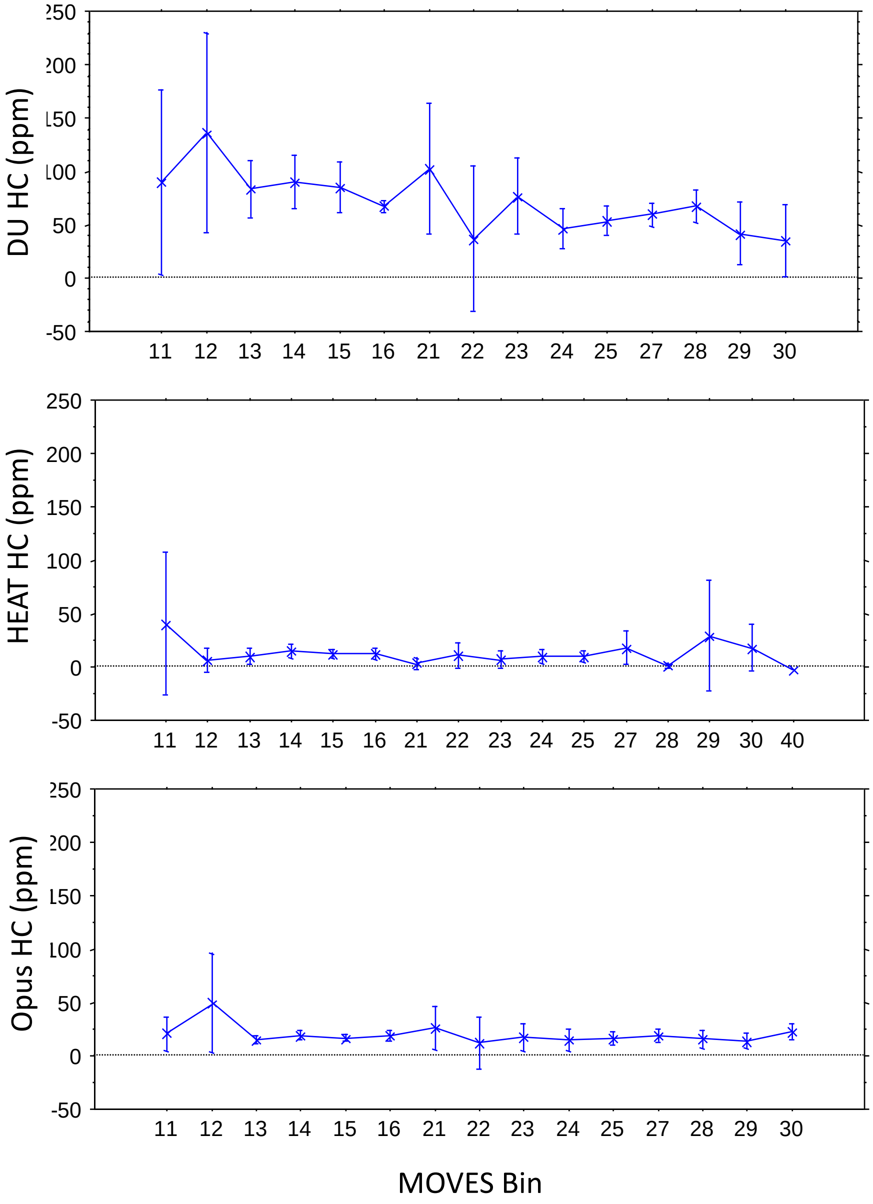

Figure 4. Average HC concentrations disaggregated by MOVES operating mode. Panels show DU HC (not offset) and Opus propane HC. The assignments of any vehicle to MOVES modes varied among RSDs depending on vehicle speed and acceleration recorded by each RSD. Data flagged invalid for HC, speed, or acceleration were excluded. Error bars indicate 95% confidence intervals.

3.4 Measurement Uncertainty

Measurement uncertainty (σ_0), as reported in previous studies and computed here, is best understood as an indicator of instrument noise (Burgard et al., 2006). Briefly, negative concentration values are fit to a double exponential (LaPlace) distribution, which is used to generate an estimate of variance (or standard deviation) around a zero concentration (for details, see Appendix C). The instrumental uncertainty associated with a concentration average for a sample of size N is $\sigma_0/N^{½}$. The concentration that is significantly greater than zero at a 95% confidence level is $\mathbf{C_0^{0.95} = 1.6282} \times \sigma_0$, obtained by integrating the LaPlace distribution from $-\infty$ to 0.95. $C_0^{0.95}$ is one possible estimator of detection limits.

Table 11 lists values of σ_0 and $C_0^{0.95}$ obtained from Phoenix data. As noted in Appendix A, the values obtained for σ_0 depend on specific details of the method. Many of the detection limits ($C_0^{0.95}$) listed in Table 11 exceed the mean concentrations listed in Table 8. In such cases, a reasonable interpretation is that the tabulated means are indeed less than species detection limits but their quantification is less reliable than the standard errors would suggest. The values of $C_0^{0.95}$ listed in Table 11 are lower than I obtained for the 2016 Chicago study (with one exception). For example, the Chicago 2016 DU and HEAT $C_0^{0.95}$ values for CO were 1991 and 127 ppm, respectively (Opus was not a participant in the Chicago RSD comparison).

A more restrictive estimate of detection limits is often used, which is three standard deviations of a blank. As previously noted, no data were available for blanks. Table C-1 compares variabilities of low-concentration standards to the calculated σ_0 values (see also Section 3.5, Table 12, and Section 4, Figure 17).

Table 11. Measurement uncertainty and detection limits. Units are ppmv. NR = not reported.

Species	**DU (AZ matched)**		**HEAT (AZ matched)**		**Opus (AZ)**	
	σ_0	$C_0^{0.95}$	σ_0	$C_0^{0.95}$	σ_0	$C_0^{0.95}$
CO	832	1354	50	82	321	523
HC	184	299	20	32	54	88
NO	70	114	25	41	38	62
NO_2	18	29	24	39	58	94
NH_3	14	23	NR	NR	NR	NR

3.5 Test Vehicle Results

The test vehicle data were compiled and analyzed by Revecorp Inc. The Revecorp analyses are the primary findings for the test vehicles. To facilitate interpretation of the full fleet comparisons, supplementary data analyses are discussed here for: (1) calibration data, (2) simulated evaporative emissions, (3) actual evaporative emissions, and (4) portable emissions monitoring system (PEMS) data.

Calibration data are shown in Figure 5 (N = 45; 10 to 15 measurements at each concentration but no values reported for blanks). Tests were conducted by releasing calibration gases from an electric vehicle (EV) at approximately 9 a.m., 12:30 p.m., and 4 p.m. The DU HC response is relatively flat (slope significantly less than 1). However, the value of the lower calibration point (255.2 ppmv propane) is less than the calculated DU detection limit shown in Table 11, so the DU system was operating below its limits of quantification for HC.

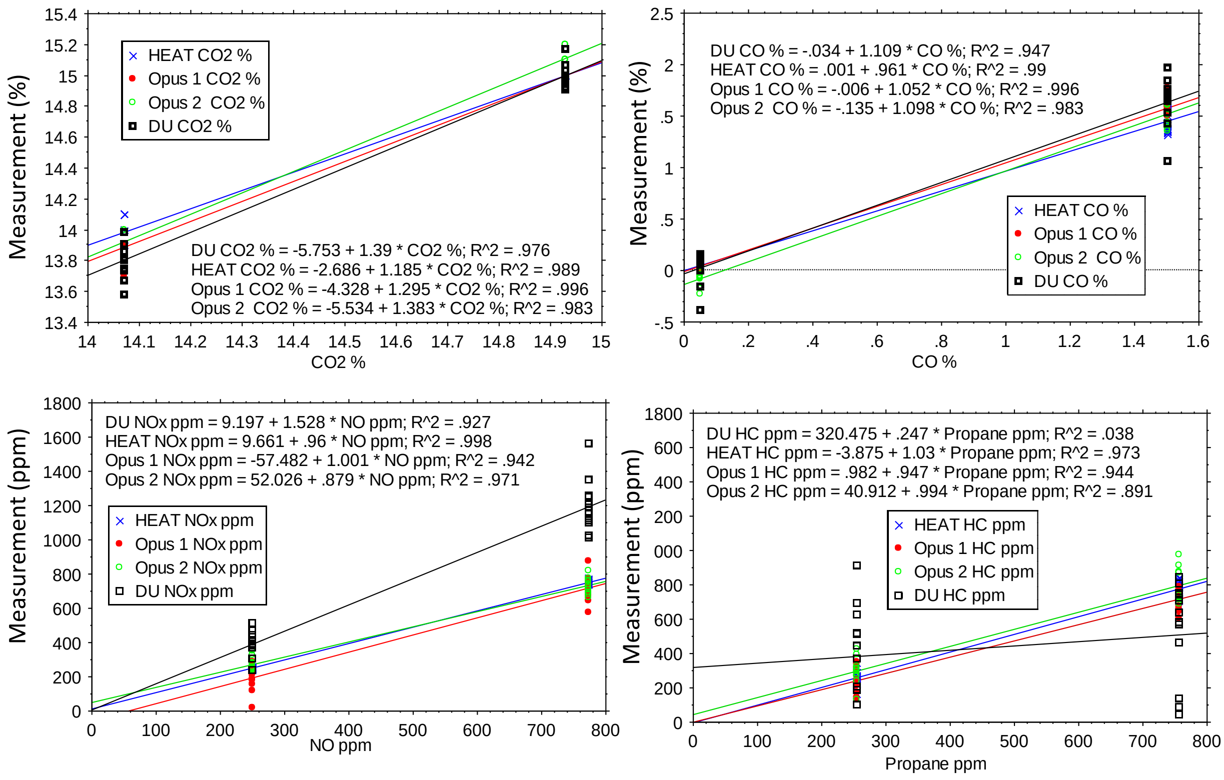

Figure 5. RSD measurements vs concentrations of calibration standards. An EV released calibration gases from a simulated tailpipe 18 inches above the ground at 30 ft^3 min^{-1}. Overlapping symbols may be hidden.

Variability in the DU response to 255.2 ppmv propane was high (standard error = 103 ppm, Table 12), indicating that a mean of 383.5 ppmv is not significantly different from the propane standard concentration of 255.2 ppmv (Table 12). The low DU responses to the 756.4 ppmv HC calibration are unexplained. The Opus and HEAT HC means were within two standard errors of the nominal propane concentrations, except the mean Opus 1 value for the 756.4 ppm standard was 717.4 ± 12.7 ppmv (Table 12).

Table 12. Mean reported HC concentrations ± 1 standard error of the mean for calibration data. Opus 1 denotes source and sensor set at standard height (12 inches); Opus 2 was set at 18 inches. Units are ppmv.

	Mean Reported HC Concentration (ppm)	
System	**255.2 ppmv propane standard**	**756.4 ppmv propane standard**
DU	383.5 ± 102.8	507.1 ± 82.0
HEAT	258.9 ± 11.0	775.1 ± 11.7
Opus 1	242.7 ± 19.4	717.4 ± 12.7
Opus 2	294.5 ± 23.4	792.4 ± 33.1

All RSD systems responded to simulated evaporative emissions (Figure 6) (see next section for discussion of evaporative emission measurements for the full fleet). Figure 6 displays results using regressions of RSD HC values versus simulated mass emission rates (g mi^{-1}), which were calculated by Revecorp from butane volume flow rates (L min^{-1}), temperature, and vehicle speed. The HEAT measurements were reported in units of g mi^{-1} and should be comparable to the test values; it is unclear why they are ~3 to 4 times higher than the calculated emissions rates (subject to change because HEAT personnel are reworking the method for calculating g mi^{-1}).

The Opus and DU measurements were reported as indices, whose conversion to g mi^{-1} would require additional information and data processing that were not part of the study. HC measurements appear to be close to the DU level of quantifiability (but were detectable). Opus ERG Index values for the blanks were positive and statistically significant (Figure 6), which implies that the ERG Index overestimated evaporative emissions (see Revecorp report and the next section of this report). The Opus 1 Score (Envirotest indicator) was recorded as discrete integer values ranging from 0 to 4 (values of zero and 1 were reported as non-detection of evaporative emissions in full-fleet data) and exhibited values of zero for all blanks (Figure 6).

The minimum butane flow rate of simulated emissions was 1 liter per min (L min^{-1}), which is equivalent to 3 to 6 g mi^{-1} butane. This minimum rate is about two orders of magnitude greater than the U.S. EPA running loss standard of 0.05 g mi^{-1} (U.S. EPA, 2016).

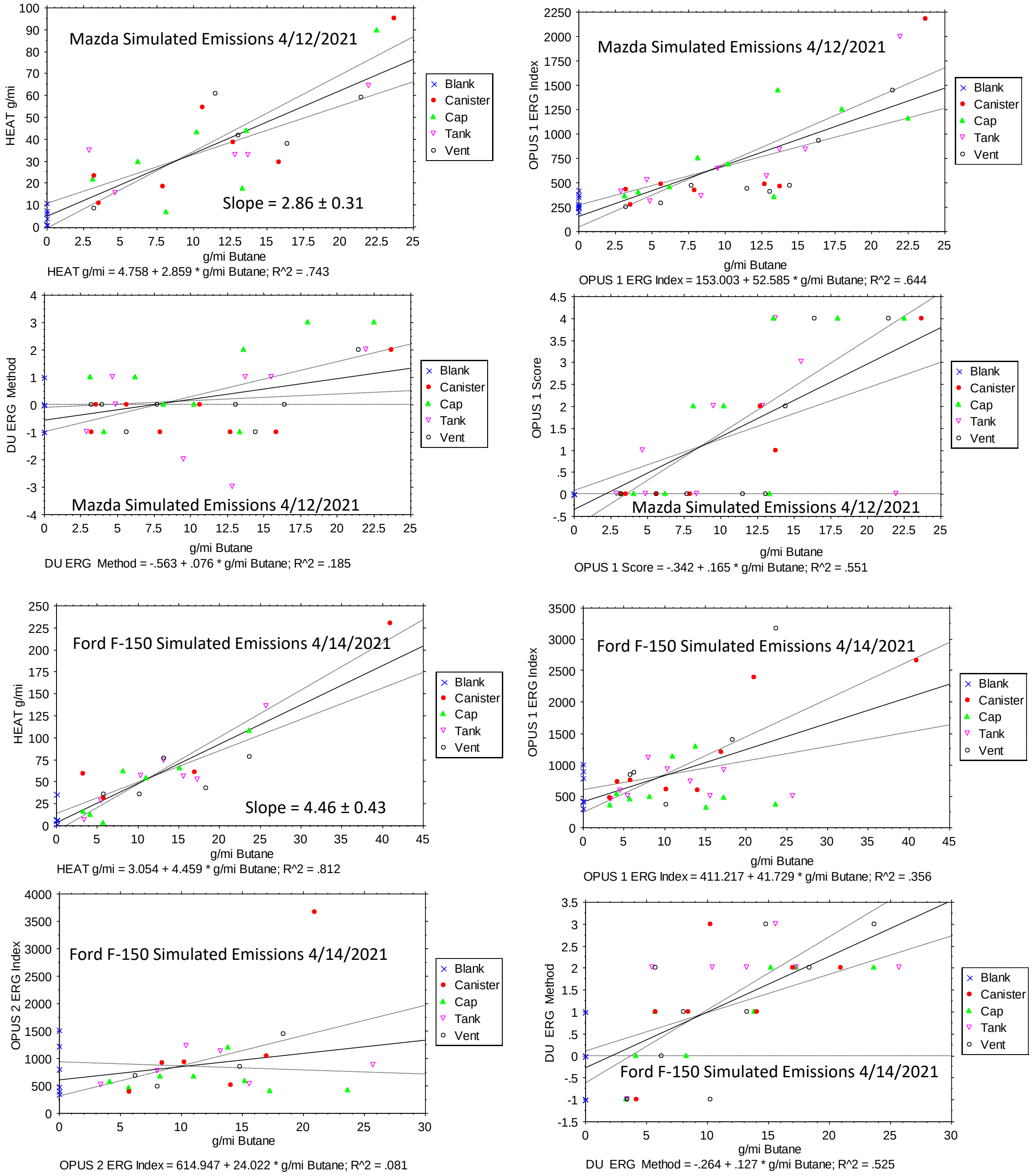

Figure 6. RSD measurements vs simulated evaporative mass emission rates. Simulated mass emission rates were calculated from butane volume flow rates of 0, 1, 2.5, and 4 L min^{-1} using temperature and vehicle speed. The DU and Opus measurements are indices whose conversion to mass flow rates requires additional information that was not part of the study. Opus 1 is a measurement made at the standard height of 12 inches above the road; Opus 2 varied from standard height to 24 inches (typically, 18 inches).

Regression slopes (observed vs. calculated) were not statistically different for different leak locations, as shown in Figure 7 for the HEAT data. Since the DU ERG Index and Opus 1 Score consisted of discrete integer values, too few unique data points were available to justify reporting leak-location regressions; the leak-location slopes were not statistically different, but the error bars were large. Results are also not shown for the Opus ERG Index since it exhibits nonzero blanks (Figure 6). Values for blanks are included in Figure 7 to reduce the uncertainties of the regression slopes; each subpanel uses the same (all) blank measurements (i.e., since all blanks represent zero simulated evaporative emissions, location was not relevant). The regression approach accounts for any residual background values, not otherwise removed by data processing algorithms, as indicated by the small positive intercepts in each panel.

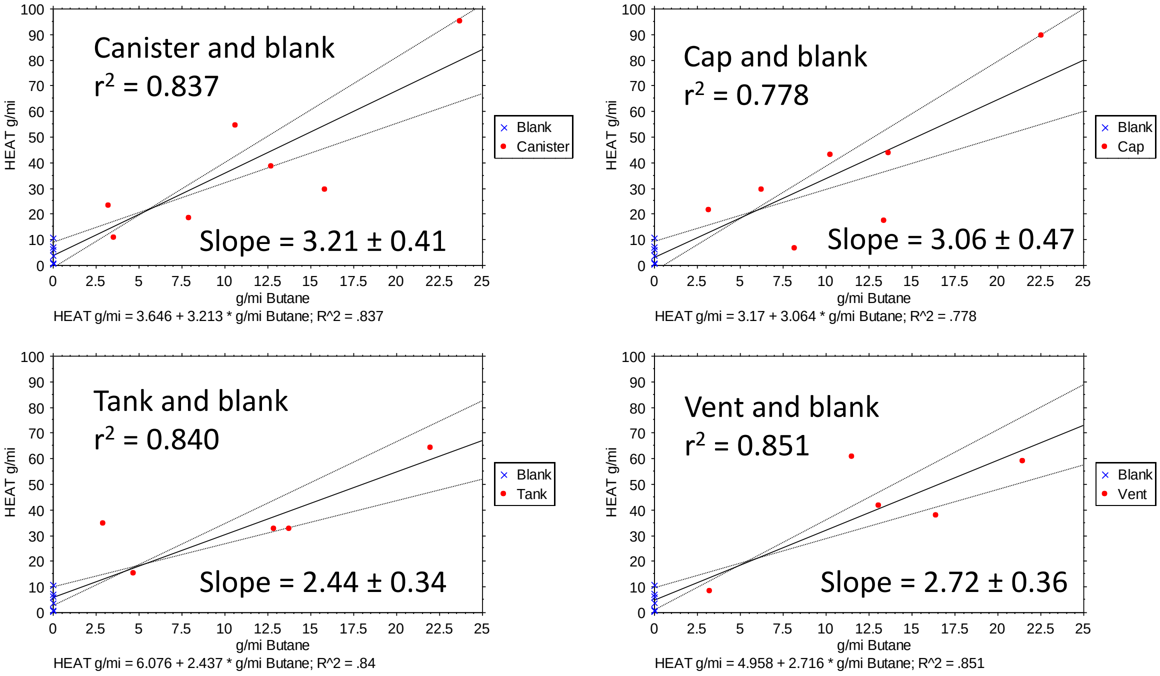

Figure 7. HEAT measurements vs simulated evaporative emissions. Simulated mass emission rates were calculated from butane volume flow rates of 0, 1, 2.5, and 4 L min^{-1} using temperature and vehicle speed.

Test of actual evaporative emissions were conducted by removing fuel caps, disconnecting vapor lines, or disconnecting purge valves. Since the mass or volume emission rates of the actual emissions are unknown, comparisons were made between RSD systems (Figure 8). The regression relationships are not obviously consistent (r^2 values of 0 to 0.25), but it is possible that absolute mass emission rates would be of approximately comparable magnitude if they were calculated for the three systems (DU, Opus 1, and Opus 2) whose data were reported as indices. Similar levels of variability were obtained for comparisons of RSD measurements of the Ford F-150 actual emissions on April 15.

A PEMS that was operated in one vehicle (Mazda 6) yielded 25 measurement records that were matched with data from the RSD systems. Since some measurements were unmatched, the number of comparison points between PEMS and RSD is less than 25 for each system (N = 22, 11, 19, and 24, respectively, for DU, HEAT, Opus 1, and Opus 2). A PEMS that was operated in the Ford F-150 yielded 31 measurement records (N = 28, 20, 26, and 12, respectively, matched to DU, HEAT, Opus 1, and Opus 2).

The test vehicle data file reports PEMS measurements in both g kg^{-1} fuel and g mi^{-1}. HEAT data are reported in g kg^{-1} fuel whereas the DU and Opus measurements are reported as concentrations (% or ppm). The DU and Opus concentrations can be converted to g kg^{-1} fuel using the equations listed in their project reports but conversions were not needed for these statistical comparisons. Converting units does not affect correlations between RSD and PEMS measurements.

For the Mazda, the DU and Opus CO concentrations agree with the PEMS data with respect to differentiating one high value from lower ones (Figure 9). At lower concentrations, the RSD CO measurements are not well correlated with the PEMS emission rate values. For the Mazda, HEAT and PEMS CO emission rates (g kg^{-1} fuel) are highly correlated over the range of CO observed but differ in magnitude (regression slope 0.57 ± 0.05) (Figure 10).

Other species measurements from the Mazda are uncorrelated and differ in magnitude (Figures 9 and 10). For the F-150, no correlations between PEMS and RSD measurements were statistically significant (r^2 = 0.02 to 0.2).

Overall, the test vehicle results for simulated emissions, actual emissions, and PEMS comparisons suggest that the uncertainties associated with lower RSD values of all measured species are high relative to the RSD measurement values, so strong correlations do not appear when the data are confined to low values.

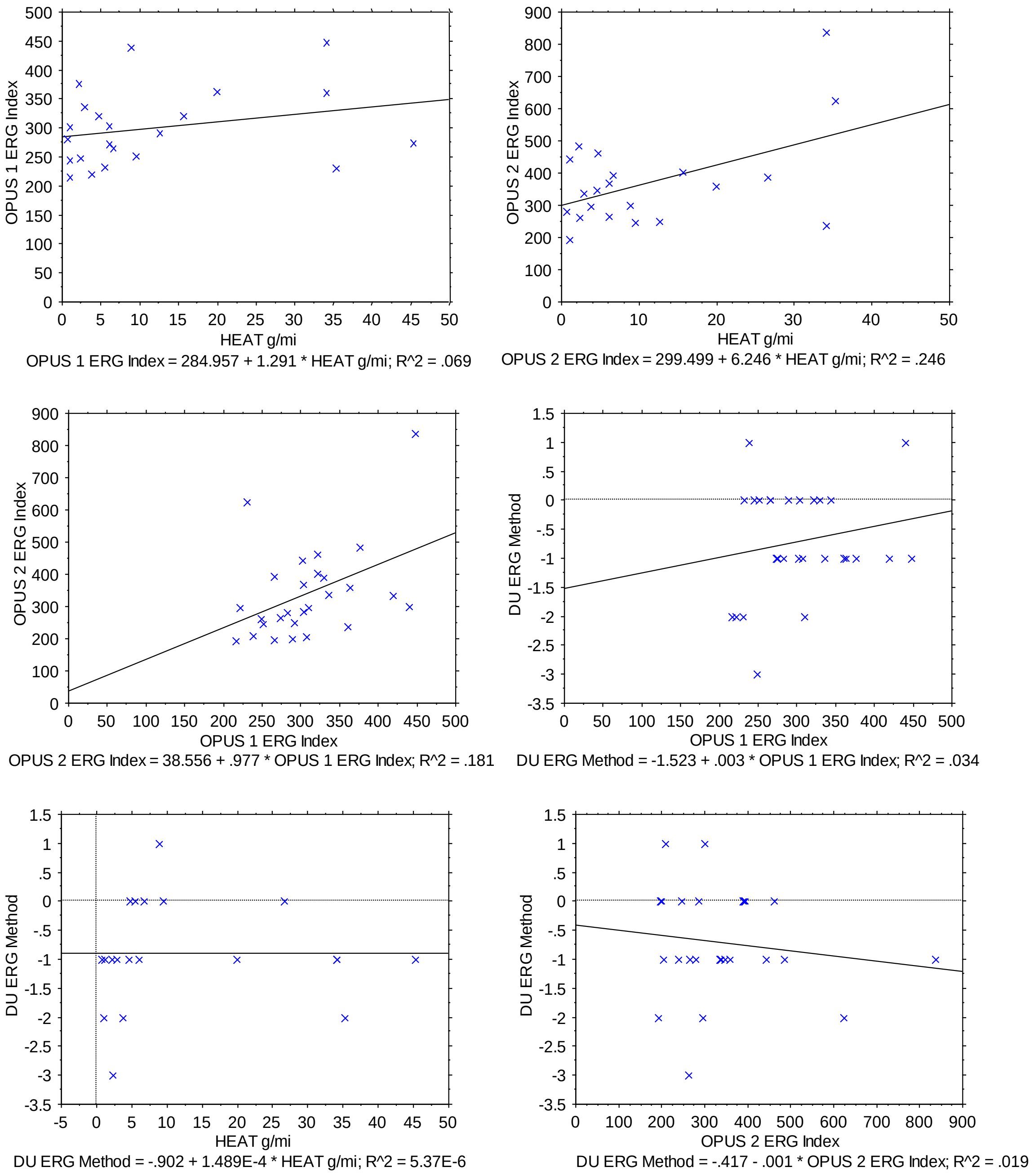

Figure 8. Comparisons of RSD measurements of actual evaporative emissions, which were conducted by removing fuel caps, disconnecting vapor lines, or disconnecting purge valves. Measurements were made on April 13 using the Mazda test vehicle.

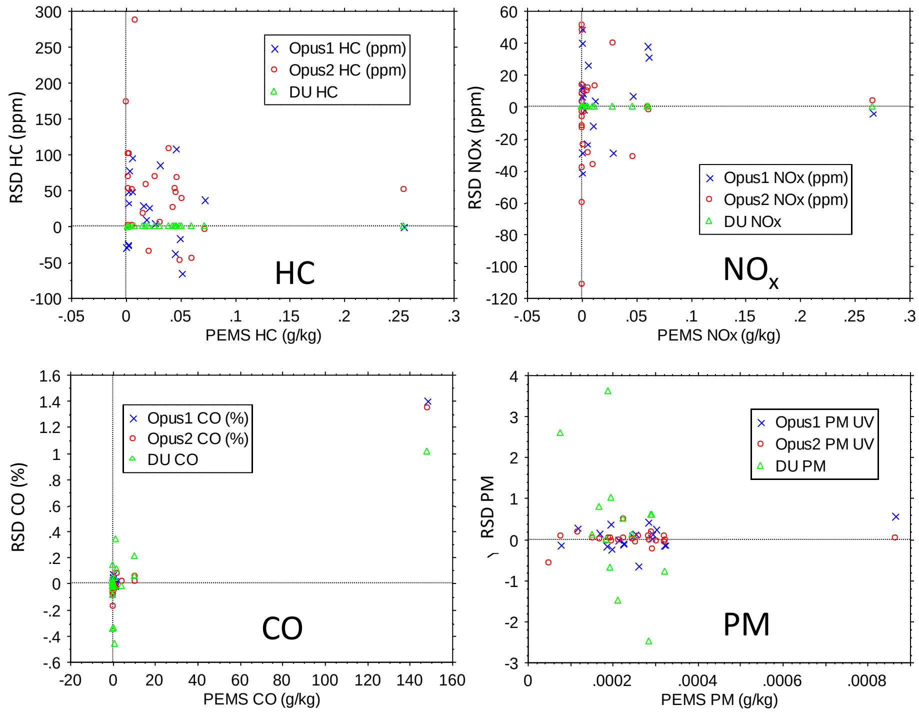

Figure 9. RSD concentrations vs. PEMS emission rates.

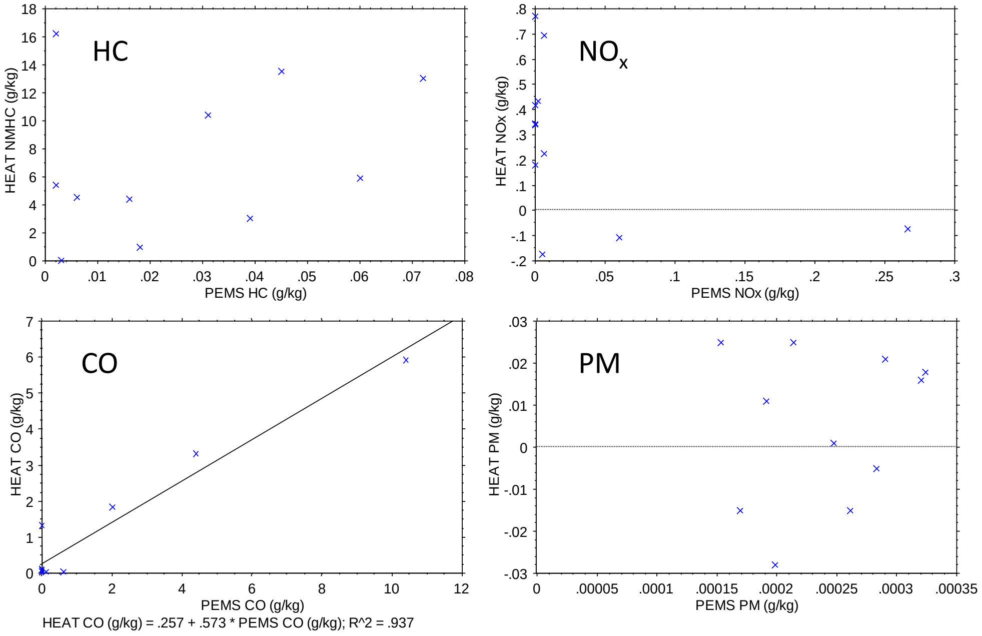

Figure 10. HEAT vs PEMS emission rates (g kg^{-1} fuel).

3.6 Full Vehicle Fleet Evaporative Emission Measurements

DU, Opus, and HEAT report evaporative emissions using different approaches and metrics. DU estimates evaporative emissions for each vehicle by quantifying outliers in the regression of HC optical absorbance (concentration × path length in units of ppm·cm) against CO_2 absorbance (Bishop et al., 2020). The metric reported is the 90th percentile of the residuals of the regressions (Bishop et al., 2020) (e.g., the 5th highest deviation from a straight-line relationship between HC against CO_2 in a set of 50 exhaust plume measurements at a measurement frequency of 0.5 s). This metric is operational, intended to identify a situation in which multiple points indicate a departure from linearity between HC and CO_2. Opus employs a similar approach, though the results are reported differently (see below). HEAT identifies evaporative emissions based on the system's downward-looking plume view, which can indicate the presence of non-tailpipe HC absorbance in a location distinct from tailpipe CO, CO_2, or NO.

The DU general data set (i.e., excluding test vehicles) includes a running loss index (RLI) and a binning variable that ranges from -5 to +5. DU also reports a RLI value associated with zero emissions (RLI_0), which ranged from 31 to 189 as a function of HC concentration and exhibited an average value of 32.6 in the 8631-vehicle merged data set. The binning variable is the difference between RLI and RLI_0 in standard deviations. Bishop et al. (2020) state that binning values of 3 or greater indicate a high probability of evaporative emissions. Only 36 of 8631 vehicles (0.4%) met this criterion. This result agrees with the DU project report, which states "80 measurements from the 18,294 (0.4%) total measurements [were] found to have an RLI_bin value greater than or equal to 3 and are suspected running loss emitters."

In the DU project report, RLI values for the newest model year were used to estimate RLI_{Noise}, which assumed that these vehicles would have negligible running losses. DU reported that $RLI_{Noise} = 31$. The average RLI_0 in the 8631-vehicle merged data set (32.6) agreed well with DU's reported RLI_{Noise}. RLI_0 averaged 32.5 for model year 2021 (count = 327) and 32.9 for model year 2022 (count = 3).

The Opus data set includes an index variable and a binning variable determined according to an Eastern Research Group (ERG) algorithm. The data are flagged to indicate if evaporative emissions were detected. Most bin values of six or greater are flagged to indicate detections. Two other detection indicators were reported in the Opus data ("overall evap detected" and "Envirotest evap detected") as well as "Envirotest evap score" (which had discrete values of 0, 1, 2, or 4). According to the ERG detection indicator, evaporative emissions were detected for 1610 vehicles (19%); in contrast, the Envirotest indicator flagged only 17 vehicles (0.2%). The ERG detection indicator therefore flagged a much higher percentage of vehicles than did either the Envirotest indicator or the DU RLI. The detections flagged by the Envirotest indicator were associated with vehicles in the ERG bins 8 through 11.

HEAT reported only 3 of valid 43,205 vehicle measurements (0.007%) had detectable evaporative emissions; only 1 of these three was in the 8631-vehicle merged data set (0.01%).

The correlation between the DU RLI and Opus ERG Index was modest ($r^2 = 0.2$).

Detections of evaporative emissions were reported for nearly all model years (Table 13).

Table 13. DU and Opus detections of evaporative emissions by model year.

Model Year	Total Vehicles	Opus ERG N Detected	Opus ERG % Detected	Opus Envirotest N Detected	Opus Envirotest % Detected	DU N Detected	DU % Detected
1955	2	1	50	1	50	0	0
1956	1	1	100	0	0	0	0
1969	1	0	0	0	0	0	0
1973	1	1	100	0	0	0	0
1986	1	0	0	0	0	0	0
1988	4	0	0	0	0	0	0
1990	1	1	100	0	0	0	0
1991	1	1	100	0	0	0	0
1992	5	1	20	0	0	0	0
1993	2	0	0	0	0	0	0
1994	8	1	13	0	0	0	0
1995	9	3	33	1	11	0	0
1996	8	4	50	0	0	0	0
1997	20	5	25	0	0	0	0
1998	31	9	29	1	3	1	3
1999	48	8	17	0	0	0	0
2000	57	16	28	1	2	1	2
2001	81	29	36	0	0	1	1
2002	90	23	26	4	4	4	4
2003	123	41	33	2	2	0	0
2004	172	56	33	0	0	2	1
2005	218	63	29	3	1	1	0
2006	254	66	26	0	0	3	1
2007	346	96	28	1	0	1	0
2008	275	73	27	0	0	2	1
2009	207	56	27	0	0	0	0
2010	263	68	26	0	0	0	0
2011	312	62	20	0	0	1	0
2012	420	67	16	0	0	0	0
2013	554	101	18	1	0	3	1
2014	563	90	16	2	0	0	0
2015	726	116	16	0	0	4	1
2016	732	98	13	0	0	0	0
2017	763	125	16	0	0	4	1
2018	695	105	15	0	0	1	0
2019	724	110	15	0	0	3	0
2020	583	69	12	0	0	3	1
2021	327	44	13	0	0	1	0
2022	3	0	0	0	0	0	0

In summary, for evaporative measurements made on non-test vehicles:

- Rates of detection of evaporative emissions were different among the three RSDs.
- The Opus ERG indicator of evaporative emissions identified a much higher percentage of vehicles (19%) having evaporative emissions than did either the Opus Envirotest indicator (0.2%) or the DU running loss indicator (RLI) (0.4%).
- HEAT reported that only 3 of 43,205 valid vehicle measurements (0.007%) had detectable evaporative emissions; only 1 of these 3 was in the 8631-vehicle merged data set (0.01%) (this vehicle exhibited the highest DU RLI and Opus ERG index).
- Modest correlation exists between the DU RLI and Opus ERG Index ($r^2 = 0.2$).
- Detections of evaporative emissions were reported for nearly all model years by DU and Opus.

3.7 Repeated Measurements

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). Of the 8631 records in the merged data set, 1936 records (22%) were measurements of vehicles passing the site more than once. The fraction of repeated measurements in the merged data set is lower than in individual data sets because the repeats in the merged data occurred only when all three RSDs captured at least one repeated pass of a vehicle.

Standard deviations of repeated vehicle passes, s_{repeat}, are shown in Figures 11, 12, and 13. These values were determined for each model year as the mean square for error in a standard one-way analysis of variance (ANOVA) in which each vehicle having two or more passes (maximum number of passes was five) represented one factor level. Vehicle counts were low (4 to 8) prior to model year 2004 and higher variability was observed by all three RSDs for those years (Figures 11, 12, and 13). Observed s_{repeat} variability was generally similar among RSDs, including higher variability observed for some later model years (e.g., CO in 2013). In such cases, usually one vehicle exhibited one high reading that was captured by all three RSDs. For example, 63 model year 2013 vehicles were sampled more than once; the higher 2013 CO s_{repeat} values are primarily due to one vehicle for which all three RSDs recorded one high CO level (8560 to 20,760 ppm) out of three passes.

The overall CO s_{repeat} values for model years 2005 through 2021 were 990 ppmv (DU), 951 ppmv (HEAT), and 947 ppmv (Opus). The overall HC s_{repeat} values for model years 2005 through 2021 were 180 ppmv (DU), 70 ppmv (HEAT), and 58 ppmv (Opus). The overall NO s_{repeat} values for model years 2005 through 2021 were 106 ppmv (DU), 71 ppmv (HEAT), and 63 ppmv (Opus).

The multiyear CO s_{repeat} values exceed measurement uncertainties (832 ppmv for DU, 50 ppmv for HEAT, and 321 ppmv for Opus, Table 11), which is expected because s_{repeat} encompasses both measurement uncertainty and day-to-day emissions variability. For HC, the multiyear s_{repeat} values equal or exceed measurement uncertainties (184 ppmv for DU, 20 ppmv for HEAT, and 54 ppmv for Opus, Table 11). For NO, the multiyear s_{repeat} values exceed measurement uncertainties (70 ppmv for DU, 25 ppmv for HEAT, and 38 ppmv for Opus, Table 11), as expected.

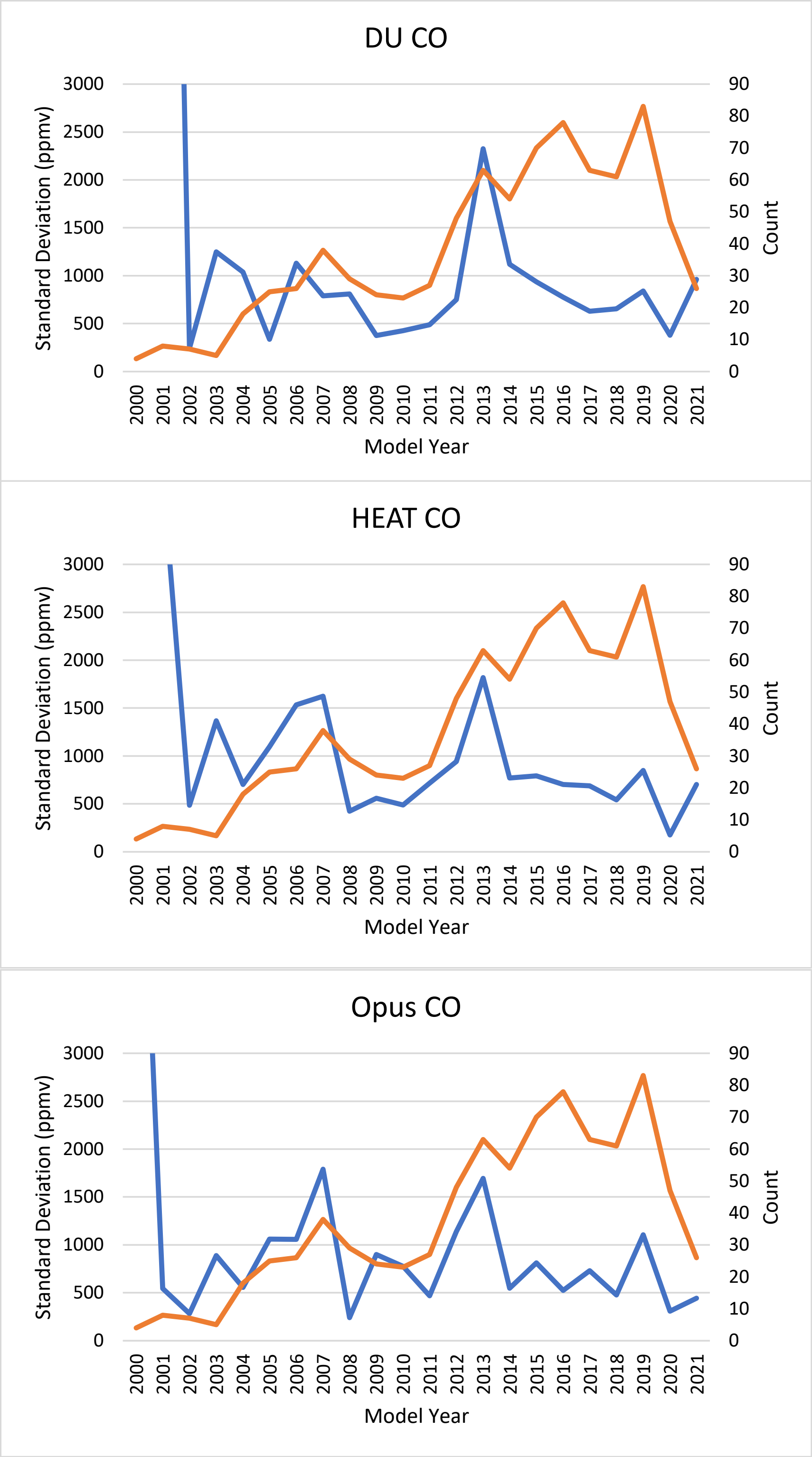

Figure 11. Variability of repeated vehicle CO emissions by model year (blue) and vehicle counts (orange). Values for 2000 and 2001 are offscale (3042 to 13,782 ppmv).

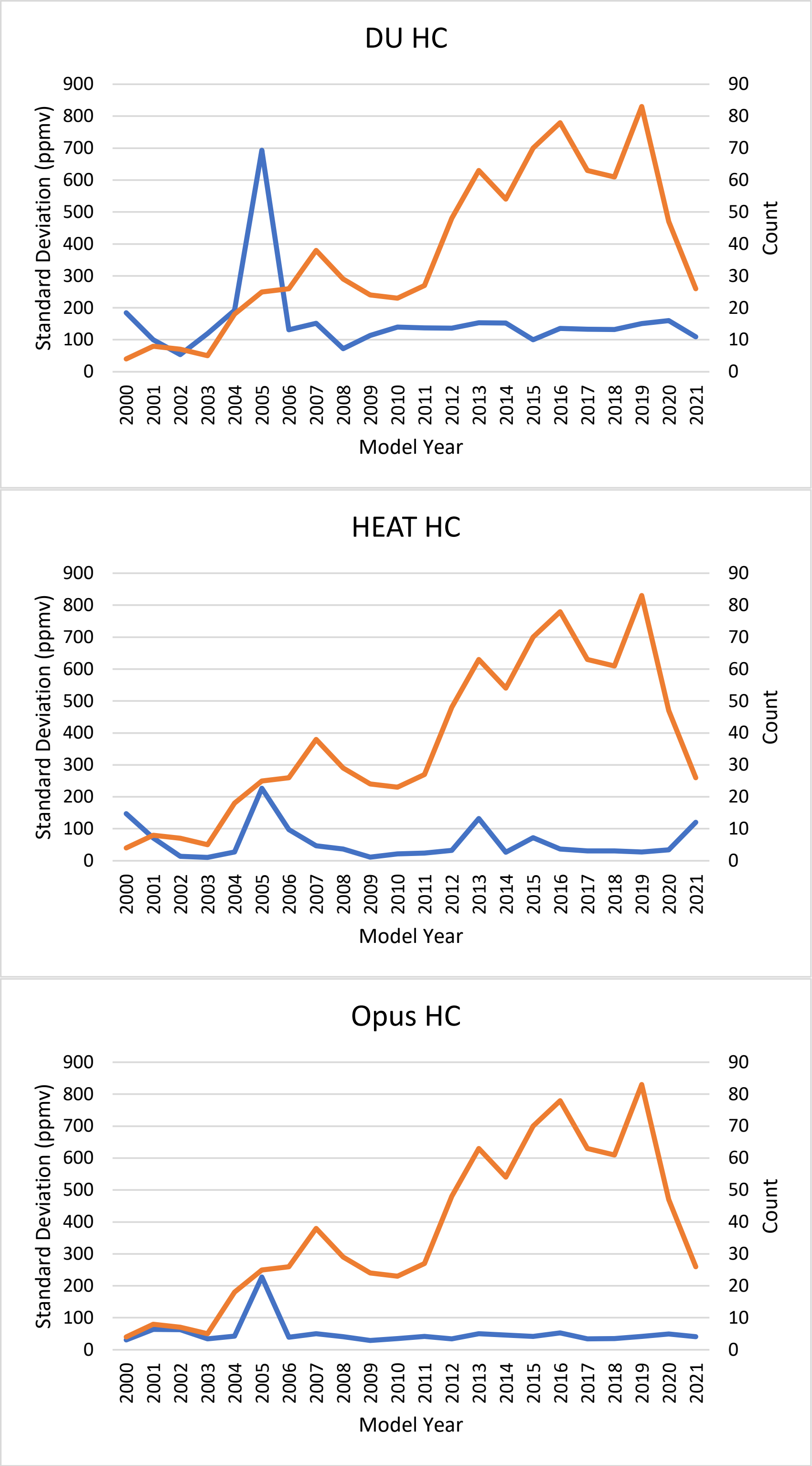

Figure 12. Variability of repeated vehicle HC emissions (blue) and vehicle counts (orange) by model year.

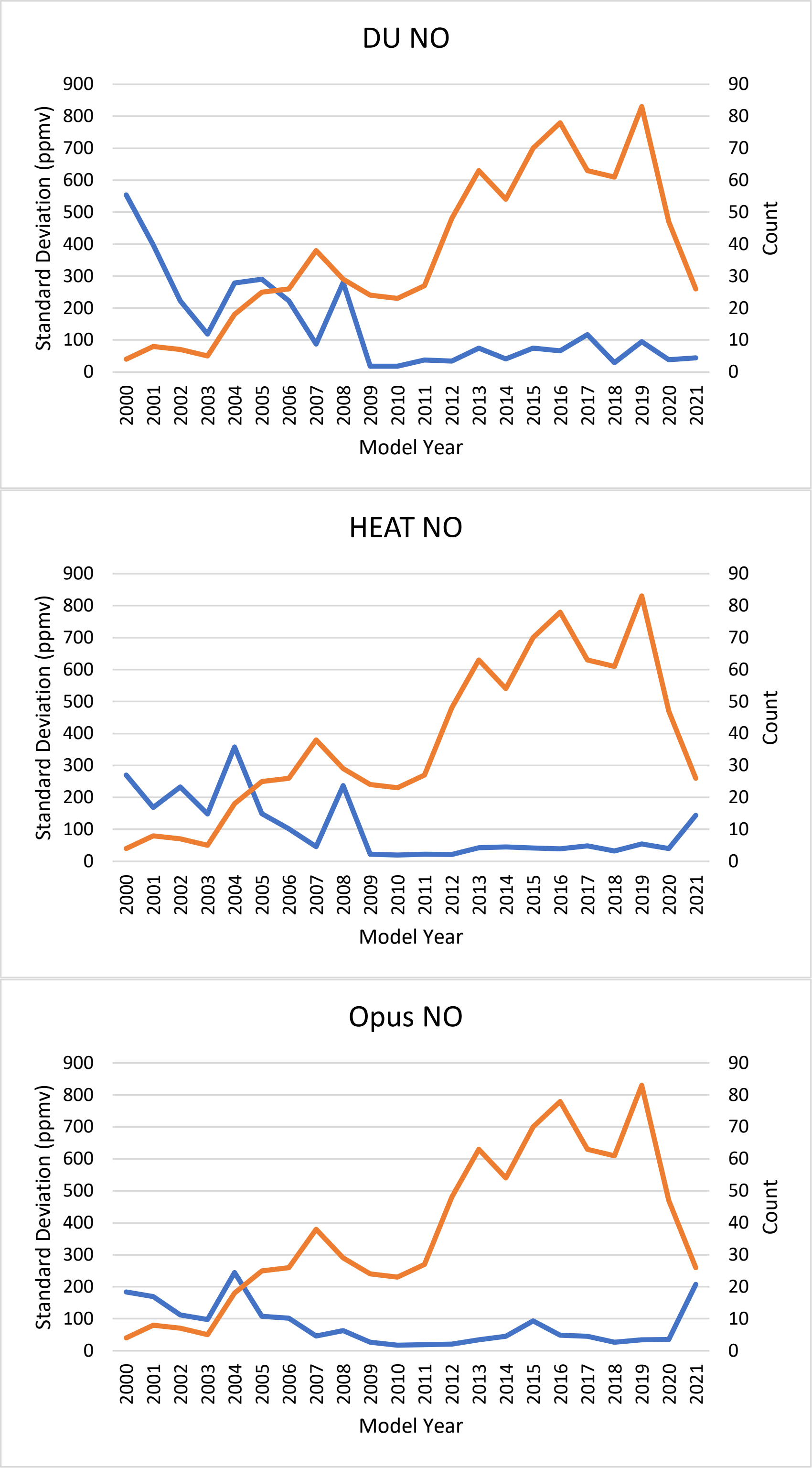

Figure 13. Variability of repeated vehicle NO emissions (blue) and vehicle counts (orange) by model year.

3.8 High Emission Vehicles

As shown in Figure 14, the frequencies of vehicle records exceeding specified high concentrations were similar among the three RSDs. This result is consistent with the frequency distributions depicted in Figure 1, which show similar distributions among RSDs for the highest concentrations. For HC and NO, the frequencies of exceedances of the lower of the two threshold values were higher for the DU system than for HEAT and Opus; this result is also consistent with the distributional results in Figure 1, in which the DU distributions are shifted to higher values relative to the other RSDs from mid-range to high concentrations. The similarities of the frequencies of values exceeding defined threshold concentrations implies that all three RSDs detected high emission vehicle passes at comparable rates. All systems detected high emissions for the vehicles involved but the exact concentrations varied somewhat.

Figure 15 shows the number of concentrations exceeding high values measured by each RSD for vehicles passing the measurement site more than once. The counts are similar among RSDs. The total count for each species is the number of times that any RSD recorded values exceeding threshold concentrations of 5000 ppmv CO or 500 ppmv HC or 1000 ppmv NO. As indicated in Figure 15, vehicles passing the site more than once were more likely to exhibit exhaust concentrations exceeding the defined thresholds only once rather than on two (or more) occasions. The number of vehicles that consistently exhibited concentrations over threshold values was small, ranging from zero to seven. As noted in the previous section, of the 8631 records in the merged data set, 1936 records (22%) were measurements of vehicles passing the site more than once. Of the 1936 records in the repeated measurements subset, 58 exceeded the CO threshold concentration, 24 exceeded the HC threshold, and 34 exceeded the NO threshold.

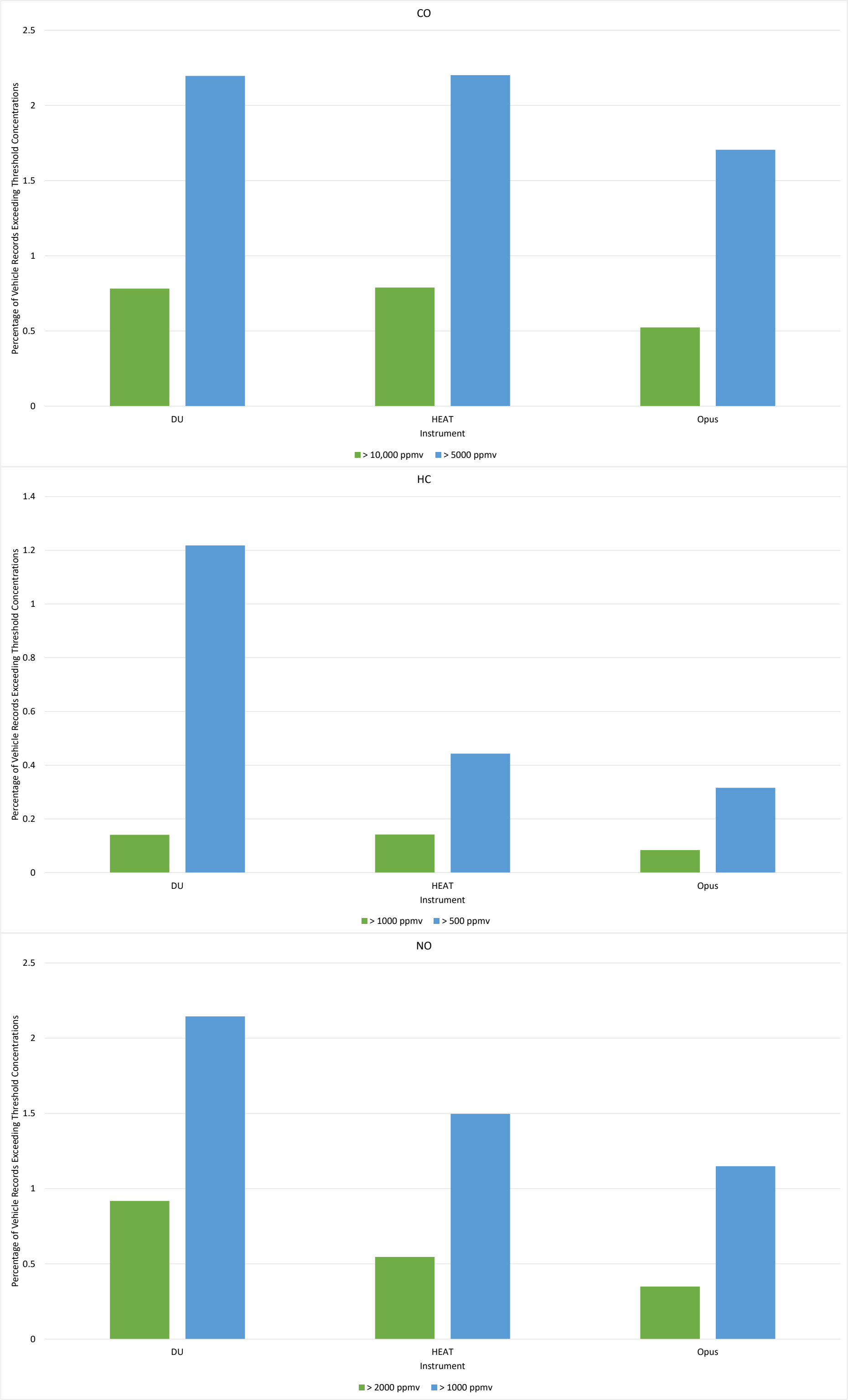

Figure 14. Frequencies of vehicle records exceeding threshold concentrations.

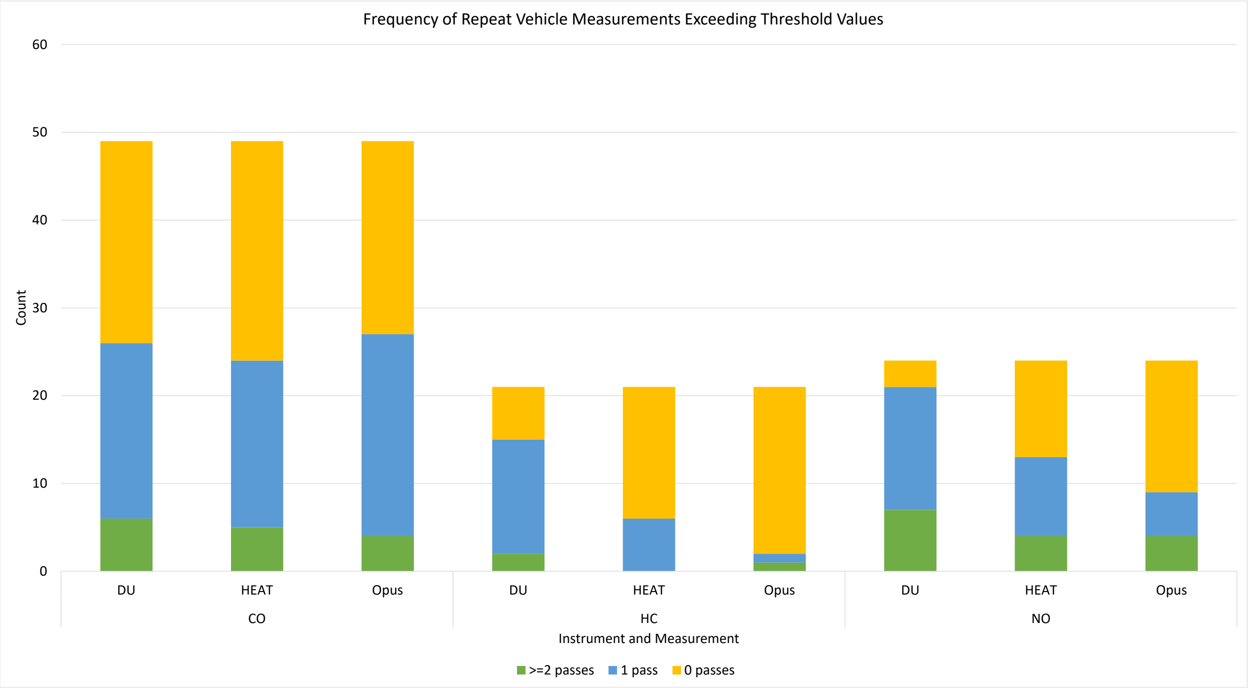

Figure 15. Number of concentrations exceeding high values measured by each RSD for vehicles passing the measurement site more than once. The total count for each species is the number of times that any RSD recorded values exceeding threshold concentrations of 5000 ppmv CO or 500 ppmv HC or 1000 ppmv NO.

4. Discussion

Mean differences between DU measurements and paired HEAT or Opus CO, HC, and NO measurements were statistically significant (Figure 16; see also Table 9). Statistical significance, by definition, means that differences between RSD systems exceeded population variability (indicated by twice the standard errors of the means in Figure 16). Differences between measurements made by different RSD systems are not explained by instrumental uncertainties, which are smaller than the population variabilities of the mean concentrations (Figure 17).

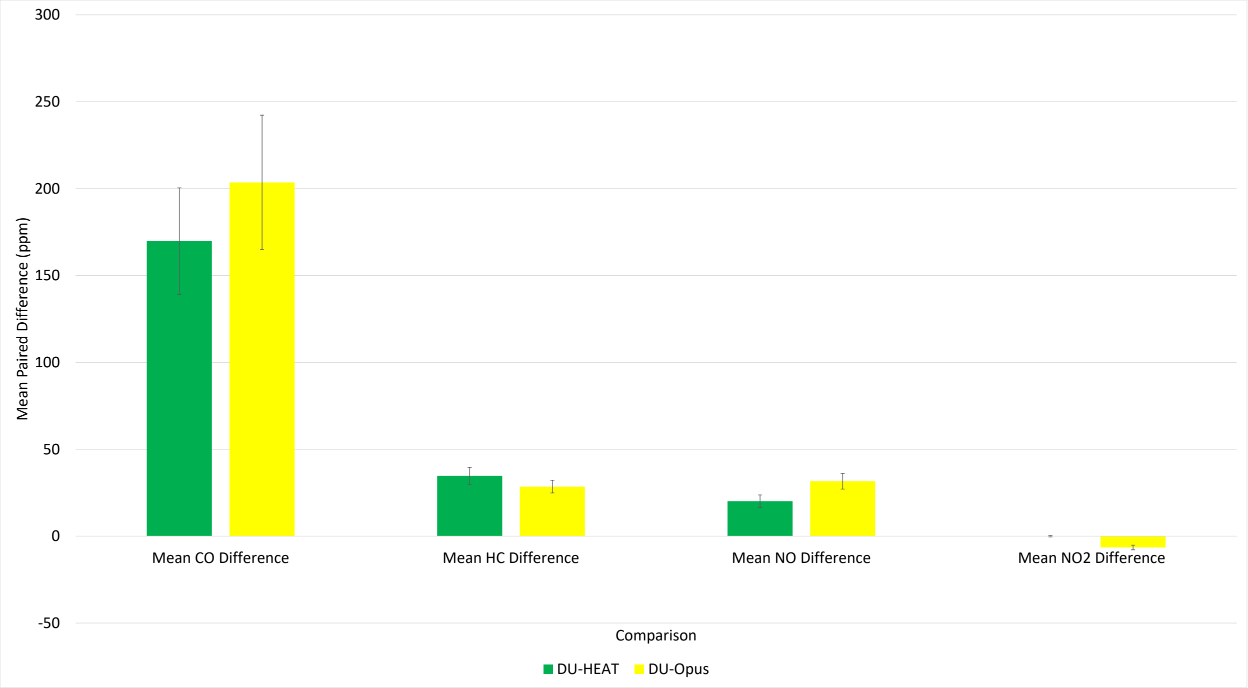

Figure 16. Mean differences in paired concentrations. Error bars are 2 S.E. of the means.

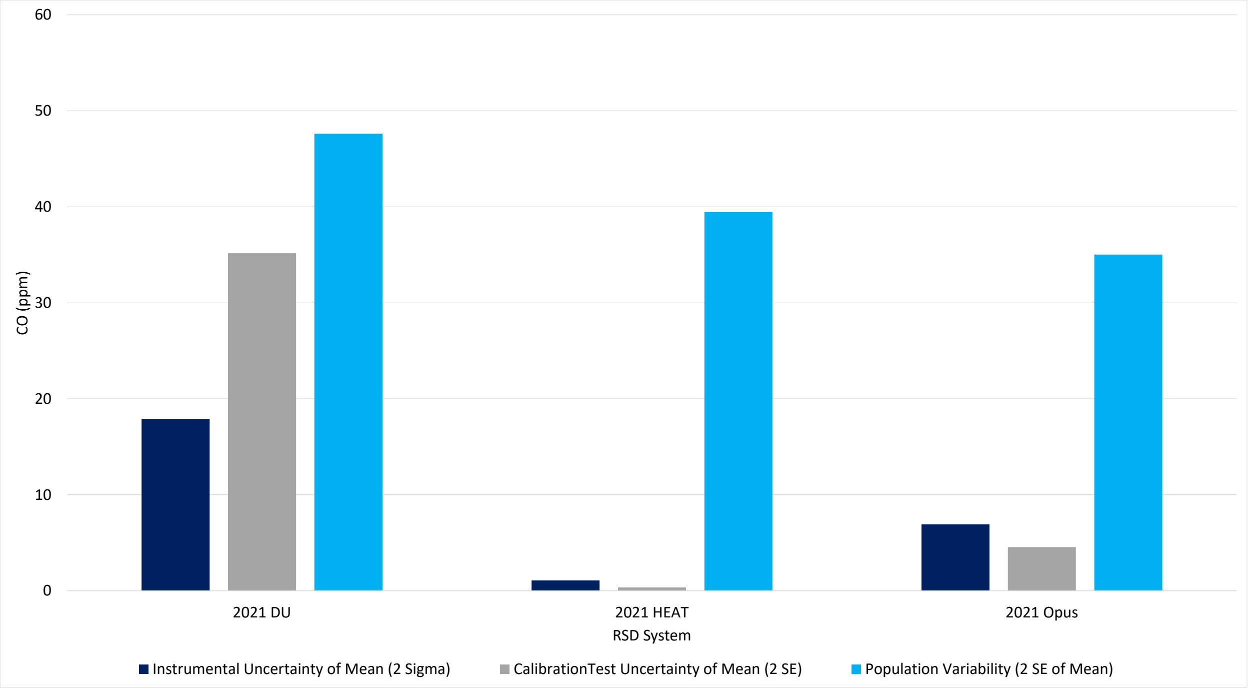

Figure 17. Uncertainties and variabilities of mean CO concentrations. Instrumental uncertainty = 2 × σ_0 / $N^{½}$ (Table 2). Calibration test variability = 2 S.D. of low concentration standard / $N^{½}$ (Figure 3; Figure C-1). Population variability = 2 S.E. of mean concentration.

Mean paired differences in the 2021 Phoenix study were larger than those obtained in the 2016 Chicago study. The 2016 and 2021 comparisons were not directionally consistent for CO and NO (Table 14). The differences between the 2016 and 2021 comparisons imply that no systematic instrumental offsets can be identified.

Both location and season differed in 2021 compared with 2016, so comparisons of mean concentrations across time are subject to uncertainty but are listed in Table 15 for completeness. For HEAT, mean concentrations in Phoenix 2021 were lower than in Chicago 2016. For DU, there was no change in mean CO and NO concentrations in 2021 compared to 2016, while mean HC concentrations either increased (based on reported HC) or remained unchanged (based on reported offset HC). An important caveat noted by the DU principal investigator is that the DU setup, data collection program, and software were different in Phoenix than in DU's previous standard E-23/116/123 studies due to measurement and estimation of evaporative emissions in Phoenix.

Table 14. Statistical summary of paired DU and HEAT vehicle exhaust differences, Chicago 2016 and Phoenix 2021: CO, HC, NO, and NO_2 ± 1 standard error of the mean excluding values flagged as invalid. Units are ppmv. *** = highly significant difference ($p < 0.0001$), ** = significant ($p<0.01$), * = significant ($p< 0.05$).

Species	2016 N Matched Vehicle Passes, Valid Flags	2016 HEAT - DU	2021 N Matched Vehicle Passes, Valid Flags	2021 HEAT - DU
CO	4728	16.5 ± 34.9	8631	-169.8 ± 15.3***
HC[a]	4728	-10.0 ± 3.0**	8629	-54.7 ± 2.5***
NO	4728	6.9 ± 3.3*	8630	-20.1 ± 1.8***
NO_2	0	NR	8553	0.05 ± 0.2

a. For DU offset HC, 2016 HEAT-DU = -3.2 ± 2.8 ppmv and 2021 HEAT-DU = -34.7 ± 2.5*** ppmv

Table 15. Statistical summary of matched DU and HEAT concentrations, Chicago 2016 and Phoenix 2021: mean exhaust concentrations ± 1 standard error of the mean. Units are % for CO_2 and ppmv for other species. NR = not reported.

Species	2016 DU	2016 HEAT	2021 DU	2021 HEAT
CO	717.1 ± 41.1	733.6 ± 30.6	732.7 ± 23.8	562.9 ± 19.7
CO_2	15.0 ± 0.003[a]	NR	15.0 ± 0.002[a]	NR
HC	44.2 ± 2.8[b]	34.2 ± 1.6	67.7 ± 2.3[b]	13.0 ± 1.1
NO	72.2 ± 4.3	79.1 ± 3.7	71.9 ± 3.7	51.8 ± 2.8
NO_2	4.2 ± 0.3	NR	1.7 ± 0.2	1.7 ± 0.2
NH_3	85.3 ± 2.3	NR	36.9 ± 1.0	NR

a. CO_2 is not an independent measurement

b. DU offset HC = 44.2 ± 2.8 ppmv in 2016 and 47.7 ± 2.3 ppmv in 2021

The larger differences between RSD measurements observed in 2021 are not explained by changes in instrumental uncertainties, which were lower in 2021 than in 2016 for both CO and HC (but somewhat higher for NO) (Table 16). For DU, mean CO, HC, and NO concentrations were lower than the calculated detection limits for individual measurements; for HEAT, mean HC concentrations were lower than the calculated detection limits (Tables 15 and 16). For large sample sizes (N = 4728 in 2016 and 8631 in 2021), the instrumental uncertainties of the means are reduced by $N^{1/2}$ ($N^{1/2}$ = 68.8 in 2016 and 93.4 in 2021), yielding expected mean-concentration replicabilities that are smaller than population variability (standard errors of the means, Table 15).

Table 16. Comparison of Chicago and Phoenix measurement uncertainty and detection limits. Units are ppmv. NR = not reported.

Species	**DU (2016)**		**DU (2021)**		**HEAT (2016)**		**HEAT (2021)**	
	σ_0	$C_0^{0.95}$	σ_0	$C_0^{0.95}$	σ_0	$C_0^{0.95}$	σ_0	$C_0^{0.95}$
CO	1223	1991	832	1354	78	127	50	82
HC	214	348	184	299	38	62	20	32
NO	30	49	70	114	13	21	25	41
NO_2	NR	NR	18	29	NR	NR	24	39
NH_3	NR	NR	14	23	NR	NR	NR	NR

For context, the average concentration differences in the 2021 study were smaller than concentration trends observed in previous DU studies between 1998 and 2021 (Figure 18). Each RSD system revealed concentration declines in 2021 compared with previous DU studies but substituting 2021 HEAT or Opus for DU concentrations would overestimate the 2021 improvements relative to past years (i.e., average HEAT and Opus concentrations tended to be lower than DU values in the 2021 study). Generalizing this result is inadvisable because it is inconsistent with the comparison of DU and HEAT measurements made in Chicago in 2016.

Future studies can add variance to mean HEAT or Opus concentrations to account for instrument differences compared with the historical record provided by past DU studies. Doing so will permit trends to be determined from multiple RSD data sources even if consistent adjustment factors cannot be established.

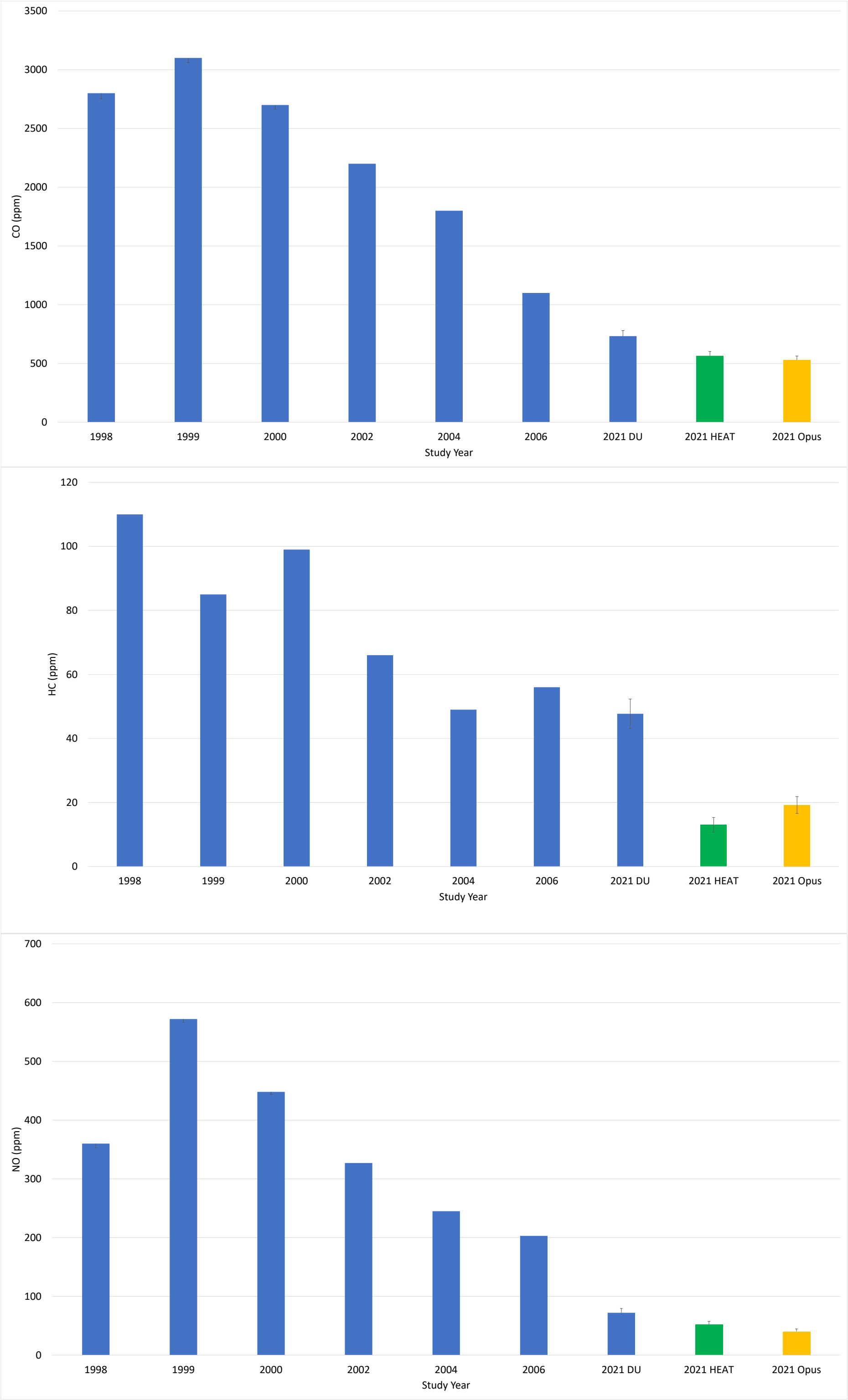

Figure 18. Trends in mean concentrations. Data for 1998 – 2006: Gary A. Bishop, Ryan Stadtmuller and Donald H. Stedman, On-Road Remote Sensing of Automobile Emissions in the Phoenix Area: Year 6, November 2006, CRC Report No. E-23-9, July 2007.

5. Conclusions

Conclusions are summarized as answers to the technical questions posed in the introduction.

How well do the measured emissions compare?

Concentration differences between 2021 DU, HEAT, and Opus paired measurements were statistically significant, larger than population variability, and larger than measurement uncertainty. Concentration differences were unrelated to differences in vehicle specific power.

Mean paired differences in the 2021 Phoenix study were larger than those obtained for a comparison of DU and HEAT measurements made in Chicago in 2016. For CO and NO, the 2021 and 2016 comparisons were not directionally consistent (i.e., DU was not consistently higher in both studies). The differences between the 2016 and 2021 results imply that no systematic RSD instrumental offsets can be identified.

The larger differences between RSD measurements observed in 2021 are not explained by changes in instrumental uncertainties, which were lower in 2021 than in 2016 for both CO and HC (but somewhat higher for NO).

For context, the average concentration differences in the 2021 study were smaller than concentration trends observed in previous DU studies between 1998 and 2021. Each RSD system revealed concentration declines in 2021 compared with previous DU studies but substituting 2021 HEAT or Opus for DU concentrations would overestimate the 2021 improvements relative to past years (i.e., 2021 HEAT and Opus concentrations tended to be lower than DU values).

Future studies can add variance to mean HEAT or Opus concentrations to account for instrument differences compared with the historical record provided by past DU studies. Doing so will permit trends to be determined from multiple RSD data sources even if consistent adjustment factors cannot be established.

How does this comparison vary across vehicle types and emissions levels?

The RSD concentration differences were evident throughout ranges of concentrations and were not uniform across concentrations. For CO and HC, DU concentrations were higher between about the 20th to 30th and the 80th to 90th percentiles, contributing to higher averages for DU compared with HEAT and Opus. The distributions of NO were very similar among RSD systems, despite the concentration differences being statistically significant. For NO_2, DU and HEAT distributions differed, but average differences were not statistically significant. The Opus NO_2 distribution differed from DU and HEAT. Opus exhibited higher NO_2 and lower NO than either DU or HEAT. These differences were more extreme for diesel than gasoline vehicles.

How well can the three measurement systems detect classes of vehicles, such as high emitters?

All systems detected high emissions at similar frequencies (~0.1 to 2 % of vehicle passes, varying with species and concentration threshold), even when the reported concentrations varied among systems. The similarities of the frequencies of values exceeding defined threshold concentrations implies that all three RSDs detected high emissions at comparable rates. For vehicles passing the site more than once (1936 records), the three RSDs recorded similar numbers of exceedances of high concentrations: 1 to 23 vehicles exhibited concentrations over threshold values on one occasion and 0 to 7 vehicles did on two passes.

How do the measurement variabilities of the three systems compare?

Instrument uncertainty (σ_0), as reported in previous studies, was computed by fitting negative concentration values to a double exponential (LaPlace) distribution to generate an estimate of variance (or standard deviation) around a zero concentration. For CO, HC, and NO, the DU system had the highest variabilities while the HEAT RSD had the lowest. CO variabilities were 832 ppmv (DU), 321 ppmv (Opus), and 50 ppmv (HEAT). HC variabilities were 184 ppmv (DU), 54 ppmv (Opus), and 20 ppmv (HEAT). NO variabilities were 70 ppmv (DU), 38 ppmv (Opus), and 25 ppmv (HEAT). NO_2 variabilities were 18 ppmv (DU), 52 ppmv (Opus), and 24 ppmv (HEAT). Because the means were determined from large sample sizes (N = 8631), which reduced the instrumental uncertainties of the means by $N^{1/2}$, mean-concentration replicabilities were smaller than population variability (as represented by the standard errors of the means).

A second estimate of measurement uncertainty was computed as two standard deviations of the lowest concentration released from the test vehicle calibrations. Small (N = 10 to 15 measurements at each concentration; no data reported for blanks) sample sizes limit conclusions. Uncertainties determined from the test vehicle calibrations were the same order of magnitude as instrumental uncertainties determined from the full data set; many were within a factor of two and some agreed within 5%.

Standard deviations of repeated vehicle passes, s_{repeat}, were similar among RSDs and comparable to σ_0 values (approximately equal to, or larger than, σ_0, which is expected because s_{repeat} encompasses both measurement uncertainty and day-to-day emissions variability).

How well can the systems measure at low levels? What are the limits of detection?

Because they were determined from a large sample, mean concentrations less than detection limits are reasonably quantified. For DU, mean CO, HC, and NO concentrations were lower than the detection limits for individual measurements. For Opus, mean HC and NO concentrations were lower than the calculated detection limits. For HEAT, mean HC concentrations were lower than the calculated detection limits. Individual measurements less than detection limits may not be well quantified but are reliably low. The individual sample concentration that is significantly greater than zero at a 95% confidence level is $C_0^{0.95} = 1.6282 \times \sigma_0$, obtained by integrating the LaPlace distribution from $-\infty$ to 0.95 for the instrument variabilities listed for the preceding question.

What is the fraction of valid readings for the systems?

Capture rates (ratio of valid to attempted measurements) could not be determined from any of the Phoenix data sets because they included only valid vehicle measurements. The DU CRC project report E-119-3 states that the data capture rate was very high (94% to 96%, depending on the species). HEAT reported 43,205 valid measurements out of 53,063 attempts, or a capture rate of 81%. Opus reported 33,434 valid measurements but did not characterize their capture rates.

The DU project report indicates that the match rate (percent of submitted plates matched by the State) was 86% and states that this is lower than typical match rates (high 90s%) for unknown reasons. In the HEAT data set, 91% of the vehicles (39,375 out of 43,205) were registered in AZ; 77% of the AZ vehicles (71% of the full HEAT data record) were matched by the State. In the Opus data set, 88% of the vehicles (29,338 out of 33,434) were registered in AZ; the State matched 24,310 Opus measurements (83% of the AZ vehicles, 73% of the full Opus data record).

How well do the systems record vehicle speed and acceleration?

All pairwise RSD differences in speed, acceleration, and vehicle specific power (VSP) were highly significant ($p < 0.0001$). Many of these differences were not large relative to the averages, though. The speed differences ranged from 4 to 9% of the average speeds. The differences for the DU system compared to either of the other two systems were 19 to 38% higher for DU acceleration and 9 to 22% higher for DU VSP. Real speed and acceleration differences need not reflect measurement differences, since they could occur due to increasing vehicle speeds along the entrance ramp (Opus preceded HEAT, which preceded DU, where distances between systems were as small as possible, ~3 to 8 m).

How effectively can the systems cover a range of vehicle specific power bins (as used in EPAs MOVES model)?

All RSD systems captured a range of MOVES operating modes. Average CO and NO concentrations showed little variability across MOVES operating modes. Average HC concentrations also showed little variability across MOVES operating modes, though DU values tended to be higher in the lowest modes.

The matched data set consists of 140 diesel and 8491 gasoline-fueled vehicles. When restricted to vehicles with Opus VSP flags of “valid”, the data set was reduced to 5725 vehicle passes. The Opus VSP validity flag eliminated the highest (~15%) and lowest (~5%) VSP values and yielded a different set of vehicle counts for MOVES bins compared with DU and HEAT, as well as a different VSP distribution. However, based on results presented in the Opus project report, the Opus VSP validity flags only indicate if VSP values fell within the range defined by the federal test procedure; flags do not describe the validity of the VSP measurements. VSP was recalculated for all RSD systems using a consistent formula for each and utilizing all vehicles passes with valid measurements of speed and acceleration.

What additional information can be ascertained when considering the control/calibration vehicle(s)?

The test vehicle data were compiled and analyzed by Revecorp Inc. The Revecorp analyses are the primary findings for the test vehicles. To facilitate interpretation of the full fleet comparisons, supplementary data analyses were carried out here for: (1) calibration data, (2) simulated evaporative emissions, (3) actual evaporative emissions, and (4) portable emissions measurement system (PEMS) data.

The results of the tests were informative but limited by small sample sizes (N = 11 to 15 calibration measurements for each RSD at each of two pollutant concentrations; N = 11 to 25 comparisons with PEMS for each RSD and pollutant) and incomparable units of measurement (e.g., indices versus g mi^{-1}).

Calibration tests were conducted by releasing calibration gases from an electric vehicle (EV) at approximately 9 a.m., 12:30 p.m., and 4 p.m. These calibrations (audits) were in addition to the usual system calibrations conducted by RSD operators. For CO, CO_2, and NO_x, results were consistent among RSDs and between RSD measurements and nominal test concentrations but with some DU NO_x values higher than nominal. Mean HEAT HC values and three of four Opus means were within two standard errors of the nominal propane concentrations. The DU response to 255.2 ppmv propane was high and its response to 756.2 ppmv propane was low. Variability in the DU response to 255.2 ppmv propane was high (standard error = 103 ppm), however, indicating that the mean of 383.5 ppmv was not significantly different from the 255.2 ppmv propane standard. Since this lower calibration point (255.2 ppmv propane) was less than the calculated DU detection limit, the calibrations were testing the DU system below its limits of quantification. The DU RSD system was designed and developed when fleet-average emissions were much higher than today.

The calibration tests provided useful supplementary information on measurement uncertainties.

A PEMS was operated in one vehicle, yielding 25 measurement records that were matched with data from the RSD systems. Since some measurements were unmatched, the number of comparison points between PEMS and RSD was less than 25 for each system (N = 22, 11, 19, and 24, respectively, for DU, HEAT, Opus 1, and Opus 2). Differences in reported measurement units, coupled with small sample size, make comparisons of RSD to PEMS data inconclusive. The test vehicle data file reported PEMS measurements in g kg^{-1} fuel and g mi^{-1}. HEAT data were reported in g kg^{-1} fuel whereas the DU and Opus measurements were reported as concentrations (% or ppm). HEAT and PEMS CO emission rates (g kg^{-1} fuel) were highly correlated over the range of CO observed but differed in magnitude (regression slope = 0.57 ± 0.05). The DU and Opus CO concentrations agreed with the PEMS data with respect to differentiating very high values from low ones. At lower concentrations, no RSD measurements were well correlated with the PEMS emission rate values. The results suggest that the uncertainties associated with lower species concentrations were relatively high compared with the measurement values, so strong correlations did not appear for low values.

All RSD systems responded to simulated evaporative emissions. The results for simulated evaporative emissions were analyzed using regressions of RSD HC values versus simulated mass emission rates (g mi^{-1}), which were calculated from butane volume flow rates (L min^{-1}), temperature, and vehicle speed. The regression approach accounts for any residual background values, not otherwise removed by data processing algorithms, as indicated by small positive intercepts. Regression slopes were not statistically different for different leak locations.

The minimum butane flow rate of simulated emissions was 1 liter per min (L min^{-1}), which is equivalent to 3 to 6 g mi^{-1} butane. This minimum rate is about two orders of magnitude greater than the U.S. EPA running loss standard of 0.05 g mi^{-1} for light-duty vehicles and trucks (U.S. EPA, 2016).

The HEAT measurements were reported in units of g mi^{-1} and should be comparable to the test values; it is unclear why they are ~3 to 4 times higher than the calculated emissions rates (subject to change because HEAT personnel are reworking the method for calculating g mi^{-1}). The HC measurements appeared to be close to the DU level of quantifiability (but were detectable). However, the Opus and DU measurements were reported as indices, whose conversion to g mi^{-1} would require additional information and data processing that were not part of the study.

Tests of actual evaporative emissions were conducted by removing fuel caps, disconnecting vapor lines, or disconnecting purge valves. Unlike the calibrations and simulated emissions, the mass or volume emission rates of the actual emissions were unknown. Therefore, comparisons were made between RSD systems. Regression relationships were not especially consistent (r^2 values of 0 to 0.25), but it is possible that absolute mass emission rates would be similar if they were calculated for the systems (DU and Opus) whose data were reported as indices.

In summary:

- Measurements of evaporative emissions were not quantitative as indicated by:
    - Little agreement among RSDs in both full fleet and test vehicles,
    - Unreliable qualitative distinctions (presence/absence inconsistent among RSDs).
- The Opus ERG evaporative index (but not the Opus Envirotest indicator) overestimated evaporative emissions:
    - Fraction of evaporative detections in full fleet much higher than HEAT or DU,
    - Positive values for blanks in simulated emissions (see Revecorp report),
    - Positive intercept when measurements regressed against calculated emissions.

References

Bishop, G. A. 2022. On-Road Remote Sensing of Automobile Emissions in the Phoenix Area: Spring 2021. Coordinating Research Council, Inc., 5755 North Point Parkway, Suite 265, Alpharetta, Georgia 30022. Contract No. E-119-3, https://crcao.org/wp-content/uploads/2022/04/DU-Phoenix-2021-Final-report-v1.pdf (last access, December 28, 2022).

Bishop, G. A. and D. H. Stedman. 1996. Measuring the emissions of passing cars. *Acc. Chem. Res*. 29, 489-495, DOI: 10.1021/ar950240x.

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.

Bishop, G. A., T. H. DeFries, J. A. Sidebottom, and J. M. Kemper. 2020. A vehicle exhaust remote sensing device method to screen vehicles for evaporative running loss emissions. *Environ. Sci. Technol*. 54, (22): 14627-14634. DOI: 10.1021/acs.est.0c05433.

Blanchard, C. 2018. Hager Environmental and Atmospheric Technologies (HEAT) and Denver University (DU) Remote Sensing Device (RSD) Data Mining. CRC Report E-119-2. Coordinating Research Council, Inc., 5755 North Point Parkway, Suite 265, Alpharetta, Georgia, 30022, CRC Report E-119-2, http://crcsite.wpengine.com/wp-content/uploads/2019/05/Final-Report-CRC-119-2-August-2018.pdf (last access April 16, 2020).

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.

Gruening, C., M. Clairotte, P. Bonnel, B. Giechaskiel, V. Valverde, A. Zardini, M. Carriero. 2019. Potential of Remote Sensing Devices (RSDs) to screen vehicle emissions - Assessment of RSD measurement performance, EUR 29871 EN, Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-79-11820-6, doi:10.2760/277092, JRC117894. https://ec.europa.eu/jrc/en/publication/potential-remote-sensing-devices-rsds-screen-vehicle-emissions (last access April 16, 2020).

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).

Hager, J. S. 2022. On-Road Remote Sensing of Automobile Emissions in the Phoenix Area: Spring 2021. Coordinating Research Council, Inc., 5755 North Point Parkway, Suite 265, Alpharetta, Georgia 30022. Contract No. E-119-3, https://crcao.org/wp-content/uploads/2022/08/Arizona_2021_CRC_final.pdf (last access, December 28, 2022).

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).

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.

Klausmeier, R. and N. Vescio. 2022. On-Road Remote Sensing of Automobile Emissions in the Phoenix Area: Spring 2021, Fleet Evaluation Analysis. Coordinating Research Council, Inc., 5755 North Point Parkway, Suite 265, Alpharetta, Georgia 30022. Contract No. E-119-3, https://crcao.org/wp-content/uploads/2022/08/E-119-3-Final-Report-OPUSv3.pdf (last access, December 28, 2022).

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

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., 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).

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 April 26, 2022).

U.S. Environmental Protection Agency (EPA). 2016. Light-Duty Vehicles, Light-Duty Trucks, and Medium-Duty Passenger Vehicles: Tier 2 Evaporative Emission Standards. https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P100O9ZV.pdf (last access, March 17, 2023).

Appendix A: Vehicle Specific Power and MOVES Operating Modes

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). Separate equations have been proposed for light-duty and heavy-duty vehicles. Hager (2018) uses the original version of the LDV 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 $tonne^{-1}$.

The U.S. EPA defines 23 vehicle operating modes in terms of speed and VSP (Table A-1). Two modes (0 = deceleration/braking and 1 = idle) are not relevant to the RSD study location.

Table A-1. MOVES operating modes.

VSP (kW tonne^{-1})	**Vehicle Speed (mph)**		
	$1 \leq v < 25$	$25 \leq v < 50$	≥ 50
< 0	11	21	33
0 ≤VSP <3	12	22	33
3 ≤VSP <6	13	23	33
6 ≤VSP <9	14	24	35
9 ≤VSP <12	15	25	35
VSP ≥12	16	26	36
12 ≤VSP <18	16	27	37
18 ≤VSP <24	16	28	38
25 ≤VSP <30	16	29	39
VSP ≥30	16	30	40

Appendix B: Types of Vehicles Sampled and Results by Type

Table B-1. Distribution of vehicles in matched data set.

Style	Diesel	Gasoline	Totals
1/2 TON PICKUP	0	33	33
1/2 TON VAN	1	3	4
2 DOOR CAB & CHASSIS	2	1	3
2 DOOR CARGO VAN	0	10	10
2 DOOR CONVERTIBLE	0	44	44
2 DOOR COUPE	0	177	177
2 DOOR CUTAWAY	0	5	5
2 DOOR HATCHBACK	1	50	51
2 DOOR INCOMPLETE PICKUP	0	8	8
2 DOOR PICKUP	4	294	298
2 DOOR SEDAN	0	3	3
2 DOOR STATION WAGON	0	16	16
2 DOOR STRAIGHT TRUCK	0	1	1
3 DOOR CARGO VAN	3	167	170
3 DOOR COUPE	0	4	4
3 DOOR PASSENGER VAN	0	31	31
3 DOOR PICKUP	0	12	12
3 DOOR STATION WAGON	0	1	1
3/4 TON PICKUP	3	3	6
3/4 TON VAN	1	2	3
4 DOOR CARGO VAN	6	8	14
4 DOOR COUPE	0	3	3
4 DOOR HARDTOP	0	2	2
4 DOOR HATCHBACK	0	331	331
4 DOOR INCOMPLETE PICKUP	0	6	6
4 DOOR PASSENGER VAN	6	318	324
4 DOOR PICKUP	77	898	975
4 DOOR SEDAN	11	2684	2695
4 DOOR STATION WAGON	19	3228	3247
5 DOOR CARGO VAN	0	52	52
5 DOOR HATCHBACK	0	3	3
5 DOOR PASSENGER VAN	0	12	12
5 DOOR STATION WAGON	0	3	3
BUS	0	2	2
CAB & CHASSIS	0	2	2
COUPE	0	2	2
DUMP TRUCK	1	0	1
HATCHBACK	0	3	3
PASSENGER VAN	0	2	2

PICKUP	0	3	3
SCHOOL BUS	0	1	1
SEDAN	0	2	2
SERVICE BODY TRUCK	1	15	16
STAKE TRUCK	0	1	1
STATION WAGON	0	7	7
STRAIGHT TRUCK	1	0	1
TANK	0	1	1
TRUCK	3	31	34
TRUCK TRACTOR	0	1	1
VAN	0	5	5
Totals	140	8491	8631

Table B-2. Mean differences between paired RSD measurements for diesel and gasoline vehicles. SE = standard error of the mean.

Species	**HEAT - DU**			**Opus - HEAT**			**Opus - DU**		
	Total	**Diesel**	**Gas**	**Total**	**Diesel**	**Gas**	**Total**	**Diesel**	**Gas**
CO	-169.8	-153.0	-170.1	-33.8	32.3	-34.9	-203.6	-120.7	-204.9
CO SE	15.3	66.2	15.5	11.8	45.7	12.0	19.3	68.3	19.6
HC	-54.7	12.7	-55.8	6.2	-46.4	7.1	-48.5	-33.7	-48.8
HC SE	2.4	33.8	2.4	1.5	31.4	1.5	1.8	13.1	1.8
NO	-20.1	-134.4	-18.2	-11.5	-46.1	-11.0	-31.6	-180.5	-29.2
NO SE	1.8	20.6	1.8	1.2	16.0	1.2	2.3	30.5	2.2
NO_2	0.1	2.9	0.0	6.4	67.6	5.4	6.5	70.5	5.5
NO_2 SE	0.2	2.4	0.2	0.7	15.6	0.6	0.7	15.6	0.6

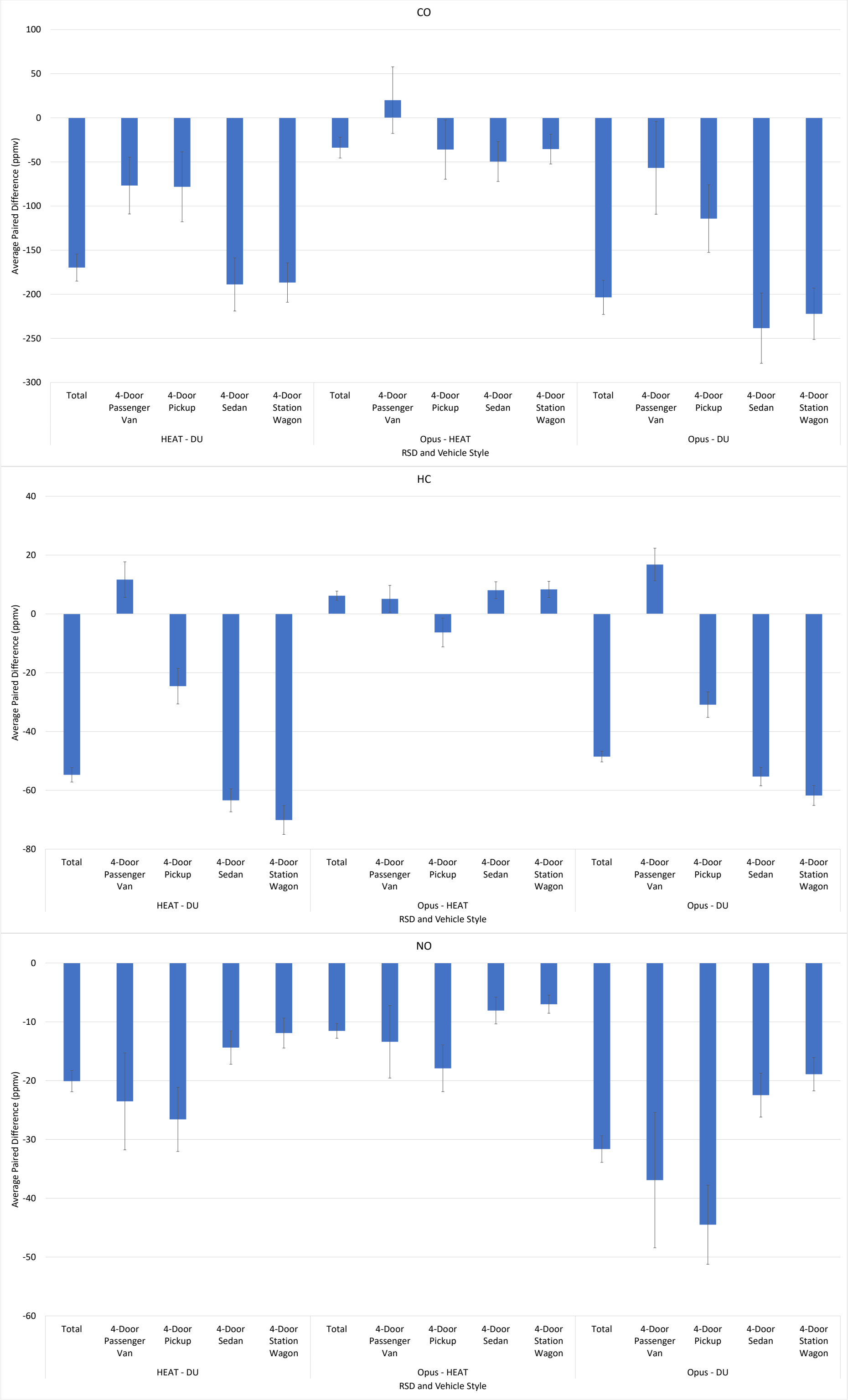

Figure B-1. RSD concentration differences by vehicle type.

Appendix C: Instrument Uncertainty

Instrument uncertainty (σ_0), as reported in previous studies and recomputed here, is determined by fitting negative concentration values to a double exponential (LaPlace) distribution, which is then 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). The regression results vary depending on the number and width of bins used to define concentration intervals. Figure C-1 shows the results for CO.

In addition to characterizing instrumental noise, calculating σ_0 provides one way to estimate detection limits. Integrating the previous equation from $-\infty$ to 0.95 yields the concentration that is significantly greater than zero at a 95% confidence level. This value is $\mathbf{C_0^{0.95} = 1.6282 \times \sigma_0}$.

The variability of measurements of a known concentration provides a second approach to estimating instrument uncertainties. Estimates of the variabilities of the lowest calibration values were determined using the test vehicle data, as illustrated in Figure C-1. The test vehicle results are limited because the sample size was small (10 to 15 passes at each concentration). Table C-1 summarizes the results in comparison with corresponding $C_0^{0.95}$ values. The large difference in sample sizes ($N = 8631$ for $C_0^{0.95}$ values versus 10 to 15 for test vehicles) suggests that the $C_0^{0.95}$ values are better estimates of measurement uncertainty. Uncertainties determined from the test vehicle calibrations are the same order of magnitude as $C_0^{0.95}$ values and many are within a factor of two; some agree within 5%.

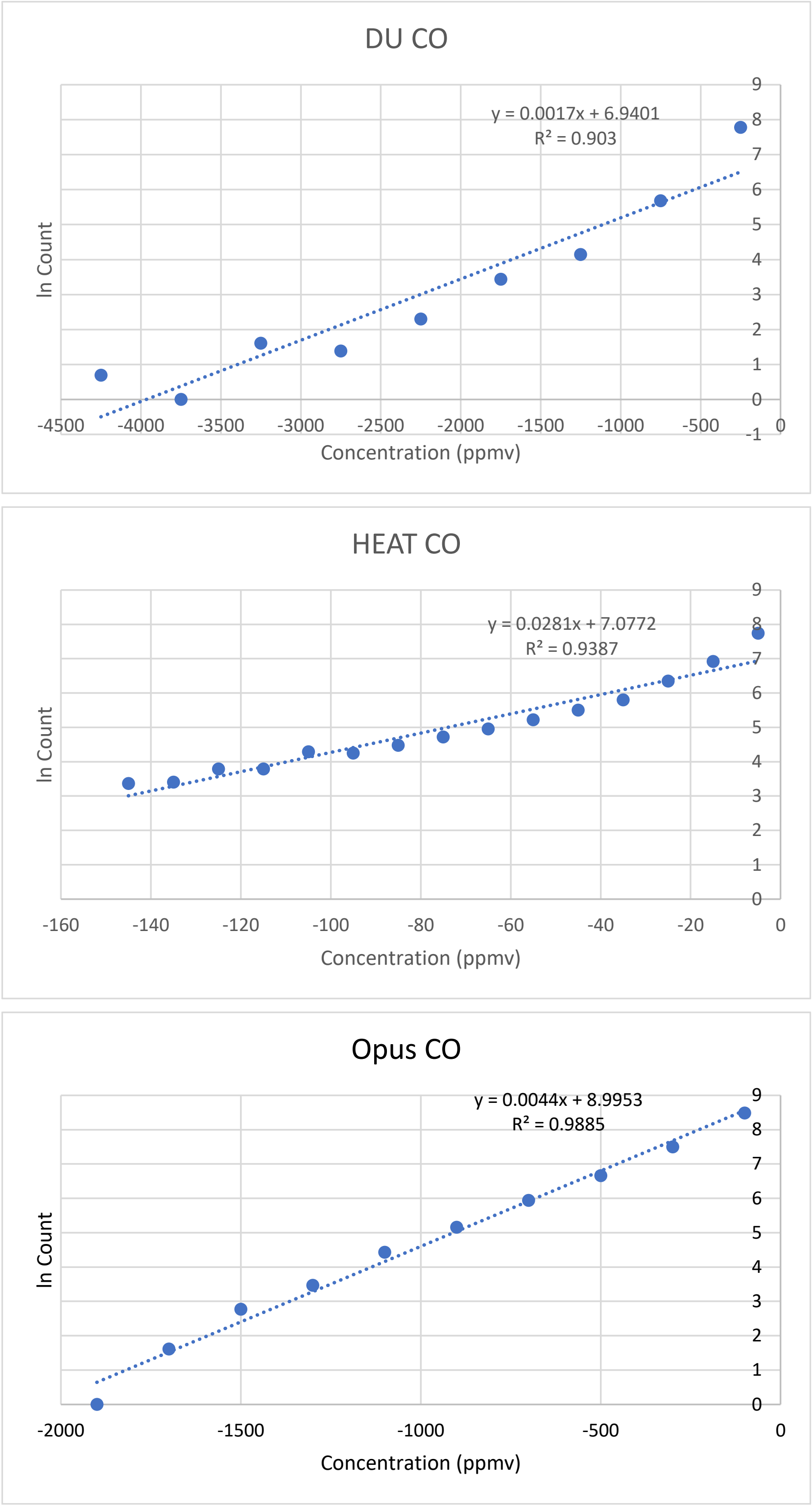

Figure C-1. Regression of the number of negative CO values within concentration intervals versus the midpoints of the intervals.

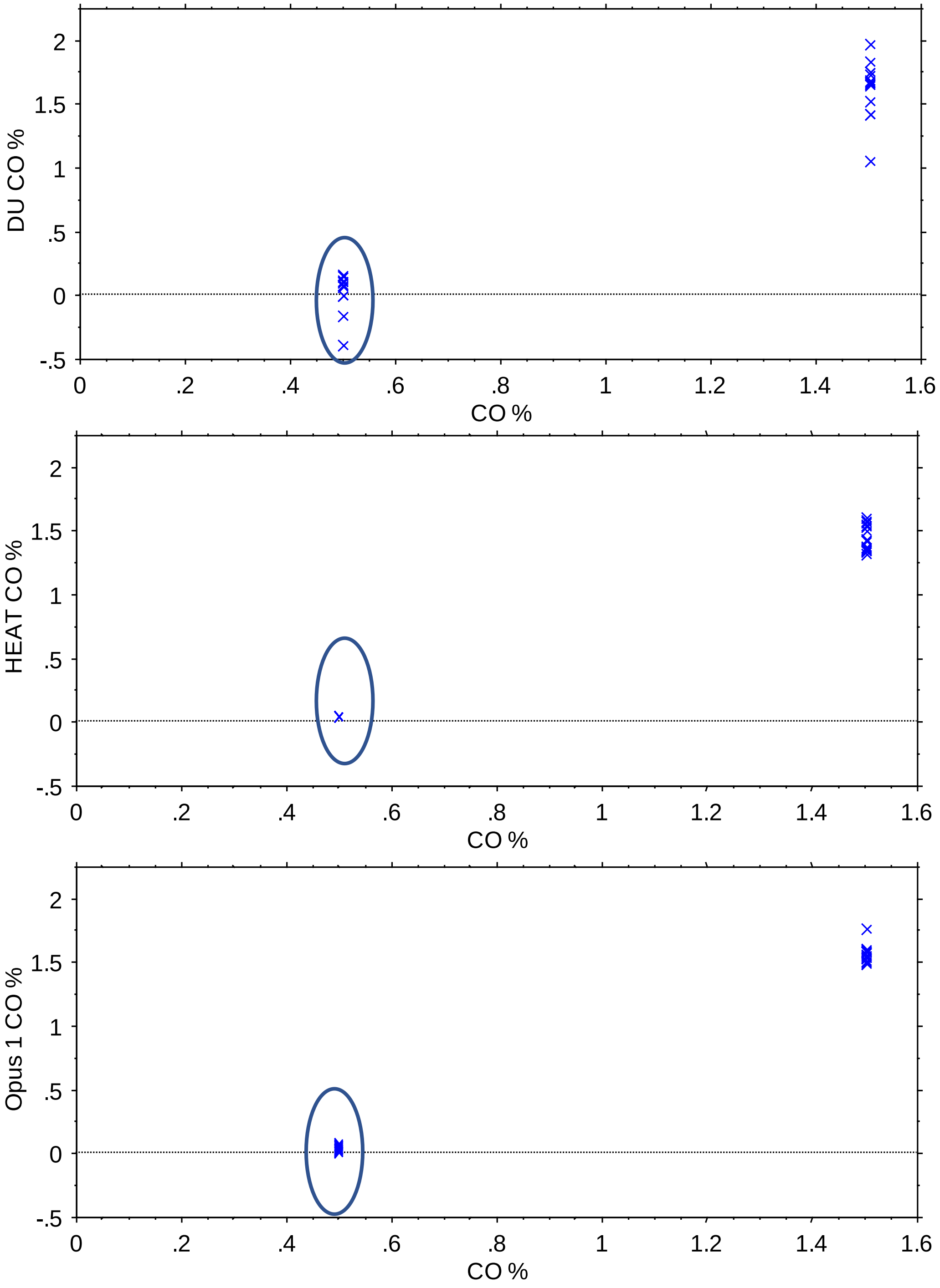

Figure C-2. Variability of measurements of a fixed, known concentration observed in test vehicles, used as an estimate of precision of individual measurements.

Table C-1. Comparison of measurement uncertainty determined from full data set ($\sigma_{0,}$ N = 8631) and from the lowest calibration standard released by test vehicles (s_{low}, N = 11 to 15). Units are ppmv. NR = not reported.

Species	DU		HEAT		Opus	
	σ0	Slow	σ0	Slow	σ0	Slow
CO	832	163	50	16	321	212
HC	184	341	20	41	54	70
NO	70	74[a]	25	5	38	35[a]
NO_2	18	74[a]	24	5	58	35[a]

a. Reported for NO_x