An analysis of real-world exhaust emission control deterioration in the California light-duty gasoline vehicle fleet

Tao Zhan a,*, Chris R. Ruehl a, Gary A. Bishop b, Seyedehsan Hosseini c, John F. Collins a, Seungju Yoon a, Jorn D. Herner a

a California Air Resources Board, 1001 I St, Sacramento, CA, 95831, USA
b Department of Chemistry and Biochemistry, University of Denver, Denver, CO, 80208, USA
c California Air Resources Board, 9480, Telstar Avenue, Ste. #4 El Monte, CA, 91731, USA

HIGHLIGHTS

- Analysis for multiple-model-year groups may produce unrealistic deterioration rates.
- Deterioration analysis for individual model years is more reliable.
- Pre-LEV vehicles’ CO and HC: no strong and consistent deterioration with age.
- Pre-LEV vehicles’ NO: consistent deterioration with age.
- LEV I vehicles’ CO, HC, and NO: consistent deterioration with age.

GRAPHICAL ABSTRACT

- Pre-LEV (1989 model year as example): CO & HC - no significant deterioration with age
- LEV I (1994 model year as example): All species - linear deterioration with age

ARTICLE INFO

Keywords:
Light-duty vehicle
Exhaust emission control
Deterioration
Remote sensing
Measurement
Modeling

ABSTRACT

The understanding of vehicle exhaust emission control deterioration is important for developing accurate emission inventories and for evaluating emission reduction regulation effectiveness. We show that a common practice of conducting deterioration analysis for multiple-model-year groups may produce unrealistic deterioration rates, because the samples’ model year constitution usually varies with age. As a more reliable approach, we analyze the deterioration for individual model years. We apply this approach to light-duty gasoline vehicle exhaust emission measurements conducted using remote sensing devices at a West Los Angeles location during seven campaigns between 1999 and 2015. Vehicles of 1985–1993 model years (termed herein as “pre-LEV” model years for predating California’s Low-Emission Vehicle regulations), most of them older than 10 years when sampled, did not exhibit strong and consistent deterioration with age for carbon monoxide (CO) and hydrocarbon (HC) emissions, whereas their nitric oxide (NO) emissions appeared to deteriorate approximately linearly with age. The 1994–2003 model-year vehicles, subject to California’s original Low-Emission Vehicle standards (LEV I standards) and up to about 20 years old when sampled, deteriorated approximately linearly with age for the three species. Our limited measurement records for the 2004–2014 model-year vehicles, subject to California’s second-generation Low-Emission Vehicle standards (LEV II standards) and up to about 10 years old when sampled, did not exhibit consistent deterioration with age. Here we propose an empirical model to describe those deterioration patterns for the pre-LEV and LEV I vehicles.

* Corresponding author.
E-mail address: Tao.Zhan@arb.ca.gov (T. Zhan).

https://doi.org/10.1016/j.atmosenv.2019.117107
Received 17 April 2019; Received in revised form 30 October 2019; Accepted 1 November 2019
Available online 2 November 2019
1352-2310/Published by Elsevier Ltd.
1. Introduction

Automobile manufacturers adopted three-way catalyst (TWC) emission control systems for light-duty vehicles (LDVs) in the United States (U.S.) around 1980, to promote simultaneous reduction of oxides of nitrogen (NOx) and oxidation of carbon monoxide (CO) and unburnt hydrocarbons (HC) in vehicle exhaust (Calvert et al., 1993; Taylor, 1993; Shelef and Graham, 1994; Shelef and McCabe, 2000). Since then, the automotive industry has improved the durability of TWC emission control systems by adopting monolithic catalyst supports in place of pellet supports, optimizing catalyst composition or pre-treatment, and developing engine control systems to keep the exhaust close to stoichiometrically balanced compositions (Taylor, 1984, 1993; Shelef and McCabe, 2000). TWC emission control performance gradually degrades due to catalyst or oxygen sensor aging, catalyst poisoning by contaminants in fuel or engine oil, catalyst or sensor malfunctioning or failure, or tampering (Taylor, 1984; Barth et al., 2000; Christou et al., 2007; Borken-Kleefeld and Chen, 2015). Over time, this deterioration can cause the in-use emission rates to be higher than the emission levels when new.

Recognizing the deterioration issues, the United States Environmental Protection Agency (U.S. EPA) and the California Air Resources Board (CARB) have enacted increasingly stringent exhaust emission control and durability requirements in their emission standards. And various U.S. states have implemented inspection and maintenance (I/M) programs (also commonly known as Smog Check programs) to identify and mitigate high-emitting older vehicles that usually have malfunctioning or significantly compromised emission controls. In addition, some states have vehicle repair and scrap incentive programs to help repair or retire the older and generally higher-emitting portion of their fleets.

Researchers have examined fleet-level emission control deterioration rates to establish reliable emission inventories and to verify the effectiveness of regulatory programs and industry actions to improve emission control durability. From a physical point of view, emission control deterioration manifests as the increase of emission factors (EF) over a period of emission control operation. But because the period of emission control operation is usually impractical to obtain for a large number of on-road vehicles, odometer reading or vehicle age is often used as a surrogate, with the former being more accurate. Odometer readings, however, are often unavailable for on-road measurements. In those cases, vehicle age may be the only feasible surrogate for emission control operation period.

On-road emission measurements using remote sensing devices (RSD) have been used to evaluate emission deterioration. Bishop and Stedman (2008) conducted RSD measurements in four U.S. cities and reported the emission control deterioration using linear regression for model years 1986–2003. Their results suggest negligible deterioration for both 1990 and earlier model year vehicles, and 2000 and later model year vehicles. Borken-Kleefeld and Chen (2015) examined about 110,000 RSD emission measurements collected at an uphill location in Zurich, Switzerland, over a 13-year period, and reported that the relative deterioration for CO and NOx were lower than suggested by the European Environment Agency’s emission inventory guidebook for Euro 1 and 2 cars (roughly corresponding to the 1994 to 1999 model years), and higher than suggested by the guidebook for Euro 3 and 4 cars (roughly corresponding to the 2000 to 2008 model years).

Portable emission measurement systems (PEMS) (Huo et al., 2012) or vehicle dynamometer tests (Wenzel, Ross, Sawyer; Bikas and Zervas, 2007; U.S. EPA; Warila et al., 2016; Zhang et al., 2017; EMFAC2017) have also been used to collect emissions measurement data for deterioration analysis. Compared with these two methods, RSD is capable of obtaining large samples of instantaneous emission rates cost-effectively and without prompting alteration of driving mode or emission device functionality. On the other hand, RSD measurements result in much shorter sampling durations as compared with PEMS or dynamometer tests; and due to its unobtrusive nature, RSD does not directly measure some important engine operational parameters that help explain emission characteristics, such as engine speed and torque, and aftertreatment temperature. Additionally, RSD measurements usually do not record vehicle odometer readings, although they often record vehicle license plates, which can be used to obtain the model years and subsequently, to estimate vehicle ages at the time of the measurement.

Regardless of the experimental method used to collect data (RSD, dynamometer, or PEMS test), many existing studies have evaluated emission control deterioration for vehicle sample bins, each of which spans multiple model years. For example, researchers have grouped all sampled Euro 1 vehicles as one bin and all sampled Euro 2 vehicles as another bin (Borken-Kleefeld and Chen, 2015; Huo et al., 2012; Bikas and Zervas, 2007; Zhang et al., 2017). However, this conventional practice of multiple-model year grouping may confound genuine deterioration (EF increase due to usage-induced emission control performance degradation) with the uncompensated impact of changing model year constitution with age or with odometer reading.

For the California LDV fleet, this study analyzes deterioration, using RSD measurements collected during seven field campaigns between 1999 and 2015 at a West Los Angeles location. To address the aforementioned potential issue associated with a multiple-model year grouping method, this study analyzes deterioration for each individual model year, providing a more reliable framework for emission deterioration assessment.

2. Materials and methods

2.1. Experimental method

The University of Denver (DU) conducted roadside LDV exhaust emission measurements using RSD at the uphill ramp (road grade of 2°) from southbound South La Brea Avenue to eastbound Interstate-10 in West Los Angeles, California in 1999, 2001, 2003, 2005, 2008, 2013, and 2015 (Bishop and Stedman, 2008, 2014, 2015; Bishop et al., 2010, 2016). All the campaigns included the measurements of CO, HC, and nitric oxide (NO), and the 2008 and later campaigns added nitrogen dioxide (NO2), sulfur dioxide (SO2), and ammonia (NH3). Table S1 provides an overview of the basic information of the measurement campaigns.

To avoid measurement drifts (deviation of measured values from corresponding true values) caused by instrument electronics or ambient conditions, quality assurance calibrations were performed at least twice daily in the field. The calibrations used gas puffs released sequentially from three multi-species cylinders (CO, CO2, propane, and NO in the first cylinder; NH3 and propane in the second; and CO2 and NO2 in the third) to adjust the measured gas ratios to those certified by cylinder manufacturers. These calibrations accounted for day-to-day instrument sensitivity variations and ambient CO2 level variations caused by other local sources, atmospheric pressure fluctuations, and instrument path length differences.

DU used a video system to capture a freeze-frame image of each vehicle’s license plate. During post-processing, the license plate numbers were used to query into the California Department of Motor Vehicles database to obtain basic vehicle information such as vehicle identification number (VIN), make, model, and model year. DU also used two pairs of infrared detectors to measure the speed and acceleration of the passing vehicles. All instruments were placed uphill to a traffic meter light, which was operational during all but the 2013 campaign.

2.2. Data treatment and filtering

Although RSD instrument measures exhaust emissions from vehicles of all fuel types, we limit the analysis to gasoline vehicle samples in this study. To establish deterioration trends, it is desirable to parse out from

T. Zhan et al.
the analysis the impact of engine load and driving mode on exhaust emission rates. To characterize engine load and driving mode, we calculated vehicle specific power (VSP) using the formula proposed by Jimenez et al. (1999), assuming negligible headwind impacts. The VSP values for all except the 2013 campaign appear to follow very similar normal distributions (Fig. S1 (d)). A closer look suggests that individual model years of different campaigns except the 2013 campaign seem to follow similar normal distributions (see Fig. S1 (e) and (f) for an illustration). This makes sense, because VSP is largely reflective of drivers’ driving patterns and can be expected to be independent of vehicles’ characteristics such as model year and age. Because of the similar VSP distributions except for the 2013 campaign, we can expect minimal engine load and driving mode impact on the longitudinal emission measurement record.

Because the traffic meter light was not operational during the 2013 campaign, the 2013 measurements concentrated in lower VSP bins and had higher fraction of negative VSP values than the other campaigns (Fig. S1 (d)). To mitigate the emission impact from the different VSP distributions in 2013, we limit the analysis to a moderate VSP range of 2–25 kW/tonne. This is because the EFs did not appear to drastically vary with VSP between zero and 25 kW/tonne (Fig. S1 (a) through (c)). We chose to set the lower VSP bound at 2 kW/tonne rather than zero to ensure positive engine load. Excluding samples outside the VSP range of 2–25 kW/tonne retained 74% of the valid samples for the 2013 campaign, as opposed to retaining 92%–97% of the valid samples for the other campaigns.

We conducted a regressions analysis using the average emissions values of the samples corresponding to each model year and age combination rather than directly using the individual samples, in order to treat all ages equally regardless of their associated sample counts. An additional benefit from this method is that the higher exclusion for the 2013 samples does not cause statistical bias in the analysis.

During each campaign, a fraction of the sampled vehicles passed through the measurement site multiple times, and hence were measured more than once. We averaged the multiple measurements for those reappearing vehicles, in order to prevent overrepresentation of reappearing vehicles in this fleet deterioration analysis. This analysis excludes model year and age combinations with less than 30 samples, because a smaller sample size is generally considered insufficient to apply Central Limit Theorem to deduce confidence intervals for the standard error of the mean when the underlying distribution for the population is unknown. By analyzing a collection of around 130,000 RSD measurements of Euro 4, 5, 6a, and 6b diesel passenger vehicles, Chen et al. (2019) suggests that the desirable minimum sample sizes for robust estimation of the population means range from 80 to 280, depending on desired certainty level and emission certification standard (Chen et al., 2019). As shown in Table S3, the vast majority of our model year and age combinations exceed the above minimum sample size of 80, and, in fact, most of our sample sizes exceed 280, which is consistent with the above suggested desirable minimum sample sizes.

Due to sample size constraints, we conduct the analysis for the light-duty vehicle category as a whole, rather than for passenger cars and light-duty trucks separately.

We categorize the RSD measurements into three vehicle emission standard groups as follows.

- Pre-LEV group is the group certified to emission standards prior to California’s original Low-Emission Vehicle program (LEV I emission standards); in this study, we limit this group to model years 1985–1993;
- LEV I group is the group certified to LEV I emission standards, consisting of all the vehicles of model years 1994–2003;
- LEV II group is the group certified to California’s second-generation Low-Emission Vehicle program (LEV II emission standards), consisting of all the vehicles of model years 2007–2013 and the vehicles of LEV I – LEV II transitional model years (model years 2004–2006)

that have been certified to the LEV II standards as determined by the VIN decoding method described below.

During the LEV II phase-in period, the volume of new vehicles certified to LEV I standards dwindled from model year 2004–2005 and from model year 2005–2006. For sampled vehicles of this transitional period, we used a commercial VIN decoder to obtain detailed vehicle specifications for each VIN, including make, model, model year, engine size, powertrain type, transmission type, and so on. Then we queried the U.S. EPA fuel economy database (www.fueleconomy.gov) to identify a representative vehicle that has the same vehicle specifications or the closest match, and assigned the emission standard (LEV I or LEV II) for that representative vehicle to the sampled vehicle. By doing so, we were able to separate LEV I and LEV II groups for the transitional 2004–2006 model years, and included in the analysis the LEV II vehicles of this transitional period, which we would otherwise have been unable to identify. We choose not to include the LEV I vehicles of these transitional model years. This is because they were only measured in the three or four most recent campaigns, resulting in shorter and hence less conclusive deterioration trends than the earlier LEV I model years.

LDV exhaust emission regulations date back to model year 1965 in California and model year 1968 at the U.S. federal level (Calvert et al., 1993). But we choose 1985 as the earliest model year in our pre-LEV group to reduce the fleet average emission impact from vehicles equipped with carburetors, the fuel delivery technology that was being phased out and replaced by closed-loop fuel injection in the 1980s.

To calculate the ages of the sampled vehicles, we assume that a new vehicle starts operating in the calendar year preceding the model year in which it is produced. For example, a 2003 model year vehicle would have been one year old in calendar year 2003. This assumption may be more accurate for the 1999 through 2005 campaigns, which took place in the fall, than the later campaigns, which took place in the spring.

2.3. Deterioration analysis for multiple-model year groups versus for individual model years

For statistical significance, emission control deterioration analyses are usually conducted for an aggregate of vehicles, rather than for each individual vehicle. A common approach of aggregating vehicles for emission control deterioration analysis in the literature is to group multiple model years of samples subject to the same emission standard into one bin (for example, model years 1994 through 2003 could be grouped into a LEV I bin). However, it is not likely for the usually limited number of measurement campaigns to cover all model years within the bin for every age point. If we take the LEV I bin for the current RSD data as an example, age of 1 year corresponds to model years 1999 (as measured in 1999), 2001 (as measured in 2001), and 2003 (as measured in 2003), hence skewed toward the later LEV I model years; whereas age of 10 years corresponds to model years 1994 (as measured in 2003), 1996 (as measured in 2005), and 1999 (as measured in 2008), hence skewed towards the earlier LEV I model years. Furthermore, it is even less likely for the sample fractions for each model year to be constant across the age range. For these reasons, the model year constitution (both the model years represented (see Table S2) and their corresponding sample fractions) within a bin would vary with age. SI Part 2 discusses this issue in more details. This issue of varying model year constitution with age would confound genuine deterioration patterns so long as different model years within the same bin have different EFs, which is often the case. Fig. S2 illustrates the unrealistic “nominal” deterioration patterns resulting from this conventional approach.

To avoid the potential impact of varying model year constitution with age, this study analyzes emission deterioration for individual vehicle model years rather than for multiple-model year groups.
3. Deterioration analysis results and discussion

3.1. Deterioration patterns

The youngest age for pre-LEV vehicles in our dataset is 7 years, which corresponds to the 1993 model year vehicles sampled during our first campaign in 1999. Most of the pre-LEV vehicles in the dataset were between 10 and 25 years old when sampled (Fig. 1). The CO and HC emission factors for 1987 and earlier model years exhibit significant fluctuations and, in some cases, downward trends (Fig. 2 (a) and (b)). This is probably because their smaller sample sizes as compared with the newer model years (Table S3) make their average emission factors more sensitive than the later pre-LEV model years to the occasional high emitter occurrences in the samples. As shown in Fig. S2.5 (a) and (b), those variations are within the very large statistical uncertainties (denoted by the 95% confidence intervals for the standard errors of the mean values). Therefore, the average CO and HC EFs do not exhibit strong and consistent deterioration.

These observations are substantiated quantitatively by the annualized deterioration for each model year (Fig. 3 (a) and (b)). In Fig. 3, we fit the points for each model year in Fig. 2 with a linear regression. The slope of such a regression represents the particular model year’s annualized emission control deterioration. Fig. 3 also includes the 95% confidence intervals for standard errors of those regression slopes. SI Part 3 provides the correlation coefficients for those regressions. For pre-LEV model years, the CO and HC deterioration rates do not consistently lie above or below the zero lines, and the 95% confidence intervals for most of those model years overlap with the zero lines (Fig. 3 (a) and (b)). This suggests that pre-LEV CO and HC emissions do not have statistically significant deteriorations with age.

These findings on deterioration trends are corroborated by a U.S. EPA analysis using dynamometer emission test data (U.S. EPA), which suggests that CO and HC EFs for the same model years leveled off after vehicles were older than about 12 years. It is worth noting that the U.S. EPA analysis shows approximately linear deterioration when vehicles were younger than 10 years (U.S. EPA), an age range generally not covered in our dataset (Fig. 1).

In contrast, pre-LEV NO emission factors increase approximately linearly with age (Fig. 2 (c)), and the increasing trend appears to be largely beyond statistical uncertainty (Fig. S2.5 (c)). The annualized NO deteriorations are consistently positive, largely beyond statistical uncertainty (Fig. 3 (c)). This differs from the findings in the U.S. EPA analysis, which suggests that NO EF for each of those pre-LEV model years leveled off after vehicles became older than about 12 years (U.S. EPA). Similar to the pre-LEV CO and HC deterioration patterns, NO emission factor fluctuated more for 1987 and older model years as compared with later pre-LEV model years. Note that the NO emission factors are expressed in mass of NO, instead of NO2.

Unlike the U.S. EPA analysis (U.S. EPA), our data do not cover the first 10–12 years of age for most pre-LEV model years. Therefore, our finding that CO and HC EFs did not increase with age whereas NO EFs increased with age is only applicable to ages older than 10–12 years. Directly extrapolating the CO and HC trends toward age zero would suggest constant EFs versus age (no deterioration), which would contradict the increasing CO and HC EFs until about 12 years old as shown in the U.S. EPA analysis (U.S. EPA).

To shed light on the pre-LEV vehicle deterioration during the first 10 years of age, we compared the CO and HC EFs for model years 1987–1991 at around 10 years old in our dataset with the EFs for the same model years when they were measured during a 1991 RSD campaign at a South El Monte location in the Los Angeles area (Stedman et al., 1994). The comparison appears to suggest substantial CO EF increase and some HC EF increase during the vehicles’ first 10 years of age (Fig. 2 (a) and (b)). We caution, however, that the South El Monte sampling location had relatively flat terrain, as opposed to the uphill West Los Angeles location; and because the 1991 campaign did not measure vehicle speed and acceleration, the above comparison cannot account for the potential impact of driving conditions on emissions. The 1991 campaign did not measure NO emissions to enable a similar comparison for NO.

For LEV I model years, the EFs of all three species consistently increase with age (Fig. 2 (d) through (f)), and the increasing trends appear to be beyond statistical uncertainties (Fig. S2.5 (d) through (f)). Their deterioration rates for the three species are positive, usually beyond statistical uncertainty (Fig. 3), suggesting consistent deterioration patterns for this group. As compared with pre-LEV model years, most LEV I model years have much smaller statistical uncertainties for the slopes (Fig. 3), because their regressions include data for all seven sampling campaigns and because their average emission factors change more consistently with ages. Their relatively high correlation coefficients for LEV I model years (SI Part 3) confirm that linear regressions fit their EF trends well. It is worth noting that, despite having more appreciable CO and HC deteriorations than pre-LEV vehicles, LEV I vehicles are estimated to have lower EF levels than pre-LEV vehicles even after 20 years of aging, thanks to their much lower EFs when new (see SI Part 4 for a detailed discussion). Perhaps unsurprisingly, RSD data collected during a 2016 Chicago area campaign showed higher EFs than data from a 2006 RSD campaign at the same location for all model years subject to the U.S. EPA Tier 1 standards (counterpart of CARB LEV I standards) (Bishop and Haugen, 2018), suggesting deterioration similar to the LEV I vehicles observed in this study.

Our measurement record for LEV II vehicles is limited. The 2003 campaign measured very few 2004 model year vehicles, resulting in too large uncertainties for the data to be useful for this analysis. Most LEV II vehicles were younger than 10 years old when measured during the four subsequent campaigns in 2005, 2008, 2013, and 2015. Except for the somewhat increasing trends for a couple model years (CO emissions of 2004 and 2005 model years, and HC and NO emissions of 2004 model year), the record has yet to show consistent EF patterns (Fig. 2 (g) through (i)) beyond statistical uncertainties (Fig. S2.5 (g) through (i)). We emphasize that each LEV II model year in Fig. 2 (g) through (i) was measured during only 2 to 4 campaigns (Fig. 1). Consequently, the uncertainties for the EF versus age linear regressions for LEV II model years are either excessively large or not quantifiable (Fig. 3). Additional campaigns in the future would extend the measurement record to enable more robust and conclusive analyses for LEV II vehicles. In comparison, Warila et al. examined I/M emission measurement data collected in Denver, Colorado since 2012 (Warila et al., 2016). That analysis populated the 2004 and subsequent model year samples that belonged to Bin 5 of the U.S. EPA Tier 2 emission standards (roughly equivalent to the LEV certification category under the CARB LEV II emission standards), and found that NOX emission factors across all examined percentiles (5,
20, 50, 80, and 95 percentiles) increased with odometer reading and with age.

3.2. Possible causes for deterioration patterns

The finding that pre-LEV vehicles older than 10 years deteriorated in NO emission performance but not in CO and HC emission performances is noteworthy. Here we propose three potential contributing factors. First, pre-LEV vehicles were driven less as they became older, making the age-based deterioration less pronounced, even if mileage-based deterioration was significant. Because it is impractical to obtain odometer readings at the time of the unobtrusive RSD measurements, we examined this factor by comparing the RSD data to another large vehicle emission dataset, for which odometer readings are available. That data, measured by the California Bureau of Automotive Repair (BAR) using chassis dynamometers set up on roadside for quasi-randomly selected on-road vehicles (Austin, McClement, Roeschen), does not show consistent age-based emission deterioration for CO, HC, and NO emissions for pre-LEV model years, although their mileage-based deteriorations were significant (Fig. S4). We postulate that this discrepancy is partly attributable to the fact that the odometer readings of those sample vehicles increased at lower rates after they became 15–20 years old. On the other hand, the pre-LEV vehicle records in the entire California Smog Check database (around 32 million records) show a similar pattern of slower mileage accumulation with increased age. If we assume that the randomly collected RSD samples represent, to some extent, the California on-road fleet, then the RSD samples may have shared a similar mileage accumulation pattern. SI Part 5 includes a more detailed discussion of this factor.

Second, California’s Smog Check program might have contributed to the CO and HC deterioration patterns for pre-LEV vehicles. Earlier versions of the Smog Check program were focused on CO and HC emissions (R-97 emission inspection, 1996). Therefore, the Smog Check program might have helped keep fleet average CO and HC emission factors approximately unchanged by removing high-CO emitters and high-HC emitters from the fleet. On the other hand, however, the NO emission factors did not level off, even though the Smog Check program imposed mandatory NO measurements starting in 1997 (R-97 emission inspection, 1996). Therefore, Smog Check as a possible contributing factor does not seem to explain the NO deterioration pattern for pre-LEV vehicles.

Third, we speculate that the different development timelines for the aftertreatment technologies for the three species may have played a role. In the 1970s, in response to the more aggressive CO and HC emission reduction requirements as compared with NO\textsubscript{X} requirements in U.S. EPA and CARB’s vehicle emission standards (Calvert et al., 1993), manufacturers adopted oxidation catalytic converters to control CO and HC emissions starting around 1975, while relying on non-catalyst technologies (mainly exhaust gas recirculation) to control NO\textsubscript{X} emissions. Manufacturers started to adopt TWC in late 1970s (Calvert et al., 1993; Taylor, 1993) to more effectively control NO\textsubscript{X} emissions for compliance with the strengthened emission requirements. Therefore, by the mid-1980s, manufacturers might have been able to address CO and HC emission control durability for aged vehicles (older than 10 years) based on accumulated experience with those technologies. It is also possible that they might have had a less thorough understanding of the newer NO\textsubscript{X} emission control technologies, and were less able to abate the deterioration of those technologies for aged vehicles.

To explain the deterioration patterns for LEV I vehicles, we postulate that manufacturers were on fast learning curves during LEV I model years; they relied on newly improved catalytic converters to control emissions to comply with increasingly stringent emission standards at that time. So the degradation of those catalytic converters themselves would be expected to increase the emission factors when vehicles aged. As a result, all three species for the LEV I vehicles exhibit consistent deterioration. On the other hand, the deterioration rates generally
This could be due to manufacturers’ incremental improvements to the performance and deterioration of emission control or the phase-in of fuel injection to replace carburetor for more accurate fuel delivery. As discussed in previous sections, CO and HC deteriorations with age were negligible (Fig. S6 (b) and (d)), whereas NO deteriorations were consistently positive (Fig. S6 (f)).

Based on the above observations, we use a simple empirical model to describe the pre-LEV vehicle deterioration. First, the model predicts the EFs at the baseline age of 10 years:

\[
EF(MY, \text{Age} = 10) = \text{slope}_{\text{pre-LEV}} \times (MY - 1985) + \text{intercept}_{\text{pre-LEV}}
\]

where EF is the emission factor (in g/kg fuel), MY and Age are the model year and the age of the vehicle (in years), and the regressed values for \( \text{slope}_{\text{pre-LEV}} \) and \( \text{intercept}_{\text{pre-LEV}} \) are listed in Table 1.

Then, the model predicts aged vehicles’ EFs, assuming linear deterioration for each model year:

\[
EF(MY, \text{Age}) = \text{DetRate} \times (\text{Age} - 10) + EF(MY, \text{Age} = 10)
\]

where DetRate is the average deterioration rate (in g/kg fuel/year) for all pre-LEV model years, also listed in Table 1. Note that the model treats the lack of deterioration as a special case of linear deterioration, in which the deterioration rate is zero. Therefore, Table 1 sets DetRate values for CO and HC at zero.

The model parameters suggest very different deterioration rates than the “nominal” deterioration trends produced by the multiple-model year grouping approach. For example, the empirical model assumes a pre-LEV NO deterioration rate of 0.33 g NO/kg fuel/year (Table 1 and Fig. S6 (f)), as compared with the “nominal” deterioration rate of 0.63 g NO/kg fuel/year (Fig. S2 (c)).

Despite the different underlying approaches, the model is capable of reproducing, with reasonable accuracies, the “nominal” deterioration trends, when we weight the model-predicted EFs by the sample sizes for the corresponding model year and age combinations, as shown in Fig. S8.

4.2. LEV I model years

For each LEV I model year, we fit the EF versus age relationship in Fig. 2 (d), (e), and (f) with a linear regression. We consider the slope of the regression line to be the emission control deterioration rate, and the y-intercept to be the extrapolated new vehicle EF. As seen in Fig. S7 (a), (c), and (e), the extrapolated new vehicle EFs for all three species decreased exponentially with model year. The decreases are not surprising, given that the LEV I regulation required every manufacturer to meet an increasingly stringent fleet average non-methane organic gas...
(NMOG) standard, which, in effect, required the then-new model year fleet to increasingly shift to cleaner emission categories under LEV I (i.e. more LEV- and ULEV-category vehicles, rather than more TLEV- and LEV-category vehicles). Incremental improvements to emission control performance also could have contributed to this decreasing trend. In addition, Fig. S7 (b), (d), and (f) show that the deterioration rates decreased approximately linearly with model year. Therefore, we model the new vehicle EFs and deterioration rates using the following equations:

$$EF_{(MY, Age = 0)} = \text{coeff} \times \text{base}^{MY-1994}$$

$$\text{DetRate}(MY) = \text{slope}_{LEV} \times (MY - 1994) + \text{intercept}_{LEV}$$

where the regressed values for coeff, base, slope_{LEV} and intercept_{LEV} are listed in Table 1.

Then the model predicts the EFs for aged vehicles, assuming linear deterioration for each model year:

$$EF_{(MY, Age)} = \text{DetRate}(MY) \times \text{Age} + EF_{(MY, Age = 0)}$$

As shown in Fig. S9, the model is capable of reproducing, with good accuracies, the "nominal" deterioration trends for LEV I model years, when we weight the model-predicted EFs by the sample sizes for the corresponding model year and age combinations.

5. Conclusions

We analyzed the deterioration with age for light-duty gasoline vehicle exhaust emission controls, using roadside RSD measurements conducted at a West Los Angeles location during seven campaigns between 1999 and 2015. Although odometer reading usually is a more accurate surrogate for emission control operation period than vehicle age, it is usually impractical to obtain odometer readings in truly unobtrusive on-road measurements (such as roadside RSD sampling), which are capable of obtaining large samples of real-world emissions in unaltered driving conditions. So long as odometer readings continue to be unavailable for unobtrusive on-road measurements, age-based deterioration analysis will continue to be useful. However, we showed that the "nominal" deterioration trends, generated by a common practice for age-based deterioration analysis by grouping multiple model years, may not reflect genuine deteriorations, because the samples’ model year constitution usually varies with age. As a more reliable approach, we analyzed the deterioration for individual model years. This individual-model year approach may also benefit mileage-based deterioration analyses, because the model year constitution of a vehicle sample set could vary with odometer reading, similar to how model year constitution varies with vehicle age.

Our results suggest that vehicle fleets deteriorated differently depending on their subjected emission standards and the pollutant species of interest. Specifically, pre-LEV vehicles (model years 1985–1993), most of them older than 10 years when sampled, did not exhibit strong and consistent deterioration with age for CO and HC EFs, whereas their NO emission control appeared to deteriorate approximately linearly with age. LEV I vehicles (model years 1994–2003), up to about 20 years old when sampled, deteriorated approximately linearly with age for those three species. Our limited measurement records for LEV II vehicles (model years 2004–2014), up to about 10 years old when sampled, did not exhibit consistent deterioration with age; additional measurement campaigns to extend LEV II record would enable more robust analyses.

Based on the analysis, we proposed an empirical model to describe the deterioration patterns for pre-LEV and LEV I vehicles. By weighting the model-predicted EFs by the sample sizes for the corresponding model year and age combinations, we showed that the model was capable of explaining, with reasonable accuracy, the "nominal" deterioration patterns for the sampled vehicle fleet. This capability demonstrates the model’s efficacy and the validity of its underlying individual-model year-based approach.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The data used in this work were collected during various on-road vehicle remote sensing measurement studies supported by the California Inspection and Maintenance Review Committee (IMRC), the Coordinating Research Council (CRC), and the California Air Resources Board (CARB). The authors thank the following CARB colleagues for their technical support and assistance: Sarah Carter, Sonya Collier, Benjamin Hancock, Michael Kambournes, Jeffrey Long, Steven Magbuhat, Michael McCarthy, Sam Pournazeri, Sherrie Sala-Moore, Satya Sardar, Yi Tan, Christopher Weaver.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2019.117107.

References


Supporting Information

An Analysis of Real-World Exhaust Emission Control Deterioration in the California Light-Duty Gasoline Vehicle Fleet

Tao Zhan†*, Chris R. Ruehl†, Gary A. Bishop†, Seyedehsan Hosseini‡, John F. Collins†, Seungju Yoon†, and Jorn D. Herner†

†California Air Resources Board, 1001 I St, Sacramento CA, 95831

‡Department of Chemistry and Biochemistry, University of Denver, Denver CO, 80208

‡California Air Resources Board, 9480 Telstar Avenue, Ste. #4 El Monte CA 91731

* Corresponding Author email: Tao.Zhan@arb.ca.gov

Unless stated otherwise, all tables, figures, and discussions herein are part of the authors’ original work for this analysis.
## SI Part 1. Basic measurement campaign information and effect of VSP

### Table S1: Basic information of West Los Angeles RSD measurement campaigns

<table>
<thead>
<tr>
<th>Campaign Year</th>
<th>Campaign Dates</th>
<th>Measured Pollutants</th>
<th>Observations (with matched license plate numbers)</th>
<th>Unique Vehicles</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>10/15/2001-10/19/2001</td>
<td>CO, HC, NO</td>
<td>20,319</td>
<td>11,805</td>
</tr>
<tr>
<td>2003</td>
<td>10/27/2003-10/31/2003</td>
<td>CO, HC, NO</td>
<td>20,191</td>
<td>11,926</td>
</tr>
<tr>
<td>2008</td>
<td>3/8/2008-3/14/2008</td>
<td>CO, HC, NO, NO₂, SO₂, NH₃</td>
<td>17,953</td>
<td>10,542</td>
</tr>
<tr>
<td>2013</td>
<td>4/27/2013-5/4/2013</td>
<td>CO, HC, NO, NO₂, SO₂, NH₃</td>
<td>27,247</td>
<td>12,210</td>
</tr>
<tr>
<td>2015</td>
<td>3/28/2015-4/3/2015</td>
<td>CO, HC, NO, NO₂, SO₂, NH₃</td>
<td>22,124</td>
<td>12,056</td>
</tr>
</tbody>
</table>

Fig. S1 (a), (b), and (c) show that carbon monoxide, hydrocarbon (HC), and nitric oxide (NO) emission factors are relatively insensitive to Vehicle Specific Power (VSP) between 0 and 25 kW/tonne. (d) shows that sample vehicles were concentrated on significantly lower VSP regions during the 2013 campaign than during the other campaigns. This is a result of the traffic meter light at the sampling location being inoperative during the 2013 campaign.

Fig. S1 (e) and (f) use two campaigns (1999 and 2015) to illustrate that individual model years of different campaigns except the 2013 campaign seem to follow similar normal distributions.
**Supporting Information**

Fig. S1. (a), (b), and (c) show distribution of carbon monoxide (CO), hydrocarbon (HC), and nitric oxide (NO) emission factor (EF) for samples with respect to Vehicle Specific Power (VSP), showing impact of VSP on emission factors for seven field campaigns. (d) shows distribution of sample size with respect to VSP. (e) and (f) illustrate VSP distribution for each of the predominant model years during the 1999 and the 2015 campaigns, respectively. Each data series is for one campaign. Each VSP bin encompasses 5 kW/tonne, and is represented in the plots by the bin's upper bound.
Supporting Information

SI Part 2. Potential disadvantages of multiple-model year grouping approach for deterioration analysis

A common approach to emission control deterioration analysis in literature is to conduct the analysis for the group of vehicles subject to a given emission standard, which usually encompasses multiple model years. As an example, Fig. S2 shows the “nominal” deterioration trends for pre-LEV, LEV I, and LEV II model year groups by using this approach. This approach may seem reasonable when those vehicles were certified to the same emission standard. However, this approach is flawed when, despite being certified to the same emission standard, the actual new vehicle emission factors vary from one model year to another. It is flawed, because the model year constitution is most likely not constant with age (see Table S2 for the model year constitution for pre-LEV, LEV I, and LEV II vehicle groups). Such model year constitution variation would confound the genuine deterioration trends, resulting in unrealistic “nominal” deterioration patterns (Fig. S2). For example, in the pre-LEV dataset, 9-year old vehicles are the 1991 model year measured during the 1999 campaign and the 1993 model year measured during the 2001 campaign. Their average model year (not weighted by model-year-specific sample sizes) is 1992. In contrast, vehicles of 15 years old are of the following model years: 1985 model year measured during the 1999 campaign, 1987 model year measured during the 2001 campaign, 1989 model year measured during the 2003 campaign, and 1991 model year measured during the 2005 campaign. Their average model year is 1988, significantly older than the model years for the 9-year-old group. Comparing those two age groups appears to suggest significant emission rate increases (Fig. S2). But such large increases do not constitute genuine deterioration; rather, they reflect the combined effect of genuine deterioration and the confounding factor that the Age 15 group consists of older model years than the Age 9 group and
have inherently higher emission rates when new, not due to deterioration. In fact, we will show in SI Part 5 that the nominal CO and HC emission rate deteriorations up to Age 18 for pre-LEV group (Fig. S2 (a), (b)) are mainly attributable to the varying model year constitution, whereas the nominal NO emission rate deterioration up to Age 17 (Fig. S2 (c)) may be caused by both of the above factors. Therefore, this multiple-model year method is flawed, and using it could generate unrealistic deterioration patterns.

Table S2: Variation of model year constitution present at each age for the samples measured during the seven campaigns

(a) Pre-LEV group

<table>
<thead>
<tr>
<th>Age</th>
<th>Year MY</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
<th>26</th>
</tr>
</thead>
</table>

(b) LEV I group

<table>
<thead>
<tr>
<th>Age</th>
<th>Year MY</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
</tr>
</thead>
</table>

(c) LEV II group

<table>
<thead>
<tr>
<th>Age</th>
<th>Year MY</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>2005</td>
<td>2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig. S2. Emission factors versus vehicle age for pre-LEV (model years 1985-1993), LEV I (model years 1994-2003), and LEV II (model years 2004-2013) model year groups. Each point is the mean EF for a model year group and age combination. The lines correspond to the listed regression formulae. The error bars represent the 95% confidence intervals for the standard
Supporting Information

errors of the mean EF values. These “nominal” trends appear to suggest significant deterioration with age for pre-LEV and LEV I groups. But they are actually the result of strong impact of varying model year constitution with age on genuine deterioration.

Because grouping vehicles for multiple model years does not parse out the impact of varying model year constitution with age, the main article carries out the deterioration analysis for individual model years.

**Table S3:** Samples sizes for pre-LEV, LEV I, and LEV II vehicle groups. Age and model year combinations with less than 30 samples are excluded.

(a) Pre-LEV group

| Age MY | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 |
|--------|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1985   | 275 | 205 | 123 | 63  | 38 |
| 1987   | 353 | 262 | 155 | 79  | 40 |
| 1988   | 513 | 399 | 239 | 128 | 71  | 40 |
| 1989   | 615 | 460 | 321 | 170 | 106 | 49 |
| 1990   | 587 | 475 | 413 | 233 | 124 | 54  | 43 |
| 1991   | 638 | 515 | 415 | 280 | 156 | 97  | 60 |
| 1992   | 587 | 502 | 382 | 242 | 168 | 95  | 56 |
| 1993   | 681 | 579 | 470 | 346 | 238 | 132 | 79 |

(b) LEV I group

| Age MY | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
|--------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1994   | 741 | 633 | 532 | 372 | 264 | 165 | 103 |
| 1995   | 951 | 749 | 606 | 476 | 313 | 187 | 133 |
| 1996   | 869 | 701 | 623 | 465 | 327 | 227 | 160 |
| 1997   | 1009 | 867 | 775 | 614 | 391 | 334 | 201 |
| 1998   | 1112 | 922 | 815 | 658 | 495 | 400 | 276 |
| 1999   | 1243 | 1025 | 942 | 739 | 518 | 455 | 312 |
| 2000   | 1296 | 1109 | 934 | 692 | 517 | 421 |
| 2001   | 1222 | 1064 | 874 | 723 | 554 | 440 |
| 2002   | 1126 | 1004 | 768 | 589 | 462 |
| 2003   | 1109 | 1078 | 767 | 660 | 567 |
Fig. S2.5 Carbon monoxide (CO), hydrocarbon (HC), and nitric oxide (NO) emission factors (EFs) versus age for pre-LEV (model years 1985-1993), LEV I (model years 1994-2003), and LEV II (model years 2004-2013) vehicles. Each point is the mean EF for a model year and age.
Supporting Information

combination. Each line connects the points for a particular model year. The error bars represent
the 95% confidence intervals for the standard errors of the mean EF values. For statistical
significance of the mean values, the plots exclude points with less than 30 sample.
Supporting Information

SI Part 3. How well do linear regressions fit the model year-specific emission factors?

Table S4: Linear regression correlation coefficients ($R^2$) for model year-specific emission factors

<table>
<thead>
<tr>
<th>MY</th>
<th>CO</th>
<th>HC</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1985</td>
<td>0.03</td>
<td>0.09</td>
<td>0.04</td>
</tr>
<tr>
<td>1986</td>
<td>0.65</td>
<td>0.13</td>
<td>0.16</td>
</tr>
<tr>
<td>1987</td>
<td>0.52</td>
<td>0.01</td>
<td>0.30</td>
</tr>
<tr>
<td>1988</td>
<td>0.06</td>
<td>0.01</td>
<td>0.14</td>
</tr>
<tr>
<td>1989</td>
<td>0.22</td>
<td>0.68</td>
<td>0.87</td>
</tr>
<tr>
<td>1990</td>
<td>0.37</td>
<td>0.32</td>
<td>0.59</td>
</tr>
<tr>
<td>1991</td>
<td>0.05</td>
<td>0.05</td>
<td>0.48</td>
</tr>
<tr>
<td>1992</td>
<td>0.29</td>
<td>0.00</td>
<td>0.76</td>
</tr>
<tr>
<td>1993</td>
<td>0.86</td>
<td>0.83</td>
<td>0.78</td>
</tr>
<tr>
<td>1994</td>
<td>0.90</td>
<td>0.70</td>
<td>0.74</td>
</tr>
<tr>
<td>1995</td>
<td>0.92</td>
<td>0.44</td>
<td>0.78</td>
</tr>
<tr>
<td>1996</td>
<td>0.56</td>
<td>0.45</td>
<td>0.83</td>
</tr>
<tr>
<td>1997</td>
<td>0.93</td>
<td>0.56</td>
<td>0.91</td>
</tr>
<tr>
<td>1998</td>
<td>0.84</td>
<td>0.63</td>
<td>0.94</td>
</tr>
<tr>
<td>1999</td>
<td>0.91</td>
<td>0.56</td>
<td>0.92</td>
</tr>
<tr>
<td>2000</td>
<td>0.90</td>
<td>0.71</td>
<td>0.96</td>
</tr>
<tr>
<td>2001</td>
<td>0.83</td>
<td>0.51</td>
<td>0.96</td>
</tr>
<tr>
<td>2002</td>
<td>0.89</td>
<td>0.00</td>
<td>0.78</td>
</tr>
<tr>
<td>2003</td>
<td>0.94</td>
<td>0.56</td>
<td>0.95</td>
</tr>
</tbody>
</table>

As discussed in the main text, we fit the points for each model year in Fig. 2 with a linear regression, in order to analyze the deterioration quantitatively.

As can be expected for relatively flat trends, the correlation coefficients for pre-LEV CO and HC regressions are generally small. The correlation coefficients for pre-LEV NO regressions are generally larger, which is consistent with the more pronounced deterioration for this species.

The correlation coefficients for LEV I linear regressions are generally much larger than their pre-LEV counterparts. Particularly, most of the coefficients for CO and NO are greater than 0.75, suggesting good linear correlations. The coefficients for HC are lower than CO and NO, due to more EF variations with age (Fig. 2).
SI Part 4. Would deteriorated LEV I vehicles reach pre-LEV vehicle emission levels?

The previous subsection shows that CO and HC emissions for pre-LEV model years did not show distinguishable deterioration trends, whereas the LEV I model years deteriorated linearly. Given this difference, as LEV I vehicles aged, would their emission controls deteriorate so much that their emission factors reached or even exceeded the pre-LEV group levels?

As can be visually estimated from Fig. 2 in the main text, the answer to this question is negative. To examine this quantitatively, we fit a linear regression for the emission factor (EF) versus age for each model year. Then we evaluate the regression lines at 10, 15, and 20 years of age for pre-LEV model years and 0, 5, 10, 15, and 20 years of age for LEV I model years (Fig. S3). Even after 20 years of overall increase, the ranges for LEV I EFs remain below the ranges for pre-LEV EFs for all three species. As a minor exception, the CO and HC EFs for model year 1994 exceeded model year 1993, which may be caused by data variability and uncertainties for linear regression. Therefore, it is not of concern that the more appreciable CO and HC deterioration for LEV I vehicles than for pre-LEV vehicles would cause the LEV I vehicle emissions catch up with the pre-LEV vehicle emission levels.
Fig. S3. Comparison of emission factor ranges for pre-LEV and LEV I vehicles. We fit a linear regression using the average emission factor (EF) measurements for each model year. We then evaluated the regression lines at age milestones 5 years apart to obtain the points in the figures. We did not evaluate EFs for pre-LEV model years to less than 10 years of age, because our RSD data generally did not cover that age range (Fig. 1).
An intriguing observation in this study is that CO and HC EFs of pre-LEV model years did not exhibit consistent increases (Fig. 2 (a) and (b)). This lack of consistent deterioration is not unique to this dataset. The California Bureau of Automotive Repair (BAR) has been testing the emissions of quasi-randomly selected on-road vehicles on Accelerated Simulation Mode (ASM) test procedures using roadside-deployed chassis dynamometers (Austin et al., 2009). An analysis of that data does not show significant age-based deterioration for CO and HC for pre-LEV model years either (Fig. S4 (a) and (b)). Interestingly, however, the NOx emissions in the BAR Roadside ASM data do not exhibit significant age-based deterioration (Fig. S4 (c)), as opposed to the approximately linear deterioration observed in the RSD data reported in the current study (Fig. 2 (c)).

Despite the lack of age-based deterioration, the same BAR Roadside ASM data does show significant deterioration for all three species for most pre-LEV model years (Fig. S4 (d), (e), (f)). We postulate that this discrepancy is partly attributable to the fact that older vehicles have been driven less than when new (Fig. S5 (a)), and therefore, their emission control deterioration, when evaluated with age, was slower when they became older. It is worth noting that the odometer readings in Fig. S4 (d), (e), and (f) and Fig. S5 have not been corrected for the potential odometer reading rollover issue (i.e. going from 99,999 miles to 0 mile) for 5-digit odometers that many pre-LEV vehicles were equipped with and that were mostly phased out by mid 1990s (Goh et al., 2007; CARB, 2015). This issue may account for the majority, if not all, of the curious decreases of exhaust concentrations between 0 and 150,000 miles (Fig. S4 (d), (e), and
Supporting Information

(f)), and the large odometer reading level differences between older and newer pre-LEV model years (Fig. S5).

The same examination of the mileage accumulation impact on the age-based deterioration is infeasible for the RSD data in this study, because accurate odometer readings were not available at the time of the RSD measurements. However, if we assume that the RSD sample vehicles, to some extent, represent the on-road fleet, then the statewide Smog Check database can lend us some insight. As shown in the Smog Check data (Fig. S5 (b)), the mileage accumulation became slower and slower with increasing age, similar to the mileage accumulation pattern in the BAR Roadside ASM data.

Therefore, we consider the slower mileage accumulation for older vehicles a potential factor contributing to the lack of consistent CO and HC emission deterioration for pre-LEV vehicles in the RSD data.
Fig. S4. Age-based deterioration (panels (a), (b), and (c)) and mileage-based deterioration (panels (d), (e), and (f)) of BAR Roadside ASM test data (pre-LEV model years). For panels (d), (e), and (f), measurements are grouped into the following odometer bins: <50 kmi, 50-100 kmi, 100-150 kmi, 150-200 kmi, 200-250 kmi, 250-300 kmi, and ≥300 kmi. Each point represents the mean of the measurements corresponding to the model year-age combination or the model year-odometer bin combination. Data included here were collected using ASM2525 test procedures between 2003 and 2017. Panels (b) and (e) exclude one exceptionally large HC measurement to avoid skewing the result.
Fig. S5. Odometer reading versus age for pre-LEV model years. (a) BAR roadside ASM test database (about 35,000 records); (b) BAR Smog Check database (about 32 million records).

References


SI Part 6. The empirical model for individual model years can reproduce the nominal trends observed by the multiple-model year grouping approach.

Parameterization of the empirical model

**Fig. S6.** Parameterization for pre-LEV deterioration empirical model. Panels (a), (c), and (e): CO, HC, and NO EFs for each model year at baseline age of 10 years; they are assumed to decrease linearly with model year. Panels (b), (d), and (f): CO, HC, and NO EF deterioration rates; for all model years, CO and HC deterioration rates are assumed to be zero, and NO deterioration rates are assumed to be a non-zero constant, 0.33 g/kg fuel/year.
Supporting Information

For the empirical model, we rely on quantitative as well as qualitative findings from the data analysis in the main body of the text. Although the mean values for pre-LEV HC are mostly positive, their uncertainties being overwhelmingly large is equally important. But there is no feasible mechanism to quantitatively incorporate the uncertainties in this straightforward empirical model. Therefore, we qualitatively incorporate the knowledge about the uncertainties, i.e. no strong and consistent HC deterioration for pre-LEV.
Fig. S7. Parameterization for LEV I deterioration empirical model. Panels (a), (c), and (e): CO, HC, and NO EFs for each model year at baseline age of 0 (i.e. for new vehicles); they are assumed to decrease exponentially with model year. Panels (b), (d), and (f): CO, HC, and NO EF deterioration rates; they are assumed to decrease linearly with model year.
We used our empirical model to predict the EFs for each pre-LEV model year and age combination. Then we weighted the EFs by the sample sizes for the corresponding model year and age combinations to reproduce the nominal trends observed from the multiple-model year grouping analysis approach (Fig. S2). As shown in Fig. S8, the model output resembles measurement-based nominal EF patterns. Particularly, even though the model assumes zero deteriorations for CO and HC EFs, its predictions align with the nominal trends of increasing EF up to age of 20 years, and the decreasing trends for higher ages (Fig. S8 (a) and (b)) caused by those higher ages generally including newer model years (SI Part 2). For NO, by assuming a relatively small deterioration rate of 0.33 g NO/kg fuel/year (Table 1 and Fig. S6 (f)), the model output matches the multiple-model year NO EF trend (Fig. S8 (c)), which shows a “nominal” deterioration rate of 0.67 g NO/kg fuel/year up to about age of 18 years old (see regression equation in Fig. S2 (c)). These model validation results further demonstrate that the nominal EF trends for multiple-model year groups in Fig. S2 are not genuine deteriorations.

Similarly, when we weighted the LEV I model output by the corresponding sample sizes for each LEV I model year and age combination, we were able to accurately predict the multiple-model year EF trends previously shown in Fig. S2 (Fig. S9). Genuine emission control deteriorations only account for a fraction of the increasing EF trends in Fig. S2. For example, the multiple-model year group NO EF increased 0.5 g CO/kg fuel/year (Fig. S2 (c)), whereas the genuine deterioration predicted by the model ranges between 0.1 and 0.4 g NO/kg fuel/year (Fig. S7 (f)).
Fig. S8. Empirical model predictions (dashed line) of the EF patterns for multiple-model year pre-LEV sample group, compared with measurement-based EF patterns (points). Model efficacy is evaluated with mean absolute percentage error (M.A.P.E.)
Supporting Information

Fig. S9. Empirical model predictions (dashed line) of the EF patterns for multiple-model year LEV I sample group, compared with measurement-based EF patterns (points). Model efficacy is evaluated with mean absolute percentage error (M.A.P.E.)