2020 National Emissions Inventory Technical Support Document: Onroad Mobile Sources

EPA-454/R-23-001e
January 2023

2020 National Emissions Inventory Technical Support Document: Onroad Mobile Sources

U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
Air Quality Assessment Division
Research Triangle Park, NC

Contents

List of Tables

List of Figures

5 Onroad Mobile – All Vehicles and Refueling

5.1 Sector description

Onroad mobile sources include emissions from motorized vehicles that normally operate on public roadways. This includes passenger cars, motorcycles, minivans, sport-utility vehicles, light-duty trucks, heavy-duty trucks, and buses. The sector includes emissions generated from parking areas, emissions from short-duration idle during pickups/deliveries, emissions from vehicles when they start, and emissions while the vehicles are moving. The sector also includes "hoteling" emissions, which refers to the time spent idling in a diesel long-haul combination truck during federally mandated rest periods of long-haul trips.

Onroad emissions in the 2020 NEI are comprised of emission estimates calculated based on version 3 of the MOVES model run with State, Local, and Tribal (S/L/T)-submitted activity data and other MOVES inputs when provided, except for California and tribes, for which the NEI includes submitted emissions. In cases where S/L/T submitted data are not provided, EPA-developed default activity based on data from the Federal Highway Administration (FHWA) and other data sources. EPA also developed default data for all other inputs required by MOVES which are used where S/L/T data of sufficient quality are not available.

The county-level GHG emissions included in the NEI for this category are calculated by running the MOVES model with State-, Local-, and Tribal-submitted activity data (when provided) and EPA-developed activity inputs based on data from FHWA and other sources. The Inventory of U.S. Greenhouse Gas Emissions and Sinks (US GHGI) reports CO_2 emissions for onroad sources based on national-level fuel consumption data from FHWA apportioned to vehicle categories (passenger cars, light-duty trucks, motorcycles, buses, and medium- and heavy-duty trucks) and fuel type according to ratios generated by MOVES. Therefore, the bottom-up NEI approach applied nationally will lead to differences with national totals in the US GHGI and the related state-level estimates in the GHGI by State.

5.2 Overview of Input Data Sources for 2020

EPA received new MOVES county database (CDB) submittals (1,565 databases) from S/L/T agencies and new 2020 vehicle registration data from IHS-Markit (IHS), which EPA adapted to compute vehicle populations (VPOP), vehicle age distributions, and fuel type fractions. FHWA provided vehicle-miles traveled (VMT) data by county and road type. EPA also received 2020 vehicle telematics data from StreetLight Data, Inc. (StreetLight), which EPA transformed into MOVES- and SMOKE-ready input files describing the distributions of vehicle speeds and fractions of VMT by hour, day of week, and month. The S/L/T CDBs for 2020 along with the vehicle registration data informed an analysis to identify counties with similar fleet characteristics to create representative county groups. Like the 2017 NEI, age distributions for representative counties are a population-weighted average of the member county age distributions. The 2020-specific vehicle speed and VMT distributions were used directly in SMOKE at the county-level; therefore, these data are not considered for the representative county selection for MOVES runs. The CDBs and representative county groups are discussed in Sections 5.5 and 5.6.2.1, respectively.

5.2.1 New 2020 Vehicle Populations, VMT, Age Distributions, and Fuel Type Mix

In areas where there is no acceptable S/L/T data available, the 2020 NEI onroad sector is based on 2020 vehicle population data from IHS-Markit (IHS) and 2020 VMT data from FHWA. To develop 2020 vehicle population data, EPA purchased a snapshot of vehicles in operation across the nation as of July 1, 2020 from IHS. EPA

processed the IHS vehicle registration summary to develop inputs for both MOVES (i.e., the CDB tables for vehicle population, age distributions, and fuel type fractions) and SMOKE (vehicle populations by SCC and county). IHS receives the registration records from each state's Department of Motor Vehicles (DMV) and decodes vehicle identification numbers (VINs) to assign each vehicle a MOVES source type code. The database IHS provided to EPA did not identify individual vehicles, but rather was a summary of the population in each county by parameters including vehicle make, model, model year, gross vehicle weight (GVW) class, and other descriptive information. An earlier analysis from the CRC A-115 study and the 2017 NEI found that IHS's registration data reflected higher light-duty vehicle (LDV) populations than corresponding state agency analyses of the same DMV data, and the differences grew with increasing age (older vehicles). The CRC A-115 study produced adjustment factors to mitigate the IHS overcount of older vehicles and released MOVES input datasets based on both the raw and adjusted information. To develop corresponding adjustment factors for 2020 NEI, EPA repeated the comparison of 2020 IHS and available 2020 S/L/T agency data for an area that included 15 agencies as described below.

Although 33 S/L/T agencies participated in the data submittal process, only 15 provided both LDV populations (MOVES `SourceTypeYear` table) and age distributions (MOVES `SourceTypeAgeDistribution` table) based on 2020 registration data, which was a requirement for comparison with the 2020 IHS data. Other agencies were excluded from the adjustment factor analysis because they provided only one type of local data (e.g., population but no age distribution) or data with outdated (e.g., year 2013) or unknown registration data draw dates. For the 15 areas that could be included in the analysis, EPA first combined the populations of passenger cars (source type 21) and light-duty trucks (source types 31 and 32) at the county level to remove the uncertainty of VIN decoding personal passenger vehicles as cars vs. light-duty trucks. EPA then allocated each county's LDV total source type population to vehicle model years for comparison with IHS and found that the IHS populations for 2020 were higher than the state data by 10.8 percent. Similar to prior years' comparisons, EPA again found that the discrepancies in the 2020 data between IHS and states are larger for older vehicles. Table 5-1 shows the adjustments EPA made to the 2020 IHS data prior to its use in the NEI.

EPA calculated the adjustment factors representing the fraction of population remaining in every model year, with two exceptions. The model year range from 2011 to 2020 received no adjustment and the model year 1990 received a capped adjustment that equals the adjustment for model year 1991. The adjustment factors in Table 5-1 were applied to the 2020 IHS data to create the EPA Default set of population and age distributions for the NEI.

Table 5-1: Older vehicle adjustments showing the fraction of IHS vehicle populations to retain for 2020 NEI

Model Year	LDV Adjustment Factor
pre-1991	0.722
1991	0.722
1992	0.728
1993	0.742
1994	0.754
1995	0.766
1996	0.774
1997	0.790
1998	0.787
1999	0.798
2000	0.796

Model Year	LDV Adjustment Factor
2001	0.806
2002	0.808
2003	0.828
2004	0.844
2005	0.857
2006	0.874
2007	0.892
2008	0.905
2009	0.919
2010	0.929
2011 - 2020	1

EPA also removed the county-specific fractions of antique license plate vehicles present in the registration summary from IHS, based on the assumption that antique vehicles are operated significantly less than average. States without any CDB submittals received EPA Default populations and age distributions based on the adjusted IHS data, and some states with submittals were overridden, decided on a case-by-case basis. Section 5.3 lists the submitted data that was accepted vs. replaced with EPA age distribution data for the 2020 NEI.

In addition to removing the older and antique plate vehicles from the IHS data, EPA also removed outlier age distributions that showed excessively "new" fleets, usually for light commercial trucks, in 28 counties. The most extreme example of this was a light commercial truck age distribution where over 85 percent of the commercial light-duty truck population in the county is 0 or 1 year old. This situation where the registration data reflects a young fleet occurs when the headquarters of a leasing or rental company owns a large fraction of the vehicles in the county. We dealt with these cases by preferentially excluding them from the representative county calculation of age distribution. For counties that were the only county in the representative county group, we made a substitution with an age distribution for the same source type from another county in the same metropolitan statistical area (MSA). EPA believes that these new vehicles do not represent the county's operating vehicle fleet, and the clean-up step avoids regions of artificially low LDV emissions in the NEI.

In areas where submitted vehicle population data were accepted for NEI, the relative populations of cars vs. light-duty trucks were reapportioned (while retaining the magnitude of the light-duty vehicles from the submittals) using the county-specific percentages from the IHS data. In this way, the categorization of cars versus light trucks is consistent from state to state. The county total light-duty vehicle populations were preserved through this process.

5.2.2 2020 Vehicle Speeds and VMT Distributions

The 2020 NEI year was unlike any other due to COVID effects on vehicle travel. Passenger car traffic went down, while in some areas freight shipping increased. Regions that typically showed slower speeds due to congestion experienced speeds that rose to free flow conditions during the day during certain months, especially in March, April, and May. For the onroad 2020 NEI, EPA surveyed S/L/T agencies regarding their ability to provide temporal profiles of speeds and VMT. Some states responded that they did not have the resources to provide inputs that would characterize the effects of COVID on the 2020 onroad activity inputs. In response to state needs, EPA purchased county-level telematics data from StreetLight for characterization of vehicle speed profiles and VMT

temporal distributions for 2020. Temporal profiles for speeds by road type were obtained by month, day of week, and hour. Vehicle types included personal, commercial medium-duty, and commercial heavy-duty.

StreetLight uses Location-Based Services (LBS) data from cellular phones as a surrogate for personal vehicles and in-vehicle Global Positioning System (GPS) data for medium and heavy-duty commercial trucks. These data are aggregated such that personal information is not revealed. StreetLight performs a great deal of data processing to identify and remove LBS trips data from cellular phones that are not traveling in vehicles (e.g., pedestrians and bicycles), and to pin the location and time data to a roadway network to calculate travel distance and speed.

The 2020 analysis of StreetLight telematics data built upon prior work sponsored by the Coordinating Research Council A-100 project to develop improved, local inputs of vehicle speeds and VMT distributions for use in MOVES and SMOKE based on 2015-2016 vehicle telematics data [ref 2]. Since 2016, StreetLight's sample size of vehicles grew significantly. For 2020, the VMT sample from Streetlight represented over nine percent of the continental US total VMT for 2020 estimated by the FHWA VM-2 table for 2020.

EPA accepted S/L/T agency submittals for month VMT fractions if the patterns clearly showed 2020 pandemic effects for the expected months. However, EPA used telematics-based data in all counties inside the CONUS for the hour and day VMT fractions as well as the speed distributions because of the higher resolution (by month and seven day types of week) available from StreetLight. Because the StreetLight dataset did not cover Alaska, Hawaii, the U.S. Virgin Islands, or Puerto Rico, EPA made substitutions using data from other areas. EPA assigned statewide averages for Montana to Alaska, and national averages of StreetLight to Hawaii, U.S. Virgin Islands, and Puerto Rico. Some of the data for commercial trucks were missing from StreetLight in certain months, so EPA substituted data from other months.

5.3 Sources of data and selection hierarchy

EPA calculated the onroad emissions for 2020 for all states using the most recent version of MOVES, MOVES3.0.4, with default database movesdb20220802. The sources of MOVES input data vary by area, representing a mix of local data, EPA defaults, and some MOVES defaults. In spite of the challenges of the pandemic, many state and local agencies submitted local input data for MOVES. The S/L/T agencies that submitted data for 2020 are listed below in Section 5.8. EPA used programs within the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system that use data output from MOVES to generate the emission inventories in all 50 states for each hour of the year. These emissions are summed over all hours and across road types to develop the emissions for the NEI. For the state of California, EPA used onroad emissions provided by California based on the EMFAC model. Pollutants submitted by California were retained to the extent possible, while any missing pollutants were estimated using a combination of data from MOVES and CARB.

The data selection hierarchy for 2020 favored local input data over EPA-developed information, with the exception of the three MOVES tables `hourVMTFraction`, dayVMTFraction`, and `avgSpeedDistribution` where county-level, telematics-based EPA Defaults were adopted for the NEI universally due to unique activity patterns by month during 2020. For areas that did not submit a MOVES CDB for this NEI, EPA used a 2020 CDB containing EPA defaults (e.g., IHS registration data and StreetLight telematics data) and some MOVES defaults, as described in Section 5.6.4. The selection definition for the onroad category in EIS is shown in Table 5-2.

Table 5-2: Onroad Data Category Selection Hierarchy for 2020 NEI

Priority	Dataset	Notes
1	S/L/T-supplied emissions	Coeur d'Alene Tribe, Kootenai Tribe of Idaho, Northern Cheyenne Tribe, Nez Perce Tribe, and Shoshone-Bannock Tribes of the Fort Hall Reservation of Idaho. California submitted emissions calculated with their own model (EMFAC).
2	S/L/T-supplied input data through 2020 NEI process	
3	2020EPA_ONROAD	All data from MOVES3

5.4 California-submitted onroad emissions

California is the only state agency for which an onroad *emissions* submittal was used in the 2020 NEI. California uses their own emission model, EMFAC 2017, which uses EICs instead of SCCs. For the 2017 NEI, EPA and California worked together to develop a code mapping to better match EMFAC's EICs to EPA MOVES' detailed set of SCCs that distinguish between off-network and on-network and brake and tire wear emissions. This level of detail is needed for modeling but not specifically for the NEI, because the NEI uses simplified/more aggregated SCCs than used in modeling. The mapping file was updated for the 2020 NEI by the California Air Resource Board (CARB) and applied to the EMFAC outputs prior to providing the data to EPA.

California provided CAP emissions, excluding NH_3, by county using EPA SCCs after applying the EIC to SCC mapping. For the 2020 NEI, we needed to add NH_3, CO_2, N_2O, methane, PAHs, and also onroad refueling emissions. Methane was added for onroad sources in California using MOVES-based scaling factors – for example, the ratio of emissions for methane compared to VOC from MOVES, for each county and SCC in California. For PAHs, due to differences in pollutants included in MOVES and those provided by CARB, PAH emissions were taken from MOVES rather than from CARB. Onroad refueling emissions are not part of the CARB submittal and were based on running MOVES with vehicle miles traveled (VMT) and vehicle population data provided by CARB.

NH_3, CO_2, and N_2O were added to the California onroad emissions by setting the state-wide total of emissions to the value obtained using MOVES, and then distributing the emissions to counties and SCCs using California-provided data from another pollutant. For NH_3, CO from California was used, while CO_2 and N_2O were based on the distribution of SO_2 from California. This way, the overall magnitude of emissions is based on MOVES, but the distribution of those emissions between counties and vehicles is based on California data. The factors used for these pollutants are computed by taking MOVES state total emissions divided by the CARB state total for CO or SO_2. The emissions for these pollutants are computed as follows:

$$CO_2 = SO_2 * 118085.3$$

$$N_2O = SO_2 * 1.805$$

$$NH_3 = CO * 0.0186$$

Like methane, manganese (MN) is an exception as it cannot be matched to speciation in MOVES and is therefore ratioed as follows:

MN = MOVES MN * CARB $PM_{2.5}$/MOVES $PM_{2.5}$

Another facet of the CARB data is that the SCC distributions are different in places from the original CARB submission. For instance, if the CARB data had emissions but no activity, or if they had emissions for non-MOVES fuel+vehicle type combinations (electric transit buses). In those cases, the emissions were apportioned to SCCs that could be mapped to SMOKE-MOVES. Another example is CARB submitted total combination truck emissions, rather than separate short-haul and long-haul, so again, emissions were apportioned to EPA SCCs.

Table -9 illustrates the data source used for CARB pollutants for the 2020 NEI.

5.5 Agency-submitted MOVES inputs

Many state and local agencies provided county-level MOVES inputs in the form of CDBs. This established format requirement enables EPA to more efficiently scan for errors and manage input datasets. EPA screened all submitted data using several quality assurance scripts that analyze the individual tables in each CDB to look for missing or unrealistic data values. EPA also reviewed submitted age distributions, road type VMT distributions, and monthly VMT distributions in consideration of whether to accept these data vs. county-specific EPA defaults.

5.5.1 Overview of MOVES input submissions

State and local agencies prepare complete sets of MOVES input data in the form of one CDB per county. One way agencies can ensure a correctly formatted CDB is to use the MOVES graphical user interface (GUI) county data manger (CDM) importer. With a proper template created for a single county, a larger set of counties (e.g., statewide) can be updated systematically with county-specific information if the preparer has well-organized county data and familiarity with MariaDB queries. However, there is no requirement of MariaDB experience to prepare the NEI submittal because the user can instead rely on the CDM to help build the individual CDBs one at a time. Table 5-3 lists the tables in each CDB and describes its content or purpose. Note that several of the tables are optional, which means that they may be left blank without consequence to a MOVES run's completeness of results. If an optional CDB table is populated, the data override MOVES internal calculations and produce a different result that may better represent local conditions.

Table 5-3: MOVES CDB tables

Table Name	**Description of Content**
avft	Fuel type fractions
avgspeeddistribution	Average speed distributions
county	Description of the county
dayvmtfraction	Fractions to distribute VMT between day types
fuelformulation	Fuel properties
fuelsupply	Fuel differences by month of year
fuelusagefraction	Fraction of the time that E85 vs. gasoline is used in flex-fuel engine vehicles
hotellingactivitydistribution	***Optional table*** – fraction of hoteling hours in which the power source is the main engine, diesel APU, electric APU, or engine-off
hotellingagefraction	***Optional table*** – fraction of hoteling hours by age (e.g., to account for newer trucks having more hoteling activity). Fractions should sum to 1.0.

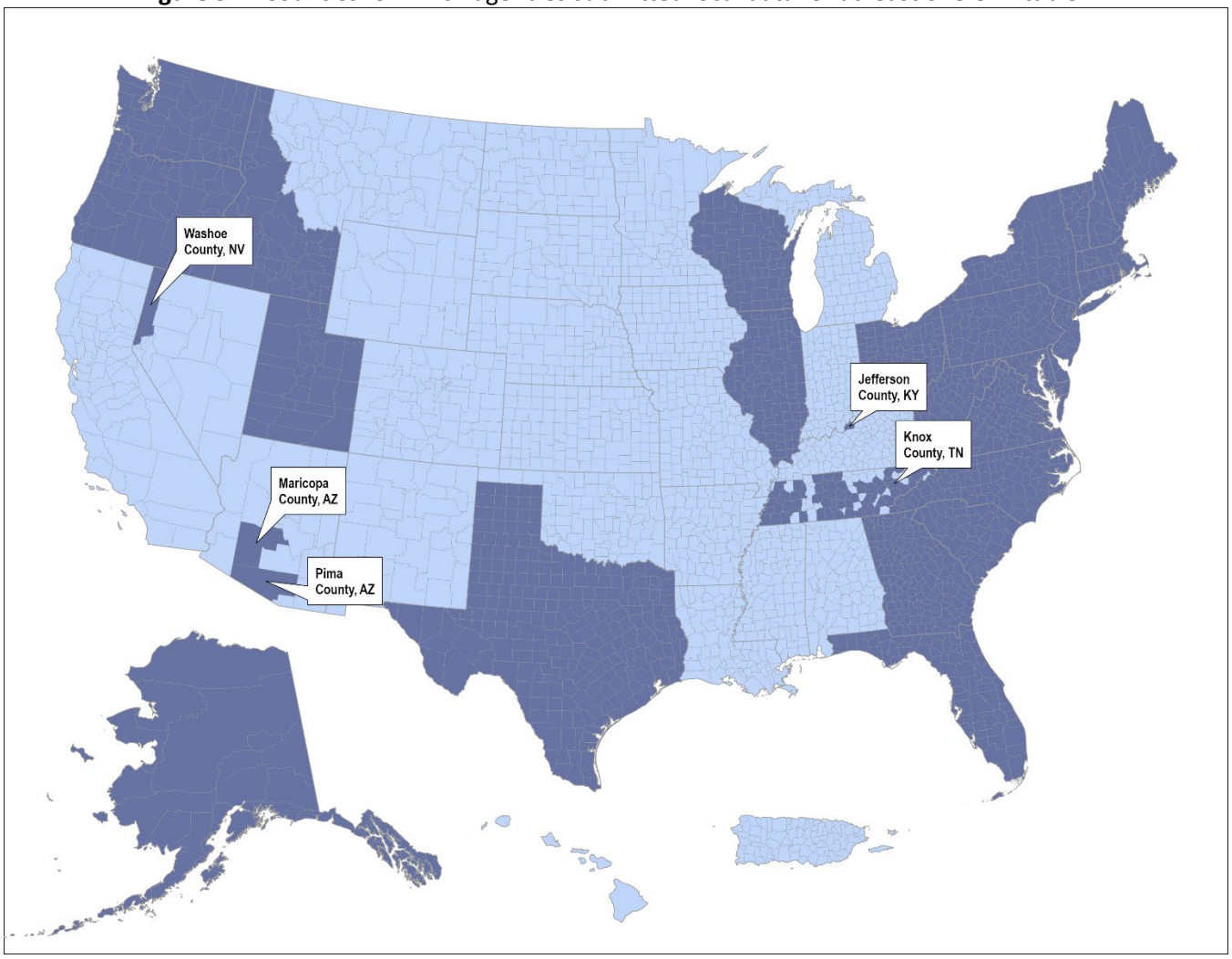
Table Name	Description of Content
hotellinghourfraction	***Optional table*** – fraction of hoteling in hours of the day. Fractions should sum to 1.0 for each day type.
hotellinghoursperday	***Optional table*** – total hours of hoteling per day, including total time spent in all of the four operating modes defined in the hotellingactivitydistribution table.
hotellingmonthadjust	***Optional table*** – adjustment factors to vary hoteling activity between different months. A factor of 1.0 for each month will model a situation where annual hoteling hours are evenly divided among months. A value of 1.1 for month ID 1 will increase the hoteling hours per day in January by 10%.
hourvmtfraction	Fractions to distribute VMT across hours in a day
hpmsvtypeday	VMT input by HPMS vehicle group, month, and day type (1 of 4 options)
hpmsvtypeyear	VMT input by HPMS vehicle group, as annual total (2 of 4 options)
idledayadjust	***Optional table*** – adjustment factors used to vary idle activity provided in the idlemodelyeargrouping table by day type (weekday or weekend day).
idlemodelyeargrouping	***Optional table*** – fraction of vehicle time operating when the speed is zero. This table is an alternative input to the totalidlefraction table. If used, idlemonthadjust and idledayadjust should also be supplied.
idlemonthadjust	***Optional table*** – adjustment factors used to vary idle activity provided in the idlemodelyeargrouping table between different months. An adjustment factor of 1.0 for each month will model the situation where the total idle fraction does not change between months.
imcoverage	Description of the inspection and maintenance program
monthvmtfraction	Fractions to distribute VMT across 12 months of the year
onroadretrofit	***Optional table*** – data for heavy-duty diesel retrofit and/or replacement program data that apply adjustments to vehicle emission rates.
roadtypedistribution	Fractions to distribute VMT across the road types
sourcetypeagedistribution	Distribution of vehicle population by age
sourcetypedayvmt	VMT input by source use type, month, and day type (3 of 4 options)
sourcetypeyear	Vehicle populations
sourcetypeyearvmt	VMT input by source use type, as annual total (4 of 4 options)
starts	***Optional table*** – starts activity, replacing the MOVES-generated starts table
startsageadjustment	***Optional table*** – numbers reflecting relative differences in the number of vehicle starts by age.
startshourfraction	***Optional table*** – fractions to distribute starts across hours in a day
startsmonthadjust	***Optional table*** – fractions to vary the vehicle starts by month of year
startsopmodedistribution	***Optional table*** – fractions to distribute the percent of engine soak-times by source type, day type, hour, and vehicle age.
Startsperday	***Optional table*** – total number of starts in a day
startsperdaypervehicle	***Optional table*** – total number of starts per vehicle in a day by source type

Table Name	Description of Content
startssourcetypefraction	***Optional table*** – fractions to distribute starts among MOVES source types
State	Description of the state
Totalidlefraction	***Optional table*** – Fraction of vehicle operating time when speed is zero
Year	Year of the database
Zone	Allocations of starts, extended idle and vehicle hours parked to the county
Zonemonthhour	Temperature and relative humidity values
Zoneroadtype	Allocation of source hours operating to the county

S/L/T agencies submitted a total of 1,565 CDBs for the 2020 NEI. Previously, agencies submitted 1,693 for the 2017 NEI, 1,816 CDBs for the 2014 NEI and 1,426 CDBs for the 2011 NEI. Agencies submitted data through the EPA Emissions Inventory System (EIS) and provided completed CDBs (i.e., each required table populated) along with documentation and a submission checklist indicating which of the CDB tables contained local data. Table 5-4 summarizes these submission checklists, showing the number of counties within each submittal for which the information was local data, as opposed to a default. Empty slots in the table indicate that the state or county did not provide local data for that particular CDB table. The grand totals of counties across all states show that VMT, population, road type distribution, and month VMT fractions were the most commonly provided local data types. Note that Table 5-4 is a select subsection of the list of CDB tables in Table 5-3. Tables not included below are tables that do not contain state specific data. For example, Year, Zone, and ZoneRoadType just list the year and geographic entity (state in this case) for the run.

Figure 5-1 shows the geographic coverage of CDB submissions where the state or local agency submitted data that was used for at least one table (dark blue). The light blue areas are counties for which the NEI uses EPA default 2020 CDBs.

Figure 5-1: Counties for which agencies submitted local data for at least one CDB table*



* Submitting areas are shown in dark blue

Table 5-4: Number of counties with submitted data, by S/L agency and select MOVES CDB table

State/County	avft	avgspeeddistribution	county	Dayvmtfraction	fuelformulation	fuelsupply	fuelusagefraction	hotellingactivitydistribution	hotellinghoursperday	hotellingmonthadjust	hourvmtfraction	hpmsvtypeyear	imcoverage	monthvmtfraction	onroadretrofit	roadtypedistribution	sourcetypeagedistribution	sourcetypedayvmt	sourcetypeyear	sourcetypeyearvmt
Alaska	29	29	29									29				29	29		29	
Arizona (Maricopa)	1	1	1	1	1	1					1	1	1	1		1	1		1	
Arizona (Pima)				1							1	1	1	1		1	1		1	
Connecticut		8		8			8				8	8	8	8		8	8*		8*	
Delaware	3						3						3	3			3		3	3
District of Columbia		1	1	1		1					1	1	1	1		1	1		1	
Florida		67									67	67	67			67	67		67	
Georgia				159							159	159	13	159		159	159		159	
Idaho	44	44		44							44	44	44	44		44	44		44	
Illinois	102		102										11				11*		102	102
Kentucky (Jefferson)	1	1	1									1				1	1		1	
Maine		16		16							16	16		16		16	16		16	
Maryland	24	24		24							24	24	24	24		24	24		24	
Massachusetts												14	14			14	14		14	
Nevada (Washoe)		1											1			1		1		
New Hampshire												10	10			10	10		10	
New Jersey		21		21							21	21	21	21	21	21	21		21	
New York				62							62	62	62	62		62	62		62	
North Carolina												100	22			100	100		100	
Ohio		88		88							88	88	7	88		88	88		88	
Oregon					36							36	4				34		34	
Pennsylvania		67		67							67	67	67	67		67	67		67	

State/County	avft	avgspeeddistribution	county	Dayvmtfraction	fuelformulation	fuelsupply	fuelusagefraction	hotellingactivitydistribution	hotellinghoursperday	hotellingmonthadjust	hourvmtfraction	hpmsvtypeyear	imcoverage	monthvmtfraction	onroadretrofit	roadtypedistribution	sourcetypeagedistribution	sourcetypedayvmt	sourcetypeyear	sourcetypeyearvmt
Rhode Island												5	5			5	5		5	
South Carolina												46								
Tennessee		63		63							63	63		63		63*	63*			
Tennessee (Knox)		1		1							1	1	1	1		1	1		1	
Texas	254	254*		254					254	254	254	254	254	254		254	254*		254	
Utah												29	29			29				
Vermont	14*						14						14	14*		14	14*		14*	14
Virginia				133								133		133		133	133		133	
Washington	1			39	1	1					39	39		39		39	39		39	39
West Virginia												1							55	54
Wisconsin												72	7	72		72	72		72	
Total	**473**	**686**	**134**	**982**	**38**	**3**	**25**	**0**	**254**	**254**	**916**	**1392**	**691**	**1071**	**21**	**1324**	**1342**	**1**	**1425**	**212**

* Partial Table Submitted (i.e., partly MOVES Default)

5.5.2 QA checks on MOVES CDB Tables

EPA reviewed lists of CDB data errors and warnings flagged by the NEI quality assurance script packaged with MOVES3. The quality assurance script reports the potential errors by compiling a list into a summary Excel file. The list of potential errors includes the CDB name, table name, a numeric error code, and in some cases the suspect data value or sum of values. EPA reviewed all potential errors, identified which ones needed to be addressed, and then coordinated with the responsible state/local agency to clarify whether the data were correct or needed revision.

The EPA MOVES team designed the NEI quality assurance script to identify not only the types of errors that would cause MOVES to crash (e.g., missing or badly formatted tables) but also those that would give erroneous results. Aside from review of quality assurance script results, EPA prepared and reviewed graphs of submitted age distributions, month VMT fractions, and road type distributions for consideration of where to override submittals with EPA default information specific to year 2020.

Many of the 1,565 submitted CDBs required at least one update due to missing or incorrect data, incorrect table formatting, or excess data (more than required), which was removed prior to use. The missing or incorrect data included the following problems:

- Age distribution represented a different data year than 2020 (i.e., LDV recession "dip" shifted by several years)
- Population data errors resulting in too-low estimates; several states did not compare within the ballpark of IHS registered vehicle populations
- Missing VMT data
- Incorrect column order for the CDB table`IMCoverage`
- Expected VMT tables required for MOVES3 (SourceTypeDayVMT, SourceTypeYearVMT, and HPMSVtypeDay) were missing
- Values sum to 0 for some source types in the `RoadTypeDistribution` table
- Removal of I/M program that did not exist in 2020

EPA resolved each of the above data problems by coordinating with state/local agencies individually. In some cases, the agency preferred to submit a corrected CDB, which EPA reviewed again to verify the intended correction. In other cases, the agency provided EPA with instructions for a spot correction to a table or simply accepted EPA's proposed update. EPA also corrected minor formatting problems with the database tables. In some cases, tables had missing data fields and/or table keys; the missing fields did not house important content, but their presence is required for MOVES to run. EPA's final decisions on the data source (submittal vs. EPA-developed information) for age distribution, speed distribution, and hourly VMT fractions can be found in the documentation spreadsheet "2022 Documentation of CDB Input Data_20230118.xlsx" posted with the 2020 NEI supplemental data files.

The following tribal onroad emissions were submitted and used in the 2020 NEI: Coeur d'Alene Tribe, Kootenai Tribe of Idaho, Northern Cheyenne Tribe, Nez Perce Tribe, and Shoshone-Bannock Tribes of the Fort Hall Reservation of Idaho.

EPA used CDBs constructed with EPA-generated data for counties where agencies did not submit input data. EPA developed new 2020 estimates of VMT, vehicle population, and hoteling at the county- and SCC-level for use in the subsequent SMOKE-MOVES processing step and inserted these data into the CDBs where states did not provide data. The SMOKE files contain this information at the resolution of SCC, which includes the source type, fuel type, and road type. When inserted into the CDB table for source type VMT (sourceTypeYearVMT), we sum

over the fuel and road type. Similarly, for population, we sum over the SCC fuel type to aggregate population to the source type level for the CDB table containing population (sourceTypeYear). In contrast, the hoteling activity detail is much more disaggregated in the two MOVES tables (hotellingHours and hotellingActivityDistribution) compared to the SMOKE FF10 hoteling file. The script that inserts these data into the set of "all CDBs" (ReverseFF10_Script_20230118.plx) is listed in *or_scripts_2020.zip*. States and counties with CDBs that included EPA-generated activity and projected CDBs are those indicated by light blue shading in Figure 5-1. Table 5-5 below lists the sources of default information by MOVES CDB table. The spreadsheet "2022 Documentation of CDB Input Data_20230118.xlsx" provides specific information about where state-supplied data were used versus default data. Additional detail on processing steps in the IHS data to create `AVFT` and `SourceTypeAgeDistribution` is provided below in Table 5-5.

Table 5-5: Source of EPA-developed information for key data tables in MOVES CDBs

CDB Table	Default content for 2020 NEI
Avft	2020 IHS registration data
Avgspeeddistribution	StreetLight telematics data
County	MOVES3 default altitude, barometric pressure, and urban/rural county type
Dayvmtfraction	StreetLight telematics data
Fuelformulation	Based on EPA estimates for each county from 2020 refinery gate batch data
Fuelsupply	Based on EPA estimates for each county from 2020 refinery gate batch data
Fuelusagefraction	MOVES3 default E85 usage
Hotellingactivitydistribution	MOVES3 default APU vs. Main Engine fractions
Hotellingagefraction	Empty by default
Hotellinghourfraction	Empty by default
Hotellinghoursperday	2020 EPA estimates of hoteling based on 2020 VMT
Hotellingmonthadjust	Flat profile that only accounts for the number of days in each month
Hourvmtfraction	StreetLight telematics data
Hpmsvtypeday	Empty by default
Hpmsvtypeyear	Empty by default
Idledayadjust	Empty by default
Idlemodelyeargrouping	Empty by default
Idlemonthadjust	Same data as Monthvmtfraction
Imcoverage	MOVES3
Monthvmtfraction	StreetLight telematics data statewide averages assigned to source types except for source type 62, which instead used a flat profile that only accounts for the number of days in each month
Onroadretrofit	Empty by default
Roadtypedistribution	MOVES3 default distributions of VMT across four road types by county
Sourcetypeagedistribution	2020 IHS registration data
Sourcetypedayvmt	Empty by default
Sourcetypeyear	2020 IHS registration data
Sourcetypeyearvmt	2020 VMT based on FHWA data
Starts	Empty by default
Startsageadjustment	Empty by default

CDB Table	Default content for 2020 NEI
Startshourfraction	Empty by default
Startsmonthadjust	Same data as Monthvmtfraction
Startsopmodedistribution	Empty by default
Startsperday	Empty by default
Startsperdaypervehicle	Empty by default
State	MOVES3 default idle region ID
Zonemonthhour	2020 meteorology data averaged by county
Emissionratebyage	The `emissionratebyage` tables for some LEV states were populated using appropriate data described in the guidance for states adopting California emission standards. These were provided to MOVES as separate databases from the CDB.

5.5.3 Preparation of `AVFT` and `SourceTypeAgeDistribution` CDB Tables

As mentioned above in Section 5.2.1, national vehicle population data from IHS for 2020 were used to derive updated age distributions adjusted to remove older vehicles (MOVES `sourceTypeAgeDistribution` table) and fuel type splits by source type and model year (MOVES `AVFT` table) in the CDBs. These data were computed at the county level for the set of "all CDBs" and were a weighted average over county groups for the set of representative CDBs used in the MOVES runs for NEI. In both cases, EPA preferred to use local data where they were found to be acceptable. Local data were used preferentially and supplemented with EPA-developed information where needed. In the EPA-developed data, the source registration data does not reliably distinguish between short-haul and long-haul activity, and so source types 52 and 53 (single unit trucks) have the same age distributions, as do source types 61 and 62 (combination unit trucks). In addition, all age distributions for long-haul trucks (source types 53 and 62) are a national average, because these vehicles are expected to travel long distances from the county where they are registered.

5.5.4 Transformation of StreetLight Telematics Data Summaries into Hour/Day/Month Distributions of VMT and Speed Distribution inputs for MOVES and SMOKE

EPA purchased year 2020 vehicle activity from StreetLight and converted the information into MOVES and SMOKE model inputs that are unique by month. EPA leveraged prior work conducted during the CRC project A-100 for the new data request to StreetLight and the data processing into NEI inputs.

Raw Data Format

Table 5-6 shows an example of five lines of data from StreetLight's delivery to EPA. The table footnotes describe the scope of possible categories in each column. In total, StreetLight generated over 630 million rows of data in the format below.

Table 5-6: Sample Rows of StreetLight Vehicle Telematics Summary Data

County FIPS[A]	Vehicle Type [B]	Road Type [C]	Year-Month [D]	Day Type [E]	Hour of Day [F]	Speed Bin [G]	Total Segment Length (ft) [H]	Total Segment Time (sec) [I]	Counts in this Combination [J]
1001	PERS	Rural Restricted	202001	F	0	10	102.01	6	1
1001	PERS	Rural Restricted	202001	F	0	20	80.71	3	1
1001	PERS	Rural Restricted	202001	F	0	25	115.85	3	1
1001	PERS	Rural Restricted	202001	F	0	45	6028.69	92	10
1001	PERS	Rural Restricted	202001	F	0	50	25878.3	353	19

[A] County FIPS: the numeric FIPS code for each county in the contiguous 48 states.

[B] Vehicle type: personal (PERS) vehicles, commercial medium-duty trucks, or commercial heavy-duty trucks.

[C] Road Type: the four MOVES road types, combination of Urban/Rural and Restricted/Unrestricted access.

[D] Year-Month: Year and month of analysis in YYYYMM format.

[E] Day Type: M, Tu, W, Th, F, Sa, Su.

[F] Hour of Day: 0 to 23, representing the hour of the day.

[G] Speed Bin: 2.5, 5, 10, 15, …90, 95, 100+ miles per hour (mph). The first 16 bins correspond to MOVES speed bins, and the final 5 are for higher speeds above 75 mph up to 100+ mph.

[H] Cumulative travel distance for all vehicles traveling within the specific speed bin on roadway segments of the specified road type in the county, occurring in the month, day type of week, and hour. Units are in feet, rounded to 2 decimal places.

[I] Time corresponding to the cumulative travel distance defined above. Units are in seconds, and values are reported in whole seconds (no decimals).

[J] Counts in the Combination refers to the number of unique road segments included on the data row. It is an indicator of sample size but does not reflect vehicle volumes.

Data Coverage

The vehicle telematics summary datasets included the lower 48 states. EPA substituted finished model-ready profiles of Montana statewide averages to cover the Alaska boroughs and nationwide averages to cover Hawaii, Puerto Rico, and the U.S. Virgin Islands.

The underlying sample size of the 2020 dataset was significantly improved from the prior effort by the CRC A-100 study because the number of devices in the data sample from Streetlight increased dramatically over the 5-year period since the CRC study, especially for the LBS data that represents personal vehicles.

Because of the growth in number of devices and resulting better coverage in the final model-ready profiles, EPA did not need to implement broad geographic grouping of counties/states, as was done previously with the telematics data used in the prior NEI. While there were not data coverage issues on the same scale as the CRC study, the nature of the SMOKE-MOVES and representative county approach to the on-road NEI still requires that activity profiles exist for all categories (i.e., all road types, vehicle classes, hours, day types), even including freeways in counties that have minimal to no freeways, or urban roads in a rural county (or vice versa – rural roads in an urban county).

Gap Filling

EPA developed two decision tree flowcharts to establish procedures for filling gaps, one for the VMT distributions and a different process for the speed distributions. In general, EPA preferred to let the local data from each county stand on its own, representing only itself, even in low data areas with resulting "noisy" data profiles that may have missing hours of data. For the VMT distributions with missing hours, EPA set those hours

values to zero (0), interpreting the missing coverage as low or no vehicle activity. As telematics data samples continue to grow into the future, the instances of missing coverage are expected to lessen. In contrast for the speed distributions, EPA did not allow missing hours of data in the modeling profiles due to the potential for data loss in the SMOKE-MOVES system and representative county approach. The decision flowchart is named 2020_StreetLight_Grouping_Decision_Charts.pdf and is included with the supporting data listed in Table -13.

MOVES and SMOKE Input File Development

EPA did not use any S/L/T agency data submittals for hourVMTFraction, dayVMTFraction, or avgSpeedDistribution, on the basis that 2020 was a unique year where the activity varied by month. January and February months look a lot different than March through December, but two of the data tables (hourVMTFraction and avgSpeedDistribution) are annual average in the CDB submittal framework. Therefore, EPA developed month-specific data for use with MOVES and SMOKE and made the information available with the supporting data, listed in Table -13.

EPA reviewed the S/L/T agency data submittals for monthVMTFraction and found that many states provided VMT distributions that reflected the actual conditions in 2020. For these agencies, EPA used the submitted monthVMTFraction. Elsewhere, EPA developed statewide averages for personal vehicles separately from commercial trucks based on the StreetLight data sample normalized by the number of devices by month. Due to sample size differences by month, the commercial truck data's monthVMTFraction had an unrealistic spike in the profile in the month of May. EPA corrected this by dropping that month and instead, interpolating it from April and June. Table 5-7 shows the S/L/T agencies where EPA used the submitted monthVMTFractions.

Table 5-7: Agency Submittal MonthVMTFraction used in the 2020 NEI

S/L/T Agency	EPA Accepted MonthVMTFraction Submittals by Source Type *
Maricopa Co., AZ	11,21,31,32
Delaware	11,21,31,32
Georgia	11, 21, 31, 32, 41, 42, 43, 51, 52, 53, 54, 61, 62
Idaho	21, 31, 32, 41, 42, 43, 51, 52, 53, 54, 61, 62
Maine	11, 21, 31, 32, 41, 42, 43, 51, 52, 53, 54, 61, 62
Maryland	11,21,31,32
New Jersey	11,21,31,32
New York	11,21,31,32
Knox Co., TN	11,21,31,32
Vermont	11
Wisconsin	11, 21, 31, 32, 41, 42, 43, 51, 52, 53, 54, 61, 62

* Source types not listed received StreetLight-based EPA Default MonthVMTFraction.

5.5.5 Default California emission standards

EPA populated an alternative MOVES database table 'EmissionRateByAge' in the CDBs for states that have adopted emission standards from California's Low Emission Vehicle (LEV) program. Table 5-8 shows states that adopted the California standards and the year the program began in each state. We developed these tables to be consistent with EPA guidance for LEV modeling provided on the EPA web site [ref 3]. The LEV database is included with MOVES_Input_DBs.zip that is available with the supporting data described in Table -13.

.

Table 5-8: States adopting California LEV standards and start year

FIPS State ID	State Name	LEV Program Start Year
06	California	1994
08	Colorado	2022
09	Connecticut	2008
10	Delaware	2014
23	Maine	2001
24	Maryland	2011
25	Massachusetts	1995
27	Minnesota	2025
32	Nevada	2025
35	New Mexico	2026
34	New Jersey	2009
36	New York	1996
41	Oregon	2009
42	Pennsylvania	2008
44	Rhode Island	2008
50	Vermont	2000
53	Washington	2009

5.6 Calculation of Emissions

5.6.1 Preparation of onroad emissions data for the continental U.S.

The 2020 NEI includes onroad emissions for every county. The same approach was used for counties inside the continental U.S. and in the outlying states and territories: the first step is to run MOVES at the county level to produce "lookup" tables of emission rates for "representative counties," using scripts designed to integrate MOVES with the SMOKE modeling system (i.e., SMOKE-MOVES). The SMOKE-MOVES approach adapted for NEI leverages gridded hourly temperature and relative humidity information available from meteorological modeling used for air quality modeling. This set of programs was developed by EPA and is also used by states and regional planning organizations to compute onroad mobile source emissions for regional air quality modeling. SMOKE-MOVES requires emission rate lookup tables generated by MOVES that differentiate emissions by process (running, start, vapor venting, etc.), vehicle type, road type, temperature, speed, hour of day, etc.

To generate the MOVES emission rates for counties in each state across the U.S., EPA used an automated process to run MOVES to produce emission factors by temperature and speed for a set of representative counties to which every other county is mapped in SMOKE, as detailed below. Using the lookup tables of MOVES emission rates, SMOKE selected appropriate emissions rates for each county, hourly temperature, SCC, and speed bin and multiplied the emission rate by activity (VMT, vehicle population, or hoteling hours) to produce emissions. These calculations were done for every county, grid cell, and hour in the continental U.S. and aggregated by county and SCC for use in the 2020 NEI. The MOVES "RunSpec" files (that provide settings for the representative county MOVES runs) are provided in the supplementary materials (see 2020_RepCounty_Runspecs.zip in Table -13). MOVES was run with two special input databases: a LEV table (see

Section 5.5.5) and a database to keep MOVES from making adjustments to NOx based on humidity levels (see Section 5.6.3 for more details). The databases are included in MOVES_Input_DBs.zip as described in Table -13.

SMOKE-MOVES tools are incorporated into recent versions of SMOKE and can be used with different versions of the MOVES model. For the 2020 NEI, EPA used the latest publicly released version at the time: MOVES3.0.4 with default database movesdb20220802 [ref 4]. Creating the NEI onroad mobile source emissions with SMOKE-MOVES requires numerous steps, as described in the sections below:

- Determine which counties will be used to represent other counties in the MOVES runs (see Section 5.6.2.1).
- Determine which months will be used to represent other month's fuel characteristics (see Section 5.6.2.2).
- Create representative CDB inputs needed for the MOVES runs (see Section 5.6.6).
- Create inputs needed both by MOVES and by SMOKE, including a list of temperatures and activity data (see Section 5.6.4).
- Run MOVES to create emission factor tables (see Section 5.6.9).
- Run SMOKE to apply the emission factors to activity data to calculate emissions (see Section 5.6.10).
- Aggregate the results at the county-SCC level for the NEI, summaries, and quality assurance (see Section 5.6.11).
- Added DIESEL-PM10 and DIESEL-PM25 by copying the PM_{10} and $PM_{2.5}$ pollutants (respectively; exhaust emissions only) as DIESEL-PM pollutants for all diesel SCCs. See Section 5.6.11.

Some things to note about the 2020 NEI that are also true of the 2017 NEI are:

SMOKE adjusts NO_X emission factors to account for humidity impacts on the pollutant using the hourly, gridded met data. To support this feature, MOVES was run with relative humidity adjustments to NO_X turned off (see nonoxadj_moves3 from MOVES_InputDbs.zip in Table -13).

- SMOKE reads in the distribution of vehicle speeds by 16 speed bins by 24 hours for weekday and weekend day types.

Some notes about the treatment of specific pollutants are as follows:

- Manganese/7439965 includes the brake and tire contribution.
- Gasoline with 85 percent ethanol (E85) was tracked as a separate fuel.
- Brake and tire PM were tracked separately from exhaust processes, although all non-refueling processes were combined into broader SCCs prior to loading into EIS.

Onroad pollutants by source are listed in Table 5-9.

Table 5-9: Onroad pollutants and sources for 2020 NEI

CAS	poll desc	poll category	data source for California
100414	Ethylbenzene	VOC HAP	CARB
100425	Styrene	VOC HAP	CARB
106423	Xylenes (mixed isomers)	VOC HAP	CARB
106990	Butadiene, 1,3-	VOC HAP	CARB

CAS	poll desc	poll category	data source for California
107028	Acrolein	VOC HAP	CARB
108383	Xylenes (mixed isomers)	VOC HAP	CARB
108883	Toluene	VOC HAP	CARB
110543	Hexane	VOC HAP	CARB
120127	Anthracene	PAH	MOVES
123386	Propionaldehyde	VOC HAP	CARB
129000	Pyrene	PAH	MOVES
1330207	Xylenes (mixed isomers)	VOC HAP	CARB
18540299	Chromium VI	Metal	CARB
191242	Benzo[g,h,i,]Perylene	PAH	MOVES
193395	Indeno[1,2,3-c,d]Pyrene	PAH	MOVES
205992	Benzo[b]Fluoranthene	PAH	MOVES
206440	Fluoranthene	PAH	MOVES
207089	Benzo[k]Fluoranthene	PAH	MOVES
208968	Acenaphthylene	PAH	MOVES
218019	Chrysene	PAH	MOVES
50000	Formaldehyde	VOC HAP	CARB
50328	Benzo[a]Pyrene	PAH	MOVES
53703	Dibenzo[a,h]Anthracene	PAH	MOVES
540841	Trimethylpentane, 2,2,4-	VOC HAP	CARB
56553	Benz[a]Anthracene	PAH	MOVES
71432	Benzene	VOC HAP	CARB
7439965	Manganese	Metal	MOVES manganese * CARB PM2.5/MOVES PM2.5
7439976	Mercury, Unspeciated	Metal	CARB
7440020	Nickel	Metal	CARB
7440382	Arsenic	Metal	CARB
75070	Acetaldehyde	VOC HAP	CARB
83329	Acenaphthene	PAH	MOVES
85018	Phenanthrene	PAH	MOVES
86737	Fluorene	PAH	MOVES
91203	Naphthalene	VOC HAP	CARB
95476	O-xylene	VOC HAP	CARB
CH4	Methane	GHG	MOVES CH4 * CARB VOC/MOVES VOC
CO	Carbon Monoxide	CAP	CARB
CO2	Carbon Dioxide	GHG	MOVES state total, allocated to county-SCC by CARB SO2
DIESEL-PM10	Diesel PM10	HAP	CARB
DIESEL-PM25	Diesel PM2.5	HAP	CARB
EC	elemental carbon	speciated PM	CARB w/MOVES speciation
N2O	Nitrous Oxide	GHG	MOVES state total, allocated to county-SCC by CARB SO2
NH3	Ammonia	CAP	MOVES state total, allocated to county-SCC by CARB SO2
NO3	particulate nitrate	speciated PM	CARB w/MOVES speciation
NOX	Nitrogen oxides	CAP	CARB
OC	organic carbon	speciated PM	CARB w/MOVES speciation

CAS	poll desc	poll category	data source for California
PM10-PRI	Particulate matter, 10 microns and less	CAP	CARB
PM25-PRI	Particulate matter, 2.5 microns and less	CAP	CARB
PMFINE	pmfine	speciated PM	CARB w/MOVES speciation
SO2	Sulfur Dioxide	CAP	CARB
SO4	particulate sulfate	speciated PM	CARB w/MOVES speciation
VOC	Volatile organic compounds	CAP	CARB

5.6.2 Representative counties and fuel months

5.6.2.1 *Representative counties*

Although EPA develops a CDB for each county in the nation, we only run MOVES for a subset of these to mitigate the computation time and cost. The representative county approach is also supported by the concept that the majority of the important emissions-determining differences among counties can be accounted for by assigning counties to groups with similar properties such as fleet age, a shared I/M program, and shared fuel controls (e.g., low RVP for summer gasoline). The county used to provide emission rates covering other counties is called the "representative county." The MCXREF file listed in Table -13 provides the mapping of each county to its representative county. Usually, the same MCXREF file is used for all MOVES processes.

In the SMOKE-MOVES framework, temperature- and speed-specific data from the representative county emission factor lookup tables are multiplied with the activity data for each of the counties within the corresponding county group. The activity data specific to individual counties in the inventory includes VMT, vehicle population, hoteling hours, hourly speed distributions, starts, and off-network idling hours (ONI).

EPA analyzed the 2020 submitted CDBs, the new 2020 age distributions derived from CRC A-115, and some MOVES data for non-submitting areas, to group similar counties and select representative counties for 2020. In line with previous modeling platforms, the MOVES input data considered for county grouping included state, altitude, fuel region, presence of an inspection and maintenance (I/M) program, and light-duty vehicle average age.

1. **State.** Only counties within the same state were allowed to be in the same representative county group.
2. **Altitude.** The altitude of each county came from the MOVES database `county` table. EPA assigned altitude values of low (most counties) or high, based on a barometric pressure cutoff of 25.8402 inches of mercury. Approximately 200 counties were considered high altitude by this metric. Only counties sharing the same altitude rating were grouped together.
3. **Fuel Region.** "Fuel region" refers to a region of counties sharing similar gasoline fuel properties. For example, those within a state's reformulated gasoline (RFG) area. The data source was the MOVES3 default database movesdb20220802.
4. **IM Bin.** The IM bin is a value of either "0" (no IM) or "1" (has IM) to indicate whether the county is part of an inspection & maintenance program area in 2020. The data source for presence of an I/M program was primarily the 2020 submittals for the NEI. If a county did not positively identify an I/M program in a submittal or did not have a submittal, the yes/no determination comes from the MOVES database `IMCoverage` table for year 2020.

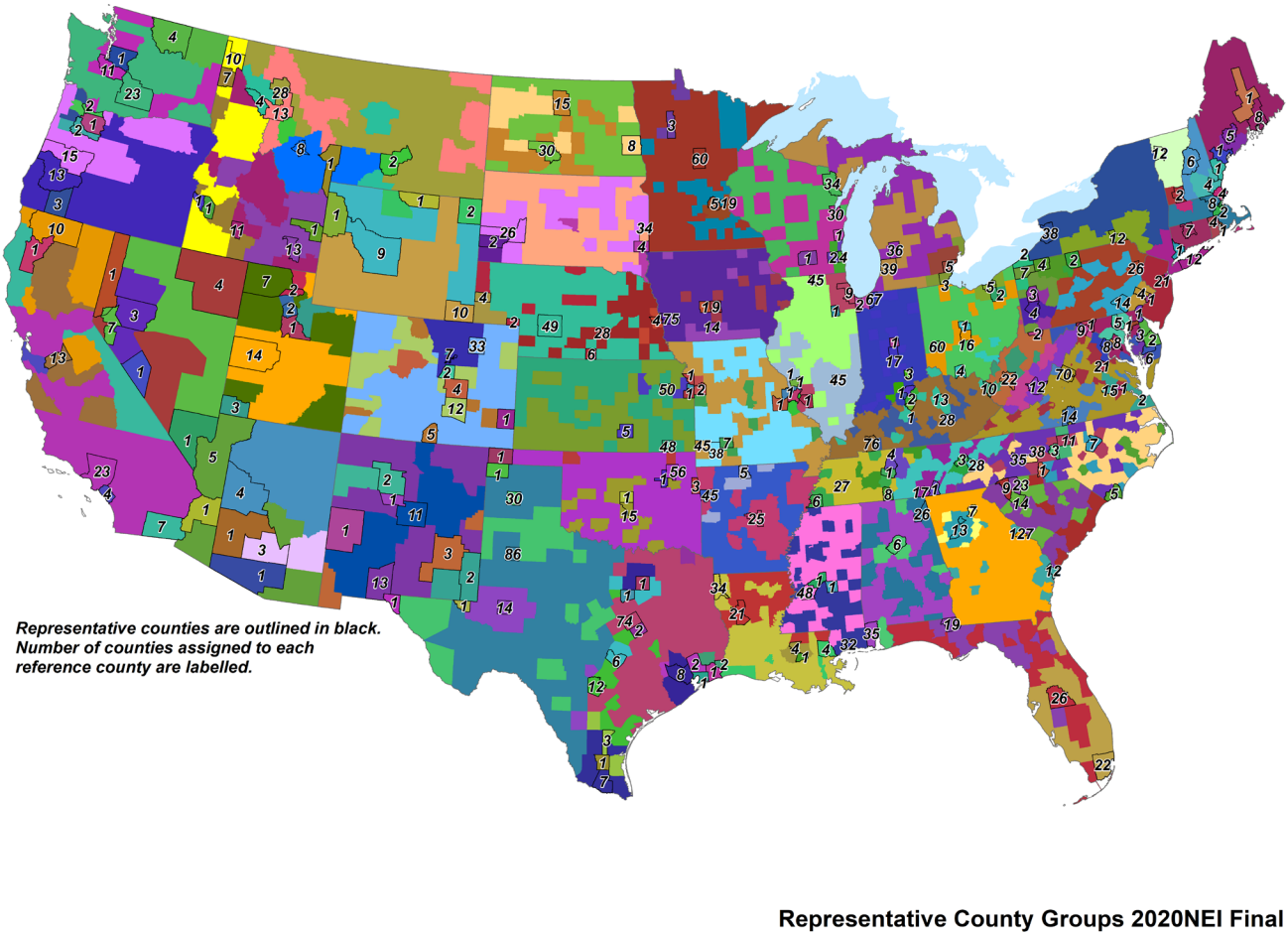
5. Mean Light-Duty Age. The age distribution of light-duty vehicles (LDVs) including passenger cars, passenger trucks, and light commercial trucks, were combined into a single population-weighted average age by county, reflecting the number of years old of the average LDV in 2020. The mean age was then binned into the six categories listed below. Only counties that share the same bin were allowed to be in the same representative county group. The source of the data was submitted age distributions that EPA accepted for use in NEI, supplemented elsewhere by the adapted 2020 IHS data.

Bin	Description (Mean age in number of years old in 2020)
1	0.0 ≤ Mean Age < 7.0
2	7.0 ≤ Mean Age < 9.0
3	9.0 ≤ Mean Age < 11.0
4	11.0 ≤ Mean Age < 13.0
5	13.0 ≤ Mean Age < 15.0
6	15.0 ≤ Mean Age

6. State requests. In the past, several agencies provided comments to EPA on the selection of representative counties for their states; however, for the 2020 NEI only Georgia requested changes, which EPA implemented.

7. After grouping similar counties, the county with the highest VMT in each group was selected as the representative county. Figure -2 displays a map of the representative counties by state and their corresponding county groups. The MCXREF file listed in Table -13 provides the mapping of each specific county to its representative county and a map showing the visualization of the county groups are provided. A spreadsheet that includes the data used in the development of the representative counties is included with the supporting data described in (2020_Representative_Counties_Analysis_20220720.xlsx).

Figure 5-2: Representative County groups for the 2020 NEI



5.6.2.2 *Fuel Months*

A "fuel month" indicates when a particular set of fuel properties should be used in a MOVES simulation. Similar to the representative county, the fuel month reduces the computational time of MOVES by using a single month to represent a set of months during which a specific fuel has been used in a representative county. Because there are winter fuels and summer fuels, EPA used January to represent October through April and July to represent May through September. For example, if the grams/mile exhaust emission rates in January are identical to February's rates for a given representative county, and temperature (as well as other factors), then we use a single fuel month to represent January and February. In other words, only one of the months needs to be modeled through MOVES to obtain the necessary emission factors. The hour-specific VMT, temperature and other factors for February are still used to calculate emissions in February, but the emission factors themselves do not need to be created, since one month can sufficiently represent the other month. The fuel months used for each representative county are provided in the MFMREF file in the supplementary materials (see Table -13 for access information).

5.6.2.3 *Fuels*

For the 2020 NEI, fuel property information came from the MOVES3 default database movesdb20220802. The fuels information was derived from refinery production compliance data, market fuel survey data, and known federal and local regulatory requirements. For a national inventory such as the NEI, this approach provides a more consistent and comprehensive result with respect to fuel use and fuel impacts on emission rates. More details on development of the MOVES fuel supply is available in this MOVES technical support document: Fuel Supply Defaults: Regional Fuels and the Fuel Wizard in MOVES3 [ref 5].

For 2020 the nationwide fuel supply assumed 100% market share E10 ethanol blends in gasoline except for Alaska which assumes E0. All diesel was assumed to be 6 ppm sulfur, and onroad diesel was 100% market share B5 biodiesel blends nationwide.

5.6.3 Temperature and humidity

Ambient temperature can have a large impact on emissions. Low temperatures are associated with high start emissions for many pollutants. High temperatures and high relative humidity are associated with greater running emissions due to the increase in the heat index and resulting higher engine load for air conditioning. High temperatures also are associated with higher evaporative emissions.

The 12-km gridded meteorological input data for the entire year of 2020 covering the continental U.S. were derived from simulations of version 4.1.1 of the Weather Research and Forecasting Model (WRF), Advanced Research WRF core [ref 6]. The WRF Model is a mesoscale numerical weather prediction system developed for both operational forecasting and atmospheric research applications. The Meteorology-Chemistry Interface Processor (MCIP) [ref 7] was used as the software for maintaining dynamic consistency between the meteorological model, the emissions model, and air quality chemistry model.

EPA applied the SMOKE program Met4moves [ref 8] to the gridded, hourly meteorological data (output from MCIP) to generate a list of the maximum temperature ranges, average relative humidity, and temperature profiles that are needed for MOVES to create the emission-factor lookup tables. “Temperature profiles” are arrays of 24 temperatures that describe how temperatures change over a day, and they are used by MOVES to estimate vapor venting emissions. The hourly gridded meteorological data (output from MCIP) was also used directly by SMOKE (see Section 5.6.10).

The temperature lists were organized based on the representative counties and fuel months as described in Section 5.6.2. Temperatures were analyzed for all of the counties that are mapped to the representative counties, i.e., for the county groups, and for all the months that were mapped to the fuel months. EPA used Met4moves to determine the minimum and maximum temperatures in a county group for the January fuel month and for the July fuel month, and the minimum and maximum temperatures for each hour of the day. Met4moves also generated temperature profiles using the minimum and maximum temperatures and 5 °F intervals. In addition to the meteorological data, the representative counties and the fuel months, Met4moves uses spatial surrogates to determine which grid cells from the meteorological data have roads and uses the WRF temperature and relative humidity data from those areas. For example, if a county had a mountainous area with no roads, the grid cells with no roads would be excluded from the meteorological processing. The spatial surrogates used for the 2020 NEI were based on activity data such as link-based VMT for the year 2017, as well as NLCD land use for the year 2019, with the goal of better characterizing the spatial variability of the onroad mobile source emissions.

For the 2020 NEI, MOVES was run with the database nonoxadj_moves3 (part of MOVES_Input_DBs.zip in Table 13) to prevent the model from making adjustments to NOx based on humidity levels. Instead, gridded hourly humidity values are used in SMOKE-MOVES to compute NOx adjustments to the unadjusted emissions output from MOVES.

Met4moves computes the range of temperatures needed by each representative county for each fuel month (i.e., 5-month summer season or 7-month winter season). When the emission factors are applied by SMOKE, the appropriate temperature bin and fuel month are used to compute the emissions. EPA used a 5 °F temperature bin size for RatePerDistance (RPD), RatePerVehicle (RPV), RatePerHour (RPH), RatePerHourONI (RPHO), and RatePerStart (RPS).

Met4moves can be run in daily or monthly mode for producing SMOKE input. In monthly mode, the temperature range is determined by looking at the range of temperatures over the whole month for that specific grid cell. Therefore, there is one temperature range per grid cell per month. While in daily mode, the temperature range is determined by evaluating the range of temperatures in that grid cell for each day. The output for the daily mode is one temperature range per grid cell per day and is a more detailed approach for modeling the vapor venting RatePerProfile (RPP) based emissions. EPA ran Met4moves in daily mode for the 2020 NEI. The temperature data output from Met4moves (2020NEI_RepCounty_Temperatures.zip) are provided with the supporting data in Table -13. The resulting temperatures for the representative counties are provided in the supplementary materials (see Table -13 for access information). The gridded, hourly temperature data used are publicly available only upon request and with provision of a disk media to copy these very large datasets.

5.6.4 VMT, vehicle population, speed, hoteling, starts, and ONI activity data

The activity data used to compute onroad mobile source emissions for the 2020 NEI uses EPA-computed data where state/local agencies did not provide their own data or where provided data did not pass quality assurance checks. These "default" (but county-specific) data were derived from Federal Highway Administration Data (FHWA) information including the published *Highway Statistics 2020* [ref 9], along with county-level VMT data that is then allocated to vehicle type, fuel type, and road type. Some additional data sources were also used. The development of the default data is described in detail in 2020NEI_default_onroad_activity_approach.pdf, which is provided with the supporting data in Table -13.

As discussed above, SMOKE combines the MOVES emission factors for each representative county with county-specific VMT, population, and hoteling data to compute the emissions for each individual county. These activity data are provided to SMOKE in a flat file format, and the source of the data varies according to area of the country and depending on whether the state/local agency submitted data for the 2020 NEI. The final activity data used are a combination of submitted data and EPA-developed data and are provided with the supporting data in (2020NEI_onroad_activity_final.zip).

For the counties for which an agency submitted a CDB (the dark blue areas shown previously in Figure 5-1), EPA ran scripts to extract the agency-submitted data from the CDBs and reformatted it into the flat file text file format that can be input to SMOKE (i.e., FF10). For the non-submitting areas of the U.S. (light blue areas in Figure 5-1), the EPA VMT, population, and hoteling were used. The 2020 default speeds are from the StreetLight telematics data. The CDBs use a distribution of speeds specific to hour, vehicle and road type, and weekday/weekend day types. SMOKE uses these same data, but the 16 speed bin distributions are averaged by hour, SCC, county, and weekday/weekend days. The speed data used for the 2020 NEI

(2020NEI_speed_spdist.zip) are included with the supporting data in Table -13. The FF10 creation scripts that read submitted CDBs are described separately by activity type below.

5.6.4.1 *VMT FF10 file creation*

The FF10-generation scripts read VMT flexibly from either the MOVES CDB table `sourceTypeYearVMT`, which contains annual VMT organized by MOVES source type, or `HPMSVtypeYear`, which contains annual VMT by groups of MOVES source types. The scripts disaggregate the VMT into fuel type, model year, and road type using a combination of other CDB tables as well as some MOVES default tables. First, the annual VMT is divided into model year using the CDB table with age distribution and the MOVES default database table containing relative annual mileage accumulation by age (`SourceTypeAge`). The scripts use these tables to create travel fractions for each source type and model year that sums to one (1) by source type.

Next, the VMT is further divided into fuel type categories of gasoline, diesel, CNG, E85, and electric vehicles – preferentially by using submitted MOVES CDB tables `AVFT` to determine the split of engine-fuel types by model year and `FuelUsageFraction` to determine the percent of flex-fuel engines that actually use E85. Flex-fuel engines refer to those capable of operating on either E85 or conventional gasoline, the percentage of which could be a function of local availability of the alternative fuel. Because the AVFT and FuelUsageFraction tables are optional tables in a MOVES CDB, they were not always populated in a submitted database. In cases where data were not provided, the FF10-generation scripts automatically default to MOVES national distributions of fuel types and/or E85 availability, using the `SampleVehiclePopulation` and `FuelUsageFraction` tables of the model default database to fill the missing data. It is worth noting that several states do not have any VMT (or vehicle population) associated with flex-fuel vehicles because they submitted data indicating either no flex-fuel vehicle population or zero E85 fuel supply in the CDB tables.

Finally, the FF10-generation scripts read the CDB table `RoadTypeDistribution` to further split VMT (by fuel type) into the four MOVES road types (urban and rural, restricted and unrestricted access). The scripts aggregate VMT across model years to the SCC level (i.e., MOVES source type, fuel type, and road type) and reports annual and monthly VMT (using the `MonthVMTFraction` CDB table) for each SCC in each county into a consolidated list. Additional processing was performed to develop the final VMT data that includes both annual and monthly totals. First state-submitted monthly profiles were applied where they were available and valid (as obtained from the MonthVMTFraction table in the CDBs). Streetlight data were used elsewhere for all vehicle types except 62s, which were treated as flat monthly where state-submitted data were not used.

5.6.4.2 *Population FF10 file creation*

The FF10-generation script that creates the SMOKE vehicle population (i.e., VPOP) data operates similarly to the VMT script just described, except that the calculations do not use travel fractions to disaggregate population by model year. First, the script reads the CDB `SourceTypeYear` table, which contains 2020 population by MOVES source type and divides it into model years based on the submitted CDB `SourceTypeAgeDistribution` table. For each vehicle model year, the scripts apportion vehicle populations to fuel types using the submitted CDB tables `AVFT' and `FuelUsageFraction`, or, if no data were provided, uses the national default corresponding data tables described in Section 5.6.4.1.

The FF10 scripts then aggregate population from the model year level back up to the SCC level (MOVES source type and fuel type, and the road type 1). The CreateFF10 script and the Reverse FF10 script that pull activity data in and out of CDBs are included with the or_scripts_2020.zip file that is included with the supporting data described in Table -13.

After the vehicle population and VMT data were finalized, the population and VMT were compared by county and source type to look for inconsistencies between the two datasets. Specifically, counties and source types with an unreasonably high miles per year per-vehicle average (VMT divided by VPOP) were identified and addressed. For counties and source types with a VMT/VPOP ratio above the threshold in Table 5-10, the vehicle population was increased so that the new VMT/VPOP ratio would equal the maximum allowable ratio. The thresholds used were based on the 90th to 95th percentile of VMT/VPOP ratio for each source type. The vehicle populations were adjusted to produce reasonable VMT/VPOP ratios because MOVES can output unrealistic emission factors when the VMT/VPOP ratios are unreasonably high.

Table 5-1010: Maximum allowable miles-per-year per-vehicle average by source type

MOVES source type	Source type description	Maximum VMT/VPOP ratio (miles per year)
11	Motorcycle	7,500
21	Passenger Car	31,000
31	Passenger Truck	31,000
32	Light Commercial Truck	31,000
41	Other Bus	130,000
42	Transit Bus	90,000
43	School Bus	30,000
51	Refuse Truck	60,000
52	Single Unit Short-haul Truck	45,000
53	Single Unit Long-haul Truck	60,000
54	Motor Home	7,000
61	Combination Short-haul Truck	150,000
62	Combination Long-haul Truck	150,000

5.6.4.3 *Speed FF10 file creation*

SMOKE uses speed data for all counties to lookup the appropriate VMT-based emission factors by speed bin and SCC. The FF10 "SPEED" input for SMOKE is one of two speed-related inputs; the other, described below, contains hourly speed distributions by SCC and county, separately for weekdays and weekends. The FF10 speed file for SMOKE contains a single daily average speed by SCC and county as an annual average and for each of the 12 months.

Because the hourly speed distributions described in the next section cover all counties and SCCs, the FF10 "SPEED" input for SMOKE is not used as part of the emissions calculations. However, SMOKE still requires an FF10 SPEED file exist, even if it is not used. Because of this, a new and complete FF10 SPEED file was not generated for 2020, and we instead used an older FF10 SPEED file in SMOKE processing.

5.6.4.4 *Speed Distribution*

The SPDIST file is generated by reformatting the MOVES `avgSpeedDistribution` CDB table into a form that can be accepted by SMOKE. The speed distribution (SPDIST) input for SMOKE is optional. Out of the three possible ways to model vehicle speeds in SMOKE, SPDIST provides the highest resolution to best match vehicle activity with the lookup tables of emission factors, which for the running processes are listed by MOVES 16 speed bins. The SPDIST file lists the fraction of time in each hour spent in each of the 16 speed bins, for weekday and weekend day types, by county, source type, and road type. MOVES provides distinct emission factors for each of the 16 speed bins, and the SPDIST tells SMOKE-MOVES how to weight each of the speed bins when computing the total emissions. For example, if the SPDIST specifies 55% of time is spent in speed bin 8 and 45% of time is spent in speed bin 9 for a particular county, hour/day, and SCC, the emission factors for those two speed bins

are weighted according to those ratios. The SMOKE-MOVES calculations also take unit conversions into account, as the SPDIST fractions are per unit time, while RPD emission factors are per unit distance.

For 2020NEI, to more accurately reflect the variation of average speeds from month to month throughout the year 2020, month-specific SPDIST files were generated. Speed data from the Streetlight dataset were used to generate hourly speed profiles by county, SCC, and month. The StreetLight data were converted into SMOKE format, gapfilled so that all counties and SCCs were covered, and modified as needed based on quality assurance checks. For example, average speed data in November and December for medium and heavy-duty vehicles was insufficient for hours 16 through 19, so those speeds in November and December were replaced with average speed data from October. To cover gaps in speed distributions (missing hours on low-data-coverage road types in certain counties) EPA grouped across urban and rural roads within the county for a given roadway "access" type, but never combining speed information to mix restricted (e.g., highways) with unrestricted (e.g., arterials and local roads).

During quality assurance review of the grouped speed profiles, it was apparent that the speeds were sometimes too different between urban and rural roads within the same county. Instead of urban/rural grouping in these cases, data were substituted from adjacent months for the same MOVES road type and county. For example, if December data on rural restricted roads in a county was lacking, the November data for rural restricted roads were sometimes a better match than December urban+rural restricted roads. The judgement of what grouping decision was the "better match" was informed from review of average hourly speed profiles for all twelve months on a single plot, with separate plots for each road type and county. EPA substituted months of speed profiles as needed. This month substitution logic was only applied to speed distributions. VMT fractions are allowed to go to zero (0) for some hours in low-data situations, while speed data may not be zero from a modeling standpoint.

5.6.4.5 *Hoteling FF10 file creation.*

Hoteling activity refers to the time spent idling in a diesel long-haul combination truck during federally-mandated rest periods for long-haul trips. Drivers may spend these rest periods with the main engine on, a smaller auxiliary power unit (APU) engine on, plugged into an electric source if available, or simply leave the engine off. MOVES tracks the emissions from hoteling using the main engine idling versus those from APUs separately. SMOKE reads each type of hoteling hours by SCC and matches them to the appropriate MOVES emission factor from the `RatePerHour' lookup table.

Submitting agencies have the option to directly provide MOVES with the number of hoteling hours (via the 'hotellingHours' table) and the percent of trucks by model year that use APUs (the 'hotellingActivityDistribution' table). These CDB tables are optional. When they are present, the FF10-generation scripts read them and translate them into the FF10 formats for SMOKE. If they are empty, the FF10-generation scripts calculate the hoteling consistently with the methodology used internally to MOVES when these tables are empty. Thus, the scripts multiply the VMT for diesel-fueled long-haul combination truck VMT on restricted access roads (urban and rural together) and with the national average rate of hoteling. For the 2020 NEI, the national average rate of hoteling was estimated by EPA to be 0.007248 hours per mile. The scripts use the submitted fractions of APU usage where available and rely on MOVES defaults otherwise.

For the 2020 NEI, EPA calculated all hoteling hours from the final VMT by SCC and county. These hoteling hours were inserted into the final set of "all CDBs" released with the modeling platform (see Section 5.8). The representative CDBs were not updated, nor do they need these data to generate hoteling emission factors. For the 2020 NEI, an adjustment to hoteling was made to address concerns raised by stakeholders about hoteling hours being artificially concentrated in areas with large amounts of combination truck VMT, but which were not

necessarily areas that trucks stopped to take long rest breaks. This is particularly an issue in heavily traveled urban areas. The hoteling hours per county were compared to the number of truck stop spaces identified in the Shapefile on which the surrogate that spatially allocates hoteling emissions to grid cells is based. This Shapefile was created collaboratively with states during the development of the 2011 NEI and updated during subsequent NEI efforts. In the analysis, for each county, the maximum number of hoteling hours per year that could be supported by the number of specified parking spaces was computed using the formula:

*max hours / year = number of spaces * 24 hours / day * 365 days / year*

This assumes that all spaces are filled at all hours of the day. The maximum number of hours was subtracted from the number of hours assigned to that county to determine if the county was over-allocated with hoteling hours as compared to the known spots. For the remaining over-allocated counties, no analysis was performed and a factor to adjust the hoteling hours down to match the max hours per year for each county was computed and applied, although it was assumed that any county can support a minimum of 105,120 hoteling hours (i.e., 12 spaces' worth). No adjustments to hoteling hours were made in counties for which hoteling hours were substantially under-allocated as compared to the number of available spots. Ideally, hoteling hours would be properly allocated to counties by someone familiar with traffic patterns in the local area. The spreadsheet used for this analysis (2020nei hotelling workbook.xlsx) is listed in Table -13.

5.6.4.6 *Off-network idling hours FF10 file creation.*

After creating VMT inputs for SMOKE-MOVES, additional work needs to be done to generation Off-network idle (ONI) activity. ONI is defined in MOVES as time during which a vehicle engine is running idle and the vehicle is somewhere other than on the road, such as in a parking lot, a driveway, or at the side of the road. This engine activity contributes to total mobile source emissions but does not take place on the road network. Examples of ONI activity include:

- light duty passenger vehicles idling while waiting to pick up children at school or to pick up passengers at the airport or train station,
- single unit and combination trucks idling while loading or unloading cargo or making deliveries, and
- vehicles idling at drive-through restaurants.

Note that ONI does not include idling that occurs on the road, such as idling at traffic signals, stop signs, and in traffic—these emissions are included as part of the running and crankcase running exhaust processes on the other road types. ONI also does not include long-duration idling by long-haul combination trucks (hoteling/extended idle), as that type of long duration idling is accounted for in other MOVES processes.
ONI activity is calculated based on VMT. For each representative county, the ratio of ONI hours to onroad VMT (on all road types) is calculated using the MOVES ONI Tool by source type, fuel type, and month. These ratios are then multiplied by each county's total VMT (aggregated by source type, fuel type, and month) to develop the ONI activity data.

5.6.4.7 *Starts FF10 file creation.*

The NEI accounts for start emissions separately from running emissions because the quantity and profile of the pollutants that vehicle engines generate are significantly different than when the running engine is fully warm. SMOKE uses the number of starts activity for all counties by SCC and matches it with the appropriate MOVES emission factor from the `RatePerStart` lookup table. The SMOKE FF10 file contains both an annual total and monthly values for the number of starts. EPA used the MOVES3 default approach to generate total starts for the year and distributed them to the twelve months using the same pattern as for the VMT FF10 file. In this way, the

vehicle starts in the NEI reflect the sharp decline in activity beginning in March 2020. EPA estimated the annual total starts by running MOVES in inventory mode at the county scale for all counties using vehicle population, age distribution, and fuel type mix consistent with the VPOP FF10 file for SMOKE. The MOVES run specification files to generate starts activity outputs included only Total Energy Consumption in the pollutant list for runtime efficiency.

5.6.5 Public release of the NEI county databases

Two sets of 2020 CDBs are available for download: (1) seeded CDBs, which have been altered to produce emission rates for all sources, roads and processes to account for represented counties that may have different distributions than their representative county, and (2) unseeded CDBs intended to be used with MOVES Inventory mode calculations. The unseeded CDBs are available for all U.S. counties, but the seeded CDBs are only available for the representative counties. See Table -13 for access details.

5.6.6 Seeded CDBs

The seeded county databases can be used with MOVES to generate emission factor lookup tables for SMOKE-MOVES. In order to create representative county CDBs for MOVES runs for SMOKE-MOVES modeling, EPA performed a "seeding" step, whereby values of zero (0) were updated to a small value of 1e-15. This seeding ensures that the lookup tables will be fully populated regardless of whether the representative county itself included activity for all of the categories covered. Seeding is necessary because counties mapping to the representative county may require an emission factor that would otherwise be missing. Note that the seeded CDBs each contain activity data for *all* of the counties represented by the CDB, not for a single county. The scripts used to develop the seeded CDBs are included in the or_scripts_2020.zip file described in Table -13.

5.6.7 Unseeded CDBs

In contrast to the seeded CDBs, the unseeded CDBs do not have any seeding performed on them and include activity data only for the individual county. This set of CDBs is true to the local conditions and could be used for MOVES inventory mode runs. The unseeded CDBs merge the databases that were agency-submitted with the default CDBs for 2020 that include updates based on StreetLight telematics data. The unseeded CDB tables `SourceTypeYearVMT`, `SourceTypeYear`, `HotellingHoursPerDay`, and `HotellingActivityDistribution`, `monthVMTFraction`, `idleMonthAdjust`, and `startsMonthAdjust` are consistent with the SMOKE-ready files of 2020 VMT, population, hoteling, ONI, and starts. Because totals of ONI and starts in the SMOKE-ready files relied on MOVES3 defaults, the totals were not put back into the unseeded set of all CDBs; only their monthly variation was put into `idleMonthAdjust` for ONI and `startsMonthAdjust` for starts, because these relied on 2020 specific VMT data by month. Activity data can be taken in and out of the unseeded individual county CDBs using the CreateFF10 and ReverseFF10 scripts included in the or_scripts2020.zip file described in Table -13.

5.6.8 Supplemental MOVES tables for Month-Specific MOVES Inventory Runs in Year 2020

EPA populated the unseeded set of CDB tables `avgSpeedDistribution`, `hourVMTFraction`, `dayVMTFraction`, and `monthVMTFraction` with profiles based on StreetLight telematics data. The data resolution for the speed distributions and hour VMT fractions is annual because the CDB table structure does not include a month field. For this reason, EPA released two supplemental data file types covering monthly, county-level versions of the CDB tables `avgSpeedDistribution` and `hourVMTFraction`. EPA split the files by state because of their large file size. MOVES modelers may use these supplemental files to model any month-by-month impacts of the COVID-19 pandemic during year 2020.

5.6.9 Run MOVES to create emission factors

EPA ran MOVES for each representative county using January fuels and July fuels for the range of temperatures spanned by the represented county group and set of months associated with each fuel set (January and July). A runspec generator script created a series of runspecs (MOVES jobs) based on the outputs from Met4moves temperature information for all months of the year. Specifically, the script used a 5-degree temperature bin with the minimum and maximum temperature ranges from Met4moves and used the idealized diurnal profiles from Met4moves to generate a series of MOVES runs that captured the full range of temperatures for the county group for the months assigned to each fuel. The MOVES runs resulted in six emission factors tables for each representative county and fuel month: rate per distance (RPD), rate per vehicle (RPV), rate per hour (RPH), rate per profile (RPP), rate per start (RPS), and rate per hour for ONI (RPHO). After the MOVES runs were completed, the post-processor script Moves2smk converted the MySQL tables into emission factor (EF) files that can be read by SMOKE. For more details on Moves2smk, see the SMOKE documentation [ref 10]. The post-processor scripts are available in 2020nei_or_postprocessing_jars.zip as described in Table -13.

5.6.10 Run SMOKE to create emissions

To prepare the NEI emissions, EPA first generated emissions at an hourly resolution using more detailed SCCs than are found in the NEI (i.e., by road type and aggregate processes). The SMOKE-MOVES program Movesmrg performs this function by combining activity data, meteorological data, and emission factors to produce gridded, hourly emissions. EPA ran Movesmrg for each of the sets of emission factor tables (RPD, RPV, RPH, RPP, RPS, and RPHO). During the Movesmrg run, the program used the hourly, gridded temperature (for RPD, RPV, RPH, RPS, and RPHO) or daily, gridded temperature profile (for RPP) to select the proper emissions rates and compute emissions. These calculations were done for all counties and SCCs in the SMOKE inputs, covering the continental U.S., as well as separate runs covering outlying areas (e.g., Alaska and Hawaii).

The emissions processes in RPD model the on-roadway driving emissions. This includes the following emission processes: vehicle exhaust, evaporation, evaporative permeation, refueling, brake wear, and tire wear. For RPD, the activity data is monthly VMT, monthly speed (i.e., SMOKE variable of SPEED), and hourly speed distributions (i.e., SPDIST in SMOKE). The SMOKE program Temporal takes temporal profiles specific to vehicle type and road type and distributes the monthly VMT to day of the week and hour. Movesmrg reads the speed distribution data for that county and SCC and the temperature from the gridded hourly (MCIP) data and uses these values to look-up the appropriate emission factors (EFs) from the representative county's EF table. It then multiplies this EF by temporalized and gridded VMT for that SCC to calculate the emissions for that grid cell and hour. This is repeated for each pollutant and SCC in that grid cell. The default diurnal and weekly VMT temporal profiles are based on StreetLight telematics data.

The emission processes in RPV model the parked or "off-network" emissions other than exhaust emissions from vehicle starts. This includes evaporative and evaporative permeation emission processes. For RPV, the activity data is vehicle population (VPOP). Movesmrg reads the temperature from the gridded hourly data and uses the temperature plus SCC and the hour of the day to look up the appropriate EF from the representative county's EF table. It then multiplies this EF by the gridded VPOP for that SCC to calculate the emissions for that grid cell and hour. This repeats for each pollutant and SCC in that grid cell.

The emissions processes in RPH model the parked emissions for combination long-haul trucks (source type 62) that are hoteling. This includes the following modes: extended idle and APUs. For RPH, the activity data is monthly hoteling hours. The SMOKE program Temporal takes a temporal profile and distributes the monthly hoteling hours to day of the week and hour. Movesmrg reads the temperature from the gridded hourly (MCIP)

data and uses these values to look-up the appropriate emission factors from the representative county's EF table. It then multiplies this EF by temporalized and gridded HOTELING hours for that SCC to calculate the emissions for that grid cell and hour. This is repeated for each pollutant and SCC in that grid cell.

The emission processes in RPP model the parked emissions for vehicles that are key-off. This includes the mode vehicle evaporative (fuel vapor venting). For RPP, the activity data is VPOP. Movesmrg reads the gridded diurnal temperature range (Met4moves' output for SMOKE). It uses this temperature range to determine a similar idealized diurnal profile from the EF table using the temperature min and max, SCC, and hour of the day. It then multiplies this EF by the gridded VPOP for that SCC to calculate the emissions for that grid cell and hour. This repeats for each pollutant and SCC in that grid cell.

In MOVES3, the emission processes in rate-per-start (RPS) are separated from RPV emissions, unlike in MOVE2014. The RPS emissions include start exhaust and crankcase start exhaust emissions.

A new process in MOVES3 called rate-per-hour-off network idling (RPHO) represents emissions that occur idling during deliveries and the pick-up and drop-off of passengers.

The result of the Movesmrg processing is hourly data as well as daily reports for each of the processing streams (RPD, RPV, RPH, RPP, RPS, and RPHO). The results include emissions for every county in the continental U.S.

5.6.10.1 Spatial Surrogates

For the onroad sector, the on-network (RPD) emissions were spatially allocated differently from other off-network processes (e.g., RPV, RPP, RPHO). Surrogates for on-network processes are based on AADT data and off network processes (including the off-network idling included in RPHO) are based on land use surrogates as shown in Table 5-11. Emissions from the extended (i.e., overnight) idling of trucks were assigned to surrogate 205, which is based on locations of overnight truck parking spaces. The total of the gridded emissions for each county and hour are summed to develop the NEI.

Table 5-11: Off-network Mobile Source Surrogates

Source type	Source Type name	**Surrogate ID**	**Description**
11	Motorcycle	307	NLCD All Development
21	Passenger Car	307	NLCD All Development
31	Passenger Truck	307	NLCD All Development
32	Light Commercial Truck	308	NLCD Low + Med + High
41	Other Bus	306	NLCD Med + High
42	Transit Bus	259	Transit Bus Terminals
43	School Bus	508	Public Schools
51	Refuse Truck	306	NLCD Med + High
52	Single Unit Short-haul Truck	306	NLCD Med + High
53	Single Unit Long-haul Truck	306	NLCD Med + High
54	Motor Home	304	NLCD Open + Low
61	Combination Short-haul Truck	306	NLCD Med + High
62	Combination Long-haul Truck	306	NLCD Med + High

5.6.11 Post-processing to create an annual inventory

For the purposes of the NEI, EPA needed emissions data by county, SCC, and pollutant. EPA ran SMOKE-MOVES at a more detailed level including road type and emission processes (e.g., extended idle) and summed over road types and processes to create the more aggregate NEI SCCs. EPA developed and used a set of scripts to combine the emissions from the six sets of reports and from all days to create the annual inventory. The post processing scripts are named aq_cb6_saprc_20220825 and nata_20220825. They are available in the documentation (see Section 5.8).

Five speciated $PM_{2.5}$ pollutants (i.e., PEC, POC, PNH4, PSO3, and PMFINE) were added to the NEI data for summary purposes. Note that air quality modeling uses a finer breakdown of these pollutants. DIESEL-PM10 and DIESEL-PM25 were also added by copying the PM_{10} and $PM_{2.5}$ pollutants (respectively) as DIESEL-PM pollutants for all diesel SCCs. See Section 5.6.1 for more details.

5.7 Summary of quality assurance methods

EPA performed a series of checks and comparisons against both the inputs and the resulting emissions to quality assure the onroad inventory. These checks are in addition to the ones described on the underlying CDBs. The following is a list of the more significant checks that were performed:

- Review of IHS data prior to becoming EPA Defaults
    - EPA identified missing motorhomes and motorcycles, as wells misclassified trucks.
    - IHS provided additional data to correct this.
- Review of StreetLight data prior to becoming EPA Defaults
    - EPA generated plots of monthly/daily/hourly VMT and average hourly speeds, comparing trends by month, vehicle type, hour, and day type. EPA performed limited gap-filling where necessary.
    - EPA reviewed summaries of gap-filling required and identified that StreetLight's delivery was missing partial commercial truck data in several hours of certain months. EPA made appropriate month substitutions to remedy the problem.
    - EPA determined month VMT distributions from StreetLight for the "personal" vehicles matched expectations for 2020, as well as published nationwide FHWA VMT trends. However, for the commercial trucks, EPA made an adjustment to the month of May to remove an artificial spike in that month's VMT. EPA filled May commercial truck VMT fractions by interpolating April and June. The May spike in the raw commercial truck data was due to natural variation in the StreetLight data sample size, not a data error.
- Review of S/L/T agency MOVES inputs
    - EPA created plots of age distributions to check that the distributions looked like year 2020 and that population totals reasonably matched with IHS; where discrepancies existed, EPA contacted the S/L/T agency for clarification of the registration data year (age distributions) and/or revised population estimates.
    - EPA reviewed plots of month VMT fractions to determine whether a clear pandemic effect could be seen in months starting in March/April. Where submittals did not show this, EPA Default (StreetLight) data were used by state.
    - Previously discussed QA script and findings (Section 5.5.2).

- The 2020 NEI emissions were compared to the 2017 and 2019ge emissions to make sure that all SCCs, counties, and pollutants were covered and as a general quality assurance of the emissions.

- Comparisons of 2020 with 2017 and 2019ge emissions were done using spreadsheets that compared emissions from the three years using various groupings, including but not limited to county-level, the first 6 digits of the SCC (fuel + MOVES source type), and grouping by light-duty and heavy-duty.
- Maps of county-level CAP and select HAP emissions were prepared for each MOVES source type and rate (e.g., RPD), including maps of the difference between 2020 emissions versus 2017 and 2019ge emissions.

The maps and spreadsheets helped to identify areas with suspect activity data or emission factors, and EPA followed up on any suspect areas to investigate further and resolve problems if any were found. Folders containing a number of QA maps, plots, and summaries are referenced as part of the supporting data in Table -13.

5.8 Supporting data

Onroad 2020 emissions were developed by EPA primarily using input data submitted by state and local agencies and secondarily using EPA-developed input data, except for the state of California where California-provided emissions were used for most pollutants. Table 5-12 provides the submittal history of these county databases. The onroad scripts and data files used in the calculations are listed in Table -13. The files and datasets listed in are available on the 2020 NEI Supplemental Data FTP site.

Table 5-12: Agency submittal history for Onroad Mobile Inputs and emissions

Agency Organization	Onroad CDB Submission Date (MM/DD/YYYY)	Onroad Emissions Submission Date (MM/DD/YYYY)	Notes
Alaska Department of Environmental Conservation	02/03/2022		
Connecticut Bureau of Air Management	01/03/2022		
California Air Resources Board		04/06/2022	
Coeur d'Alene Tribe		01/21/2022	
Department of Energy and Environment (Washington D.C.)	01/06/2022		
Delaware Department of Natural Resources	02/18/2022		
Florida Department of Environmental Protection	02/04/2022		
Georgia Department of Natural Resources	09/17/2021		
Idaho Department of Environmental Quality	10/27/2021		
Illinois EPA	12/20/2021	12/18/2021	
Knox County (Tennessee) Department of Air Quality Management	01/21/2022		

Agency Organization	Onroad CDB Submission Date (MM/DD/YYYY)	Onroad Emissions Submission Date (MM/DD/YYYY)	Notes
Kootenai Tribe of Idaho		01/25/2022	
Louisville (Kentucky) Metro Air Pollution Control District	12/22/2021		
Maine Department of Environmental Protection	01/20/2022		
Maricopa County (Arizona) Air Quality Department	12/16/2021		
Maryland Department of the Environment	01/03/2022		
Massachusetts Department of Environmental Protection	02/06/2022		
New Hampshire Department of Environmental Services	12/20/2021		
New Jersey Department of Environment Protection	01/19/2022		
New York Department of Environmental Conservation	01/21/2022		
North Carolina DEQ, Division of Air Quality	01/31/2022		
Northern Cheyenne Tribe		11/15/2021	
Ohio EPA	04/05/2022		
Oregon Department of Environmental Quality	03/24/22		
Nez Perce Tribe		01/25/2022	
Pennsylvania Department of Environmental Protection	02/03/2022		
Pima Association of Governments (Tuscon, Arizona)	01/18/2022		
Rhode Island Department of Environmental Management	02/09/2022		EPA constructed the Rhode Island CDBs from spreadsheets provided by RIDEM.
Shoshone-Bannock Tribes of the Fort Hall Reservation of Idaho		01/25/2022	
South Carolina Department of Health and Environmental Control	01/11/2022		

Agency Organization	Onroad CDB Submission Date (MM/DD/YYYY)	Onroad Emissions Submission Date (MM/DD/YYYY)	Notes
Tennessee department of Environmental Conservation	01/31/2022		
Texas Commission on Environmental Quality	12/28/2021		
Utah Division of Air Quality	02/14/2022		
Vermont Department of Environmental Conservation	01/31/2022		
Virginia Department of Environmental Quality	01/14/2022		
Washington State Department of Ecology	02/03/2022		
Washoe County (Nevada) Health District, Air Quality Management Division	02/02/2022		
West Virginia Division of Air Quality	01/05/2022		
Wisconsin Department of Natural Resources	01/22/2022		

Table 5-13: Onroad Mobile data file references for the 2020 NEI

	File Name	Description
1	NEI2020_default_onroad_activity_approach.docx	Describes method used for EPA default VMT, VPOP, data used in counties for which data were not submitted by S/L/T agencies.
2	Folder *CDBs_for_all_counties* contains 2020_CDBs_stateXX.zip where XX is the two-digit state FIPS code	"*Unseeded*" CDBs for all counties in the U.S. archived separately by state. These may not produce fully populated emission rates tables across all categories without "seeding". Activity data and age distributions are specific to each county and not aggregated.
3	Folder *CDBs_for_rep_counties* contains 2020_RepCDBs_Seeded_12oct2022.zip	"*Seeded*" CDBs for representative counties in the continental U.S. used to develop 2020 NEI. These should produce fully populated rates tables because values of zero in the MOVES input tables have been updated to small numbers (1e-15). Age distributions and AVFT are vehicle-population-weighted across all represented counties. VMT and population are summed across all represented counties.

	File Name	Description
4	Folder *CDBs_for_rep_counties* contains 2020_RepCounty_Runspecs.zip	The MOVES3 run specifications (runspecs) for the representative counties for running MOVES in emissions rate mode for SMOKE-MOVES.
5	Folder *CDBs_for_rep_counties* contains 2020_RepCounty_ZMH_Databases.zip	The input databases containing the meteorology (ZoneMonthHour) table for each MOVES runspec.
6	2020NEI_onroad_activity_final_20230112.zip	All three data types are in FF10 format for SMOKE and are a combination of EPA estimates, agency submittals, and corrections: 1. Vehicle population by county and SCC covering every county in the U.S., 2. VMT annual and monthly by county and SCC covering every county in the U.S., and 3. Hoteling hours annual and monthly by county covering every county in the U.S. including hours of extended idle and hours of auxiliary power units for combination long-haul trucks only. 4. Off-network idle hours by county and SCC. 5. Starts by county and SCC.
7	2020NEI_RepCounty_Temperatures.zip	The temperature and relative humidity bins for running MOVES to create the full range of emissions factors necessary to run SMOKE-MOVES and the ZMH files used to run MOVES. Generated by running the SMOKE Met4moves program.
8	MFMREF_2020NEI_28jul2022_v0	Fuels cross reference (MFMREF) is a table that maps representative fuel months to calendar months for each representative county. The MFMREF file is an input to SMOKE.
9	MCXREF_2020NEI_28jul2022_v0 representative_county_groups_2020nei_final.png	County cross reference file (MCXREF) is a table that shows every US county along with the representative county used as its surrogate. The MCXREF is an input to SMOKE. A map showing the county groups is also available.

	File Name	Description
10	2020NEI_spdist.zip	These data are in FF10 format for SMOKE and are a combination of EPA estimates, agency submittals, and corrections: 1. Average speed in miles per hour, annual and monthly values, by county and SCC covering every county in the U.S. and 2. Weekend and weekday hourly speed distributions (SPDIST) in miles per hour, by county and SCC covering every county in the U.S.
11	The archive *or_scripts_2020.zip* includes the FF10 generation scripts: 1_CreateFF10database_20220331.sql 2_PopulateFF10_fromMOVES3CDB_v0_20220331.sql	FF10 generation scripts read CDB tables and produce SMOKE-formatted activity input files for use in SMOKE-MOVES. The SMOKE activity files include VMT, vehicle population, hoteling hours, and starts. However, for the 2020 NEI, only VMT and population were extracted from the CDBs.
12	The archive *or_scripts_2020.zip* contains the script ReverseFF10_Script_20230118.plx	The reverse FF10 script populates CDBs from SMOKE-formatted activity files VMT, vehicle population, and hotelling hours to fill the MOVES CDB tables SourceTypeYearVMT, SourceTypeYear, HotellingHours, HotellingActivityDistribution, HotellingMonthAdjust, IdleMonthAdjust, startsMonthAdjust, monthVMTFraction, and roadtypeDistribution.
13	Folders with QA / review products: *age_distribution_plots* *streetlight_plots* *draft_NEI_onroad* *emissions_and_activity_maps* *summaries*	Plots, maps, and summaries for quality assurance and data visualization are available in several folders to assist interested parties in better understanding the data.
14	2020_Documentation_of_CDB_-Input_Data_20230118.xlsx	Spreadsheet that shows how state-submitted and default data were merged together to prepare 2017 NEI.
15	2020_Representative_Counties_Analysis_20220720.xlsx	Spreadsheet of representative county characteristics.
16	2020_StreetLight_Grouping_Decision_Charts.docx	Documentation showing process to group data behind the VMT distributions and speed distributions.

	File Name	Description
17	2020nei hotelling by county versus truck stop parking.xlsx	Spreadsheet documenting computation of adjustment factors applied to hoteling hours where there were more hours assigned than the available truck stop parking spaces could support.
18	The archive *2020nei_or_postprocessing_jars.zip* includes the scripts postprocess_aq_cb6_saprc_20220825.jar postprocess_nata_20220825.jar postprocess_invmode_speciation_20210519.jar	MOVES lookup table post-processing scripts that can create emission factor tables for various chemical mechanisms and purposes (e.g., the NEI).
19	The archive *or_scripts_2020.zip* includes the script and meteorological data tables: UpdateMet_and_Fuels_20230117.plx 2020nei_month_hour_for_nonroad_rerun	Perl script that inserts met data into set of "all CDBs" intended for inventory mode. The representative CDBs do not use this data. The 2020 met data is listed in the MySQL database `2020nei_month_hour_for_nonroad_rerun ` and is the same ZoneMonthHour table used in nonroad. This script also replaces any existing fuel supply, formulations, and E85 usage fractions with MOVES3 defaults.
20	The archive *or_scripts_2020.zip* includes the representative county seeding scripts: SeedingScript_ERG.sql `seed` seedCDBs.py	These items can be used to seed a set of representative CDBs so that they produce complete lookup tables. SeedingScript_ERG.sql is a MySQL script that turns 0 values into small values of 1e-15. The MySQL database `seed` is required by the script. The python script seedCDBs.py is a wrapper to run the MySQL script "SeedingScript_ERG.sql" on a batch of CDBs. This script also updates the version of the CDB name to the current date (YYYYMMDD format). The CDB naming convention is `c01015y2017_YYYYMMDD` for county 1015 calendar year 2017.
21	2017NEI_California_onroad_HAP_augmentation_factors.csv	Factors used to augment the California Air Resources Board submitted criteria pollutant data with HAPs.
22	The archive *MOVES_Input_DBs.zip* includes databases LEV\lev_XX_20220824 (where XX is the two-digit state ID) and nonoxadj_moves3	Databases used when running MOVES include LEV* that represents where California LEV rules apply and nonoxadj_moves3.zip which causes MOVES not to make humidity-based adjustments to NOx emissions, so that they can instead by applied using hourly, grid-cell based humidity values.

	File Name	Description
23	Moves_hourvmtfraction_monthly_streetlight_stateXX.csv Moves_avgspeeddistribution_monthly_streetlight_stateXX.csv (where XX is the two-digit state ID)	Due to large file sizes, these files are not posted to EPA's FTP site. Please contact the EPA NEI team to request these files (Godfrey.janice@epa.gov).

5.9 References for onroad mobile

1. Coordinating Research Council. 2019. Developing Improved Vehicle Population Inputs for the 2017 National Emissions Inventory. Report No. A-115.
2. Coordinating Research Council. 2017. Improvement of Default Inputs for MOVES and SMOKE-MOVES: Final Report. Report No. A-100.
3. U.S. EPA, Tools to Develop or Convert MOVES Inputs, LEV and early NLEV modeling information for MOVES2014-20141022.
4. U.S. EPA, MOVES3: Latest Version of MOtor Vehicle Emission Simulator (MOVES).
5. U.S. EPA, MOVES Onroad Technical Reports.
6. The Weather Research & Forecasting Model, Jatin Kala, Louis Marelle, J. Shpund, Jordan Schnell, Robert Gilliam, Tim Juliano, and Maria Frediani, National Center for Atmospheric Research, Mesoscale and Microscale Meteorology Division, Boulder CO, August 2022.
7. Meteorology-Chemistry Interface Processor (MCIP) version 5.3.3.
8. User's Guide for SMOKE, including MOVES integration tools.
9. Federal Highway Administration, Highway Statistics 2020.
10. Scripts that interface between SMOKE and MOVES, MOVES Utility Scripts and SMOKE-MOVES.

United States Environmental Protection Agency	Office of Air Quality Planning and Standards Air Quality Assessment Division Research Triangle Park, NC	Publication No. EPA-454/R-23-001e January 2023