Model Evaluation

David Choi, Darrell Sonntag, and James Warila
MOVES Review Work Group
March 1, 2017
Agenda

• Background on MOVES model evaluation
• Context of current MOVES evaluation
• MOVES2014a comparisons to
  – Inspection/Maintenance (I/M) & remote sensing data (RSD)
  – Tunnel studies
• Summary
MOVES Evaluation

• Why?
  – A key recommendation in the National Research Council’s review of EPA’s mobile source modeling program\(^1\)
  – A key element of EPA’s quality assurance guidance for developing models\(^2\)
  – A critical component of EPA’s development and upkeep of MOVES

• Objectives
  – To assess model performance in accurately estimating current emission inventories and forecasting emission trends
  – To identify areas in clear need of improvement
  – To guide future work and research needs
MOVES Evaluation (cont’d)

• Priorities
  – Major sources of emissions (e.g., light-duty gasoline, heavy-duty diesel)
  – Areas where significant independent data/studies available

• Assessment
  – If systematic bias is observed across multiple data sources, it is indicative of model underperformance
  – If the model predictions are generally within the variability of independent measurements, it gives confidence that the model is predicting real-world emissions reasonably well

• Improper means of evaluation
  – Comparisons against measurements based only on a few vehicles
  – Not sufficiently customizing MOVES inputs to account for the measurement conditions (i.e., fleet composition, vehicle activity)
Types of Evaluation

• Emission rates
  – Using dynamometer, RSD, and PEMS measurements
    • Large samples with best chance to capture rare high emitters
    • Known operating conditions (i.e., pre-conditioned IM240 drive cycle)
  – Comparing MOVES predictions to such measurements is the most controlled comparison
    • Activity and fleet variables such as vehicle mix and vehicle age are known for a given study
    • Eliminates sources of significant variability inherent in comparisons to ambient monitor data, and even in tunnel and roadside measurements
Types of Evaluation (cont’d)

• “Localized composite” emissions
  – Using composite emission measurements from tunnel or roadside emission monitors where
    • Vehicle emissions are predominant
    • Vehicle activity and fleet mix can be accounted for to some degree
  – Provides a snapshot of overall model performance, for the narrow operating conditions represented at the specific location

• Regional air quality
  – Evaluation of air quality model results (CMAQ) vs. air quality monitor data

• “Macro-scale” fuel consumption
  – Comparison of “bottom-up” fuel consumption to “top-down” fuel tax data
FUEL TAX DATA

Regional Onroad Emissions

Start, Evap Emissions

Localized Composite Emissions

VMT, population, activity estimates

operating mode & fleet mix estimates

Emission Rates

AIR QUALITY MONITOR DATA

CMAQ

Other Sector Emissions

TUNNEL, ROADSIDE MONITOR DATA

DYNO, RSD, PEMS DATA
History

• EPA’s evaluation work on MOVES began with MOVES2004, focused on fuel consumption

• For MOVES2010a, we evaluated model performance using several methods and found that:

  – Emission rate comparisons against multiple data showed no systematic bias for both light-duty and heavy-duty vehicles

  – Tunnel comparisons showed

    • MOVES predictions were higher than the observed for LD, but MOVES compared well for HD
    • MOVES trends over time are consistent with observations

<table>
<thead>
<tr>
<th>Evaluation Type</th>
<th>Analysis</th>
</tr>
</thead>
</table>
| Emission Rates       | **Light-Duty**
                       | Atlanta RSD
                       | CRC E-23 Chicago RSD
                       | Chicago I/M Dyno
                       | Kansas City Study Dyno
                       | NCSU PEMS (NC State) |
|                      | **Heavy-Duty**
                       | CRC E-55 Dyno
                       | HD in-use compliance
                       | Houston drayage       |
| Localized Composite  | Caldecott Tunnel - range analysis
                       | Van Nuys Tunnel (Fujita, et. al)
                       | Borman Expressway     |
| Fuel                 | FHWA Fuel Sales 2000-2007                                      |
• Several recent studies\textsuperscript{3,4} have shown differences between air quality model estimates and monitored values for nitrogen oxides suggesting AQ models appear to overestimate NOx

• Staff across EPA are investigating various aspects of the issue
MOVES is just one complex part of the modeling system:
NOx Evaluation Efforts for MOVES2014a

• Focus on light-duty gasoline passenger cars and trucks
  – Most evidence\(^5\) suggests that MOVES under-predicts NOx for HD diesel

• Focus on running exhaust emissions
  – Due to lack of significant sources of independent data for start emissions
  – Running exhaust emissions contribute over 80% of NOx emissions from onroad gasoline and diesel
Comparison to Denver I/M Data

• **GOAL:** compare MOVES **BASE RATES** to external data
  – Taken from input database
    • No modifications or adjustments (humidity, I/M compliance, etc.)

• **SCOPE:** running emissions for
  – Light-duty cars and trucks
    • Tier 2 vehicles (in MY 2010-2016)
      – Bins 8, 5, 4, 3, 2
    • Tier 1 cars (in MY 1996-2000)

• **BASIS:** NO\textsubscript{X} emissions on IM240 cycle
  – Denver I/M: measurements
  – Using random sample
    • CY 2008-2015
  – MOVES2014a: simulate IM240 using modal rates
    • Average by age
Denver I/M Data (cont’d)

- Light-duty cars
  - Tier 2 (Bin 5 and equivalent) meeting 70 mg/mi NOx FTP standard
- Distribution spans over 3 orders of magnitude

Note the log scale
Preliminary Results for Tier 2 Cars: MOVES2014a Rates vs. Denver I/M

Tier 2 Passenger Cars

MOVES:
Simulated IM240 by age, for MY 2010-2016

Denver:
Mean IM240 by age, for “Bin-5” (70 mg/mi NOx FTP std)

MOVES rates appear lower than corresponding I/M results.
Preliminary Results for Tier 1 Cars: MOVES2014a Rates vs. Denver I/M

Tier 1 Passenger Cars

MOVES:
Simulated IM240 by age, for MY 1996-2000

Denver:
Mean IM240 by age, for Tier 1
(600 mg/mi NOx FTP std)

MOVES rates appear higher than corresponding I/M results.
Limitations & Areas for More Work

• **Sample sizes (for each age level)**
  – T1: 10 – 370 vehicles
  – T2: 240 – 2,460 vehicles
  – Larger samples probably give a more representative comparison

• **Fuel properties**
  – Data collected over 8 years
  – Fuels changing over time

• **Temperature**
  – Don’t expect effect for hot-running operation

• **Altitude** (adjust if appropriate)

• **Potential positive bias due to “clean screen”**
  – Vehicles screened by remote sensing
  – “Clean” vehicles exempted from inspection
Comparisons to Remote Sensing Data

- University of Denver collected a series of remote sensing data, funded by Coordinating Research Council
  - Measurement sites in Arizona, California, Colorado, Illinois, Maryland, Nebraska, Nevada, Pennsylvania, Texas, Oklahoma, Utah, and Washington
  - Typically collected at on-ramps during weekdays
- Remote sensing measured the ratios of CO, HC, NO*, to CO2 in the exhaust and reported the percent concentrations of pollutants
- RSD databases include
  - Measurement conditions (i.e., speed, acceleration, temperature, and humidity)
  - Vehicle information (i.e., Vehicle Identification Number (VIN))
  - Flags for invalid measurements

* Recent RSD data include NO2 concentrations, as well as NO concentrations
RSD Data

- Current analysis includes RSD data that were collected over multiple years at the same location
  - Phoenix, AZ, Denver, CO, Chicago, IL, and Tulsa, OK
  - In calendar years 1999 to 2007 and 2013 to 2015
  - Total number of measurements: ~400,000

<table>
<thead>
<tr>
<th>RSD Sites</th>
<th>Number of Measurements (light-duty cars and trucks combined)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phoenix, AZ</td>
<td>95,266</td>
</tr>
<tr>
<td>Chicago, IL</td>
<td>107,007</td>
</tr>
<tr>
<td>Denver, CO</td>
<td>127,518</td>
</tr>
<tr>
<td>Tulsa, OK</td>
<td>64,658</td>
</tr>
</tbody>
</table>
MOVES Runs

- **MOVES project scale** used to simulate the measurement conditions, as much as possible

- County inputs include:

<table>
<thead>
<tr>
<th>Input</th>
<th>Time &amp; Location-Specific</th>
<th>MOVES Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating Mode Distribution</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Age Distribution</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Fuel Properties</td>
<td>X*</td>
<td></td>
</tr>
<tr>
<td>Meteorology</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Inspection/Maintenance</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

*With the exception of sulfur, MOVES defaults were used for all fuel properties.*

- Pollutants – nitric oxide (NO) and total energy consumption
Project Scale vs. MOVES National Scale

- MOVES national scale runs using the default inputs result in significantly higher emissions than the project scale runs
  - MOVES can show clear over-prediction when not properly modeled
- Highlights the importance of modelling the measurement conditions as much as possible using the project scale when evaluating MOVES

![Graph: LDV – Chicago](image-url)
Sample Results – Comparisons to RSD

• Showing illustrative results
  – Only light-duty passenger cars
  – For select calendar years

• Comparisons for light-duty trucks similar to passenger cars

• RSD sites differ in age distributions, operating mode distributions, presence of I/M programs, altitude, etc.
Sample Results: MOVES2014a vs. RSD for CY2013-2015

- MOVES2014a lower for Tier 2 vehicles
- For Tier 1 vehicles, MOVES2014a generally within the variability of the data
Sample Results: MOVES2014a vs. RSD for CY2005

- MOVES2014a lower or within the variability of the data for Tier 2 vehicles
- MOVES2014a higher than RSD for Tier 1 vehicles
Next Steps – Comparisons to RSD

• Analyze other available RSD datasets
• Understand variations between RSD sites and calendar years
• Evaluate fuel consumption in MOVES
  – Since comparisons made in fuel-based emission rates
Tunnel Comparisons

- **Caldecott Tunnel**
  - 1 km long tunnel in Oakland, California
  - 4% uphill grade (eastbound)
  - 3 separate two-lane traffic bores
  - Bore 2 is limited to light-duty vehicles (switches direction with flow of commuter traffic)

- **UC-Berkeley derived fleet-average emission rates from their most recent campaign (2010)**
  - Measured pollution concentrations: 4-6 pm, 8 weekdays, July 2010

*Picture from Dallmann et al. 2013*
MOVES2014a Comparison to Caldecott Tunnel

• MOVES default runs
  – Run at National-scale
• MOVES project-scale used to model the tunnel conditions
  – Created inputs from measurements conditions, e.g.
    • 4% grade
    • CA standards
      – Section 177/LEV inputs for CA standards on 1994+ vehicles
  – Lower, midpoint, and upper bound for uncertain inputs
    • Age distributions
    • Driving cycles
    • Fuel sulfur levels
Tunnel Comparison - Preliminary Results

- **High end estimate**
  - More aggressive driving
  - Older age distribution

- **Midpoint estimate**

- **Low end estimate**
  - Lower fuel sulfur
  - Smoother driving
  - Younger age distribution

- **Default inputs for 2010 national aggregation, urban freeways only**
- **Default inputs for 2010 national aggregation, all roads and processes**

**Caldecott measurements in July 2010 reported by Dallmann et al. 2013**

**EMFAC2014 for Contra Costa County, all roads and processes**

**Data Points:**
- Caldecott
- MOVES project
- MOVES default
- EMFAC

**Oxides of Nitrogen (NOx)**

- Gasoline

**Y-axis:** g/kg-fuel

**X-axis:**
- Caldecott
- MOVES project
- MOVES default
- EMFAC
Tunnel Comparisons - Observations and Limitations

• Observations
  – Key sources of uncertainty for project-level runs
    • For NOx g/kg-fuel: age distributions, LEV inputs
    • For NOx grams: age distributions, LEV inputs, driving cycles
  – In the case of Caldecott Tunnel, using national defaults produced substantially higher emission rates than using project-level inputs

• Limitations
  – Caldecott tunnel gasoline measurements have tended to be lower than other remote sensing studies$^8,9$
  – MOVES data is not based on CA vehicles or fuels, e.g.
    • Section 177/LEV inputs do not account for differences in CA and National vehicle program for pre-1994 vehicles
• When doing comparisons to RSD and tunnel measurements, it is important to properly model the measurement conditions

• We will be continuing and refining our comparison of MOVES2014a to I/M, RSD, and tunnel measurements

• Additional work exploring other aspects of the air quality modeling system is also ongoing
Acknowledgements

• Gary Bishop and Don Stedman at University of Denver
• Coordinating Research Council (E-23, E-106)
• Colorado Department of Public Health and Environment (CDPHE)
• Eastern Research Group (ERG)
• Timothy Dallmann, Brian McDonald and Robert Harley from University of California-Berkeley
References


