Quantification of Uncertainty in MOVES: Issues and Proposal

Chad Bailey, MPH

MOVES Day

November 6, 2002

Overview

- **What do we mean by uncertainty?**
- **Methods under consideration**
  - (Parametric) Bootstrap Simulation
  - Propagation of Error
- **MOVES implementation issues**
  - Empirical Binning
  - PERE-based emission rates
- **Proposal for MOVES GHG**

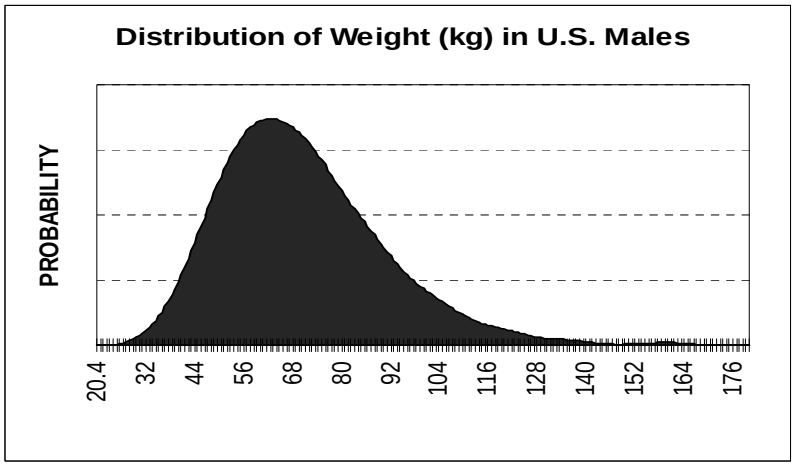
What do we mean by uncertainty?

- **Uncertainty is *not* variability!**
  - Variability is a part of the system (i.e. tailpipe emissions), where uncertainty is what we don't know about the system
  - Both uncertainty and variability can be described by probability distributions
- **Probability distribution around different parameters in the distribution (ie. Mean, standard deviation, percentiles)**

Sources of Uncertainty

- **Scenario**
- **Model structure**
  - Bin definitions
  - Incomplete or incorrect formulation
- **Inputs (parametric)**

Example of Variability vs. Uncertainty (lognormal, mean=70)



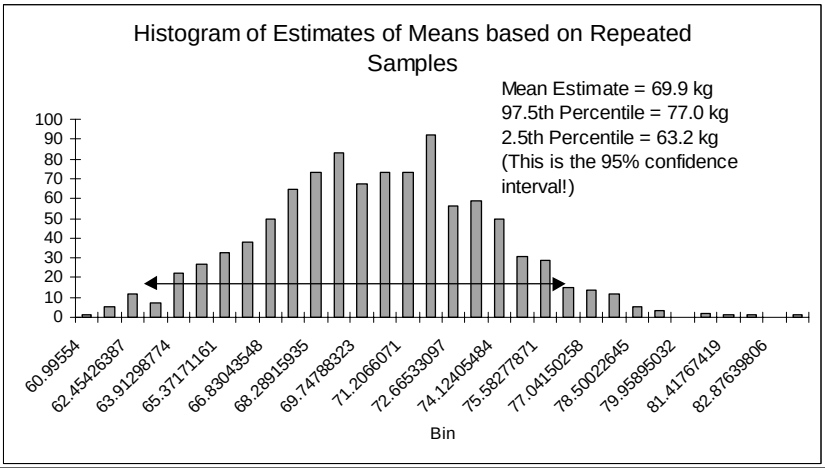
Hypothetical Exercise

- What if we randomly sampled 30 men from the U.S. population? (say, in this room?)
- What could we say about the true mean of the U.S. male population?
- We repeated the sample 500 times, what would it do to our estimate of the mean?

Table of First Ten Samples

Name	Minimum	Mean	Maximum
Sample 1	36.45314	67.49744	96.036
Sample 2	43.92532	69.50186	103.3376
Sample 3	36.65549	74.40438	120.6124
Sample 4	42.31688	74.38654	110.9498
Sample 5	28.87856	69.42949	111.174
Sample 6	40.18633	67.32497	146.2084
Sample 7	44.55801	68.46724	115.0667
Sample 8	41.34025	66.80198	95.56773
Sample 9	40.308	64.6759	115.5476
Sample 10	34.90135	67.77705	115.4475

Means are normally distributed



What does this mean?

- **We can estimate the mean well, but based on a limited sample, we have "uncertainty"**
- **The original distribution we sampled from can be considered the "variability" - based on a lognormal distribution**
- **The distribution of means can be considered the "uncertainty" - it forms a normal distribution (CLT)**

Model Applications

- **Each MOVES input is variable**
  - Empirical Binning: Emission Rates in Bins
  - Physical Model: Parameters (weight, fuel rate)
- **Same concepts apply to characterizing uncertainty in model outputs as the example of weight estimates in U.S. males**

Methods Under Consideration

- (Parametric) Bootstrap Simulation
- Error Propagation

Parametric Bootstrap Simulation

- First, fit a parametric distribution to data in our model inputs (i.e. emission rates within each VSP bin)
- Second, randomly sample values from the input distributions and calculate output
- Repeat to calculate mean and confidence intervals

Pros/Cons of Bootstrap

- **Pros**
  - Consideration of uncertainty when distributions based on very little data
  - Uncertainty associated with inputs with non-normal error (VSP distribution)
  - Allows user-defined inputs to be included
- **Cons**
  - Computationally intense!
  - Consideration of appropriate sample sizes
  - Treatment of averaging times (sensitivity analysis?)

Bootstrap Application

- **Empirical Binning**
  - Fit distribution to input data, repeatedly sample
- **PERE/Physical Model**
  - Assign parametric distribution to model parameters
  - Resample to estimate variability in each VSP bin, fit parametric distribution
  - Bootstrap simulation based on PERE output distributions

Analytical Propagation of Error

- **Basic Theory**
  - Uncertainties in model inputs contribute to output uncertainty in proportionally to sensitivity of outputs to each input…

Analytical Propagation of Error

- **Basic Theory**
  - Sum uncertainties, weighted by sensitivity of outputs to each input

    ε represents uncertainty

$$\varepsilon_{\text{out}}^{\ 2} = \left(\varepsilon_{\text{input1}}\right)^2 \frac{\partial(OUT)}{\partial(\text{input1})} + \left(\varepsilon_{\text{input2}}\right)^2 \frac{\partial(OUT)}{\partial(\text{input2})} + \ldots$$

Propagation of Error (Continued)

- **Frey et al. (2002) recommends using fraction of time spent in each mode as weighting factor (W_i) for uncertainty (U):**

$$U_{total} = \sqrt{\sum_{i}^{n} \left(U_i * W_i\right)}$$

Implementation in MOVES

- **Requirements**
  - Errors must be normally distributed (often not the case with small n and large standard error in mean)
  - Quantify uncertainties associated with each input
  - Quantitative sensitivity analysis to determine weights in error propagation equation (or use alternative weighting such as time)

Pros/Cons of Analytical Propagation

- **Pros**
  - From user perspective, computationally streamlined
  - Simpler to program into model code
- **Cons**
  - Greater front-end analysis
    - uncertainties “hard coded” into model
    - may require greater user front-end analysis
  - Will not allow uncertainties that are not normal to be included in propagation (supplement w/bootstrap?)
  - Activity uncertainty not addressed

Comparison of Methods

	Computer Time (+)	**Computer Time (-)**
Flexibility of Method (+)		**Bootstrap**
Flexibility of Method (-)	**Propagation of Error**	

General Considerations

- **Averaging time for uncertainty analysis?**
  - Does # seconds in VSP bin during a cycle (such as US06) require weighting of 1-Hz uncertainties?
- **Activity uncertainty**
  - Bootstrap simulation
- **Contribution of uncertainty in inventory from non-MOVES inputs?**
  - Travel demand models, macroscale VMT estimates, fuel, I/M effectiveness
  - Requires greater interface with other models

Recommendation

- **MOVES GHG (first implementation)**
  - Error propagation when normality assumptions fulfilled
  - Bootstrap simulation to estimate uncertainty when assumptions of normality violated
  - Model validation should be priority
- **Later MOVES implementations**
  - Explore incorporation of other sources of uncertainty using bootstrap, other methods
  - Regular validation exercises