Calibration and Validation

Questions and Challenges

By

Ziad A. Sabra, P.E., Ph.D.

Sabra, Wang & Associates, Inc.

&

John Halkias, P.E., Ph.D.

FHWA

Calibration and Validation

Why is this subject important to the Agency

- Policies and Decisions
- Operations and Safety
- Impact Assessment
- Build more or less lanes
- Cost/Funding
- Planning correctly (right-of-way impacts)
- Responsibility to the road user

Calibration and Validation

Common Questions

- What is Calibration, Validation and Verification?
- Isn't the Model already calibrated, so why do we need to calibrate it again?
- When do we calibrate models?
- How is Calibration different from Validation and Verification?
- Who is responsible for doing all of this?

Calibration and Validation

Verification vs Validation vs Calibration

- **Verification is the process of determining if the logic that describes the underlying mechanics of the model, as specified by the model developer, is faithfully captured by the computer code**
  - For example, if the model developer specifies that $Y = b_0X + C$, then model verification determines if the computer code computes Y using the relationship $b_0X + C$ and not whether this relationship adequately captures reality or if Y should be equal to something other than the relationship of $b_0X + C$
- **Validation is the process of determining if the model's underlying fundamental rules and relationships, formulated by the model developer, are able to adequately capture the targeted traffic behavior, as specified by the relevant theory and as demonstrated by field data**
  - For example, can the car following, lane changing and gap acceptance rules utilized by the model produce the corresponding capacities, queues, speeds, etc.
- **Calibration is the process of adjusting the model parameters, performed by the analyst, to improve the model's ability to reproduce observed local driver behavior and traffic performance characteristics**

Calibration and Validation

Why Calibrate?

- No single model can be expected to be equally accurate for all possible traffic conditions
- No single model can include the whole universe of variables
  - **Every model must be adapted to local conditions**
- **Every model was developed using a very limited amount of real data**
  - **Traffic data (i.e., vehicle trajectories) is very costly**
  - **Combining traffic characteristics with the human factor (driver behavior) is rarely done**
- All models have suggested default values for model parameters
  - **Rarely would a model be able to produce accurate results for a specific area using these default values**
- Driver Behavior and Road Characteristics Depend on
  - **Location (New York vs Florida)**
  - **Vehicle Characteristics (Horsepower, Size, etc.)**
  - **Weather Conditions (Dry, Wet, Ice, etc.)**
  - **Day or Night**
- Simulation models can adapt and replicate almost any condition if the model parameters are properly adjusted
- What is Considered Realistic and What is Not?

Calibration and Validation

When Do We Calibrate?

- The analyst should ALWAYS perform some calibration tests to ensure that the coded model accurately reproduces local traffic conditions and behavior
- The fundamental assumption of calibration is that the travel behavior models/logic/rules are essentially sound
  - There is no need to verify or validate that they produce the correct delay, speed, queue, travel time, density, flow when they are given the correct inputs
  - The only remaining task for the analyst is to adjust these models slightly so that they correctly predict local traffic conditions
  - Some perform one additional step – validating the calibrated model with untried data to ensure that the calibrated model (with a calibrated parameter set) works for different cases
    - Some call this step Validation

Calibration and Validation

Common Questions (Cont'd)

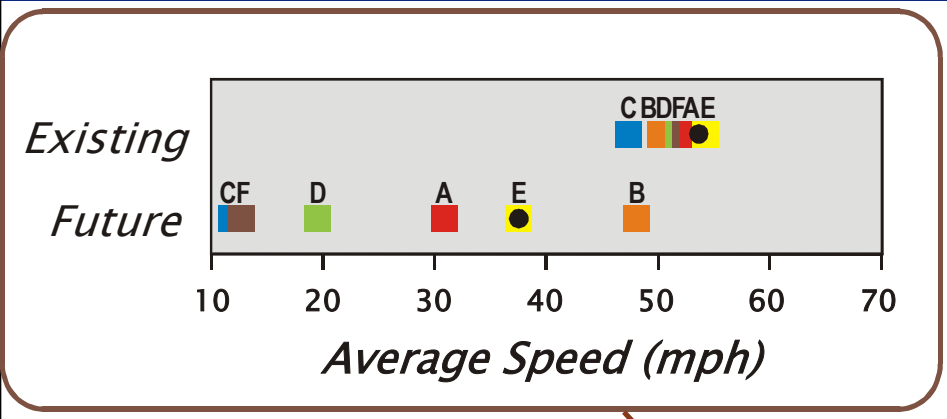
- Should models be calibrated every time we use them, and why?
- How is calibration of macroscopic models different from calibration of microscopic models?
- What if I don't calibrate the model and still use it?
  - What are the consequences

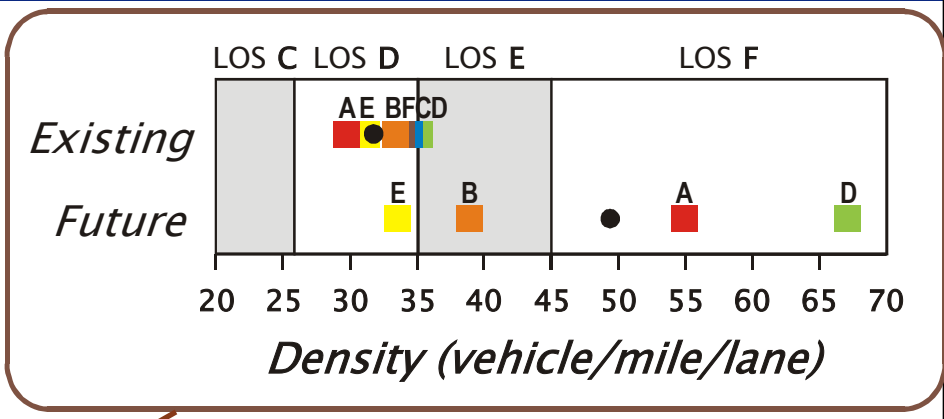
Calibration and Validation

Common Questions (Cont'd)

- Is calibration of a model for existing operation adequate for future traffic volumes and roadway conditions?
- What about calibration for new facility; e.g., Inter-County Connector (ICC) or similar?
- What about calibration for a significant change in an existing facility; e.g., I-95 Section 100 Project or similar?

Freeway Results

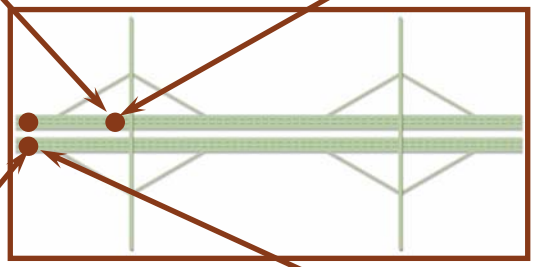


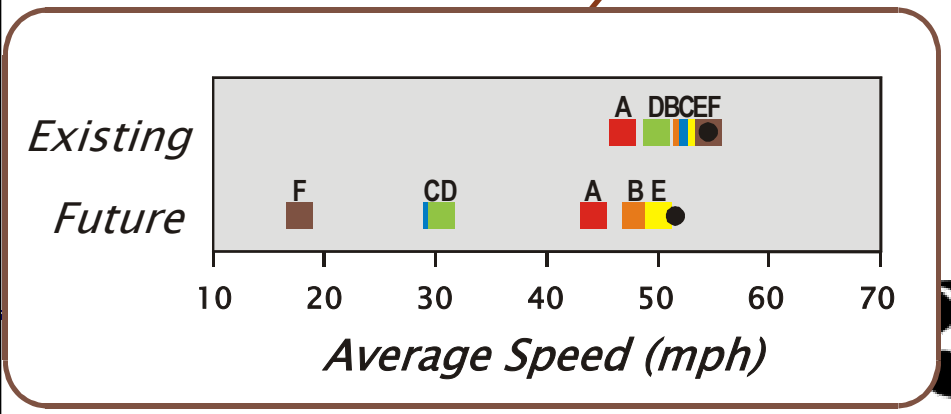


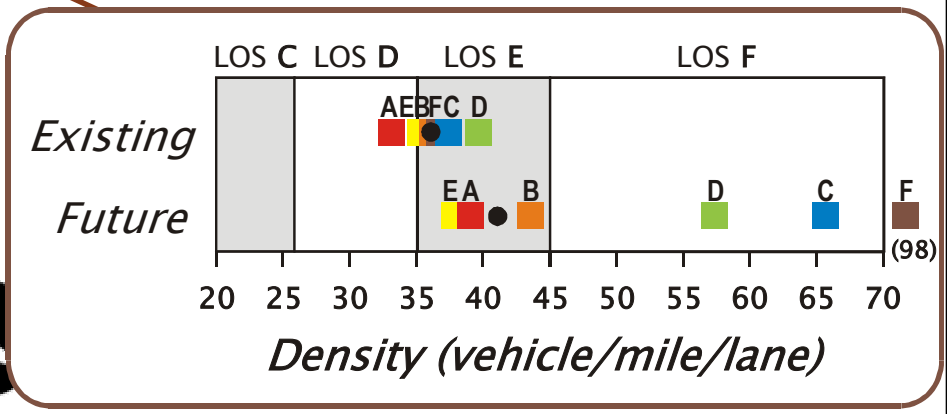
LEGEND

= HCM Chapter 22 Results

= Model Results







Calibration and Validation

Common Questions (Cont'd)

- Is it important that my results match the field results, say within 5 to 10% of actual values, or would 25% for example be acceptable?
- Should statistical analysis be used as part of calibration?

Calibration and Validation

Compare Model to Field Data

- **Calibration of Existing Models**
  - Sometimes Referred to as Existing Condition
- **Base Model is Calibrated When:**
  - Volume, Density and other Operational Observations are Satisfactorily Replicated
  - Statistical Tests Support Such a Determination

Calibration and Validation

Compare Model to Field Data

- **What is Satisfactory?**
  - A statistical test is highly recommended
  - Volumes Match
  - Density/Occupancy Match
  - Speeds Match
  - Field Observations Occur in Model
  - “False” bottlenecks are cleared
- **Post Processor**
- **Spreadsheets**

Calibration and Validation

Statistical Method to Compare Numbers

1. **Observed field data**
   $N_f = 30$, $\mu_f = 34.5$ mph, $S_f = 9.64$ mph
2. **Simulation**
   $N_s = 43$ runs, $\mu_s = 30.5$ mph, $S_s = 9.5$ mph
3. **Perform Hypothesis Test to Compare Means**
   1. **Set up your Hypotheses**
      Ho: $\mu_s = \mu_f$
      Ha: $\mu_s \neq \mu_f$
   2. **Calculate your Test Statistic (Z)**
      $Z = \mu_s - \mu_f / \sqrt{S_s^2/N_s + S_f^2/N_f} = 34.5 - 30.5 / \sqrt{(9.64)^2/30 + (9.5)^2/43}$
      $= 1.75$
   3. **Come up with your Rejection Region**
      Reject Ho if $Z_{calculated} > Z_{critical}$: Is $1.75 > 1.96$? NO
   4. **Conclusion**
      Do not Reject Ho; Not enough evidence to say that the two means are significantly different at the 95% confidence

Calibration and Validation

Common Questions (Cont'd)

- How is Validation different from Calibration?
- When do we validate models?
- Should models be validated every time we use them, and why?
- How is validation of macroscopic models different from validation of microscopic models?

Calibration and Validation

Common Questions (Cont'd)

- **Has the model been verified?**
  - If this means is the logic that describes the underlying mechanics of the model faithfully captured by the computer code?
    - Then the answer is all models when they are developed should have been verified
- **Has the model been validated?**
  - If this means are the model's underlying fundamental rules and relationships (driver behavior algorithms and logic) able to adequately capture the targeted traffic behavior, as specified by the relevant theory and as demonstrated by field data?
    - Then the answer is all models when they are developed should have been validated
  - However, if this means are the results from the model valid? Then it's a different ballgame. Valid may mean:
    - Do the results of capacities, queues, speeds, etc. follow the rules utilized by the model? - The answer is probably yes
    - Do the results reproduce or emulate real-world conditions? The answer is probably no
    - Since the model has been verified and validated, shouldn't the results match field conditions in my area? The answer is absolutely NO!

Calibration and Validation

Common Questions (Cont'd)

- Since this model has been verified and validated and has been used for a long time can you provide me with a formal certificate of the Model's validation and calibration?

Calibration and Validation

Calibration Performance Measures

- Throughput
- Bottleneck Emulation
- Speeds
- Travel Time
- Cycle Failure
- Lane Utilization
- Queues and Queue Residual

Calibration and Validation

Calibration Parameters

Macroscopic/Deterministic Models

- Saturation flow rate for throughput
- Speed especially for signal timing and TSDs

Calibration and Validation

Calibration Parameters

Microscopic Models

- Lane Change
- Gap Acceptance (Headway, Driver Behavior, Queue Discharge, etc.)
- Car following
- Speed

Where do I start and what variables within the parameters do I start with?

Calibration and Validation

What Parameters Do We Calibrate?

- **Analyst can easily get trapped in a never-ending circular process – fixing one problem only to find a new one pops up somewhere else**

- **Divide model parameters into two basic categories**
  - Parameters that analyst is reasonably certain and does not wish to adjust
  - Parameters about which the analyst is less certain and is willing to adjust
  - Analyst should avoid adjusting parameters about which he/she has no information as to the appropriate meaning or value

Calibration and Validation

What Parameters Do We Calibrate?

- Further divide the "adjustable" model parameters into:
- Those that directly impact capacity – calibrate these first
  - Freeways
    - Mean headway or car-following sensitivity factor
    - Driver reaction time
    - Critical gap for lane changing
    - Minimum separation under stop and go conditions
  - Signals
    - Queue discharge headway
    - Start up lost time
    - Gap acceptance for unprotected left turns
- Those that directly impact route choice – calibrate these second
  - The specific route choice algorithm parameters will vary by software. They generally relate to the driver's awareness of, perception of, and sensitivity to travel time, delay, and the cost of alternate routes
  - Usually involve weightings placed on the actual cost and travel time for each route. Additional parameters may be related to the familiarity of the driver with each route and the amount of error in the driver's perception of the cost and time for each route
- System performance calibration – calibrate these last

Calibration and Validation

What Parameters Do We Calibrate?

- **Adjustable parameters can be further subdivided into**:
  - Those that affect the simulation on a global basis – calibrate these first
    - Sample Global Parameters
      - Vehicle Entry Headway
      - Fleet Composition
      - Driver Behavior
  - Those that affect the simulation on a more localized basis – calibrate these "local "link-specific" parameters second to fine-tune the results
    - Sample Link Parameters
      - Free Flow Speed
      - Warning Sign Locations
      - Mean Start Up Delay

Calibration and Validation

Step 1: Capacity Calibration

- **Adjust the global and link-specific "capacity-related" parameters to best replicate local field measurements of capacity**
  - Important because capacity has a significant effect on predicted performance (i.e., delay and queues)
  - Focus on critical counts at the bottlenecks to ensure that the model reproduces these counts correctly
    - If model does not show congestion at the same bottleneck(s) as exist in the field, demands should be adjusted to force the creation of congestion at those bottlenecks
    - If network/facility under study is not congested, capacity values cannot be obtained in the field
      - The capacities for these "potential" bottlenecks can be estimated using HCM procedures
- **Even after global calibration is done satisfactorily, there will be still some locations where model performance deviates a great deal from field conditions**
  - Fine-tune the predicted capacity to match the location-specific measurements of capacity as closely as possible
    - Need to account for roadside factors that affect capacity but are not typically coded in the model (e.g., on-street parking, multiple driveways, narrow lanes or shoulders, etc.)
    - Some models have link-specific capacities (headways)
    - Link-specific adjustments should be used sparingly and carefully because they are not behavior based and are fixed adjustments that will be carried thru all future analyses

Calibration and Validation

Step 2: Route Choice Calibration

- With the correct capacities, the model should no longer have bottlenecks at the incorrect locations with too low or too high discharge rates
- The upstream and downstream traffic volumes on each of the links should now better match the observed counts
- Any remaining differences should be the result of route choice errors in the model
- Therefore, the purpose of route choice calibration is to better match the observed counts at the non-bottleneck locations of the network
  - Through O-D tables or percent turn or turn probabilities for routing traffic

Calibration and Validation

Step 3: System Performance Calibration

- After calibration to capacity and route choice are complete, calibrate the overall performance of the model
  - Compare to field measurements of travel time, queue lengths and the duration of queuing
  - Refine link free-flow speeds and link capacities to better match field conditions
    - For example, may need to adjust free-flow speed distribution (i.e., more narrow range) to reflect commuter traffic behavior
  - These changes may compromise the prior two steps: these changes should be made sparingly and very carefully
  - Analysts must be cautious when comparing field versus simulated delays and queues because these measures are typically defined differently in most simulations (within each model) as well the approaches used in field measurements

Calibration and Validation

How do I know that the Model is Calibrated?

- Did you model demand, capacity, throughput?
  - Unsaturated *vs.* Saturated Conditions
- Did you check the Route Choice/O-D Volumes?
  - CORSIM vs. VISSIM
- Can the model be trusted to represent reality under new, untried conditions, such as revised geometries?
- Is the model flawed if it produces a poor match to reality at only one (five?) of one hundred links?
- Does the model reflect the operation characteristics throughout the peak period or only for a small time period?

Calibration and Validation

How do I know that the Model is Calibrated?

- After calibration, the model may produce a very good match to the base field conditions but produce unrealistic system responses to deviations from this base case. A good example would be an attempt to match simulated and field observed traffic stream speeds. The modeler has the option to either alter the link free-flow speed, the speed-at-capacity, the mixture of traffic volume, the vehicle characteristics, and/or the speed variance; all of which produce very different behavior but match field conditions adequately for the base case condition

Calibration and Validation

How do I know that the Model is Calibrated?

- **What if real-world data is not available for roadways that do not currently exit? (e.g., a future change in geometry). Is it possible to perform a full calibration when no data exists?**
  - A modified calibration procedure where the network is calibrated to typical conditions will ensure the model of the new roadway will perform reasonably well.
  - Modeling the evacuation of a facility. There may be no historical data to calibrate the network to, and there is no way to force a real-world experimental evacuation. However, calibrating the network to perform as a typical network under saturated conditions is critical to the estimation of the effects of an evacuation

Calibration and Validation

How do I know that the Model is Calibrated?

(Cont'd)

- How much error can you tolerate?
  - ➢Sensitivity Analysis
- Were the seeding and simulation periods adequate?
- Did you use adequate sample size, i.e. number of runs?
  - ➢How many runs? Duration of Runs?
- Do you think the results are realistic?

Calibration and Validation

Determination of Sample Size

- **Beside the urban myths of 10 / 20 / 30 runs?**
- **Here is a quick process**
  - Choose your performance measures(s) (MOE)
  - Run the simulation a number of times initially (e.g.,10) to determine the mean and the standard deviation
  - Choose a confidence level (this should be done in the assumption document)
  - Choose a tolerance error (this should be done in the assumption document)
  - Compute the required sample size

Calibration and Validation

Sample Size Example

1. Assume MOE is Speed in mph
2. Run simulation (initially 10 times) with the following results:
   - Sample Mean (Xs) = 32.5 mph
   - Sample Standard Dev (Ss) = 8.5 mph
3. Assume 95% Confidence Level
4. Tolerance Error
   - Observed data; n = 30, Xs = 34.5 mph, Ss = 9.64 mph
   - Tolerable error is 10%

Calibration and Validation

Sample Size Calculations

- 95% Confidence Interval

$Xs \pm Z(Ss/\sqrt{N})$

Where: **$Z(Ss/\sqrt{N})$ = sampling error or tolerable error**

Xs = sample mean

Z = Number of standard deviations away from the mean corresponding to the required confidence level in a normal distribution

Ss = sample standard deviation

N = sample size

$32.5 \pm 1.96(8.5/\sqrt{10}) \rightarrow 32.5 \pm 5.27 \rightarrow$ 5.27 is 16.21% of the mean: too high To bring it to 10% tolerable error or 3.25 mph away from the mean:

- **Sample Size Needed**

$N = (Z)^2(Ss)^2/(E)^2 \rightarrow N = (1.96)^2(8.5)^2/(3.25)^2 = 27$

Therefore, 27 runs will be needed to be 95% confident that the sampling error or the tolerable error will not be greater than 10% of the mean speed

Calibration and Validation

Sample Size Calculations: Additional Iterations

When we run the simulation 17 more times to get to 27 runs, the mean and/or standard deviation may change. Once we have all 27 runs, repeat the computation of the confidence interval to make sure that the sampling error is 10% of the mean or lower

- Example:
  - After 27 runs, mean is 31.5 mph and standard deviation is 10.5 mph
  - New 95% confidence interval: $31.5 \pm 1.96(10.5/\sqrt{27}) \approx 31.5 \pm 3.96$
  - The new sampling error is 12.57% of the mean which is still not good enough.
  - Repeating the computation of the sample size: $N = (Z)2(Ss)^2/(E)^2 = (1.96)2(10.5)2/(3.15)2 = 43$
  - Therefore, 43 runs will be needed to be 95% confident that the sampling error or the tolerable error will not be greater than 10% of the mean speed.
  - Repeating the process of running the simulation 43 times, the new mean now is 30.5 with a standard deviation of 9.5 mph.
  - The new 95% confidence interval is: $30.5 \pm 1.96(9.5/\sqrt{43}) \approx 30.5 \pm 2.84$
  - The new sampling error now is 9.31% of the mean which is lower that the maximum tolerable error of 10% so now we are 95% confident that the sampling error or the tolerable error is not greater than 10% of the mean speed

Calibration and Validation

Challenges

- Understand what you are doing
- Cost 
- Time/Schedule
- Using the right Model
  - Does it really matter which model *vs.* how you used the model

Calibration and Validation

Challenges

- Accuracy and Expectations
- Risk/Still not satisfied with results
- Documentation
  - Did you summarize what you did?
  - Can you replicate the same process?
  - Can someone else follow what you did?

Calibration and Validation

Are Calibration and Validation covered in the New Signal Timing Manual?

- How about some other resources and references

Google it!