

Development and Testing of Comprehensive Transport and Dispersion Model Evaluation Methodologies

Steven Hanna
(GMU and HSPH)

OFCM Hanna 19 June 2003

Components of Model Evaluation

- Independent evaluation is best
- Properly pose the questions being asked
- Acquisition of field and lab data bases
- Exploratory data analysis
- Scientific evaluation
- Statistical evaluation methods and model acceptance criteria
- Communicate results to decision makers

Scientific Evaluation Criteria

- Are source code, users guide, and scientific documentation available for review?
- Are results of model evaluations given?
- Are the model physics and chemistry assumptions appropriate?
- Are limits to applicability given?
- Is sufficient guidance given that users consistently arrive at the same answer?

Outputs of model internal modules can be compared with observations using statistical evaluation methods. This is considered part of the scientific evaluation.

- Plume width or depth
- Plume rise
- Chemical reaction rate
- Speed of cloud
- Wake retention time scale

Statistical Approach

- Statistician John Tukey says to first “look and see” (i.e., look at the data)
- Only then, apply statistical methods (fractional bias etc. plus confidence limits)
- Can determine whether a given model’s performance has high confidence or whether one model’s performance is significantly different from another’s

Fractional bias (FB), geometric mean bias (MG), normalized mean square error (NMSE), geometric mean variance (VG), correlation coefficient (R), and fraction of predictions within a factor of two of observations (FAC2)

$$FB = \frac{(\overline{C_o} - \overline{C_p})}{0.5 (\overline{C_o} + \overline{C_p})}$$

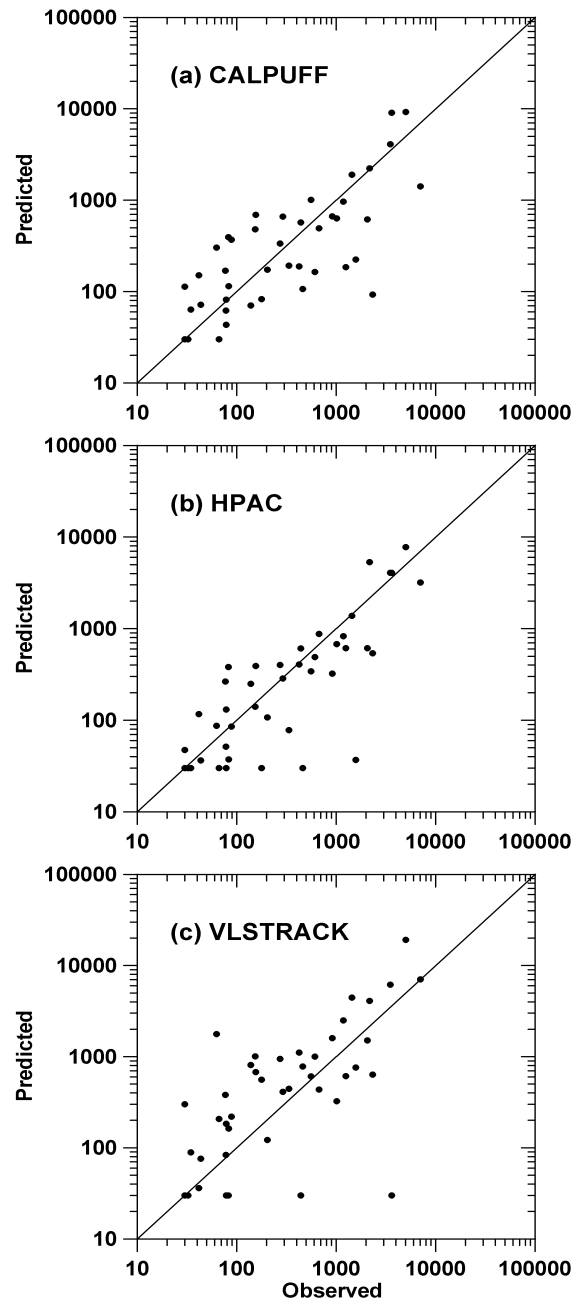
$$MG = \exp (\overline{\ln C_o} - \overline{\ln C_p})$$

$$NMSE = \frac{\overline{(C_o - C_p)^2}}{\overline{C_o} \overline{C_p}}$$

$$VG = \exp \left[\overline{(\ln C_o - \ln C_p)^2} \right]$$

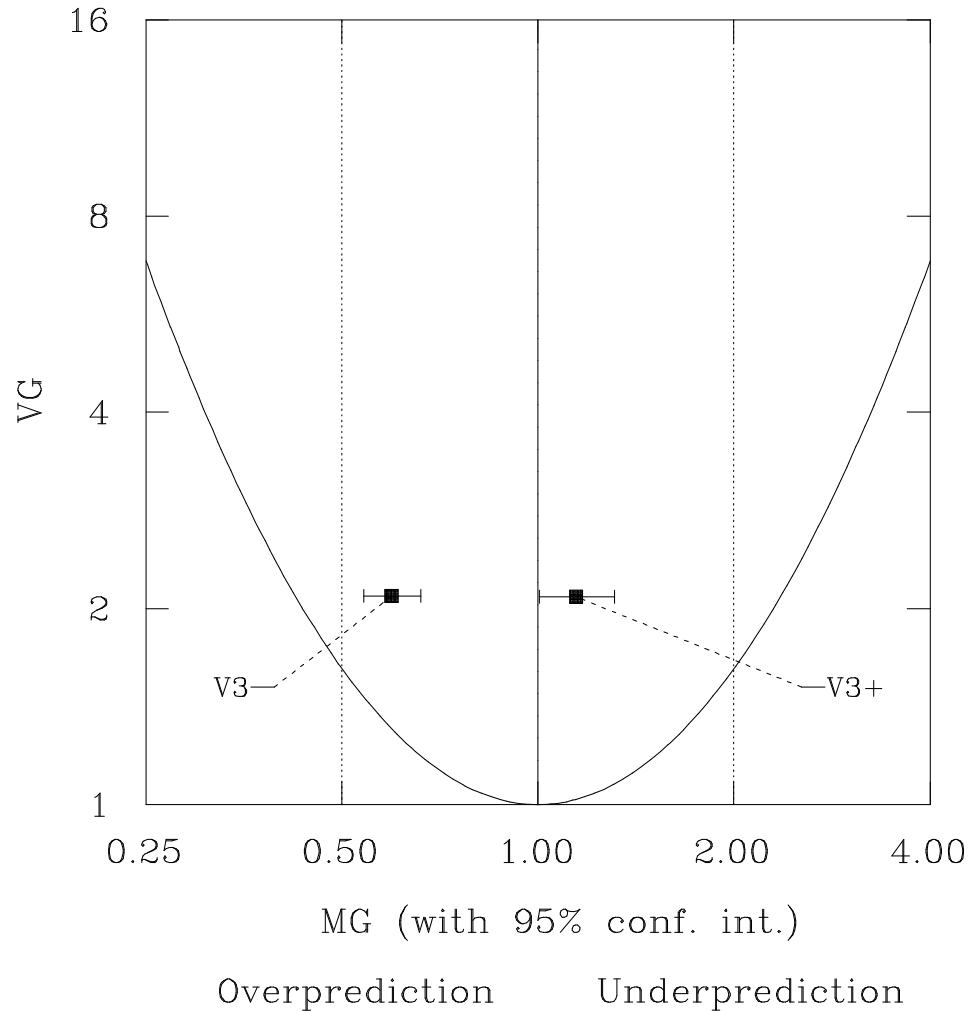
$$R = \frac{\overline{(C_o - \overline{C_o})(C_p - \overline{C_p})}}{\sigma_{C_p} \sigma_{C_o}}$$

$$FAC2 = \text{fraction of data that satisfy } 0.5 \leq \frac{C_p}{C_o} \leq 2.0$$



Scatter plots of the DP 26 maximum dosage (ppt-hr) at each sampling line for CALPUFF, HPAC, and VLSTRACK.

Plot of geometric mean MG versus geometric variance VG can identify whether one model is performing better.



Model Performance Criteria

- Where do we draw the line between an “acceptable” model and an “unacceptable” model?
- Depends on scenario (e.g. availability of emissions information and local meteorology) and which output is being evaluated. Easier – max C on domain. Tougher – Pairings in space and time.
- For max C on arc, acceptable relative mean bias is + or – 50 % and acceptable relative scatter is factor of two for research grade experiments with source well-known and on-site meteorology.

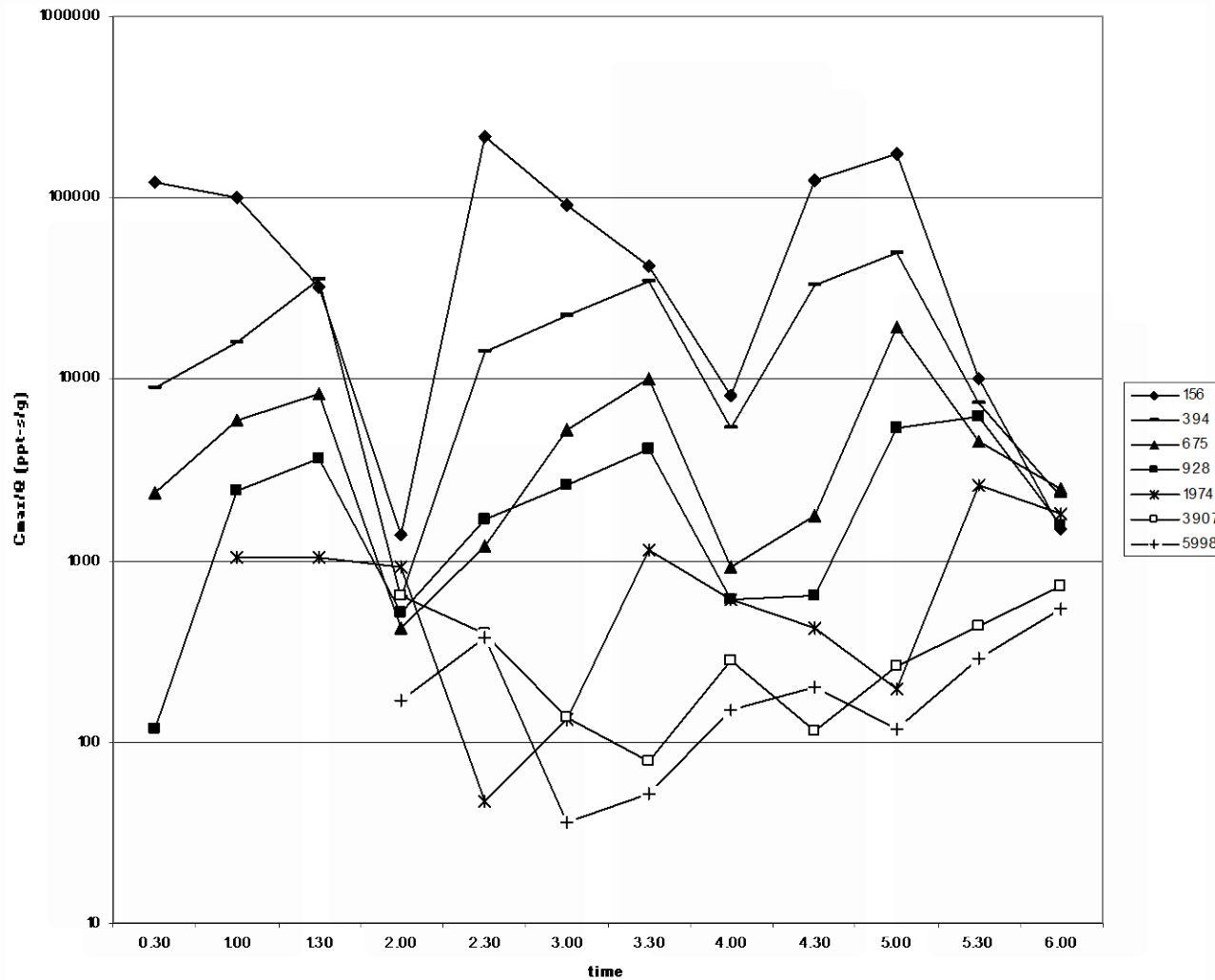
Communication of Model Evaluation Results to Decision Makers

- Brief and understandable
- Fast response emergencies versus planning exercises
- Must communicate degree of uncertainty in predictions

Model evaluation document and BOOT software updated

- Joseph Chang Ph.D. thesis at GMU
- Chang and Hanna manuscript containing a review of model evaluation procedures recently submitted to the journal *Meteorology and Atmospheric Physics*. This manuscript contains a test application to the SLC Urban 2000 data
- CFD model evaluation suggestions published in QNET newsletter

Plots of time series of concentrations



Time series of observed $C_{\max}/Q$ for IOP04, for the seven monitoring arcs, at downwind distances in meters given in the legend. SF_6 releases occurred between 0:00 and 1:00, 2:00 and 3:00, and 4:00 and 5:00.

Observed hourly-averaged Cmax/Q (in s/m³ times 10⁶) for the seven monitoring arcs and the 18 trials at Salt Lake City Urban 2000. The symbol "n/a" means that the data did not meet acceptance criteria. The bottom two rows of the table contain the averaged observed Cmax/Q for each monitoring arc, followed by the predicted Cmax/Q by the baseline urban model.

IOP	Trial	u (m/s)	Arc R (m)	Arc 1 156	Arc 2 394	Arc 3 675	Arc 4 928	Arc 5 1974	Arc 6 3907	Arc 7 5998
2	1	0.81	Cmax/Q(s/m ³) * 10 ⁶	317.7	79.6	14.2	3.58	3.91	1.97	0.47
2	2	0.61	“	421.2	103.7	2.85	n/a	n/a	n/a	n/a
2	3	0.5	“	366.1	86.8	7.26	1.39	n/a	n/a	n/a
4	1	1.13	“	606.3	120.1	35.4	16.2	5.67	2.81	1.48
4	2	0.94	“	836.1	154.7	29.5	12.7	4.39	0.94	0.95
4	3	0.76	“	573.1	186.7	60.6	21.1	11.3	2.53	2.28
5	1	0.64	“	149.6	77.3	22.2	13.8	3.87	1.19	2.98
5	2	0.91	“	249.4	80.5	19.3	13.6	4.09	1.35	1.32
5	3	1.06	“	402.2	118.3	20.1	12.8	8.09	1.5	1.25
7	1	1.01	“	520	187.7	32.4	17.9	6.08	1.98	n/a
7	2	1.04	“	200.6	25.8	28.9	9.16	10.3	2.47	2.59
7	3	1.21	“	207.6	75.8	41.8	36.4	10.5	3.13	1.85
9	1	2.69	“	129.3	49.6	44.9	4.77	n/a	0.63	0.49
9	2	2.47	“	243.7	56.8	32.5	n/a	n/a	1.06	1.01
9	3	3.23	“	115.3	37.3	11	7.56	2.63	1.5	n/a
10	1	1.51	“	158	33	10.6	4.05	2.03	1.19	n/a
10	2	2.16	“	153.4	31.9	9.84	8.1	3.45	1.87	n/a
10	3	2.31	“	72.9	22.7	4.22	1.78	1.48	0.58	n/a
AvgObs		1.39	“	317.9	84.9	23.8	11.6	5.56	1.67	1.52
AvgPred		1.39	“	229.1	52.4	21.2	12.5	3.71	1.36	0.76

Special Needs for Evaluating CFD Models Applied to Urban and Industrial Buildings

- Laboratory observations are more straightforward, with regular building arrays
- In real cities, observations are at random locations around buildings
- Separately evaluate 1) mean flow and turbulence, and 2) concentrations
- Pose specific questions to be answered (e.g., what is retention time in street canyon?)

Research Needs

- Review existing model evaluation exercises to determine model acceptance criteria for various scenarios
- Collect field data in non-ideal and non-steady conditions as well as ideal conditions
- Help to determine optimum data inputs and model complexity
- Determine best approaches for evaluating near-building models