

Performance Evaluations Metrics for Long Range Transport Models

Evaluation Paradigm for Long Range Transport Models

- LRT models play a unique role in air quality modeling. This class of models plays several roles.
 - Emergency response modeling
 - Class I increments and visibility
- Requires additional level of skill to reflect time and space considerations of LRT model use
 - Statistical measures should examine spatiotemporal pairing ability of LRT models.

Evaluation Goal

- Develop meteorological and tracer databases for evaluation of long range transport models.
- Develop a consistent and objective method for evaluating long range transport (LRT) models used by the EPA.
- Promote the best scientific application of models based upon lessons learned from evaluations and reflect this in EPA modeling guidance.

Irwin Methodology

Spatial	Temporal	Performance
Azimuth of Plume Centerline	Plume Arrival Time	Crosswind Integrated Concentration
Plume Sigma-y	Transit Time on Arc	Observed Maximum

Method:

- Compute n -hour average for each receptor on arc (ArcAverage program)
- Use trapezoidal integration to 'fit' average plume on arc (PLMFIT program)

Background from Original Performance Evaluations

- Measures employed by Irwin (1998) and EPA (1998) provide useful diagnostic information about the performance of LRT modeling systems such as CALPUFF, but they do not always lend themselves easily to spatiotemporal analysis or direct model intercomparison.
- For tracer studies such as the Great Plains Tracer Experiment, Savannah River, and INEL74 where distinct arcs of monitors were present, the Irwin evaluation approach was used.
- In addition to the Irwin methodology, EPA augmented statistical measures focusing upon spatiotemporal comparisons of model-observation pairings.

Statistical Evaluation Methodology

- The model evaluation methodology employed for this project was designed following the procedures of Mosca et al. (1998) and Draxler et al. (2001).
- Statistical measures fall into four broad categories
 - Scatter
 - Bias
 - Spatial
 - Cumulative distribution
- NOAA ARL DATEM performance evaluation program (STATMAIN) augmented by EPA with additional spatial statistics for false alarm rates, probability of detection, and threat score.

NOAA's Data Archive of Tracer Experiments and Meteorology (DATEM)

Internet Explorer cannot display this page. Air Resources Laboratory...

ARL
Air Resources Laboratory
Conducting research and development in the fields of air quality, atmospheric dispersion, and climate

NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE

Skip Top Navigation

Enter search term(s)

ARL site only All of NOAA
[Advanced Search](#)

- ARL Home
- HYSPLIT Model
- READY
- Air Quality
- Atmospheric Dispersion
 - ▶ READY
 - ▶ HYSPLIT Model
 - ▶ Volcanic Ash Modeling
 - ▶ [DATEM Project >>](#)
- Climate

DATEM

Data Archive of Tracer Experiments and Meteorology

DATEM - Data Archive of Tracer Experiments and Meteorology

Project Overview

The availability of meteorological re-analysis archives at several international centers, provides an opportunity to link high quality modern meteorological data with the data from many older long-range tracer experiments. This web site has been created to contain the experimental data, relevant reports, meteorological data, statistical analysis, and display software, all in a common format for PC or UNIX applications. This data base permits the atmospheric transport modeling community to conduct various verification and sensitivity studies and compare model results with each other on a common basis. Currently, only longer range (10's to 1000's of km downwind) experimental data are considered, consistent with the spatial and temporal resolution of the meteorological re-analysis data. User's of the data base are encouraged to share any additional analysis software that they might develop and prepare and submit additional experimental data to the archive.

Statistical Measures - Scatter

$$FOEX = \left[\frac{N_{(P_i > N_i)}}{N} - 0.5 \right] \times 100\%$$

$$FA\alpha = \left[\frac{N(y - y_0 = [x - x_0]\alpha)}{N} \right] \times 100$$

$$NMSE = \frac{1}{NPM} \sum (P_i - M_i)^2$$

$$R = \frac{\sum_i (M_i - \bar{M}) \cdot (P_i - \bar{P})}{\left[\sqrt{\sum (M_i - \bar{M})^2} \right] \left[\sqrt{\sum (P_i - \bar{P})^2} \right]}$$

- Factor of Exceedance (FOEX)
- Factor of 2/5 (FA2/5)
- Normalized Mean Square Error (NMSE)
- Pearson's Correlation Coefficient (R)

Statistical Measures – Cumulative Distribution

- Kolmogorov-Smirnov Parameter $KS = \text{MAX} |C(M_k) - C(P_k)|$

Statistical Measures - Bias

$$\bar{B} = \frac{1}{N} \sum_i (P_i - M_i)$$

- Mean Bias (B)

$$FB = 2\bar{B} / (\bar{P} + \bar{M})$$

- Fractional Bias (FB)

Spatial Statistics

$$FMS = 100 \frac{N(p > 0) \cap N(m > 0)}{N(p > 0) \cup N(m > 0)}$$

- Figure of merit in Space (FMS)

- Additional EPA Metrics

- False Alarm Rate (FAR)

- Probability of Detection (POD)

- Threat Score (TS)

$$FAR = \left(\frac{a}{a + b} \right) \times 100\%$$

$$POD = \left(\frac{b}{b + d} \right) \times 100\%$$

$$TS = \left(\frac{b}{a + b + d} \right) \times 100\%$$

a =forecast, not observed

b =forecast and observed

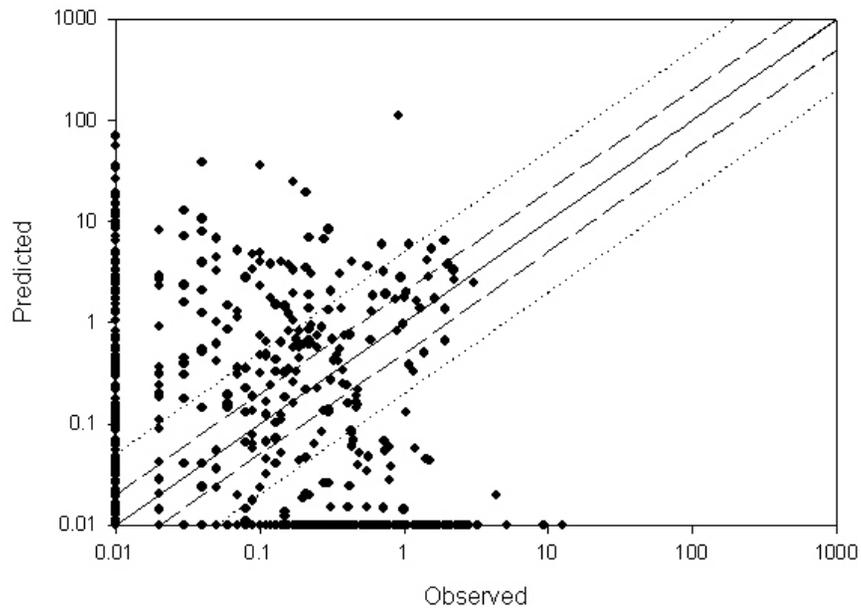
d =observed, not forecast

Statistical Measure – Final Model Rank

$$RANK = R^2 + \left(1 - \left| \frac{FB}{2} \right| \right) + \frac{FMS}{100} + \left(1 - \frac{KS}{100} \right)$$

- Model Rank (RANK) – measure of model “success” across each of the four broader statistical categories.
- Range from 0 to 4, with 0 poorest and 4 best performance.
- Unique measure allows for direct intercomparison amongst models and summarizes success of model into a single number easily relatable

Example Performance Evaluation – FLEXPART 6.2



EXP_DATA results file: gbl1_001.txt
 Model variation: 001 Tracer number: 1 Station select: All

3105 Unaveraged data points for processing
 0.00 Percentile input for zero measured
 0.00 Zero measured concentration value

0.00 Correlation coefficient
 0.06 T-value (|Slope|/Standard Error)
 269.56 Average measured concentration
 874.65 Average calculated concentration
 3.24 Ratio of calculated/measured
 118.44 Normalized mean square error
 1104 Number of pairs analyzed

605.09 Average bias [(C-M)/N]
 197.29 Lo 99 % confidence interval
 1012.90 Hi 99 % confidence interval
 1.06 Fractional bias [2B/(C+M)]

30.07 Fig of merit in space (%)

0.54 False Alarm Ratio
 0.27 Probability of Detection
 0.20 Threat score at 0.1 ngm-3 threshold

-18.30 Factor exceeding [N(C>M)/N-0.5]
 5.53 Percent C/M \pm 2
 12.77 Percent C/M \pm 5
 30.07 Percent M>0 and C>0
 55.16 Percent M>0 and C=0
 14.76 Percent M=0 and C>0

51.00 Kolmogorov-Smirnov Parameter

1.26 Final rank (C,FB,FMS,KSP)