Evaluation of Four Lagrangian Models Against the Cross-Appalachian and European Tracer Experiments

Bret A. Anderson, Roger W. Brode

2010 EPA Modeling Workshop
May 13, 2010
Portland, Oregon
Outline

- Introduction on long range transport (LRT) models and their role in regulatory air modeling
- Background on EPA evaluation program
  - Evaluation paradigm
  - Statistical framework
  - Candidate model platforms
- Review of results from European and Cross-Appalachian Tracer Experiments
EPA LRT Model Evaluation Project Goals

► 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.
► Evaluate new models as part of IWAQM Phase 3 process.
Issues with Original Evaluation Paradigm

► Provides limited diagnostic information regarding model performance, but lacks objective measures to measure model performance

► Treatment of LRT model in fashion similar to near-field dispersion models such as ISC or AERMOD, neglecting how LRT models are applied in both real-world and regulatory contexts
Statistical Evaluation Methodology

- Global statistical measures fall into four broad categories
  - Scatter
  - Bias
  - Spatial
  - Cumulative Distribution
- Temporal statistical analysis added
  - Figure of Merit in Time (FMT)
- Additional spatial performance measures added based upon Kang et al. (2007)
  - False Alarm Rate (FAR)
  - Probability of Detection (POD)
  - Threat Score (TS)
- NOAA ARL DATEM performance evaluation program (STATMAIN) augmented by EPA with additional spatial statistics for false alarm rates, probability of detection, and threat score.
Key Statistical Parameters

\[ PCC = \frac{\sum_{i}(M_i - \bar{M}) \cdot (P_i - \bar{P})}{\sqrt{\sum_{i}(M_i - \bar{M})^2} \cdot \sqrt{\sum_{i}(P_i - \bar{P})^2}} \]

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

\[ KSP = Max|C(M_k) - C(P_k)| \]

\[ FMS = \frac{A_M \cap A_P}{A_M \cup A_P} \]
Spatial Statistics

**Additional spatial statistics from Kang et al (2007)**

- **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** is number of times a condition is forecasted, but not observed ("false alarm")
- **B** is number of times a condition is correctly forecasted ("hit")
- **D** is number of times a condition was observed but not forecasted ("miss")
Analogous to the FMS is the figure of merit in time (FMT), which is calculated at a fixed location \((\bar{x})\), rather than at a fixed time as the FMS.

FMT evaluates the overlap between the measured \((M)\) and predicted \((P)\) concentrations at location and time.

The FMT is normalized to the maximum predicted or measured value at each time interval and is expressed as a percentage value in the same manner as the FMS.

\[
FMT(\bar{x}) = \frac{\sum_{j} \min\{M(\bar{x}, t_j), P(\bar{x}, t_j)\}}{\sum_{j} \max\{M(\bar{x}, t_j), P(\bar{x}, t_j)\}} \times 100\%
\]
Model Comparison Parameter

\[ RANK = |R| + \left(1 - \frac{|FB|}{2}\right) + \frac{FMS}{100} + \left(1 - \frac{KSP}{100}\right) \]

Draxler et al. (2001) introduced a model comparison parameter called RANK, a composite statistic of the four broad statistical categories (scatter, bias, spatial, and unpaired distribution).

- Allows for direct comparison of different models or perturbations of model control options in the same model system.
Evaluation Paradigm

- Inherent amount of uncertainty due to differences in technical formulations between various modeling systems.
- Use common meteorological platform with minimal diagnostic adjustments to reduce uncertainty.
  - This is a challenge when models such as SCIPUFF and CALPUFF use diagnostic wind models as primary source of 3-D meteorological data.
    - Use MM5SCIPUFF developed by Penn State and MMIF (CALPUFF) developed by EPA to couple MM5 directly to these models.
- Model control options mostly default “out-of-the-box” configuration.
  - CALPUFF configured for turbulence dispersion and puff-splitting similar to SCIPUFF, which is a deviation from its default configuration.
Issues with Statistical Evaluation

► Treatment of zero concentration pairs is an issue – Stohl (1998) finds a number of ATMES-II statistical measures are highly sensitive to inclusion of zero-zero concentration pairs.

► ATMES-II (Mosca et al (1998)) dropped all zero-zero pairs except data within ±6-hours of arrival/departure of tracer cloud at a station.

► Draxler et al (2001) retained all zero-zero pairs in their evaluation of HYSPLIT.

► EPA uses hybrid ATMES-II approach, drops all zero-zero pairs for entire global statistical analysis.
Models Under Evaluation

Two Distinct Class of Lagrangian Models
  ► Gaussian Puff Models
  ► Particle Models

Operational models used for emergency response or research purposes

► CALPUFF Version 5.8 (EPA approved version)
► MM5-FLEXPART (Version 6.2)
► HYSPLIT (Version 4.8)
► SCIPUFF (Version 2.303)
European Tracer Experiment (ETEX)

► ETEX initiated in 1992 by the European Commission (EC), International Atomic Energy Agency (IAEA), and the World Meteorological Organization (WMO) to address many questions that arose from 1986 Chernobyl accident regarding the development of LRT models.

► ETEX was designed to validate LRT models used for emergency response situations and to develop a database which could be used for model evaluation purposes.

► Two perfluorocarbon tracer (PFT) releases in October and November 1994.

► 168 monitoring sites in 17 countries with a sampling frequency of 3 hours for 90 hour duration.
MM5 Version 3.7.4 used to supply 3-D meteorological fields to LRT models

- Initialized with NNRP dataset (2.5° x 2.5° available at 6h intervals)
- Single 36 km domain, 43 vertical levels

- Physics options
  - ETA PBL
  - Kain-Fritsch II Cumulus
  - RRTM radiation
  - Dudhia Simple Ice

- Analysis nudging (above PBL for temperature and moisture)

- Performance evaluation against 3-hr observation dataset collected at 168 ETEX monitoring sites
MM5 Evaluation - Wind Speed

Observed/Predicted Windspeed

Bias Windspeed
MM5 Evaluation - Wind Direction

**Observed/Predicted Wind Direction**

- **ObsWndDir**: Observed Wind Direction
- **PrdWndDir**: Predicted Wind Direction

**Bias/Gross Error Wind Direction**

- **Bias**: Bias in wind direction
- **Gross Error**: Gross error in wind direction

The charts illustrate the comparison between observed and predicted wind directions over a period from 10/23 to 10/27, with bias and gross error analysis.
MM5 Evaluation - Temperature

10/23 10/26

ObsTemp
PrdTemp

Bias Temp
Spatiotemporal Analysis

Temporal evolution of FMS scores help elucidate issues with potential model performance.

SCIPUFF and HYSPLIT exhibit good agreement with the spatial extent of tracer cloud (50% - 60%) out to T+60 hours.

CALPUFF shows similar agreement to FLEXPART out to T+36 hours (20% - 30%), but advection errors cause FMS score to drop dramatically after this point.
Temporal Analysis - Figure of Merit in Time

[Map of Europe]

[Graph showing time series data with various markers for CALPUFF, SCIPUFF, MM5-FLEXPART, and HYSSPLIT]
# Global Statistical Analysis

<table>
<thead>
<tr>
<th>Statistic</th>
<th>CALPUFF</th>
<th>SCIPUFF</th>
<th>HYSPLIT</th>
<th>FLEXPART</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCC</td>
<td>0.17</td>
<td>0.65</td>
<td>0.64</td>
<td>0.45</td>
</tr>
<tr>
<td>FB</td>
<td>1.49</td>
<td>1.57</td>
<td>1.00</td>
<td>1.79</td>
</tr>
<tr>
<td>FOEX</td>
<td>-34.97</td>
<td>12.16</td>
<td>-18.79</td>
<td>-21.48</td>
</tr>
<tr>
<td>KSP</td>
<td>75.00</td>
<td>37.00</td>
<td>53.00</td>
<td>57.00</td>
</tr>
<tr>
<td>RANK</td>
<td>0.67</td>
<td>1.77</td>
<td>1.78</td>
<td>1.03</td>
</tr>
</tbody>
</table>
## Spatial Analysis

<table>
<thead>
<tr>
<th>Statistic</th>
<th>CALPUFF</th>
<th>SCIPUFF</th>
<th>HYSPLIT</th>
<th>FLEXPART</th>
</tr>
</thead>
<tbody>
<tr>
<td>FMS</td>
<td>13.75</td>
<td>49.93</td>
<td>40.48</td>
<td>29.28</td>
</tr>
<tr>
<td>POD</td>
<td>12.56</td>
<td>53.52</td>
<td>48.02</td>
<td>26.65</td>
</tr>
<tr>
<td>FAR</td>
<td>60.14</td>
<td>50.41</td>
<td>31.66</td>
<td>48.73</td>
</tr>
<tr>
<td>TS</td>
<td>10.56</td>
<td>34.66</td>
<td>39.28</td>
<td>21.27</td>
</tr>
</tbody>
</table>
CALPUFF Time Series

24 hours

36 hours

48 hours

60 hours
Cross-Appalachian (CAPTEX-83) Monitoring Sites and Model Domain

- 7 PFT releases from 2 sites (Dayton, OH and Sudbury, ON) between September – October, 1983
- 84 monitoring sites located across northeast US and southeast Canada.
- 3 hr and 6 hr sample intervals
Goal is to evaluate importance of grid resolution and FDDA strategies on model performance.

12 MM5 Simulations
- 80 km resolution, 16 vertical layers
  - 108-, 36-, and 12-km resolution, 33 layer
  - 108-, 36-, and 12-km resolution, 43 layer
MM5 Performance Evaluation - Wind Direction/Speed

Gross Error (degrees)

Gross Error (m/s)
CALPUFF Evaluation - Overall Simulation
RANK

CALMET Experiment

[Graph showing the distribution of ranks for different scenarios, with ranks ranging from 0.0 to 2.5]
CALMET Experiments

- 24 CALMET Experiments
  - 18-, 12-, and 4-km CALMET resolutions
    - Initial Guess Data MM5 with 80-, 36-, and 12-km resolution
    - Large (500/1000), Medium (100/500), Small (10/100) Radii of Influence for Objective Analysis
    - NOOBS

- Meteorological Evaluation Using MMIFstat
CALMET Performance Evaluation - Wind Direction/Speed

Gross Error (degrees)

Gross Error (m/s)
### Overall Model Performance

<table>
<thead>
<tr>
<th>Tracer Study</th>
<th>SCIPUFF</th>
<th>CALPUFF</th>
<th>FLEXPART</th>
<th>HYSPLIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAPTEX-1</td>
<td>2.23</td>
<td>1.81</td>
<td>2.54</td>
<td>2.35</td>
</tr>
<tr>
<td>CAPTEX-2</td>
<td>1.92</td>
<td>1.59</td>
<td>1.65</td>
<td>2.10</td>
</tr>
<tr>
<td>CAPTEX-3</td>
<td>1.25</td>
<td>1.10</td>
<td>1.49</td>
<td>1.83</td>
</tr>
<tr>
<td>CAPTEX-4</td>
<td>1.82</td>
<td>1.69</td>
<td>1.53</td>
<td>1.88</td>
</tr>
<tr>
<td>CAPTEX-5</td>
<td>1.30</td>
<td>1.20</td>
<td>1.81</td>
<td>1.93</td>
</tr>
<tr>
<td>CAPTEX-7</td>
<td>2.40</td>
<td>2.06</td>
<td>1.76</td>
<td>1.90</td>
</tr>
<tr>
<td>ETEX-1</td>
<td>1.77</td>
<td>0.67</td>
<td>1.08</td>
<td>1.78</td>
</tr>
<tr>
<td><strong>Overall RANK</strong></td>
<td><strong>1.81</strong></td>
<td><strong>1.35</strong></td>
<td><strong>1.69</strong></td>
<td><strong>1.97</strong></td>
</tr>
</tbody>
</table>
Summary

► EPA adopted ATMES-II/NOAA DATEM statistical evaluation framework for LRT model evaluations.

► 4 systems currently under evaluation (Lagrangian puff and particle systems)

► HYSPLIT performs best of all competing Lagrangian models examined in this study, with highest rank in 5 of 7 mesoscale tracer experiments.

► SCIPUFF ranked second, FLEXPART followed closely. CALPUFF lowest performance overall.