OAQPS

Office of Air Quality Planning and Standards

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

- OAQPS chose statistical framework adopted for ATMES-II experiment (Mosca et al, 1998) as implemented by Draxler et al (2001)
- 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

$$PCC = \frac{\sum_i \left(M_i - \overline{M}\right) \bullet \left(P_i - \overline{P}\right)}{\left[\sqrt{\sum \left(M_i - \overline{M}\right)^2}\right]\left[\sqrt{\sum \left(P_i - \overline{P}\right)^2}\right]}$$

$$FB = 2\overline{B} / \left(\overline{P} + \overline{M}\right)$$

$$KSP = Max\left|C\left(M_k\right) - C\left(P_k\right)\right|$$

$$FMS = \frac{A_M \cap A_P}{A_M \cup A_P}$$

- Correlation (PCC) – Scatter
- Fractional bias (FB)- Bias
- Kolomogorov – Smirnov Parameter (KSP) - Distribution
- Figure of Merit in Space (FMS) - Spatial

OAQPS

Spatial Statistics

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

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

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

- Additional spatial statistics from Kang et al (2007)
  - False Alarm Rate (FAR)
  - Probability of Detection (POD)
  - Threat Score (TS)

- *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")

OAQPS

Figure of Merit in Time (FMT)

$$FMT(\bar{x}) = \frac{\sum_j \min\{M(\bar{x}, t_j), P(\bar{x}, t_j)\}}{\sum_j \max\{M(\bar{x}, t_j), P(\bar{x}, t_j)\}} \times 100\%$$

- Analogous to the FMS is the figure of merit in time (FMT), which is calculated at a fixed location (), 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

Model Comparison Parameter

$$RANK = |R| + \left(1 - |FB/2|\right) + FMS/100 + \left(1 - KSP/100\right)$$

- 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.

OAQPS

Evaluation Paradigm

- Evaluation procedures follow logic of Chang et al (2003) regarding multi-model evaluations
- 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

OAQPS

Issues with Statistical Evaluation

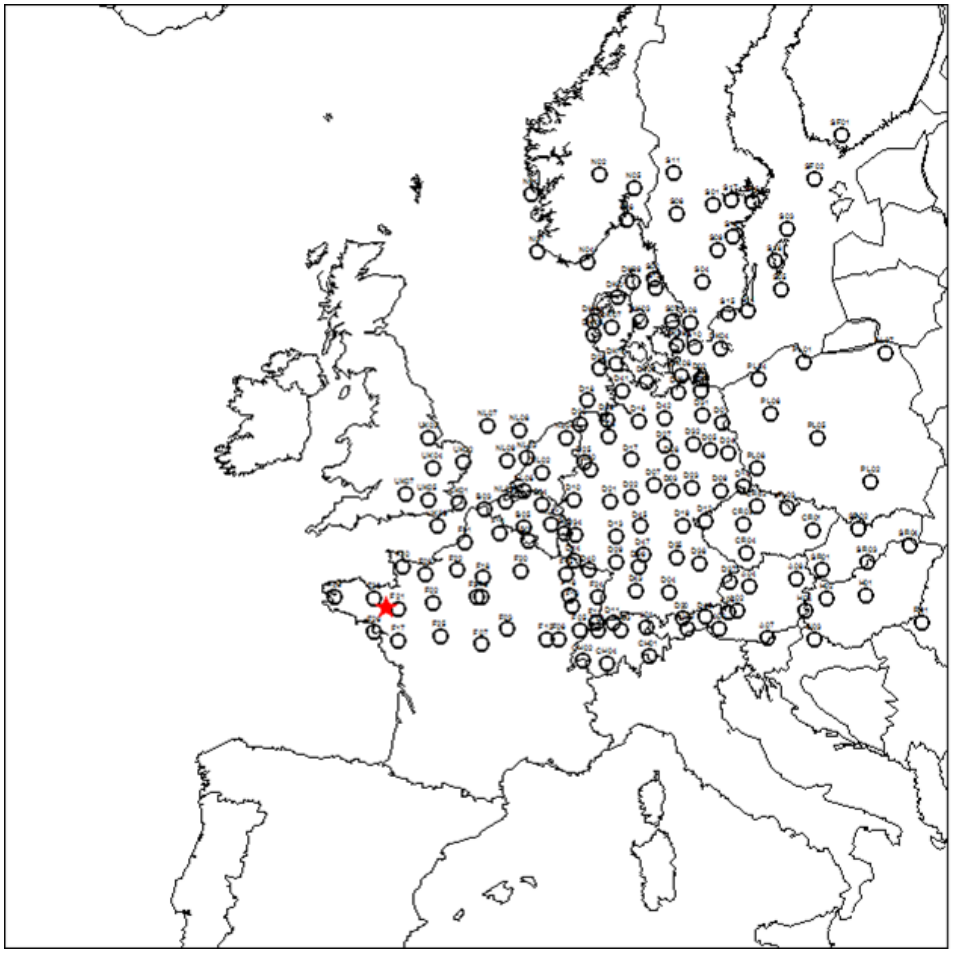
- Treatment of zero concentration pairs is an issue – Stohl (1998) finds a number of ATMES-II statistical measures are highly sensitive to inclusion of zero-zero concentration pairs.
  - ATMES-II (Mosca et al (1998)) dropped all zero-zero pairs except data within ±6-hours of arrival/departure of tracer 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)

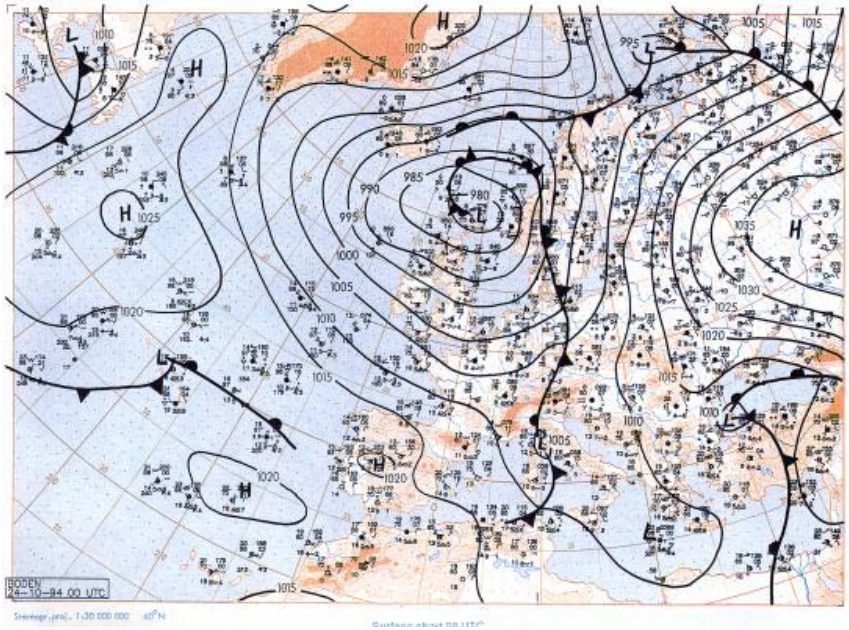
European Tracer Experiment (ETEX)



- 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 perflourocarbon tracer (PFT) releases in October and November 1994.
  - 168 monitoring sites in 17 countries with a samling frequency of 3 hours for 90 hour duration.

OAQPS

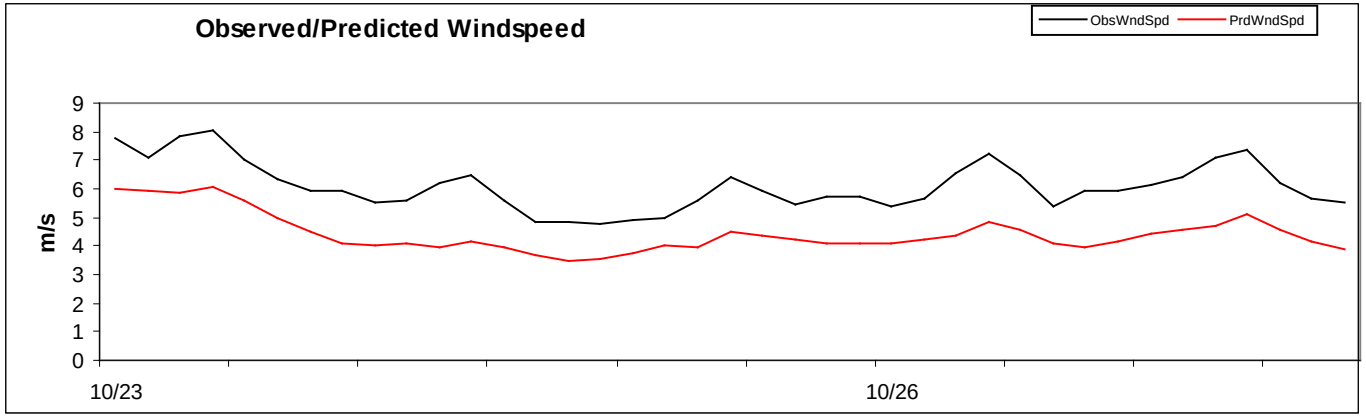
Meteorology

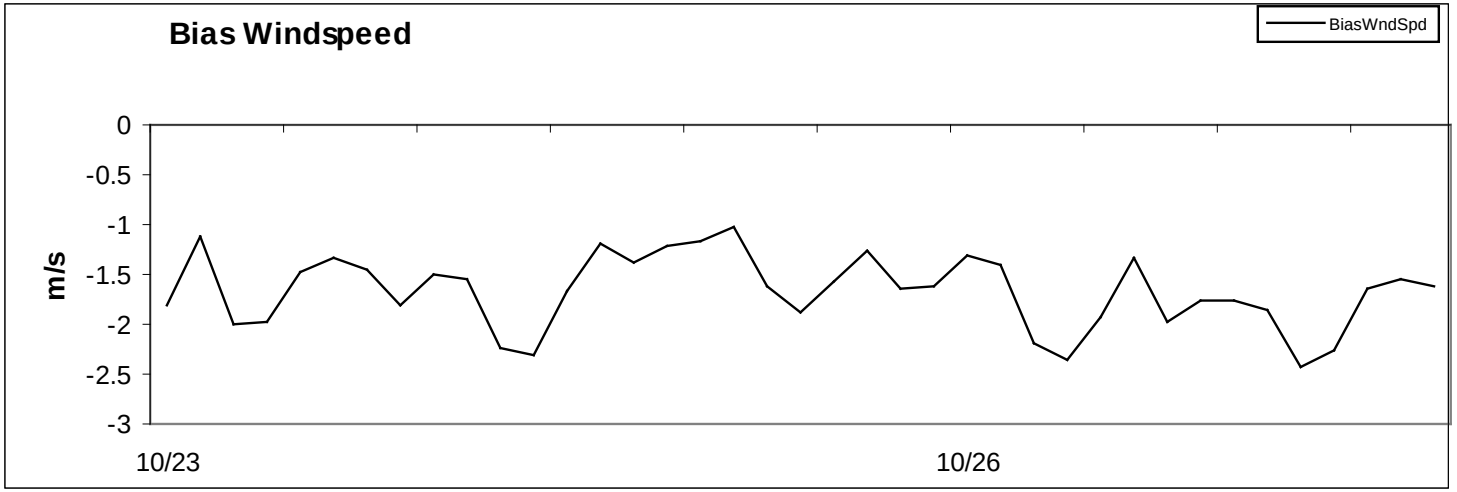


- MM5 Version 3.7.4 used to supply 3-D meteorological fields to LRT models
- Initialized with NNRP dataset (2.5º x 2.5º available at 6h intervals)
- Single 36 km domain, 43 vertical levels
- Physics options
  - ETA PBL
  - Kain-Fritsch II Cumulus
  - RRTM radiation
  - Dudhia Simple Ice
- Analysis nudging (above PBL for temperature and moisture)
- Performance evaluation against 3-hr observation dataset collected at 168 ETEX monitoring sites

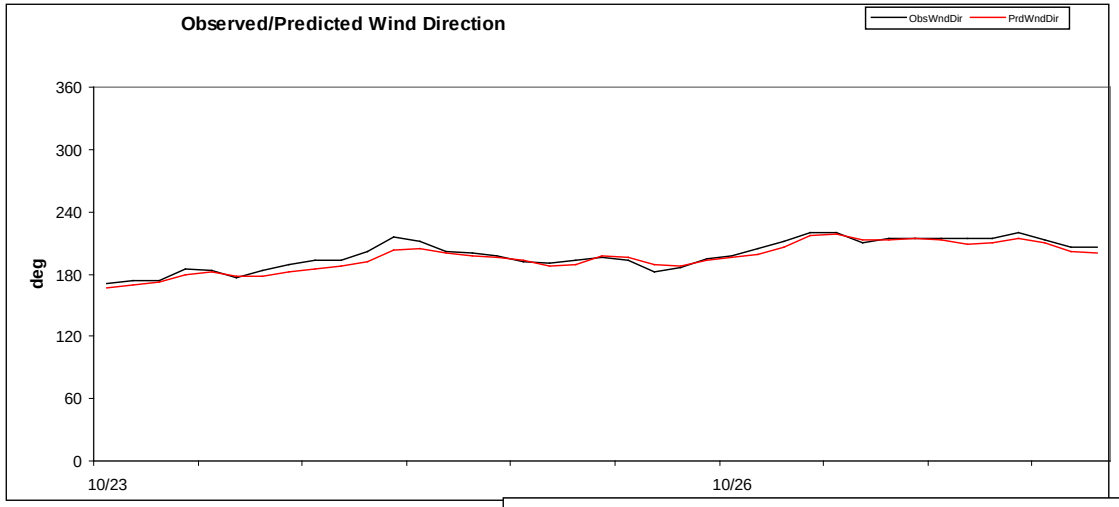
OAQPS

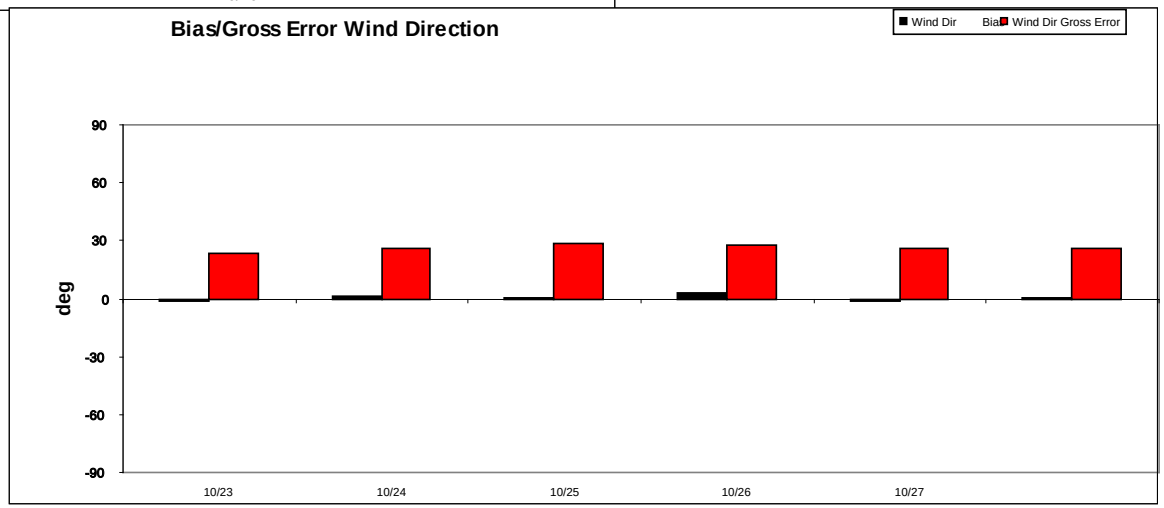
MM5 Evaluation – Wind Speed



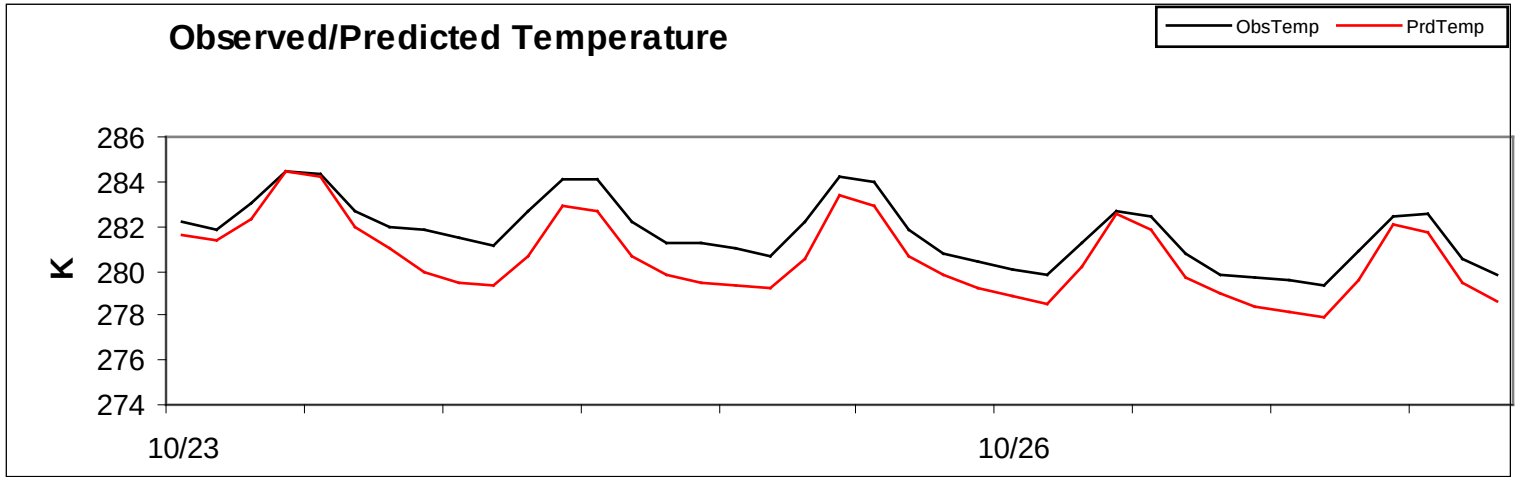


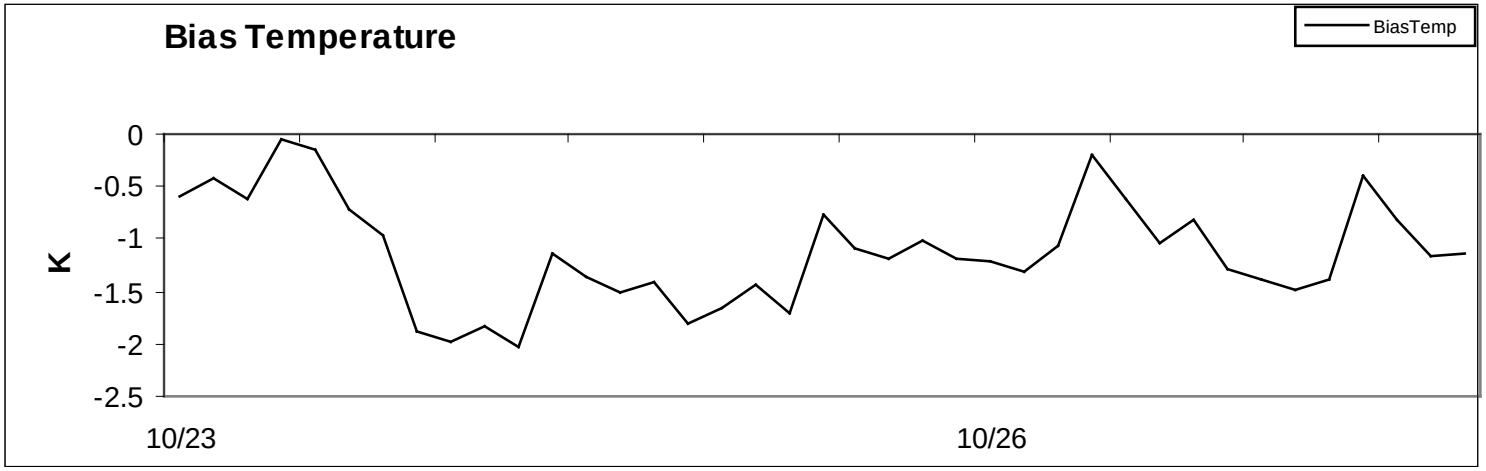
MM5 Evaluation – Wind Direction



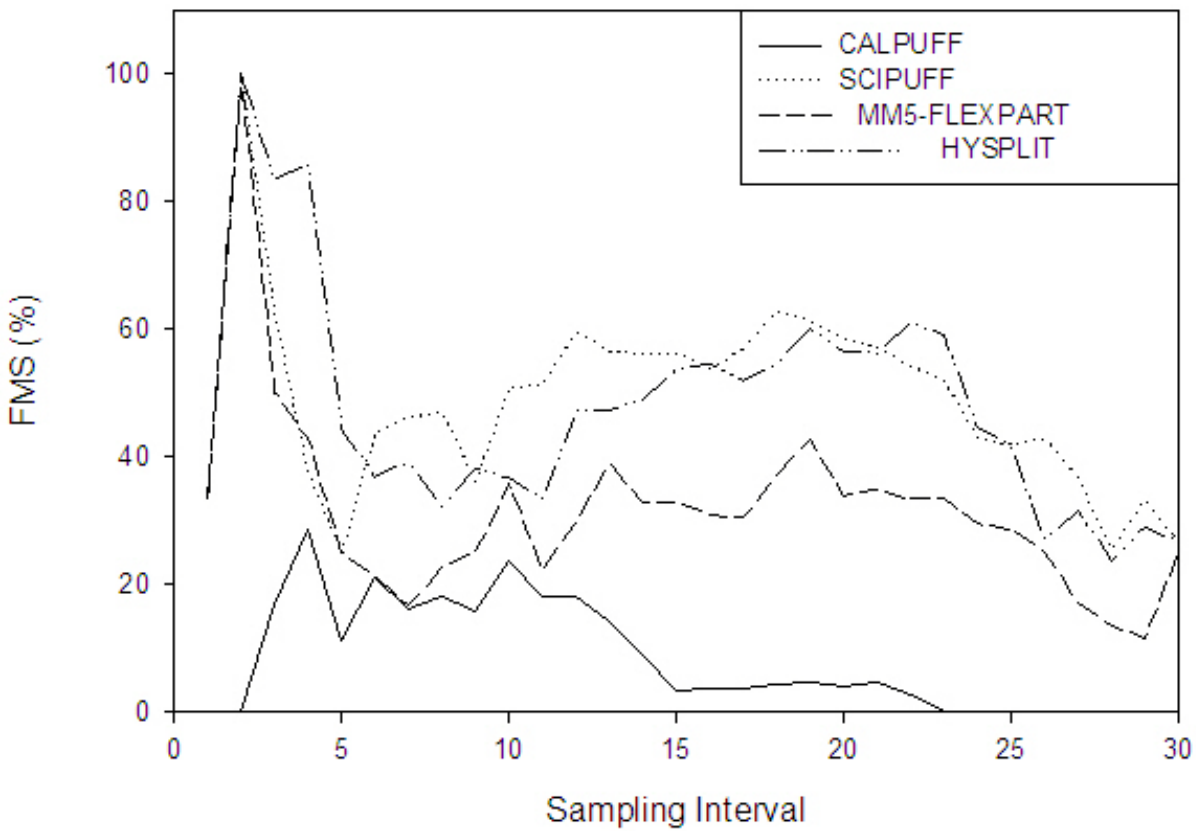


MM5 Evaluation - Temperature





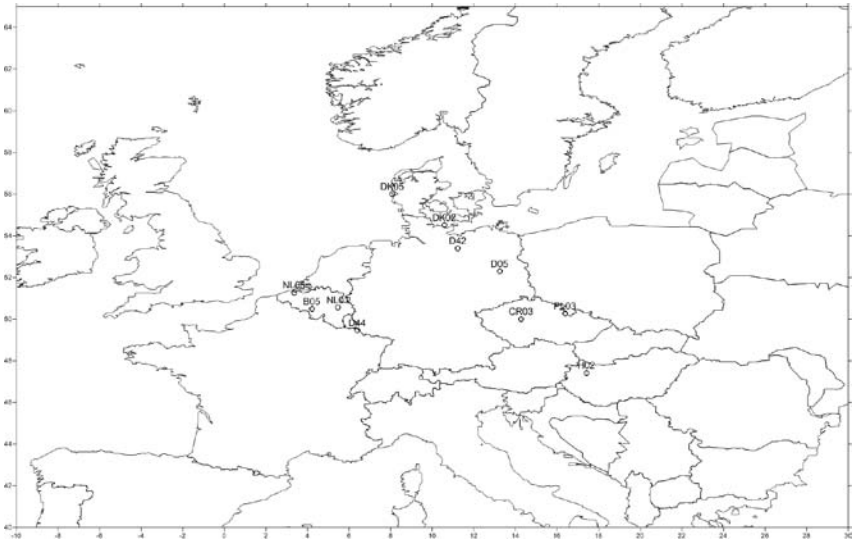
Spatiotemporal Analysis

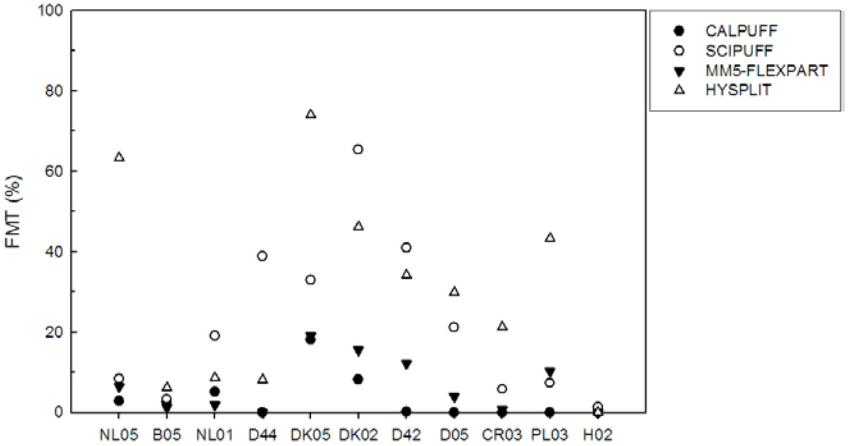


- Temporal evolution of FMS scores help elucidate issues with potential model performance
- SCIPUFF and HYSPLIT exhibit good agreement with the spatial extent of tracer cloud (50% – 60%) out to T+60 hours
- CALPUFF shows similar agreement to FLEXPART out to T+36 hours (20% - 30%), but advection errors cause FMS score to drop dramatically after this point.

OAQPS

Temporal Analysis - Figure of Merit in Time





Global Statistical Analysis

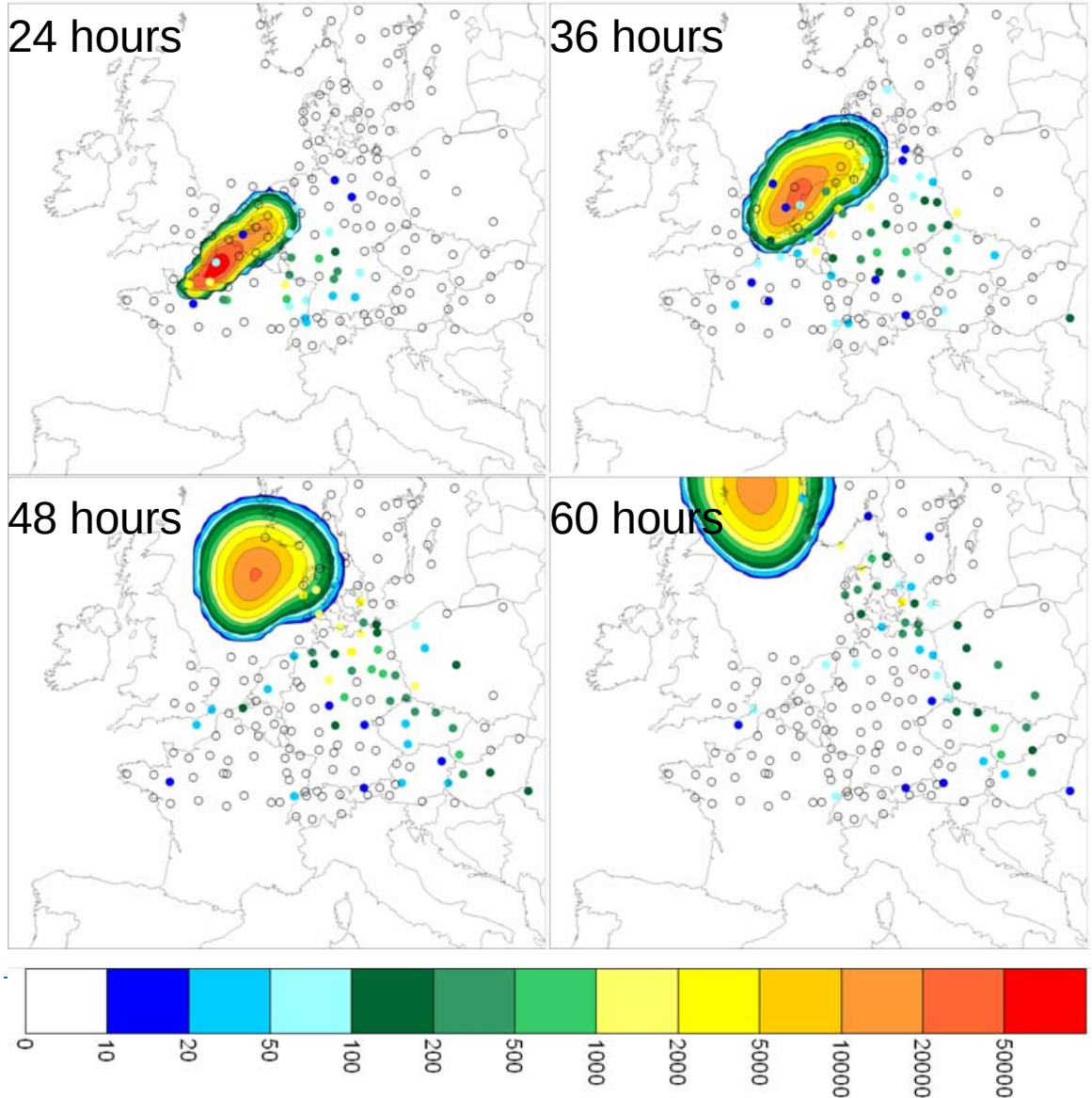
Statistic	CALPUFF	SCIPUFF	HYSPLIT	FLEXPART
PCC	0.17	0.65	0.64	0.45
FB	1.49	1.57	1.00	1.79
FOEX	-34.97	12.16	-18.79	-21.48
KSP	75.00	37.00	53.00	57.00
RANK	0.67	1.77	1.78	1.03

Spatial Analysis

Statistic	Model			
	CALPUFF	SCIPUFF	HYSPLIT	FLEXPART
FMS	13.75	49.93	40.48	29.28
POD	12.56	53.52	48.02	26.65
FAR	60.14	50.41	31.66	48.73
TS	10.56	34.66	39.28	21.27

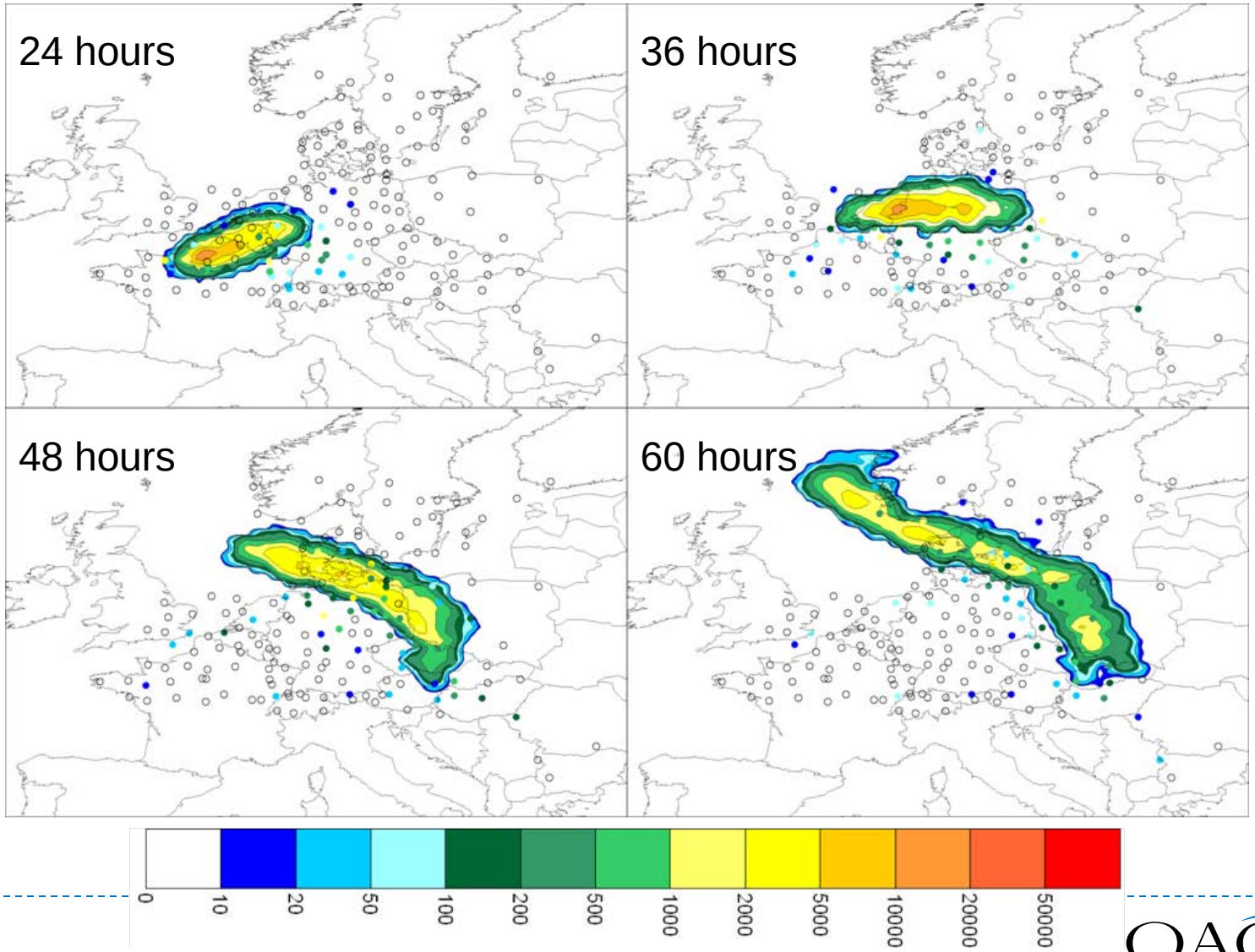
OAQPS

CALPUFF Time Series



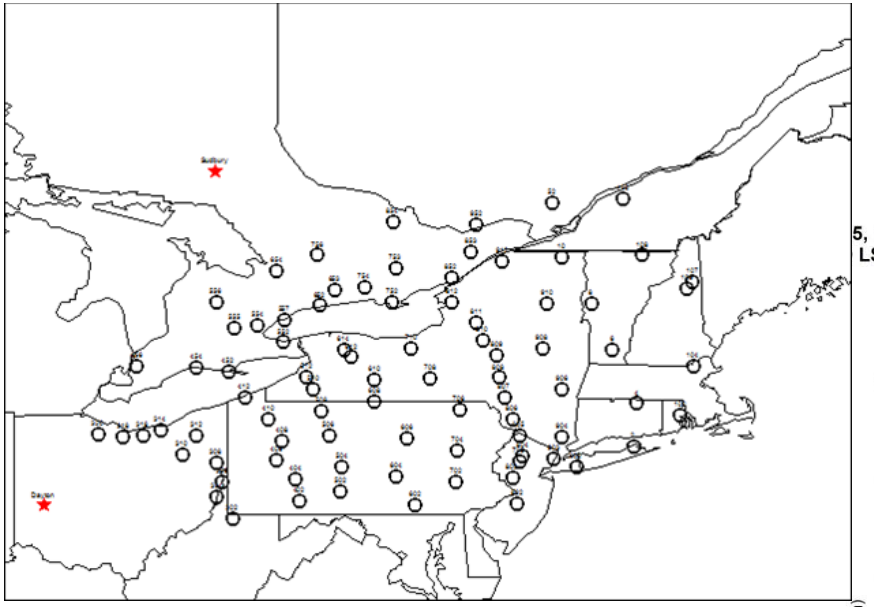
OAQPS

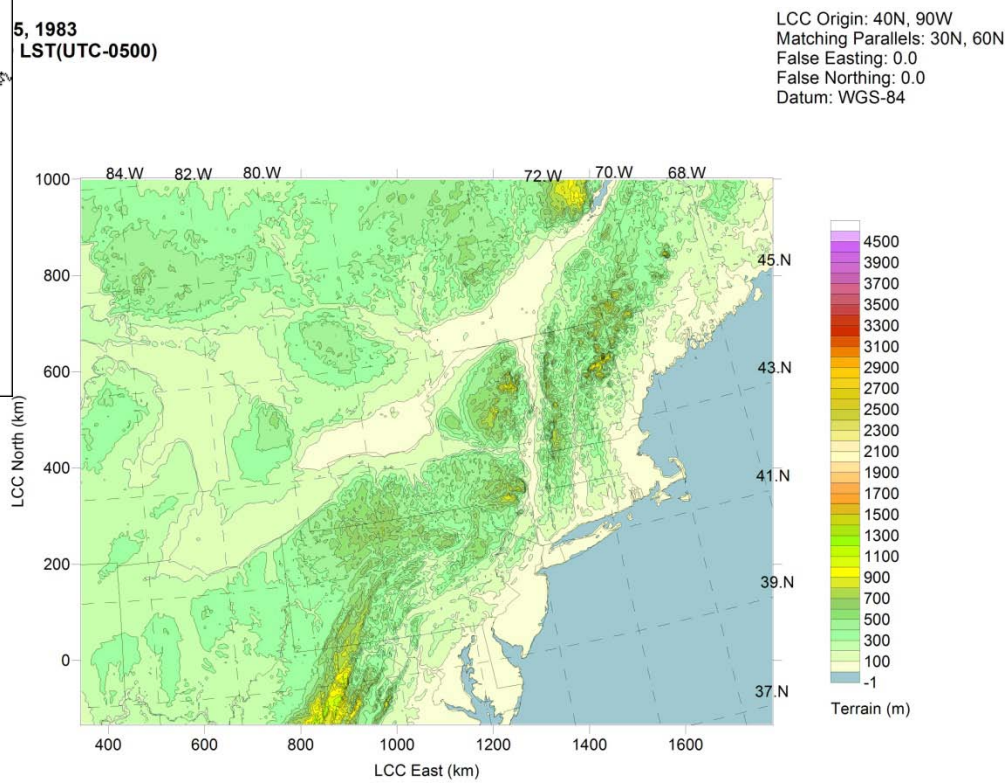
HYSPLIT Time Series



OAQPS

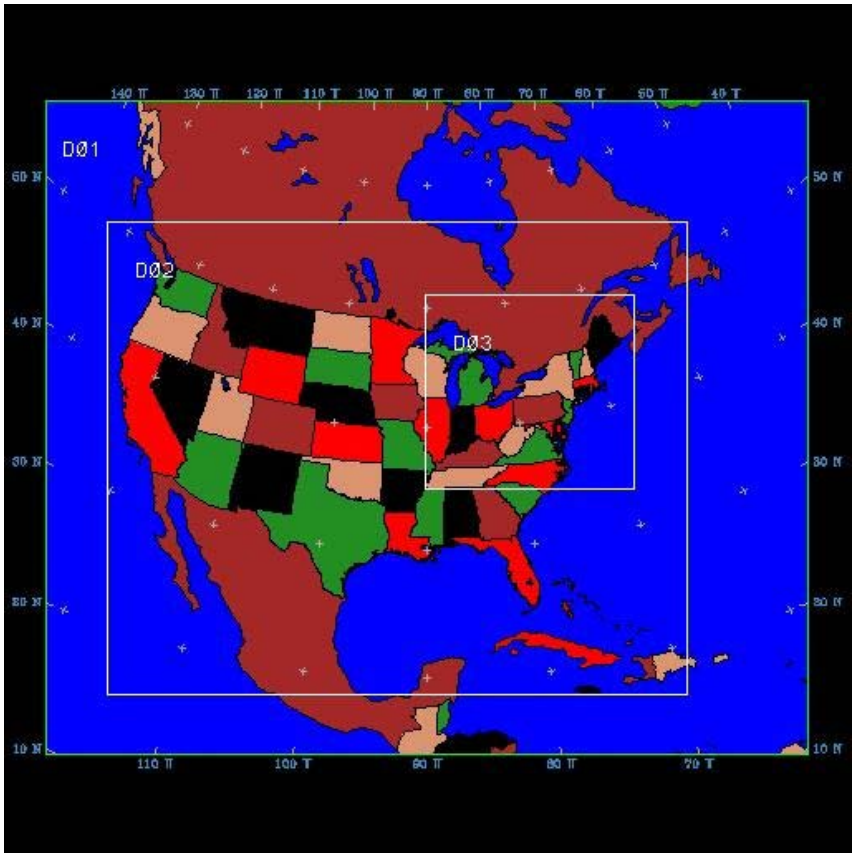
Cross-Appalachian (CAPTEX-83) Monitoring Sites and Model Domain





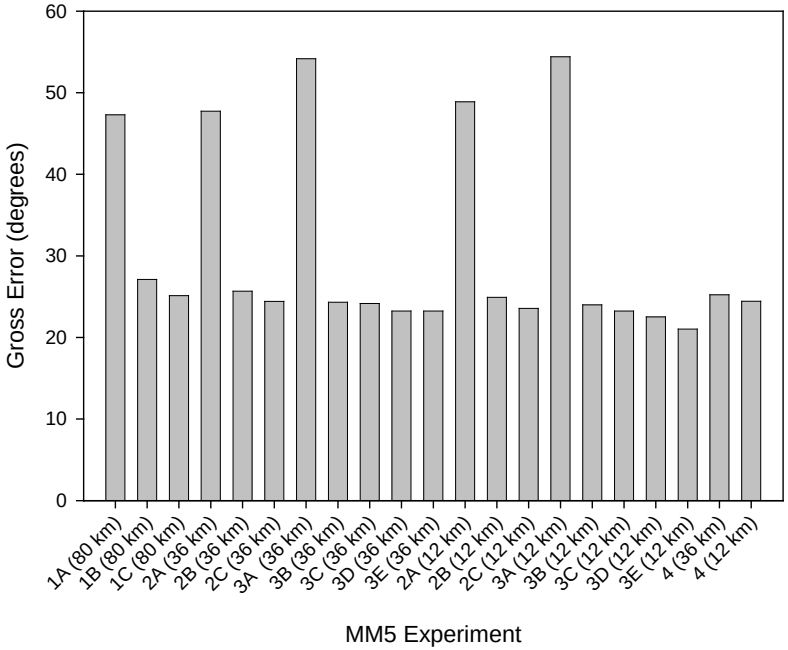
- 7 PFT releases from 2 sites (Dayton, OH and Sudbury, ON) between September – October, 1983
- 84 monitoring sites located across northeast US and southeast Canada.
- 3 hr and 6 hr sample intervals

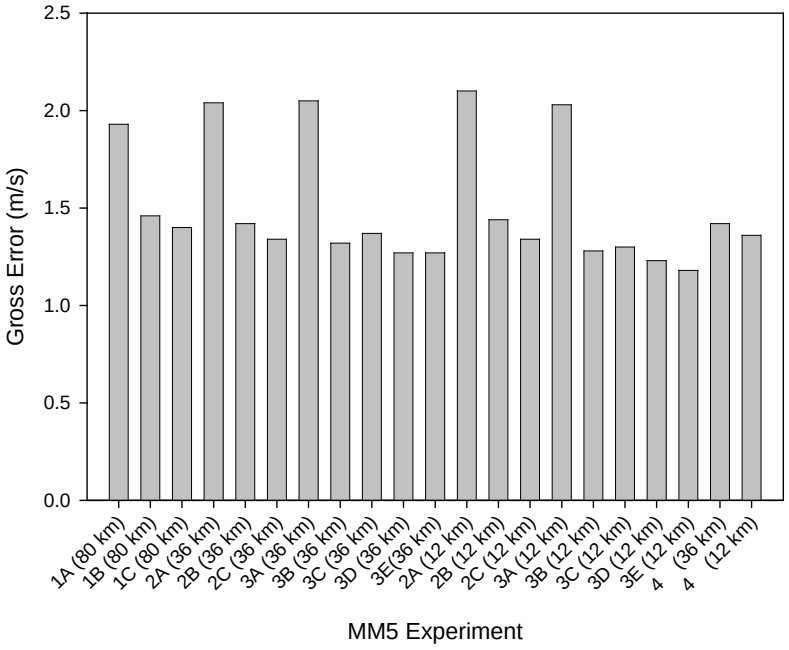
Prognostic Meteorological Model Evaluation



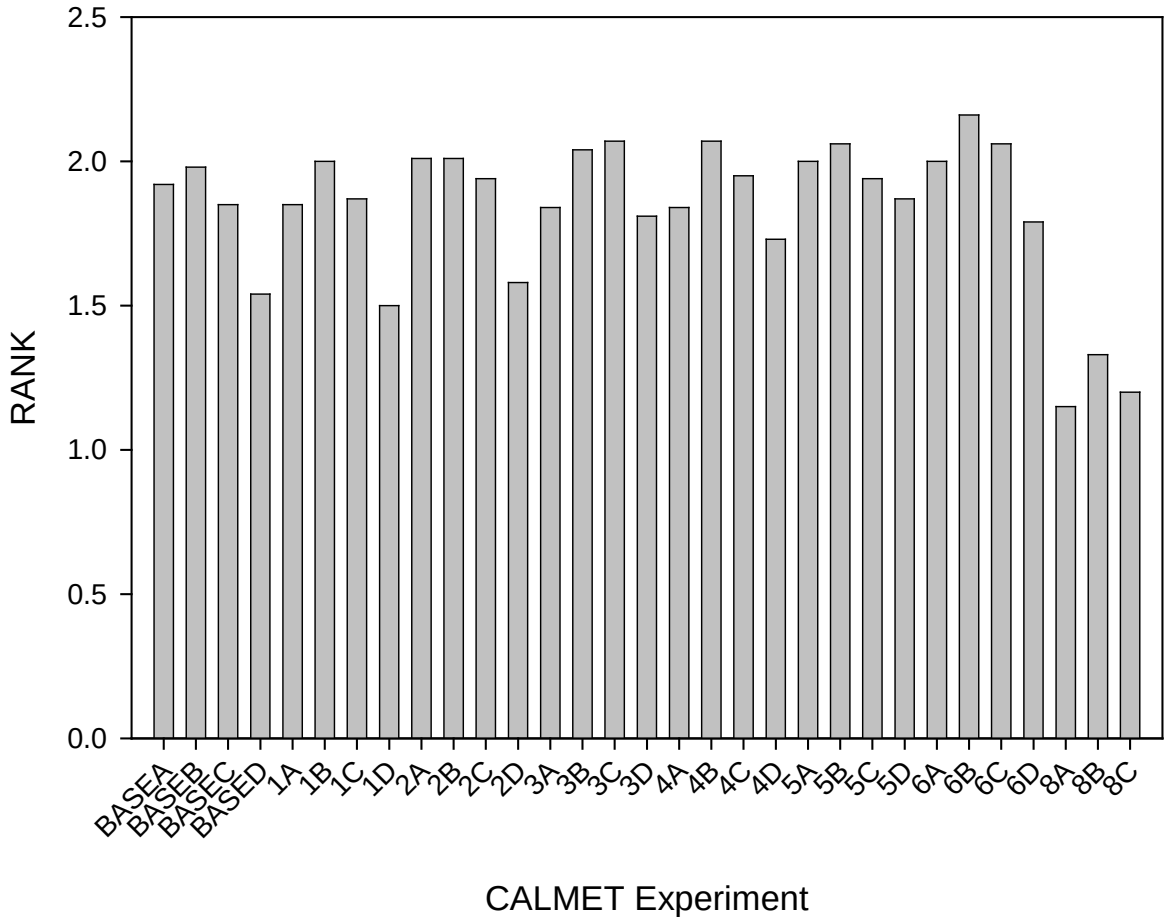
- Goal is to evaluate importance of grid resolution and FDDA strategies on model performance
- 12 MM5 Simulations
  - 80 km resolution, 16 vertical layers
    - 108-, 36-, and 12-km resolution, 33 layer
    - 108-, 36-, and 12-km resolution, 43 layer

MM5 Performance Evaluation - Wind Direction/Speed





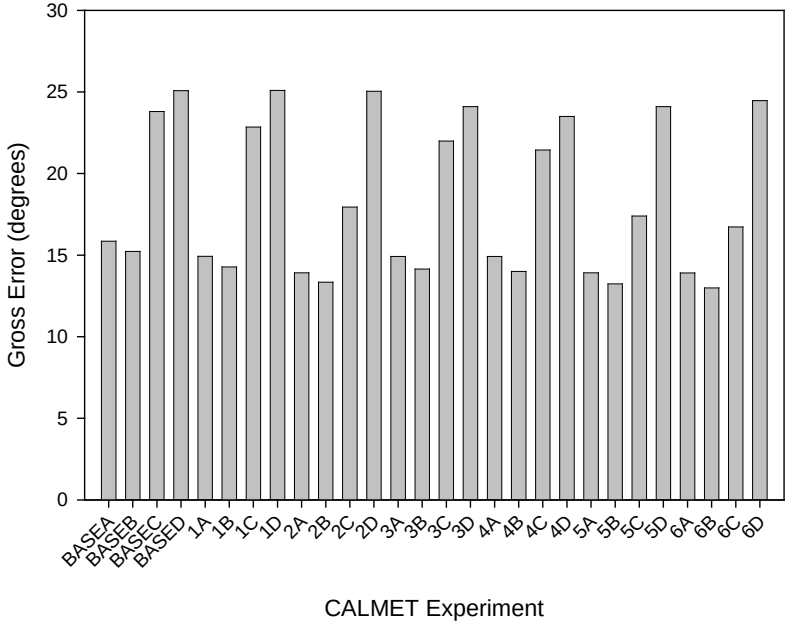
CALPUFF Evaluation – Overall Simulation RANK

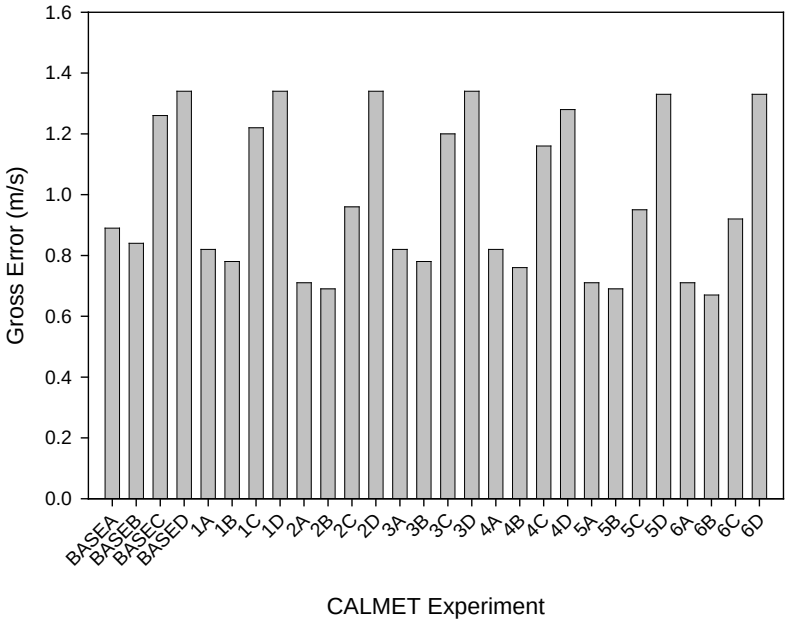


CALMET Experiments

- 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





Overall Model Performance

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

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.