MANE - VU

Mid-Atlantic/Northeast Visibility Union

Trajectory Clustering Techniques

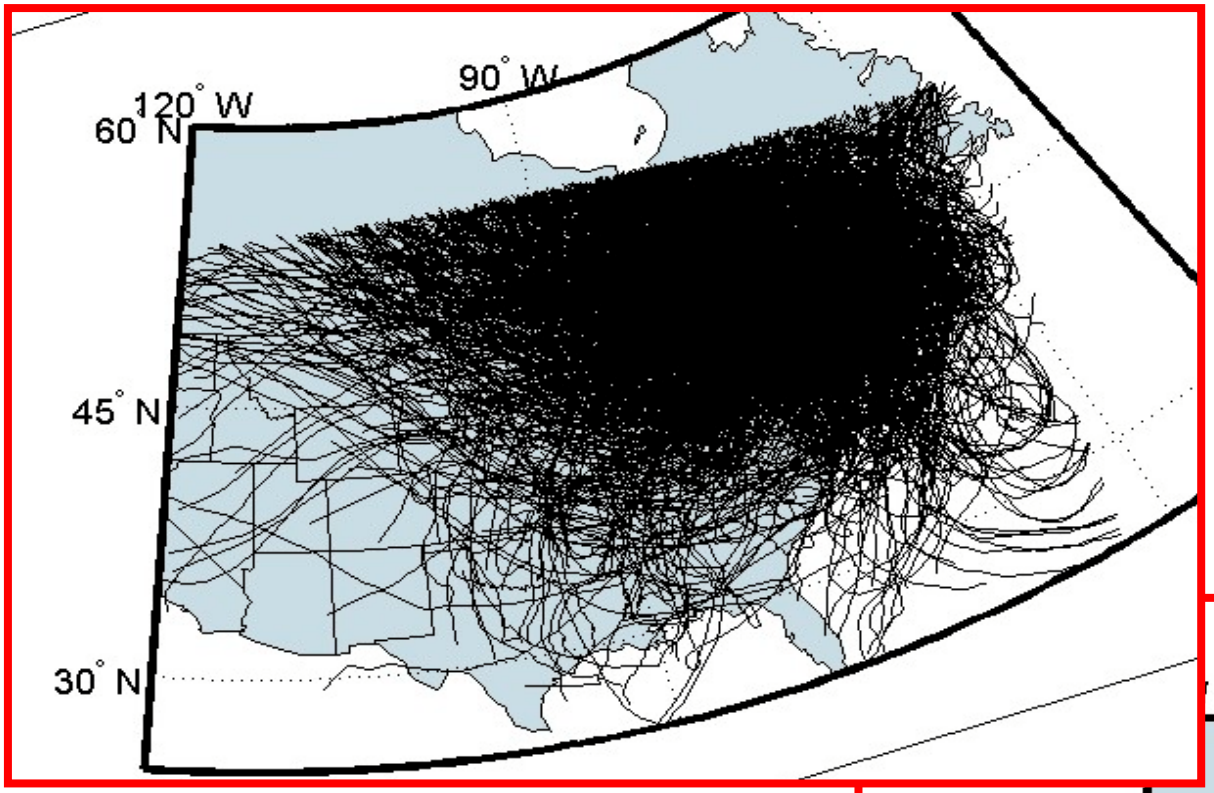
Gary Kleiman, John Graham, Iyad Kheirbek,
Nicolas Hamel, Jaime Lehner, and Ingrid Ulbrich
NESCAUM

Air and Waste Management Association
Visibility Specialty Conference
October 27, 2004

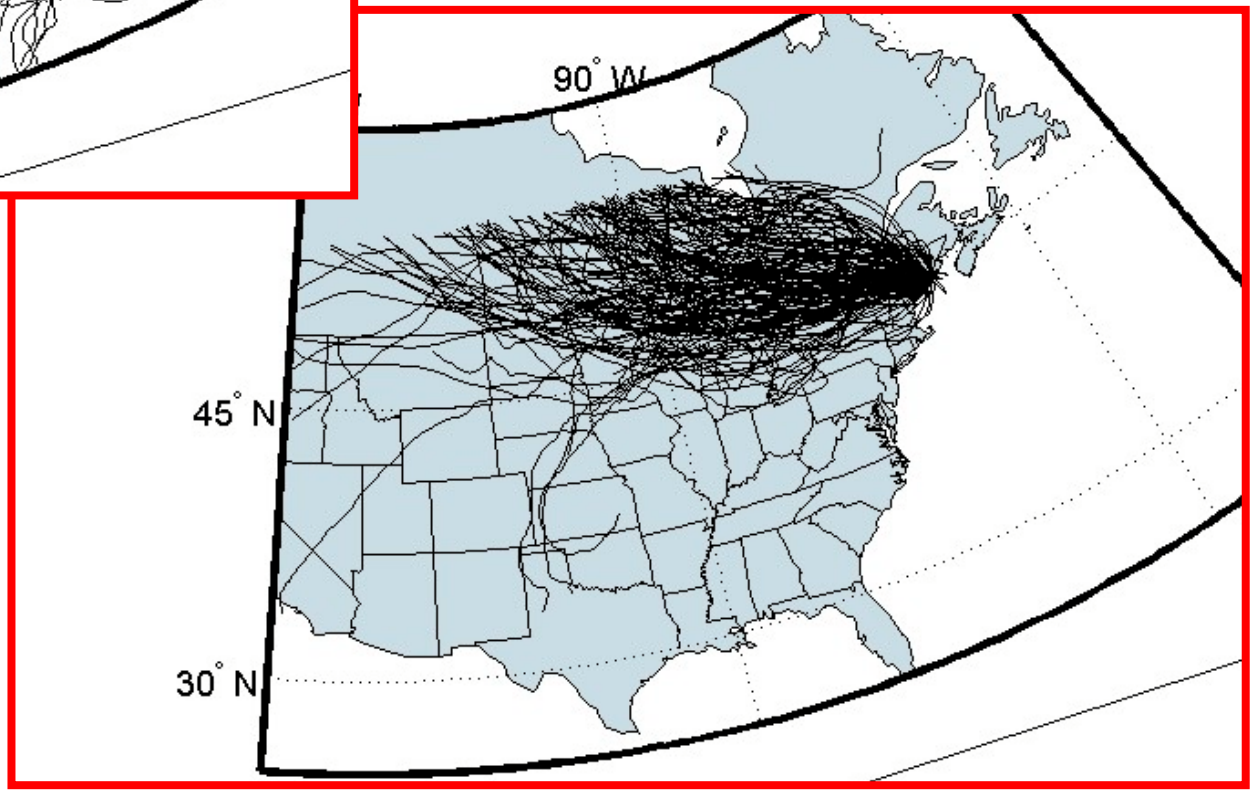
Acadia 1997

NESCAUM
Northeast States for Coordinated Air Use Management

ALL TRAJECTORIES



CLUSTER 3 TRAJECTORIES



PATH Analysis

- Patterns in Atmospheric Transport History
- Based on prior work by Jenny Moody (UVA) at Harvard Forest in Central MA
- Groups similar trajectories in 3-dimensions

PATH Algorithm

- Similarity quantified as distance between trajectories:

$$D_{ij} = \sqrt{\sum_{k=1}^{n} (Z_{ik}^{lat} - Z_{jk}^{lat})^2 + (Z_{ik}^{lon} - Z_{jk}^{lon})^2 + (Z_{ik}^{press} - Z_{jk}^{press})^2}$$

(Z represent **normalized** coordinates with 0 mean and standard deviation of 1)

Cluster Definition

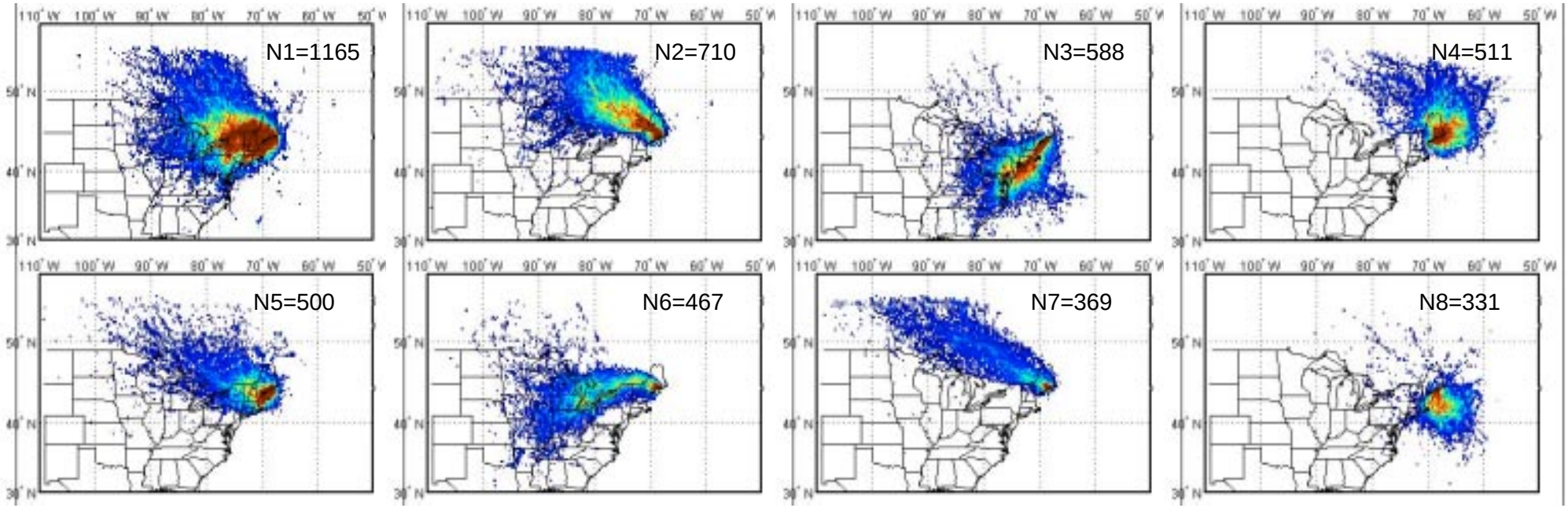
- Clusters are defined by establishing a radius of proximity R:

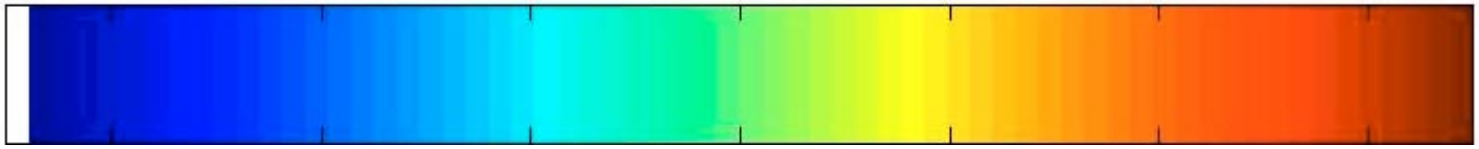
$$C(T_i) = \{ T_j \} \ni D_{ij} \leq R$$

The maximal cluster (largest number of elements) is removed from the analysis and the process is repeated

Top 8 Clusters for Acadia

Acadia, 1997-2002, 48hr, 500m, R=6



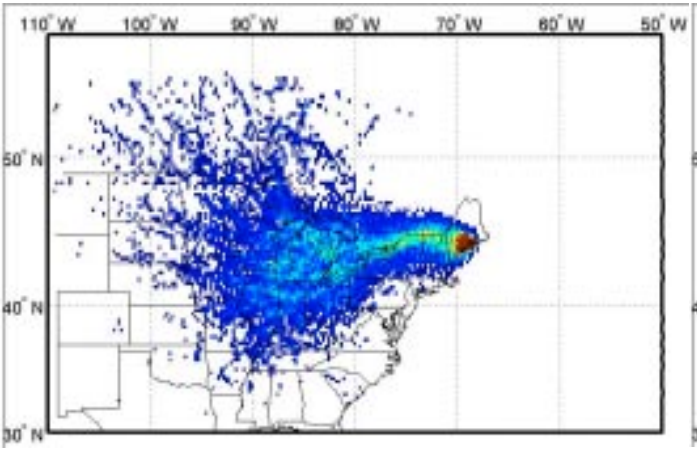


Free Parameters

- Radius of proximity
- Height of starting point
- Length of cluster calculation (e.g. 24 hr, 48 hr, 120hr)
- Temporal range (e.g. 97 vs. 97-02)
- Receptor site
- Resolution/completeness of met data

Starting height

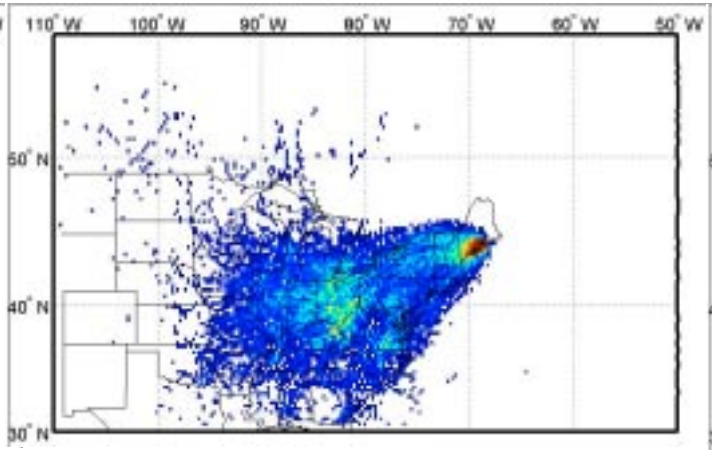
Acadia, 1997-2002, 48hr, 200/500/1000m, R=10



Cluster 5

Receptor Level= 200m

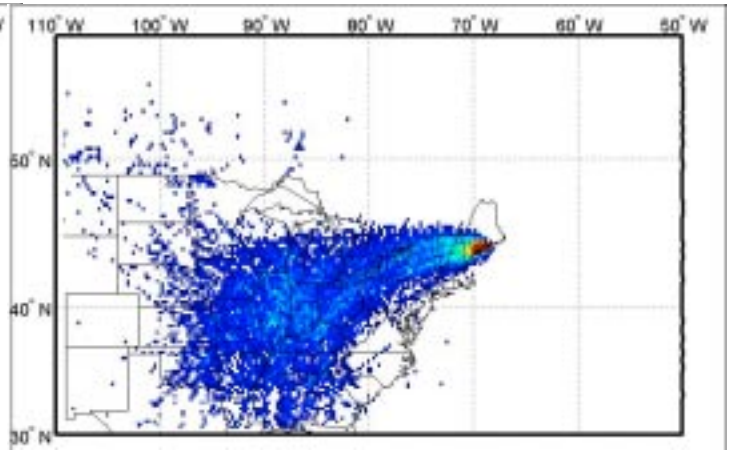
Mean Height= 892 m



Cluster 4

Receptor Level= 500m

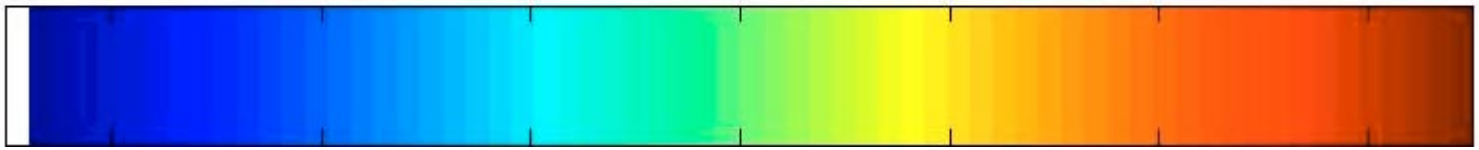
Mean Height= 867 m



Cluster 5

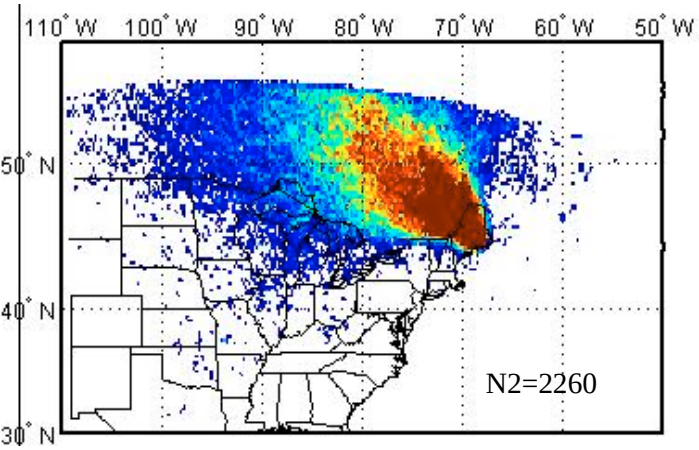
Receptor Level= 1000m

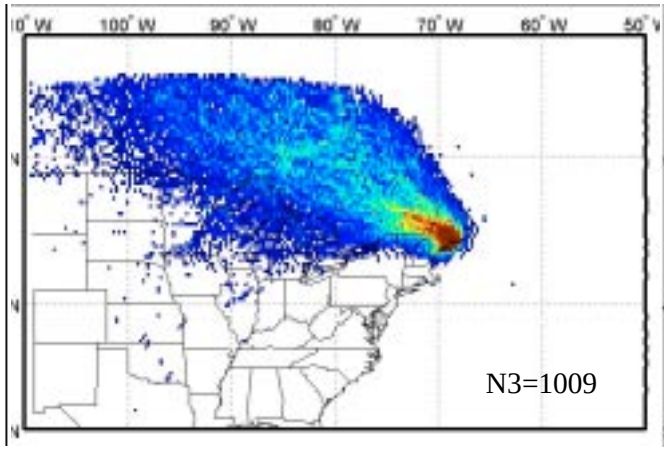
Mean Height= 1247 m

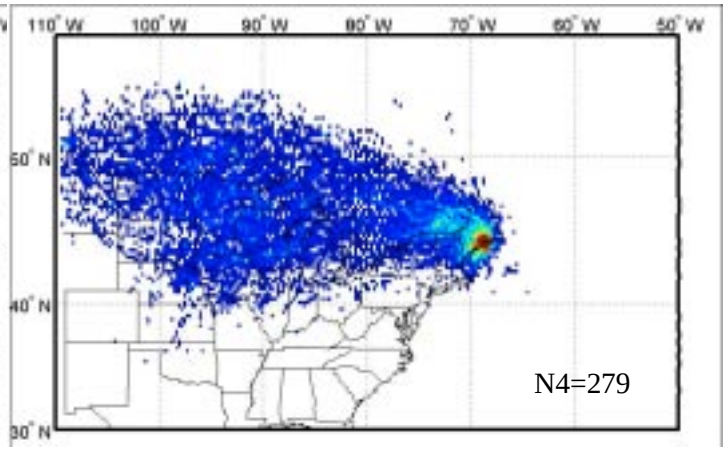


Length of cluster

Acadia, 1997-2002, 24/48/120hr, 500m, R=6,10,16







Total Trajectories in 24 hr
Analysis = 14235 (99-02)

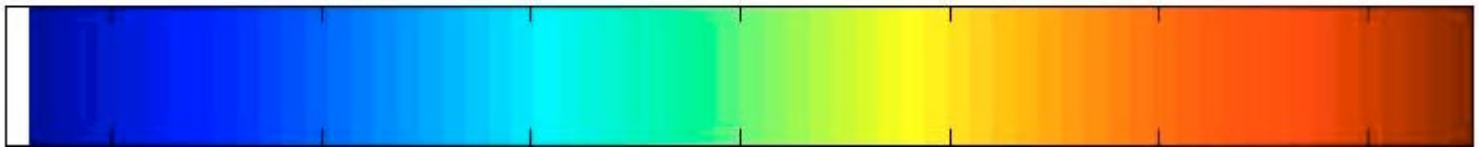
Mean Height= 1286 m

Total Trajectories in 48 hr
Analysis = 10592

Mean Height= 1916 m

Total Trajectories in 120 hr
Analysis = 4429

Mean Height= 1692 m

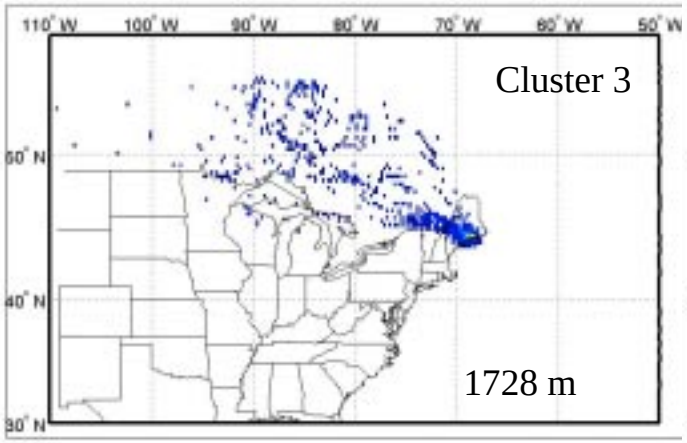


Temporal range

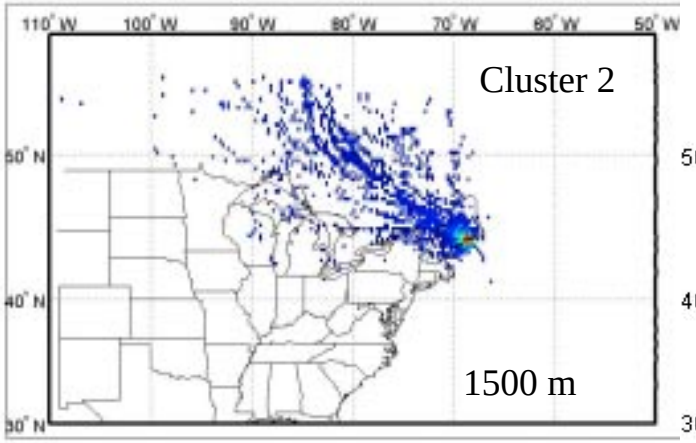
NESCAUM
Northeast States for Coordinated Air Use Management

Acadia, 97/98/99/00/01/02/97-02, 48hr, 500m, R=10

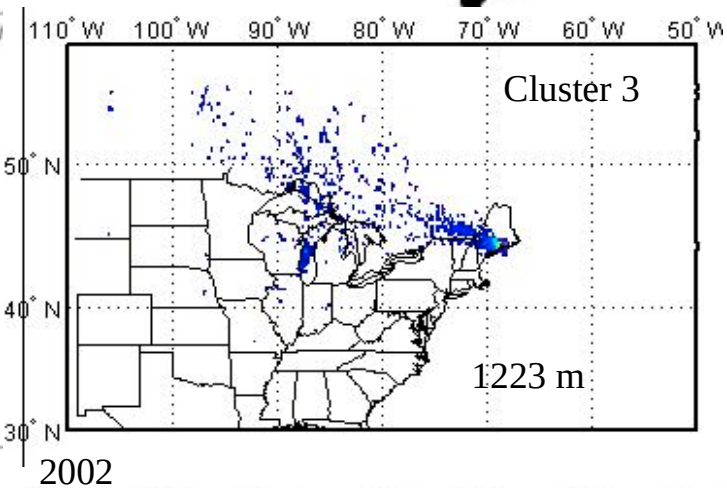
1997



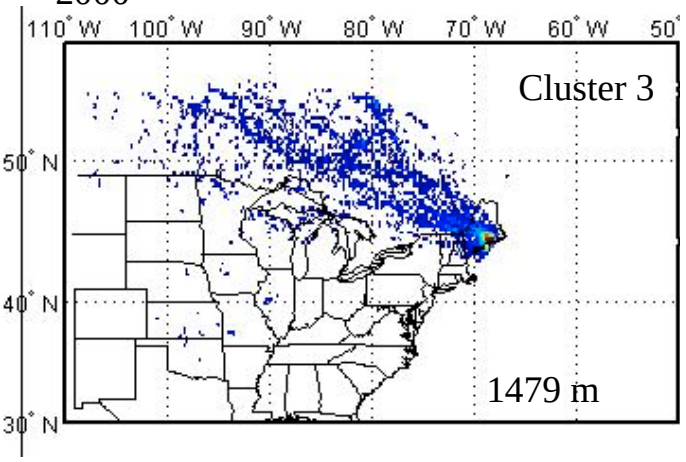
1998



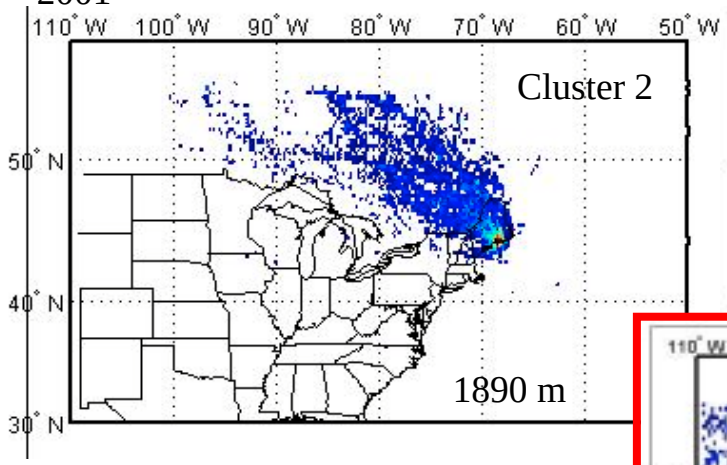
1999



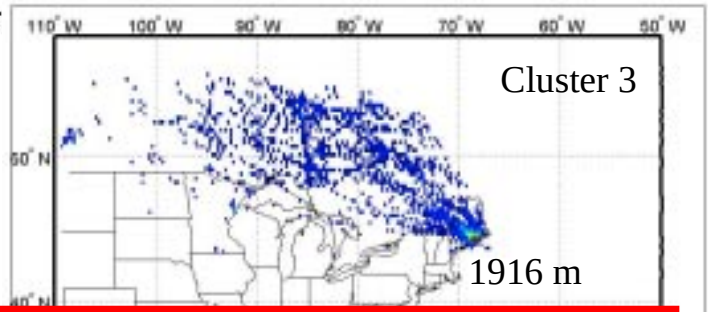
2000



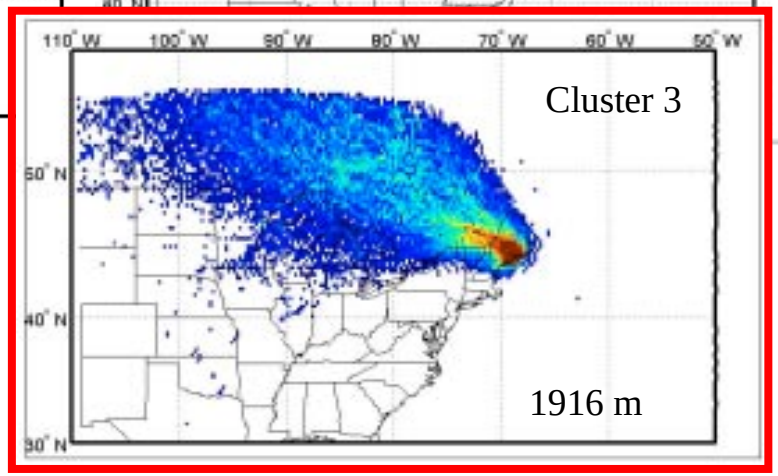
2001



2002



1997-2002

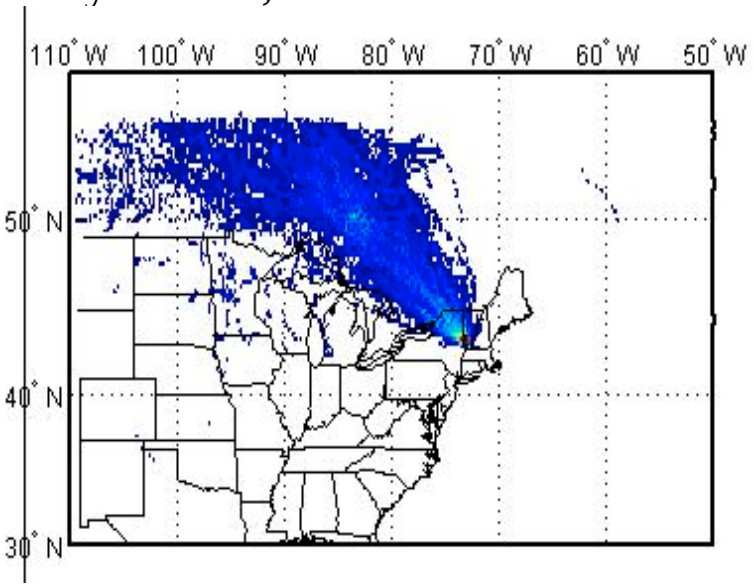


Atmospheric Modes vs. Clusters

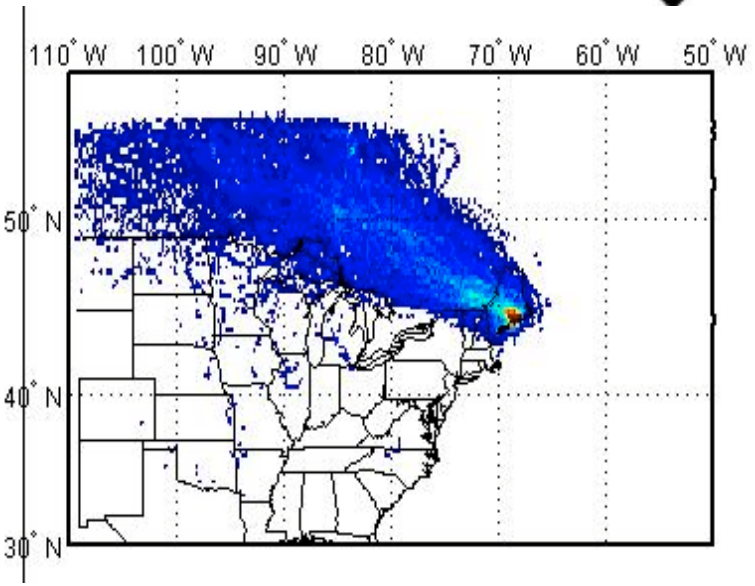
- *Clusters* are defined as groupings of similar trajectories during a specific time period
- *Atmospheric modes* describes the underlying meteorological phenomena that explain the observed clusters

Northwest Fast (NWF)

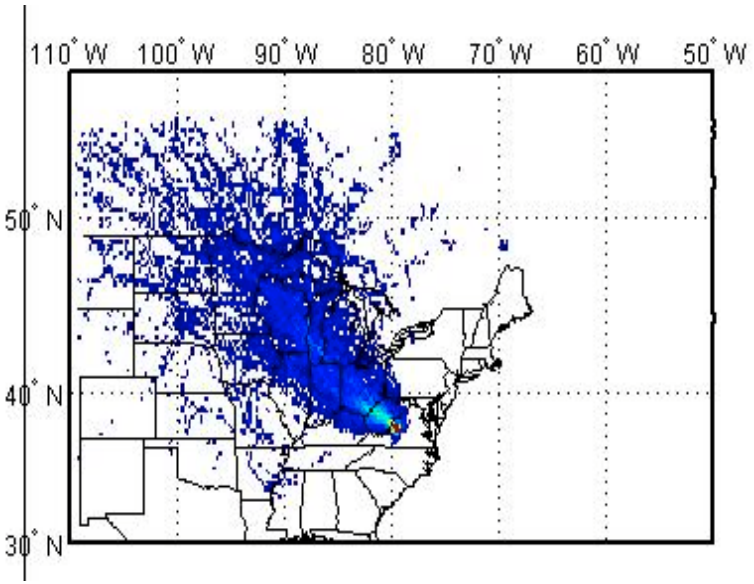
Lye Brook, VT



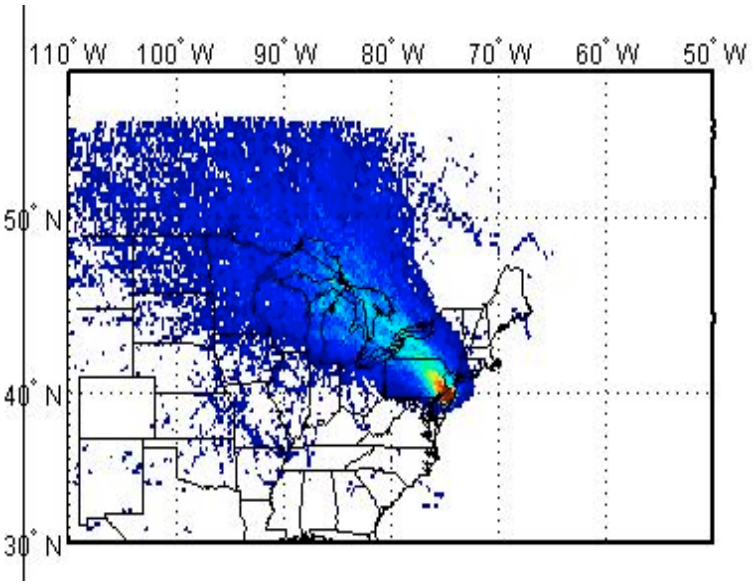
Acadia, ME



Jefferson/James River Face, VA

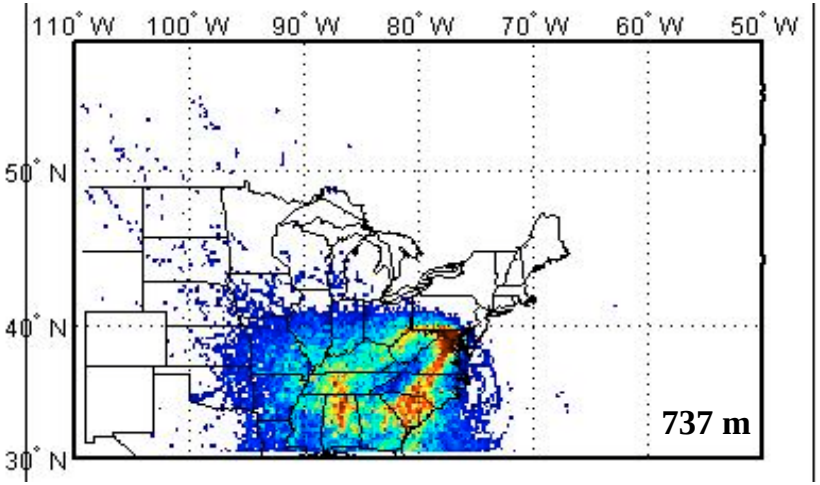


Brigantine, NJ

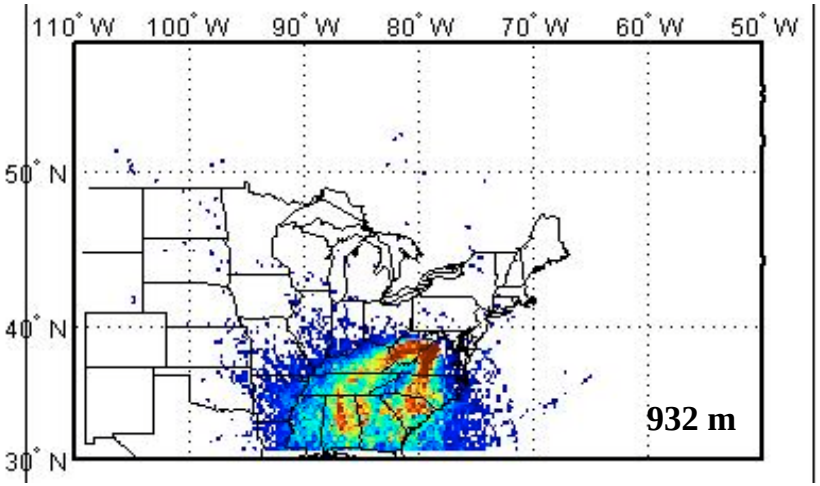


Southwest Bi-modal (SWB)

Baltimore



Shenandoah



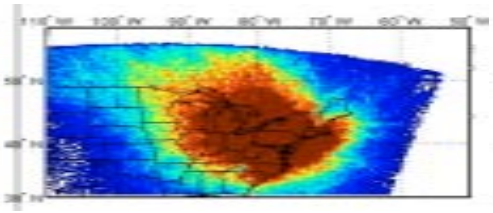
- Represents a mix of southwest coastal and southwest inland

Predominant Meteorological Pathways and Influence on Air Quality

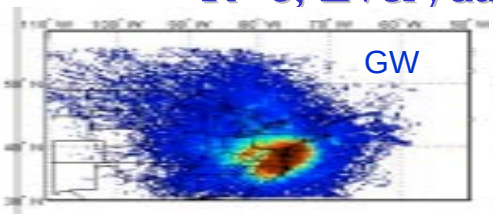
- Back-trajectory Clusters
- Pollution-Date Database
- Correlate dates to establish relationships between atmospheric modes and pollution tendencies

Brigantine 97-02 EDAS- 48 hrs

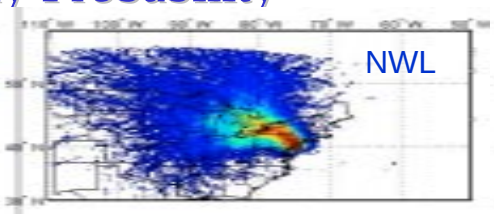
R=6, Everyday Probability



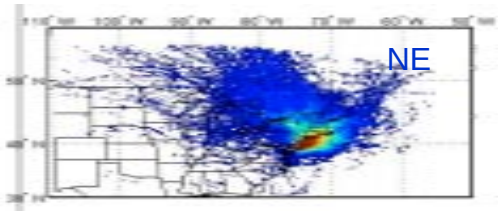
Cluster 0
N=12993 , H=1042 m



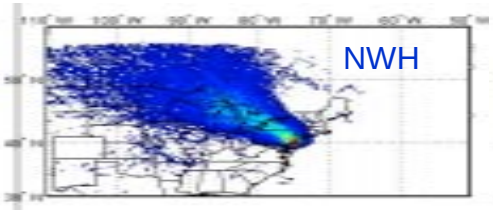
Cluster 1
N=1830 , H=479 m
15.41, 6.29, 107.92



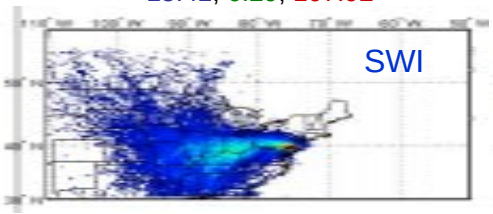
Cluster 2
N=1101, H=923.0 m
13.57, 6.10, 95.51



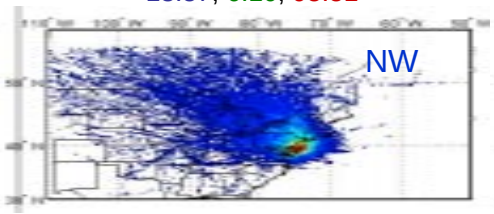
Cluster 3
N=708, H=512.1 m
8.63, 3.17, 63.58



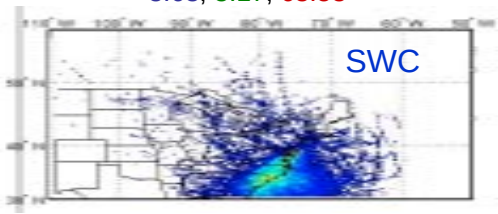
Cluster 4
N=692, H=1586.6 m
11.55, 4.10, 81.23



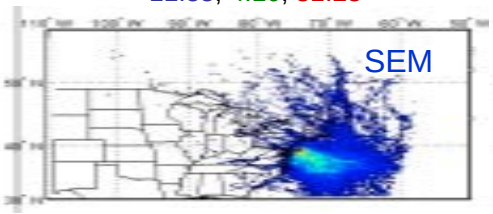
Cluster 5
N=655, H=871.2 m
14.48, 5.32, 103.44



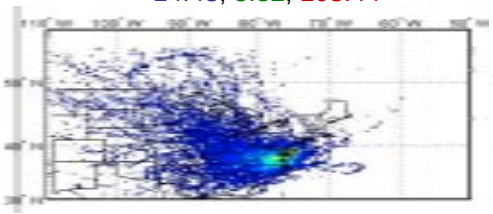
Cluster 6
N=624, H=1308.5 m
11.63, 4.01, 79.55



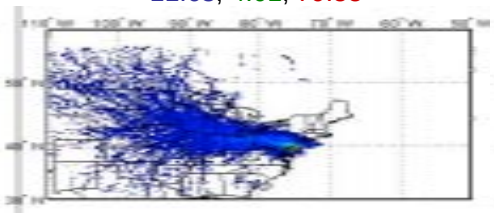
Cluster 7
N=619, H=479.9 m
10.67, 3.94, 73.66



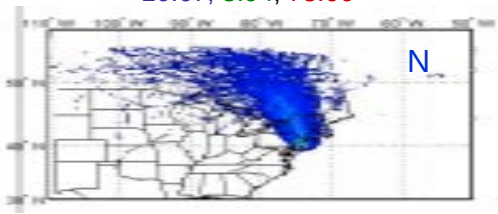
Cluster 8
N=402, H=441.8 m
7.42, 3.33, 59.62



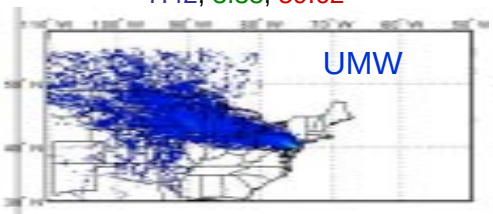
Cluster 9
N=375 , H=1319.7 m
14.99, 5.90, 107.12



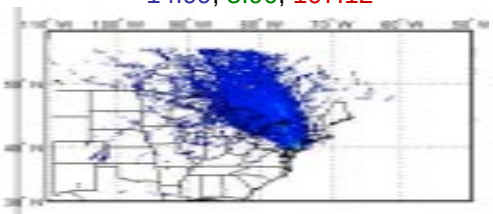
Cluster 10
N=357 , H=1800.2 m
15.49, 6.48, 114.24



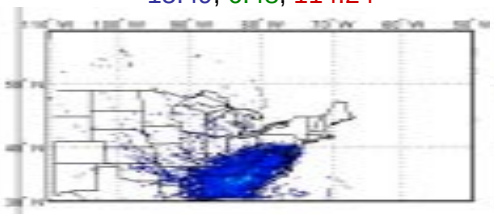
Cluster 11
N=329, H=1688.2 m
9.91, 2.07, 61.94



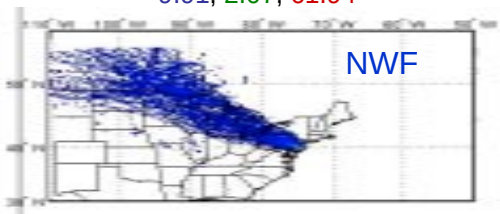
Cluster 12
N=282, H=1226.0 m
10.86, 3.32, 79.65



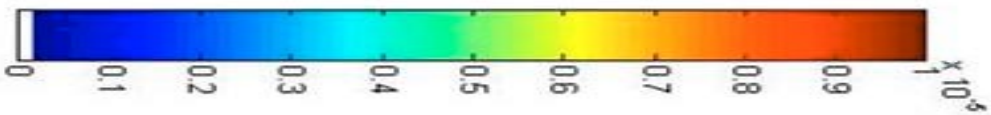
Cluster 13
N=267, H=762.4 m
7.00, 2.39, 53.80



Cluster 14
N=223, H=693.6 m
12.61, 4.26, 82.92



Cluster 15
N=215, H=1948.6 m
9.61, 3.27, 71.54



1997-2002 IMPROVE Averages:
Total $PM_{2.5}$ (µg/m^3)
$PM_{2.5}$ SO_4 (µg/m^3)
Reconstructed Extinction (Mm^{-1})

Trajectory Analysis: Multiple Metrics

Everyday Residence-time Probability

$$EP = \left(n_{ij} / N \right)$$

n_{ij} = total endpoints passing through grid cell i, j

N = total endpoints passing through all grid cells from all trajectories

High Day Residence-time Probability

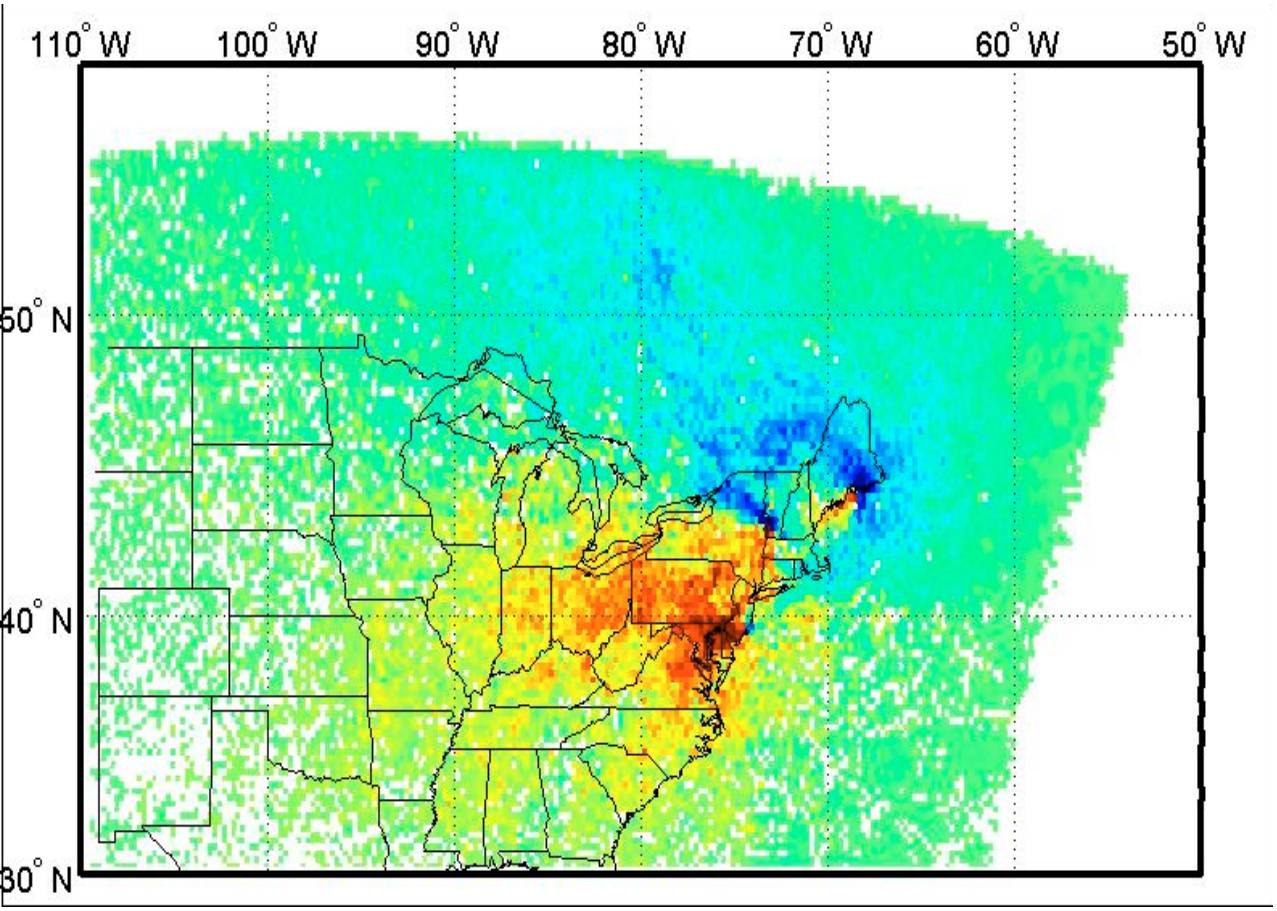
$$HP = \left(m_{ij} / M \right)$$

m_{ij} = total high day endpoints passing through grid cell i, j

M = total high day endpoints passing through all grid cells from high day trajectories

Trajectory Analysis: Multiple Metrics

Incremental Probability

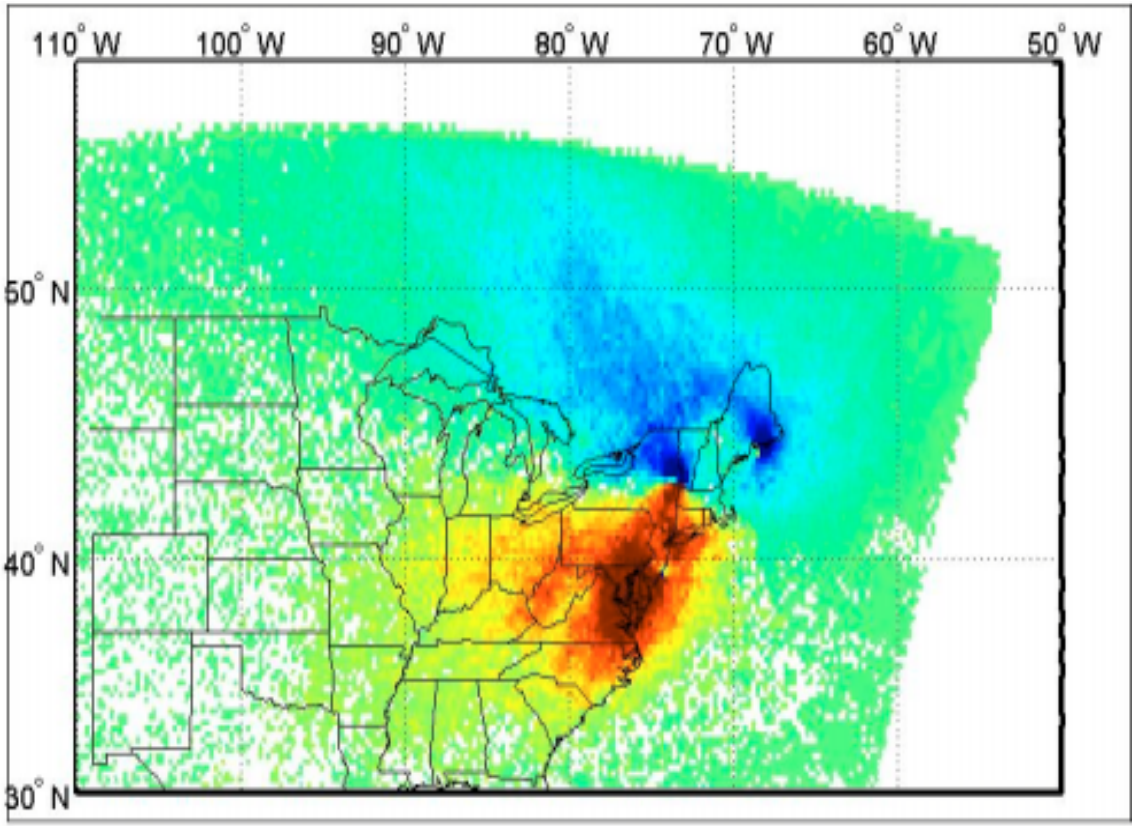


$$IP = HP - EP$$

3-site average:
Acadia, Lye Brook, and Brigantine

Trajectory Analysis: Multiple Metrics

Clustered Incremental Probability



$$CIP = \sum_{k \in K} RP_k - EP$$

K = Subset of clusters, 20% worst pollution days

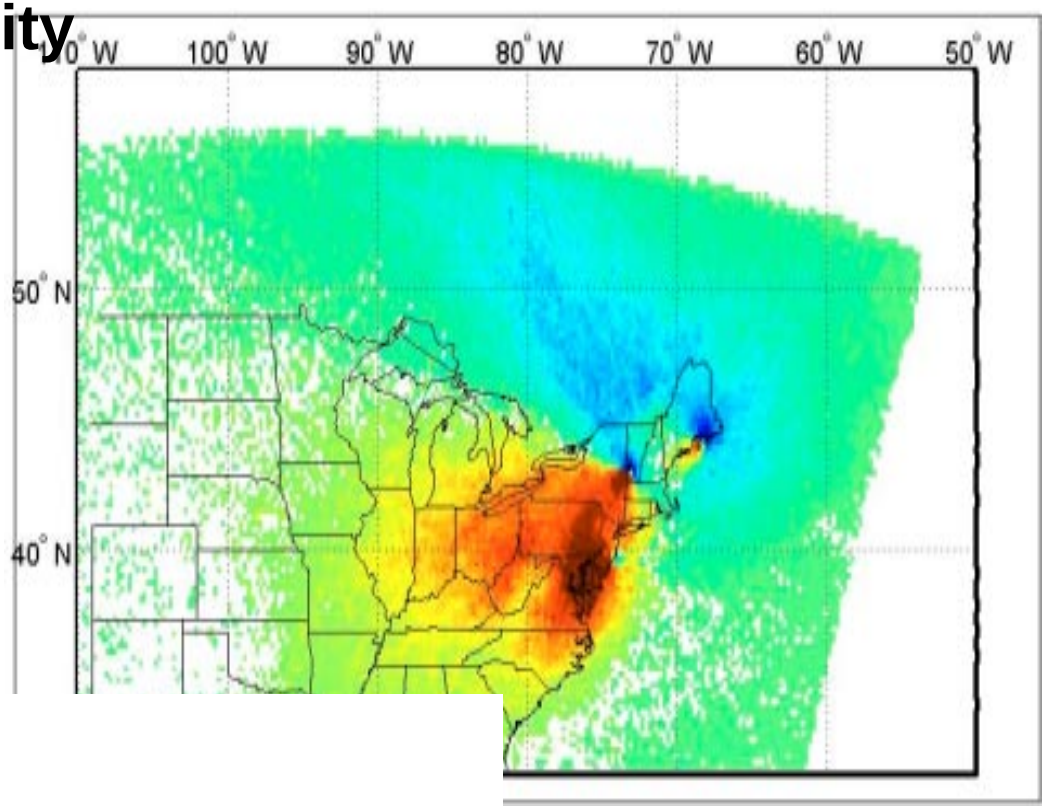
RP_k = Residence - time probability for cluster k

3-site average:
Acadia, Lye Brook, and Brigantine

Trajectory Analysis: Multiple Metrics

NESCAUM
Northeast States for Coordinated Air Use Management

Cluster-Weighted Probability



$$CWP = \frac{1}{\overline{C}}\left(\sum_{i=1}^{L}(\overline{C})_i \cdot RP_i - \overline{C} \cdot EP\right)$$

$L = \text{total number of clusters calculated}$

$(\overline{C})_i = \text{Average concentration (obs associated with cluster i)}$

$\overline{C} = \text{Average concentration (all days)}$

3-site average:
Acadia, Lye Brook, and Brigantine

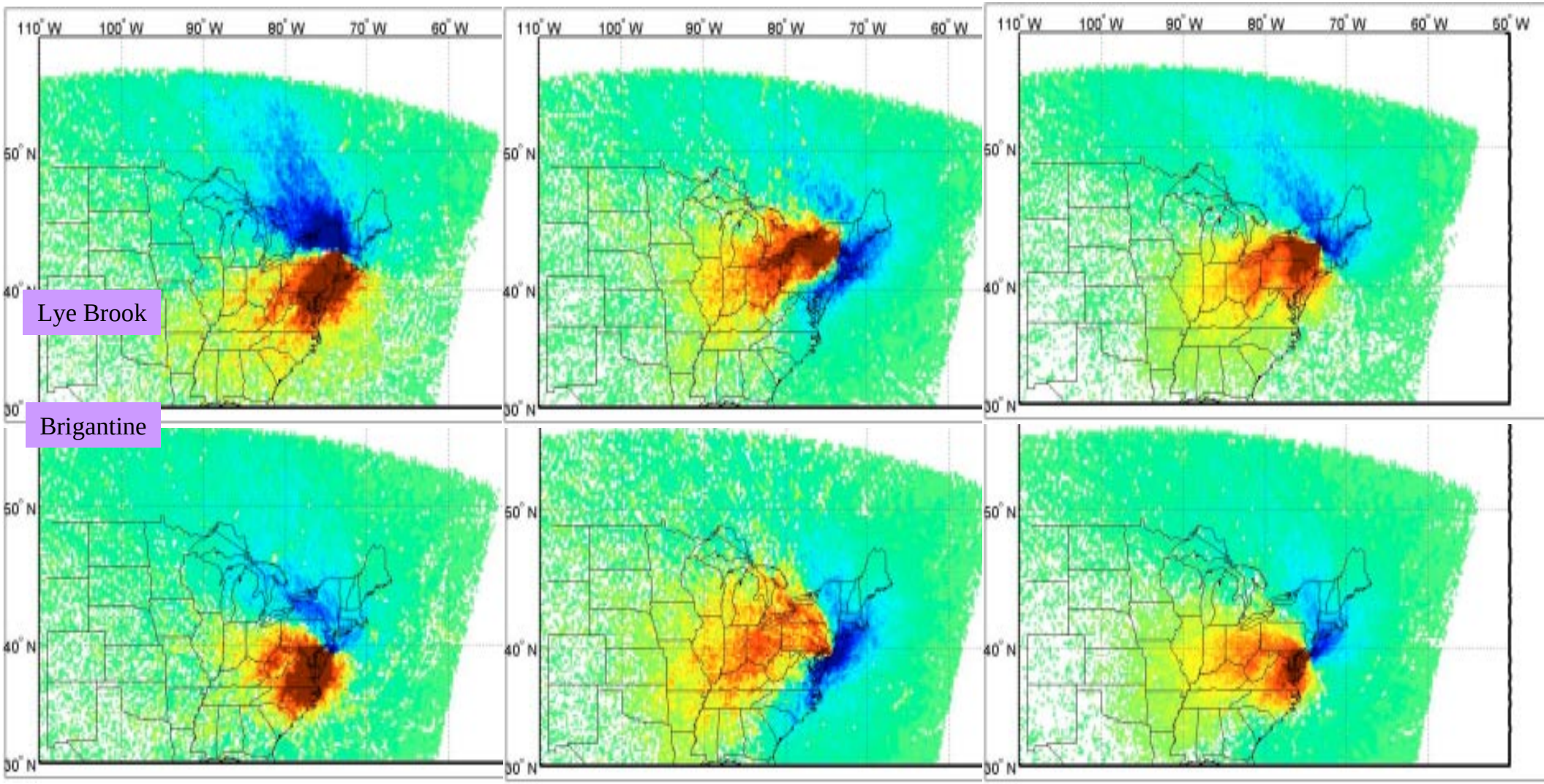
Distribution of CIP for Sulfate Transport

NESCAUM
Northeast States for Coordinated Air Use Management

Worst 20% days

Second worst 20% days

Worst 40% days



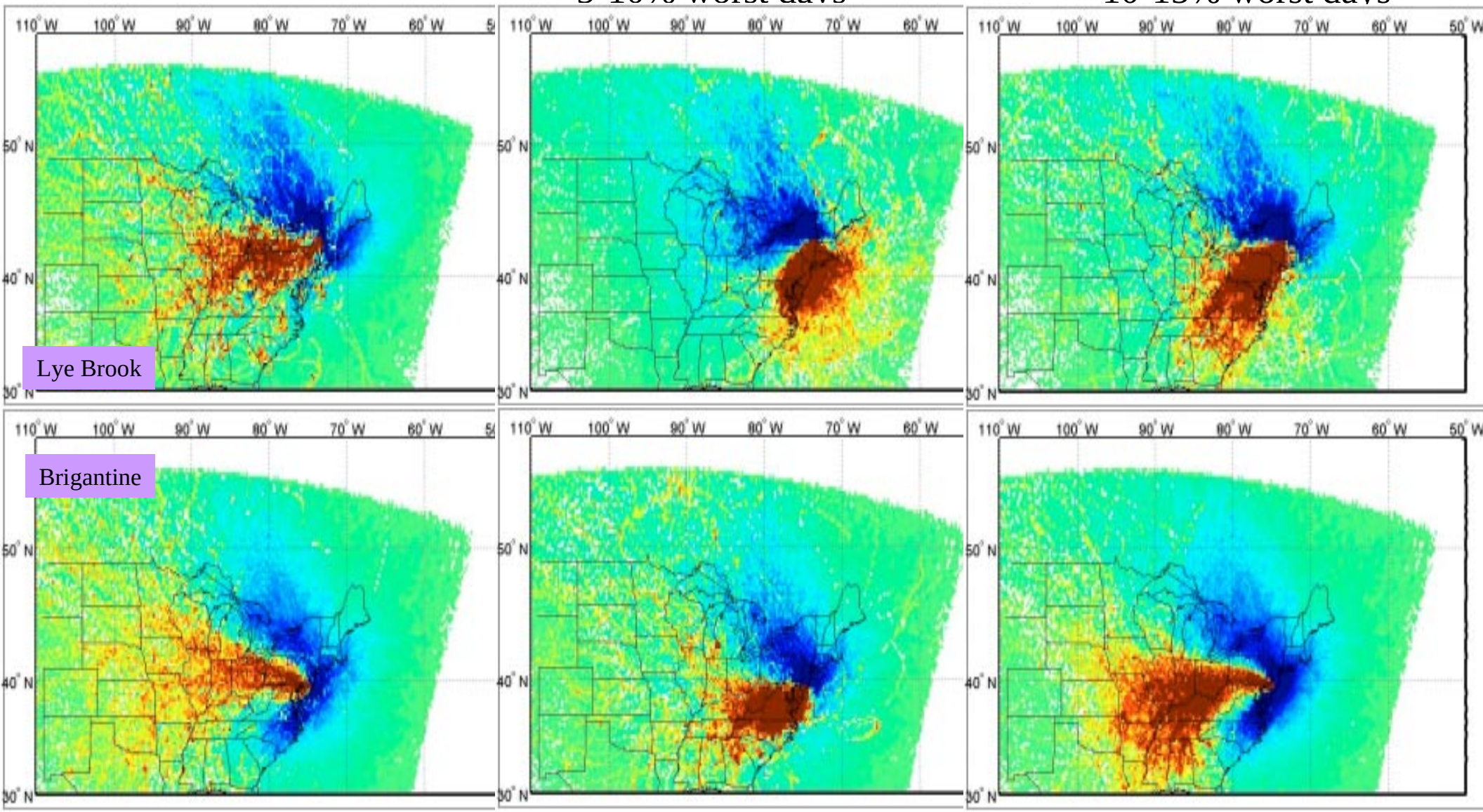
NESCAUM
Northeast States for Coordinated Air Use Management

20 Percent is a lot of days…

0-5% worst days

5-10% worst days

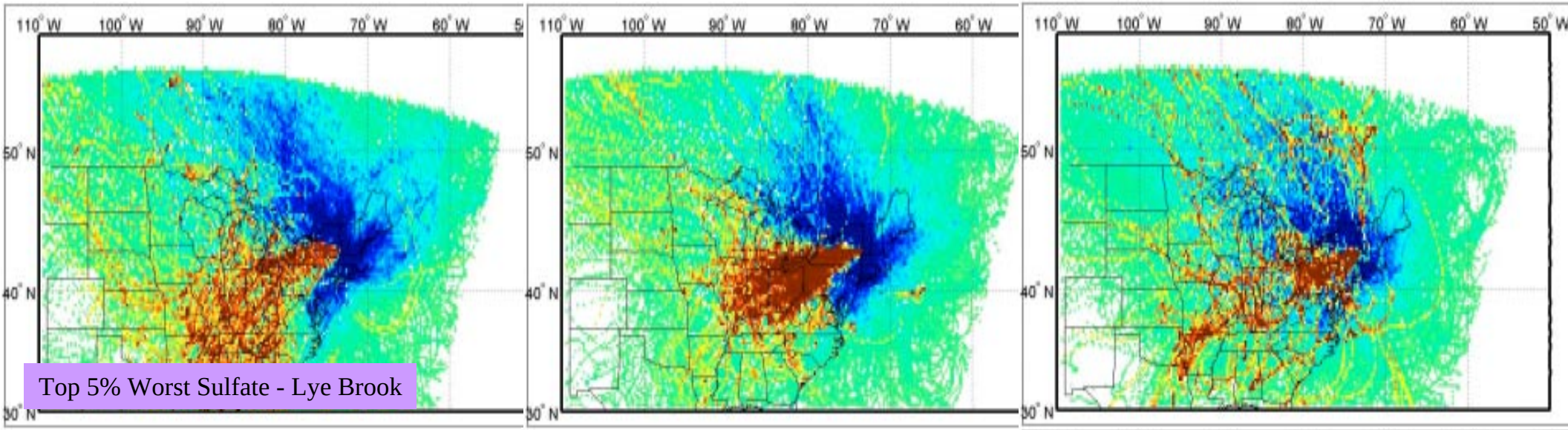
10-15% worst days



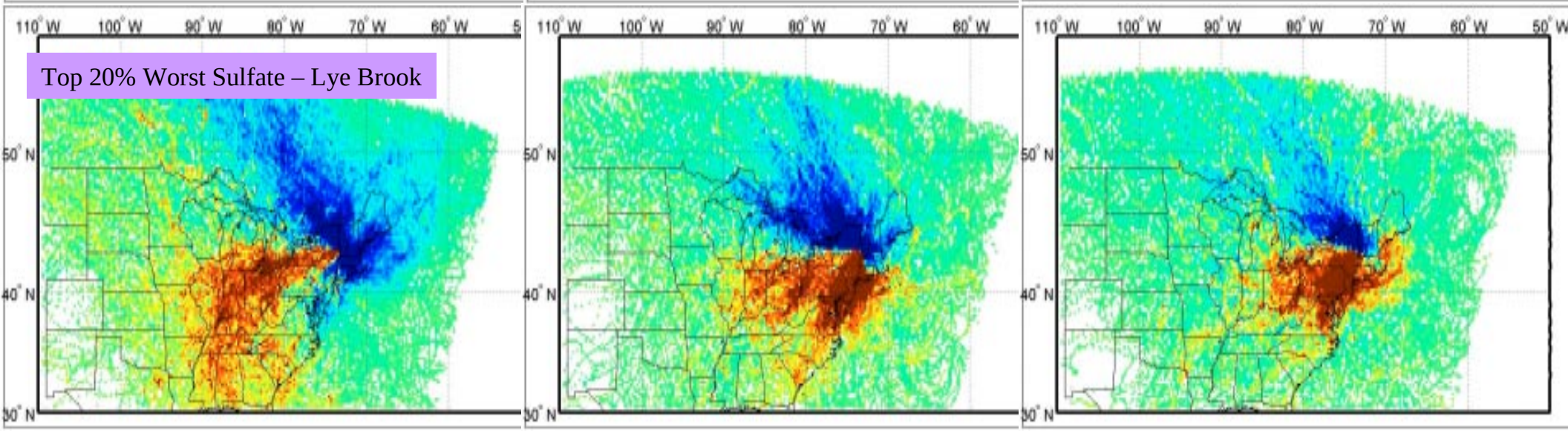
20 Percent covers a lot of seasons…

NESCAUM
Northeast States for Coordinated Air Use Management

Spring | Summer | Fall

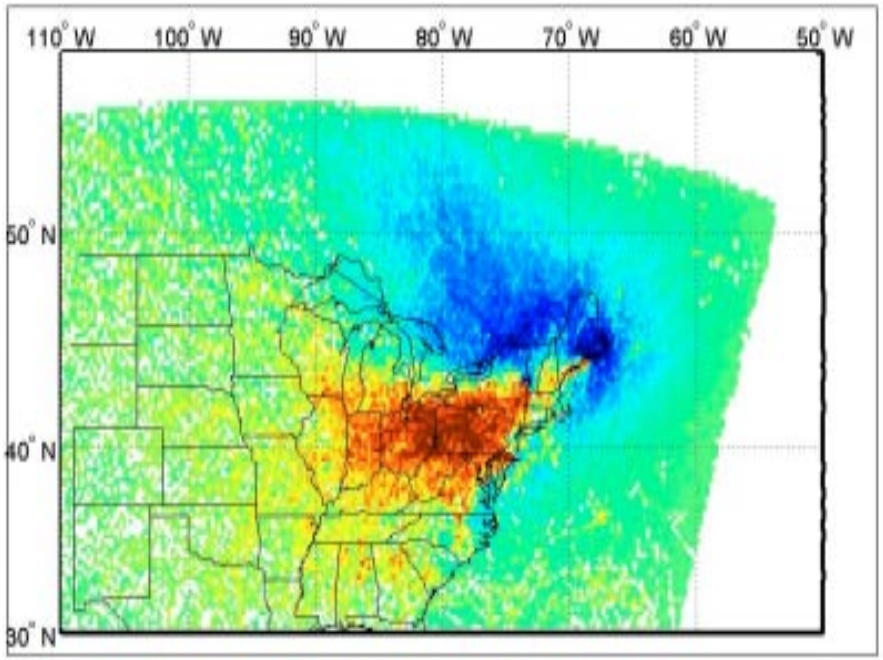


Top 5% Worst Sulfate - Lye Brook

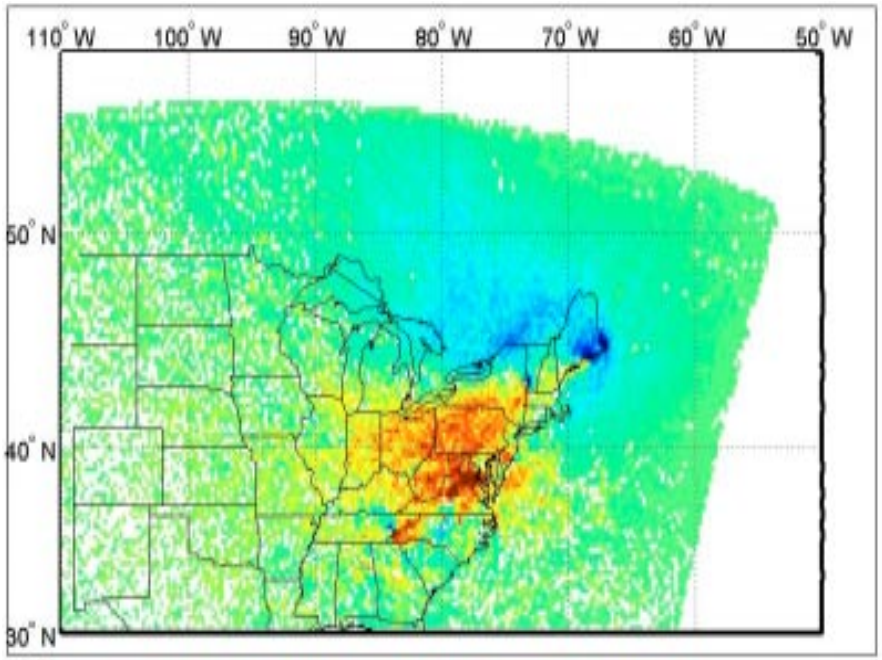


Top 20% Worst Sulfate – Lye Brook

5% worst days from the perspective of multiple sites



Acad, Brig, Lybr, Moos



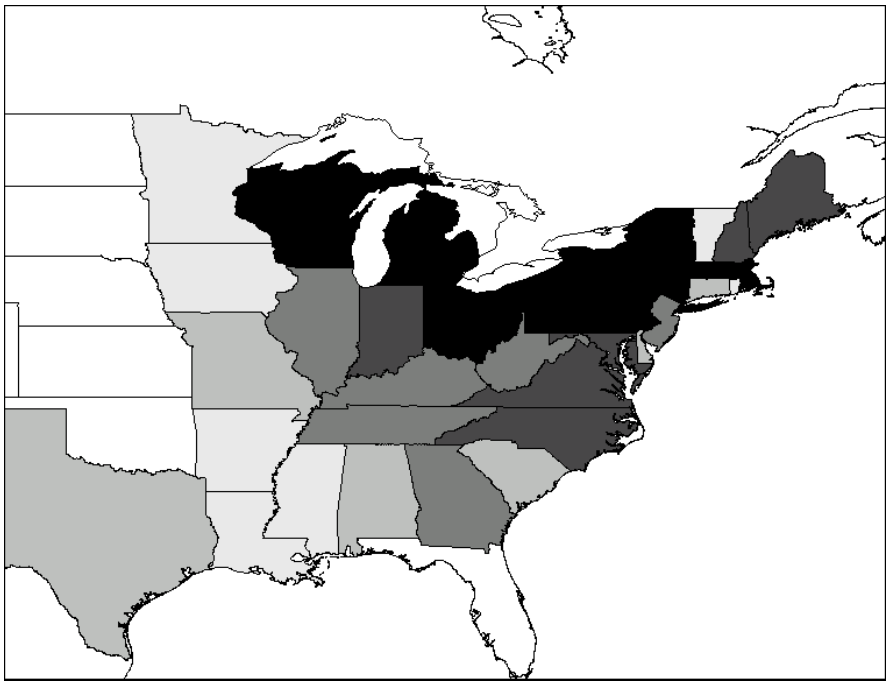
Acad, Brig, Lybr, Grsm, Moos, Shen, Wash, Jeff

REMSAD v.7.10

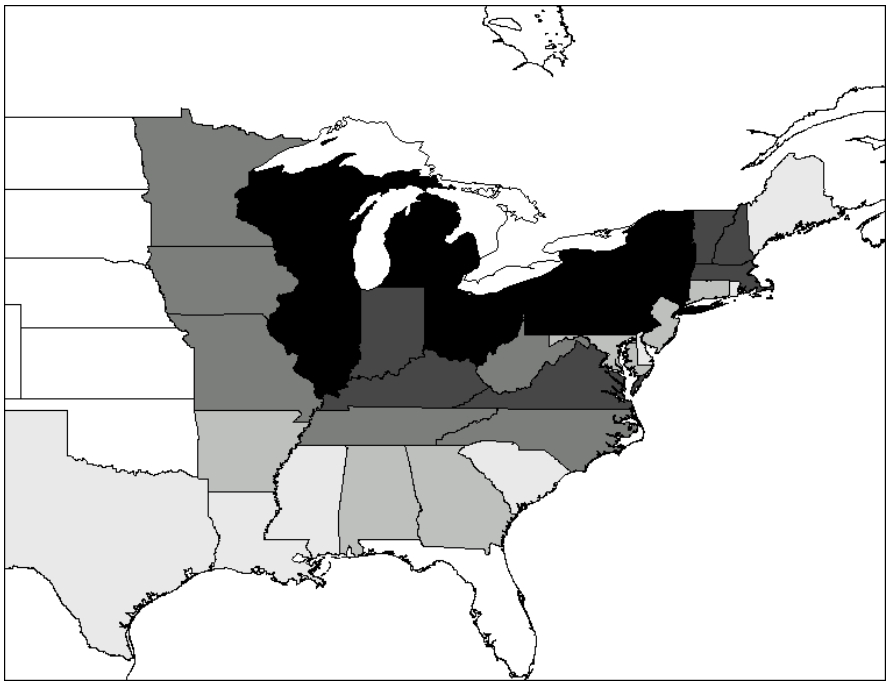
- Allows for source tagging (up to 11 tags per run)
- Meteorology and Emissions from EPA Clear Skies Act (2003) platform
- Tagged SO_2 emissions from elevated point sources in 33 Eastern U.S. states

Compare to REMSAD

REMSAD 7.10



Cluster-Weighted Probability



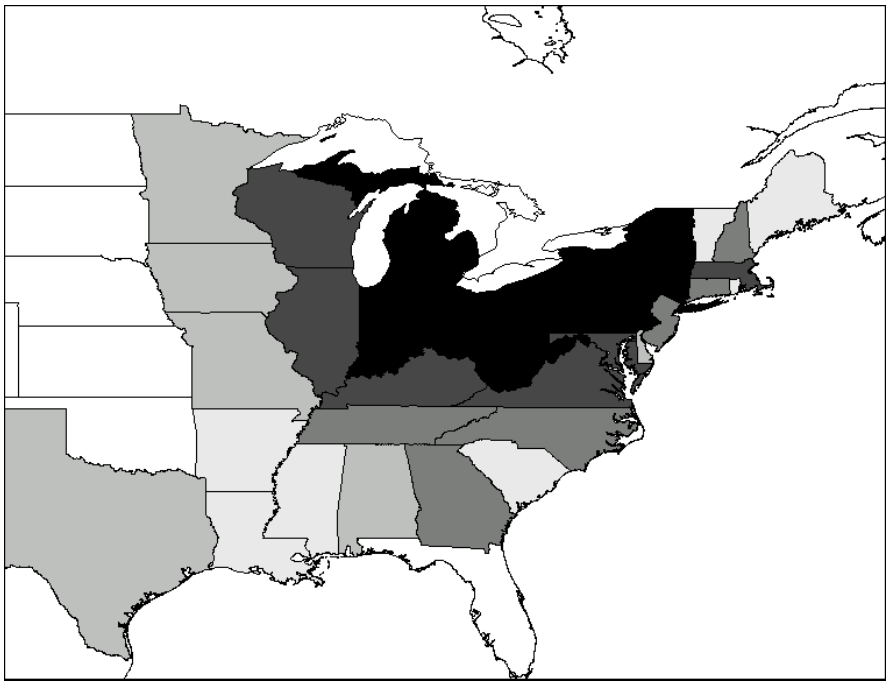
Top quintile contributor
2nd quintile contributor
3rd quintile contributor
4th quintile contributor
Bottom quintile contributor

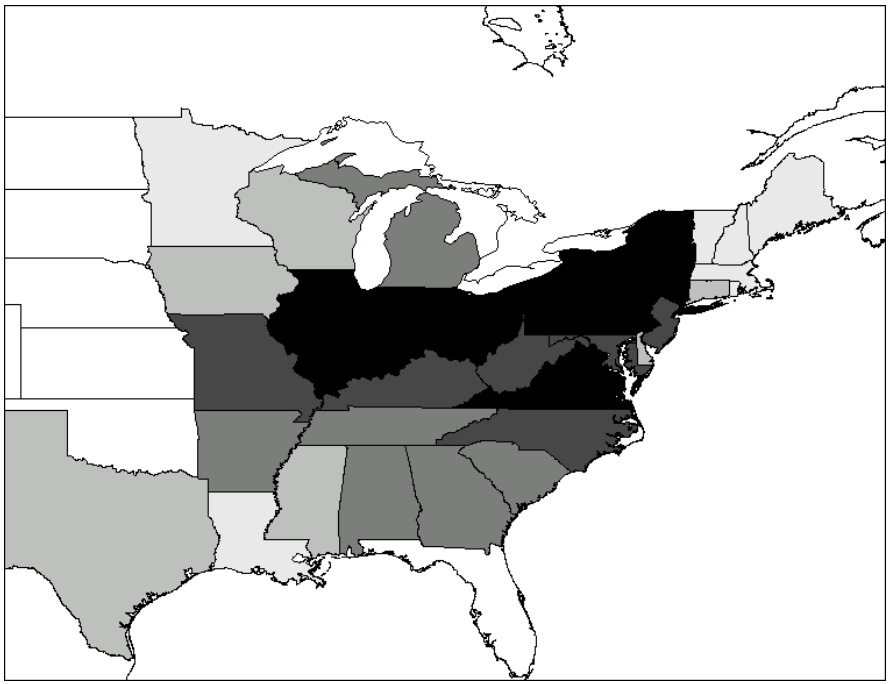
Acadia National Park, ME

Compare to REMSAD

REMSAD 7.10

Cluster-Weighted Probability





Top quintile contributor

2nd quintile contributor

3rd quintile contributor

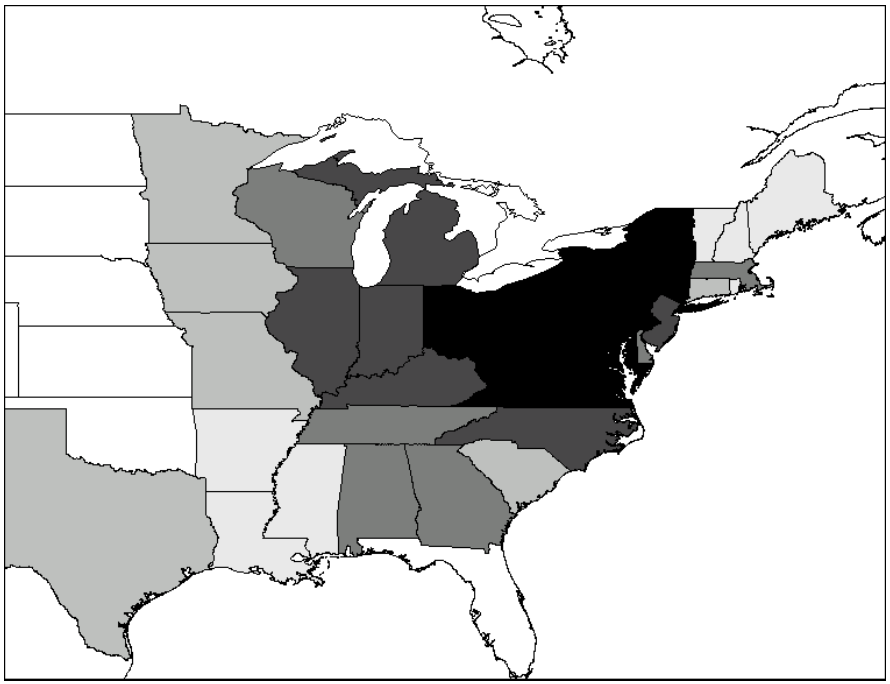
4th quintile contributor

Bottom quintile contributor

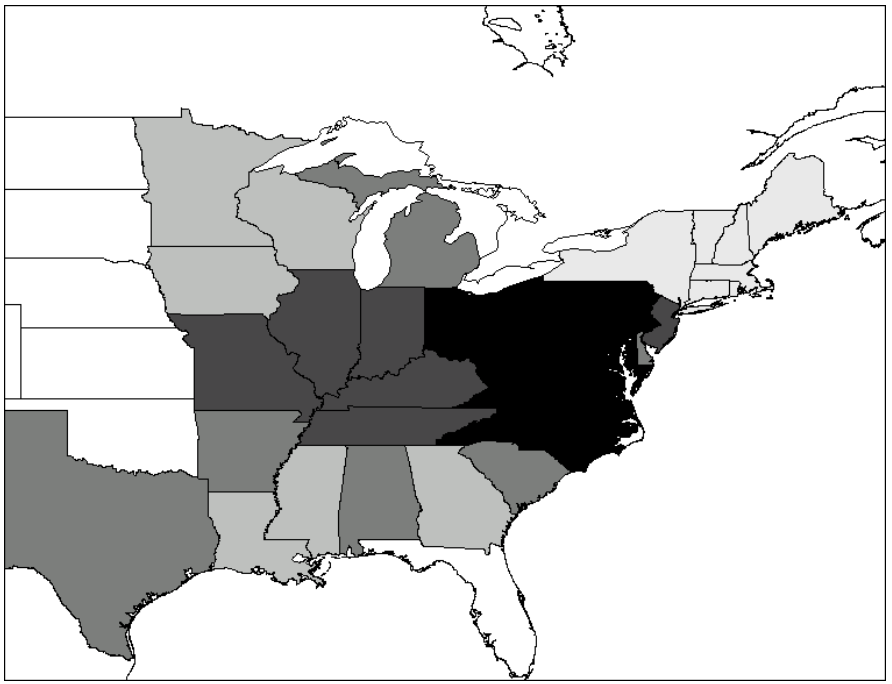
Lye Brook Wilderness, VT

Compare to REMSAD

REMSAD 7.10



Cluster-Weighted Probability



Top quintile contributor
2nd quintile contributor
3rd quintile contributor
4th quintile contributor
Bottom quintile contributor

Brigantine Wilderness, NJ