Appendix A- Multiple Site Averages Supplement

Trajectory analysis results Multiple Sites.

Equations for Different Metrics

Everyday Residence-time Probability

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

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

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

Incremental Probability

$$IP = HP - EP$$

High Day Residence-time Probability

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

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

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 = total number of clusters calculated

$(\overline{C})_i$ = Average pollutant concentration (based on observations associated with cluster i)

$\overline{C}$ = Average pollutant concentration (based on all days)

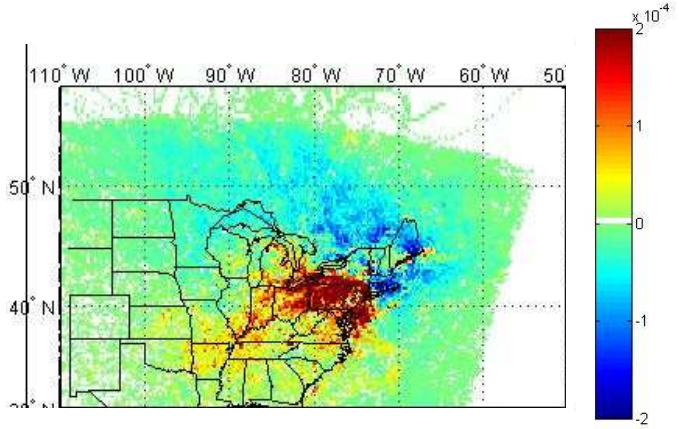
Description of Figures

- Central Trajectory (CT)- Trajectory with the largest number of nearest neighbors in the dataset.
- Frequency Based Clusters- These clusters are formed by finding the “central” trajectory which has the greatest number of neighboring trajectories within a subjectively selected radius of proximity (R). These trajectories are then removed from the dataset and the process is applied to the remaining trajectories.
- Proximity Based Clusters- Clustering relies on the frequency-based cluster groups, but forms trajectory groups based on proximity rather than frequency. In the first step, the frequency-based approach is used to identify the central trajectories that represent the most populated frequency-based clusters (approximately 10 clusters typically contain at least 98% of the trajectories in the dataset using R=12 and 120 hour back-trajectory (BT) time). These 10 central trajectories are then used to develop 10 proximity-based clusters by assigning every trajectory in the dataset to its nearest central trajectories (calculated back to 72 hours).
- Incremental Probability- Difference between the everyday probability (probability derived from all the trajectories in the dataset) and high day probability (probability derived from trajectories arriving at the site on the subset of high pollution days).
- Cluster Weighted Probability- Each PATH-derived cluster’s residence-time probability is weighted by the average sulfate (or other pollutant) value for any measurements corresponding to a trajectory which is a member of that cluster. The weighted residence-time probability is summed over *all* clusters calculated for a site. The everyday probability is subtracted from the sum of cluster-weighted probabilities to identify areas of increased (or in the case of negative values, decreased) probability of being associated with a meteorological pathway for pollutant transport.

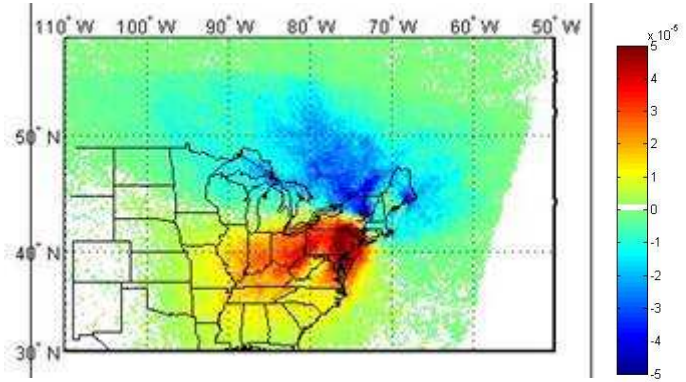
Lyebrook, Mhwk Mt, Acadia 00-04, Avg Inc. Prob. and CWP

IP Based on Top10% Sulfate, CWP calculated using Proximity Based Clusters, 500m

Top 10% IP

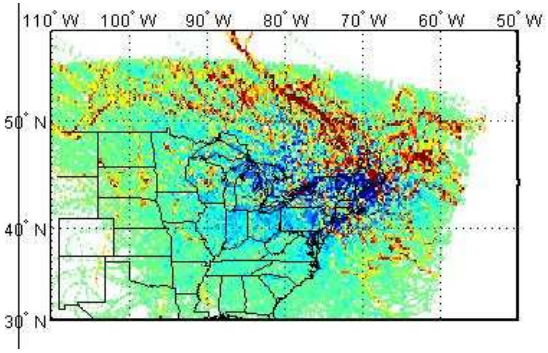


CWP

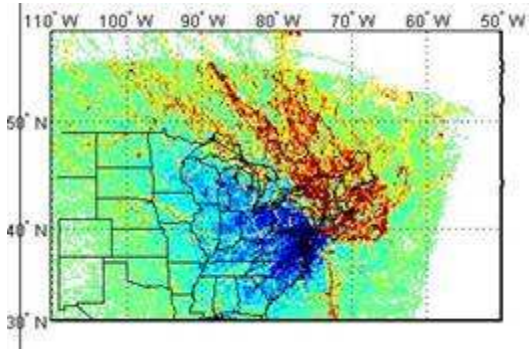


Acad, Brig, Lybr All Trajectories 00-04, Average Incremental Probability

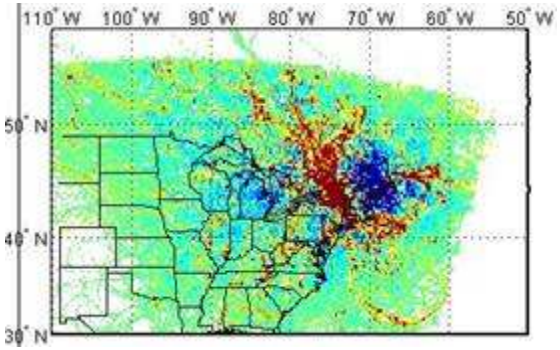
IP Based on Bottom 10% Sulfate, 500m



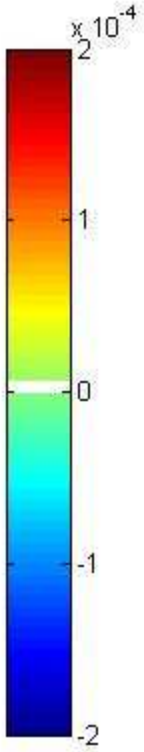
Acadia



Brigantine



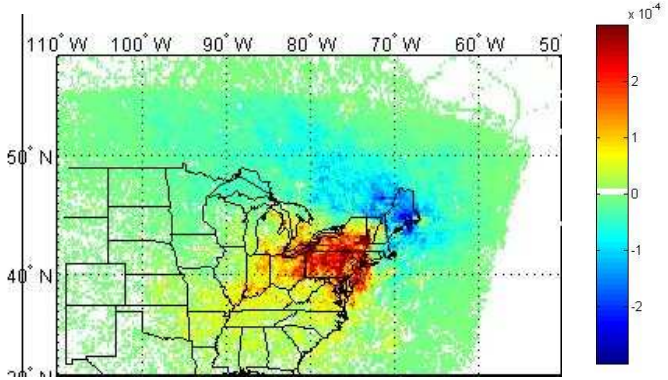
Lye Brook



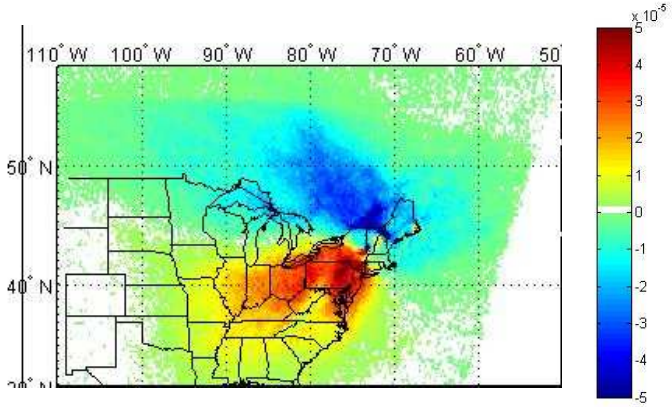
Lyebrook, Mhwk Mt, Acadia, Camp Dodge, Moosehorn 00-04, Avg Inc. Prob. and CWP

IP Based on Top10% Sulfate, CWP calculated using Proximity Based Clusters, 500m

Top 10% IP



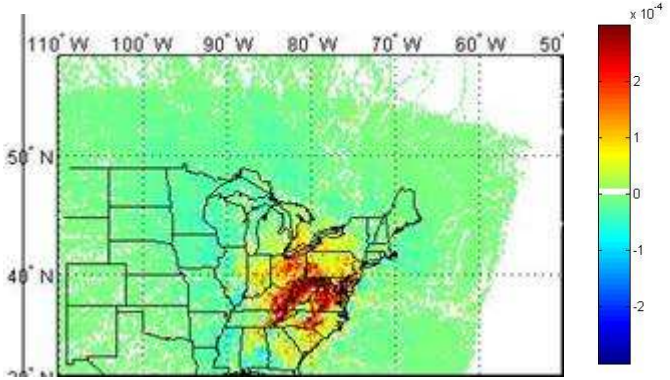
CWP



Brigantine, Shenendoah, Dolly Sods, Great Smokey Mountains 00-04, Avg Inc. Prob. and CWP

IP Based on Top10% Sulfate, CWP calculated using Proximity Based Clusters, 500m

Top 10% IP



CWP