Appendix A- New York City Supplement

Trajectory analysis results at Queens College, New York City.
Equations for Different Metrics

**Everyday Residence-time Probability**

\[
EP = \left( \frac{n_{ij}}{N} \right)
\]

- \( n_{ij} \) = total endpoints passing through grid cell i, j
- \( N \) = total endpoints passing through all grid cells from all trajectories

**High Day Residence-time Probability**

\[
HP = \left( \frac{m_{ij}}{M} \right)
\]

- \( m_{ij} \) = total high day endpoints passing through grid cell i, j
- \( M \) = total high day endpoints passing through all grid cells from high day trajectories

**Incremental Probability**

\[
IP = HP - EP
\]

**Cluster-Weighted Probability**

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

- \( L \) = 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.
NYC All Trajectories 00-04, Top 10 Clusters

Modes defined at: R= 12, 120hr BT, 500m start height, 6488 valid trajectories, 7698 invalid
Reassign Trajectories Based on 72hr BT, 500m start height, 10116 Valid Trajectories

Cluster 1
- Central Trajectory
- Frequency Based Cluster
- Proximity Based Cluster

Cluster 2
- Central Trajectory
- Frequency Based Cluster
- Proximity Based Cluster

Cluster 3
- Central Trajectory
- Frequency Based Cluster
- Proximity Based Cluster

Cluster 4
- Central Trajectory
- Frequency Based Cluster
- Proximity Based Cluster

Cluster 5
- Central Trajectory
- Frequency Based Cluster
- Proximity Based Cluster

<table>
<thead>
<tr>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
<th>Cluster 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sulfate</td>
<td>5.87</td>
<td>3.42</td>
<td>4.41</td>
<td>2.26</td>
</tr>
<tr>
<td>Bext</td>
<td>18.30</td>
<td>13.10</td>
<td>12.98</td>
<td>9.15</td>
</tr>
<tr>
<td>PM</td>
<td>3.02</td>
<td>2.30</td>
<td>1.90</td>
<td>1.93</td>
</tr>
<tr>
<td>OC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># Trajs</td>
<td>4617</td>
<td>1638</td>
<td>851</td>
<td>687</td>
</tr>
<tr>
<td># Trajs w. Poll</td>
<td>1105</td>
<td>339</td>
<td>182</td>
<td>146</td>
</tr>
</tbody>
</table>

# Trajs w. Poll | 284 | 275 | 226 | 230 | 775 |
NYC All Trajectories 00-04, Top 10 Clusters

Modes defined at: R= 12, 120hr BT, 500m start height, 6488 valid trajectories, 7698 invalid

Reassign Trajectories Based on 72hr BT, 500m start height, 10116 Valid Trajectories

Cluster 6
Frequency Based Cluster
Central Trajectory

Cluster 7
Frequency Based Cluster
Central Trajectory

Cluster 8
Frequency Based Cluster
Central Trajectory

Cluster 9
Frequency Based Cluster
Central Trajectory

Cluster 10
Frequency Based Cluster
Central Trajectory

<table>
<thead>
<tr>
<th>Sulfate</th>
<th>Bext</th>
<th>PM</th>
<th>OC</th>
<th># Trajs</th>
<th># Trajs w. Poll</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.53</td>
<td>4.97</td>
<td>15.47</td>
<td>2.89</td>
<td>317</td>
<td>72</td>
</tr>
<tr>
<td>3.95</td>
<td>2.75</td>
<td>18.04</td>
<td>2.56</td>
<td>305</td>
<td>69</td>
</tr>
<tr>
<td>3.46</td>
<td>2.73</td>
<td>12.05</td>
<td>2.33</td>
<td>279</td>
<td>69</td>
</tr>
<tr>
<td>7.20</td>
<td>3.01</td>
<td>18.79</td>
<td>2.33</td>
<td>150</td>
<td>49</td>
</tr>
<tr>
<td>2.67</td>
<td>2.44</td>
<td>9.26</td>
<td>1.53</td>
<td>110</td>
<td>37</td>
</tr>
</tbody>
</table>

Proximity Based Cluster

Frequency Proximity

Sulfate 2.44 2.67 Bext PM 9.15 10.10 OC 1.21 2.21 # Trajs 110 812 # Trajs w. Poll 17 166

Frequency Proximity

Sulfate 4.53 6.05 Bext PM 15.47 19.79 OC 2.89 3.22 # Trajs 317 770 # Trajs w. Poll 72 167

Frequency Proximity

Sulfate 4.97 3.95 Bext PM 18.04 14.50 OC 2.56 2.33 # Trajs 305 2119 # Trajs w. Poll 69 470

Frequency Proximity

Sulfate 2.75 3.46 Bext PM 12.05 12.63 OC 2.73 2.39 # Trajs 279 775 # Trajs w. Poll 49 150

Frequency Proximity

Sulfate 3.01 7.20 Bext PM 9.26 20.90 OC 1.53 3.26 # Trajs 150 658 # Trajs w. Poll 37 123
NYC All Trajectories 00-04, Top 10 Clusters
Modes defined at: R= 12, 120hr BT, 500m start height, 6488 valid trajectories, 7698 invalid
Reassign Trajectories Based on 72hr BT, 500m start height, 10116 Valid Trajectories
Best and Worst Days

Highest Sulfate (Proximity)

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Proximity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sulfate</td>
<td>3.48</td>
<td>6.42</td>
</tr>
<tr>
<td>Bext</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM</td>
<td>10.73</td>
<td>19.36</td>
</tr>
<tr>
<td>OC</td>
<td>1.97</td>
<td>3.60</td>
</tr>
<tr>
<td># Trajs</td>
<td>775</td>
<td>438</td>
</tr>
<tr>
<td># Trajs w. Poll</td>
<td>190</td>
<td>85</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Proximity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sulfate</td>
<td>5.63</td>
<td>2.26</td>
</tr>
<tr>
<td>Bext</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM</td>
<td>18.32</td>
<td>9.15</td>
</tr>
<tr>
<td>OC</td>
<td>3.07</td>
<td>2.42</td>
</tr>
<tr>
<td># Trajs</td>
<td>927</td>
<td>687</td>
</tr>
<tr>
<td># Trajs w. Poll</td>
<td>230</td>
<td>146</td>
</tr>
</tbody>
</table>

Lowest Sulfate (Proximity)

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Proximity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sulfate</td>
<td>2.67</td>
<td>2.44</td>
</tr>
<tr>
<td>Bext</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM</td>
<td>10.10</td>
<td>9.15</td>
</tr>
<tr>
<td>OC</td>
<td>2.21</td>
<td>1.21</td>
</tr>
<tr>
<td># Trajs</td>
<td>812</td>
<td>110</td>
</tr>
<tr>
<td># Trajs w. Poll</td>
<td>166</td>
<td>17</td>
</tr>
</tbody>
</table>

Sulfate- Sulfate ion Conc. (ug/m3)
Bext- Extinction (Mm-1)
PM- Particulate Matter Conc. (ug/m3)
OC- Organic Carbon Conc. (ug/m3)
Num Trajs- Number of trajectories in cluster
Num Trajs w. Poll- Number of trajectories in cluster with associated pollution measurement (Based on number of AQS ASPD samples taken during the 2000-2004 period).
NYC All Trajectories 00-04, Incremental Probability
IP Based on Top10%, 500m

Sulfate

PM
NYC All Trajectories 00-04, Cluster Weighted Probability
CWP calculated using Proximity Based Clusters, 500m

Sulfate

PM
NYC All Trajectories 00-04, Cluster Weighted Probability
Calculated using Frequency Based Clusters, 500m

Sulfate

PM