Environmental Public Health Surveillance for Exposure to Respiratory Health Hazards: A Joint NASA/CDC Project to Use Remote Sensing Data for Estimating Airborne Particulate Matter Over the Atlanta, Georgia Metropolitan Area

Dale A. Quattrochi
Douglas Rickman
NASA
Marshall Space Flight Center
Huntsville, AL

Mohammad, Al-Hamdan, William Crosson,
Maurice Estes, Jr., Ashutosh Limaye
Universities Space Research Association
National Space Science & Technology Center
Huntsville, AL

Judith Qualters
Centers for Disease Control and Prevention
Atlanta, GA
Public Health Surveillance

- Ongoing systematic collection, analysis, and interpretation of outcome-specific data used to plan, implement, and evaluate public health practice.
Surveillance Information Uses

- Monitor & detect changes in the magnitude & distribution of selected events
- Develop hypotheses for research
- Evaluate interventions
- Facilitate public health decision-making
Types of PH Surveillance

- Prevalence
  - All cases
- Incidence
  - Newly diagnosed cases
- Active
  - Health department initiated
- Passive
  - Health care provider initiated
CDC's National Environmental Public Health Tracking (EPHT) Program initiated in 2002

- Congressional funding for development and implementation of a nationwide environmental health tracking network and capacity development in environmental health at State and local health Departments
Selected EPHT Network Features

- Tools for linkage, visualization, analysis, generation of alerts, & reporting
- Internet-based
- Standards-based
- HIPAA compliant
- Access to the network is based on role & purpose
Environmental public health tracking is the ongoing, systematic collection, integration, analysis, and interpretation of data about the following factors:

- environmental hazards
- human exposure to environmental hazards
- health effects potentially related to exposure to environmental hazards

Data must be disseminated to plan, implement, and evaluate environmental public health action.
ENVIRONMENTAL PUBLIC HEALTH TRACKING

DEPARTMENT OF HEALTH AND HUMAN SERVICES
CENTERS FOR DISEASE CONTROL AND PREVENTION
SAFER • HEALTHIER • PEOPLE
### Health Effects
- Asthma
- Poisoning – heavy metal; CO; pesticides
- Cancer
- Birth Defects
- Other adverse reproductive outcome such as low birth wt, preterm birth
- Developmental disabilities
- Other chronic respiratory disease
- Multiple Sclerosis
- Cardiovascular Disease
- Systemic Lupus Erythematosus
- Amyotrophic lateral sclerosis

### Exposures/Hazards
- PCBs
- Heavy metals
- Pesticides
- Environmental tobacco smoke
- Radionuclides
- Asbestos
- Other drinking water contaminants such as trihalomethanes, PCE, TCE,
- Outdoor air contaminants such as particulate matter, ozone, CO and air toxics
- Indoor air contaminants such as mold, carbon monoxide
• Provide information regarding the 5-county Metro-Atlanta Area
  • Clayton, Cobb, DeKalb, Fulton, & Gwinnett
• Integrate environment & public health data into a local network that is part of a national network
• Take action to prevent & control environmentally related health effects
HELIx-Atlanta was developed to support current and future state and local EPHT programs to implement data linking demonstration projects which could be part of the EPHT Network.

HELIx-Atlanta is a pilot linking project in Atlanta for CDC to learn about the challenges the states will encounter.

NASA/MSFC and the CDC are partners in linking environmental and health data to enhance public health surveillance.

The use of NASA technology creates value – added geospatial products from existing environmental data sources to facilitate public health linkages.

Proving the feasibility of the approach is the main objective
- Sharing data between agencies with different missions and mindsets
- Protecting confidentiality of information
- Ensuring high quality geocoded data
- Ensuring appropriate spatial and temporal resolutions of environmental data
- Developing sound resources and methods for conducting data linkages and data analysis
RH Team Pilot Data Linkage Project:
Link environmental data related to ground-level PM$_{2.5}$ (NASA+EPA) with health data related to asthma

Goals:
1. Produce and share information on methods useful for integrating and analyzing data on asthma and PM$_{2.5}$ for environmental public health surveillance.
2. Generate information and recommendations valuable to sustaining surveillance of asthma with PM$_{2.5}$ in the Metro-Atlanta area.

Environmental Hazard Measure: Daily PM$_{2.5}$
Asthma Measure: Daily acute asthma office visits to KP-GA Medical Facilities
Time period: 2001-2003
Linkage Domain: 5-county metropolitan Atlanta
EPA Air Quality System (AQS) ground measurements

- National network of air pollution monitors
- Concentrated in urban areas, fewer monitors in rural areas
- Time intervals range from 1 hr to 6 days (daily meas. every 6th day)
- Three monitor types:
  - Federal Reference Method (FRM)
  - Continuous
  - Speciation
- FRM is EPA-accepted standard method; processing time 4-6 weeks
Sources of PM

MODIS Aerosol Optical Depth (AOD)
- AOD is a measure of the total particulate in the atmosphere
- If atmosphere is well mixed, AOD is a good indicator of surface PM$_{2.5}$
- Enhanced Spatial Coverage
- Provided on a 10x10 km grid
- Available twice per day (Terra ~10:30 AM, Aqua ~1:30 PM)
- Clear-sky coverage only
- Available since spring 2000
Estimating PM$_{2.5}$ from MODIS data

- For 2002-2003, obtain MODIS AOD and EPA AQS PM$_{2.5}$ data
- Extract AOD data for 5 AQS site locations
- Calculate daily averages from hourly AQS PM$_{2.5}$ data
- Using daily PM$_{2.5}$ averages from all 5 Atlanta AQS sites, determine statistical regression equations between PM$_{2.5}$ and MODIS AOD
- Apply regression equations to estimate PM$_{2.5}$ for each 10 km grid cell across region
MODIS AOD

• MODIS data not available every day due to cloud cover
• MODIS AOD follows seasonal patterns of PM$_{2.5}$ but not the day-to-day variability in fall and winter
• Correlations between PM$_{2.5}$ and MODIS AOD are generally high (> 0.55) for the warm season.
• The lower correlation for MODIS-Aqua in 2002 is for July-September only.
1st degree recursive B-spline in x- and y-directions

Inverse Distance Weighted (IDW)

Daily surfaces created on a 10x10 km grid

Variable number of measurements available each day

PM$_{2.5}$ Concentration
- High: 50 µg/m$^3$
- Low: 0 µg/m$^3$

EPA sites
Eliminates anomalous measurements based on a non-parametric rank-order spatial analysis
- Applied to all daily AQS PM$_{2.5}$ measurements before spatial surfaces are built
Assumption: AQS measurements are unbiased relative to the local mean, but MODIS PM$_{2.5}$ estimates may have biases.

Procedure:
1. Use a two-step B-spline algorithm to create highly smoothed versions of the MODIS and AQS PM$_{2.5}$ daily surface
2. Compute the 'Bias' as the difference between the smoothed fields
3. Subtract the bias from the MODIS PM$_{2.5}$ daily surface to give the 'bias-corrected' MODIS daily surface
MODIS and AQS data have been merged to produce final PM\textsubscript{2.5} surfaces.

**B-Spline Surfacing**

Unadjusted MODIS  Bias-adjusted MODIS  Merged

AQS only

65 µg/m\textsuperscript{3}  0 µg/m\textsuperscript{3}
a.k.a. ‘bootstrapping’ or ‘omit-one’ analysis

Objective: Estimate errors associated with daily spatial surfaces

Procedure:
1. Omitting one observation, create surface using N-1 observations
2. Compare value of surface at location of omitted observation with the observed value
3. Repeat for all observations
4. Calculate error statistics by day or site

\[ y = 0.924x + 1.356y \]
\[ R^2 = 0.949 \]
RMSD by Site

Rank Order

Cross-Validated Error Statistics by Site

Bootstrap-Observed

RMSD (micrograms/m³)

Site Number

1 11 21 31 41 51 61 71 81 91 101 111 12

1 2 3 4 5 6 7

NASA
Time Series
RMSD = 2.7 \mu g/m^3
<table>
<thead>
<tr>
<th>Surfacing Technique and Data Source</th>
<th>RMSD (All Days)</th>
<th>RMSD (Warm Season - Days 91-273)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bspline, AQS only, no QC</td>
<td>3.30</td>
<td>3.56</td>
</tr>
<tr>
<td>Bspline, AQS only, with QC</td>
<td>2.93</td>
<td>3.16</td>
</tr>
<tr>
<td>IDW, AQS only</td>
<td>2.45</td>
<td>2.69</td>
</tr>
<tr>
<td>B-Spline, merged AQS/MODIS</td>
<td>N/A</td>
<td>2.76</td>
</tr>
<tr>
<td>IDW, merged AQS/MODIS</td>
<td>N/A</td>
<td>1.61</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Surfacing Technique and Data Source</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bspline: QC vs. No QC</td>
<td>12 %</td>
</tr>
<tr>
<td>Bspline: AQS only vs. merged AQS/MODIS</td>
<td>16 %</td>
</tr>
<tr>
<td>IDW: AQS only vs. merged AQS/MODIS</td>
<td>40 %</td>
</tr>
</tbody>
</table>
### Health Data Set

#### Members

<table>
<thead>
<tr>
<th>LON</th>
<th>LAT</th>
<th>ID</th>
<th>AGE</th>
<th>GENDER</th>
<th>YEAR/MO</th>
</tr>
</thead>
<tbody>
<tr>
<td>-84.207</td>
<td>99.200</td>
<td>1</td>
<td>Child</td>
<td>M</td>
<td>200301</td>
</tr>
<tr>
<td>-84.802</td>
<td>99.359</td>
<td>2</td>
<td>Adult</td>
<td>M</td>
<td>200301</td>
</tr>
<tr>
<td>-83.798</td>
<td>99.993</td>
<td>4</td>
<td>Child</td>
<td>F</td>
<td>200301</td>
</tr>
</tbody>
</table>

#### Acute asthma office visits

<table>
<thead>
<tr>
<th>ID</th>
<th>AGE</th>
<th>LON</th>
<th>LAT</th>
<th>GENDER</th>
<th>DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1811</td>
<td>Child</td>
<td>-84.179</td>
<td>99.118</td>
<td>F</td>
<td>1/1/2003</td>
</tr>
<tr>
<td>54767</td>
<td>Adult</td>
<td>-84.625</td>
<td>99.802</td>
<td>F</td>
<td>1/1/2003</td>
</tr>
<tr>
<td>84580</td>
<td>Adult</td>
<td>-84.679</td>
<td>99.691</td>
<td>F</td>
<td>1/1/2003</td>
</tr>
</tbody>
</table>

*Simulated Data Set. F=female, M=male, A=adult, C=child.*
### Data Linkage Outputs

#### Visit counts by grid cell

<table>
<thead>
<tr>
<th>Date</th>
<th>Cell</th>
<th>PM2.5</th>
<th>FC</th>
<th>MC</th>
<th>FA</th>
<th>MA</th>
</tr>
</thead>
<tbody>
<tr>
<td>200301</td>
<td>1</td>
<td>21.74</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>200301</td>
<td>2</td>
<td>12.79</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>200301</td>
<td>3</td>
<td>12.21</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

#### PM$_{2.5}$ for each visit

<table>
<thead>
<tr>
<th>Date</th>
<th>ID</th>
<th>Lat/Lon</th>
<th>Cell</th>
<th>Cell Lat/Lon</th>
<th>County</th>
<th>State</th>
<th>Gender</th>
<th>Age</th>
<th>PM2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1</td>
<td>1811</td>
<td>99.572 -84.251</td>
<td>1944</td>
<td>99.552 -84.284</td>
<td>Coweta</td>
<td>GA</td>
<td>F</td>
<td>Child</td>
<td>21.74</td>
</tr>
<tr>
<td>1 2</td>
<td>15299</td>
<td>99.063 -83.860</td>
<td>1608</td>
<td>99.104 -83.806</td>
<td>Upson</td>
<td>GA</td>
<td>F</td>
<td>Child</td>
<td>12.79</td>
</tr>
<tr>
<td>1 2</td>
<td>15879</td>
<td>99.727 -84.369</td>
<td>2079</td>
<td>99.731 -84.403</td>
<td>Fulton</td>
<td>GA</td>
<td>M</td>
<td>Child</td>
<td>12.21</td>
</tr>
</tbody>
</table>

*Simulated Data Set. F=female, M=male, A=adult, C=child.*
Cholera Deaths Soho, London August-September, 1854

Legend
- Streets
- Walks
- Grid

Cholera Deaths Per Residence
COUNT
- 1 - 2
- 3 - 4
- 5 - 6
- 7 - 10
- 11 - 18

Integration Radius = 55m
Grid Spacing = 40m


Digital Data of Streets, Walks, and Deaths Per Residence which were used to create this surface were downloaded from the UCLA Department of Epidemiology Website at http://www.ph.ucla.edu/epidem.html.


** Digital Data of Streets, Walks, and Deaths Per Residence which were used to create this surface were downloaded from the UCLA Department of Epidemiology Website at http://www.ph.ucla.edu/epidem.html.
Cholera Deaths Soho, London August-September, 1854

Legend

- Streets
- Wells
- Deaths Per Unit Area

COUNT
0-1
1-5
5-10
10-15
15-20
20-25
25-30
30-35
35-40
40-45
45-50
50-55
55-63

Integration Radius = 55m
Grid Spacing = 40m

Legend

- Streets
- Wells
- Deaths Per Unit Area

COUNT
0-1
1-5
5-10
10-15
15-20
20-25
25-30
30-35
35-40
40-45
45-50
50-55
55-63

Integration Radius = 55m
Grid Spacing = 40m

**Digital Data of Streets, Wells, and Deaths of Residents which were used to create this surface were downloaded from the UCLA Department of Epidemiology Website at http://research.phhs.ucla.edu/epidemiology.html.
The Red Granite kerbstone marks the site of the historic BROAD STREET PUMP associated with Dr. John Snow's discovery in 1854 that Cholera is conveyed by water.
- Proven the feasibility of linking environmental data (MODIS PM$_{2.5}$ estimates and AQS) with health data (asthma)
- Developed algorithms for QC, bias removal, merging MODIS and AQS PM$_{2.5}$ data, and others...
- Negotiated a Business Associate Agreement with a health care provider to enable sharing of Protected Health Information
Member's Name, Affiliation

- (Co-Chair) Kafayat Adeniyi, Centers for Disease Control and Prevention,
- (Co-Chair) Solomon Pollard, Environmental Protection Agency (EPA), Region 4
- Mohammad Z. Al-Hamdan, National Aeronautics and Space Administration
- Rob Blake, DeKalb County Board of Health
- David Blaney, Georgia Division of Public Health
- Bill Crosson, National Aeronautics and Space Administration
- Kristen Mertz, Georgia Division of Public Health
- Amanda Sue Niskar, Centers for Disease Control and Prevention
- Dale Quattrochi, National Aeronautics and Space Administration
- Amber Sinclair, Kaiser Permanente
- Allison Stock, Centers for Disease Control and Prevention
- Denis Tolsma, Kaiser Permanente
- Linda Thomas, Environmental Protection Agency, Region 4
- Ntale Kajumba, Environmental Protection Agency, Region 4
- Carolyn Williams, Georgia Division of Public Health

Acknowledgments

- Leslie Fierro, Centers for Disease Control and Prevention
- Gabriel Rainisch, Centers for Disease Control and Prevention
- Emily Hansen
- HELIX-Atlanta Partners
www.cdc.gov/nceh/tracking