Health and Environment Linked for Information Exchange in Atlanta (HELIX-Atlanta)

Mohammad Al-Hamdan, Bill Crosson, Maury Estes, Sue Estes, Ashutosh Limaye, Dale Quattrochi, Doug Rickman
NSSTC at NASA/MSFC

Amber Sinclair, Dennis Tolsma
Kaiser-Permanente Georgia

Judy Qualters
Centers for Disease Control and Prevention

Kafayat Adeniyi
Northrop Grumman

Amanda Niskar
Fusion Cell, US Department of Health and Human Services

Partners
U.S. Environmental Protection Agency
Georgia Environmental Protection Division
Georgia Division of Public Health
Emory University
Georgia Institute of Technology

ONE step...BEYOND Workshop
European Space Agency/European Space Research Institute
Frascati, Italy, October 15-16, 2015
HELIX-Atlanta Overview

- 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
HELIX-Atlanta Challenges

- 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
Data Linkage

AQS  MODIS

Environ Data  Health Data

Linked Data

Linkage

Acute Asthma Visits

Linked Data

Aggregation

email

HELIX - Atlanta Team

NCEH

EHTB

EPA

NASA

EHTB

Kaiser Permanente

CDC

NCEH

EHTB
Sources of PM$_{2.5}$ data: EPA AQS

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 6$^{th}$ day)
- Three monitor types:
  - Federal Reference Method (FRM)
  - Continuous
  - Speciation
- FRM is EPA-accepted standard method; processing time 4-6 weeks
Sources of PM$_{2.5}$ data: MODIS

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

June 25, 2003
Data Merging

NASA MODIS only

EPA AQS only

Merged

Color scale:
0 μg/m³
65 μg/m³
- 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

PM2.5 Concentration

High: 65 μg/m³
Low: 0 μg/m³
Estimating PM$_{2.5}$ from MODIS data

- For 2000-2003, 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 - PM$_{2.5}$ Relationship

- Daily 5-site means of observed PM$_{2.5}$ and 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
PM 2.5 – MODIS AOD Correlations

- 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.

<table>
<thead>
<tr>
<th></th>
<th>April - September</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MODIS-Terra</td>
</tr>
<tr>
<td>2000</td>
<td>0.579</td>
</tr>
<tr>
<td>2001</td>
<td>0.643</td>
</tr>
<tr>
<td>2002</td>
<td>0.559</td>
</tr>
<tr>
<td>2003</td>
<td>0.661</td>
</tr>
</tbody>
</table>
Quality Control Procedure for AQS PM\textsubscript{2.5} data

- Eliminates anomalous measurements based on a non-parametric rank-order spatial analysis
- Applied to all daily AQS PM\textsubscript{2.5} measurements before spatial surfaces are built
MODIS PM$_{2.5}$ Bias Adjustment

- 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.

**Smooth MODIS**

**Smooth AQS**

**MODIS Bias**

- **Legend**
  - **Bias Value**
    - High: 10.59 µg/m$^3$
    - Low: -22.92 µg/m$^3$

- **Colorbars**
  - 0 µg/m$^3$ to 65 µg/m$^3$
  - 10.6 µg/m$^3$ to -22.9 µg/m$^3$
Merging MODIS and AQS PM$_{2.5}$ Data

MODIS and AQS data have been merged to produce final PM$_{2.5}$ surfaces.

B-Spline Surfacing

Unadjusted MODIS ➔ Bias-adjusted MODIS ➔ Merged

AQS only

Legend:
- 0 μg/m$^3$
- 65 μg/m$^3$
Merging MODIS and AQS PM$_{2.5}$ Data

**IDW Surfacing**

- MODIS Only
- AQS only
- Merged

- 65 μg/m$^3$
- 0 μg/m$^3$
Cross-Validation

- 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

![Graph showing observed versus bootstrap PM2.5 with regression line and R^2 value]

\[ y = 0.924 \times + 1.356 \]

\[ R^2 = 0.949 \]
Cross-Validation for B-Spline Surfaces

Daily Error Statistics

Bootstrap-Observed

Time Series

RMSD = 2.7 μg/m³

Rank Order
Cross-Validation for B-Spline Surfaces
Error Statistics by Site

**Bootstrap-Observed**

- **RMSD by Site**
  - **Rank Order**

![Graph showing error statistics by site with Bootstrap-Observed RMSD values.](image-url)
Cross-Validation for B-Spline Surfaces
Error Statistics by Site

RMSD by Site

Legend
RSMD
BOOT_OBS
0.61 - 1.70
1.71 - 2.79
2.80 - 3.88
3.89 - 4.97
4.98 - 6.09
## Surfacing Methods Comparison

<table>
<thead>
<tr>
<th>Surfacing Technique</th>
<th>Data Source</th>
<th>RMSD (All Days)</th>
<th>RMSD (Warm Season: April-September)</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-Spline</td>
<td>AQS only, no QC</td>
<td>3.30</td>
<td>3.56</td>
<td></td>
</tr>
<tr>
<td>B-Spline</td>
<td>AQS only, with QC</td>
<td>2.93</td>
<td>3.16</td>
<td>12% (than with no QC)</td>
</tr>
<tr>
<td>B-Spline</td>
<td>Merged AQS/MODIS</td>
<td>N/A</td>
<td>2.76</td>
<td>16% (than with AQS data only)</td>
</tr>
<tr>
<td>IDW</td>
<td>AQS only</td>
<td>2.45</td>
<td>2.69</td>
<td>15% (than B-Spline with AQS)</td>
</tr>
<tr>
<td>IDW</td>
<td>Merged AQS/MODIS</td>
<td>N/A</td>
<td>1.61</td>
<td>40% (than with AQS data only)</td>
</tr>
</tbody>
</table>
Annual Composite Surfaces

PM2.5 (ug/m³)

B-Spline

IDW
Linkage of Environmental and Health Data

Data Linkage Outputs

Acute asthma office visit counts by grid cell

<table>
<thead>
<tr>
<th>Date</th>
<th>Cell</th>
<th>PM2.5</th>
<th>Female Child</th>
<th>Male Child</th>
<th>Female Adult</th>
<th>Male Adult</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\textsubscript{2.5} for each visit

<table>
<thead>
<tr>
<th>Date</th>
<th>ID</th>
<th>Member</th>
<th>Lat/Lon</th>
<th>Cell</th>
<th>Cell</th>
<th>Lat/Lon</th>
<th>County</th>
<th>State</th>
<th>Gender</th>
<th>Age</th>
<th>PM2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>200301</td>
<td>1811</td>
<td>99.572</td>
<td>-84.25</td>
<td>1944</td>
<td>99.552</td>
<td>-84.28</td>
<td>Coweta</td>
<td>GA</td>
<td>F</td>
<td>Child</td>
<td>21.74</td>
</tr>
<tr>
<td>200302</td>
<td>15299</td>
<td>99.063</td>
<td>-83.86</td>
<td>1608</td>
<td>99.104</td>
<td>-83.81</td>
<td>Upson</td>
<td>GA</td>
<td>F</td>
<td>Child</td>
<td>12.79</td>
</tr>
<tr>
<td>200302</td>
<td>15879</td>
<td>99.727</td>
<td>-84.37</td>
<td>2079</td>
<td>99.731</td>
<td>-84.4</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.*
Public Health Surveillance

Cholera Deaths Soho, London August-September, 1854

Legend
- Streets
- Wells
- Grid

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

Integration Radius = 55m
Grid Spacing = 40m

12/15/05 10:35am
Dr. Mohammed Al Hamdan
USRA at NASA JPL
HELIK-Atlantis Project
mohammad.alhamdan@nasa.gov


**Digital data of streets, wells, and deaths per residence which were used to create this surface were downloaded from the UCLA Department of Epidemiology Web site at http://www.ph.ucla.edu/episr/home.html.*
Public Health Surveillance

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

** Digital Data of Streets, Wells, and Deaths Residents which were used to create this surface were downloaded from the UCLADepartment of Epidemiology Website at http://www.phdata.ucl.ac.uk/cholera.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.
Successes

- 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 to incorporate satellite remote sensing into the CDC Environmental Public Health Tracking Network

- Developed algorithms for health data surfacing that protects PHI which can be helpful for public health surveillance and decision makers

- Negotiated a Business Associate Agreement with a health care provider to enable sharing of Protected Health Information
Thanks!

Presenter’s Contact Information:
Mohammad Z. Al-Hamdan, PhD

mohammad.alhamdan@nasa.gov