Sensor Networks: Data Processing for Improved Spatial and Temporal Resolution

Prof. David J. Lary, Shea Caspersen, Phil Dickerson, Ethan McMahon, Robin Thottungal, David Smith, John White

Air Sensors International Conference
September 12, 2018

The views, opinions, and observations expressed in this article are of the authors and are not necessarily that of the EPA.
Agenda

- Aspects of analyzing air quality data for AQ analytics and QA of sensor networks
- Cascade of accuracies
- Representativeness
- Interactive Visualization
- Questions
The need to analyze data from air sensors

● As more sensors are deployed there’s an increased need to manage and understand the data
● We’ll present a few methods for managing the data
● These are under development
● We’d like to hear how they can be improved, combined and coordinated
The Key Ideas

- For characterization of neighborhood scale air quality extensive daily street level surveys have shown that we need a \textit{spatial resolution} of at least 0.5 km.
- \textbf{Temporal resolution} is just as critical as spatial resolution, and \textit{low latency} is helpful.
- Given the costs involved using \textit{low cost} sensors is helpful.
- With low cost sensors \textit{calibration} is a particularly critical issue.
- We also want to have \textit{size resolved} observations into the pollen and mold size range.
Aerobiological Observations

Sizes of various particles
Fine Temporal Resolution is Advantageous

- We use techniques developed over a decade for satellite validation to provide pre-deployment and real-time calibration that utilizes:
  - Machine Learning.
  - The Probability Distribution Functions (PDFs) of all observations made over various temporal & spatial scales.
- Measuring the full size distribution up to 40 microns is helpful to also identify airborne mold and pollen. 

![A Full Diurnal Cycle at 10s Resolution](image)
Cascade of accuracies

We can use different levels of accuracy. For particulates:

1. EPA certified instrument: $25,000-$50,000 (primary)
2. Medium accuracy: $2,000-$5,000 (secondary)
3. Inexpensive but useful: $200-$500 (tertiary)
Pre-deployment Calibration

A batch of ten sensors are placed in a calibration chamber for several days together with an EPA certified reference instrument.

The full aerosol size distribution is collected by the reference instrument and by the lower cost sensors, along with the temperature, pressure and humidity.

This is then used together with machine learning to provide a calibration for PM$_{1}$, PM$_{2.5}$, PM$_{10}$, Alveolic, Inhalable and Thoracic estimates.
Example Machine Learning Calibration

Calibration is greatly improved when it is multivariate, nonlinear and parametric.

Uncalibrated Data Set

Training Data Set

Independent Validation

Note the inclusion of error estimates.
Representativeness

- When performing chemical data assimilation the observational, representativeness, and theoretical uncertainties have very different characteristics.
- We routinely accurately characterize the representativeness uncertainty by studying the probability distribution function (PDF) of observations. The average deviation has been used as a measure of the width of the PDF and of the variability (representativeness uncertainty).
- The representativeness uncertainty can be markedly different from the observational uncertainty and clearly delineates mixing barriers.

\[ \sigma_{\text{rep}} = A\text{Dev}(\chi_1, \ldots, \chi_N) = \frac{1}{N} \sum_{j=1}^{N} |\chi_j - \bar{\chi}| \]

What spatial scale to use?
Required Spatial Scale Characterization With Variograms

Real-time Comparison of Neighboring Sensor PDFs
Visualization of Local Level Analytics (demonstration)

- Comparing levels of pollution in neighborhoods adjacent to freeways divided by natural & artificial buffers (e.g., vegetation barriers, soundwalls, etc.)
- Smaller spatial scale analytics
- View high volume of data in one visualization
- Interactive ingest as end users slice and dice through various aspects of the data (e.g., over time, across space, speed, statistical distribution, conditional distribution)
## Advantages of the data analysis methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Advantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representativeness</td>
<td>Quantify the variability at a given location and time</td>
</tr>
<tr>
<td>Cascade of accuracies</td>
<td>Address spatial and temporal variability using low cost sensors and tie them to reference monitors</td>
</tr>
<tr>
<td>Visualization of data in fine scale</td>
<td>See high volume of data in one graphic</td>
</tr>
</tbody>
</table>
Next Steps

● Improve the protocol for calibrating inexpensive sensors to include multiple variables, nonlinearity and parametrics.
● Characterize the temporal scale using similar methods.
● Use an open portal to store and display data from over 55 countries and over 8,000 sites.
  ○ Make this an open platform and protocol.
● Help communities make the best use of low cost sensor data.
Contacts

Prof. David J. Lary, University of Texas at Dallas
Shea Caspersen, U.S. EPA
Phil Dickerson, U.S. EPA
Ethan McMahon, U.S. EPA
Robin Thottungal, U.S. EPA
David Smith, U.S. EPA
John White, U.S. EPA
Extra Slides
Complimented by Using Aerial Vehicle Measurements

Day within EPA Air Quality Standards

Day with exceedance of EPA Air Quality Standards

Flight on Nov 18, 2014 clear skies

Flight on Dec 04, 2014 hazy/overcast
Representativeness

- When performing chemical data assimilation the observational, representativeness, and theoretical uncertainties have very different characteristics.
- We routinely accurately characterize the representativeness uncertainty by studying the probability distribution function (PDF) of observations. The average deviation has been used as a measure of the width of the PDF and of the variability (representativeness uncertainty).
- The representativeness uncertainty can be markedly different from the observational uncertainty and clearly delineates mixing barriers.

\[ \sigma_{\text{rep}} = ADev(\chi_1, \ldots, \chi_N) = \frac{1}{N} \sum_{j=1}^{N} |\chi_j - \bar{\chi}| \]

Fine Particulate Matter Size Comparison

Human hair (about 70μm wide)
Grain of sand (about 50μm wide)
PM$_{10}$ (less than 10μm wide)
PM$_{2.5}$ (less than 2.5μm wide)

μm = micrometer

HUMAN HAIR
50-70μm (microns) in diameter

PM$_{2.5}$
Combustion particles, organic compounds, metals, etc.
< 2.5μm (microns) in diameter

PM$_{10}$
Dust, pollen, mold, etc.
< 10μm (microns) in diameter

90μm (microns) in diameter
FINE BEACH SAND

Image courtesy of the U.S. EPA