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

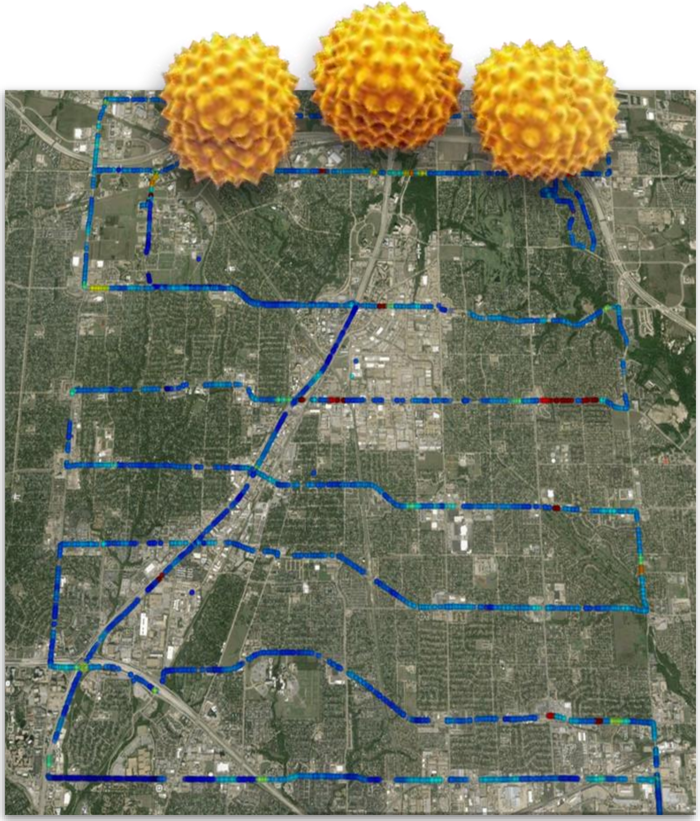
- Aspects of analyzing air quality data for AQ analytics and QA of sensor networks
- Cascade of accuracies
- Representativeness
- Interactive Visualization
- Questions

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

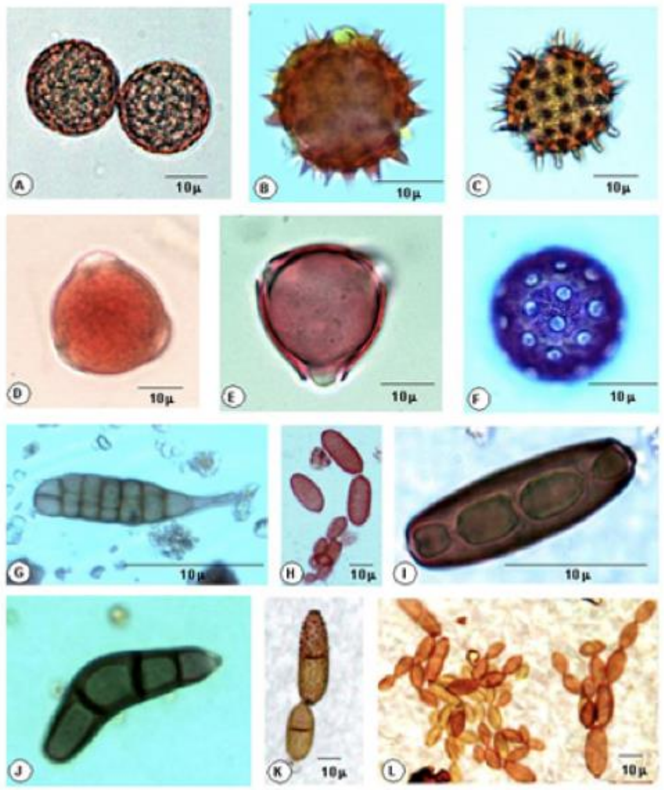
The Key Ideas

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



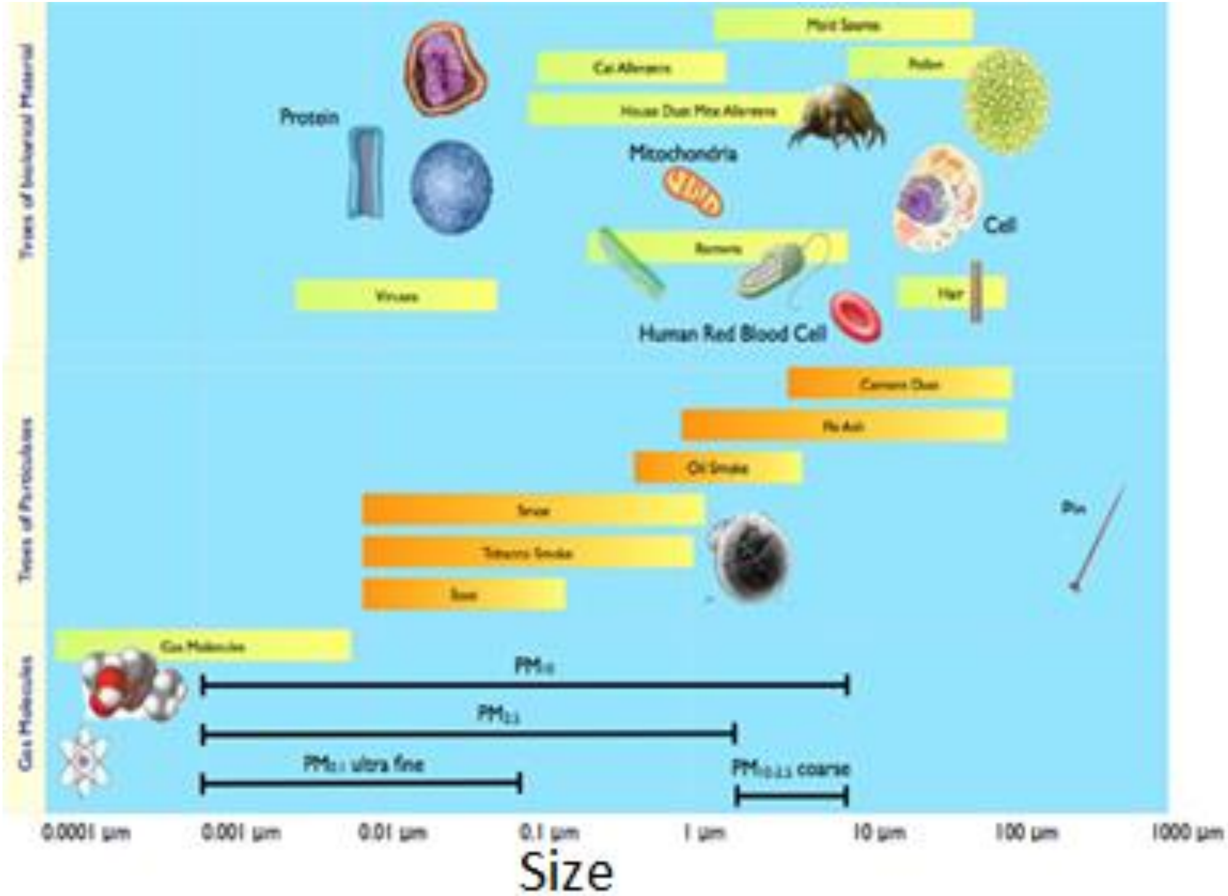
Routes of mobile sensors near the University of Texas at Dallas

Aerobiological Observations



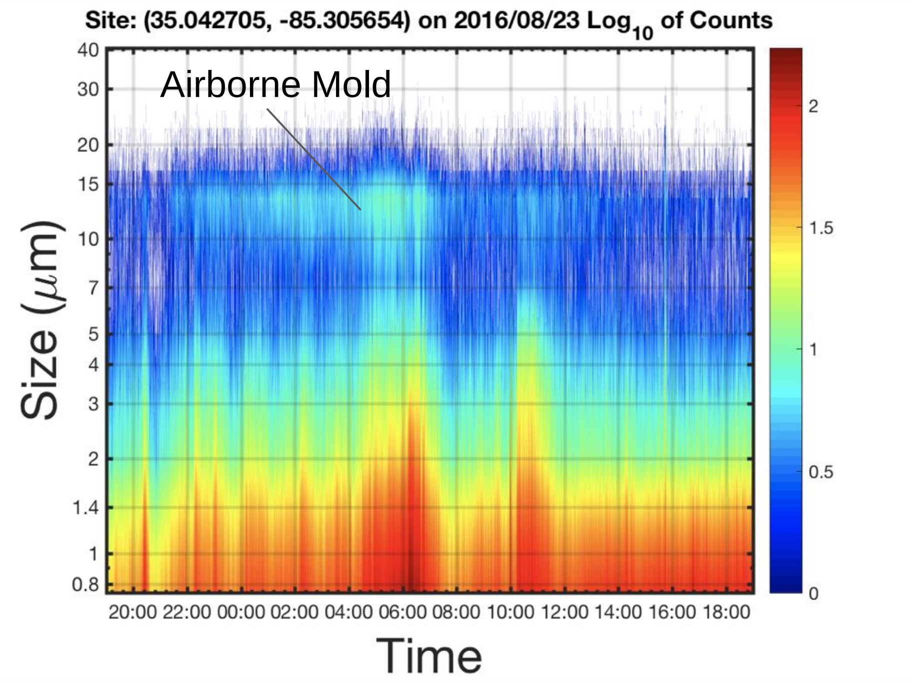
Figures 1 A-L showing most frequent Aeroallergens of Texas Panhandle: Pollens:
A. *Ambrosia artemisiifolia* (Short Ragweed), B. *Helianthus annuus* (Common Sunflower),
C. *Helianthus hirsutus* (Hairy Sunflower), D. *Solanum rostratum* (Buffalo Bur),
E. *Solanum elaeagnifolium* (Purple Nightshade), F. *Chenopodium album* (Lamb's Quarters).
Fungal spores: G. *Alternaria* sp. H. *Stachybotrys* sp. I. *Drechslera* sp. J. *Curvularia* sp.,
K. Ascospores, L. *Cladosporium* sp.

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

Cascade of accuracies

We can use different levels of accuracy. For particulates:

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

1

2

3

Pre-deployment Calibration

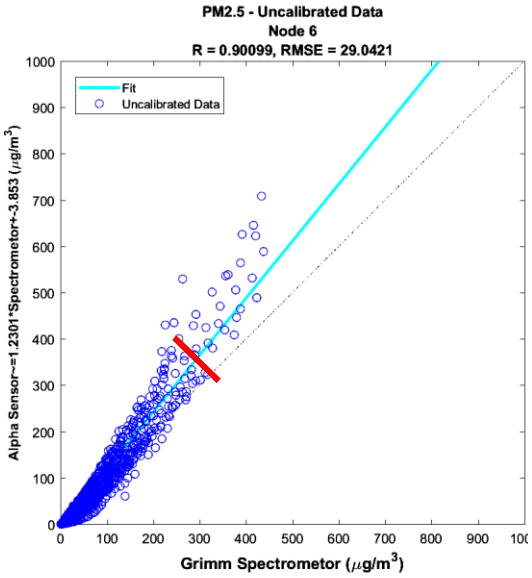
A batch of ten sensors are placed in a calibration chamber for several days together with an EPA certified reference instrument.

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.

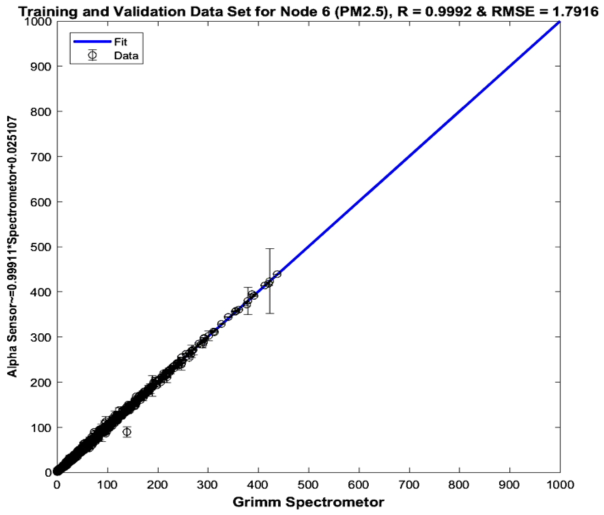
Example Machine Learning Calibration

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

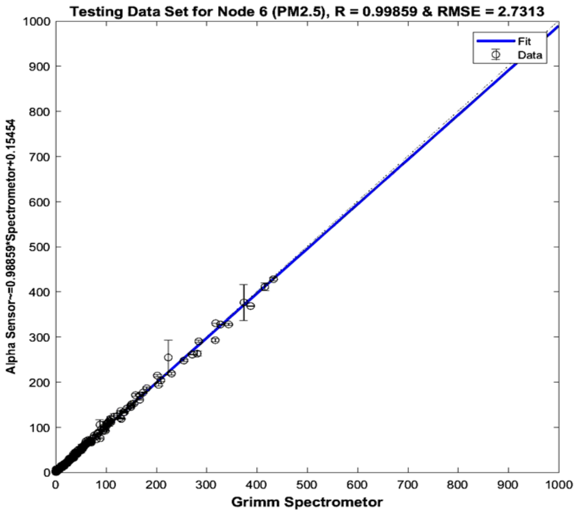


Uncalibrated Data Set

Regression Analysis with the Predicted Errors (PM2.5) for Node 6 (Updated)



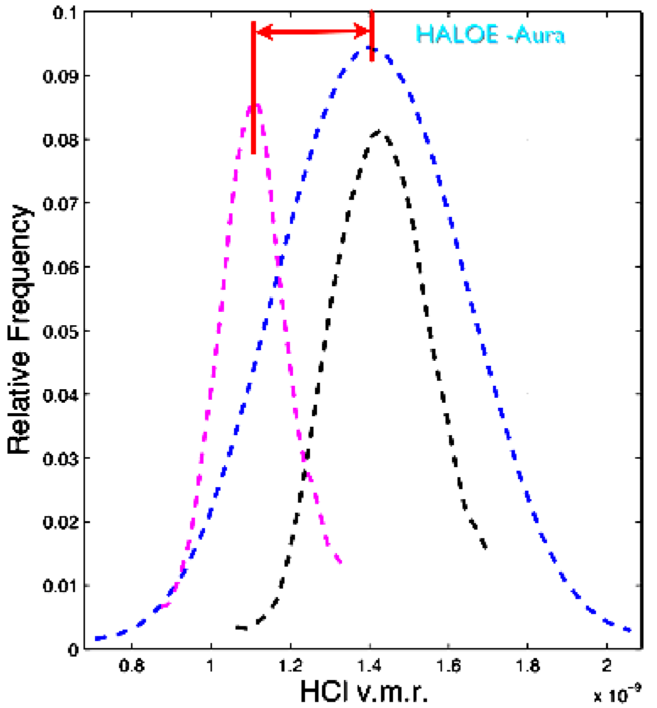
Training Data Set



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.

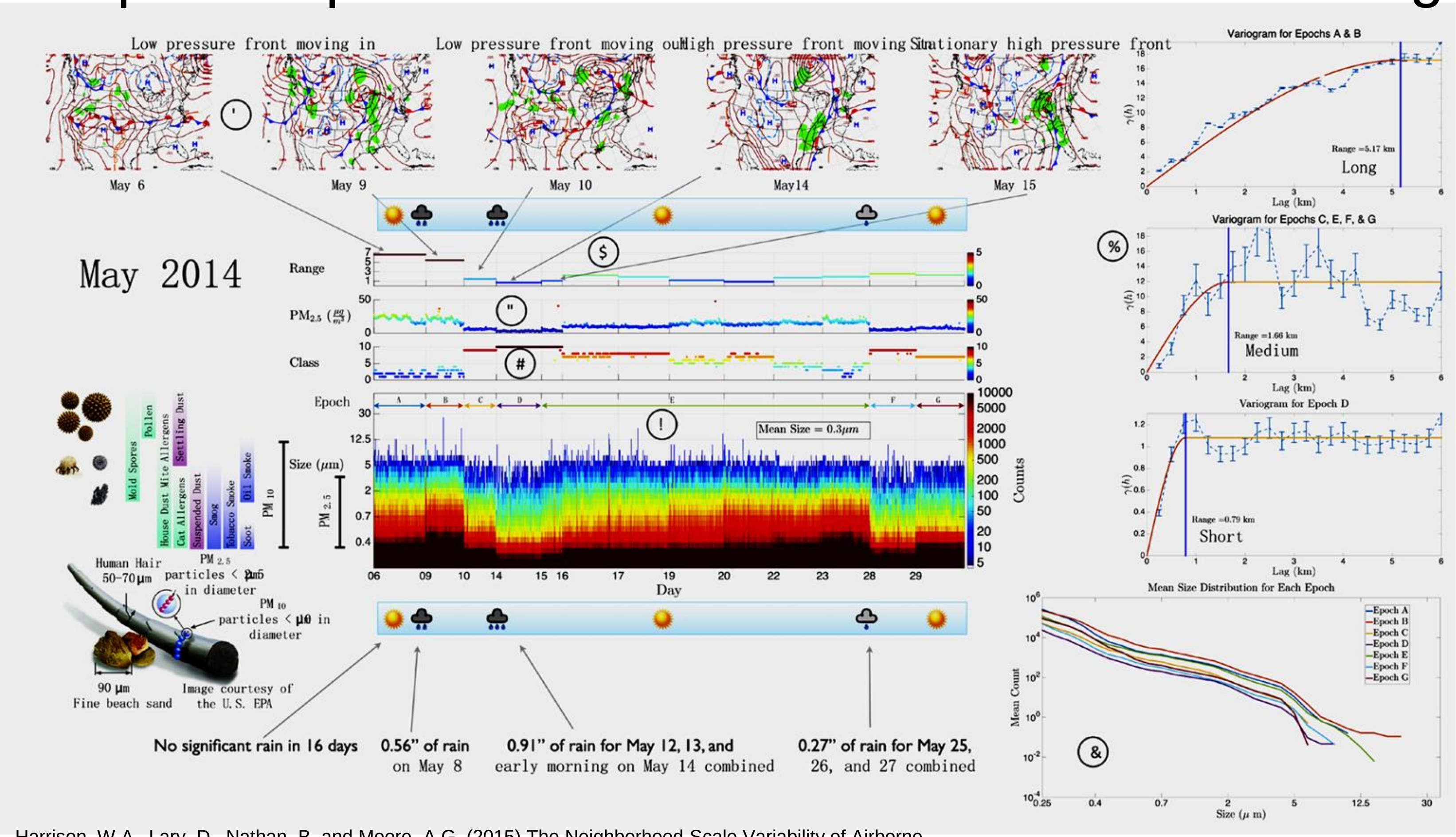


$$\sigma_{\text{rep}} = \text{ADev}(\chi_1, \ldots, \chi_N) = \frac{1}{N} \sum_{j=1}^{N} |\chi_j - \bar{\chi}|$$

Source: doi:10.1016/j.atmoscilet.2003.11.002

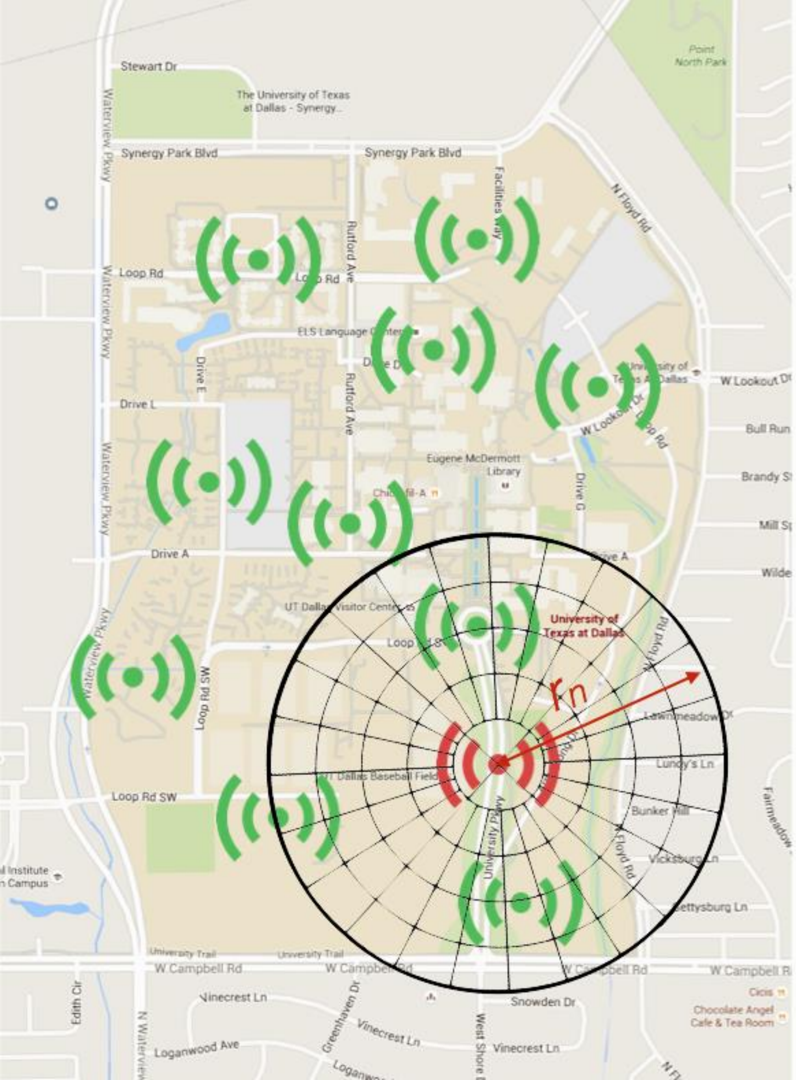
What spatial scale to use?

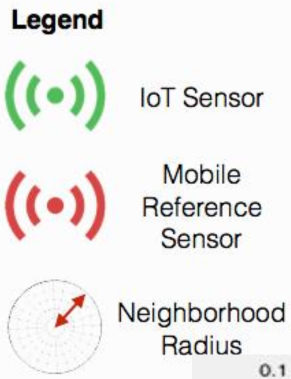
Required Spatial Scale Characterization With Variograms

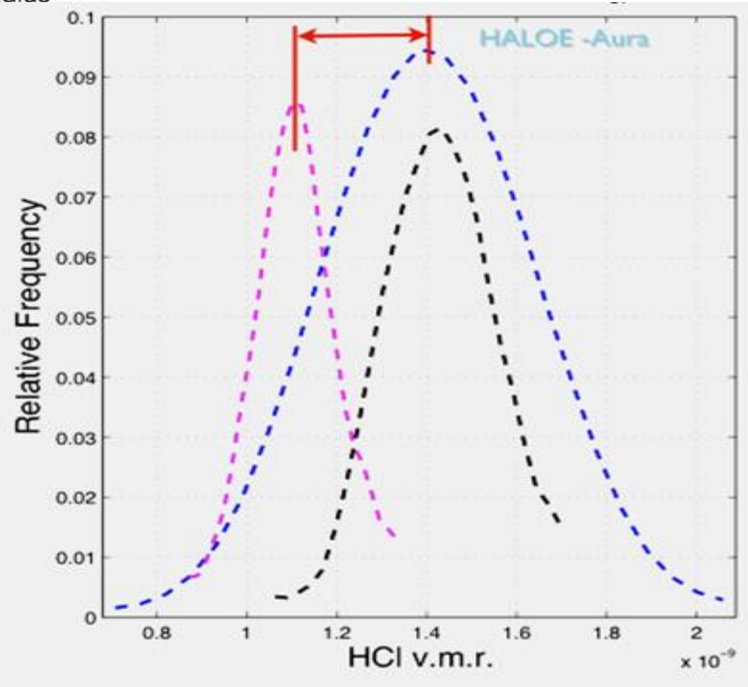


Harrison, W.A., Lary, D., Nathan, B. and Moore, A.G. (2015) The Neighborhood Scale Variability of Airborne Particulates. *Journal of Environmental Protection*, **6**, 464-476. http://dx.doi.org/10.4236/jep.2015.65045

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

Method	Advantages
Representativeness	Quantify the variability at a given location and time
Cascade of accuracies	Address spatial and temporal variability using low cost sensors and tie them to reference monitors
Visualization of data in fine scale	See high volume of data in one graphic

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

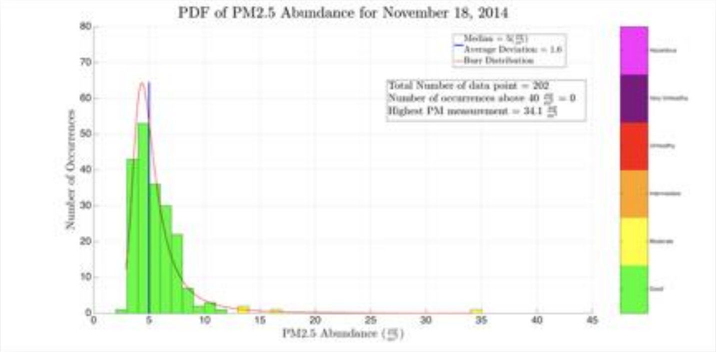
Robin Thottungal, U.S. EPA

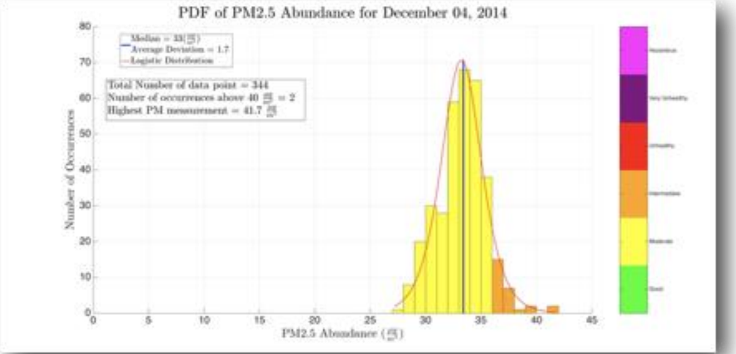
David Smith, U.S. EPA

John White, U.S. EPA

Extra Slides

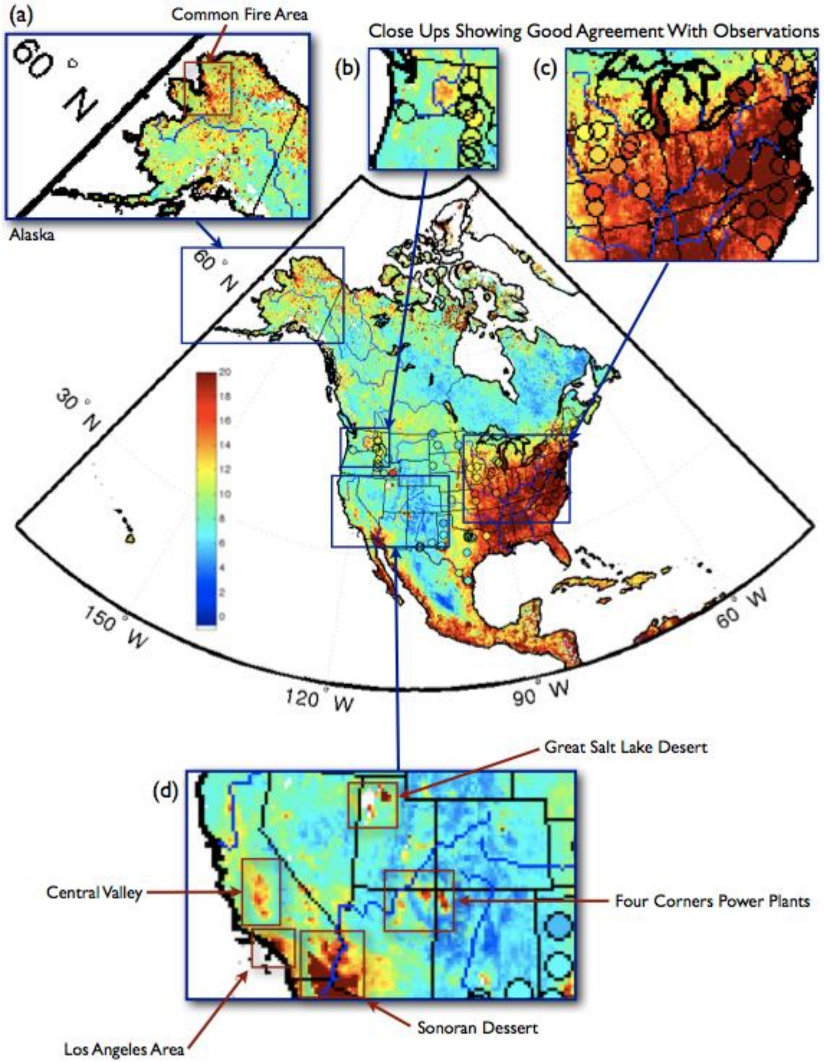
Complimented by Using Aerial Vehicle Measurements





Flight on Nov 18, 2014 clear skies

Flight on Dec 04, 2014 hazy/overcast


(a)
Common Fire Area
Close Ups Showing Good Agreement With Observations
(b)
(c)
60 N
Alaska
60 N
30 N
20
18
16
14
12
10
8
6
4
2
0
150 W
120 W
90 W
60 W
(d)
Great Salt Lake Desert
Central Valley
Four Corners Power Plants
Los Angeles Area
Sonoran Dessert

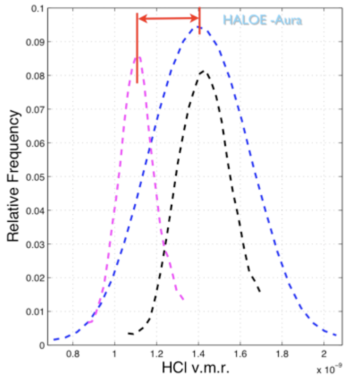

PD
Polarized
Filter
Ceder
Pollen
LD
PD

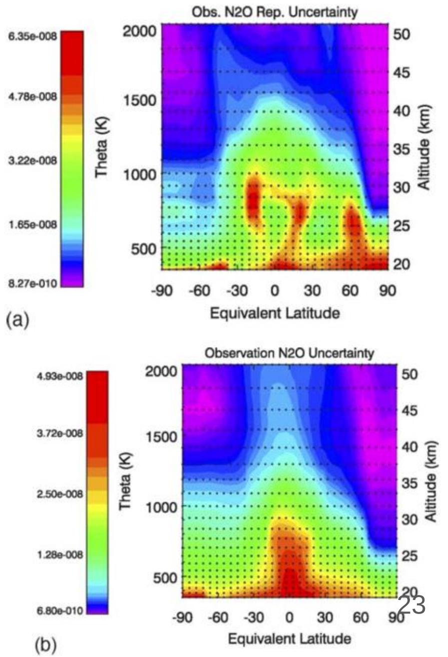
Representativeness

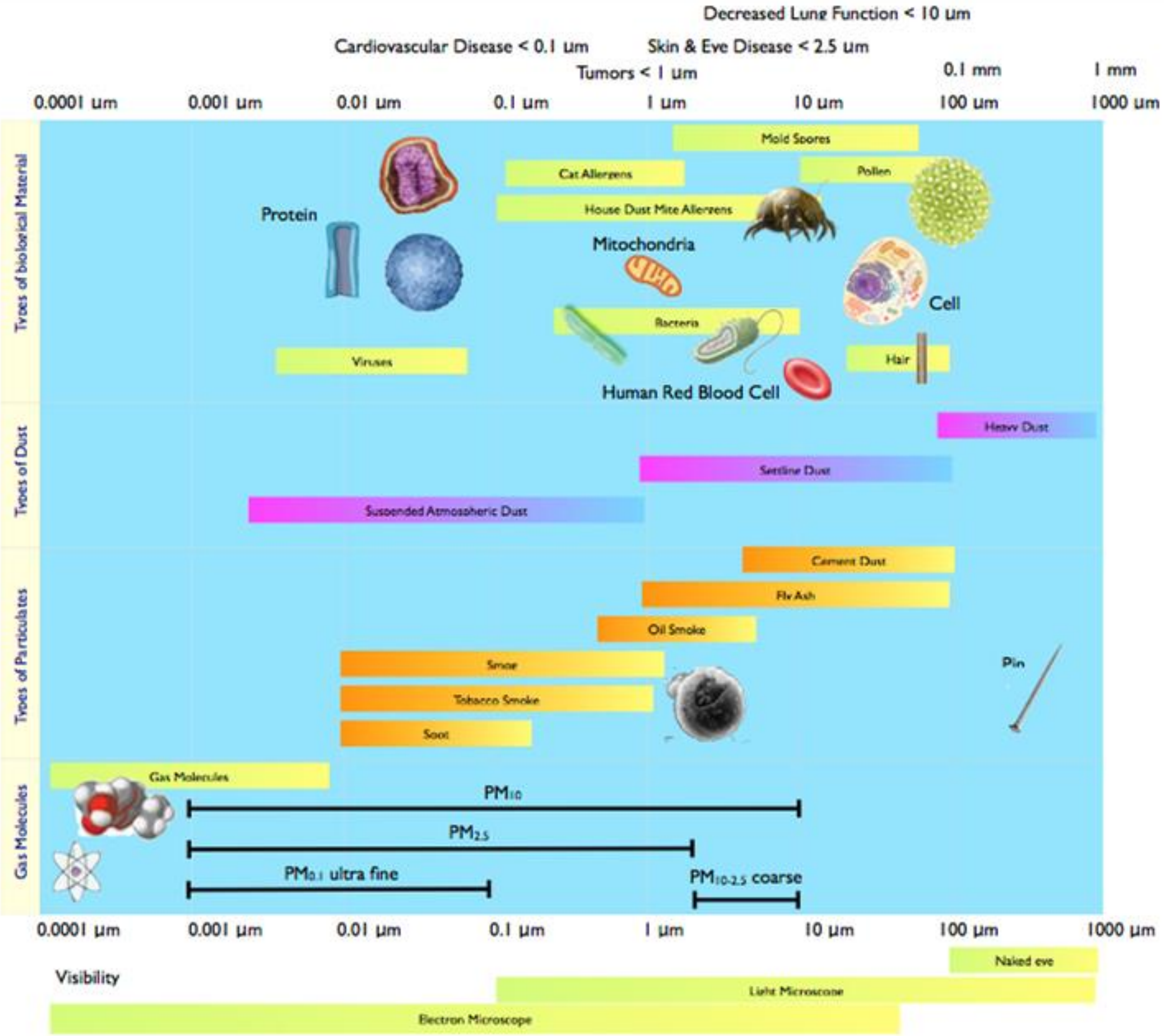
- When performing chemical data assimilation the observational, representativeness, and theoretical uncertainties have very different characteristics.
- We routinely accurately characterize the representativeness uncertainty by studying the probability distribution function (PDF) of observations. The 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

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

Source: doi:10.1016/j.atmoscilet.2003.11.002

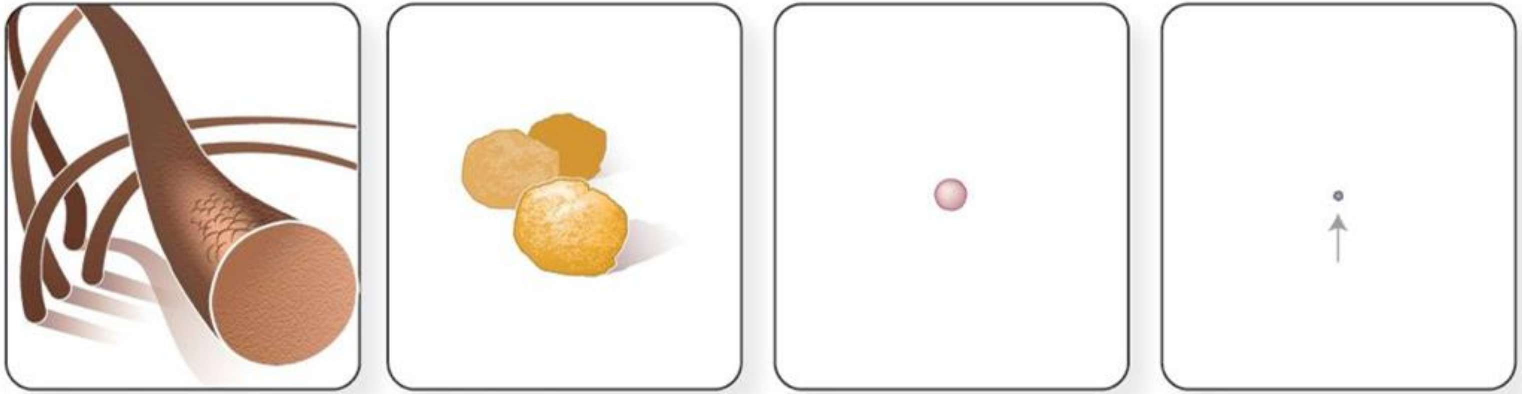






Decreased Lung Function < 10 um
Cardiovascular Disease < 0.1 um
Skin & Eye Disease < 2.5 um
Tumors < 1 um
0.1 mm
1 mm
0.0001 um
0.001 um
0.01 um
0.1 um
1 um
10 um
100 um
1000 um
Types of biological Material
Mold Spores
Cat Allergens
Pollen
House Dust Mite Allergens
Protein
Mitochondria
Cell
Bacteria
Viruses
Hair
Human Red Blood Cell
Types of Dust
Heavy Dust
Settling Dust
Suspended Atmospheric Dust
Types of Particulates
Cement Dust
Fly Ash
Oil Smoke
Smog
Pin
Tobacco Smoke
Soot
Gas Molecules
Gas Molecules
PM10
PM2.5
PM0.1 ultra fine
PM10-2.5 coarse
0.0001 µm
0.001 µm
0.01 µm
0.1 µm
1 µm
10 µm
100 µm
1000 µm
Naked eye
Visibility
Light Microscope
Electron Microscope

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

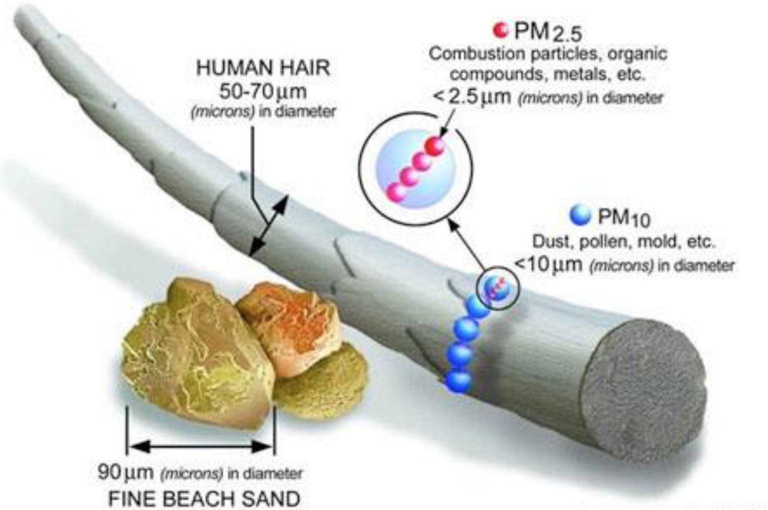


Image courtesy of the U.S. EPA