Assessing the Statistical Relationship in Carbon Measurements Between Old and New Sampling and Analysis Protocols in the Chemical Speciation Network (CSN)

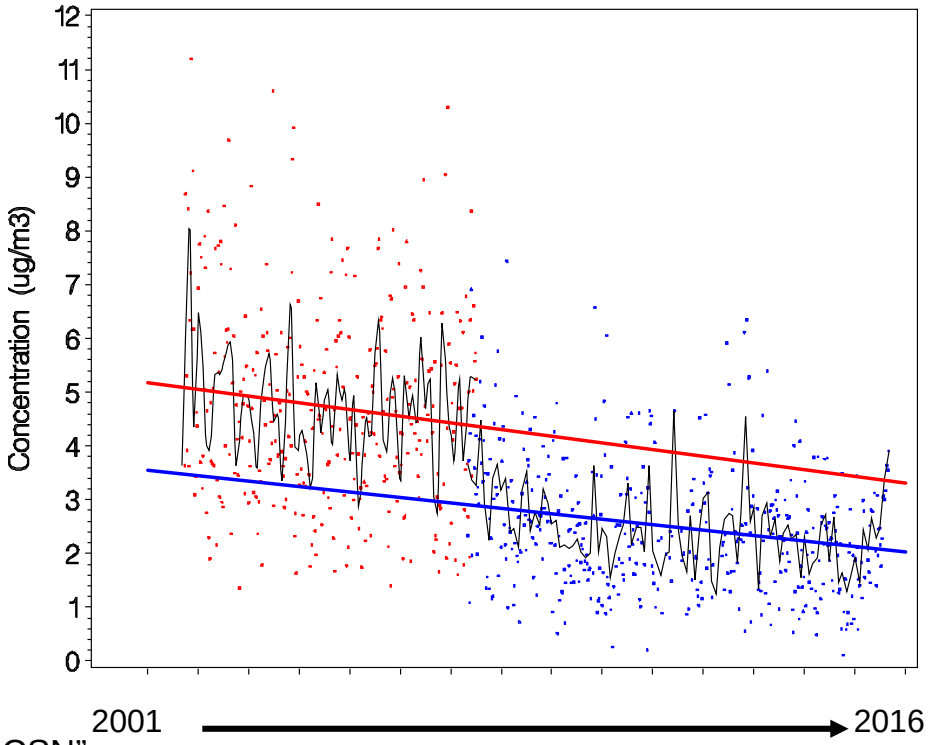
A Presentation to the 2016 National Ambient Air Monitoring Conference

Robert Lordo, Battelle
Elizabeth Landis, U.S. EPA/Office of Air Quality Planning and Standards
Joann Rice, U.S. EPA/Office of Air Quality Planning and Standards
Cheryl Triplett, Battelle

Introduction

- Since 2001**, the CSN** has measured **organic carbon (OC)** and **elemental carbon (EC)** from $PM_{2.5}$ samples taken at ~200 primarily urban sites
- The **IMPROVE network** of ~170 rural/remote sampling locations has measured OC and EC since 1989
- "**Old CSN**" protocol (2001-2010):
  - NIOSH 5040-like thermal optical transmittance (TOT) analytical method
  - One of five sampler types (predominantly Met One SASS/SuperSASS)
- Transition to "**New CSN**" protocol (2007-2009):
  - IMPROVE network's thermal optical reflectance (TOR) analytical method
  - URG-3000N sampler
  - Aimed to resolve inconsistencies in carbon sampling/analysis procedures between the (urban) CSN and (rural) IMPROVE programs, thus improving data comparability

Example Trend Plot: Organic Carbon (Winston-Salem NC site)



Red dots: Actual “Old CSN” conc.
Blue dots: Actual “New CSN” conc.
Red line: Fitted trend for “Old CSN” conc. (2001-2007)
Blue line: Fitted trend for “New CSN” conc. (2007-2015)
Black connected lines: Monthly average concs.

Overall Objective

- Use statistical techniques to generate an improved set of EC and OC measurements from the CSN and IMPROVE networks (over all monitoring years) for input to trends analysis and health risk assessments
  - For both EC and OC, assess statistical relationships between collocated “Old CSN” and “New CSN” protocols at CSN sites, and between “Old CSN” and IMPROVE protocols at IMPROVE sites
  - Determine if accounting for the change in protocol (through statistical modeling) and for data uncertainty can yield an improved historical dataset

Work done previously to relate carbon concentrations between protocols

- **Malm, WC, Schichtel, BA, and Pitchford, ML. (2011)** Uncertainties in PM2.5 gravimetric and speciation measurements and what we can learn from them. *Journal of the Air & Waste Management Association*. 61(11):1131-1149.
  - Used least-squares linear regression to adjust Old CSN carbon concentrations (MetOne SASS) to IMPROVE concentrations (to account for relative biases between systems)
  - Used data from collocated IMPROVE and CSN samplers
  - Predicted $\mathbf{EC_{IMPROVE}} = 1.3*EC_{oldCSN}$
  - Predicted $\mathbf{OC_{IMPROVE}} = (OC_{oldCSN} - 0.3*EC_{oldCSN} - a_i)/(1+b_{OC})$, where a_i is a month-specific positive organic artifact adjustment, and $1+b_{OC} = 1.2$ is a multiplicative artifact adjustment
- Our objectives were slightly different
  - Assess the relationship between Old CSN and New CSN protocols
  - Convert New CSN concentration to Old CSN concentrations for EC and OC
  - We are not considering TC as they did in formulating the models

Collocated Carbon Data Collection

At 14 CSN sites, **collocated samplers** were positioned from 2006 to 2010 to measure carbon under 2 protocols:

CSN Site	AQS ID Number	Old CSN / New CSN	Old CSN / IMPROVE
Birmingham, AL (N. Birmingham)	01-073-0023	●	●
Phoenix AZ (Supersite)	04-013-9997		●
Fresno, CA (First Street)	06-019-0008		●
Los Angeles, CA (Rubidoux W. Riverside)	06-065-8001	●	
Sacramento, CA (Del Paso Manor)	06-067-0006	●	
Denver, CO (Commerce City)	08-001-0006	●	
Atlanta, GA (South Dekalb)	13-089-0002	●	●
Chicago, IL (ComEd)	17-031-0076	●	
Detroit, MI (Allen Park)	26-163-0001	●	●
New York, NY (Bronx – IS52)	36-005-0110	●	●
New York, NY (Queens College 2)	36-081-0124	●	
Cleveland, OH (GT Craig)	39-035-0060	●	
Pittsburgh, PA (Lawrenceville)	42-003-0008		●
Seattle, WA (Beacon Hill)	53-033-0080	●	●

5 of the 14 sites included 3 samplers, utilizing the Old CSN, New CSN and IMPROVE protocols

Data Utilized for Model Fitting

- Samples collected from the 14 sites from 1/1/2006 to 12/31/2010
- Raw EC and OC measurement data stored within AQS report AMP501 for the following AQS parameter codes:
  - 88305 / 88307 – OC / EC under the old CSN (TOT) protocol
  - 88320 / 88321 – OC / EC under the IMPROVE (TOR) protocol
  - 88370 / 88380 – OC / EC under the new CSN (TOR) protocol
- Old CSN and New CSN concentration data stored within AQS are not corrected for measured blank or artifact values
- 7,300 records downloaded for Old CSN vs. New CSN data analysis
- 15,718 records downloaded for Old CSN vs. IMPROVE data analysis

Modeling Approach

- Is a linear model form adequate?

$$Y = \alpha + \beta X + [\text{other terms}] + [\text{error}]$$

  - ***Y*** denotes carbon measurement under **"old CSN"** protocol
  - ***X*** denotes carbon measurement under **"new CSN"** or **"IMPROVE"** protocol

- What are the "other terms"?

$$Y = \alpha + \beta X + [\text{other terms}] + [\text{error}]$$

  - A **site-specific** increment (positive or negative) to the overall mean α
  - A **quarter (season)-specific** increment (positive or negative) to the overall mean α

Modeling Approach (cont.)

- "Ordinary least squares" regression ignores uncertainty present in the predictor (*X*) variable

$$Y = \alpha + \beta X + [\text{other terms}] + [\text{error}]$$

  - We should be aware that *X* is subject to uncertainty as it's a measured concentration
  - If it's negligible relative to uncertainty in *Y,* we can ignore it
  - If it's not negligible, then the slope term β will be underestimated
- A *structural equations modeling approach* was used to account for uncertainty in *X*
  - This approach was applied to each of the four combinations of carbon parameter (EC and OC) and protocol associated with *X* (new CSN or IMPROVE).

Final Model

Expected (predicted) value of Y (old CSN) =

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}X_i + \text{[average site effect]} + \text{[average season effect]}$$

$$= \hat{\alpha}^* + \hat{\beta}X_i$$

$\hat{\alpha}^*$ represents the smallest possible predicted value for the New CSN carbon concentration

Converting **IMPROVE** concentration (*X*) to **Old CSN** concentration ($\hat{y}_i$):

EC: $\hat{y}_i = 0.139769 + 0.75024*(\text{IMPROVE}_i)$

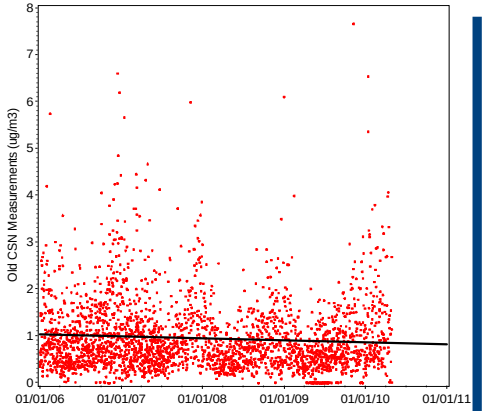
OC: $\hat{y}_i = 1.72511 + 1.30492*(\text{IMPROVE}_i)$

Converting **New CSN** concentration (*X*) to **Old CSN** concentration ($\hat{y}_i$):

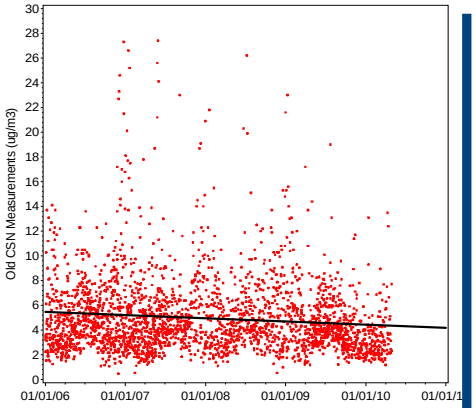
EC: $\hat{y}_i = 0.104071 + 0.92462*(\text{NEW\_CSN}_i)$

OC: $\hat{y}_i = 1.57432 + 1.15101*(\text{NEW\_CSN}_i)$

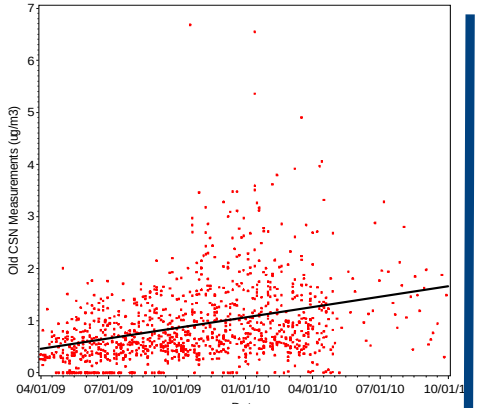
Actual (red) and Predicted (blue) "Old CSN" Concentrations (plotted versus time)



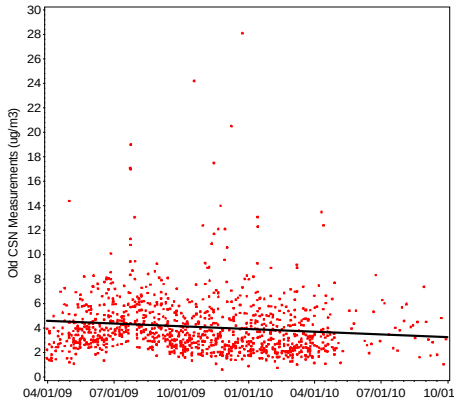
EC: *X*=IMPROVE (1/2006 – 4/2010)



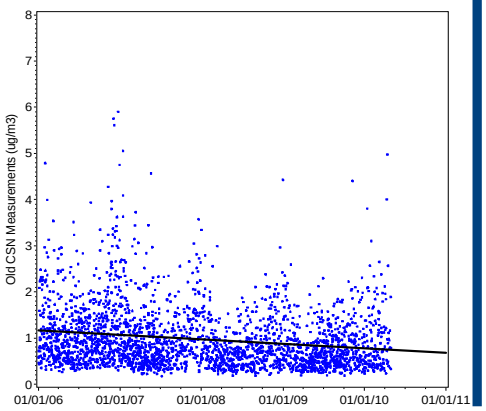
OC: *X*=IMPROVE (1/2006 – 4/2010)

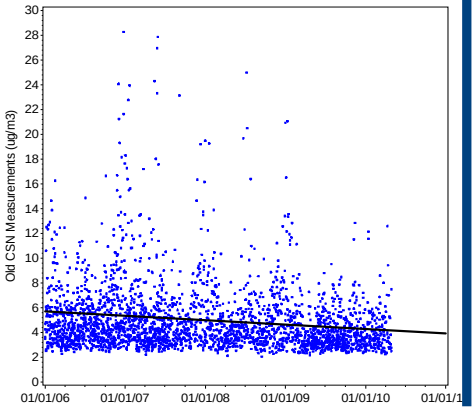


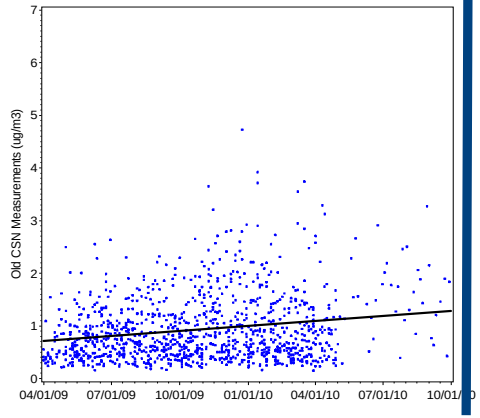
EC: *X*=New CSN (4/2009 – 10/2010)

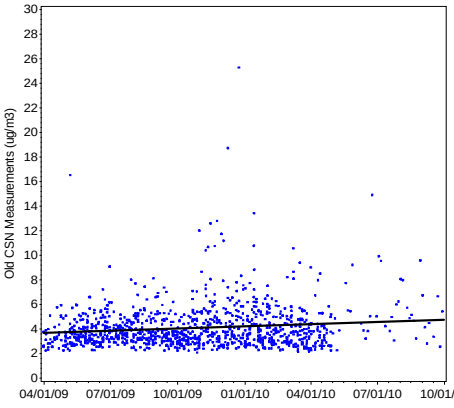


OC: *X*=New CSN (4/2009 – 10/2010)





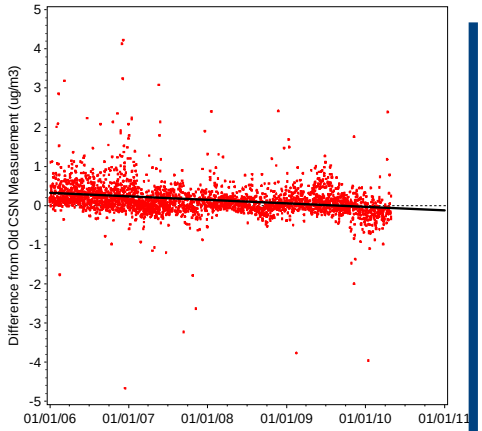




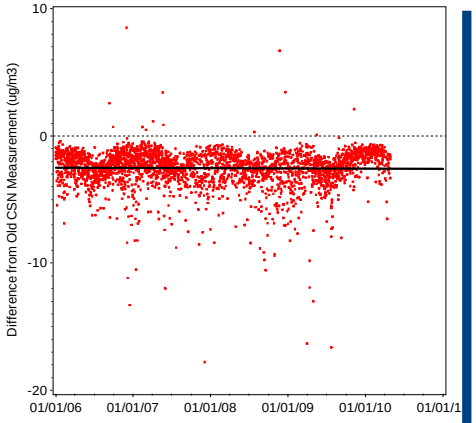
Black line = fitted time trend line

Battelle
The Business of Innovation

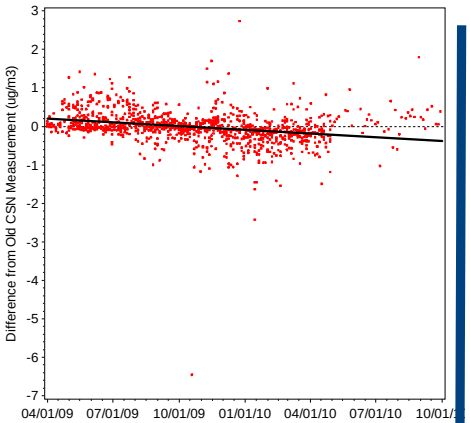
X (see below) Minus Actual (red) and Predicted (blue) "Old CSN" Concentrations (plotted versus time)



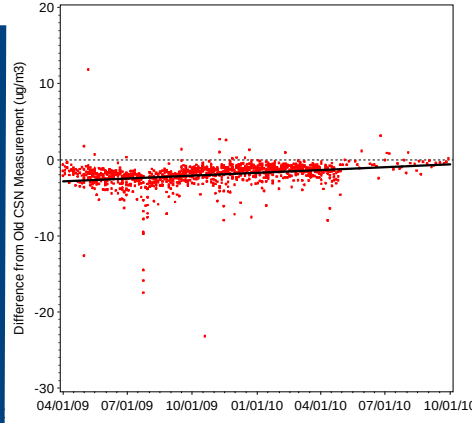
EC: *X*=IMPROVE (1/2006 – 4/2010)



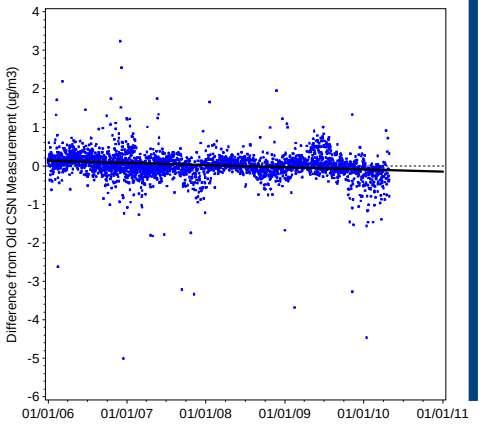
OC: *X*=IMPROVE (1/2006 – 4/2010)

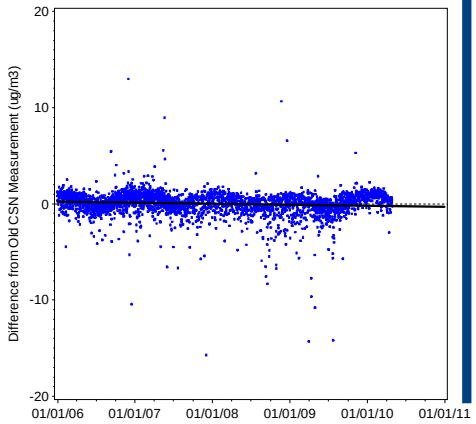


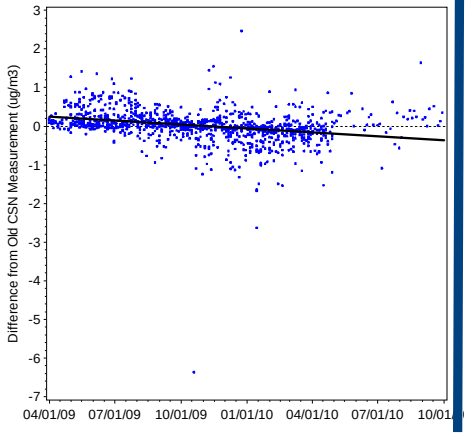
EC: *X*=New CSN (4/2009 – 10/2010)

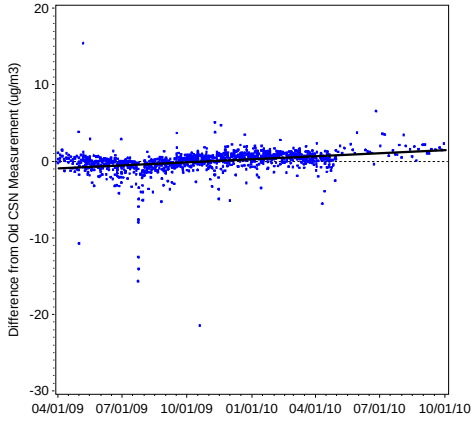


OC: *X*=New CSN (4/2009 – 10/2010)









Black line = fitted time trend line

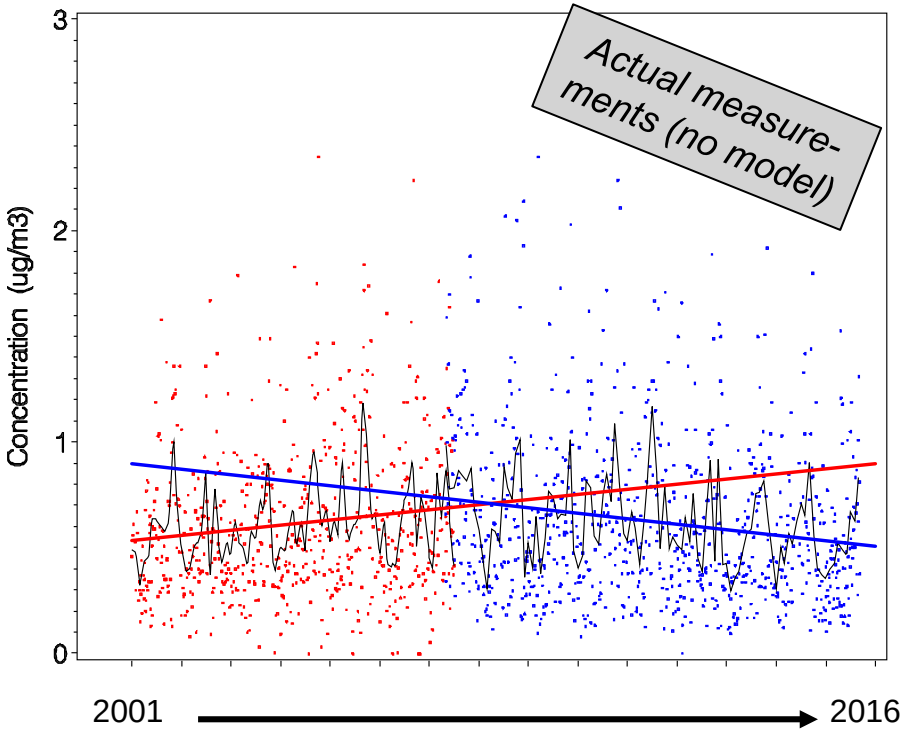
Battelle
The Business of Innovation

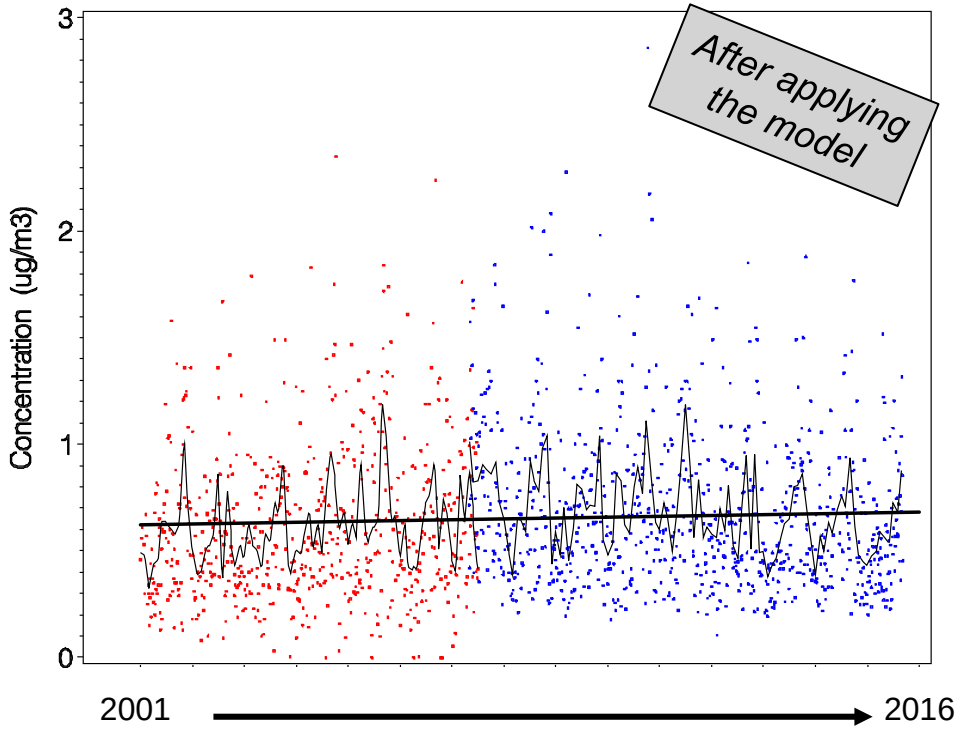
Applying the Model to Carbon Concentrations from Selected CSN Sites

CSN Site	AQS Site ID
El Cajon (San Diego Co.), CA (Floyd Smith Dr.)	06-073-0003
Washington, DC (McMillan Reservoir)	11-001-0043
Indianapolis, IN (Washington Park)	18-097-0078
Boston, MA (Dudley Square)	25-025-0042
St. Louis, MO (Blair Street)	29-510-0085
Las Vegas, NV (Jerome Mack MS)	32-003-0540
Winston Salem, NC (Hattie Ave.)	37-067-0022
Houston, TX (Deer Park)	48-201-1039
Salt Lake City, UT (Hawthorne)	49-035-3006

These sites are health effects research cities of interest

Example Trend Plot: Elemental Carbon (Indianapolis IN site)

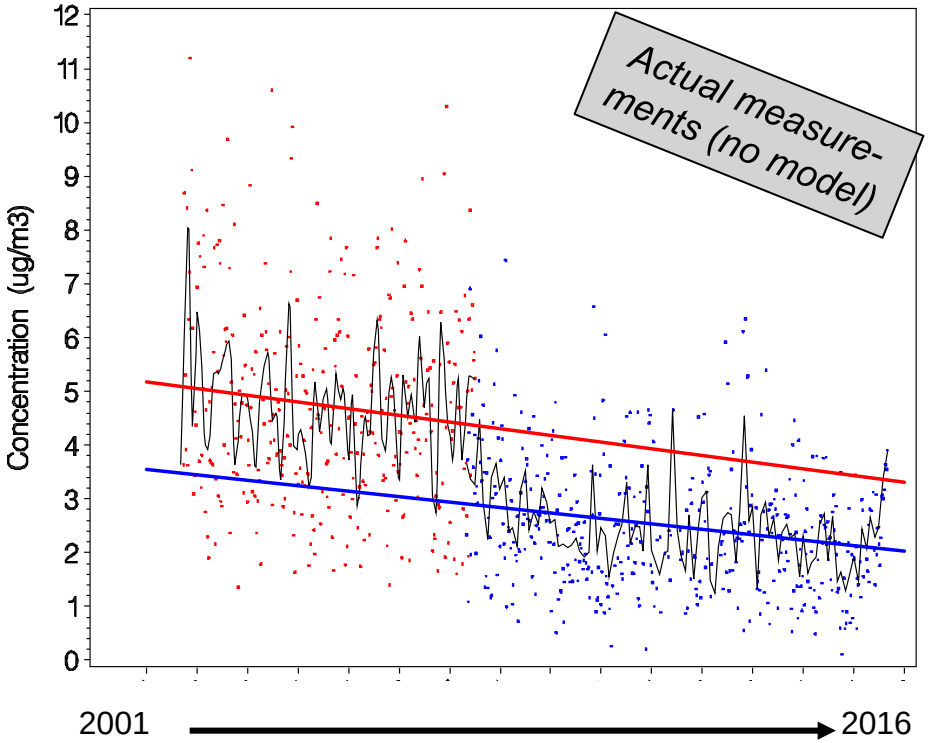


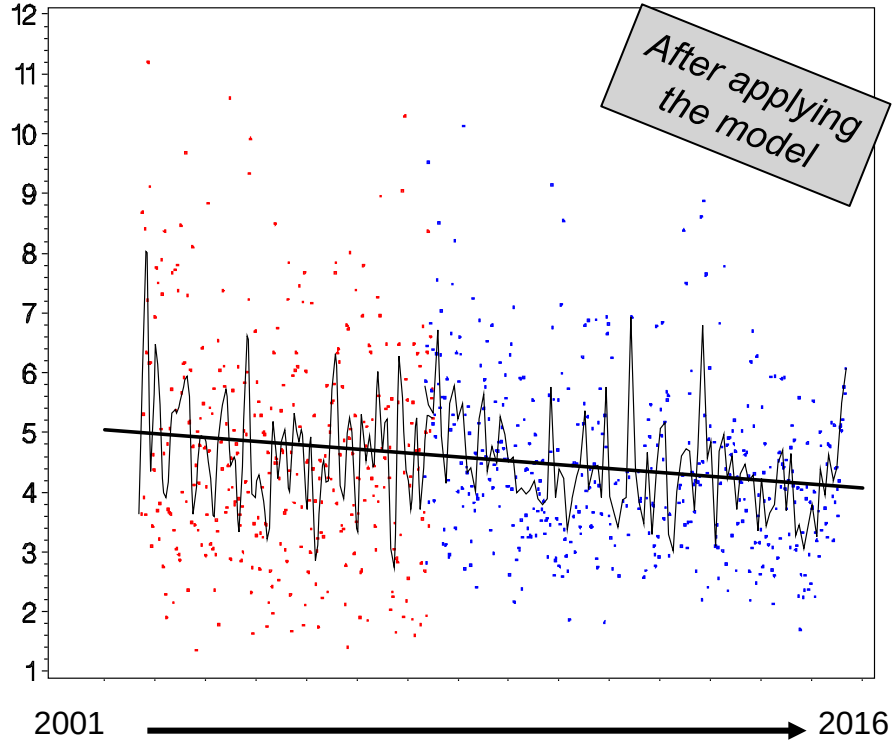


Red dots: Actual "Old CSN" conc.
Blue dots: Actual "New CSN" conc.
Red line: Fitted trend for "Old CSN" conc. (2001-2007)
Blue line: Fitted trend for "New CSN" conc. (2007-2015)
Black connected lines: Monthly average concs.

Red dots: Actual "Old CSN" conc.
Blue dots: Predicted "Old CSN" conc. from New CSN conc.
Thick black line: Fitted trend (2001-2015)
Black connected lines: Monthly average concs.

Example Trend Plot: Organic Carbon (Winston-Salem NC site)





Red dots: Actual "Old CSN" conc.
Blue dots: Actual "New CSN" conc.
Red line: Fitted trend for "Old CSN" conc. (2001-2007)
Blue line: Fitted trend for "New CSN" conc. (2007-2015)
Black connected lines: Monthly average concs.

Red dots: Actual "Old CSN" conc.
Blue dots: Predicted "Old CSN" conc. from New CSN conc.
Thick black line: Fitted trend (2001-2015)
Black connected lines: Monthly average concs.

Key Findings in Trend Plots

- For Elemental Carbon:
  - Shifts due to the protocol change are slight
  - Overall trends were flat in general
  - Most common type of trend is a slight increase, followed by a slight decrease
- For Organic Carbon:
  - Shifts due to the protocol change are much more noticeable than for elemental carbon
  - Model adjustment is critical to proper characterization of the trend in OC over the history of the CSN (2001 – 2015)
  - Most sites show a declining trend over time – the trend is steeper in the earlier years

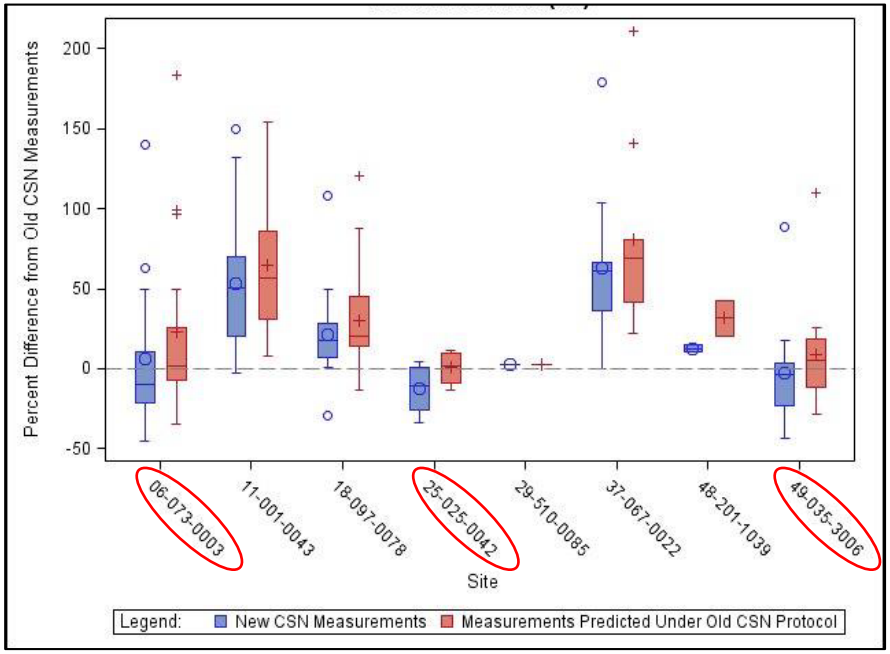
Conclusions

- An adjustment of “New CSN” protocol data to be more comparable to “Old CSN” protocol data is more critical for OC than for EC.
- The model yields some improvements in predicting OC and EC trends, but more so with OC
  - However, accounting for specific quarter (season) effects, rather than average quarter effects, would be expected to result in more accurate trend prediction.
- We should continue to assess the effectiveness of the model adjustment on characterizing trend for other sites.

BACKUP

Percent Difference from Actual "Old CSN" Concentrations: Elemental Carbon

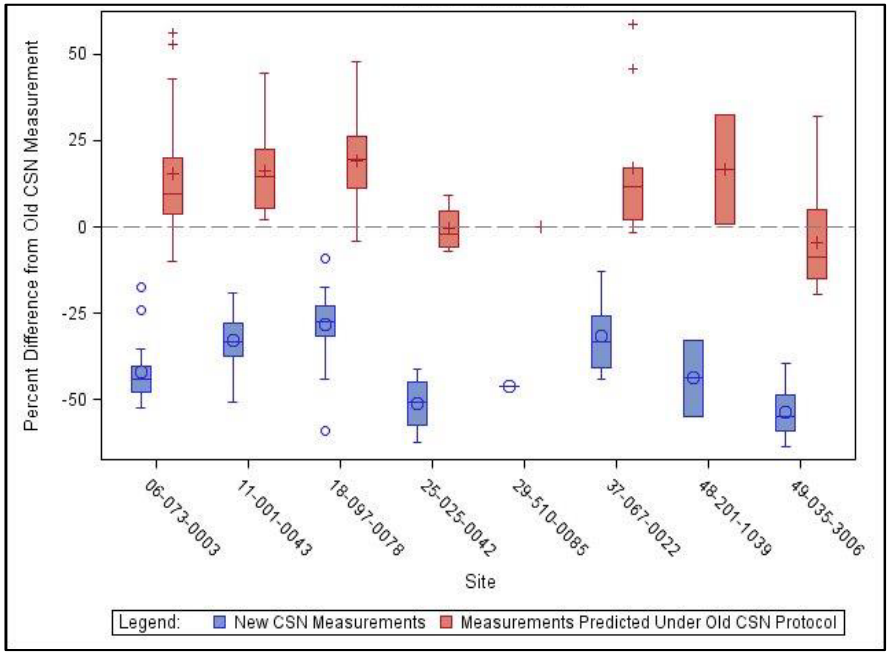
Data represented in these boxplots cover dates where both "Old CSN" and "New CSN" concs. were collected at the specified site (1 to 18 days)



- **Blue boxplots: Actual New CSN concentrations**
- **Red boxplots: Model-predicted Old CSN concentrations**
- Mixed results across these sites
  - For 3 sites (circled), the 0 reference line falls more within the red boxplot, indicating better agreement with the actual "Old CSN" concs on a given day upon fitting the model
  - For other sites, both boxplots fall primarily above the 0 reference line, indicating higher values under the Old CSN protocol in both cases

Percent Difference from Actual "Old CSN" Concentrations: Organic Carbon

Data represented in these boxplots cover dates where both "Old CSN" and "New CSN" concs. were collected at the specified site (1 to 18 days)



- **Blue boxplots: Actual New CSN concentrations**
  - Old CSN protocol consistently yields ~25-50% lower concentrations compared to the New CSN protocol at each site on a given day
  - Increased flow rate with the URG vs MetOne sampler, along with the filter face velocity, yield lower OC concentrations from the New CSN sampler
- **Red boxplots: Model-predicted Old CSN concentrations**
  - The model applies a slight over-adjustment to the New CSN measurements, but it is an improvement over no adjustment at all