

Investigation of OM/OC Using Multiple Regression

IMPROVE Carbon Trends Workshop, 2010

*Heather Simon, Prakash Bhawe, Jenise Swall, Neil Frank,
and Bill Malm*



Acknowledgements: Marc Pitchford, Chuck McDade,
Lowell Ashbaugh, Warren White, Venkatesh Rao,
Steve McDow, Chris Nolte

Outline

- Motivation and background
- Regression methodology
 - Physical interpretation of coefficients
 - Pitfalls of multiple regression analysis
 - Measurement uncertainty
 - Colinearity between covariates
- Results

Importance of OM/OC

- Key uncertainty in almost any PM mass closure calculation
- Correlated with aerosol properties (i.e. hygroscopicity and light extinction)
- Major source of error in air quality model evaluations
 - Routine monitoring networks measure OC
 - Most models predict OM
- Organic Mass (OM): total organic mass in PM
- Organic Carbon (OC): the portion of OM made up of carbon

Current measurement techniques for OM/OC

- GC/MS speciation (Turpin and Lim, 2001)
- HR-ToF-MS (Aiken et al., 2008; sun et al., 2009)
- FTIR spectroscopy (Russell et al., 2003 and 2009; many others)
- Sequential extraction (El-Zanan et al., 2005; Lowenthal et al., 2009; Polidori et al., 2008)
- Coupled thermal gravimetric and chemical analyses (Chen et al., 2007)
- Mass closure using STN data - SANDWICH (Frank, 2006)
- IMPROVE network data analysis
 - Mass closure (El-Zanan et al., 2005)
 - Regression (Malm and Hand, 2007)

Measurement techniques for OM/OC using IMPROVE data

- Mass closure - El -Zanan et al., 2005

$$[OM] = PM_{2.5} - ([(NH_4)_2SO_4] + [NH_4NO_3] + [SOIL] + [EC] + [TraceElements])$$

- Assumptions include sulfate neutralization, no particle-bound water, no nitrate volatilization

- Regression - Malm and Hand, 2007

$$[PM_{2.5}] = \beta_0 + \beta_1[OC] + \beta_2[(NH_4)_2SO_4] + \beta_3[NH_4NO_3] + \beta_4[SOIL] + \beta_5[EC] + \beta_6[SS] + \varepsilon$$

- Does not rely on assumptions about 1) the presence of unmeasured components (ammonium and water), 2) the amount of nitrate volatilization, or 3) the accuracy of the IMPROVE soil equation.

- We extend Malm and Hand's regression technique

Methods

- Use 2002-2008 data from IMPROVE network
- Samples were split up by site and quarter
- Sites that averaged less than 15 samples/quarter were not analyzed : 153 sites * 4 quarters = 612 regression analyses

$$[PM_{2.5}]_i = \beta_{OC}[OC]_i + \beta_{sulf}[(NH_4)_2SO_4]_i + \beta_{nit}[NH_4NO_3]_i + \beta_{soil}[SOIL]_i + [EC]_i + 1.2[K_{non}]_i + 1.8[Cl^-]_i + \varepsilon_i$$

$$[SOIL] = 3.48[Si] + 1.63[Ca] + 2.42[Fe] + 1.94[Ti]$$

$$[K_{non}] = K - 0.6[Fe]$$

- β_{OC} , β_{sulf} , and β_{nit} were allowed to vary by quarter. β_{soil} was held constant on an annual basis
- No filtering of sampling data within a site/quarter grouping

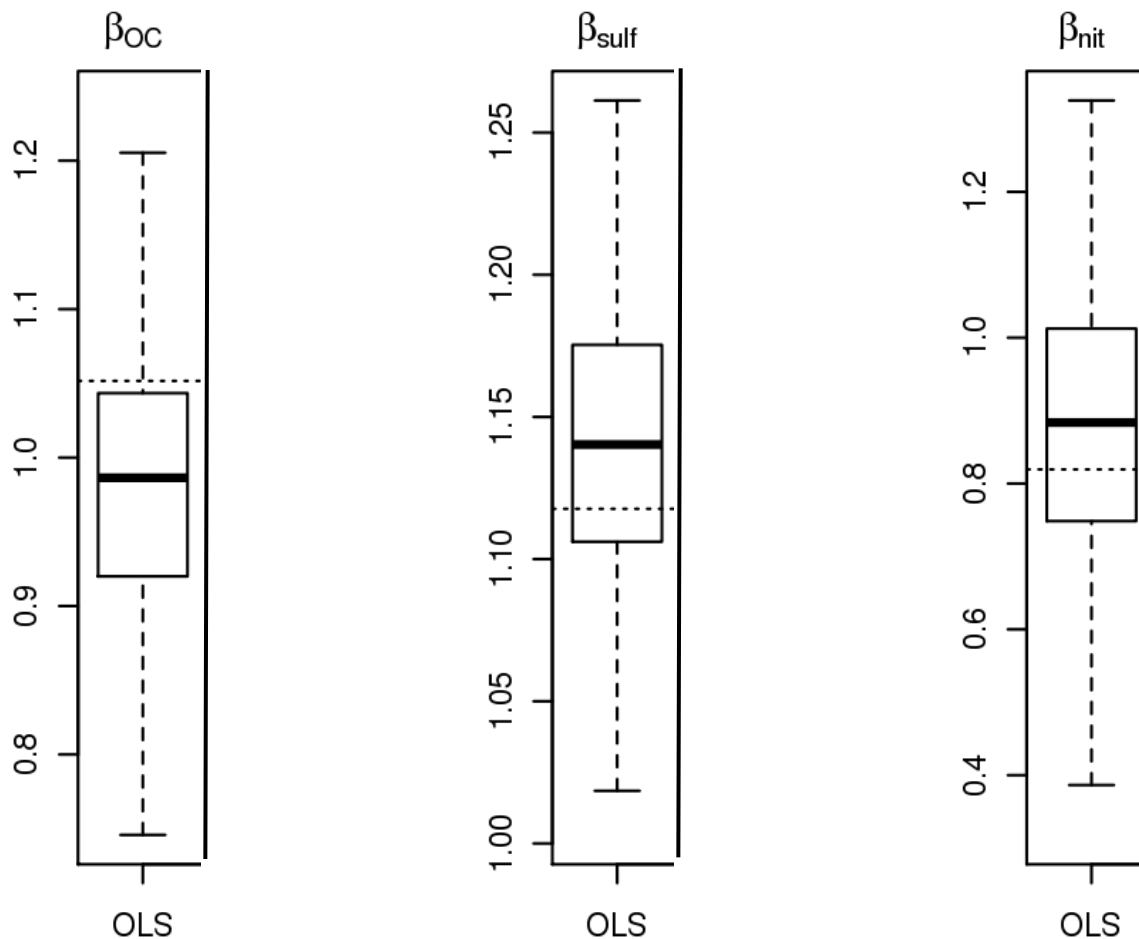
Pitfalls of multiple regression analysis

- Model selection
 - Does the regression equation capture all elements of the system?
 - Are explanatory variables linearly related to the response variable?
- Dataset selection – datasets should be selected such that β_{OC} , β_{sulf} , β_{nit} , and β_{soil} are expected to be relatively constant
- Measurement uncertainty in explanatory variables
- Colinearity of explanatory variables

Effects of measurement uncertainty

- Measurement uncertainty in explanatory variables biases results
- With one explanatory variable - the coefficient is biased towards zero (Fuller, 1987; Saylor et al., 2006; White, 1998).
- With multiple explanatory variables, biases depend on:
 - relative uncertainties in explanatory variables
 - correlation between explanatory variables
 - correlation between measurement errors.
- White (1998) examined this problem with two independent explanatory variables for which measured values and measurement uncertainties were not correlated
- We perform evaluation of bias within more complex conditions of our present study

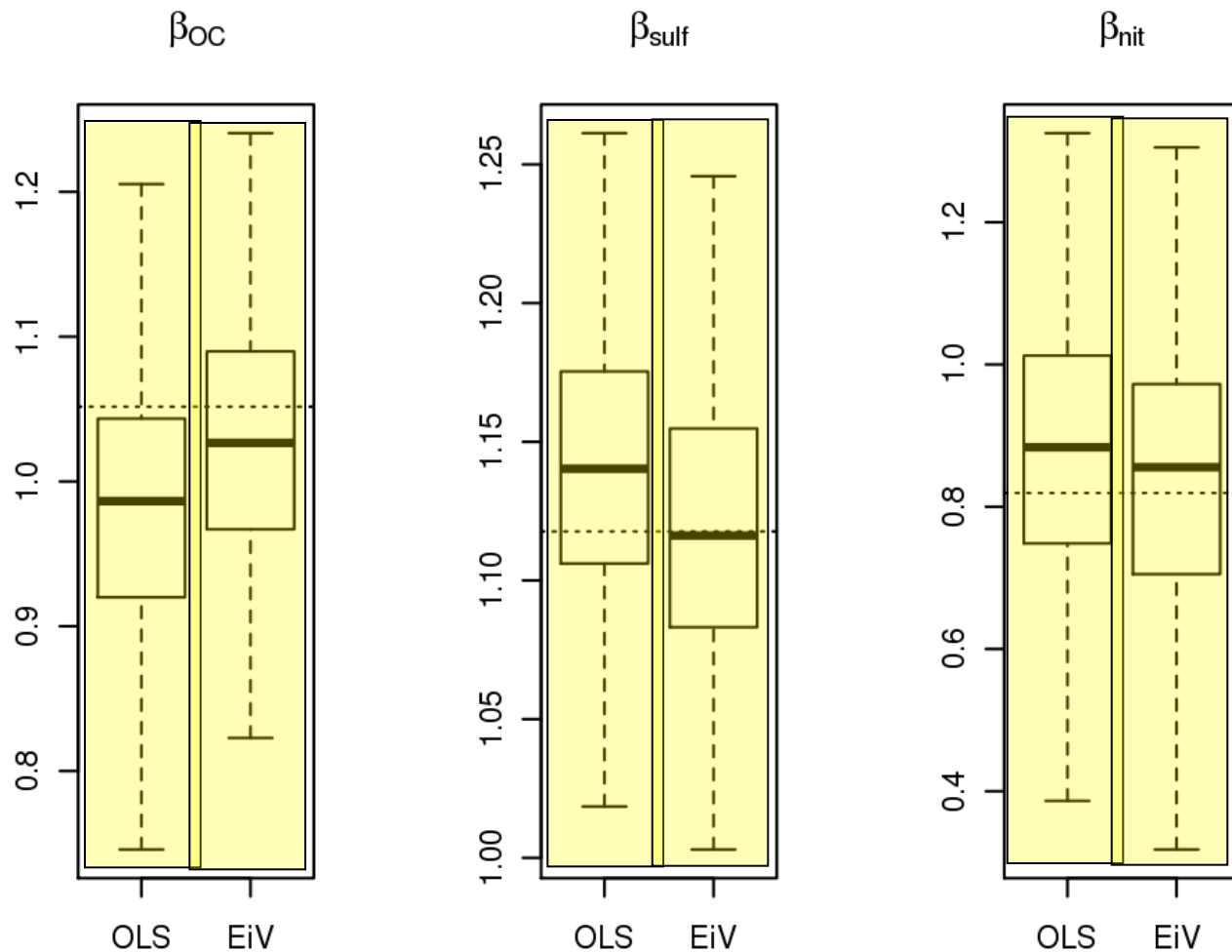
Effects of measurement uncertainty



Effects of measurement uncertainty

- To correct for these biases, we use an errors-in-variables (EiV) model based on Fuller (1987)
- Use reported uncertainty for each measurement
- Simplifying assumptions
 - Measurement errors are independent among all explanatory variables
 - Measurement error is normally distributed
 - Accurate measurement uncertainty estimates
- This method leads to a reduction in bias

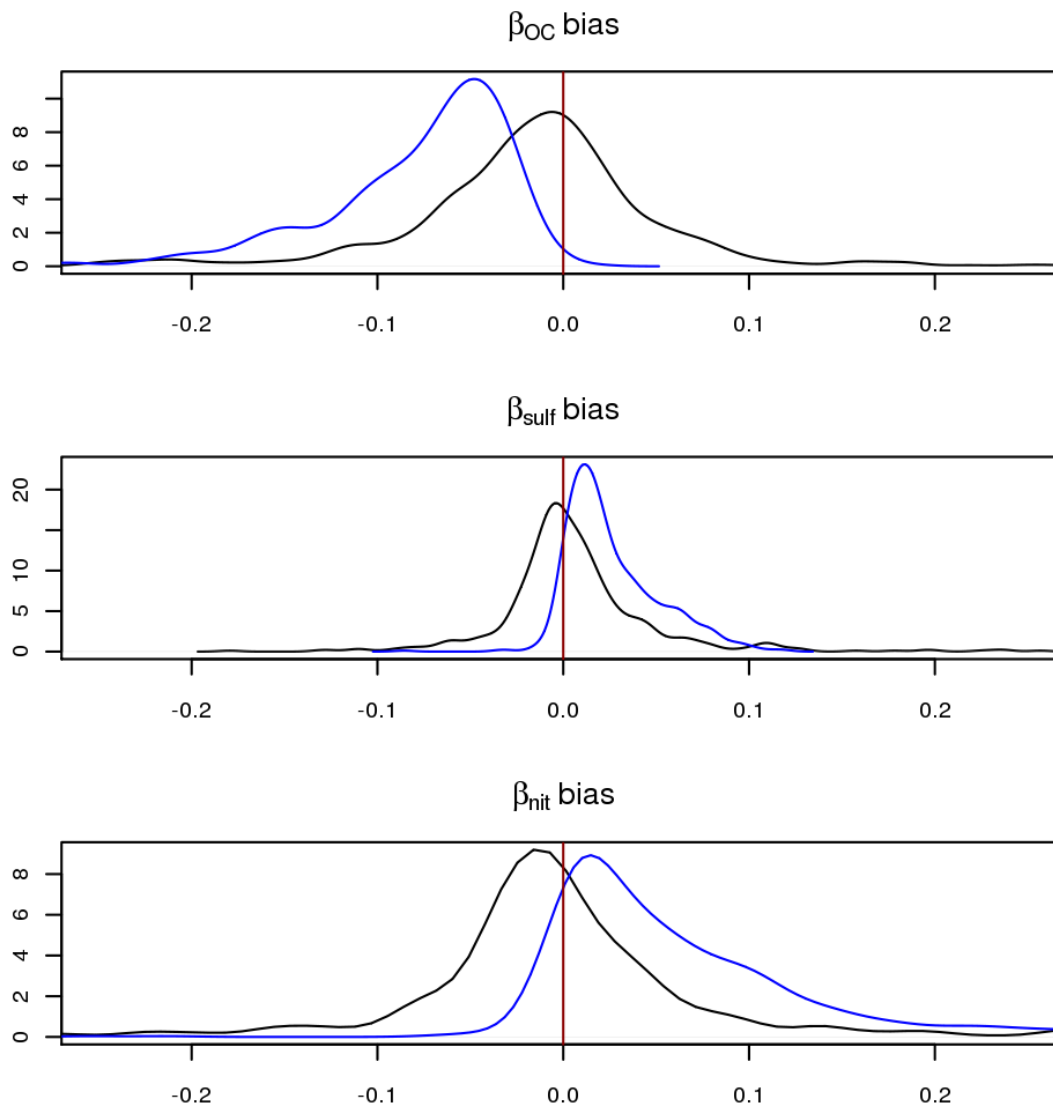
Effects of measurement uncertainty



- OLS = ordinary least squares regression
- EiV = errors in variables regression

Gila Wilderness, NM – quarter 1

Effects of measurement uncertainty



OLS

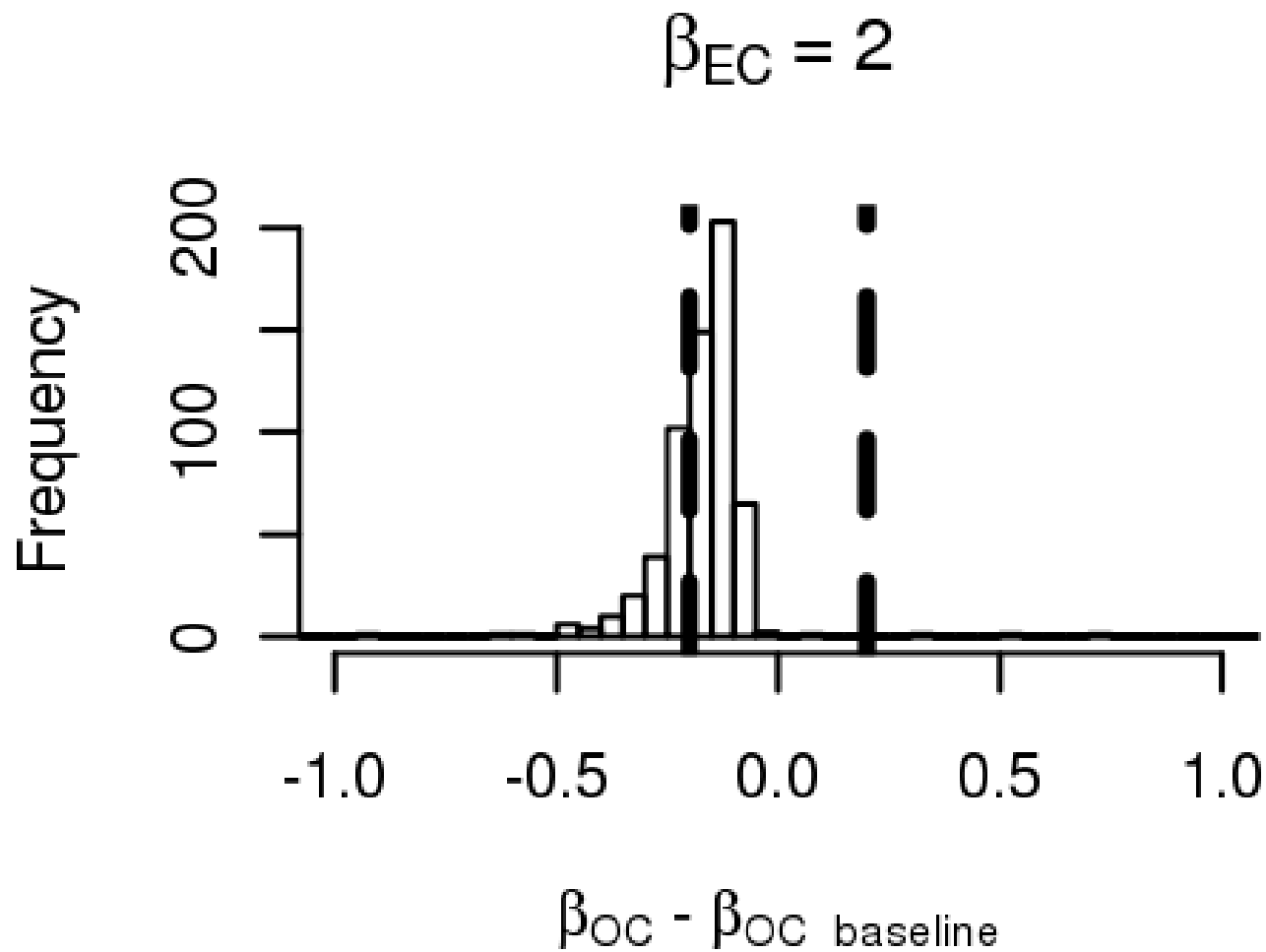
EiV

Median bias across all regressions

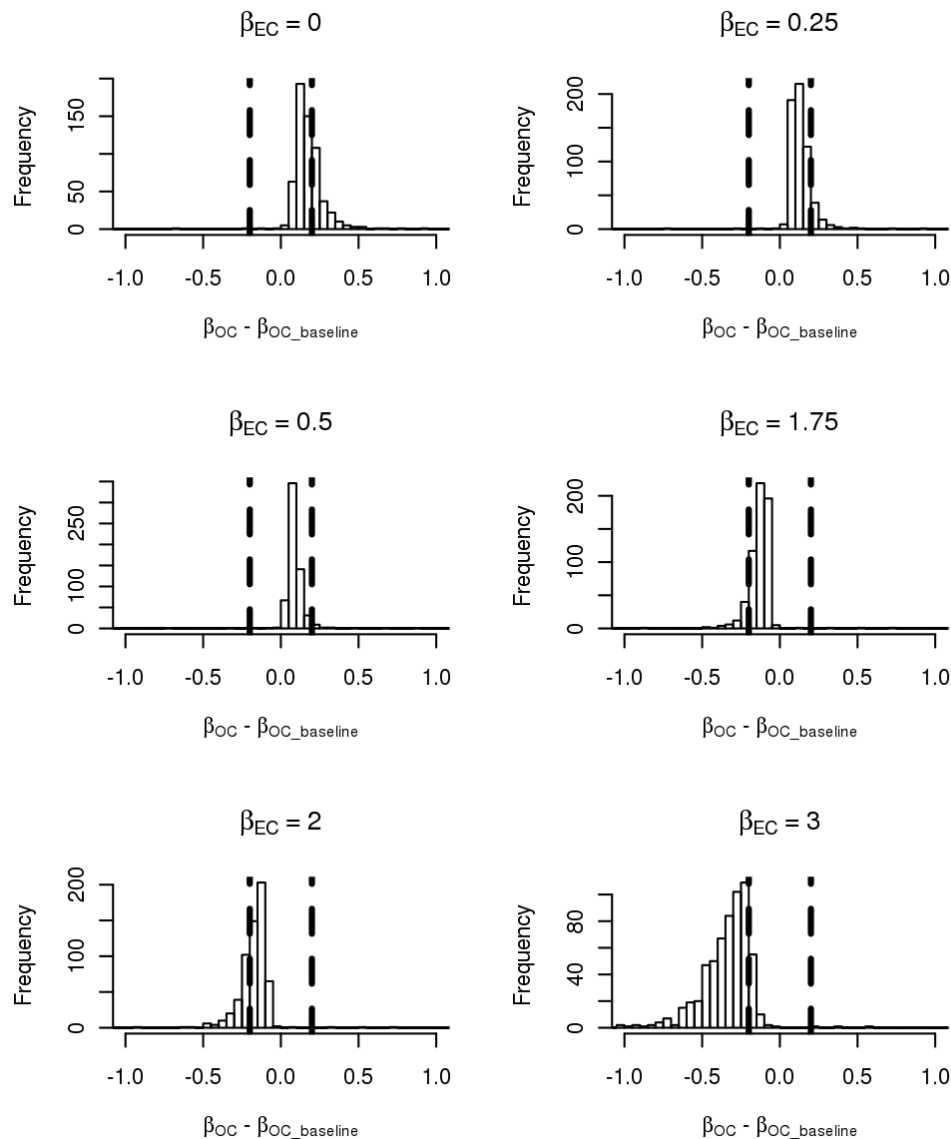
Pitfalls of multiple regression analysis

- Model selection
 - Does the regression equation capture all elements of the system?
 - Are explanatory variables linearly related to the response variable?
- Dataset selection – datasets should be selected such that β_{OC} , β_{sulf} , β_{nit} , and β_{soil} are expected to be relatively constant
- Measurement uncertainty in explanatory variables
- Colinearity of explanatory variables

Effects of colinearity among explanatory variables: EC and OC



Effects of colinearity among explanatory variables: EC and OC



Effects of colinearity among explanatory variables: EC and OC

- We fix $\beta_{EC} = 1$, due to high colinearity between OC and EC
- Sensitivity study using TC as an explanatory variable instead of OC, shows that $\beta_{EC} = 1$ should not significantly bias results
- We eliminate any datasets for which $r_{\text{Pearson}} > 0.85$ between any two explanatory variables (OC, sulfate, nitrate, soil).

Methods

- Use 2002-2008 data from IMPROVE network
- Samples were split up by site and quarter
- Sites that averaged less than 15 samples/quarter were not analyzed : 153 sites * 4 quarters = 612 regression analyses

$$[PM_{2.5}]_i = \beta_{OC}[OC]_i + \beta_{sulf}[(NH_4)_2SO_4]_i + \beta_{nit}[NH_4NO_3]_i + \beta_{soil}[SOIL]_i + [EC]_i + 1.2[K_{non}]_i + 1.8[Cl^-]_i + \varepsilon_i$$

- Out of 612 Errors-in-Variables regressions:
 - 511 high confidence results
 - 19 eliminated due to $r_p > 0.85$ or high correlation between ε_i and PM constituents
 - 35 flagged for suspicious temporal trends in ε_i
 - 61 regressions had one or more unrealistic β coefficient

Methods

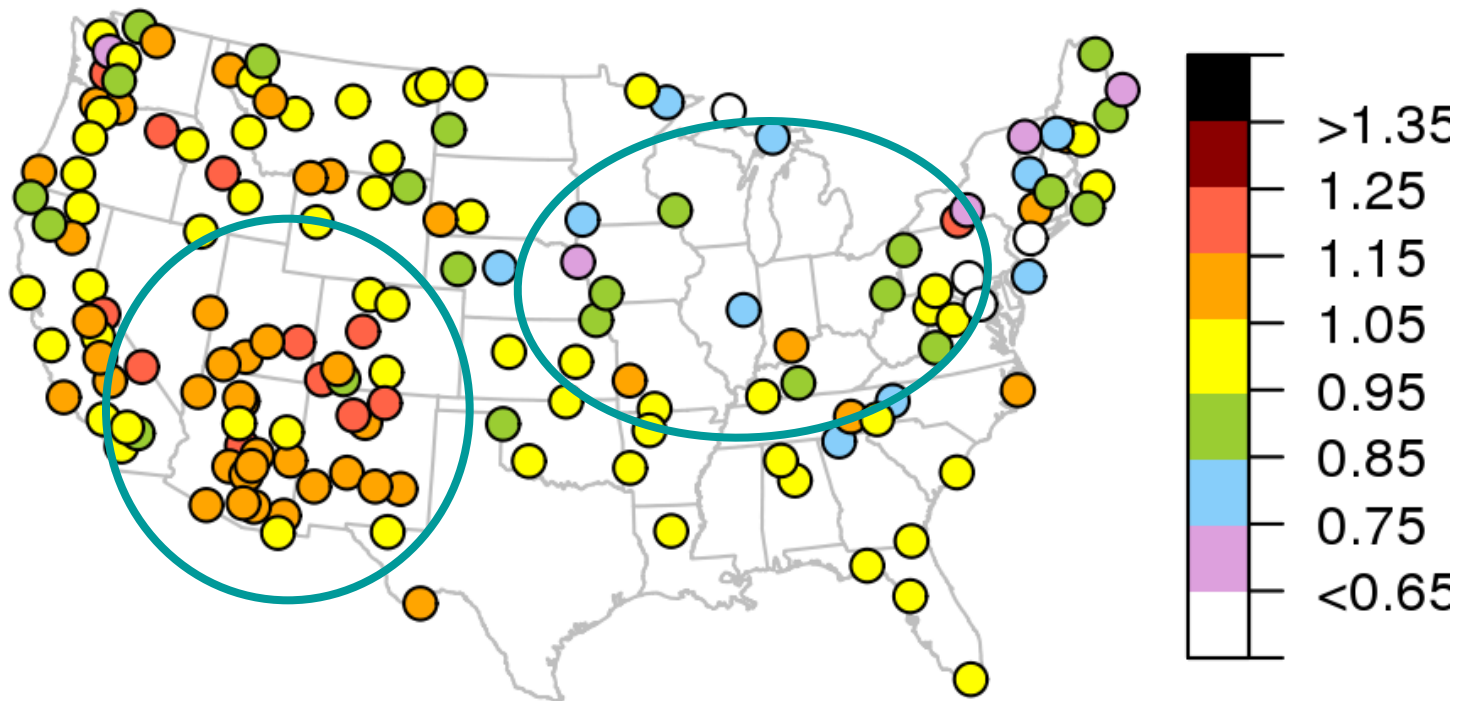
- Use 2002-2008 data from IMPROVE network
- Samples were split up by site and quarter
- Sites that averaged less than 15 samples/quarter were not analyzed : 153 sites * 4 quarters = 612 regression analyses

$$[PM_{2.5}]_i = \beta_{OC}[OC]_i + \beta_{sulf}[(NH_4)_2SO_4]_i + \beta_{nit}[NH_4NO_3]_i + \beta_{soil}[SOIL]_i + [EC]_i + 1.2[K_{non}]_i + 1.8[Cl^-]_i + \varepsilon_i$$

-If $\beta_{soil} \neq 1$, soil composition different from that used to create the IMPROVE soil equation

-For varying soil compositions, $\beta_{soil} = 0.41 - 1.63$

β_{soil} results

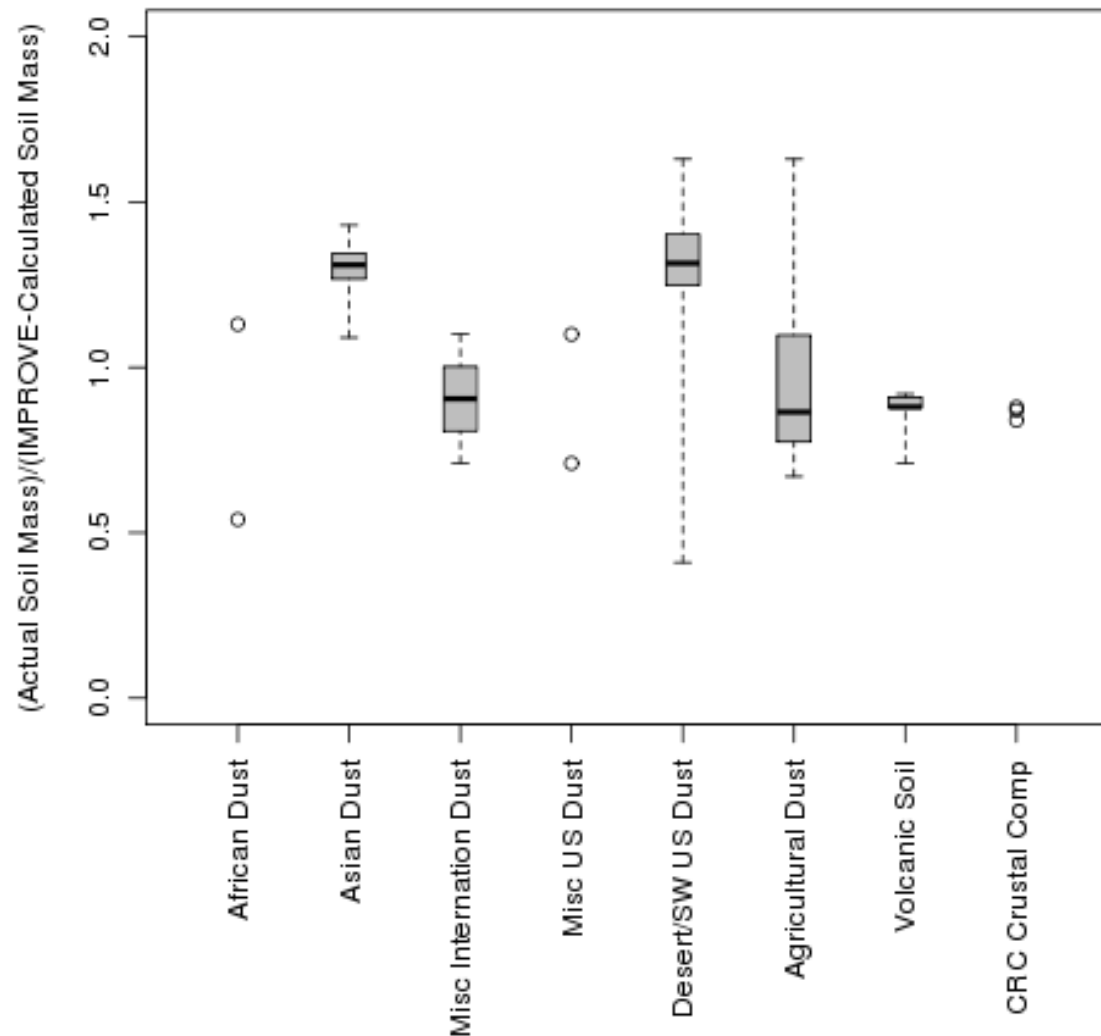


- Desert soil should have β_{soil} in the range of 1.25 – 1.40
- Agricultural soil should have β_{soil} in the range of 0.78-1.10

Simon et al. (2010)

β_{soil} results

- Soil composition profiles for different soil types were identified in EPA's SPECIATE database and in the literature
- These profiles were used to calculate over/under-estimates from the IMPROVE soil equation for different soil types



Methods

- Use 2002-2008 data from IMPROVE network
- Samples were split up by site and quarter
- Sites that averaged less than 15 samples/quarter were not analyzed : 153 sites * 4 quarters = 612 regression analyses

$$[PM_{2.5}]_i = \beta_{OC}[OC]_i + \beta_{sulf}[(NH_4)_2SO_4]_i + \beta_{nit}[NH_4NO_3]_i + \beta_{soil}[SOIL]_i + [EC]_i + 1.2[K_{non}]_i + 1.8[Cl^-]_i + \varepsilon_i$$

$-\beta_{sulf} < 1$: partly neutralized sulfate

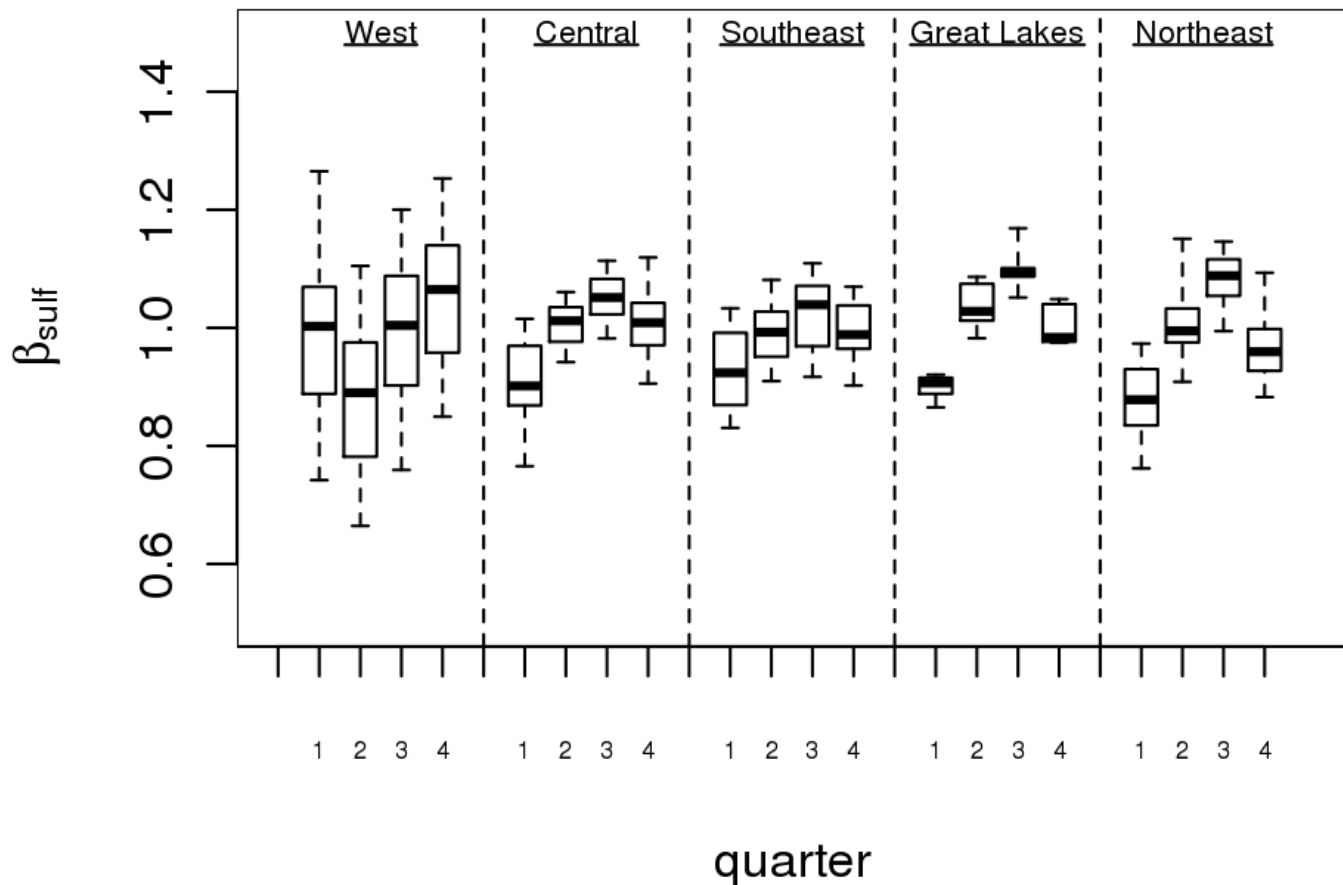
-for $(NH_4)HSO_4$, $\beta_{sulf} = 0.87$:

-for H_2SO_4 , $\beta_{sulf} = 0.74$

$-\beta_{sulf} > 1$: hydrated sulfate

-At highest RH measured in the lab, $\beta_{sulf} = 1.53$

β_{sulf} results



- Except in the west, β_{sulf}
 - >1 in summer (hydrated)
 - <1 in winter (not fully neutralized)

Q1 = JFM
Q2 = AMJ
Q3 = JAS
Q4 = OND

Methods

- Use 2002-2008 data from IMPROVE network
- Samples were split up by site and quarter
- Sites that averaged less than 15 samples/quarter were not analyzed : 153 sites * 4 quarters = 612 regression analyses

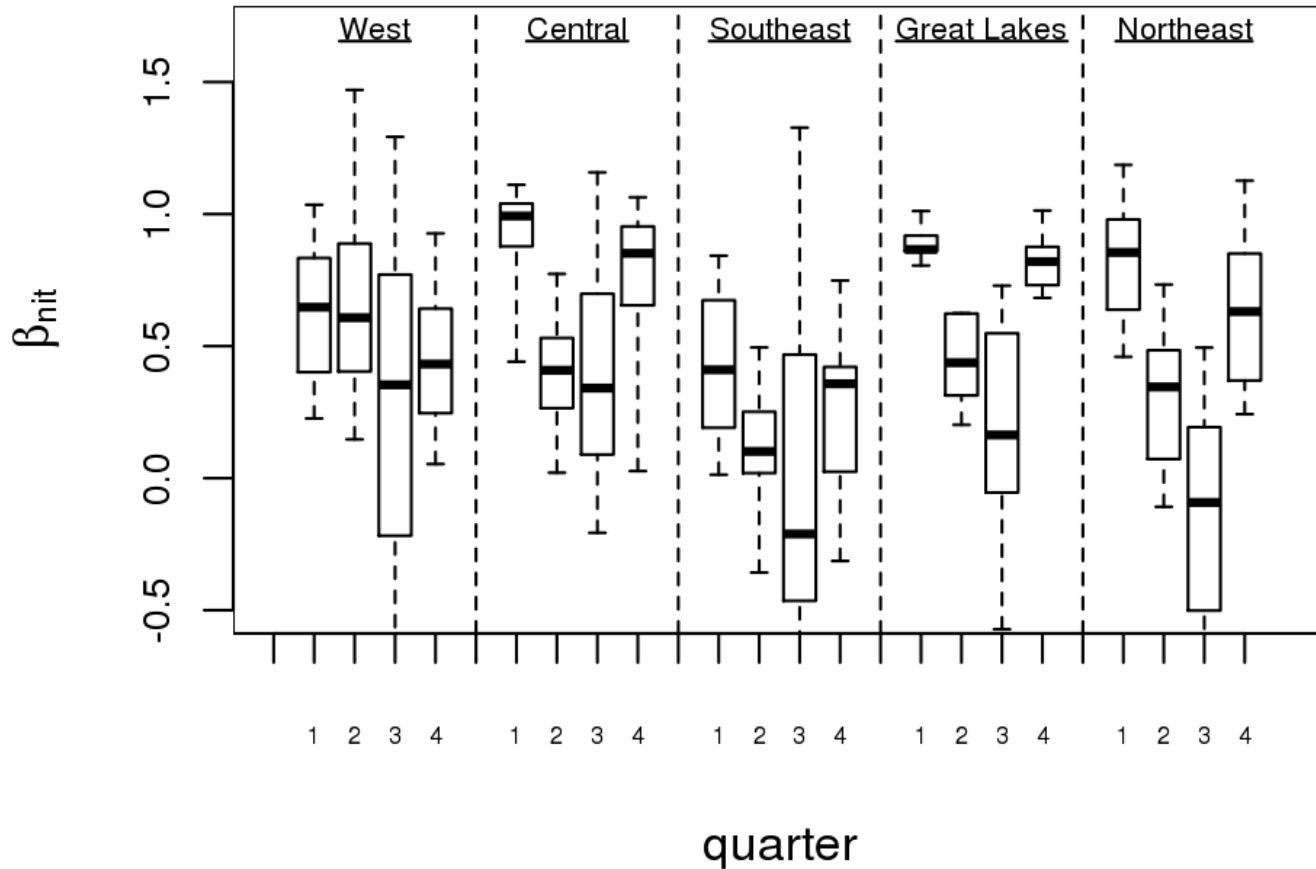
$$[PM_{2.5}]_i = \beta_{OC}[OC]_i + \beta_{sulf}[(NH_4)_2SO_4]_i + \beta_{nit}[NH_4NO_3]_i + \beta_{soil}[SOIL]_i + [EC]_i + 1.2[K_{non}]_i + 1.8[Cl^-]_i + \varepsilon_i$$

$-\beta_{nit} < 1$: partial or total nitrate volatilization.

$-\beta_{nit} > 1$: hydrated NH_4NO_3 or $NaNO_3$.

-At highest RH measured in the lab, $\beta_{nit} = 1.35$

β_{nit} results



- lower in warm regions (Southeast)
 - lower in summer than in winter
 - large variability in summer – possibly due to low nitrate concentrations
- } Nitrate volatilizes at higher temps

Methods

- Use 2002-2008 data from IMPROVE network
- Samples were split up by site and quarter
- Sites that averaged less than 15 samples/quarter were not analyzed : 153 sites * 4 quarters = 612 regression analyses

$$[PM_{2.5}]_i = \beta_{OC}[OC]_i + \beta_{sulf}[(NH_4)_2SO_4]_i + \beta_{nit}[NH_4NO_3]_i + \beta_{soil}[SOIL]_i + [EC]_i + 1.2[K_{non}]_i + 1.8[Cl^-]_i + \varepsilon_i$$

$$-\beta_{OC} = OM/OC$$

-POA : OM/OC = 1.2 (vehicle exhaust) – 1.8 (wood smoke)

-SOA : OM/OC = 1.4 - 2.7

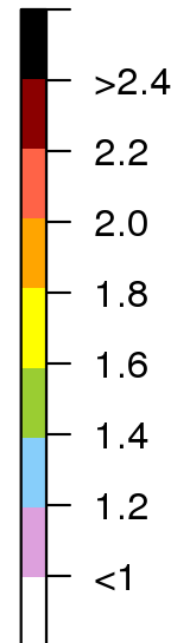
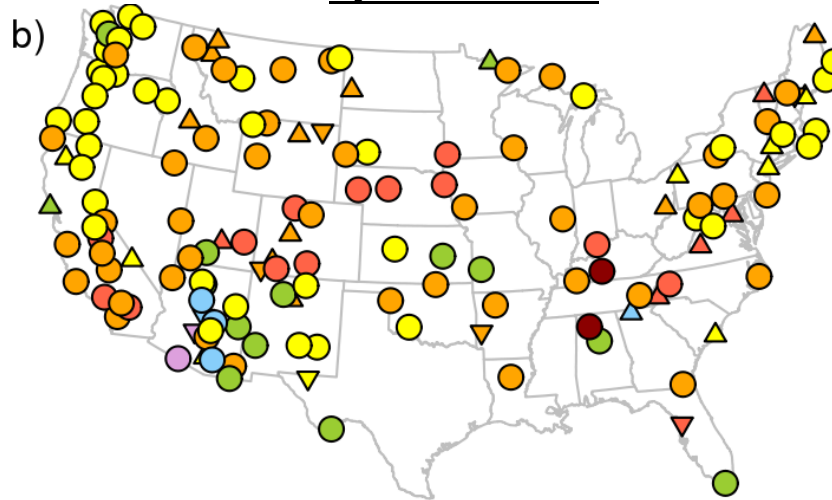
-Oxidative aging increases OM/OC

-By definition, OM/OC > 1

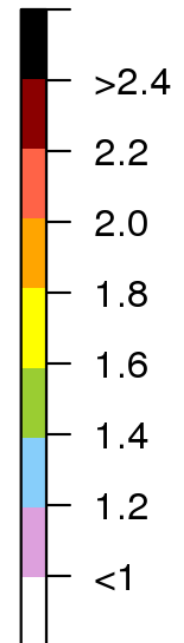
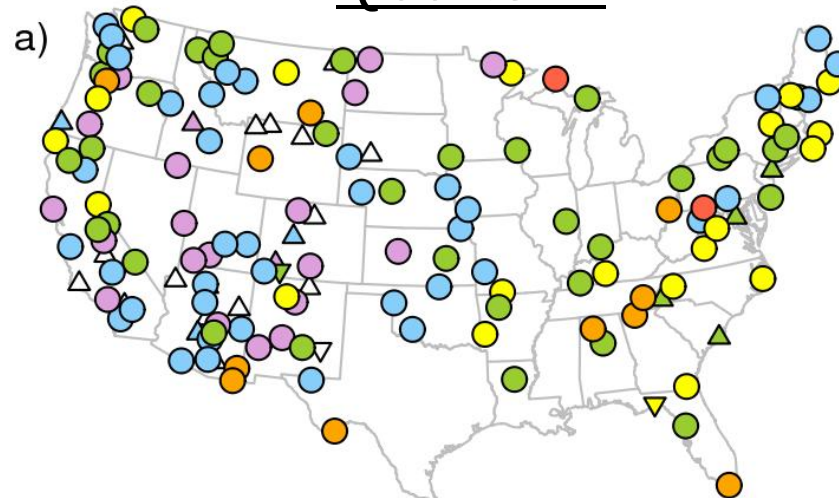
-IMPROVE has assumed OM/OC = 1.4 and 1.8

β_{OC} results

Quarter 3

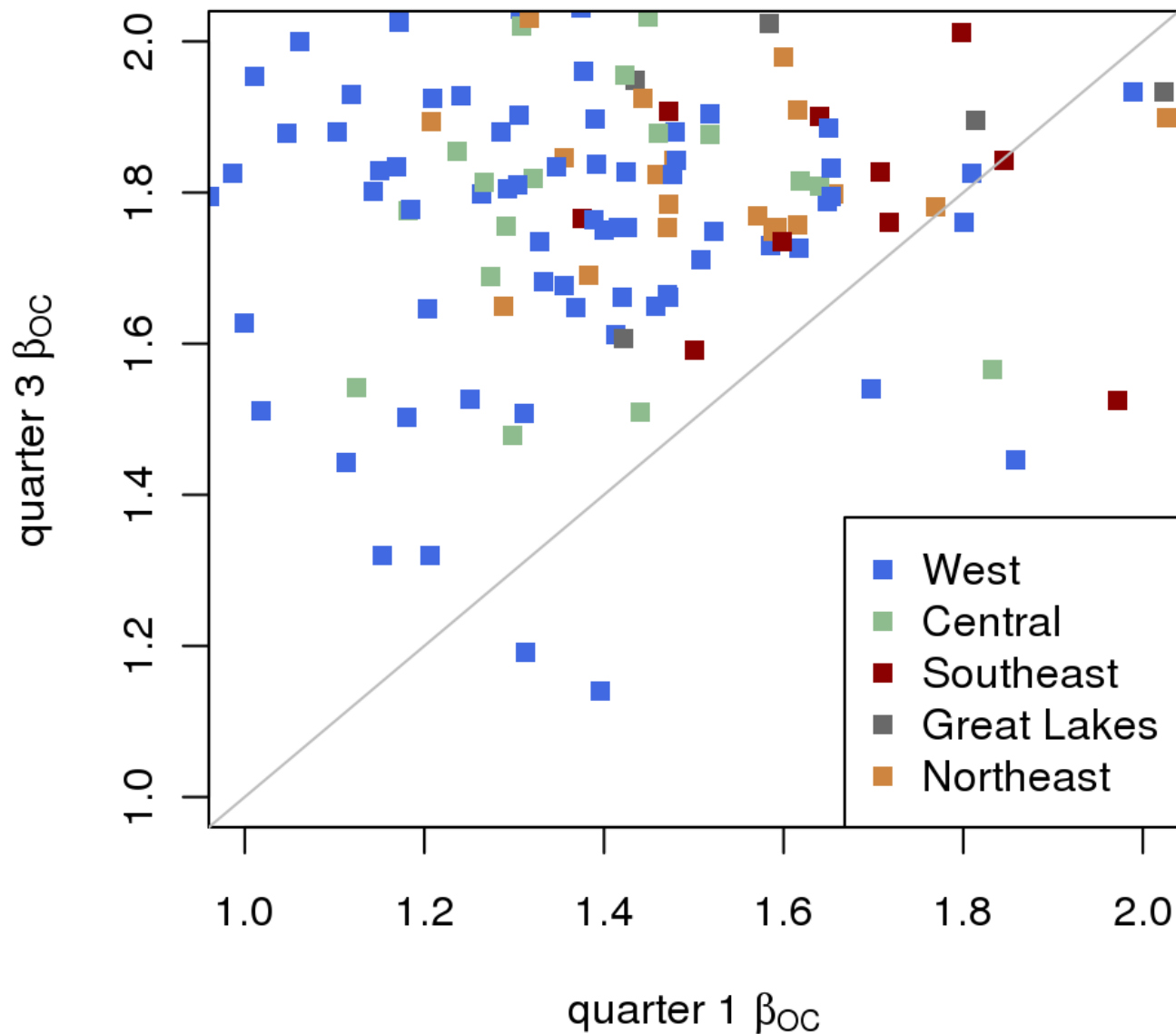


Quarter 1



- Summer (Q3) values are fairly consistent across the US
 - Median values in all regions between 1.80 and 1.87
- Winter (Q1): OM/OC higher in East than in west.
Median values:
 - West: 1.32
 - Central: 1.29
 - Southeast: 1.51
 - Great Lakes: 1.58
 - Northeast: 1.64
- Summer values higher than winter values

β_{OC} results



Conclusions

- We have created a comprehensive dataset of OM/OC covering a large spatial and temporal extent
- Key changes from previous work:
 - Account for measurement uncertainty
 - Eliminate EC as an explanatory variable
- Regression analysis estimates reasonable spatial and temporal trends in β_{OC} , β_{sulf} , β_{nit} , and β_{soil}
- Future analysis will compare these results to other OM/OC determinations

Implications for mass closure at IMPROVE sites

- Mass closure assumptions supported by this work:
 - OM/OC
 - We find summertime inter-quartile OM/OC = 1.72 – 1.91
 - IMPROVE OM/OC = 1.8 is consistent with this finding
 - β_{soil} Values close to 1 support IMPROVE soil equation
- Mass closure assumptions not supported by this work:
 - We find wintertime inter-quartile OM/OC = 1.20 - 1.58
 - Lower OM/OC values could be applied to IMPROVE OC measurements taken during winter months
 - Sulfate neutralization varies seasonally: characterizing ammonium concentrations at IMPROVE sites could improve our ability to achieve mass closure
 - During summer months, sulfate appears to be hydrated.
 - Should laboratory RH be routinely reported?
 - Nitrate volatilization varies widely between sites and seasons.
 - Could shipping samples cold help?