

**Determining the  
spatial and seasonal  
variability in OM/OC  
ratios across the US**

H. Simon et al.

# Determining the spatial and seasonal variability in OM/OC ratios across the US using multiple regression

H. Simon<sup>1</sup>, P. V. Bhave<sup>1</sup>, J. L. Swall<sup>1</sup>, N. H. Frank<sup>2</sup>, and W. C. Malm<sup>3</sup>

<sup>1</sup>US EPA, National Exposure Research Laboratory, Atmospheric Modeling and Analysis Division, Research Triangle Park, NC, USA

<sup>2</sup>US EPA, Office of Air Quality Planning and Standards, Research Triangle Park, NC, USA

<sup>3</sup>National Park Service, Colorado State University/Cooperative Institute for Research in the Atmosphere, Fort Collins, CO, USA

Received: 28 September 2010 – Accepted: 4 October 2010 – Published: 21 October 2010

Correspondence to: H. Simon (simon.heather@epa.gov)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

## Abstract

Data from the Interagency Monitoring of Protected Visual Environments (IMPROVE) network are used to estimate organic mass to organic carbon (OM/OC) ratios across the United States by extending previously published multiple regression techniques.

5 Our new methodology addresses common pitfalls of multiple regression including measurement uncertainty, colinearity of covariates, and dataset selection. As expected, summertime OM/OC ratios are larger than wintertime values across the US with all regional median OM/OC values tightly confined between 1.8 and 1.95. Further, we find that OM/OC ratios during the winter are distinctly larger in the eastern US than in  
10 the West (regional medians are 1.58, 1.64, and 1.85 in the great lakes, southeast, and northeast regions, versus 1.29 and 1.32 in the western and central states). We find less spatial variability in long-term averaged OM/OC ratios across the US (90% of our multiyear regressions predicted OM/OC ratios between 1.37 and 1.94) than previous studies (90% of OM/OC estimates from a previous regression study fell between 1.30 and 2.10). We attribute this difference largely to the inclusion of EC as a covariate in previous regression studies. Due to the colinearity of EC and OC, we believe that up to one-quarter of the OM/OC estimates in a previous study are biased low. In addition to estimating OM/OC ratios, our technique reveals trends that may be contrasted with conventional assumptions regarding nitrate, sulfate, and soil across the IMPROVE network. For example, our regressions show pronounced seasonal and spatial variability in both nitrate volatilization and sulfate neutralization and hydration.

## 1 Introduction

Atmospheric measurements have shown that organic mass (OM) is a major component of fine particulate matter (PM<sub>2.5</sub>), comprising over 50% of ambient PM<sub>2.5</sub> in some  
25 locations (Jimenez et al., 2009; Murphy et al., 2006; Zhang et al., 2007). OM can be divided broadly into two components: organic carbon (OC), and all other mass which

## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



**Determining the spatial and seasonal variability in OM/OC ratios across the US**

H. Simon et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

we will hereafter refer to as non-carbon organic mass (NCOM). NCOM is the largest component of ambient  $PM_{2.5}$  that is not routinely measured. To achieve mass closure in source testing and ambient aerosol measurements, an OM/OC ratio (denoted as  $k$  in some earlier literature) is often multiplied by measured OC to estimate total OM. This ratio is primarily affected by the oxygen content in the organic aerosol (Pang et al., 2006), although hydrogen, nitrogen, and sulfur also make small contributions to the NCOM.

The first estimate of OM/OC was made by White and Roberts (1977) who calculated an average ratio of 1.4 for specific organic compounds measured in Los Angeles. This value was used widely until Turpin and Lim (2001) analyzed a larger dataset to show that OM/OC is generally higher than 1.4. In recent years a range of techniques have been applied to quantify OM/OC including gas chromatography/mass spectroscopy (GC/MS) (Turpin and Lim, 2001; Yu et al., 2005), high resolution time of flight aerosol mass spectrometry (HR-ToF-AMS) (Aiken et al., 2008; Chan et al., 2010; Sun et al., 2009), Fourier Transform Infrared (FTIR) spectroscopy (Gilardoni et al., 2007; Kiss et al., 2002; Liu et al., 2009; Reff et al., 2007; Russell, 2003; Russell et al., 2009), sequential extraction followed by gravimetric weighing and thermal optical measurement of carbon (El-Zanan et al., 2005, 2009), and coupled thermal gravimetric and chemical analyses (Chen and Yu, 2007). Those studies have contributed substantially to our understanding of NCOM in many laboratory and field settings, but none of the techniques have been applied over a broad temporal and spatial range.

Two large US monitoring networks measure a range of  $PM_{2.5}$  constituents, including OC but not OM. A technique for computing OM from these networks could yield a comprehensive dataset of OM/OC ratios covering a large spatial and temporal extent. Frank (2006) developed the SANDWICH method to estimate OM from measurements across the urban-centric Chemical Speciation Network (CSN). He calculated total OM as  $PM_{2.5}$  minus the sum of other components (sulfate, nitrate, ammonium, water, crustal material, and elemental carbon – EC), while making adjustments for particle-bound water (not measured directly) and nitrate volatilization. Unfortunately,

the uncertainty in OC data collected at CSN sites prior to some major network changes in 2008 is comparable to the uncertainty in OM/OC ratios (Watson et al., 2008). Therefore, although the SANDWICH technique is useful for estimating total OM, CSN data are not yet adequate for estimating OM/OC over large multiyear periods.

5 The Interagency Monitoring of Protected Visual Environments (IMPROVE) network tracks visibility degradation in national parks and wilderness areas via routine measurements of PM<sub>2.5</sub> mass and composition (Malm et al., 1994). The network began with 36 monitoring sites in 1988, and currently reports data from 178 remote and 13 urban sites across the continental US, Hawaii, Alaska and the Virgin Islands  
10 (<http://vista.cira.colostate.edu/improve/Data/IMPROVE/AsciiData.aspx>). PM<sub>2.5</sub> is collected on filters for a 24-h period (midnight to midnight) every third day. The filters are subjected to a gravimetric analysis that measures total mass and various chemical analyses that measure bulk composition. In addition to these direct measurements, the network reports a reconstructed fine mass (RCFM) concentration which is  
15 a weighted sum of selected chemical constituents. RCFM was first calculated using Eqs. (1) and (2) (Malm et al., 1994).

$$\text{RCFM} = (\text{NH}_4)_2 \text{SO}_4 + \text{SOIL} + \text{EC} + \text{OM} \quad (1)$$

$$\text{SOIL} = 2.20 \text{ Al} + 2.49 \text{ Si} + 1.63 \text{ Ca} + 2.42 \text{ Fe} + 1.94 \text{ Ti} \quad (2)$$

Ammonium sulfate ((NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>) was calculated as 4.125 × S (sulfur was measured by  
20 Particle Induced X-ray Emission – PIXE – until 2002), SOIL was calculated with Eq. (2) (assuming the soil in PM<sub>2.5</sub> samples mimics the average composition of sedimentary rock), and OM was calculated as 1.4 × OC. Our notation differs slightly from the original publication (Malm et al., 1994). Changes to the RCFM equation since 1994 include the  
25 addition of more components (ammonium nitrate (NH<sub>4</sub>NO<sub>3</sub>), non-soil potassium, and sea salt), modification of Eq. (2) to eliminate Al, and an increase of OM/OC from 1.4 to 1.8 (McDade, 2008).

Although a network-wide OM/OC ratio is commonly used to compute RCFM, a few studies have estimated site-specific OM/OC ratios from IMPROVE data. El-Zanan et

**Determining the spatial and seasonal variability in OM/OC ratios across the US**

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



al. (2005) describe a mass closure technique for calculating OM/OC,

$$\frac{OM}{OC} = \frac{PM_{2.5} - ((NH_4)_2 SO_4 + NH_4 NO_3 + EC + SOIL + Other)}{OC} \quad (3)$$

in which “Other” is the sum of sodium, chlorine, and trace elements measured by x-ray fluorescence (XRF) that are not associated with soil (Lowenthal and Kumar, 2003).

Unfortunately, there are many uncertainties associated with a mass closure analysis of IMPROVE data. First, assumptions must be made about two unmeasured PM<sub>2.5</sub> components: ammonium and particle-bound water. Since ammonium is not routinely measured at IMPROVE sites, sulfate and nitrate are commonly assumed to be fully neutralized by ammonium. Estimation of water mass is complicated by the fact that filter samples are shipped at ambient conditions and weighed in a laboratory where relative humidity (RH) is not controlled. Second, nitrate measurements are made from particles collected on nylon filters, to which nitrate adheres well, whereas PM<sub>2.5</sub> weights are determined from Teflon filters, from which nitrate is known to volatilize (Hering and Cass, 1999). The amount of volatilization from the Teflon filter depends on which cation the nitrate is bound to as well as the temperature and RH during sampling, shipping, and analysis. Third, the IMPROVE soil equation relies on assumptions about the abundance and oxidation states of various trace elements. Since soil composition is spatially heterogeneous, this equation may not accurately estimate the soil contribution in all regions. Finally, OC measurement artifacts contribute additional uncertainty because OC is measured from quartz filters while OM is derived from gravimetric measurements on Teflon filters. Differing tendencies among these two filter materials at retaining OM and/or adsorbing semi-volatile organic gases may affect OM/OC estimates.

Malm and collaborators recently developed a multiple regression technique to estimate OM/OC from 1988–2003 IMPROVE data (Hand and Malm, 2006; Malm et al., 2005; Malm and Hand, 2007). They fit six coefficients in Eq. (4) using ordinary least squares (OLS) regression at each monitoring site. Some notation in Eq. (4) has been changed from that of Malm and Hand (2007) for consistency with the present study.

## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



$$\text{PM}_{2.5,i} = \beta_0 + \beta_{\text{OC}} \text{OC}_i + \beta_{\text{sulf}} (\text{NH}_4)_2 \text{SO}_{4,i} + \beta_{\text{nit}} \text{NH}_4 \text{NO}_{3,i} + \beta_{\text{soil}} \text{SOIL}_i + \beta_{\text{EC}} \text{EC}_i + \beta_{\text{seasalt}} \times 1.8 \text{Cl}_i^- + \varepsilon_i \quad (4)$$

The subscript,  $i$ , represents a daily sample and  $\beta_0$  represents a site-specific intercept. The remaining  $\beta$  coefficients represent ratios of the mass associated with a given  $\text{PM}_{2.5}$  component on the Teflon filter when it was weighed to the mass of that same component determined (or estimated) via chemical analysis of a (possibly) separate filter. The residual error ( $\varepsilon_i$ ) denotes the difference between the measured  $\text{PM}_{2.5}$  mass and the estimated mass (based on fitted coefficients and measured chemical components) for a particular daily sample. The coefficient of most interest to us is  $\beta_{\text{OC}}$  because it represents OM/OC. This technique circumvents many of the assumptions needed for mass closure. For example,  $\beta_{\text{OC}}$  is insensitive to the degree of sulfate neutralization since the relative abundance of ammonium would mainly affect  $\beta_{\text{sulf}}$ . However, OC measurement artifacts can certainly introduce bias in  $\beta_{\text{OC}}$ .

In this paper we develop a nationwide dataset of seasonally- and spatially-varying OM/OC ratios across the IMPROVE network by extending the methodology of Malm and Hand (2007) while addressing some common pitfalls in multiple regression. We discuss quantitative insights regarding the measurement artifacts associated with  $\text{PM}_{2.5}$  components other than OC (i.e. nitrate volatilization and water associated with particulate sulfate), which are ancillary benefits of our methodology. Finally, spatial and temporal trends in OM/OC are reported and examined.

## 2 Methodology

Figure 1 shows a schematic of our methodology, with complete details provided in this section.

### Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## 2.1 General equation and dataset selection

We begin with three minor modifications to Eq. (4) from Malm and collaborators (Malm and Hand, 2007; Hand and Malm, 2006; Malm et al., 2005). First, we eliminate the intercept term ( $\beta_0$ ) and reduce the number of explanatory variables (i.e., covariates) to four that constitute the majority of  $\text{PM}_{2.5}$  and have large uncertainty in their coefficient: OC,  $(\text{NH}_4)_2\text{SO}_4$ ,  $\text{NH}_4\text{NO}_3$ , and SOIL (Eq. 5).

$$\text{PM}_{2.5,i} = \beta_{\text{OC}} \text{OC}_i + \beta_{\text{sulf}} (\text{NH}_4)_2 \text{SO}_{4,i} + \beta_{\text{nit}} \text{NH}_4 \text{NO}_{3,i} + \beta_{\text{soil}} \text{SOIL}_i + \text{EC}_i + 1.8 \times \text{Cl}_i^- + 1.2 \times \text{KNON}_i + \varepsilon_i \quad (5)$$

$$\text{KNON} = K - 0.6 \times \text{Fe} \quad (6)$$

$$\text{SOIL} = 3.48 \times \text{Si} + 1.63 \times \text{Ca} + 2.42 \times \text{Fe} + 1.94 \times \text{Ti} \quad (7)$$

In contrast to Eq. (4), we assume that EC has no artifact and set its coefficient to 1 because treating EC as a separate explanatory variable can bias  $\beta_{\text{OC}}$  (see Sect. 3.3 and Supplement Sect. S3). Similar to Eq. (4), we estimate sea salt as  $1.8 \times \text{Cl}^-$  (Pitchford et al., 2007; White, 2008) but do not treat it as an explanatory variable. Its coefficient should not greatly affect the fit of the regression because it is a relatively minor component at most monitoring sites. Second, we add KNON to Eq. (5) for consistency with the newest IMPROVE RCFM formula (McDade, 2008). KNON represents non-soil potassium (e.g., from wood burning) and is calculated using Eq. (6). The KNON coefficient is fixed at 1.2, the molar mass ratio of potassium oxide to potassium. Although KNON is influenced by soil composition (i.e., soil K/Fe ratio may deviate from 0.6), it contributes a small enough mass to total  $\text{PM}_{2.5}$  that fixing its coefficient should not adversely affect the regression as a whole. Third, we use an updated IMPROVE soil equation (compare Eqs. 2 and 7) which eliminates aluminum from the calculation because Al is not reliably measured by the IMPROVE XRF analysis (McDade, 2008).

We downloaded the IMPROVE data from <http://views.cira.colostate.edu/web/DataWizard/> on 6 January 2010, and analyzed the measurements collected at 186 continental US sites between 1 January 2002 and 31 December 2008. All analysis was

### Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



performed using the R statistical software package (R Development Core Team, 2010). Like Malm and Hand (2007), we segregate the data by monitoring site. In addition, we segregate data by season: quarter 1 (Jan, Feb, Mar), quarter 2 (Apr, May, Jun), quarter 3 (Jul, Aug, Sep), and quarter 4 (Oct, Nov, Dec), because we expect the coefficients (i.e., OM/OC and nitrate volatilization) to vary seasonally. However, we could not justify the seasonal variability in soil coefficients estimated from our initial analyses. For instance, the variability in  $\beta_{\text{soil}}$  was not correlated to Asian dust plumes or other seasonally varying dust sources. We therefore hold the soil coefficient constant throughout the year by first performing a multiyear regression at each site using all data from 2002–2008 and then fixing  $\beta_{\text{soil}}$  in each quarter-specific regression to the  $\beta_{\text{soil}}$  value obtained from the multiyear regression at that given site.

Within site and quarter-specific datasets, the only data filter that we apply is completeness. If a major component in Eq. (5) (i.e.,  $\text{PM}_{2.5}$ , OC, S,  $\text{NO}_3^-$ , Si, Ca, Fe, Ti, or EC) is missing from a single site and sample, we eliminate the whole date from that site. Missing data values for  $\text{Cl}^-$  and K are set to 0. All concentrations reported as negative values are left as is. Finally, sites that do not have an average of at least 15 days of complete data per quarter (i.e., 105 samples for each quarter over the 7 year measurement period) for all four quarters are eliminated from the analysis. This criterion eliminates thirty-three sites. As shown in Fig. 1, we perform one multiyear and four quarter-specific regressions for each of the remaining 153 monitoring sites (i.e., 765 separate regressions).

## 2.2 Physical interpretation of coefficients

When interpreting the coefficients in Eq. (5), it is important to note that all results may be affected by changes in measurement techniques and variability in the ambient conditions. Therefore, readers are cautioned against over-interpreting results from a single regression and instead are encouraged to use these results to understand spatial and temporal trends in the coefficients. For each  $\text{PM}_{2.5}$  component, the regression coefficient represents the ratio of retained mass associated with that component on the



Teflon filter (used for gravimetric  $PM_{2.5}$  analysis) to the mass of that component derived from chemical analysis. Here we describe how values different than 1 may be interpreted and set bounds on physically reasonable values for each coefficient.

The OC coefficient,  $\beta_{OC}$ , should represent the OM/OC ratio. We set its lower bound on this value to 1, representing pure graphitic carbon with no associated hydrogen, oxygen, or nitrogen mass. We set the upper bound to 5, which would occur for a carbonate ion ( $CO_3^{2-}$ ). It is possible to have a higher OM/OC for some organic sulfates, but it is unlikely that these compounds would contribute enough mass to raise the overall OM/OC above 5. Typical OM/OC ratios for primary organic emissions are around 1.25 in vehicle exhaust and 1.7 in wood smoke emissions (Reff et al., 2009). Measurements of OM/OC from laboratory-generated secondary organic aerosol (SOA) range from 1.4–2.7 (Kleindienst et al., 2007). Although we interpret  $\beta_{OC}$  as equivalent to OM/OC, this coefficient may also be skewed by two types of OC measurement artifact: negative artifacts occur when organic aerosol collected on the filter volatilizes before chemical analysis and positive artifacts occur when organic vapors adsorb to the filter surface (McDow and Huntzicker, 1990; Turpin et al., 1994).  $\beta_{OC}$  will be influenced further by differences in the sampling artifact on quartz filters (used to measure OC) versus Teflon filters.

A soil coefficient not equal to 1 could represent soil compositions differing from the average sediment used to develop Eqs. (2) and (7).  $\beta_{soil}$  represents the actual soil mass in the  $PM_{2.5}$  sample divided by the soil mass calculated from Eq. (7). Simon et al. (2010) report that this ratio can range from 0.41 to 1.63 based on soil compositions in the literature, so these bounds are applied to  $\beta_{soil}$ .

A sulfate coefficient below unity would indicate that the assumption of dry ammonium sulfate over-estimates total sulfate mass. Incomplete neutralization could cause such an over-estimate. The molar mass of ammonium bisulfate ( $NH_4HSO_4$ ) and sulfuric acid ( $H_2SO_4$ ) are 87% and 74% of the  $(NH_4)_2SO_4$  molar mass. Therefore, 0.74 would seem like a reasonable lower bound for  $\beta_{sulf}$ . However, the sulfate mass in our regression is calculated from an XRF sulfur measurement which can detect organo-sulfur atoms. A

## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

**Determining the spatial and seasonal variability in OM/OC ratios across the US**

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

conservative lower bound could be calculated assuming that all carbonaceous mass associated with organic molecules would be included in the  $\beta_{OC}$ . Surratt et al. (2008) report that up to 20% of sulfur may be contained in these organic compounds, so we set a lower bound for  $\beta_{sulf}$  at 0.59 ( $0.74 \times 0.8$ ) to capture an admittedly extreme scenario in which all inorganic sulfate is in the form of sulfuric acid and 20% of the total sulfur is contained in organic compounds. A sulfate coefficient above 1 would indicate that there is extra mass associated with the particulate sulfate. This extra mass could come from water if the aerosol remains hydrated during gravimetric analysis. During the history of the IMPROVE network, RH in the gravimetric measurement laboratory was only recorded intermittently. We obtained laboratory measurements of RH during the gravimetric analysis of filters collected from September 2003 to May 2005 and from May to December of 2008 (C. McDade, personal communication, 2009). The maximum reasonable  $\beta_{sulf}$  is estimated from the 99th percentile of those measurements (i.e., 52% RH). At this humidity, the AIM model (Wexler and Clegg, 2002) (available at <http://www.aim.env.uea.ac.uk/aim/aim.php>) computes hydrated  $(NH_4)_2SO_4$  to have 53% more mass than dry  $(NH_4)_2SO_4$  and hydrated  $NH_4HSO_4$  to have 32% more mass than dry  $NH_4HSO_4$ . Therefore, 1.53 is a reasonable upper bound for  $\beta_{sulf}$ .

Nitrate coefficients less than 1 likely represent volatilization of  $NH_4NO_3$  from the Teflon filter prior to gravimetric analysis. Because a value of 0 (complete nitrate volatilization) would imply no relationship between nitrate mass and  $PM_{2.5}$  mass, a slightly negative  $\beta_{nit}$  value caused by measurement error is just as likely as a slightly positive  $\beta_{nit}$  value. Consequently, for each regression performed, we set the lowest reasonable value for  $\beta_{nit}$  as 1.5 standard errors below 0 (calculation of standard errors is described in the Supplement, Sect. S1). There are 93 site/quarter groupings exhibiting negative  $\beta_{nit}$  values within 1.5 standard errors of 0. To show that these negative values really represent slight variations around 0, we repeated each of these regressions without the nitrate term and found that  $\beta_{OC}$  and  $\beta_{sulf}$  coefficients changed by less than 3% on average (no  $\beta_{OC}$  and six  $\beta_{sulf}$  coefficients changed by more than 5%). A nitrate coefficient greater than 1 indicates that the assumption of dry  $NH_4NO_3$  underestimates

nitrate mass on the Teflon filter. This would occur either if the cation had a larger molar mass than ammonium (e.g. Na) or if there were water associated with the nitrate during weighing. Again a maximum reasonable value for  $\beta_{\text{nit}}$  was determined by computing increases in water mass at 52% RH with the AIM model for both  $\text{NH}_4\text{NO}_3$  and  $\text{NaNO}_3$ .

5 This analysis shows that 35% extra mass can be added to the nitrate, so 1.35 is a reasonable upper bound for  $\beta_{\text{nit}}$ .

### 2.3 Effects of measurement uncertainty

Despite the advantages of regression described in Sect. 1, several pitfalls can occur with ordinary least squares (OLS) multiple regression analysis. One such pitfall is that OLS regression assumes that explanatory variables are measured without error. This assumption conflicts with the reality of our application, in which there is measurement uncertainty associated with all explanatory variables: OC,  $(\text{NH}_4)_2\text{SO}_4$ ,  $\text{NH}_4\text{NO}_3$ , and SOIL. Measurement uncertainty in the explanatory variables can introduce bias in the estimates of regression coefficients. For regressions with a single explanatory variable, the coefficient is biased towards zero when the explanatory variable measurement is uncertain (Fuller, 1987; Saylor et al., 2006; White, 1998). With multiple explanatory variables, bias in the various coefficients exhibits a more complex dependency on factors such as the relative uncertainties in various components, the correlation between explanatory variables, and the correlation between measurement errors. White (1998) examined this problem in a simplified case with two independent explanatory variables for which neither the measured values nor the measurement uncertainties were correlated. For that case, he showed that the coefficient for the explanatory variable with lower uncertainty was artificially inflated while the one with higher uncertainty was diminished.

25 To evaluate this bias within the more complex conditions of the present study, we analyzed synthetic datasets that mimic our data. Assuming that the actual values for each measurement were exactly equal to the reported value, we created 200 synthetic datasets for each site- and quarter-specific dataset that represent “observed” data with

## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



error in the explanatory variables. Errors were added by perturbing the reported values of OC, sulfate, nitrate, and  $PM_{2.5}$  using the reported uncertainty and assuming that “observed” values would be normally distributed around the actual value. For each site- and quarter-specific dataset, we then performed an OLS regression on the reported dataset and the 200 synthetic datasets. The reported dataset is considered the “truth” in this exercise, so OLS regression yields “true” coefficients for comparison with the results from our synthetic datasets. Results from one such analysis for a regression with typical OLS biases (Gila Wilderness in New Mexico during quarter 1) are shown in the left half of each plot in Fig. 2. The dotted lines represent the “true” coefficients and the box plot shows the distribution of values that would be obtained from the 200 synthetic datasets. Although the true value could be accurately estimated from some synthetic datasets,  $\beta_{OC}$  is typically under-estimated while  $\beta_{sulf}$  and  $\beta_{nit}$  are over-estimated in this example.

To overcome the biases associated with the OLS assumption of error-free explanatory variables, a class of methods has been developed to explicitly account for the existence of such errors; these are often collectively called measurement error models or errors-in-variables (EiV) models. Such methods typically assume that for all observations of each covariate, the errors are independent, identically distributed and follow a normal distribution with mean zero and a fixed (possibly unknown) standard deviation. In the IMPROVE data, the standard deviation is not fixed because we have a different estimated error associated with each observation of a given covariate, which we take as the standard deviation of the error distribution. To accommodate this added complexity, we turn to an advanced measurement error model described by Fuller (1987) (Sect. 3.1.2). The following discussion is based entirely on Fuller’s work, conforming to his original notation as much as is feasible.

To begin, we define  $Y_t$  as the value of the response variable for observation  $t$ , such that  $t = 1, 2, \dots, n$ , with  $n$  representing the number of observations. For the multiyear regression, this response is given by  $PM_{2.5} - (1.2 \times KNON + 1.8 \times CI^- + EC)$ , and for the quarter-specific regression it is

## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

PM<sub>2.5</sub> - (1.2 × KNON + 1.8 × Cl<sup>-</sup> + EC + β<sub>soil</sub> × SOIL). The row vector  $X_t$  contains the observed values of the explanatory variables associated with observation  $t$ . The first element is the observed value of OC, the next element corresponds to (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, the third is that for NH<sub>4</sub>NO<sub>3</sub>, and the fourth is the observation for SOIL. (In the quarter-specific regression case, the SOIL component is omitted.) Note that the order of these explanatory variables mimics the order they take in Eq. (5) and is preserved in the various mathematical representations of their coefficients, errors, etc. which follow.

Additionally, we let  $\sum_{uutt}$  represent the covariance matrix associated with the covariates for observation  $t$ . Since we are assuming that errors are independent of each other, this is a diagonal matrix. The elements along the diagonal contain the variance (square of the error standard deviation) associated with the explanatory variables, in the specified order. As an initial estimate for the regression coefficients, we use the method-of-moments estimator, the column vector  $\tilde{\beta}$ , given by Eq. (8)

$$\tilde{\beta} = \left[ n^{-1} \sum_{t=1}^n (X_t' X_t - \sum_{uutt}) \right]^{-1} \left[ n^{-1} \sum_{t=1}^n X_t' Y_t \right] \quad (8)$$

Having obtained this initial estimate, we work to refine it, as outlined by Fuller (1987). We define for each observation  $t$  the matrix  $\sum_{aatt}$ . This is also a diagonal matrix, with the elements along the diagonal consisting of the variance for the response followed by the variances for the explanatory variables in the specified order. We take the square of the reported measurement uncertainty for each chemical constituent in a particular sample as its variance. (Note that the  $\sum_{uutt}$  featured in Eq. 8 is simply a submatrix of  $\sum_{aatt}$ .) We also let  $Z_t$  represent the row vector containing the observed response and the observed explanatory variables for each  $t$ ; i.e.,  $Z_t = (Y_t, X_t)$ . We then define the matrices  $\mathbf{M}$  and  $\mathbf{A}$  as

$$\mathbf{M} = \sum_{t=1}^n \sum_{aatt} \text{ and } \mathbf{A} = \sum_{t=1}^n Z_t' Z_t$$

## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



With these defined, we can now obtain an estimate of the variance associated with the regression error, denoted  $\sigma_{qq}$ . We first solve for the eigenvalues of the matrix product  $\mathbf{M}^{-1}\mathbf{A}$ . If the minimum of these eigenvalues is less than one, then  $\tilde{\sigma}_{qq}$  is 0. Otherwise,  $\tilde{\sigma}_{qq}$  is given by Eq. (9):

$$\tilde{\sigma}_{qq} = \sum_{t=1}^n \left[ (n-k)^{-1} (Y_t - \mathbf{X}_t \tilde{\boldsymbol{\beta}})^2 - n^{-1} (1, -\tilde{\boldsymbol{\beta}}') \sum_{\text{aatt}} (1, -\tilde{\boldsymbol{\beta}}')' \right] \quad (9)$$

Both  $\tilde{\boldsymbol{\beta}}$  and  $\tilde{\sigma}_{qq}$  are then used to obtain an estimate of the error associated with the linear relationship between the observed (with error) response and covariates,  $\tilde{\sigma}_{vtt}$  (Eq. 10):

$$\tilde{\sigma}_{vtt} = \tilde{\sigma}_{qq} + \sigma_{wvtt} + \tilde{\boldsymbol{\beta}}' \sum_{\text{uutt}} \tilde{\boldsymbol{\beta}} \quad (10)$$

where  $\sigma_{wvtt}$  is the variance of the measurement error associated with the response at time  $t$ . To obtain our final estimate,  $\hat{\boldsymbol{\beta}}$ , of the regression coefficients, we combine the previous elements to obtain Eq. (11):

$$\hat{\boldsymbol{\beta}} = \left[ \sum_{t=1}^n \tilde{\sigma}_{vtt}^{-1} (\mathbf{X}'_t \mathbf{X}_t - \sum_{\text{uutt}}) \right]^{-1} \sum_{t=1}^n \tilde{\sigma}_{vtt}^{-1} \mathbf{X}'_t Y_t \quad (11)$$

Here  $\hat{\boldsymbol{\beta}}$  is a column vector containing our estimates of  $\beta_{OC}$ ,  $\beta_{\text{sulf}}$ ,  $\beta_{\text{nit}}$ , and  $\beta_{\text{soil}}$  (for the multiyear regression). Fuller (1987) also provides an estimator for the covariance matrix  $\hat{\boldsymbol{\beta}}$ . We use the diagonal elements of this matrix to obtain the standard errors for our estimated regression coefficients. In the interest of brevity, we leave further discussion of this variance estimate to the Supplement (Sect. S1). In addition, sample R code used to perform these regressions is also supplied in Supplement Sect. S1.

We recognize that our method includes several assumptions. Perhaps most notable is the assumption that the measurement errors are independent among all the covariates and the response measured at a given date and location. The method could be

**Determining the spatial and seasonal variability in OM/OC ratios across the US**

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



extended to include information about the correlation between measurement errors, if such were known. This would result in non-diagonal matrices  $\sum_{uutt}$  and  $\sum_{aatt}$ . Another key assumption is that the measurement error distributions are normal. If we did not believe this to be a reasonable assumption, we would have to explore more complex statistical models which allow for nonnormal measurement errors, which are currently a subject of statistical research.

To demonstrate that this new technique improves the bias in coefficients, we repeat our previous analysis of the synthetic datasets using the EiV regression methodology. The results for quarter 1 data from Gila Wilderness are shown in the right-hand box plots of Fig. 2. Clearly, the EiV method yields coefficients that are much closer to the “truth” than the OLS methodology. To confirm this result at other sites and quarters, Fig. 3 shows the distribution of bias across all site- and quarter-specific datasets. Substantial bias in  $\beta_{OC}$  (under-prediction),  $\beta_{sulf}$  (over-prediction), and  $\beta_{nit}$  (over-prediction) are seen for the OLS regression, but these biases are greatly mitigated with the EiV technique. White (1986) provides a similar analysis of regression performance using measurements from the 1981–1982 Western Regional Air Quality Study. That analysis, which included three explanatory variables (sum of ionic sulfate, nitrate, and ammonium; organic carbon; sum of silicon dioxide and calcium oxide), also found that correcting for measurement uncertainty reduces bias in the coefficients.

Although the EiV methodology shows improved results, it should be noted that an additional source of error arises if the measurement uncertainties are biased themselves. Hyslop and White (2008) report some systematic biases in the measurement uncertainty from XRF, ion chromatography, and TOR carbon measurements at IMPROVE sites. If future updates to the IMPROVE data include substantial changes to uncertainty estimates for these components, it may warrant some repetition of the present work. For all subsequent analyses discussed in this paper, we apply the EiV method (instead of OLS).

## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)



## 2.4 Statistical identification of high-confidence regressions

After applying the EiV method to each multiyear and quarter-specific dataset, it is tempting to begin examining spatial and temporal patterns in the regression coefficients. However, as emphasized by Malm and Hand (2007), “Regression coefficients are vulnerable to a variety of systematic and random errors”. In this subsection, we establish some empirical guidelines for flagging or eliminating datasets that do not conform to Eq. (5). As summarized in the lower half of Fig. 1, these guidelines are subsequently applied to identify regression results that can be used with high confidence for applications such as air quality model evaluations, source-apportionment analyses, epidemiology studies, and radiative calculations.

### 2.4.1 Multicollinearity among explanatory variables

One requirement of our regression method is that all explanatory variables be independent of each other. If any two  $\text{PM}_{2.5}$  components are linearly related, the dataset is not suitable for regression analysis because the technique may over-estimate one coefficient and under-estimate another as a result of having too many degrees of freedom. To identify such datasets, Pearson correlation coefficients ( $r_p$ ) are calculated for all couplings of the four explanatory variables (OC, sulfate, nitrate, and soil) in each site- and quarter-specific dataset. To establish a guideline for the maximum acceptable  $r_p$  in the present application, we examine all datasets which have  $|r_p|$  values greater than 0.65 between any two explanatory variables. Specifically, we look for cases in which the coefficient on one of the highly correlated explanatory variables appears to be over-estimated while the other appears to be under-estimated. For example, sulfate and nitrate from 4th quarter measurements at the Puget Sound monitoring site are correlated with an  $r_p = 0.86$ . In that regression,  $\beta_{\text{sulf}} = 0.83$  (lower end of its physically reasonable range) and  $\beta_{\text{nit}} = 1.28$  (higher end of its range). We acknowledge that this analysis is neither exhaustive nor foolproof since (1) coefficients that appear to be skewed may actually be accurate results, and (2) some regressions that are affected

## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



by co-linearity may not be identifiable if the estimated coefficients fall in the middle of the range of physically reasonable values. However, this analysis provides empirical guidance on acceptable  $r_p$  values. A summary of our analysis is shown in Fig. 4 and a cutoff of maximum  $r_p$  of 0.85 was selected. Seven quarter-specific datasets are eliminated from our analysis based on this criterion (refer to Supplement Table S3).

## 2.4.2 Assessing the fit of the regression model

A second requirement for accurate regressions is that the equation used to fit coefficients is physically realistic. Based on our knowledge of ambient aerosols in the US, we believe Eq. (5) includes all the essential  $PM_{2.5}$  components. However, if the true coefficient for EC, KNON, or  $Cl^-$  is substantially different from our fixed coefficients, the regression could be adversely affected. In addition, if the actual soil coefficient varies greatly throughout the year, then our assumption of temporally-invariant  $\beta_{soil}$  could also degrade the regression results. Finally, if the relationship between  $PM_{2.5}$  mass and any major chemical component is nonlinear, our regression analysis will be inaccurate. For instance, if OC artifact corrections were biased low at high concentrations and vice versa, OC concentrations would be positively biased in clean conditions and negatively biased in polluted conditions so the relationship between reported OC and total  $PM_{2.5}$  would be nonlinear.

To identify cases influenced by one or more of these phenomena, we examined the residual errors ( $\varepsilon_i$  in Eq. 5) for each dataset. Spearman rank order correlation coefficients ( $r_S$ ) are calculated between the residual error ( $\varepsilon_i$ ) values and each  $PM_{2.5}$  component in Eq. (5) (OC, EC, sulfate, nitrate, KNON,  $Cl^-$ , and SOIL). Any high correlation may indicate that Eq. (5) is an inadequate representation of  $PM_{2.5}$  at the given site/quarter. Following this analysis, a criterion of  $|r_S| > 0.4$  is imposed to eliminate 12 quarter-specific datasets that are likely affected by the problems discussed above (refer to Supplement Table S3): 9 for high  $r_S$  values between residual error and chloride, one for high  $r_S$  between residual error and SOIL (Domelands Wilderness, California, quarter 4), one for high  $r_S$  between residual error and KNON (Sawtooth National

## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Forest, Idaho, quarter 4), and one for high  $r_s$  between residual error and both OC and EC (Mount Hood, Oregon, quarter 2). Figure 5 shows four examples of these correlations. Most of the datasets with large  $|r_s|$  values between chloride and residual error contain many negative chloride values. White (2008) reports that negative chloride values in 2002 and 2003 were caused by variability in filter blanks (a change of filter suppliers in 2004 corrected this problem). These sites exemplify how measurement variability can adversely affect regression analysis.

### 2.4.3 Dataset selection

A third key element to obtaining meaningful regression coefficients from IMPROVE measurements is appropriate segregation of data. For this analysis, data are grouped by season and monitoring site with the intention that samples taken within each subset should yield fairly constant regression coefficients. However, sites that are strongly influenced by time-varying sources may not match our intent and therefore may not be ideal input for our regression analysis. For instance, a monitoring site that is impacted intermittently by large wildfires and heavy diesel traffic will exhibit varying OM/OC ratios that would violate our assumption of constant  $\beta$  coefficients by quarter.

To check for temporal trends or irregularities during our 7 year study period, residual error values were binned by year and examined for each site- and quarter-specific dataset. This analysis was designed to identify three possible problems: (1) a one-time abrupt change in  $\varepsilon_i$  which could indicate a change in measurement methods, (2) a monotonic temporal trend in  $\varepsilon_i$  which could indicate changing aerosol characteristics at the site, possibly due to the implementation of regulatory controls on emissions, and (3) a single year which showed vastly different  $\varepsilon_i$  from other years indicating that a distinct and infrequent event (forest fire, abnormal meteorology etc.) affected the monitoring site.

Visual inspection of the data shows no evidence of problem 1 in these datasets. There was a change in EC and OC measurement equipment between 2004 and 2005

## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



(White, 2007) as well as a coincident change in the calibration of the XRF sulfur measurements (White, 2009a). (Advisories about other such changes to IMPROVE data can be found at [http://vista.cira.colostate.edu/improve/Data/QA\\_QC/Advisory.htm](http://vista.cira.colostate.edu/improve/Data/QA_QC/Advisory.htm).) Despite these changes in OC, EC, and sulfur, no shift in residual values is apparent between 2004 and 2005 for the network as a whole. Figure 6 shows  $\varepsilon_j$  by year at Sipsy Wilderness (a site with one of the highest median OC concentrations). The change in  $\varepsilon_j$  between 2004 and 2005 is no greater than other inter-annual variations. Therefore, we do not believe that this measurement change has an observable effect on our analysis although the change in measurement techniques does add some uncertainty to our final results.

Seven site- and quarter-specific datasets show evidence of a temporal trend (i.e., problem 2) in which mean residual values or the inter-quartile range of residual values either increase or decrease monotonically from 2002–2008. Those regressions are retained in our analysis, but not regarded as high-confidence results. An example of one such dataset (Yosemite National Park during quarter 4) is shown in Fig. 7 and all seven are listed in Supplement Table S5. Further investigation of these datasets by people with expert knowledge of those specific sites would be worthwhile.

Finally, sites affected by an infrequent event are identified if they fit one of two criteria: the inter-quartile range of  $\varepsilon_j$  in any year does not overlap with the inter-quartile ranges from any other year; or the year with the broadest inter-quartile range was greater than two times the second broadest inter-quartile range. Examples of these two phenomena are also shown in Fig. 7. These regressions were re-run without the errant year and the results from both the full and abridged datasets and are reported in Table S5 of the Supplement (Sect. S5). In general, removing the outlier year has a modest effect on the regression coefficients: of the 28 cases flagged, only 9  $\beta_{OC}$  values, 4  $\beta_{sulf}$  values, 12  $\beta_{nit}$  values, and 1  $\beta_{soil}$  value change by more than 0.1. We regard the 10 cases in which all four coefficients were relatively unperturbed as high-confidence results. These values are reported in both Supplement Tables S4 and S5. In cases where coefficients changed more substantially (gray shading in Supplement Table S5), the

**Determining the spatial and seasonal variability in OM/OC ratios across the US**

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



outlier year appears to substantially skew the regression results. Again, we believe a detailed investigation of those data by site-specific experts would be worthwhile.

### 3 Results

Table S2 in the Supplement (Sect. S5) shows our multiyear regression results. Supplement Tables S4, S5, and S6 (Sect. S5) show coefficients with standard error values, normalized mean errors (NME), and normalized mean biases (NMB) for all quarter-specific regressions. Supplement Table S4 includes regressions for which we have a high degree of confidence (see Sects. 2.2 and 2.4). Supplement Table S5 includes all regression results for datasets flagged for temporal trends and irregularities in  $\varepsilon_i$  (see Sect. 2.4.3). Supplement Table S6 includes results from all regressions in which one or more coefficients fall outside our physically reasonable ranges (see Sect. 2.2). NME and NMB are calculated using Eqs. (12) and (13). NMB and NME values are generally small (mean NMB for all regressions in Supplement Tables S4, S5, and S6 = -0.2%, maximum absolute NMB = 2.6%, mean NME = 8.5%, maximum NME = 22.6%) indicating that the data fit Eq. (5) quite well.

$$\text{NME} = \left( \frac{\sum_{i=1}^n |\varepsilon_i|}{\sum_{i=1}^n \text{PM}_{2.5,i}} \right) \times 100\% \quad (12)$$

$$\text{NMB} = \left( \frac{\sum_{i=1}^n \varepsilon_i}{\sum_{i=1}^n \text{PM}_{2.5,i}} \right) \times 100\% \quad (13)$$

## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



### 3.1 Physically unreasonable results

Only 7 of the multiyear regressions (< 5%) have a coefficient that is physically unreasonable. Of these, 2 have  $\beta_{\text{soil}}$  values (0.21 and 0.27) falling below those of known soil profiles (see Sect. 2.2). Both low  $\beta_{\text{soil}}$  values come from urban IMPROVE sites (New York City and Washington DC). In these locations, there are likely non-soil sources of Si, Ca, Fe, or Ti. For instance, residential wood combustion is a major source of all four elements, on-road vehicle exhaust is a major source of Si, Ca, and Fe, and surface coating operations are a major source of Ti (Reff et al., 2009). In urban areas where such sources may dominate, Eq. (7) would overestimate total soil mass and might yield an erroneously low value of  $\beta_{\text{soil}}$ . The other 5 problematic multiyear regressions have low  $\beta_{\text{nit}}$  values, for which the cause is unclear. We are nevertheless able to extract high-confidence values of  $\beta_{\text{OC}}$  at these sites by using the multiyear  $\beta_{\text{soil}}$  value in our quarter-specific regressions.

In addition, 61 quarter-specific regressions (10%) have at least one physically unreasonable coefficient. The total number of quarter-specific regressions with problematic coefficients is greatest in quarter 1 ( $n = 21$ ) and quarter 3 ( $n = 22$ ) and least in quarters 2 and 4 ( $n = 13$  and  $n = 5$  respectively). Problematic soil coefficients from the multiyear regressions account for 8 of the 61 problematic quarter-specific regressions (2 in each quarter).

Overall, 20 out of the 61 regressions with physically unrealistic coefficients are due to  $\beta_{\text{OC}}$  values less than unity, 17 of which occur in quarter 1. These low  $\beta_{\text{OC}}$  values may be caused by errors in OC artifact correction. Although the low  $\beta_{\text{OC}}$  values predominantly occur in quarter 1, this may be due to the fact that  $\beta_{\text{OC}}$  values are lower in quarter 1 than in other quarters (median  $\beta_{\text{OC}}$  in quarter 1, 2, 3, and 4 are 1.39, 1.83, 1.81, and 1.59 respectively). Therefore, a slight low bias would push more estimated OM/OC ratios below 1 in the winter than in other seasons. The sensitivity of  $\beta_{\text{OC}}$  values to IMPROVE OC artifact correction techniques is discussed in more detail in Sects. 3.3 and S3 in the Supplement.

## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Eighteen of the 61 problematic regressions are due to negative  $\beta_{\text{nit}}$  values that are more than 1.5 standard errors below zero. Fourteen out of these eighteen bad  $\beta_{\text{nit}}$  values occur in quarter 3. There are two possible explanations for the high number of problematic nitrate results in quarter 3. First, nitrate concentrations are generally low in the summer. In quarter 3, network-wide median nitrate concentrations were only 3% of median  $\text{PM}_{2.5}$  (annual median nitrate concentrations were, on average, 6% of median  $\text{PM}_{2.5}$  concentrations and quarter 1 median nitrate concentrations were, on average, 11% of median  $\text{PM}_{2.5}$  concentrations). When the mass of an explanatory variable is low compared to the mass of other  $\text{PM}_{2.5}$  components, the model fit is not very sensitive to large changes in that coefficient. Second, these problematic nitrate coefficient estimations may be due to the large number of complete nitrate volatilization cases in quarter 3. The lower-bound for negative  $\beta_{\text{nit}}$  values (1.5 standard errors below 0) may be too conservative, and we would therefore be flagging regressions in which nitrate volatilization is 100% (i.e.  $\beta_{\text{nit}}$  is essentially 0) as problematic regressions.

The third largest error comes from high  $\beta_{\text{nit}}$  values: 13 regressions estimate  $\beta_{\text{nit}} > 1.35$ . In general these data points have higher than average standard errors (the mean nitrate standard error for these regressions is 0.50 while the mean nitrate standard error for all site-specific regressions is 0.21). These large standard error values indicate that  $\beta_{\text{nit}}$  estimates are highly uncertain for these regressions. Some less frequent problems include 3 regressions with  $\beta_{\text{sulf}} < 0.59$  and one with  $\beta_{\text{sulf}} > 1.53$ . There are three cases shown in Supplement Tables S5 and S6 for which a regression with problematic coefficient(s) also had an outlier year in residual values, but eliminating the outlier year did not fix the regression coefficients in any of those cases.

### 3.2 Spatial and temporal trends in $\beta_{\text{soil}}$ , $\beta_{\text{sulf}}$ and $\beta_{\text{nit}}$

Figure 8 shows spatial trends in soil coefficients. Much of the country has  $\beta_{\text{soil}}$  values near 1, confirming that the IMPROVE soil equation does a reasonable job of estimating soil concentrations. Some notable departures from this are high values displayed in orange and red in the southwestern US and lower values (green, blue, purple, and



white) in much of the Midwest. Both of these are consistent with the calculated  $\beta_{\text{soil}}$  values for different soil types described by Simon et al. (2010) who report that  $\beta_{\text{soil}}$  values for desert soil range between 1.25 and 1.4 and  $\beta_{\text{soil}}$  values for agricultural soil range between 0.78 and 1.10.

5 In order to evaluate spatial and temporal trends for  $\beta_{\text{sulf}}$  and  $\beta_{\text{nit}}$ , multiyear and quarter-specific regression results are grouped by region, matching the regional planning organizations designated by the EPA to address regional haze (EPA, 2010). From this point forward, states included in WRAP, CENRAP, LADCO, MANE-VU, and VIS-TAS will be referred to as the western region, the central region, the great lakes region,  
10 the northeastern region, and the southeast region respectively.

Maps of  $\beta_{\text{sulf}}$  during each quarter are given in the Supplement (Figs. S6–S9). Figure 9 shows a summary of  $\beta_{\text{sulf}}$  values for all quarters and regions. This figure includes data from all regressions reported in Supplement Tables S4, S5, and S6. Apart from in the western region,  $\beta_{\text{sulf}}$  follows a seasonal trend in which values are lowest in the winter (median values in the central region, the southeast region, the great lakes region,  
15 and the northeast region are 0.90, 0.92, 0.91, and 0.88 respectively) and highest in the summer (median values in the central region, the southeast region, the great lakes region, and the northeast region are 1.05, 1.04, 1.09, and 1.09 respectively). The median wintertime values less than 1 suggest that sulfate is dry and not-fully neutralized  
20 in quarter 1. The summertime values greater than 1 suggest wet sulfate. Further analysis presented in the Supplement suggests that the trends predicted by the regression analysis can indeed be reasonably explained by the seasonal variation in laboratory RH where samples were weighted and in the degree of sulfate neutralization.

Quarter-specific maps of all  $\beta_{\text{nit}}$  are given in the Supplement (Figs. S10–S13). Figure 10 depicts several temporal and spatial trends. The regions which have the most dramatic seasonal temperature variations (the central region, the great lakes region,  
25 and the northeast region) also have the most dramatic variation in median  $\beta_{\text{nit}}$  values. In addition,  $\beta_{\text{nit}}$  values are lower (i.e. higher percentages of nitrate is volatilized from the Teflon filter) in locations and in seasons where temperature is higher. For example,

## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



the southeast is warmer, on average, than the rest of the country throughout the entire year. Median nitrate coefficients in this region are lower than median nitrate coefficients in all other regions in every quarter. Similarly, summer nitrate coefficients are lower than winter nitrate coefficients in all regions. Finally we estimate that any site whose  $\beta_{\text{nit}}$  value is within 1.5 standard deviations of 0 has total nitrate volatilization. The number of sites falling into this category increase from 6 in the winter to 71 in the summer, again showing that more nitrate volatilizes in warmer months. Since nitrate volatilization is governed by the temperature-dependent nitrate equilibrium (Hering and Cass, 1999), this behavior is expected. In addition, there is a large range of  $\beta_{\text{nit}}$  values in quarter 3 which may be due, in part, to low nitrate concentrations. This large variation may indicate that, although we think the seasonal variation demonstrated by these analyses is real, the regression model is not capable of precisely estimating  $\beta_{\text{nit}}$  in the summer months.

### 3.3 OM/OC results

The analysis of spatial and temporal trends in sulfate, nitrate, and soil coefficients shows that they can all be reasonably explained by known physical or chemical processes. These results lend credence to the ability of this regression technique to identify real trends in  $\text{PM}_{2.5}$  characteristics and build confidence in the OM/OC ratio estimations. Table 1 summarizes the distribution of  $\beta_{\text{OC}}$  values among sites in all regions for all quarters. Table 1 and Fig. 11a shows that wintertime OM/OC ratios are generally higher in the eastern portion of the country than in the western portion of the country. Median  $\beta_{\text{OC}}$  values in the great lakes, southeast, and northeast are 1.58, 1.64, and 1.51 respectively while mean  $\beta_{\text{OC}}$  values for the west and central regions are 1.29 and 1.32 respectively. Higher OM/OC ratios may be a result of high residential wood smoke concentrations in densely populated areas. In addition, high values in the southeast may be due to secondary organic aerosol formation, which has been shown to occur in this region even during winter months (Yu et al., 2007). For quarter 1 regressions, there is the most site-to-site variability within the Western and Central

## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



regions. Ninety percent of  $\beta_{OC}$  values lie between 0.67 and 1.76 in the West and between 1.18 and 1.64 in the Central US. In contrast, 90% of  $\beta_{OC}$  values lie in a much smaller range in the Eastern half of the country (1.43 to 1.98, 1.44 to 1.87, and 1.29 to 1.78 for the Great Lakes, Southeast, and Northeast regions).

5 Figure 11b suggests that OM/OC ratios in the summer do not vary substantially by region: median  $\beta_{OC}$  values are 1.87, 1.81, 1.93, 1.81, and 1.80 in the southeast, northeast, great lakes region, central states, and west. The range of  $\beta_{OC}$  values is also quite consistent between regions for quarter 3 (see Table 1).

Seasonal variations in  $\beta_{OC}$  can also be seen in Table 1 and Fig. 12, which show  $\beta_{OC}$  values are higher in the summer than in the winter. Regressions at only 12 sites yielded higher  $\beta_{OC}$  values in the wintertime than in the summer (regressions for both quarter 1 and quarter 3 were performed at 146 sites). This is consistent with higher SOA concentrations in the summer and more oxidative aging due to more photolysis and thus higher OH concentrations. In addition to lower wintertime values, there is also a larger spread of  $\beta_{OC}$  values in the winter than in other seasons: in quarter 1 we estimate that 90% of  $\beta_{OC}$  values fall between 0.79 and 1.84; in quarter 3 we estimate that 90% of  $\beta_{OC}$  values fall between 1.44 and 2.08. Maps for quarter 2 and quarter 4 results are given in Figs. S14 and S15 of the Supplement.

Differences between our methods and those used by Malm and Hand (2007) are summarized in Table 2. Our changes made to the technique first developed by Malm and Hand (2007) have resulted in substantial changes in OM/OC estimates. Figure 13 shows a comparison of  $\beta_{OC}$  estimates from multiyear regressions in our work to  $\beta_{OC}$  estimates reported by Malm and Hand (2007). Within each region, the  $\beta_{OC}$  estimates from our regressions show less site-to-site variability than estimates from Malm and Hand (2007). In addition, although we find median values in each region similar to those reported by Malm and Hand (2007), our low  $\beta_{OC}$  values are higher than theirs in the great lakes and southeast regions: regression results reported by Malm and Hand (2007) show that  $\beta_{OC}$  values were 1.4 and 1.3 for sites in the 5th percentile of the great lakes and southeast regions while our  $\beta_{OC}$  estimates were 1.7 and 1.5 for sites in

## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

**Determining the spatial and seasonal variability in OM/OC ratios across the US**

H. Simon et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

the 5th percentile in those regions.  $\beta_{OC}$  estimates at sites at the 5th percentile in other regions were comparable in the two studies. Finally, high  $\beta_{OC}$  values in the west and central regions were lower in our multiyear regressions than in the regressions reported by Malm and Hand (2007):  $\beta_{OC}$  values were 2.1 for sites in the 95th percentile of the west and central regions according to Malm and Hand (2007), while  $\beta_{OC}$  values were 1.9 for sites in the 95th percentile of the west and central regions in our study.

To isolate the main cause of these different results, we perform a series of regressions, beginning with the methods described by Malm and Hand (2007) and incrementally changing one parameter with each regression. The original dataset used by Malm and Hand (2007) was downloaded in December of 2004. The download date is important because historic IMPROVE data are updated on the website as QA issues are identified. One such change occurred in November of 2009 when chlorine values were readjusted because the original blank correction was found to be too low (White, 2009b). Our regressions that use the methods of Malm and Hand (2007) match the results reported in that study and will be referred to as the baseline regressions. When comparing the baseline regressions to our final results, 61% of  $\beta_{OC}$  estimates change by more than 0.1 and 37% change by more than 0.2. The three parameters that have the largest effect on  $\beta_{OC}$  values are the dataset download date, the years analyzed (i.e. 1988–2003 vs. 2002–2008), and the choice of covariates. The use of the EiV rather than OLS affects the  $\beta_{OC}$  estimates to a smaller degree. The use of S vs.  $SO_4^{2-}$  to calculate ammonium sulfate and the choice of soil equation has almost no effect on the  $\beta_{OC}$  estimates. The large effect of the years analyzed may indicate a change in this value over time (about 64% of the sites have higher  $\beta_{OC}$  values when using 2002–2008 data than when using 1988–2003 data). In addition, this may indicate that changes to measurement protocols and hardware which occurred during these time periods have influenced the results. Both the effect of the dataset used (download date) and the years analyzed on our results indicate that this analysis may be sensitive to changes in measurement and data processing methodology. We further analyze which part of the covariate changes causes the largest difference in  $\beta_{OC}$



IMPROVE's current OC artifact correction. If OC artifact is not properly corrected on both the Teflon and quartz filters, then the  $\beta_{OC}$  value will be influenced. This could occur if (1) IMPROVE's back-up filter method does not completely capture all positive artifact on quartz filters, (2) Teflon filters have non-negligible positive artifact, or (3) the magnitude of negative artifact differs on the quartz and Teflon filters. An in-depth exploration of OC artifact is beyond the scope of this paper, but the uncertainties associated with these issues should be kept in mind when interpreting the regression results presented here.

#### 4 Summary and future work

This work has helped to develop a robust technique for estimating OM/OC ratios that can be applied to a comprehensive dataset, such as the IMPROVE monitoring network data. The ability of this technique to estimate physically reasonable spatial and seasonal trends in  $\beta_{OC}$ ,  $\beta_{sulf}$ ,  $\beta_{nit}$ , and  $\beta_{soil}$  builds confidence in the results. Furthermore, major improvements on the technique reported by Malm and Hand (2007) include the use of an errors-in-variables regression and the elimination of EC as an explanatory variable. We believe that these two changes provide more realistic results and have eliminated substantial biases that occurred in approximately 1/4 of the regressions performed by Malm and Hand (2007). Comparison with other OM/OC estimation methods will be the subject of future work.

In addition, this work has helped to identify further areas of research for IMPROVE data. First, our analysis shows the effect of the neutralization state on predicted mass associated with sulfate. These results suggest that sulfate is often not fully neutralized. Further studies to characterize ammonium concentrations could improve our ability to achieve mass closure with IMPROVE data. Second, nitrate volatilization appears to vary substantially by site and season. A measurement study could be performed to verify the nitrate volatilization estimates made here. In addition, samples could be shipped in refrigerated conditions to prevent nitrate volatilization during transport. At a

### Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



minimum, these results demonstrate the importance of recording temperature and RH conditions to which filters are exposed during sampling, transport, and measurement. Finally, this work has identified general temporal and spatial trends in OM/OC ratios showing that during the winter they are larger in the eastern US than in the West and that summertime values are larger than wintertime values across the US. Considering this work, and that of Malm and Hand (2007) and El-Zanan et al. (2005), the IMPROVE steering committee should re-evaluate the current practice of using a fixed OM/OC ratio when calculating reconstructed fine mass.

**Supplementary material related to this article is available online at:**

**<http://www.atmos-chem-phys-discuss.net/10/24651/2010/acpd-10-24651-2010-supplement.pdf>.**

*Acknowledgements.* The authors would like to thank Warren White for helpful information on IMPROVE chloride measurements and feedback on this project, Chuck McDade for providing IMPROVE laboratory RH data, Lowell Ashbaugh for providing IMPROVE OC back-up filter measurement data, Doug Lowenthal for describing details of his mass closure methods, and Mark Pitchford, Venkatesh Rao, and Steve McDow for their helpful feedback on this work. The United States Environmental Protection Agency through its Office of Research and Development funded and managed the research described here. It has been subjected to Agency's administrative review and approved for publication.

## References

- Aiken, A., Decarlo, P. F., Kroll, J. H., et al.: O/C and OM/OC ratios of primary, secondary, and ambient organic aerosols with high-resolution time-of-flight aerosol mass spectrometry, *Environ. Sci. Technol.*, 42, 4478–4485, doi:10.1021/es703009q, 2008.
- Chan, T. W., Huang, L., Leaitch, W. R., Sharma, S., Brook, J. R., Slowik, J. G., Abbatt, J. P. D., Brickell, P. C., Liggio, J., Li, S.-M., and Moosmüller, H.: Observations of OM/OC and specific attenuation coefficients (SAC) in ambient fine PM at a rural site in central Ontario, Canada, *Atmos. Chem. Phys.*, 10, 2393–2411, doi:10.5194/acp-10-2393-2010, 2010.

## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





**Determining the spatial and seasonal variability in OM/OC ratios across the US**

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Chen, X. and Yu, J. Z.: Measurement of organic mass to organic carbon ratio in ambient aerosol samples using a gravimetric technique in combination with chemical analysis, *Atmos. Environ.*, 41, 8857–8864, doi:10.1016/j.atmosenv.2007.08.023, 2007.
- 5 El-Zanan, H. S., Lowenthal, D. H., Zielinska, B., et al.: Determination of the organic aerosol mass to organic carbon ratio in IMPROVE samples, *Chemosphere*, 60, 485–496, doi:10.1016/j.chemosphere.2005.01.005, 2005.
- El-Zanan, H. S., Zielinska, B., Mazzoleni, L. R., et al.: Analytical determination of the aerosol organic mass-to-organic carbon ratio, *J. Air Waste Manage.*, 59, 58–69, doi:10.3155/1047-3289.59.1.58, 2009.
- 10 EPA regional planning organizations: <http://epa.gov/visibility/regional.html#thefive>, last access: 14 October, 2010.
- Frank, N. H.: Retained nitrate, hydrated sulfates, and carbonaceous mass in Federal Reference Method fine particulate matter for six eastern US cities, *J. Air Waste Manage.*, 56, 500–511, 2006.
- 15 Fuller, W. A.: *Measurement error models*, John Wiley & Sons, New York, USA, 1987.
- Gilardoni, S., Russell, L. M., Sotooshian, A., et al.: Regional variation of organic functional groups in aerosol particles on four US east coast platforms during the International Consortium for Atmospheric Research on Transport and Transformation 2004 campaign, *J. Geophys. Res.-Atmos.*, 112, D10S27, doi:10.1029/2006jd007737, 2007.
- 20 Hand, J. L. and Malm, W. C.: Review of the IMPROVE equation for estimating ambient light extinction coefficients – final report, Colorado State University, CIRA, 146, 2006.
- Hering, S. and Cass, G.: The magnitude of bias in the measurement of PM<sub>2.5</sub> arising from volatilization of particulate nitrate from teflon filters, *J. Air Waste Manage.*, 49, 725–733, 1999.
- 25 Hyslop, N. P. and White, W. H.: An evaluation of interagency monitoring of protected visual environments (IMPROVE) collocated precision and uncertainty estimates, *Atmos. Environ.*, 42, 2691–2705, doi:10.1016/j.atmosenv.2007.06.053, 2008.
- Jimenez, J. L., Canagaratna, M. R., Donahue, N. M., et al.: Evolution of organic aerosols in the atmosphere, *Science*, 326, 1525–1529, doi:10.1126/science.1180353, 2009.
- 30 Kiss, G., Varga, B., Galambos, I., et al.: Characterization of water-soluble organic matter isolated from atmospheric fine aerosol, *J. Geophys. Res.-Atmos.*, 107, 8339, doi:10.1029/2001jd000603, 2002.

---

**Determining the spatial and seasonal variability in OM/OC ratios across the US**

---

H. Simon et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Kleindienst, T. E., Jaoui, M., Lewandowski, M., et al.: Estimates of the contributions of biogenic and anthropogenic hydrocarbons to secondary organic aerosol at a southeastern US location, *Atmos. Environ.*, 41, 8288–8300, doi:10.1016/j.atmosenv.2007.06.045, 2007.

Liu, S., Takahama, S., Russell, L. M., Gilardoni, S., and Baumgardner, D.: Oxygenated organic functional groups and their sources in single and submicron organic particles in MILA-GRO 2006 campaign, *Atmos. Chem. Phys.*, 9, 6849–6863, doi:10.5194/acp-9-6849-2009, 2009.

Lowenthal, D. H. and Kumar, N.:  $PM_{2.5}$  mass and light extinction reconstruction in IMPROVE, *J. Air Waste Manage.*, 53, 1109–1120, 2003.

Malm, W. C., Sisler, J. F., Huffman, D., et al.: Spatial and seasonal trends in particle concentration and optical extinction in the United-States, *J. Geophys. Res.-Atmos.*, 99, 1347–1370, 1994.

Malm, W. C., Day, D. E., Carrico, C., et al.: Intercomparison and closure calculations using measurements of aerosol species and optical properties during the Yosemite Aerosol Characterization Study, *J. Geophys. Res.-Atmos.*, 110, D14302, doi:10.1029/2004jd005494, 2005.

Malm, W. C. and Hand, J. L.: An examination of the physical and optical properties of aerosols collected in the IMPROVE program, *Atmos. Environ.*, 41, 3407–3427, doi:10.1016/j.atmosenv.2006.12.012, 2007.

McDade, C. E.: IMPROVE standard operating procedure, Crocker Nuclear Laboratory, University of California, Davis, CASOP 351-2, 258, 2008.

McDow, S. R. and Huntzicker, J. J.: Vapor adsorption artifact in the sampling of organic aerosol – face velocity effects, *Atmos. Environ. A Gen.*, 24, 2563–2571, 1990.

Murphy, D. M., Cziczo, D. J., Froyd, K. D., et al.: Single-particle mass spectrometry of tropospheric aerosol particles, *J. Geophys. Res.-Atmos.*, 111(15), D23s32, doi:10.1029/2006jd007340, 2006.

Pang, Y., Turpin, B. J., and Gundel, L. A.: On the importance of organic oxygen for understanding organic aerosol particles, *Aerosol Sci. Tech.*, 40, 128–133, doi:10.1080/02786820500423790, 2006.

Pitchford, M., Malm, W. C., Schichtel, B., et al.: Revised algorithm for estimating light extinction from IMPROVE particle speciation data, *J. Air Waste Manage.*, 57, 1326–1336, doi:10.3155/1047-3289.57.11.1326, 2007.

## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



R Development Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, <http://www.R-project.org>, last access: 14 October 2010, ISBN 3-900051-07-0, Vienna, Austria, 2010.

Reff, A., Turpin, B. J., Offenberg, J. H., et al.: A functional group characterization of organic PM<sub>2.5</sub> exposure: Results from the RIOPA study, *Atmos. Environ.*, 41, 4585–4598, doi:10.1016/j.atmosenv.2007.03.054, 2007.

Reff, A., Bhawe, P. V., Simon, H., et al.: Emissions inventory of PM<sub>2.5</sub> trace elements across the United States, *Environ. Sci. Technol.*, 43, 5790–5796, doi:10.1021/es802930x, 2009.

Russell, L. M.: Aerosol organic-mass-to-organic-carbon ratio measurements, *Environ. Sci. Technol.*, 37, 2982–2987, doi:10.1021/es026123w, 2003.

Russell, L. M., Takahama, S., Liu, S., et al.: Oxygenated fraction and mass of organic aerosol from direct emission and atmospheric processing measured on the R/V Ronald Brown during TEXAQS/GoMACCS 2006, *J. Geophys. Res.-Atmos.*, 114, D00F05, doi:10.1029/2008jd011275, 2009.

Saylor, R. D., Edgerton, E. S., and Hartsell, B. E.: Linear regression techniques for use in the EC tracer method of secondary organic aerosol estimation, *Atmos. Environ.*, 40, 7546–7556, doi:10.1016/j.atmosenv.2006.07.018, 2006.

Simon, H., Beck, L., Bhawe, P. V., et al.: The development and use of EPA's SPECIATE database, *Atmos. Pollut. Res.*, 1, 196–206, 2010.

Sun, Y., Zhang, Q., Macdonald, A. M., Hayden, K., Li, S. M., Liggio, J., Liu, P. S. K., Anlauf, K. G., Leaitch, W. R., Steffen, A., Cubison, M., Worsnop, D. R., van Donkelaar, A., and Martin, R. V.: Size-resolved aerosol chemistry on Whistler Mountain, Canada with a high-resolution aerosol mass spectrometer during INTEX-B, *Atmos. Chem. Phys.*, 9, 3095–3111, doi:10.5194/acp-9-3095-2009, 2009.

Surratt, J. D., Gomez-Gonzalez, Y., Chan, A. W. H., et al.: Organosulfate formation in biogenic secondary organic aerosol, *J. Phys. Chem. A*, 112, 8345–8378, doi:10.1021/jp802310p, 2008.

Turpin, B. J., Huntzicker, J. J., and Hering, S. V.: Investigation of organic aerosol sampling artifacts in the Los-Angeles basin, *Atmos. Environ.*, 28, 3061–3071, 1994.

Turpin, B. J. and Lim, H. J.: Species contributions to PM<sub>2.5</sub> mass concentrations: Revisiting common assumptions for estimating organic mass, *Aerosol Sci. Tech.*, 35, 602–610, 2001.

**Determining the spatial and seasonal variability in OM/OC ratios across the US**

H. Simon et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Watson, J. G., Chow, J. C., Chen, L.-W. A., et al.: Assessment of carbon sampling artifacts in the IMPROVE, STN/CSN, and SEARCH networks, Desert Research Institute, Reno, NV, 2008.

Wexler, A. S. and Clegg, S. L.: Atmospheric aerosol models for systems including the ions  $\text{H}^+$ ,  $\text{NH}_4^+$ ,  $\text{Na}^+$ ,  $\text{SO}_4^{2-}$ ,  $\text{NO}_3^-$ ,  $\text{Cl}^-$ ,  $\text{Br}^-$ , and  $\text{H}_2\text{O}$ , *J. Geophys. Res.-Atmos.*, 107(14), 4207, doi:10.1029/2001jd000451, 2002.

White, W. H. and Roberts, P. T.: Nature and origins of visibility-reducing aerosols in Los-Angeles air basin, *Atmos. Environ.*, 11, 803–812, 1977.

White, W. H.: On the theoretical and empirical-basis for apportioning extinction by aerosols – a critical-review, *Atmos. Environ.*, 20, 1659–1672, 1986.

White, W. H.: Statistical considerations in the interpretation of size-resolved particulate mass data, *J. Air Waste Manage.*, 48, 454–458, 1998.

White, W. H.: IMPROVE data advisory: Shift in EC/OC split with 1 January 2005 TOR hardware upgrade, 2007.

White, W. H.: Chemical markers for sea salt in IMPROVE aerosol data, *Atmos. Environ.*, 42, 261–274, doi:10.1016/j.atmosenv.2007.09.040, 2008.

White, W. H.: IMPROVE data advisory, Inconstant bias in XRF sulfur, 2009a.

White, W. H.: IMPROVE data advisory: Under-correction of chloride concentrations for filter blank levels – historical advisory, Applies to downloads before 23 November 2009, 2009b.

Yu, L. E., Shulman, M. L., Kopperud, R., et al.: Fine organic aerosols collected in a humid, rural location (Great Smoky Mountains, Tennessee, USA): Chemical and temporal characteristics, *Atmos. Environ.*, 39, 6037–6050, doi:10.1016/j.atmosenv.2005.06.043, 2005.

Yu, S., Bhave, P. V., Dennis, R. L., et al.: Seasonal and regional variations of primary and secondary organic aerosols over the Continental United States: Semi-empirical estimates and model evaluation, *Environ. Sci. Technol.*, 41, 4690–4697, doi:10.1021/es061535g, 2007.

Zhang, Q., Jimenez, J. L., Canagartna, M. R., et al.: Ubiquity and dominance of oxygenated species in organic aerosols in anthropogenically-influenced Northern Hemisphere midlatitudes, *Geophys. Res. Lett.*, 34, L13801, doi:10.1029/2007gl029979, 2007.

**Determining the spatial and seasonal variability in OM/OC ratios across the US**

H. Simon et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

⏪ ⏩

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

**Table 1.** Summary of  $\beta_{OC}$  distributions among sites for each quarter and region.

Region	Quarter	$B_{OC}$					Number of regressions
		5th percentile	25th percentile	50th percentile	75th percentile	95th percentile	
West	1	0.67	1.06	1.29	1.42	1.76	89
West	2	1.36	1.66	1.81	1.90	2.14	86
West	3	1.33	1.66	1.80	1.88	2.04	85
West	4	1.22	1.43	1.57	1.68	1.88	86
Central	1	1.18	1.27	1.32	1.52	1.64	21
Central	2	1.59	1.69	1.78	1.87	2.10	21
Central	3	1.51	1.72	1.81	1.92	2.07	19
Central	4	1.37	1.45	1.53	1.64	1.90	21
Great Lakes	1	1.43	1.44	1.58	1.81	1.98	5
Great Lakes	2	1.83	1.83	1.94	1.95	1.97	5
Great Lakes	3	1.67	1.90	1.93	1.95	2.01	5
Great Lakes	4	1.31	1.31	1.48	1.61	1.61	5
Southeast	1	1.44	1.58	1.64	1.80	1.87	17
Southeast	2	1.50	1.76	1.89	2.00	2.16	16
Southeast	3	1.47	1.75	1.87	2.08	2.25	16
Southeast	4	1.42	1.60	1.67	1.75	1.83	17
Northeast	1	1.29	1.43	1.51	1.60	1.78	20
Northeast	2	1.23	1.74	1.87	2.01	2.09	19
Northeast	3	1.69	1.76	1.81	1.90	2.03	20
Northeast	4	1.07	1.49	1.57	1.67	1.85	16
all	1	0.79	1.20	1.39	1.58	1.84	153
all	2	1.39	1.69	1.83	1.94	2.15	148
all	3	1.44	1.72	1.81	1.91	2.08	146
all	4	1.24	1.44	1.59	1.68	1.87	146
all	all	1.10	1.44	1.66	1.83	2.06	593



## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



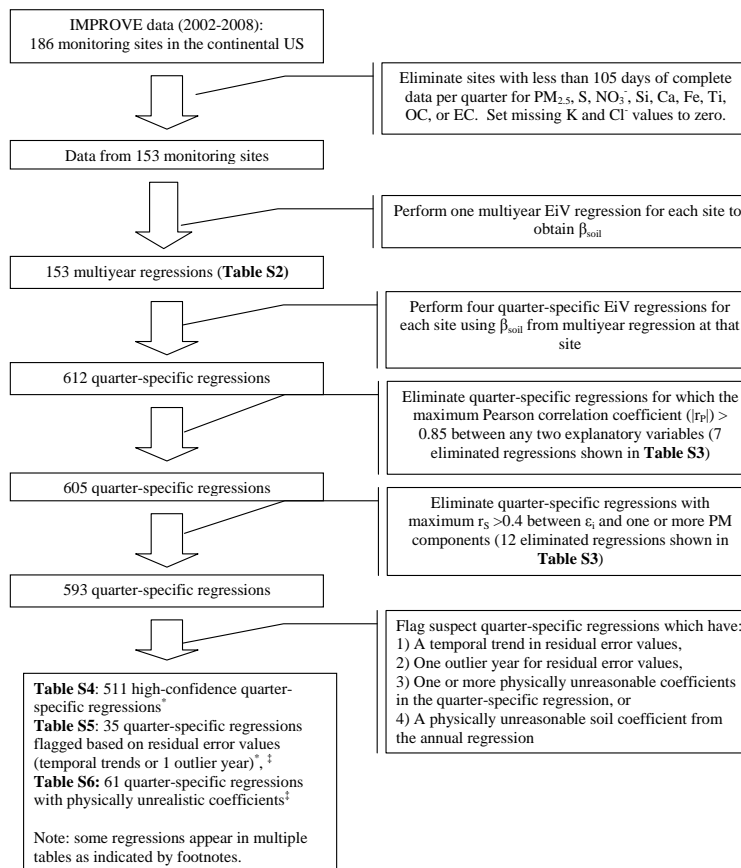
**Table 2.** Differences between our methodology and that of Malm and Hand (2007).

Methodological Aspect	Malm and Hand (2007)	This work
IMPROVE dataset	Download date: 3 Dec 2004 Years analyzed: 1988–2003	Download date: 6 Jan 2010 Years analyzed: 2002–2008
Data segregated by	Monitoring site	Monitoring site for $\beta_{\text{soil}}$ Monitoring site and quarter for all other coefficients
Regression type	Ordinary least squares	Errors-in-variables
Response variable	PM <sub>2.5</sub>	PM <sub>2.5</sub> – (1.2 KNON + 1.8 Cl <sup>–</sup> + EC)
Intercept ( $\beta_0$ )	Included	Excluded
Explanatory variables	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> , NH <sub>4</sub> NO <sub>3</sub> , OC, EC, soil, sea salt*	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> , NH <sub>4</sub> NO <sub>3</sub> , OC, soil
Calculation of explanatory variables	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> = 1.37 × SO <sub>4</sub> <sup>2–</sup> (SO <sub>4</sub> <sup>2–</sup> measured by ion chromatography) SOIL from Eq. (2)	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> = 4.125 × S (S measured by XRF) SOIL from Eq. (7)

\* Note: Sea salt was not used as an explanatory variable for sites with a small number of dates that reported Cl<sup>–</sup> concentrations: ADPI1, AGTI1, AREN1, BALD1, BOAP1, BRLA1, CACR1, CADI1, CAPI1, CEBL1, CHER1, CHOI1, COHU1, CRES1, CRMO1, DEVA1, DOME1, ELDO1, ELLI1, FOPE1, GAMO1, GRGU1, HALE1, HEGL1, HOOV1, IKBA1, JARI1, JOSH1, LASU1, LIGO1, LIVO1, LOST1, MACA1, MELA1, MING1, MKGO1, MOM1, MONT1, NEBR1, NOCH1, PMRF1, QUCI1, QURE1, QUVA1, SAFO1, SAGA1, SAGU1, SAMA1, SAPE1, SENE1, SHRO1, SIKE1, SIPS1, SPOK1, SWAN1, TALL1, THBA1, THRO1, ULBE1, WHRI1, WICA1, WIMO1, ZION1.

## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.



**Fig. 1.** Flow diagram outlining regression methodology used in this work.

\* Table includes 10 regressions originally flagged for an outlier year, but for which excluding that year does not change coefficients.

‡ Table includes 4 regressions with both physically unrealistic coefficients and low confidence temporal trends.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

◀ ▶

◀ ▶

Back Close

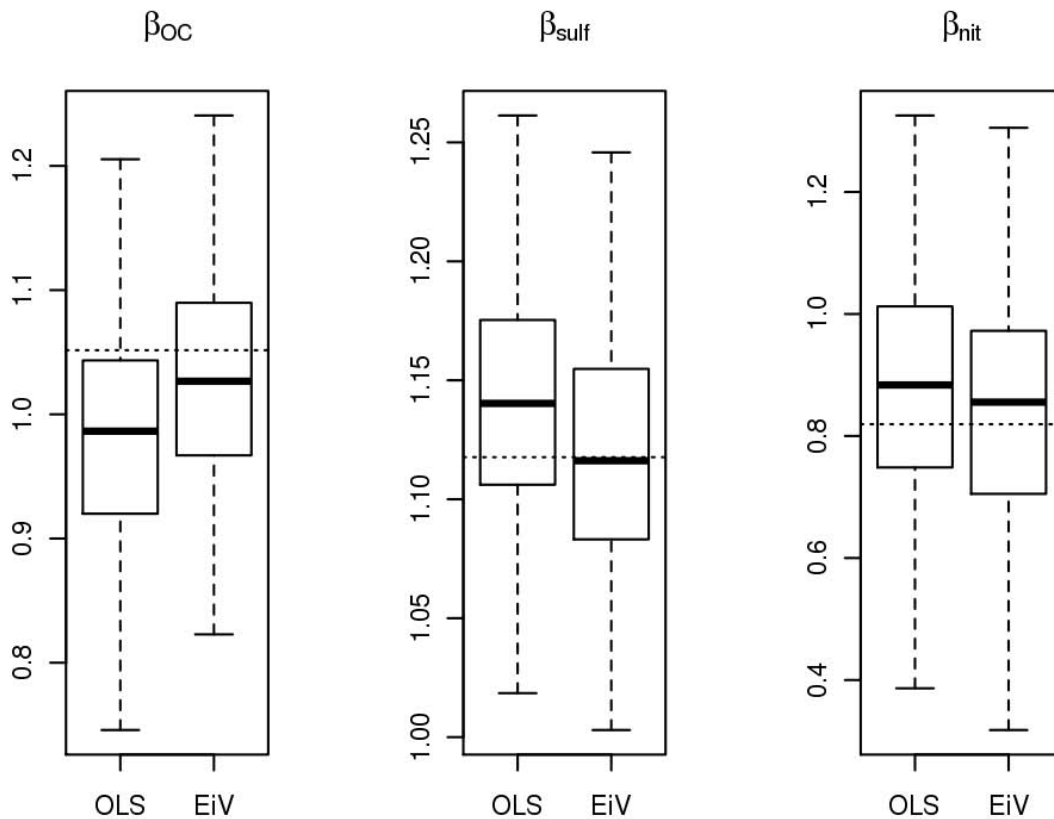
Full Screen / Esc

Printer-friendly Version

Interactive Discussion







**Fig. 2.** Bias in regression coefficients caused by measurement error in synthetic datasets representative of Gila Wilderness, NM in quarter 1. Horizontal dotted lines represent the “true” value of each coefficient. The left box in each panel illustrates bias for OLS regressions and the right box shows a greatly reduced bias after implementing the errors-in-variables (EIV) regression method.

**Determining the spatial and seasonal variability in OM/OC ratios across the US**

H. Simon et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

⏪ ⏩

◀ ▶

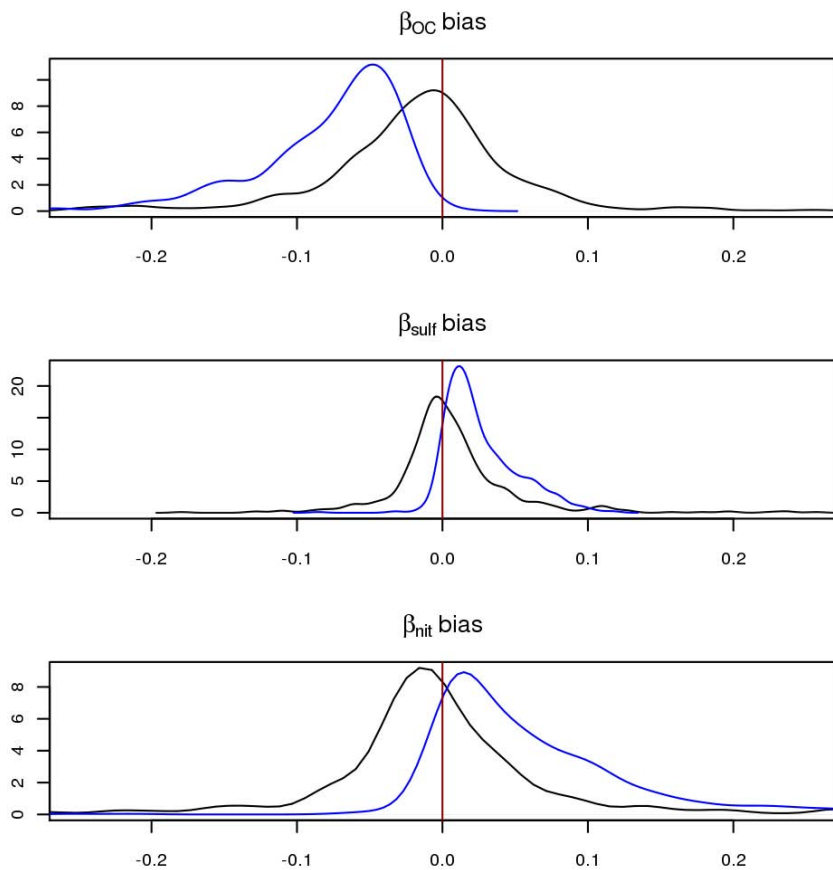
Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





**Fig. 3.** Distribution of bias in regression coefficients for quarter-specific regressions at all IMPROVE sites. For each technique, we compute the median bias from 200 synthetic datasets at each site/quarter using ordinary least squares (blue) and errors-in-variables regression (black) and plot the distribution of those median values across all regressions. The red vertical line shows zero bias.

**Determining the spatial and seasonal variability in OM/OC ratios across the US**

H. Simon et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

⏪ ⏩

◀ ▶

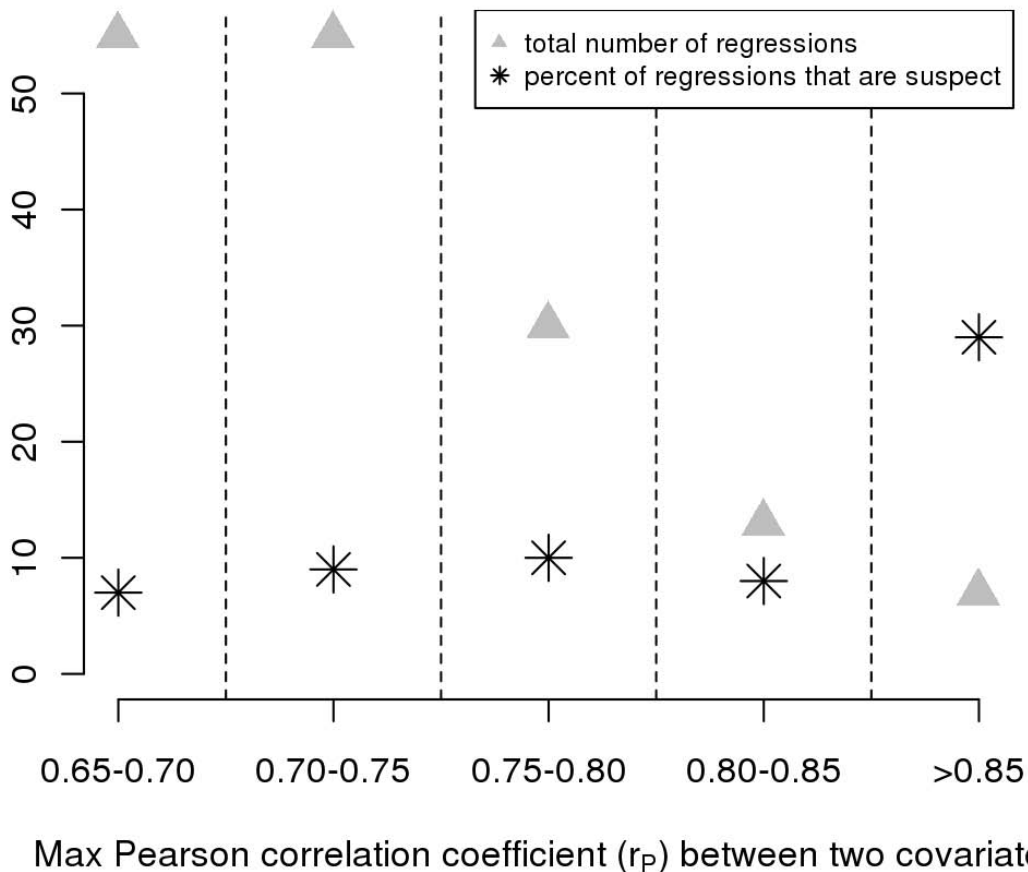
Back Close

Full Screen / Esc

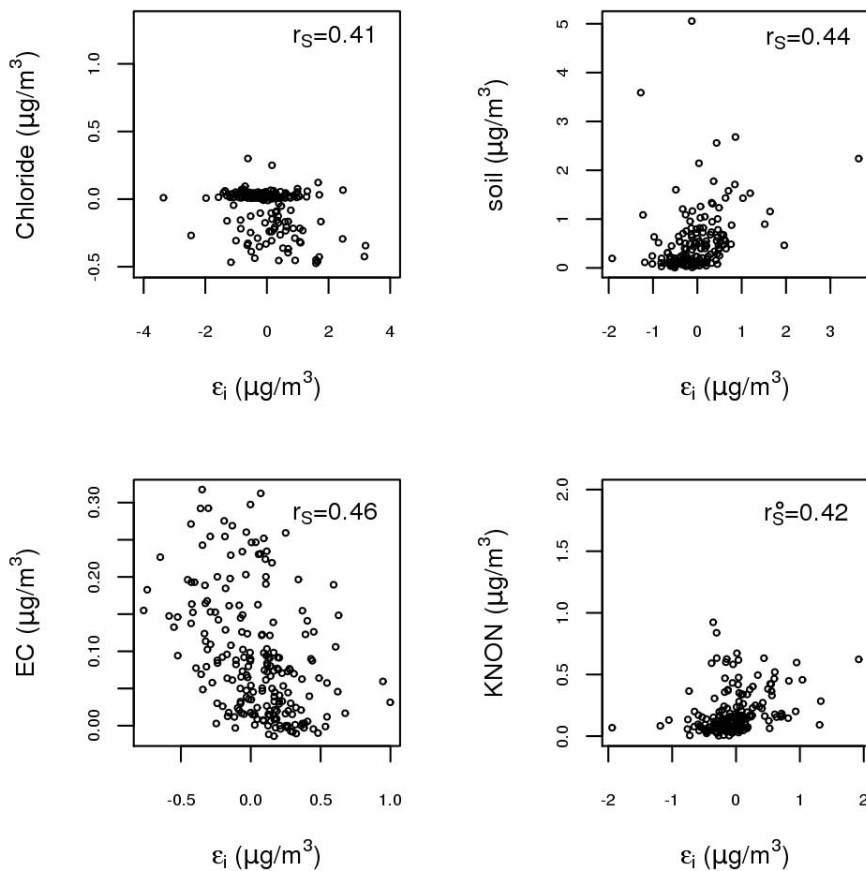
Printer-friendly Version

Interactive Discussion





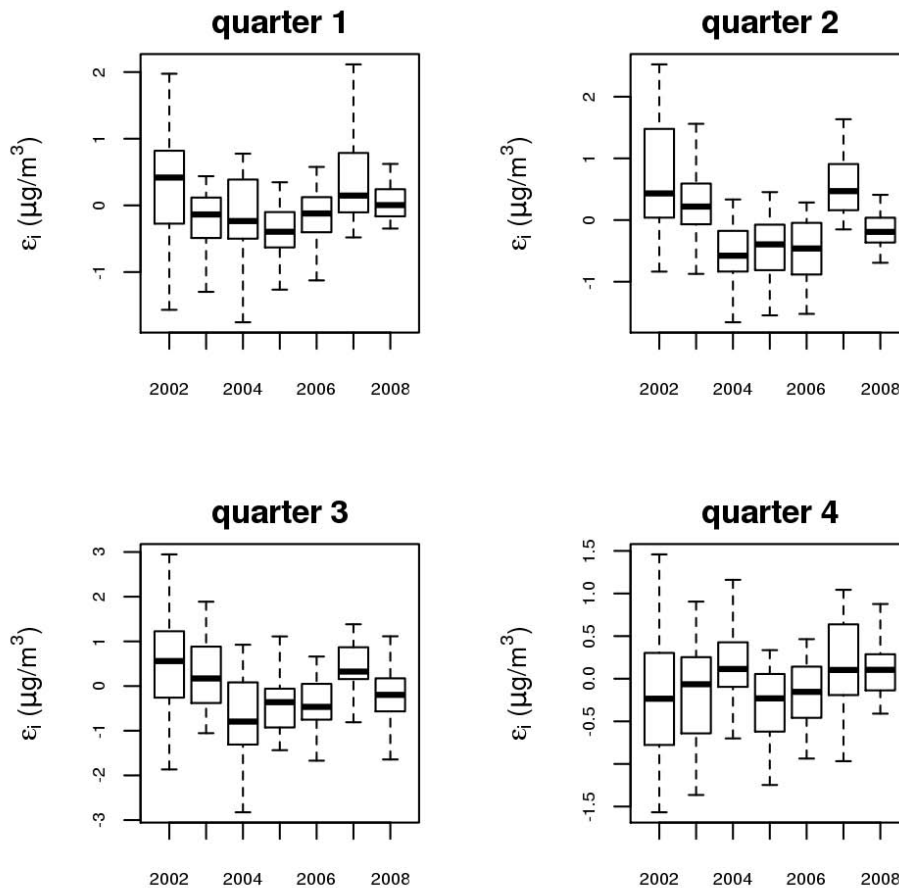
**Fig. 4.** Empirical selection of a threshold for identifying regressions which may be biased due to multicollinearity. See Sect. 2.4.1 for an explanation of what constitutes a regression that is “suspect”.



**Fig. 5.** Example datasets in which residual error ( $\epsilon_i$ ) exhibits a high correlation with a PM<sub>2.5</sub> component: chloride at Hercules-Glades, Missouri – quarter 2 (upper left), EC at Mount Hood, Oregon – quarter 2 (lower left), SOIL at Domelands Wilderness, California – quarter 4 (upper right), and KNON at Sawtooth National Forest, Idaho – quarter 4 (lower right).

**Determining the spatial and seasonal variability in OM/OC ratios across the US**

H. Simon et al.

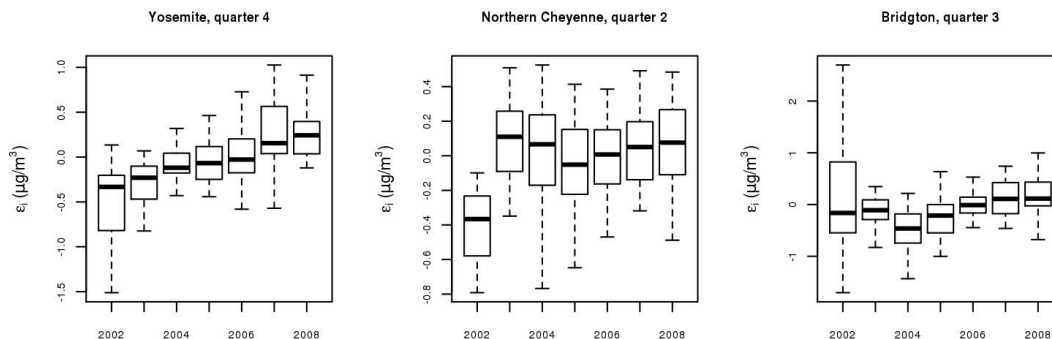


**Fig. 6.** Lack of systematic change in residual error values ( $\varepsilon_i$ ) between 2004 and 2005 at the Sipsy Wilderness site.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

## Determining the spatial and seasonal variability in OM/OC ratios across the US

H. Simon et al.



**Fig. 7.** Residual error values ( $\varepsilon_i$ ) from quarter 4 at Yosemite National Park show a monotonically increasing trend between 2002 and 2008 (left).  $\varepsilon_i$  in quarter 2 from Northern Cheyenne show that the inter-quartile range for 2002 does not overlap with other years (center). There is a greater spread in  $\varepsilon_i$  during quarter 3 at Bridgton, Maine in 2002 than in all other years (right).

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

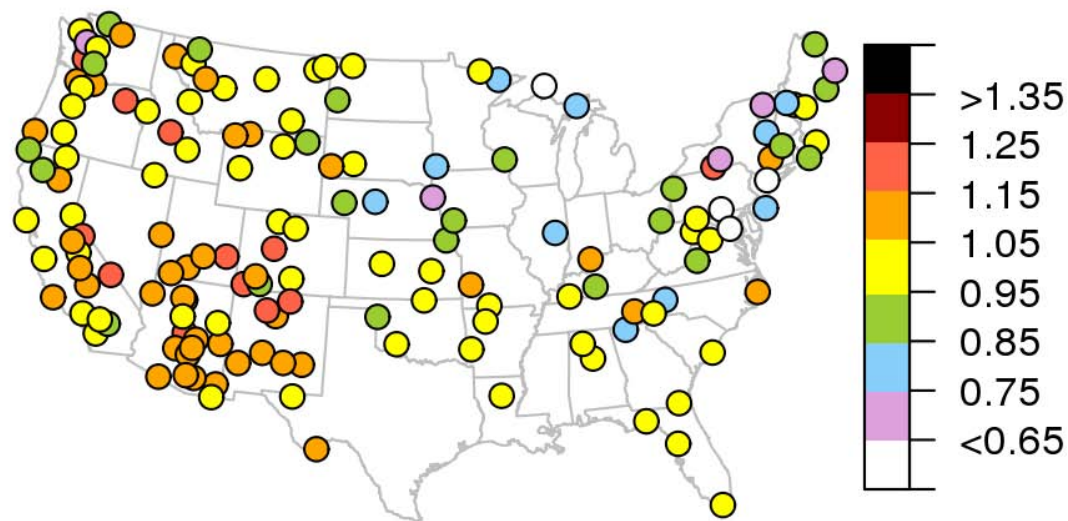
Full Screen / Esc

Printer-friendly Version

Interactive Discussion

**Determining the spatial and seasonal variability in OM/OC ratios across the US**

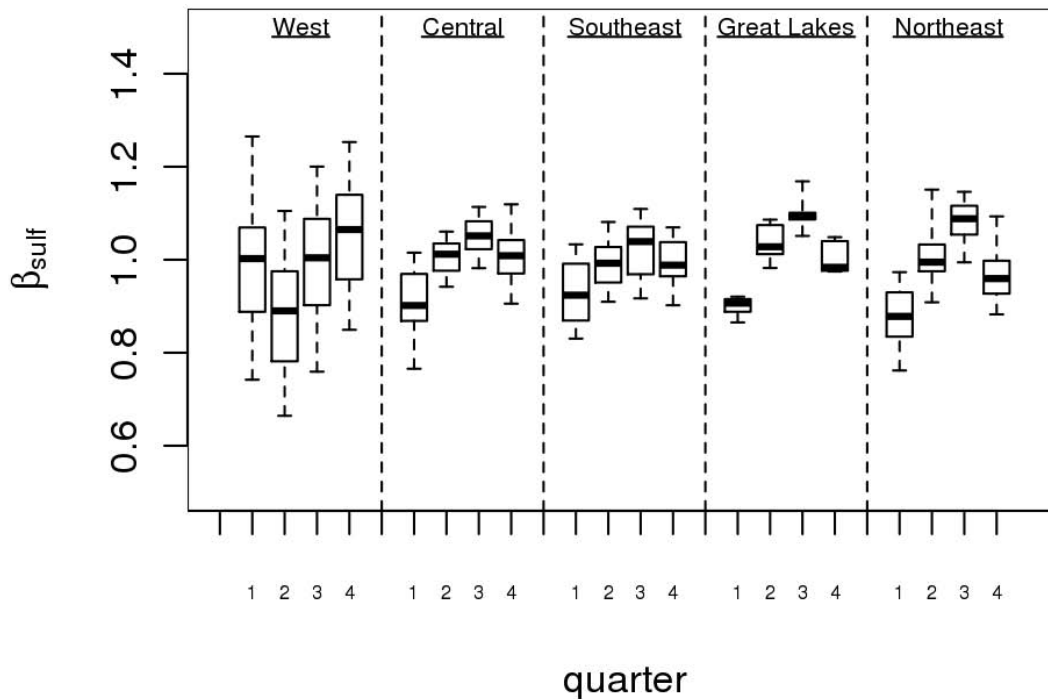
H. Simon et al.

**Fig. 8.**  $\beta_{\text{soil}}$  at 153 IMPROVE sites.[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)



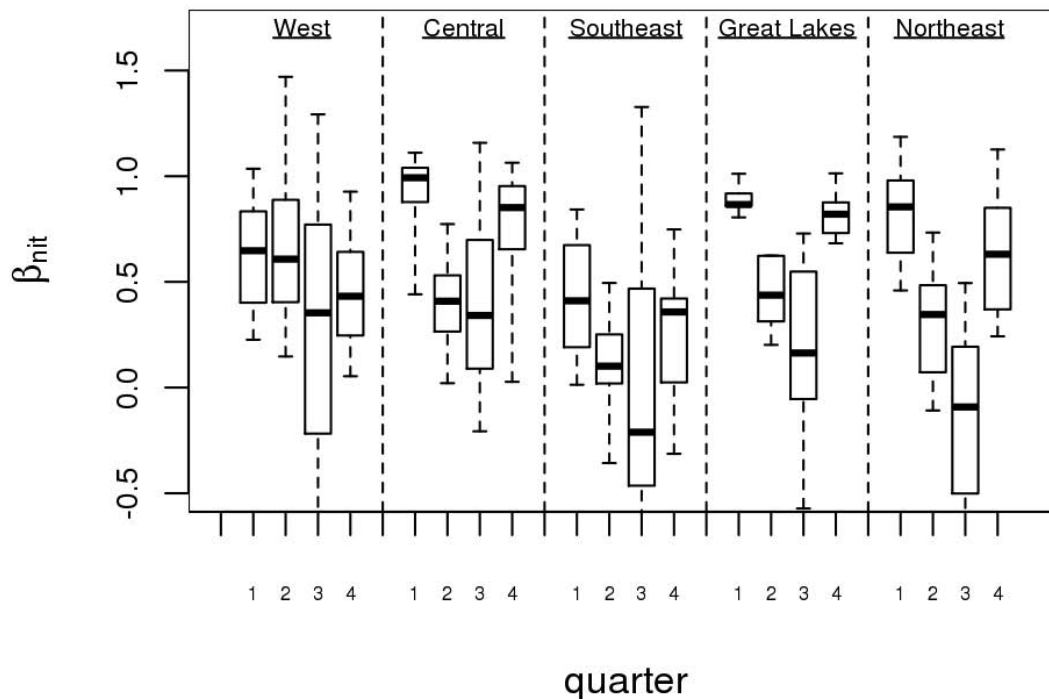
**Determining the spatial and seasonal variability in OM/OC ratios across the US**

H. Simon et al.

**Fig. 9.** Spatial and temporal trends in  $\beta_{\text{sulf}}$ .[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

**Determining the spatial and seasonal variability in OM/OC ratios across the US**

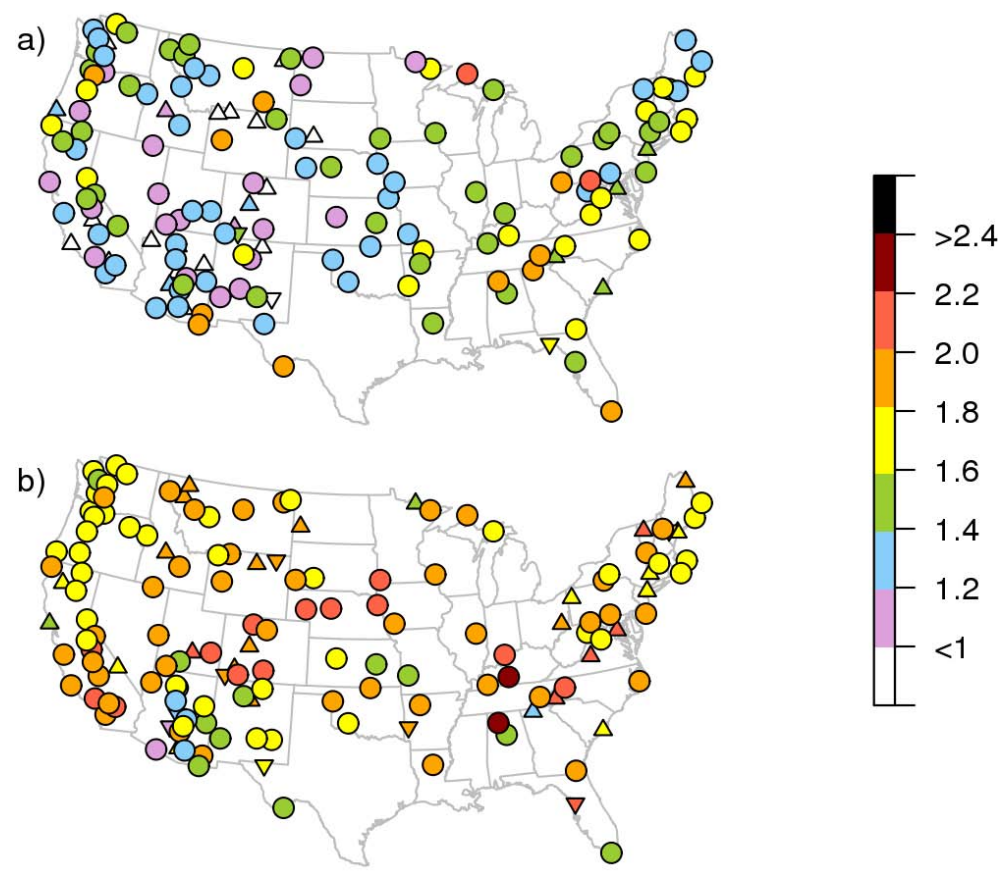
H. Simon et al.

**Fig. 10.** Spatial and temporal trends in  $\beta_{\text{nit}}$ .[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

## Determining the spatial and seasonal variability in OM/OC ratios across the US

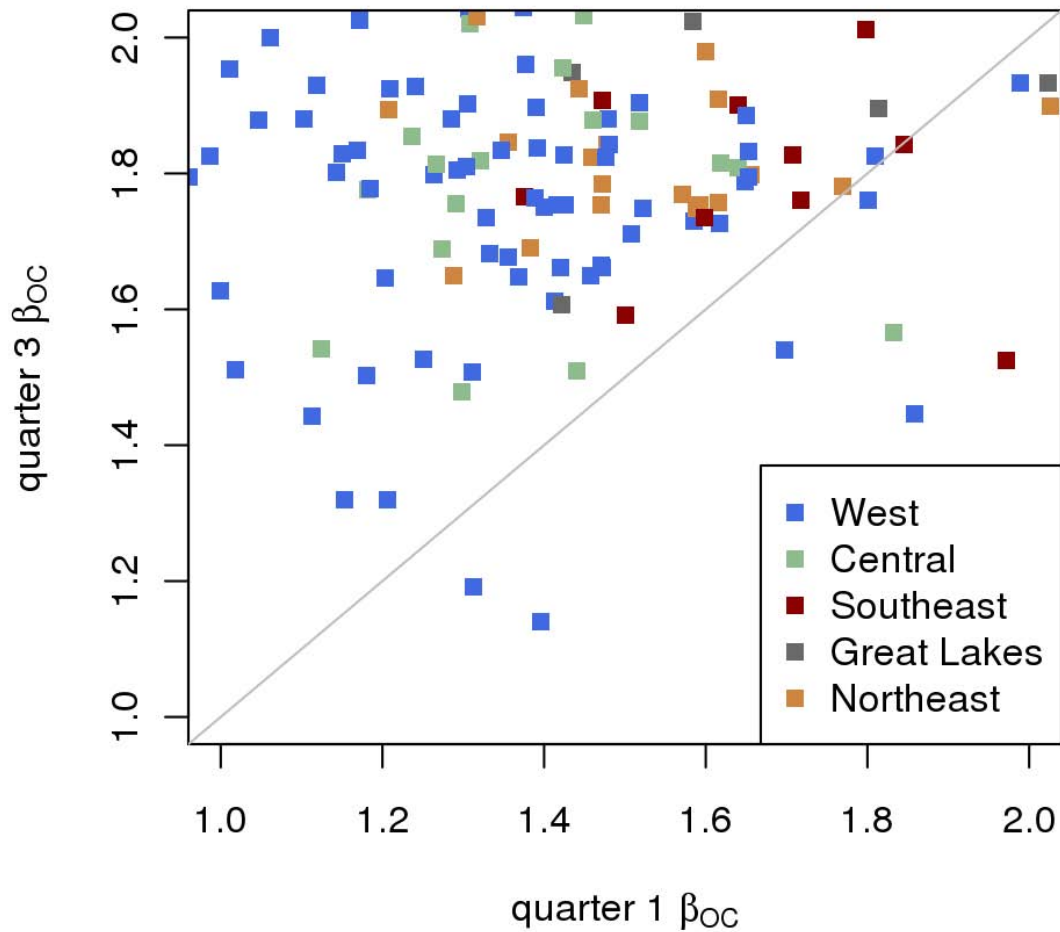
H. Simon et al.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	



**Fig. 11.**  $\beta_{OC}$  values for quarter 1 (top) and quarter 3 (bottom). High confidence results are depicted by circles, regressions with questionable residual trends are depicted by downward facing triangles, and regressions with a questionable coefficient are depicted by upward facing triangles.





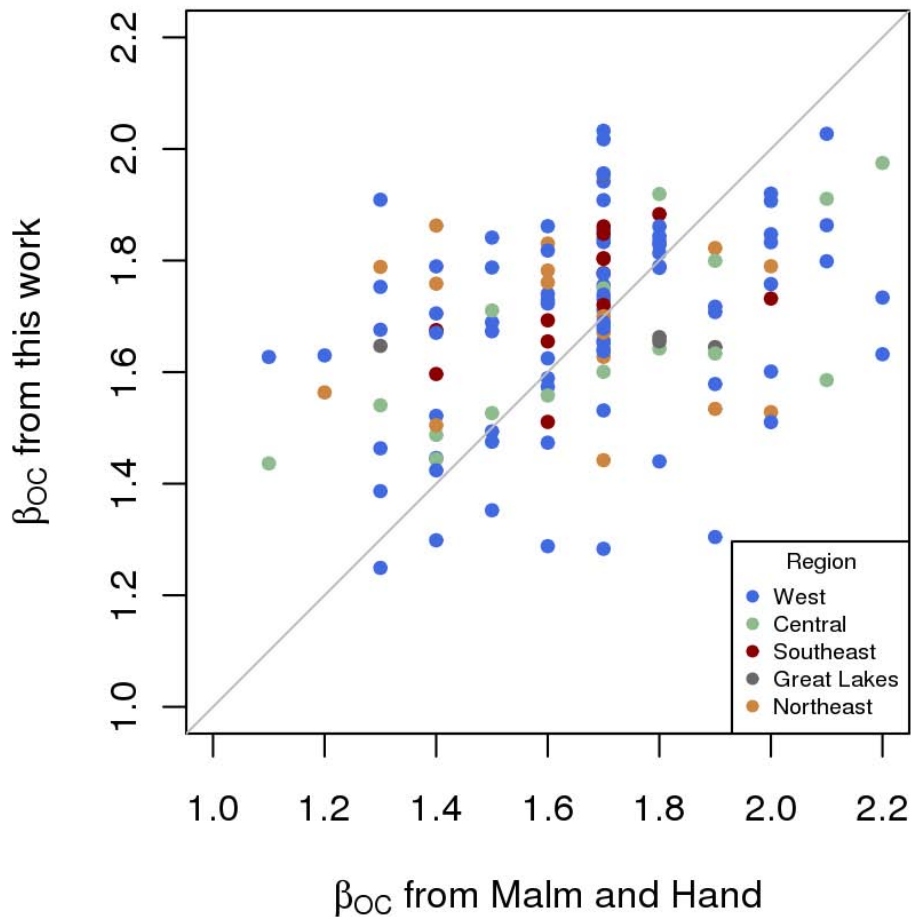
**Fig. 12.** Comparison of  $\beta_{OC}$  values for quarters 1 and 3.

**Determining the spatial and seasonal variability in OM/OC ratios across the US**

H. Simon et al.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
⏪	⏩
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	





**Fig. 13.** Comparisons of  $\beta_{OC}$  values reported by Malm and Hand (2007) to multiyear  $\beta_{OC}$  values from this work.

**Determining the spatial and seasonal variability in OM/OC ratios across the US**

H. Simon et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

⏪ ⏩

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

