Appendix G

Revised IMPROVE Algorithm for Estimating Light Extinction from Particle Speciation Data

Revised IMPROVE Algorithm for Estimating Light Extinction from Particle Speciation Data

Section II. Overview of the Revised Algorithm

(IMPROVE technical subcommittee for algorithm review)

The recommended revised algorithm is shown in the equation below with revised terms in bold font. The total sulfate, nitrate and organic carbon compound concentrations are each split into two fractions, representing small and large size distributions of those components. Though not explicitly shown in the equation, the organic mass concentration used in this new algorithm is 1.8 times the organic carbon mass concentration, changed from 1.4 times carbon mass concentration as used for input for the current IMPROVE algorithm. New terms have been added for sea salt (important for coastal locations) and for absorption by NO_2 (only used where NO_2 data are available). Site-specific Rayleigh scattering is calculated for the elevation and annual average temperature of each of the IMPROVE monitoring sites as shown in the Table A at the end of the document.

$$
\begin{aligned}
b_{ext} \approx\ & \times \times [\quad] + \times \times [\quad] \\
& + \times \times [\quad] + \times \times [\quad] \\
& + \times [\quad] + \times [\quad] \\
& + 10 \times [Elemental\ \ Carbon] \\
& + 1 \times [Fine\ \ Soil] \\
& + \times \times [\quad] \\
& + 0.6 \times [Coarse\ \ Mass] \\
& + \\
& + \times [\quad]
\end{aligned}
$$

The apportionment of the total concentration of sulfate compounds into the concentrations of the small and large size fractions is accomplished using the following equations.

$$[Large\ \ Sulfate] = \frac{[Total\ \ Sulfate]}{20\mu g / m^3} \times [Total\ \ Sulfate],\ \ for\ [Total\ \ Sulfate] < 20\mu g / m^3$$

$$[Large\ \ Sultate] = [Total\ \ Sulfate]\ \ for\ [Total\ \ Sulfate] \geq 20\ \mu g\ /\ m^3$$

$$[Small\ Sulfate] = [Total\ Sulfate] - [Large\ \ Sulfate]$$

The same equations are used to apportion total nitrate and total organic mass concentrations into the small and large size fractions.

Sea salt is calculated as 1.8 x [*Chloride*], or 1.8 x [*Chlorine*] if the chloride measurement is below detection limits, missing or invalid. The algorithm uses three water growth adjustment term as shown in the Figure 2 and Table 1. They are for use with the small size distribution and the large size distribution sulfate and nitrate compounds and for sea salt ($f_S(RH)$, $f_L(RH)$ and $f_{SS}(RH)$ respectively).

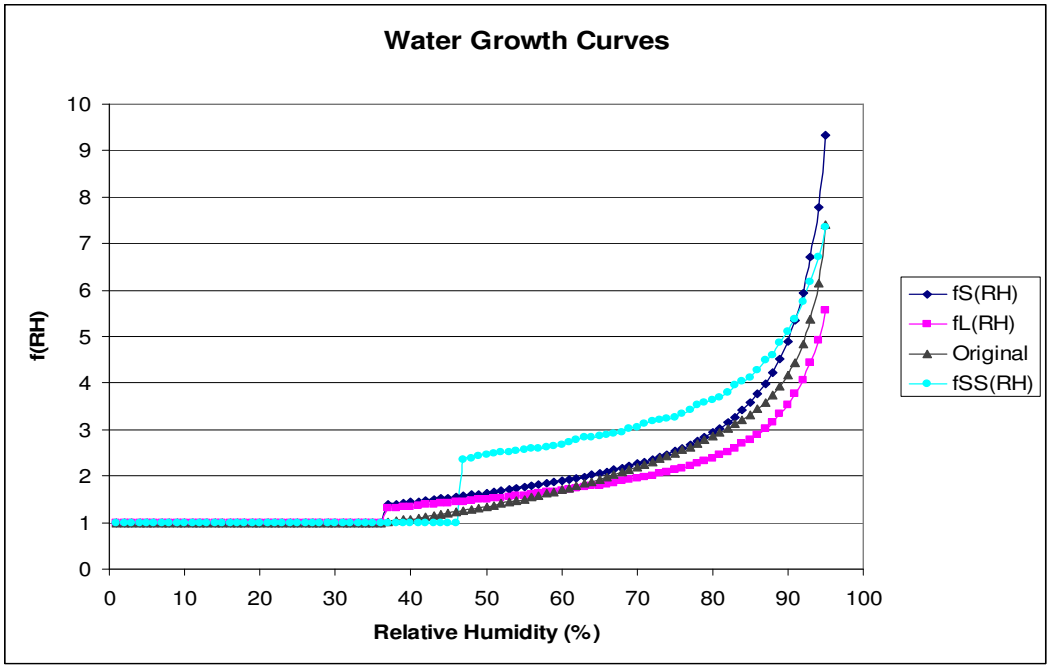

Figure 1. Water growth curves for small and large size distribution sulfate and nitrate, sea salt and the original IMPROVE algorithm sulfate and nitrate.

Table 1. f(RH) for small and large size distribution sulfate and nitrate, an sea salt.

RH (%)	$f_S(RH)$	$f_L(RH)$	$f_{SS}(RH)$	RH (%)	$f_S(RH)$	$f_L(RH)$	$f_{SS}(RH)$	RH (%)	$f_S(RH)$	$f_L(RH)$	$f_{SS}(RH)$
0 to 36	1.00	1.00	1.00	**56**	1.78	1.61	2.58	**76**	2.60	2.18	3.35
37	1.38	1.31	1.00	**57**	1.81	1.63	2.59	**77**	2.67	2.22	3.42
38	1.40	1.32	1.00	**58**	1.83	1.65	2.62	**78**	2.75	2.27	3.52
39	1.42	1.34	1.00	**59**	1.86	1.67	2.66	**79**	2.84	2.33	3.57
40	1.44	1.35	1.00	**60**	1.89	1.69	2.69	**80**	2.93	2.39	3.63
41	1.46	1.36	1.00	**61**	1.92	1.71	2.73	**81**	3.03	2.45	3.69
42	1.48	1.38	1.00	**62**	1.95	1.73	2.78	**82**	3.15	2.52	3.81
43	1.49	1.39	1.00	**63**	1.99	1.75	2.83	**83**	3.27	2.60	3.95
44	1.51	1.41	1.00	**64**	2.02	1.78	2.83	**84**	3.42	2.69	4.04
45	1.53	1.42	1.00	**65**	2.06	1.80	2.86	**85**	3.58	2.79	4.11
46	1.55	1.44	1.00	**66**	2.09	1.83	2.89	**86**	3.76	2.90	4.28
47	1.57	1.45	2.36	**67**	2.13	1.86	2.91	**87**	3.98	3.02	4.49
48	1.59	1.47	2.38	**68**	2.17	1.89	2.95	**88**	4.23	3.16	4.61
49	1.62	1.49	2.42	**69**	2.22	1.92	3.01	**89**	4.53	3.33	4.86
50	1.64	1.50	2.45	**70**	2.26	1.95	3.05	**90**	4.90	3.53	5.12
51	1.66	1.52	2.48	**71**	2.31	1.98	3.13	**91**	5.35	3.77	5.38
52	1.68	1.54	2.50	**72**	2.36	2.01	3.17	**92**	5.93	4.06	5.75
53	1.71	1.55	2.51	**73**	2.41	2.05	3.21	**93**	6.71	4.43	6.17

54	1.73	1.57	2.53	**74**	2.47	2.09	3.25	**94**	7.78	4.92	6.72
55	1.76	1.59	2.56	**75**	2.54	2.13	3.27	**95**	9.34	5.57	7.35

Algorithm Performance Evaluation

Performance of the current and proposed new algorithm for estimating extinction can be assessed in a number of ways each of which serves to answer different questions. Reduction of the biases in light scattering estimates at the extremes (i.e. underestimation of the high values and over estimation of the low values) when compared to nephelometer measurements was one of the most compelling reasons for development of a new algorithm, so comparisons of bias for the current and proposed new algorithm are one way to evaluate performance.

The fractional bias for each sample period was calculated as the difference in light scattering (i.e. estimated b_{sp} minus the measured b_{sp}) divided by the measured light scattering. These biases were then averaged in each quintile to indicate the bias in those five subsets of the data from the lowest to the highest light scattering values. Two different approaches to this grouping by quintiles were performed, referred to as criteria 1 and 2.

Criterion 1 used the measured light scattering to determine which sample periods were in each quintile. Since we think of the nephelometer as the better measure of light scattering, bias by this criterion better addresses the question of algorithm performance with regards to the haze conditions. Criterion 2 uses the algorithm-estimated light extinction to determine which sample periods were in each quintile. The Regional Haze Rule index is based on the highest and lowest haze levels as determined by the algorithm, so criterion 2 better addresses the haze rule application of the algorithm. Tables 2 through 5 show the bias results by both criteria for the current and new algorithm for sites averaged by RPO.

Table 2. Averaged fractional bias by RPO for the current IMPROVE algorithm with quintiles based on measured light scattering (criterion 1). Bold font highlights the bias values that are lower than corresponding values in Table 3.

RPO	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Average
CEN	0.67	0.18	0.10	**0.02**	-0.11	0.17
MANE	0.93	0.27	0.19	0.10	**0.01**	0.28
VISTAS	0.59	0.21	0.11	**0.02**	-0.13	0.16
WRAP	1.07	0.37	0.18	0.07	**-0.08**	0.32

Table 3. Averaged fractional bias by RPO for the current new proposed algorithm with quintiles based on measured light scattering (criterion 1). Bold font highlights the bias values that are lower than corresponding values in Table 2.

RPO	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Average
CEN	**0.51**	**0.08**	**0.02**	-0.03	**-0.08**	**0.10**
MANE	**0.74**	**0.14**	**0.06**	**0.01**	-0.02	**0.17**
VISTAS	**0.50**	**0.16**	0.11	0.06	**-0.01**	0.16
WRAP	**0.84**	**0.25**	**0.08**	**0.01**	-0.10	**0.21**

Table 4. Averaged fractional bias by RPO for the current IMPROVE algorithm with quintiles based on estimated light scattering (criterion 2). Bold font highlights the bias values that are lower than corresponding values in Table 5.

RPO	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Average
CEN	0.42	0.19	0.18	0.09	**-0.01**	0.17
MANE	0.35	0.27	0.30	0.33	**0.15**	0.28
VISTAS	0.38	0.19	0.14	**0.08**	**0.01**	0.16
WRAP	0.58	0.37	0.26	0.25	**0.15**	0.32

Table 5. Averaged fractional bias by RPO for the current new proposed algorithm with quintiles based on estimated light scattering (criterion 2). Bold font highlights the bias values that are lower than corresponding values in Table 4.

RPO	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Average
CEN	**0.23**	**0.11**	**0.08**	**0.06**	0.02	**0.10**
MANE	**0.15**	**0.14**	**0.15**	**0.19**	0.21	**0.17**
VISTAS	**0.25**	**0.14**	**0.08**	0.13	0.21	0.16
WRAP	**0.37**	**0.19**	**0.19**	**0.16**	0.17	**0.21**

These tables show that the new algorithm has lower fractional bias than the current IMPROVE algorithm in all but the haziest conditions (i.e. quintile 5) regardless of the criterion used to sort the data into quintiles. By criterion 1, the two algorithms perform about the same for haziest days except for the sites in the southeastern U.S. (i.e. the VISTA RPO), where the new algorithm has much lower bias (1% compared to 13%). Using criterion 2, the current algorithm has consistently lower bias compared with the new algorithm for the haziest days (i.e. quintile 5) for each of the RPOs. This seeming paradox is the result of the somewhat greater imprecision of the new algorithm compared to the current algorithm, which results in somewhat larger errors in selecting worst haze sample periods for the new algorithm compared with the current algorithm.

Scatter plots (Figures 1 and 3) of light scattering estimates from the current and new proposed algorithms versus nephelometer data for all available data at 21 monitoring sites are one way to view the overall performance differences between the two. These figures show that the bias at the extremes is reduced using the new algorithm compared to the original IMPROVE algorithm (i.e. the points tend to be better centered on the one-to-one line). They also show that the somewhat reduced precision of the new algorithm compared to the original IMPROVE algorithm (i.e. points are more broadly scattered).

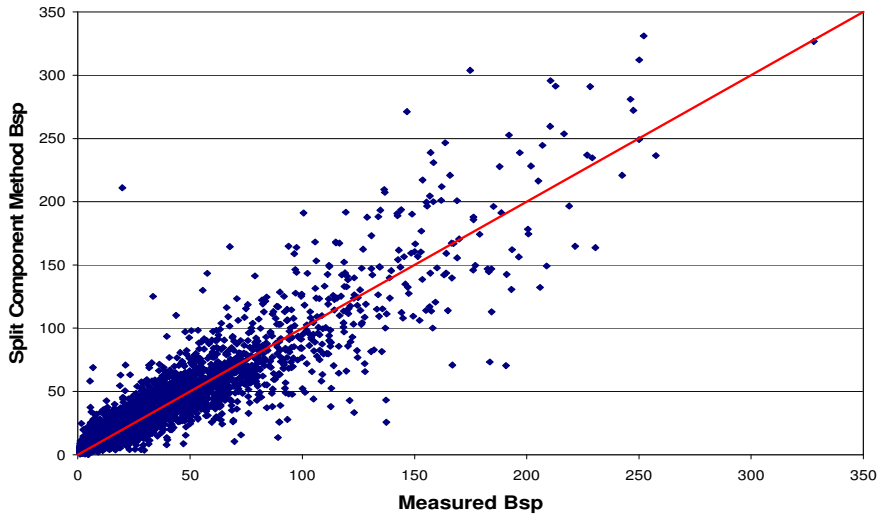

Figure 2. Scatter plot of the recommended revised algorithm estimates of light scattering versus measured light scattering.

Similar pairs of scatter plots were prepared for each individual monitoring site (available in the appendix). Figures 4 and 5 are example plots for Shenandoah and Grand Canyon National Parks. The logarithmic scales on these plots exaggerate the scatter for low values compared to high values. The individual-site scatter plots have the 80th percentile values indicated on the graphs for the predicted and measured values by horizontal and vertical lines respectively. Points that are to the right of the vertical line have nephelometer values that are among the 20% worst light scattering for that monitoring sites. Points that are above the horizontal line have algorithm determined values that are among the 20% worst estimated light scattering for that monitoring site.

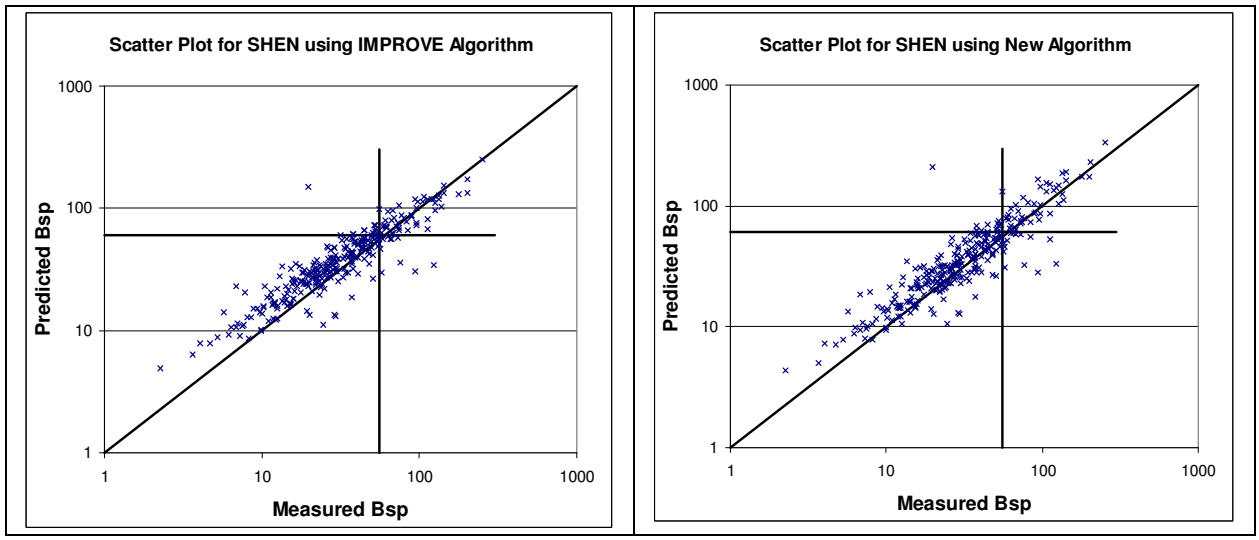

Figure 4. Scatter plots of the current IMPROVE and new recommended algorithm estimates of light scattering versus measured light scattering for Shenandoah National Park. Horizontal and vertical lines are at the 80th percentile for estimated and measured light scattering.

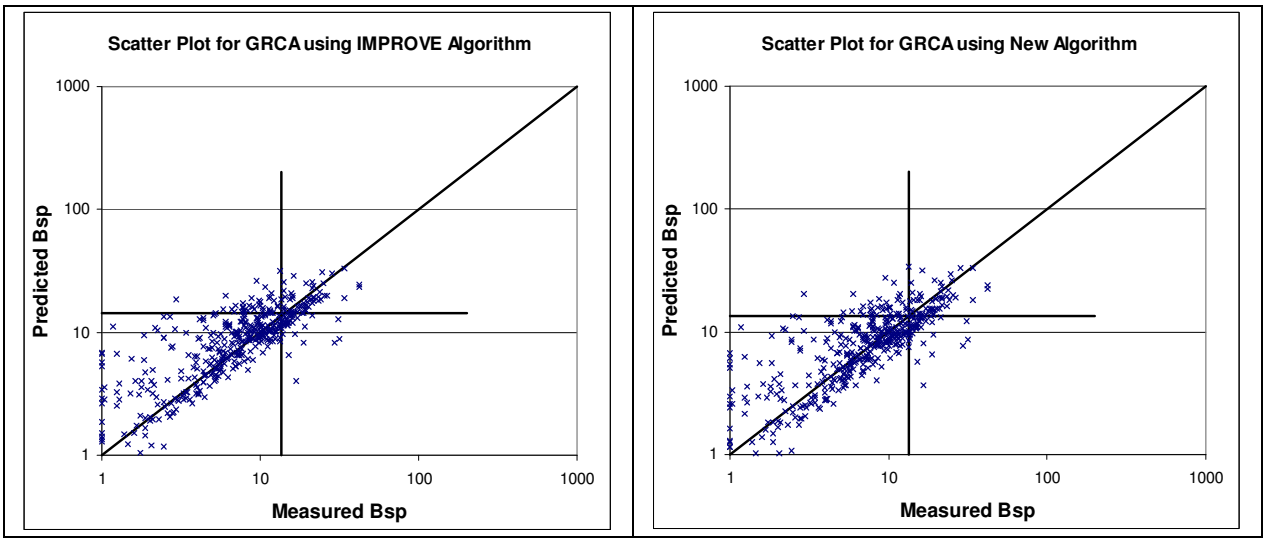

Figure 5. Scatter plots of the current IMPROVE and new recommended algorithm estimates of light scattering versus measured light scattering for Grand Canyon National Park. Horizontal and vertical lines are at the 80th percentile for estimated and measured light scattering.

The proposed new algorithm performs noticeably better with respect to having data points more centered on the one-to-one line at the high and low haze level extremes than the current IMPROVE algorithm for Shenandoah National Park, which is typical for the high haze level locations in the southeast U.S. A large number of the measured worst haze sample periods are correctly identified by both algorithms (these are the points above and to the right of the two 80th percentile lines). The differences between the two algorithms for Grand Canyon National Park and most of the other less hazy locations are not apparent in these scatter plots.

The final approach for evaluating the relative performance of the two algorithms is to compare the average composition of the best haze days and the worst haze days as selected using each algorithm and using the measured light scattering. Table 6 and 7 contain the average composition by RPO for days selected as best and worst by these three methods. Similar results for each of the 21 nephelometer monitoring locations are shown in tables in the appendix.

Table 6. Mean light scattering and percent $PM_{2.5}$ composition for the five major components for 20% best days as determined by measurement, the current IMPROVE algorithm and the proposed new algorithm.

RPO		Mean Bsp (Mm-1)	Percent Ammonium Sulfate	Percent Ammonium Nitrate	Percent OCM	Percent Soil	Percent Coarse	Percent EC
	Measured	6.8	19	2	15	8	54	2
CENRAP	IMPROVE	9.1	20	2	15	7	53	2
	NEW	8.1	21	3	16	7	51	2
	Measured	6.1	22	3	22	4	47	3
MANEVU	IMPROVE	8.4	21	3	22	4	47	3
	NEW	7.4	22	4	22	4	45	3
	Measured	13.8	25	7	21	4	40	3
VISTAS	IMPROVE	18.4	25	7	21	4	40	3
	NEW	17.0	25	8	21	4	39	3

	Measured	3.4	13	3	18	8	55	3
WRAP	IMPROVE	5.2	14	3	19	8	53	3
	NEW	4.5	15	3	19	8	52	3

Table 7. Mean light scattering and percent $PM_{2.5}$ composition for the five major components for 20% worst days as determined by measurement, the current IMPROVE algorithm and the proposed new algorithm.

RPO		Mean Bsp (Mm-1)	Percent Ammonium Sulfate	Percent Ammonium Nitrate	Percent OCM	Percent Soil	Percent Coarse	Percent EC
	Measured	76	34	6	19	5	34	2
CENRAP	IMPROVE	67	34	6	19	5	33	2
	NEW	72	34	6	19	6	34	2
	Measured	61	36	6	23	3	30	3
MANEVU	IMPROVE	61	36	6	22	3	30	3
	NEW	63	35	6	23	3	31	3
	Measured	120	46	5	21	3	22	2
VISTAS	IMPROVE	106	47	4	21	3	23	2
	NEW	127	47	3	22	3	22	2
	Measured	36	15	6	27	7	42	3
WRAP	IMPROVE	33	14	6	27	6	44	3
	NEW	33	13	6	27	6	44	3

These tables demonstrate that the composition associated with the best and worst haze days are not very sensitive to the method of identifying the sample periods that fit in best and worst categories. Some of the individual sites (e.g. Grand Canyon) have somewhat larger variations in the composition between measurement-selected days compared to algorithm-selected days, though there's little difference between the average composition comparing the two algorithms on the best and worst days. The contributions to light extinction by the various components were not explicitly calculated, but are inherently somewhat different because of the explicit differences in the two algorithms.

In summary, the proposed new algorithm for estimating haze reduces the biases compared to measurements at the high and low extremes. This is most apparent for the hazier eastern sites. The composition of days selected as best and worst by the current and the new algorithm are very similar, and similar to days selected by measurements. Most of the reduction of bias associated with the new algorithm is attributed to the use of the split component extinction efficiency method for sulfate, nitrate and organic components that permitted variable extinction efficiency depending on the component mass concentration. Though not subject to explicit performance testing, the proposed new algorithm also contains specific changes from the current algorithm that reflect a better understanding of the atmosphere as reflected in the more recent scientific literature (e.g. change to 1.8 from 1.4 for organic compound mass to carbon mass ratio) and a more complete accounting for contributors to haze (e.g. sea salt and NO_2 terms), and use of site specific Rayleigh scattering terms to reduce elevation-related bias.