

Impact of Missing Data on Worst Days at Midwest Northern Class 1 Areas

Donna Kenski

Midwest RPO

March 12, 2007 (revised 6/19/07)

Midwest RPO has identified a number of days during 2000-2004 where data capture at the Class 1 monitors was incomplete (typically, coarse mass and soil are missing species). The missing data causes the days to be excluded from the IMPROVE regional haze baseline calculations. However, the light extinction due to the remaining measured species is significant (above the 80th percentile). If these days were included in the baseline worst days, they would have a small but measurable effect on the average worst-day deciview calculation.

Table 1 summarizes these days, the measured and missing species, and the light extinction values for the non-missing components. Similarly, Table 2 summarizes the impact on 5-year baseline values if these days are retained in the regulatory dataset. Values increase from 0.25 to 0.88 dv. Coarse mass and soil are the components most frequently missing, yet they account for a small fraction of light extinction in this region of the country. It makes sense to include these days in our baseline calculations, even though the RH guidelines call for them to be excluded, because they appear to be largely dominated by anthropogenic sources. Only one of these days is driven by high organic carbon, which might indicate nonanthropogenic aerosol from wildfires. Model evaluation and control programs need to target precisely this subset of days that have the highest anthropogenic influence. Retaining these days in our dataset assures that they will receive adequate scrutiny.

The RPO Data Workgroup, which includes representatives from VIEWS and IMPROVE, have agreed that the guidelines were not meant to exclude these types of days and that retaining them is a more reasonable approach. VIEWS maintains a 'substitute' dataset that is meant to incorporate various changes to the 'regulatory' dataset, and this is a candidate change we propose to be incorporated. This substitute dataset includes other changes like the use of Voyageurs data for some missing Boundary Waters data, for example.

Table 1. Northern Class 1 Areas: Days that exceed p80, despite missing components (2000-2004)

site	date	Esoil	Eamm_no3	ELAC	ECM	Eamm_so4	EOMC	p80	dv
BOWA1	1/15/2000		75.4	4.1		20.2	8.1	16.7	24.7
BOWA1	7/22/2002	0.5		2.2	3.8	40.8	8.7	15.9	19.0
BOWA1	9/8/2002		2.5	4.4		118.5	19.8	15.9	27.5
BOWA1	9/9/2003	0.5		2.7	3.2	28.7	9.4	16.0	17.1
BOWA1	9/12/2003	0.3		1.8	2.0	29.8	8.2	16.0	16.7
BOWA1	3/25/2004		44.5	3.2		28.7	21.4	14.7	23.9
ISLE1	2/23/2000		83.0	4.9		33.7	11.4	16.4	26.7
ISLE1	8/5/2000	0.4		3.6	7.9	22.7	13.0	16.4	17.8
ISLE1*	8/12/2000	0.3		2.8	7.8	20.2	10.2	16.4	16.7
ISLE1	3/29/2001		53.8	4.0		36.7	6.9	18.0	24.3
ISLE1	4/1/2001		5.2	3.1		53.6	5.2	18.0	20.7
ISLE1	9/8/2002		1.6	4.1		132.9	18.9	16.9	28.3
ISLE1	2/26/2003	0.5	61.4		4.4	16.2		15.9	22.5
ISLE1	3/16/2003		140.3	6.2		51.6	12.3	15.9	31.0
ISLE1	7/26/2003		3.8	5.8		50.1	21.3	15.9	22.3
ISLE1	8/19/2003		3.4	5.4		62.0	21.5	15.9	23.4
ISLE1	9/9/2003		3.7	5.4		88.1	15.4	15.9	25.2
ISLE1	9/12/2003		2.2	6.7		299.7	11.7	15.9	35.0
ISLE1	3/25/2004		47.6	5.2		58.0	15.5	15.7	26.3
SENE1	3/22/2000	0.2	25.7		0.8	48.3		19.5	21.6
SENE1	12/12/2001		105.9	5.8		60.8	17.7	22.0	30.1
SENE1	9/8/2002		4.1	6.4		351.6	19.5	21.6	36.7
VOYA2	1/15/2000		67.2	3.8		17.7	8.5	17.1	23.9
VOYA2	8/25/2003		0.6	6.7		17.8	57.5	16.4	22.5
VOYA2	3/25/2004		64.9	4.8		32.6	21.8	15.4	26.1

*This day is above the 80th percentile as calculated by RHR Guidelines, but below the new 80th percentile when recalculated including days with missing data.

Table 2. Average of 20% worst days, with and without missing data days

	Average Worst Day DV, per RHR	Average Worst Day DV, with Missing Data Days ^a	Difference
BOWA	19.59	19.86	0.27
ISLE	20.74	21.59	0.85
SENE	24.16	24.38	0.22
VOYA	19.27	19.48	0.21

^aDays were selected for inclusion in this revised baseline if the extinction resulting from the sum of measured species (despite missing species) was equal or greater than the 80th percentile value as calculated according to the Regional Haze Tracking Progress Guidelines. A new 80th percentile was then calculated based on the revised dataset.

To make these calculations, two data files were combined. Data for Boundary Waters are the substituted values from Scott Copeland's analysis, given in the file bowav_nia_substituted.xls found on VIEWS

(<ftp://vista.cira.colostate.edu/public/airquality/data/aerosol/improve/substitute/bowa/>). Data for other sites are from the VIEWS raw data file, daily budgets 3_07.csv (ftp://vista.cira.colostate.edu/public/IMPROVE_RHR_Budgets_2005).