RMSE for information content metric of IMPROVE data for RHR Trends

***DRAFT – This information is not endorsed by
the IMPROVE Steering Committee***

Scott Copeland

10/7/13

Fundamental Question:
What is the value in monitoring spatial density vs. temporal frequency vs. particulate parameters for tracking RHR trends in CIAs?

Answer: We need a metric that can be compared uniformly throughout the network and normalized to cost savings.

Always Remember the Big Picture

Excerpts from Poirot, 2012
Ancillary Uses of IMPROVE Monitoring Data – Potentially Affected by Funding Reductions

A. Development and refinement of sampling and analytical methods [1-15]

B. Assessment of sampling, analytical and data processing artifacts, errors and uncertainties [16-31]

C. Comparison, evaluation and synthesis of methods and data across measurement networks [32-39]

D. Characterization of aerosol formation mechanisms, composition and morphology [40-57]

E. Improved understanding of aerosol vs. optical relationships [58-89]

F. Assessment of long-term temporal and spatial patterns and trends [90-107]

G. Regional and historical background for short-term intensive field and/or tracer studies [108-130]

H. Performance evaluation and refinement of regional and global air quality models [131-153]

I. Input data for multivariate mathematical and/or back trajectory receptor models [154-195]

J. Evaluation of "natural" source impacts (smoke, dust, sea salt, etc.) and regional air quality events [196-215]

K. Assessment of sources, composition, optical & radiative properties of carbonaceous aerosols [216-233]

L. Assessment of transboundary and intercontinental aerosol transport influences [234-245]

M. Comparisons to and synthesis with remote sensing, modeling and surface observation data [246-251]

N. Inverse modeling / development, confirmation and refinement of emissions estimates [252-257]

O. Assessment of human health and/or environmental impacts of specific aerosol species [258-267]

P. Assessment of sources of potentially toxic trace elements [268-274]

Q. Evaluation of single source impacts and control strategies [275-278]

Excerpts from Moore, 2012
## *Data Requirements for Tracking Progress Under the Regional Haze Rule*

Step	Responsible Agency	Reference *	Comment
Choice of Deciview as metric	n/a	40 CFR 51.308 (d)(1) GTP Sec. 1.6	Identifies deciview as tracking metric, but does not define it mathematically.
Identify best/worst days	IMPROVE program	GTP Sec. 2.2, Step 8	The "best" and "worst" days are here defined as the cleanest (lowest) 20% and dirtiest (highest) 20% daily deciview values, per complete year (as determined in GTP Step 7).
		GTP Sec. 4.2	Provides the calculation for determining how to identify the 20% best/worst days.
Calculation of the 5-yr deciview metric	IMPROVE program	GTP Sec. 2.2, Step 9	Calculation of annual average deciviews for best/worst days (as determined in GTP Step 8).
		GTP Sec. 2.2, Step 10	Calculation of 5-yr average deciviews from the annual best/worst average deciviews for complete years (as determined in GTP Step 9). It is these 5-yr average deciview values for best/worst days which are to be used for setting the baseline and tracking progress.
		GTP Sec. 4.3	Repeat of above discussion. Also defines 5-yr progress periods as "2005-2009, 2010-2014, etc."

No Coarse Mass (Remove Module D)

		RHR Haziest dv	No MT Haziest dv	Squared Error
CANY1	2000	11.18	10.09	1.19
CANY1	2001	10.95	10.03	0.83
CANY1	2002	12.69	11.18	2.27
CANY1	2003	11.78	10.43	1.83
CANY1	2004	9.63	8.78	0.72
CANY1	2005	10.56	9.53	1.05
CANY1	2006	10.51	9.26	1.54
CANY1	2007	11.39	10.42	0.95
CANY1	2008	11.12	8.73	5.72
CANY1	2009	11.51	9.61	3.60
CANY1	2010	10.57	9.27	1.69
CANY1	2011	9.80	8.26	2.37
			RMSE	1.407
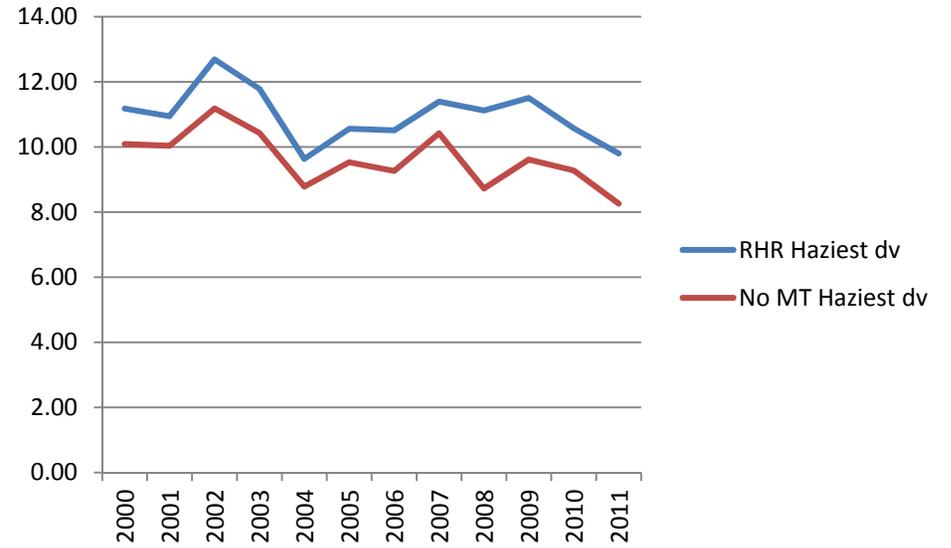
No MT Haziest dv
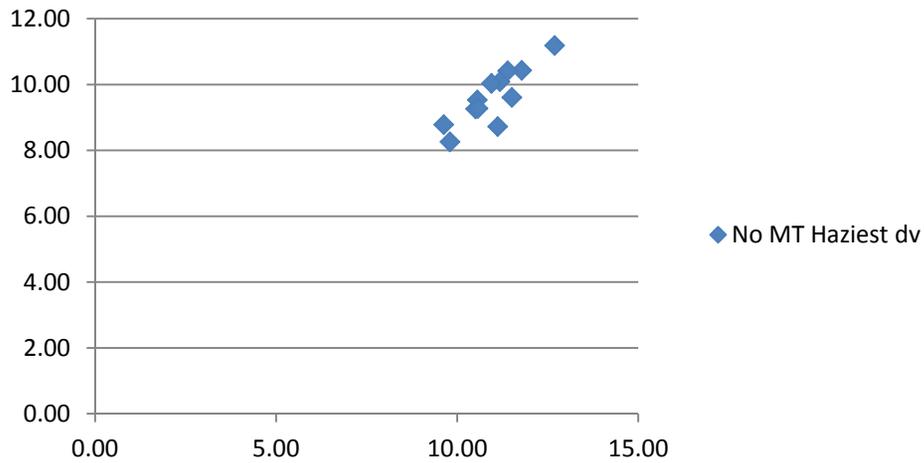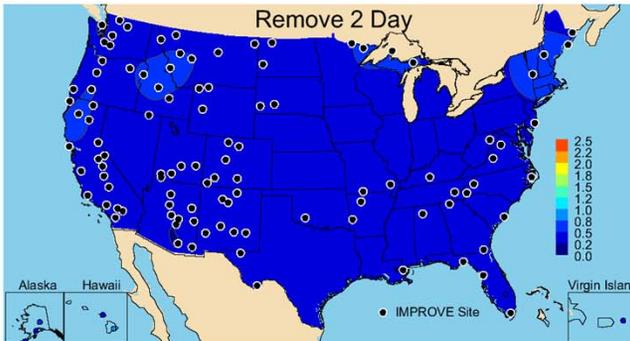
No Coarse Mass (Remove Module D)

site	n	rmse
HAVO1	11	0.126
SHRO1	7	0.138
DOSO1	12	0.140
MACA1	12	0.148
LIGO1	10	0.175
SENE1	12	0.182
COHU1	8	0.183
GRSM1	12	0.183
SIPS1	11	0.205
JARI1	11	0.206

Remove n Samples per month at Every Site

site	One_p_rmse	two_p_rmse	three_p_rmse	four_p_rmse	five_p_rmse
ACAD1	0.34	0.52	0.62	0.80	1.01
AGTI1	0.17	0.26	0.37	0.47	0.57
BADL1	0.20	0.29	0.38	0.49	0.61
BALD1	0.22	0.34	0.43	0.60	0.73
BAND1	0.29	0.44	0.62	0.76	0.93
BIBE1	0.23	0.35	0.46	0.56	0.74
BLIS1	0.28	0.43	0.54	0.76	0.92
BOAP1	0.22	0.33	0.45	0.60	0.66
BOWA1	0.37	0.58	0.73	0.98	1.24
BRCA1	0.27	0.40	0.53	0.65	0.82
BRID1	0.23	0.34	0.46	0.58	0.74
BRIG1	0.28	0.40	0.52	0.70	0.80
BRIS1	0.18	0.25	0.34	0.45	0.53

Remove Sites (Error compared to selected nearby site with best regression)

site1	site2	n	rmse
DOME1	RAFA1	7	0.426
GRSM1	COHU1	8	0.427
SIAN1	IKBA1	7	0.552
JARI1	SHEN1	11	0.602
UPBU1	CACR1	8	0.661
UPBU1	HEGL1	9	0.711
SNPA1	MORA1	10	0.822
TONT1	IKBA1	10	0.833
TONT1	SIAN1	6	0.834
MELA1	THRO1	10	0.838
DOSO1	SHEN1	12	0.844
WHRI1	MOZI1	11	0.859
SAMA1	OKEF1	8	0.866

No MT RMSE

Remove 1 Day

Remove 2 Day

Remove 3 Day

Remove 4 Day

Remove 5 Day

Site Pairs RMSE
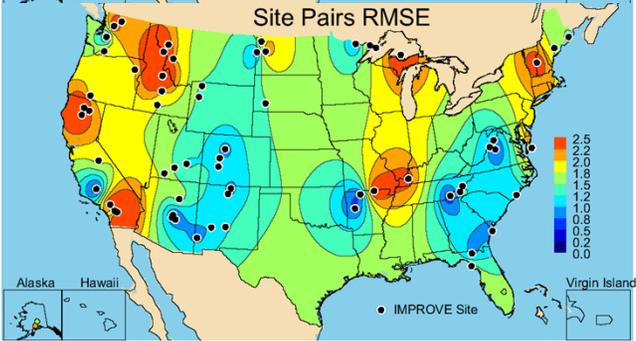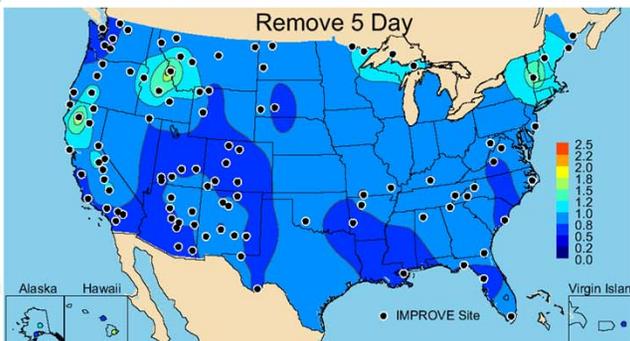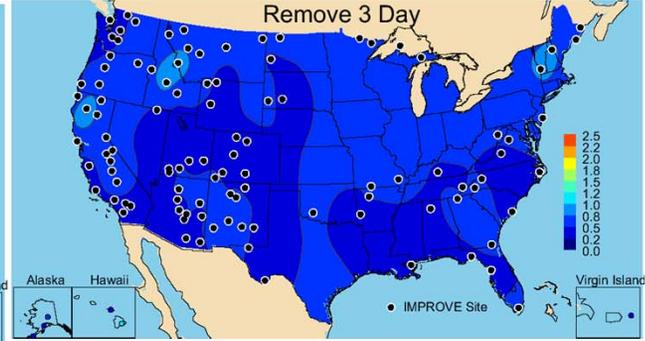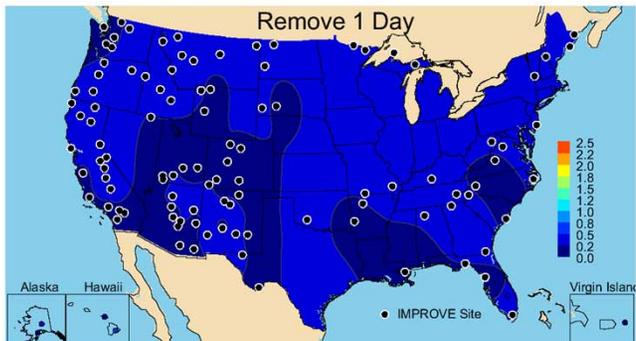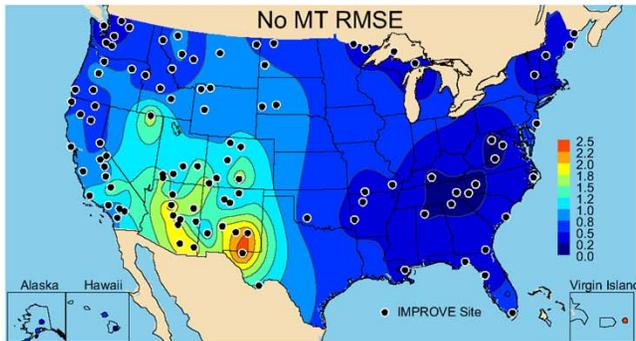
Jenny Hand 2013

ROUGH Projected Cost Savings

Remove 11 IMPROVE sites (very non linear) ≈

Eliminate 1 sample per month network wide ≈

4 * (Remove Module D at 60 sites)