IMPROVE – Data and Metrics

Virtual Meeting, October 14, 2020

Scott Copeland, CIRA/USDA Forest Service

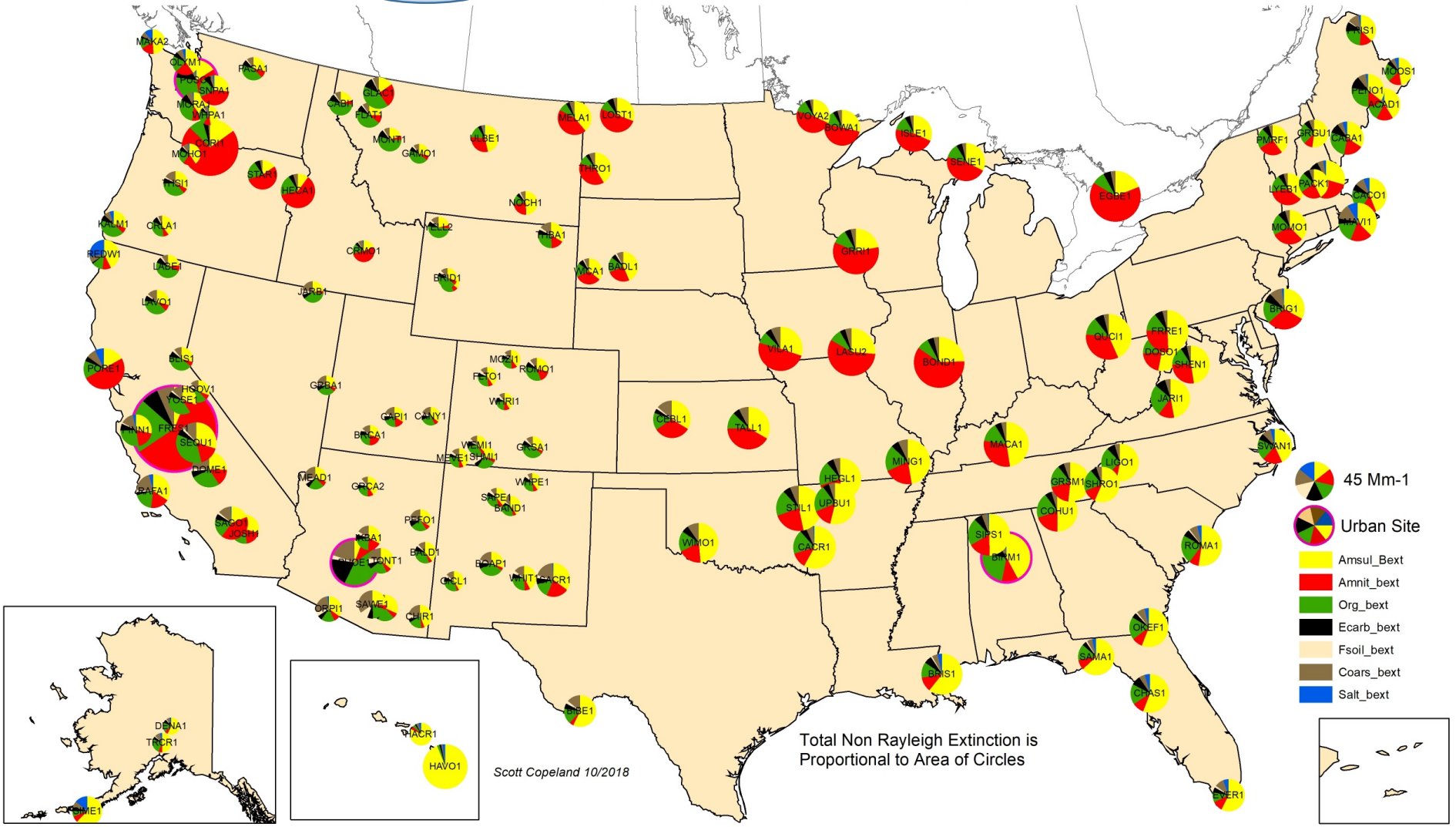
IMPROVE Steering Committee Chair

scott.copeland@colostate.edu

- Preliminary 2019 IMPROVE Data has been delivered to CIRA
- Three data reprocessing efforts since last year's meeting
  -Small data redelivery.
  -Multiple Patching implemented (up to 2 per sample day).
  -WRAP data substitution effort.
- All these changes are very confusing, so beginning with 10/19, all changes to the data processing are documented in a publicly available Powerpoint (appended to this presentation).

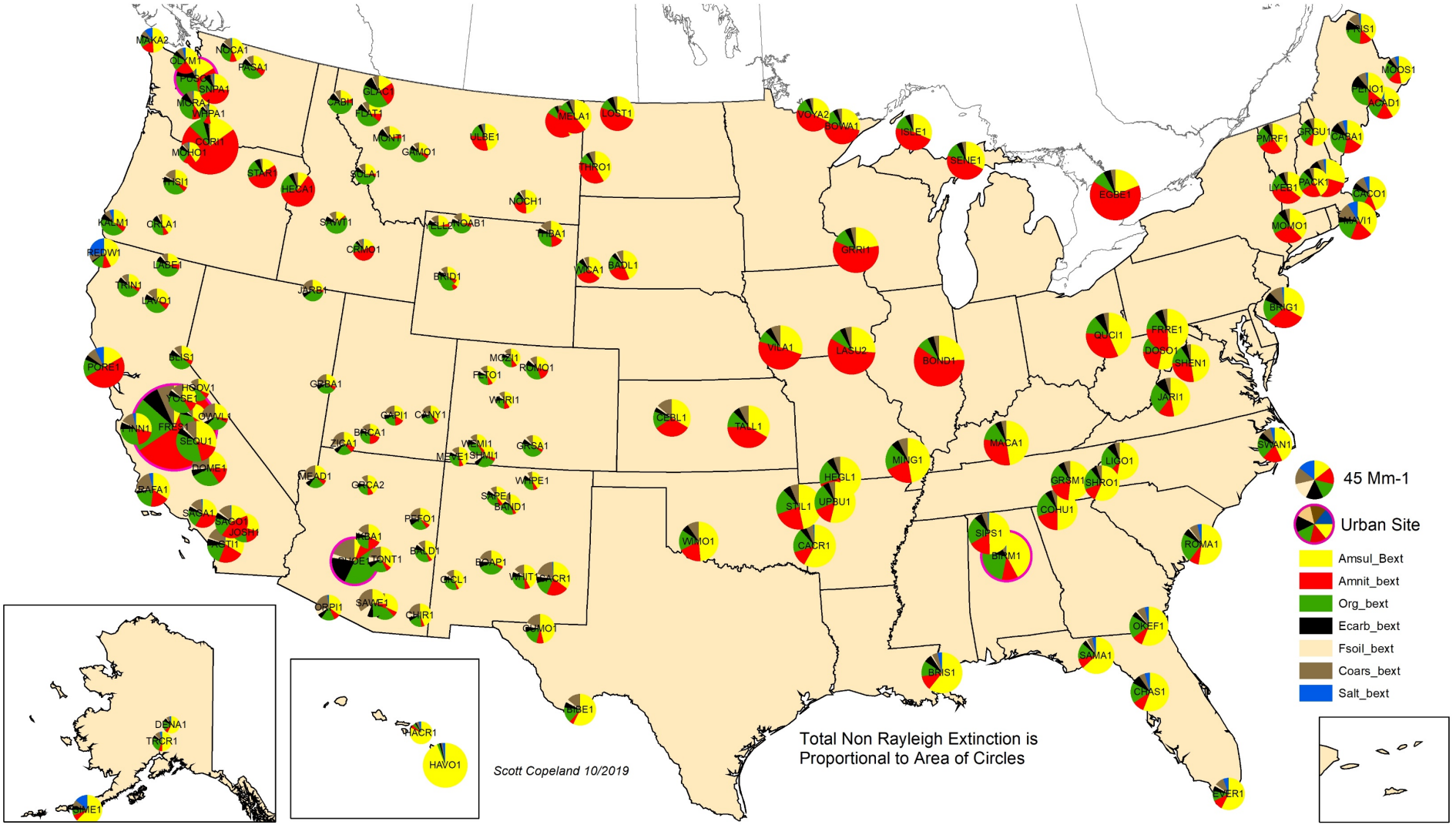
IMPROVE Data - 2017 Second IMPROVE Algorithm

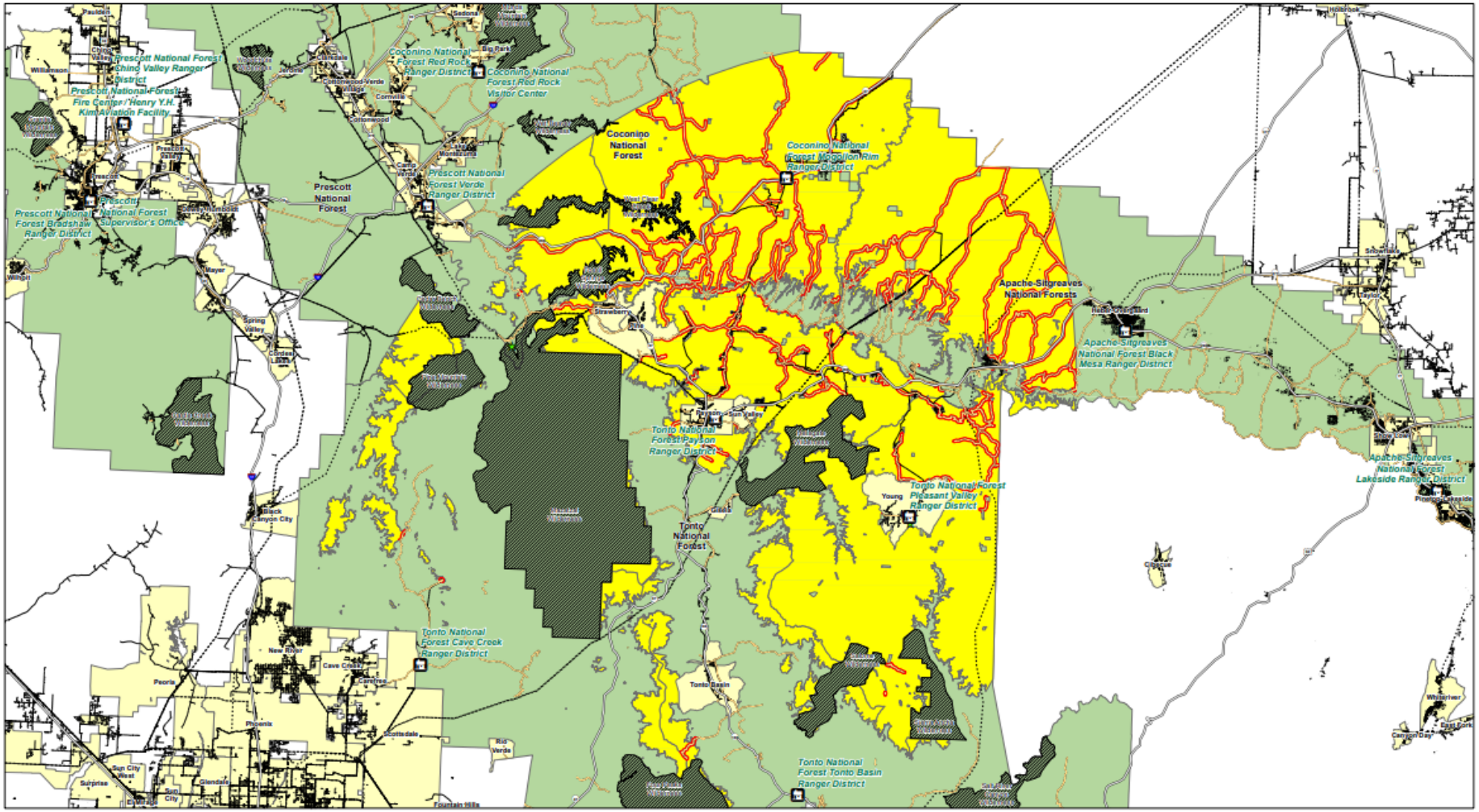
DRAFT Non Rayleigh Mean of 20% Most Impaired

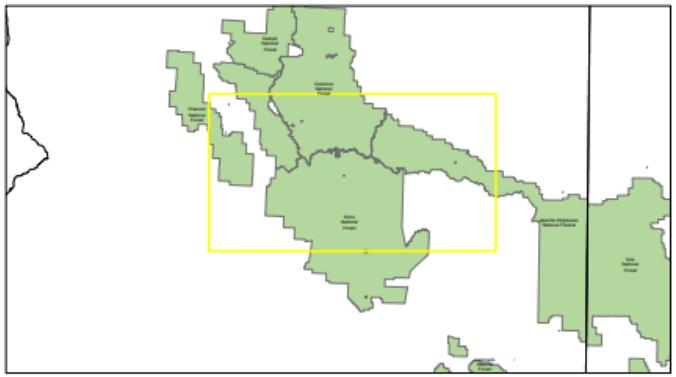


IMPROVE Data - 2017 Second IMPROVE Algorithm

Non Rayleigh Mean of 20% Most Impaired







Possible IMPROVE Site Locations

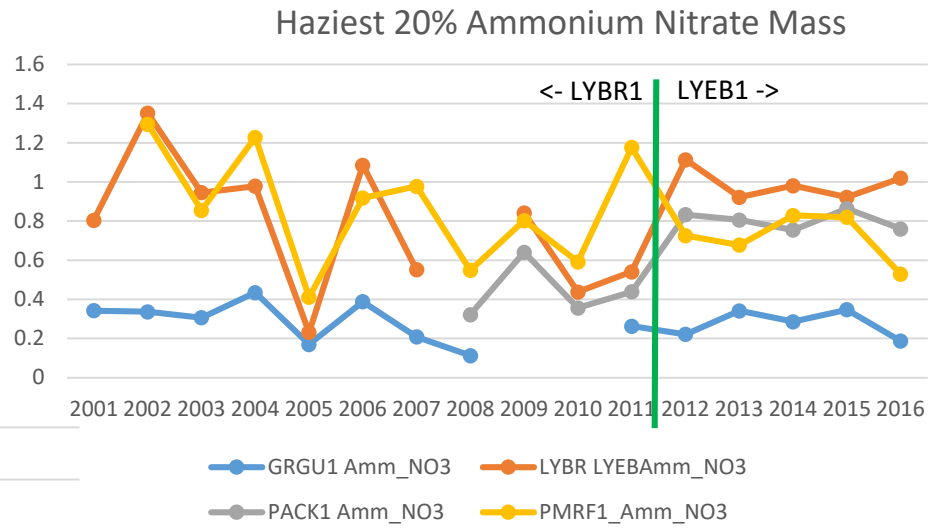
- FS Office Locations
- Current IMPROVE Sites
- Major Roads
- FS Roads - Maintenance Lvl 3-5
- Transmission Power Lines
- Feeder Power Lines
- Cities
- National Wilderness Area
- Possible IMPROVE Site within 800m of Class 3-5 FS Road
- Other Possible IMPROVE Site Locations
- Administrative FS Lands

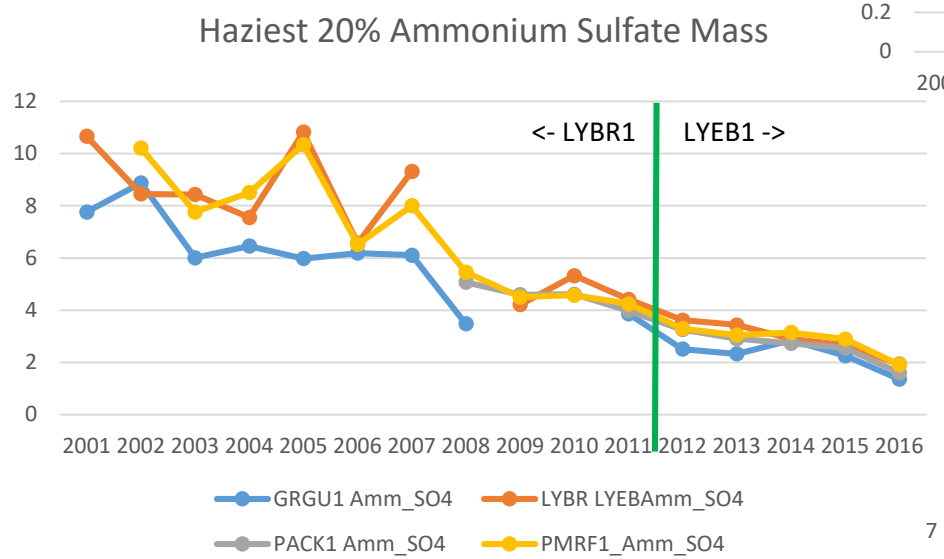
N

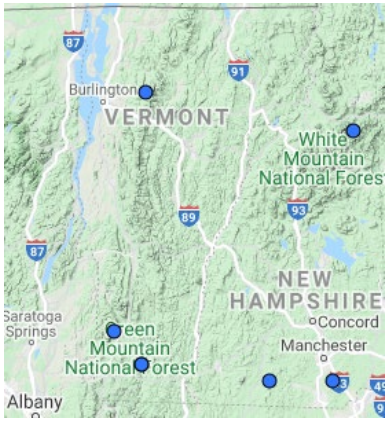
0 10 20 40 Miles

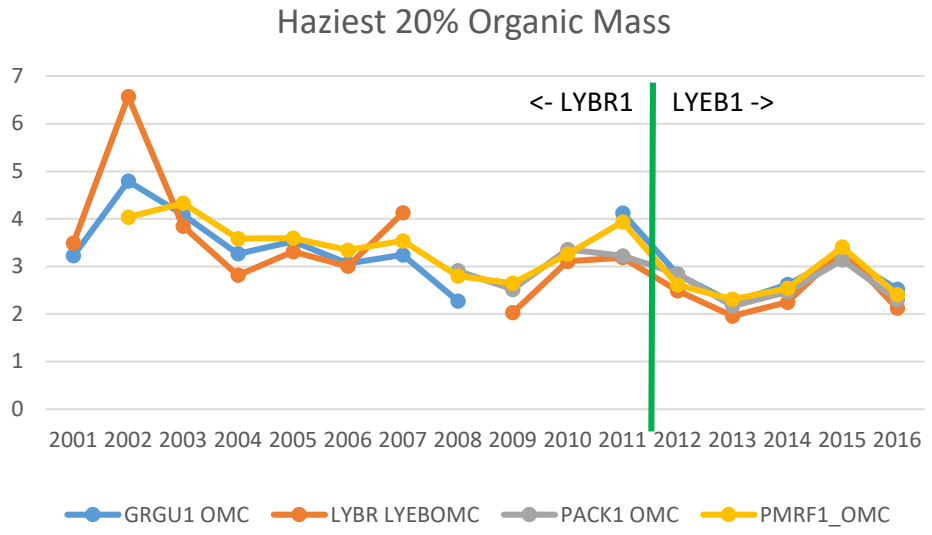
Document Path: T:\FS\NFS\R03\Program\7140Geometronics\GISstaff\Projects\IMPROVE_SiteSelection\MXD\IMPROVE_Analysis_v3.mxd

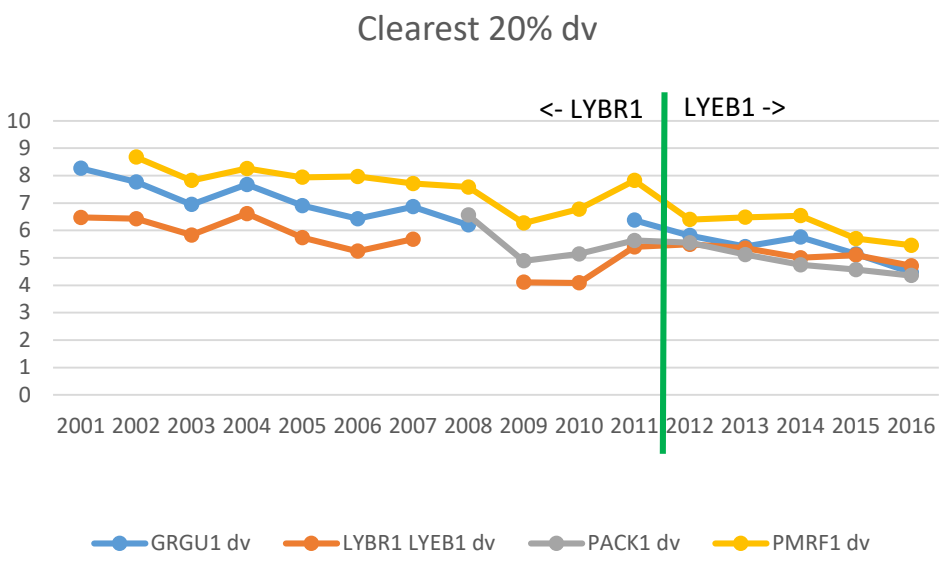
From my Steering Committee meeting presentation in Ely, 2017










Clearest 20% dv
<- LYBR1
LYEB1 ->
10
9
8
7
6
5
4
3
2
1
0
2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016
GRGU1 dv
LYBR1 LYEB1 dv
PACK1 dv
PMRF1 dv

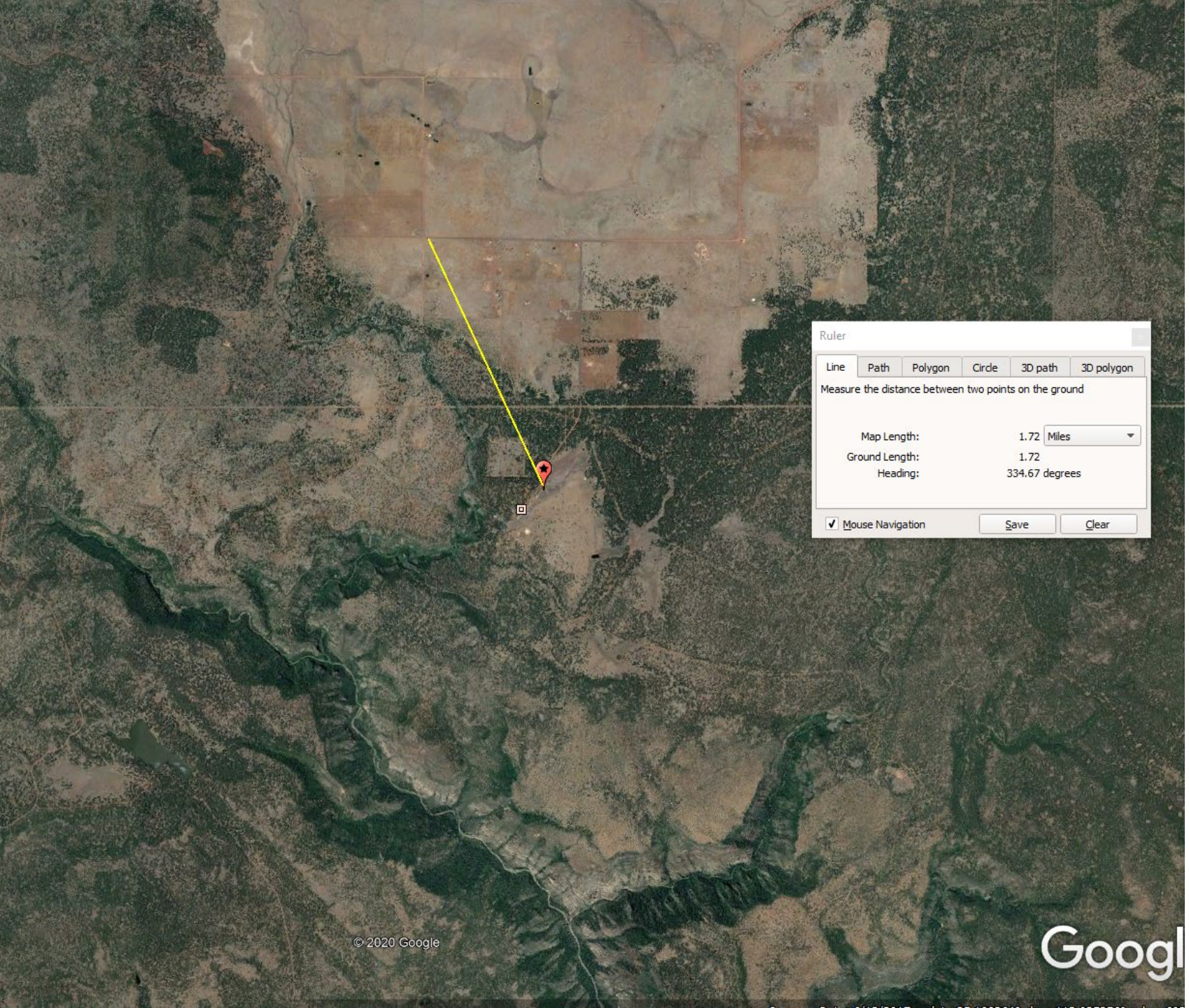
PROCEDURES FOR SITE SELECTION

Prepared by
Crocker Nuclear Laboratory
University of California
One Shields Avenue
Davis, CA 95616-8569
(530) 752-4106

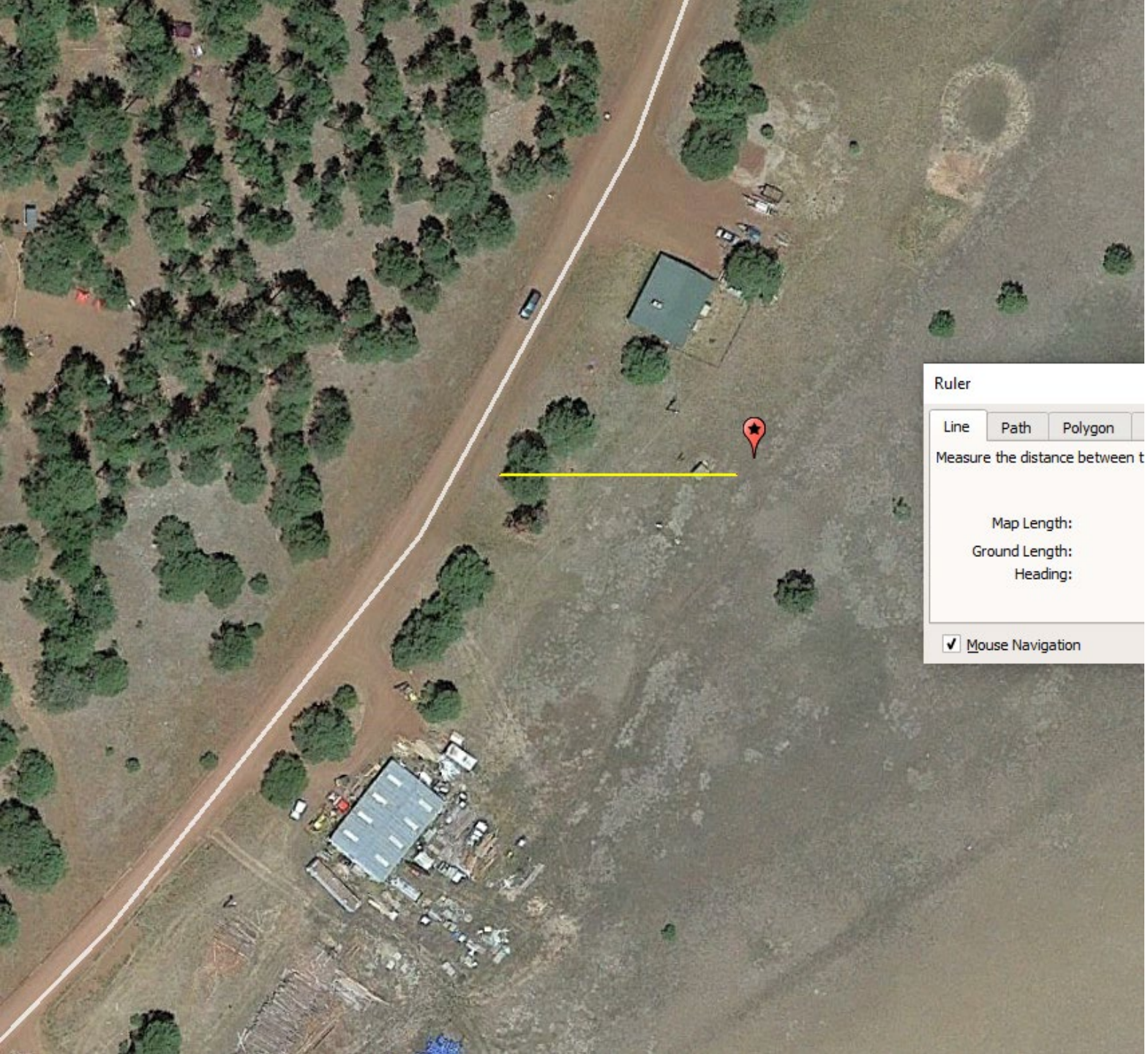
February 24, 1999

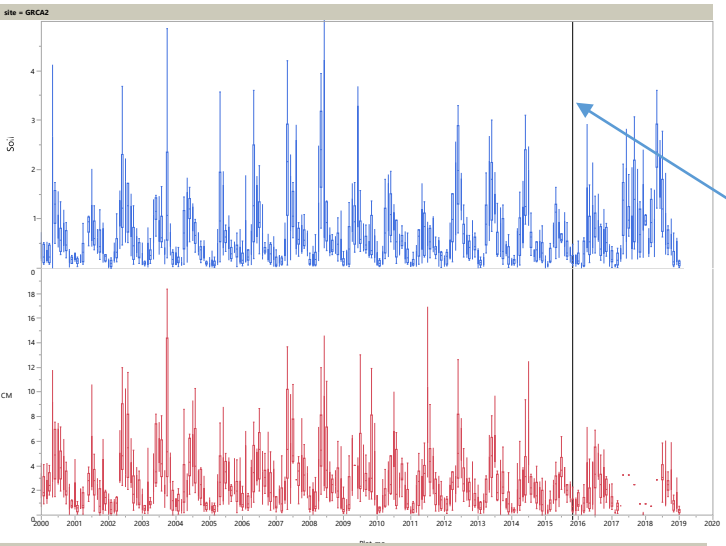
3. The site must avoid all local sources of pollution.
   a. Automotive Sources: vehicle usage, distance between road and sampler
* <10,000 vehicles per day >25m between road and sampler.
* 10,000-20,000 vehicles per day >50m between road and sampler.
* 20,000-40,000 vehicles per day >75m between road and sampler.
* >40,000 vehicles per day >100m between road and sampler.

   b. Combustion Sources
      Avoid any areas influenced by diesel generator emissions, wood smoke, or incinerators.
   c. Dust Sources
      At least 400m from a large potential source of dust, such as a landfill, agricultural operations, or an unpaved road with more than 400 cars per day.


Ruler
Line
Path
Polygon
Circle
3D path
3D polygon
Measure the distance between two points on the ground
Map Length:
1.72
Miles
Ground Length:
1.72
Heading:
334.67 degrees
Mouse Navigation
Save
Clear
© 2020 Google


Thomas Loop Rd
Ruler
Line
Path
Polygon
Circle
3D path
3D polygon
Measure the distance between two points on the ground
Map Length:
44.63
Yards
Ground Length:
44.64
Heading:
269.61 degrees
Mouse Navigation
Save
Clear
527
© 2020 Google


Ruler
Line
Path
Polygon
Measure the distance between t
Map Length:
Ground Length:
Heading:
Mouse Navigation



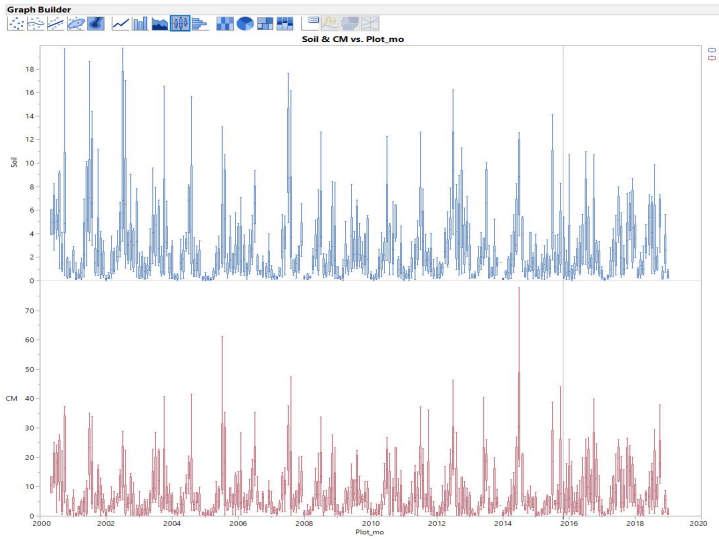
Soil

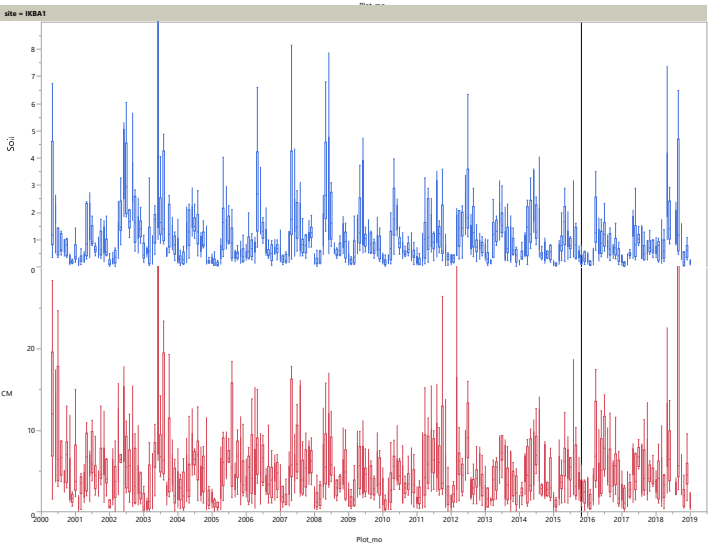
CM

GRCA2

SYCA1 & SYCA2

Vertical Line in each plot indicates end of SYCA1 and begin SYCA2 in 10/2015.



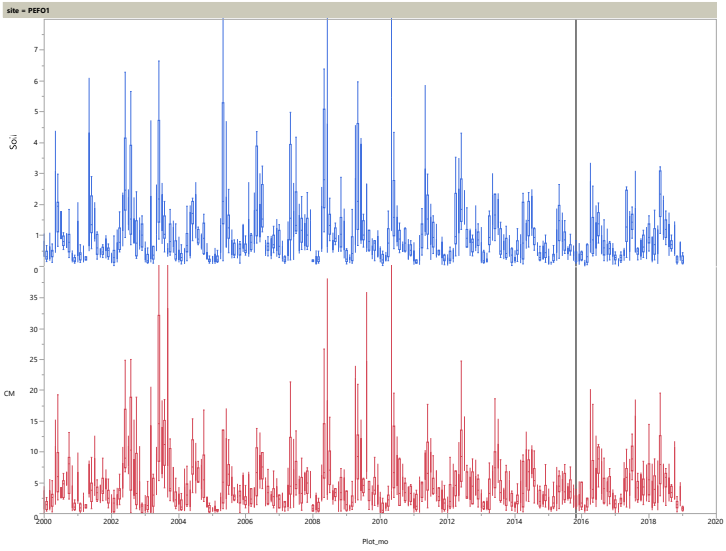


IKBA1

PEFO1

GRCA2, SYCA1&2, and IKBA1 show somewhat similar increases for Soil and CM after the relocation.

Sea salt also compares well across the site relocation (not shown).



Ryan Templeton <Templeton.Ryan@azdeq.gov> | Copeland,Scott; Disha Gadre | 11/6/2019

Re: 2018 IMPROVE Data

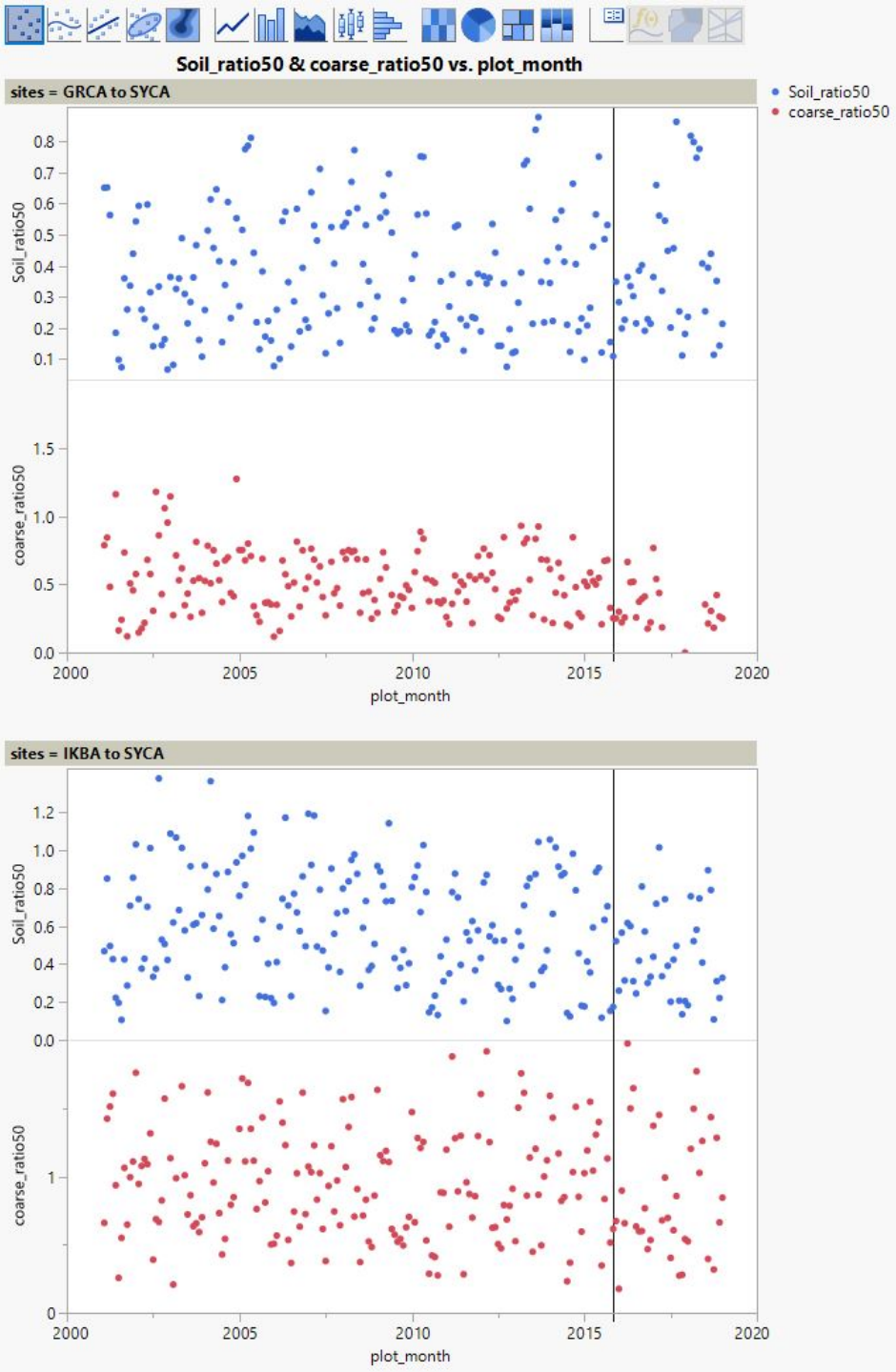
If there are problems with how this message is displayed, click here to view it in a web browser.

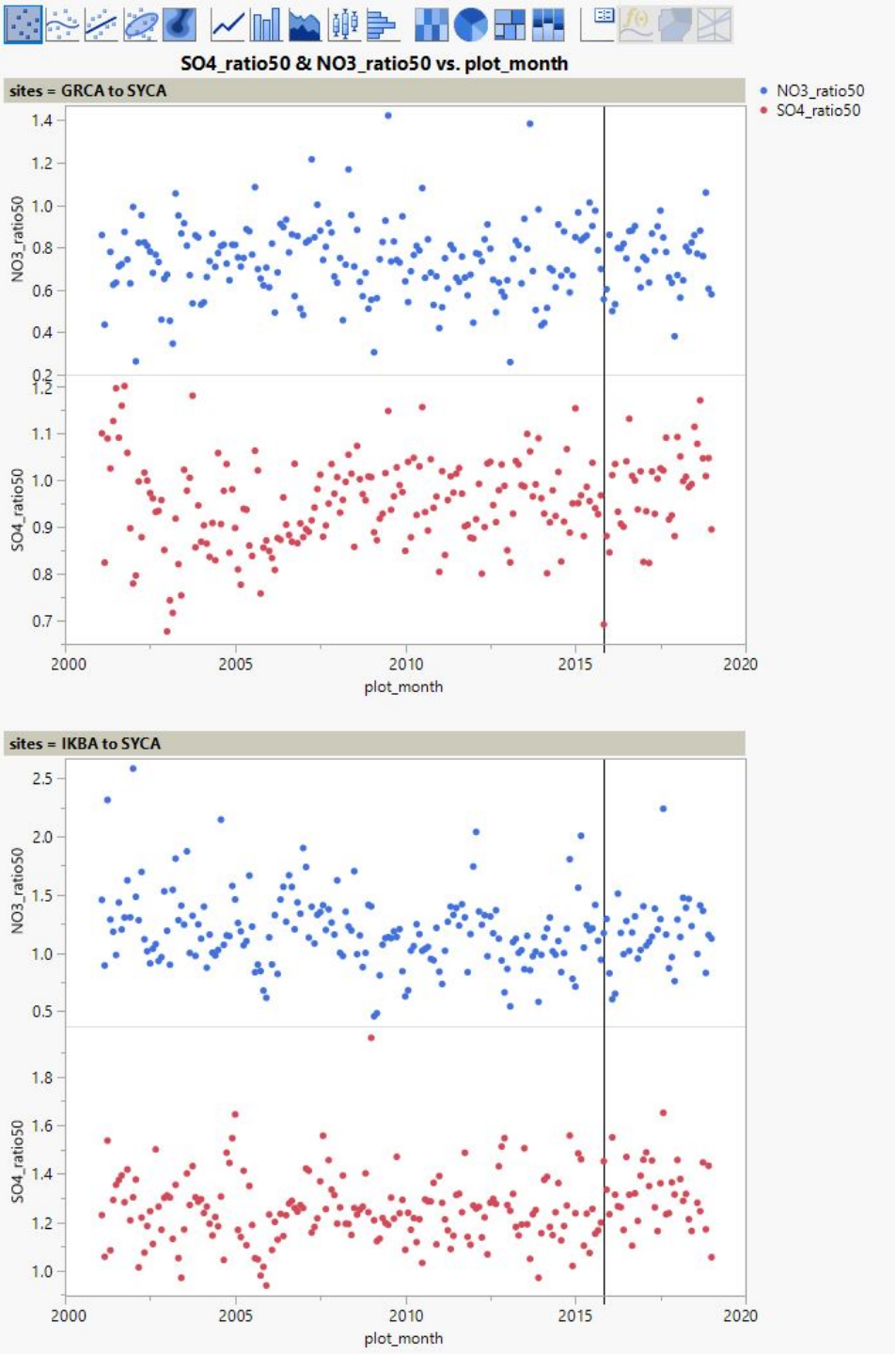
Hey Scott,

That sounds reasonable and in-line with Air Resource Specialist's suggested approach. Please feel free to move forward with combining the 2 sites datasets. Thanks.

Ryan Templeton, P.E.
Senior Environmental Engineer
Ph: 602-771-4230

azdeq.gov


Soil_ratio50 & coarse_ratio50 vs. plot_month
sites = GRCA to SYCA
Soil_ratio50
coarse_ratio50
Soil_ratio50
coarse_ratio50
plot_month
sites = IKBA to SYCA
Soil_ratio50
coarse_ratio50
plot_month

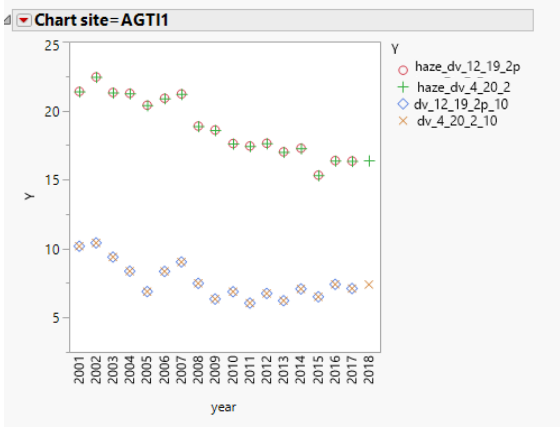

SO4_ratio50 & NO3_ratio50 vs. plot_month
sites = GRCA to SYCA
NO3_ratio50
SO4_ratio50
NO3_ratio50
SO4_ratio50
plot_month
sites = IKBA to SYCA
NO3_ratio50
SO4_ratio50
plot_month

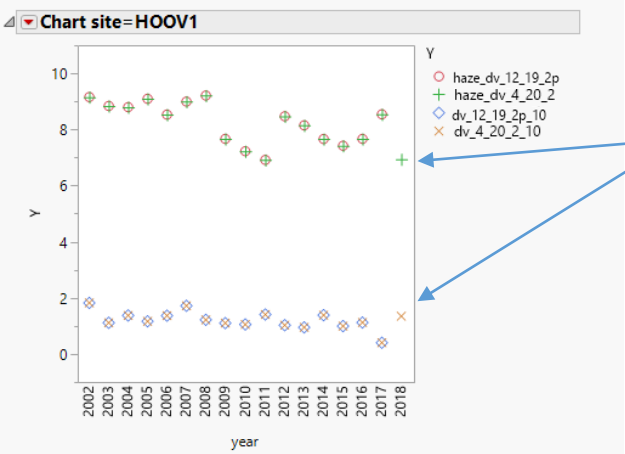
Changes to IMPROVE RHR Calculations and Metrics since 12/2019 version

* *affects AGTI1, HOOV1, IKBA1, WHPE1, SWAN1 and LTCC1 only**

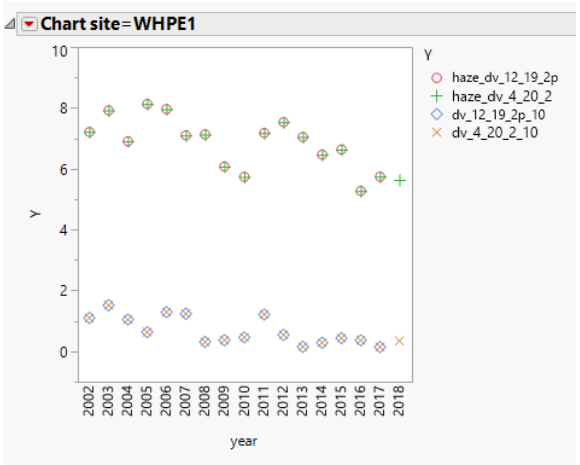
Scott Copeland 4_23_2020

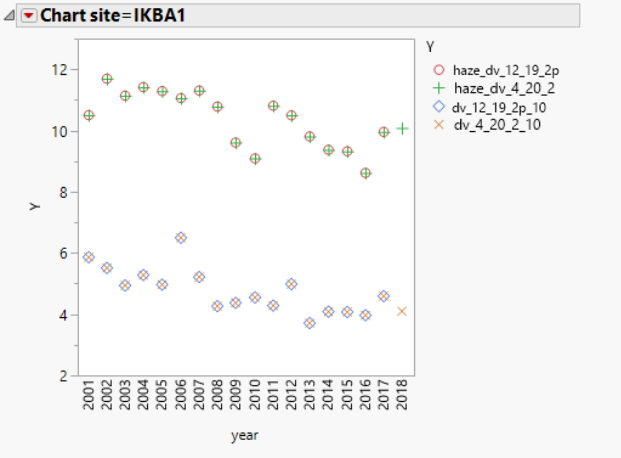
Data Substitution performed by Air Resource Specialists, Inc. for WRAP at four western IMPROVE sites in 2018.





Clearest and Most Impaired Days are now calculated for these years at these sites.





The problem: Insect debris clogged inlets of samplers at Chassahowitzka FL, Mingo MO, and Swanquarter NC in the 2003-2005 timeframe.

- Much data at all three sites was invalidated, but the QA tests failed to find a problem with the module C data at Swanquarter.

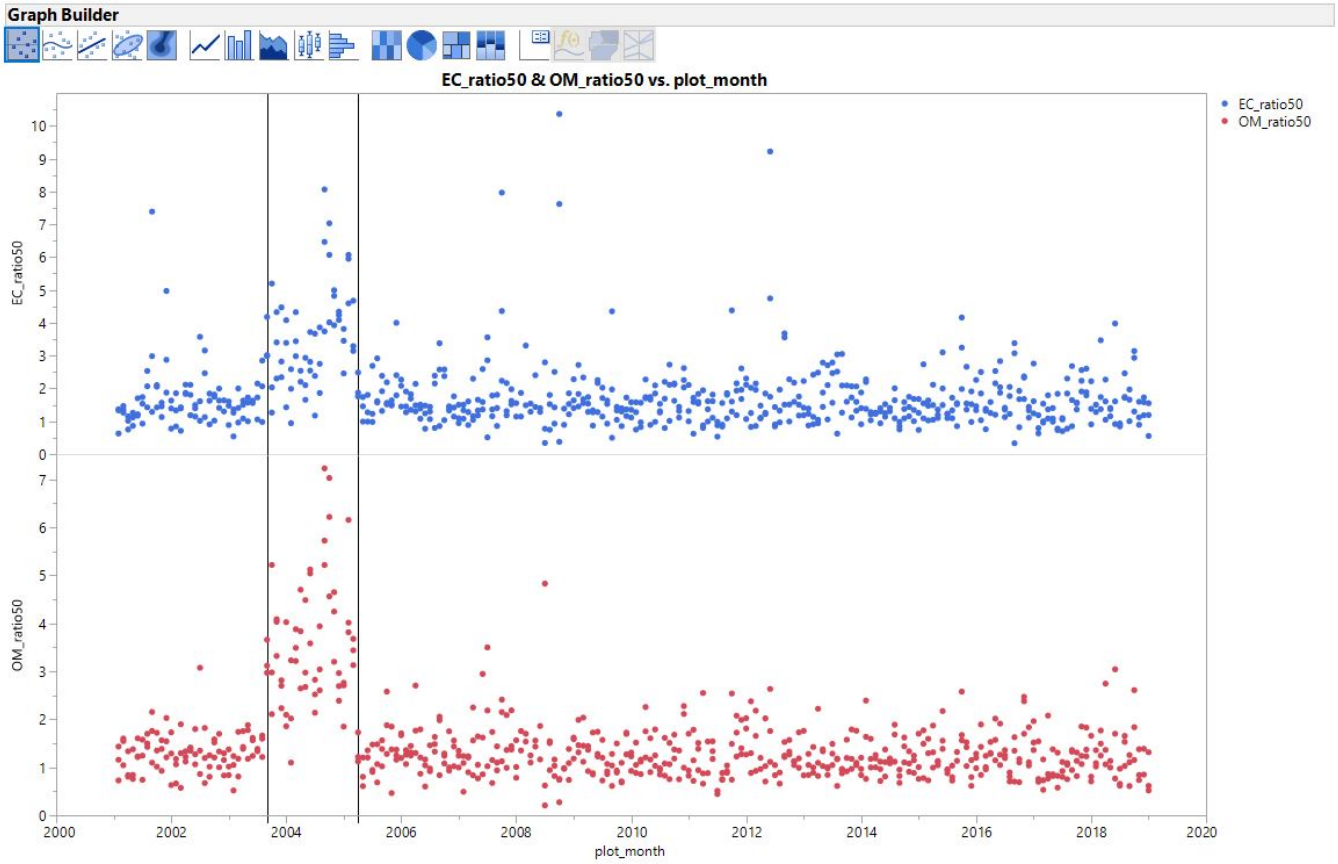
Some inlets clogged with insect debris, discovered spring 2005

CROCKER NUCLEAR LABORATORY
University of California, Davis

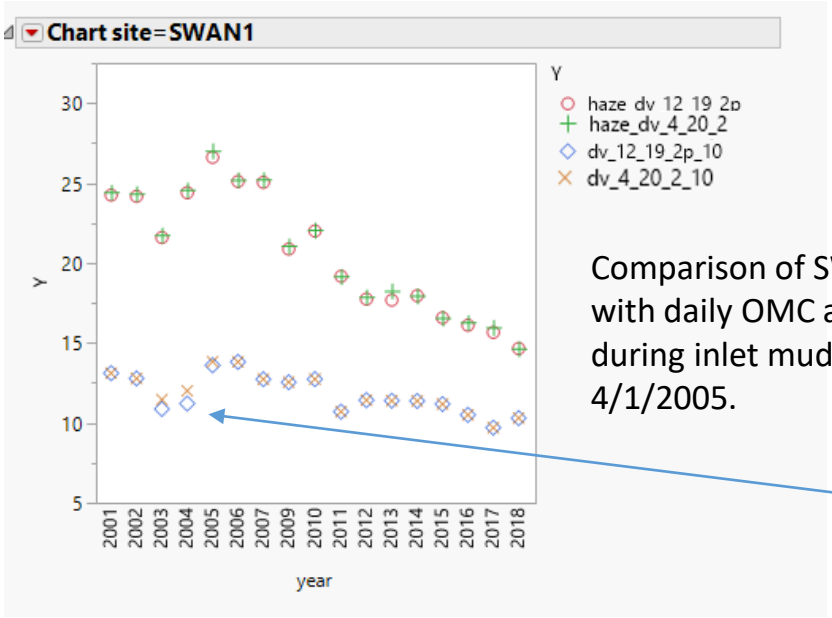
Slide from IMPROVE Steering Committee meeting in fall 2005

This plot shows the problematic data. These are the median ratios of OMC and LAC at ROMA1, BRIG1, and LIGO1 to OMC and LAC at SWAN1. Three data points in each plot per month, 1 for each site, each point is the median for that month of OMC_{SITE}/OMC_{SWAN} or LAC_{SITE}/LAC_{SWAN}.

Period with abnormally high ratio is August 2003 through March 2005, shown by vertical lines.



I reran all the RHR metric codes, **with each daily OMC or LAC value being multiplied by 2.5** from 8/1/2003 to 4/1/2005. Rerun includes e95 calculation, baseline, NC2, 2064 endpoints, etc.



Comparison of SWAN1 MID dv done with daily OMC and LAC scaled by 2.5 during inlet mud clogging 8/1/2003-4/1/2005.

Using the 2.5 factor barely changed the MID baseline and 2064 endpoints. There was a slight increase in baseline clearest (visible in plot, but no table provided).

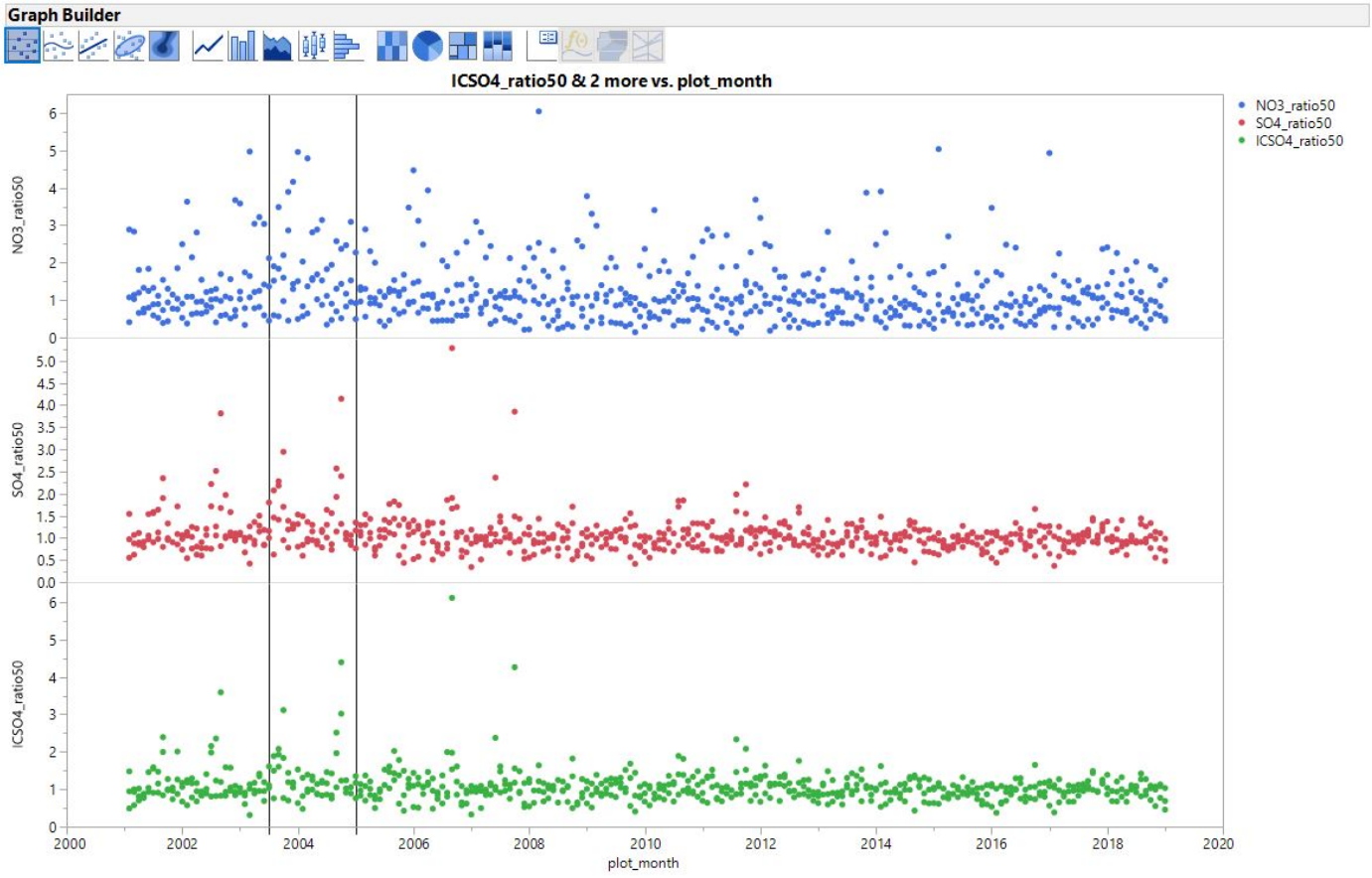
2001-2004 MID Baseline

site	year	EAmm_SO4	EAmm_NO3	EOMC	ELAC	ESoil	ECM	ESea_Salt	haze_dv
SWAN1 Before Carbon Adjust	2001-2004	82.212	3.709	8.192	2.677	0.605	3.286	0.157	23.604
SWAN1 After Carbon Adjust	2001-2004	82.066	3.654	9.898	3.219	0.611	3.284	0.171	23.790

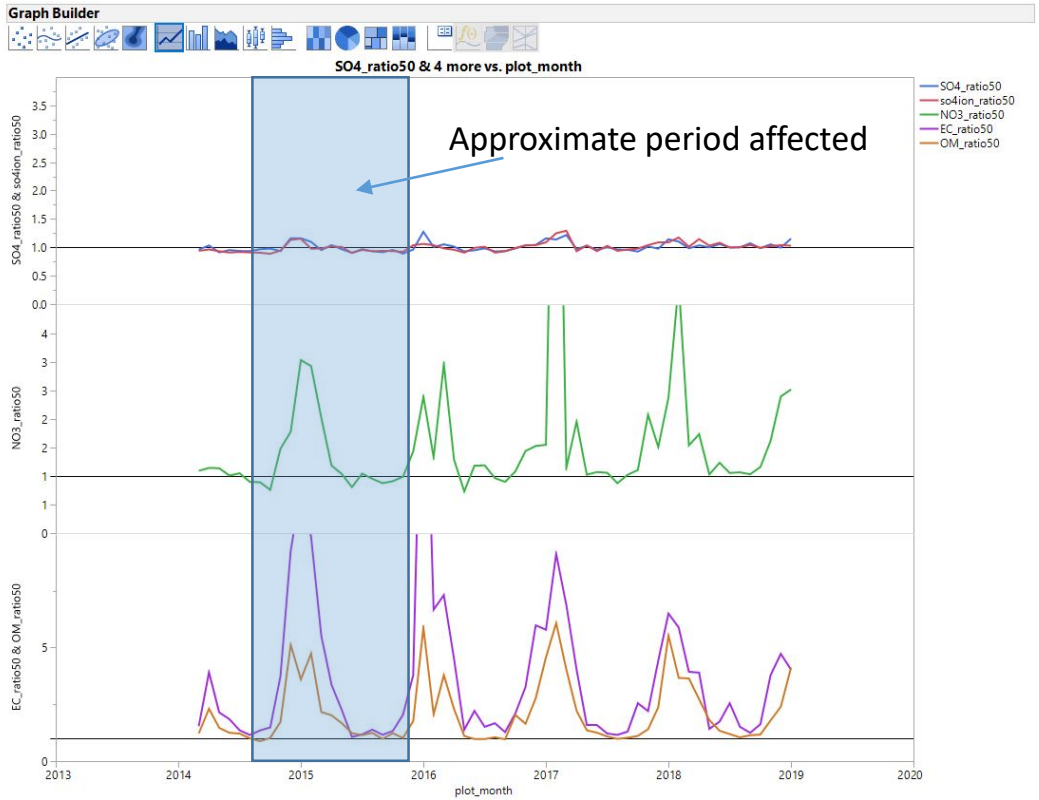
2064 Endpoints

site	end_point_eamm_so4	end_point_eamm_no3	end_point_esea_salt	end_point_eomc	end_point_elac	end_point_esoil	end_point_ecm	end_point_episodic_routine_dv
SWAN1 Before Carbon adjust	3.392	0.871	0.514	8.412	0.475	0.507	1.810	10.005
SWAN1 After Carbon adjust	3.362	0.869	0.515	8.064	0.720	0.501	1.797	9.924

The Module A data for that period was flagged as invalid, and subsequently filled in by data substitution. This plot shows why I don't think the Module B data at SWAN1 was affected significantly (hence requiring substituting the NO3- data). Same plot of ratios to nearby sites, except done for IC NO3-, IC SO4– and the derived variable Amm_SO4 which is a mixture of XRF S and SO4-. The vertical lines are near, but a bit off of the true period affected.



Site Affected: Lake Tahoe Community College (LTCC1) (not a regional haze tracking site)
Issue: 7/1/2014 - 11/17/2015 - the B module stack was "not fully lowered" in the T during that period.
Note that if the air at the stack “T” is like the air at the inlet and no extra nitric acid is getting around the denuder, we would expect the data to be basically the same since it still goes through the cyclone.
After examining the period affected and comparing with nearby BLIS1 data, and at the request of an analyst working for the sponsoring agency, I’ve passed the data through as valid for daily extinction and RHR metric calculations.



Changes to IMPROVE RHR Calculations and Metrics after changing only “patching” routine

Scott Copeland 12_19_2019

Results are graphically displayed in
“Compare 12_2019_1p with 12_2019_2p.pptx”
Available here
https://drive.google.com/open?id=0Bxfj1vyyXeDYWVpfeUo4NEYtTU0

Extended Patching

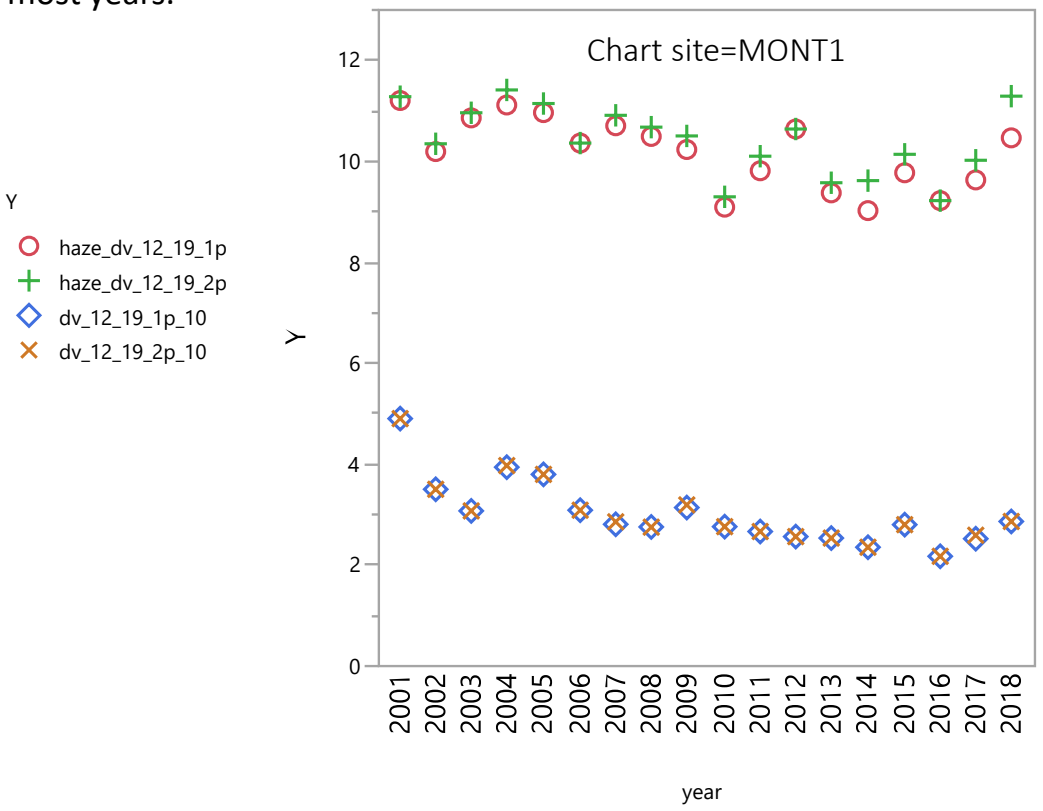
- RHR Guidance based “patching” algorithm was extended to allow up to two missing components of aerosol extinction at one site on one date to be replaced per the guidance technique, rather than just one.
- The test is performed for each missing component independently, not combinations of missing components as was described in the 2003 guidance.
- This change is a response to a request by LADCO analysts in 2017 to allow a subset of important sample dates to be used in progress determinations and is a reasonable extension of the guidance technique.
- This change is the minimum modification to the patching algorithm that can still recover most of the important sample dates at the LADCO sites.
- About 2200 additional sample dates are recovered at various sites since 2000.
- At the IMPROVE Steering Committee meeting in Petaluma and at the National Regional Haze meeting in St Louis, I recommended allowing three components per day to be replaced, but changed to two after a careful inspection of the results, especially at high elevation western sites.
- Every patched or substituted value is now reported explicitly with each data version for each affected site and date. See All_patches_12_19_2p.csv and subs_rhr_5_2019.csv on Google Drive.

	Reported Data	Clearest Days	Most Impaired Days	2000-2004 Baseline	Natural Conditions Estimates	2064 endpoints
Significance of change	n/a	mostly minor-some moderate-one high	mostly minor-some moderate-one high	minor	minor	minor
Number of sites affected	-	potentially all	potentially all	next slide	next slide	potentially all

More about the baseline period

- Below left are all the class I area sites that had at least one observation changed in the baseline period. All of them could have changes to baseline values, natural_conditions2 values, endpoint values, and impairment metrics.
  - In most cases the changes are very small.
- The largest changes by a wide margin were at MONT1 (Bob Marshall, and Mission Mountains Wilderness) in Montana.
  - The large change at MONT1 results from a few added high fire days that changed the min95 carbon threshold, which then changed the impairment values in most years.

ACAD1	AGTI1	BAND1	BOWA1	
BRCA1	BRET1	BRIG1	CHAS1	CHIR1
CRLA1	DENA1	EVER1	GICL1	GRCA2
GRGU1	GRSM1	GUMO1	HALE1	
HAVO1	HEGL1	HOOV1	ISLE1	KAIS1
KALM1	LAVO1	MELA1	MEVE1	
MONT1	MOOS1	OKEF1	PASA1	PMRF1
PORE1	REDW1	SAMA1	SAWT1	
SENE1	SHEN1	SHRO1	SIPS1	SNPA1
SULA1	SWAN1	THSI1	TRCR1	TRIN1 TUXE1
UPBU1	VOYA2	WHPA1	WHPE1	
YELL2	YOSE1			



Historical changes after this slide

Changes to IMPROVE RHR Calculations and Metrics since 10/2019 version before changing patching routine

Scott Copeland 12_19_2019

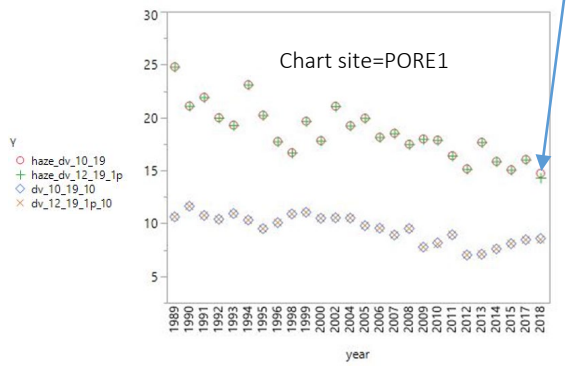
Results are graphically displayed in
"Compare 10_2019 with 12_2019_1p.pptx"
Available here
https://drive.google.com/open?id=0Bxfj1vyyXeDYWVpfeUo4NEYtTU0

UCD limited data redelivery

- Several sample dates with erroneously missing XRF values at 8 sites for parts of 2018 were restored.
- Missing IC data for three days at AGTI1 in 2018 were restored – all three days remain invalid for RHR metrics.
- LOST1 Lostwood NWR, ND - 08/26/2017-12/31/2018, A module only.
  WICA1 Wind Cave NP, SD - 07/1/2018-12/31/2018, B module only.
  -> Status flags were set to "SA". Data from LOST and WICA are still used provisionally for RHR metric calculation.

Cumulative effects of all three issues:

	Reported Data	Clearest Days	Most Impaired Days	2000-2004 Baseline	Natural Conditions Estimates	2064 endpoints
Significance of change	minor	very minor	Moderate at PORE1, minor everywhere else	n/a	n/a	n/a
Number of sites affected	AGTI1, BLIS1, BRCA1, MORA1, PORE1, REDW1, SEQU1, WHPA1, YOSE1	9	9	-	-	-



Modification of legacy status flags at four sites

"SA"* codes were changed to "TU"* codes for all four modules at 4 sites in 2002 or 2006 after a review of my notes indicated that the "SA" code had been assigned before the "TU" code was in use. "TU" is used to indicate timers that are set up to 24 hours off of the correct time. These reflagged observations are now used for RHR metrics.

At AGTI1, SULA1 and KAIS1, adding these days allowed 2006 to meet the completeness requirements.

	Reported Data	Clearest Days	Most Impaired Days	2000-2004 Baseline	Natural Conditions Estimates	2064 endpoints
Significance of change	high	high	high	minor	minor	minor
Number of sites affected	AGTI1, SULA1, KAIS1, 2006 and BALD1 2002	AGTI1, SULA1, KAIS1, 2006 and BALD1 2002	AGTI1, SULA1, KAIS1, 2006 and BALD1 2002	BALD1	BALD1	AGTI1, SULA1, KAIS1, and BALD1

*"SA" means "Sampling Anomaly" and is used to describe a wide range of issues including obvious impacts from unusual sources or sampler issues. Data is normally delivered to FED but not used for RHR calculations.

*"TU" is used to indicate controller clocks not set to the correct time. Since the "TU" flag was created, these observations have been considered valid for RHR metrics. Care should be taken to remove them from analyses that depend directly on the timing of the sample, e.g. comparisons with other monitors.

Typo

Corrected a long standing typo in the RHR data patching algorithm. The effect was a small random error in determining whether sea_salt patching was allowed. Only sea_salt patching was affected. Only applicable to a small fraction of observations where sea salt wasn't already valid.

	Reported Data	Clearest Days	Most Impaired Days	2000-2004 Baseline	Natural Conditions Estimates	2064 endpoints
Significance of change	n/a	very minor	very minor	very minor	very minor	very minor
Number of sites affected	-	potentially all	potentially all	potentially all	potentially all	potentially all

Addition of new combined haze tracking site

- RHR metrics now include a “SYCA_RHTS” site code which is the combined trending site that treats SYCA1 and SYCA2(which began 24OCT2015) as a single site with no data modification.
- This was based on an analysis that I performed and shared with AZDEQ of the two sites’ data compared with nearby sites.
- 2015 is now a “complete” year for the hybrid site.

	Reported Data	Clearest Days	Most Impaired Days	2000-2004 Baseline	Natural Conditions Estimates	2064 endpoints
Significance of change	n/a	High	high	n/a	n/a	n/a
Number of sites affected	-	SYCA_RHTS 2015 only	SYCA_RHTS 2015 only	-	-	-

Addition of EPA 2019 results to my default 2064 endpoint file

- I am now including the output of the 2019 EPA modeling effort in my 2064 endpoint file for every class I IMPROVE site.
- This is for convenience in evaluating the numbers for myself and the community.
- I haven't evaluated the numbers and how they interact with the glide slopes yet.
- The most recent version of that file is "endpoint_2064_imp_g_90_12_19_2p.csv"

Changes to IMPROVE RHR Calculations and Metrics between 5/2019 and 10/2019 versions

Modified chloride calculation

- Chloride algorithm updated prior to 10_19 version to reflect dramatically reduced IC chloride artifact after 2003.

	Reported Data	Clearest Days	Haziest Days	Most Impaired Days	Natural Conditions Estimates	2064 endpoints
Significance of change	very minor	very minor	very minor	very minor	very minor	very minor
Number of sites affected	all	all	all	all	all	all

Chloride=chlf/**1000**;

Chlorine=clf/**1000**;

Sea_Salt=**1.8***Chloride;

The old way

if ((year le **2003**) and (chlf < chlf_mdl or chlf=**.**)) then Sea_Salt=**1.8***Chlorine;

if ((year gt **2003**) and (chlf < chlf_mdl or chlf=**.**) and (clf ne **.**)) then Sea_Salt=**1.8***Chlorine;

A "." represents a missing observation

The new way