

Appendix B
Screening Approaches – Methods and
Results

Ramboll Areas of Influence Analysis

Determining Areas of Influence – CenSARA Round Two Regional Haze

Final Report

Prepared for:
Michael Vince
Central States Air Resource Agencies
Association (CenSARA)

Prepared by:
Uarporn Nopmongcol, Ross Beardsley,
Jeremiah Johnson, and Susan Kemball-Cook
Ramboll
7250 Redwood Blvd, Suite 105 Novato, CA
94945
www.ramboll.com
P-415-899-0700
F-415-899-0707

November 2018

CONTENTS

1.0 INTRODUCTION	1
2.0 APPROACH	2
2.1 Data Sources	2
2.1.1 IMPROVE Data	2
2.1.2 Emission Data.....	3
2.2 Back-Trajectory Modeling	4
2.3 Area of Influence Analysis	5
2.3.1 Residence Time Analysis	5
2.3.2 Distance-weighted Analysis	5
2.3.3 Extinction Weighted Residence Time.....	5
2.3.4 EWRT Plot Combined with Distance-Weighted Emissions	6
2.4 Point Source Emissions Contributions.....	7
3.0 GRAPHIC PRODUCTS	8
4.0 SUMMARY AND RECOMMENDATIONS	15
4.1 Deliverables	15
4.2 Uncertainties and limitations.....	15
4.3 Recommendations.....	16
5.0 REFERENCES.....	17

TABLES

Table 2-1. Class I Areas of Interest.	2
--	---

FIGURES

Figure 2-1. Annual point source NO _x emissions for the CenSARA for 2016 and 2028 in tons/year.	6
Figure 2-2. Annual point source SO ₂ emissions for the CenSARA for 2016 and 2028 in tons/year.	7
Figure 3-1. Example residence time plot for 20% worst visibility days in 2012-2016 for Caney Creek based on trajectories with 100-m, 200-m, 500-m, and 1,000-m end height.....	9

Figure 3-2. Example distance-weighted residence time plot for 20% worst visibility days in 2012-2016 for Caney Creek based on trajectories with 100-m, 200-m, 500-m, and 1,000-m end height.....10

Figure 3-3. Example sulfate extinction-weighted residence time (EWRT) plot for 20% worst visibility days in 2012-2016 for Caney Creek based on trajectories with 100-m, 200-m, 500-m, and 1,000-m end height.11

Figure 3-4. Example nitrate extinction-weighted residence time (EWRT) plot for 20% worst visibility days in 2012-2016 for Caney Creek based on trajectories with 100-m, 200-m, 500-m, and 1,000-m end height.12

Figure 3-5. Example sulfate EWRT combined with distance-weighted (SO₂) emissions (2016 on the left and 2028 on the right) plot for 20% worst visibility days in 2012-2016 for Caney Creek based on trajectories with 100-m (top) and 500-m (bottom) end height. Contour boundaries based on the sulfate EWRT greater than 0.1% (lighter green) or 0.5% (darker green) are also shown.13

Figure 3-6. Example nitrate EWRT combined with distance-weighted (NO_x) emissions (2016 on the left and 2028 on the right) plot for 20% worst visibility days in 2012-2016 for Caney Creek based on trajectories with 100-m (top) and 500-m (bottom) end height. Contour boundaries based on the nitrate EWRT greater than 0.1% (lighter green) or 0.5% (darker green) are also shown.14

1.0 INTRODUCTION

Regional haze, as defined in the Regional Haze Rule at 40 CFR 51.300, is “visibility impairment that is caused by the emission of air pollutants from numerous sources located over a wide geographic area. Such sources include, but are not limited to, major and minor stationary sources, mobile sources, and area sources”. The Regional Haze Rule required states to submit initial state implementation plans (SIPs) to the US Environmental Protection Agency (EPA) by December 2007. SIPs contain enforceable measures for reducing concentrations of pollutants that cause visibility impairment including fine particulate matter (PM_{2.5}), and PM_{2.5} precursors such as oxides of nitrogen (NO_x), sulfur dioxides (SO_x), and volatile organic compounds (VOC). The second phase (Round 2) SIPs are due in July 2021.

A primary step in developing the SIPs is to characterize the sources that lead to visibility impairment at Class I areas. Back-trajectory receptor models are useful tools for identifying potential regional source locations impacting visibility and have been used to facilitate regional haze planning. This project used a back-trajectory model together with air quality measurement data and emission inventories to identify the geographic areas and emission sources with a high probability of contributing to anthropogenically impaired visibility at Class I areas within CenSARA¹ and nearby states.

The purpose of this work is to identify the geographic areas and emission sources with a high probability of contributing to anthropogenically impaired visibility at CenSARA Class I areas. Ramboll carried out residence time analysis using back-trajectory modeling and extended the analysis using emission, visibility extinction, and distance weighting approach.

This report summarizes our approach and provides examples of graphical results developed in this work. Section 2 of this report describes data sources and our approach. Section 3 presents examples of residence time graphics. We provide additional graphics separately to accompany this report. Summary and recommendations are provided in Section 4.

¹ The Central States Air Resource Agencies Association (CenSARA) promotes the exchange of information related to air quality among states, tribes, local and federal agencies and others and includes the states of Arkansas, Iowa, Kansas, Louisiana, Missouri, Nebraska, Oklahoma and Texas.

2.0 APPROACH

2.1 Data Sources

2.1.1 IMPROVE Data

The latest Regional Haze Rule Summary Data daily impairment values include daily IMPROVE PM_{2.5} components and coarse PM concentration measurements, light extinction values, and visibility impairment parameters. The data flag the 20% most anthropogenically impaired days during the 5-year period from 2012 to 2016². The data include “patched” values (historical seasonal median values are used to fill in missing values following procedures described in EPA’s Guidance for Tracking Progress Under Regional Haze Rule³) so that data were available for each day of the 2012-2016 period of this study. Daily impairment data were downloaded from the IMPROVE website⁴.

Table 2-1 lists the Class I Areas included in the analysis.

Table 2-1. Class I Areas of Interest.

IMPROVE Site (FLM)	IMPROVE Site Code	State	Latitude	Longitude	Operation Dates
Big Bend N.P. (NPS)	BIBE1	TX	29.3027	-103.178	3/1988 - Present
Guadalupe Mountains N.P. (NPS)	GUMO1	TX	31.833	-104.8094	3/1988 - Present
Wichita Mountains Wilderness (FWS)	WIMO1	OK	34.7323	-98.713	3/2001 - Present
Caney Creek Wilderness Area (FS)	CACR1	AR	34.4544	-94.1429	6/2000 - Present
Upper Buffalo Wilderness Area (FS)	UPBU1	AR	35.8258	-93.203	12/1991 - Present
Breton Wilderness Area (FWS)	BRIS1 ⁵	LA	30.1086	-89.7617	1/2008 - Present
Hercules-Glades Wilderness Area (FS)	HEGL1	MO	36.6138	-92.9221	3/2001 - Present
Mingo Wilderness Area (FWS)	MING1	MO	36.9717	-90.1432	5/2000 - Present
Great Sand Dunes Wilderness Area (NPS)	GRSA1	CO	37.7249	-105.5185	5/1988 - Present
Rocky Mountain National Park (NPS)	ROMO1	CO	40.2783	-105.5457	9/1990 - Present
Salt Creek Wilderness Area (FWS)	SACR1	NM	33.4598	-104.4042	4/2000 - Present
White Mountain Wilderness Area (FS)	WHIT1	NM	33.4687	-105.5349	1/2002 - Present

² The Round 2 regional haze rule SIP baseline is 2013-2017. However, the IMPROVE most anthropogenically impaired days data for 2017 are not yet available, so 2012-2016 will be used.

³ <https://www3.epa.gov/ttnamti1/files/ambient/visible/tracking.pdf>

⁴ <http://vista.cira.colostate.edu/Improve/rhr-summary-data/>

⁵ Note change in monitor location from planning period 1 – previously BRET1, now BRIS1

IMPROVE Site (FLM)	IMPROVE Site Code	State	Latitude	Longitude	Operation Dates
Wheeler Peak Wilderness Area (FS)	WHPE1	NM	36.5854	-105.452	8/2000 - Present
Voyageurs NP #2 (NPS)	VOYA2	MN	48.4126	-92.8286	11/1999 - Present
Boundary Waters Canoe Area (FS)	BOWA1	MN	47.9466	-91.4955	8/1991 - Present
Seney (FWS)	SENE1	MI	46.2889	-85.9503	11/1999 - Present
Isle Royale (NPS)	ISLE1	MI	47.4596	-88.1491	11/1999 - Present
Mammoth Cave NP (NPS)	MACA1	KY	37.1318	-86.1479	9/1991 - Present
Sipsey Wilderness (FS)	SIPS1	AL	34.3433	-87.3388	3/1992 - Present
Wind Cave (NPS)	WICA1	SD	43.5576	-103.4838	12/1999 - Present
Badlands NP (NPS)	BADL1	SD	43.7435	-101.9412	3/1988 - Present
Theodore Roosevelt (NPS)	THRO1	ND	46.8948	-103.3777	12/1999 - Present
Lostwood (FWS)	LOST1	ND	48.6419	-102.4022	12/1999 - Present

2.1.2 Emission Data

For this study, we used 2016 and 2028 emission inventory data to determine the potential impact from sources of SO₂ and NO_x emissions (precursors of sulfate [SO₄] and nitrate [NO₃], respectively). Industrial sources, including electric generating unit (EGU) and other industrial point (non-EGU) sources, are major contributors to both SO₂ and NO_x emissions. Industrial emissions released at elevated stack heights can potentially be transported very far downwind impacting visibility in the Class I areas. We analyzed potential visibility impacts from EGU and non-EGU sources.

The EPA's National Emissions Inventories (NEI) are comprehensive and detailed estimate of air emissions of criteria pollutants, criteria precursors, and hazardous air pollutants from air emissions sources. The NEIs are generated using EPA approved methods and are publicly available. These inventories have been used in determining compliance with the NAAQS and for policy development and community planning. The 2011v6.3 modeling platform⁶ is based on the 2011NEI version 2 and includes projected future years of 2017, 2023, and 2028. There are multiple modeling cases available through the 2011NEI platform. These modeling cases are indexed alphabetically beginning with a as EPA introduces emission updates. The modeling cases 2011ek and 2017ek supported the Final Cross-State Air Pollution Rule (CSAPR) Update, a rule related to interstate transport for the 2008 Ozone National Ambient Air Quality Standards (NAAQS). Updates to the platform were made to support preliminary modeling of interstate

⁶ <https://www.epa.gov/air-emissions-modeling/2011-version-63-platform>

transport for the 2015 Ozone NAAQS with cases 2011el and 2023el, and preliminary modeling for the assessment of reasonable progress for regional haze with cases 2011el and 2028el. A complete description of the inventory and preparation procedures for these data is available in the NEI2011v6.3 Technical Support Document (EPA, 2015).

EPA is developing a 2016 modeling platform based on the 2014NEI⁷. The 2016NEI version alpha incorporates the 2016 Clean Air Markets Division (CAMD) hourly Continuous Emissions Monitor (CEM) data for EGU sources. The 2016NEI version beta will update non-EGU point sources and include a projected future year of 2028. Currently, the 2016NEI version alpha is available from EPA's FTP site⁸ while the beta version is being finalized.

We compiled facility-specific emissions based on the 2016NEI version alpha for the current year and the 2011NEI modeling case 2028el for the future year⁹. This emission database was provided to CenSARA for review and we updated the inventory based on the feedback received.

2.2 Back-Trajectory Modeling

The National Oceanic and Atmospheric Administration (NOAA) Air Resources Laboratory (ARL) administers an archive of meteorological model forecast and reanalysis datasets prepared by the National Weather Service (NWS) National Centers for Environmental Prediction (NCEP) which can be used as inputs for the Hybrid-Single Particle Lagrangian Integrated Trajectory (HYSPLIT) back-trajectory model¹⁰. HYSPLIT is one of the most commonly used atmospheric transport and dispersion models in the atmospheric sciences community (Stein et al., 2015; Fleming et al., 2012). The gridded meteorological dataset selected for this work is the North American Model (NAM)¹¹ sigma-pressure hybrid dataset (NAMS) which has 12 km horizontal spatial resolution covering the continental US and most of Canada and Mexico. The NAMS dataset offers the finest spatial and temporal resolution (i.e., hourly) available for this study's modeling period. We obtained the daily NAMS meteorological data from the NOAA ARL FTP server¹². There were six days in the modeling period for which the NAMS hourly data was not available. The NAM 3-hourly data which also has 12 km resolution was used to fill gaps¹³.

We ran HYSPLIT model for each of the 20% most anthropogenically impaired days to develop back trajectories for the IMPROVE site in each of the selected Class I areas. We generated 72-hour back trajectories arriving at each of the IMPROVE sites at 06:00, 12:00, 18:00 and 24:00

⁷ https://www.wrapair2.org/pdf/2017-12-12_2016%20Emissions%20Modeling%20Platform%20Development%20Plan_V1.0.pdf

⁸ <ftp://newftp.epa.gov/air/emismod/2016/alpha/2016fd/emissions/>

⁹ Our Quality Assurance Project Plan (QAPP) proposed using the 2011 NEI modeling case 2017ek for non-EGU sources. After consulting with CenSARA, the 2016NEI alpha inventory is deemed more representative of the 2016 conditions.

¹⁰ <https://ready.arl.noaa.gov/archives.php>

¹¹ <http://www.nco.ncep.noaa.gov/pmb/products/nam/>

¹² <ftp://arlftp.arlhq.noaa.gov/nams>

¹³ Trajectories ending on the following days used NAM 3-hourly data obtained from <ftp://arlftp.arlhq.noaa.gov/pub/archives/nam12> as the NAMS data was missing: 07/26/2012, 07/27/2012, 01/12/2015, 01/13/2015, 01/15/2015 and 01/16/2015.

local time for trajectory ending altitudes of 100 m, 200 m, 500 m, and 1,000 m. We used model vertical velocity option in HYSPLIT to simulate vertical motion.

2.3 Area of Influence Analysis

This section describes multiple metrics used to characterize areas and emission sources that lead to visibility impairment at Class I areas.

2.3.1 Residence Time Analysis

Based on the HYSPLIT back trajectories for the 20% most anthropogenically impaired days in 2012-2016, we developed back-trajectory residence time plots for each IMPROVE site. The residence time is the cumulative time that trajectories reside in a specific geographical area (e.g., a grid cell of a modeling domain) and are usually normalized to display percentage of total trajectory time:

$$\tau_{ij} = \frac{1}{NT} \sum_{k=1}^N \tau_{ijk}$$

where τ_{ijk} is the residence time of the k^{th} trajectory at the grid cell (i, j) , N is the total number of trajectories, and T is the duration of each trajectory (72 hours in this analysis).

2.3.2 Distance-weighted Analysis

An alternative method is to weight the residence times by the distance of the grid cell from the receptor (d_{ij}). This approach is based on Source Contribution Function (SCF), which is defined as the residence time normalized by an idealized residence time that would exist if all air masses arrived at the receptor following a straight trajectory with constant speed and equal probability from all directions. A SCF with a value greater than 1 corresponds to a transport pattern that is much more likely than if air arrived from all directions with equal probability. This idealized residence time is always inversely proportional to d_{ij} ¹⁴. Therefore,

$$SCF_{ij} = d_{ij}\tau_{ij}$$

This formulation is designed to compensate for the bias of residence time toward the receptor site due to the receptor site being the point from which all trajectories originate.

2.3.3 Extinction Weighted Residence Time

EPA's previous analysis of contributions of individual PM components to total extinction on the 20% most anthropogenically impaired days during 2010-2014 showed that sulfate (SO₄) and nitrate (NO₃) are two major PM components that account for a large fraction of the anthropogenic visibility impairment at these Class I areas¹⁵. To define geographical areas with a high probability of influencing visibility (i.e. the area of influence) at each of the IMPROVE sites

¹⁴ d_{ij} is calculated in meters in the Lambert conformal conic projection

¹⁵ https://www.epa.gov/sites/production/files/2016-07/documents/technical_support_document_for_draft_guidance_on_regional_haze.pdf

that has impairment due to SO₄ and NO₃, extinction weighted residence time (EWRT) plots were generated separately for SO₄ and NO₃

$$EWRT_{ij} = \sum_{k=1}^N b_{ext_k} \tau_{ijk}$$

where b_{ext_k} is the extinction coefficient attributed to the pollutant (SO₄ or NO₃) measured upon arrival of the k^{th} trajectory at the IMPROVE site. The gridded EWRT values are normalized to display the percentage of the domain total EWRT.

2.3.4 EWRT Plot Combined with Distance-Weighted Emissions

To determine the potential impact from sources of SO₂ and NO_x emissions (precursors of SO₄ and NO₃, respectively), the EWRT values for SO₄ and NO₃ calculated in Section 2.3.3 were combined with emissions (Q) from sources of SO₂ and NO_x, respectively. CenSARA states chose to focus on EGU and non-EGU point sources since these sources comprise major fractions of the NO_x and SO₂ emissions inventory. To incorporate the effects of dispersion, deposition and chemical transformation along the path of the trajectories, emissions were inversely weighted by the distance (d) between the centers of the grid cell emitting the emissions and the grid cell containing the IMPROVE site. Each grid cell has a horizontal resolution of 36 km x 36 km. In the case that the monitoring grid cell also contains emissions (i.e., d is zero), we set the distance to half of the grid cell size (i.e., 18 km).

$$\frac{Q_{ij}}{d_{ij}} EWRT_{ij}$$

The EWRT value combined with distance-weighted emissions for each grid cell were normalized by the domain total, and then plotted for both 2016 and 2028 emissions. Figure 2-1 and 2-2 display the gridded point source NO_x and SO₂ emissions from the 2016 and 2028 inventories.

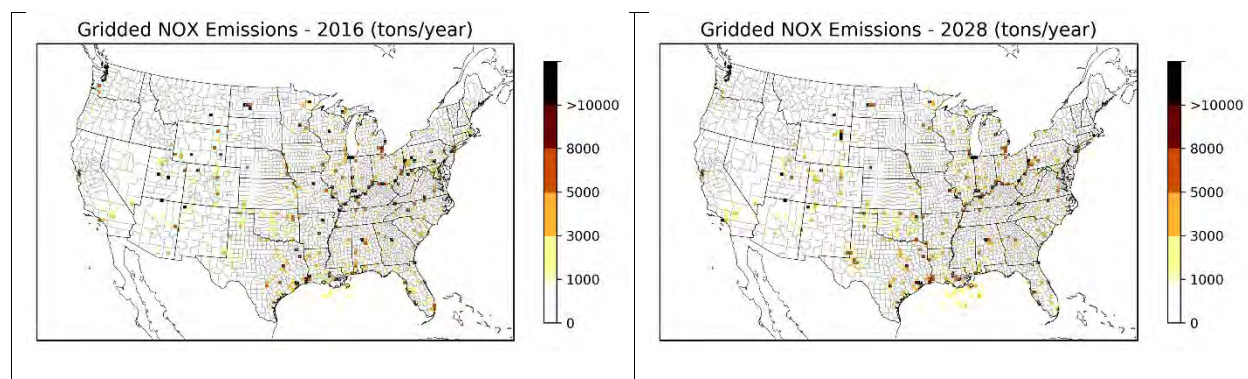


Figure 2-1. Annual point source NO_x emissions for the CenSARA for 2016 and 2028 in tons/year.

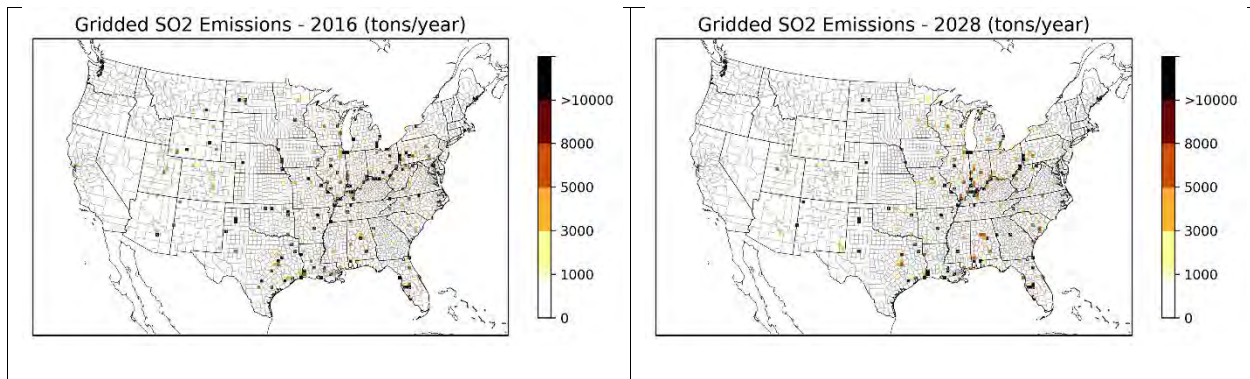


Figure 2-2. Annual point source SO₂ emissions for the CenSARA for 2016 and 2028 in tons/year.

2.4 Point Source Emissions Contributions

We examined source contributions from each facility to visibility impairment at each Class I area by matching the extinction weighted residence time (described in Section 2.3.3) with the facility-level emissions over distance of the 2016 and 2028 point source inventories. The resulting dataset is presented as Excel spreadsheets (provided separately) that contain the following information: Facility ID/Name, State/County/Federal Information Processing Standard (FIPS) code, North American Industry Classification System (NAICS) code, Industry description, SO₂/NO_x emissions in tons per year (Q), Distance in km (d) between the facility center and each IMPROVE site, Q/d for SO₂/NO_x, EWRT for SO₄/NO₃ and EWRT*(Q/d).

3.0 GRAPHIC PRODUCTS

We prepared images of the residence time plots for each Class I area listed in Table 2-1 for all altitudes. We first mapped the back-trajectories to EPA's 12 km continental U.S. (CONUS) domain¹⁶, aggregated to 36-km resolution, and added image smoothing to reduce image noise¹⁷. The images are centered around each IMPROVE site and include outlines of states and counties. This section presents examples for the residence time analysis described in Section 2 for the Caney Creek (CACR) Class I area located in Arkansas.

The interpretation of these results can be made qualitatively and quantitatively. The RHR has no specific guidance on threshold values for residence time. We chose a color scale that offers a reasonable range for normalized percentages across selected Class I areas and altitudes. As an aid to analysis, contour boundaries were added to identify regions with scaled residence time values greater than 0.05%, 0.1%, 0.2%, 0.5%, and 1%. States may select a specific cut-off (e.g., >0.5%) to identify Areas of Influence (Aoi).

In the Caney Creek examples, the unweighted residence time plots (Figure 3-1) suggest influences from southerly air masses followed by northeasterly air masses on the 20% most impaired days in 2012-2016. The influencing air mass directions are evident in the distance-weighted residence time plots (Figure 3-2). The similarity of the unweighted (Figure 3-1) and the sulfate extinction-weighted (Figure 3-3) residence time plots imply that the 20% most impaired days are largely driven by high sulfate concentrations. Nonetheless, nitrate also contributes to visibility impairment at this site and is primarily associated with northwesterly and northeasterly air masses (Figure 3-4). The potential impact from SO₂ emission sources can be determined using the sulfate EWRT combined with distance-weighted SO₂ emissions displayed in Figure 3-5. Similarly, nitrate EWRT combined with distance-weighted NO_x emissions plots are shown Figure 3-6. Both types of figures include EWRT three contour boundaries (shown in green) to help define the SO₂ (or NO_x) Aoi as those areas with EWRT greater than 0.1% or 0.5%. For this specific site, the results using 100-m and 200-m end height are similar suggesting areas and sources near the monitor while the results using 500-m and 1000-m end heights also capture areas and sources further away.

¹⁶ 12US2 domain has a lower-left corner at (-2412000 m., -1620000 m.) and 396x246 grid cells. Projection is Lambert-Conformal, with Alpha = 33°, Beta = 45° and Gamma = -97°, with a center of X = -97° and Y = 40°.

¹⁷ Based on Gaussian Filter,

https://docs.scipy.org/doc/scipy/reference/generated/scipy.ndimage.gaussian_filter.html

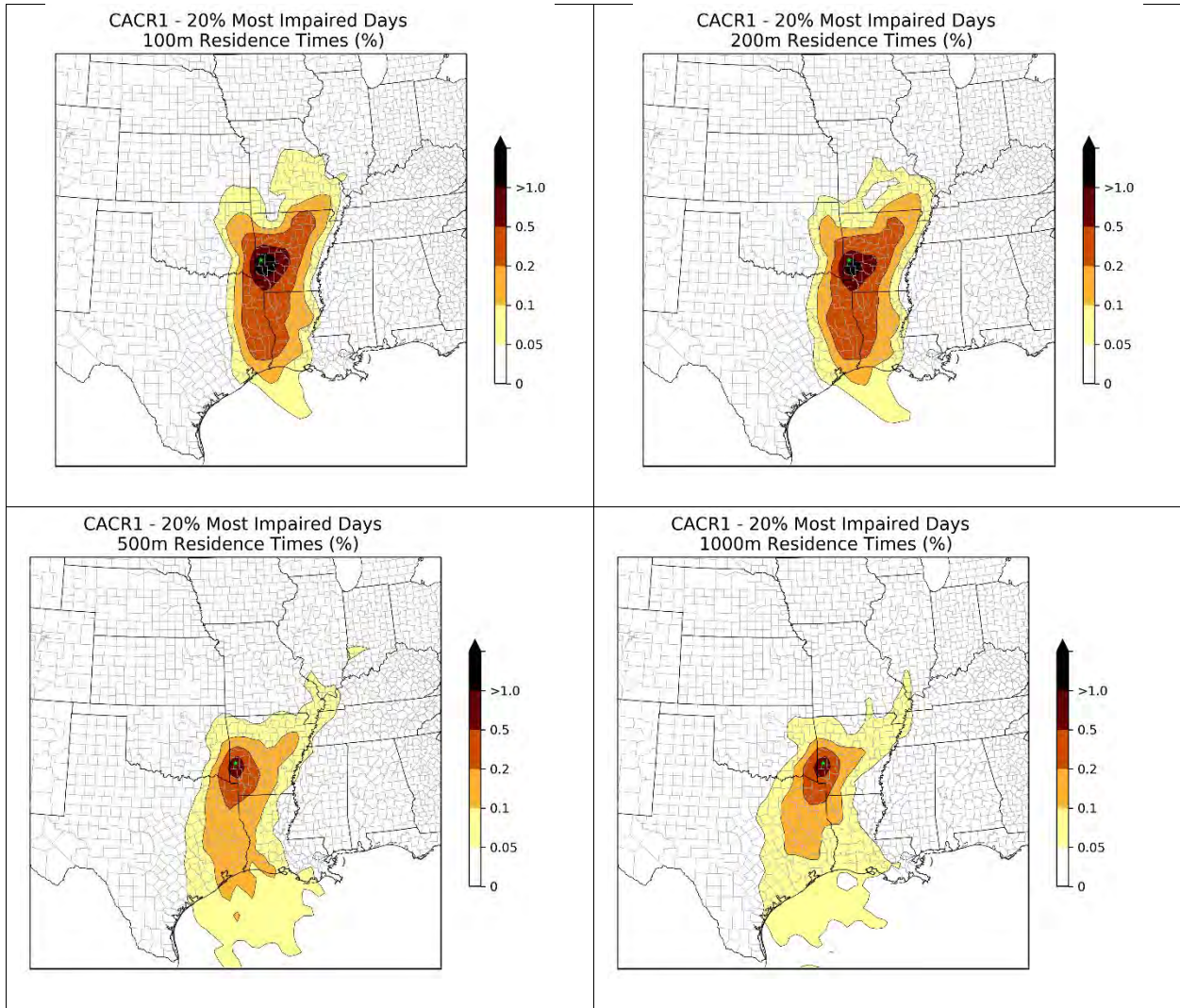


Figure 3-1. Example residence time plot for 20% worst visibility days in 2012-2016 for Caney Creek based on trajectories with 100-m, 200-m, 500-m, and 1,000-m end height.

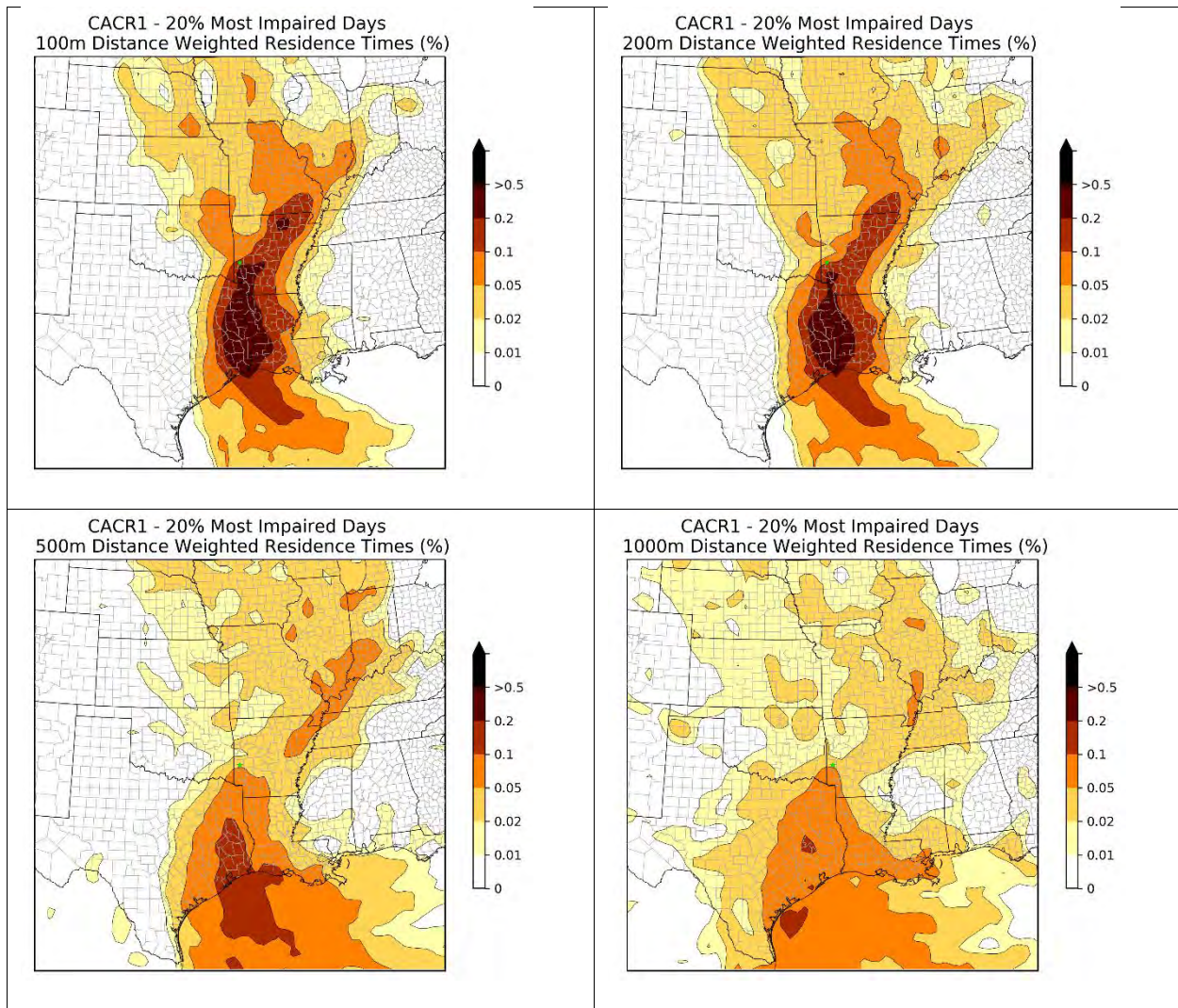


Figure 3-2. Example distance-weighted residence time plot for 20% worst visibility days in 2012-2016 for Caney Creek based on trajectories with 100-m, 200-m, 500-m, and 1,000-m end height.

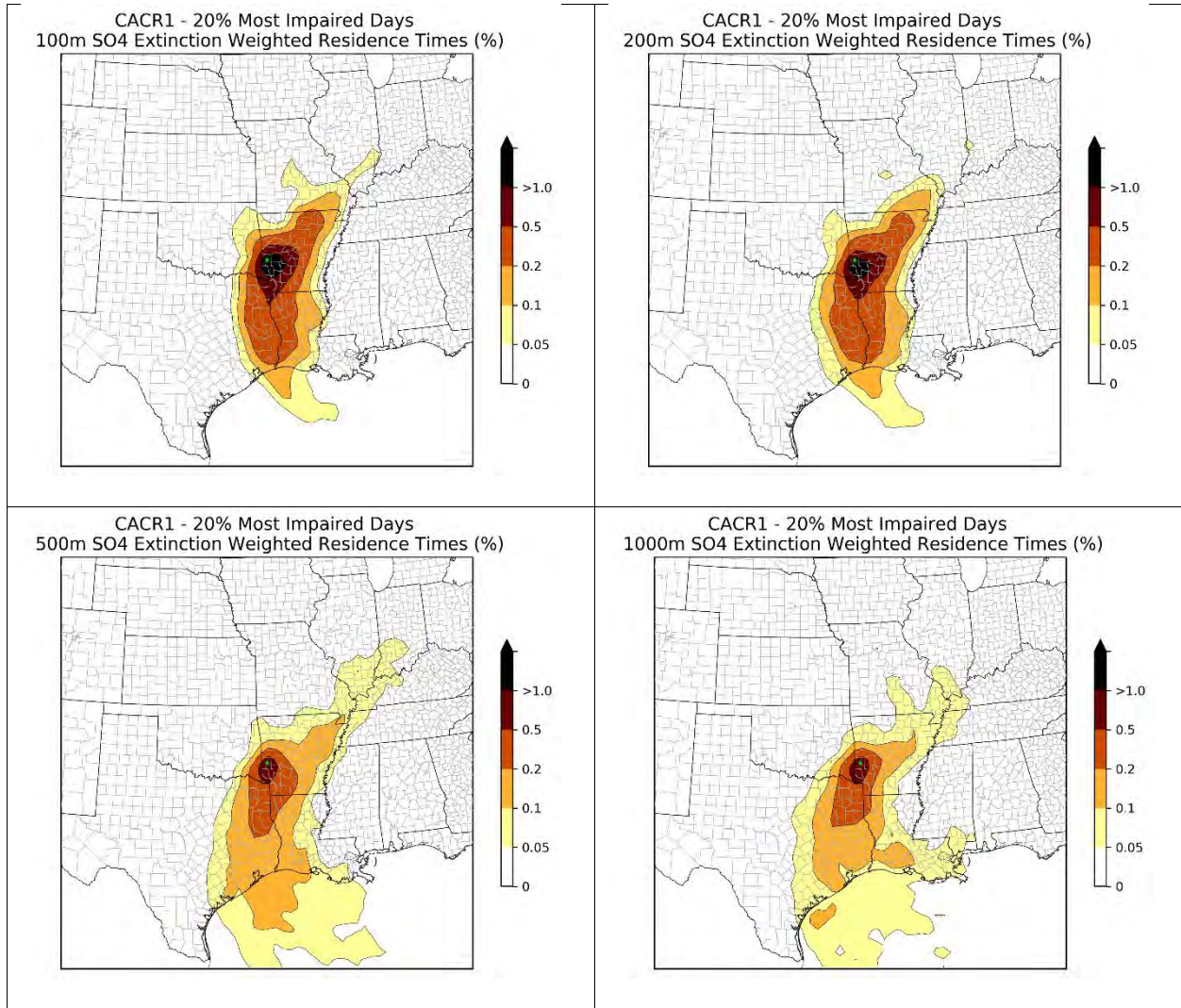


Figure 3-3. Example sulfate extinction-weighted residence time (EWRT) plot for 20% worst visibility days in 2012-2016 for Caney Creek based on trajectories with 100-m, 200-m, 500-m, and 1,000-m end height.

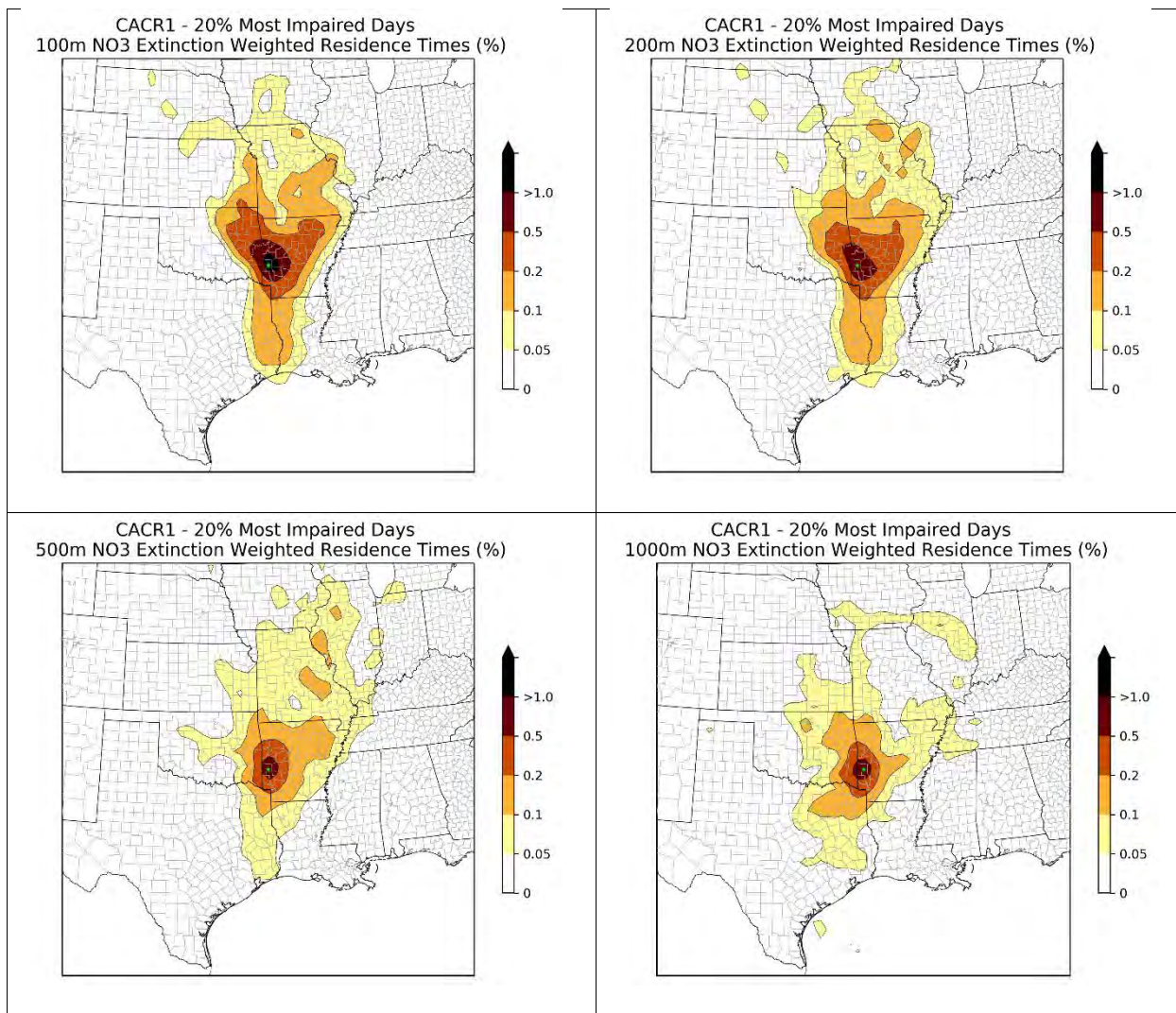


Figure 3-4. Example nitrate extinction-weighted residence time (EWRT) plot for 20% worst visibility days in 2012-2016 for Caney Creek based on trajectories with 100-m, 200-m, 500-m, and 1,000-m end height.

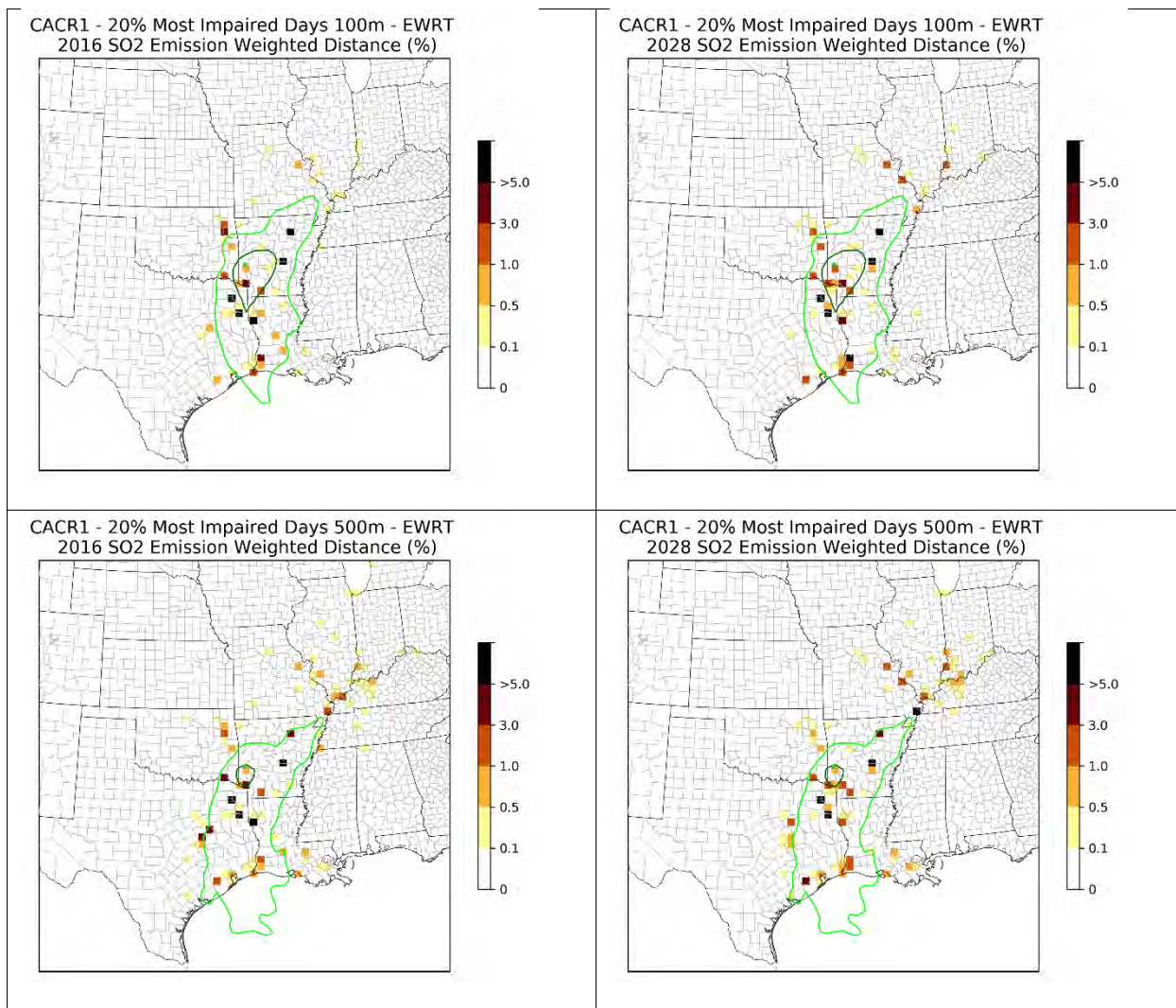


Figure 3-5. Example sulfate EWRT combined with distance-weighted (SO₂) emissions (2016 on the left and 2028 on the right) plot for 20% worst visibility days in 2012-2016 for Caney Creek based on trajectories with 100-m (top) and 500-m (bottom) end height. Contour boundaries based on the sulfate EWRT greater than 0.1% (lighter green) or 0.5% (darker green) are also shown.

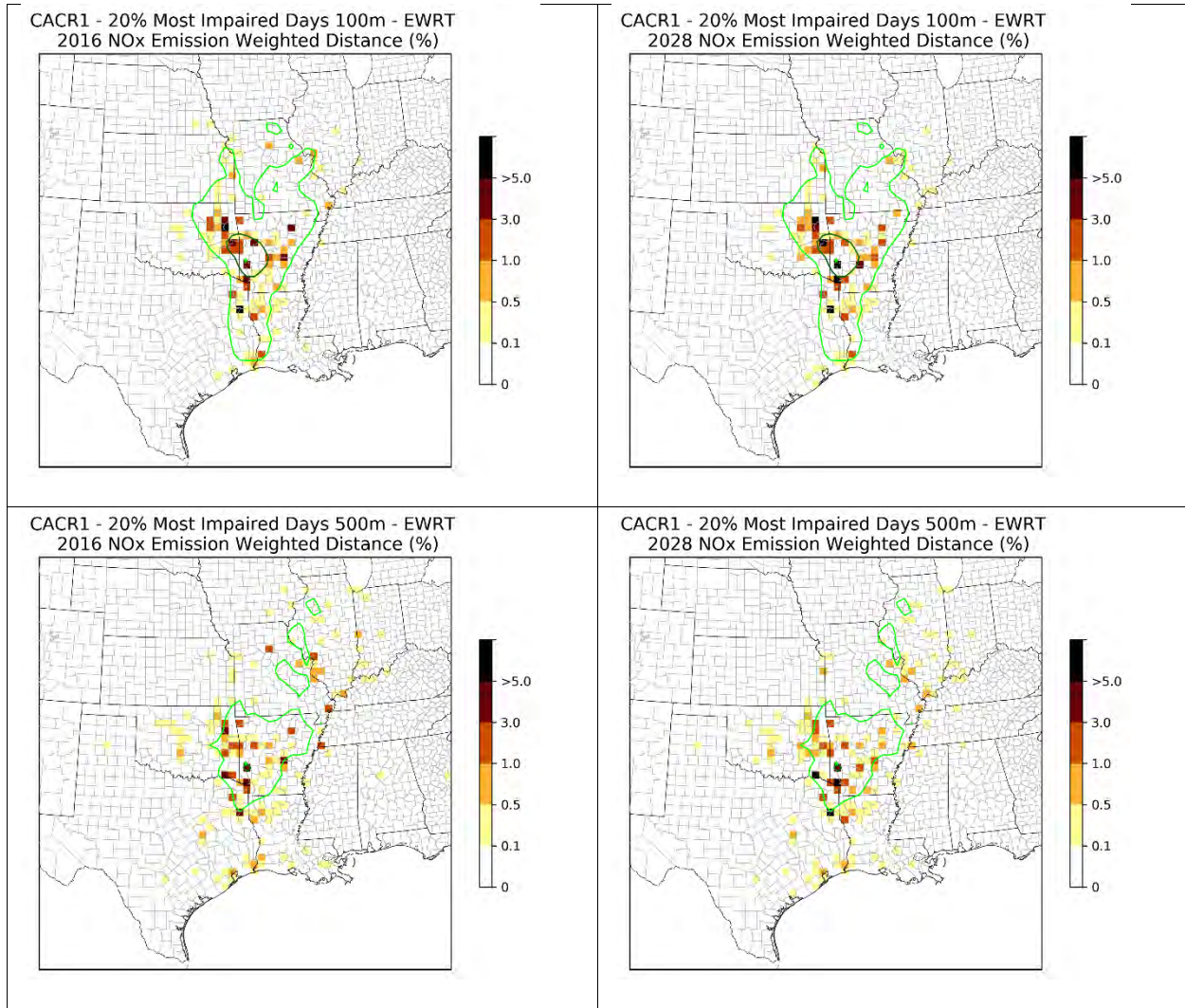


Figure 3-6. Example nitrate EWRT combined with distance-weighted (NO_x) emissions (2016 on the left and 2028 on the right) plot for 20% worst visibility days in 2012-2016 for Caney Creek based on trajectories with 100-m (top) and 500-m (bottom) end height. Contour boundaries based on the nitrate EWRT greater than 0.1% (lighter green) or 0.5% (darker green) are also shown.

4.0 SUMMARY AND RECOMMENDATIONS

In support of CenSARA's Area of Influence (Aol) analysis, Ramboll generated HYSPLIT back trajectories for IMPROVE sites in CenSARA and neighboring states. Back trajectory analyses use interpolated measured or modeled meteorological fields to estimate the most likely central path of air masses that arrive at a receptor at a given time. Back trajectories account for the impact of wind direction and wind speed on delivery of emissions to the receptor, but do not account for chemical transformation and dispersion of emissions. We generated 72-hour back trajectories arriving at each of the IMPROVE sites at 06:00, 12:00, 18:00 and 24:00 local time for trajectory ending altitudes of 100 m, 200 m, 500 m, and 1,000 m.

Based on the five years of individual back trajectories on the most 20 percent impaired visibility days, we mapped trajectory paths into 36-km x 36-km horizontal grid cells and generated residence time data for each IMPROVE site. We then extended the analysis using emission, visibility extinction, and distance weighting approaches. States can use these values to further determine control strategy development for individual Class I areas.

4.1 Deliverables

Our deliverables in this project includes:

- A Final project report (this document)
- Images of the weighted and unweighted residence time for each Class I area (.png electronic format)
- Excel Spreadsheets that show source contributions from each facility to visibility impairment at each Class I separately for the year 2016 and 2028

4.2 Uncertainties and limitations

Some of the uncertainties and limitations of the Aol analysis include:

- The choice of trajectory setup (i.e., ending time, ending altitudes, meteorology, vertical motion) affects the trajectories generated and the final Aol analysis. The gridded meteorological data file used for the HYSPLIT computation is a discrete representation of a continuous field. How the modeled representative of a nearby measurement depend upon local effects as well as the larger scale gradients of the variable and how well a gridded field can represent the underlying continuous field. The impact of receptor height (or end height) on an individual trajectory is also important. Low-ending trajectories represent air parcels nearer to ground level and high-ending trajectories may represent more accurate boundary layer flow above the local terrain.
- We quality assured the trajectory output by calculating the length of each 72-hour back trajectories (or maximum distance between each hourly trajectory segment and the trajectory endpoint) and examine if the distribution of trajectory lengths is reasonable (e.g., checking any obviously incorrect or unphysical trajectories). Some of the trajectory lengths at each 10-min timestep are longer than a 12-km grid cell (corresponding to wind speed of >

20 m/s), but they are occurring at higher altitudes that would be above mixing heights. These instances were deemed plausible so were not discarded.

- We adopted the latest emission estimates available. However, uncertainties related to emission inventories are expected, especially for the future year emission estimates. We did not evaluate uncertainty in the emission inventories used in this study.
- All residence time plots are displayed as relative values (i.e., percentage) based on all grid cells within the CONUS domain. The emission-weighted residence time plots account for EGU and non-EGU point emissions in the US, but they exclude emissions from non-point and international sources. Caution should be made when interpreting the emission-weighted results for Class I areas that are near the international borders, such as Big Bend (BIBE). The EWRT contour boundaries should be considered together with the emission-weighted values.
- The back trajectories are based on a single air parcel transport pattern and does not fully account for three-dimensional transport and dispersion patterns and chemical transformation that can influence the transport and formation of visibility impairing particulate matter species.

4.3 Recommendations

Initial recommendations are as follows:

- Our results may be sensitive to the horizontal grid resolutions chosen (12 km x 12 km trajectories aggregated to 36 km x 36 km in the analysis). Further aggregation of horizontal grid cells (e.g., 50-km) may help tailor the presentations of the AoI to meet the CenSARA states' needs.
- Given that air parcels are expected to be exposed to emissions only when they are below mixing heights, the AoI analysis can be refined by excluding those trajectory segments that are above mixing height. HYSPLIT can optionally output mixing height estimates to facilitate this refinement.

5.0 REFERENCES

- EPA. 2015. 2011 National Emissions Inventory, version 3, Technical Support Document. U.S. Environmental Protection Agency, Research Triangle Park, NC. Aug. https://www.epa.gov/sites/production/files/2016-09/documents/2011v6_3_2017_emismod_tsd_aug2016_final.pdf)
- Fleming, Z.L., Monks, P.S. and Manning, A.J., 2012. Untangling the influence of air-mass history in interpreting observed atmospheric composition. *Atmospheric Research*, 104, pp.1-39, <https://www.sciencedirect.com/science/article/pii/S0169809511002948>
- Stein, A.F., Draxler, R.R, Rolph, G.D., Stunder, B.J.B., Cohen, M.D., and Ngan, F., (2015). NOAA's HYSPLIT atmospheric transport and dispersion modeling system, *Bull. Amer. Meteor. Soc.*, 96, 2059-2077, <http://dx.doi.org/10.1175/BAMS-D-14-00110.1>

Screening Approaches – Methods and Results

Q/D method

Under this method only SO₂ and NO_x emissions and distance were considered for sorting and selecting sources. Back-trajectory and residence-time analysis were not included. The following steps were completed for source selection using the Q/D method.

1. The Ramboll's spreadsheet ([facilityemis.EWRT.qd2028.alltraj.xlsx](#)) is used as the basis for screening sources using the Q/D method.
2. Under the “summary” tab, MING1 is selected for cell C2 to start the screening for Mingo wilderness area.
3. On columns X, Y, and Z, Row 4, arbitrary thresholds for Q/D based on NO_x, SO₂, and sum of them are added. Also, new filters from row 6 are added. The filters allow to initially select sources per desired thresholds for either NO_x, SO₂, both together, or sum of them. All sources with a total (NO_x + SO₂)/D of 10 (tons/km) or higher are selected. The 10 (tons/km) threshold was selected arbitrarily and tentatively to capture a reasonable number of high-contributing sources given the air program's budget and staff constraints. The selected sources are then copied to a new spreadsheet ([MING1.xlsx](#)) for the next steps.
4. Column L (sum of both emissions) is added and sorted in descending order. Next, column M (cumulative emissions/d) is created.
5. In the next step, a group of sources (with highest total Q/D) that have a total of 80 percent of sum of column L are selected.
6. The sources highlighted in bright yellow together contribute 3,536 (tons/km) or about 80 percent of the contributions (4,434 tons/km) of all sources that have a Q/D of 10 (tons/km) or higher for Mingo. The list includes a total of 70 sources, of which, the following 11 sources are located in Missouri. These sources have the largest amounts of Q/D for Mingo among all Missouri sources.

New Madrid Power Plant-Marston

Ameren Missouri-Labadie Plant

Ameren Missouri-Rush Island Plant

Noranda Aluminum Inc-New Madrid

Thomas Hill Energy Center Power Division-Thomas Hill

Mississippi Lime Company-Ste. Genevieve

Ameren Missouri-Sioux Plant

Appendix B - Screening Approaches – Methods and Results

Doe Run Company-Buick Smelter
River Cement Co. Dba Buzzi UNICEM USA-Selma Plant
City Utilities of Springfield Missouri-John Twitty Energy Center
Buzzi UNICEM USA-Cape Girardeau

In this spreadsheet, there are only three other sources from Missouri:

Holcim (US) Inc-Ste. Genevieve Plant
University of Missouri (Mu)-Power Plant
Lhoist North America Of Missouri-Ste. Genevieve

A similar analysis was conducted for HELG ([HEGL1.xlsx](#)), and the following sources from Missouri were identified.

Ameren Missouri-Labadie Plant
Thomas Hill Energy Center Power Division-Thomas Hill
Ameren Missouri-Rush Island Plant
City Utilities of Springfield Missouri-John Twitty Energy Center
New Madrid Power Plant-Marston
Ameren Missouri-Sioux Plant
Mississippi Lime Company-Ste. Genevieve
Noranda Aluminum Inc-New Madrid
University of Missouri (MU)-Power Plant

The following two sources had smaller effects on HEGL

Doe Run Company-Buick Smelter
Hawthorn

Back-Trajectory and Residence Time Analyses

Appendix B - Screening Approaches – Methods and Results

Under this method, [facilityemis.EWRT.qd2028.alltraj.xlsx](#) was used to screen sources based on EWRT*Q/D time for NO₃ and SO₄.

Initially, arbitrary thresholds of 0.05% for both EWRT*NO_x and EWRT*SO₂ were used for filtering out the sources. In the next step, a cumulative (NO_x and SO₂) threshold of 80% contribution was set to screen the final list of sources as follows, which are identified for MING and HEGL, respectively:

MING:

- Ameren Missouri-Labadie Plant
- Ameren Missouri-Rush Island Plant
- Ameren Missouri-Sioux Plant
- Buzzi UNICEM USA-Cape Girardeau
- Mississippi Lime Company-Ste. Genevieve
- New Madrid Power Plant-Marston
- Noranda Aluminum Inc-New Madrid

HEGL:

- Ameren Missouri-Labadie Plant
- Ameren Missouri-Rush Island Plant
- Branson
- James River Power Plant
- John Twitty Energy Center
- Mississippi Lime Company-Ste. Genevieve
- New Madrid Power Plant-Marston
- Noranda Aluminum Inc-New Madrid
- Thomas Hill

To create the above lists, the following steps were completed:

1. Select Exclusion Thresholds for NO_x and SO₂ Area of Influence
 - a. Change EWRT NO_x threshold to 0.05%.
 - b. Change EWRT SO₂ threshold to 0.05%.

Appendix B - Screening Approaches – Methods and Results

- 2 Select Class I Area of Interest
 - a. Select from drop-down box

- 3 Select Records for EWRT - NO_x > threshold selected in Step 1
 - a. Click on Filter drop-down for column R and uncheck "NO"
 - b. Select cells in Columns B - Q starting with Row 6

 - c. Create a new worksheet tab for the Class I Area

 - d. Copy Selected fields (Ctrl-C)
 - e. Paste into Class I Area tab (Ctrl-V)
 - f. Click on Filter drop-down for column and recheck "NO"

- 4 Select Records for EWRT - SO₂ > threshold selected in Step 1
 - a. Click on Filter drop-down for column S and uncheck "NO"
 - b. Select cells in Columns B - Q starting with Row 6
 - c. Copy Selected fields (Ctrl-C)
 - d. Paste into Class I Area tab (Ctrl-V) below data from Step 3
 - e. Click on Filter drop-down for column and recheck "NO".

Perform the following procedures in the Individual Class I Area Tab

- 5 De-duplicate pasted data
 - a. Select all columns and rows
 - b. Use the Remove Duplicates Data Tool to remove duplicate facilities

- 6 Create Column Q
 - a. Title Column Q "Combined EWRT*Q/D"
 - b. $Q(n) = O(n) + P(n)$ for $n > 1$

- 7 Sort data by Column Q - Largest to Smallest

- 8 Create Column R
 - a. Title Column R "% of AOI Source Impact"
 - b. $R(n) = Q(n) / (\text{Total of Column Q})$ for $n > 1$
 - c. Format as Percent with zero decimal places

- 9 Create Column S
 - a. Title Column S "Cumulative Combined EWRT*Q/D"
 - b. $S(2) = Q(2), S(n) = S(n-1) + Q(n)$ for $n > 2$

- 10 Create Column T
 - a. Title Column T "Cumulative % of AOI Source Impact"
 - b. $T(n) = S(n) / (\text{Total of Column Q})$ for $n > 1$
 - c. Format as Percent with zero decimal places

- 11 Choose a threshold for selecting sources

Since the two pollutants and their cumulative thresholds are arbitrary values to be determined before using this method, Missouri ran the method using different combinations of thresholds. The following list of in-state sources is an instance of such runs, with individual thresholds of 0.05% and a cumulative threshold of 50%, for MING.

New Madrid Power Plant-Marston
Ameren Missouri-Labadie Plant
Ameren Missouri-Rush Island Plant
Noranda Aluminum Inc-New Madrid

And these sources are screened for HEGL:

Ameren Missouri-Labadie Plant
City Utilities of Springfield Missouri-John Twitty Energy Center
Ameren Missouri-Rush Island Plant

New Madrid Power Plant-Marston

Thomas Hill Energy Center Power Division-Thomas Hill

In the absence of a science-based approach to selecting optimal values for thresholds, under this method, different scenarios were examined to discover the sensitivity of selected sources to different levels of thresholds. The results are summarized in Table B.1.

Table B.1. Summary of the sensitivity analysis of thresholds for MING

Case Number	EWRT Threshold NO3 (%)	EWRT Threshold SO4 (%)	Cumulative			Total
			Impact Threshold (%)	Missouri Sources	Other Sources	
1	0.01	0.01	50	4	3	7
2	0.01	0.05	50	4	2	6
3	0.01	0.1	50	4	3	7
4	0.05	0.01	50	4	2	6
5	0.05	0.05	50	4	1	5
6	0.05	0.1	50	4	1	5
7	0.1	0.01	50	3	1	4
8	0.1	0.05	50	3	1	4
9	0.1	0.1	50	3	1	4
10	0.01	0.01	80	10	33	43
11	0.01	0.05	80	10	26	36
12	0.01	0.1	80	10	29	39
13	0.05	0.01	80	10	28	38
14	0.05	0.05	80	9	18	27
15	0.05	0.1	80	8	18	26
16	0.1	0.01	80	9	15	24
17	0.1	0.05	80	9	19	28
18	0.1	0.1	80	7	12	19

From the above sensitivity analysis, the following results were obtained:

1. The method is significantly more sensitive to the cumulative percentage threshold (i.e., 50% vs. 80%) than it is to the individual EWRT*Q/D thresholds (e.g., 0.05%).
2. EWRT*Q/D thresholds of 0.01% and 0.05%, for both pollution species, result in nearly the same number of sources. Therefore, the model is not considerably sensitive to these variables at this range (0.01% - 0.05%).
3. Increasing the EWRT*Q/D thresholds, for both or either species, to 0.1% results in a small, but noteworthy, decrease in the number of screened sources.
4. Overall, the method is more sensitive to thresholds regarding the out-of-state sources than it is for the in-state sources. That is, by changing the thresholds, the number of out-of-state screened sources changes more sharply than the number of in-state sources.
5. Regardless of the species' thresholds, the cumulative 50% threshold always screens more in-state sources than it screens out-of-state sources, while the 80% threshold always screens more out-of-state sources than it screens in-state sources.
6. The effect of thresholds of the species on the number of screened sources is negligible compared with the much sharper effect of the cumulative threshold.
7. The extreme cases among the above 18 alternatives are, expectedly, cases 9 and 10 for the most compromising and the most conservative approaches with a total of 4 and 43 sources, respectively.

Cluster Analysis

To demonstrate the capabilities of this method for source selection, Missouri sources are clustered for one of the two Missouri class I area, Hercules-Glades wilderness area, below, using both 2016 and 2028 emissions rates. The following clustering scheme has been considered that matches the screening objectives.

It should be noted that the below cluster definitions are presented to make it easier for readers to interpret the meaning of clusters and they have no effects on the mathematical implications of the

algorithms and the number and order of sources in each cluster. The only parameter that affect the distribution of sources between different clusters is the number of clusters, which here was set to 4. The following definitions are rough estimates for the contributions of sources in each cluster. As a result, one might find a source in cluster 1 that has a nitrate contribution that is slightly smaller than that of another source in cluster 3 (see Figure 2 below for an example). Such examples do not contradict with the underlying rationale and reason for using cluster analysis for screening, because the proposed cluster definitions below consider the overall (combined) effect of both pollutants. That is, in general, a source that falls in cluster one has a higher combined contribution of both pollutants than a source in cluster 2 or 3, which is evident from inspecting Figures 1-2 and also examining the EWRT*Q/D analysis spreadsheets.

Cluster 1: Sources with both High SO₄ and High NO₃ contributions

Cluster 2: Sources with High SO₄ and Low NO₃ contributions

Cluster 3: Sources with High NO₃ and Low SO₄ contributions

Cluster 4: Sources with both Low NO₃ and Low SO₄ contributions

Figures 1 and 2 illustrates source clusters using 2016 and 2028 emissions data, respectively.

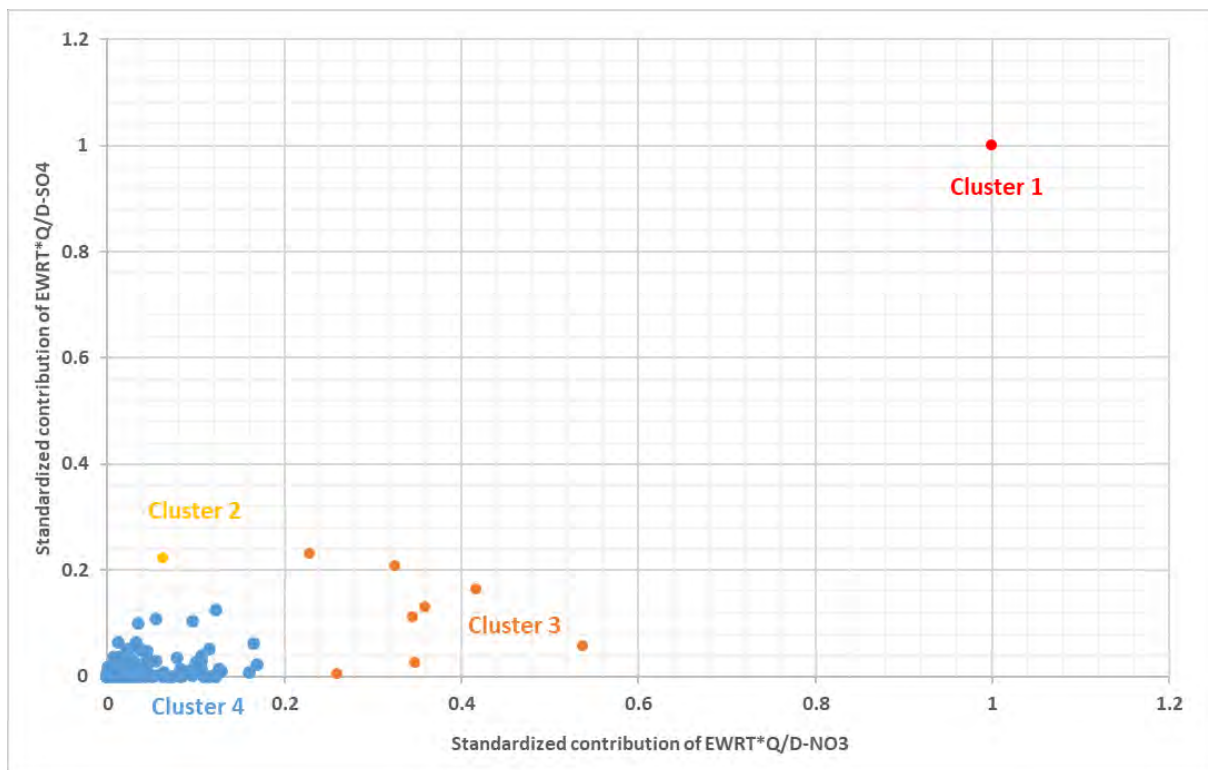


Figure 1. Cluster Analysis using 2016 emissions— Hercules Glades Wilderness Area

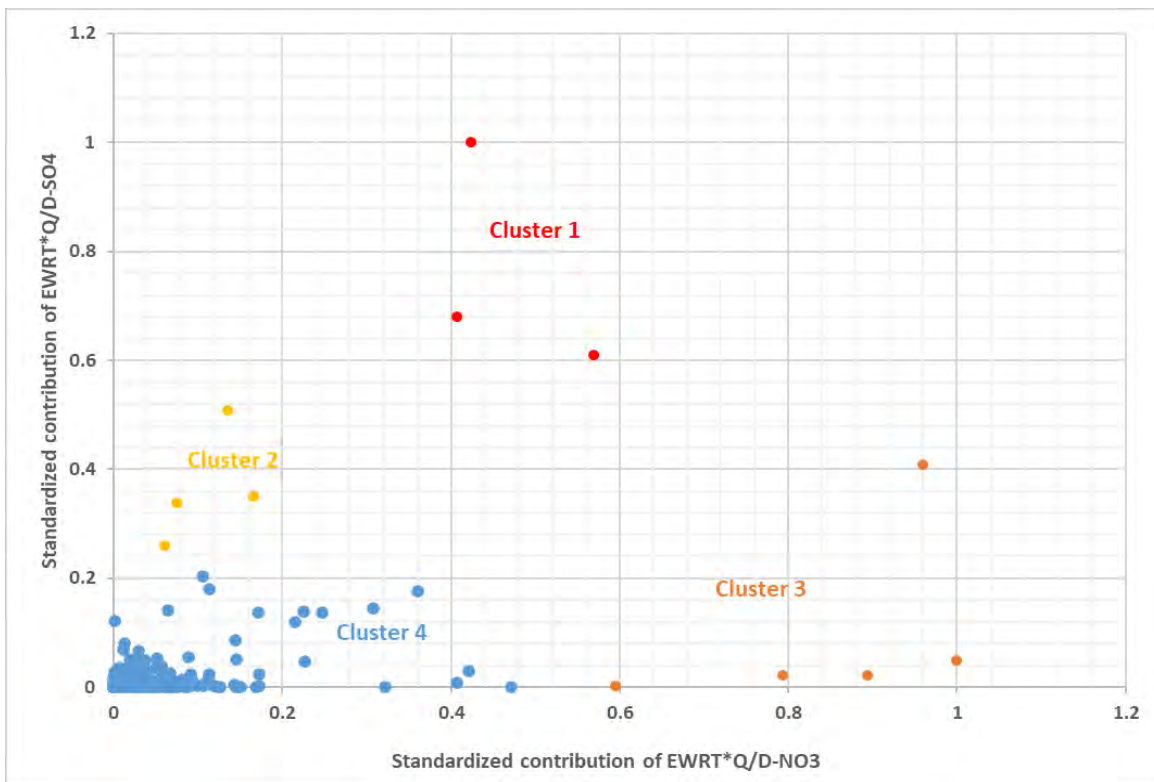


Figure 2. Cluster Analysis using 2028 emissions— Hercules Glades Wilderness Area

Using 2016 emissions data, the following clusters have been identified as potential sources for four-factor analysis.

- Cluster 1 Entergy Arkansas Inc-Independence Plant

- Cluster 2 Ameren Missouri Rush Island Plant

- Cluster 3 Thomas Hill Energy Center Power Division Thomas Hill
 City Utilities of Springfield Missouri John Twitty Energy Center
 Flint Creek Power Plant (SWEPCO)
 New Madrid Power Plant Marston
 Muskogee Gnrtnng Sta
 KCP&L - LA CYGNE
 Ameren Missouri Labadie Plant
 Entergy Ark-White Bluff

Appendix B - Screening Approaches – Methods and Results

As a result, a total of 10 sources, from both Missouri and other states, affecting Hercules-Glades have been identified in the above list.

Using the 2028 data, the following 12 sources are clustered as follows:

- Cluster 1
 - Entergy Arkansas Inc-Independence Plant
 - Entergy Ark-White Bluff
 - Ameren Missouri-Labadie Plant

- Cluster 2
 - Ameren Missouri-Rush Island Plant
 - New Madrid Power Plant-Marston
 - FUTUREFUEL Chemical Company
 - Sikeston Power Station

- Cluster 3
 - City Utilities of Springfield Missouri-John Twitty Energy Center
 - Flint Creek Power Plant (SWEPCO)
 - Chouteau Coal Fired Complex
 - Branson
 - City Utilities of Springfield Missouri-James River Power Plant

Overview of the Clustering Algorithm

Clustering or cluster analysis is a widely-used method of data mining for the purpose of pattern recognition. Clustering is among the unsupervised methods of pattern recognition. By an unsupervised method, one means that the data analyzer does not have any prior hypothesis or pre-specified models for the data, but intends to understand the general characteristics or the structure of the *high-dimensional* data. A supervised method means that the investigator wants to confirm the validity of a hypothesis/model or a set of assumptions, given the available data. Clustering is also called an “unlabelled” method. In pattern recognition, data analysis is concerned with predictive modelling: given some training data, we want to predict the behavior of the ‘unseen’ test data. This task is also referred to as learning. Often, a clear distinction is made between learning problems that are (i) supervised

(classification) or (ii) unsupervised (clustering), the first involving only labelled data (training patterns with known labels), while the latter involving only unlabeled data

K-means Clustering Algorithm

K-means is the mostly used clustering algorithm in the research literature. In k-means clustering, we are given a set of n data points in d -dimensional space, R^d , and an integer k -- the problem is to determine a set of k points in R^d , called centers, so as to minimize the mean squared distance from each data point to its nearest center. Let $X = \{x_i, i = 1, \dots, n\}$ be the set of n d -dimensional points to be clustered into a set of K clusters, $C = \{c_k, K = 1, \dots, K\}$. K-means algorithm finds a partition such that the squared error between the empirical mean of a cluster and the points in the cluster is minimized. Let μ_k be the mean of the cluster c_k . The squared error between μ_k and the points in cluster c_k is defined as:

$$J(c_k) = \sum_{x_i \in c_k} \|x_i - \mu_k\|^2$$

The goal of K-means is to minimize the sum of the squared error over all K clusters,

$$J(c) = \sum_{k=1}^K \sum_{x_i \in c_k} \|x_i - \mu_k\|^2$$

A major challenge in cluster analysis is the estimation of the optimal number of clusters. The algorithm requires three user-specified parameters: number of clusters K , cluster initialization, and distance metric. The most critical choice is K . While no perfect mathematical criterion exists, a number of heuristics and discussions are available for choosing K . Typically, K-means is run independently for different values of K and the partition that appears to be the most “meaningful” one to the domain expert is selected. Different initializations can lead to different final clustering because K-means only converges to “local minima”. One way to overcome the local minima is to run the K-means algorithm, for a given K , with multiple different initial partitions and choose the partition with the smallest squared error.