Atmospheric and Environmental Research

131 Hartwell Avenue
Lexington, Massachusetts
02421-3126
USA
Tel: +1 781 761-2288
Fax: +1 781 761-2299
www.aer.com

TECHNICAL MEMO

Investigating the Impact of Meteorology on O_3 and $PM_{2.5}$ Trends, Background Levels, and NAAQS Exceedances

Task 3: Estimating Background O_3 and $PM_{2.5}$

TCEQ Contract No. 582-15-50415
Work Order No. 582-15-54118-01
Deliverable 3.1
Revision 1.1

Prepared by:

Atmospheric and Environmental Research, Inc. (AER)
131 Hartwell Ave.
Lexington, MA 02466
Correspondence to: malvarad@aer.com

Prepared for:

Erik Gribbin
Texas Commission on Environmental Quality
Air Quality Division
Building E, Room 342S
Austin, Texas 78711-3087

June 15, 2015

Document Change Record

Revision	Revision Date	Remarks
1.0	**29 May 2015**	**Version submitted to TCEQ with deliverable**
1.1	**15 June 2015**	**Revised to use forward-averaging for MDA8 calculations**

Table of Contents

List of Figures

List of Tables

1 Introduction

This technical memo documents the files provided to TCEQ to complete Deliverable 3.1 of Work Order No. 582-15-54118-01. As stated in the Work Plan, this deliverable is:

Deliverable 3.1: Daily estimates of regional background O_3 (May through October) and $PM_{2.5}$ (all year) by the TCEQ s method for 2005-2014 for the Group 1 and Group 2 metropolitan areas, as well as the state of Texas. Estimates will be provided in comma separated variable (csv) format.

The Group 1 and Group 2 metropolitan areas are listed in Table 1.

Table 1. Urban areas of interest to this study.

Group 1 Urban Areas	**Group 2 Urban Areas**
Dallas/Fort Worth (DFW)	Beaumont/Port Arthur (BPA)
Houston/Galveston/Brazoria (HGB)	Tyler-Longview-Marshall (TLM)
San Antonio (SA)	
Austin/Round Rock (ARR)	

Section 2 of this memo briefly outlines the technical approach used to prepare the files in the deliverable (provided via email to Erik Gribbin of TCEQ as a gzipped tar file: p1952_deliverable_3_1_R1_1.tar.gz) and Section 3 describes the format of the files. Section 4 briefly outlines the quality assurance steps that have been performed. Further details and analysis of the results will be included in the project Final Report.

2 Technical Approach

As described in the Work Plan, our approach follows the TCEQ method described in Berlin et al. (2013). This method requires: the selection of background sites; the calculation of the maximum daily 8-hour average (MDA8) for ozone (O_3) and the daily average of fine particulate matter ($PM_{2.5}$) at each site; estimating a preliminary background value as the lowest of the valid values for the background sites; and then further investigations to ensure the values are appropriate background estimates. These steps are described in detail below.

2.1 Selection of Background Sites

The initial data for our analysis was provided by Erik Gribbin of TCEQ, which consisted of hourly-average measurements of O_3 and $PM_{2.5}$ at several sampling sites surrounding the urban areas of interest. After consultation with TCEQ, we selected “background” monitor sites near the edge of each urban area. These background sites were chosen to be at a significant distance from major pollutant emission sources. The AQS site numbers for the selected background sites for each urban area are given in Table 2. Note that for the ARR and TLM areas, most or all of the available urban sites are considered potential “background” sites due to the limited number of sampling sites available.

Table 2. AQS site numbers for the selected background sites for each urban area.

Urban Area	Total # of Sites	# of Background Sites	AQS Site Numbers of Background Sites
DFW	28	11	481210034, 481211032, 481215008, 481391044, 482210001, 482311006, 482510003, 482570005, 483491051, 483670081, 484390075
HGB	69	31	480390618, 480390619, 480391003, 480391004, 480391016, 480710013, 481570696, 481670697, 481671034, 481675005, 482010029, 482010066, 482010552, 482010553, 482010554, 482010555, 482010556, 482010557, 482010558, 482010559, 482010560, 482010561, 482010563, 482010617, 482011042, 482011050, 482910699, 483390078, 483390698, 483395006, 483739991
SA	15	8	480290059, 480290501, 480290502, 480910503, 480910505, 481870504, 481870506, 481875004
ARR	12	8	482090675, 480210684, 481490001, 482090614, 482091675, 484530020, 484910690, 484916602
BPA	17	5	482450022, 482450101, 482450628, 483611001, 483611100
TLM	4	4	484230007, 481830001, 482030002, 480370004

To calculate the background MDA8 O_3 for the State of Texas as a whole, we used two approaches, the first using data from TCEQ sites near the Texas border, and the second using data from sites in the US EPA Clean Air Status and Trends Network (CASTNET)[1]. The CASTNet sites used to calculate Texas background O_3 are listed in Table 3; a csv file (CASTNet_site_info.csv) with the latitudes, longitudes, and elevations of the CASTNet sites is included in the deliverable.

To calculate the background daily average $PM_{2.5}$ for the State of Texas, we used data from sites in the Interagency Monitoring of Protected Visual Environments (IMPROVE)[2] network near the Texas border, as TCEQ sites near the Texas border rarely made $PM_{2.5}$ measurements. The IMPROVE sites used to calculate Texas background O_3 are listed in Table 3; a csv file

[1] U.S. Environmental Protection Agency Clean Air Markets Division, *Clean Air Status and and Trends Network (CASTNET),* Table OZONE_8HR_DMAX, last updated 2015-04-06. Available at www.epa.gov/castnet. Accessed 2015-04-09.

[2] *Interagency Monitoring of Protected Visual Environments (IMPROVE) network*, Table EPA PM2.5 Mass FRM – Daily, Available at http://vista.cira.colostate.edu/improve/Data/IMPROVE/improve_data.htm. Accessed 2015-04-09.

(IMPROVELocTable.csv) with the latitudes, longitudes, and elevations of the IMPROVE sites is included in the deliverable.

Table 3. Sites used to calculate background O_3 and $PM_{2.5}$ for the State of Texas as a whole.

Pollutant (Network)	# of Background Sites	IDs of Background Sites
O_3 (TCEQ)	23	484790017, 484790016, 484790313, 482150043, 480610006, 482730314, 483550025, 482450101, 481675005, 480391003, 482030002, 480370004, 482311006, 483670081, 480650007, 480650004, 480650005, 481351014, 481350003, 481410058, 800060003, 481410029, 481410057
O_3 (CASTNet)	10	CHA467, PET427, MEV405, CHE185, CAD150, CVL151, SUM156, EVE419, PAL190, BBE401
$PM_{2.5}$ (IMPROVE)	12	BOAP1, SAAN1, WHIT1, GUMO1, SACR1, BIBE1, ELLI1, WIMO1, CACR1, SIKE1, HOUS1, BRET1

2.2 Calculation of MDA8 O_3 and daily average $PM_{2.5}$ Values for Each Site

We developed a python script (calc_bkgrd_ozone.py)[3] that calculated the MDA8 O_3 (ppbv) for all of the monitoring sites in the six urban areas. The MDA8 for a site was calculated as follows:

1. A running 8-hour average was calculated for each hour, averaged over that hour and the following seven hours. At least 6 hours in this 8-hour range had to have valid O_3 measurements for the 8-hour average to be considered valid.
2. The largest of each of the calculated 8-hour averages in a day was selected as the MDA8 for that day.

A similar script (calc_pm.py) was used to calculate daily average $PM_{2.5}$ values from the available hourly data. This average was calculated as follows:

1. If more than one $PM_{2.5}$ instrument was active for a site, the reported hourly values were averaged.
2. A daily average $PM_{2.5}$ value was then calculated for each site. At least 18 hours of that day had to have valid $PM_{2.5}$ measurements for the daily average to be considered valid.

For the background values for the State of Texas as a whole, the MDA8 values for the TCEQ sites in Table 2 were calculated as above. The CASTNet and IMPROVE data was already provided as appropriately averaged values. The scripts calc_TX_bkgrd_ozone.py and calc_TX_bkgrd_PM25.py were used to process the data.

[3] All listed scripts will be supplied to TCEQ with the project's Final Report.

2.3 Estimating Preliminary Background Values

The lowest of the daily MDA8 O_3 values in the background sites for each urban area were selected as our preliminary background estimates. In addition, the maximum and minimum MDA8 O_3 values for all urban sites in the area were also calculated.

Similarly, the lowest of the daily average $PM_{2.5}$ values in the background sites for each urban area were selected as our preliminary background estimates. In addition, the maximum and minimum daily average $PM_{2.5}$ values for all urban sites in the area were also calculated.

For the State of Texas as a whole, we calculated separate background values for O_3 from the TCEQ sites and the CASTNet sites. In both cases the minimum valid MDA8 value was used. The minimum valid IMPROVE $PM_{2.5}$ value was used as the $PM_{2.5}$ background estimate for Texas.

2.4 Linear Regression Test and Outlier Analysis

We investigated the preliminary background estimates for each urban area by performing a linear regression of the preliminary background values (x) versus the maximum values (y) of MDA8 O_3 and daily average $PM_{2.5}$ using the R software package (using scripts bckgd_fit_o3.R and bckgd_fit_pm.R). For example, Figure 1 shows a scatterplot of the background MDA8 O_3 value versus the maximum MDA8 O_3 value for the HGB area (the other fit figures are included in Appendix A). The solid black line is the linear fit, and the dotted and dashed black lines are the upper and lower 95% (or 2σ) confidence intervals, respectively. In this example, 89 of the 1834 valid data points (4.9%) have maximum MDA8 O_3 values that fall above the upper confidence internal of the linear fit, suggesting that these background estimates are lower than would be expected given the maximum values seen in the urban area. Table 4 gives the number of such points for each urban area and pollutant. All such data points are identified in the csv files in a column called “high_flag”, with a value of TRUE meaning that day was above the upper 95% confidence interval for that day (see also Section 3). Given the skewed distribution of both the background and maximum MDA8 O_3 and daily average $PM_{2.5}$, very few points were identified below the lower confidence interval of the fit (one MDA8 O_3 value for Dallas, six $PM_{2.5}$ values for Austin, and three $PM_{2.5}$ values for Houston) and so these points are not flagged in the csv files.

In addition, for some days only one monitoring site within the urban area had a valid MDA8 O_3 or daily average $PM_{2.5}$ value, so that the maximum and preliminary background estimates were identical. These data points are identified in the csv files in a column called “eq_flag”, with a value of TRUE meaning that day only had a single site with valid data, and so the maximum and background estimates are equal (see also Section 3). For example, this is true of all background $PM_{2.5}$ estimates for the Tyler-Longview-Marshall area (TLM), as there was only a single site with valid $PM_{2.5}$ data (see Table 4). ***While we have included these data points in our background estimates for completeness, we strongly recommend that users be careful about including these points in their analyses, as they may bias the results of, for example, the average difference between the maximum and background values.***

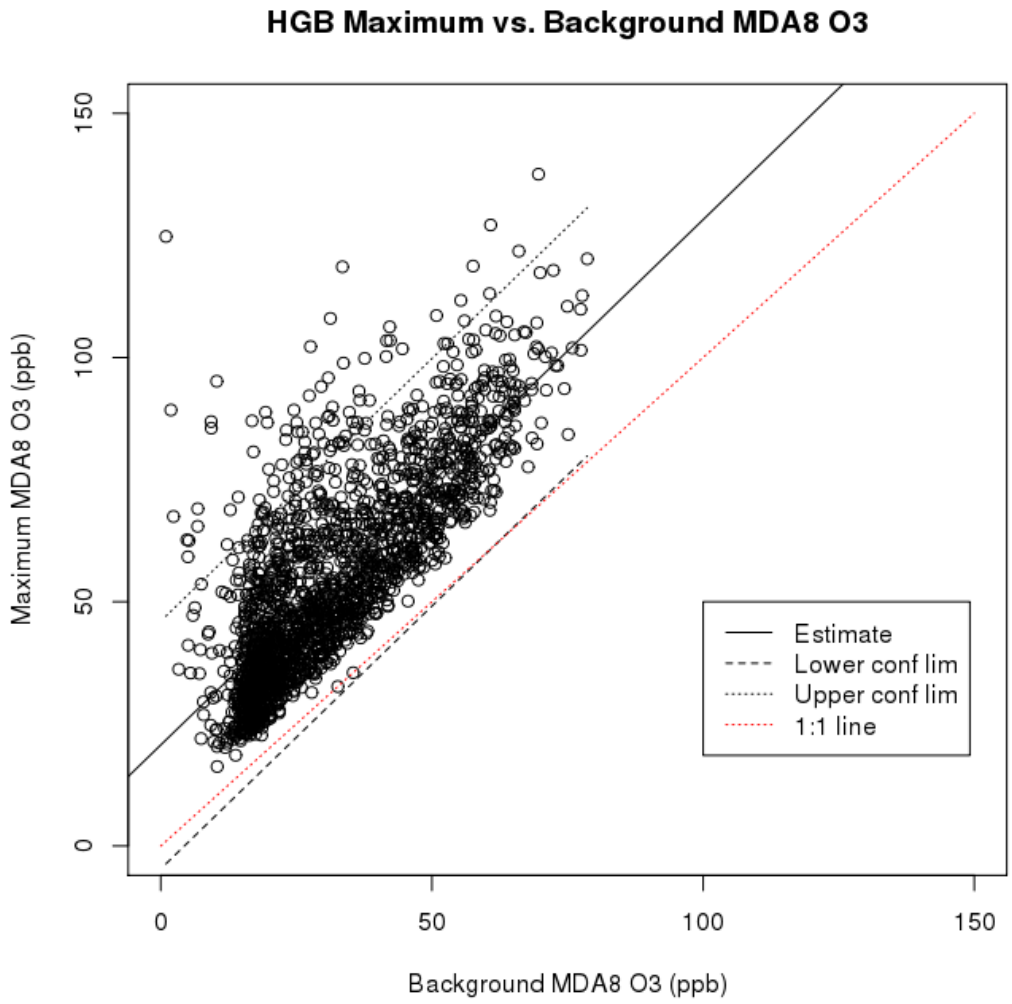

Figure 1. Maximum versus background MDA8 O_3 values for the HGB area.

Table 4. Number of background points quality flagged for each urban area and pollutant.

Urban Area	MDA8 O_3				Daily Average $PM_{2.5}$			
	# high	# eq	# final	# replaced	# high	# eq	# final	# replaced
DFW	78	0	46	0	118	21	68	0
HGB	89	2	61	43	149	11	101	75
SA	67	117	37	0	129	490	0	0
ARR	72	217	20	0	172	100	0	0
BPA	78	14	65	49	112	226	0	0
TLM	93	30	0	0	NA	3509	NA	NA

Similar to Berlin et al. (2013), we performed further analysis of the points that were above the 95% confidence interval of the fit (e.g., where high_flag = TRUE). First, we identified the subset of those points where (a) where high_flag = TRUE AND (b) at least one other background site in the urban area had a valid MDA8 O_3 or daily average $PM_{2.5}$ value for that day AND (c) the valid values at the other background sites were all more than 10% larger than the preliminary background estimate. Note that the later two criteria have to be true for replacing the preliminary background estimate with a value from a different background site to make a significant impact on any subsequent analysis. Data points that met all three criteria are flagged in the csv files in a column called "final_flag", with a value of TRUE meaning that the above criteria were satisfied (see also Section 3). The number of points with final_flag = TRUE for each urban area is shown in Table 4. For these points, we have included the AQS site number and the MDA8 O_3 or daily average $PM_{2.5}$ value for the background site with the second largest value in the csv files as an alternate background estimate.

However, we only replaced the preliminary background value if:

1. The final_flag = TRUE
2. The estimate was for the HGB or BPA areas, as these areas near the Gulf of Mexico could plausibly have times when the gulf/lake breeze front affects some of the outlying background sites, but does not affect the urban area as a whole.
3. The preliminary background site was between the city and the Gulf of Mexico (or the city and Sabine Lake). These sites are given in Table 5.

Table 5. Background sites that were replaced if final_flag = TRUE

Urban Area	AQS Site Numbers of Background Sites that could be replaced
HGB	480391016, 480390618, 481671034, 480390619
BPA	482450628, 482450101

The total number of background sites where data was replaced with the second highest value is given in Table 4. These replacements do not significantly impact the statistics of the background estimates.

For the State of Texas as a whole, we did not perform a similar sort of analysis, as there is difficulty in deciding what is the appropriate maximum value to use for the state as a whole. The preliminary background estimates are thus identical to the final background estimates.

3 File Descriptions

The data contained in the csv files included in the deliverable are described below. All files are in comma-separated-value (csv) format unless otherwise stated.

3.1 Urban Area Ozone Files (file name = *_flagged_O3_v3.csv, six files in total)

Column Descriptions:

1. **Date:** In YYYYMMDD format. Note only dates in the ozone season (May-October) will have valid values.
2. **AQS_Code_max:** AQS site number of the site in the urban area with the maximum MDA8 ozone.

3. **O3_max..ppbv.:** Maximum of the valid MDA8 ozone (ppbv) values for all sites in the urban area.
4. **AQS_Code_min_max:** AQS site number of the site with the minimum valid MDA8 ozone (ppbv) for ***all*** sites in an urban area. Note this may not be equal to the background estimate (Columns 6 and 7), as that is the minimum for the ***background*** sites only.
5. **O3_min_max..ppbv.:** Minimum of the valid MDA8 ozone (ppbv) values for ***all*** sites in an urban area. Note this may not be equal to the background estimate (Columns 6 and 7), as that is the minimum for the ***background*** sites only.
6. **AQS_Code_min:** AQS site number of the preliminary background estimate.
7. **O3_min_bkgrd..ppbv.**: The preliminary background estimate, calculated as the minimum valid MDA8 ozone (ppbv) for the background sites in an urban area.
8. **high_flag:** TRUE if this day was above the 95% confidence interval for a linear fit of the preliminary background MDA8 ozone value (x, Column 7) against the maximum MDA8 ozone value (y, Column 3). See Section 2.4.
9. **eq_flag:** TRUE if this day only had one valid MDA8 ozone value for the urban area, and so the preliminary background MDA8 ozone value (x, Column 7) and the maximum MDA8 ozone value (y, Column 3) are equal. ***We strongly recommend that users be careful about including the points flagged as TRUE in their analysis, as they may bias the results of, for example, the average difference between the maximum and background values.*** See Section 2.4.
10. **final_flag:** TRUE if (a) high_flag (Column 8) is TRUE, AND (b) at least one other background site in the urban area had a valid MDA8 ozone value for that day, AND (c) the valid values at the other background sites were all more than 10% larger than the preliminary background estimate. See Section 2.4.
11. **X2nd.Highest.MDA8.AQS.Code:** If final_flag = TRUE, this contains the AQS site number of the background site with the second lowest MDA8 value.
12. **X2nd.Highest.MDA8..ppbv.:** If final_flag = TRUE, this contains the second lowest MDA8 value of the background sites.
13. **Final.MDA8.Background.AQS.Code:** AQS site number of the final MDA8 background estimate (ppbv), with some HGB and BPA values replaced as described in Section 2.4.
14. **Final.MDA8.Background..ppbv.:** final MDA8 background estimate (ppbv), with some HGB and BPA values replaced as described in Section 2.4.

3.2 Urban Area $PM_{2.5}$ Files (file name = *_flagged_PM_v2.csv, six files in total)

Column Descriptions:

1. **Date:** In YYYYMMDD format.
2. **AQS_Code_max:** AQS site number of the site in the urban area with the maximum daily average $PM_{2.5}$ value.
3. **PM2.5_max..ug.m.3.):** Maximum of the valid daily average $PM_{2.5}$ values ($\mu g\ m^{-3}$) for all sites in the urban area.
4. **AQS_Code_min_max:** AQS site number of the site with the minimum valid daily average $PM_{2.5}$ values ($\mu g\ m^{-3}$) for ***all*** sites in an urban area. Note this may not be equal to the background estimate (Columns 6 and 7), as that is the minimum for the ***background*** sites only.

5. **PM2.5_min_max..ug.m.3:** Minimum of the valid daily average $PM_{2.5}$ values ($\mu g\ m^{-3}$) for ***all*** sites in an urban area. Note this may not be equal to the background estimate (Columns 6 and 7), as that is the minimum for the ***background*** sites only
6. **AQS_Code_min:** AQS site number of the preliminary background estimate.
7. **PM2.5_min_bkgrd..ug.m.3.:** The preliminary background estimate, calculated as the minimum valid daily average $PM_{2.5}$ values ($\mu g\ m^{-3}$) for the background sites in an urban area.
8. **high_flag:** TRUE if this day was above the $95^{\%}$ confidence interval for a linear fit of the preliminary background MDA8 ozone value (x, Column 7) against the maximum MDA8 ozone value (y, Column 3). See Section 2.4.
9. **eq_flag:** TRUE if this day only had one valid MDA8 O3 value for the urban area, and so the preliminary background daily average $PM_{2.5}$ values ($\mu g\ m^{-3}$) (x, Column 7) and the maximum daily average $PM_{2.5}$ values ($\mu g\ m^{-3}$) (y, Column 3) are equal. ***We strongly recommend that users be careful about including the points flagged as TRUE in their analysis, as they may bias the results of, for example, the average difference between the maximum and background values.*** See Section 2.4.
10. **final_flag:** TRUE if (a) high_flag (Column 8) is TRUE, AND (b) at least one other background site in the urban area had a valid daily average $PM_{2.5}$ value for that day, AND (c) the valid values at the other background sites were all more than 10% larger than the preliminary background estimate.
11. **X2nd.Highest.PM2.5.AQS.Code:** If final_flag = TRUE, this contains the AQS site number of the background site with the second lowest daily average $PM_{2.5}$ value.
12. **X2nd.Highest.PM2.5..ug.m.3.:** If final_flag = TRUE, this contains the AQS site number of the background site with the second lowest daily average $PM_{2.5}$ value.
15. **Final.PM2.5.Background.AQS.Code:** AQS site number of the final daily average $PM_{2.5}$ background estimate ($\mu g\ m^{-3}$), with some HGB and BPA values replaced as described in Section 2.4.
16. **Final.PM2.5.Background..ug.m.3.:** Final daily average $PM_{2.5}$ background estimate ($\mu g\ m^{-3}$), with some HGB and BPA values replaced as described in Section 2.4.

3.3 Texas O_3 Background (file name = TX_State_bkgrd_O3_calc.csv)

Column Descriptions:

1. **Date:** In YYYYMMDD format. Note only dates in the ozone season (May-October) will have valid values.
2. **AQS_Code_min:** AQS site number of the preliminary background estimate using TCEQ sites near the Texas border.
3. **O3_min(bkgrd):** The preliminary background estimate, calculated as the minimum valid MDA8 ozone (ppbv) for the TCEQ sites near the Texas border.
4. **CASTNet ID:** CASTNet ID code for the CASTNet site near the Texas border with the lowest MDA8 O_3 value for the day.
5. **CASTNet O3_min (bkgrd):** The CASTNet-based background estimate, calculated as the minimum valid MDA8 ozone (ppbv) for the CASTNet sites near the Texas border.

3.4 Texas $PM_{2.5}$ Background (file name = TX_State_PM_calc.csv)

Column Descriptions:

1. **Date:** In YYYYMMDD format. Note that as these estimates are based on IMPROVE network data, data is only available one out of every 3 days.
2. **IMPROVE ID:** IMPROVE ID code of the IMPROVE site near the Texas border with the lowest daily average $PM_{2.5}$ value.
3. **PM_min(bkgrd, ug/m^3):** The background estimate for the State of Texas, calculated as the minimum valid daily average $PM_{2.5}$ values ($\mu g\ m^{-3}$) for the background sites in an urban area.

4 Quality Assurance Steps

In addition to the analyses described in Section 2.4, other quality assurance checks were made. First, all scripts used in this project were independently inspected to ensure they were calculating properly, and any errors noted in early versions were fixed. Second, the statistics of the background and maximum values for each urban area were investigated to ensure that they were reasonable and did not change unexpectedly between file versions.

In addition, as we continue our work on Task 3, we will be performing further analysis of the spatial and temporal trends in the estimates of regional background O_3 and $PM_{2.5}$ (Deliverable 3.2). If that analysis uncovers any errors in the csv files, we will correct those and provide TCEQ with corrected csv files as part of our final report.

5 References

Berlin, S. R., A. O. Langford, M. Estes, M. Dong, and D. D. Parrish (2013) Magnitude, decadal changes, and impact of regional background ozone transported into the greater Houston, Texas, area, *Environ. Sci. & Technol.*, 47, 13985-92.

Appendix A Linear Fit Figures

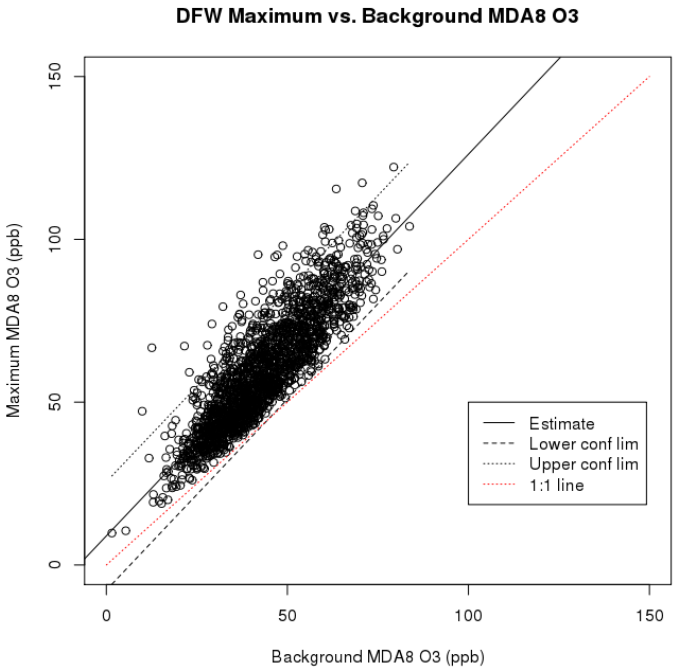

Figure 2. Maximum versus background MDA8 O_3 values for the DFW area.

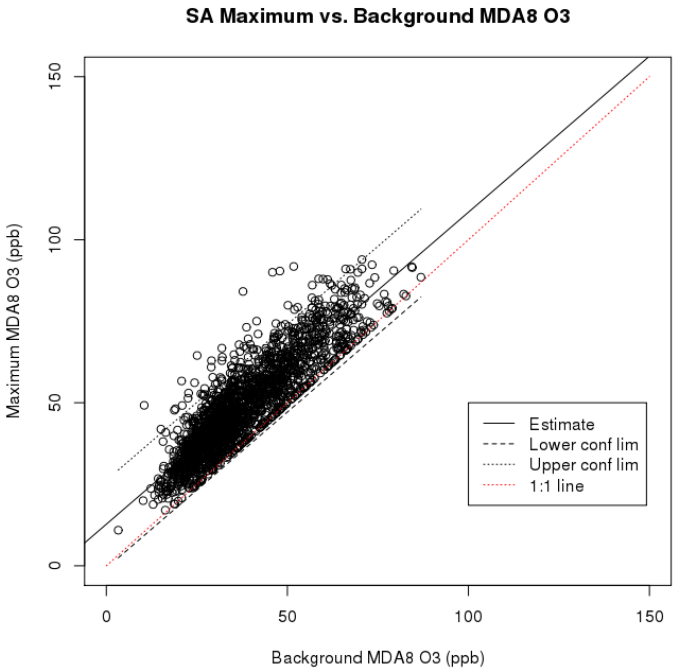

Figure 3. Maximum versus background MDA8 O_3 values for the SA area.

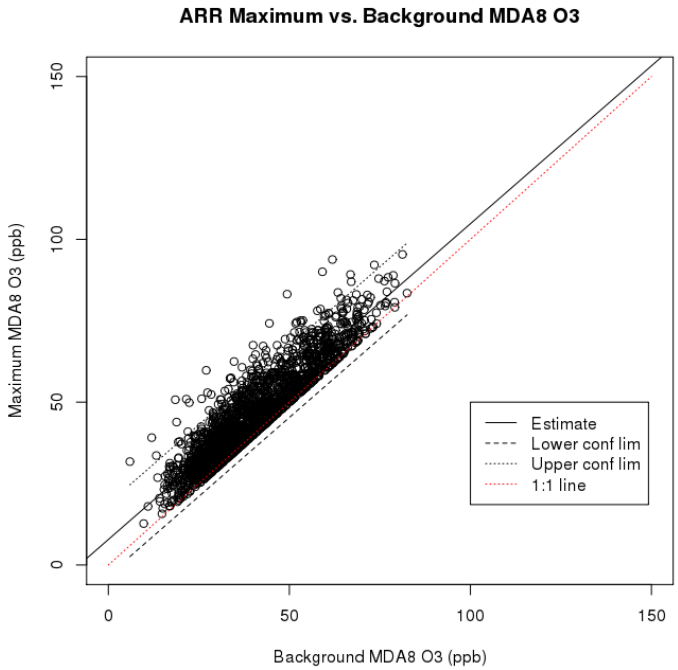

Figure 4. Maximum versus background MDA8 O_3 values for the ARR area.

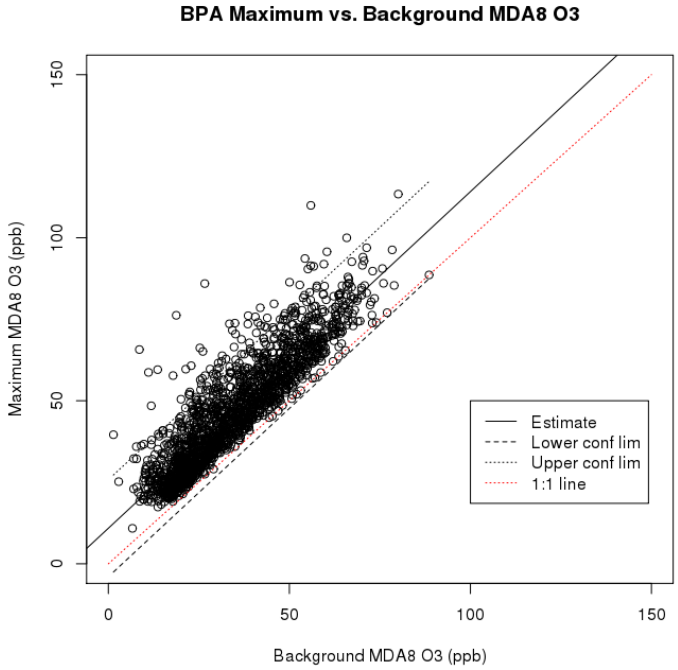

Figure 5. Maximum versus background MDA8 O_3 values for the BPA area.

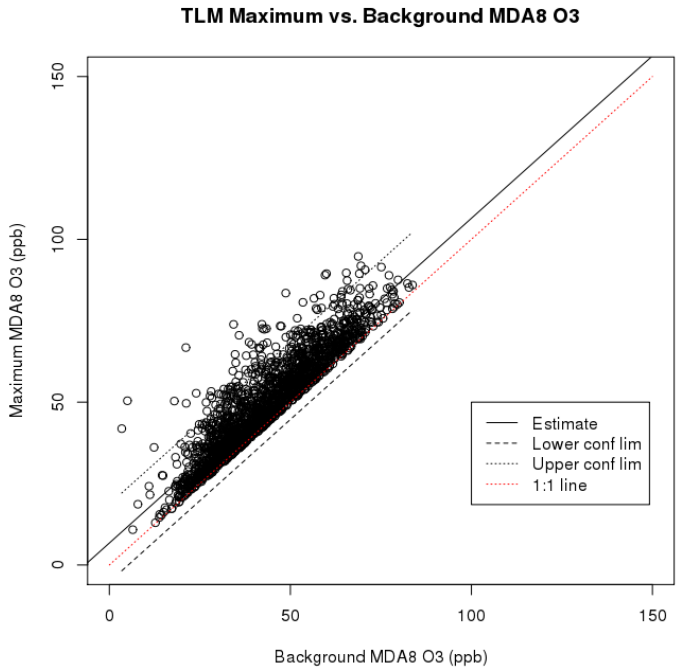

Figure 6. Maximum versus background MDA8 O_3 values for the TLM area.

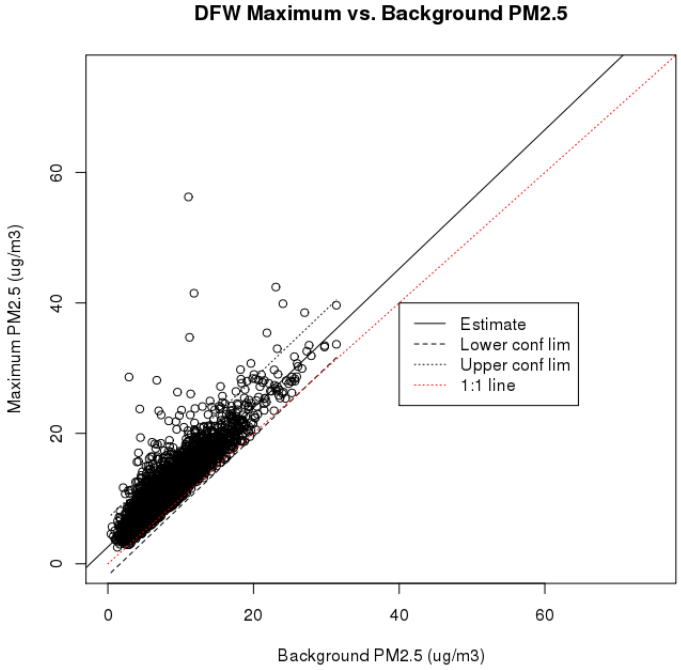

Figure 7. Maximum versus background daily average $PM_{2.5}$ values for the DFW area.

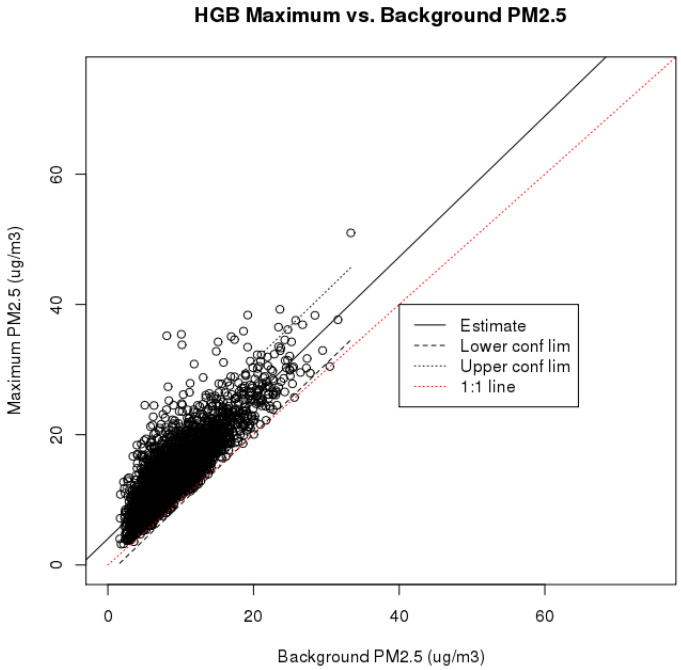

Figure 8. Maximum versus background daily average $PM_{2.5}$ values for the HGB area.

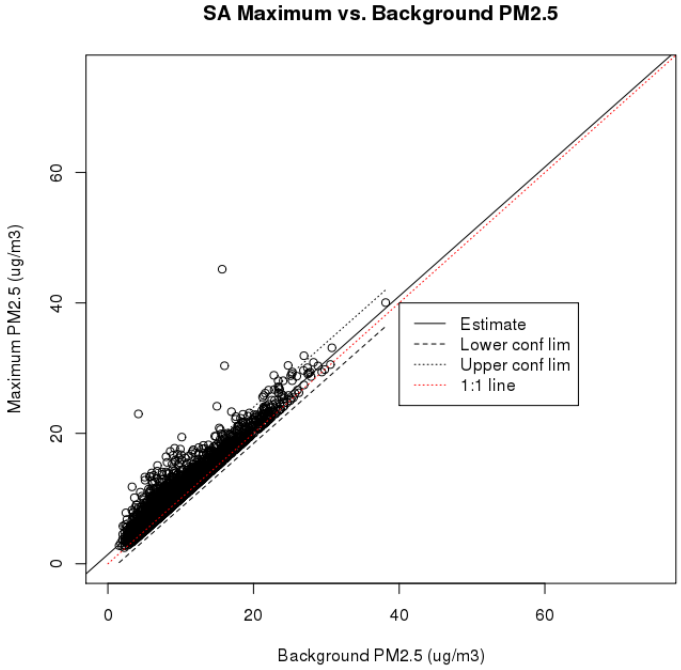

Figure 9. Maximum versus background daily average $PM_{2.5}$ values for the SA area.

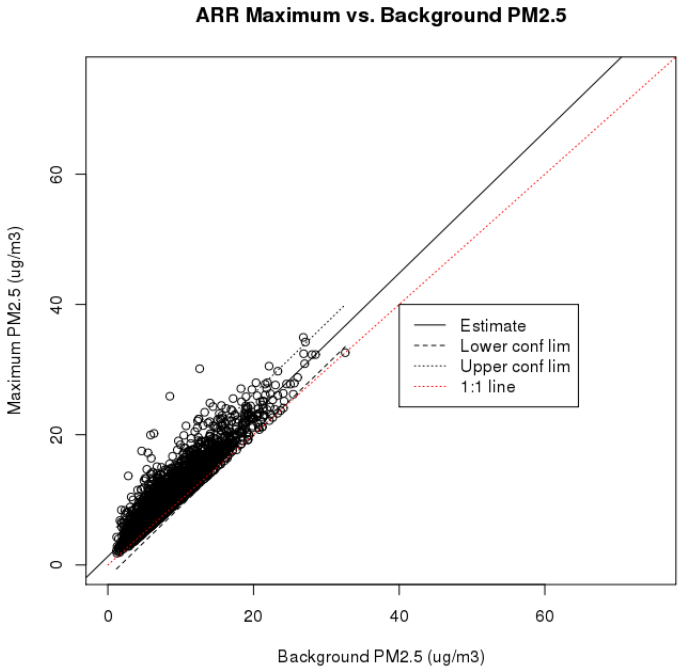

Figure 10. Maximum versus background daily average $PM_{2.5}$ values for the ARR area.

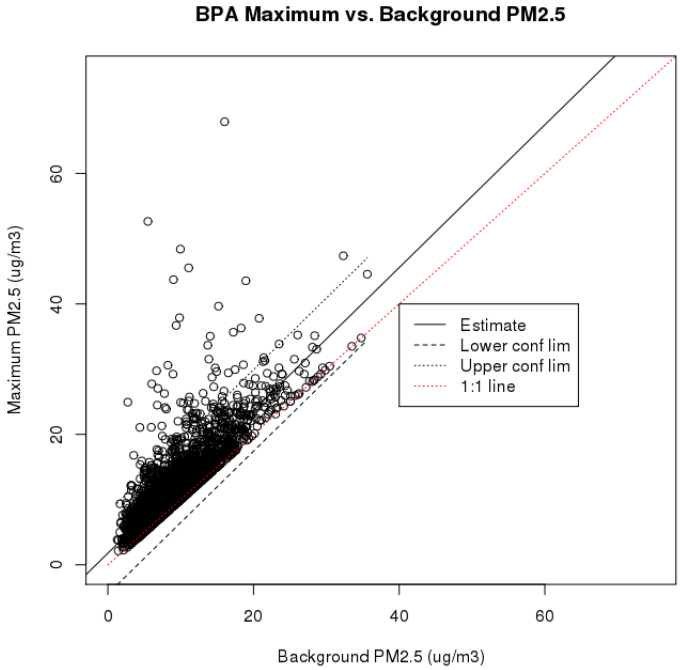

Figure 11. Maximum versus background daily average $PM_{2.5}$ values for the BPA area.

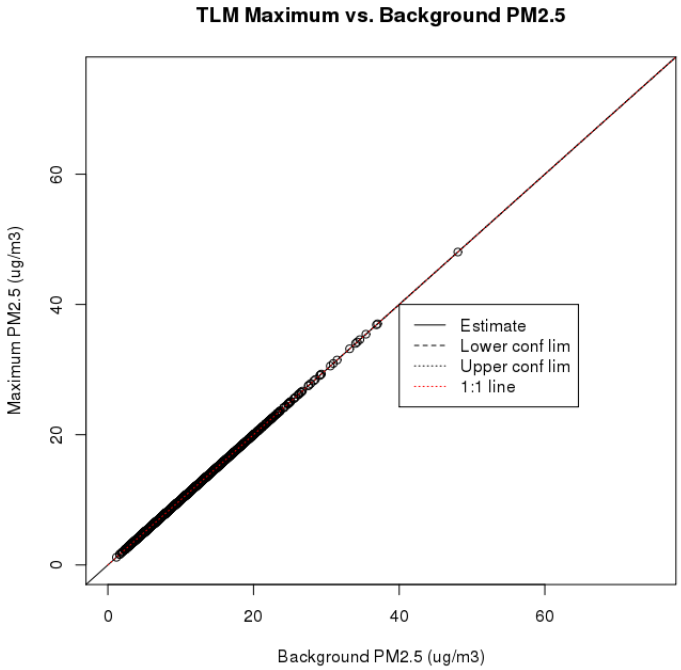

Figure 12. Maximum versus background daily average $PM_{2.5}$ values for the TLM area.