FINAL REPORT

Meteorologically Corrected Ozone, SO_2, and $PM_{2.5}$ Trends

TCEQ Contract No. 582-19-90498
Work Order No. 582-22-31570-011

Deliverable 6.2
Revision 2.0

Prepared by:

Marikate Mountain
Atmospheric and Environmental Research, Inc. (AER)
131 Hartwell Ave.
Lexington, MA 02421
Correspondence to: mmountai@aer.com

Prepared for:

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

December 6, 2022

Document Change Record

Revision	Revision Date	Remarks
0.1	**June 26, 2022**	**Internal Version for Review**
1.0	**June 30, 2022**	**Delivery to TCEQ**
2.0	**Dec 6, 2022**	**Update trend calculation in section 2.3**

TABLE OF CONTENTS

Executive Summary 17
1. Introduction 19
1.1 Project Objectives 19
1.2 Purpose and Background 19
1.2.1 Trends in O_3, $PM_{2.5}$, and SO_2 19
1.2.2 Regional Background Concentrations of O_3, $PM_{2.5}$, and SO_2 19
1.2.3 Synoptic- and Urban-scale Meteorological Controls on O_3, $PM_{2.5}$, and SO_2 ... 20
1.3 Report Outline 20
2 Task 3: Effects of Meteorology on O_3, $PM_{2.5}$, and SO_2 & 21
Task 4: Estimating Background O_3, $PM_{2.5}$, and SO_2 21
2.1 Input Data and Processing 22
2.1.1 TCEQ Monitor Data 22
2.1.1.1 MDA8 O_3 23
2.1.1.2 $PM_{2.5}$ 23
2.1.1.3 SO2 23
2.1.2 IGRA Radiosonde Data 24
2.1.3 NCDC Integrated Surface Data 24
2.1.4 NAM-12 Meteorological Data 25
2.1.5 HYSPLIT Back Trajectories 25
2.2 Generalized Additive Model 26
2.2.1 GAMs Description 27
2.2.2 MDA8 O_3 GAM Results 28
2.2.2.1 Total MDA8 O_3 GAM Results 28
2.2.2.2 Background MDA8 O_3 GAM Results 42
2.2.3 $PM_{2.5}$ GAM Results 57
2.2.3.1 Total PM2.5 GAM Results 57
2.2.3.2 Background $PM_{2.5}$ GAM Results 72
2.2.4 SO_2 GAM Results 87
2.2.4.1 Total SO_2 GAM Results 87
2.2.4.2 Background SO_2 GAM Results 102
2.2.5 Cross Validation Analysis 113

2.3 Meteorologically Adjusted Trends of O_3, $PM_{2.5}$, and SO_2 115
2.3.1 Meteorologically Adjusted Trends for Total and Background MDA8 O_3 115
2.3.2 Meteorologically Adjusted Trends for Total and Background $PM_{2.5}$ 120
2.3.3 Meteorologically Adjusted Trends for Total and Background SO_2 124
2.4 Temporal Trends of Background MDA8 O_3, $PM_{2.5}$, and SO_2 129
2.4.1 Temporal Trends of Background O_3 129
2.4.2 Temporal Trends of Background $PM_{2.5}$ 136
2.4.3 Temporal Trends of Background SO_2 142
3 Task 5: The Role and Importance of Synoptic or Mesoscale Meteorological Conditions in Creating High Ozone, $PM_{2.5}$, and SO_2 148
3.1 Synoptic Scale Conditions 148
3.2 Urban-scale conditions 153
3.3 Discussion 161
4 Quality Assurance Steps and Reconciliation with User Requirements 161
4.1 Task 3: Effects of Meteorology on Ozone, PM2.5, and SO2 162
4.2 Task 4: Estimating Background Ozone, PM2.5, and SO2 162
4.3 Task 5: The Role and Importance of Synoptic or Mesoscale Meteorological Conditions in Creating High Ozone, PM2.5, and SO2 163
5 Conclusions 163
6 Recommendations 165
7 References 167
Appendix A File Descriptions and Process Flow 169
A.1 Process Flow 169
A.2 File Descriptions 170
A.2.1 Input data *(./data/)* 170
A.2.1.1 IGRA Data (./data/IGRA2) 170
A.2.1.2 NCDC data (./data/NCDC/) 171
A.2.1.3 Monitoring data (./data/TCEQ/) 171
A.2.1.4 Meteorological data (./data/TCEQ/) 171
A.2.2 Data Processing Scripts (./scripts/) 171
A.2.3. HYSPLIT Files (./HYSPLIT/) 172
A.2.4 Processed Input Data Files in CSV Format (./csv_files/) 173
A.2.4.1 Intermediate CSV Files (./csv_files/Intermediate/) 173
A.2.4.2 Final CSV Files (./csv_files/final/) 173

A.2.5 GAM (./full_gam_fits/) 173
A.2.6 Synoptic and Mesoscale Meteorological Analysis (./MetAnalysis/) 173

LIST OF TABLES

Table 1. Projected Schedule for TCEQ Work Order No. 582-22-31570-011 21
Table 2. TCEQ monitor sites from which meteorological quantities were calculated. 22
Table 3. IGRA Sites 24
Table 4. NCDC Surface Sites 24
Table 5. Starting points for HYSPLIT back-trajectories. 25
Table 6. Dates with missing NAM files 25
Table 7. Meteorological parameters used in the GAMs. The column name is given in italics. ... 27
Table 8. Performance of GAMs for Total MDA8 O_3 28
Table 9. Performance of GAMs for Background MDA8 O_3. 43
Table 10. Performance of GAMs for Total $PM_{2.5}$. 58
Table 11. Performance of GAMs for background $PM_{2.5}$. 73
Table 12. Performance of GAMs for Total SO_2. 88
Table 13. Performance of GAMs for background SO_2 103
Table 14. Cross validation analysis 114
Table 15. Total MDA8 O_3 original trends and meteorologically adjusted linear trends. 115
Table 16. Background MDA8 O_3 original trends and meteorologically adjusted linear trends. 116
Table 17. Total $PM_{2.5}$ original trends and meteorologically adjusted linear trends 120
Table 18. Background $PM_{2.5}$ original trends and meteorologically adjusted linear trends. 120
Table 19. Total SO_2 original trends and meteorologically adjusted linear trends. 124
Table 20. Background SO_2 original trends and meteorologically adjusted linear trends. 125
Table 21. 90^{th} percentiles of observations during 2012-2021 148

LIST OF FIGURES

Figure 1. Smooth functions for the total MDA8 O_3 GAM fit in the area of Austin, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the MDA8 O_3 in ppbv from its mean value. 29

Figure 2. GAM evaluation plots for total MDA8 O_3 in the area of Austin, TX. 30

Figure 3. Smooth functions for the total MDA8 O_3 GAM fit in the area of Beaumont/Port Arthur, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the MDA8 O_3 in ppbv from its mean value. 31

Figure 4. GAM evaluation plots for total MDA8 O_3 in the area of Beaumont/Port Arthur, TX.. 32

Figure 5. Smooth functions for the total MDA8 O_3 GAM fit in the area of Corpus Christi, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the MDA8 O_3 in ppbv from its mean value. 33

Figure 6. GAM evaluation plots for total MDA8 O_3 in the area of Corpus Christi, TX. 34

Figure 7. Smooth functions for the total MDA8 O_3 GAM fit in the area of Dallas/Fort Worth, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the MDA8 O_3 in ppbv from its mean value. 35

Figure 8. GAM evaluation plots for total MDA8 O_3 in the area of Dallas/Fort Worth, TX. 36

Figure 9. Smooth functions for the total MDA8 O_3 GAM fit in the area of El Paso, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the MDA8 O_3 in ppbv from its mean value. 37

Figure 10. GAM evaluation plots for total MDA8 O_3 in the area of El Paso, TX..... 38

Figure 11. Smooth functions for the total MDA8 O_3 GAM fit in the area of Houston/Galveston/Brazoria, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the MDA8 O_3 in ppbv from its mean value. 39

Figure 12. GAM evaluation plots for total MDA8 O_3 in the area of Houston/Galveston/Brazoria, TX. 40

Figure 13. Smooth functions for the total MDA8 O_3 GAM fit in the area of San Antonio, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the MDA8 O_3 in ppbv from its mean value. 41

Figure 14. GAM evaluation plots for total MDA8 O_3 in the area of San Antonio, TX..... 42

Figure 15. Smooth functions for the background MDA8 O_3 GAM fit in the area of Austin, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background MDA8 O_3 in ppbv from its mean value. 44

Figure 16. GAM evaluation plots for background MDA8 O_3 in the area of Austin, TX. 45

Figure 17. Smooth functions for the background MDA8 O_3 GAM fit in the area of Beaumont/Port Arthur, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background MDA8 O_3 in ppbv from its mean value. 46

Figure 18. GAM evaluation plots for background MDA8 O_3 in the area of Beaumont/Port Arthur, TX..... 47

Figure 19. Smooth functions for the background MDA8 O_3 GAM fit in the area of Corpus Christi, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background MDA8 O_3 in ppbv from its mean value. 48

Figure 20. GAM evaluation plots for background MDA8 O_3 in the area of Corpus Christi, TX. 49

Figure 21. Smooth functions for the background MDA8 O_3 GAM fit in the area of Dallas/Fort Worth, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background MDA8 O_3 in ppbv from its mean value. 50

Figure 22. GAM evaluation plots for background MDA8 O_3 in the area of Dallas/Fort Worth, TX. 51

Figure 23. Smooth functions for the background MDA8 O_3 GAM fit in the area of El Paso, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background MDA8 O_3 in ppbv from its mean value. 52

Figure 24. GAM evaluation plots for background MDA8 O_3 in the area of El Paso, TX. 53

Figure 25. Smooth functions for the background MDA8 O_3 GAM fit in the area of Houston/Galveston/Brazoria, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background MDA8 O_3 in ppbv from its mean value. 54

Figure 26. GAM evaluation plots for background MDA8 O_3 in the area of Houston/Galveston/Brazoria, TX. 55

Figure 27. Smooth functions for the background MDA8 O_3 GAM fit in the area of San Antonio, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background MDA8 O_3 in ppbv from its mean value. 56

Figure 28. GAM evaluation plots for background MDA8 O_3 in the area of San Antonio, TX. 57

Figure 29. Smooth functions for the total daily average $PM_{2.5}$ GAM fit in the area of Austin, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the total daily average $PM_{2.5}$ in µg/m3 from its mean value. 59

Figure 30. GAM evaluation plots for total daily average $PM_{2.5}$ in the area of Austin, TX. 60

Figure 31. Smooth functions for the total daily average $PM_{2.5}$ GAM fit in the area of Beaumont/Port Arthur, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the total daily average $PM_{2.5}$ in µg/m3 from its mean value. 61

Figure 32. GAM evaluation plots for total daily average $PM_{2.5}$ in the area of Beaumont/Port Arthur, TX. 62

Figure 33. Smooth functions for the total daily average $PM_{2.5}$ GAM fit in the area of Corpus Christi, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the total daily average $PM_{2.5}$ in µg/m3 from its mean value. 63

Figure 34. GAM evaluation plots for total daily average $PM_{2.5}$ in the area of Corpus Christi, TX. 64

Figure 35. Smooth functions for the total daily average $PM_{2.5}$ GAM fit in the area of Dallas/Fort Worth, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the total daily average $PM_{2.5}$ in µg/m3 from its mean value. 65

Figure 36. GAM evaluation plots for total daily average $PM_{2.5}$ in the area of Dallas/Fort Worth, TX. 66
Figure 37. Smooth functions for the total daily average $PM_{2.5}$ GAM fit in the area of El Paso, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the total daily average $PM_{2.5}$ in µg/m3 from its mean value..... 67
Figure 38. GAM evaluation plots for total daily average $PM_{2.5}$ in the area of El Paso, TX..... 68
Figure 39. Smooth functions for the total daily average $PM_{2.5}$ GAM fit in the area of Houston/Galveston/Brazoria, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the total daily average $PM_{2.5}$ in µg/m3 from its mean value. 69
Figure 40. GAM evaluation plots for total daily average $PM_{2.5}$ in the area of Houston/Galveston/Brazoria, TX. 70
Figure 41. Smooth functions for the total daily average $PM_{2.5}$ GAM fit in the area of San Antonio, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the total daily average $PM_{2.5}$ in µg/m3 from its mean value. 71
Figure 42. GAM evaluation plots for total daily average $PM_{2.5}$ in the area of San Antonio, TX. 72
Figure 43. Smooth functions for the background $PM_{2.5}$ GAM fit in the area of Austin, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background $PM_{2.5}$ in µg/m3 from its mean value. 74
Figure 44. GAM evaluation plots for background $PM_{2.5}$ in the area of Austin, TX..... 75
Figure 45. Smooth functions for the background $PM_{2.5}$ GAM fit in the area of Beaumont/Port Arthur, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background $PM_{2.5}$ in µg/m3 from its mean value. 76
Figure 46. GAM evaluation plots for background $PM_{2.5}$ in the area of Beaumont/Port Arthur, TX. 77
Figure 47. Smooth functions for the background $PM_{2.5}$ GAM fit in the area of Corpus Christi, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background $PM_{2.5}$ in µg/m3 from its mean value. 78
Figure 48. GAM evaluation plots for background $PM_{2.5}$ in the area of Corpus Christi, TX..... 79
Figure 49. Smooth functions for the background $PM_{2.5}$ GAM fit in the area of Dallas/Fort Worth, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background $PM_{2.5}$ in µg/m3 from its mean value. 80
Figure 50. GAM evaluation plots for background $PM_{2.5}$ in the area of Dallas/Fort Worth, TX. . 81
Figure 51. Smooth functions for the background $PM_{2.5}$ GAM fit in the area of El Paso, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background $PM_{2.5}$ in µg/m3 from its mean value..... 82
Figure 52. GAM evaluation plots for background $PM_{2.5}$ in the area of El Paso, TX. 83
Figure 53. Smooth functions for the background $PM_{2.5}$ GAM fit in the area of Houston/Galveston/Brazoria, TX. The y-axis scale is the scale of the "linear predictor", i.e.,

the deviation of the natural logarithm of the background $PM_{2.5}$ in µg/m3 from its mean value. 84
Figure 54. GAM evaluation plots for background $PM_{2.5}$ in the area of Houston/Galveston/Brazoria, TX. 85
Figure 55. Smooth functions for the background $PM_{2.5}$ GAM fit in the area of San Antonio, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background $PM_{2.5}$ in µg/m3 from its mean value. 86
Figure 56. GAM evaluation plots for background $PM_{2.5}$ in the area of San Antonio, TX. 87
Figure 57. Smooth functions for the maximum daily 1-hour average SO_2 GAM fit in the area of Austin, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the maximum daily 1-hour average SO_2 in ppbv from its mean value. 89
Figure 58. GAM evaluation plots for maximum daily 1-hour average SO_2 in the area of Austin, TX. 90
Figure 59. Smooth functions for the maximum daily 1-hour average SO_2 GAM fit in the area of Beaumont/Port Arthur, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the maximum daily 1-hour average SO_2 in ppbv from its mean value. 91
Figure 60. GAM evaluation plots for maximum daily 1-hour average SO_2 in the area of Beaumont/Port Arthur, TX. 92
Figure 61. Smooth functions for the maximum daily 1-hour average SO_2 GAM fit in the area of Corpus Christi, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the maximum daily 1-hour average SO_2 in ppbv from its mean value. 93
Figure 62. GAM evaluation plots for maximum daily 1-hour average SO_2 in the area of Corpus Christi, TX. 94
Figure 63. Smooth functions for the maximum daily 1-hour average SO_2 GAM fit in the area of Dallas/Fort Worth, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the maximum daily 1-hour average SO_2 in ppbv from its mean value. 95
Figure 64. GAM evaluation plots for maximum daily 1-hour average SO_2 in the area of Dallas/Fort Worth, TX. 96
Figure 65. Smooth functions for the maximum daily 1-hour average SO_2 GAM fit in the area of El Paso, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the maximum daily 1-hour average SO_2 in ppbv from its mean value. 97
Figure 66. GAM evaluation plots for maximum daily 1-hour average SO_2 in the area of El Paso, TX. 98
Figure 67. Smooth functions for the maximum daily 1-hour average SO_2 GAM fit in the area of Houston/Galveston/Brazoria, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the maximum daily 1-hour average SO_2 in ppbv from its mean value. 99

Figure 68. GAM evaluation plots for maximum daily 1-hour average SO_2 in the area of Houston/Galveston/Brazoria, TX. 100
Figure 69. Smooth functions for the maximum daily 1-hour average SO_2 GAM fit in the area of San Antonio, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the maximum daily 1-hour average SO_2 in ppbv from its mean value. 101
Figure 70. GAM evaluation plots for maximum daily 1-hour average SO_2 in the area of San Antonio, TX. 102
Figure 71. Smooth functions for the background SO_2 GAM fit in the area of Beaumont, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background SO_2 in ppbv from its mean value. 104
Figure 72. GAM evaluation plots for background SO_2 in the area of Beamont, TX. 105
Figure 73. Smooth functions for the background SO_2 GAM fit in the area of Dallas/Fort Worth, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background SO_2 in ppbv from its mean value. 106
Figure 74. GAM evaluation plots for background SO_2 in the area of Dallas/Fort Worth, TX... 107
Figure 75. Smooth functions for the background SO_2 GAM fit in the area of El Paso, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background SO_2 in ppbv from its mean value. 108
Figure 76. GAM evaluation plots for background SO_2 in the area of El Paso, TX. 109
Figure 77. Smooth functions for the background SO_2 GAM fit in the area of Houston/Galveston/Brazoria, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background SO_2 in ppbv from its mean value. 110
Figure 78. GAM evaluation plots for background SO_2 in the area of Houston/Galveston/Brazoria, TX. 111
Figure 79. Smooth functions for the background SO_2 GAM fit in the area of San Antonio, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background SO_2 in ppbv from its mean value. 112
Figure 80. GAM evaluation plots for background SO_2 in the area of San Antonio, TX. 113
Figure 81. Original and meteorologically adjusted annual averages for total and background MDA8 O_3 trends in Austin, TX. 116
Figure 82. Original and meteorologically adjusted annual averages for total and background MDA8 O_3 trends in Beaumont/Port Arthur, TX. 117
Figure 83. Original and meteorologically adjusted annual averages for total and background MDA8 O_3 trends in Corpus Christi, TX. 117
Figure 84. Original and meteorologically adjusted annual averages for total and background MDA8 O_3 trends in Dallas/Fort Worth, TX. 118
Figure 85. Original and meteorologically adjusted annual averages for total and background MDA8 O_3 trends in El Paso, TX. 118

Figure 86. Original and meteorologically adjusted annual averages for total and background MDA8 O_3 trends in Houston/Galveston/Brazoria, TX. 119
Figure 87. Original and meteorologically adjusted annual averages for total and background MDA8 O_3 trends in San Antonio, TX. 119
Figure 88. Original and meteorologically adjusted annual averages for total and background $PM_{2.5}$ trends in Austin, TX. 121
Figure 89. Original and meteorologically adjusted annual averages for total and background $PM_{2.5}$ trends in Beaumont/Port Arthur, TX. 121
Figure 90. Original and meteorologically adjusted annual averages for total and background $PM_{2.5}$ trends in Corpus Christi, TX. 122
Figure 91. Original and meteorologically adjusted annual averages for total and background $PM_{2.5}$ trends in Dallas/Fort Worth, TX. 122
Figure 92. Original and meteorologically adjusted annual averages for total and background $PM_{2.5}$ trends in El Paso, TX. 123
Figure 93. Original and meteorologically adjusted annual averages for total and background $PM_{2.5}$ trends in Houston/Galveston/Beaumont, TX. 123
Figure 94. Original and meteorologically adjusted annual averages for total and background $PM_{2.5}$ trends in San Antonio, TX. 124
Figure 95. Original and meteorologically adjusted annual averages for total SO_2 trends in Austin, TX. Background SO_2 data was not available in Austin. 125
Figure 96. Original and meteorologically adjusted annual averages for total and background SO_2 trends in Beaumont/Port Arthur, TX. 126
Figure 97. Original and meteorologically adjusted annual averages for total SO_2 trends in Corpus Christi, TX. Background SO_2 data was not available in Corpus Christi. 126
Figure 98. Original and meteorologically adjusted annual averages for total and background SO_2 trends in Dallas/Fort Worth, TX. 127
Figure 99. Original and meteorologically adjusted annual averages for total and background SO_2 trends in El Paso, TX. 127
Figure 100. Original and meteorologically adjusted annual averages for total and background SO_2 trends in Houston/Galveston/Brazoria, TX. 128
Figure 101. Original and meteorologically adjusted annual averages for total and background SO_2 trends in San Antonio, TX. 128
Figure 102. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background MDA8 O_3 for Austin, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range. 129
Figure 103. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background MDA8 O_3 for Beaumont/Port Arthur, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the

whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range..... 130

Figure 104. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background MDA8 O_3 for Corpus Christi, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range. 131

Figure 105. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background MDA8 O_3 for Dallas/Fort Worth, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range..... 132

Figure 106. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background MDA8 O_3 for El Paso, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range. 133

Figure 107. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background MDA8 O_3 for Houston/Galveston/Brazoria, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range..... 134

Figure 108. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background MDA8 O_3 for San Antonio, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range. 135

Figure 109. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background $PM_{2.5}$ for Austin, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range. 136

Figure 110. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background $PM_{2.5}$ for Beaumont/Port Arthur, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range..... 137

Figure 111. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background $PM_{2.5}$ for Corpus Christi, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range. 138

Figure 112. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background $PM_{2.5}$ for Dallas/Fort Worth, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range. 139

Figure 113. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background $PM_{2.5}$ for El Paso, TX estimated using the TCEQ method. The line is the mean. Box edges

show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range. 140

Figure 114. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background $PM_{2.5}$ for Houston/Galveston/Brazoria, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range..... 141

Figure 115. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background $PM_{2.5}$ for San Antonio, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range. 142

Figure 116. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background SO_2 for Beaumont/Port Arthur, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range. 143

Figure 117. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background SO_2 for Dallas/Fort Worth, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range. 144

Figure 118. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background SO_2 for El Paso, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range. 145

Figure 119. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background SO_2 for Houston/Galveston/Brazoria, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range..... 146

Figure 120. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background SO_2 for San Antonio, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range. 147

Figure 121. NCEP/NCAR Reanalysis composite mean maps of 500mb geopotential height (m) for dates where a majority of urban areas exceed local 90th percentile thresholds. (Upper left: MDA8 O3 max, upper right: MDA8 O3 background, lower left: $PM_{2.5}$ max, lower right: $PM_{2.5}$ background)..... 150

Figure 122. NCEP/NCAR Reanalysis composite anomaly maps of 500mb geopotential height (m) for dates where a majority of urban areas exceed local 90th percentile thresholds. (Upper left: MDA8 O3 max, upper right: MDA8 O3 background, lower left: $PM_{2.5}$ max, lower right: $PM_{2.5}$ background) 151

Figure 123. NCEP/NCAR Reanalysis composite mean maps of 500mb winds (m/s) for dates where a majority of urban areas exceed local 90th percentile thresholds. (Upper left: MDA8

O3 max, upper right: MDA8 O3 background, lower left: $PM_{2.5}$ max, lower right: $PM_{2.5}$ background) 152

Figure 124. NCEP/NCAR Reanalysis composite anomaly maps of 500mb winds (m/s) for dates where a majority of urban areas exceed local 90th percentile thresholds. (Upper left: MDA8 O3 max, upper right: MDA8 O3 background, lower left: $PM_{2.5}$ max, lower right: $PM_{2.5}$ background) 153

Figure 125. NARR Reanalysis composite mean maps of mean sea level pressure (Pa) for dates where a majority of coastal urban areas exceed local 90th percentile thresholds. (Upper left: MDA8 O_3 max, upper right: MDA8 O_3 background, middle left: $PM_{2.5}$ max, middle right: $PM_{2.5}$ background, lower left: SO_2 max, lower right: SO_2 background)..... 155

Figure 126. NARR Reanalysis composite mean maps of mean sea level pressure (Pa) for dates where most inland urban areas exceed local 90th percentile thresholds. (Upper left: MDA8 O_3 max, upper right: MDA8 O_3 background, middle left: $PM_{2.5}$ max, middle right: $PM_{2.5}$ background, lower left: SO_2 max, lower right: SO_2 background) 156

Figure 127. NARR Reanalysis composite anomaly maps of 2m temperature (K) for dates where most coastal urban areas exceed local 90th percentile thresholds. (Upper left: MDA8 O_3 max, upper right: MDA8 O_3 background, middle left: $PM_{2.5}$ max, middle right: $PM_{2.5}$ background, lower left: SO_2 max, lower right: SO_2 background) 157

Figure 128. NARR Reanalysis composite anomaly maps of 2m temperature (K) for dates where most inland urban areas exceed local 90th percentile thresholds. (Upper left: MDA8 O_3 max, upper right: MDA8 O_3 background, middle left: $PM_{2.5}$ max, middle right: $PM_{2.5}$ background, lower left: SO_2 max, lower right: SO_2 background) 158

Figure 129. NARR Reanalysis composite anomaly maps of 1000mb winds (m/s) for dates where most coastal urban areas exceed local 90th percentile thresholds. (Upper left: MDA8 O_3 max, upper right: MDA8 O_3 background, middle left: $PM_{2.5}$ max, middle right: $PM_{2.5}$ background, lower left: SO_2 max, lower right: SO_2 background) 159

Figure 130. NARR Reanalysis composite anomaly maps of 1000mb winds (m/s) for dates where most coastal urban areas exceed local 90th percentile thresholds. (Upper left: MDA8 O_3 max, upper right: MDA8 O_3 background, middle left: $PM_{2.5}$ max, middle right: $PM_{2.5}$ background, lower left: SO_2 max, lower right: SO_2 background) 160

Figure 131. Flow chart showing the processing from the original data sources (green boxes) to the final CSV file (red box) that is used as input for the GAM scripts..... 169

Figure 132. Flow chart showing the processing from the input CSV file generated at the end of Figure 131 (red box) to the GAM output files (yellow box). 170

List of Acronyms

AER – Atmospheric and Environmental Research
BG – Background
CAMx – Comprehensive Air Quality Model with Extensions
CMAQ – Community Multi-scale Air Quality Model
COMB – COMBined dataset for El Paso and Cuidad Juárez
CSV – Comma Separated Value
CTM – Chemical Transport Model
ELP – Dataset for El Paso only
GAM – Generalized Additive Model
GEO-CAPE – GEOstationary Coastal and Air Pollution Events
GLM – Generalized Linear Model
HYSPLIT – Hybrid Single Particle Lagrangian Integrated Trajectory Model
IQR – Inter-Quartile Range
LIDORT – Linearized Discrete ORdinate Radiative Transfer model
MDA8 – maximum daily 8-hour average ozone
MODIS – Moderate Resolution Imaging Spectroradiometer
MT – Map Type
NAAQS – National Ambient Air Quality Standards
NCEP – National Centers for Environmental Prediction
O_3 – Ozone
OE – Optimal Estimation
OLS – Ordinary Least Squares
OMI – Ozone Monitoring Instrument
OSSE – Observing System Simulation Experiment
PC1 – Principle Component 1
PCA – Principle Component Analysis
$PM_{2.5}$ – Particulate Matter with diameter below 2.5 microns
ppbv – Parts Per Billion by Volume
QAPP – Quality Assurance Project Plan
RH – Relative Humidity
TCEQ – Texas Commission on Environmental Quality
TES – Tropospheric Emission Spectrometer
URBE – Un-Biased Risk Estimator
UTC – Coordinated Universal Time

Executive Summary

The purpose of this project was (a) to determine the effects of meteorology on trends in O_3, $PM_{2.5}$, and SO_2 by developing new generalized additive models (GAM) for O_3, $PM_{2.5}$, and SO_2 concentrations to selected meteorological variables for a ten year period (2012-2021) for seven urban areas in Texas (Austin, Beaumont/Port Arthur, Corpus Christi, Dallas/Fort Worth, El Paso, Houston/Galveston/Brazoria, and San Antonio), (b) to estimate the regional background concentrations of O_3, $PM_{2.5}$, and SO_2 for those same seven urban areas, and (c) to investigate the synoptic and urban-scale meteorological conditions that are associated with high concentrations of background and total O_3, $PM_{2.5}$, and SO_2 in El Paso.

As the formation and loss of pollutants such as O_3, $PM_{2.5}$, and SO_2 are strongly influenced by meteorology, inter-annual trends in these pollutants represent a combination of changes due to inter-annual variability in meteorology and changes due to air quality policy actions and other economic and societal trends. Statistical techniques are thus used to account for the effect that meteorological variations have on the trends of O_3, $PM_{2.5}$, and SO_2 so that the adjusted trends can be used to assess the effectiveness of air quality policy. A common approach to performing this "meteorological adjustment" is to use a generalized additive model (GAM) (Wood, 2006) to describe the potentially nonlinear relationship between measured urban O_3 (maximum daily 8-hour average, or MDA8), $PM_{2.5}$ (daily average), or SO_2 (maximum daily 1-hour average) concentrations and selected meteorological variables taken from an array of candidate meteorological variables (e.g., Camalier et al., 2007). While TCEQ has had such a model developed for other urban areas in Texas for the 2005-2014 time period (Alvarado et al., 2015), updated models are required to accurately reflect more recent years.

Daily surface concentrations of O_3, $PM_{2.5}$, and SO_2 in urban areas can be considered as the sum of O_3, $PM_{2.5}$, and SO_2 produced within the urban area (either through primary emissions of $PM_{2.5}$ and SO_2 or through secondary chemical production of O_3, $PM_{2.5}$, and SO_2) and a "regional background" that is transported into the urban area by the large-scale circulation. Accurate estimates of this regional background are critical to determining the potential for further reductions in O_3, $PM_{2.5}$, and SO_2 concentrations in urban areas through control of local emissions of primary $PM_{2.5}$ and SO_2 and the precursors of O_3, $PM_{2.5}$, and SO_2.

We also performed basic research into what synoptic- and urban-scale meteorological conditions are important in explaining and forecasting high concentrations of O_3, $PM_{2.5}$, and SO_2 in the El Paso urban area. The goal was to identify necessary and/or sufficient meteorological conditions that lead to NAAQS exceedances or other high concentration events (e.g., above 90th percentile) for these pollutants. Meteorological conditions leading to both high regional background levels and high total levels of O_3, $PM_{2.5}$, and SO_2 were identified. Meteorological re-analyses were used in this task, in conjunction with the map composite tools developed by the NOAA Physical Sciences Laboratory.

The seven GAMs for total MDA8 O_3 generally explain 53-72% of the deviance (i.e. variability), while the seven GAMs for background MDA8 O_3 generally explain 51-66% of the deviance, consistent with the results of Camalier et al. (2007) and Alvarado et al. (2015). The GAMs also generally show good fits with normally-distributed residuals and little dependence of the residual variance on the predicted value. The seven GAMs for total $PM_{2.5}$ only explain 14-34% of the deviance (i.e. variability) and the seven GAMs for background $PM_{2.5}$ only explain 21-38% of the deviance. They generally show a poor fit with long, positive residual tails and a strong dependence of the variance of the residuals on the predicted value. The seven GAMs for total SO_2

explain 14-56% of the deviance (i.e. variability) and the five GAMs for background SO_2 explain 31-62% of the deviance. They generally show a poor fit with long, positive residual tails and a strong dependence of the variance of the residuals on the predicted value. GAMs for background SO_2 in Austin and Corpus Christi were not run due to a lack of background data.

After meteorological adjustment via the GAMs fit to total and background MDA8 O_3, total and background $PM_{2.5}$, and total and background SO_2, no trends in pollutant metrics between 2012-2021 were observed to be significant at a 95% confidence level.

Background MDA8 O_3 is fairly constant with month between May and August for Dallas/Fort Worth and El Paso, but has a July minimum in Austin, Beaumont/Port Arthur, Corpus Christi, Houston/Galveston/Brazoria, and San Antonio. In contrast, background $PM_{2.5}$ peaks in June and July. The range of values for a given month or year is large for all cities, with Corpus Christi having the largest $PM_{2.5}$ spread and the most outliers. Background SO_2 values in Beaumont/Port Arthur, Dallas/Forth Worth, Houston/Galveston/Brazoria, and San Antonio have a large range with many outliers, and displaying no discernible seasonal trend. However, El Paso values have a much smaller spread and show a minimum in summer months.

We calculated the 90^{th} percentiles for total and background MDA O_3, total and background $PM_{2.5}$, and total and background SO_2 for each of the seven urban areas. We found dates that exceeded this threshold in the majority of the urban areas, in the majority of coastal urban areas, or in the majority of inland urban areas. We created composite mean and composite anomaly maps for these categories of dates to analyze the synoptic-scale and mesoscale meteorological patterns associated with high pollutant days. The differences in synoptic and mesoscale patterns in the anomaly fields in section 3, are consistent with our understanding that high value days for different pollutants are driven by a combination of meteorological conditions and source placement, which vary by pollutant.

We recommend that future work focus on:

- Quantifying the impact of the relative sparsity of SO_2 observations on the robustness of our conclusions.
- Creating similar tools to those of NOAA's physical sciences laboratory to use with a higher resolution model, such as NAM-12km fields, and analyze the mean composite and composite anomalies at a finer scale.
- The heteroskedastic patterns observed in the GAM residuals plots for $PM_{2.5}$ and SO_2 would benefit from further study. Determining how the log of the concentrations is distributed and performing weighted regressions are possible avenues for further study.
- Developing GAMs to provide forecasts of air quality for each of the seven urban areas.
- Comparing these GAMs derived from monitor network data with similar GAMs fit to meteorological and chemical data from 3D Eulerian air quality models like CAMx and CMAQ to determine if these models accurately represent the dependence of O_3, $PM_{2.5}$, and SO_2 concentrations, and the probability of high O_3, $PM_{2.5}$, and SO_2 events, on meteorology. Differences discovered between the two sets of GAMs could point towards missing physics or incorrect parameterizations in the current Eulerian air quality models.

1. Introduction

1.1 Project Objectives

The objectives of this project were to:

- Determine the effects of meteorology on trends in O_3, $PM_{2.5}$, and SO_2 by developing new generalized additive models (GAM) that relate O_3, $PM_{2.5}$, and SO_2 concentrations to selected meteorological variables for seven urban areas in Texas: Austin, Beaumont/Port Arthur, Corpus Christi, Dallas/Fort Worth, El Paso, Houston/Galveston/Brazoria, and San Antonio.
- Estimate the regional background concentrations of O_3, $PM_{2.5}$, and SO_2 for those same seven urban areas in Texas.
- Investigate the synoptic and urban-scale meteorological conditions that are associated with high concentrations of background and total O_3, $PM_{2.5}$, and SO_2 in the seven urban areas in Texas.

Table 1 summarizes the main tasks and deliverables for the project.

1.2 Purpose and Background

1.2.1 Trends in O_3, $PM_{2.5}$, and SO_2

As the formation and loss of pollutants such as O_3, $PM_{2.5}$, and SO_2 are strongly influenced by meteorology, inter-annual trends in these pollutants represent a combination of changes due to inter-annual variability in meteorology and changes due to air quality policy actions and other economic and societal trends. Statistical techniques are thus used to account for the effect that meteorological variations have on the trends of O_3, $PM_{2.5}$, and SO_2 so that the adjusted trends can be used to assess the effectiveness of air quality policy. A common approach to performing this "meteorological adjustment" is to use a generalized additive model (GAM, Wood, 2006) to describe the potentially non-linear relationship between measured O_3 (maximum daily 8-hour average, or MDA8), $PM_{2.5}$ (daily average), or SO_2 (maximum daily 1-hour average) and selected meteorological variables (e.g., Camalier et al., 2007). In this project, AER derived GAMs for urban O_3, $PM_{2.5}$, and SO_2 for Austin, Beaumont/Port Arthur, Corpus Christi, Dallas/Fort Worth, El Paso, Houston/Galveston/Brazoria, and San Antonio following the procedures used in two previous projects (WO #582-18-81763-07 and WO #582-15-54118-01, Alvarado et al., 2015). To the extent possible, the variables used in the meteorological adjustments were kept similar so that the adjusted trends in different urban areas in Texas could be compared. AER used these models to account for the effect that meteorological variations have on the trends of O_3, $PM_{2.5}$, and SO_2.

1.2.2 Regional Background Concentrations of O_3, $PM_{2.5}$, and SO_2

Daily surface concentrations of O_3, $PM_{2.5}$, and SO_2 in urban areas can be considered as the sum of O_3, $PM_{2.5}$, and SO_2 produced within the urban area (either through primary emissions of $PM_{2.5}$ and SO_2 or through secondary chemical production of O_3, $PM_{2.5}$, and SO_2) and a "regional background" that is transported into the urban area. Accurate estimates of this regional background are critical to determining the potential for further reductions in O_3, $PM_{2.5}$, and SO_2 concentrations in urban areas through control of local emissions of primary $PM_{2.5}$ and SO_2 and the precursors of O_3, $PM_{2.5}$, and SO_2.

In this project, AER determined daily regional background estimates of O_3, $PM_{2.5}$, and SO_2 for a ten-year period (2012-2021) for Austin, Beaumont/Port Arthur, Corpus Christi, Dallas/Fort Worth, El Paso, Houston/Galveston/Brazoria, and San Antonio using the TCEQ method (i.e., the

lowest value observed at defined "background" sites near the border of the area of interest, Berlin et al., 2013). AER then used the background estimates to investigate the spatial and temporal trends of regional background O_3, $PM_{2.5}$, and SO_2.

1.2.3 Synoptic- and Urban-scale Meteorological Controls on O_3, $PM_{2.5}$, and SO_2

There are a variety of synoptic- and urban-scale meteorological conditions, some of which are important in explaining and forecasting high concentrations of O_3, $PM_{2.5}$, and SO_2 in the seven urban areas. The goal of this task was to identify necessary and/or sufficient meteorological conditions that lead to NAAQS exceedances or other high concentration events (e.g., above 90th percentile) for these pollutants. Meteorological conditions leading to both high regional background levels and high total levels of O_3, $PM_{2.5}$, and SO_2 were identified. Meteorological re-analyses were used in this task, in conjunction with the map composite tools developed by the NOAA Physical Sciences Laboratory.

1.3 Report Outline

This Draft Report documents the methods and pertinent accomplishments of this project, including comprehensive overviews of each task, a summary of the data collected and analyzed during this work, key findings, shortfalls, limitations and recommended future tasks. It satisfies Deliverable 6.1 of the Work Plan for Work Order No. 582-22-31570-011:

Deliverable 6.1: Draft Report delivered electronically via file transfer protocol or e-mail in Microsoft Word format and PDF format

Deliverable Due Date: June 1, 2022

This report contains three sections that describe the methods and major findings for Task 3 (Effects of Meteorology on O_3, $PM_{2.5}$, and SO_2, Section 2), Task 4 (Estimating Background O_3, $PM_{2.5}$, and SO_2, Section 3) and Task 5 (The Role and Importance of Synoptic or Mesoscale Meteorological Conditions in Creating High O_3, $PM_{2.5}$, and SO_2 Days, Section 4). Section 5 discusses the Quality Assurance performed for the project, including answers to the assessment questions from the Quality Assurance Project Plan (QAPP). Section 6 summarizes the conclusions of our study, and Section 7 lists our recommendations for further research. In addition, Appendix A describes the files that are included in the final deliverable package (Deliverables 3.1, 4.1, and 5.1).

Table 1. Projected Schedule for TCEQ Work Order No. 582-22-31570-011

Milestones	Planned Date
Task 1 - Work Plan	
1.1: TCEQ-approved Work Plan	January 6, 2022
1.2: TCEQ-approved QAPP	January 6, 2022
Task 2 – Progress Reports	
2.1: Monthly Progress Reports	Monthly with invoice
Task 3 – Effects of Meteorology on Ozone, $PM_{2.5}$, and SO_2	
3.1: A dataset in *.csv file format of the model estimates of ozone MDA8, daily average $PM_{2.5}$, and maximum daily one-hour average SO_2 alongside the observed in each urban area modeled	June 1, 2022
Task 4 – Estimating Background Ozone, $PM_{2.5}$, and SO_2	
4.1: A *.csv formatted dataset for each urban area modeled containing the estimated daily regional background estimate for each pollutant	June 1, 2022
Task 5 – The Role and Importance of Synoptic or Mesoscale Meteorological Conditions in Creating High Ozone, $PM_{2.5}$, and SO_2	
5.1: A short technical memo describing any synoptic or mesoscale meteorological conditions or variables found to play an important role in predicting high ozone, $PM_{2.5}$, or SO_2 conditions in the seven urban areas and why those conditions or variables are important	June 1, 2022
Task 6 – Draft and Final Reports	
6.1: Draft Report	June 1, 2022
6.2: Final Report	June 30, 2022

2 Task 3: Effects of Meteorology on O_3, $PM_{2.5}$, and SO_2 & Task 4: Estimating Background O_3, $PM_{2.5}$, and SO_2

As described in the Work Plan, AER derived GAMs for O_3, $PM_{2.5}$, and SO_2 for selected monitoring sites near seven urban areas: Austin, Beaumont/Port Arthur, Corpus Christi, Dallas/Fort Worth, El Paso, Houston/Galveston/Brazoria, and San Antonio. For O_3, only data during the O_3 season (March to October) was analyzed, but $PM_{2.5}$ and SO_2 data for the entire year was analyzed. The O_3 season was expanded beyond the May to October period used in Alvarado et al. (2015) as the mean O_3 concentrations in May were higher than those in October, and extending the season to March gave a more symmetric variation of O_3 concentrations across the season.

AER fit the data to the eight meteorological parameters that were determined to give the best fit based on a previous project (Alvarado et al., 2015) and our recent work on air quality forecasting with GAMs in Texas urban areas (Pernak et al., 2017). We also ran HYSPLIT back-trajectories

for each of the seven urban areas, following the approach of Alvarado et al. (2015), but as the date range of interest for this project is later than that in Alvarado et al. (2015) and thus higher-resolution meteorological data is available for the whole period, we used the 12-km resolution NAM-12 meteorology to drive HYSPLIT instead of the 32-km resolution data from NARR. This should result in more accurate estimates of the path of background air impacting the each of the seven urban areas.

One of the dangers of using GAMs to perform the meteorological adjustment of pollutant trends is the possibility of "over-fitting," where some of the variability that is actually due to changes in air quality policy is accounted for in the GAM by the meteorological variables. AER explored the potential errors from over-fitting via cross validation. In cross validation, some of the data (the testing set) is removed before building the GAM. The remaining data (the training set) is used to derive the GAM parameters. The testing set can then be used to test the performance of the GAM in predicting "unseen" data (e.g., Starkweather et al., 2011).

2.1 Input Data and Processing

2.1.1 TCEQ Monitor Data

The TCEQ provided AER with air quality and meteorological monitoring data covering a ten-year period (2012-2021) from the air quality monitoring network operated by the TCEQ, its grantees, or local agencies whose data is stored in the Texas Air Monitoring Information System (TAMIS) in and near each of the seven urban areas in Texas. AER then used previously built Python scripts that processed the TCEQ air quality and meteorological data and calculated the average (daily, morning, afternoon, etc.) and derived quantities (e.g., deviations from 10-year monthly averages) needed for the GAMs. Following Camalier et al. (2007) and previous projects (Alvarado et al., 2015; Pernak et al., 2017), these average and derived meteorological quantities were calculated using a single surface site in the center of the urban area combined with the nearest radiosonde location available. The single selected surface sites for each urban area are given in Table 2.

Table 2. TCEQ monitor sites from which meteorological quantities were calculated.

Urban Area	Site #	Latitude (deg)	Longitude (deg)
Houston/Galveston/Brazoria	482011035	29.73374	-95.2576048
Dallas/Fort Worth	484391002	32.80582	-97.3565229
San Antonio	480290055	29.40729	-98.431251
Austin/Round Rock	484530014	30.35494	-97.7617291
Beaumont/Port Arthur	482450009	30.03647	-94.0710877
El Paso	481410044	31.76569	-106.455232
Corpus Christi	483550025	27.76534	-97.4342604

Two additional python scripts (*calc_GLM_all.py* and *calc_GLM_NCDC.py*) were used to calculate the potential meteorological predictors. The TCEQ monitor data, Integrated Global Radiosonde Archive data (IGRA, Section 2.1.2) and the integrated surface data (ISD) of the National Climatic Data Center (NCDC, Section 2.1.3), along with the previously calculated MDA8 O_3, $PM_{2.5}$, and SO_2 maximum and minimum concentrations and parameter from the HYSPLIT

back trajectories (Section 2.1.5), were merged by a final script (*merge_param_all_Camalier.py*). This script then outputs the final CSV file used in the GAMs.

2.1.1.1 MDA8 O_3

We developed a python script (*calc_o3.py)* that calculated the MDA8 O_3 (ppbv) for all of the monitoring sites. Background sites were chosen based on their distance from the approximate center point of the study. The MDA8 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.
3. The maximum and minimum of the valid MDA8 O_3 values for all sites in the urban area were determined.
4. The minimum of the valid MDA8 O_3 values for the selected background sites were determined as the daily background concentration for that area.

2.1.1.2 $PM_{2.5}$

A similar script (*calc_pm25.py*) was used to calculate daily average $PM_{2.5}$ values from the available hourly data. Background sites were chosen based on their distance from the approximate center point of the study. 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.
3. The maximum and minimum of the valid $PM_{2.5}$ values for all sites in the urban area were determined.
4. The minimum of the valid $PM_{2.5}$ values for the selected background sites were determined as the daily background concentration for that area.

2.1.1.3 SO2

Another similar script (*calc_so2.py*) was used to calculate maximum daily 1-hour average SO_2 values from the available hourly data. Background sites were chosen based on their distance from the approximate center point of the study. This maximum was calculated as follows:

1. If more than one SO_2 instrument was active for a site, the reported hourly values were averaged.
2. A maximum daily 1-hour average SO_2 value was then calculated for each site. At least 18 hours of that day had to have valid SO_2 measurements for the daily maximum to be considered valid.
3. The maximum and minimum of the valid SO_2 values for all sites in the urban area were determined.
4. The minimum of the valid SO_2 values for the selected background sites were determined as the daily background concentration for that area.

2.1.2 IGRA Radiosonde Data

The Integrated Global Radiosonde Archive (IGRA Version 2) provided upper atmosphere data used to derive the meteorological predictors for the GAMs. These data can be downloaded at ftp://ftp.ncdc.noaa.gov/pub/data/igra. The relevant measurements include the geopotential height, temperature, and dewpoint depression at several altitudes with - 99999 values as missing. Table 3 describes the sites selected for this case based on proximity to the center of each of the seven urban areas and having continuous data for the 2012-2021 period.

Table 3. IGRA Sites

Urban Area	ID	Station Name	Lat.	Lon.
Dallas/Fort Worth	72249	FORT WORTH	32.8	-97.3
Houston/Galveston/Brazoria	72240	LAKE CHARLES	30.12	-93.22
San Antonio	72261	DEL RIO	29.37	-100.92
Austin/Round Rock	72261	DEL RIO	29.37	-100.92
Beaumont/Port Arthur	72240	LAKE CHARLES	30.12	-93.22
El Paso	72364	SANTA TERESA	31.8728	-106.6981
Corpus Christi	72251	CORPUS CHRISTI/INT.	27.7789	-97.5056

2.1.3 NCDC Integrated Surface Data

We have also added data from the integrated surface data (ISD) of the National Climatic Data Center (NCDC) to our dataset. Due to the new format for these files, the script calc_GLM_NCDC.py required modifications for our use. We used the NCDC data to get estimates of surface pressure and relative humidity, as this data was not generally available in the TCEQ dataset. The NCDC sites used are described in Table 4 below. These sites were selected because they are the closest site to the center of their respective urban areas and have continuous data for the 2012-2021 period.

Table 4. NCDC Surface Sites

Urban Area	USAFWBAN_ID	Station Name	Lat.	Lon.
DFW	722590 03927	DALLAS/FT WORTH INTERNATIONAL	32.898	-97.019
HGB	722430 12960	G BUSH INTERCONTINENTAL AP/HOU	29.98	-95.36
SAT	722530 12921	SAN ANTONIO INTERNATIONAL AIRP	29.544	-98.484
ARR	722544 13958	AUSTIN-CAMP MABRY ARMY NATIONA	30.321	-97.76
BPA	722410 12917	SOUTHEAST TEXAS REGIONAL AIRPO	29.951	-94.021
ELP	722700 23044	EL PASO INTERNATIONAL AIRPORT	31.811	-106.376
CCH	722510 12924	CORPUS CHRISTI INTERNATIONAL AIRP	27.7742	-97.5122

2.1.4 NAM-12 Meteorological Data

The higher spatial resolution North American Mesoscale Forecast System (NAM) 12-km data was used in this project instead of the North American Reanalysis (NARR) meteorological data used in Alvarado et al. (2015), as the NAM-12 data is available for the entire period of interest here on a 6 hourly 12-km grid. The NAM is one of the primary vehicles by which NCEP's Environmental Modeling Center provides mesoscale guidance to public and private sector meteorologists. It is prepared using the Weather Research and Forecasting (WRF) model initialized with a 6-h Data Assimilation (DA) cycle with hourly analysis updates. The NAM data can be downloaded from NOAA's public server at ftp://arlftp.arlhq.noaa.gov/pub/archives/nam12.

2.1.5 HYSPLIT Back Trajectories

We ran 24-hour HYSPLIT back-trajectories for the 2012-2021 period. These back-trajectories were calculated using the 12 km horizontal resolution NAM data, as these data were available in a form suitable to drive HYSPLIT for our entire study period (2012-2021). As in Camalier et al. (2007), these back-trajectories are calculated assuming an initial height of 300 m above ground level (AGL) and are started at noon local solar time. The starting point for the back-trajectories sites are given in Table 5. The HYSPLIT model (Draxler and Hess, 1997, 1998) is available for download from the HYSPLIT website (http://ready.arl.noaa.gov/HYSPLIT.php). The performance of HYSPLIT driven with NAM meteorological fields has been evaluated with tracer release studies (e.g., Hegarty et al., 2013).

Table 5. Starting points for HYSPLIT back-trajectories.

Urban Area	Site Name	Site #	Lat (deg)	Lon (deg)
Houston/Galveston/Brazoria	Clinton	482011035	29.7337409	-95.2576048
Dallas/Fort Worth	Fort Worth Northwest	484391002	32.8058182	-97.3565229
San Antonio	San Antonio Northwest	480290032	29.5150543	-98.6201886
Austin/Round Rock	Austin North Hills Drive	484530014	30.3549371	-97.7617291
Beaumont/Port Arthur	Beaumont Downtown	482450009	30.0364651	-94.0710877
El Paso	El Paso International Airport		31.811	-106.376
Corpus Christi	Corpus Christi West	483550025	27.7653364	-97.4342604

The endpoints of the back-trajectories were used to calculate the 24-hour transport direction and distance for each urban area for the 2012-2021 period. This was done using the R functions *bearing* and *distMeeus* from the *geosphere* package (see the script *./HYSPLIT/calc_trajec.R*, described in Section A.2.3). The function *bearing* gets the initial bearing (direction; azimuth) to go from point 1 to point 2 following the shortest path (a Great Circle). The function *distMeeus* calculates the shortest distance between two points (i.e., the 'great-circle-distance' or 'as the crow flies') using the WGS84 ellipsoid.

Note the NAM-12km meteorology files were not available or were incomplete for certain dates. As a result, HYSPLIT could not be run for dates requiring these files as listed in Table 6.

Table 6. Dates with missing NAM files

Dates with missing/incomplete	dates with no HYSPLIT

NAM-12km files	back-trajectories
8/17/2013	8/16-17/2013
11/29/2015	11/28-29/2015
3/20/2017	3/20/2017
9/8-9/2020	9/7-9/2020
1/21/2021	1/20-21/2021

The HYSPLIT back-trajectories used in the model development appear reasonable and are generally consistent with the surface wind speed and direction measured near the center of the area. The HYSPLIT back-trajectory distance is generally correlated with the urban area average surface wind speed with a linear correlation coefficient (R) of 0.44 to 0.59, depending on the urban area. The peak frequency of the daily average wind direction compared to the peak frequency of the HYSPLIT back-trajectory bearings is: Austin: 170°, 150°; Beaumont/Port Arthur: 170°, 150°; Corpus Christi: 150°, 150°; Dallas/Fort Worth: 170°, 170°; El Paso: 270°, 290°; Houston/Galveston/Brazoria: 170°, 150°; San Antonio: 150°, 150.

2.2 Generalized Additive Model

The easiest way to understand the GAM approach is to contrast it with two related, but simpler, approaches: ordinary linear models and generalized linear models. In an ordinary linear model (e.g., Wood, 2006, p. 12), the model equation is:

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} \quad \mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{I_n}\sigma^2)$$

where $\boldsymbol{\mu}$ is a vector of the expected values of the observation vector, $\mathbf{y}$, (both of dimension N_{obs}), which is assumed to be normally distributed around the expected values with a constant variance of σ^2. $\mathbf{X}$ is a matrix of predictor variables (dimension N_{obs} by N_{preds}), and $\boldsymbol{\beta}$ is the (initially unknown) vector of best-fit coefficients for the predictor variables. Note that this functional form is not as limited as it first appears. For example, known non-linear functions of the predictor variables (e.g., $x_i^2, \sin\frac{x_i^2}{x_j^3}$) can be used as new predictor variables, and the observation vector $\mathbf{y}$ can be similarly transformed to make it normally distributed (e.g., taking the logarithm of a log-normally distributed observation).

However, ordinary linear models have two inherent limitations. The first is the requirement that the observation be distributed according to a normal distribution. This rules out the use of ordinary linear models to predict observations that follow other distributions, such as when you wish to predict the probability that the result of an experiment will be true or false based on a set of predictors (e.g., logistic regression), and thus your observations are expected to follow a binomial distribution. Generalized linear models (GLM) (Wood, 2006, p. 59) relax this normality requirement so that distributions of any exponential family (Poison, Binomial, Gamma, Normal) can be used, as well as a set of "link" functions – smooth, monotonic functions of the expected value vector $\boldsymbol{\mu}$.

The second limitation of ordinary (and generalized) linear models is that they require that the functional dependence of the observation on the predictor variables be specified ahead of time, with only the linear coefficients $\boldsymbol{\beta}$ of those functions allowed to vary. This makes these approaches less useful where the functional form of the response is not known, or where it might be highly complex. In this case, a generalized *additive* model can be used (Wood, 2006, p. 121). The response of each predictor variable is expected to be a non-linear but smooth function constructed

as a linear sum of group of simpler basis functions of the predictor. By fitting the coefficients of these basis functions, one can estimate the previously unknown smooth function of the predictor. Cubic splines are generally used as the basis functions, as this ensures the resulting smooth function is continuous up to the second derivative.

In our procedure, we fit the total MDA8 O_3 value, the daily average $PM_{2.5}$ value, and the maximum daily 1-hour average for each urban area using the GAM function in the *mgcv* package (Wood, 2006) in R (R Core Team, 2015). The GAM can be written as follows:

$$g(\mu_i) = \beta_o + f_1(x_{i,1}) + f_2(x_{i,2}) + \cdots f_n(x_{i,n}) + f_p(D_i) + W_d + Y_k$$

where *i* is the *i*th day's observation, $g(\mu_i)$ is the "link" function (here, a log link is used), $x_{i,j}$ are the *n* meteorological predictors fit, with the corresponding $f_j(x_{i,j})$ being a (initially unknown) smooth function of $x_{i,j}$ made from a cubic-spline basis set. Following Camalier et al. (2007), three non-meteorological predictors are also included: a smooth function $f_p(D_i)$ of the Julian day of the year (D_i); a factor for the day of the week W_d and a factor for the year Y_k. As we are only fitting O_3 data during the O_3 season (March-October), $f_p(D_i)$ is built with a non-periodic cubic spline basis for O_3, but for $PM_{2.5}$ and SO_2 a periodic cubic spline basis is used. To reduce the possibility of over-fitting the data, we set the "gamma" parameter to 1.4 for these fits, as recommended by Wood (2006).

2.2.1 GAMs Description

GAMs were developed (gam_CV.R) at each of the 7 urban areas for O_3 MDA8, background O_3 MDA8, daily average $PM_{2.5}$, background $PM_{2.5}$, maximum daily 1-hour average SO_2, and background SO_2. Due to a lack of background data, the GAM fitting for background SO_2 concentrations could not be performed for Corpus Christi and Austin. The meteorological parameters used were informed by past work. In a previous project (Alvarado et al., 2015) AER described three different GAMs that related meteorological variables to measured MDA8 O_3 and $PM_{2.5}$. Starting with the meteorological parameters suggested by Camalier et al. (2007) and comparing results using different meteorological parameters, we found some were more significant than others, and selected the variables that were highly significant for most of the Texas urban areas studied as our common set of predictor variables (Alvarado et al., 2015). Further work on using GAMs to forecast O_3 in Texas urban areas found that using the water vapor density in g/m^3 as a predictor gave better performance than using dew point or relative humidity (Pernak et al., 2016, 2017). Thus, for the seven urban areas of interest here, we used the predictors identified by Alvarado et al. (2015) but with the humidity variable replaced with the water vapor density (Table 7). This was done to keep the results of the meteorological adjustment consistent for the different Texas urban areas to allow comparisons of trends between the areas. In all but one urban area, the difference between morning temperature at 925mb and the surface was used to estimate the impact of atmospheric stability. However, the surface in El Paso is commonly at a lower pressure, so we used the difference in the 700mb temperature instead to represent the lower atmospheric stability more accurately.

Table 7. Meteorological parameters used in the GAMs. The column name is given in italics.

Afternoon mean temperature (°C, *afternoon_mean_T, 1-4 PM CST*)
Diurnal temperature change (°C, *diurnal_T*)

Daily average wind speed (m/s, *daily_ws*)
Daily average wind direction (degrees clockwise from North, *daily_wd*)
Daily average water vapor density (g/m^3, *SWVP*)
Morning surface temperature difference (1200 UTC) (temperature at 925 mb–temperature at surface at 1200 UTC) (ºC, *T_dif_925mb*) or for El Paso: (temperature at 700 mb–temperature at surface at 1200 UTC) (ºC, *T_dif_700mb*)
Transport direction (degrees clockwise from North, *HYSPLIT_Bearing*)
Transport distance (m, *HYSPLIT_dist*)

2.2.2 MDA8 O_3 GAM Results

2.2.2.1 Total MDA8 O_3 GAM Results

Table 8 organizes the total MDA8 O_3 GAM results for each of the seven urban areas. This includes figure numbers for the results, the percentage deviance of the MDA8 O_3 values explained by the GAM model, and the level statistical significance of the meteorological predictors in table 7. Odd number figures are the smooth functions from the GAM fit of the natural logarithm of the total MDA8 O_3 values to the meteorological predictors. 95% confidence intervals are shown in red in the smooth functions. The day of year (*doy*) function is also shown. The percentage deviance values can be compared to the Camalier et al. (2007) results, which showed the predictive power of their models (measured by the R^2 statistic) to be between 0.56 and 0.80 for the cities in that study. Even number figures are the standard GAM evaluation plots (made with the *gam.check* function in the R mgcv package). The plots for all seven urban areas indicate a good fit, as the model residuals are roughly normally distributed and show no trend versus predicted value. The variance of the residuals is lower for low values of the predictor, but this reflects the fact that the measured MDA8 O_3 values cannot go below 0.

Table 8. Performance of GAMs for Total MDA8 O_3

Urban Area	Figure for Smooth Functions	Figure for GAM Evaluation	% Deviance Explained	Significance of Meteorological Predictors
Austin	1	2	68.6	8 terms at α=0.001
Beaumont/Port Arthur	3	4	63.6	7 terms at α=0.001 1 term at α=0.01
Corpus Christi	5	6	64.5	7 terms at α=0.001 1 term at α=0.01
Dallas/Fort Worth	7	8	72.6	7 terms at α=0.001 1 term at α=0.01
El Paso	9	10	53.9	7 terms at α=0.001 1 term at α=0.05
Houston/Galveston/Brazoria	11	12	67	8 terms at α=0.001
San Antonio	13	14	69.8	8 terms at α=0.001

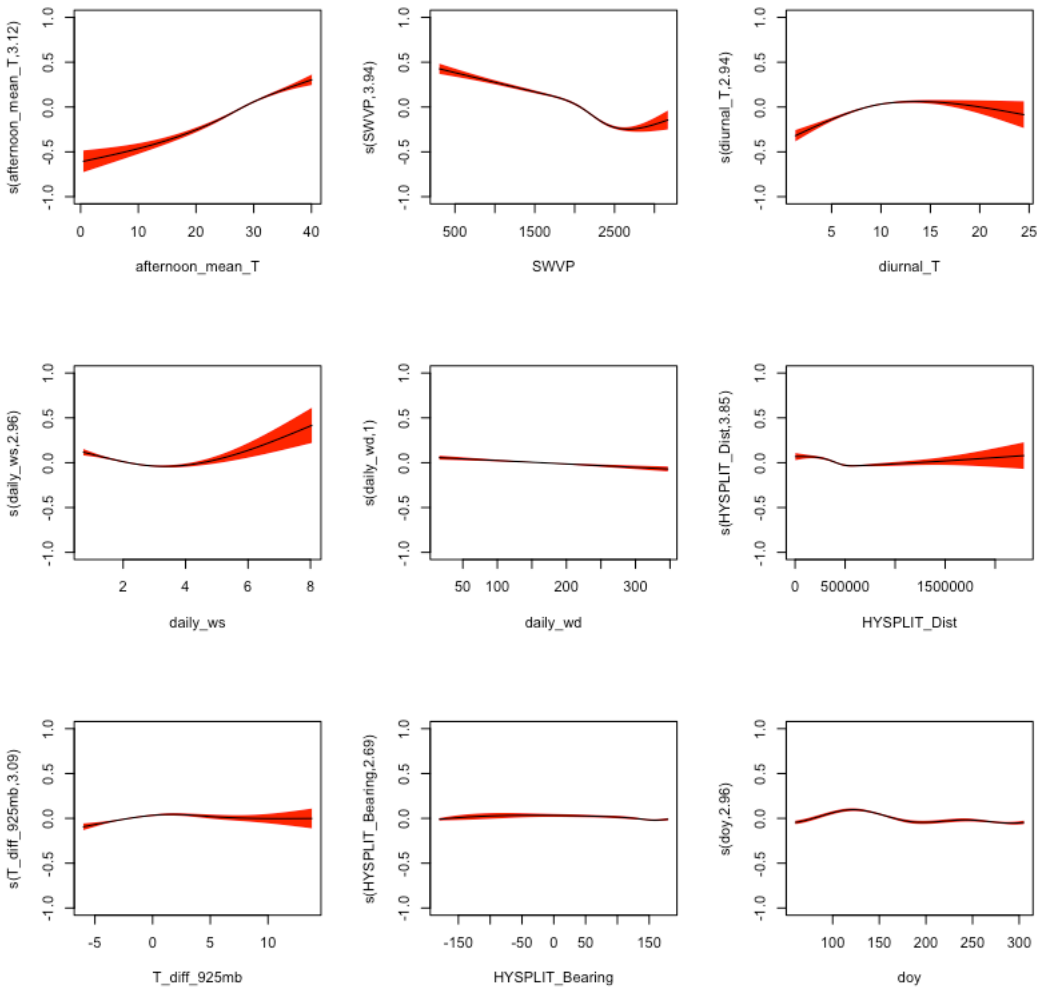

Figure 1. Smooth functions for the total MDA8 O_3 GAM fit in the area of Austin, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the MDA8 O_3 in ppbv from its mean value.

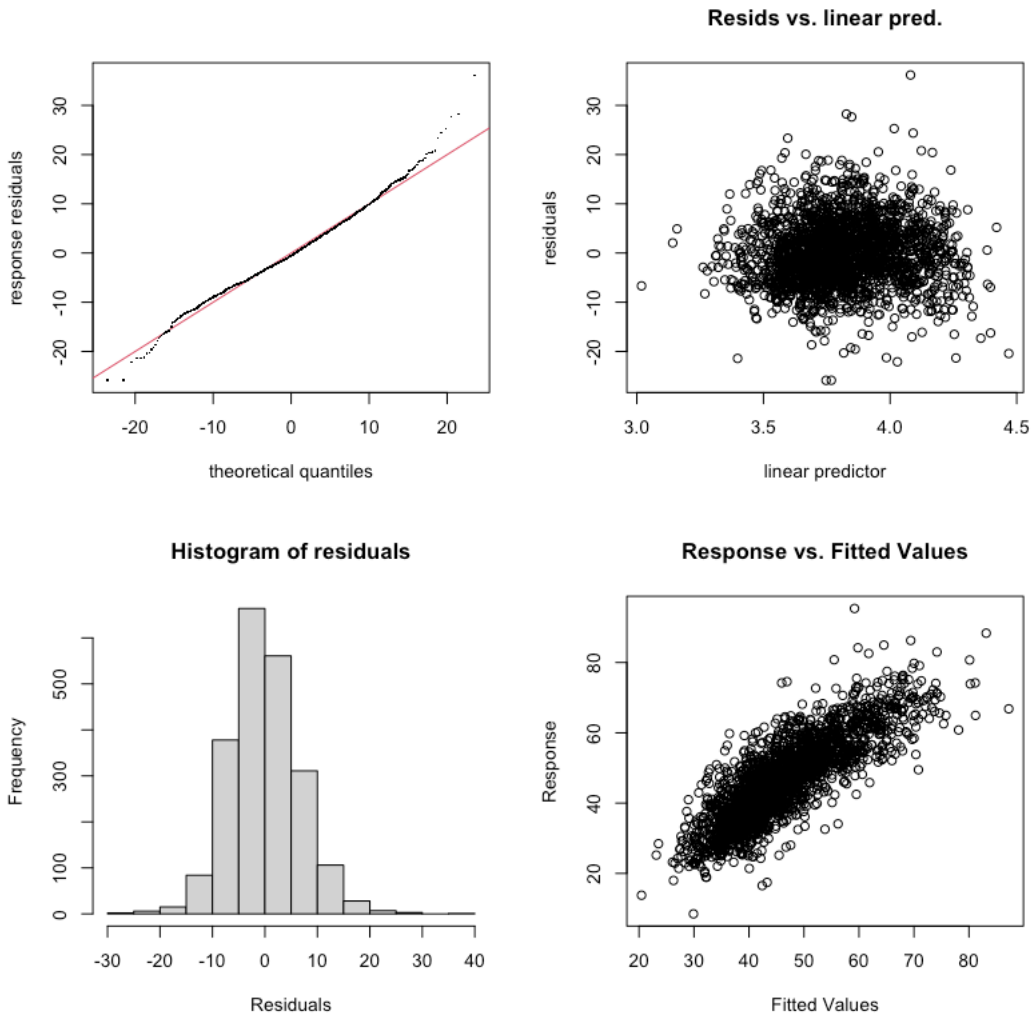

Figure 2. GAM evaluation plots for total MDA8 O_3 in the area of Austin, TX.

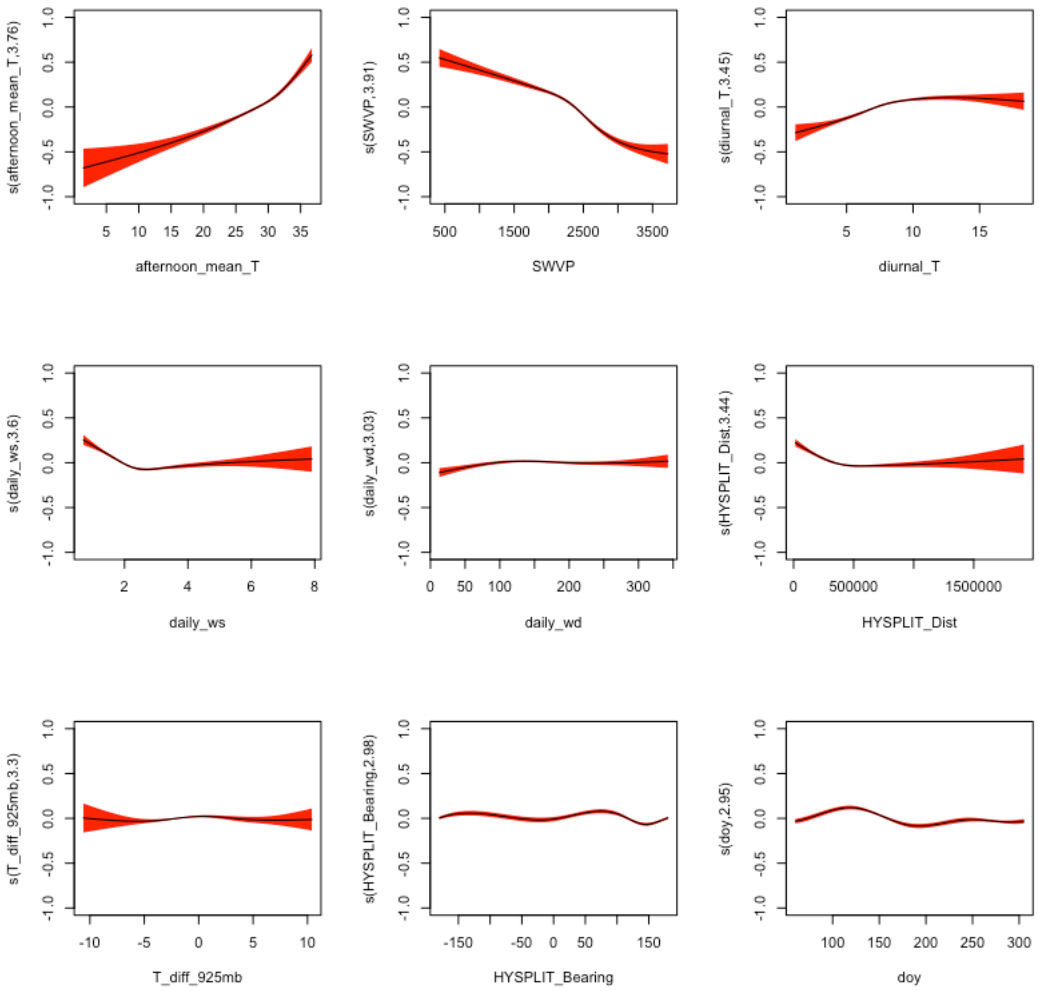

Figure 3. Smooth functions for the total MDA8 O_3 GAM fit in the area of Beaumont/Port Arthur, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the MDA8 O_3 in ppbv from its mean value.

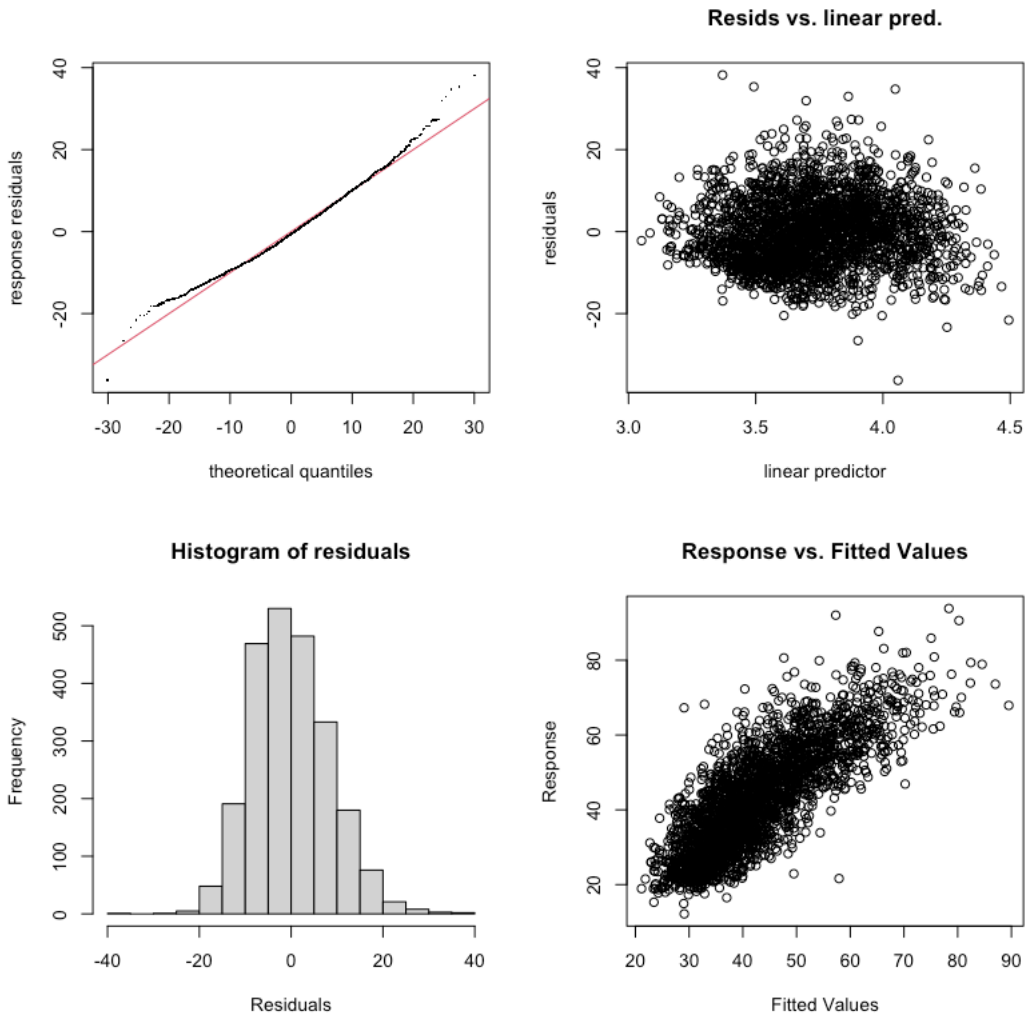

Figure 4. GAM evaluation plots for total MDA8 O_3 in the area of Beaumont/Port Arthur, TX.

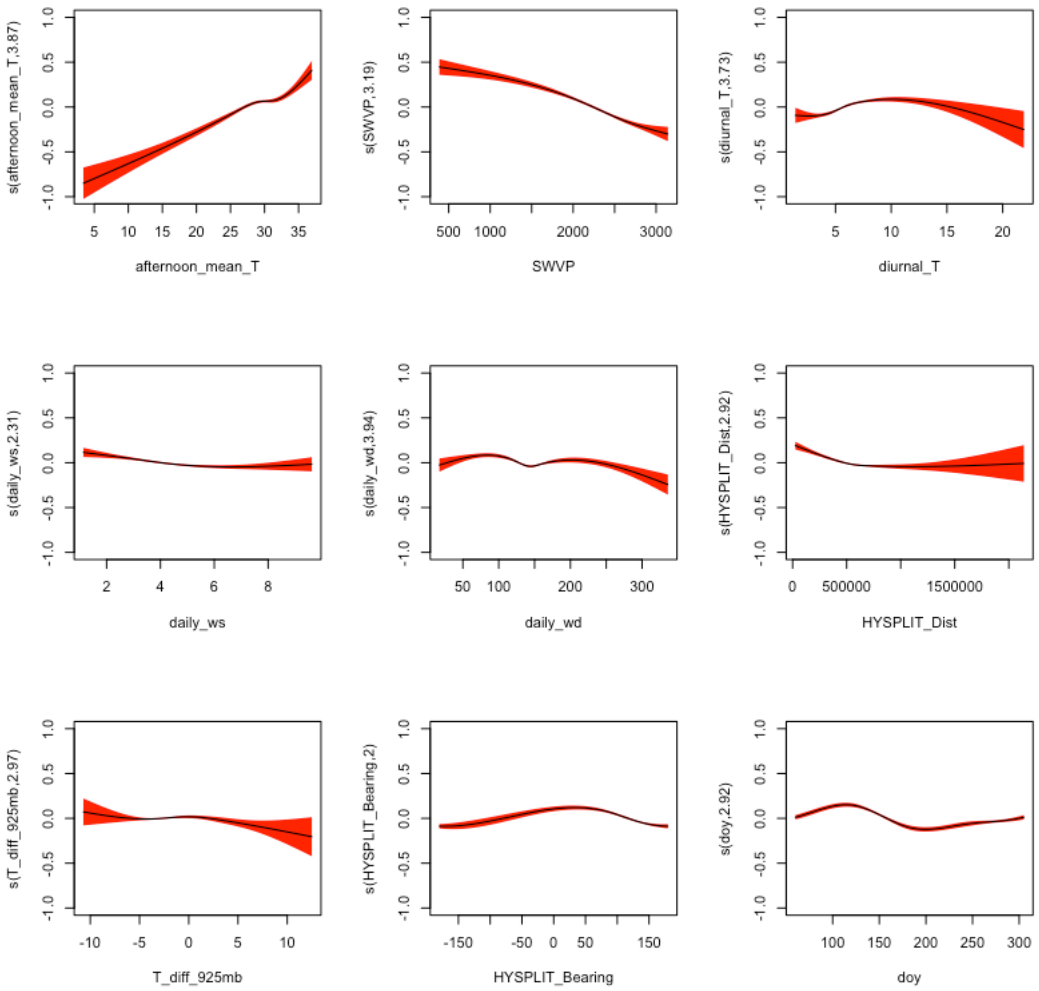

Figure 5. Smooth functions for the total MDA8 O_3 GAM fit in the area of Corpus Christi, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the MDA8 O_3 in ppbv from its mean value.

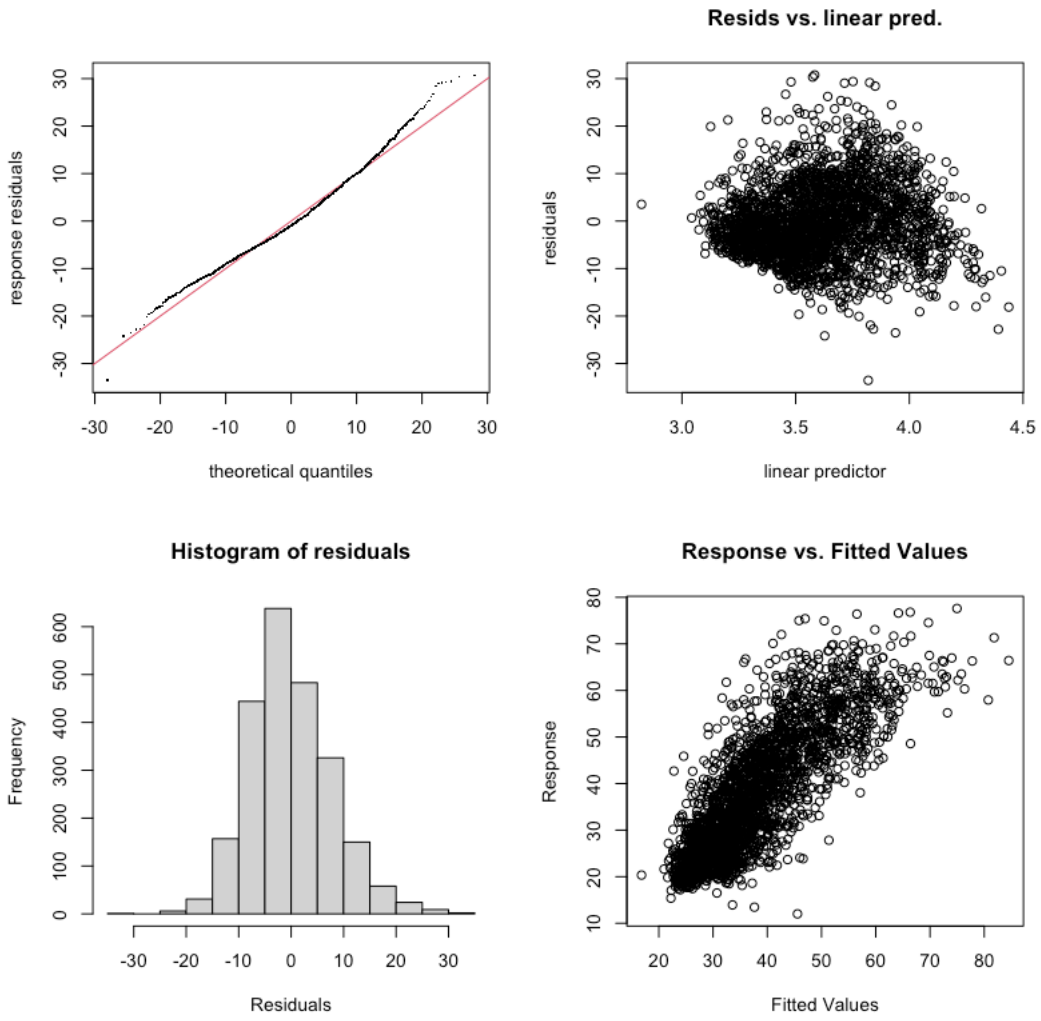

Figure 6. GAM evaluation plots for total MDA8 O_3 in the area of Corpus Christi, TX.

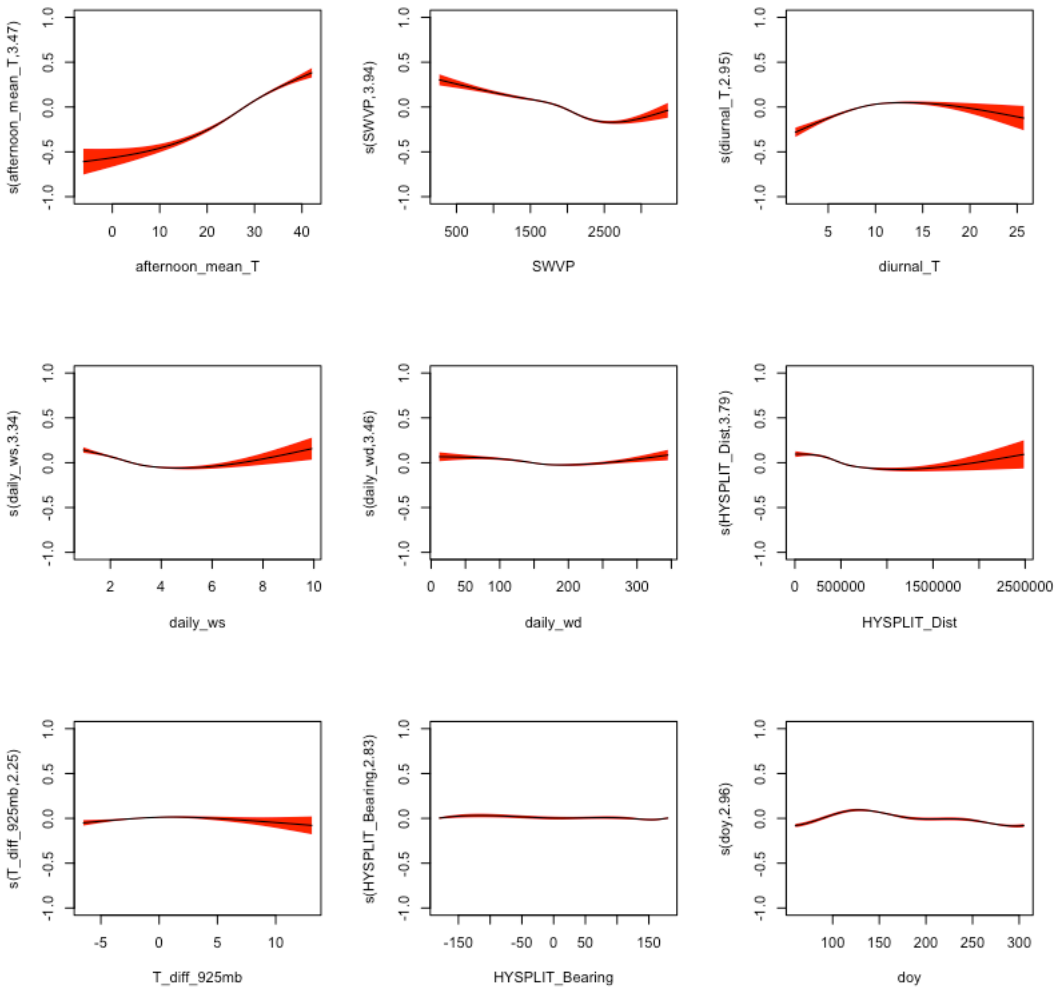

Figure 7. Smooth functions for the total MDA8 O_3 GAM fit in the area of Dallas/Fort Worth, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the MDA8 O_3 in ppbv from its mean value.

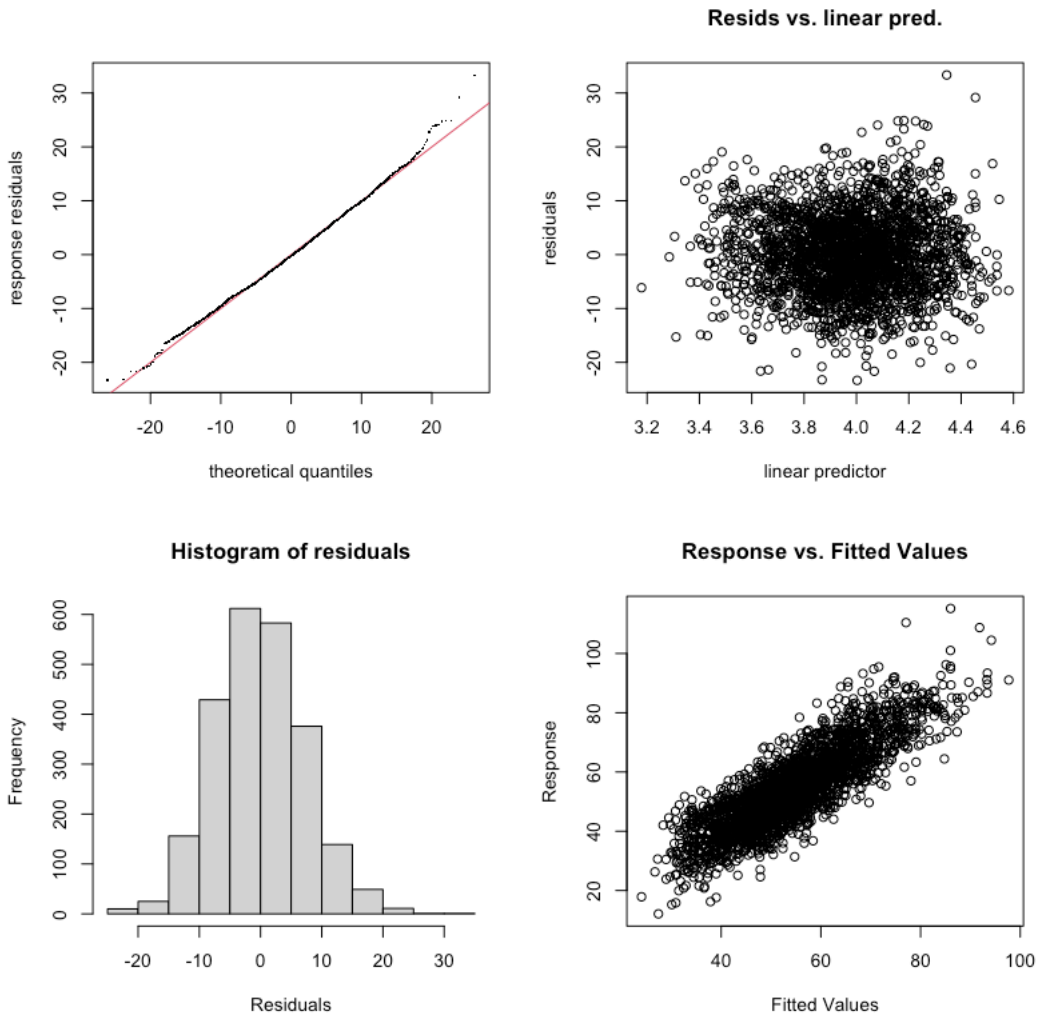

Figure 8. GAM evaluation plots for total MDA8 O_3 in the area of Dallas/Fort Worth, TX.

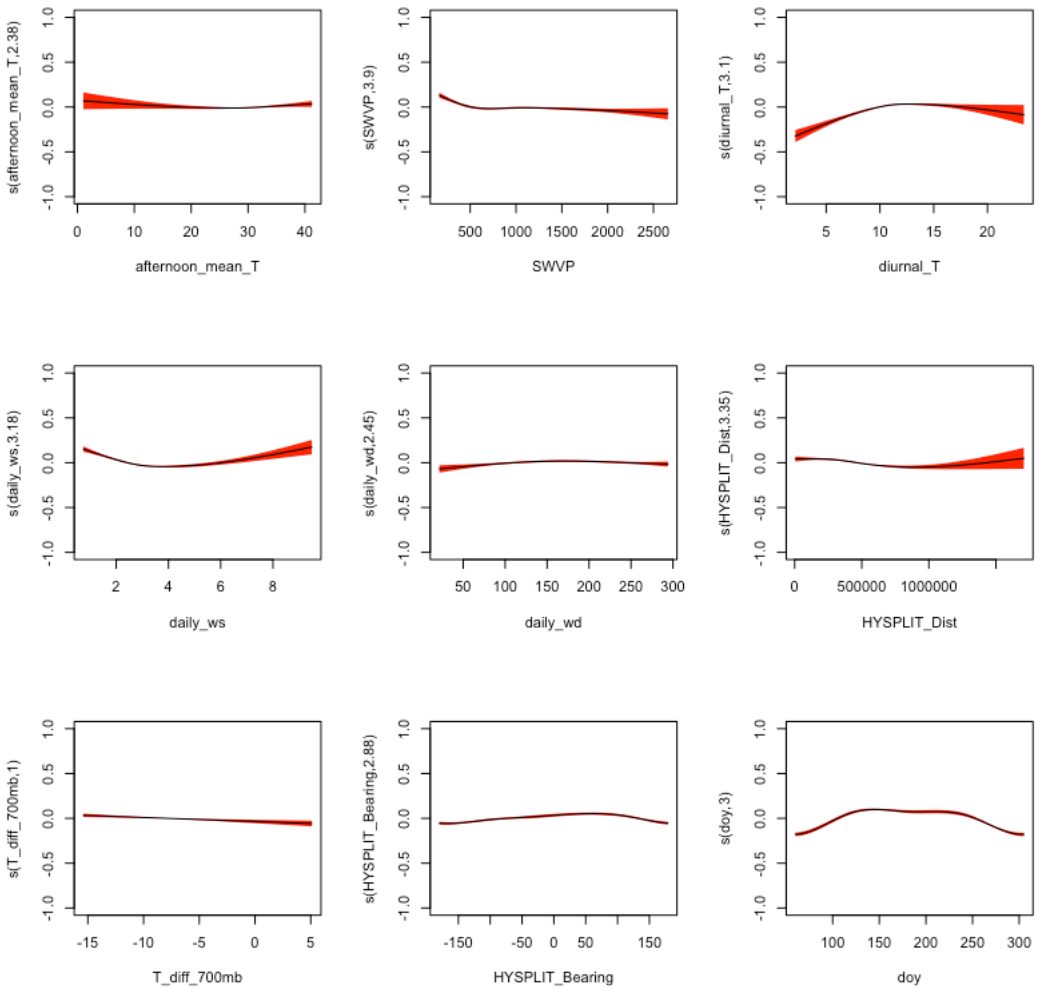

Figure 9. Smooth functions for the total MDA8 O_3 GAM fit in the area of El Paso, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the MDA8 O_3 in ppbv from its mean value.

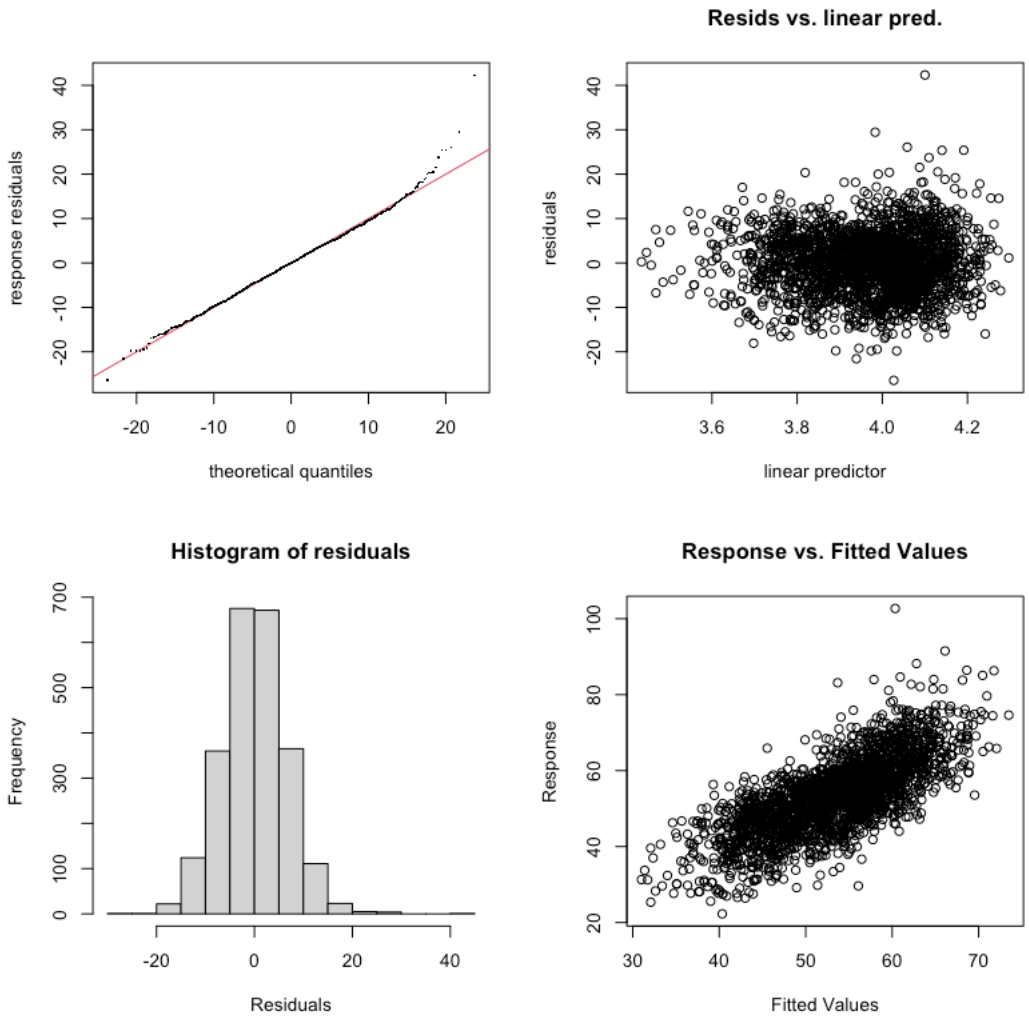

Figure 10. GAM evaluation plots for total MDA8 O_3 in the area of El Paso, TX.

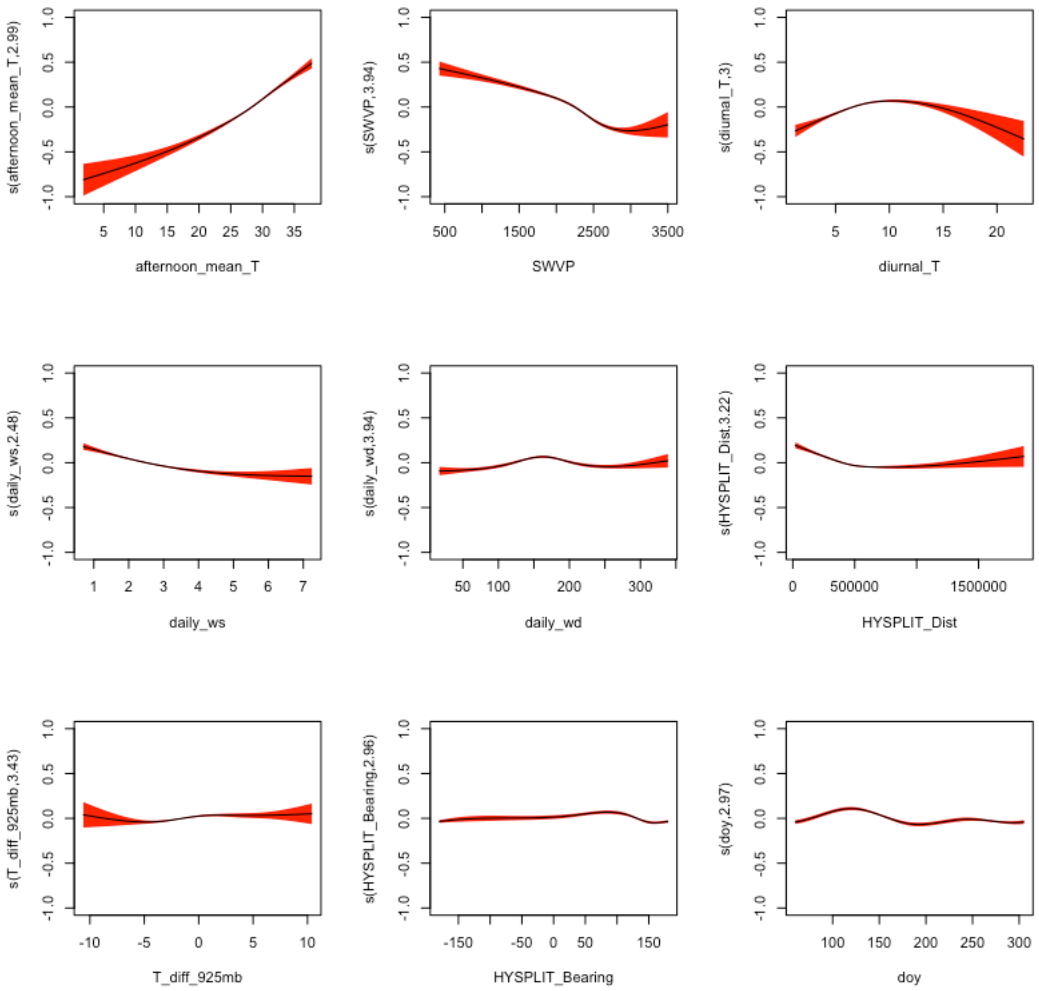

Figure 11. Smooth functions for the total MDA8 O_3 GAM fit in the area of Houston/Galveston/Brazoria, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the MDA8 O_3 in ppbv from its mean value.

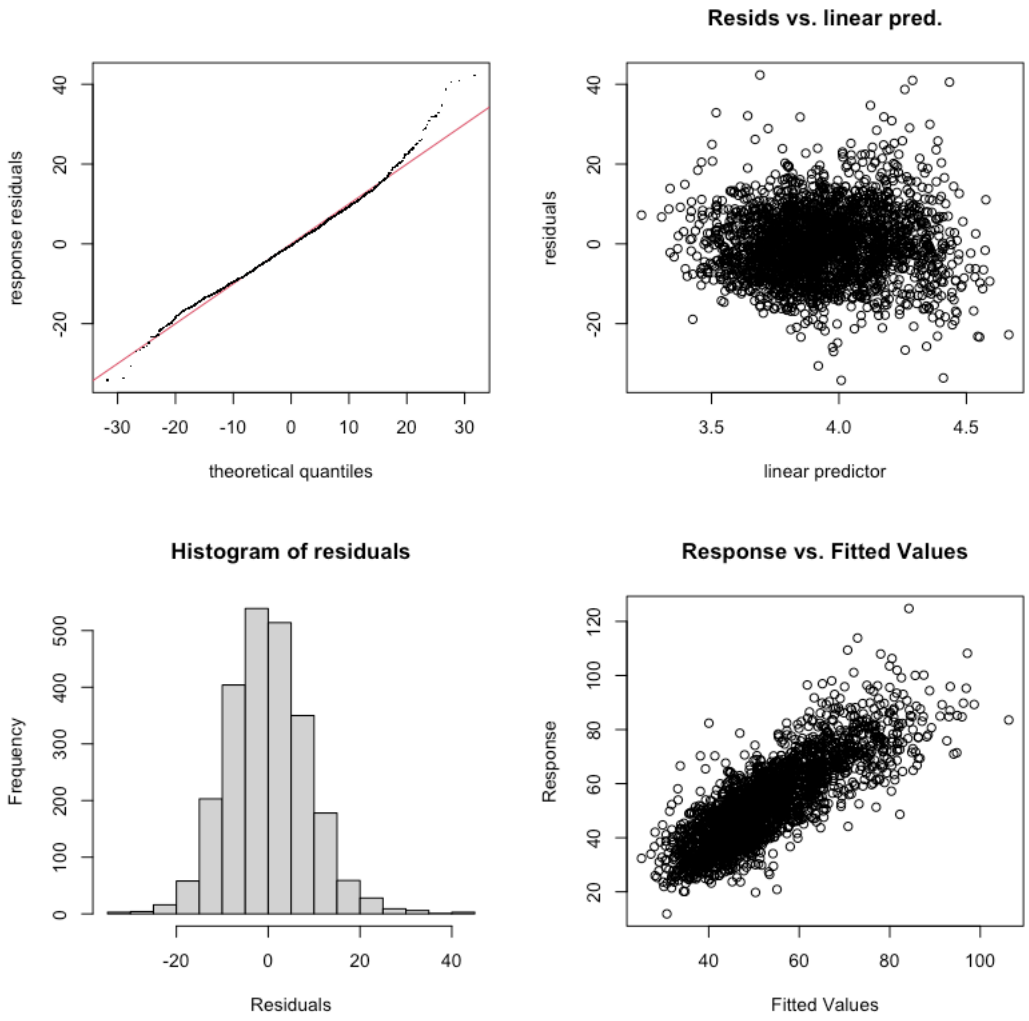

Figure 12. GAM evaluation plots for total MDA8 O_3 in the area of Houston/Galveston/Brazoria, TX.

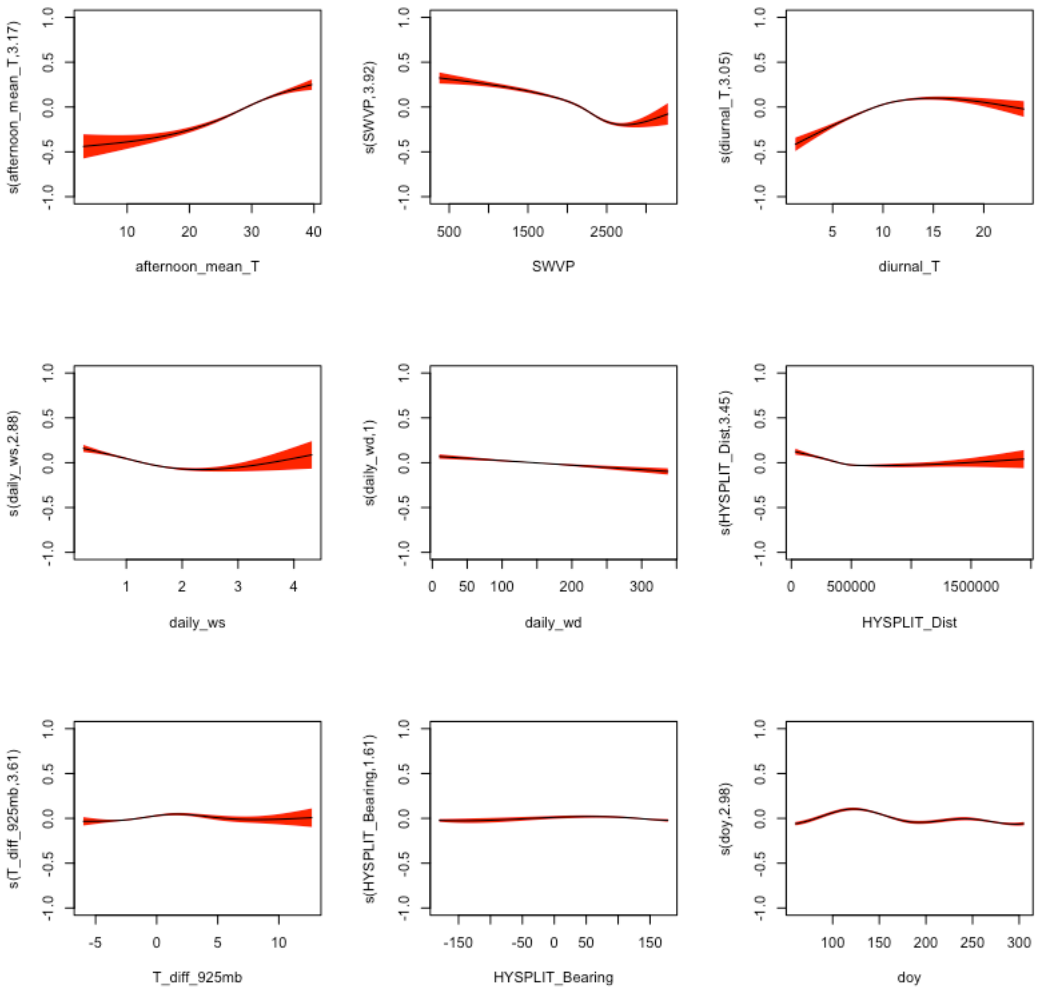

Figure 13. Smooth functions for the total MDA8 O_3 GAM fit in the area of San Antonio, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the MDA8 O_3 in ppbv from its mean value.

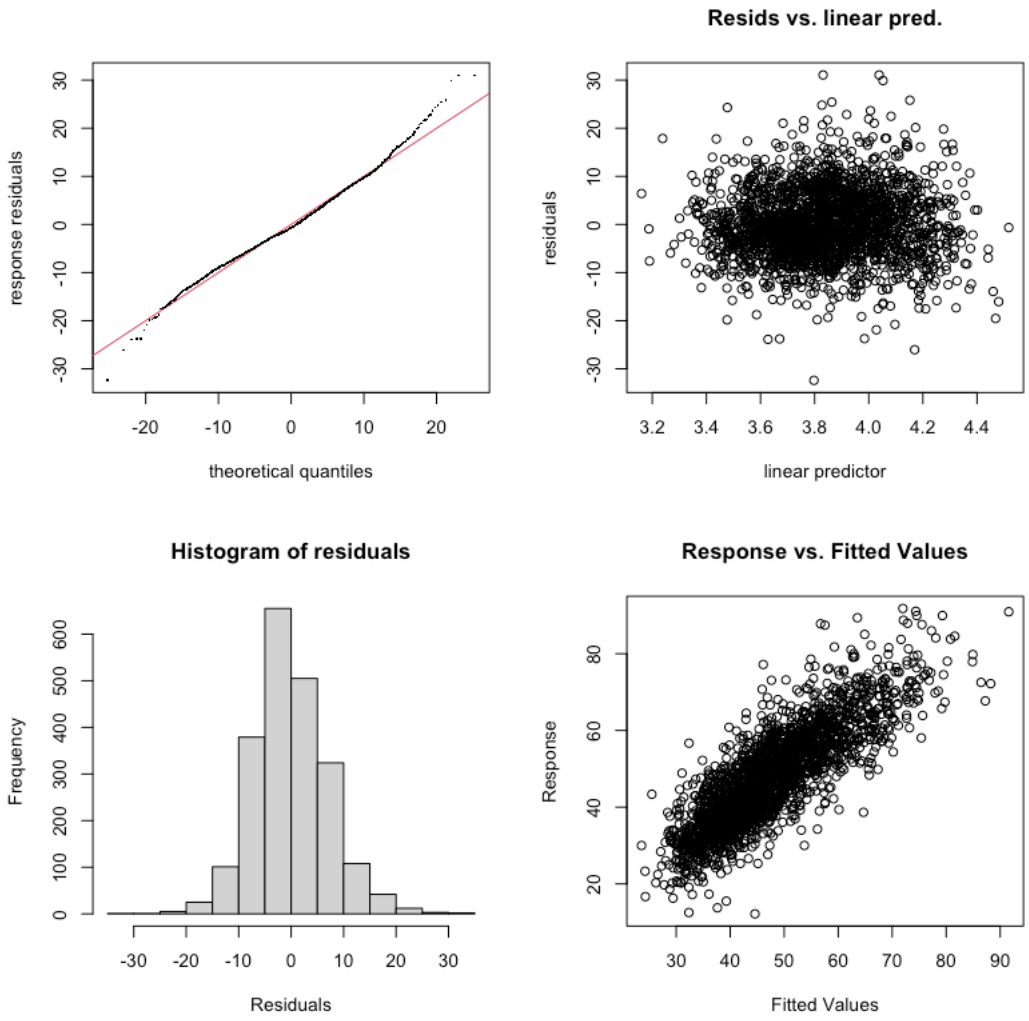

Figure 14. GAM evaluation plots for total MDA8 O_3 in the area of San Antonio, TX.

2.2.2.2 Background MDA8 O_3 GAM Results

Table 9 organizes the background MDA8 O_3 GAM results for each of the seven urban areas. This includes figure numbers for the results, the percentage deviance of the background MDA8 O_3 values explained by the GAM model, and the level statistical significance of the meteorological predictors in table 7. Odd number figures are the smooth functions from the GAM fit of the natural logarithm of the background MDA8 O_3 values to the meteorological predictors. 95% confidence intervals are shown in red in the smooth functions. The day of year (*doy*) function is also shown. The percentage deviance values can be compared to the Camalier et al. (2007) results, which showed the predictive power of their models (measured by the R^2 statistic) to be between 0.56 and 0.80 for the cities in that study. Even number figures are the standard GAM evaluation plots (made with the *gam.check* function in R). The plots for all seven urban areas indicate a good fit, as the model residuals are roughly normally distributed and show no trend versus predicted value. The

variance of the residuals is lower for low values of the predictor, but this reflects the fact that the measured MDA8 O_3 values cannot go below 0.

Table 9. Performance of GAMs for Background MDA8 O_3.

Urban Area	Figure for Smooth Functions	Figure for GAM Evaluation	% Deviance Explained	Significance of Meteorological Predictors
Austin	15	16	66.3	7 terms at α=0.001 1 term at α=0.01
Beaumont/Port Arthur	17	18	60	7 terms at α=0.001 1 term at α=0.05
Corpus Christi	19	20	59.7	8 terms at α=0.001
Dallas/Fort Worth	21	22	60.5	5 terms at α=0.001 1 term at α=0.01 1 term at α=0.1 1 term at α=1
El Paso	23	24	51	7 terms at α=0.001 1 term at α=0.05
Houston/Galveston/Brazoria	25	26	63.4	7 terms at α=0.001 1 term at α=0.01
San Antonio	27	28	65.9	7 terms at α=0.001 1 term at α=0.01

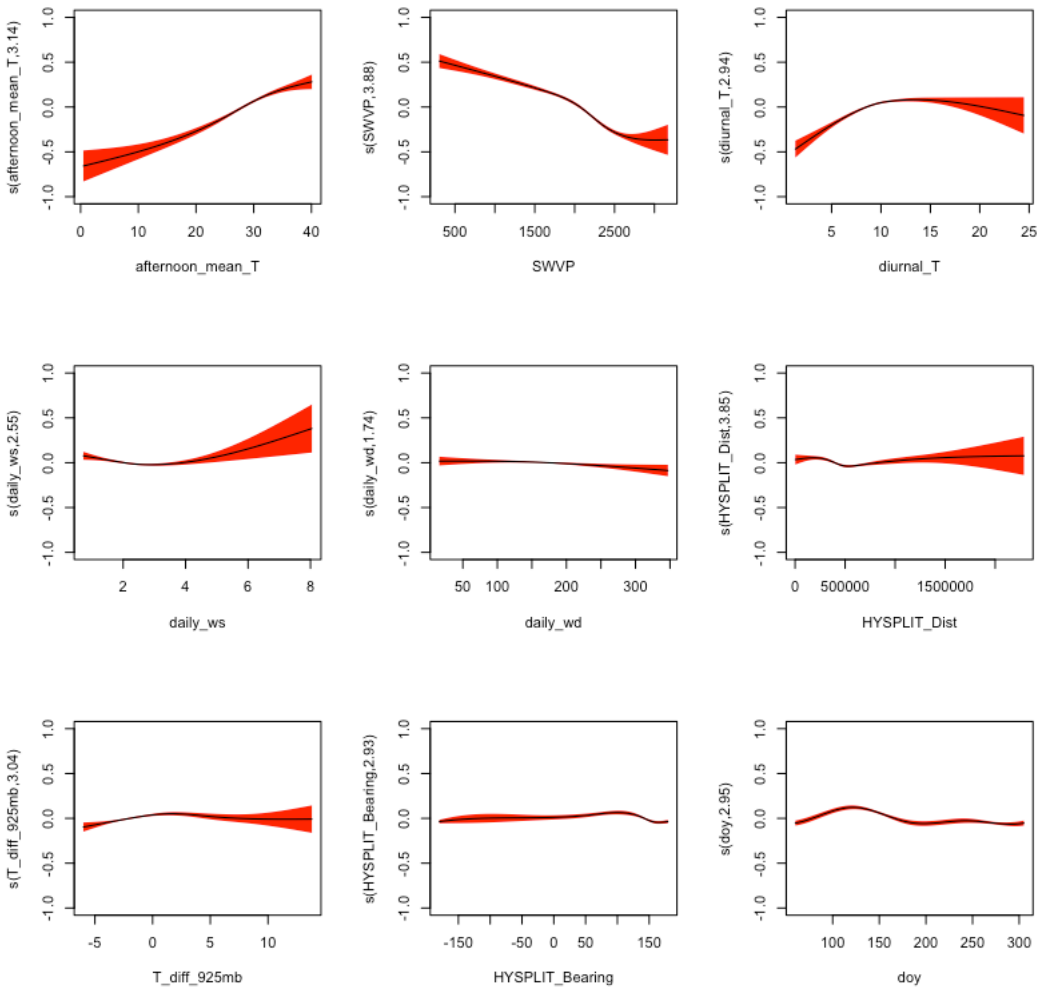

Figure 15. Smooth functions for the background MDA8 O_3 GAM fit in the area of Austin, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background MDA8 O_3 in ppbv from its mean value.

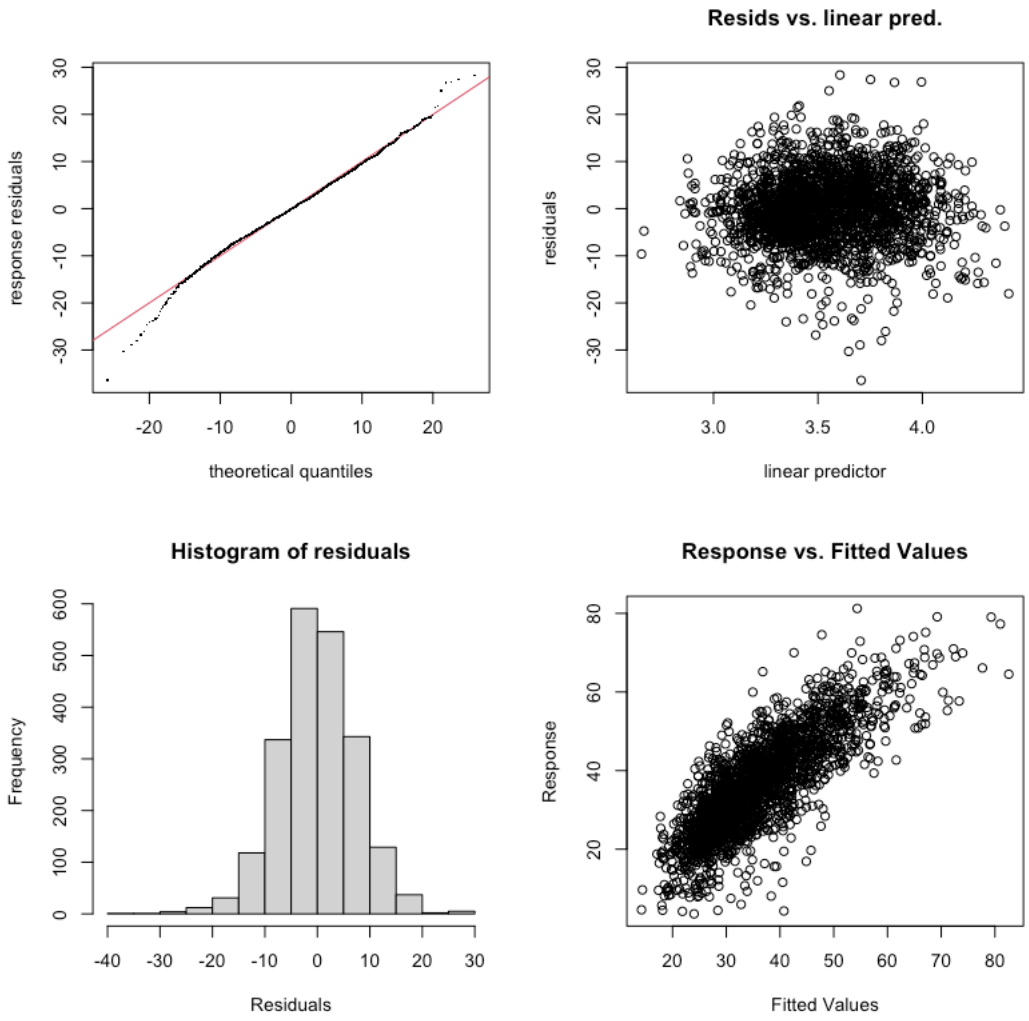

Figure 16. GAM evaluation plots for background MDA8 O_3 in the area of Austin, TX.

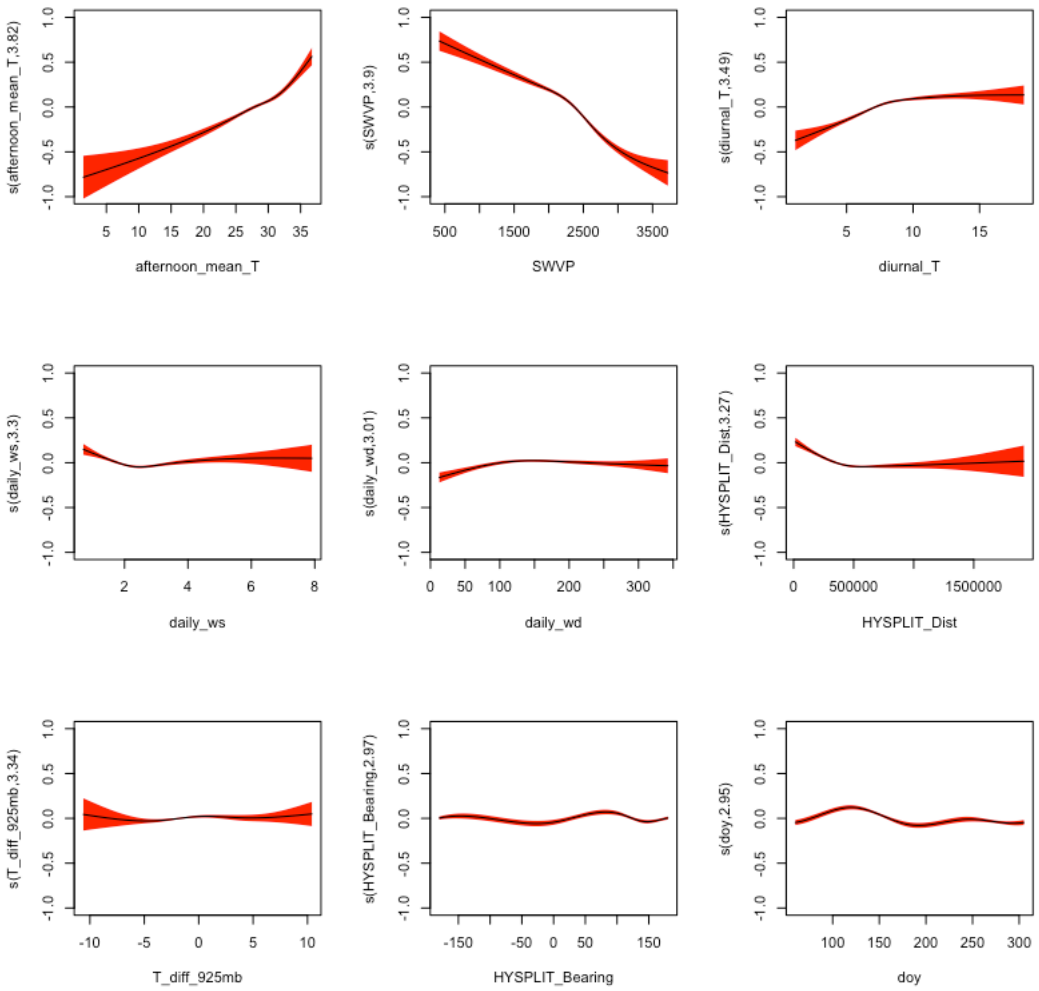

Figure 17. Smooth functions for the background MDA8 O_3 GAM fit in the area of Beaumont/Port Arthur, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background MDA8 O_3 in ppbv from its mean value.

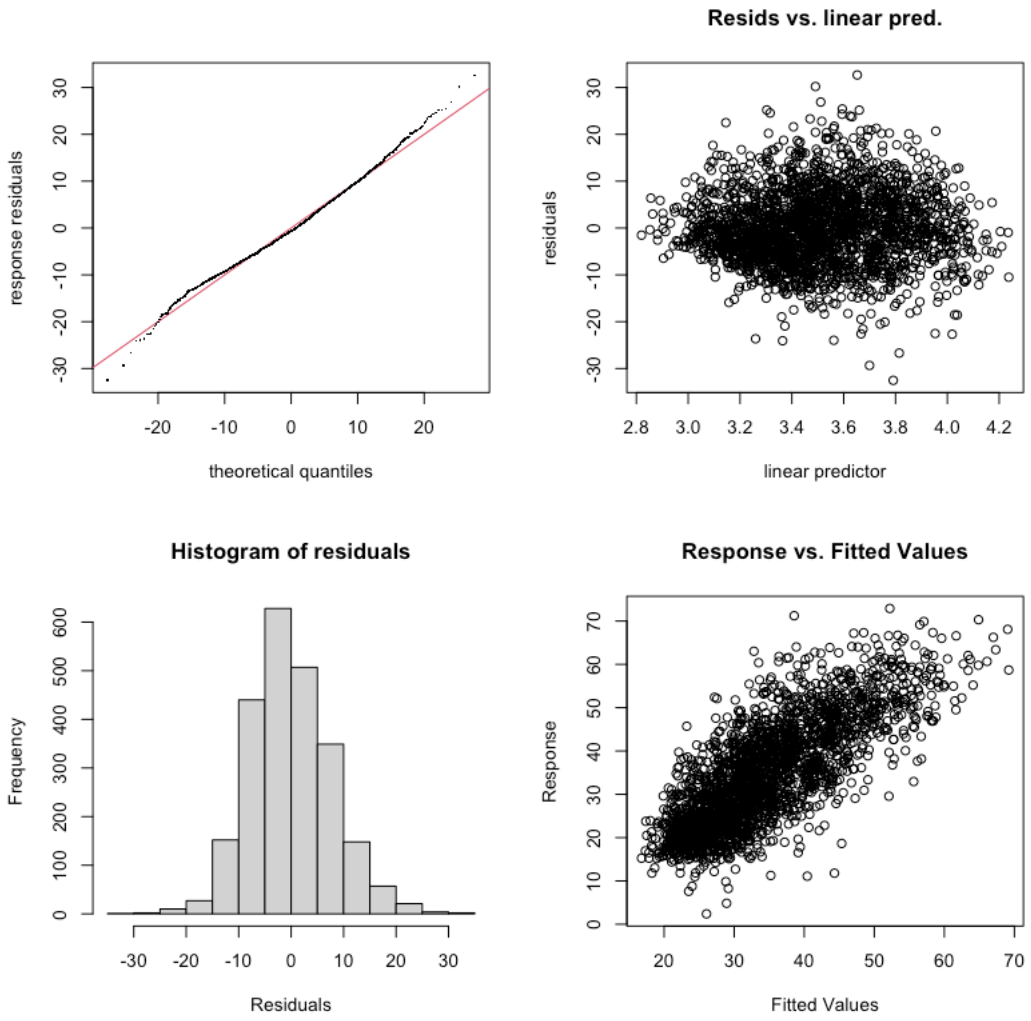

Figure 18. GAM evaluation plots for background MDA8 O_3 in the area of Beaumont/Port Arthur, TX.

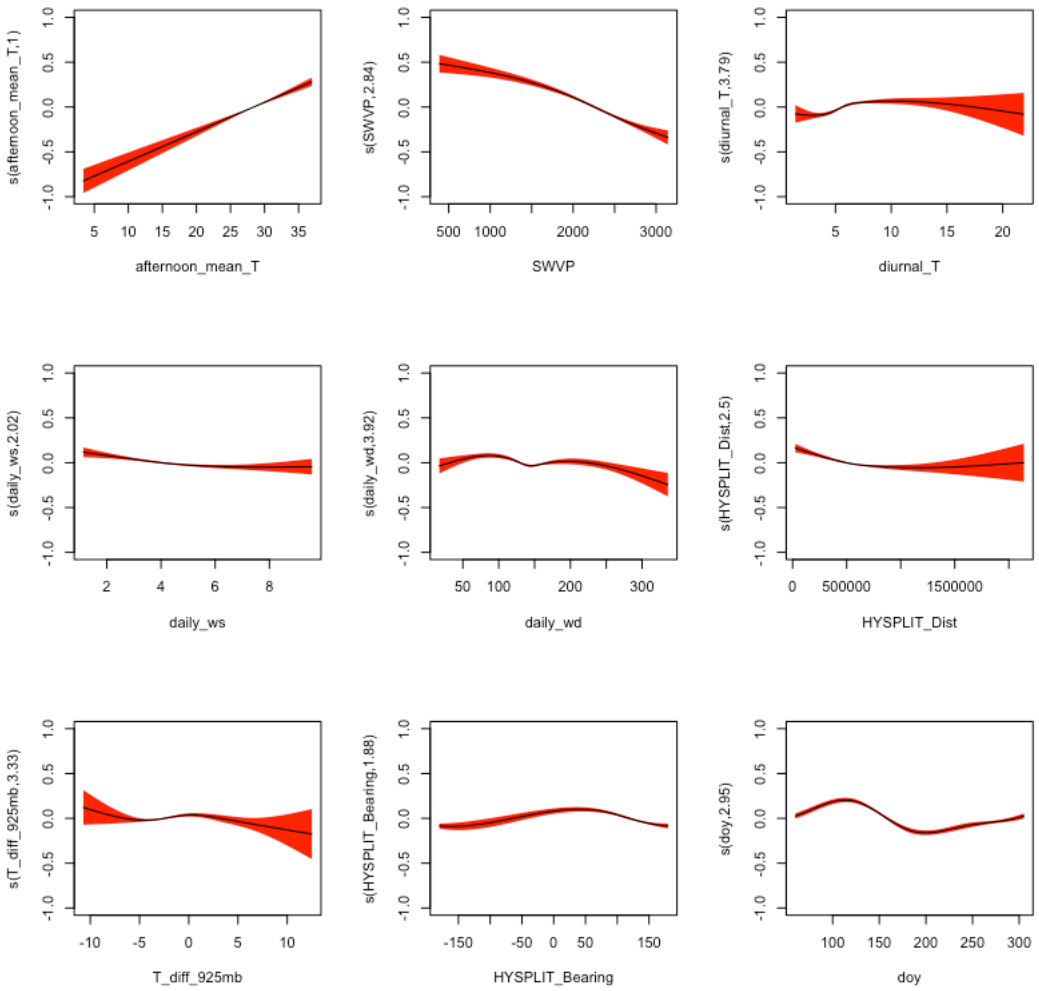

Figure 19. Smooth functions for the background MDA8 O_3 GAM fit in the area of Corpus Christi, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background MDA8 O_3 in ppbv from its mean value.

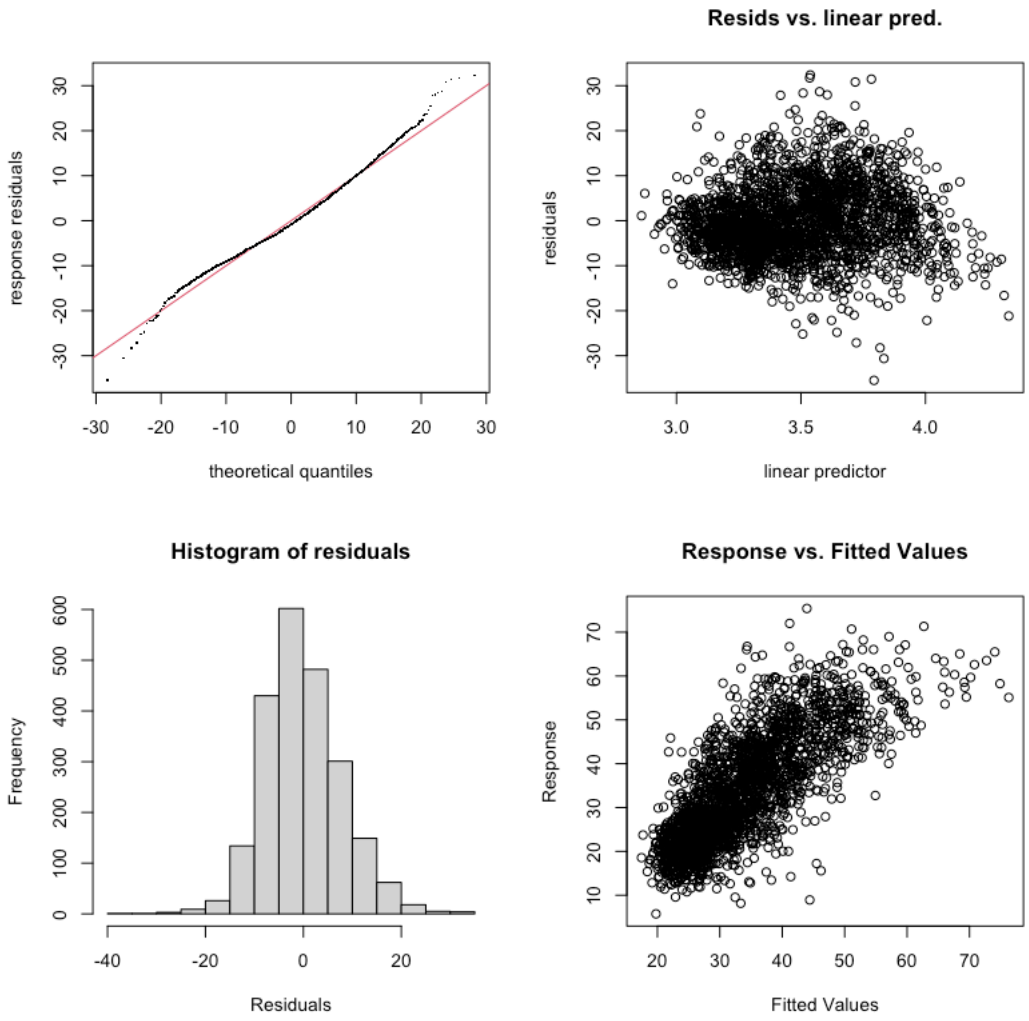

Figure 20. GAM evaluation plots for background MDA8 O_3 in the area of Corpus Christi, TX.

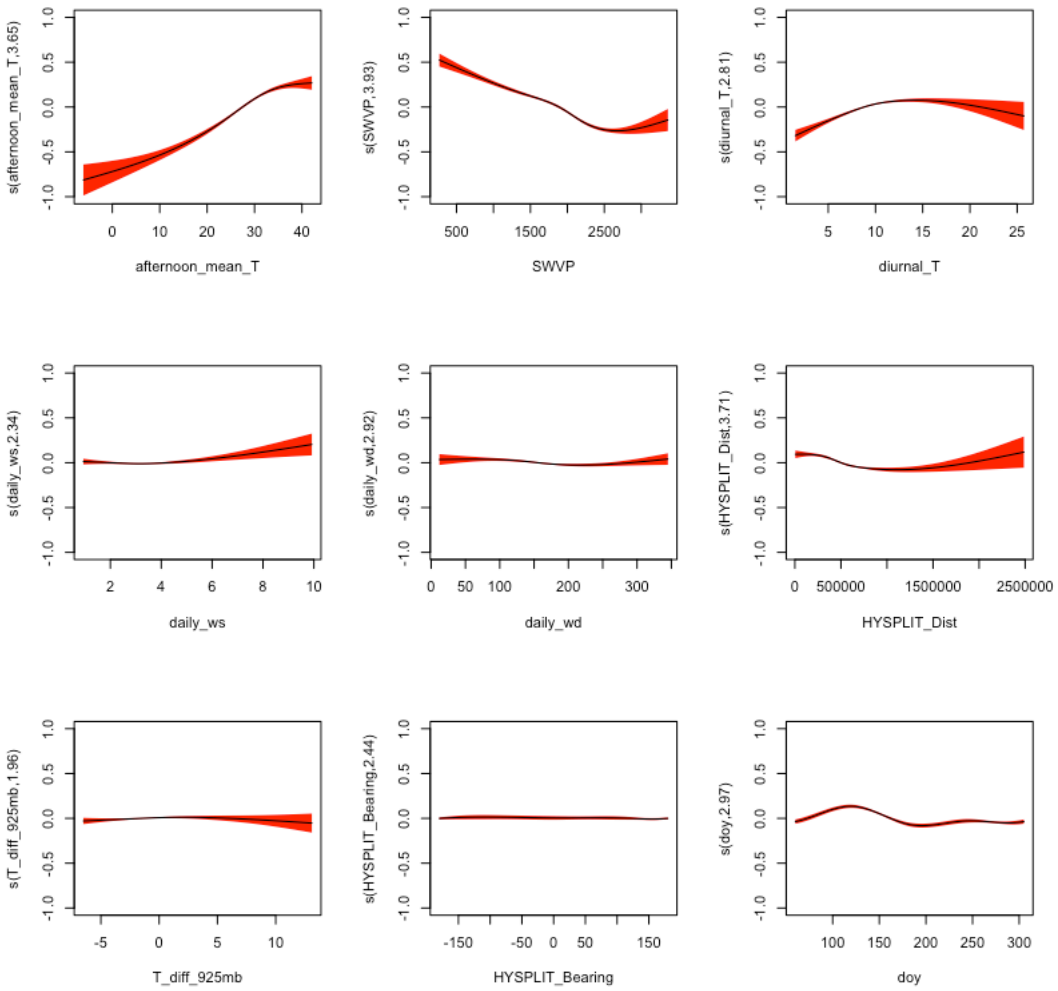

Figure 21. Smooth functions for the background MDA8 O_3 GAM fit in the area of Dallas/Fort Worth, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background MDA8 O_3 in ppbv from its mean value.

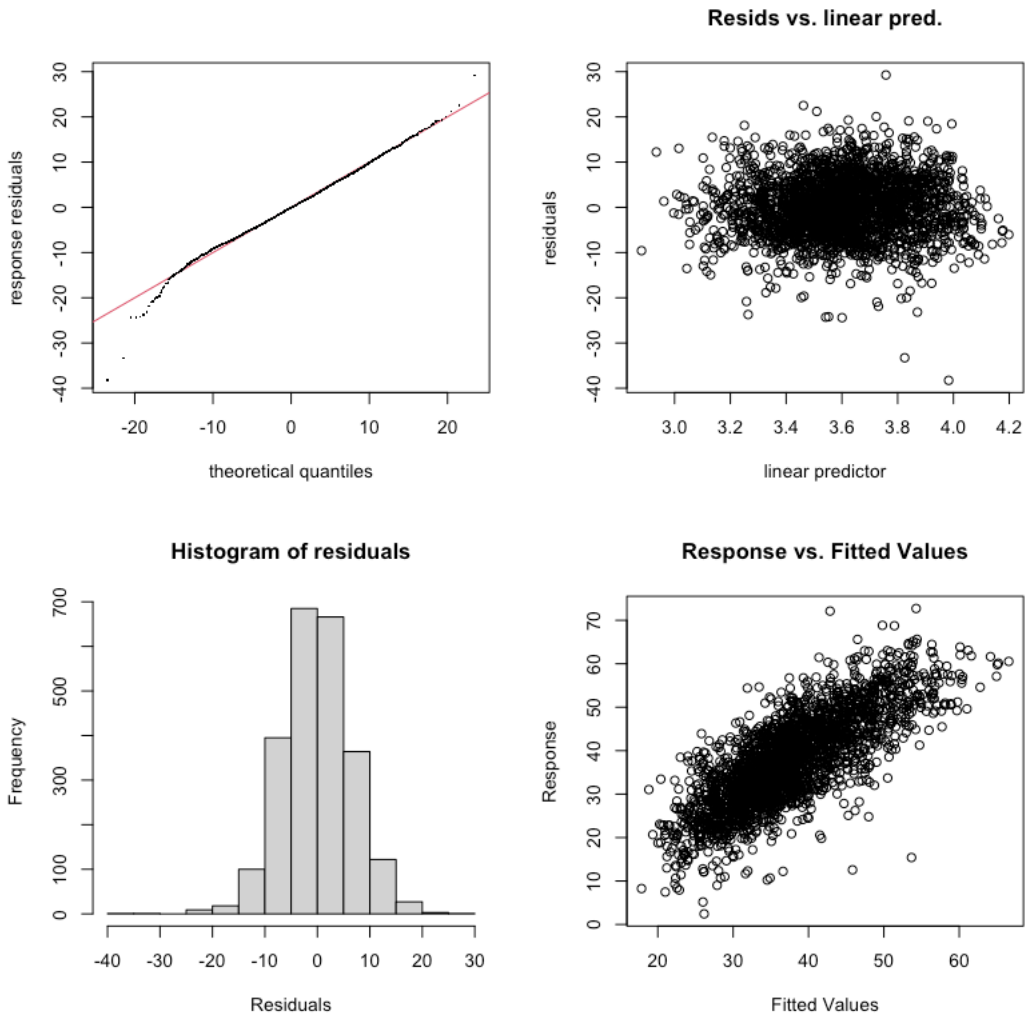

Figure 22. GAM evaluation plots for background MDA8 O_3 in the area of Dallas/Fort Worth, TX.

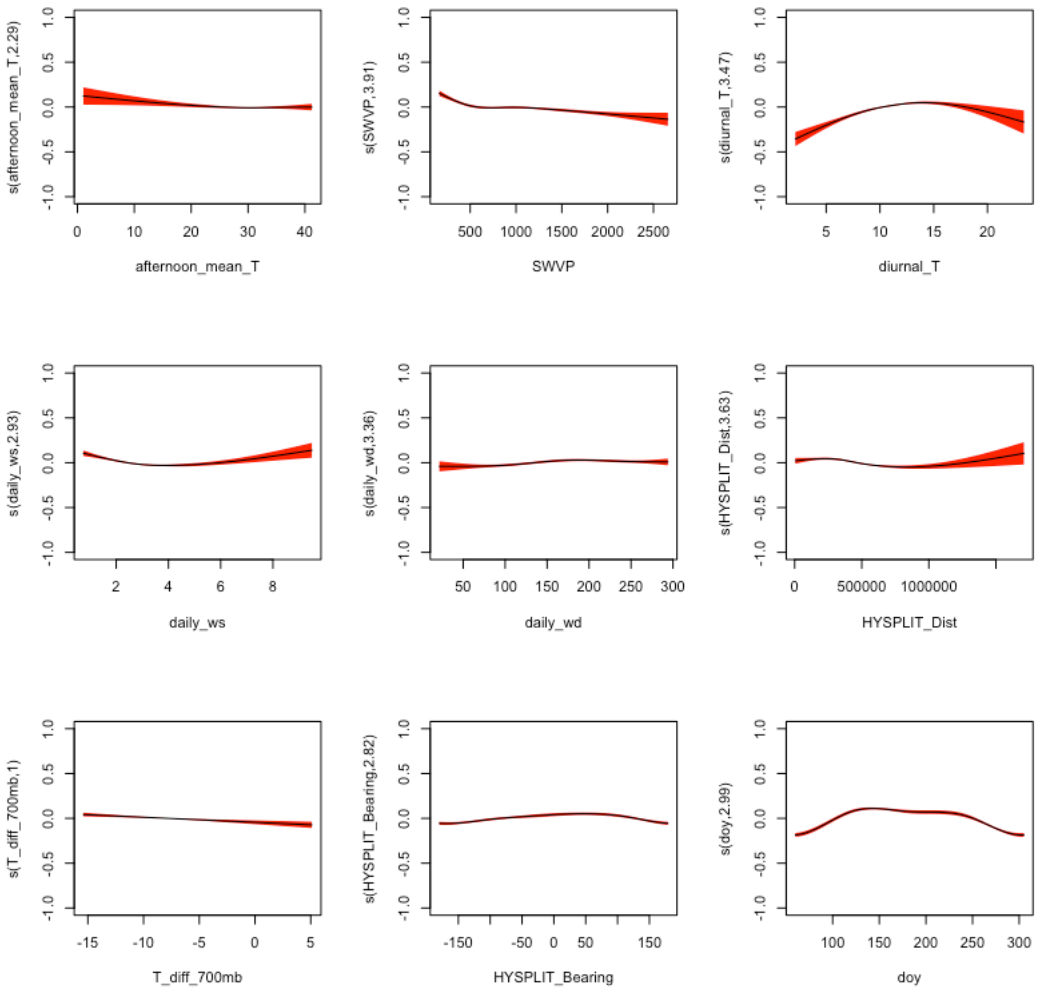

Figure 23. Smooth functions for the background MDA8 O_3 GAM fit in the area of El Paso, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background MDA8 O_3 in ppbv from its mean value.

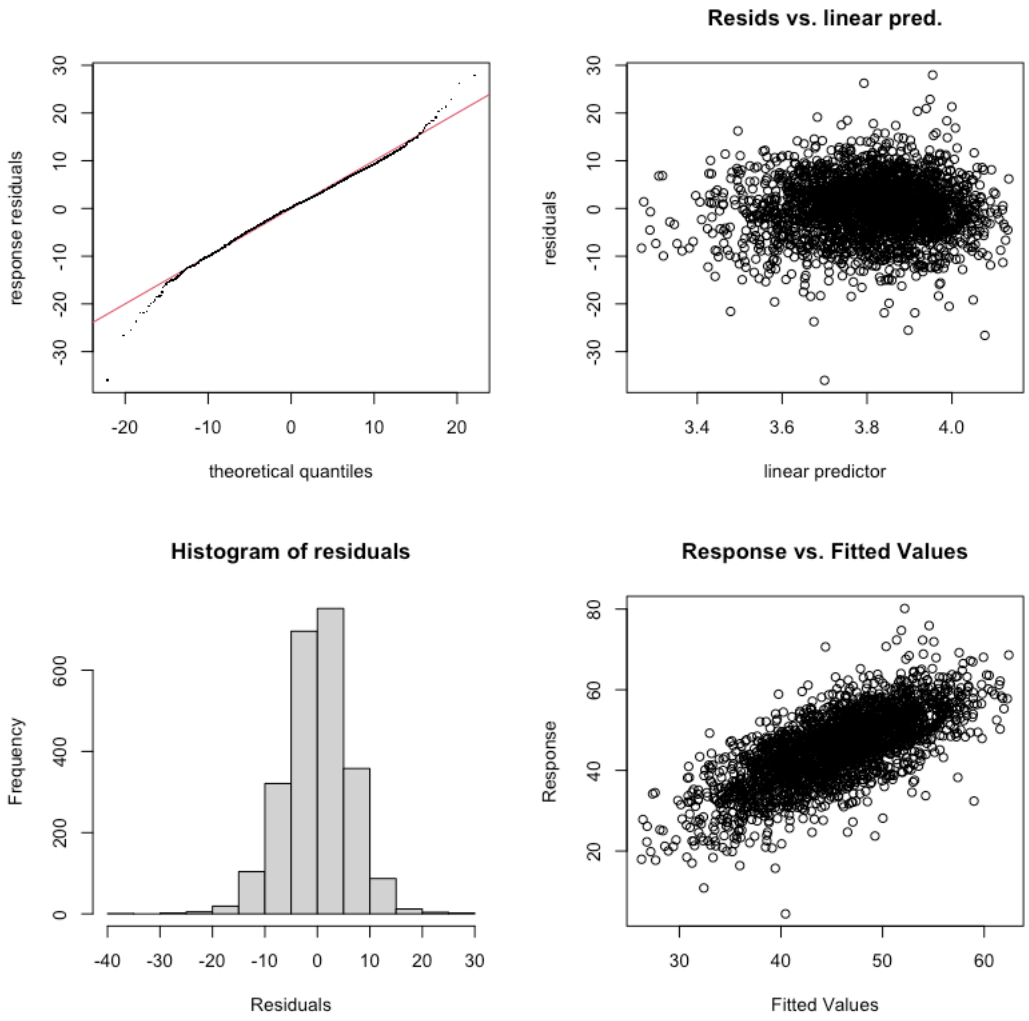

Figure 24. GAM evaluation plots for background MDA8 O_3 in the area of El Paso, TX.

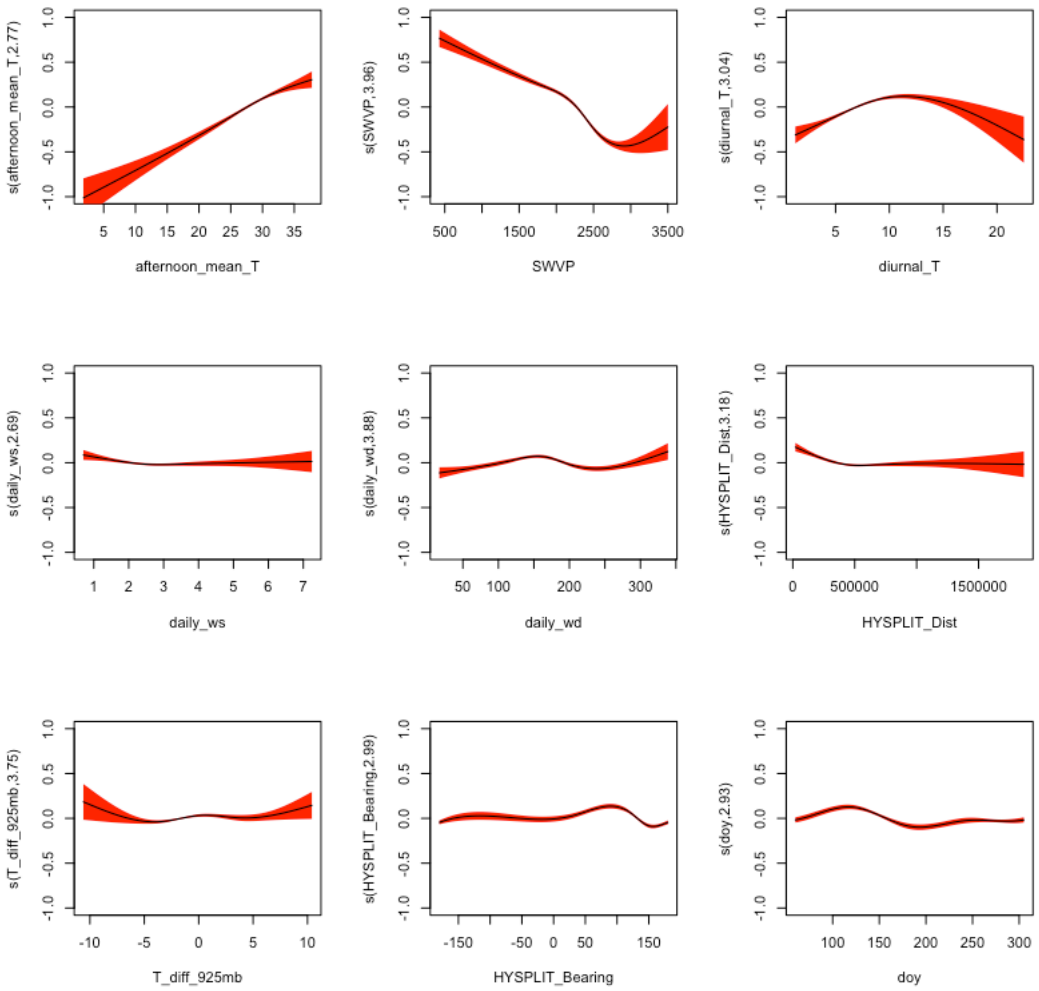

Figure 25. Smooth functions for the background MDA8 O_3 GAM fit in the area of Houston/Galveston/Brazoria, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background MDA8 O_3 in ppbv from its mean value.

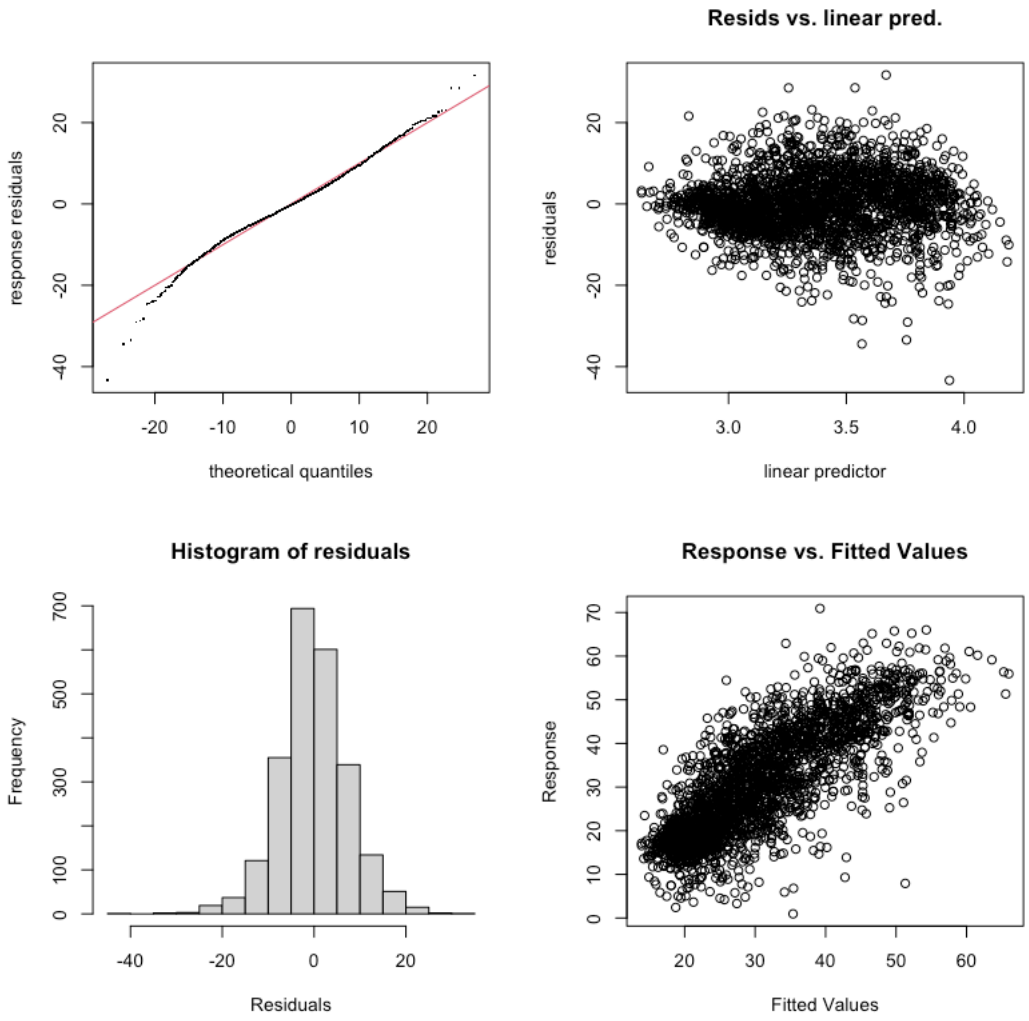

Figure 26. GAM evaluation plots for background MDA8 O_3 in the area of Houston/Galveston/Brazoria, TX.

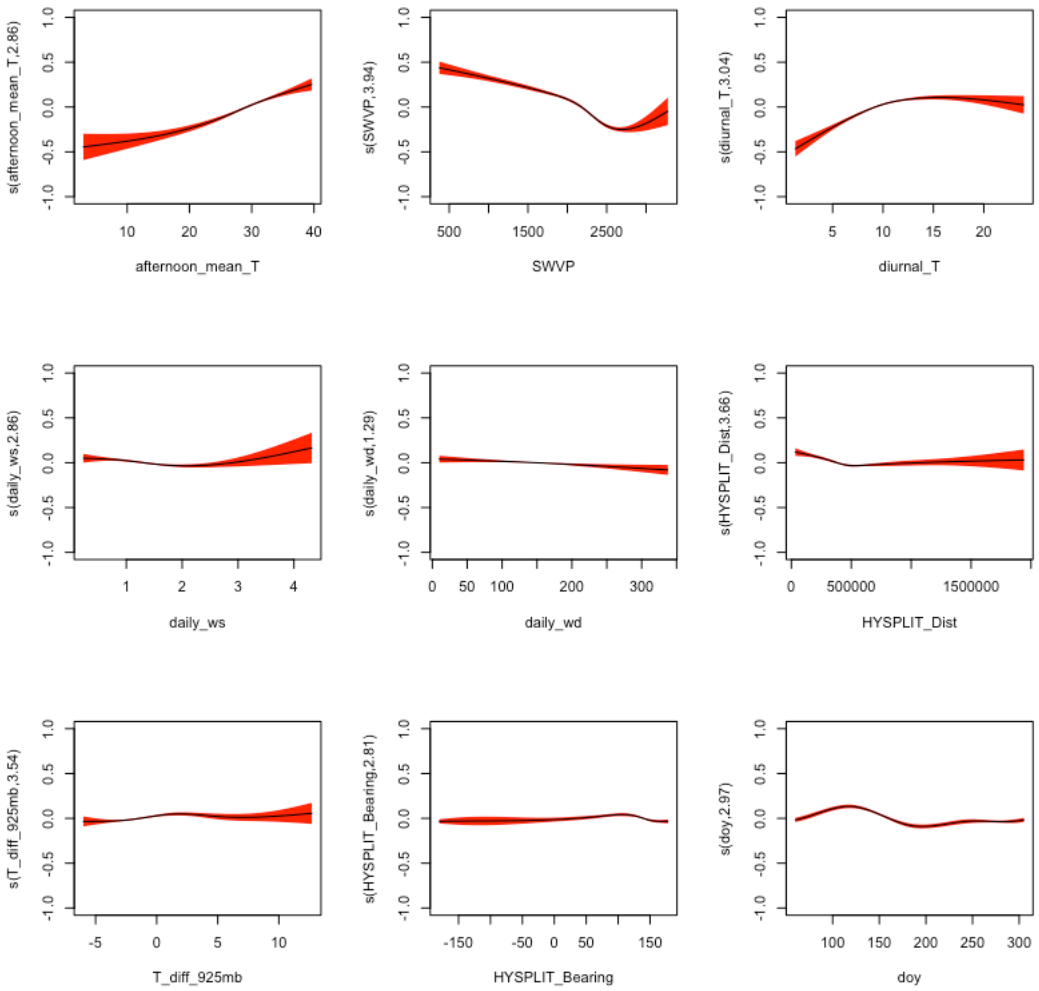

Figure 27. Smooth functions for the background MDA8 O_3 GAM fit in the area of San Antonio, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background MDA8 O_3 in ppbv from its mean value.

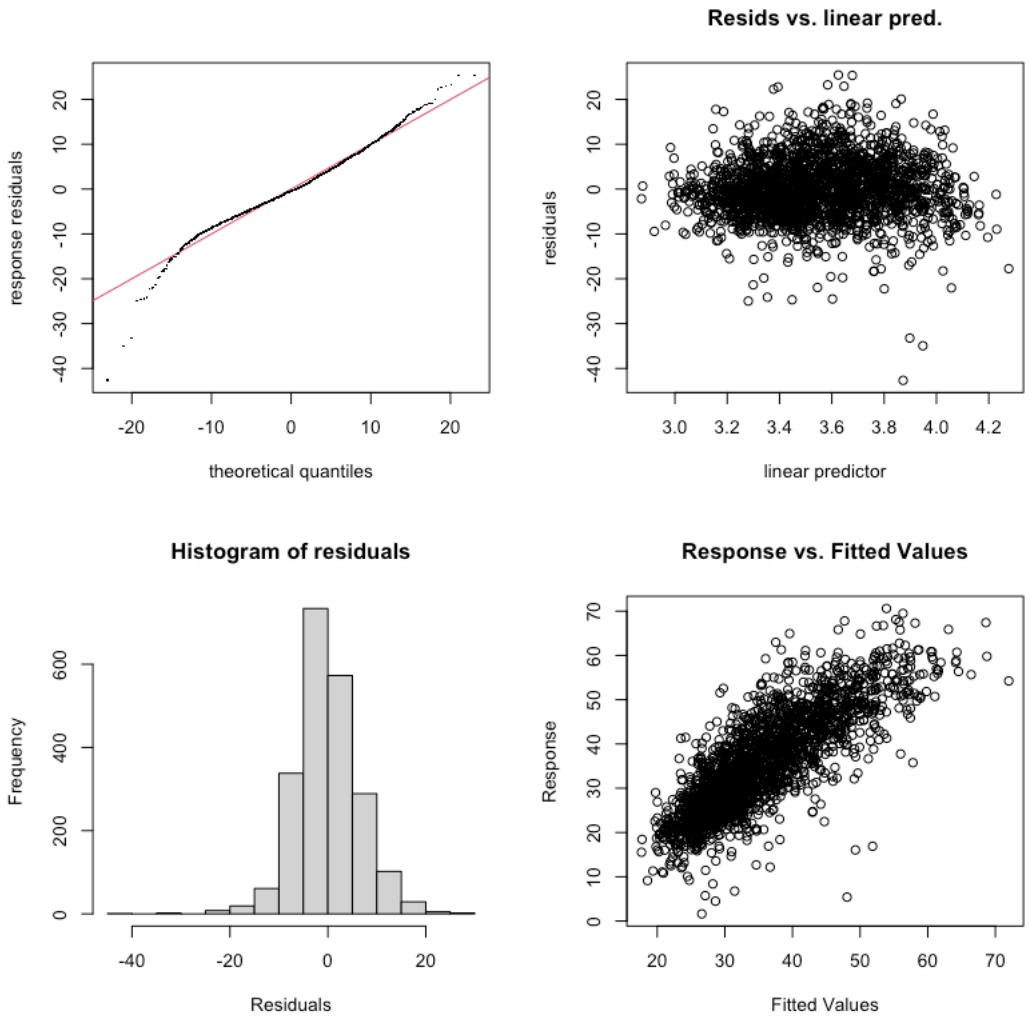

Figure 28. GAM evaluation plots for background MDA8 O_3 in the area of San Antonio, TX.

2.2.3 $PM_{2.5}$ GAM Results

2.2.3.1 Total PM2.5 GAM Results

Table 10 organizes the total $PM_{2.5}$ GAM results for each of the seven urban areas. This includes figure numbers for the results, the percentage deviance of the total $PM_{2.5}$ values explained by the GAM model, and the level statistical significance of the meteorological predictors in table 7. Odd number figures are the smooth functions from the GAM fit of the natural logarithm of the total $PM_{2.5}$ values to the meteorological predictors. 95% confidence intervals are shown in red in the smooth functions. The day of year (*doy*) function is also shown. The percentage deviance values can be compared to the Camalier et al. (2007) results, which showed the predictive power of their models (measured by the R^2 statistic) to be between 0.56 and 0.80 for the cities in that study. Even number figures are the standard GAM evaluation plots (made with the *gam.check* function in R). The low percentage deviance values and the plots and for all seven urban areas generally indicate a poor fit (as the model residuals do not follow a normal distribution). We do not think these fits

are providing much accurate information on the variability of $PM_{2.5}$ under different meteorological conditions.

Table 10. Performance of GAMs for Total $PM_{2.5}$.

Urban Area	Figure for Smooth Functions	Figure for GAM Evaluation	% Deviance Explained	Significance of Meteorological Predictors
Austin	29	30	21.4	7 terms at α=0.001 1 term at α=0.05
Beaumont/Port Arthur	31	32	19.1	6 terms at α=0.001 1 term at α=0.1 1 term at α=1
Corpus Christi	33	34	34.2	8 terms at α=0.001
Dallas/Fort Worth	35	36	26.1	4 terms at α=0.001 1 term at α=0.01 1 term at α=0.05 1 term at α=0.1 1 term at α=1
El Paso	37	38	32	6 terms at α=0.001 1 term at α=0.05 1 term at α=0.1
Houston/Galveston/Brazoria	39	40	14.7	4 terms at α=0.001 1 term at α=0.01 1 term at α=0.1 2 terms at α=1
San Antonio	41	42	19	5 terms at α=0.001 2 terms at α=0.01 1 term at α=1

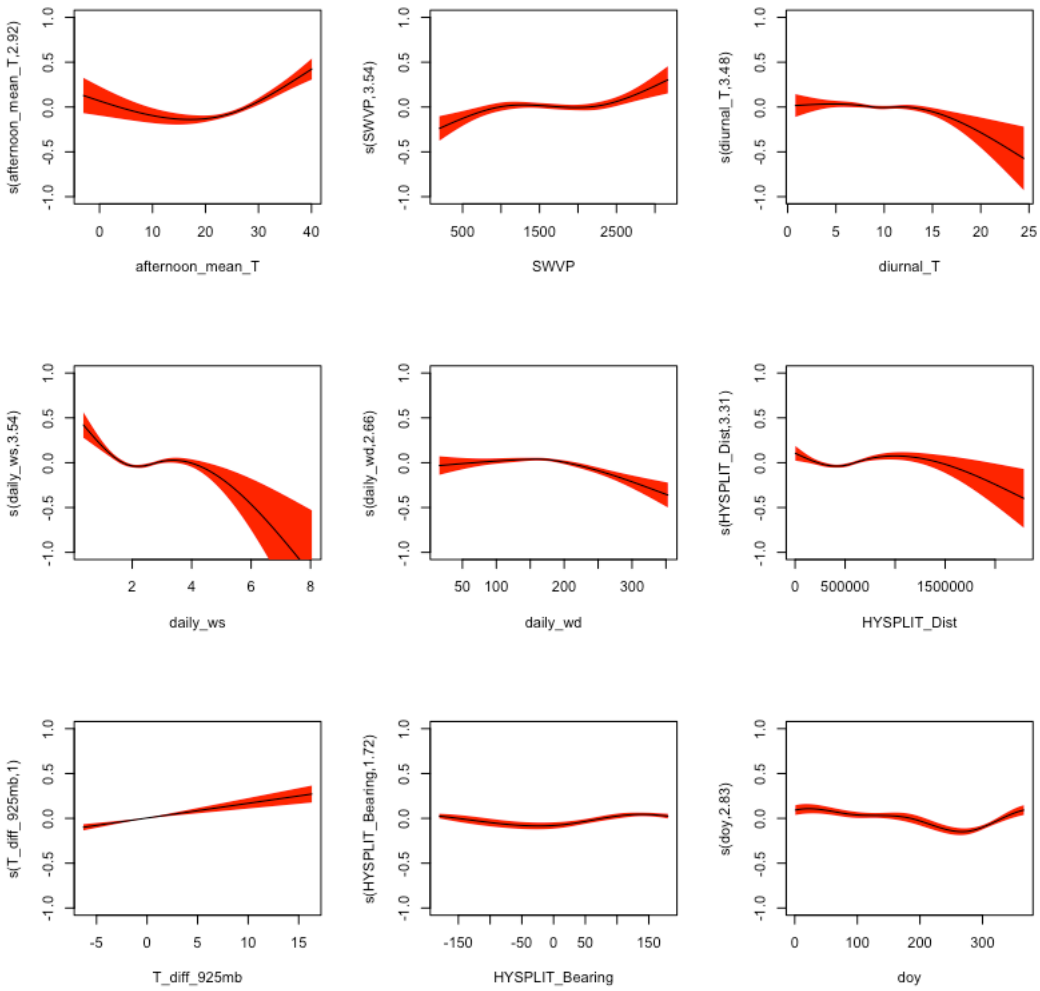

Figure 29. Smooth functions for the total daily average $PM_{2.5}$ GAM fit in the area of Austin, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the total daily average $PM_{2.5}$ in µg/m3 from its mean value.

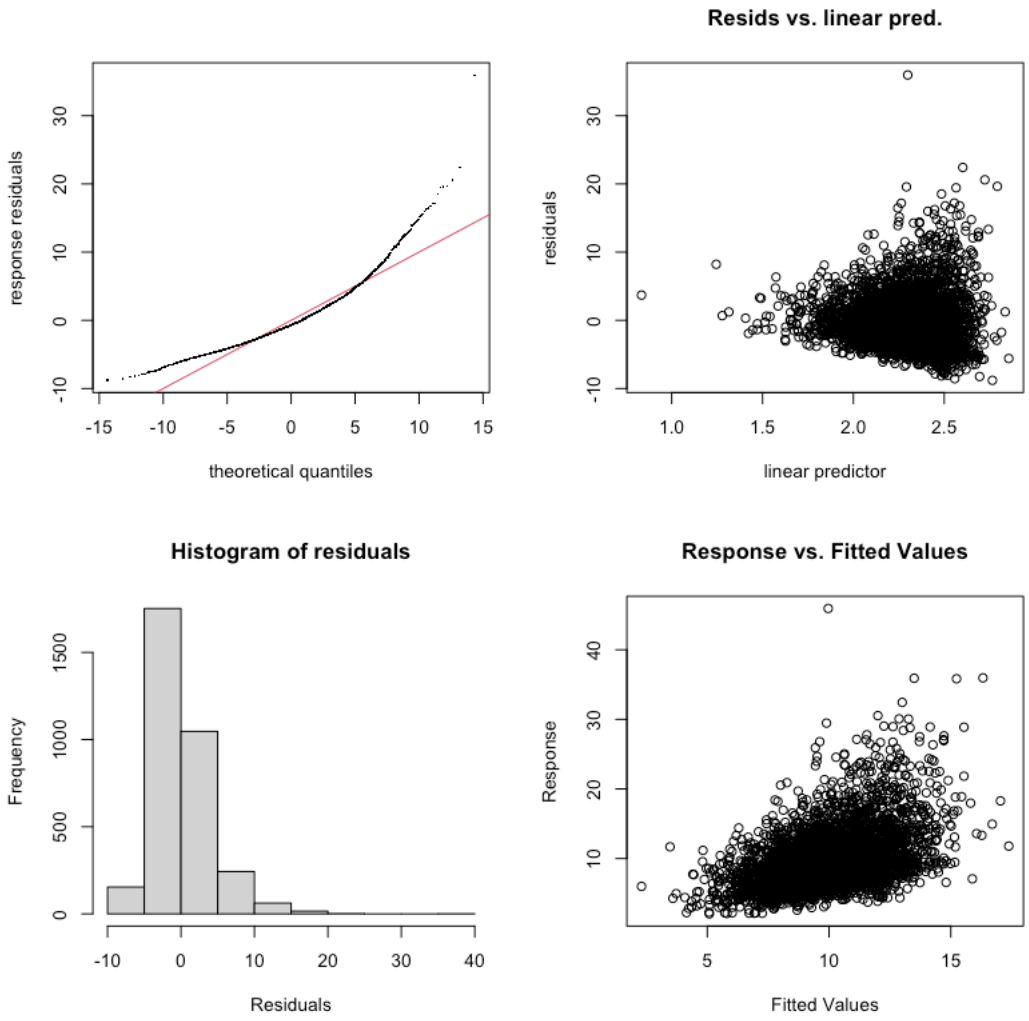

Figure 30. GAM evaluation plots for total daily average $PM_{2.5}$ in the area of Austin, TX.

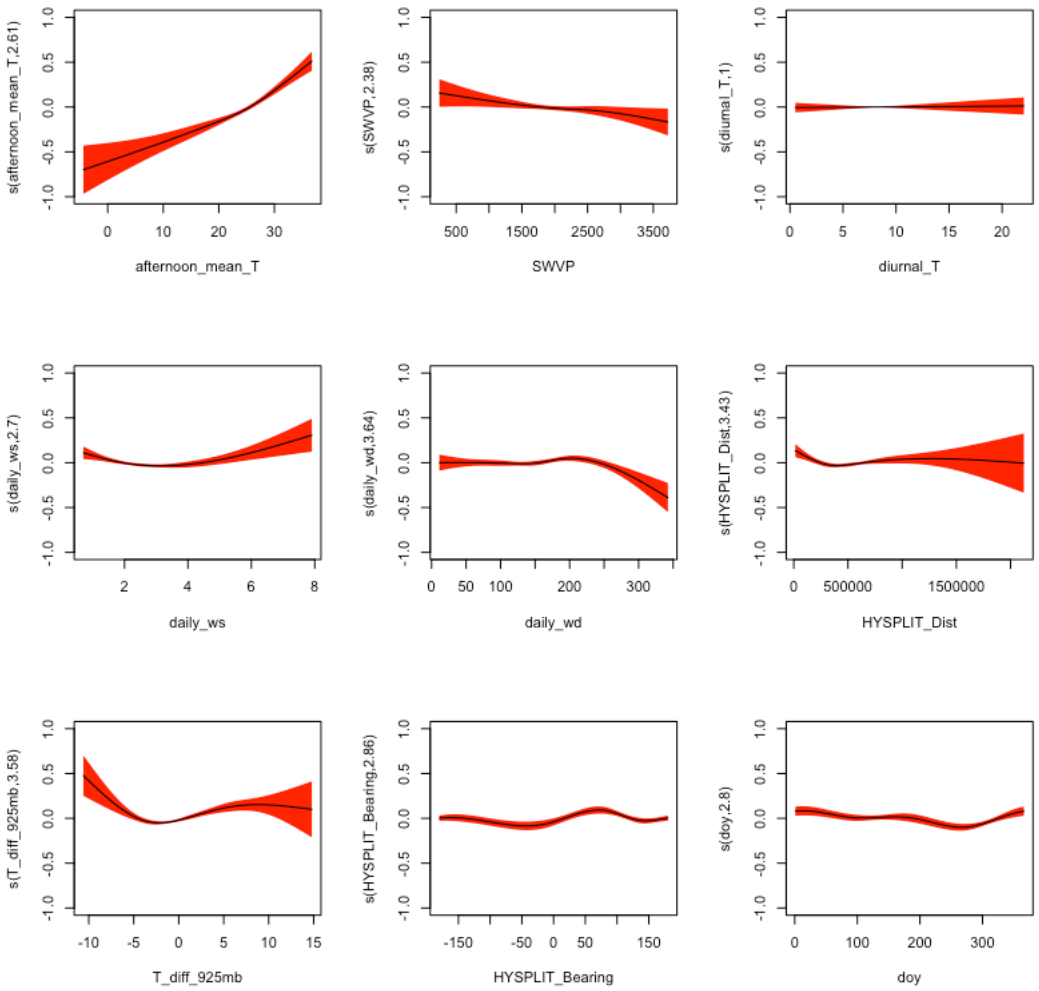

Figure 31. Smooth functions for the total daily average $PM_{2.5}$ GAM fit in the area of Beaumont/Port Arthur, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the total daily average $PM_{2.5}$ in µg/m3 from its mean value.

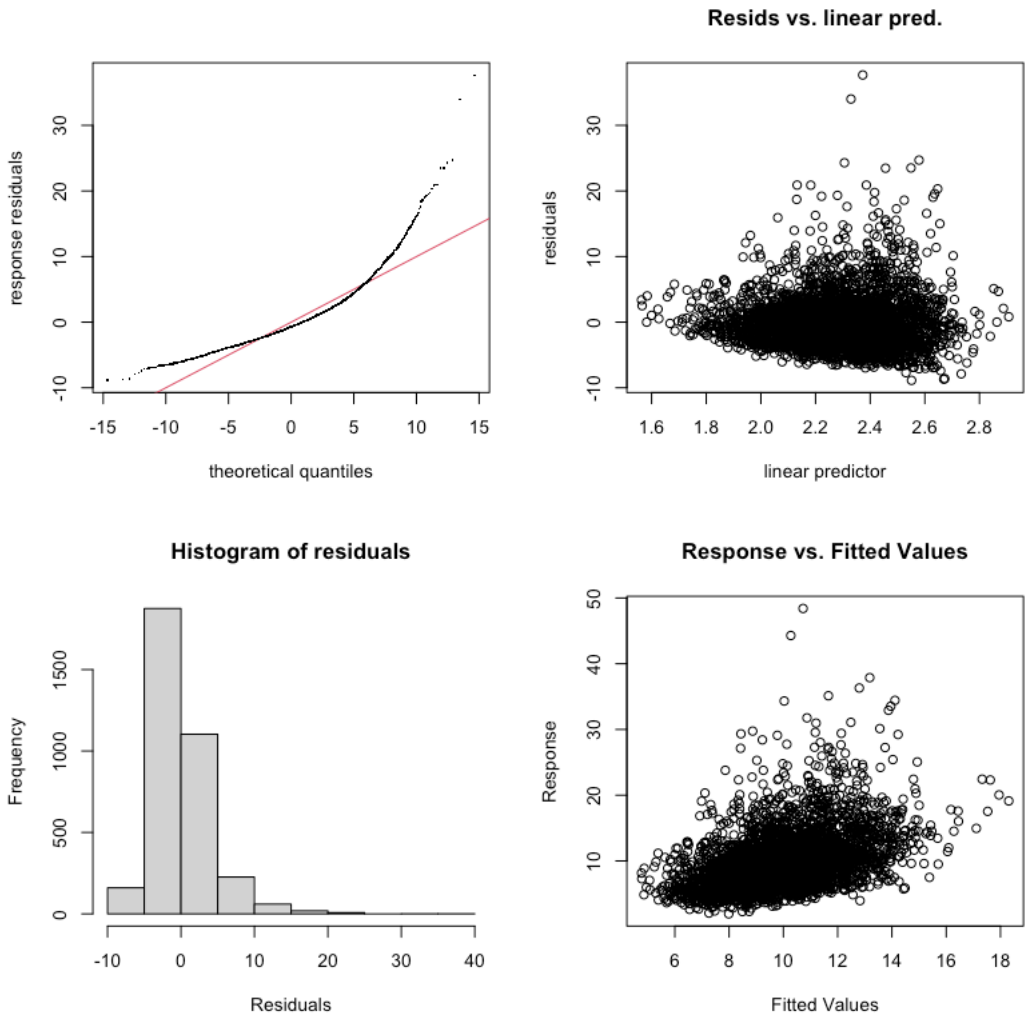

Figure 32. GAM evaluation plots for total daily average $PM_{2.5}$ in the area of Beaumont/Port Arthur, TX.

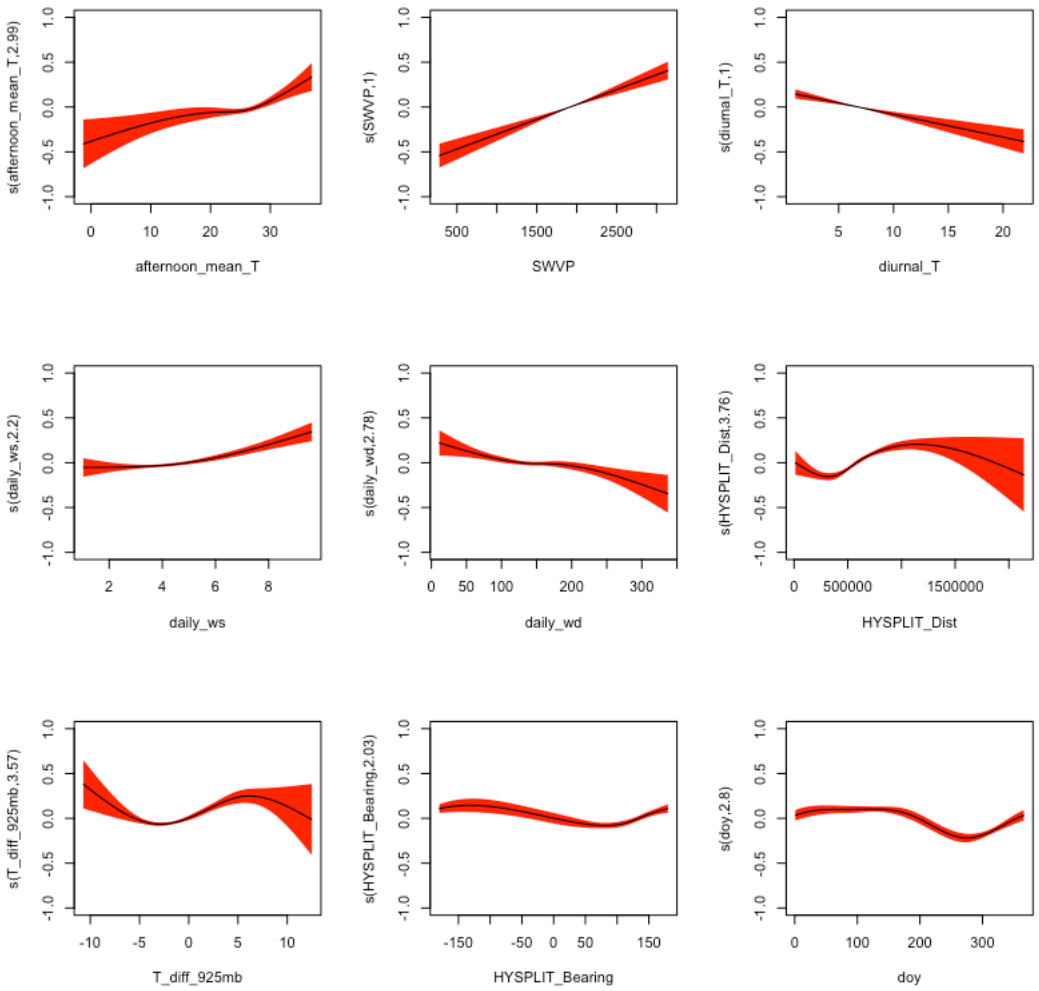

Figure 33. Smooth functions for the total daily average $PM_{2.5}$ GAM fit in the area of Corpus Christi, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the total daily average $PM_{2.5}$ in µg/m3 from its mean value.

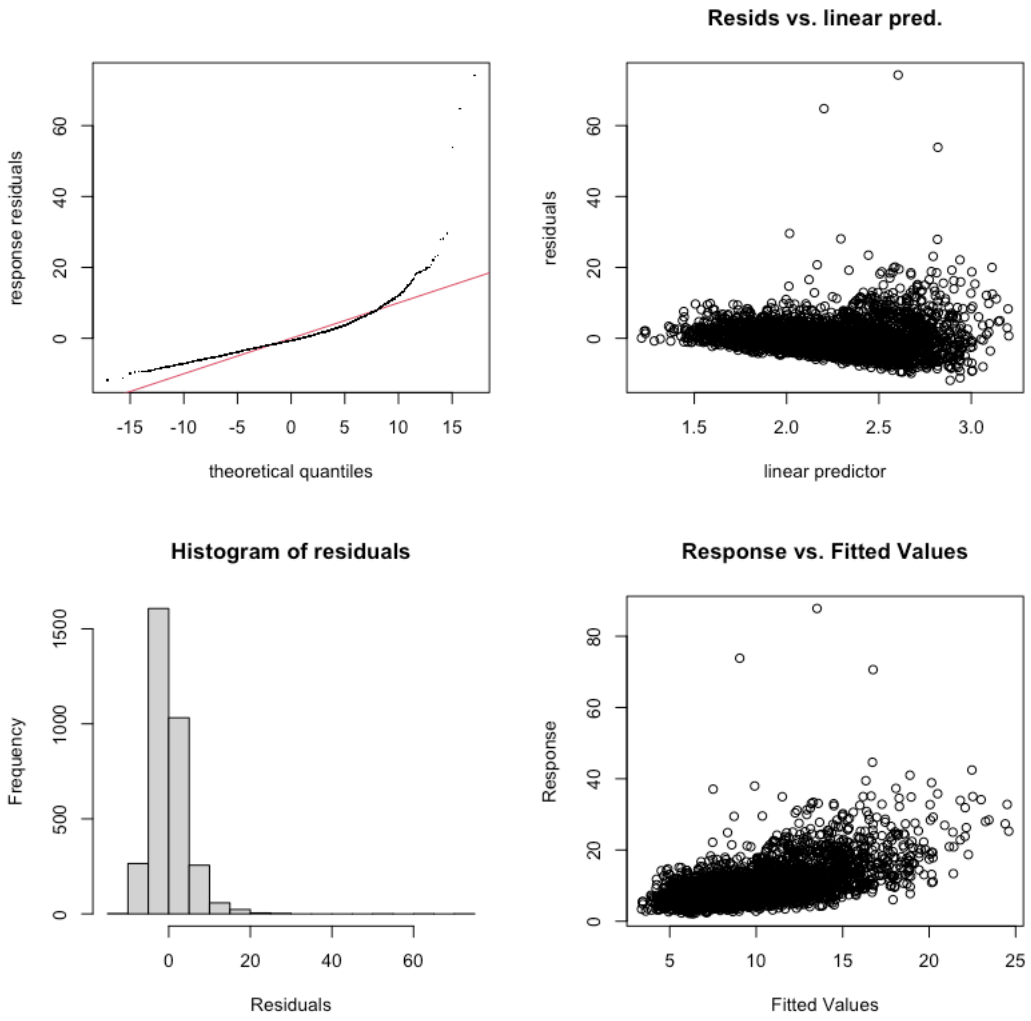

Figure 34. GAM evaluation plots for total daily average $PM_{2.5}$ in the area of Corpus Christi, TX.

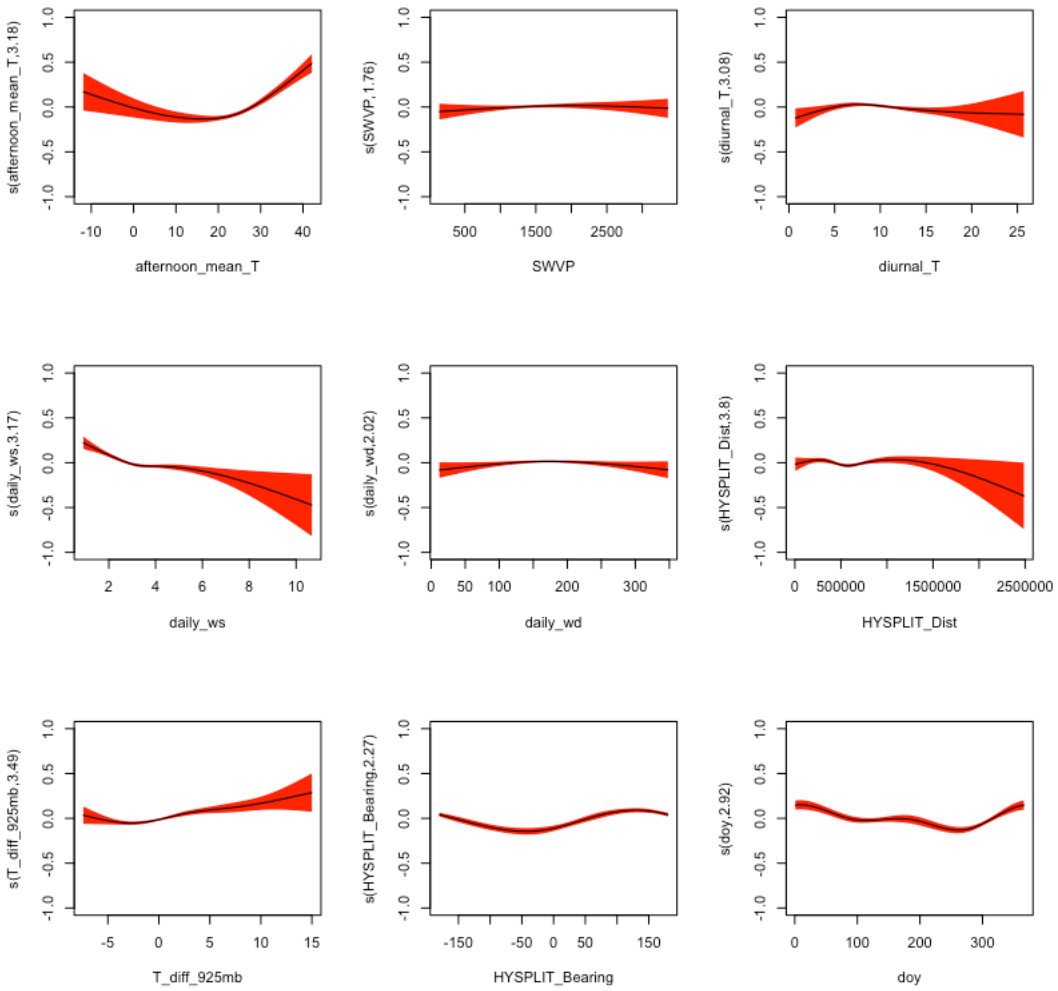

Figure 35. Smooth functions for the total daily average $PM_{2.5}$ GAM fit in the area of Dallas/Fort Worth, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the total daily average $PM_{2.5}$ in µg/m3 from its mean value.

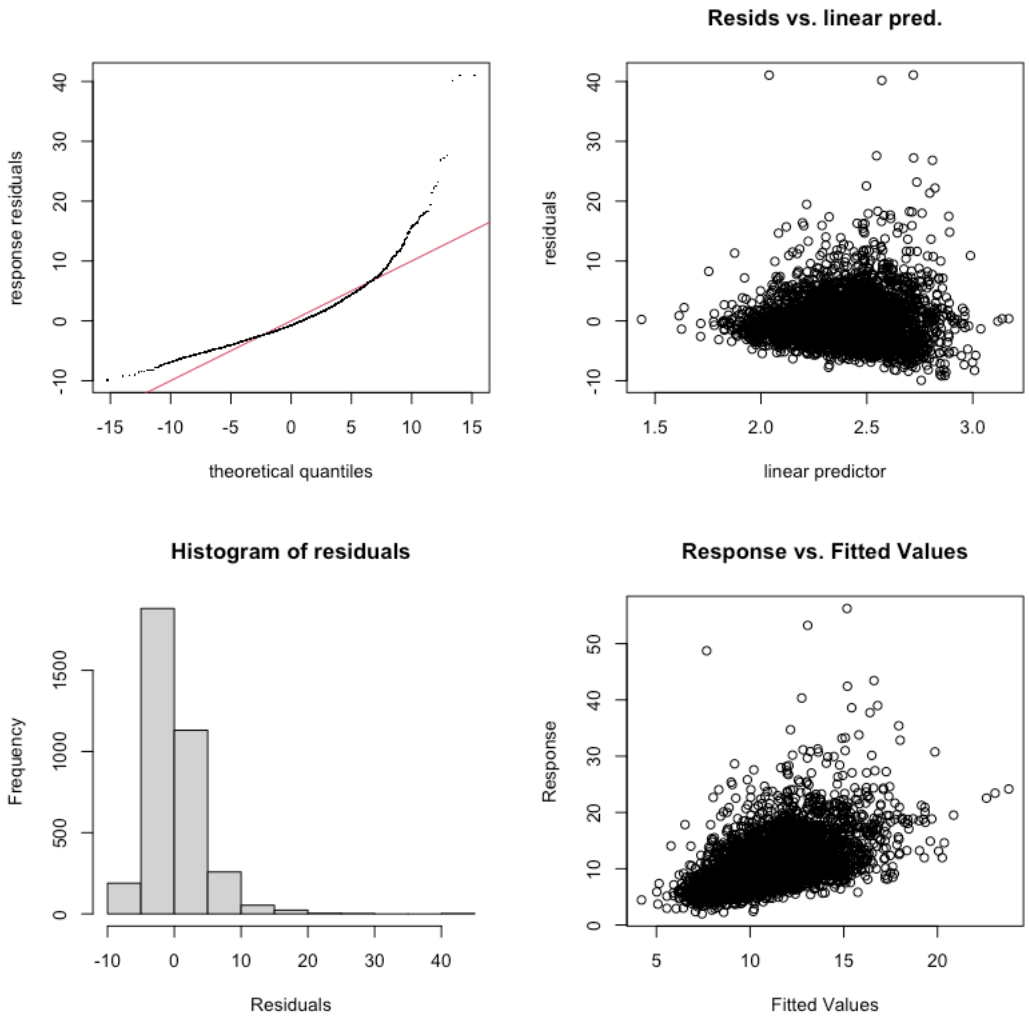

Figure 36. GAM evaluation plots for total daily average $PM_{2.5}$ in the area of Dallas/Fort Worth, TX.

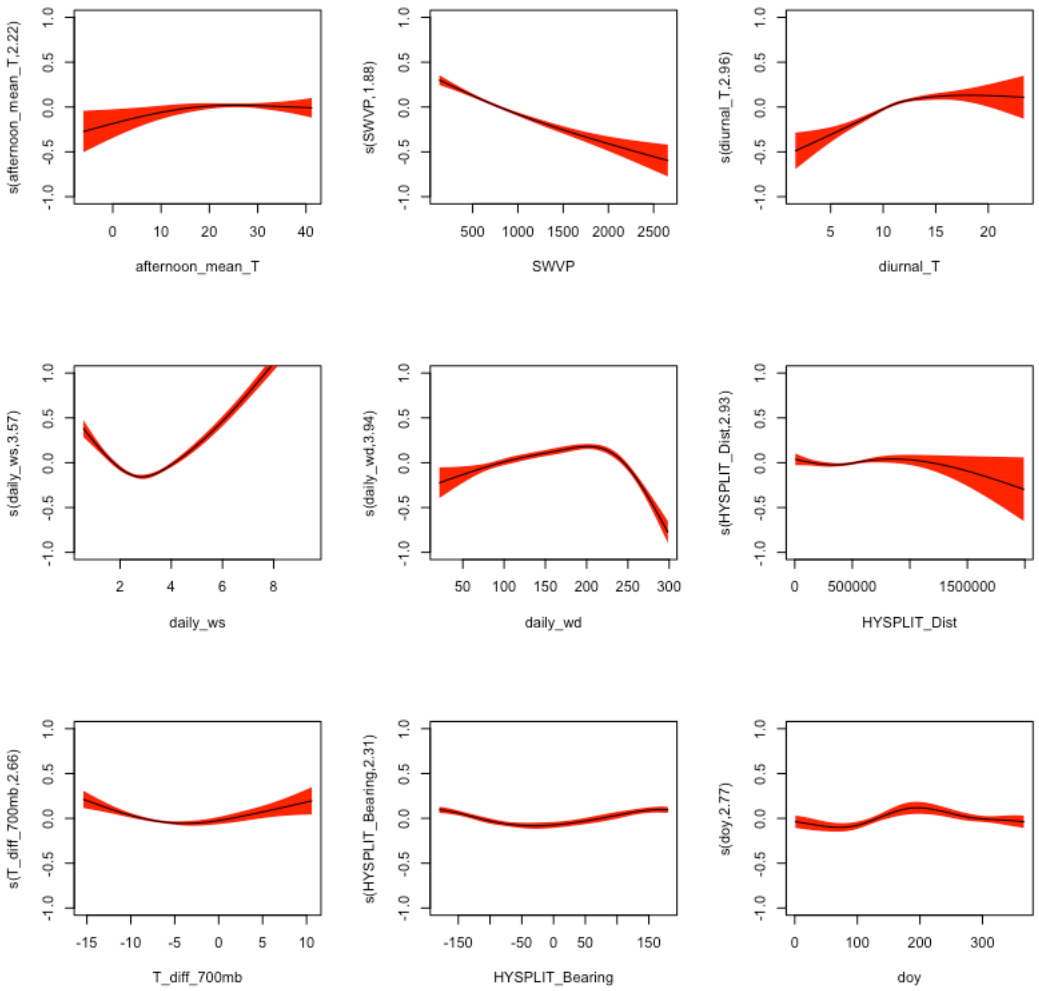

Figure 37. Smooth functions for the total daily average $PM_{2.5}$ GAM fit in the area of El Paso, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the total daily average $PM_{2.5}$ in µg/m3 from its mean value.

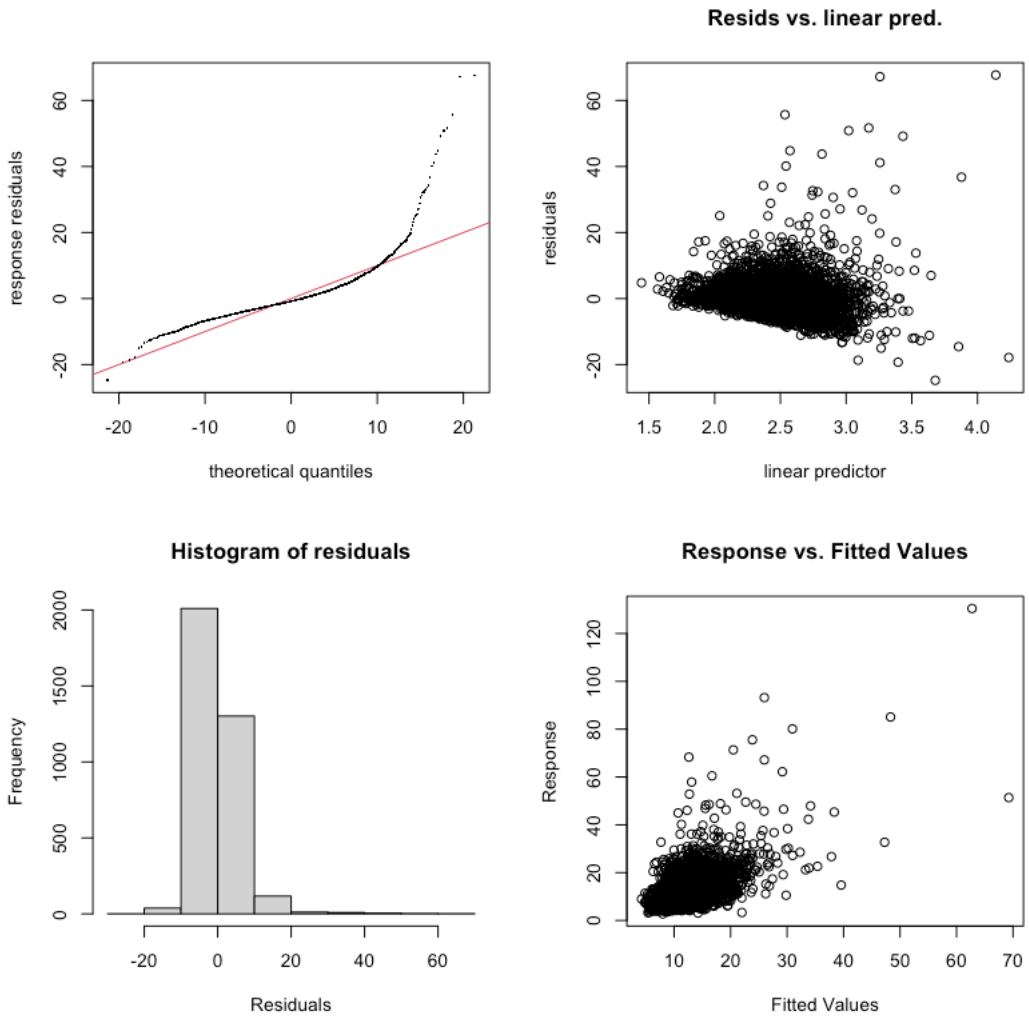

Figure 38. GAM evaluation plots for total daily average $PM_{2.5}$ in the area of El Paso, TX.

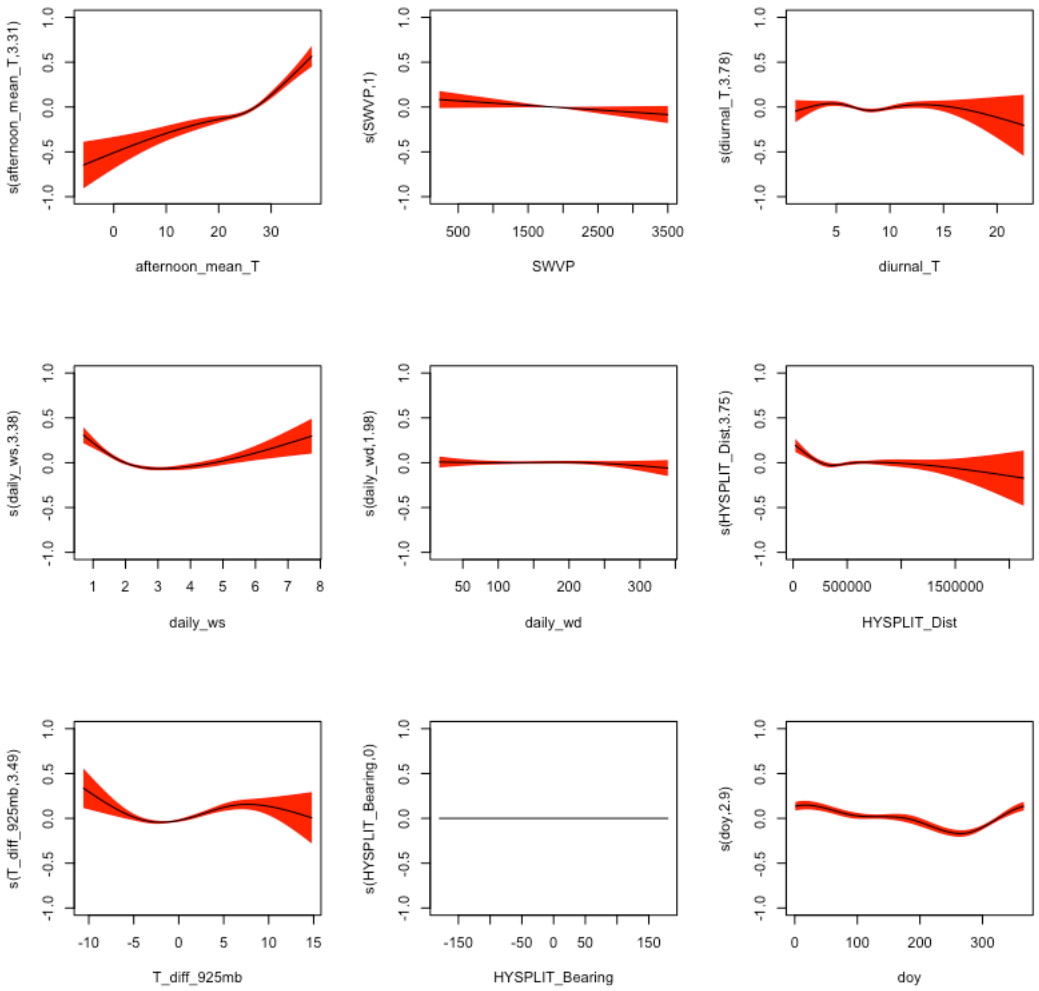

Figure 39. Smooth functions for the total daily average $PM_{2.5}$ GAM fit in the area of Houston/Galveston/Brazoria, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the total daily average $PM_{2.5}$ in µg/m3 from its mean value.

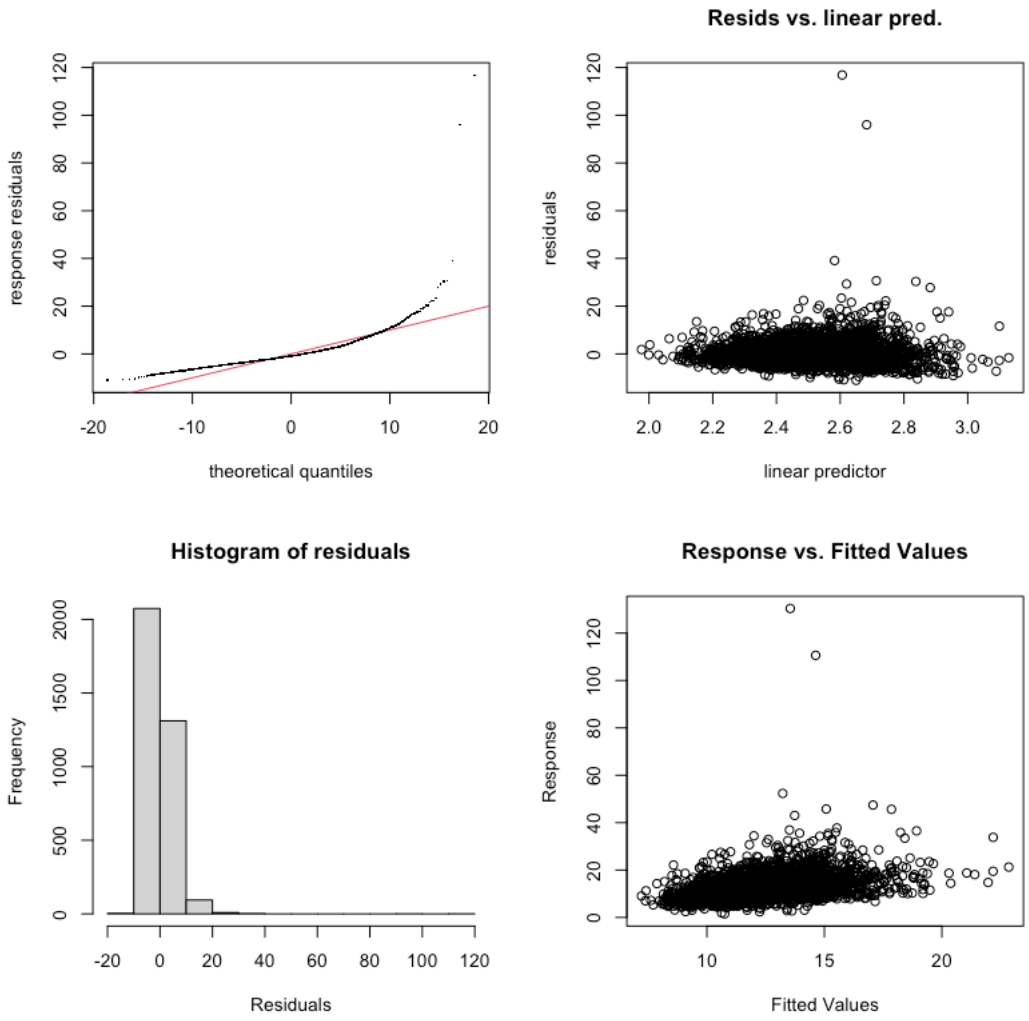

Figure 40. GAM evaluation plots for total daily average $PM_{2.5}$ in the area of Houston/Galveston/Brazoria, TX.

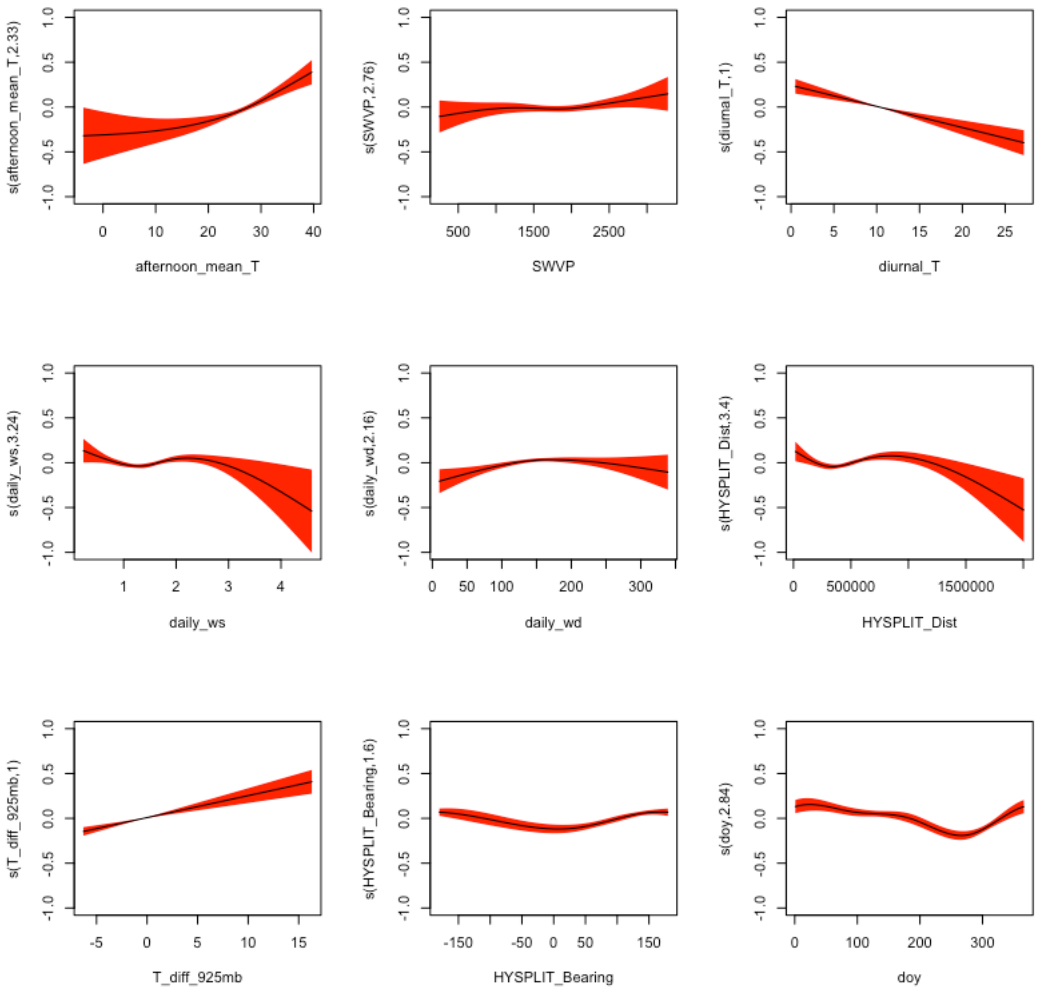

Figure 41. Smooth functions for the total daily average $PM_{2.5}$ GAM fit in the area of San Antonio, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the total daily average $PM_{2.5}$ in µg/m3 from its mean value.

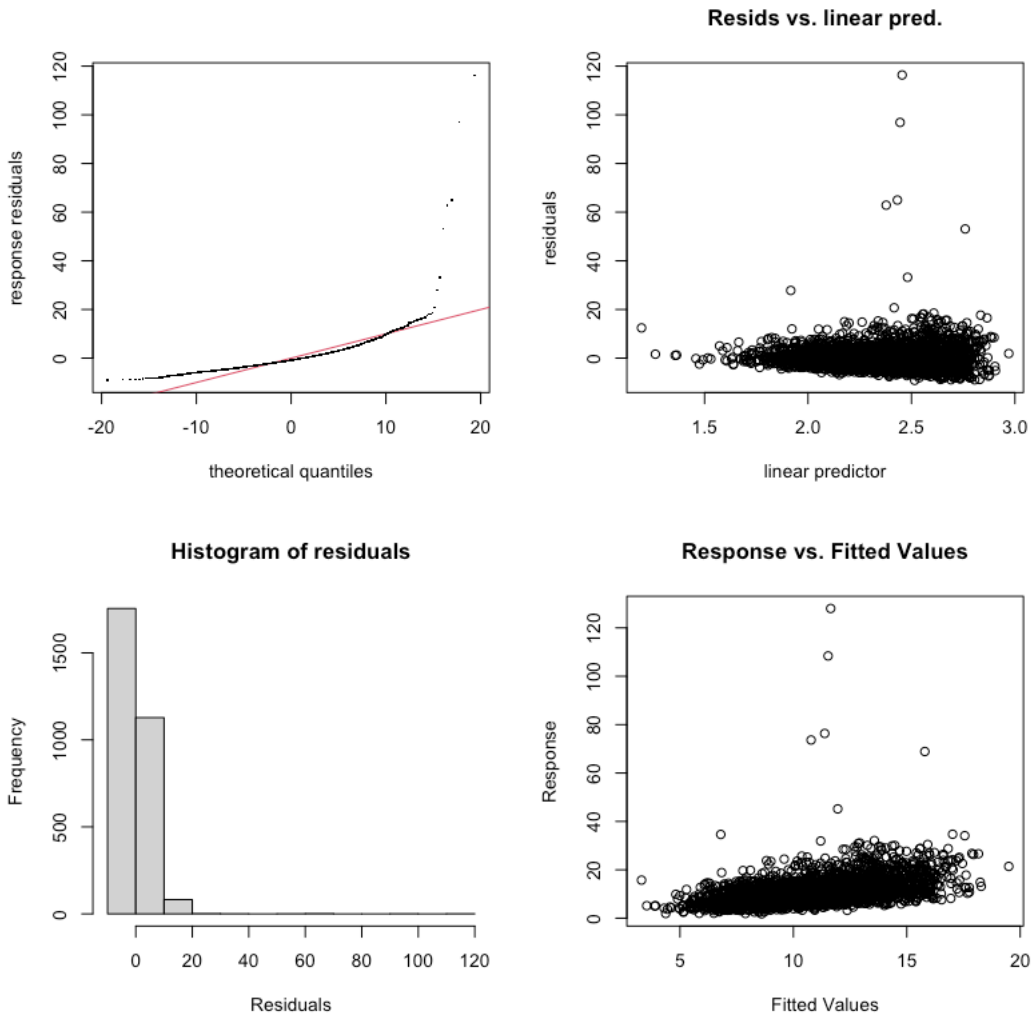

Figure 42. GAM evaluation plots for total daily average $PM_{2.5}$ in the area of San Antonio, TX.

2.2.3.2 Background $PM_{2.5}$ GAM Results

Table 11 organizes the background $PM_{2.5}$ GAM results for each of the seven urban areas. This includes figure numbers for the results, the percentage deviance of the background $PM_{2.5}$ values explained by the GAM model, and the level statistical significance of the meteorological predictors in table 7. Odd number figures are the smooth functions from the GAM fit of the natural logarithm of the background $PM_{2.5}$ values to the meteorological predictors. 95% confidence intervals are shown in red in the smooth functions. The day of year (*doy*) function is also shown. The percentage deviance values can be compared to the Camalier et al. (2007) results, which showed the predictive power of their models (measured by the R^2 statistic) to be between 0.56 and 0.80 for the cities in that study. Even number figures are the standard GAM evaluation plots (made with the *gam.check* function in R). Although slightly better than the total $PM_{2.5}$ GAM models, the low percentage deviance values and the plots and for all seven urban areas generally indicate a poorer fit (as the model residuals do not follow a normal distribution). We do not think these fits are providing

much accurate information on the variability of background $PM_{2.5}$ under different meteorological conditions.

Table 11. Performance of GAMs for background $PM_{2.5}$.

Urban Area	Figure for Smooth Functions	Figure for GAM Evaluation	% Deviance Explained	Significance of Meteorological Predictors
Austin	43	44	33.2	5 terms at α=0.001 2 terms at α=0.01 1 term at α=0.05
Beaumont/Port Arthur	45	46	24.7	5 terms at α=0.001 2 terms at α=0.01 1 term at α=1
Corpus Christi	47	48	38.4	7 terms at α=0.001 1 term at α=0.01
Dallas/Fort Worth	49	50	38.5	3 terms at α=0.001 2 terms at α=0.01 2 terms at α=0.05 1 term at α=1
El Paso	51	52	21.9	4 terms at α=0.001 2 terms at α=0.05 2 terms at α=1
Houston/Galveston/Brazoria	53	54	31.6	4 terms at α=0.001 4 terms at α=1
San Antonio	55	56	30.2	6 terms at α=0.001 1 term at α=0.01 1 term at α=0.1

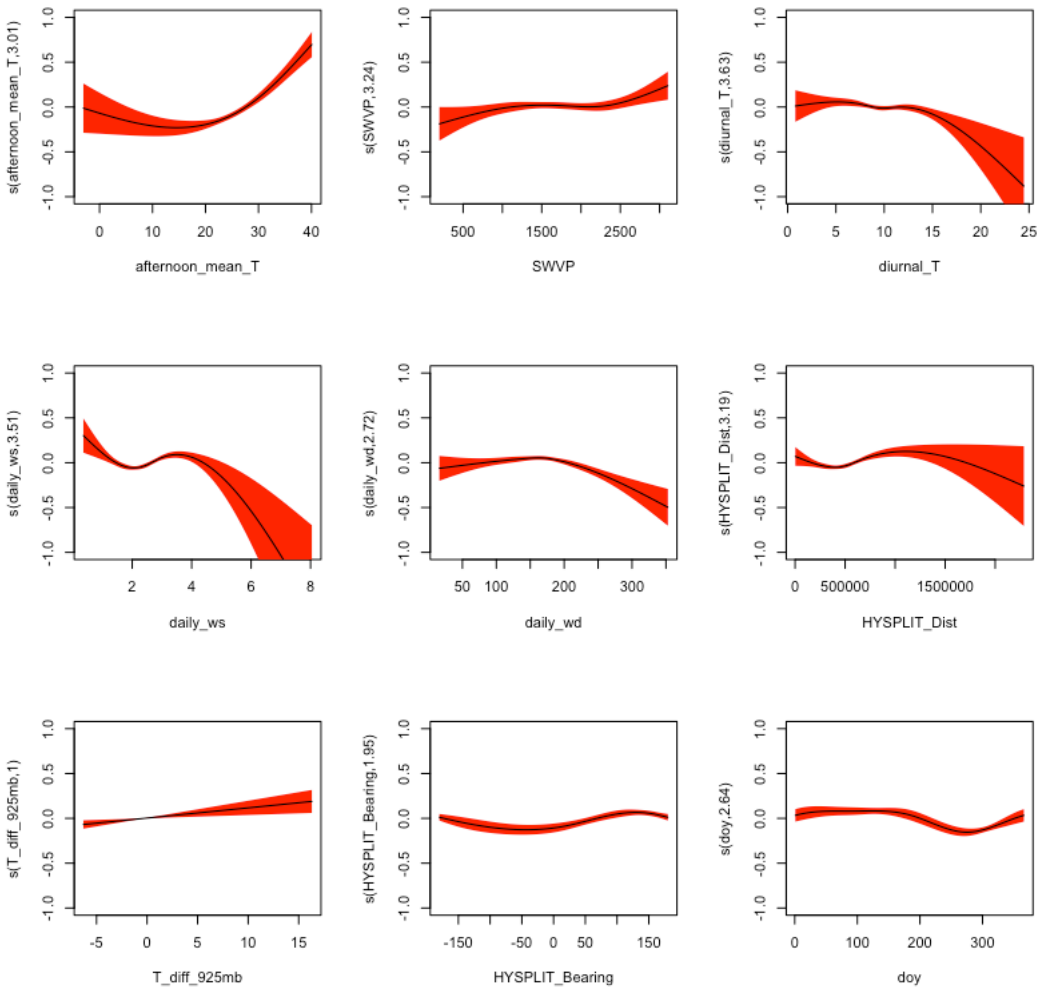

Figure 43. Smooth functions for the background $PM_{2.5}$ GAM fit in the area of Austin, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background $PM_{2.5}$ in µg/m3 from its mean value.

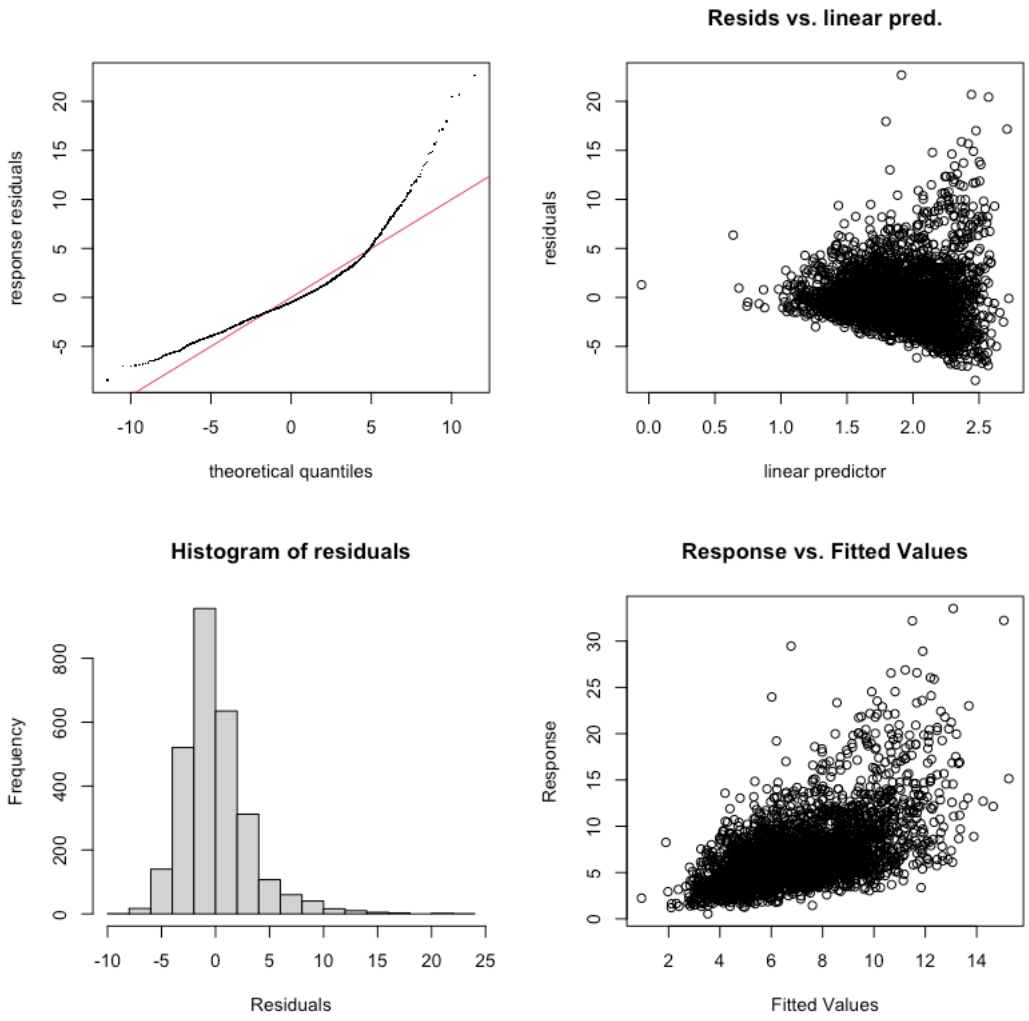

Figure 44. GAM evaluation plots for background $PM_{2.5}$ in the area of Austin, TX.

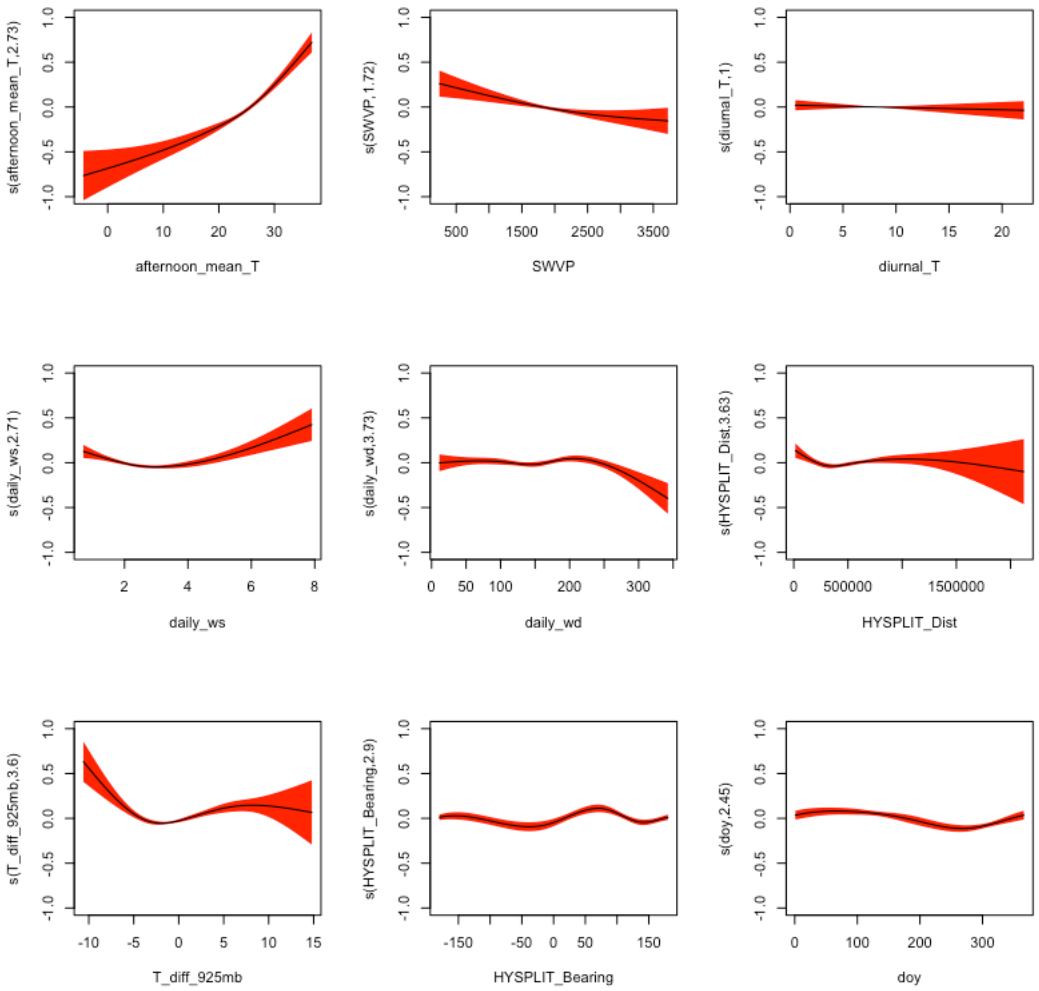

Figure 45. Smooth functions for the background $PM_{2.5}$ GAM fit in the area of Beaumont/Port Arthur, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background $PM_{2.5}$ in µg/m3 from its mean value.

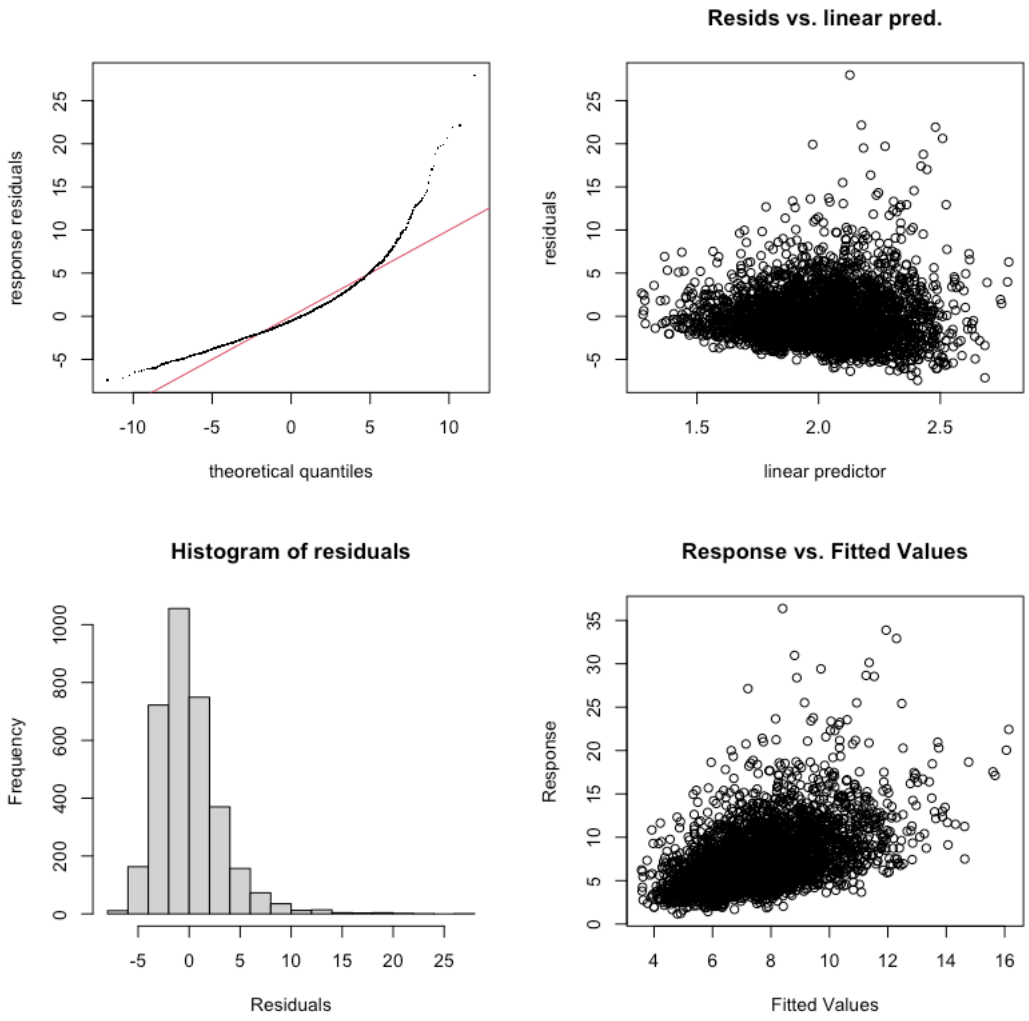

Figure 46. GAM evaluation plots for background $PM_{2.5}$ in the area of Beaumont/Port Arthur, TX.

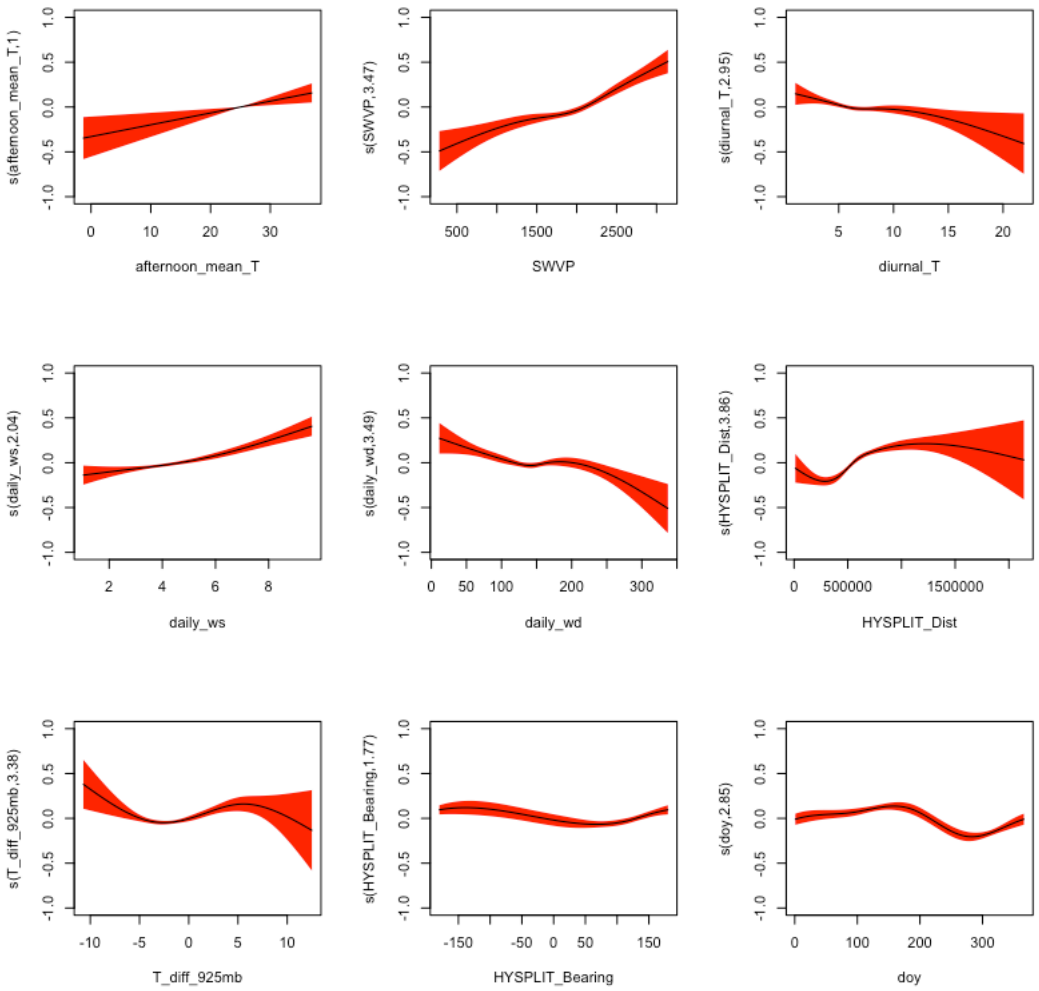

Figure 47. Smooth functions for the background $PM_{2.5}$ GAM fit in the area of Corpus Christi, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background $PM_{2.5}$ in µg/m3 from its mean value.

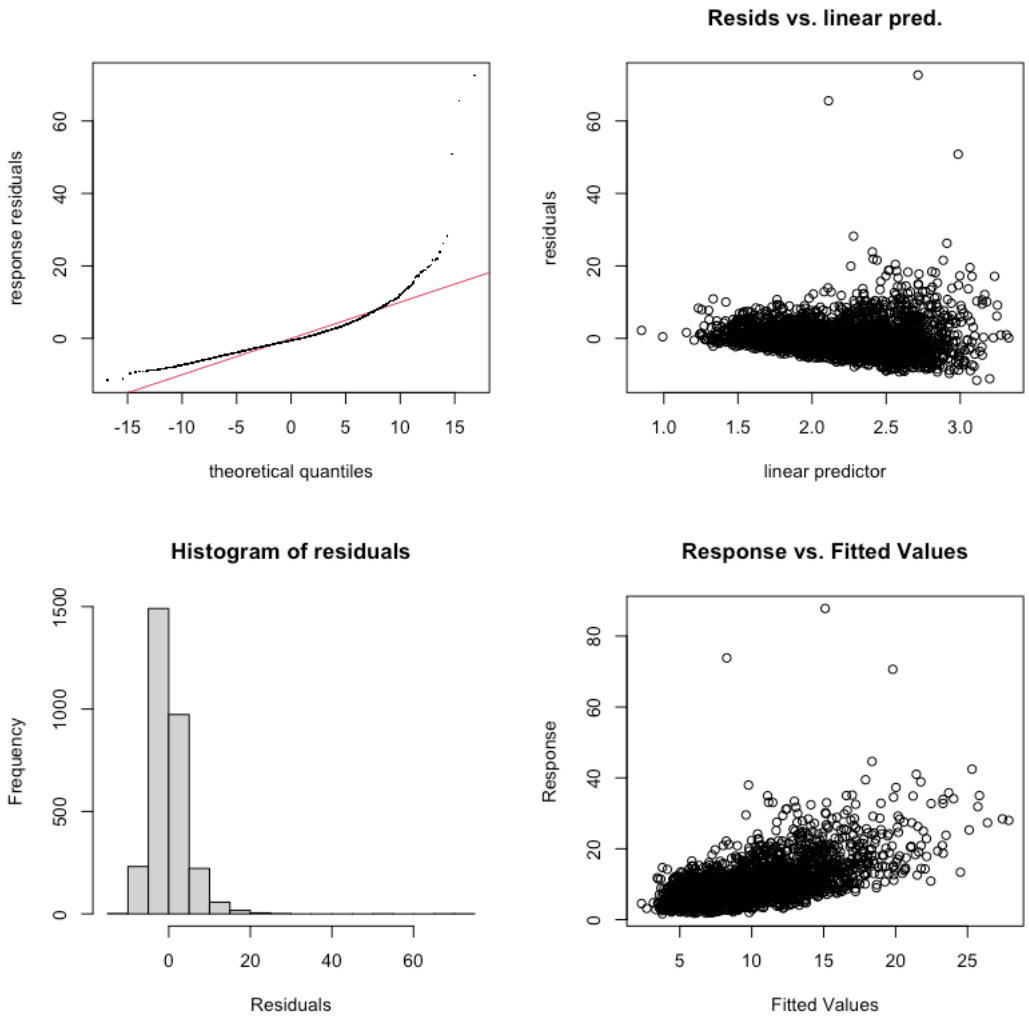

Figure 48. GAM evaluation plots for background $PM_{2.5}$ in the area of Corpus Christi, TX.

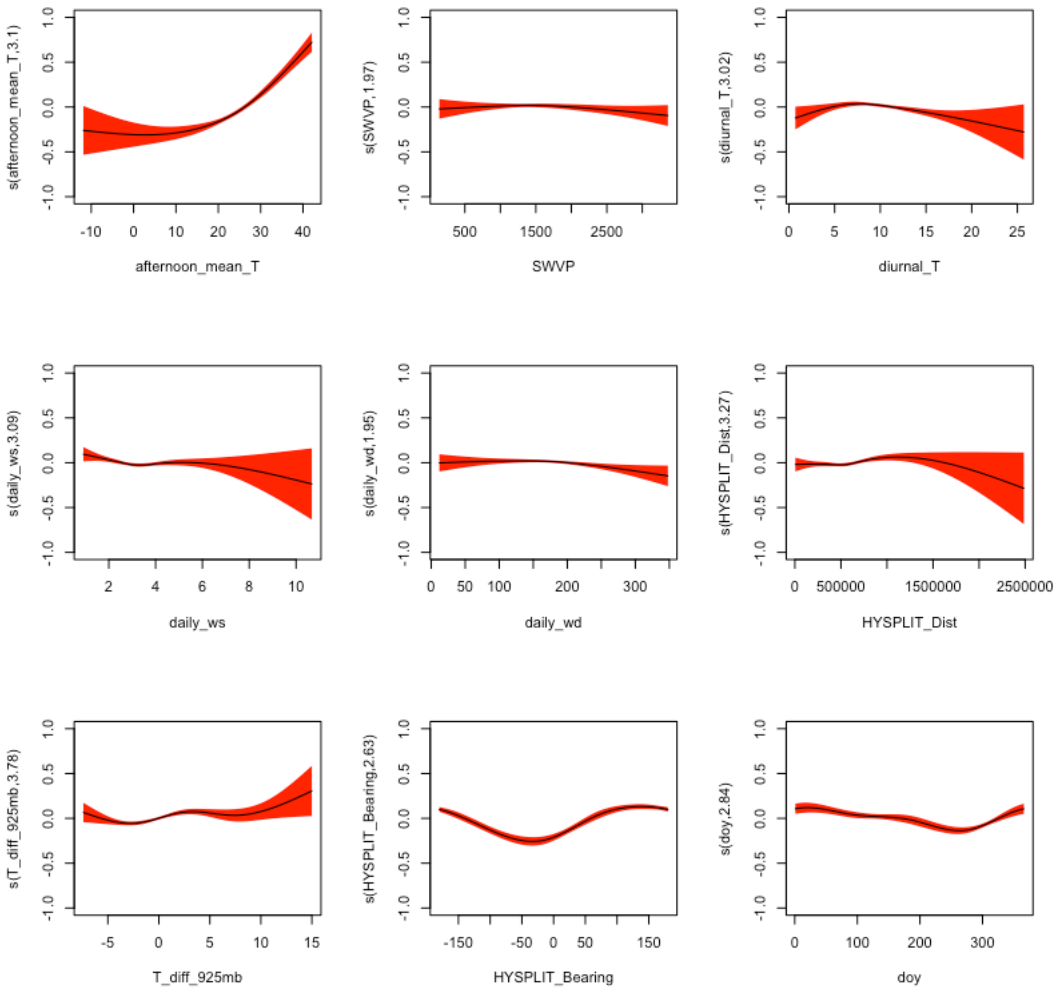

Figure 49. Smooth functions for the background $PM_{2.5}$ GAM fit in the area of Dallas/Fort Worth, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background $PM_{2.5}$ in µg/m3 from its mean value.

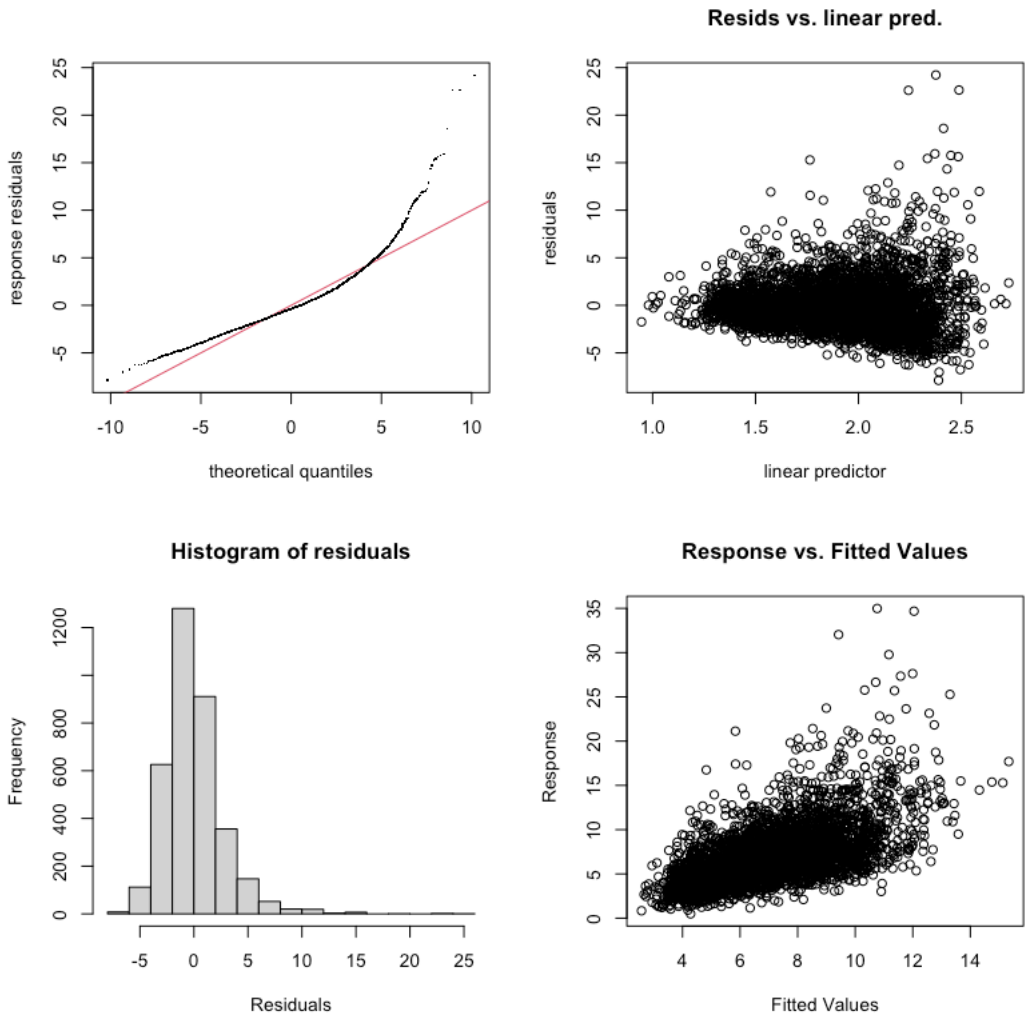

Figure 50. GAM evaluation plots for background $PM_{2.5}$ in the area of Dallas/Fort Worth, TX.

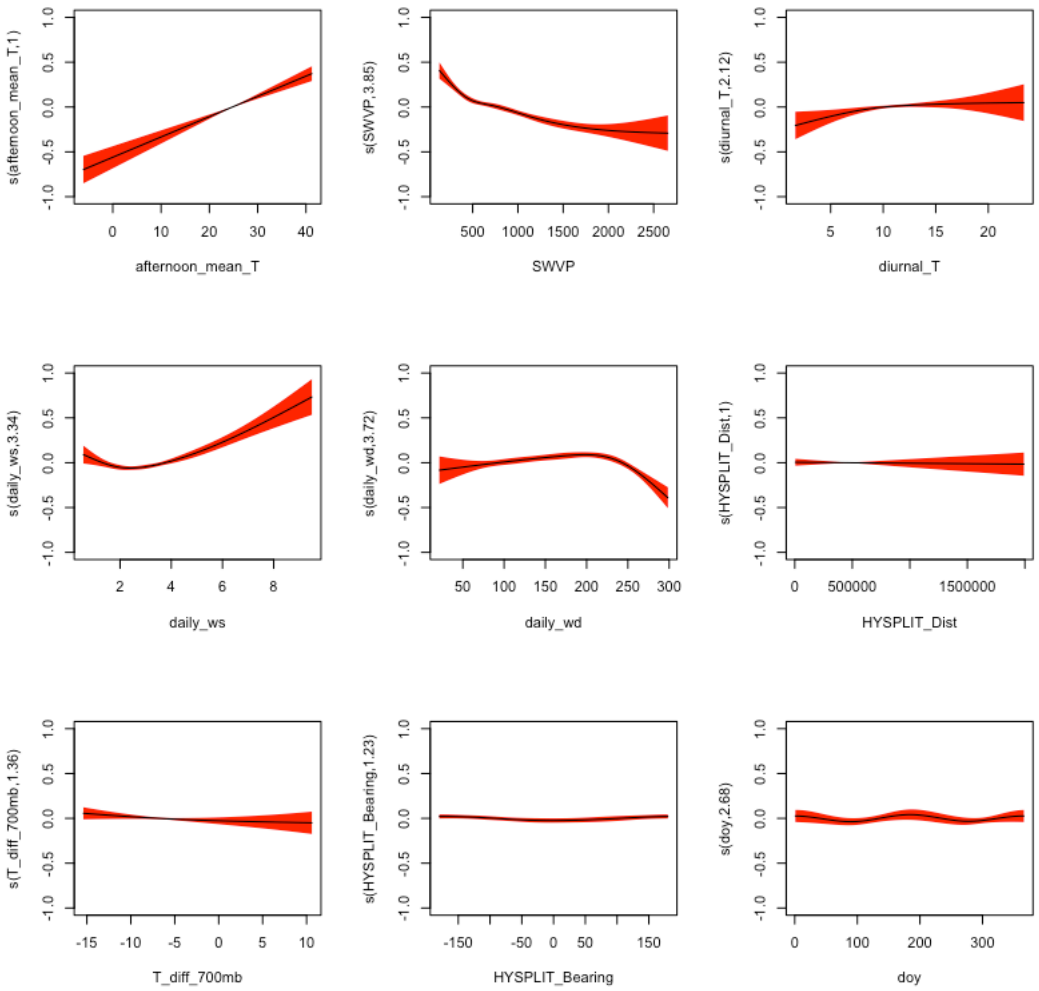

Figure 51. Smooth functions for the background $PM_{2.5}$ GAM fit in the area of El Paso, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background $PM_{2.5}$ in µg/m3 from its mean value.

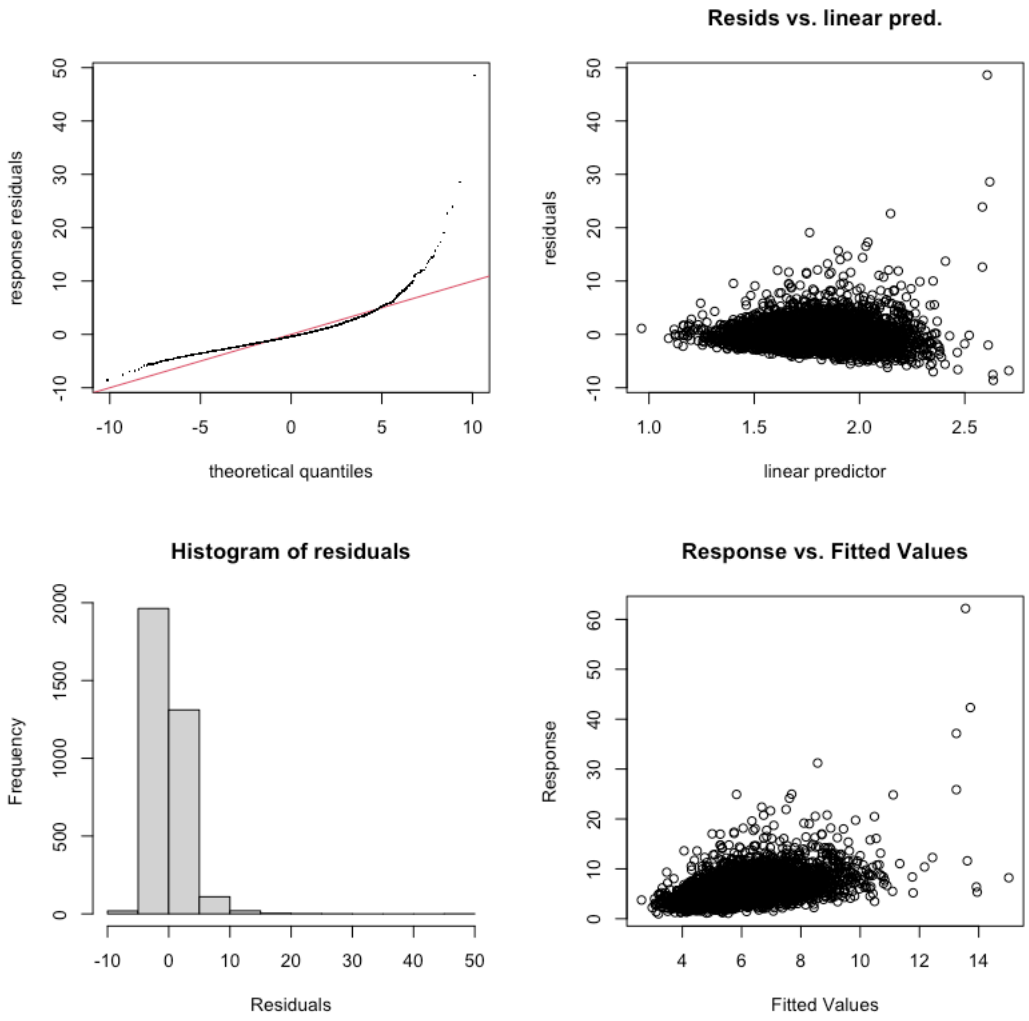

Figure 52. GAM evaluation plots for background $PM_{2.5}$ in the area of El Paso, TX.

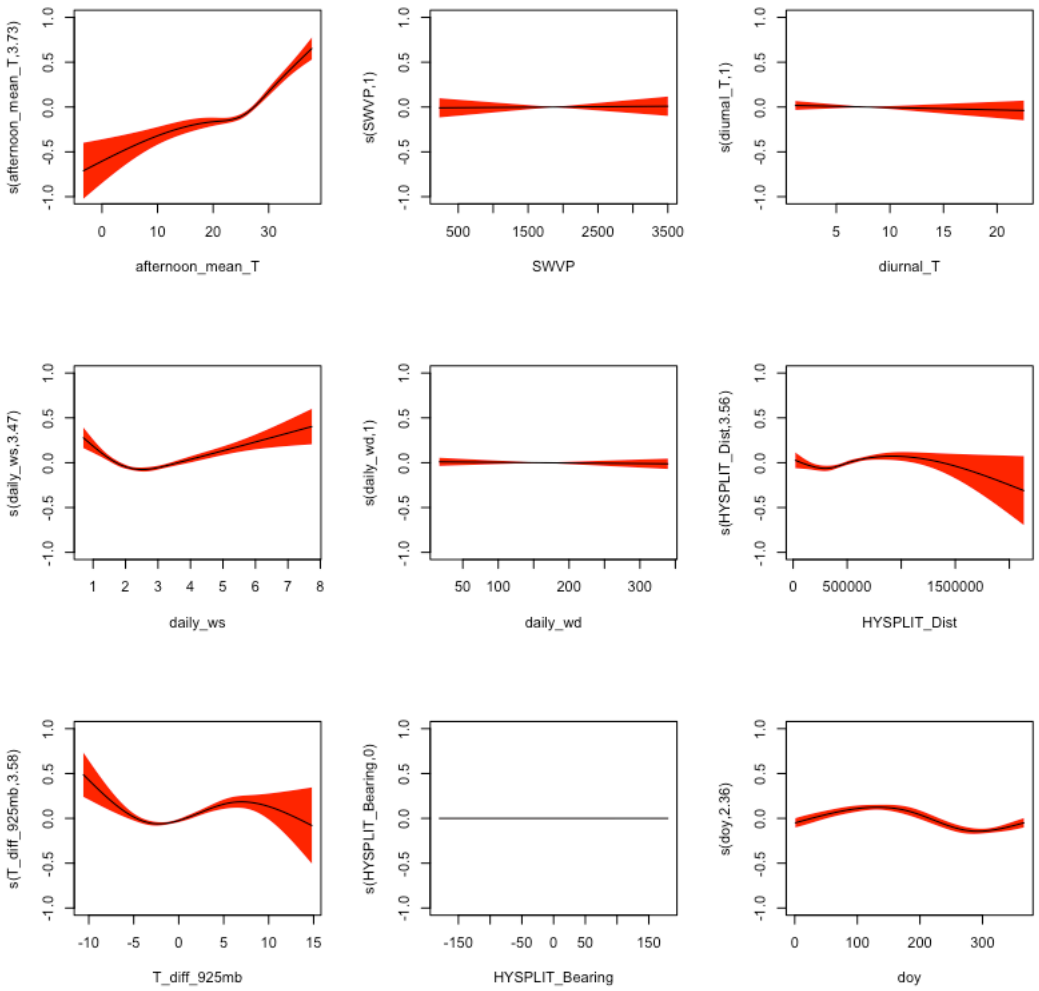

Figure 53. Smooth functions for the background $PM_{2.5}$ GAM fit in the area of Houston/Galveston/Brazoria, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background $PM_{2.5}$ in µg/m3 from its mean value.

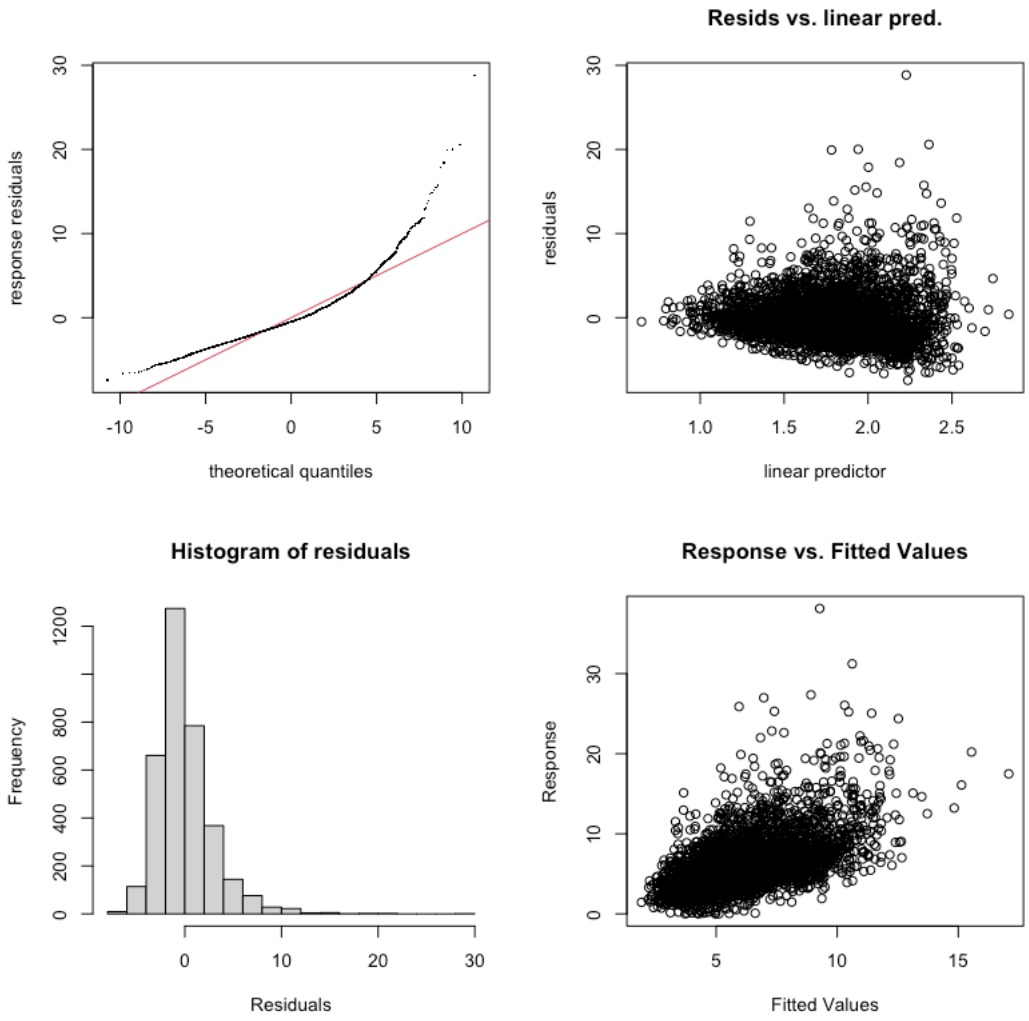

Figure 54. GAM evaluation plots for background $PM_{2.5}$ in the area of Houston/Galveston/Brazoria, TX.

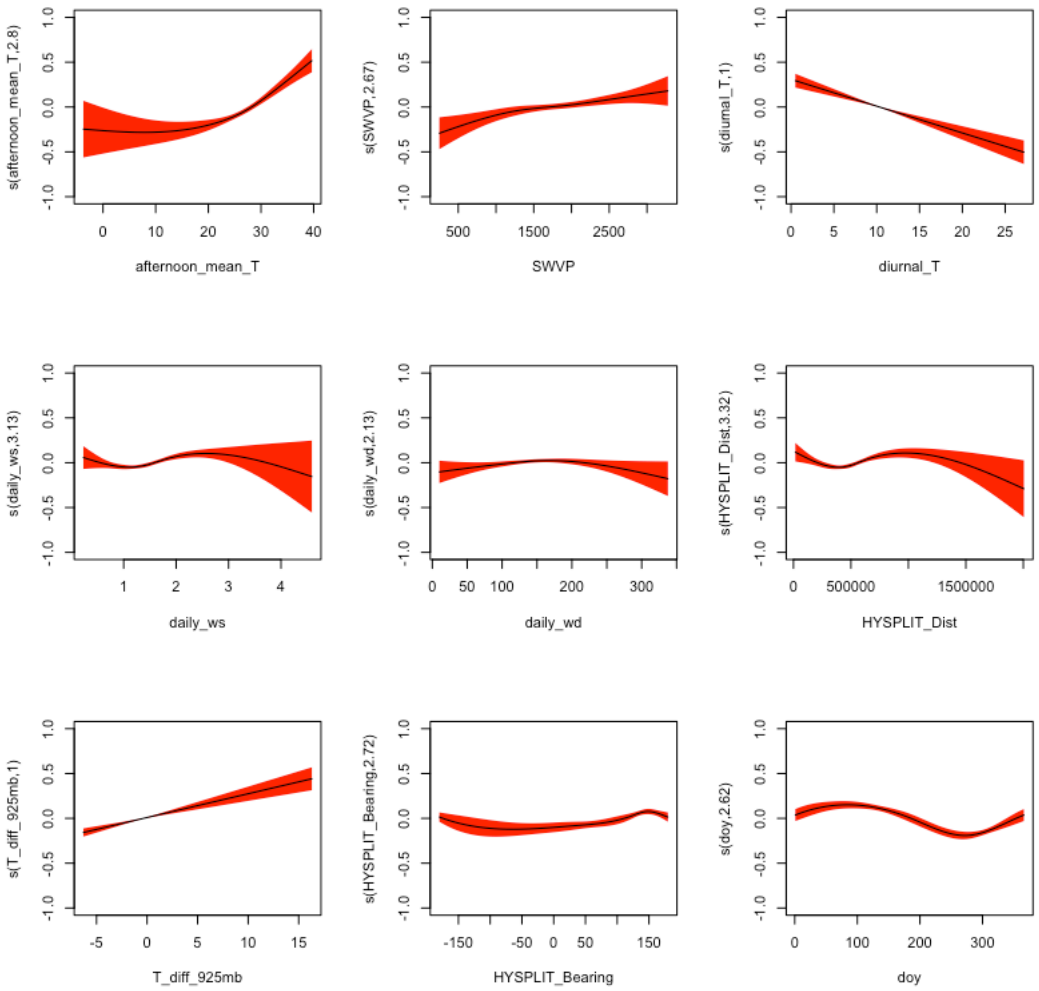

Figure 55. Smooth functions for the background $PM_{2.5}$ GAM fit in the area of San Antonio, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background $PM_{2.5}$ in µg/m3 from its mean value.

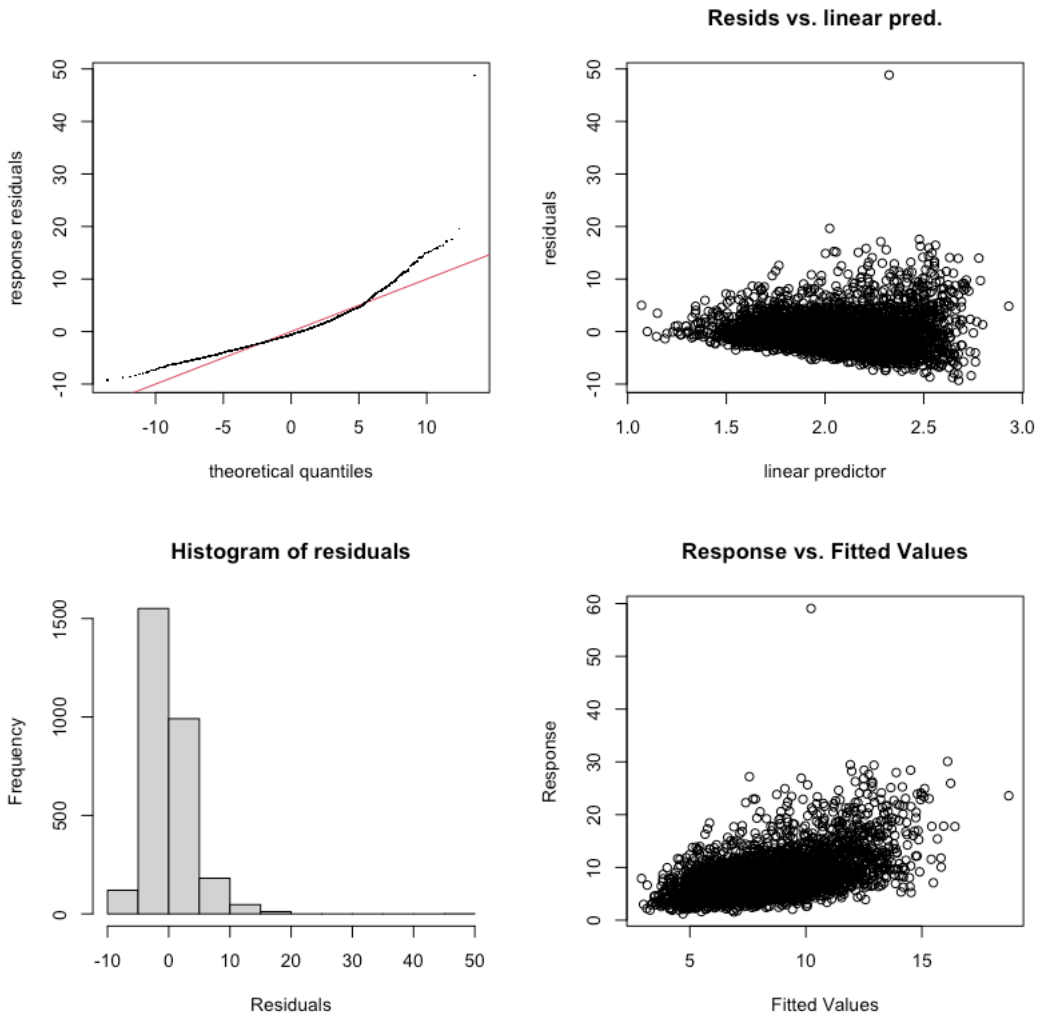

Figure 56. GAM evaluation plots for background $PM_{2.5}$ in the area of San Antonio, TX.

2.2.4 SO_2 GAM Results

2.2.4.1 Total SO_2 GAM Results

Table 12 organizes the total SO_2 GAM results for each of the seven urban areas. This includes figure numbers for the results, the percentage deviance of the total SO_2 values explained by the GAM model, and the level statistical significance of the meteorological predictors in table 7. Odd number figures are the smooth functions from the GAM fit of the natural logarithm of the total SO_2 values to the meteorological predictors. 95% confidence intervals are shown in red in the smooth functions. The day of year (*doy*) function is also shown. The percentage deviance values can be compared to the Camalier et al. (2007) results, which showed the predictive power of their models (measured by the R^2 statistic) to be between 0.56 and 0.80 for the cities in that study. Even number figures are the standard GAM evaluation plots (made with the *gam.check* function in R). The low percentage deviance values and the plots and for all seven urban areas indicate a poor fit (as the model residuals do not follow a normal distribution and show a trend versus predicted

value). We do not think these fits are providing much accurate information on the variability of total SO_2 under different meteorological conditions.

Table 12. Performance of GAMs for Total SO_2.

Urban Area	Figure for Smooth Functions	Figure for GAM Evaluation	% Deviance Explained	Significance of Meteorological Predictors
Austin	57	58	38.1	7 terms at α=0.001 1 term at α=1
Beaumont/Port Arthur	59	60	26	6 terms at α=0.001 1 term at α=0.01 1 term at α=1
Corpus Christi	61	62	32.4	7 terms at α=0.001 1 term at α=0.01
Dallas/Fort Worth	63	64	22.5	6 terms at α=0.001 1 term at α=0.1 1 term at α=1
El Paso	65	66	56.9	7 terms at α=0.001 1 term at α=0.01
Houston/Galveston/Brazoria	67	68	14.4	3 terms at α=0.001 1 term at α=0.1 4 terms at α=1
San Antonio	69	70	45.7	6 terms at α=0.001 2 terms at α=1

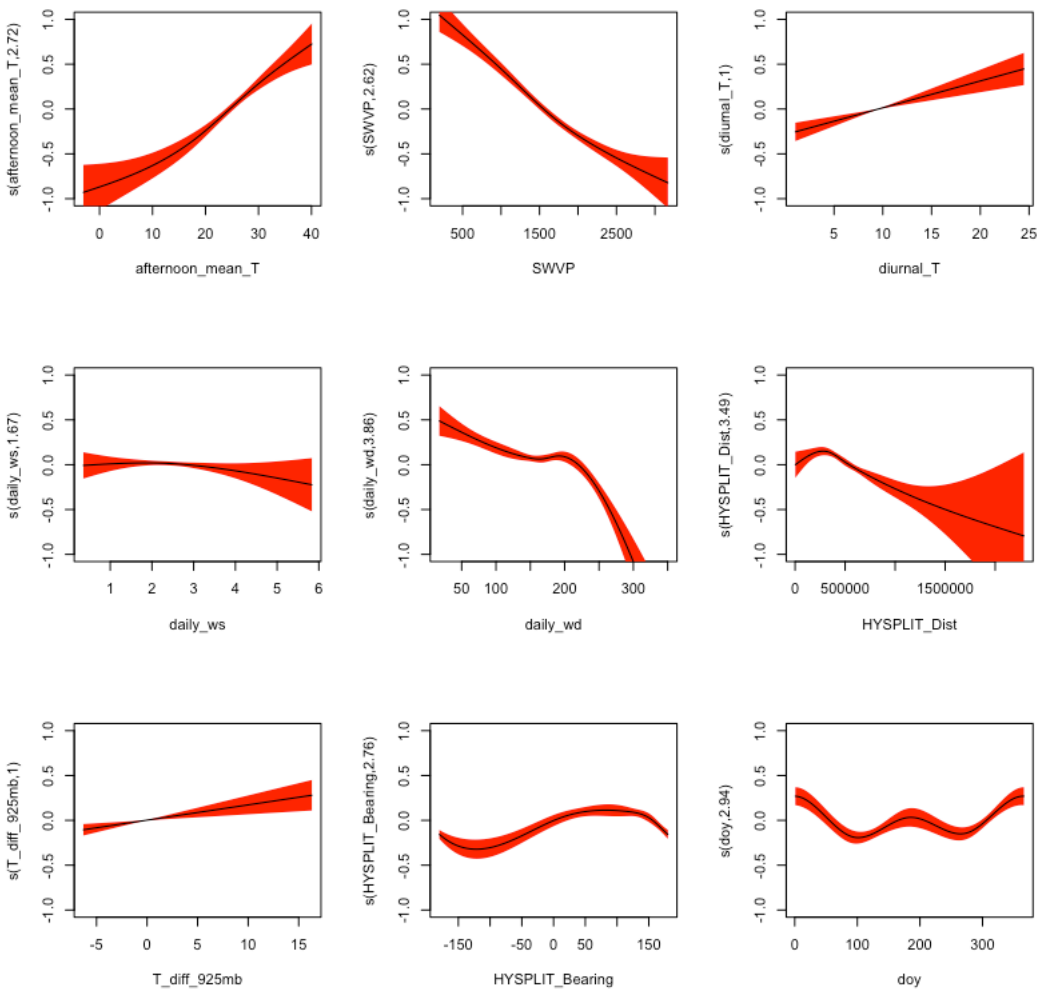

Figure 57. Smooth functions for the maximum daily 1-hour average SO_2 GAM fit in the area of Austin, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the maximum daily 1-hour average SO_2 in ppbv from its mean value.

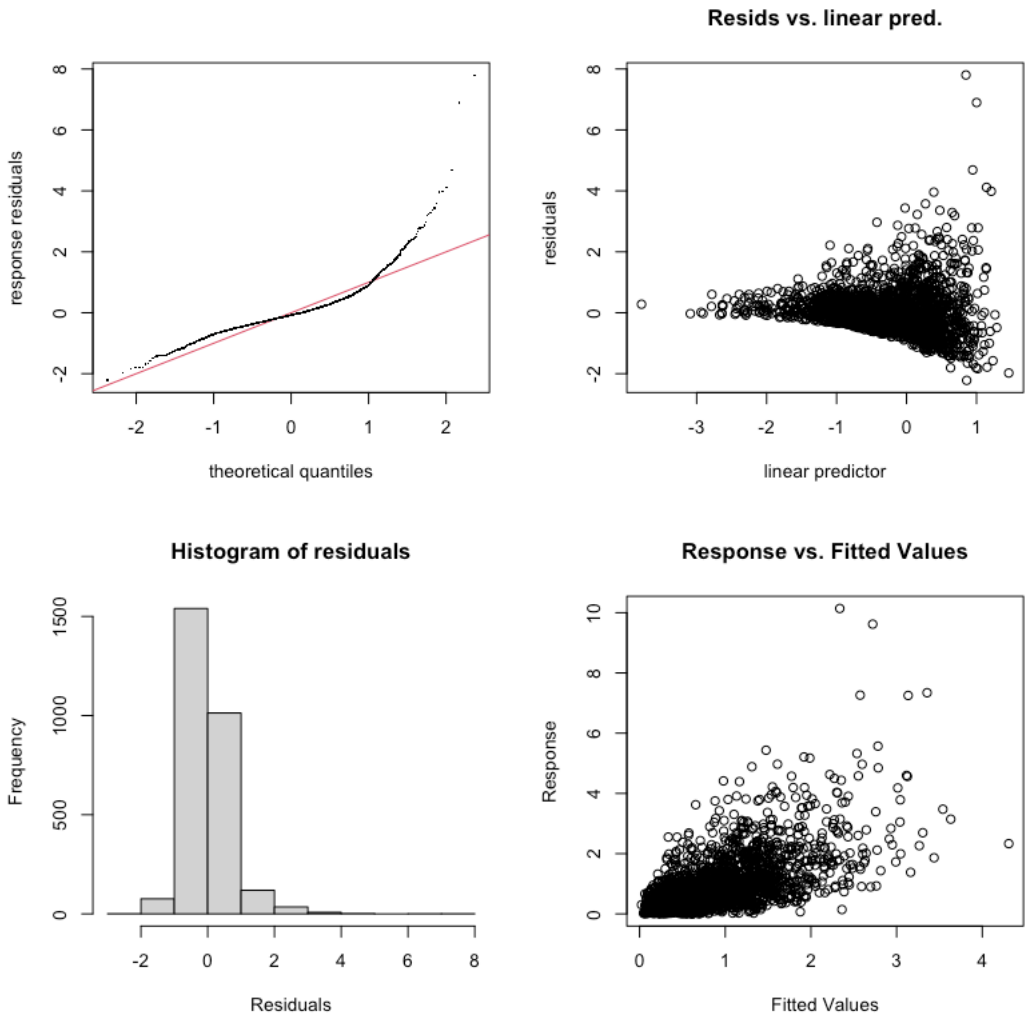

Figure 58. GAM evaluation plots for maximum daily 1-hour average SO_2 in the area of Austin, TX.

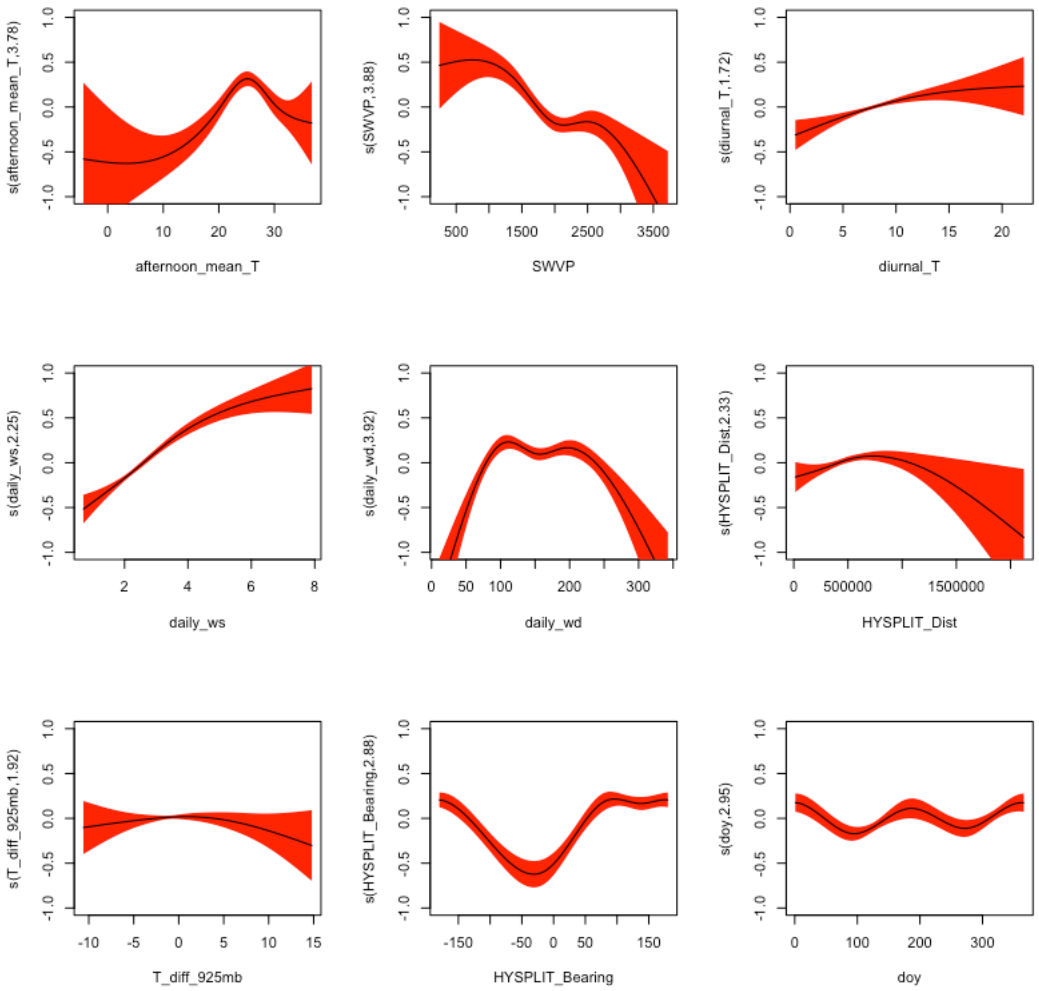

Figure 59. Smooth functions for the maximum daily 1-hour average SO_2 GAM fit in the area of Beaumont/Port Arthur, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the maximum daily 1-hour average SO_2 in ppbv from its mean value.

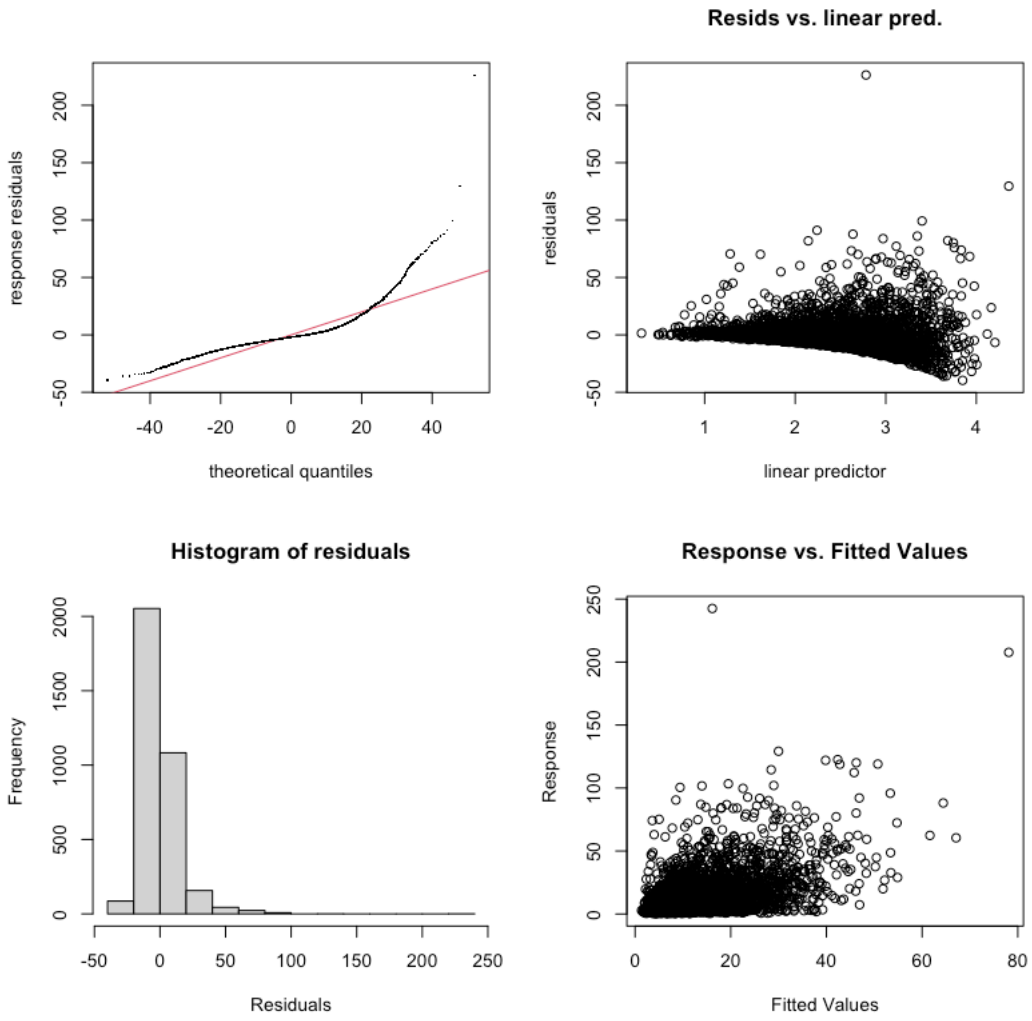

Figure 60. GAM evaluation plots for maximum daily 1-hour average SO_2 in the area of Beaumont/Port Arthur, TX.

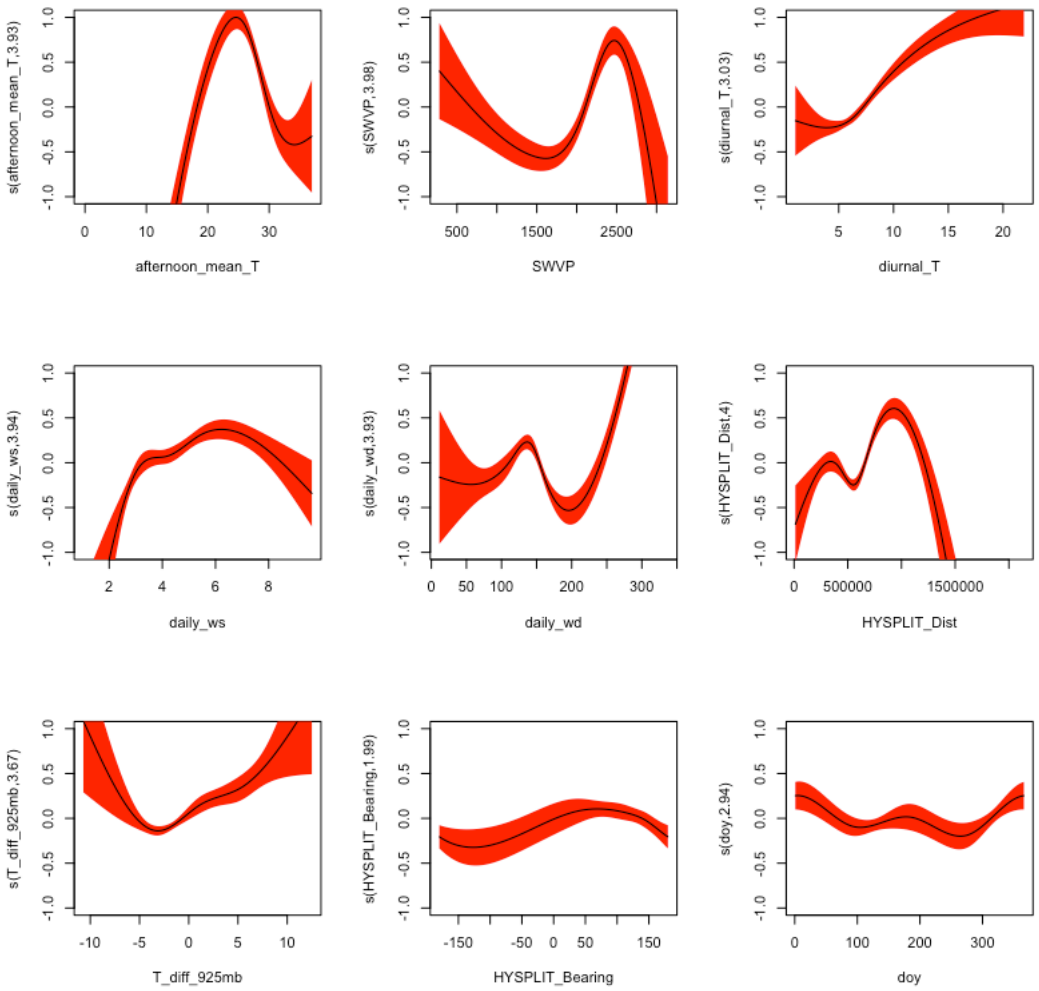

Figure 61. Smooth functions for the maximum daily 1-hour average SO_2 GAM fit in the area of Corpus Christi, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the maximum daily 1-hour average SO_2 in ppbv from its mean value.

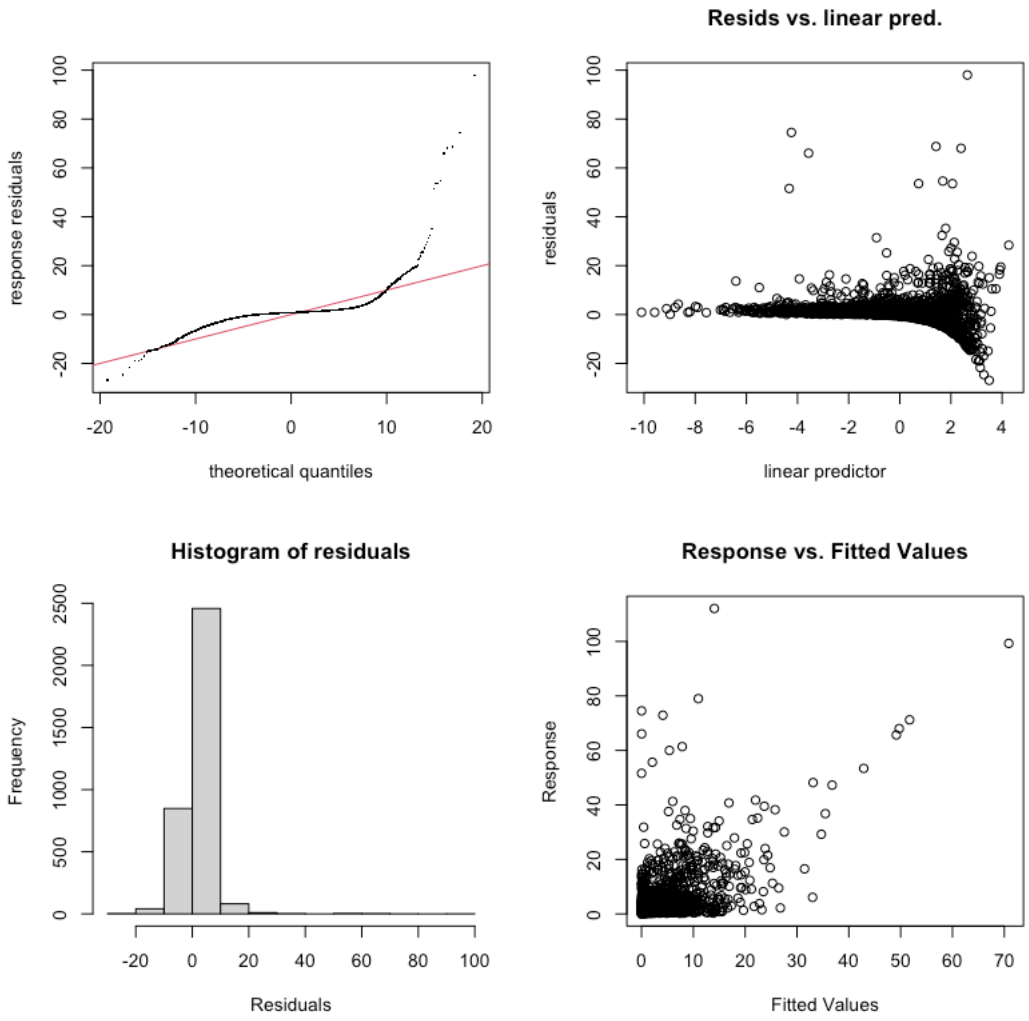

Figure 62. GAM evaluation plots for maximum daily 1-hour average SO_2 in the area of Corpus Christi, TX.

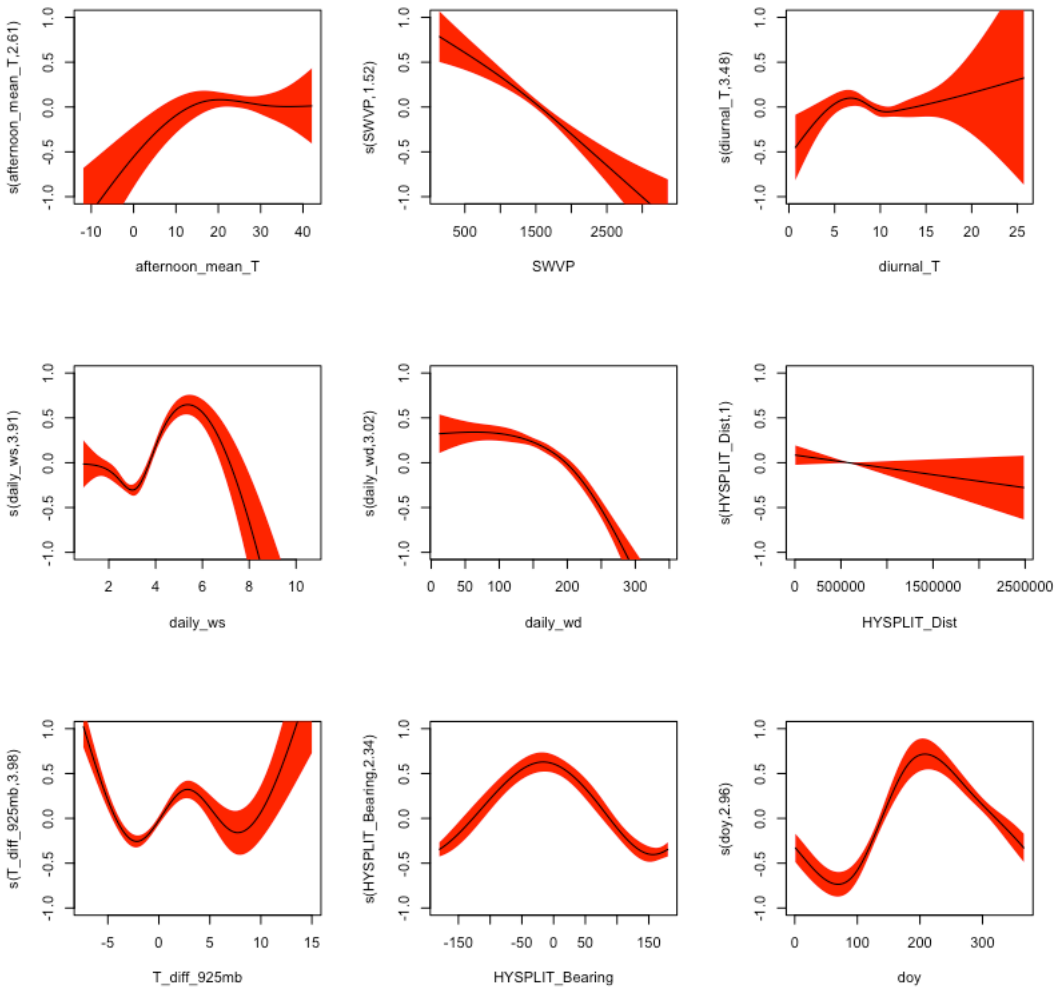

Figure 63. Smooth functions for the maximum daily 1-hour average SO_2 GAM fit in the area of Dallas/Fort Worth, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the maximum daily 1-hour average SO_2 in ppbv from its mean value.

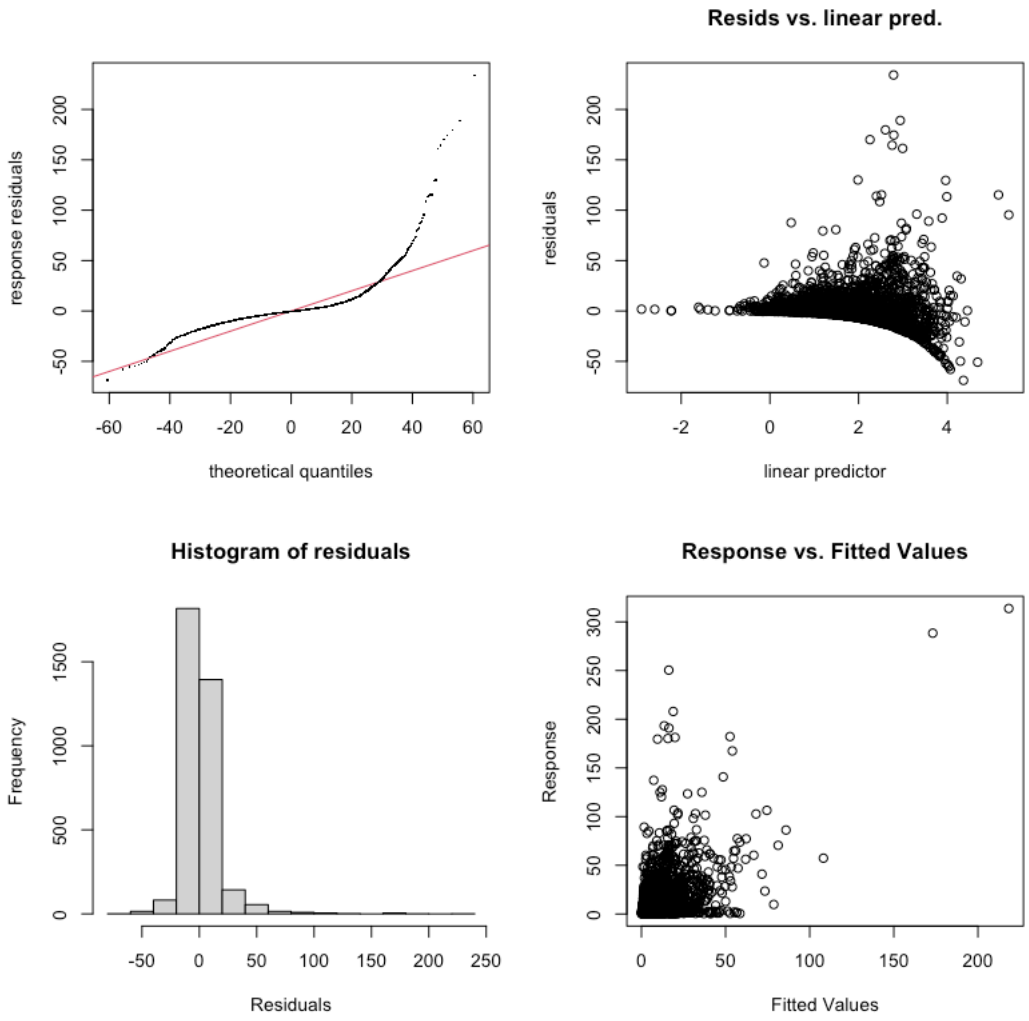

Figure 64. GAM evaluation plots for maximum daily 1-hour average SO_2 in the area of Dallas/Fort Worth, TX.

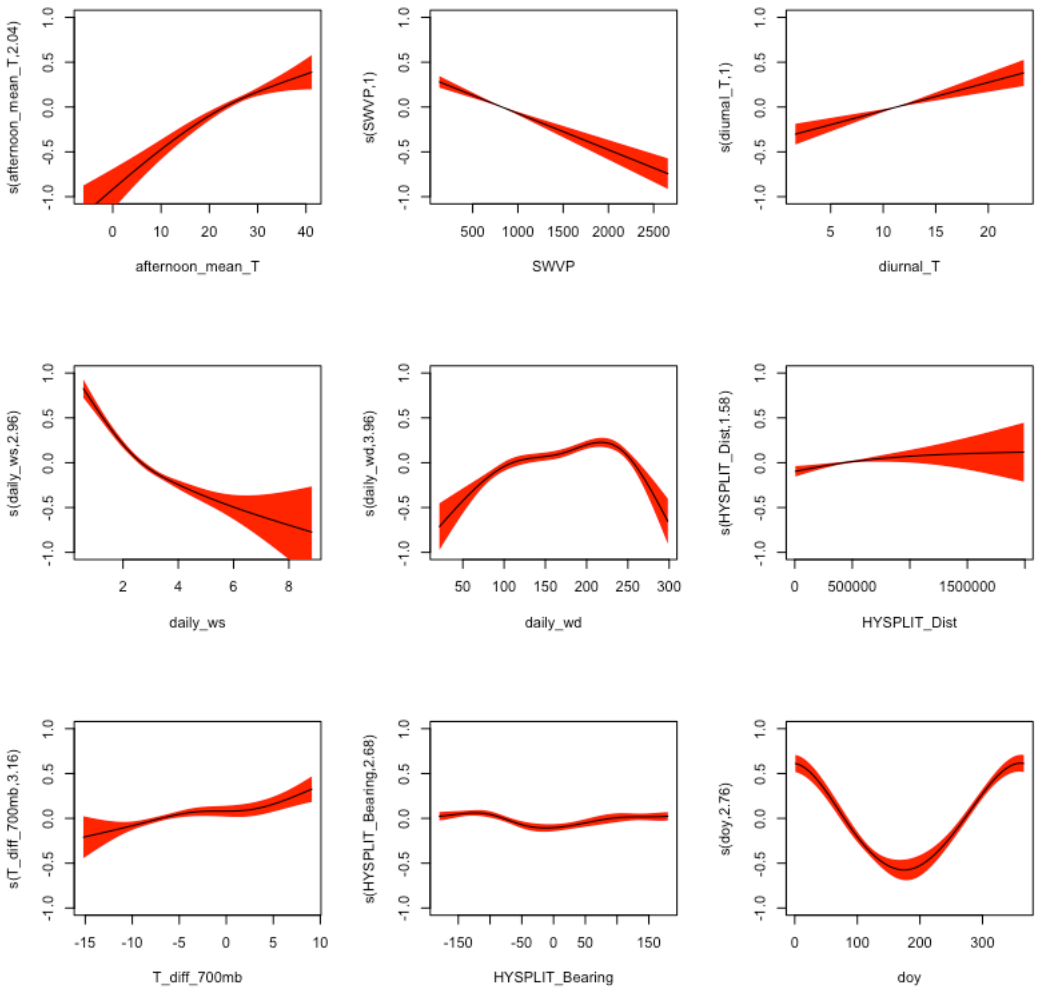

Figure 65. Smooth functions for the maximum daily 1-hour average SO_2 GAM fit in the area of El Paso, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the maximum daily 1-hour average SO_2 in ppbv from its mean value.

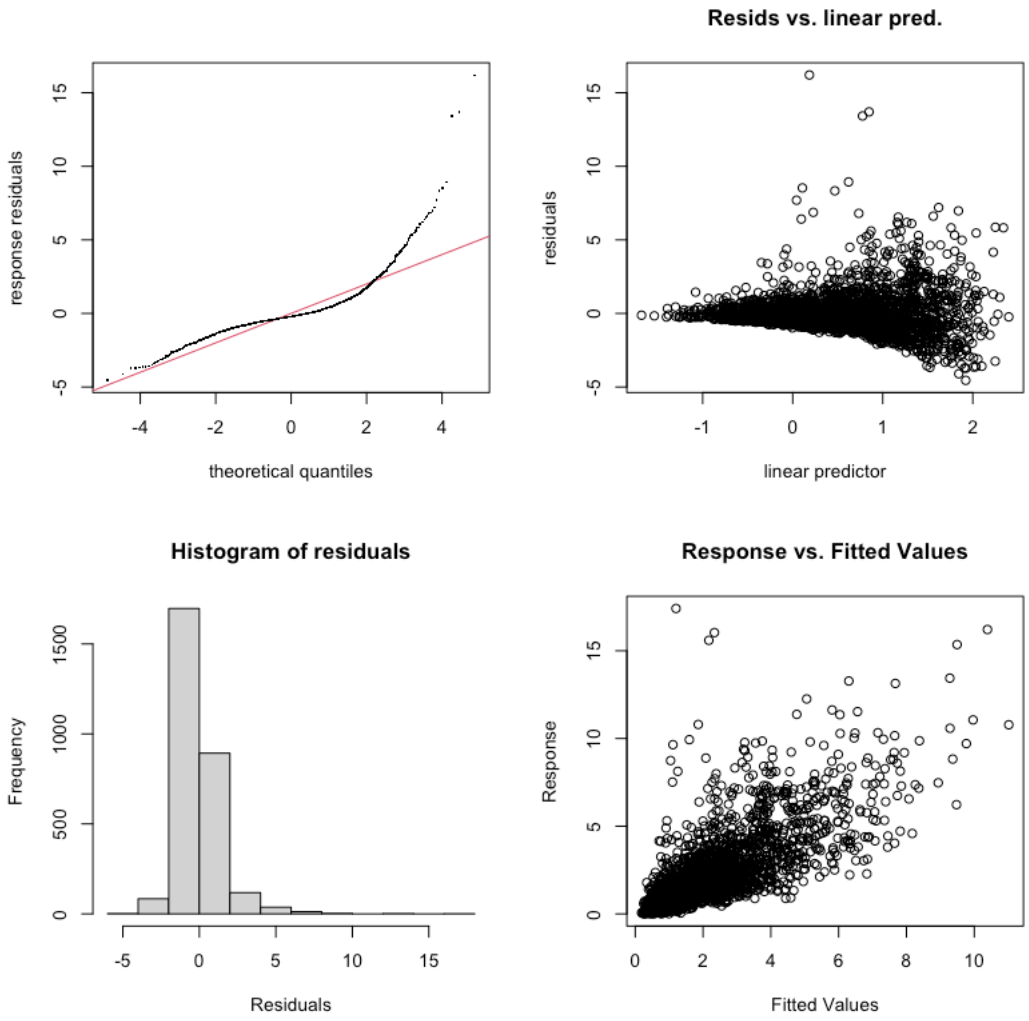

Figure 66. GAM evaluation plots for maximum daily 1-hour average SO_2 in the area of El Paso, TX.

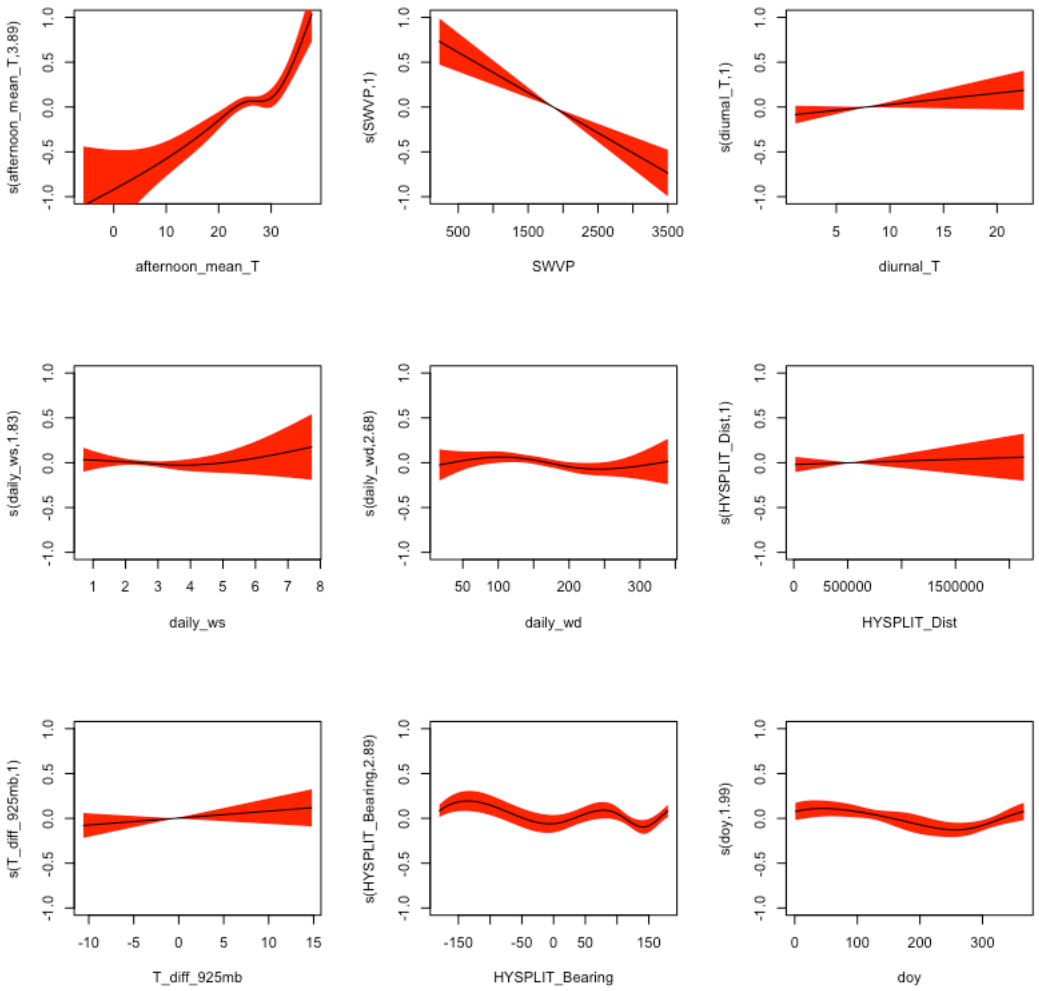

Figure 67. Smooth functions for the maximum daily 1-hour average SO_2 GAM fit in the area of Houston/Galveston/Brazoria, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the maximum daily 1-hour average SO_2 in ppbv from its mean value.

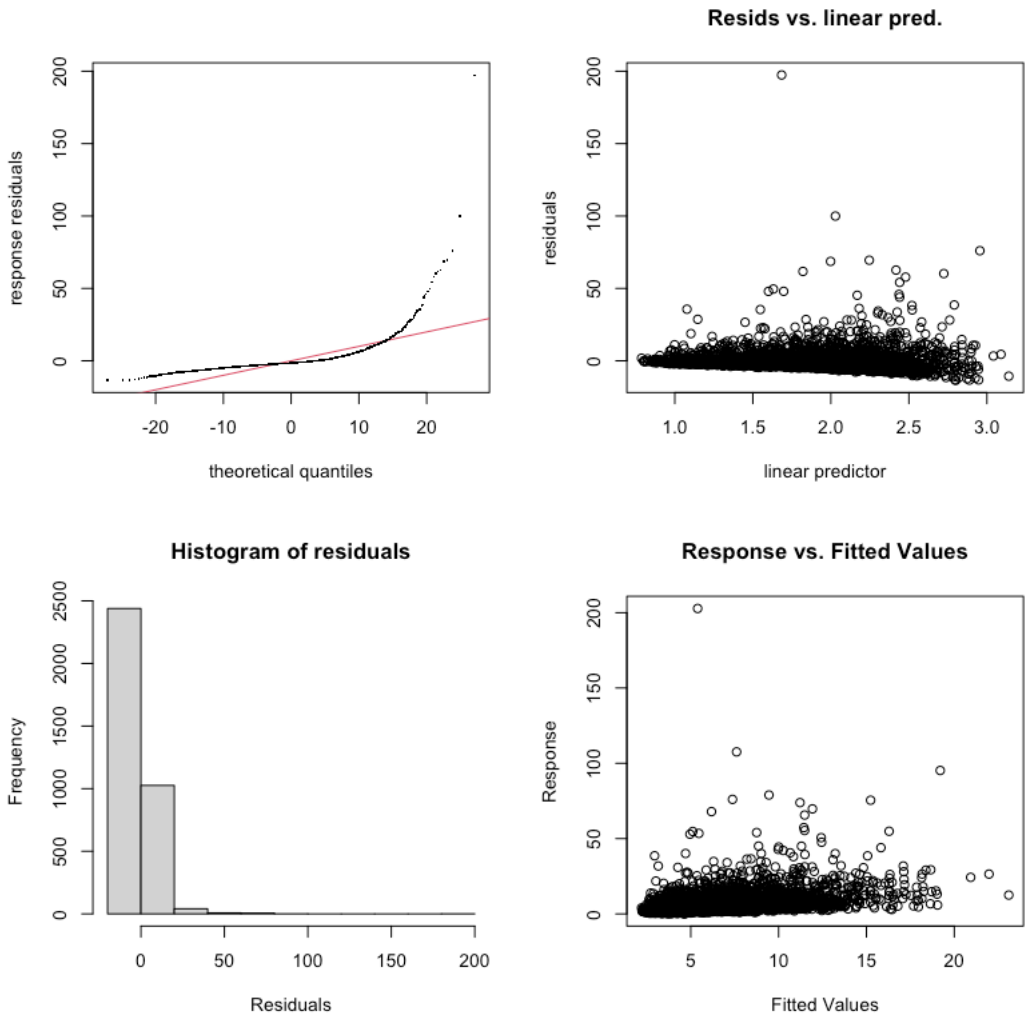

Figure 68. GAM evaluation plots for maximum daily 1-hour average SO_2 in the area of Houston/Galveston/Brazoria, TX.

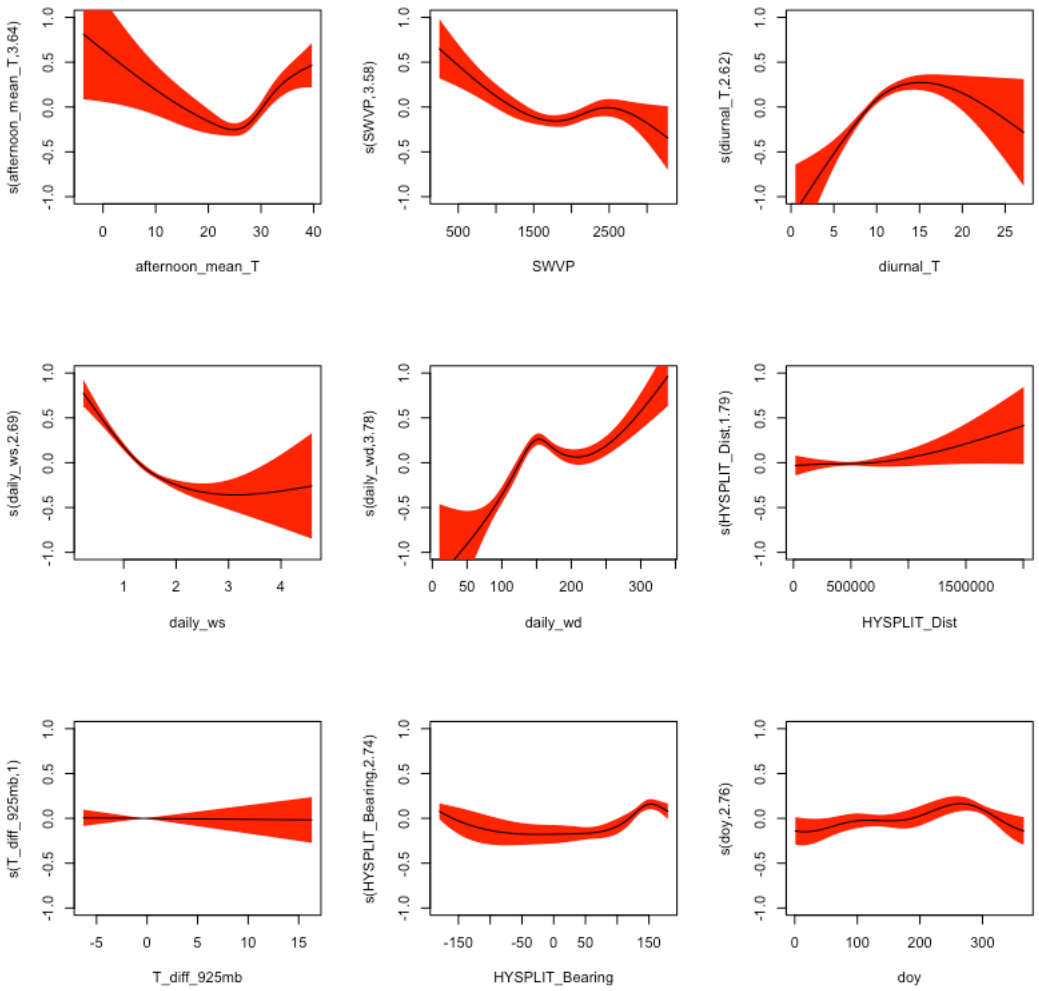

Figure 69. Smooth functions for the maximum daily 1-hour average SO_2 GAM fit in the area of San Antonio, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the maximum daily 1-hour average SO_2 in ppbv from its mean value.

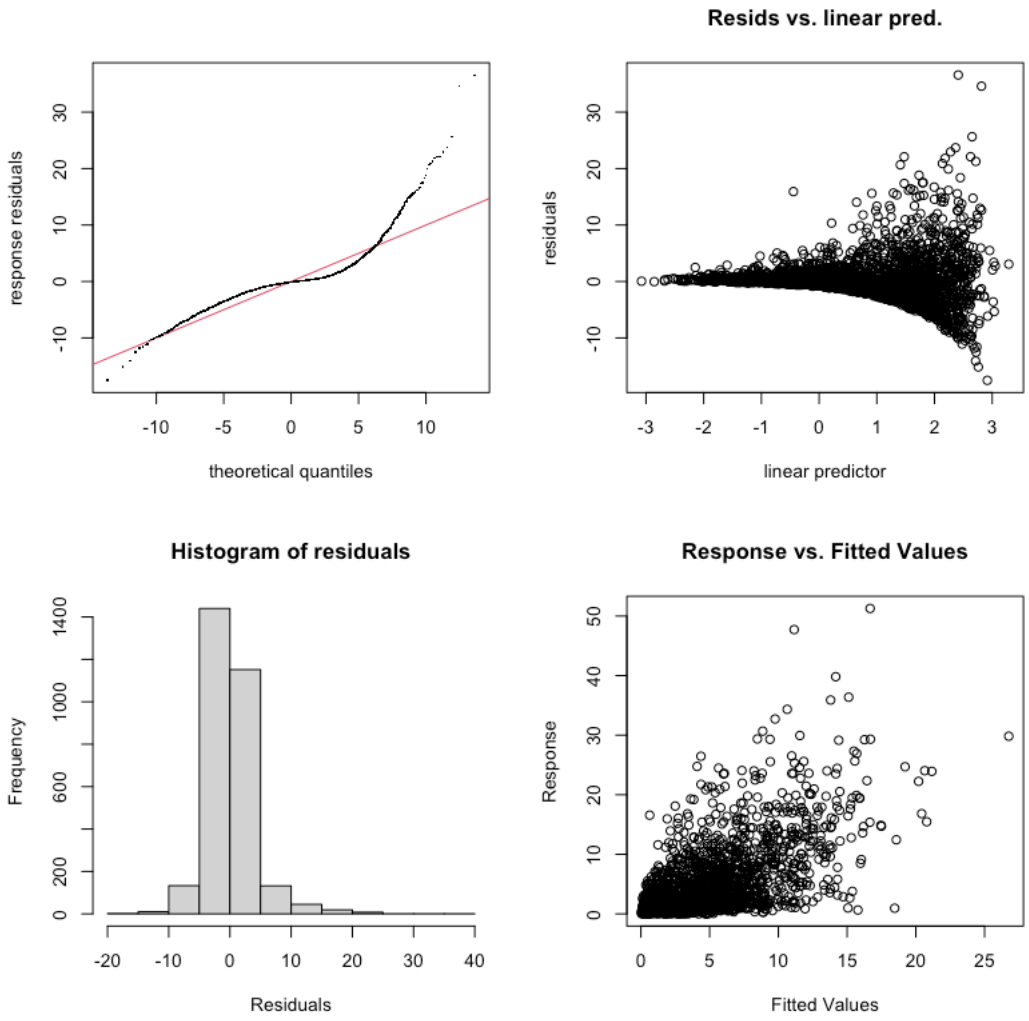

Figure 70. GAM evaluation plots for maximum daily 1-hour average SO_2 in the area of San Antonio, TX.

2.2.4.2 Background SO_2 GAM Results

Table 13 organizes the background SO_2 GAM results for each of the seven urban areas. Note that there was no background SO_2 data for Austin and Corpus Christi and so no GAM model results. The table includes figure numbers for the results, the percentage deviance of the background SO_2 values explained by the GAM model, and the level statistical significance of the meteorological predictors in table 7. Odd number figures are the smooth functions from the GAM fit of the natural logarithm of the background SO_2 values to the meteorological predictors. 95% confidence intervals are shown in red in the smooth functions. The day of year (*doy*) function is also shown. The percentage deviance values can be compared to the Camalier et al. (2007) results, which showed the predictive power of their models (measured by the R^2 statistic) to be between 0.56 and 0.80 for the cities in that study. Even number figures are the standard GAM evaluation plots (made with the *gam.check* function in R). The low-mid percentage deviance values and the

plots and for all seven urban areas indicate a poor-fair fit. However, the model residuals do not follow a normal distribution and show a trend versus predicted value. We do not think these fits are providing much accurate information on the variability of background SO_2 under different meteorological conditions.

Table 13. Performance of GAMs for background SO_2

Urban Area	Figure for Smooth Functions	Figure for GAM Evaluation	% Deviance Explained	Significance of Meteorological Predictors
Austin	N/A	N/A	N/A	N/A
Beaumont/Port Arthur	71	72	51.3	7 terms at α=0.001 1 term at α=1
Corpus Christi	N/A	N/A	N/A	N/A
Dallas/Fort Worth	73	74	62.7	8 terms at α=0.001
El Paso	75	76	44.8	7 terms at α=0.001 1 term at α=1
Houston/Galveston/Brazoria	77	78	39.3	7 terms at α=0.001 1 term at α=1
San Antonio	79	80	31.8	8 terms at α=0.001

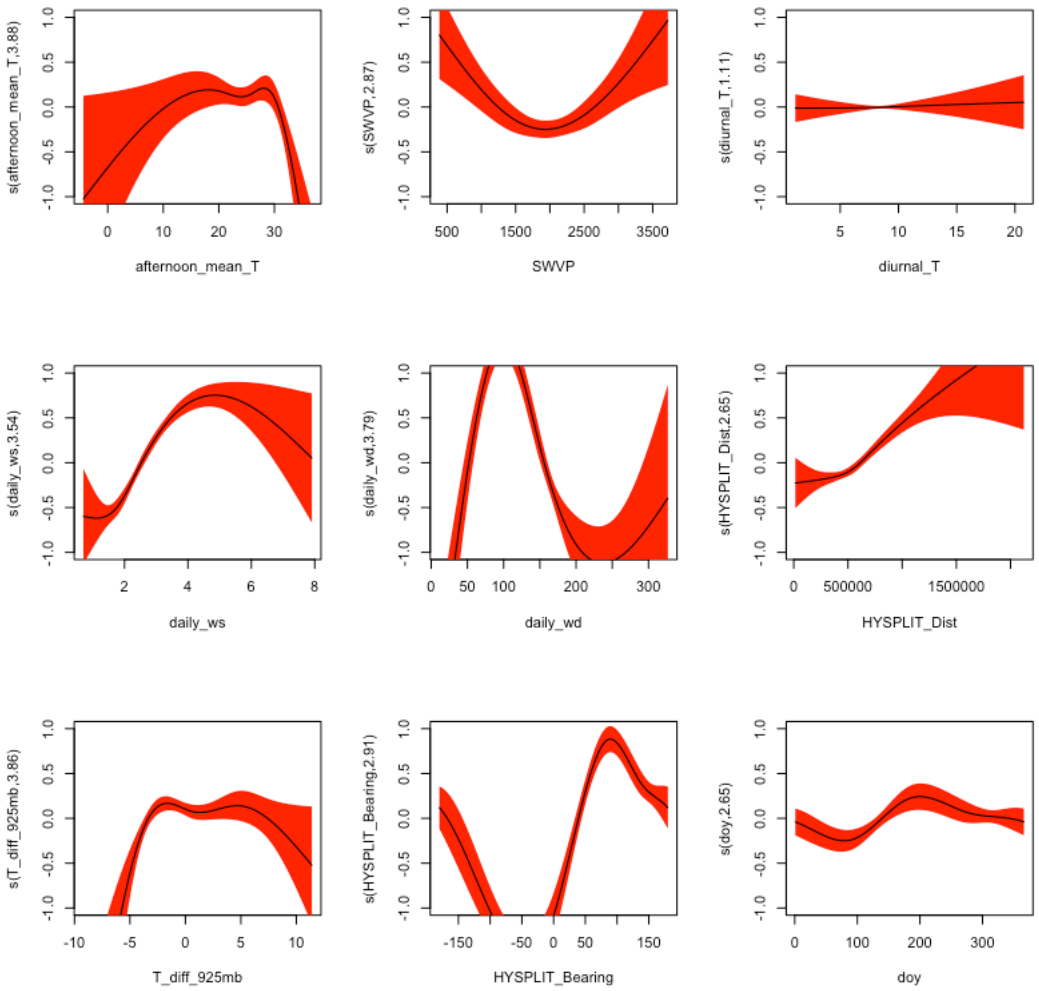

Figure 71. Smooth functions for the background SO_2 GAM fit in the area of Beaumont, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background SO_2 in ppbv from its mean value.

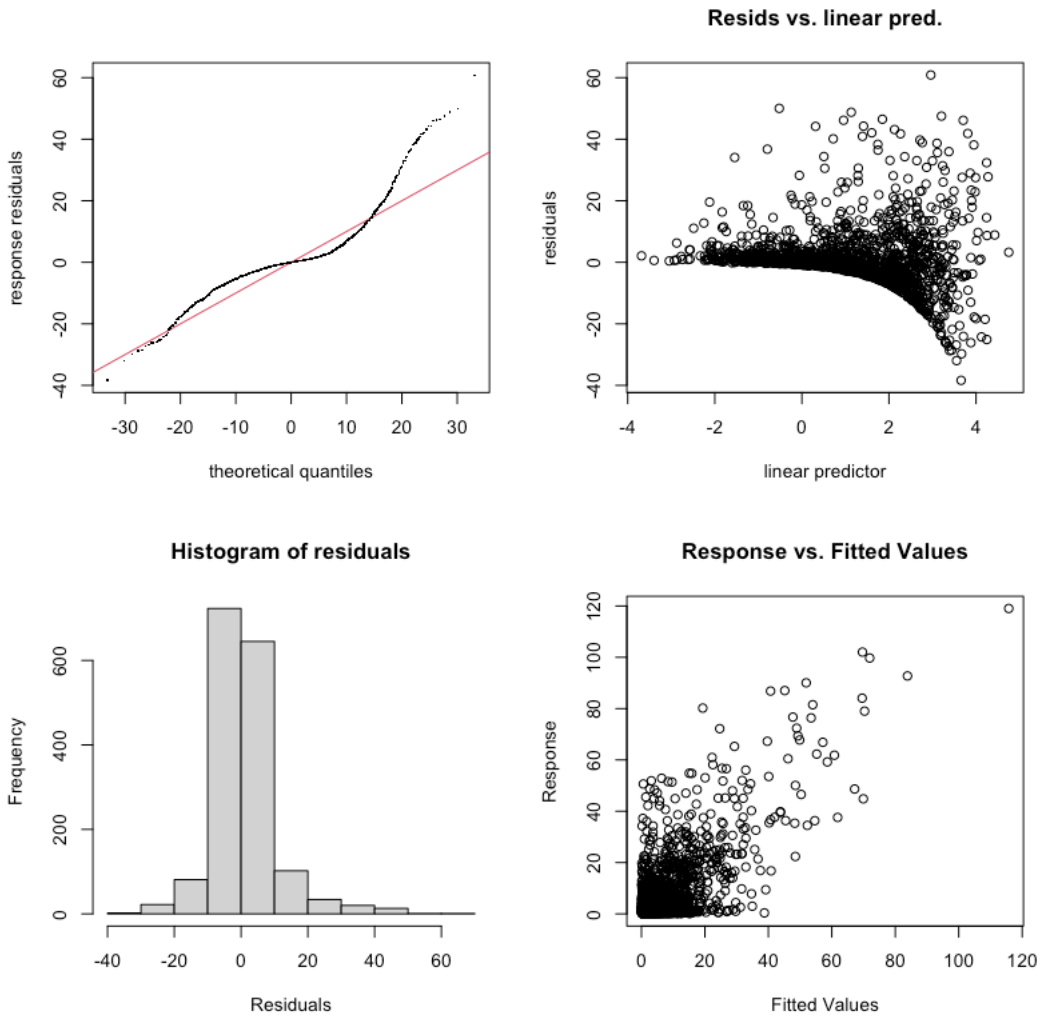

Figure 72. GAM evaluation plots for background SO_2 in the area of Beamont, TX.

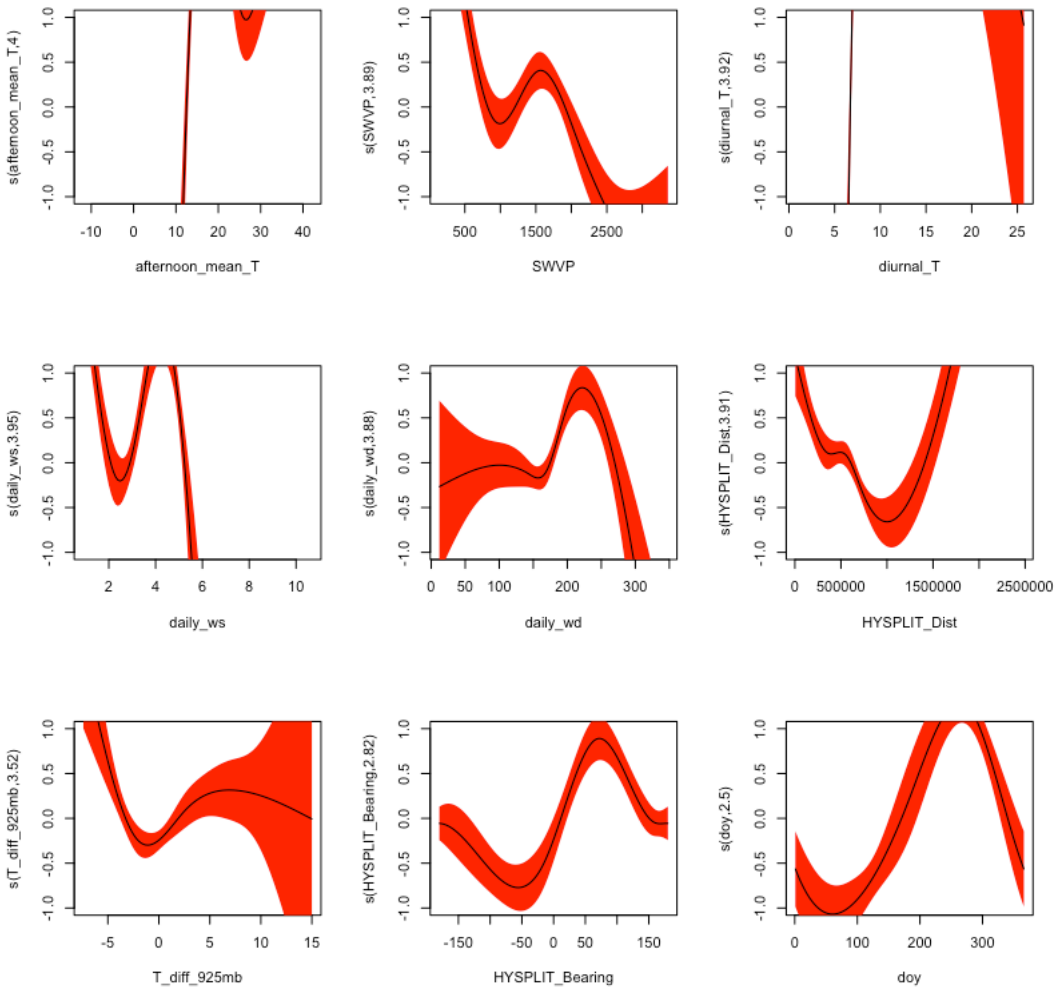

Figure 73. Smooth functions for the background SO_2 GAM fit in the area of Dallas/Fort Worth, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background SO_2 in ppbv from its mean value.

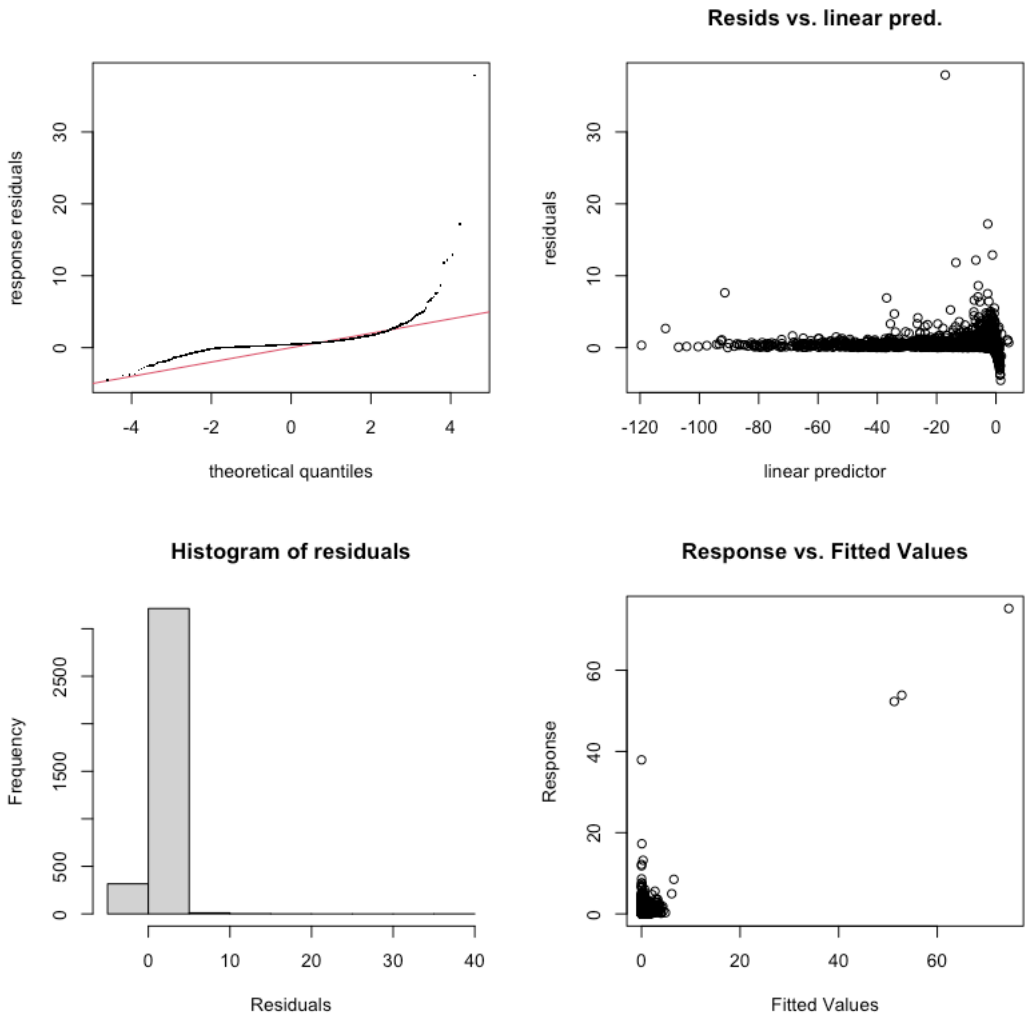

Figure 74. GAM evaluation plots for background SO_2 in the area of Dallas/Fort Worth, TX.

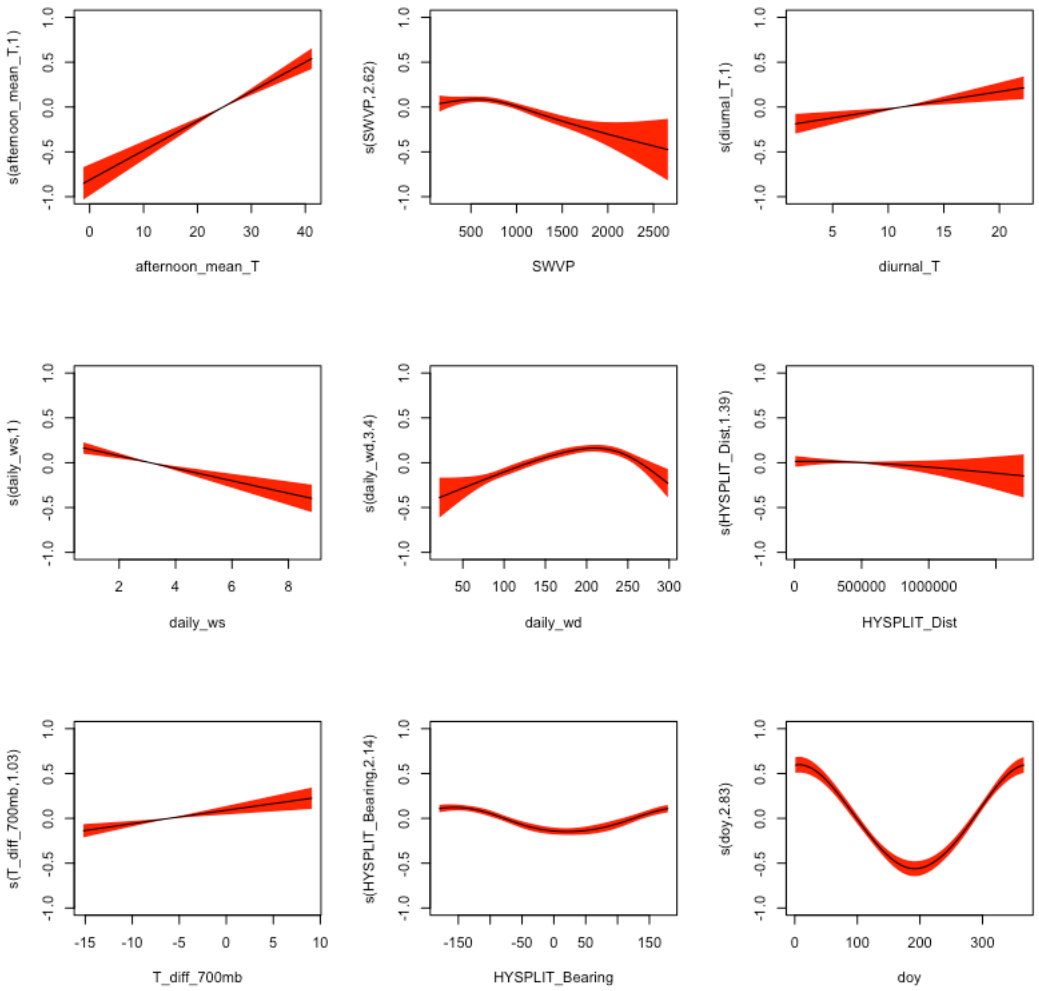

Figure 75. Smooth functions for the background SO_2 GAM fit in the area of El Paso, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background SO_2 in ppbv from its mean value.

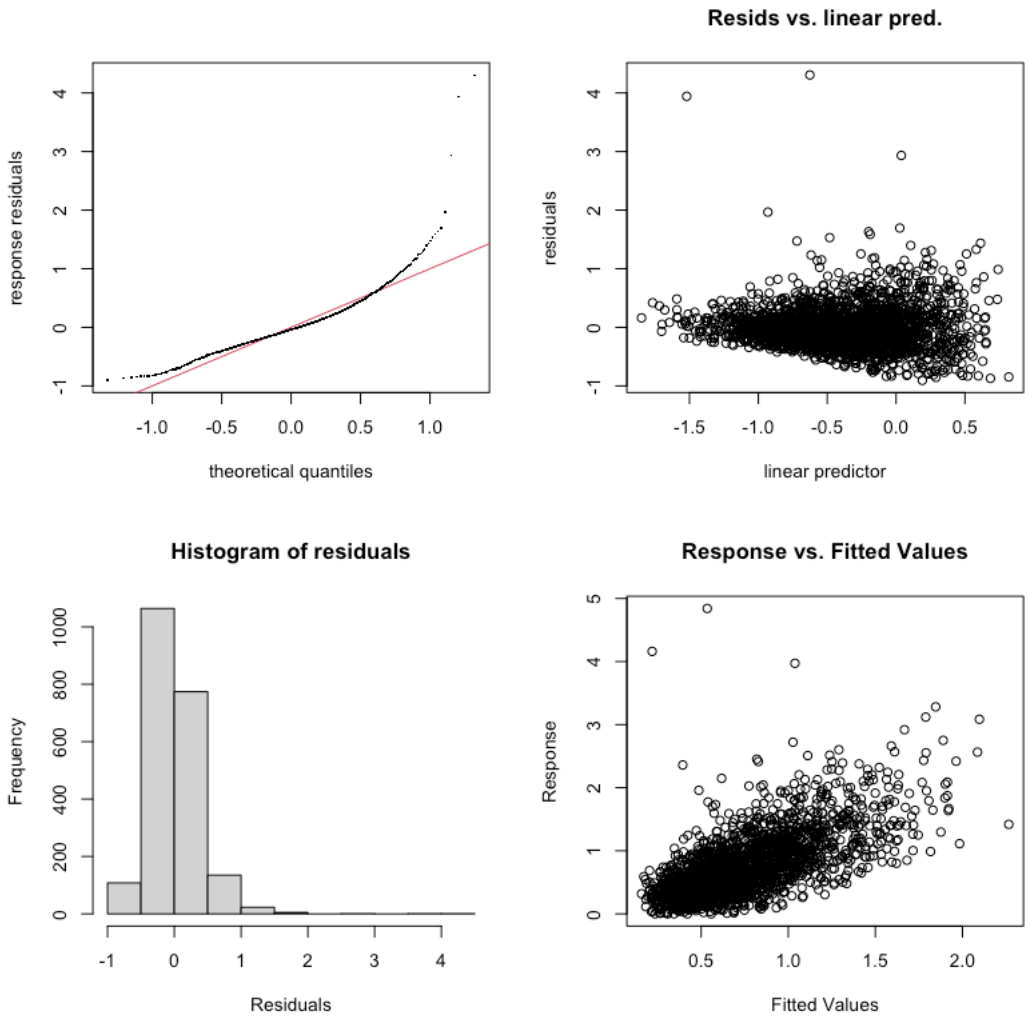

Figure 76. GAM evaluation plots for background SO_2 in the area of El Paso, TX.

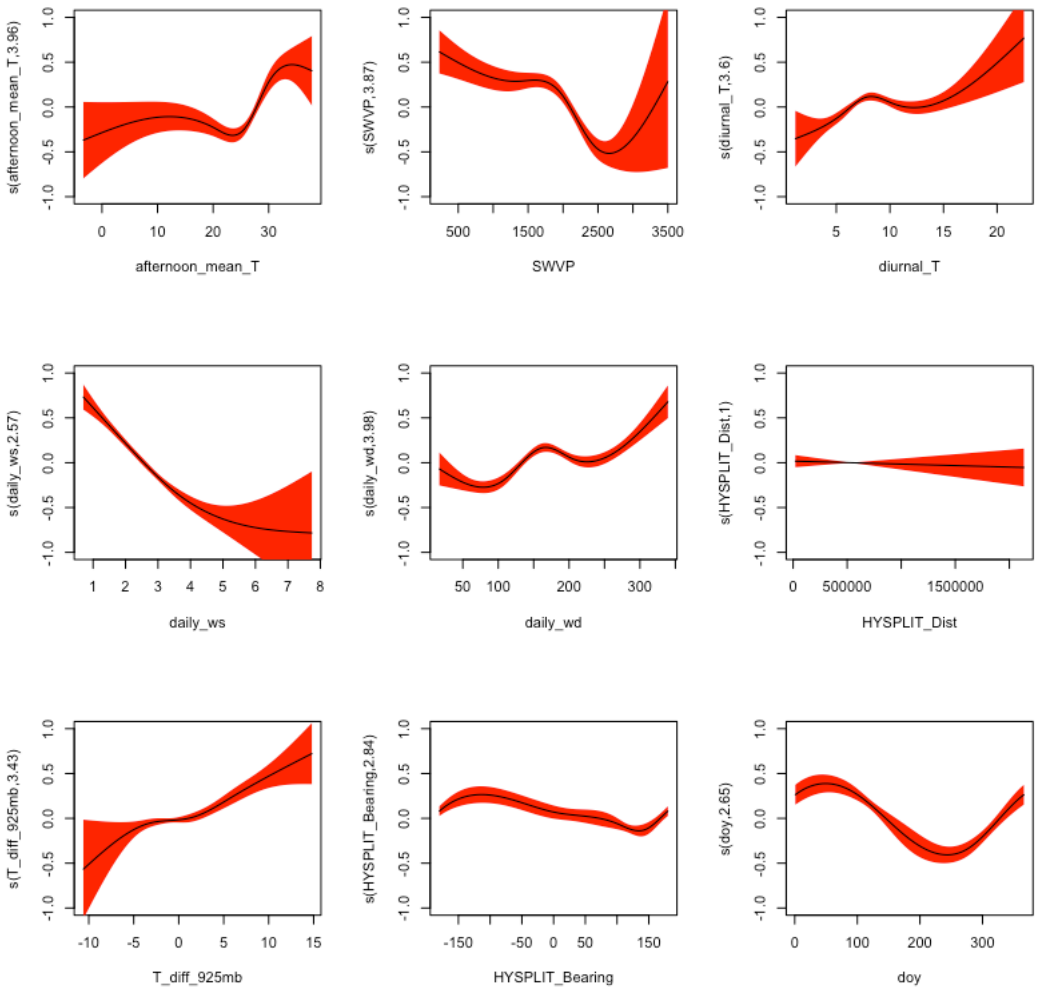

Figure 77. Smooth functions for the background SO_2 GAM fit in the area of Houston/Galveston/Brazoria, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background SO_2 in ppbv from its mean value.

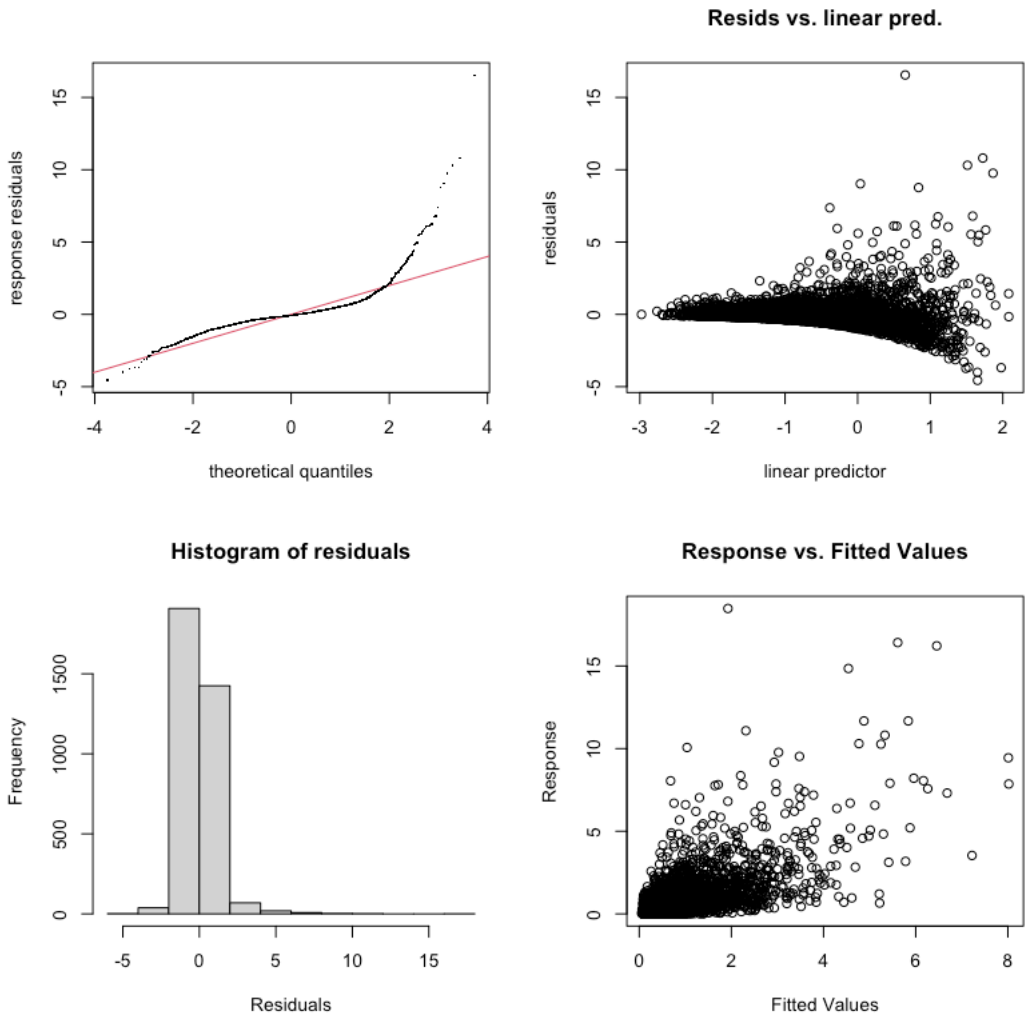

Figure 78. GAM evaluation plots for background SO_2 in the area of Houston/Galveston/Brazoria, TX.

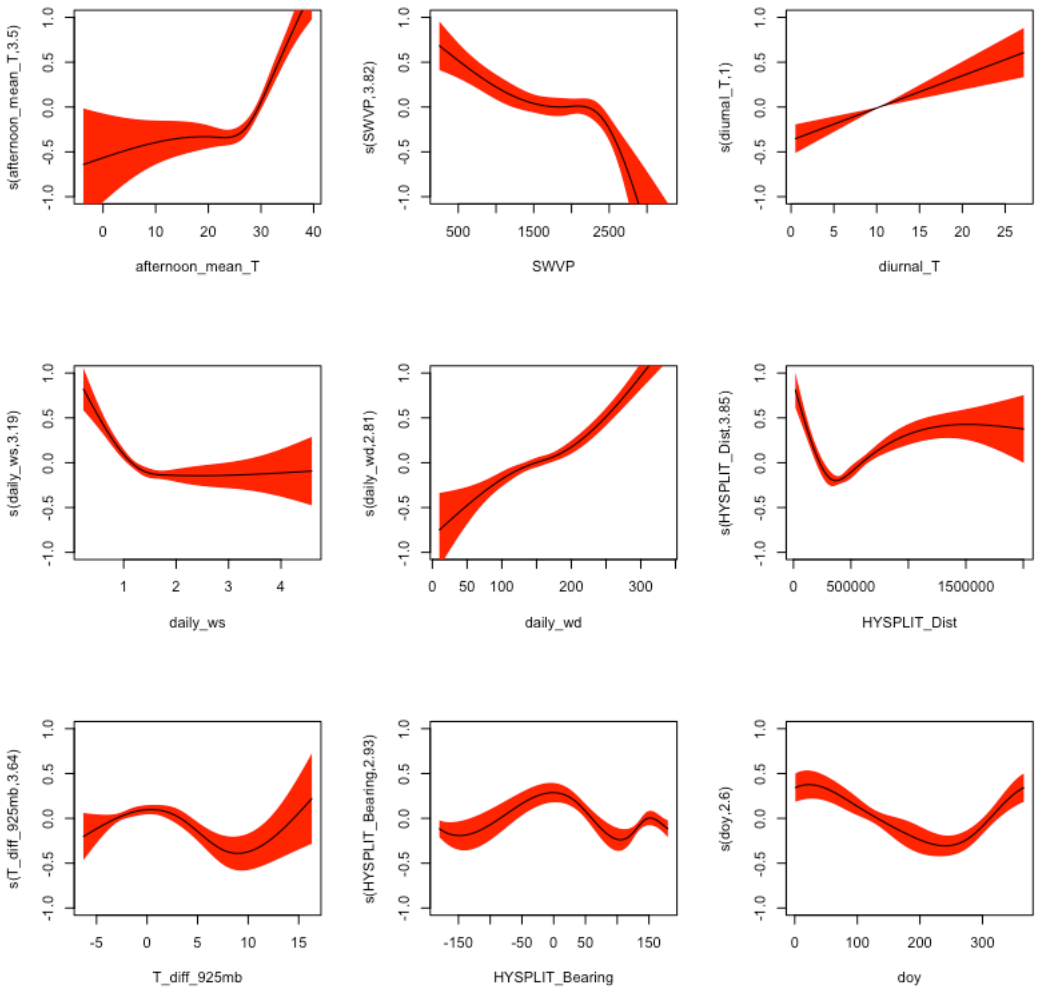

Figure 79. Smooth functions for the background SO_2 GAM fit in the area of San Antonio, TX. The y-axis scale is the scale of the "linear predictor", i.e., the deviation of the natural logarithm of the background SO_2 in ppbv from its mean value.

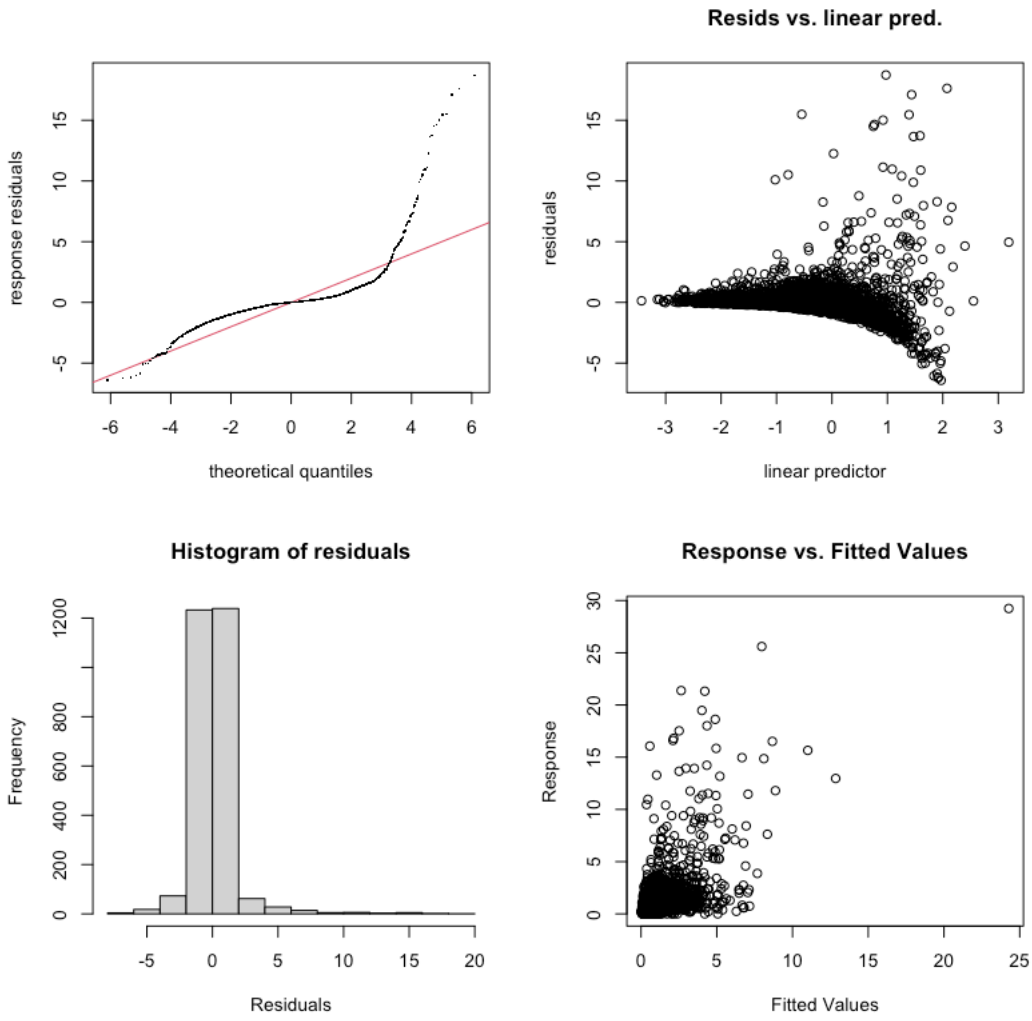

Figure 80. GAM evaluation plots for background SO_2 in the area of San Antonio, TX.

2.2.5 Cross Validation Analysis

In order to test for over-fitting in our GAMs, as well as to test the robustness of our results for the functional relationships between the meteorological predictors and pollutants, we performed a 10-fold cross-validation for each GAM. We used the "CVgam" function in the "gamclass" R package. Table 14 shows the results of the cross validation for each GAM. The "GAMscale" is the mean of the squares of the errors of the original GAM fits. The "CV-mse-GAM" is the mean of the squares of the errors calculated for the 10% of data not included in the fit for each of the 10 cross-validation GAMs. While it is expected that the CV-mse-GAM will be larger than the GAMscale, a large difference between these values would suggest that the GAMs are over-fitting the data, as the performance is much poorer on the data not included in the fit during the cross-validation.

Table 14. Cross validation analysis

Urban Area	GAM	GAMscale	CV-mse-GAM	% Change
Austin	MDA8 O_3	44.01311	45.33445	3.0
	background O_3	53.27466	55.17987	3.6
	average daily $PM_{2.5}$	15.98153	16.19972	1.4
	background $PM_{2.5}$	10.61685	10.8224	1.9
	max daily SO_2	0.4850608	0.4961835	2.3
	background SO_2	N/A	N/A	N/A
Beaumont/ Port Arthur	MDA8 O_3	71.56887	74.37709	3.9
	background O_3	60.79124	62.92314	3.5
	average daily $PM_{2.5}$	16.39413	16.66939	1.7
	background $PM_{2.5}$	10.37982	10.50091	1.2
	max daily SO_2	220.5402	224.0471	1.6
	background SO_2	136.1601	140.607	3.3
Corpus Christi	MDA8 O_3	59.7399	61.29711	2.6
	background O_3	60.61019	62.68148	3.4
	average daily $PM_{2.5}$	23.70982	24.14094	1.8
	background $PM_{2.5}$	23.59579	23.87506	1.2
	max daily SO_2	33.48989	33.95746	1.4
	background SO_2	N/A	N/A	N/A
Dallas/ Fort Worth	MDA8 O_3	57.71213	59.17019	2.5
	background O_3	43.95739	45.05525	2.5
	average daily $PM_{2.5}$	17.61873	17.84149	1.3
	background $PM_{2.5}$	7.95054	8.048999	1.2
	max daily SO_2	313.1894	318.985	1.9
	background SO_2	4.081353	4.145063	1.6
El Paso	MDA8 O_3	46.15778	47.24758	2.4
	background O_3	38.83021	39.78439	2.5
	average daily $PM_{2.5}$	36.74953	37.79743	2.9
	background $PM_{2.5}$	7.888733	8.032832	1.8
	max daily SO_2	2.091772	2.133603	2.0
	background SO_2	0.1513528	0.1563242	3.3
Houston/ Galveston/ Brazoria	MDA8 O_3	81.26525	83.26797	2.5
	background O_3	57.24772	58.69582	2.5
	average daily $PM_{2.5}$	26.2529	26.56265	1.2
	background $PM_{2.5}$	8.988189	9.094864	1.2

	max daily SO_2	56.08964	56.96203	1.6
	background SO_2	1.227698	1.260116	2.6
San Antonio	MDA8 O_3	51.19005	52.63645	2.8
	background O_3	41.59651	42.76332	2.8
	average daily $PM_{2.5}$	29.58574	30.11392	1.8
	background $PM_{2.5}$	14.84524	15.07341	1.5
	max daily SO_2	17.54138	17.76601	1.3
	background SO_2	3.451773	3.578984	3.7

2.3 Meteorologically Adjusted Trends of O_3, $PM_{2.5}$, and SO_2

We used the results from the GAMs to determine the meteorologically adjusted trends in total and background MDA8 O_3, total and background $PM_{2.5}$, and total and background SO_2 for Dallas-Fort Worth, Houston-Galveston-Brazoria, San Antonio, Austin, Beaumont-Port Arthur, El Paso, and Corpus Christi (except those without SO_2 background observations in Austin and Corpus Christi). In this procedure, we use the Y_k terms from the GAM equation to determine the relative difference between the annual averages after meteorology has been taken into account. Our equation for the annual averages is thus

$$g(\mu_K) = \beta_0 + Y_K + c_0$$

where k is the kth year's average and c_o is a constant. The constant c_o is needed because of how R treats factor variables. In order to have an identifiable model, one of the factor levels, in this case the year 2012, must be set to have a value of $Y_k = 0$. However, when the year 2012 is the year with the largest annual average values in the original data set, it results in Y_k values that are predominantly less than 0, leading to meteorologically adjusted annual averages that do not have the same 10-year average as the original data set. To avoid this issue, we add a constant c_o to the meteorologically adjusted annual averages so that the 10-year averages in the original and meteorologically adjusted trend data are identical. The value of the meteorologically adjusted linear trends over 2012-2021 is relatively insensitive to the value of c_o.

2.3.1 Meteorologically Adjusted Trends for Total and Background MDA8 O_3

The total and background MDA8 O_3 original trends and meteorologically adjusted linear trends are provided in Tables 15 and 16. The trend estimates are determined by ordinary least squares (OLS) linear regression of the annual averages. The original and meteorologically adjusted annual averages for total and background MDA8 O_3 trends are shown in Figure 81-87. No statistically significant trends with time are observed for 2012-2021 either before or after meteorological adjustment.

Table 15. Total MDA8 O_3 original trends and meteorologically adjusted linear trends.

Urban Area	Original	Met Adjusted
Austin	-0.170 ± 0.242 ppb/yr	-0.260 ± 0.455 ppb/yr
Beaumont/Port Arthur	-0.490 ± 0.340 ppb/yr	-0.170 ± 0.323 ppb/yr

Corpus Christi	-0.410 ± 0.287 ppb/yr	0.030 ± 0.260 ppb/yr
Dallas/Fort Worth	-0.470 ± 0.306 ppb/yr	-0.060 ± 0.200 ppb/yr
El Paso	0.350 ± 0.291 ppb/yr	0.150 ± 0.227 ppb/yr
Houston/Galveston/Brazoria	-0.220 ± 0.233 ppb/yr	-0.080 ± 0.272 ppb/yr
San Antonio	-0.470 ± 0.406 ppb/yr	-0.390 ± 0.310 ppb/yr

Table 16. Background MDA8 O_3 original trends and meteorologically adjusted linear trends.

Urban Area	Original	Met Adjusted
Austin	-1.360 ± 0.655 ppb/yr	-1.200 ± 0.521 ppb/yr
Beaumont/Port Arthur	-0.300 ± 0.412 ppb/yr	0.040 ± 0.401 ppb/yr
Corpus Christi	0.490 ± 0.550 ppb/yr	0.870 ± 0.420 ppb/yr
Dallas/Fort Worth	-0.550 ± 0.343 ppb/yr	-0.220 ± 0.242 ppb/yr
El Paso	0.710 ± 0.386 ppb/yr	0.570 ± 0.336 ppb/yr
Houston/Galveston/Brazoria	-0.090 ± 0.267 ppb/yr	-0.010 ± 0.327 ppb/yr
San Antonio	-0.590 ± 0.311 ppb/yr	-0.590 ± 0.269 ppb/yr

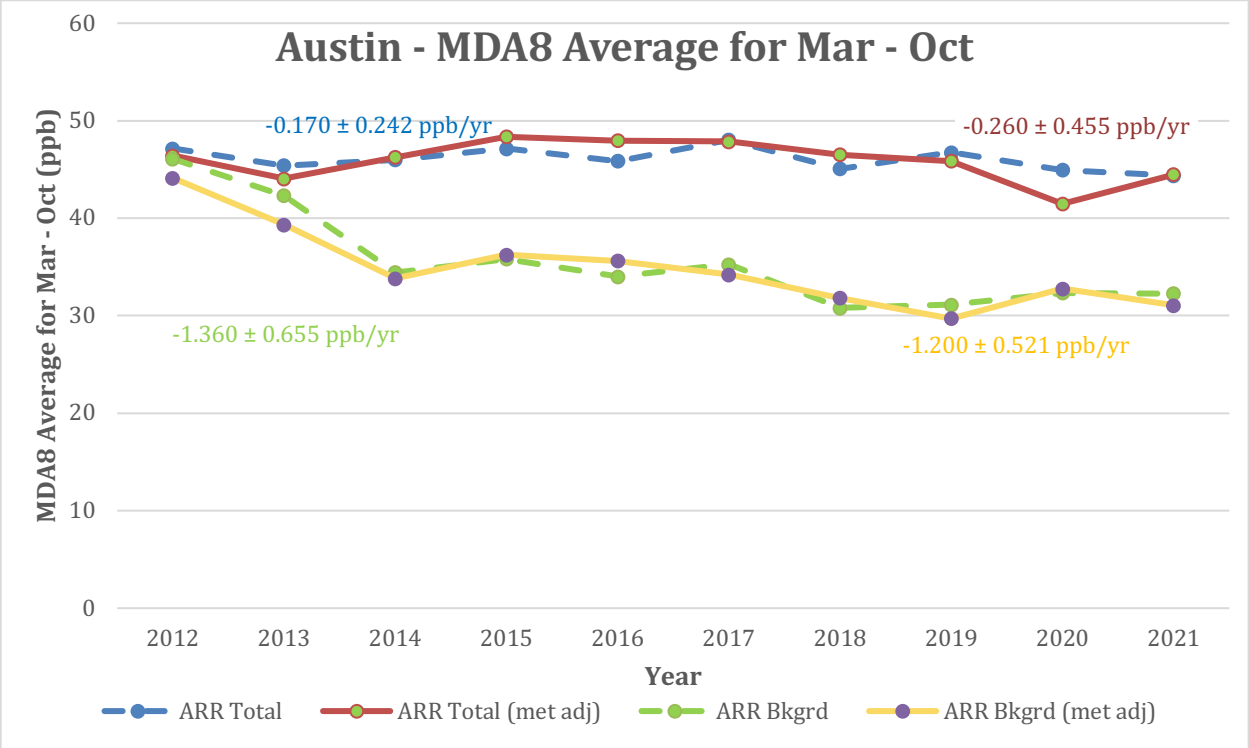

Figure 81. Original and meteorologically adjusted annual averages for total and background MDA8 O_3 trends in Austin, TX.

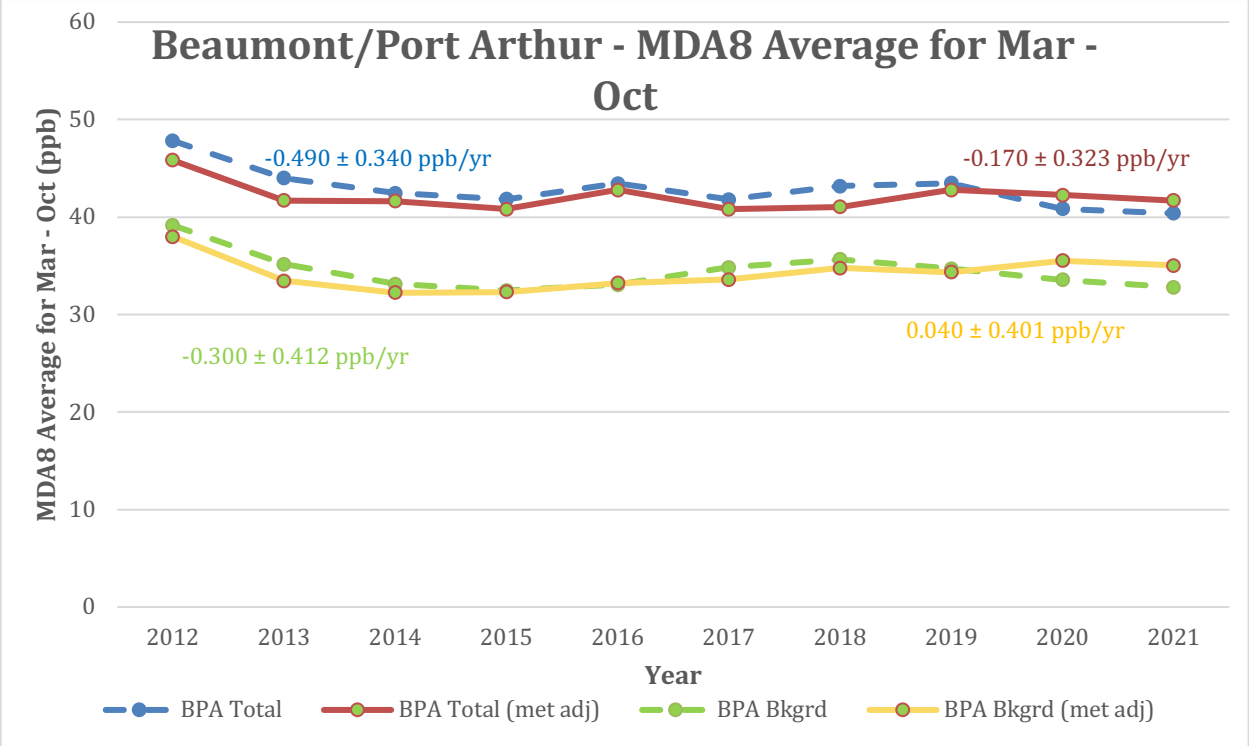

Figure 82. Original and meteorologically adjusted annual averages for total and background MDA8 O_3 trends in Beaumont/Port Arthur, TX.

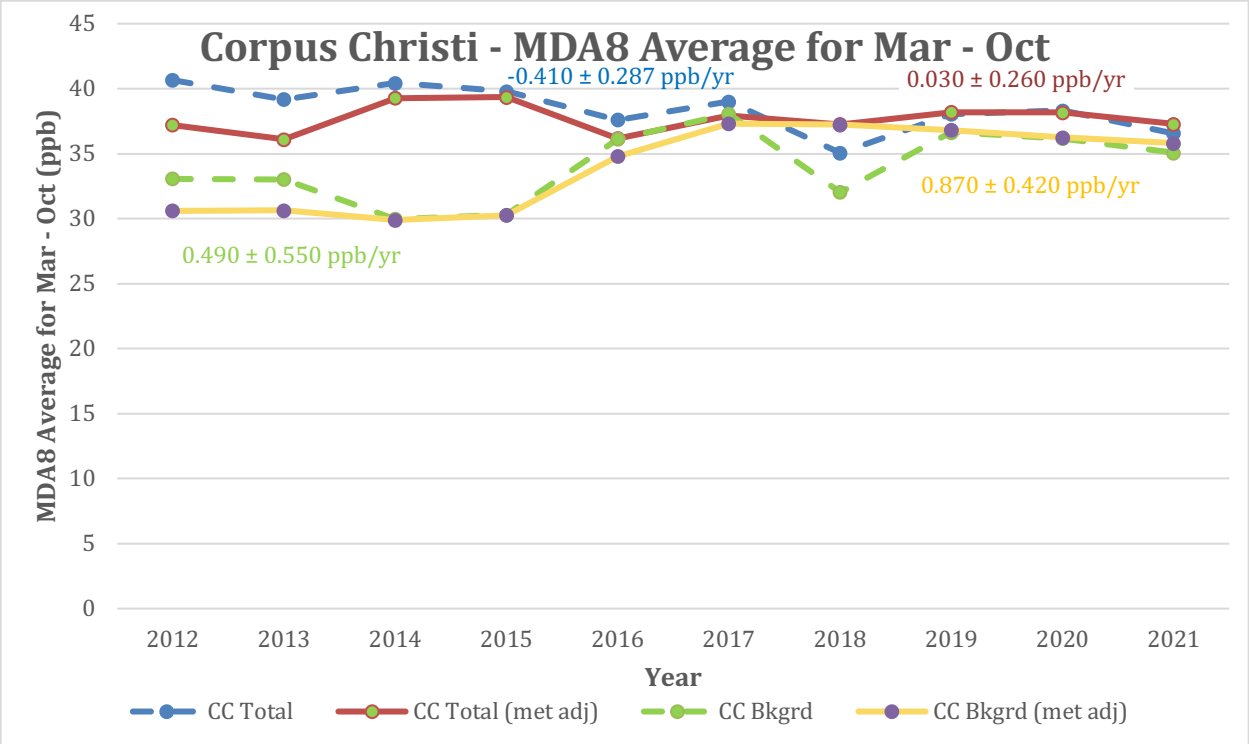

Figure 83. Original and meteorologically adjusted annual averages for total and background MDA8 O_3 trends in Corpus Christi, TX.

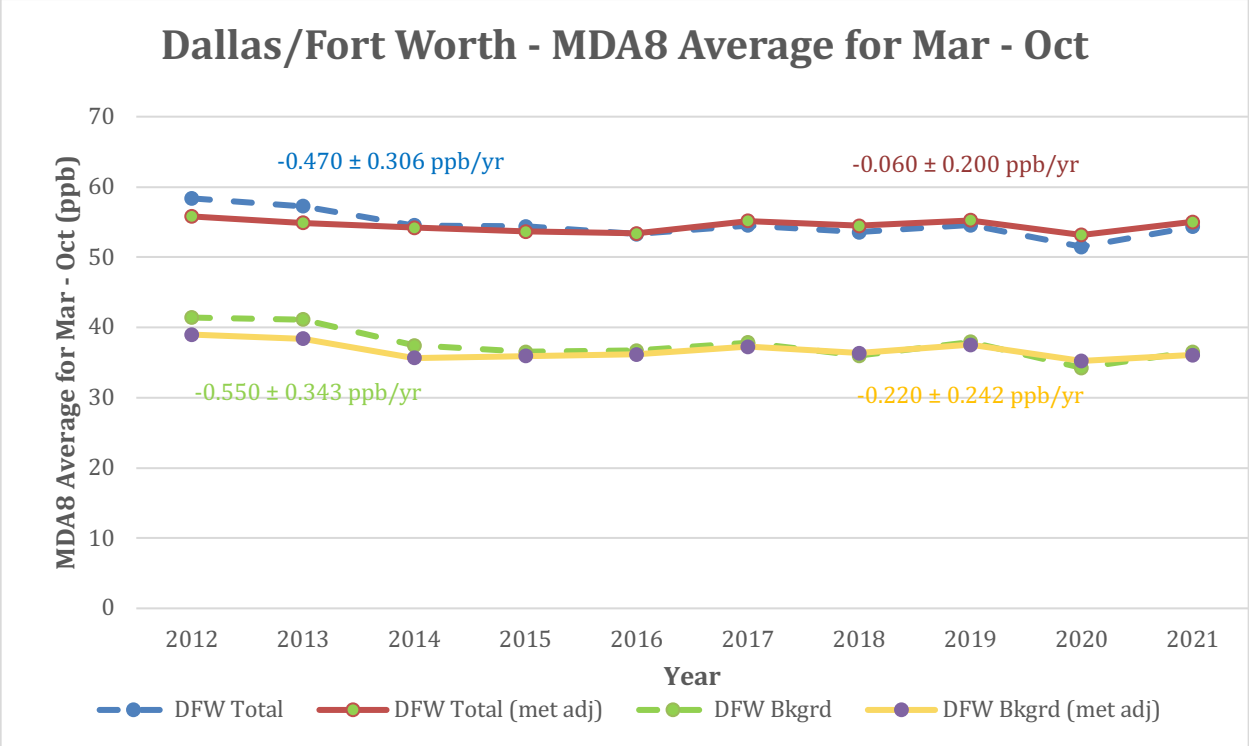

Figure 84. Original and meteorologically adjusted annual averages for total and background MDA8 O_3 trends in Dallas/Fort Worth, TX.

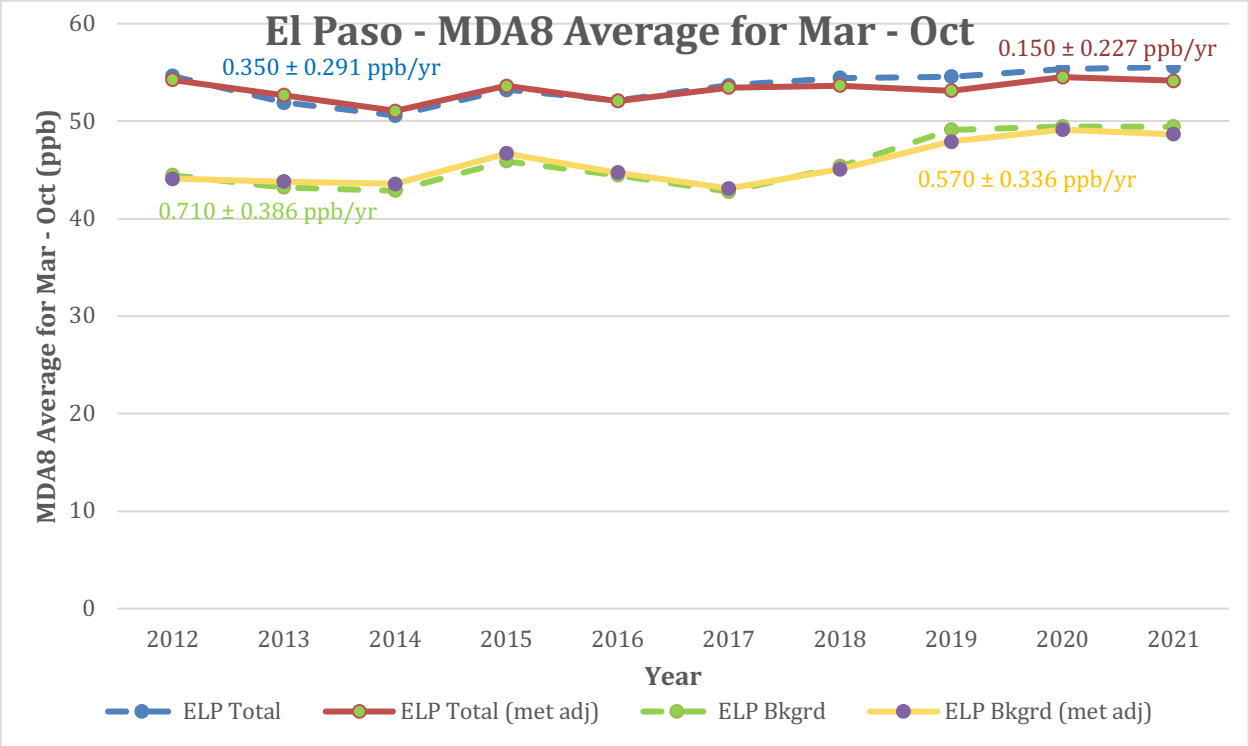

Figure 85. Original and meteorologically adjusted annual averages for total and background MDA8 O_3 trends in El Paso, TX.

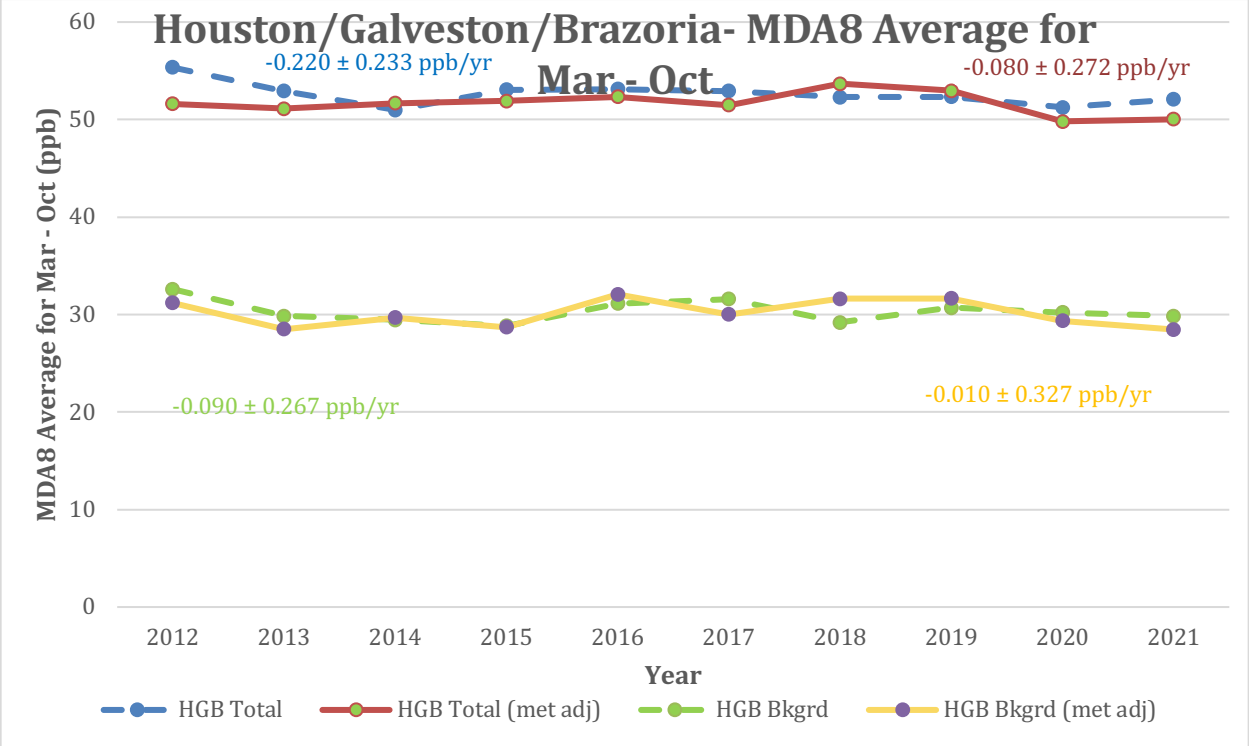

Figure 86. Original and meteorologically adjusted annual averages for total and background MDA8 O_3 trends in Houston/Galveston/Brazoria, TX.

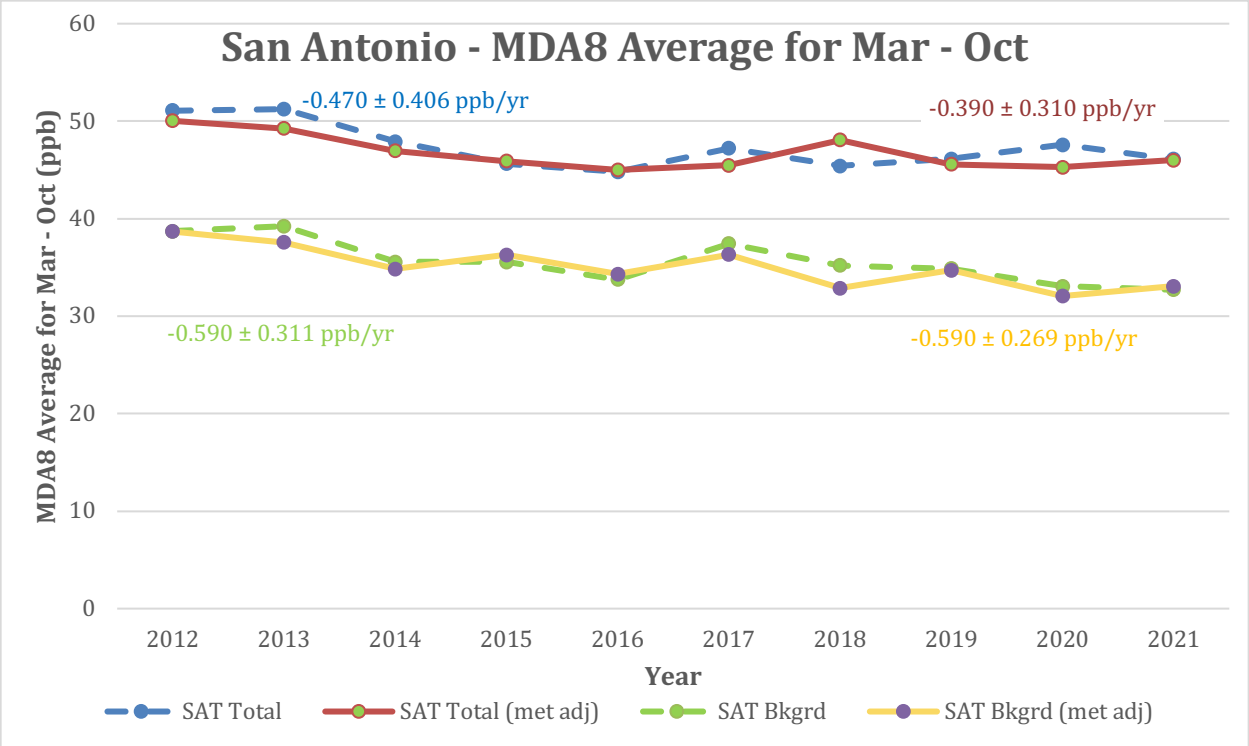

Figure 87. Original and meteorologically adjusted annual averages for total and background MDA8 O_3 trends in San Antonio, TX.

2.3.2 Meteorologically Adjusted Trends for Total and Background $PM_{2.5}$

The total and background $PM_{2.5}$ original trends and meteorologically adjusted linear trends are provided in Tables 17 and 18. The trend estimates are determined by ordinary least squares (OLS) linear regression of the annual averages. The original and meteorologically adjusted annual averages for total and background $PM_{2.5}$ trends are shown in Figure 88-94. No statistically significant trends with time are observed for 2012-2021 either before or after meteorological adjustment.

Table 17. Total $PM_{2.5}$ original trends and meteorologically adjusted linear trends.

Urban Area	Original	Met Adjusted
Austin	0.040 ± 0.077 µg/m3/yr	0.010 ± 0.103 µg/m3/yr
Beaumont/Port Arthur	-0.030 ± 0.165 µg/m3/yr	-0.060 ± 0.159 µg/m3/yr
Corpus Christi	0.180 ± 0.165 µg/m3/yr	0.020 ± 0.151 µg/m3/yr
Dallas/Fort Worth	-0.190 ± 0.205 µg/m3/yr	-0.140 ± 0.189 µg/m3/yr
El Paso	-0.090 ± 0.229 µg/m3/yr	-0.160 ± 0.206 µg/m3/yr
Houston/Galveston/Brazoria	-0.050 ± 0.127 µg/m3/yr	-0.050 ± 0.153 µg/m3/yr
San Antonio	0.130 ± 0.215 µg/m3/yr	0.180 ± 0.192 µg/m3/yr

Table 18. Background $PM_{2.5}$ original trends and meteorologically adjusted linear trends.

Urban Area	Original	Met Adjusted
Austin	N/A	N/A
Beaumont/Port Arthur	0.010 ± 0.153 µg/m3/yr	0.000 ± 0.144 µg/m3/yr
Corpus Christi	0.190 ± 0.208 µg/m3/yr	0.030 ± 0.172 µg/m3/yr
Dallas/Fort Worth	-0.170 ± 0.065 µg/m3/yr	-0.130 ± 0.075 µg/m3/yr
El Paso	0.050 ± 0.133 µg/m3/yr	0.050 ± 0.148 µg/m3/yr
Houston/Galveston/Brazoria	-0.150 ± 0.272 µg/m3/yr	-0.190 ± 0.265 µg/m3/yr
San Antonio	-0.170 ± 0.111 µg/m3/yr	-0.160 ± 0.107 µg/m3/yr

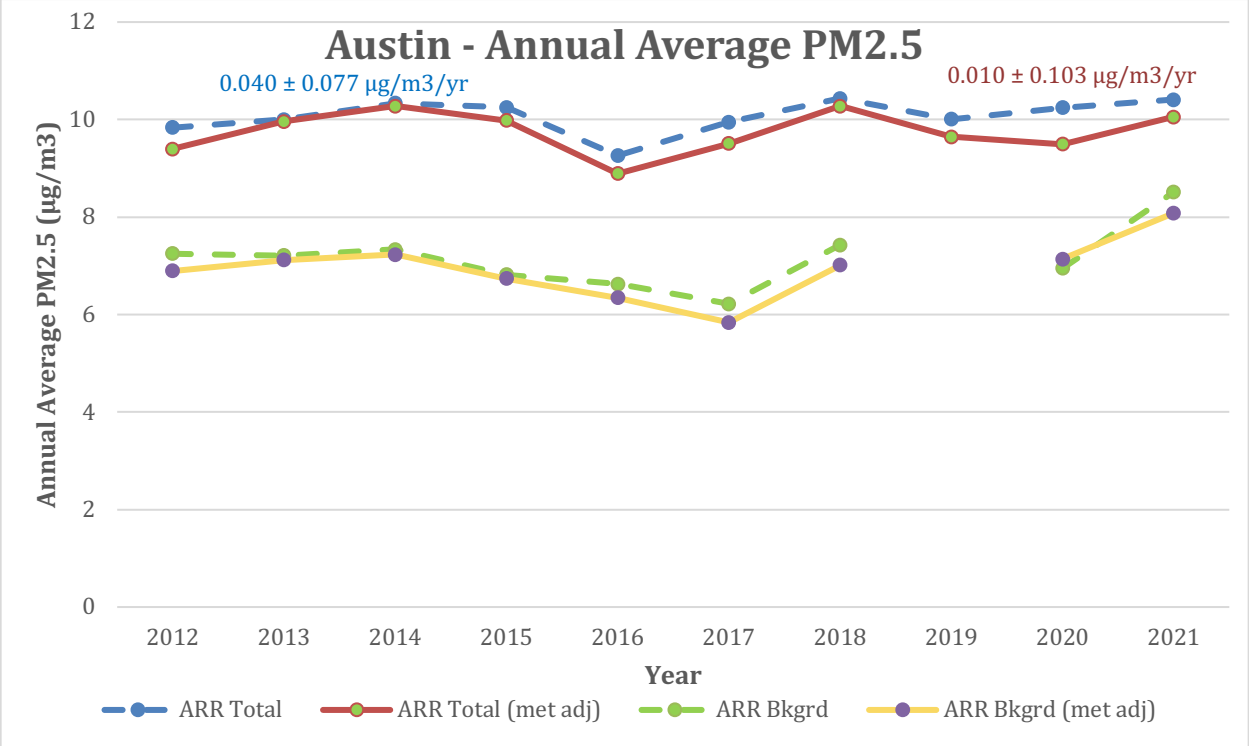

Figure 88. Original and meteorologically adjusted annual averages for total and background $PM_{2.5}$ trends in Austin, TX.

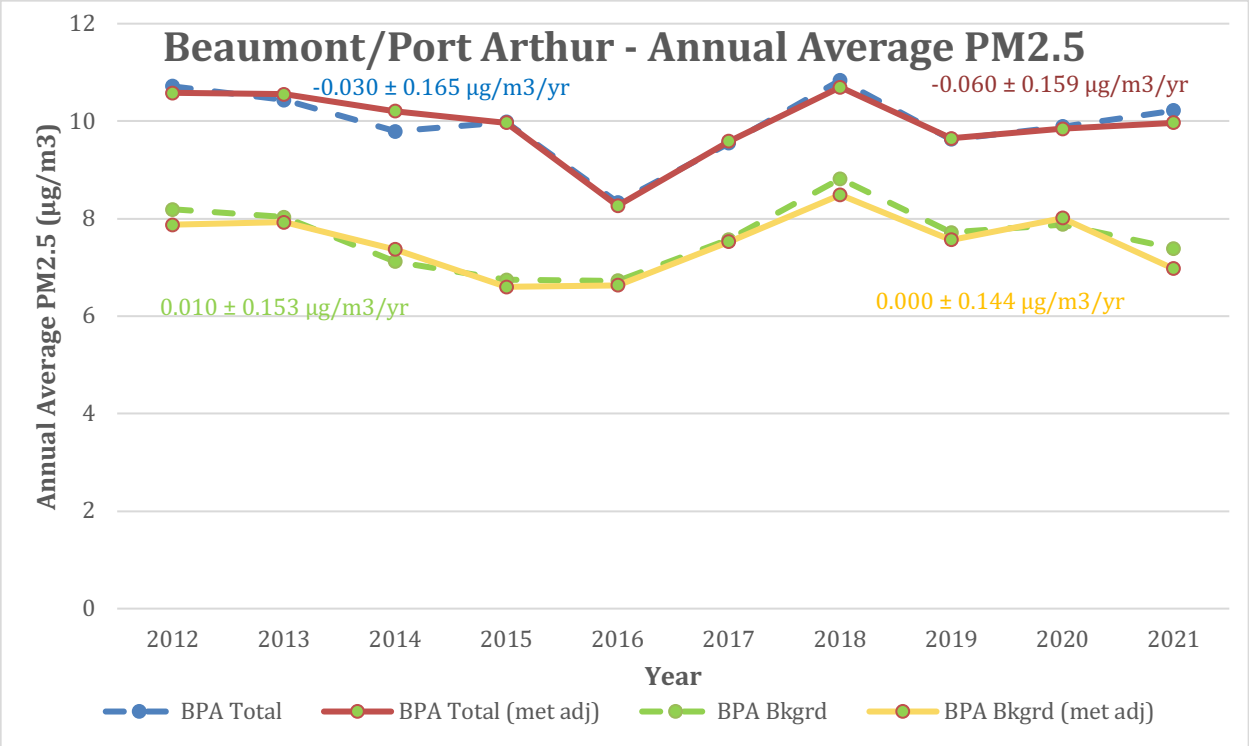

Figure 89. Original and meteorologically adjusted annual averages for total and background $PM_{2.5}$ trends in Beaumont/Port Arthur, TX.

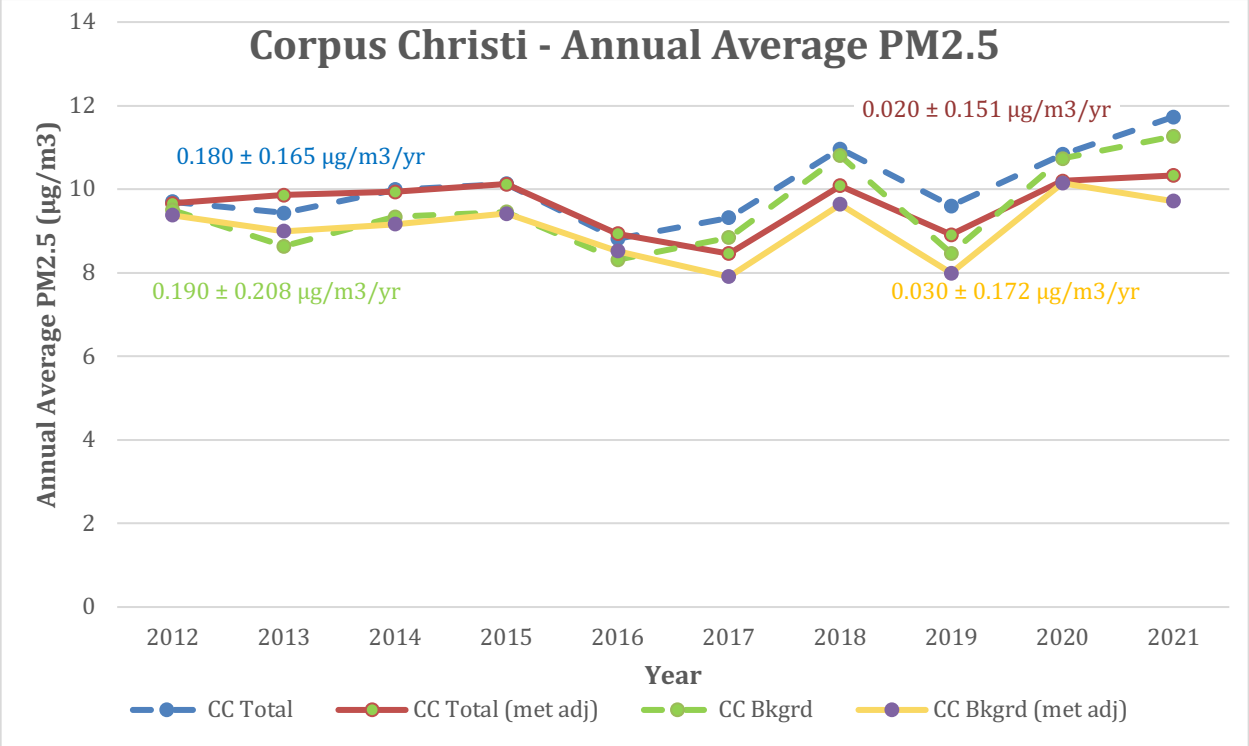

Figure 90. Original and meteorologically adjusted annual averages for total and background $PM_{2.5}$ trends in Corpus Christi, TX.

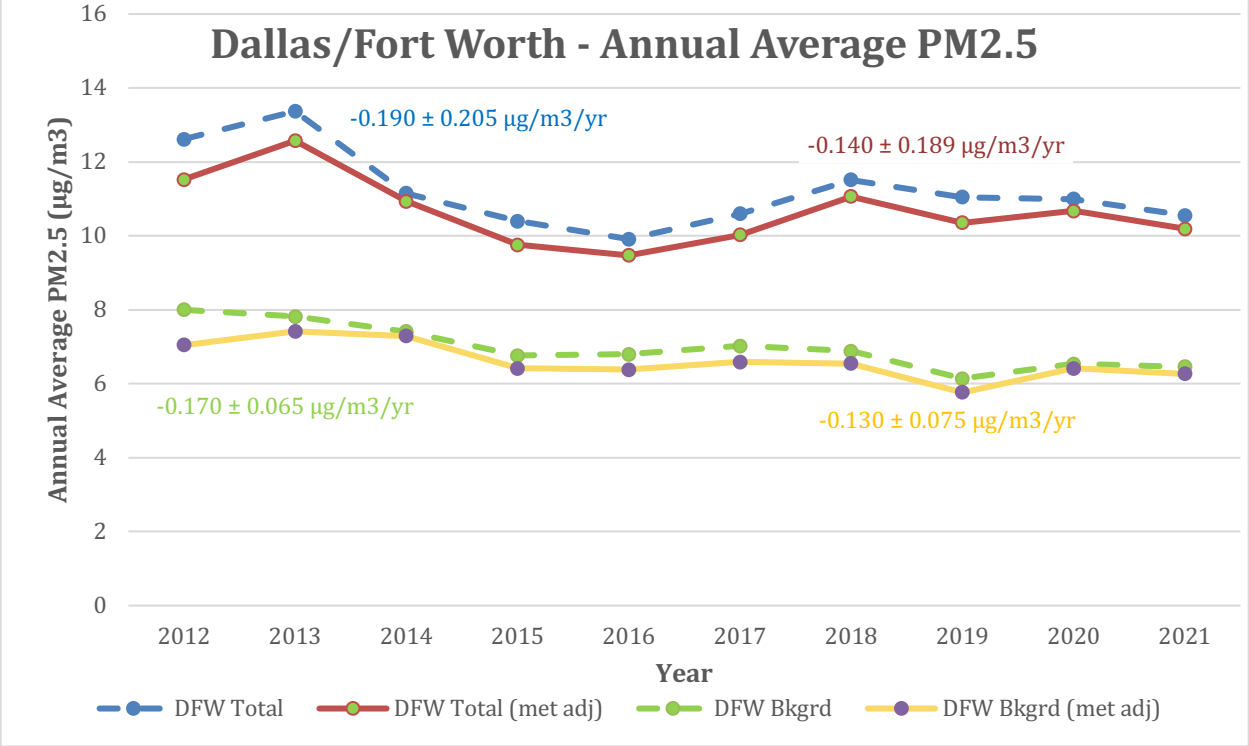

Figure 91. Original and meteorologically adjusted annual averages for total and background $PM_{2.5}$ trends in Dallas/Fort Worth, TX.

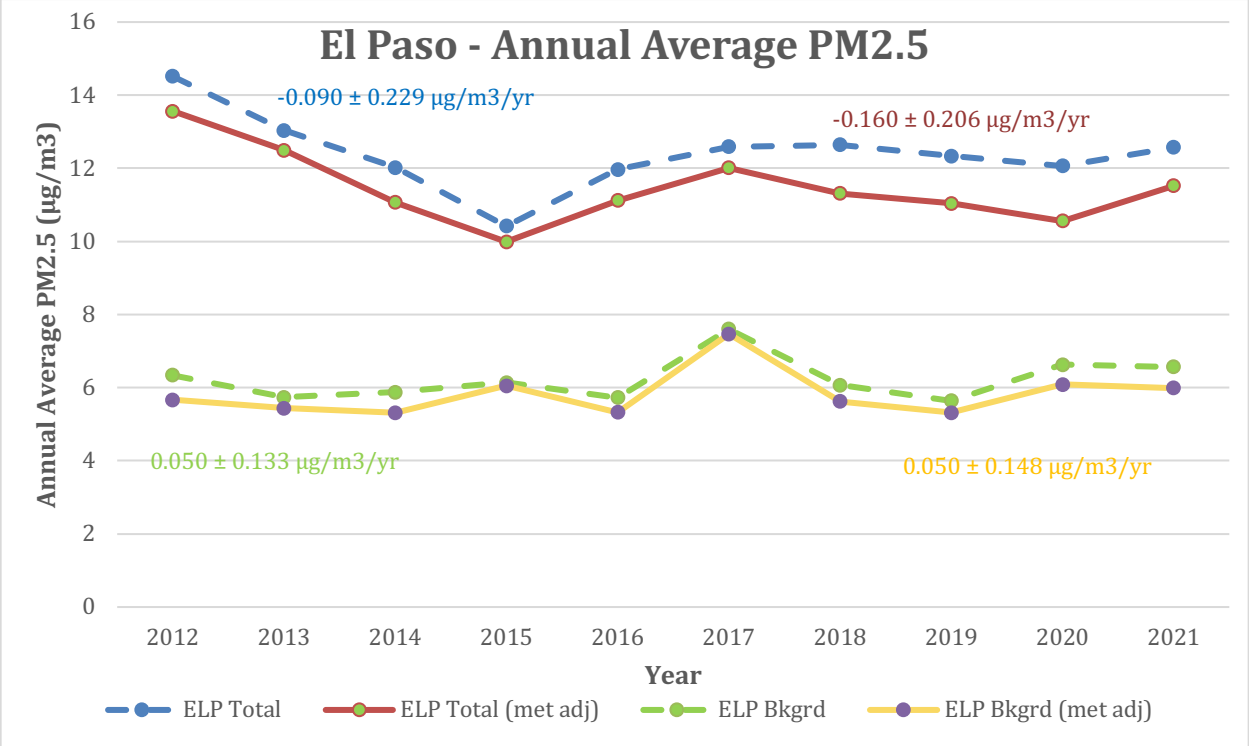

Figure 92. Original and meteorologically adjusted annual averages for total and background $PM_{2.5}$ trends in El Paso, TX.

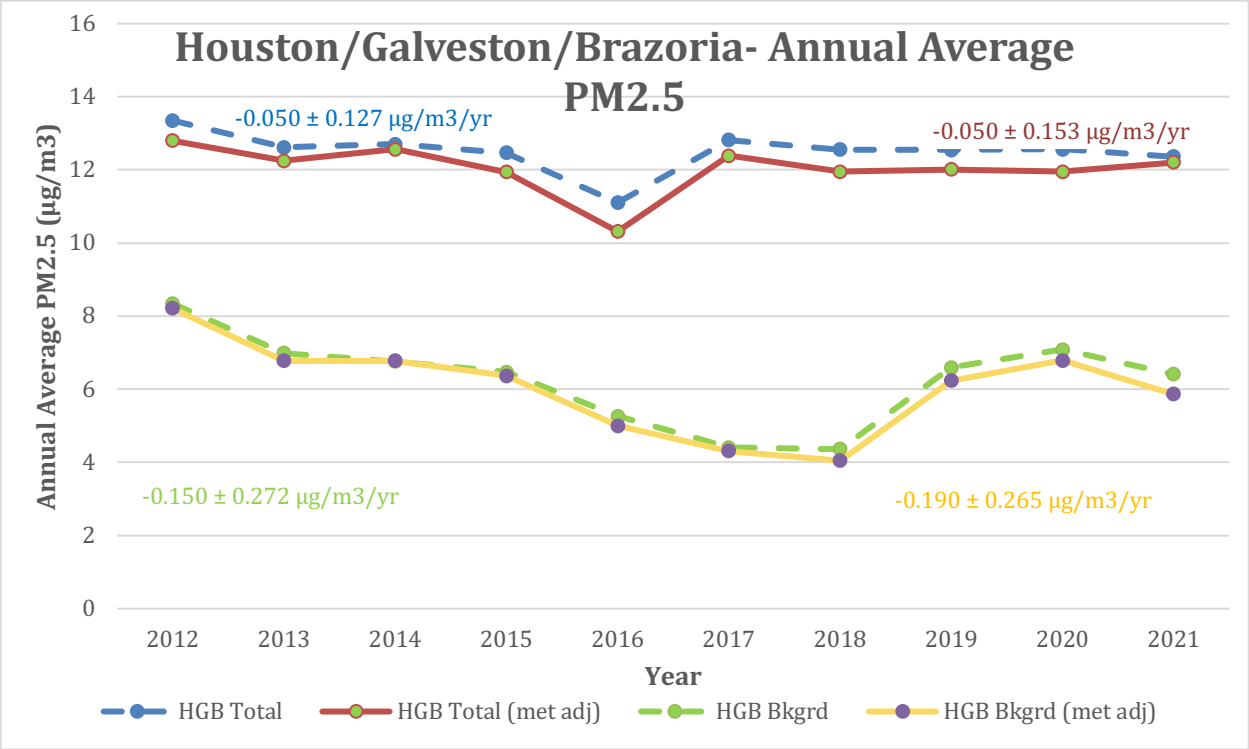

Figure 93. Original and meteorologically adjusted annual averages for total and background $PM_{2.5}$ trends in Houston/Galveston/Beaumont, TX.

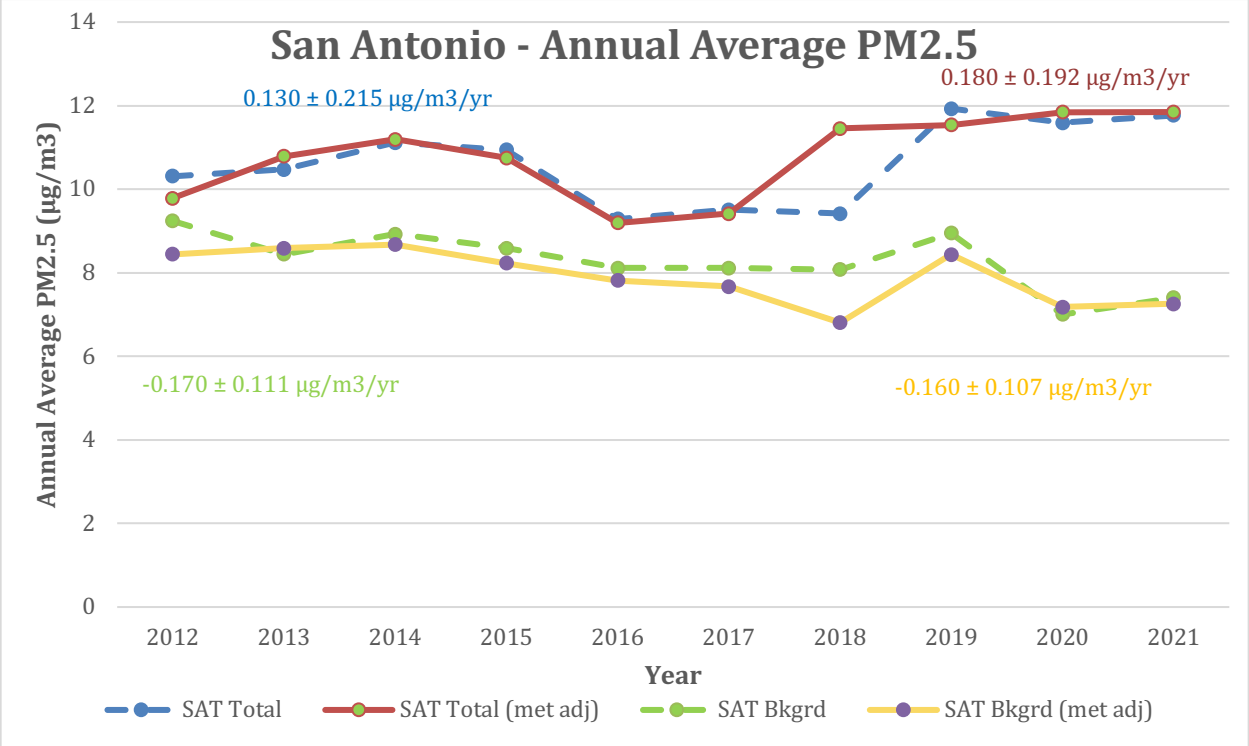

Figure 94. Original and meteorologically adjusted annual averages for total and background $PM_{2.5}$ trends in San Antonio, TX.

2.3.3 Meteorologically Adjusted Trends for Total and Background SO_2

The total and background SO_2 original trends and meteorologically adjusted linear trends are provided in Tables 19 and 20. The trend estimates are determined by ordinary least squares (OLS) linear regression of the annual averages. The original and meteorologically adjusted annual averages for total and background SO_2 trends are shown in Figure 95-101. No statistically significant trends with time are observed for 2012-2021 either before or after meteorological adjustment.

Table 19. Total SO_2 original trends and meteorologically adjusted linear trends.

Urban Area	Original	Met Adjusted
Austin	-0.120 ± 0.035 ppb/yr	-0.090 ± 0.034 ppb/yr
Beaumont/Port Arthur	0.110 ± 0.523 ppb/yr	0.200 ± 0.567 ppb/yr
Corpus Christi	-0.670 ± 0.361 ppb/yr	-0.130 ± 0.070 ppb/yr
Dallas/Fort Worth	1.250 ± 0.582 ppb/yr	1.140 ± 0.605 ppb/yr
El Paso	-0.120 ± 0.221 ppb/yr	-0.120 ± 0.053 ppb/yr
Houston/Galveston/Brazoria	-0.750 ± 0.289 ppb/yr	-0.730 ± 0.247 ppb/yr
San Antonio	-0.500 ± 0.415 ppb/yr	-0.470 ± 0.305 ppb/yr

Table 20. Background SO_2 original trends and meteorologically adjusted linear trends.

Urban Area	Original	Met Adjusted
Austin	N/A	N/A
Beaumont/Port Arthur	-0.360 ± 0.589 ppb/yr	-0.140 ± 0.224 ppb/yr
Corpus Christi	N/A	N/A
Dallas/Fort Worth	0.010 ± 0.046 ppb/yr	0.000 ± 0.000 ppb/yr
El Paso	-0.010 ± 0.027 ppb/yr	0.000 ± 0.012 ppb/yr
Houston/Galveston/Brazoria	-0.100 ± 0.105 ppb/yr	-0.110 ± 0.086 ppb/yr
San Antonio	-0.170 ± 0.116 ppb/yr	-0.190 ± 0.047 ppb/yr

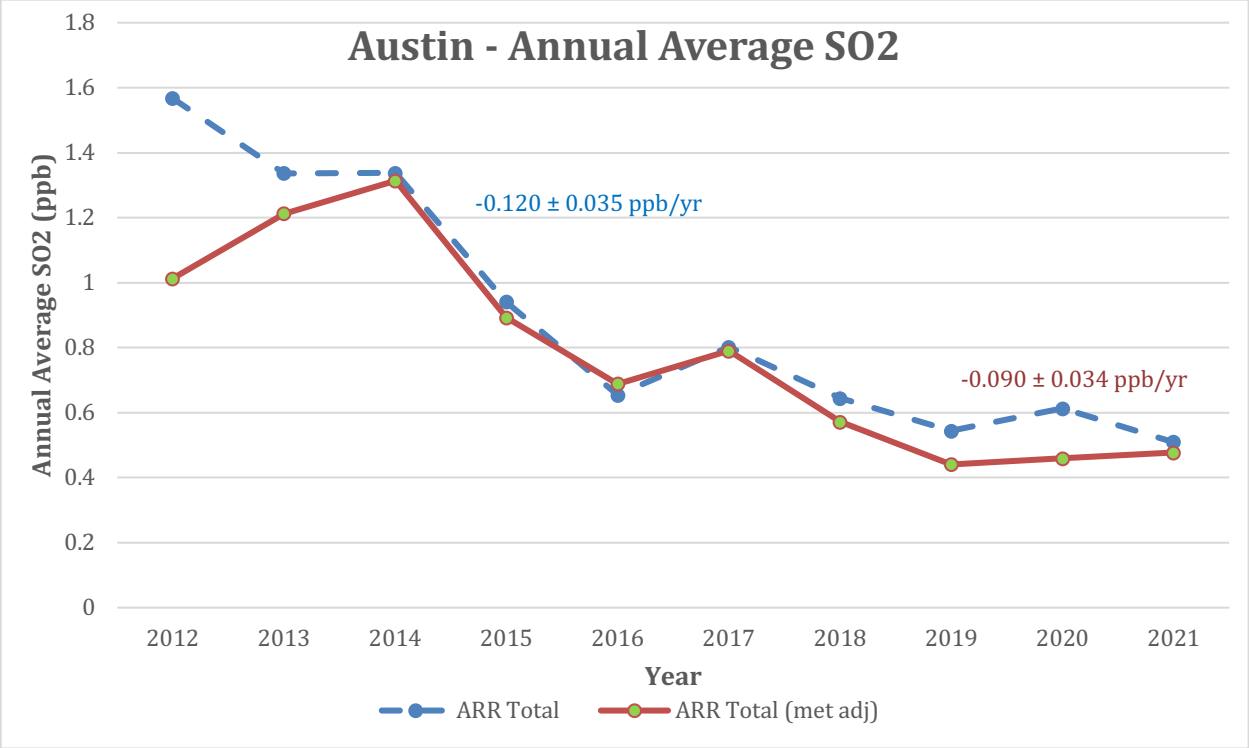

Figure 95. Original and meteorologically adjusted annual averages for total SO_2 trends in Austin, TX. Background SO_2 data was not available in Austin.

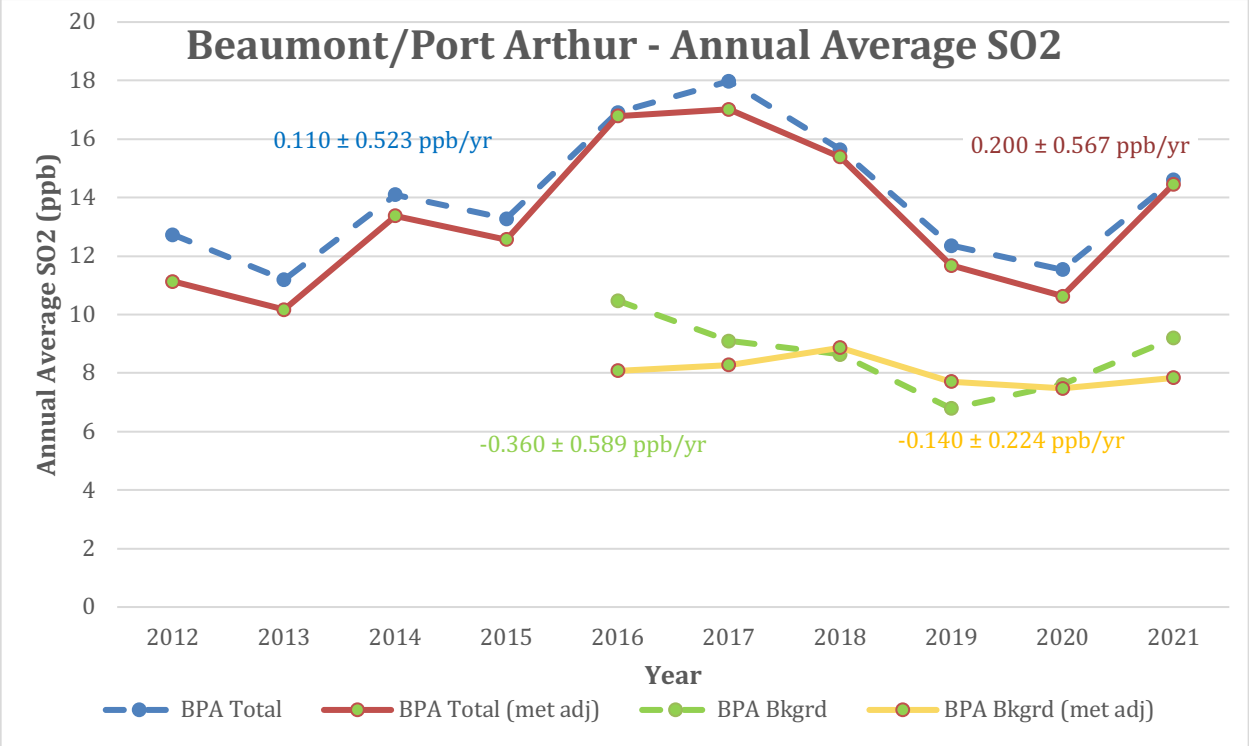

Figure 96. Original and meteorologically adjusted annual averages for total and background SO_2 trends in Beaumont/Port Arthur, TX.

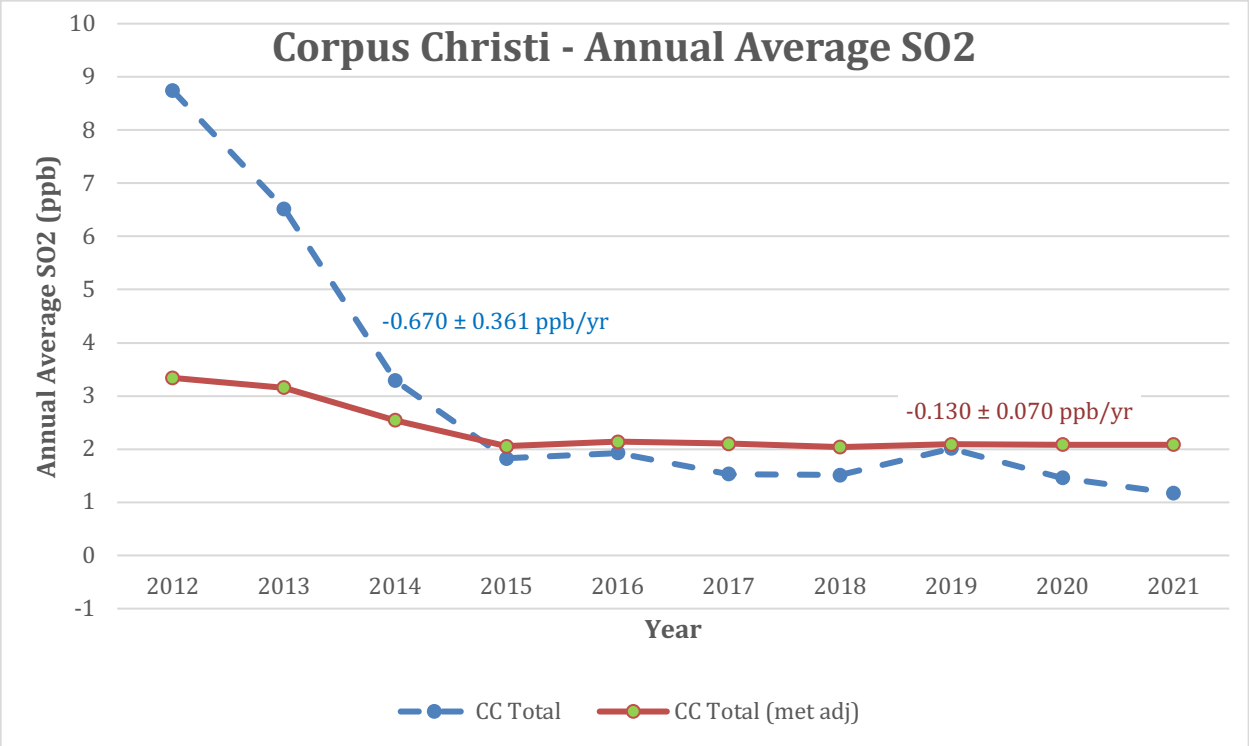

Figure 97. Original and meteorologically adjusted annual averages for total SO_2 trends in Corpus Christi, TX. Background SO_2 data was not available in Corpus Christi.

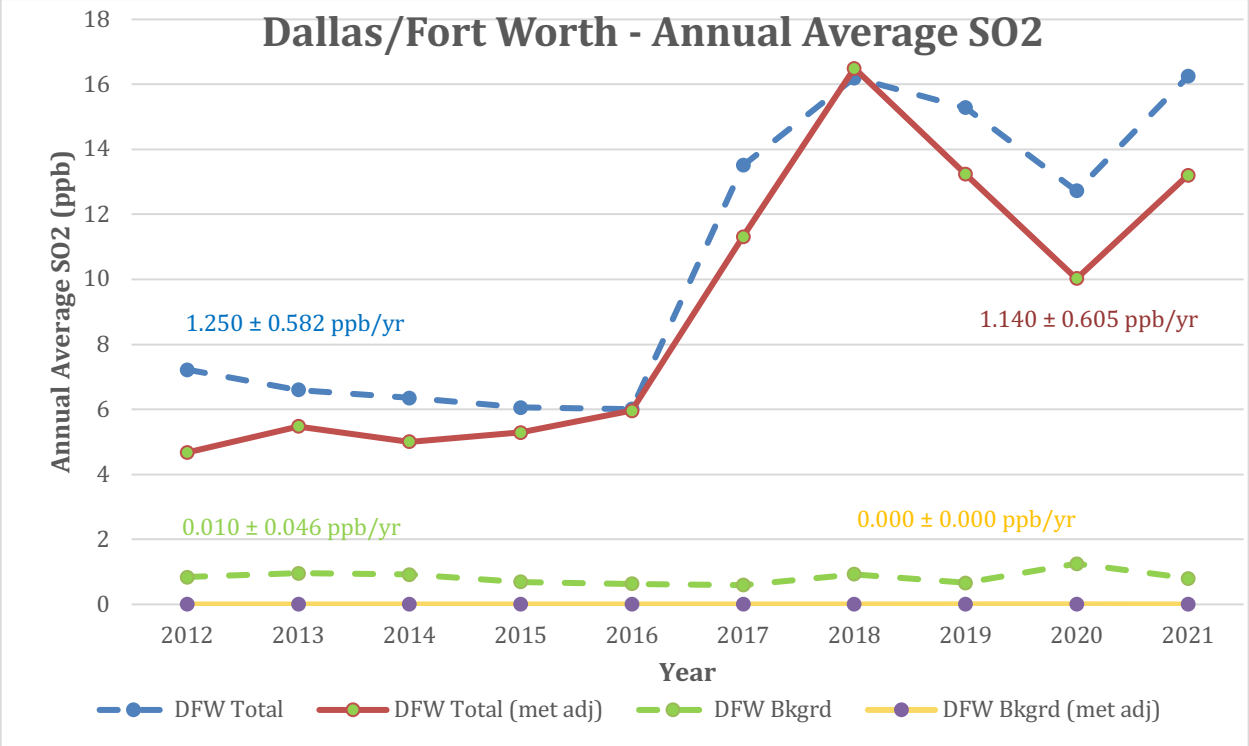

Figure 98. Original and meteorologically adjusted annual averages for total and background SO_2 trends in Dallas/Fort Worth, TX.

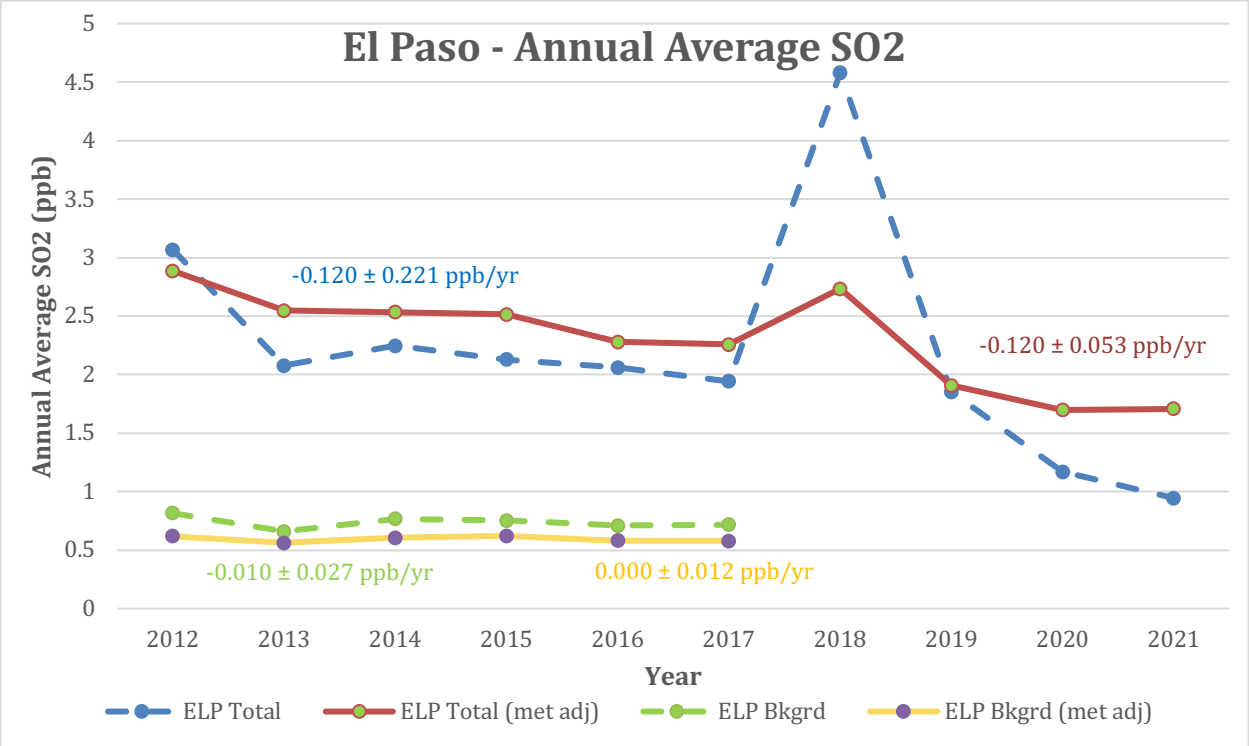

Figure 99. Original and meteorologically adjusted annual averages for total and background SO_2 trends in El Paso, TX.

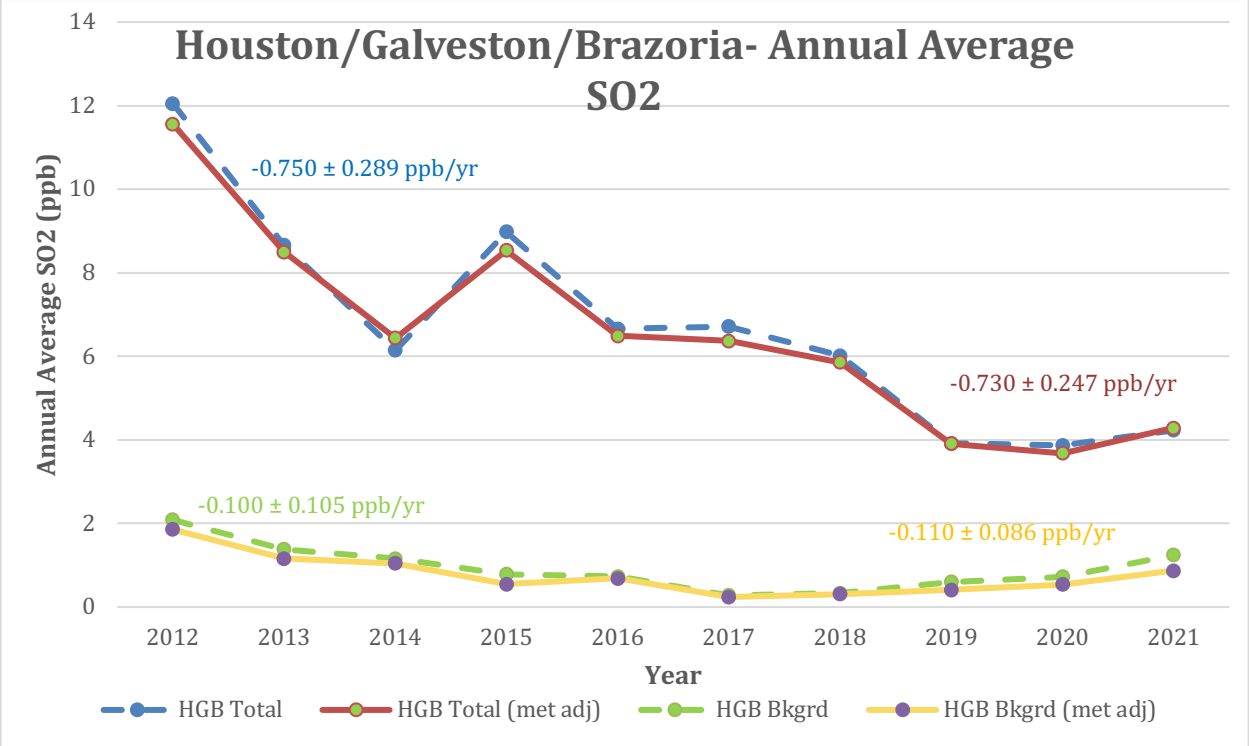

Figure 100. Original and meteorologically adjusted annual averages for total and background SO_2 trends in Houston/Galveston/Brazoria, TX.

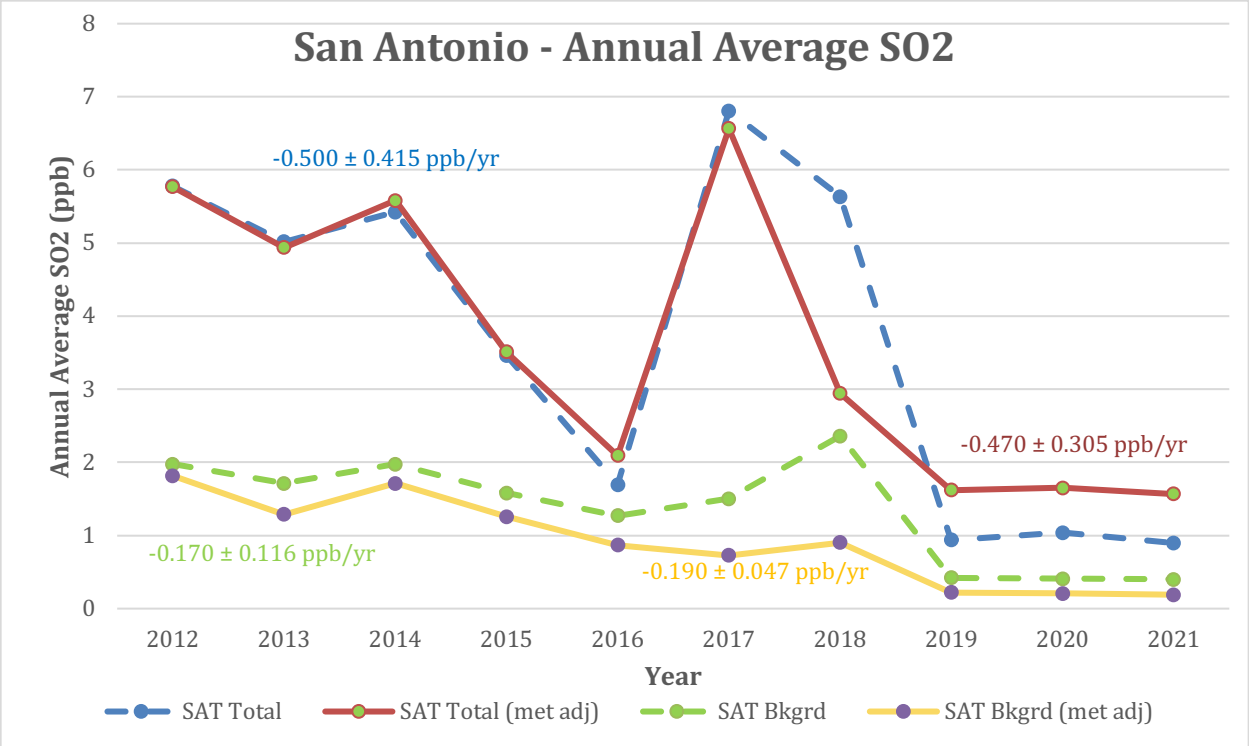

Figure 101. Original and meteorologically adjusted annual averages for total and background SO_2 trends in San Antonio, TX.

2.4 Temporal Trends of Background MDA8 O_3, $PM_{2.5}$, and SO_2

2.4.1 Temporal Trends of Background O_3

Figures 102-108 show the seasonal (top) and annual (bottom) trends in the background MDA8 O_3.

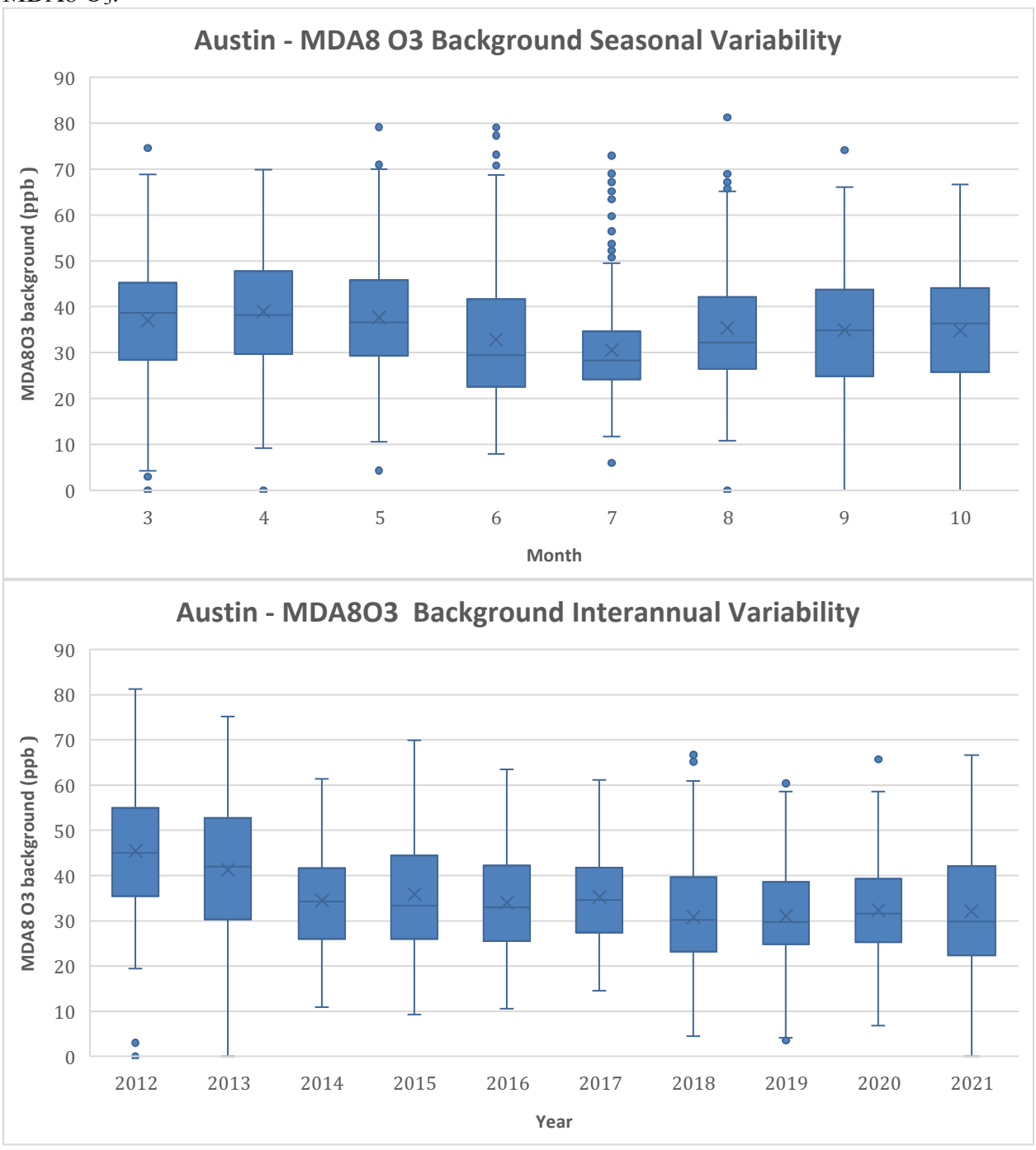

Figure 102. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background MDA8 O_3 for Austin, TX estimated using the TCEQ method. The line is the mean. Box edges

show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range.

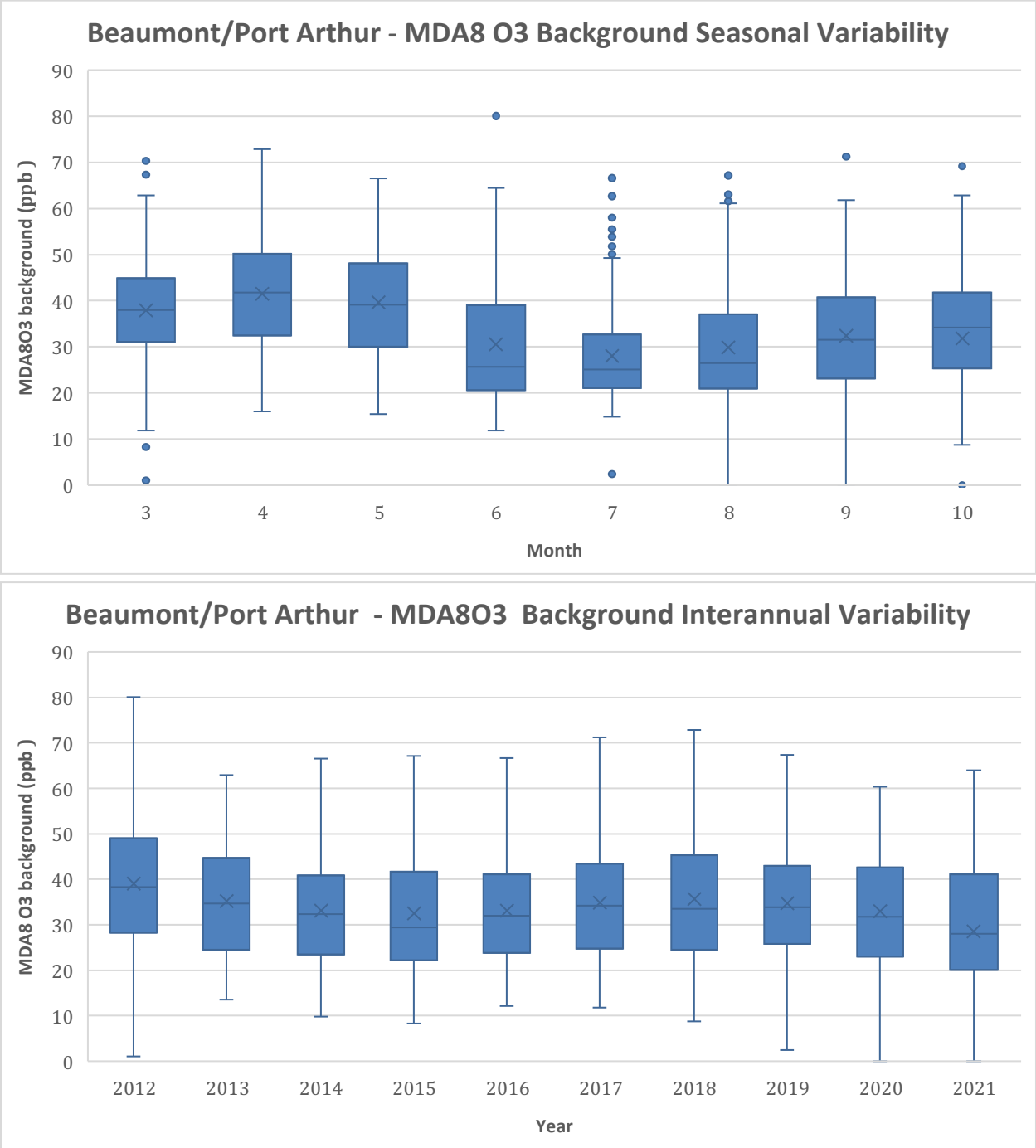

Figure 103. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background MDA8 O_3 for Beaumont/Port Arthur, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range.

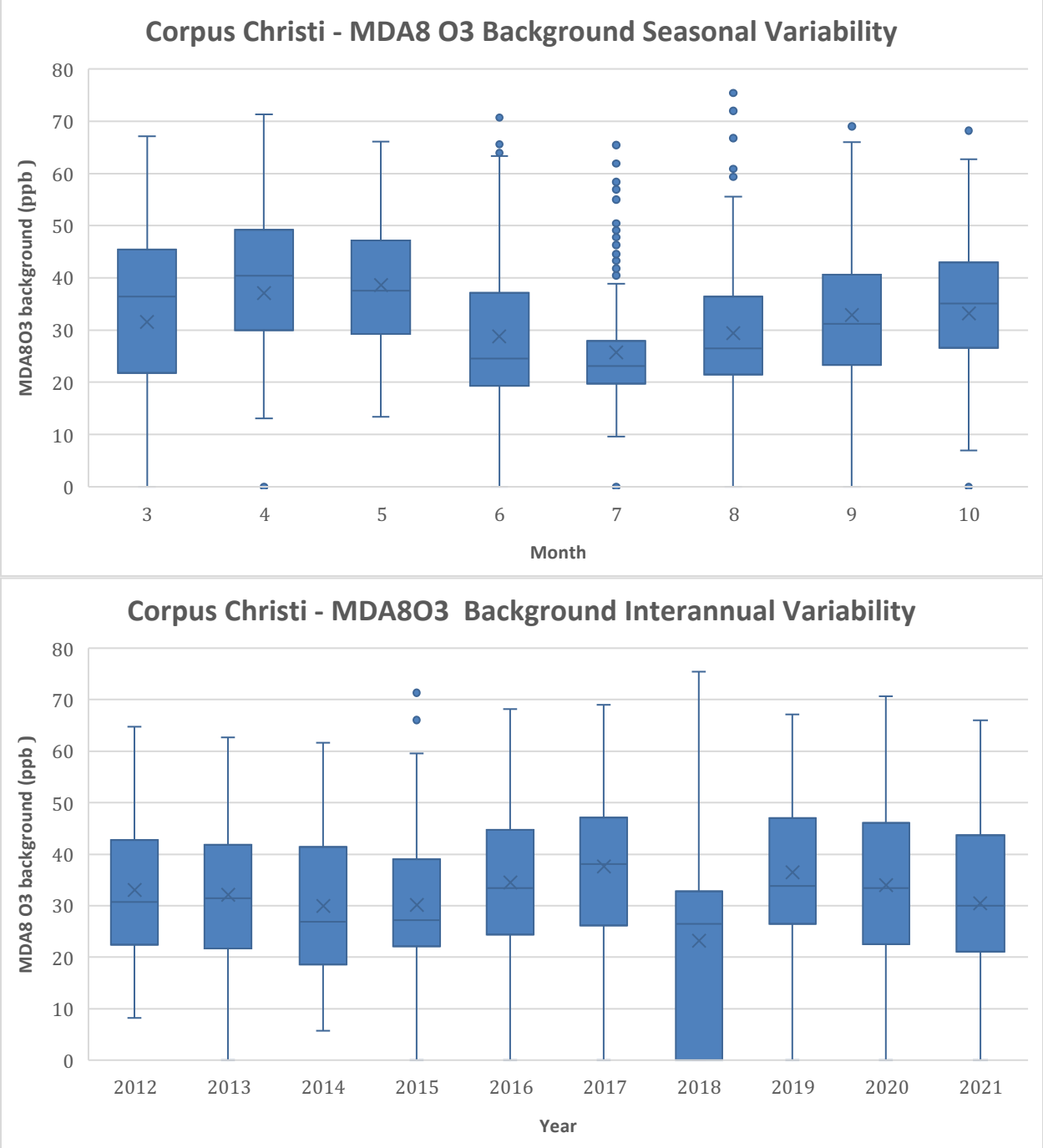

Figure 104. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background MDA8 O_3 for Corpus Christi, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range.

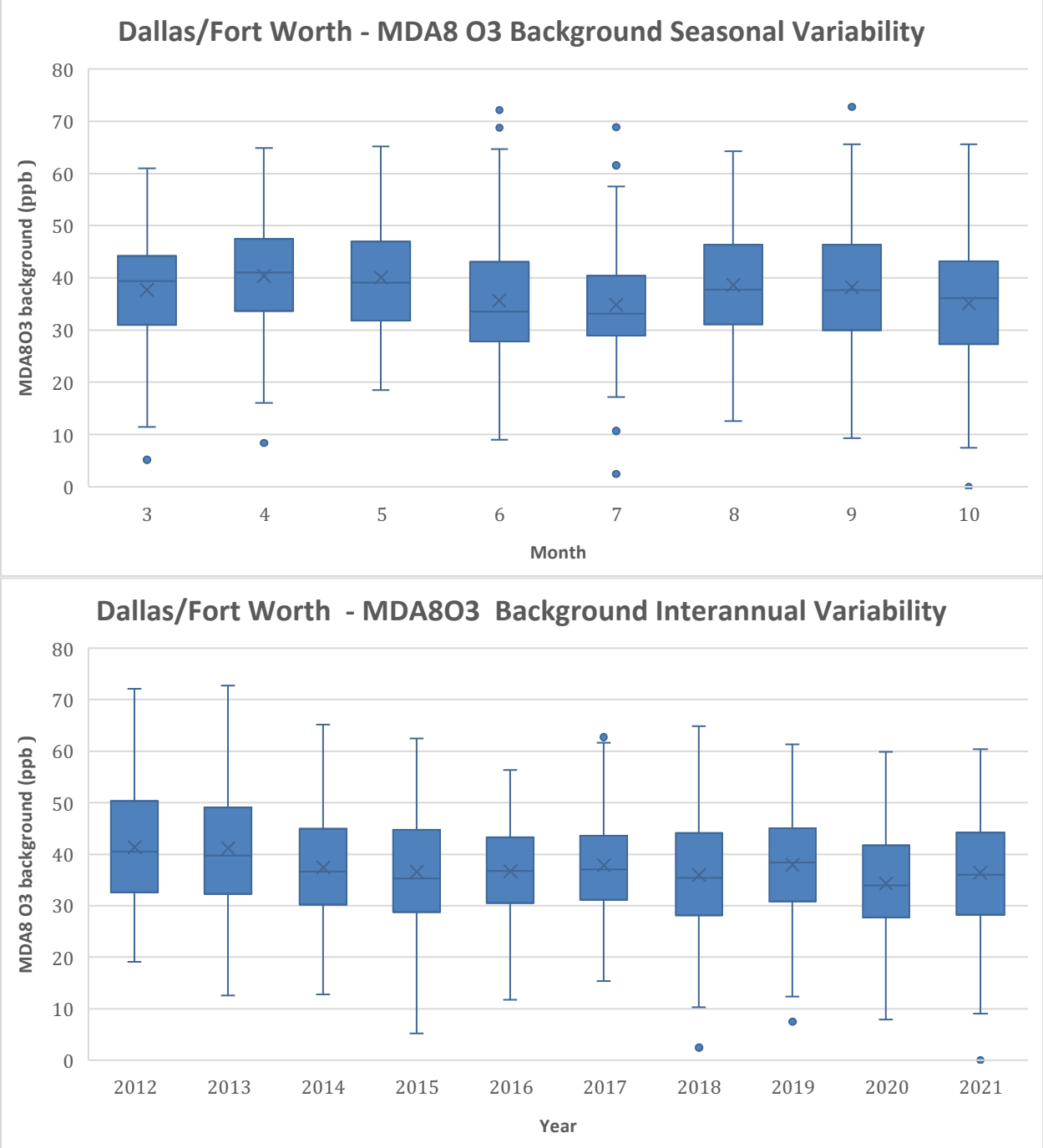

Figure 105. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background MDA8 O_3 for Dallas/Fort Worth, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range.

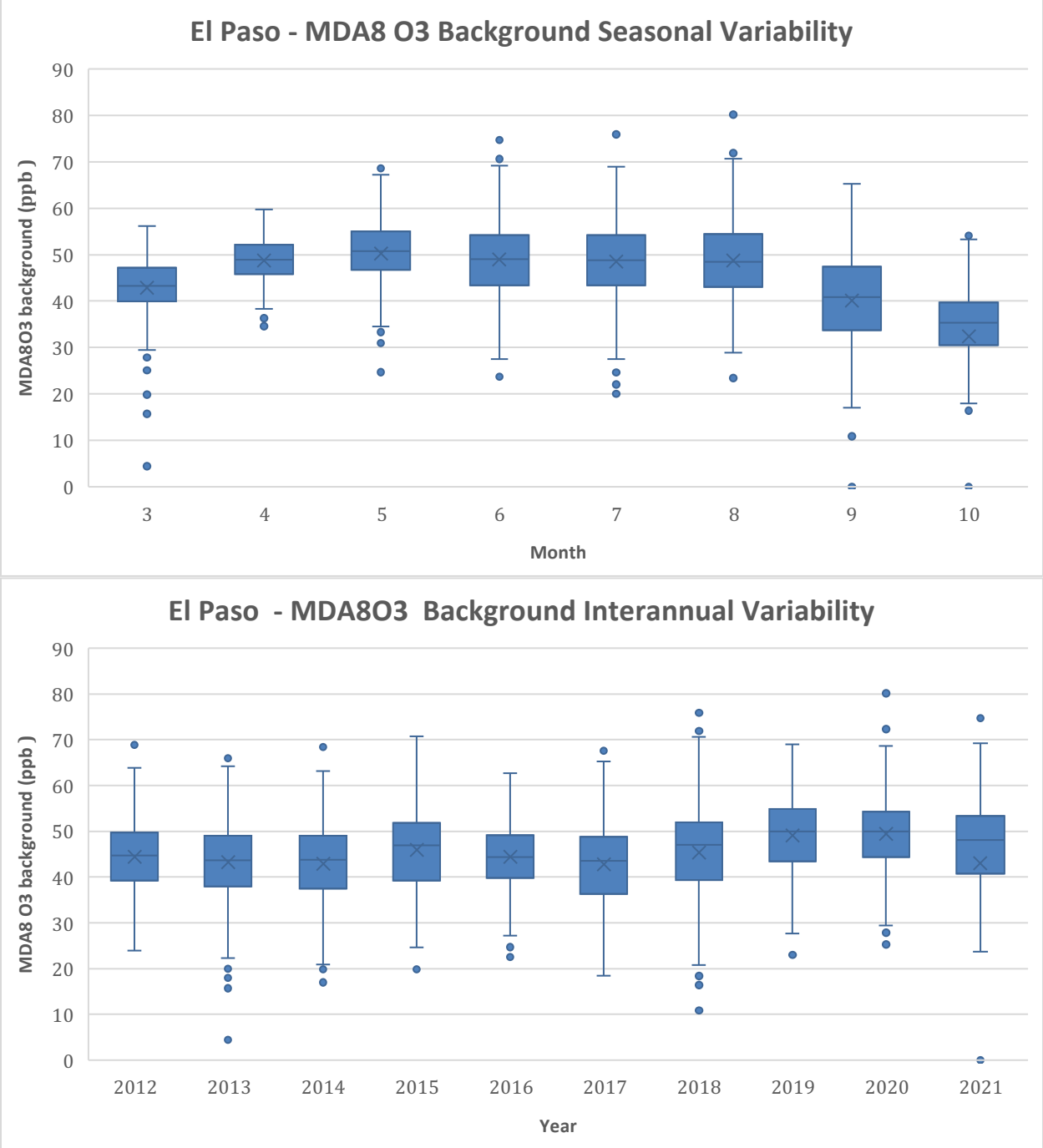

Figure 106. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background MDA8 O_3 for El Paso, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range.

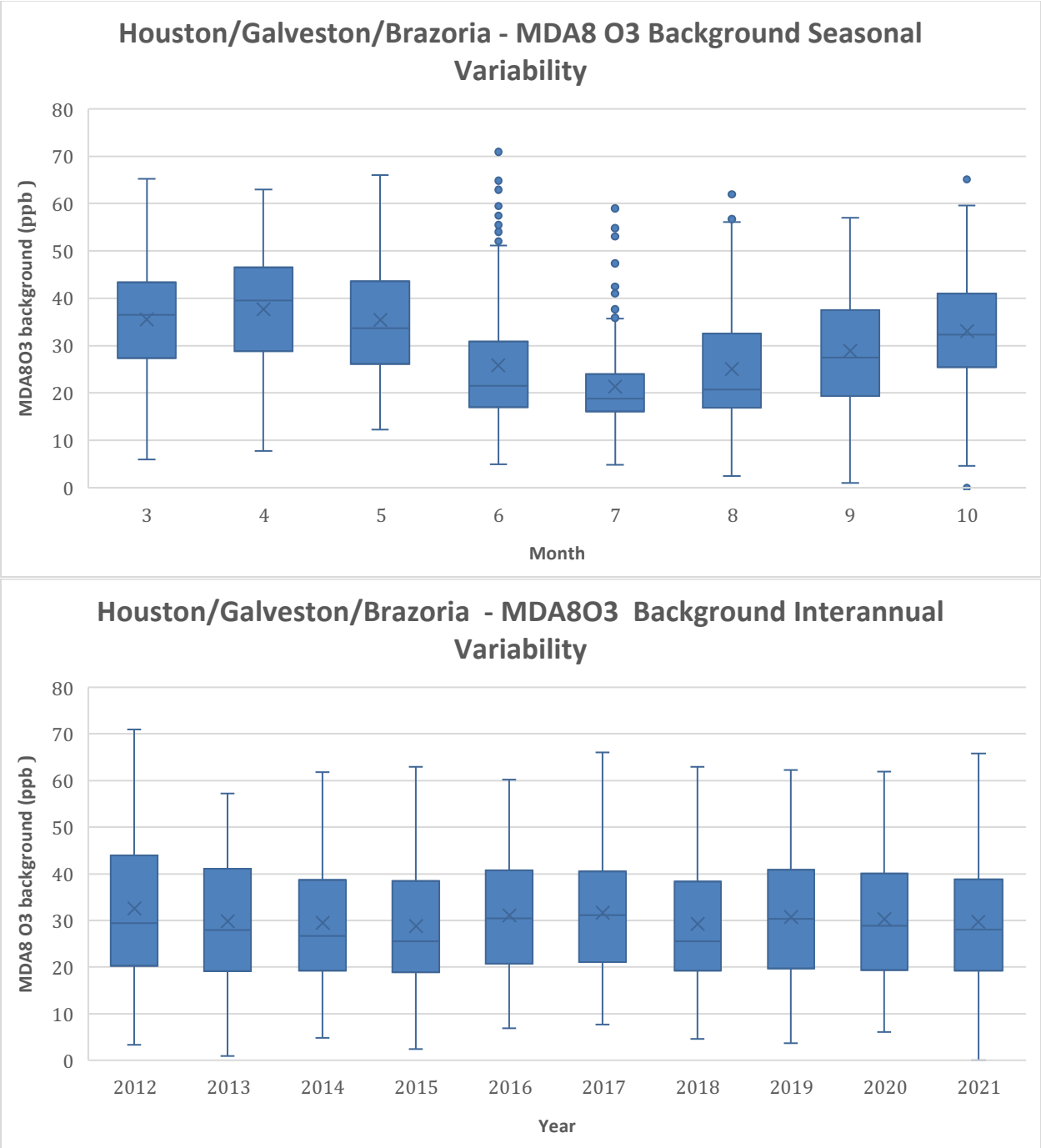

Figure 107. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background MDA8 O_3 for Houston/Galveston/Brazoria, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range.

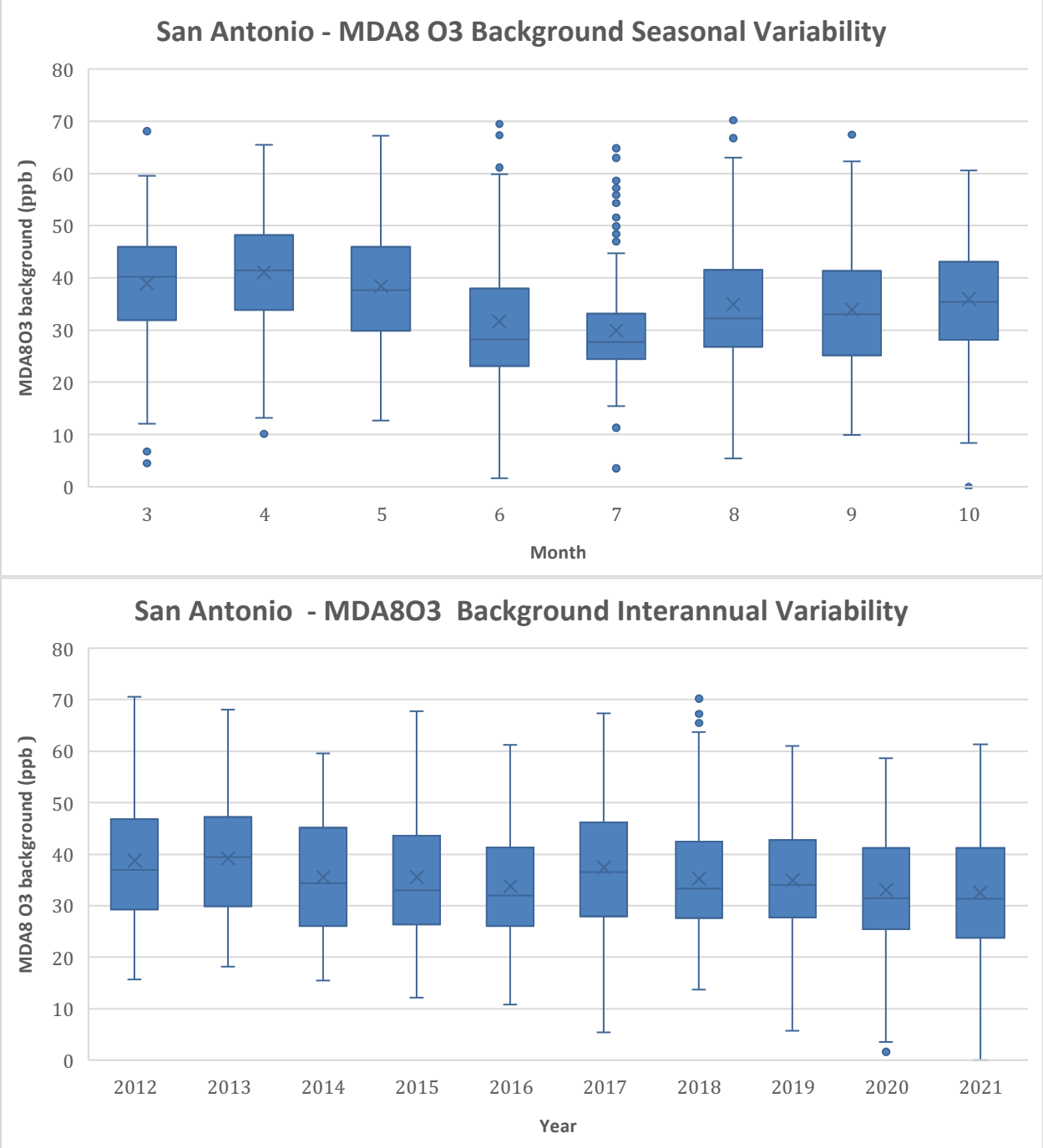

Figure 108. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background MDA8 O_3 for San Antonio, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range.

2.4.2 Temporal Trends of Background $PM_{2.5}$

Figures 109-115 show the seasonal (top) and annual (bottom) trends in the background $PM_{2.5}$.

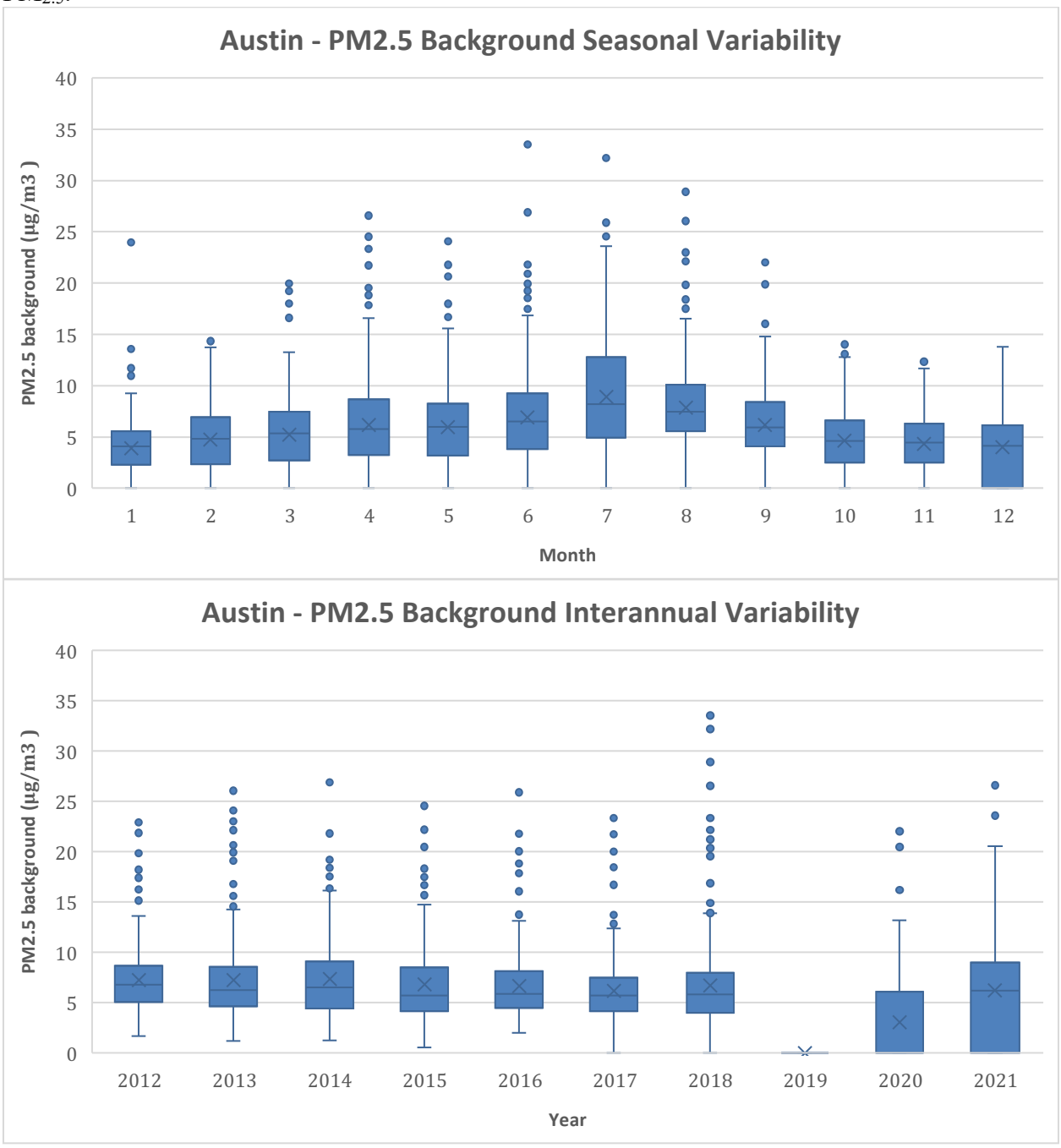

Figure 109. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background $PM_{2.5}$ for Austin, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range.

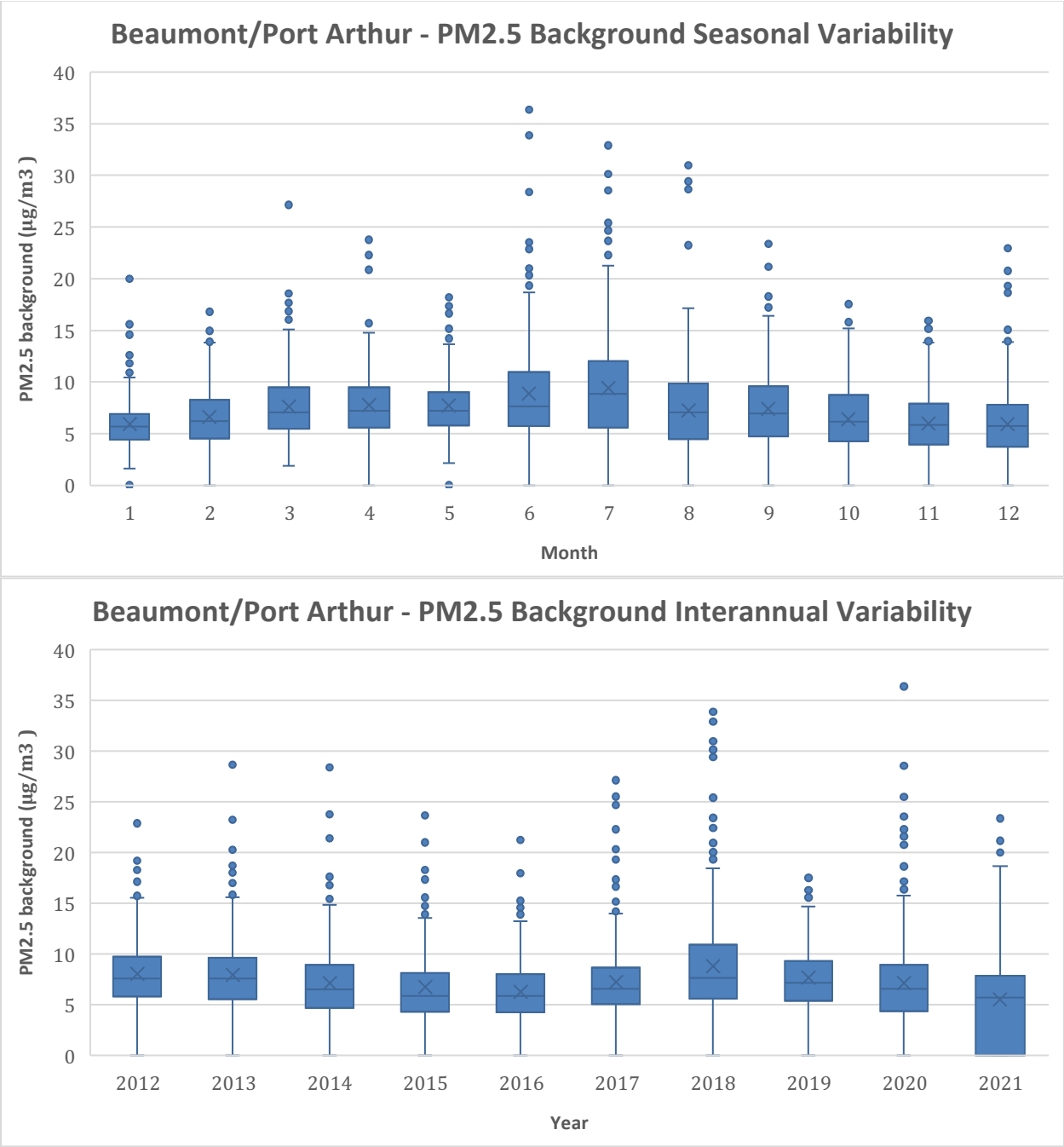

Figure 110. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background $PM_{2.5}$ for Beaumont/Port Arthur, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range.

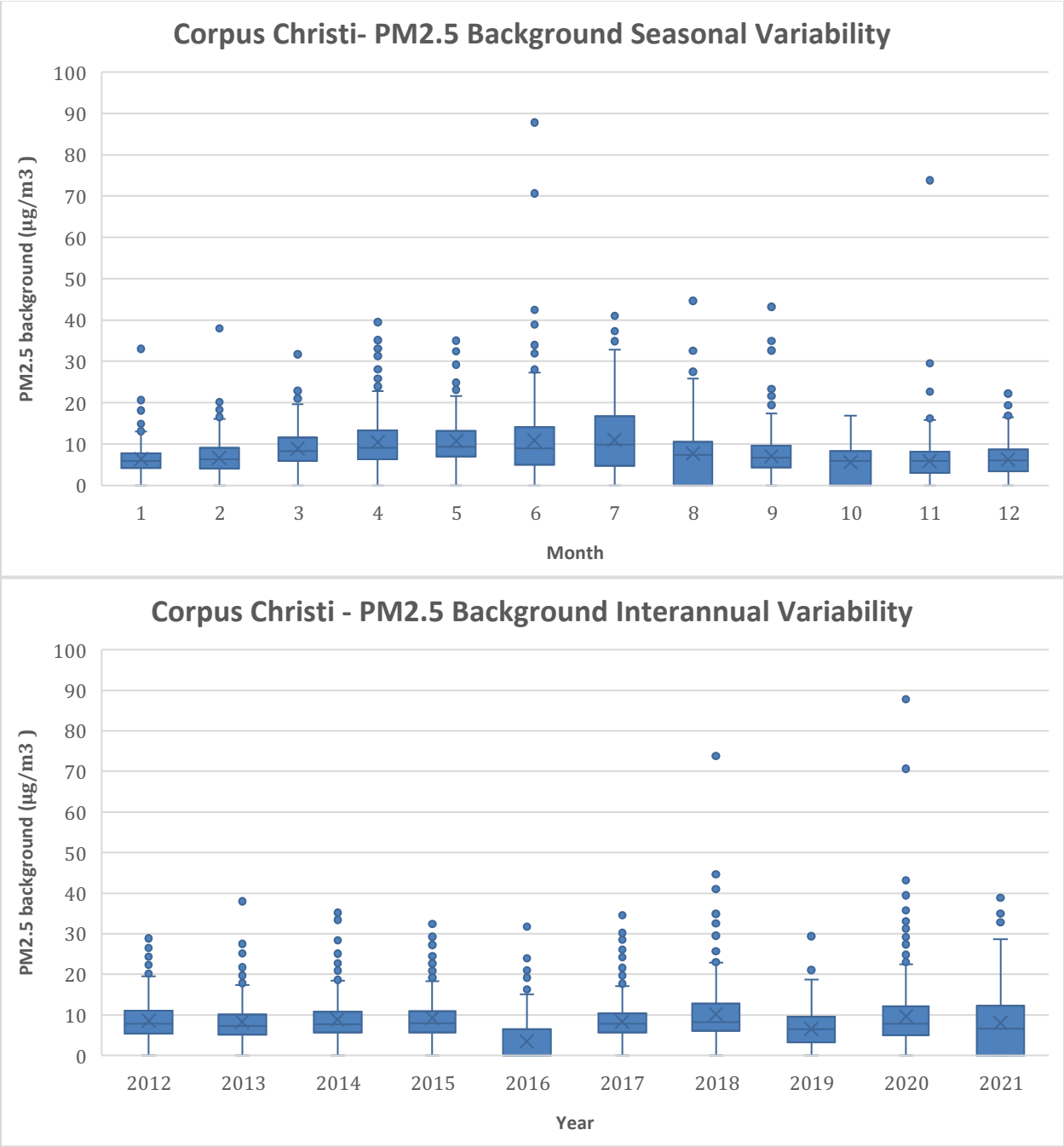

Figure 111. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background $PM_{2.5}$ for Corpus Christi, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range.

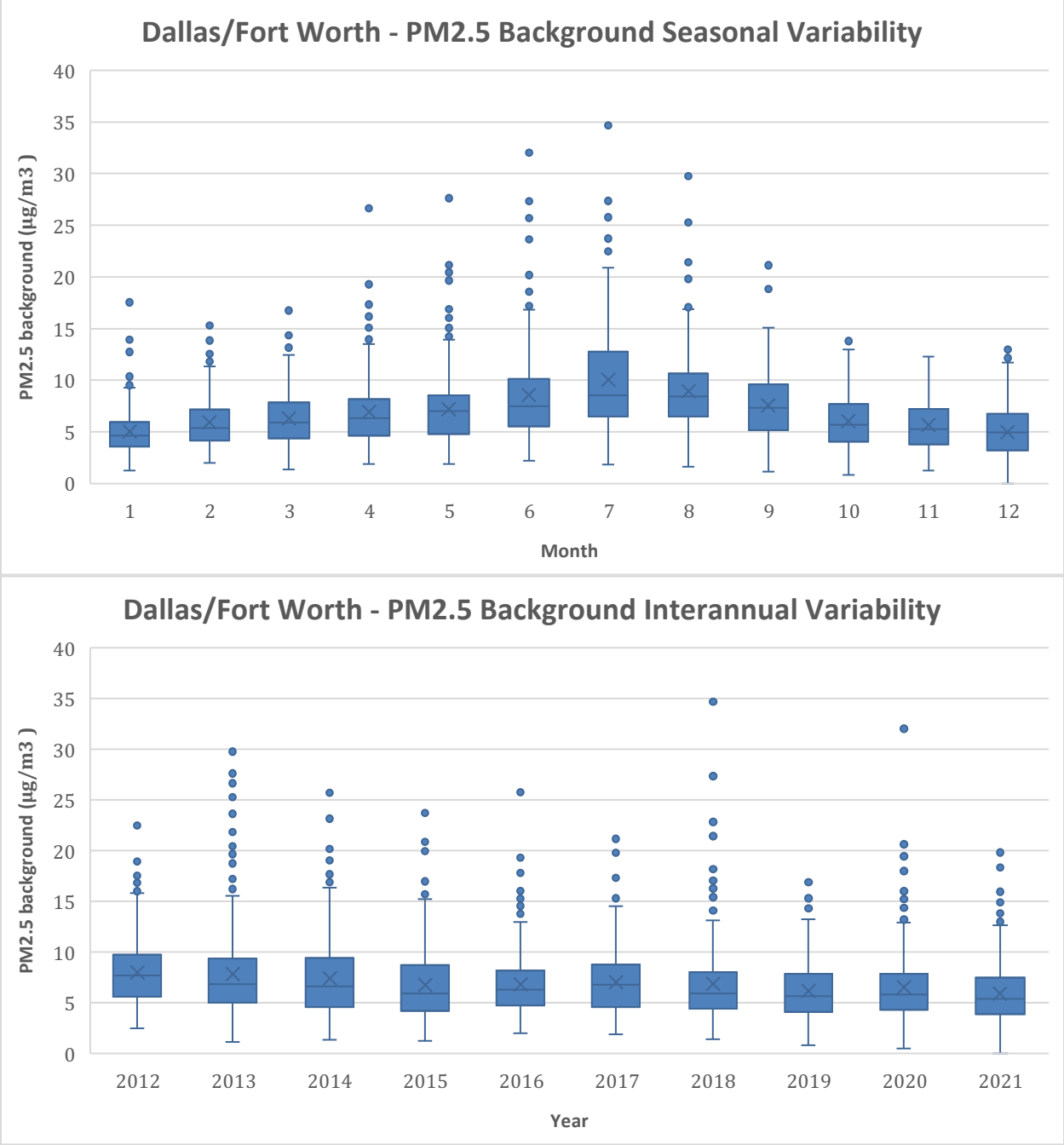

Figure 112. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background $PM_{2.5}$ for Dallas/Fort Worth, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range.

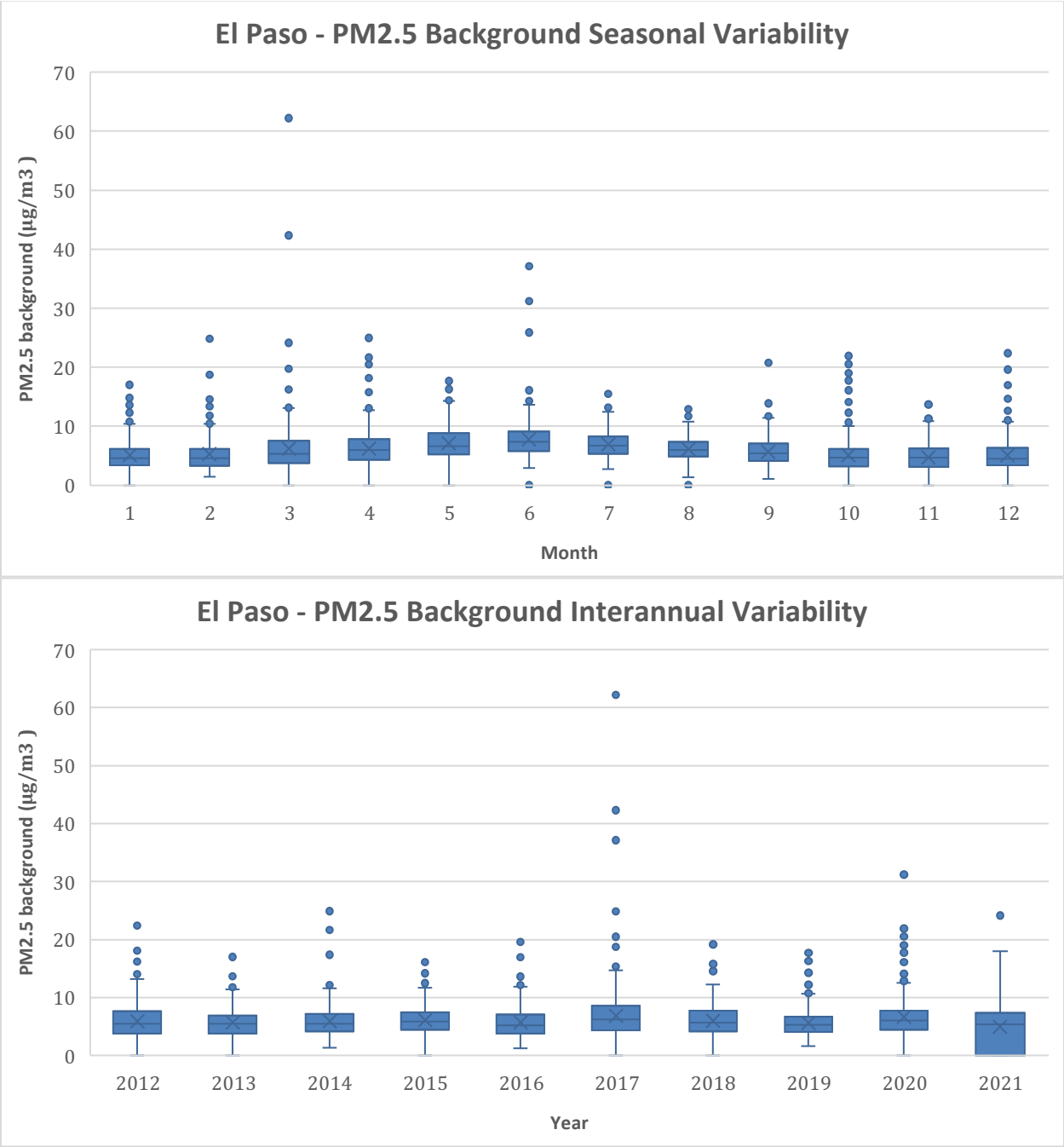

Figure 113. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background $PM_{2.5}$ for El Paso, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range.

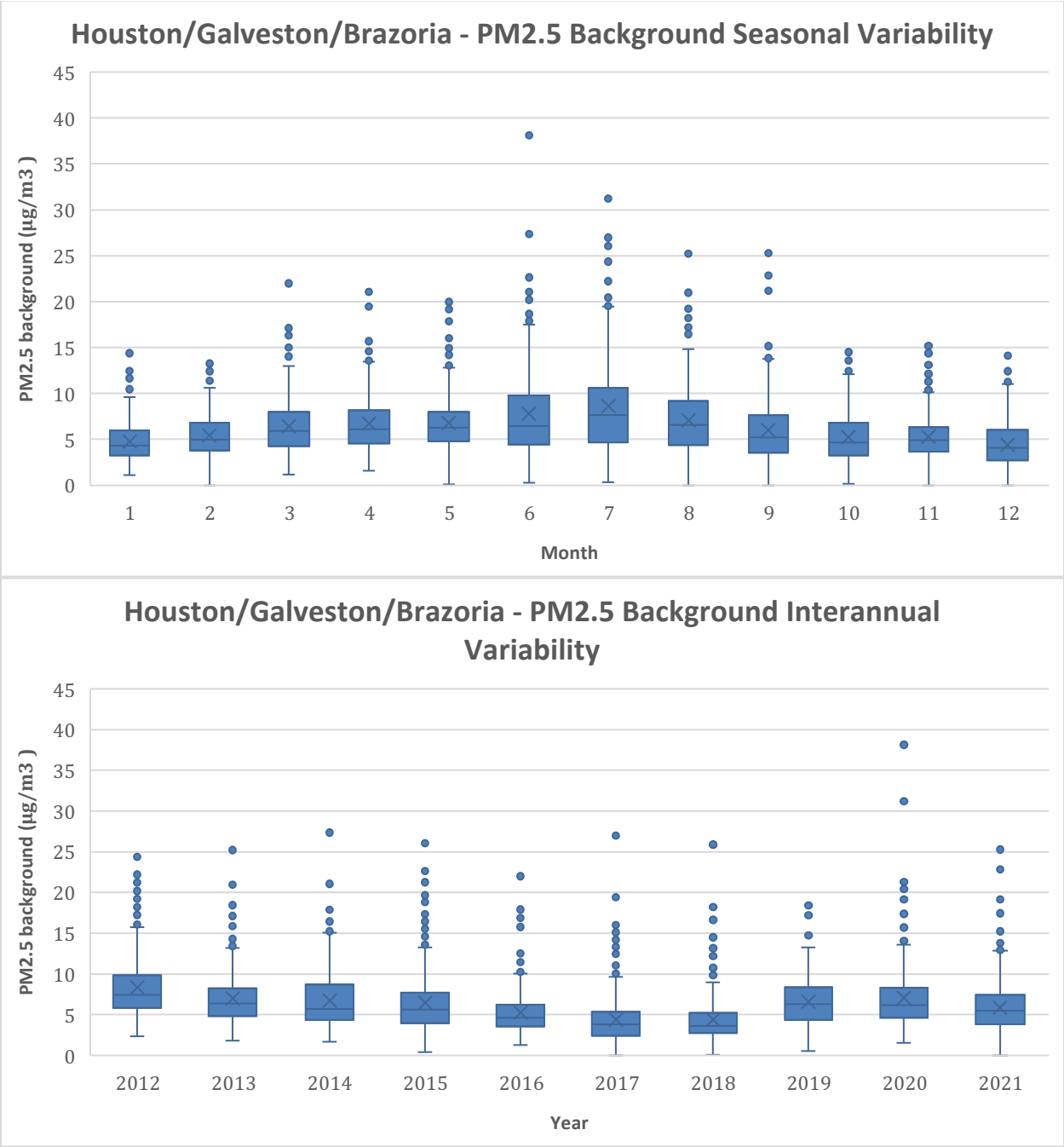

Figure 114. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background $PM_{2.5}$ for Houston/Galveston/Brazoria, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range.

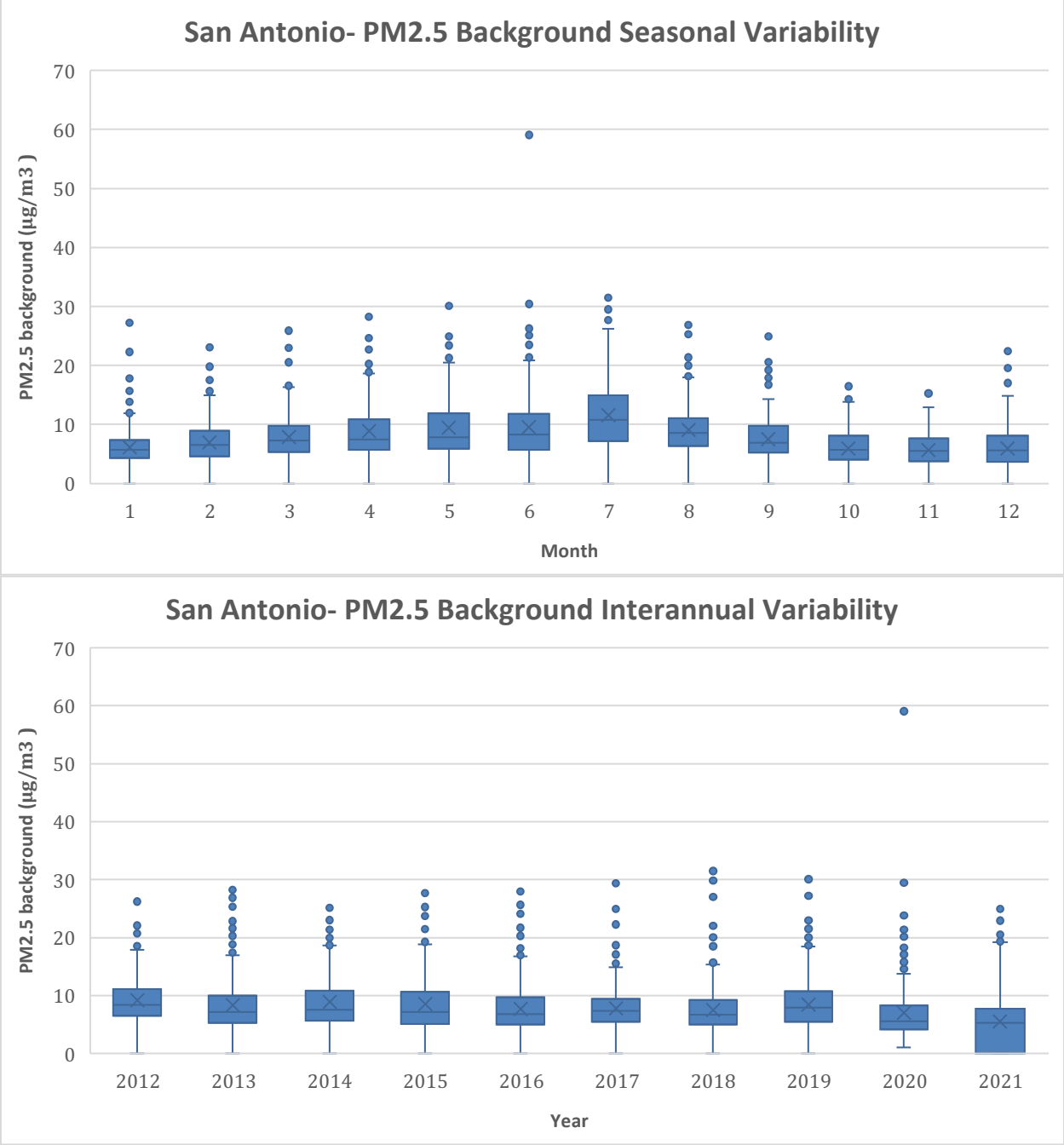

Figure 115. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background $PM_{2.5}$ for San Antonio, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range.

2.4.3 Temporal Trends of Background SO_2

Figures 116-120 show the seasonal (top) and annual (bottom) trends in the background SO_2.

Figure 116. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background SO_2 for Beaumont/Port Arthur, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range.

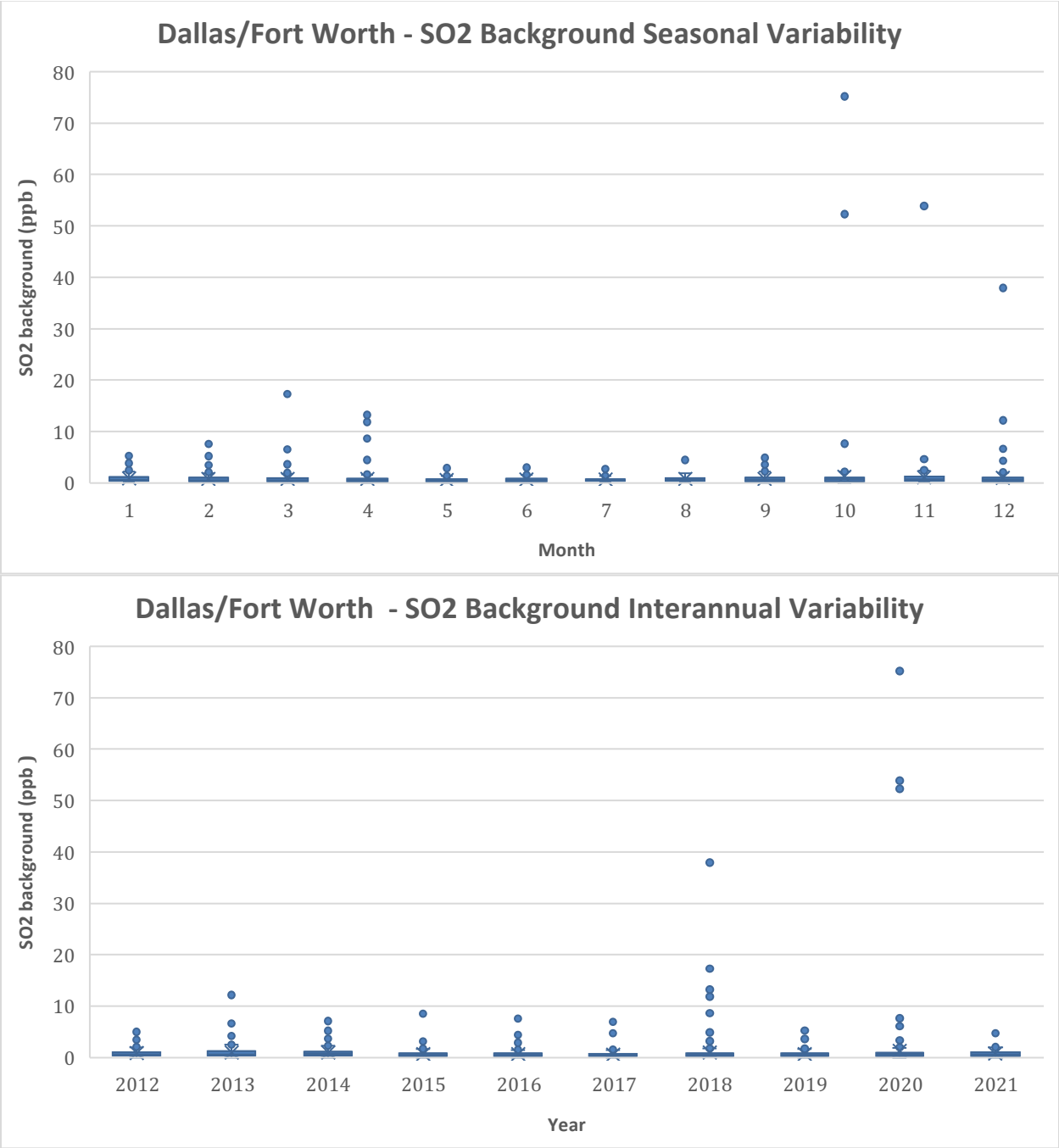

Figure 117. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background SO_2 for Dallas/Fort Worth, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range.

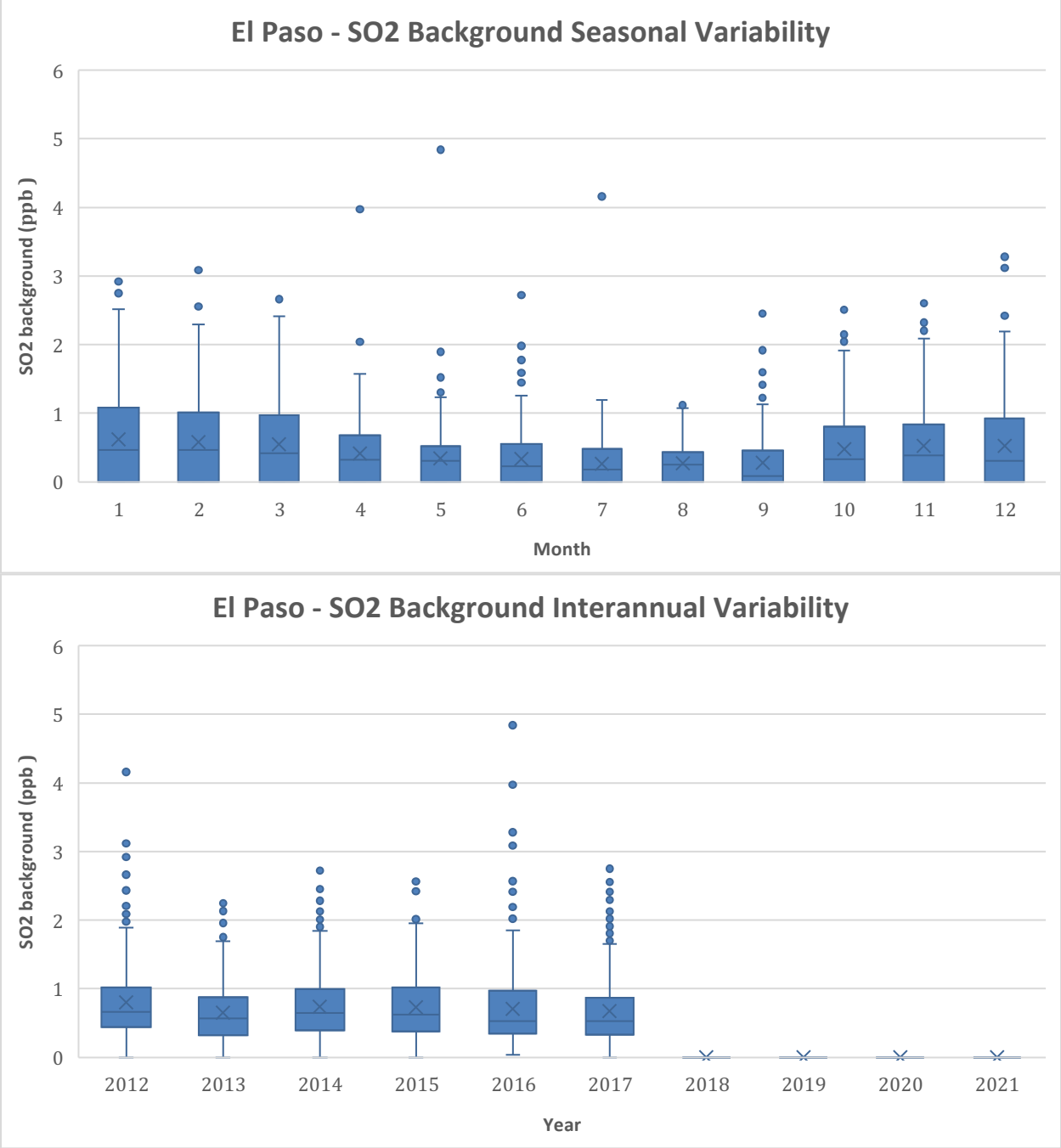

Figure 118. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background SO_2 for El Paso, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range.

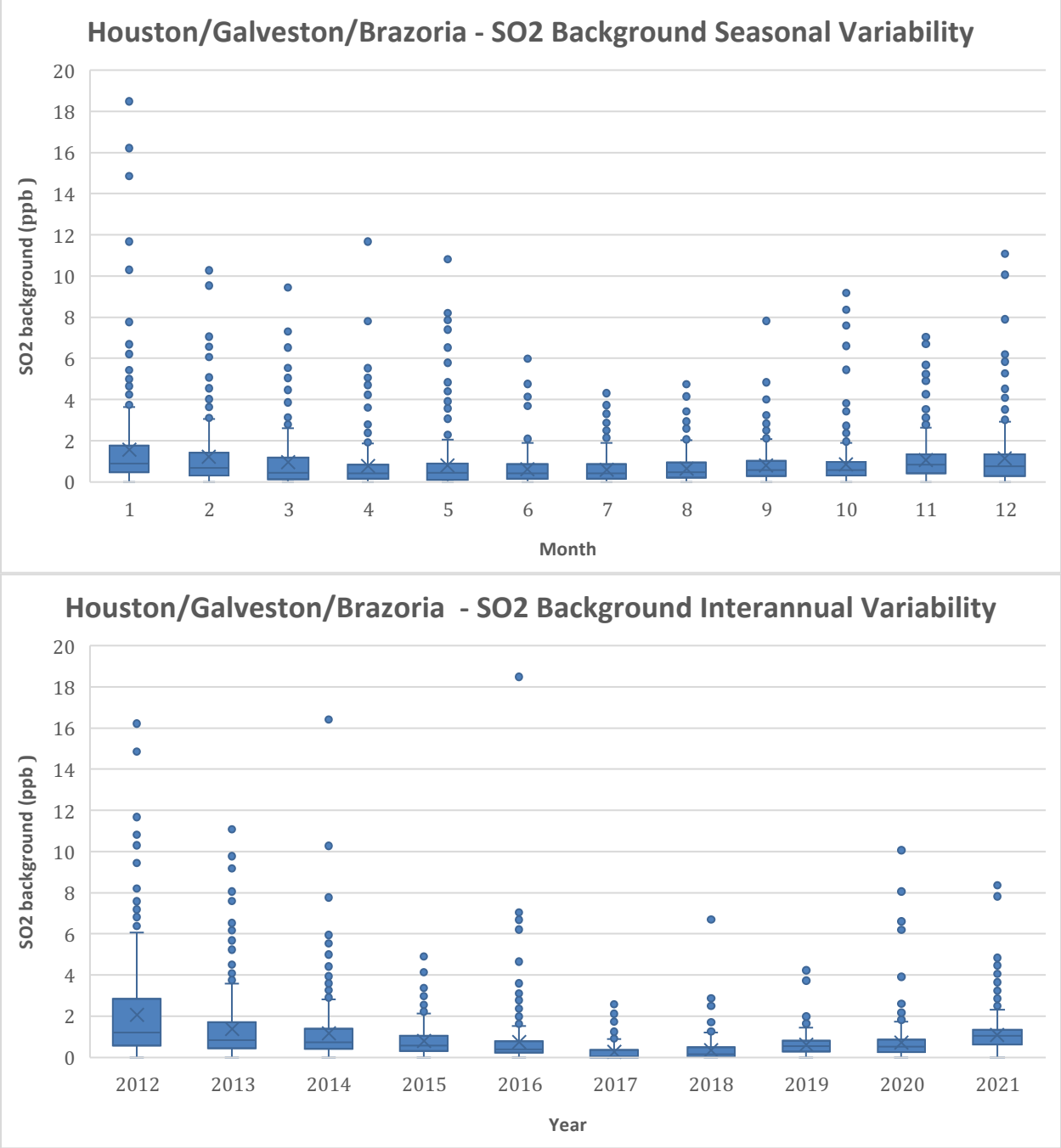

Figure 119. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background SO_2 for Houston/Galveston/Brazoria, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range.

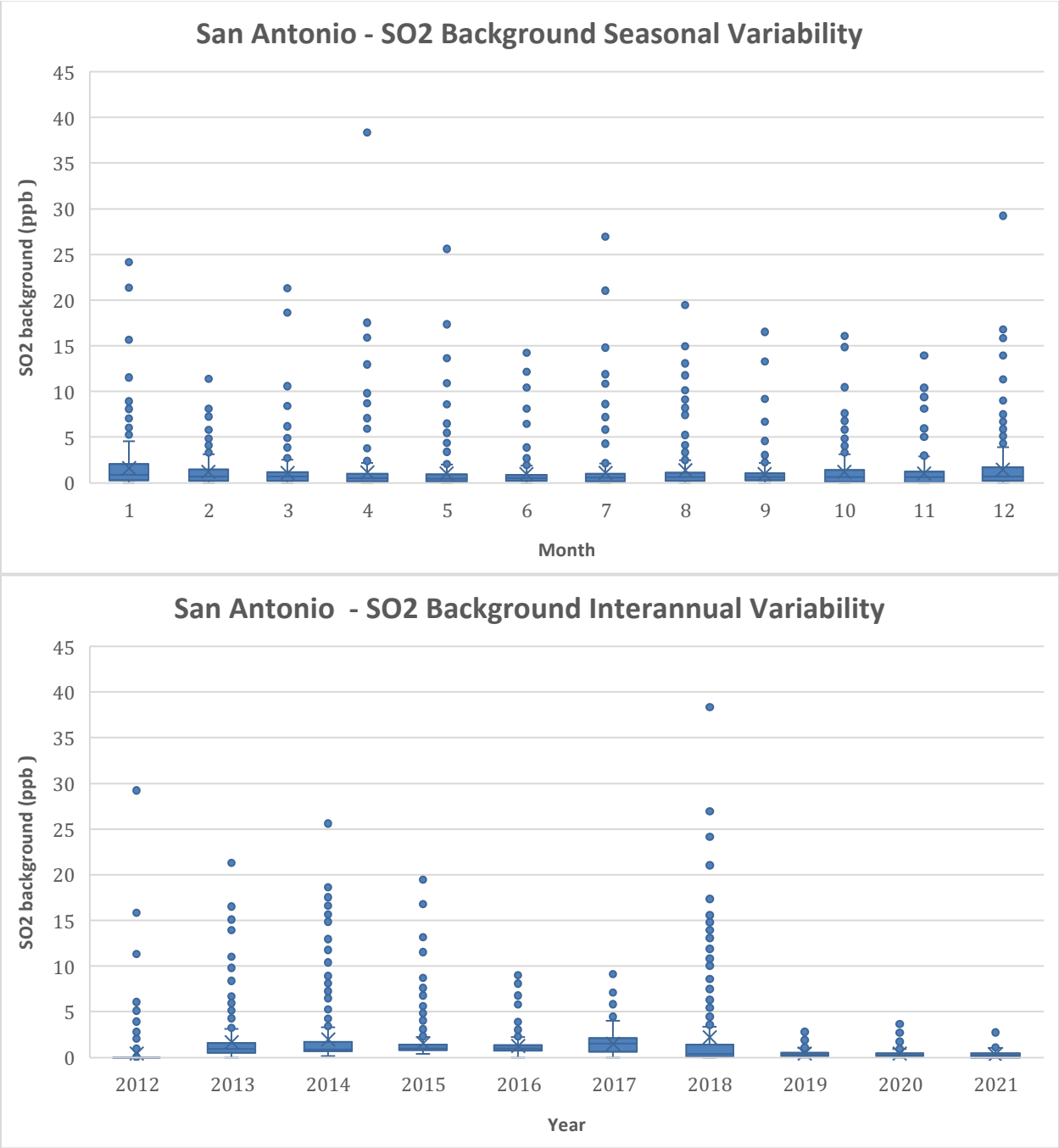

Figure 120. Box-and-whisker plots seasonal (top) and annual (bottom) trends in the background SO_2 for San Antonio, TX estimated using the TCEQ method. The line is the mean. Box edges show the 25th and 75th percentiles (Inter-Quartile Range, or IQR), the whiskers show the data range up to ±1.5*IQR and the circles show data points outside that range.

3 Task 5: The Role and Importance of Synoptic or Mesoscale Meteorological Conditions in Creating High Ozone, $PM_{2.5}$, and SO_2

For task 5, we performed basic research into what synoptic-scale and urban-scale meteorological conditions are important in explaining and forecasting high concentrations of ozone, $PM_{2.5}$, and SO_2 in the seven urban areas: Austin, Beaumont/Port Arthur, Corpus Christi, Dallas/Fort Worth, El Paso, Houston/Galveston/Brazoria, and San Antonio. To determine if there was a relationship between synoptic or urban conditions and the likelihood of high total or background pollutant values, we first needed a quantitative definition of a "high" value of each metric. To find the days of interest with high concentrations, we calculated the 90th percentile in the 2012-2021 period for each location and pollutant, as seen in Table 21. Our analysis focuses on those days that were above the 90th percentile thresholds for each respective area and pollutant.

Table 21. 90th percentiles of observations during 2012-2021.

Area	MDA8_O_3_max (ppb)	MDA8_O_3_bkgrd (ppb)	$PM_{2.5}$_max (µg/m3)	$PM_{2.5}$_bkgrd (µg/m3)	SO_2_max (ppb)	SO_2_bkgrd (ppb)
Austin	62.24436	52.42711	15.66670	11.74331	1.85625	NA
Beaumont	62.13155	51.74977	15.33110	12.15476	32.65264	23.21766
Corpus Christi	57.88818	51.74479	17.04170	16.41670	5.53910	NA
Dallas/FW	73.73854	51.64120	17.20857	11.37919	24.82016	1.55356
El Paso	66.21195	56.45025	19.65370	9.62058	4.70706	1.40210
Houston	74.00184	47.92650	18.65464	10.58770	13.76440	2.01618
San Antonio	64.59844	50.99904	16.63140	13.78882	10.44644	2.46952

3.1 Synoptic Scale Conditions

For the analysis of the synoptic-scale conditions, we looked at dates where the pollutant fell above these 90th percentile thresholds across a majority of the 7 urban areas. Here we defined majority as at least 6 urban areas, except for the background SO_2 analysis which had data availability issues, for which we required at least 4 urban areas to exceed the threshold. We used the NOAA physical sciences laboratory tools (https://psl.noaa.gov/data/composites/day/) for the NCEP/NCAR Reanalysis (Kalnay et al 1996) to create composite mean and composite anomaly plots for dates that exceeded the 90th percentile thresholds in Table 21 across the majority of the urban areas. SO_2 did not have any dates where the majority of the urban areas exceeded the 90th percentile thresholds for the SO_2 total or SO_2 background. This makes sense due to the reactive nature of SO_2, which would cause SO_2 to have more of an urban signal, rather than a synoptic signal. The results for MDA8 O_3 and $PM_{2.5}$ are presented in Figures 121-124. Figure 121 shows the 500mb geopotential height composite mean plots from the NCEP/NCAR Reanalysis for dates where a majority of urban areas exceed local 90th percentile thresholds. Figure 122 shows the composite anomalies associated with Figure 121. Figure 123 shows the 500mb wind composite mean plots from the NCEP/NCAR Reanalysis for dates where a majority of urban areas exceed local 90th percentile thresholds. Figure 124 shows the composite anomalies associated with Figure 123.

The NOAA Physical Sciences Laboratory specifies that the composite mean is simply the average field for the list of days provided. The composite anomaly based on the 1981-2010

climatology is calculated for each day. The average of all days is then calculated. For vector wind, the anomaly of the u and v components are calculated separately for each day. They are each averaged and then the resulting u and v anomalies are plotted. The wind speed is the wind speed of the anomalies. If the u anomaly component is -4 and the v component is -3 then the resulting wind anomaly vector points southwest and has a magnitude of +5.

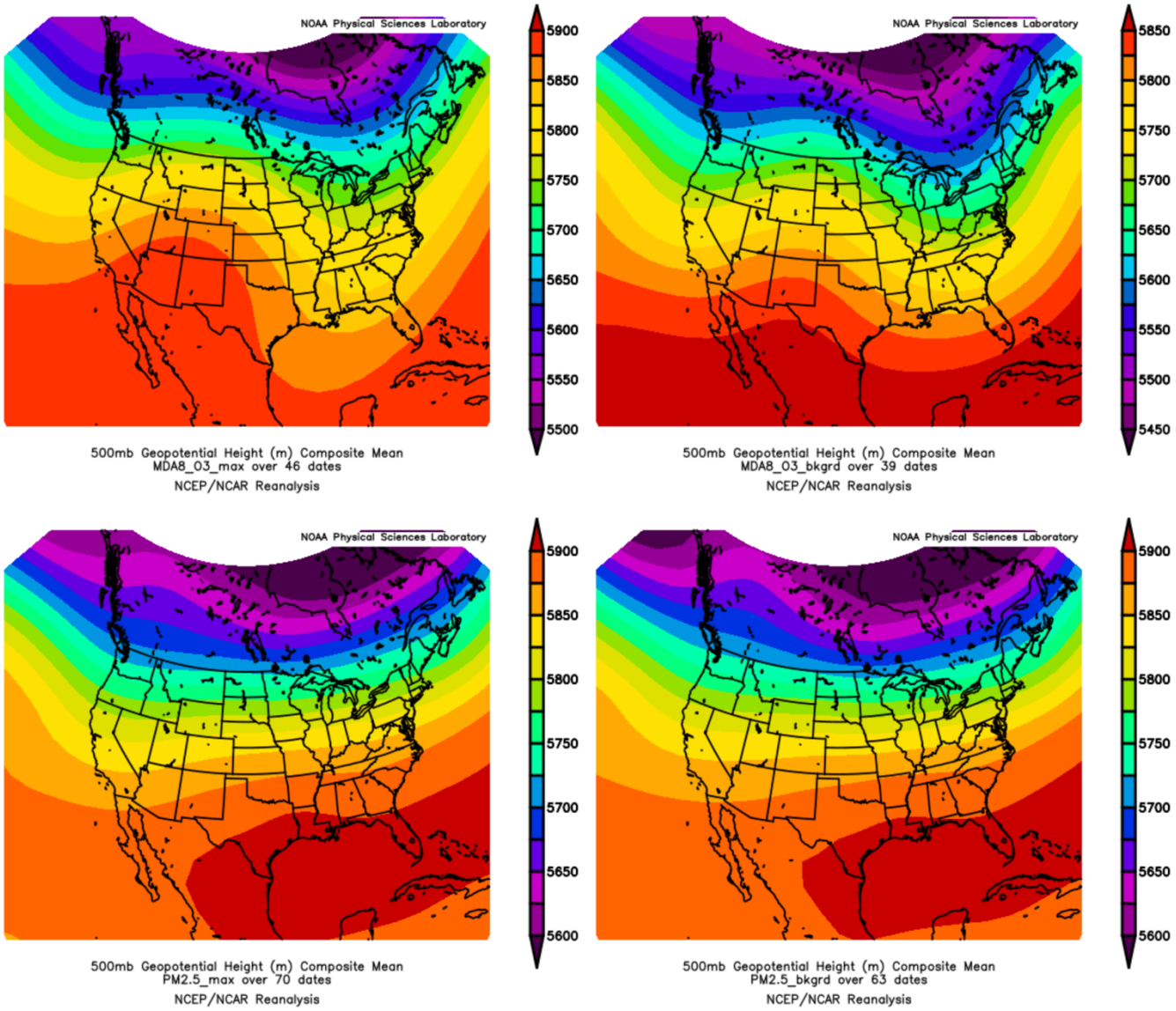

Figure 121. NCEP/NCAR Reanalysis composite mean maps of 500mb geopotential height (m) for dates where a majority of urban areas exceed local 90th percentile thresholds. (Upper left: MDA8 O3 max, upper right: MDA8 O3 background, lower left: $PM_{2.5}$ max, lower right: $PM_{2.5}$ background)

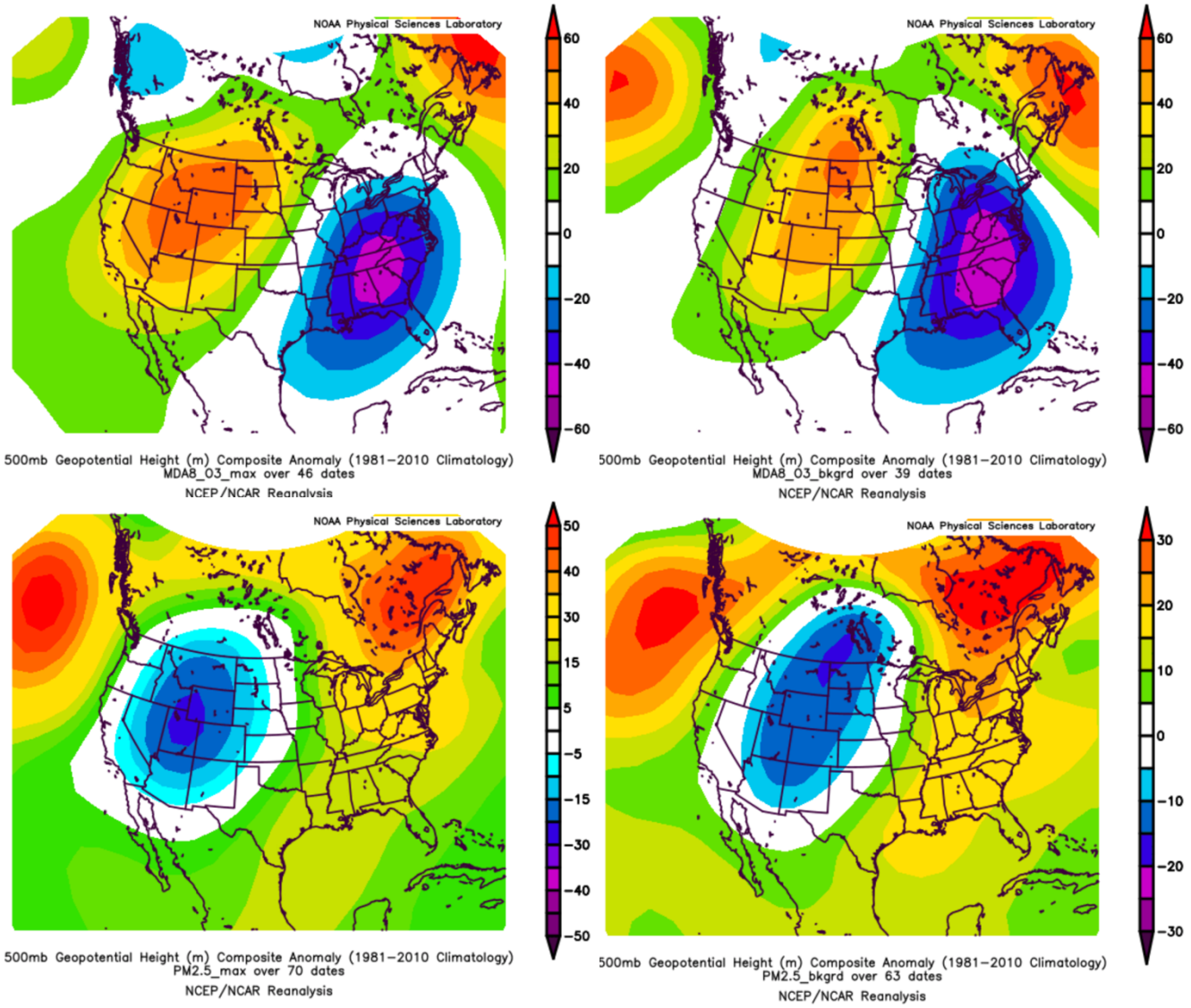

Figure 122. NCEP/NCAR Reanalysis composite anomaly maps of 500mb geopotential height (m) for dates where a majority of urban areas exceed local 90th percentile thresholds. (Upper left: MDA8 O3 max, upper right: MDA8 O3 background, lower left: $PM_{2.5}$ max, lower right: $PM_{2.5}$ background)

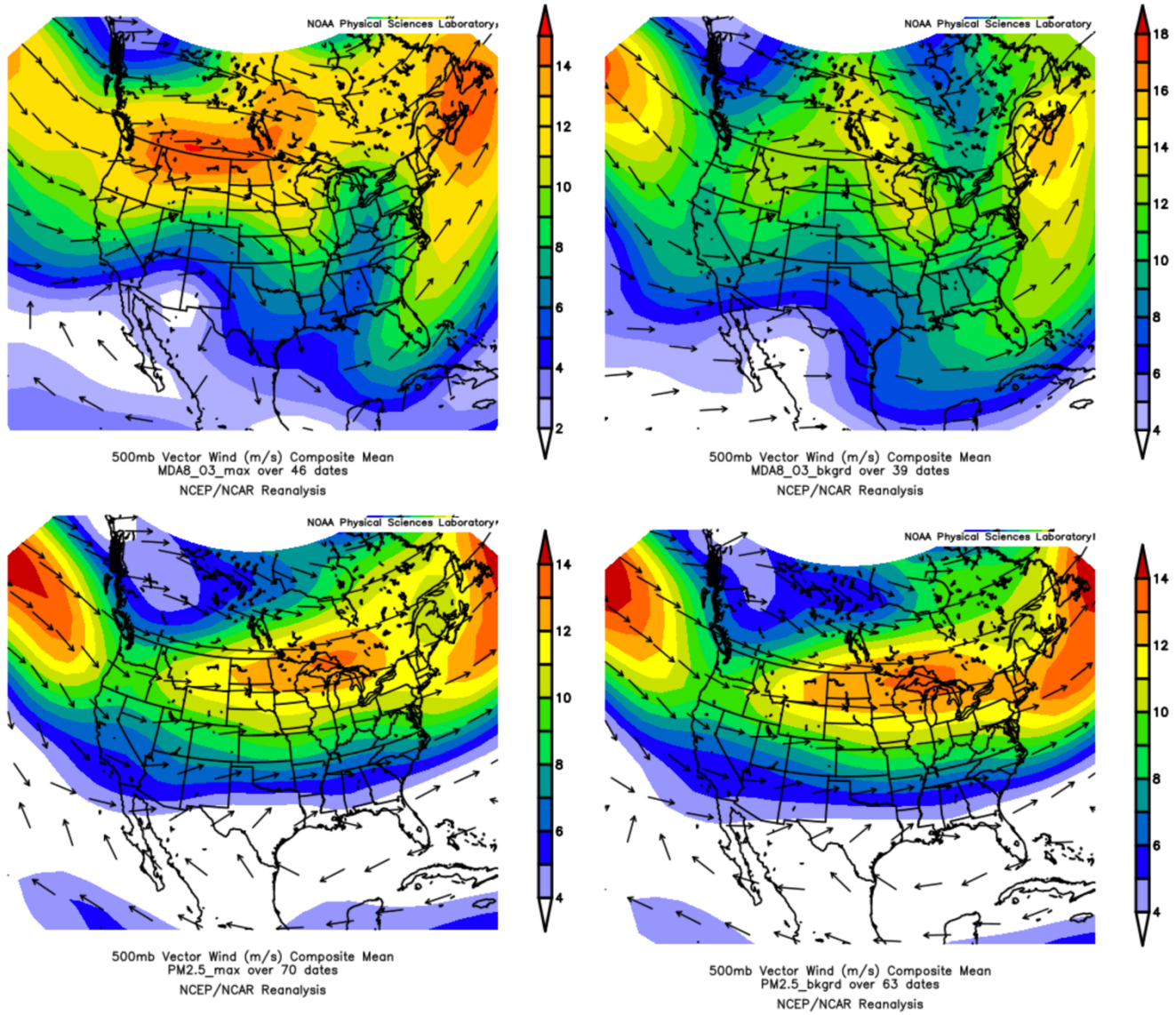

Figure 123. NCEP/NCAR Reanalysis composite mean maps of 500mb winds (m/s) for dates where a majority of urban areas exceed local 90th percentile thresholds. (Upper left: MDA8 O3 max, upper right: MDA8 O3 background, lower left: $PM_{2.5}$ max, lower right: $PM_{2.5}$ background)

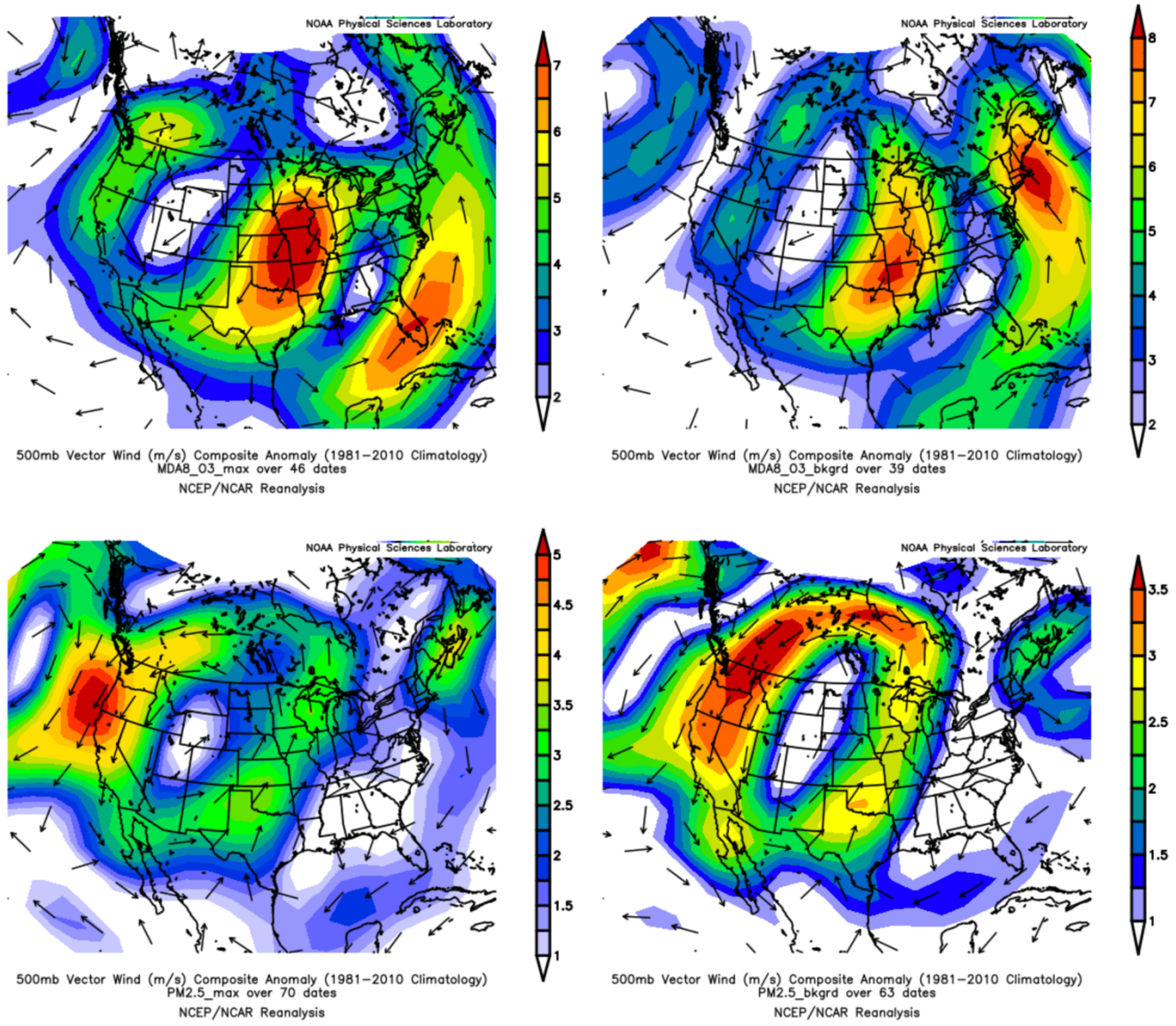

Figure 124. NCEP/NCAR Reanalysis composite anomaly maps of 500mb winds (m/s) for dates where a majority of urban areas exceed local 90th percentile thresholds. (Upper left: MDA8 O3 max, upper right: MDA8 O3 background, lower left: $PM_{2.5}$ max, lower right: $PM_{2.5}$ background)

3.2 Urban-scale conditions

For the analysis of the urban-scale conditions, we looked at dates where the pollutant fell above these 90th percentile thresholds in the majority of coastal urban areas and separately in the majority of the inland urban areas. The coastal urban areas include Beaumont/Port Arthur, Corpus Christi, and Houston/Galveston/Brazoria. The inland urban areas include Austin, Dallas/Fort Worth, El Paso, and San Antonio. A majority is defined as two or more for the coastal areas and three or more for the inland areas. For the analysis of SO_2 background data, Corpus Christi and Austin are

not available. So, the majority is defined as one or more coastal areas and two or more inland areas for this pollutant. Once the dates of interest were identified, they were screened for overlap with the dates of interest in the synoptic analysis. In order to focus on those dates that exceeded the threshold due to urban or meso-scale conditions rather than synoptic conditions, any overlapping synoptic dates were removed.

We used the NOAA physical sciences laboratory tools (https://psl.noaa.gov/cgi-bin/data/narr/plotday.pl/) for the NARR Reanalysis (Fedor Mesinger et. al., 2005) to create composite mean and composite anomaly plots for dates that exceeded the 90th percentile thresholds in Table 21 across the majority of the coastal/inland areas. We chose the NARR Reanalysis over the NCEP/NCAR Reanalysis used for the synoptic-scale due to its higher resolution (0.3deg x 0.3deg), which is more appropriate for finer scales. (The capabilities of the PSL tools developed for the NARR Reanalysis did not always display well for use with the synoptic scale.)

The results for MDA8 O_3, $PM_{2.5}$, and SO_2 are presented in Figures 125-130. Figures 125 and 126 show composite mean plots for mean sea level pressure for coastal and inland dates, respectively. Figures 127 and 128 show composite anomaly plots for 2m temperature for coastal and inland dates, respectively. Figures 129 and 130 show composite anomaly plots for 1000mb winds.

Note for this study we did not explore urban-scale meteorological indices (e.g., functions of the city average pressure, temperature, and winds, either from monitors or from the 12-km North American Mesoscale Forecast System (NAM-12) to explain and predict poor air quality events in the urban areas. The forecasting ability of the NAM-12km meteorological data is reflected in the GAM results. The NARR composite map analysis presented here for the coastal/inland dates of interest is also exploring surface pressure, temperature and winds at the meso-scale.

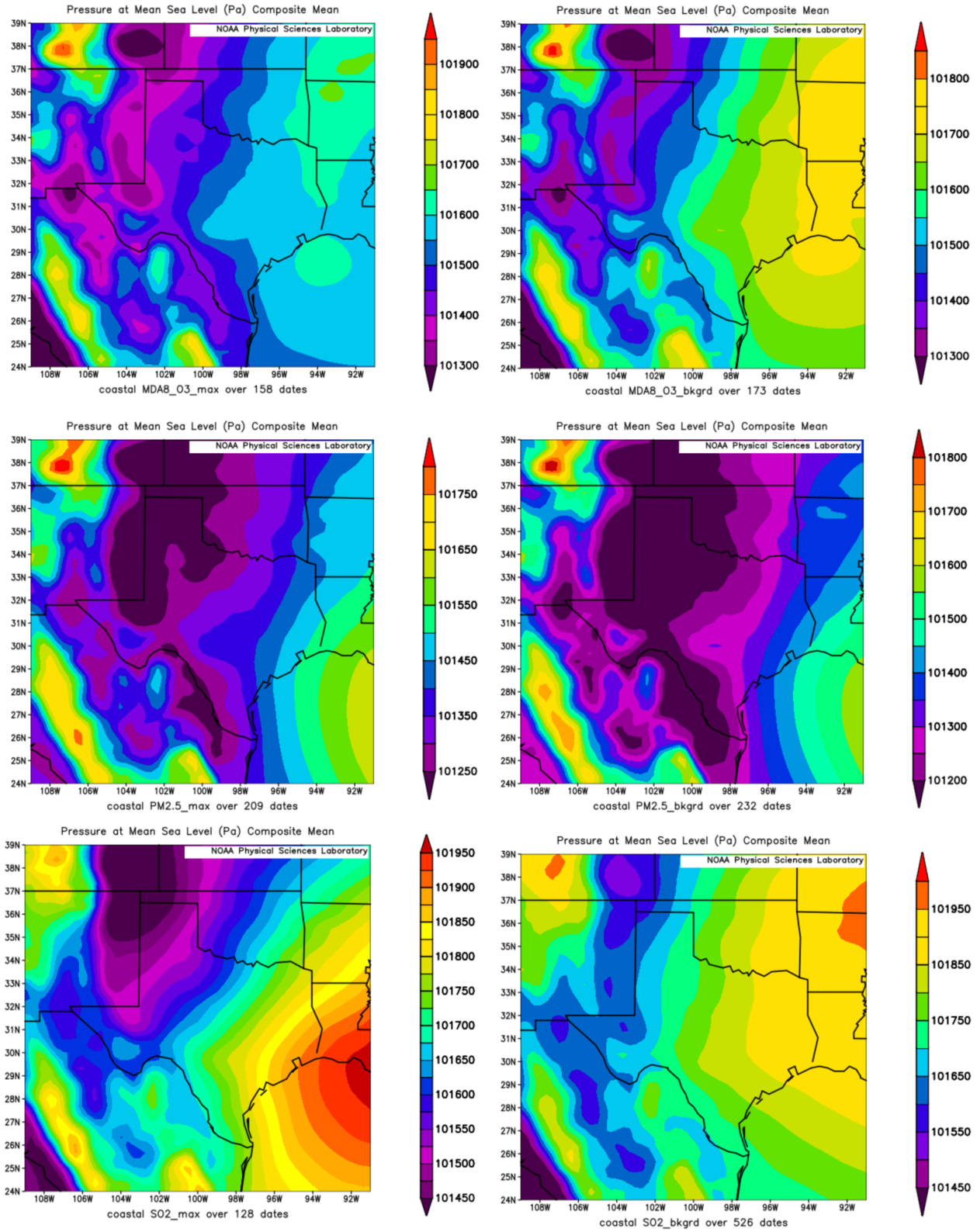

Figure 125. NARR Reanalysis composite mean maps of mean sea level pressure (Pa) for dates where a majority of coastal urban areas exceed local 90th percentile thresholds. (Upper left: MDA8 O_3 max, upper right: MDA8 O_3 background, middle left: $PM_{2.5}$ max, middle right: $PM_{2.5}$ background, lower left: SO_2 max, lower right: SO_2 background)

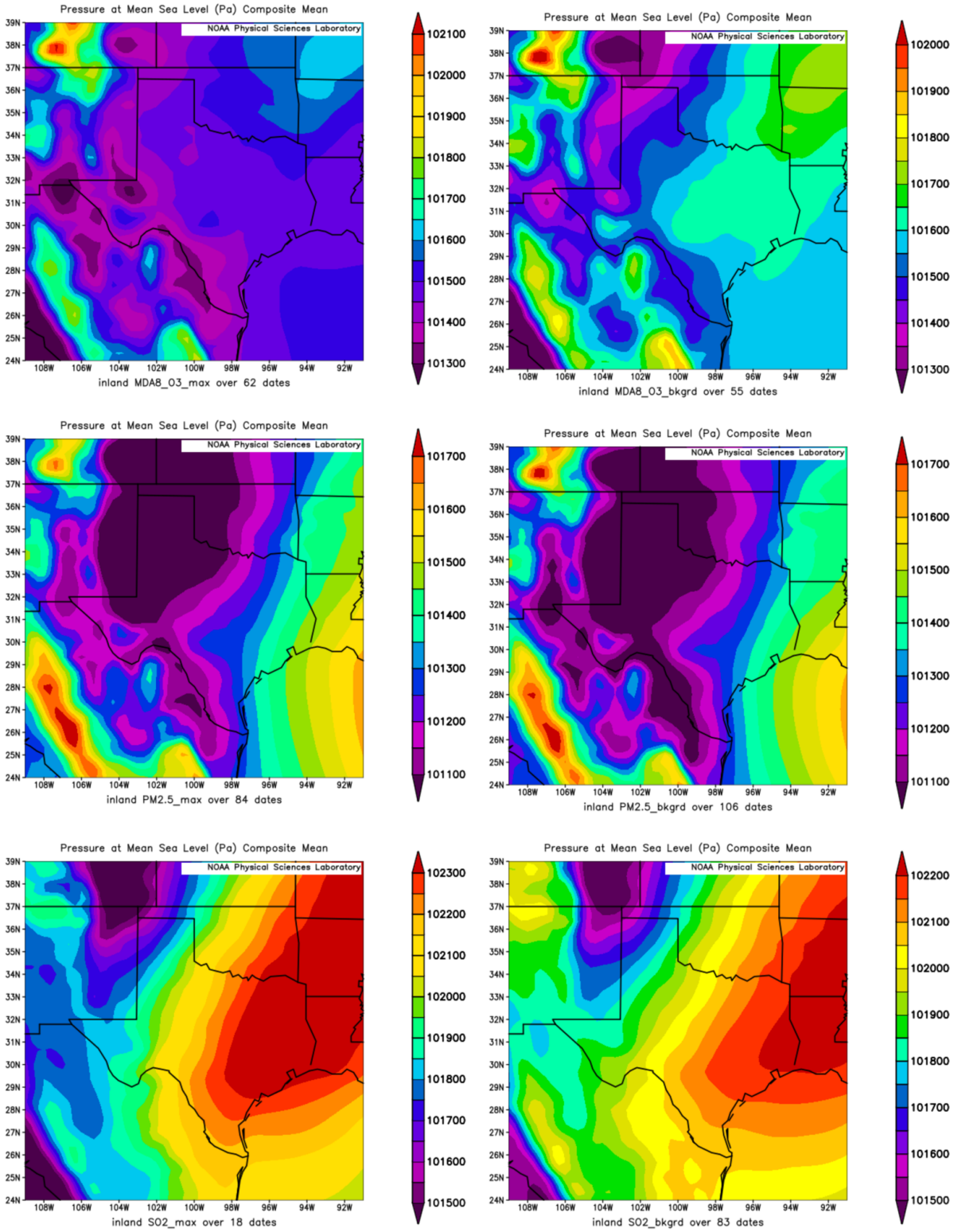

Figure 126. NARR Reanalysis composite mean maps of mean sea level pressure (Pa) for dates where most inland urban areas exceed local 90th percentile thresholds. (Upper left: MDA8 O_3 max, upper right: MDA8 O_3 background, middle left: $PM_{2.5}$ max, middle right: $PM_{2.5}$ background, lower left: SO_2 max, lower right: SO_2 background)

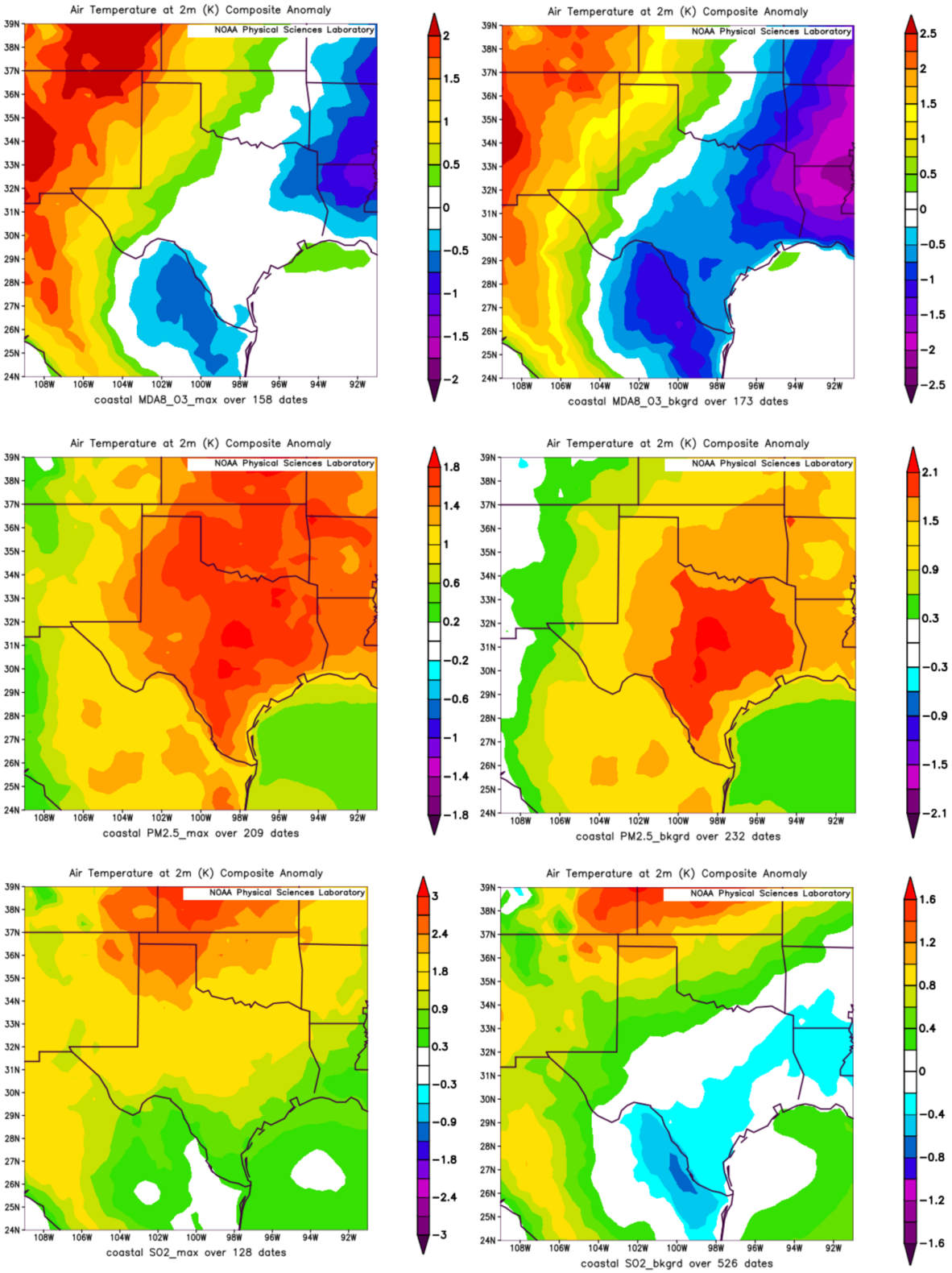

Figure 127. NARR Reanalysis composite anomaly maps of 2m temperature (K) for dates where most coastal urban areas exceed local 90th percentile thresholds. (Upper left: MDA8 O_3 max, upper right: MDA8 O_3 background, middle left: $PM_{2.5}$ max, middle right: $PM_{2.5}$ background, lower left: SO_2 max, lower right: SO_2 background)

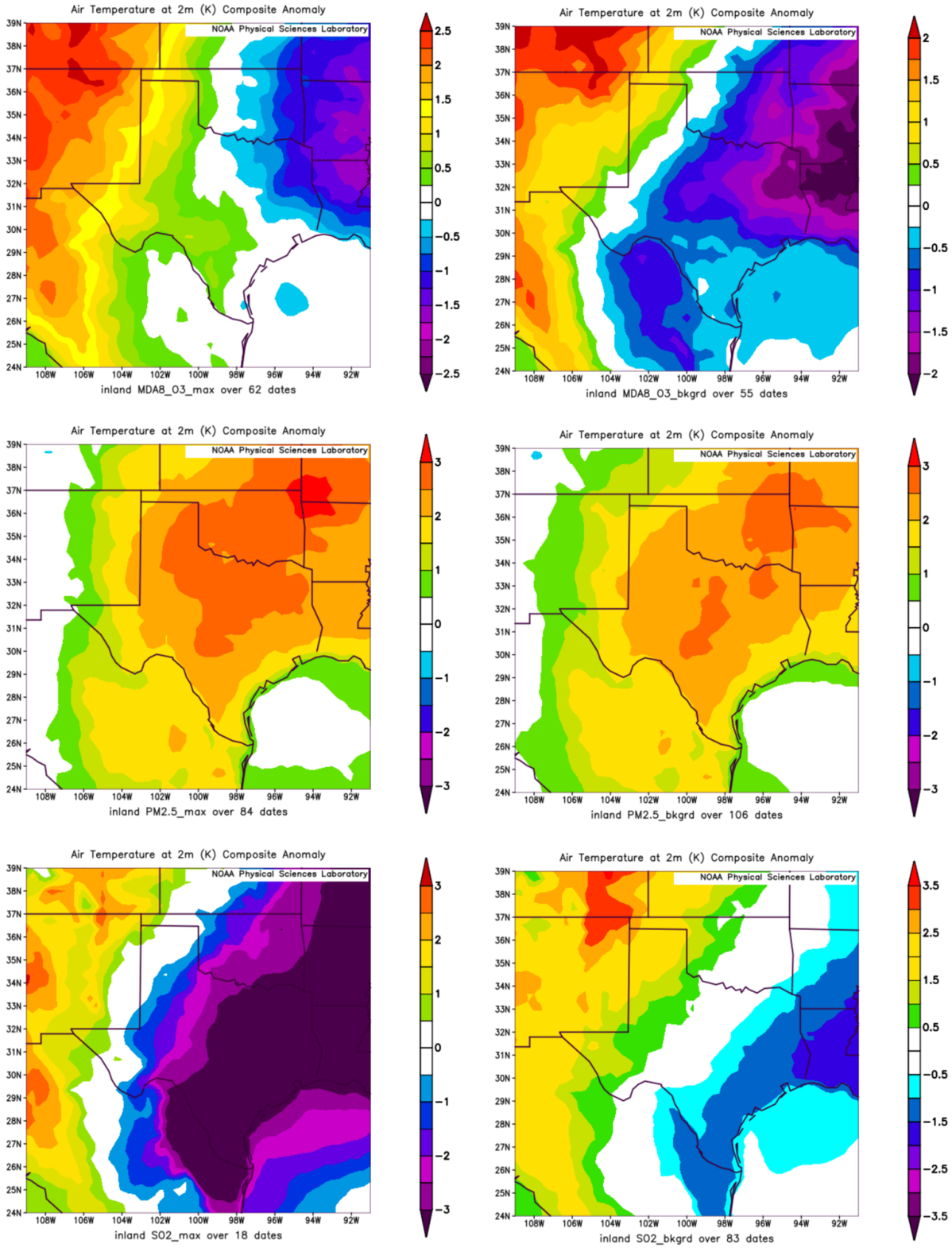

Figure 128. NARR Reanalysis composite anomaly maps of 2m temperature (K) for dates where most inland urban areas exceed local 90th percentile thresholds. (Upper left: MDA8 O_3 max, upper right: MDA8 O_3 background, middle left: $PM_{2.5}$ max, middle right: $PM_{2.5}$ background, lower left: SO_2 max, lower right: SO_2 background)

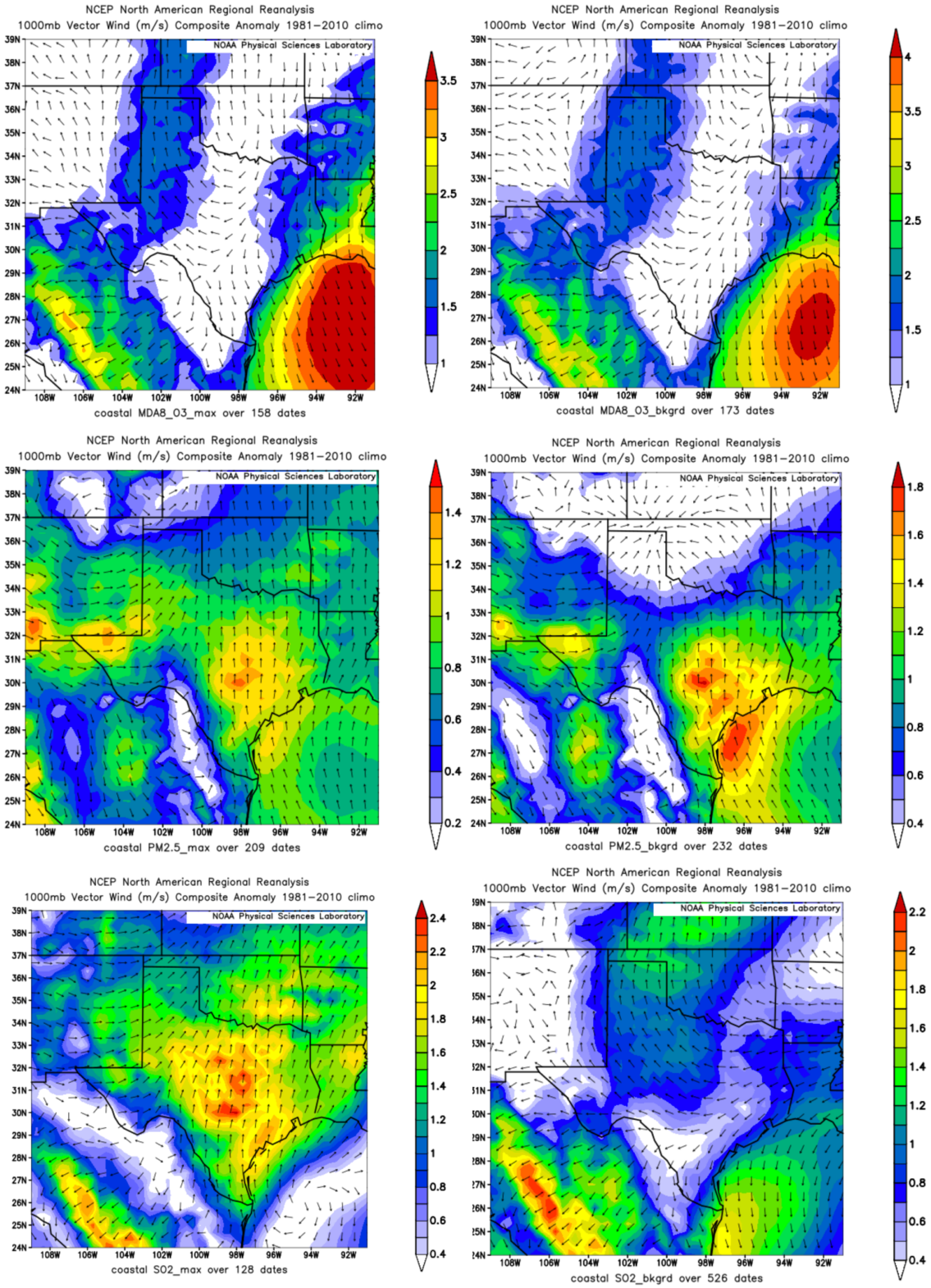

Figure 129. NARR Reanalysis composite anomaly maps of 1000mb winds (m/s) for dates where most coastal urban areas exceed local 90th percentile thresholds. (Upper left: MDA8 O_3 max, upper right: MDA8 O_3 background, middle left: $PM_{2.5}$ max, middle right: $PM_{2.5}$ background, lower left: SO_2 max, lower right: SO_2 background)

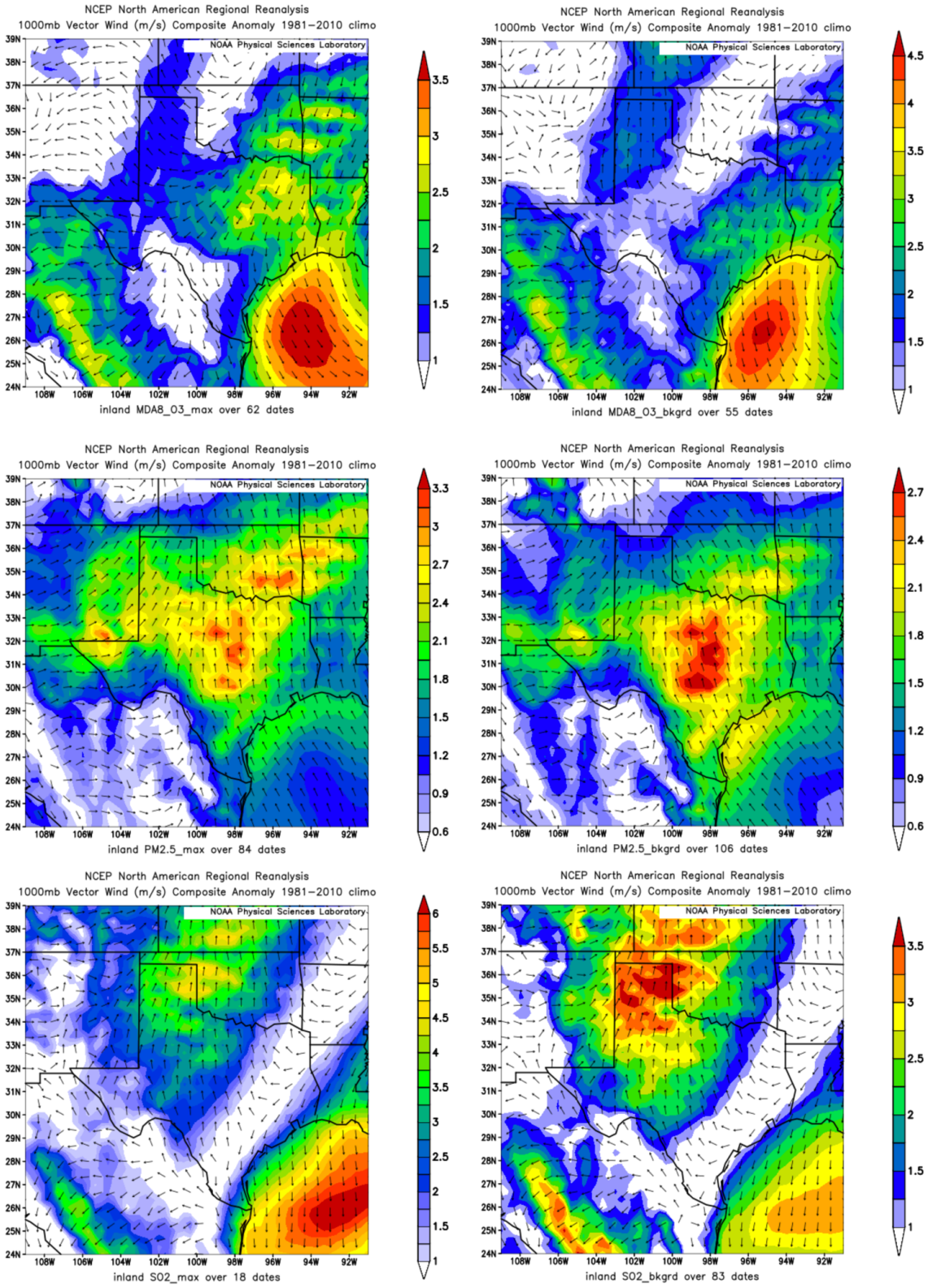

Figure 130. NARR Reanalysis composite anomaly maps of 1000mb winds (m/s) for dates where most coastal urban areas exceed local 90th percentile thresholds. (Upper left: MDA8 O_3 max, upper right: MDA8 O_3 background, middle left: $PM_{2.5}$ max, middle right: $PM_{2.5}$ background, lower left: SO_2 max, lower right: SO_2 background)

3.3 Discussion

The synoptic-scale composite maps are shown in Figures 121-124. In Figure 121, we see slight differences in the composite 500mb geopotential height. However, when we look at the scale of the associated 500mb geopotential height anomalies in Figure 122, we see that they appear at most to be 30-60m. So, while there are some differences in the synoptic patterns for the means and anomalies, the magnitude of these differences is not large. In Figure 123 are the composite mean 500mb wind patterns. All 4 panels show, the maximum 500mb winds to be well north of Texas, leaving lighter winds aloft over Texas. The pattern of composite anomalies in Figure 124 is interesting but given where the anomalies are when overlaid with the composite mean maps, it seems that they may be largely due to differences in wind vectors cancelling each other out.

The urban to meso-scale composite maps are shown in Figures 125-130. Figures 125 and 126 show composite mean plots for mean sea level pressure. While all the plots show a general pattern of slightly higher pressures to the south and east with lower pressures to the north and west, the magnitudes of these differences are very small, and on the whole, not a notable difference. Figures 127 and 128 show composite anomalies for 2m temperatures for coastal and inland areas, respectively. Both figures show similar patterns for each pollutant. High O_3 and SO_2 days at the coast and inland show cooler temperatures than climatology to the southeast and warmer to the northwest. High $PM_{2.5}$ days show higher temperatures than climatology over Texas itself. However, the fact that the patterns are similar between the coastal and inland figures, suggests that the coastal or inland location of an urban area is irrelevant. Figures 129 and 130 show the composite anomaly winds at 1000mb for coastal and inland urban areas, respectively. Overall, it's important to note that the magnitude of these anomalies is relatively small. High O_3 days at the coast and inland show slightly elevate winds over the Gulf. Days with high $PM_{2.5}$ at the coast and inland show slightly elevated winds in the interior of Texas. Three out of the four SO_2 figures for the coast and inland show elevated winds in Northern Texas and in the Gulf while the fourth figure show elevated winds over interior Texas. Again, the fact that the patterns are similar between the coastal and inland figures, suggests that the coastal or inland location of an urban area is irrelevant but rather there are more general conditions that make an urban area more susceptible to have a high pollutant day.

In general, high O_3 days tend to be associated with higher-than-normal winds from the north, with high pressure anomalies over the Rockies and low-pressure anomalies over the Southeast US. This may be due to an accumulation of O_3 precursors: Figure 124 suggests that synoptic transport from California, north across the Rockies, and then south into Texas is more common on high ozone days. In contrast, high $PM_{2.5}$ days tend to have higher than normal winds from the south, suggesting sea salt from the Gulf of Mexico or dust/smoke from Mexico and Central America may be contributing to high $PM_{2.5}$ days. Thus, the conditions and sources leading to high $PM_{2.5}$ might be very different from those leading to high O_3.

4 Quality Assurance Steps and Reconciliation with User Requirements

All work on the project was done in accordance with the Quality Assurance Project Plan (QAPP). All scripts and data files used in this project were inspected by team members different from the original author to ensure they were correct, and any errors noted in early versions were fixed. Other required evaluations are contained within the report. In addition, if further analysis

or feedback from the TCEQ uncovers any errors in the provided files, we will correct those and provide the TCEQ with corrected files.

In addition, the QAPP listed several questions that needed to be addressed for each project task. These questions are addressed below.

4.1 Task 3: Effects of Meteorology on Ozone, $PM_{2.5}$, and SO_2

- *Do the relationships between meteorological variables and O_3, $PM_{2.5}$, and SO_2 described in the developed GAMs make physical sense given our conceptual models of O_3, $PM_{2.5}$, and SO_2 emissions, chemistry, and transport?*

  The functional dependencies in the GAMs between the predictors related to temperature, water vapor density, wind speed, vertical stability, and HYSPLIT bearing are all qualitatively consistent with our conceptual understanding of *O_3, $PM_{2.5}$, and SO_2* emissions, chemistry, and transport.
- *Are these relationships consistent with the scientific literature?*

  Our GAMs for MDA8 O_3 here are consistent with those found for eastern US cities by Camalier et al. (2007) and for other Texas urban areas by Alvarado et al. (2015).
- *Are the HYSPLIT back-trajectories used in the model development reasonable? How sensitive are these trajectories to the initial location?*

  The HYSPLIT back-trajectories used in the model development appear reasonable and generally consistent with the measured surface wind speed and direction.
- *How well does the GAM reproduce the testing sets in the cross-validation evaluation?*

  The ten-fold cross-validation showed that the GAMs fit the data withheld from the training about as well as they fit the training data, giving little evidence of over-fitting (Section 2.2.5).
- *Does the cross-validation evaluation of the models show evidence of over-fitting?*

  As noted in Section 2.2.5, there is no evidence of over-fitting in the MDA8 O_3, $PM_{2.5}$, or SO_2 predictions.
- *Under what conditions are the GAMs expected to be valid? What conditions give exceptionally large residuals?*

  Strictly speaking, the GAMs are only expected to be valid during the periods for which they were fit, and when the data is taken from the sources and sites noted in this memo. Extrapolations to other times and monitoring locations may be problematic, and the GAMs ability in this regard has not been assessed in this project. We have not identified any set of necessary or sufficient conditions that lead to large residuals in the GAMs.

4.2 Task 4: Estimating Background Ozone, $PM_{2.5}$, and SO_2

- *Are the derived background estimates, and their spatial and temporal variation, consistent with our conceptual models of O_3, $PM_{2.5}$, and SO_2 emissions, chemistry, and transport?*

  The overall trends of background O_3, $PM_{2.5}$, and SO_2 are decreasing, consistent with our understanding of reduction of pollutant emissions (primarily NO_x and SO_2) over this time period. While US emissions of NOx are decreasing, which should

decrease regional background O_3, in some areas, the influence of Mexican emissions may be keeping the trend effectively zero. Background O_3 has a minimum in July and a maximum in April for urban areas near the Gulf of Mexico, consistent with the seasonal shifts in synoptic conditions.

- *Are these estimates consistent with the scientific literature?*

  The MDA O_3 values derived here are consistent with our previous work in other urban areas in Texas (Alvarado et al, 2015), while the $PM_{2.5}$ values in this study have somewhat larger uncertainties. Future research could examine the cause of the change.

- *What are the uncertainties in the background estimates, and under what conditions are they valid?*

  The major uncertainties in the background estimates calculated using the TCEQ method are, first, that they assume the regional background can be estimated as the lowest value observed at a selected number of sites around the urban area. This neglects the fact that urban areas in Texas and Mexico likely influence each other's "background", and so our background estimates cannot be interpreted as estimates of what the concentrations would be with all Texas or Mexican sources removed.

4.3 Task 5: The Role and Importance of Synoptic or Mesoscale Meteorological Conditions in Creating High Ozone, $PM_{2.5}$, and SO_2

- *Are identified synoptic and mesoscale meteorological controls on extreme and background concentrations of O_3, $PM_{2.5}$, and SO_2 consistent with our conceptual understanding of O_3, $PM_{2.5}$, and SO_2 emissions, chemistry, and transport?*

  In the synoptic analysis, while O_3 and $PM_{2.5}$ did have dates where the majority of the urban areas exceeded the 90^{th} percentile thresholds, the SO_2 total or SO_2 background did not. This aligns with our conceptual understanding due to the reactive nature of SO_2, which would cause SO_2 to have more of an urban signal, rather than a synoptic signal. The differences in synoptic and mesoscale patterns in the anomaly fields in section 3, are consistent with our understanding that high value days for different pollutants are driven by a combination of meteorological conditions and source placement, which vary by pollutant.

- *Are these estimates consistent with the scientific literature?*

  The GAM results are discussed above. The results of the logistic regressions of section 2.3 are reasonable.

5 Conclusions

Here we summarize the conclusions of our analysis, with reference to the corresponding report section.

- For each of the seven urban areas, we calculated estimates of total and background MDA8 O_3, total and background daily average $PM_{2.5}$, and total and background maximum daily 1-hour average of SO_2 for the period 2012-2021. The background estimates were calculated using the TCEQ method described in Berlin et al. (2013) (Section 2.1.1).

- We fit Generalized Additive Models (GAMs) relating meteorological variables to total and background MDA8 O_3, total and background daily average $PM_{2.5}$, and total and background maximum daily 1-hour average of SO_2 for each of the seven urban areas.
    - The seven GAMs for total MDA8 O_3 generally explain 53-72% of the deviance (i.e. variability), consistent with the results of Camalier et al. (2007) and Alvarado et al. (2015). The GAMs also generally show good fits with normally-distributed residuals and little dependence of the residual variance on the predicted value (Section 2.2.2.1).
    - The seven GAMs for background MDA8 O_3 generally explain 51-66% of the deviance (i.e. variability), consistent with the results of Camalier et al. (2007) and Alvarado et al. (2015). The GAMs also generally show good fits with normally-distributed residuals and little dependence of the residual variance on the predicted value (Section 2.2.2.2).
    - The seven GAMs for total $PM_{2.5}$ only explain 14-34% of the deviance (i.e. variability), and generally show a poor fit with long, positive residual tails and a strong dependence of the variance of the residuals on the predicted value (Section 2.2.3.1).
    - The seven GAMs for background $PM_{2.5}$ only explain 21-38% of the deviance (i.e. variability), and generally show a poor fit with long, positive residual tails and a strong dependence of the variance of the residuals on the predicted value (Section 2.2.3.2).
    - The seven GAMs for total SO_2 explain 14-56% of the deviance (i.e. variability), and generally show a poor fit with long, positive residual tails and a strong dependence of the variance of the residuals on the predicted value (Section 2.2.4.1).
    - The five GAMs for background SO_2 explain 31-62% of the deviance (i.e. variability), and generally show a poor fit with long, positive residual tails and a strong dependence of the variance of the residuals on the predicted value (Section 2.2.4.2). GAMs for background SO_2 in Austin and Corpus Christi were not run due to a lack of background data.
    - Heteroskedasticity was observed in some of the GAM evaluation plots for $PM_{2.5}$ and SO_2 models. For example, this can be seen in the upper right panel of Figure 44 where the variance of the residual increases with the linear predictor. This suggests the logarithm of the $PM_{2.5}$ and SO_2 concentrations were not normally distributed, which may mean that the significance of the predictors in the GAMs are overestimated. Future work could determine if transforming the predicted variable to one with a normal distribution or using weighted regressions would be a potential area for improvement in the models.
- After meteorological adjustment via the GAMs fit to total and background MDA8 O_3, total and background $PM_{2.5}$, and total and background SO_2, no trends in pollutant metrics between 2012-2021 were observed to be significant at a 95% confidence level (Section 2.3).

- Background MDA8 O_3 is fairly constant with month between May and August for Dallas/Fort Worth and El Paso, but has a July minimum in Austin, Beaumont/Port Arthur, Corpus Christi, Houston/Galveston/Brazoria, and San Antonio (Section 2.4.1).
- In contrast, background $PM_{2.5}$ peaks in June and July. The range of values for a given month or year is large for all cities, with Corpus Christi having the largest $PM_{2.5}$ spread and the most outliers (Section 2.4.2).
- Background SO_2 values in Beaumont/Port Arthur, Dallas/Forth Worth, Houston/Galveston/Brazoria, and San Antonio have a large range with many outliers, and displaying no discernible seasonal trend. However, El Paso values have a much smaller spread and show a minimum in summer months (section 2.4.3).
- We calculated the 90th percentiles for total and background MDA O_3, total and background $PM_{2.5}$, and total and background SO_2 for each of the seven urban areas. We found dates that exceeded this threshold in the majority of the urban areas, in the majority of coastal urban areas, or in the majority of inland urban areas. We created composite mean and composite anomaly maps for these categories of dates to analyze the synoptic-scale and mesoscale meteorological patterns associated with high pollutant days. The differences in synoptic and mesoscale patterns in the anomaly fields in section 3, are consistent with our understanding that high value days for different pollutants are driven by a combination of meteorological conditions and source placement, which vary by pollutant.

6 Recommendations

There are several questions raised by the results of our current study that would benefit from further investigation. In general, there weren't many SO_2 observation sites in the urban areas. In fact, Corpus Christi and Austin had no background data available for SO_2, which precluded any investigation of the SO_2 background in those areas. Future work could quantify the impact of the relative sparsity of SO_2 observations on the robustness of our conclusions.

In addition, the mean composite and composite anomaly mapping tools from the NOAA physical sciences laboratory provided a new way to analyze the meteorological impacts on high pollutant days. However, the NARR Reanalysis is at 0.3deg resolution (~33km). It would be valuable to take another look at the urban areas using an even finer scale model, such as the NAM-12km fields made available by NOAA since 2003. Implementing similar tools as those provided by NOAA's physical sciences laboratory would be required, since we are unaware of tools to use with NAM-12km fields.

The heteroskedastic patterns observed in the GAM residuals plots for $PM_{2.5}$ and SO_2 would benefit from further study. Determining how the log of the concentrations is distributed and performing weighted regressions are possible avenues for further study.

The GAMs developed in this study to relate meteorological predictors to the concentrations of total O_3, $PM_{2.5}$, and SO_2 as well as the logistic GAMs used to determine necessary and sufficient conditions for high O_3, $PM_{2.5}$, and SO_2 events, could be developed to provide accurate forecasts of air quality for each of the seven urban areas.

Finally, these GAMs derived from monitor network data should be compared with similar GAMs fit to meteorological and chemical data from 3D Eulerian air quality models like CAMx and CMAQ to determine if these models accurately represent the dependence of O_3, $PM_{2.5}$, and SO_2 concentrations, and the probability of high O_3, $PM_{2.5}$, and SO_2 events, on meteorology.

Differences discovered between the two sets of GAMs could point towards missing physics or incorrect parameterizations in the current Eulerian air quality models.

7 References

Alvarado, M. J., C. R. Lonsdale, M. E. Mountain, and J. D. Hegarty (2015), Investigating the Impact of Meteorology on O_3 and $PM_{2.5}$ Trends, Background Levels, and NAAQS Exceedances, Final Report to Texas Commission on Environmental Quality (TCEQ) for Work Order No. 582-15-54118-01 under TCEQ Contract No. No. 582-15-50415, August 31.

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.

Camalier, L., Cox, W., and Dolwick, P. (2007), The effects of meteorology on ozone in urban areas and their use in assessing ozone trends, *Atmos. Environ.*, 41, 7127-7137.

Draxler, R. R. and G. D. Hess (1997), Description of the HYSPLIT_4 modeling system. NOAA Tech. Memo. ERL ARL-224, 24 pp.

Draxler, R. R. and G. D. Hess (1998), An overview of the HYSPLIT_4 modeling system for trajectories, dispersion, and deposition, *Aust. Meteorol. Mag., 47,* 295-308.

Fedor Mesinger et. al, 2005: North American Regional Reanalysis: A long-term, consistent, high-resolution climate dataset for the North American domain, as a major improvement upon the earlier global reanalysis datasets in both resolution and accuracy, BAMS.

Fu, D., Worden, J. R., Liu, X., Kulawik, S. S., Bowman, K. W., and Natraj, V.: Characterization of ozone profiles derived from Aura TES and OMI radiances, *Atmos. Chem. Phys., 13*, 3445-3462, doi:10.5194/acp-13-3445-2013, 2013.

Hegarty, H. R. R. Draxler, A. F. Stein, J. Brioude, M. Mountain, J. Eluszkiewicz, T. Nehrkorn, F. Ngan, and A. Andrews (2013), Evaluation of Lagrangian Particle Dispersion Models with Measurements from Controlled Tracer Releases. *J. Appl. Meteor. Climatol.*, *52*, 2623–2637, doi: http://dx.doi.org/10.1175/JAMC-D-13-0125.1

Hegarty, J., H. Mao, *and* R. Talbot *(*2007*),* Synoptic controls on summertime surface ozone in the northeastern United States*,* J. Geophys. Res.*,* 112*, D14306, doi:*10.1029/2006JD008170.

Kalnay, E. and Coauthors, 1996: The NCEP/NCAR Reanalysis 40-year Project. Bull. Amer. Meteor. Soc., 77, 437-471.

Langford, A. O., J. Brioude, O. R. Cooper, C. J. Senff, R. J. Alvarez II, R. M. Hardesty, B. J. Johnson, and S. J. Oltmans (2012), Stratospheric influence on surface ozone in the Los Angeles area during late spring and early summer of 2010, J. Geophys. Res., 117, D00V06, doi:10.1029/2011JD016766.

Pernak, R., and M. Alvarado (2017), *Data Driven Ozone Forecasting Phase II*, Final Report to Texas Commission on Environmental Quality (TCEQ) for Work Order No. 582-17-71680-05 under TCEQ Contract No. No. 582-15-50414, June 29.

R Core Team (2015), R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org.

Starkweather, J. (2011), Cross Validation techniques in R: A brief overview of some methods, packages, and functions for assessing prediction models, available at https://www.unt.edu/rss/class/Jon/Benchmarks/CrossValidation1_JDS_May2011.pdf

van Donkelaar, A., R. V. Martin, R. J. D. Spurr, E. Drury, L. A. Remer, R. C. Levy, and J. Wang (2013), Optimal estimation for global ground-level fine particulate matter concentrations, J. Geophys. Res. Atmos., 118, 5621–5636, doi:10.1002/jgrd.50479.

Wood, S. N. (2006), *Generalized Additive Models: An Introduction with R*, part of the "Texts in Statistical Science" series, Chapman & Hall/CRC, New York.

Appendix A File Descriptions and Process Flow

A.1 Process Flow

This section describes all of the files included in the deliverable. Figure 131 is a flow chart showing the processing from the initial data sources to the final CSV file used as input for the GAMs. Figure 132 shows the scripts that use the CSV file produced at the end of Figure 131 to produce and evaluate the GAMs.

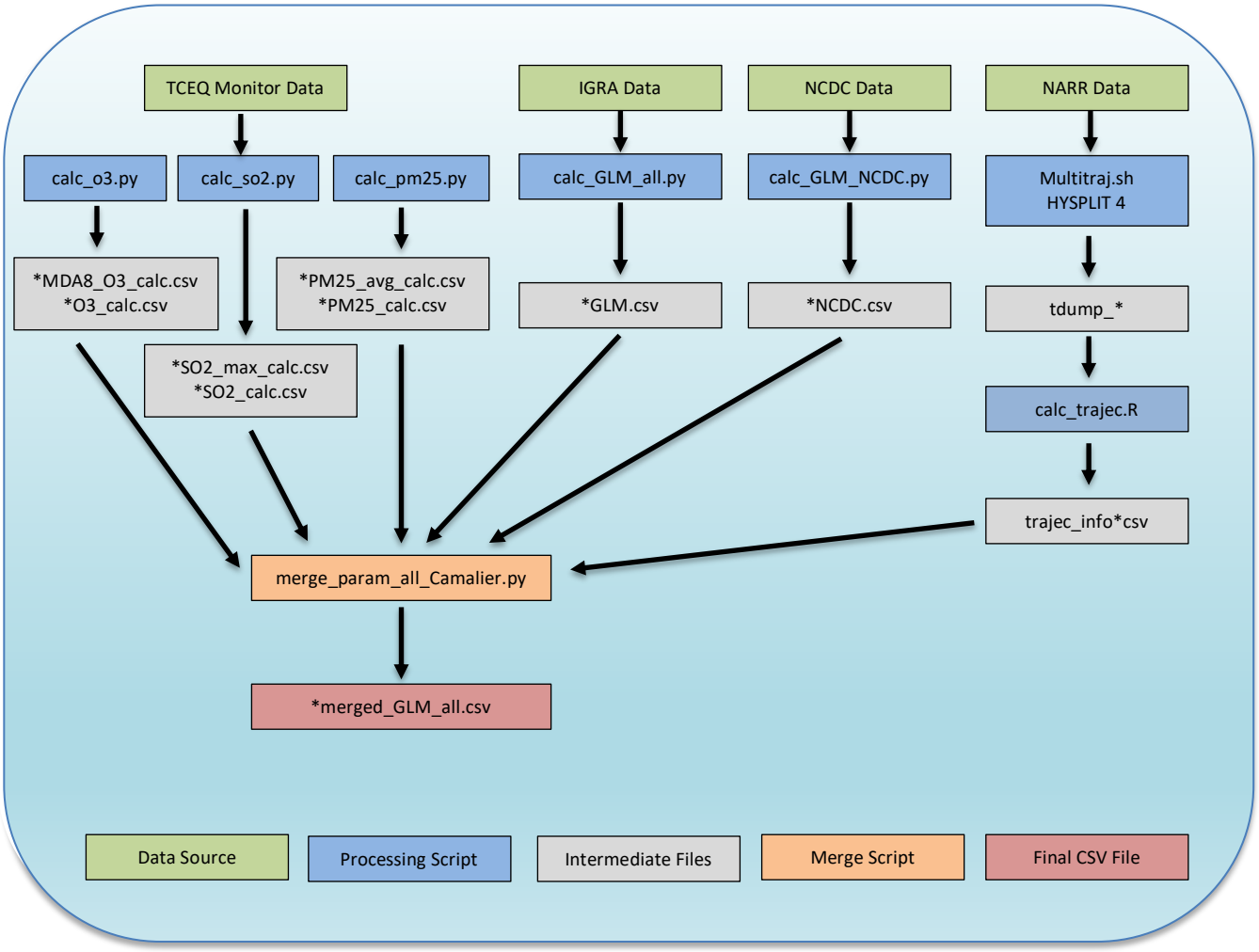

Figure 131. Flow chart showing the processing from the original data sources (green boxes) to the final CSV file (red box) that is used as input for the GAM scripts.

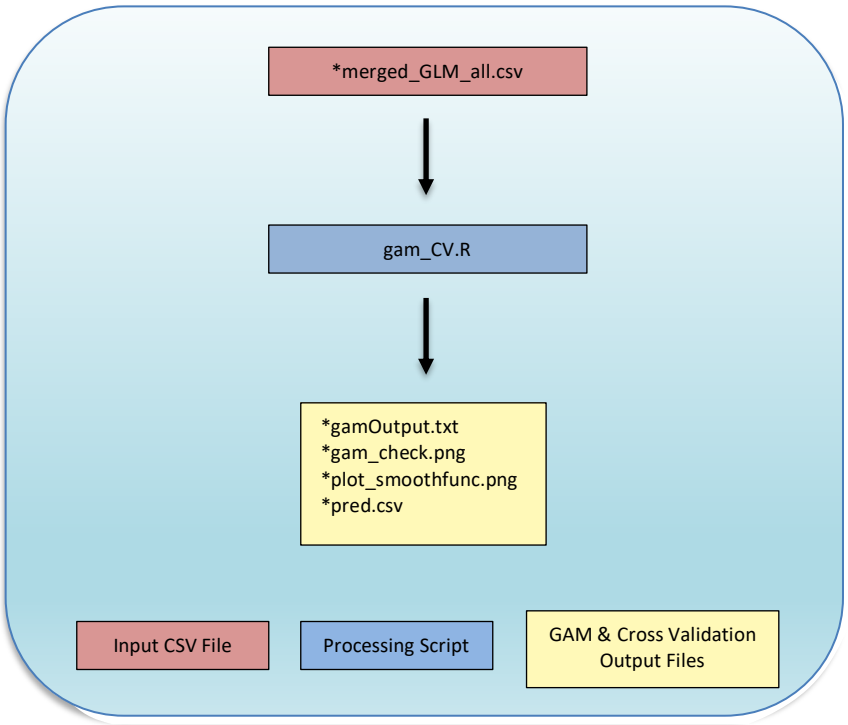

Figure 132. Flow chart showing the processing from the input CSV file generated at the end of Figure 131 (red box) to the GAM output files (yellow box).

Note that all R scripts below were run using R version 4.1.3 (2022-03-10) and package mgcv v1.8-39 on a 64bit MacOS platform running MacOS Monterey with 2.3 GHz Quad-Core Intel® Core™ i7 Processor and 32 GB RAM. All python scripts were run using Python v2.7.14. The HYSPLIT runs were performed using a K shell (ksh) script on a Linux server running CentOS Linux release 7.9.2009 (Core) (x86-64) with 4 AMD Opteron 6168 CPUs (12-core @ 1.9 GHz) and 2.6 GB RAM per core. The Microsoft Excel spreadsheets were made using Microsoft Excel for Mac v16.62. All scripts should run on any Linux or Windows OS system with the correct versions of R, Python, and Microsoft Excel installed.

A.2 File Descriptions

A.2.1 Input data *(./data/)*

This directory contains the raw IGRA2, NCDC, and monitor data provided by TCEQ including the station meteorological data for the seven urban areas in this project.

A.2.1.1 IGRA Data (./data/IGRA2)

The Integrated Global Radiosonde Archive (IGRA Version 2) provided upper atmosphere data used to derive some of the meteorological predictors. Table 3 describes the data files used: USM*.txt along with a readme.txt file that describes the data format and measurements. The relevant measurements include the geopotential height, temperature and dew point depression at several altitudes with - 99999 values as missing.

A.2.1.2 NCDC data (./data/NCDC/)

This directory contains the National Climatic Data Center (NCDC) Integrated Surface Data (ISD) used to get estimates of surface pressure and relative humidity, as this data was not generally available in the TCEQ dataset. The raw data (*###.csv*) for the seven urban areas is sorted into separate subdirectories. The raw data file contains hourly data from 2012-2021. The data format for each field is described in detail in the isd-format-document.pdf. Station details are provided in Table 4.

A.2.1.3 Monitoring data (./data/TCEQ/)

This directory includes the monitoring data from the seven urban areas as provided by the TCEQ for the period of 2012-2021. All monitoring stations are listed in Table 2.

- *"*o3_12_21.txt"* – Hourly O_3 monitor data.
- *"*pm25_all_1hr_12_21.txt"* – Hourly $PM_{2.5}$ monitor data.
- *"*so2_12_21.txt"* – Hourly SO_2 monitor data.

A.2.1.4 Meteorological data (./data/TCEQ/)

This directory also includes the meteorological data from the seven urban area's monitoring stations, as stated in Table 2.

- *"*met_12_21.txt"* – Hourly meteorological monitor data.

A.2.2 Data Processing Scripts (./scripts/)

- *./scripts/calc_o3.py* : This script reads in the ozone monitor data provided by the TCEQ, as described above, to calculate the MDA8 O_3 for a single urban area at time. Requires the file *lookup_table_array.txt*. After filtering out non-data the script derives the maximum and minimum MDA8 for all sites in each urban area, as well as the minimum (background) MDA8 O_3 value for all selected background sites according to the technique described in Section 2.1. The selected background sites are listed in the script. The produced CSV files are inputs to the script *./scripts/merge_param_all_Camalier.py* (described below), which will combine O_3 daily values with the GAM parameters.
- *./scripts/calc_pm25.py* : This script reads in the $PM_{2.5}$ monitor data provided by the TCEQ and finds the daily average for $PM_{2.5}$ for the sites in each urban area. The produced CSV files are inputs to the script *./scripts/merge_param_all_Camalier.py* (described below), which will combine $PM_{2.5}$ daily values with the GAM parameters.
- *./scripts/calc_so2.py* : This script reads in the SO_2 monitor data provided by the TCEQ and finds the maximum daily 1-hour average for SO_2 for the sites in each urban area. The produced CSV files are inputs to the script *./scripts/merge_param_all_Camalier.py* (described below), which will combine SO_2 daily values with the GAM parameters.
- *./scripts/calc_GLM_all.py* : This script reads in TCEQ monitor site and IGRA2 (upper atmosphere) measurements to derive daily GAM parameters described in the script itself. It performs all the necessary conversions (e.g. Fahrenheit to Celsius, mph to m/s) and derivations (e.g., wind direction u component, dewpoint to RH based on August-Roche-Magnus approximation), to compile the full list of daily meteorological predictors, except those from the NCDC (described below). See the script for full details on all conversions and derivations. These output files are used in *./scripts/merge_param_all_Camalier.py*.

- *./scripts/calc_GLM_NCDC.py* : This script reads in the NCDC data to derive daily meteorological predictors indicated as an NCDC parameter. It performs all the necessary conversions (ex. Fahrenheit to Celsius) and derivations (ex. Apparent Temperature according to the National Digital Forecast Database). See the script for full details on all conversions and derivations. These output files are used in *./scripts/merge_param_all_Camalier.py*.
- *./scripts/merge_param_all_Camalier.py* : This script reads in all intermediate files described above. This includes the daily and background values for O_3, $PM_{2.5}$, and SO_2; daily values for all meteorological predictors; the HYSPLIT bearing and distance (*../HYSLPIT/trajec_info.csv*). It aligns the date for all files, checks for missing data and replaces with 'nan' if there is no data. It creates the final merged files used in the GAM scripts.

A.2.3. HYSPLIT Files (./HYSPLIT/)

- *./tdump_nam_* :* An intermediate text file generated from the *./multitraj.sh* script for running HYSPLIT (previously delivered to TCEQ). * is a 3-letter code indicating the urban area. The first line in each file lists the 3-letter city code and the latitude and longitude of the trajectory origin. The starting back trajectory elevation is always 300 m above ground level (agl) and not included in these files. The rest of the lines are the endpoint time and location data, one line per endpoint. The lines include the following:
    - Trajectory run - will always be 1 in this application, ignore
    - Trajectory number – will always be 1 in this applications, ignore
    - YEAR – 2-digit format
    - Month
    - Day
    - Hour – always 18 UTC
    - Minute – always 0
    - Second –always 0
    - Trajectory age – always -24 (indicating a 24 hour back trajectory)
    - Latitude
    - Longitude- west is negative
    - Elevation- meters AGL
    - Pressure – hPa
- *./calc_trajec.R :* This R script takes the 24 hour back-trajectory endpoint files from the ./HYSPLIT/ directory and calculates the distance and bearing from the starting point to the end point of the trajectory using the R functions bearing and distMeeus from the geosphere package. The function bearing gets the initial bearing (direction; azimuth) to go from point 1 to point 2 following the shortest path (a Great Circle). The function distMeeus calculates the shortest distance between two points (i.e., the 'great-circle-distance' or 'as the crow flies') using the WGS84 ellipsoid.
- *./trajec_info_*.csv :* CSV file produced by *./calc_trajec.R* that contains the distance and bearing for the back trajectories. These files are used as inputs by *./scripts/merge_param_all_Camalier.py*

A.2.4 Processed Input Data Files in CSV Format (./csv_files/)

A.2.4.1 Intermediate CSV Files (./csv_files/Intermediate/)

These files include the meteorological data derived from the NCDC, TCEQ, and IGRA2 datasets described in Section 2.1.

A.2.4.2 Final CSV Files (./csv_files/final/)

The **merged_GLM_all.csv* files are created by *./scripts/merge_param_all_Camalier.py,* which combines all daily meteorological predictors with the daily values for O_3, $PM_{2.5}$, and SO_2. The file includes daily values from 2012-2021, with missing values indicated by 'nan'. These files are used as inputs by the GAM scripts described below.

A.2.5 GAM (./full_gam_fits/)

This directory contains the files for the GAMs discussed in Section 2.2.

- *./gam_CV.R:* This script reads in the final csv files and fits a GAM to each file and a given pollutant. It produces the following in the subdirectory /*gam_out*:
    - **gamOutput.txt*: A log of final model diagnostics: summary and cross validation.
    - **gam_check.png*: gam.check plot
    - **plot_smoothfunc.png*: smooth variable function plots
    - * *pred.csv*: A file containing the daily inputs to the GAM and the output value predicted by the GAM for a given pollutant.

  This script also does a cross validation analysis after each GAM fit using the CVgam function of the "mgcv" R package.
- *./met_adj_trends.xlsx:* This script excel file calculates and plots the meteorologically adjusted trends.

A.2.6 Synoptic and Mesoscale Meteorological Analysis (./MetAnalysis/)

This directory contains the files for the analysis of synoptic and mesoscale meteorological conditions on high pollutant days in section 3.

- *./MetAnalysis.src:* This script takes the final CSV file for each urban area and calculates the 90th percentile concentration value for a given pollutant in the 2012-2021 period. It then groups the urban areas into categories: all, inland, coastal. It finds days where the 90th percentile threshold was exceeded in all or a majority of the urban areas in that category. It outputs the identified dates for given criteria into separate text files for each category.
- *dates_out/*txt*: Output from *MetAnalysis.src*. Files containing a list of dates that exceed the 90th percentile concentration values for given criteria. The first line is the number of dates in the file. Following lines contain 1 date per line. The format is compatible with that expected by the NOAA physical sciences laboratory tools described in section 3. These files can be uploaded to the NOAA ftp site and then selected in the options where the composite and anomaly maps are generated. For instructions see https://psl.noaa.gov/data/composites/day/.